NONPARAMETRIC ANALYSIS OF BIVARIATE CENSORED DATA

By

EDWARD ANTHONY POPOVICH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1983

To my parents, this is as
much yours as it is mine.

ACKNOWLEDGEMENTS

This dissertation completes the work toward my Doctor of Philosophy degree. However, I would be remiss in failing to mention the contributions of many others without whom this work would not be possible. I would like to thank my committee for their guidance and support. Dr. Dennis Wackerly was especially helpful due to his expert proofreading. Even though it goes without saying that the major advisor has the most important direct impact on the actual work, I will further say that Dr. P.V. Rao went beyond the call of duty. The endless hours given by Dr. Rao will not be forgotten. Moreover, he is more than just my advisor, he is a respected friend and shall always remain so. I would like to thank the Department of Statistics for allowing me to be a part of this excellent group. To my family, especially my parents, I will always be grateful for their support during my years at the University of Florida. I would like to thank Debra Lynn Smith for her understanding and for being there when I needed her the most. To all my friends, especially my roommates during the last two years, John L. Jackson, Jr. and George Tanase, Jr., I would like to express my gratitude for just being my friends. I would also like to thank my typist, Donna Ornowski, for the many hours, including several evenings, she gave in order to do such an excellent job. Without those mentioned above, this work could not have been completed, therefore, in a real way this degree is as much theirs as it is mine.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

NONPARAMETRIC ANALYSIS
OF
BIVARIATE CENSORED DATA

By

Edward Anthony Popovich

December, 1983

Chairman: Pejaver V. Rao
Major Department: Statistics

A class of statistics is proposed for the problem of testing for location difference using censored matched pair data. The class consists of linear combinations of two conditionally independent statistics where the conditioning is on the number, N_1, of pairs in which both members are uncensored and the number, N_2, of pairs in which exactly one member is uncensored. Since every member of the class is conditionally distribution-free under the null hypothesis, H_o: no location difference, the statistics in the proposed class can be utilized to provide an exact conditional test of H_o for all N_1 and N_2.

If n denotes the total number of pairs, then under suitable conditions the proposed test statistics are shown to have asymptotic normal distributions as n tends to infinity. As a result, large sample tests can be performed using any member of the proposed class.

A method that can be used to choose one test statistic from the proposed class of test statistics is outlined. However, the resulting test statistic depends on the underlying distributional forms of the populations from which the bivariate data and censoring variables are sampled.

Simulation results indicate that the powers of certain members in the class are as good as and, in some cases, better than the power of a test for H_0 proposed by Woolson and Lachenbruch in their paper titled "Rank Tests for Censored Matched Paris" appearing on pages 597-606 of Biometrika in 1980. Also, unlike the test of Woolson and Lachenbruch, the critical values for small samples can be tabulated for the tests in the new class. Consequently, members of the new class of tests are recommended for testing the null hypothesis.

CHAPTER ONE

INTRODUCTION

1.1 What is Censoring?

In many experiments it is important to compare two or more treatments by observing the times to "failure" of objects subjected to the different treatments. For example, one might be interested in comparing two drugs to see if there is a difference in the length of time the drugs (treatments) are effective. The time period in which a treatment is effective is called the survival time or the lifetime of that treatment. Unfortunately, in many experiments the survival times for all treatments under study are not observable due to a variety of reasons. If for a particular subject the treatment is still effective at the time of the termination of the experiment, then the corresponding survival time is known to be longer than the observation time of that subject. In that case the survival time is unknown and said to be right censored at the length of time the subject was under observation. It is also possible, that at the time of first observation of a treated subject, the treatment is already ineffective so that the corresponding survival time is not observable, but is known to be less than the observation time of that subject. In that case the survival time which is unknown is said to be left censored at the observation time. It should be noted that censored data can arise in cases where responses are not measured on a time-scale. Miller (1981) gives an example from

Leavitt and Olshen (1974) where the measured response is the amount of money paid on an insurance claim on a certain date. For this example, an uncensored response would be the amount paid on the total claim while a right-censored response would be the amount paid to date if the total claim is unknown.

Censoring can arise in different situations determined by the experimental restrictions on the observation time. For example, Type I censoring occurs when the period of observation for each subject is preset to be a certain length of time T; the survival times longer than T would be right-censored while those survival times less than T would be observed and are known as failure times. Type II censoring occurs when the period of observation will terminate as soon as the r^{th} failure is observed where r is a predetermined integer greater than or equal to one. If n is the total number of subjects in the experiment, then the n-r survival times which are unobserved are right-censored at the observation time T. Random censorship is a third form of censoring and in many cases it is a generalization of Type I censoring. Random censoring occurs when each subject has its own period of observation which need not be the same for all subjects. This could occur, for instance, when the duration of the experiment is fixed, but the subjects do not necessarily enter at the beginning of the experiment. In that case each subject could have its own observation time. When the time of entry is random, the observation time will also be random. Clearly, if a subject's survival time is greater than the subject's observation time, then a right-censored survival time results; otherwise, a failure time is observed for that subject. In the random censoring case, left-censored survival times could also occur if at the time of first observation of the subject a failure has already occurred.

1.2 Censored Matched Pairs

This dissertation is concerned with the analysis of the survival times of two treatments on matched pairs of subjects. The problem of testing for the difference in locations of two treatments using censored matched pair data was examined by Woolson and Lachenbruch (1980) under the assumption that both members of a pair have equal but random observation times and that the observation times for different pairs are independent. They utilized the concept of the generalized rank vector introduced by Kalbfleisch and Prentice (1973) and also used by Prentice (1978) to derive a family of rank tests for the matched pairs problem. The tests are developed by imitating the derivation of the locally most powerful (LMP) rank tests for the uncensored case (Lehmann, 1959). In spite of the fact that the Woolson-Lachenbruch family of rank tests (W-L tests) is an intuitively reasonable generalization of the LMP rank tests for the uncensored case, the complicated nature of these tests makes it difficult to investigate their theoretical properties. For example, it is not clear that these tests are LMP in the censored case. Also, the asymptotic normality of these tests, as claimed by Woolson and Lachenbruch, is strictly valid only when N_1 (the number of pairs in which both members are uncensored) and N_2 (the number of pairs in which exactly one member is censored) are regarded as nonrandom and tending to infinity simultaneously. It is not known whether the asymptotic normality holds (unconditionally) as the total sample size tends to infinity.

For small sample sizes exact critical values for the W-L test statistics could be based on their (conditional on N_1 and N_2) permutation distributions. However, the determination of these critical values becomes progressively impractical when either N_1 or N_2 is greater than 8 and the other is greater than 2. An example of the Woolson and Lachenbruch test ststistic is presented in Chapter Four.

The objective of this dissertation is to propose a class of statistics for testing the null hypothesis, H_o: no difference in treatment effects. Any member of this class is an alternative to the W-L statistic for the censored matched pair problem. The statistics in the proposed class are not only computationally simple, but under H_o are also distribution-free (conditional on N_1 and N_2). Furthermore, the critical values of these statistics are easily tabulated.

Chapter Two begins with the general setup of the problem and some preliminary results which are needed in Chapter Three. The class of linear combinations of two conditionally distribution-free and independent statistics for testing H_o is proposed in Chapter Three and it is shown that the asymptotic normality of the corresponding test statistics hold unconditionally as the total sample size n tends to infinity. Simulation studies are used in Chapter 4 to compare the test proposed by Woolson and Lachenbruch (1980) to the tests proposed in this dissertation.

CHAPTER TWO

TWO CONDITIONAL TESTS FOR CENSORED MATCHED PAIRS

2.1 Introduction and Summary

In this chapter two statistics for testing the null hypothesis H_o, that there is no difference in treatment effects, using censored matched pair data are developed. Either of these statistics can be utilized to test for a shift in location, and together they provide conditionally independent tests of H_o conditional on the numbers of observed pairs in which both members of a pair are uncensored (N_1) and in which exactly one member of a pair is censored (N_2).

The two statistics are proposed in Section 2.4 and it is shown that the exact conditional distributions as well as the conditional asymptotic distributions of the statistics are easily specified. Some notations and assumptions are stated in Section 2.2, while Section 2.3 contains preliminary results needed in the development of Section 2.4.

2.2 Notation and Assumptions

This section begins with a set of notation which will be used hereafter throughout. Let $\{(X_i^o, Y_i^o, C_i);\ i=1,2,\ldots,n\}$ denote independent random variables distributed as (X^o, Y^o, C). In addition, (X^o, Y^o) and C are referred to as the true value of the treatment responses and the censoring variable, respectively. The actual observed value of the treatment responses is (X,Y) where $X \equiv \min(X^o, C)$

and $Y \equiv \min(Y^o, C)$. In addition to (X,Y) one also observes the value of the random variable $(\delta(X), \delta(Y))$ where

$$\delta(X) = \begin{cases} 1 & X = X^o \\ 0 & X = C \end{cases}$$

and

$$\delta(Y) = \begin{cases} 1 & Y = Y^o \\ 0 & Y = C \end{cases}.$$

Note here that the indicator random variable $\delta(X)$ (respectively, $\delta(Y)$) indicates whether or not X (respectively, Y) represents a true uncensored response.

Let

$$N_1 = \sum_{i=1}^{n} \delta(X_i)\, \delta(Y_i) ,$$

$$N_{2L} = \sum_{i=1}^{n} \delta(X_i)[1 - \delta(Y_i)] ,$$

$$N_{2R} = \sum_{i=1}^{n} [1 - \delta(X_i)]\, \delta(Y_i) ,$$

$$N_2 = N_{2L} + N_{2R} ,$$

and

$$N_3 = \sum_{i=1}^{n} [1 - \delta(X_i)][1 - \delta(Y_i)]$$

$$= n - N_1 - N_2 .$$

Clearly, N_1, N_2, and N_3 denote the number of pairs (X_i, Y_i) in which both members are uncensored, exactly one member is censored, and both members are censored, respectively. Similar interpretations for N_{2L} and N_{2R} may be given.

The test procedures under consideration here are based on the observed differences $\{D_i = X_i - Y_i \ ; \ i=1,\ldots,n\}$. It is said an observed difference D_i falls into category C1 if $\delta(X_i)\delta(Y_i)=1$, into category C2 if $\delta(X_i)[1-\delta(Y_i)] + [1-\delta(X_i)]\delta(Y_i)=1$, and into category C3 if $[1-\delta(X_i)][1-\delta(Y_i)] = 1$. It is evident that N_k is the number of pairs (X_i, Y_i) in category Ck, k=1,2,3. Also, as noted by Woolson and Lachenbruch (1980) a D_i in category C1 equals $X_i^o - Y_i^o \equiv D_i^o$; furthermore, a D_i in category C2 either satisfies $D_i > D_i^o$ (i.e., D_i^o is left censored at D_i) if $\delta(X_i)[1-\delta(Y_i)] = 1$ or satisfies $D_i < D_i^o$ (i.e., D_i^o is right censored at D_i) if $[1-\delta(X_i)]\delta(Y_i) = 1$. The order relationship between D_i and D_i^o for those D_i in category C3 is unknown.

Additional notation is needed for the N_1 and N_2 pairs (X_i, Y_i) in categories C1 and C2, respectively. Let (X_{1i}, Y_{1i}), $i = 1, \ldots, N_1$, denote the N_1 pairs in category C1 and let (X_{2j}, Y_{2j}), $j = 1, \ldots, N_2$ denote the N_2 pairs in category C2. Moreover, let D_{1i} and D_{2j} be defined as follows:

$$D_{1i} \equiv X_{1i} - Y_{1i} \quad , i = 1, \ldots, N_1,$$

and (2.2.1)

$$D_{2j} \equiv X_{2j} - Y_{2j} \quad , j = 1, \ldots, N_2.$$

Now a set of assumptions, which are used later, are stated.

Assumptions

(A1): (X_i^o, Y_i^o) are independent and identically distributed (i.i.d.) as a continuous bivariate random variable (X^o, Y^o).

(A2): There exists a real number Θ such that $(X^o, Y^o+\Theta)$ is an exchangeable pair.

(A3): $C_1, C_2, \ldots, C_n$ are i.i.d. as a continuous random variable C, and is independent of (X^o, Y^o).

(A4): C_i is independent of (X_i^o, Y_i^o).

(A5): Let $P_\Theta\{\cdot\}$ indicate probability calculated with the value of Θ satisfying (A2). Then

$P_\Theta\{D_i \in C3\} \equiv P_3(\Theta) \equiv P_\Theta\{X^o > C, Y^o > C\} < 1.$

Note that in view of (A2), X^o and $Y^o + \Theta$ have identical marginal distributions and that (A5) implies $P_\Theta\{N_3 = n\} < 1$. Therefore, under (A5), with positive probability, there will be at least one D_i in either category C1 or C2 or both. If the probability of observing a D_i in category C1 is positive, then $F_\Theta(x)$ will be used to denote the distribution function of D_{1i}. In other words,

$$
\begin{aligned}
F_\Theta(x) &\equiv P_\Theta\{D_{1i} \leq X\} \\
&= P_\Theta\{D_1 \leq x \mid D_1 \in C1\}.
\end{aligned} \tag{2.2.2}
$$

Finally, because a test of H_o: $\Theta = \Theta_o$ utilizing D_i, i=1,2,...,n is a test of H_o:$\Theta = 0$ utilizing $D_i' = D_i - \Theta_o$ we may, without loss of generality, take $\Theta_o = 0$.

2.3 Preliminary Results

It is obvious that if $N_2=N_3=0$ with probability one, then the problem under consideration is the usual paired sample problem. Since under (A2), the D_i are symmetrically distributed about Θ, there exist a number of linear signed rank tests appropriate for testing H_o: $\Theta = 0$ (Randles and Wolfe, 1979) of which the well-known Wilcoxon signed rank test is an important example. However, if $N_2 > 0$ with positive probability, then the situation is different because a D_i^o in C2 will be either left or right censored at D_i and not all the D_i^o will equal D_i. Lemmas 2.3.1 and 2.3.2 below state some properties of the D_i belonging to C1 and C2.

Lemma 2.3.1:

Let $P_\Theta\{D_i \varepsilon C1\} \equiv P_1(\Theta) \equiv P_\Theta\{X^o < C, Y^o < C\}$.

Suppose $\Theta = 0$ and assumptions (A1), (A2) and (A4) hold. Suppose further that $P_1(0) > 0$, and set $\psi(X) = 1$ or 0 depending on whether $X > 0$ or $X \leqq 0$, respectively.

(a) The distribution of D_i conditional on $D_i \varepsilon C1$ is symmetric about 0, That is, for every real number a

$$P_0\{D_i \leq -a \mid D_i \varepsilon C1\}$$
$$= P_0\{D_i \geq a \mid D_i \varepsilon C1\} .$$

(b) The random variables $|D_i|$ and $\psi(D_i)$ are conditionally independent. That is, for every real number $a \geqq 0$ and $u = 0, 1$

$$P_0\{|D_i| \leq a, \psi(D_i) = u \mid D_i \varepsilon C1\}$$

$$= P_0\{|D_i| \leq a \mid D_i \varepsilon C1\} P_0\{\psi(D_i) = u \mid D_i \varepsilon C1\} .$$

Proof:

To prove (a) consider

$$P_0\{D_i \leq -a \mid D_i \varepsilon C1\} P_0\{D_i \varepsilon C1\}$$

$$= P_0\{X_i^0 - Y_i^0 \leq -a,\ X_i^0 \leq C_i,\ Y_i^0 \leq C_i\}.$$

By (A2), X_i^o and Y_i^o are exchangeable, therefore, the quantity above equals

$$P_0\{Y_i^0 - X_i^0 \leq -a,\ X_i^0 \leq C_i,\ Y_i^0 \leq C_i\}$$

$$= P_0\{X_i^0 - Y_i^0 \geq a,\ X_i^0 \leq C_i,\ Y_i^0 \leq C_i\}$$

$$= P_0\{D_i \geq a \mid D_i \varepsilon C1\} P_0\{D_i \varepsilon C1\},$$

which completes the proof of (a).

To prove (b) consider for $a > 0$,

$$P_0\{|D_i| \leq a, \psi(D_i) = 1 \mid D_i \varepsilon C1\}$$

$$= P_0\{0 < D_i \leq a \mid D_i \varepsilon C1\}.$$

From part (a) the conditional distribution of D_i is symmetric about 0. Therefore, the quantity on the right hand side of the equation equals

$$\tfrac{1}{2}[P_0\{0 < D_i \leq a \mid D_i \varepsilon C1\} + P_0\{-a \leq D_i < 0 \mid D_i \varepsilon C1\}]$$

$$= \tfrac{1}{2}[(P_0\{D_i \leq a \mid D_i \varepsilon C1\} - P_0\{D_i \leq 0 \mid D_i \varepsilon C1\})$$

$$+ (P_0\{D_i < 0 \mid D_i \varepsilon C1\} - P_0\{D_i < -a \mid D_i \varepsilon C1\})]$$

$$= \tfrac{1}{2}[P_0\{D_i \le a \mid D_i \varepsilon\ C1\} - P_0\{D_i = 0 \mid D_i \varepsilon\ C1\}$$

$$- P_0\{D_i < -a \mid D_i \varepsilon\ C1\}]$$

$$= \tfrac{1}{2}P_0\{ -a \le D_i \le a \mid D_i \varepsilon\ C1\} - \tfrac{1}{2}P_0\{D_i = 0 \mid D_i \varepsilon\ C1\}.$$

Since under (A1) (X^o, Y^o) has continuous bivariate distribution, then for all θ $P_\theta\{X^o - Y^o = 0\} = 0$, and since $D_i \varepsilon$ C1 implies $D_i = D_i{}^o$, then the quantity above equals

$$\tfrac{1}{2}P_0\{ -a \le D_i \le a \mid D_i \varepsilon\ C1\}$$
$$= \tfrac{1}{2}P_0\{ |D_i| \le a \mid D_i \varepsilon\ C1\} . \tag{2.3.1}$$

Now, from part (a)

$$P_0\{D_i > 0 \mid D_i \varepsilon\ C1\} = P_0\{\psi(D_i) = 1 \mid D_i \varepsilon\ C1\} = \tfrac{1}{2}$$

and

$$P_0\{D_i \le 0 \mid D_i \varepsilon\ C1\} = P_0\{\psi(D_i\ ) = 0 \mid D_i \varepsilon\ C1\} = \tfrac{1}{2} .$$

Consequently, from (2.3.1) it can be seen that

$$P_0\{|D_i| \le a,\ \psi(D_i\ ) = 1 \mid D_i \varepsilon\ C1\}$$

$$= \tfrac{1}{2}\ P_0\{|D_i| \le a \mid D_i \varepsilon\ C1\}$$

$$= P_0\{|D_i| \le a \mid D_i \varepsilon\ C1\}\ P_0\{\psi(D_i) = 1 \mid D_i \varepsilon\ C1\} .$$

Similarly,

$$P_0\{|D_i| \le a,\ \psi(D_i) = 0 \mid D_i \varepsilon\ C1\}$$

$$= P_0\{ |D_i| \le a \mid D_i \varepsilon\ C1\} - P_0\{|D_i| \le a,\ \psi(D_i) = 1 \mid D_i \varepsilon\ C1\}$$

$$= P_0\{ |D_i| \le a \mid D_i \varepsilon C1\}$$

$$- P_0\{|D_i| \le a \mid D_i \varepsilon C1\} P_0\{\psi(D_i) = 1 \mid D_i \varepsilon C1\}$$

$$= P_0\{|D_i| \le a \mid D_i \varepsilon C1\} [1 - P_0\{\psi(D_i) = 1 \mid D_i \varepsilon C1\}]$$

$$= P_0\{|D_i| \le a \mid D_i \varepsilon C1\} P_0\{\psi(D_i) = 0 \mid D_i \varepsilon C1\} .$$

This completes the proof of (b). □

Lemma 2.3.2: Let $D_{11}, \ldots, D_{1N_1}$ be as defined in equation (2.2.1). Then conditional on $N_i = n_i$, $i=1,2,3$, $D_{11}, \ldots, D_{1N_1}$, and N_{2R} are mutually independent. That is, for real numbers $d_1,\ldots,d_{n_1}$, and integer n_{2R}

$$P_\theta\{D_{1j} \le d_j,\ j=1,\ldots,N_1, \text{ and } N_{2R} = n_{2R} \mid N_i = n_i,\ i=1,2,3\}$$

$$= \prod_{j=1}^{n_1} P_\theta \{D_{1j} \le d_j \mid N_i = n_i,\ i=1,2,3\} P_\theta \{N_{2R} = n_{2R} \mid N_i=n_i,\ i=1,2,3\}.$$

Proof:

Let

$$P_1(d,\theta) = P_\theta\{D = X^o - Y^o \le d, \max (X^o, Y^o) \le C\},$$

$$P_{2R}(\theta) = P_\theta \{Y^o \le C < X^o \},\quad P_{2L}(\theta) = P_\theta \{X^o \le C < Y^o\},$$

and

$$P_2(\theta) = P_{2R}(\theta) + P_{2L}(\theta) = P_\theta\{D_i \varepsilon C2\} .$$

Recall that from assumption (A5)

$$P_3(\theta) = P_\theta \{\min(X^o,Y^o) > C\} = P_\theta \{D_i \varepsilon C3\}.$$

Also, as a result of the definition of $P_1(\theta)$ in Lemma 2.3.1

$$P_1(\theta) = P_1(\infty, \theta) = P_\theta\{D_i \varepsilon C1\}.$$

For arbitrary integers n_1, $n_2 = n_{2R} + n_{2L}$ and n_3 such that $n = n_1 + n_2 + n_3$ and for an arbitrary permutation

$$(i_1, \ldots, i_{n_1}, i_{n_1+1}, \ldots, i_{n_1+n_{2R}}, i_{n_1+n_{2R}+1}, \ldots,$$
$$i_{n_1+n_2}, i_{n_1+n_2+1}, \ldots, i_n)$$

of

$$(1,2, \ldots , n)$$

under assumptions (A1), (A3), and (A4) the following identity results.

$$\begin{aligned} P_\theta\{D_{1i_j} \leq d_j\ , j=1, \ldots, n_1\ ;\ &[1 - \delta(X_{i_j})]\delta(Y_{i_j}) = 1, \\ &j = n_1 + 1, \ldots, n_1 + n_{2R}\ ;\ \delta(X_{i_j})[1 - \delta(Y_{i_j})] = 1, \\ &j = n_1 + n_{2R} + 1, \ldots, n_1 + n_2; \\ &[1 - \delta(X_i)][1 - \delta(Y_i)] = 1,\ j = n_1 + n_2 + 1, \ldots, n\} \end{aligned}$$

$$= \left\{ \prod_{j=1}^{n_1} P_1(d_j, \theta) \right\} [P_{2R}(\theta)]^{n_{2R}} [P_{2L}(\theta)]^{n_{2L}} [P_3(\theta)]^{n_3}\ . \qquad (2.3.2)$$

Therefore, summing (2.3.2) over the

$$\binom{n}{n_1, n_{2R}, n_{2L}, n_3}$$

distinct ways of dividing n subscripts into four groups of n_1, n_{2R}, n_{2L}, and n_3 subscripts it follows that

$$P_\Theta \{D_{1j} \leq d_j,\ j=1\ldots,n_1, \text{ and } N_{2R} = n_{2R},\ N_{2L} = n_{2L},\ N_3 = n_3\}$$

$$= \binom{n}{n_1,n_{2R},n_{2L},n_3} \left\{ \prod_{j=1}^{n_1} P_1(d_j,\Theta) \right\} [P_{2R}(\Theta)]^{n_{2R}} [P_{2L}(\Theta)]^{n_{2L}} [P_3(\Theta)]^{n_3} .$$

Now since (N_1, N_2, N_3) has a trinomial distribution with probabilities $P_i\ (\Theta)$, $i=1,2,3$, then

$$P_\Theta\{N_i = n_i\ ,\ i=1,2,3\} = \binom{n}{n_1,\ n_2,\ n_3} [P_1(\Theta)]^{n_1} [P_2(\Theta)]^{n_2} [P_3(\Theta)]^{n_3} .$$

Therefore,

$$P_\Theta\{D_{11} \leq d_1,\ldots,\ D_{1N_1} \leq d_{N_1};\ N_{2R} = n_{2R} \,|\, N_i = n_i,\ i=1,2,3\}$$

$$= \frac{\binom{n}{n_1,n_{2R},n_{2L},n_3} \left(\prod_{j=1}^{n_1} P_1(d_j,\Theta)\right) (P_{2R}(\Theta))^{n_{2R}} (P_{2L}(\Theta))^{n_{2L}} (P_3(\Theta))^{n_3}}{\binom{n}{n_1,n_2,n_3} (P_1(\Theta))^{n_1} \ (P_2(\Theta))^{n_2} \ (P_3(\Theta))^{n_3}}$$

$$= \binom{n_2}{n_{2R}} \left\{ \prod_{j=1}^{n_1} \frac{P_1(d_j,\Theta)}{P_1(\Theta)} \right\} \left\{ \frac{P_{2R}(\Theta)}{P_2(\Theta)} \right\}^{n_{2R}} \left\{ \frac{P_{2L}(\Theta)}{P_2(\Theta)} \right\}^{n_{2L}} \qquad (2.3.3)$$

which shows that the joint conditional distribution of D_{1j} , N_{2R} and N_{2L} is a product of the marginals, thereby completing the proof of Lemma 2.3.2. □

Lemmas 2.3.1 and 2.3.2 will be utilized in the next section to develop two test statistics for testing H_o: $\Theta = 0$.

2.4 Two Tests for Difference in Location

Lemma 2.3.1 suggests that one could use the observed differences in C1 to define a conditional test (conditional on $N_1 = n_1$) of H_o: $\Theta = 0$. As a consequence of the symmetry of the conditional null distribution of the D_{1j}, $j=1,2,\ldots, n_1$, any test for the location of a symmetric population is a reasonable candidate for a test of H_o: $\Theta = 0$. For example, one might consider using a linear signed rank test. The well-known Wilcoxon signed rank test (Randles and Wolfe, 1979, p. 322) is an obvious choice due to its high efficiency for a wide variety of distributions and the availability of tables of critical values for a large selection of sample sizes. Accordingly, consider the statistic $T_{1n} \equiv T_{1n}(0)$ where

$$T_{1n}(\Theta) = \sum_{j=1}^{N_1} \psi\,(D_{1j} - \Theta)\; R_j^+\,(\Theta)$$

with R_j^+ (Θ) denoting the rank of $|D_{1j} - \Theta|$ among

$$|D_{11} - \Theta|, \quad |D_{12} - \Theta|, \ldots, \quad |D_{1N_1} - \Theta| .$$

Clearly, the D_i in C2 also contain information about H_o. Lemma 2.3.2 implies that conditional on $N_1 = n_1$, $N_2 = n_2$, any test based on N_{2R} and N_{2L} will be independent of the test based on T_{1n}. Since, as will subsequently be shown, under H_o

$$E\{N_{2R} - N_{2L} \mid N_i = n_i,\ i=1,2,3] = 0,$$

a second statistic for testing H_o is T_{2n} where

$$T_{2n} = N_{2R} - N_{2L} = 2N_{2R} - N_2 \quad .$$

Some properties of the conditional distributions of T_{1n} and T_{2n} follow directly from Lemmas 2.3.1 and 2.3.2. These properties are summarized in Lemma 2.4.1.

<u>Lemma 2.4.1</u>: Let $F_\Theta(\cdot)$ be defined as in equation (2.2.2). Also, let $E_\Theta(\cdot)$ and $Var_\Theta(\cdot)$ denote the expectation and variance with respect to $P_\Theta(\cdot)$. Under (A1), (A2), (A3), and (A4):

(a) Conditional on $N_i = n_i$, i=1,2,3 , T_{1n} and T_{2n} are independent.

(b) $E_\Theta\, [T_{1n} \mid N_i = n_i,\ i=1,2,3] = \mu_{1n}(\theta)$,

and

$$Var_\theta[T_{1n} \mid N_i = n_i,\ i=1,2,3] = \sigma^2_{1n}(\theta) + 0(n_1{}^2)$$

where

$$\mu_{1n}(\theta) = \binom{n_1}{2} \int_{-\infty}^{\infty} 1 - F_\Theta(-x)\, dF_\Theta(x) \quad + n_1 \quad [1 - F_\Theta(0)] \quad ,$$

$$\sigma^2_{1n}(\theta) = 4\, n_1^{-1} \binom{n_1}{2}^2 \quad \{ \int_{-\infty}^{\infty} [1 - F_\Theta(-x)]^2 dF_\Theta(x)$$

$$-[\int_{-\infty}^{\infty} 1 - F_\Theta(-x) dF_\Theta(x)]^2\} \quad .$$

(c) $E_\Theta\, [T_{2n} \mid N_i = n_i,\ i=1,2,3] = \mu_{2n}(\theta)$

and

$$Var_\Theta\, [T_{2n} \mid N_i = n_i,\ i=1,2,3] = \sigma^2_{2n}(\theta)$$

where

$$\mu_{2n}(\theta) = n_2 \; [P_2(\theta)]^{-1} \; [P_{2R}(\theta)-P_{2L}(\theta)]$$

and

$$\sigma^2_{2n}(\theta) = 4 \; n_2[P_2(\theta)]^{-2}[P_{2R}(\theta)][P_{2L}(\theta)] \; .$$

Proof:

The proof of (a) follows directly from Lemma 2.3.2 since T_{1n} is only a function of D_{1j}, j=1, ..., N_1, and T_{2n} is only a function of N_{2R}.

The proof of (b) follows by noting that

$$T_{1n} = \sum_{j=1}^{n_1} \psi(D_{1j}) \quad R_j^+(0)$$

$$= \sum\sum_{1 \le i < j \le n_1} \psi(D_{1i} + D_{1j}) + \sum_{j=1}^{n_1} \psi(D_{1j})$$

$$= \binom{n_1}{2} U_{1,n_1} + n_1 \, W_{1,n_1}$$

where

$$U_{1,n_1} = \binom{n_1}{2}^{-1} \sum\sum_{1 \le i < j \le n_1} \psi(D_{1i} + D_{1j})$$

$$W_{1,n_1} = n_1^{-1} \sum_{j=1}^{n_1} \psi(D_{1j})$$

are two U-statistics (Randles and Wolfe, 1979, page 83). Since U_{1,n_1} and W_{1,n_1} are unbiased estimates $P_\Theta(D_{1i} + D_{1j} > 0)$ and $P_\Theta(D_{1j} > 0)$, respectively, then

$$\mu_{1n}(\theta) = \binom{n_1}{2} P_\theta(D_{1i} + D_{1j} > 0) + n_1 \cdot P_\theta(D_{1j} > 0).$$

Upon noting

$$P_\theta(D_{1i} + D_{1j} > 0) = \int_{-\infty}^{\infty} P_\theta(D_{1i} > - D_{1j} \mid D_{1j} = x) \, d\, P_\theta(D_{1j} \leq x)$$

$$= \int_{-\infty}^{\infty} [1 - F_\theta(-x)]dF_\theta(x)$$

and

$$P_\theta(D_{1j} > 0) = 1 - F_\theta(0),$$

the expression for $\mu_{1n}(\theta)$ follows.

Now,

$$\mathrm{Var}_\theta\, [T_{1n} \mid N_i = n_i, i=1,2,3]$$

$$= \binom{n_1}{2}^2 \mathrm{Var}_\theta\, U_{1,n_1} + n_1^2 \;\mathrm{Var}_\theta\, W_{1,n_1}$$

$$+ 2 \binom{n_1}{2} n_1 \,\mathrm{Cov}_\theta(U_{1,n_1}, W_{1,n_1})$$

$$= \binom{n_1}{2}^2 \mathrm{Var}_\theta\, U_{1,n_1} + n_1^2 \,\mathrm{Var}_\theta\, W_{1,n_1}$$

$$+ 2 \binom{n_1}{2} n_1 \;\rho_{n_1}(\theta) \,[\mathrm{Var}_\theta\, U_{1,n_1}]^{\frac{1}{2}} \,[\mathrm{Var}_\theta\, W_{1,n_1}]^{\frac{1}{2}}$$

where $\rho_{n_1}(\Theta) = \mathrm{Corr}[U_{1,n_1}, W_{1,n_1}]$ is the correlation coefficient between U_{1,n_1} and W_{1,n_1}.

As a result of Lemma A part (iii) of Serfling (1980, p. 183) it follows that

$$\mathrm{Var}_\Theta\, U_{1,n_1} = (2)^2\, n_1^{-1}\, \{E_\Theta[\Psi(D_{11} + D_{12})\Psi(D_{11} + D_{13})] -$$

$$[E_\Theta \Psi(D_{11} + D_{12})]^2\} + O(n_1^{-2})$$

$$= 4n_1^{-1}\, \{P_\Theta(D_{11} + D_{12} > 0,\ D_{11} + D_{13} > 0)$$

$$- [P_\Theta(D_{11} + D_{12} > 0)]^2\} + O(n_1^{-2})$$

$$= 4n_1^{-1}\{ \int_{-\infty}^{\infty} [1 - F_\Theta(-x)]^2\, dF_\Theta(x)$$

$$- [\int_{-\infty}^{\infty} 1 - F_\Theta(-x)\, dF_\Theta(x)]^2\} + O(n_1^{-2}) \quad .$$

Also,

$$\mathrm{Var}_\Theta\, W_{1,n_1} = n_1^{-1}\{E_\Theta[\Psi(D_{11})]^2 - [E_\Theta\, \Psi(D_{11})]^2\}$$

$$= O(n_1^{-1}) \quad .$$

Since $\rho_{n_1}(\theta) = 0(1)$, then it follows that

$$\mathrm{Var}_\Theta\,[T_{1n}\,|\,N_i = n_i, i=1,2,3]$$

$$= 4\,n_1^{-1}\,\binom{n_1}{2}^2\,\{\int_{-\infty}^{\infty}[1 - F_\Theta(-x)]^2\,dF_\Theta(x)$$

$$- [\int_{-\infty}^{\infty}[1 - F_\Theta(-x)\,dF_\Theta(x)]^2\} + \binom{n_1}{2}^2 \cdot 0(n_1^{-2})$$

$$+ n_1^{\,2} \cdot 0(n_1^{\,-1})$$

$$+ 2\binom{n_1}{2} n_1\,[\mathrm{Var}_\Theta\,U_{1,n_1}]^{\frac{1}{2}}\,[0(n_1^{\,-1})]^{\frac{1}{2}}$$

$$= \sigma_{1n}^2(\theta) + \binom{n_1}{2}^2 \cdot 0(n_1^{\,-2}) + n_1^{\,2} \cdot 0(n_1^{\,-1})$$

$$+ 2\binom{n_1}{2} n_1\,[0(n_1^{\,-1})][0(n_1^{\,-\frac{1}{2}})]$$

$$= \sigma_{1n}^2(\theta) + 0(n_1^{\,2}) + 0(n_1)$$

$$+ 0\,(n_1^{3/2})$$

$$= \sigma_{1n}^2(\theta) + 0(n_1^{\,2})\ .$$

The proof of (c) follows from Lemma 2.3.2 when it is noted that

$$P_\theta\{N_{2R} = n_{2R} \mid N_i = n_i,\ i=1,2,3\}$$

$$= \binom{n_2}{n_{2R}} \left\{\frac{P_{2R}(\theta)}{P_2(\theta)}\right\}^{n_{2R}} \left\{\frac{P_{2L}(\theta)}{P_2\ (\theta)}\right\}^{n_{2L}}$$

which indicates that the conditional distribution of N_{2R} is Binomial with parameters n_2 and $[P_2(\theta)]^{-1}\ P_{2R}(\theta)$.

Now,

$$E_\theta\ [T_{2n} \mid N_i = n_i, i=1,2,3]$$

$$= E_\theta[N_{2R} - N_{2L} \mid N_i = n_i, i=1,2,3]$$

$$= n_2\ [P_2\ (\theta)]^{-1}\ [P_{2R}\ (\theta) - P_{2L}\ (\theta)]$$

$$= \mu_{2n}\ (\theta)$$

and

$$\mathrm{Var}_\theta\ [T_{2n} \mid N_i = n_i, i=1,2,3]$$

$$= \mathrm{Var}_\theta\ [2N_{2R} - N_2 \mid N_i = n_i, i=1,2,3]$$

$$= 4\ n_2\ [P_2\ (\theta)]^{-1}\ [P_{2R}(\theta)]\ [P_{2L}(\theta)]$$

$$= \sigma^2_{2n}(\theta)$$

which completes the proof of Lemma 2.4.1. □

Note that if $\theta = 0$ then $F_\theta(x)$ is symmetric by part (a) of Lemma 2.3.1 and as a result

$$\mu_{1n}(0) = \binom{n_1}{2} \int_{-\infty}^{\infty} 1 - F_0(-x)dF_0(x) + n_1[1 - F_0(0)]$$

$$= \binom{n_1}{2} \int_{-\infty}^{\infty} F_0(x)\, dF_0(x) \quad + \quad (1/2) \cdot n_1$$

$$= (1/2)\, n_1(n_1 -1) \int_0^1 u\, du + (1/2)\, n_1$$

$$= (1/4)\, n_1(n_1 -1) + (1/2)\, n_1$$

$$= \frac{n_1\,(n_1 + 1)}{4} \quad . \tag{2.4.1}$$

Similarly,

$$\sigma^2_{1n}(0) = \frac{4}{n_1} \binom{n_1}{2}^2 \{(1/3) - (1/2)^2\}$$

$$= \sigma^2_{n_1} + 0(n_1^2) \tag{2.4.2}$$

where

$$\sigma^2_{n_1} = \frac{n_1(n_1+1)(2n_1+1)}{24}$$

is the null variance of the Wilcoxon signed rank statistic based on n_1 observations.

Also, since under (A2) $P_{2R}(0) = P_{2L}(0) = (1/2)\ P_2(0)$,

$$\mu_{2n}\ (0) = n_2 \cdot \frac{P_{2R}(0) - P_{2L}(0)}{P_2(0)}$$

$$= 0 \quad . \tag{2.4.3}$$

Similarly,

$$\sigma_{2n}^2\ (0) = 4n_2 \cdot \frac{P_{2R}(0) \cdot P_{2L}(0)}{[P_2(0)]^2}$$

$$= n_2 \quad . \tag{2.4.4}$$

In Lemma 2.4.2 it is shown that, conditional on $N_1 = n_1$ and $N_2 = n_2$, T_{1n} and T_{2n} have asymptotic normal distributions.

Lemma 2.4.2:

(a) $\lim\limits_{n_1\to\infty} P_\theta\{[\sigma_{1n}(\theta)]^{-1}(T_{1n} - \mu_{1n}(\theta)) \le x \mid N_i = n_i, i=1,2,3\} = \Phi(x)$

where

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp[-y^2/2]\ dy.$$

(b) $\lim\limits_{n_2\to\infty} P_\theta\{[\sigma_{2n}(\theta)]^{-1}(T_{2n} - \mu_{2n}(\theta)) \le x \mid N_i = n_i, i=1,2,3\} = \Phi(x)$.

Proof:

(a) From the proof of Lemma 2.4.1

$$T_{1n} = \binom{n_1}{2} U_{1,n_1} + n_1 W_{1,n_1} .$$

Therefore, it follows that

$$n_1^{1/2} \binom{n_1}{2}^{-1} [T_{1n} - \mu_{1n}(\theta)] =$$

$$n_1^{1/2}[U_{1,n_1} - E_\Theta U_{1,n_1}] + n_1^{3/2}\binom{n_1}{2}^{-1}[W_{1,n_1} - E_\Theta W_{1,n_1}] .$$

Now,

$$W_{1,n_1} = n_1^{-1} \sum_{j=1}^{n_1} \Psi(D_{1j})$$

and since $E_\Theta[\Psi(D_{1j})]^2 = P_\Theta(D_{1j} > 0)$ is finite, then by corollary 3.2.5 of Randles and Wolfe (1979) W_{1,n_1} converges in quadratic mean to $E_\Theta W_{1,n_1} = P_\Theta(D_{1j} > 0)$. Also, $n_1^{3/2} \binom{n_1}{2}^{-1} \to 0$ as $n_1 \to \infty$.

Therefore,

$$\text{p-lim } n_1^{3/2} \binom{n_1}{2}^{-1} [W_{1,n_1} - E_\Theta W_{1,n_1}] = 0$$

as $n_1 \to \infty$, showing that

$$n_1^{1/2} \binom{n_1}{2}^{-1} [T_{1n} - \mu_{1n}(\theta)]$$

and

$$n_1^{\frac{1}{2}} \, [U_{1,n_1} - E_\Theta U_{1,n_1}]$$

have the same limiting distribution as $n_1 \to \infty$.

By taking

$$r^2 \zeta_1 = n_1 \binom{n_1}{2}^{-2} \sigma_{1n}^2(\theta)$$

in Theorem 3.3.13 of Randles and Wolfe (1979), it is seen that

$$\frac{U_{1,n_1} - E_\Theta(U_{1,n_1})}{\binom{n_1}{2}^{-1} \sigma_{1n}(\Theta)}$$

has an asymptotic normal distribution.

Note that $\sigma_{1n}^2(\theta) = 0(n_1^3)$ so that $\mathrm{Var}_\Theta\,[T_{1n} \mid N_i = n_i,\ i=1,2,3] \cdot \sigma_{1n}^{-2}(\theta)$ $= 1 + 0(n_1^{-1})$ which implies that $\mathrm{Var}_\Theta\,[T_{1n} \mid N_i = n_i,\ i=1,2,3] \cdot \sigma_{1n}^{-2}(\theta)$ tends to one as $n_1 \to \infty$.

Therefore,

$$\frac{n_1^{\frac{1}{2}} \binom{n_1}{2}^{-1} [T_{1n} - \mu_{1n}(\theta)]}{n_1^{\frac{1}{2}} \binom{n_1}{2}^{-1} \sigma_{1n}(\Theta)} = \frac{T_{1n} - \mu_{1n}(\Theta)}{\sigma_{1n}(\Theta)}$$

has a limiting standard normal distribution which completes the proof of (a).

The proof of (b) follows since the conditional distribution of N_{2R} is binomial with parameters n_2 and $[P_2(\theta)]^{-1} P_{2R}(\theta)$. □

Lemma 2.4.3 follows from the conditional independence of T_{1n} and T_{2n} shown in Lemma 2.4.1.

<u>Lemma 2.4.3</u>:

$$\lim_{\min(n_1,n_2)\to\infty} P_\theta\left\{\frac{T_{1n} - \mu_{1n}(\theta)}{\sigma_{1n}(\theta)} \leq x, \frac{T_{2n} - \mu_{2n}(\theta)}{\sigma_{2n}(\theta)} \leq y \,\middle|\, N_i = n_i,\ i=1,2,3\right\}$$

$$= \Phi(x)\Phi(y) .$$

The conditional independence of T_{1n} and T_{2n} indicates that two conditionally independent tests of H_o can be employed using the same data set. In the next chapter a test for H_o based on a combination of T_{1n} and T_{2n} is considered.

CHAPTER THREE

A CLASS OF TESTS FOR TESTING FOR DIFFERENCE IN LOCATION

3.1 Introduction and Summary

In Chapter Two the general setup was given for the problem of testing for the difference in locations of two treatments applied to matched pairs of subjects when random censoring is present. In addition, two conditionally independent test statistics were proposed for testing H_o: $\theta=0$. The main focus of this chapter is on a class of test statistics consisting of certain linear combinations of these two test statistics.

In Section 3.2 a class of test statistics for testing H_o: $\theta=0$ is proposed and some properties of the conditional distributions (conditional on N_1 and N_2) of the members in this class are stated. In particular it is shown that the statistics in the proposed class are conditionally distribution-free. The asymptotic null distributions of these statistics are shown to be normal in Section 3.3. Section 3.4 contains a worked example to demonstrate the use of one member of this class for testing H_o when the sample size is small. In Section 3.5 a method to choose one test statistic from the class of test statistics is suggested.

3.2. An Exact Test for Difference in Location

In this section a class of exact conditional tests of H_o based on T_{1n} and T_{2n} is proposed.

Let

$$T_{1n}^*(N_1) = [N_1(N_1 + 1)(2N_1 + 1)/24]^{-\frac{1}{2}} [T_{1n} - N_1(N_1 + 1)/4]$$

$$T_{2n}^*(N_2) = N_2^{-\frac{1}{2}} T_{2n}$$

and $\{L_n\}$ be a sequence of random variables satisfying:

Condition I

(a) L_n is only a function of N_1 and N_2,

(b) $0 \leqq L_n \leqq 1$ with probability one,

(c) There exists a constant L such that p-lim $L_n = L$ as $n \to \infty$.

The proposed test statistic is

$$T_n(N_1,N_2) = (1-L_{1n})^{\frac{1}{2}} T_{1n}^*(N_1) + L_n^{\frac{1}{2}} T_{2n}^*(N_2) \quad .$$

Note that under H_o and conditional on $N_i=n_i$, i=1,2, $T_{1n}^*(N_1)$ has a distribution which is the same as the distribution of a standardized Wilcoxon signed rank statistic based on n_1 observations (Randles and Wolfe, 1979), while $T_{2n}^*(N_2)$ is distributed as an independent standardized Binomial random variable with parameters n_2 and ½.

Therefore, $T_n(N_1,N_2)$ is conditionally distribution-free under H_o and a test of H_o can be based on $T_n(N_1,N_2)$.

Hereafter, for the sake of simplicity $T_n(n_1,n_2)$ will denote the statistic $T_n(N_1,N_2)$ conditioned on $N_i=n_i$, $i=1,2$. Useful properties of $T_n(n_1,n_2)$ are stated in Lemma 3.2.1.

Lemma 3.2.1: Under H_o: $\theta=0$ and (A1), (A2), (A3), and (A5),

(a) $E_0\,[T_n(n_1,n_2)] = 0$,

(b) $\mathrm{Var}_0\,[T_n(n_1,n_2)] = 1$,

(c) $T_n(n_1,n_2)$ is symmetrically distributed about 0.

Proof:

Under H_o, and conditional on $N_i=n_i, i=1,2$, $T_{1n}^*(N_1)$ has the same distribution as a standardized Wilcoxon signed rank statistic, and $T_{2n}^*(N_2)$ has the same distribution as a standardized Binomial random variable. Also, conditional on $N_i = n_i$, $i=1,2$, L_n is nonrandom. Therefore,

$$E_0[T_{1n}^*(N_1)\,|\,N_i = n_i, i=1,2,3] = 0 \ ,$$

$$E_0[T_{2n}^*(N_2) \mid N_i = n_i, i=1,2,3] = 0 \ ,$$

$$\mathrm{Var}_0\ [T_{1n}^*(N_1) \mid N_i = n_i, i=1,2,3] = 1 \quad ,$$

and

$$\mathrm{Var}_0[T_{2n}^*(N_2) \mid N_i = n_i, i=1,2,3] = 1 \quad .$$

As a result (a) follows directly, and by Lemma 2.4.1(a) it is clear that

$$\begin{aligned} \mathrm{Var}_0\ [T_n(n_1,n_2)] &= (1-L_n)\ \mathrm{Var}_0\ [T_{1n}^*(N_1) \mid N_i = n_i, i=1,2,3] \\ &\quad + L_n \mathrm{Var}_0[T_{2n}^*(N_2) \mid N_i=n_i, i=1,2,3] \\ &= (1 - L_n) + L_n, \end{aligned}$$

from which (b) follows.

(c) It is known (Randles and Wolfe, 1979) that the Wilcoxon signed rank statistic and the binomial random variable with $p=\frac{1}{2}$ are symmetrically distributed about their respective means. Therefore, it follows that the conditional null distributions of

$T_{1n}^*(N_1)$ and $T_{2n}^*(N_2)$ are symmetric about 0 under H_o. Since $T_{1n}^*(N_1)$ and $T_{2n}^*(N_2)$ are conditionally independent, then the proof of Lemma 3.2.1 is complete. $\square$

The conditional distribution of $T_n(N_1,N_2)$ is a convolution of two well-known discrete distributions (i.e., the Wilcoxon signed rank and Binomial distributions). The critical values for tests based on $T_n(n_1,n_2)$

are easily obtained for small and moderate values of n_i,i=1,2. Note that the tables given in the appendix are applicable to the test statistic

$$T_n(n_1,n_2) = (.5)^{\frac{1}{2}} T_{1n}^*(n_1) + (.5)^{\frac{1}{2}} T_{2n}^*(n_2)$$

for $n_1 = 1,\ldots,10$ and $n_2 = 1,2,\ldots,15$ at the .01,.025,.05, and .10 levels of significance. Tables of critical values for other choices of $T_n(n_1,n_2)$ can be similarly produced but they depend on the choice of L_n.

For large values of n_i,i=1,2, a test of H_o can be based on the asymptotic distribution of $T_n(N_1,N_2)$ as seen in Section 3.3.

3.3 The Asymptotic Distribution of the Test Statistic

In this section it will be shown that the null distribution of $T_n(N_1,N_2)$ converges in distribution to a standard normal distribution as n tends to infinity, thereby showing that $T_n(N_1,N_2)$ is asymptotically distribution-free. This fact can be utilized to derive a large sample test of H_o.

The main result of this section is Theorem 3.3.3 which will be proved after establishing several preliminary results. The result stated in Lemma 3.3.1 is a generalization of Theorem 1 of Anscombe (1952).

Lemma 3.3.1: Let $\{T_{n_1,n_2}\}$ for $n_1=1,2,\ldots$, $n_2=1,2,\ldots$ be an array of random variables satisfying conditions (i) and (ii).

Condition (i): There exists a real number γ, an array of positive numbers $\{\omega_{n_1, n_2}\}$, and a distribution function $F(\cdot)$ such that

$$\lim_{\min(n_1,n_2)\to\infty} P\{T_{n_1,n_2} - \gamma \leq x\, \omega_{n_1,n_2}\} = F(x) ,$$

at every continuity point x of $F(\cdot)$.

Condition (ii): Given any $\varepsilon > 0$ and $\eta > 0$ there exist

$\nu \equiv \nu(\varepsilon,\eta)$ and $c \equiv c(\varepsilon,\eta)$ such that whenever $\min(n_1,n_2) > \nu$ then

$$P\{|T_{n_1',n_2'} - T_{n_1,n_2}| < \varepsilon\, \omega_{n_1,n_2} \text{ for all } n_1',n_2' \text{ such that}$$

$$|n_1'-n_1| < cn_1, \quad |n_2'-n_2| < cn_2\} > 1-\eta.$$

Let $\{n_r\}$ be an increasing sequence of positive integers tending to infinity and let $\{N_{1r}\}$ and $\{N_{2r}\}$ be random variables taking on positive integer values such that

$$\text{p-lim } (\lambda_i n_r)^{-1} N_{ir} = 1$$

as $r\to\infty$ for $0 < \lambda_i < 1$, $i=1,2$.

Then at every continuity point x of $F(\cdot)$,

$$\lim_{r\to\infty} P\{T_{N_{1r},N_{2r}} - \gamma < x\, \omega_{[\lambda_1 n_r],[\lambda_2 n_r]}\} = F(x)$$

where [x] denotes the greatest integer less than or equal to x.

Proof:

For given $\varepsilon>0$ and $\eta>0$ let ν^* and c satisfy condition (ii). Also, let ν^{**} be chosen such that for $\min(\lambda_1 n_r, \lambda_2 n_r) > \nu^{**}$

$$P\{ |N_{ir} - \lambda_i n_r| < c\, \lambda_i n_r\} > 1 - \frac{\eta}{2}\,, \qquad i=1,2.$$

Then for $\min(\lambda_1 n_r, \lambda_2 n_r) > \max(\nu^*, \nu^{**}) \equiv \nu$

$$\begin{aligned} P\{ &|N_{1r} - \lambda_1 n_r| < c\lambda_1 n_r \text{ and } |N_{2r} - \lambda_2 n_r| < c\lambda_2 n_r\} \\ &= P\{ |N_{1r} - \lambda_1 n_r| < c\lambda_1 n_r\} + P\{|N_{2r} - \lambda_2 n_r| < c\lambda_2 n_r\} \\ &\quad - P\{|N_{1r} - \lambda_1 n_r| < c\lambda_1 n_r \text{ or } |N_{2r} - \lambda_2 n_r| < c\lambda_2 n_r\} \\ &\geq (1 - \frac{\eta}{2}) + (1 - \frac{\eta}{2}) - 1 \\ &= 1 - \eta\,. \end{aligned} \tag{3.3.1}$$

Let $n_{1r} = [\lambda_1 n_r]$ and $n_{2r} = [\lambda_2 n_r]$ and define the events A_r, B, O_r, D_r, and E_r as follows:

$$A_r: |N_{1r} - n_{1r}| < cn_{1r}\,, |N_{2r} - n_{2r}| < cn_{2r} \text{ and } |T_{N_{1r},N_{2r}} - T_{n_{1r},n_{2r}}| < \varepsilon\omega_{n_{1r},n_{2r}}\,,$$

$$B: |T_{n_1',n_2'} - T_{n_1,n_2}| < \varepsilon\,\omega_{n_1,n_2} \text{ for all } n_i',\ i=1,2, \text{ such that}$$

$$n_i > \nu, \text{ and } |n_i' - n_i| < cn_i\,,\ i=1,2\,,$$

O_r: $|N_{1r} - n_{1r}| < cn_{1r}$ and $|N_{2r} - n_{2r}| < cn_{2r}$,

D_r: $T_{N_{1r},N_{2r}} - \gamma \leq x\, \omega_{n_{1r},n_{2r}}$ and A_r ,

E_r: $T_{N_{1r}, N_{2r}} - \gamma \leq x\, \omega_{n_{1r},n_{2r}}$ and $A_r^{\ c}$,

where $A_r^{\ c}$ denotes the complement of A_r.

Note that for given Θ such that $\min(n_{1r},n_{2r}) > \nu$, the event B implies $|T_{n_1',n_2'} - T_{n_{1r}, n_{2r}}| < \varepsilon\, \omega_{n_{1r},n_{2r}}$ for all n_i', i=1,2, such that $n_i > \nu$ and $|n_i' - n_i| < cn_i$, i=1,2. Since A_r is equal to O_r and $|T_{N_{1r},N_{2r}} - T_{n_{1r},n_{2r}}| < \varepsilon\, \omega_{n_{1r},n_{2r}}$ if follows that B and O_r imply A_r.

Hence, for given r such that $\min(n_{1r},n_{2r}) > \nu$,

$$\begin{aligned} P\{A_r\} &\geq P\{B \text{ and } O_r\} \\ &= P\{B\} - P\{B \text{ and } O_r^{\ c}\} \\ &\geq P\{B\} - P\{O_r^{\ c}\}. \end{aligned}$$

By condition (ii), $P\{B\} > 1-\eta$ and by (3.3.1) for given r such that $\min(n_{1r},n_{2r}) > \nu$, $P\{O_r\} > 1-\eta$, which implies $P\{O_r^{\ c}\} < \eta$. Therefore, if $\min(n_{1r},n_{2r}) > \nu$, then

$$P\{A_r\} \geq 1-2\eta \qquad (3.3.2)$$

which implies

$$P\{A_r^{\,c}\} \leq 2\eta. \tag{3.3.3}$$

Now,

$$P\{D_r\} = P\{T_{N_{1r}, N_{2r}} - \gamma \leq x\, \omega_{n_{1r}, n_{2r}} \text{ and } A_r\}$$

$$\leq P\{T_{N_{1r}, N_{2r}} - \gamma \leq x\, \omega_{n_{1r}, n_{2r}} \text{ and}$$

$$|T_{N_{1r}, N_{2r}} - T_{n_{1r}, n_{2r}}| < \varepsilon\, \omega_{n_{1r}, n_{2r}}\}$$

$$= P\{T_{N_{1r}, N_{2r}} + (T_{n_{1r}, n_{2r}} - T_{N_{1r}, N_{2r}}) - \gamma \leq$$

$$x\omega_{n_{1r}, n_{2r}} + (T_{n_{1r}, n_{2r}} - T_{N_{1r}, N_{2r}})$$

$$\text{and } |T_{N_{1r}, N_{2r}} - T_{n_{1r}, n_{2r}}| < \varepsilon\, \omega_{n_{1r}, n_{2r}}\} \;.$$

Since

$$|T_{N_{1r}, N_{2r}} - T_{n_{1r},n_{2r}}| < \varepsilon\, \omega_{n_{1r}, n_{2r}}$$

implies

$$x\, \omega_{n_{1r},n_{2r}} + (T_{n_{1r},n_{2r}} - T_{N_{1r},N_{2r}}) < (x + \varepsilon)\omega_{n_{1r},n_{2r}} \;,$$

it follows that

$$P(D_r) \leq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x + \varepsilon)\omega_{n_{1r},n_{2r}}\} \;.$$

Now, D_r and E_r are mutually exclusive events such that $D_r \cup E_r$ is the event $T_{N_{1r}, N_{2r}} - \gamma \leq x\,\omega_{n_{1r},n_{2r}}$, so that

$$P\{D_r \cup E_r\} = P\{D_r\} + P\{E_r\}$$

$$\leq P\{T_{n_{1r}, n_{2r}} - \gamma \leq (x+\varepsilon)\,\omega_{n_{1r}, n_{2r}}\} + P\{E_r\}$$

$$\leq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x+\varepsilon)\,\omega_{n_{1r},n_{2r}}\} + P\{A_r^{c}\} \quad .$$

Therefore, from (3.3.3), for min $(n_{1r},n_{2r}) > \nu$,

$$P\{T_{N_{1r},N_{2r}} - \gamma \leq x\omega_{n_{1r},n_{2r}}\}$$

$$\leq P\{T_{n_{1r}, n_{2r}} - \gamma \leq (x+\varepsilon)\,\omega_{n_{1r}, n_{2r}}\} + 2\eta \quad . \qquad (3.3.4)$$

Now, for $\min(n_{1r},n_{2r})>\nu$

$$P\{D_r\} = P\{\, T_{N_{1r}, N_{2r}} - \gamma \leq xw_{n_{1r}, n_{2r}} \text{ and } A_r\}$$

$$= P\{\, T_{N_{1r}, N_{2r}} + (T_{n_{1r}, n_{2r}} - T_{N_{1r}, N_{2r}}) - \gamma \leq$$

$$x\omega_{n_{1r},n_{2r}} + (T_{n_{1r},n_{2r}} - T_{N_{1r},N_{2r}}) \text{ and } A_r\} \quad .$$

Now since the event A_r implies

$$-\,\varepsilon\omega_{n_{1r},n_{2r}} \leq T_{n_{1r},n_{2r}} - T_{N_{1r},N_{2r}} \quad ,$$

then the event

$$\{T_{n_{1r},n_{2r}} - \gamma \leq x\omega_{n_{1r},n_{2r}} - \varepsilon\omega_{n_{1r},n_{2r}} \text{ and } A_r\}$$

implies the event

$$\{T_{n_{1r},n_{2r}} - \gamma \leq x\omega_{n_{1r},n_{2r}} + (T_{n_{1r},n_{2r}} - T_{N_{1r},N_{2r}}) \text{ and } A_r\} \ .$$

Therefore,

$$P\{D_r\} \geq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}} \text{ and } A_r\}$$

$$= P\{T_{n_{1r},\ n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},\ n_{2r}}\}$$

$$+ P\{A_r\} - P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}} \text{ or } A_r\} \ .$$

Therefore, from (3.3.2), if $\min(n_{1r},n_{2r}) > \nu$, then

$$P(D_r) \geq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}}\} + (1-2\eta) - 1$$

$$\geq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}}\} - 2\eta \ .$$

As a result, for given r such that $\min(n_{1r},n_{2r})>\nu$

$$P\ \{T_{N_{1r},N_{2r}} - \gamma \leq x\omega_{n_{1r},n_{2r}}\}$$

$$= P\{D_r\} + P\{E_r\}$$

$$\geq P\{D_r\}$$

$$\geq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}}\} - 2\eta\ . \qquad (3.3.5)$$

In view of equations (3.3.4) and (3.3.5) it follows that for given r such that min $(n_{1r},n_{2r}) > \nu$

$$P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}}\} - 2\eta$$

$$\leq P\{T_{N_{1r},N_{2r}} - \gamma \leq x\ \omega_{n_{1r},n_{2r}}\}$$

$$\leq P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\ \omega_{n_{1r},n_{2r}}\} + 2\eta. \qquad (3.3.6)$$

Furthermore, condition (i) implies

$$\lim_{r\to\infty} P\{T_{n_{1r},n_{2r}} - \gamma \leq (x-\varepsilon)\omega_{n_{1r},n_{2r}}\} = F(x-\varepsilon)$$

and

$$\lim_{r\to\infty} P\{T_{n_{1r},n_{2r}} - \gamma \leq (x+\varepsilon)\omega_{n_{1r},n_{2r}}\} = F(x+\varepsilon)$$

at all continuity points $(x-\varepsilon)$ and $(x+\varepsilon)$ of $F(\cdot)$ Therefore, if x is a continuity point of $F(\cdot)$, then $(x-\varepsilon)$ and $(x+\varepsilon)$ can be chosen to be continuity points of $F(\cdot)$ since the set of continuity points is dense. Consequently,

$$\lim_{r\to\infty} P\{T_{N_{1r},N_{2r}} - \gamma \leq x\omega_{n_{1r},n_{2r}}\} = F(x) \quad ,$$

which completes the proof of Lemma 3.3.1. □

Lemma 3.3.2 below follows from Lemmas 2.4.1, 2.4.3, and 3.2.1.

Lemma 3.3.2: Let $L_n = L_n(N_1,N_2)$ be defined as in Condition I. Under H_o: $\theta=0$, and (A1), (A2), (A3), and (A4) ,

$$P_0\{T_n(N_1,N_2) \le x \mid N_i = n_i,\ i=1,2,3\} \text{ tends to } \Phi(x)$$

as min $(n_1,n_2) \to\infty$ subject to the condition

$$\lim_{\min(n_1,n_2)\to\infty} L_n(n_1,n_2) = L \quad .$$

Proof:

Lemma 2.4.3 shows that $\sigma_{1n}^{-1}(\theta)[T_{1n}-\mu_{1n}(\theta)]$ and $\sigma_{2n}^{-1}(\theta)[T_{2n}-\mu_{2n}(\theta)]$ possess conditional distributions, conditional on $N_i=n_i$, $i=1,2,3$, which converge to standard normal distributions as $\min(n_1,n_2)$ tends to infinity. From equations (2.4.1) and (2.4.2) it is seen that

$$\mu_{1n}(0) = \frac{n_1(n_1+1)}{4} \quad ,$$

and

$$\lim_{n\to\infty} \sigma_{1n}^2(0) \cdot \frac{n_1(n_1+1)(2n_1+1)}{24} = 1 \quad .$$

Therefore under H_o, $T_{1n}^*(N_1)$ has a conditional distribution which is asymptotically standard normal. Similarly, under H_o: $\theta=0$, $T_{2n}^*(N_2)$ has a conditional distribution which is asymptotically standard normal. As a consequence of Lemmas 2.4.3 and 3.2.1, Lemma 3.3.2 holds. □

The following Theorem contains the main result of this section. The proof makes use of the result of Sproule (1974) stated as Lemma 3.3.3.

Lemma 3.3.3: Suppose that

$$U_n = \binom{n}{r}^{-1} \sum_{\beta \varepsilon B} f(X_{\beta_1}, \ldots, X_{\beta_r})$$

where B = set of all unordered subsets of r integers chosen without replacement from the set of integers $\{1,2, \ldots, n\}$ is a U-statistic of degree r with a symmetric kernel $f(\cdot)$. Let $\{n_r\}$ be an increasing sequence of positive integers tending to ∞ as $r \to \infty$ and $\{N_r\}$ be a sequence of random variables taking on positive integer valves with probability one. If $E\{f(X_1,\ldots X_r)\}^2 < \infty$,

$$\lim_{n \to \infty} \operatorname{Var}(n^{\frac{1}{2}} U_n) = r^2 \zeta_1 > 0 ,$$

and

$$\operatorname*{p-lim}_{r \to \infty} n_r^{-1} N_r = 1,$$

then

$$\lim_{r \to \infty} P\{(U_{N_r} - EU_{N_r}) \leq N_r^{-\frac{1}{2}} x (r^2 \zeta_1)^{\frac{1}{2}}\} = \Phi(x).$$

Proof:

It should be noted that in order to prove Lemma 3.3.3, only condition C2 of Anscombe (1952) needs to be verified since condition C1 is valid under the hypothesis. If C1 and C2 of Anscombe (1952) are valid then Theorem 1 of Anscombe (1952), and thus Lemma 3.3.3, is valid. This proof is contained in the proof of Theorem 6 of Sproule (1974). □

<u>Theorem 3.3.4</u>: Under H_o: $\theta=0$, and (A1),(A2), (A3), (A4), (A5), and for every real value x,

$$\lim_{n\to\infty} P_0\{T_n(N_1,N_2) \leq x\} = \Phi(x).$$

<u>Proof</u>:

Let $T_{n_1, n_2} \equiv T_n(n_1,n_2)$. Then Lemma 3.3.2 shows that $\{T_{n_1,n_2}\}$ for $n_1=1,2,\ldots$, $n_2=1,2,\ldots$, satisfies condition (i) of Lemma 3.3.1 with $\gamma=0$ and $\omega_{n_1, n_2} = 1$. Note that under H_o,

$$P_i(0) = \underset{n\to\infty}{\text{p-lim}}\; n^{-1}N_i \equiv \lambda_i \quad .$$

Therefore, from assumption (A5) it can be seen that $\lambda_i > 0$ for at least one i=1,2. Since $P_i(0) = 0$ implies $P_0\{N_i > 0\} = 0$, then $T_n(N_1,N_2)$ is equivalent to one $T^*_{in}(N_i)$ if $P_i(0) = 0$ for exactly one i=1,2. In that case by straightforward application of Theorem 1 of Anscombe (1952) and Lemma 3.3.3 in a manner similar to the argument which follows in the case where $\lambda_i > 0$ for i=1,2 , Theorem 3.3.4 follows. Consequently, it will be assumed in the following argument that $\lambda_i > 0$ for i=1,2. Now all that is needed is to show condition (ii) of Lemma 3.3.1 is satisfied to prove Theorem 3.3.3.

Lemma 2.4.2 shows that $\sigma^{-1}_{1n}(\theta)\ [T_{1n}-\mu_{1n}(\theta)]$ has a limiting standard normal distribution. As a result $T^*_{1n}\ (n_1)$ has a limiting standard normal distribution under H_o: $\theta=0$. Now, under H_o it can be seen that

$$T^*_{1n}(n_1) = (3n_1)^{\frac{1}{2}}(U_{1,n_1}-E_\theta U_{1,n_1}) + o_p(n_1^{-\frac{1}{2}}) , \qquad (3.3.7)$$

where U_{1,n_1} is the U-statistic defined in the proof of Lemma 2.4.1(b). As a consequence of Lemma 3.3.3, U_{1,n_1} satisfies condition C2 of Anscombe (1952). In other words, for given $\varepsilon > 0$ and $\eta > 0$, there exists ν_1 and $d_1 > 0$ such that for any $n_1 > \nu_1$

$$P\{(3n_1)^{\frac{1}{2}} \left| U_{1,n_1'} - U_{1,n_1} \right| < \varepsilon \text{ for all } n_1' \text{ such that } \left| n_1' - n_1 \right| < d_1 n_1\}$$

$$\geq 1 - \eta .$$

As a result of (3.3.7) it can be seen that $T^*_{1n}(n_1)$ satisfies condition C2 of Anscombe (1952). In other words, for given $\varepsilon_1 > 0$ and $\eta > 0$ there exists ν_1 and $d_1 > 0$ such that for any $n_1 > \nu_1$

$$P\{ \left| T^*_{1n}(n_1') - T^*_{1n}(n_1) \right| < \varepsilon \text{ for all } n_1' \text{ such}$$

$$\text{that} \left| n_1' - n_1 \right| < d_1 n_1\} \geq 1 - \eta. \tag{3.3.8}$$

Now let $D_{21}, \ldots, D_{2n_2}$ be as defined in equation (2.2.1), let D^o_{2i} denote the true difference corresponding to D_{2i}, and let $I[D_{2i} < D^o_{2i}] = 1$ or 0 according to whether $D_{2i} < D^o_{2i}$ or $D_{2i} \geq D^o_{2i}$.

Note that conditional on $N_2 = n_2$,

$$T_{2n} = 2n_{2r} - n_2$$

$$= 2 \sum_{i=1}^{n_2} I\,[D_{2i} < D^o_{2i}] - n_2$$

$$= 2n_2\, U_{2,\, n_2} - n_2$$

where

$$U_{2,n_2} = n_2^{-1} \sum_{i=1}^{n_2} I[D_{2i} < D_{2i}^o]$$

is the U-statistic estimator of $P_\Theta[D_i < D_i^o \mid D_i \in C2]$.

Now by a similar argument used above for $T_{1n}^*(n_1)$, shows that U_{2,n_2}

satisfies condition C2 of Anscombe (1952). Thus under H_o, $T_{2n}^*(n_2)$ satisfies condition C2. In other words, for given $\varepsilon_2 > 0$ and $\eta > 0$, there exists ν_2 and $d_2 > 0$ such that for any $n_2 > \nu_2$

$$P\{ \left| T_{2n}^*(n_2') - T_{2n}^*(n_2) \right| < \varepsilon_2 \text{ for all } n_2' \text{ such that}$$

$$\left| n_2' - n_2 \right| < d_2 n_2 \} \geq 1 - \eta. \tag{3.3.9}$$

Let $\varepsilon > 0$ and $\eta > 0$ be given and let ν_1, ν_2, d_1, d_2 satisfy (3.3.8) and (3.3.9). Let $\nu \equiv \max(\nu_1, \nu_2)$ and $c \equiv \min(d_1, d_2)$. Consider

$T'_{n_1,n_2} = (1 - L)^{\frac{1}{2}} T_{1n}^*(n_1) + L^{\frac{1}{2}} T_{2n}^*(n_2)$, then

$$P\{ \left| T'_{n_1',n_2'} - T'_{n_1,n_2} \right| < 2\varepsilon \text{ for all } n_1', n_2' \text{ such that}$$

$$\left| n_1' - n_1 \right| < cn_1, \ \left| n_2' - n_2 \right| < cn_2 \}$$

$$\geq P\{(1-L)^{\frac{1}{2}}\left|T_{1n}^{*}(n_1') - T_{1n}^{*}(n_1)\right|$$

$$+ L^{\frac{1}{2}}\left|T_{2n}^{*}(n_2') - T_{2n}^{*}(n_2)\right| < 2\,\varepsilon \text{ for all } n_1', n_2' \text{ such that}$$

$$|n_1' - n_1| < cn_1,\ |n_2' - n_2| < cn_2\}$$

$$\geq P\{(1-L)^{\frac{1}{2}}\left|T_{1n}^{*}(n_1') - T_{1n}^{*}(n_1)\right| < \varepsilon \text{ and}$$

$$L^{\frac{1}{2}}\left|T_{2n}^{*}(n_2') - T_{2n}^{*}(n_2)\right| < \varepsilon \qquad \text{for all}$$

$$n_1', n_2' \text{ such that } |n_1' - n_1| < cn_1, |n_2' - n_2| < cn_2\}$$

$$= P\{(1-L)^{\frac{1}{2}}\left|T_{1n}^{*}(n_1') - T_{1n}^{*}(n_1)\right| < \varepsilon \text{ for all } n_1'$$

$$\text{such that } |n_1' - n_1| < cn_1\}$$

$$+ \ P\{L^{\frac{1}{2}}\left|T_{2n}^{*}(n_2') - T_{2n}^{*}(n_2)\right| < \varepsilon \qquad \text{for all } n_2'$$

$$\text{such that } |n_2' - n_2| < cn_2\}$$

$$- \ P\{(1-L)^{\frac{1}{2}}\left|T_{1n}^{*}(n_1') - T_{1n}^{*}(n_1)\right| < \varepsilon \qquad \text{or}$$

$$L^{\frac{1}{2}}\left|T_{2n}^{*}(n_2') - T_{2n}^{*}(n_2)\right| < \varepsilon \qquad \text{for all } n_1', n_2'$$

$$\text{such that } |n_1' - n_1| < cn_1 \text{ and } |n_2' - n_2| < cn_2\}$$

$$\geq P\{(1 - L)^{\frac{1}{2}} \left| T_{1n}^{*}(n_1') - T_{1n}^{*}(n_1) \right| < \varepsilon \quad \text{for all } n_1'$$

$$\text{such that } \left| n_1' - n_1 \right| < cn_1\}$$

$$+ \quad P\{L^{\frac{1}{2}} \left| T_{2n}^{*}(n_2') - T_{2n}^{*}(n_2) \right| < \varepsilon \quad \text{for all } n_2'$$

$$\text{such that } \left| n_2' - n_2 \right| < cn_2\}$$

$$- 1 \quad . \tag{3.3.10}$$

Assume $0 < L < 1$ for, if not, then by straightforward application of Lemma 3.3.3 and Anscombe's Theorem 1 (1952)

$$T_n'(N_1,N_2) = (1 - L)^{\frac{1}{2}} T_{1n}^{*}(N_1) + L^{\frac{1}{2}} T_{2n}^{*}(N_2)$$

has an asymptotic normal distribution as $n \to \infty$.

Now using inequalities (3.3.8) and (3.3.9) and applying them to (3.3.10) with $\varepsilon = \min \{\varepsilon_1(1 - L)^{-\frac{1}{2}}, \varepsilon_2 L^{-\frac{1}{2}}\}$, then

$$P\{ \left| T'_{n_1', n_2'} - T'_{n_1, n_2} \right| < 2\varepsilon \quad \text{for all } n_1', n_2'$$

$$\text{such that } \left| n_1' - n_1 \right| < cn_1, \ \left| n_2' - n_2 \right| < cn_2\}$$

$$\geq (1 - \eta) + (1 - \eta) - 1 = 1-2\eta \ .$$

Therefore, T'_{n_1, n_2} satisfies condition (ii) of Lemma 3.3.1 so that the Theorem is valid for $T_n'(N_1,N_2) = (1 - L)^{\frac{1}{2}} T_{1n}^{*}(N_1) + L^{\frac{1}{2}} T_{2n}^{*}(N_2)$. To see that the theorem is true if L is replaced by L_n consider

$$T_n(N_1,N_2) - T_n'(N_1,N_2) =$$

$$(1 - L_n)^{\frac{1}{2}} T_{1n}^*(N_1) + L_n^{\frac{1}{2}} T_{2n}^*(N_2) - (1 - L)^{\frac{1}{2}} T_{1n}^*(N_1) - L^{\frac{1}{2}} T_{2n}^*(N_2)$$

$$= [(1 - L_n)^{\frac{1}{2}} - (1 - L)^{\frac{1}{2}}] T_{1n}^*(N_1) + (L_n^{\frac{1}{2}} - L^{\frac{1}{2}}) T_{2n}^*(N_2). \quad (3.3.11)$$

Since $T_{1n}^*(N_1)$ and $T_{2n}^*(N_2)$ converge in distribution to standard normal random variables, then $T_{1n}^*(N_1)$ and $T_{2n}^*(N_2)$ are $O_p(1)$ (Serfling, 1980, p. 8). Also, since p-lim $L_n = L$ as $n \to \infty$, by Condition I, then $[(1 - L_n)^{\frac{1}{2}} - (1-L)^{\frac{1}{2}}]$ and $(L_n^{\frac{1}{2}} - L^{\frac{1}{2}})$ are $o_p(1)$.

Consequently (3.3.11) shows that

$$T_n(N_1 - N_2) - T_n'(N_1,N_2) = o_p(1) \quad ,$$

thus proving Theorem 3.3.4.

As a consequence of the results obtained in Sections 3.2 and 3.3 it is clear that a distribution-free test of H_o: $\theta=0$ could be based on $T_n(N_1, N_2)$. For small sample sizes exact tests utilizing the conditional distribution of $T_n(N_1,N_2)$ is possible and for large sample sizes the asymptotic normality of $T_n(N_1,N_2)$ can be used. In the next section the use of $T_n(N_1, N_2)$ is illustrated with a worked example.

3.4 A Small-Sample Example

To illustrate the use of $T_n(N_1,N_2)$ developed in Sections 3.2 and 3.3, the data set considered by Woolson and Lachenbruch (1980), which is a slightly modified data set of Holt and Prentice (1974), will be considered in this section. The data set is reproduced below.

TABLE 3.1

DAYS OF SURVIVAL OF SKIN GRAFTS ON BURN PATEINTS

Patient	1	2	3	4	5	6	7	8	9	10	11
Survival of close match ($\exp(X_i)$)	37	19	57+	93	16	22	20	18	63	29	60+
Survival of Poor Match ($\exp(Y_i)$)	29	13	15	26	11	17	26	21	43	15	40

Note. 57+ and 60+ denote right-censored observations at 57 and 60.

The model considered by Woolson and Lachenbruch (1980) is

$$\exp(X_i^{o}) = \phi\, V_{1i}\, W_i$$

$$\exp(Y_i^{o}) = V_{2i}\, W_i \tag{3.4.1}$$

where $\phi > 0$ is an unknown parameter, V_{1i} and V_{2i} are independent and identically distributed nonnegative random variables and W_i is an independent nonnegative random variable for all i. Taking the natural logarithms of (3.4.1) yields

$$X_i^{o} = \theta + \log V_{1i} + \log W_i$$

$$Y_i^{o} = \log V_{2i} + \log W_i \qquad (3.4.2)$$

where $\theta \equiv \log \phi$. It is clear that a test of H_o': $\phi = 1$ is equivalent to a test of H_o: $\theta=0$. Now considering X_i-Y_i the data set is given in Table 3.2.

TABLE 3.2

LOGARITHMS OF THE DIFFERENCE IN SURVIVAL TIME OF THE SKIN GRAFTS

Patient	1	2	3	4	5	6
$X_i - Y_i$	0.2436	0.3795	1.3550+	1.2745	0.3747	0.2578

Patient	7	8	9	10	11
$X_i - Y_i$	-0.2624	-0.1542	0.3819	0.6592	0.4055+

As can be seen from the data, $N_1= 9$, $N_{2L} = 0$, $N_{2R}=2$, and $N_2=2$. It follows by some elementary calculations that

$$T_{1n} = 40, \quad T_{2n} = 2, \quad T_{1n}^{*}(9) = 2.073, \text{ and } \quad T_{2n}^{*}(2) = 1.414.$$

The test statistic $T_n(9,2)$ can be calculated now if L_n is known. Assuming that nothing more is known about the underlying distributions, a reasonable choice for L_n would be $L_n = .5$ and, consequently, $(1 - L_n) = .5$. Note that for this choice equal weight is given to

$T_{1n}^{*}(N_1)$ and $T_{2n}^{*}(N_2)$ in the calculation of $T_n(N_1, N_2)$. Now,

$$T_n(9,12) = (.5)^{\frac{1}{2}}\, T_{1n}^{*}(9) + (.5)^{\frac{1}{2}}\, T_{2n}^{*}(2)$$

$$= 2.466.$$

Comparing this result with the critical values appearing in the tables given in the Appendix it can be seen that

$$P_0\{T_n(9,2) \geq 2.298\} = 0.0093,$$

so that the observed value of 2.466 is significant at the .01 level. As shown in Woolson and Lachenbruch (1980) the W-L test statistic of logistic scores yields the large-sample test statistic $Z = 2.49$, which indicates significance at the .01 level. Also, the exact level for the W-L test statistic based on the permutation distribution of the logistic scores is 11/2048 or 0.00537, again showing significance at the .01 level. Note that the value of 8/2048 as given in Woolson and Lachenbruch (1980) is incorrect due to a minor computational error. In the next section a method to choose L_n will be outlined.

3.5 Choosing the Test Statistic

In section 3.4 an example was presented which illustrated the testing of H_o: $\theta=0$ by using a linear combination of $T_{1n}^{*}(N_1)$ and $T_{2n}^{*}(N_2)$. In this section a method for selecting the linear combination with the maximum conditional Pitman Asymptotic Relative Efficiency (PARE) is outlined. The main result of this section is contained in Theorem 3.5.3 which is proved after establishing some necessary preliminary results.

The following lemma is a restatement of the result appearing in Callaert and Janssen (1978).

Lemma 3.5.1:

Let U_n be a U-statistic of degree r=2 as defined in Lemma 3.3.3 such that

$$\lim_{n\to\infty} \mathrm{Var}_\theta(n^{\frac{1}{2}} U_n) = 4\ \zeta_1, > 0.$$

If $\mu_3 = E_\theta \left| f(X_1, X_2) \right|^3 < \infty$, then there exists a constant M such that for all $n \geq 2$

$$\sup_{-\infty < x < \infty} \left| P\{n^{\frac{1}{2}}(4\zeta_1)^{-\frac{1}{2}} (U_n - E_\theta U_n) \leq x\} - \Phi(x) \right|$$

$$\leq M\ \mu_3\ \zeta_1^{-3} n^{-\frac{1}{2}} \quad .$$

Proof:

The proof is given by Callaert and Janssen (1978). □

Lemma 3.5.2:

Let $\mu_{in}(\theta)$ and $\sigma^2_{in}(\theta)$, i=1,2, be as defined in Lemma 2.4.1.

and

$$\sigma_1{}^2(\theta) = P_\theta[D_{11} + D_{12} > 0,\ D_{11} + D_{13} > 0]$$

$$- \{P_\theta[D_{11} + D_{12} > 0]\}^2$$

$$\sigma_2{}^2(\theta) = [P_2(\theta)]^{-2}[P_{2R}(\theta)]\ [P_{2L}(\theta)] \quad .$$

Suppose that for some $x > 0$

(i) $$\inf_{0 \le \theta \le \delta} \sigma_1^2(\theta) \equiv \sigma_1^2 > 0 ,$$

(ii) $$\inf_{0 \le \theta \le \delta} \sigma_2^2(\theta) \equiv \sigma_2^2 > 0 .$$

Then

$$\lim_{n_1 \to \infty} P_\theta\{\sigma_{in}^{-1}(\theta)[T_{in} - \mu_{in}(\theta) \le x \mid N_1 = n_1, N_2 = n_2\} = \Phi(x),$$

$i=1,2$ uniformly in $\theta \in [0,\delta]$ for every x.

<u>Proof:</u>

The proof follows from Lemma 3.5.1 concerning the rate of convergence to normality of the distribution of a U-statistic.

In Lemma 2.4.2(a) it was shown that

$$n_1^{\frac{1}{2}} \binom{n_1}{2}^{-1}[T_{1n} - \mu_{1n}(\theta)] = n_1^{\frac{1}{2}}[U_{1,n_1} - E_\theta U_{1,n_1}]$$

$$+ n_1^{3/2}\binom{n_1}{2}^{-1}[W_{1,n_1} - E_\theta W_{1,n_1}] \quad (3.5.1)$$

where

$$U_{1,n_1} = \binom{n_1}{2}^{-1} \sum_{1 \le i \le j \le n_1} \Psi(D_{1j} + D_{1j}) ,$$

and

$$W_{1,n_1} = n_1^{-1} \sum_{j=1}^{n_1} \Psi(D_{1j}) .$$

It now will be shown that $[W_{1,n_1} - E_\Theta W_{1,n_1}]$ converges to 0 uniformly in $\Theta \in [0, \delta]$, and that the distribution of $n_1^{\frac{1}{2}}[U_{1,n_1} - E_\Theta U_{1,n_1}]$ converges to a normal distribution uniformly in $\Theta \in [0, \delta]$. In view of equation (3.5.1) this will complete the proof of Lemma 3.5.2, for i=1. □

It is clear that W_{1,n_1} is an unbiased estimator for $P_\Theta[D_{11} > 0]$ and has variance equal to

$$\begin{aligned} Var_\theta(W_{1,n_1}) &= n_1^{-1} P_\Theta[D_{11} > 0] [1 - P_\Theta(D_{11} > 0)] \\ &= n_1^{-1} P_\Theta[D_{11} > 0] P_\Theta[D_{11} \leq 0] \\ &\leq (4n_1)^{-1} \quad . \end{aligned}$$

Therefore, by Chebyshev's inequality (Randles and Wolfe, 1979)

$$\begin{aligned} P_\Theta\{ \left| W_{1,n_1} - E_\Theta W_{1,n_1} \right| > \varepsilon\} &\leq \frac{Var_\Theta W_{1,n_1}}{\varepsilon^2} \\ &\leq \frac{1}{4n_1 \varepsilon^2} \end{aligned}$$

which implies that $[W_{1,n_1} - E_\Theta W_{1,n_1}]$ converges in probability to 0 uniformly in Θ. Since $n_1^{3/2} \binom{n_1}{2}^{-1}$ converges to zero as n_1 tends to infinity, then the second term on the right-hand side of equation (3.5.1) converges to zero uniformly in $\Theta \in [0,\delta]$.

Let $\mu_3 \equiv E_\Theta[\Psi(D_{11} + D_{12})]^3$. Now, from Lemma 3.5.1 it follows that for some positive constant M and $n_1 \geq 2$

$$\sup_{-\infty < x < \infty} \left| P_\Theta\{2^{-1} n_1^{\frac{1}{2}} \sigma_1^{-1}(\theta)[U_{1,n_1} - E_\Theta U_{1,n_1}] \leq x \;\middle|\; N_1 = n_1, N_2 = n_2\} - \Phi(x) \right|$$

$$\leq M\mu_3 \, \sigma_1(\theta)]^{-3} \, n_1^{-\frac{1}{2}} \quad . \tag{3.5.2}$$

Note that $\mu_3 \leq 1$, by definition of the function $\Psi(\cdot)$. Further, by assumption (i) the right-hand side of equation (3.5.2) is less than or equal to $M \, \sigma_1^{-3} n_1^{-\frac{1}{2}}$ for $0 \leq \Theta \leq \delta$. Consequently, the distribution of $n_1[U_{1,n_1} - E_\Theta U_{1,n_1}]$ converges in law to a normal distribution as n_1 tends to infinity uniformly in $\Theta \varepsilon [0,\delta]$, thus proving the Lemma for i=1. The proof for i=2 follows similarly except that assumption (ii) is needed for the application of Lemma 3.5.1 to the U-statistic U_{2,n_2} where U_{2,n_2} is defined as in the proof of Lemma 3.4.1(c). □

The main concern of this section is choosing a test statistic

$$T_{\gamma n} = (1 - \lambda) \, T_{1n}^*(N_1) + \lambda \, T_{2n}^*(N_2) \tag{3.5.3}$$

where $0 < \lambda < 1$. The next Theorem provides a method for selecting λ

such that the corresponding test procedure given by

$$\begin{array}{ll} \text{Reject} & H_o \text{ if } T_{\lambda n} > t , \\ \text{Do not reject } H_o & \text{otherwise,} \end{array} \qquad (3.5.4)$$

where t is a constant, has the maximum PARE relative to all test procedures corresponding to $T_{\lambda n}$.

<u>Theorem 3.5.3</u>:

Let $\mu_{in}(\theta)$ and $\sigma_{in}^2(\Theta)$, i=1,2, be as defined in Lemma 2.4.1.

Suppose the following conditions hold:

(i) For some $\delta > 0$ and $\Theta \in [0,\delta]$,

μ_{in} (θ) is differentiable with respect to Θ and

$$\mu'_{in}(0) = \frac{d}{d\theta} \mu_{in}(\theta)\Big|_{\Theta = 0} > 0, \ i=1,2.$$

(ii) There exist positive constrants k_i, i=1,2,

such that

$$k_i = \lim_{n_i \to\infty} [\mu'_{in}(0)]^{-1} [n_i^{\frac{1}{2}} \sigma_{in}(0)], \ i=1,2.$$

(iii) $\lim\limits_{n_i \to\infty} [\mu'_{in}(0)]^{-1} [\mu'_{in}(\theta_n)] = 1$, i=1,2,

where $\Theta_n = 0(n_i^{-\frac{1}{2}})$.

(iv) $\lim_{n_i \to \infty} \sigma_{in}^{-1}(0)\, \sigma_{in}(\theta_n) = 1$, i=1,2,

where $\theta_n = O(n_i^{-\frac{1}{2}})$.

(v) For some $\delta > 0$, suppose conditions (i) and (ii) of Lemma 3.5.1 hold. Then among all tests for testing H_o: $\theta = 0$ against H_a: $\theta > 0$ having the form (3.5.4) the test corresponding to the test statistic $T_{\gamma n}$ where

$$\gamma = (k_1 + k_2)^{-1} k_1$$

has maximum conditional PARE (as min $(n_1, n_2) \to \infty$) relative to all tests of the form (3.5.4).

Proof:

In view of Theorem 10.2.3 of Serfling (1980) the proof is complete if the Pitman conditions P(1) through P(6) (see Serfling (1980), p. 317 and p. 323) are verified for T_{1n} and T_{2n} conditional on $N_1 = n_1$, $N_2 = n_2$. Conditions P(2) through P(5) follow directly from assumptions (i), (ii), (iii) and (iv), while P(1) follows from assumption (v) and Lemma 3.5.2. To verify condition P(6) note that by Lemma 2.4.2 the conditional asymptotic joint distribtuion of (T_{1n}, T_{2n}) is bivariate normal and from Lemma 3.5.2 it follows that

$$\sup_{0 \leq \theta \leq \delta} \sup_{\substack{-\infty < x < \infty \\ -\infty < y < \infty}} \Big| P_\theta\{\sigma_{1n}^{-1}(\theta)[T_{1n} - \mu_{1n}(\theta)] \leq x,$$

$$\sigma_{2n}^{-1}(\theta)[T_{2n} - \mu_{2n}(\theta)] \leq y \,\Big|\, N_1 = n_1, N_2 = n_2\}$$

$$- \Phi(x)\Phi(y) \Big| \to 0$$

as min $(n_1, n_2) \to \infty$. Thus, the proof of Theorem 3.5.3 follows by taking $\rho = 0$ in Theorem 10.2.3 of Serfling (1980). □

Selection of an optimum test based on Theorem 3.5.3 requires verification of assumptions (i) through (v) and the computation of constants k_1 and k_2. Unfortunately, implementation of this method of selection appears intractable because of the general nature of the functional form of $F_\Theta(\cdot)$ given in equation (2.2.2). It is not clear in general what assumptions concerning $F_\Theta(\cdot)$ can be considered reasonable. However, as is shown below in the special case where X^o and Y^o obey the usual nonparametric analogue of the paired sample model, the form of $F_\Theta(\cdot)$ is considerably simplified.

Suppose X^o and Y^o can be modelled as follows:

$$
\begin{aligned}
X^o &= \Theta + B + E_1 \\
Y^o &= \quad\;\; B + E_2
\end{aligned}
\qquad (3.5.5)
$$

where Θ is a constant, B is a random variable representing the "pairing" variable, and E_i , i=1,2 are i.i.d. according to the distribution function $F(\cdot)$. Furthermore, assume the censoring variable C has distribution function $H(\cdot)$ while C-B has distribution function $G(\cdot)$. Let

$$
\begin{aligned}
P_1(x) &= P[E_1 \leq (C-B) - x,\ E_2 \leq (C-B)] \\
&= \int_{-\infty}^{\infty} F(c - x)\, F(c)\, d\, G(c) \quad .
\end{aligned}
$$

Note that

$$\lambda_1 = P_\Theta\,[X^o \leq C,\ Y^o \leq C]$$

$$= P\,[E_1 \leq (C - B) - \Theta,\ E_2 \leq (C - B)]$$

$$= P_1(\theta)$$

and, as indicated in the proof of Theorem 3.3.4, λ_1, can be assumed to be greater than 0. Therefore,

$$F_\Theta(x) = P_\Theta[X^o - Y^o \leq x \mid X^o \leq C,\ Y^o \leq C]$$

$$= \frac{A\,(x,\Theta)}{P_1(\theta)}$$

where

$$A(x,\ \Theta) = P_\Theta[X^o - Y^o \leq x,\ X^o \leq C,\ Y^o \leq C]$$

$$= P\,[E_1 - E_2 \leq x - \Theta,\ E_1 \leq (C - B) - \Theta,\ E_2 \leq (C - B)]$$

$$= \int_{-\infty}^{\infty} P[E_1 - E_2 \leq x - \Theta,\ E_1 \leq c - \Theta,\ E_2 \leq c]dG(c)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{c} P[E_1 \leq e + x - \Theta,\ E_1 \leq c - \theta]dF(e)dG(c)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{c} F[\min(e + x,\ c) - \theta]dF(e)dG(c).$$

From an application of results such as Theorems A.2.3 and A.2.4 in Randles and Wolfe (1979) it now follows that if $F(\cdot)$ is absolutely continuous and its derivative $f(\cdot)$ is bounded, then both $P_1(\theta)$ and $A(x,\theta)$ are differentiable functions of Θ. Furthermore, $F_\Theta(\cdot)$ is also a differentiable and bounded function of Θ. Therefore, the representation for $\mu_{1n}(\theta)$ and $\sigma^2_{1n}(\theta)$ given in Lemma 2.4.1(b) can possibly be utilized for checking the desired assumptions. For example, assumptions (iii) and (iv) follow directly, but verification of assumptions (i), (ii), and (v) will require more conditions on the specific form of $F(\cdot)$, a problem yet to be solved. Similarly, assumptions (iii) and (iv) are easily verified for T_{2n}, because the representation

$$P_{2R}(\theta) = P_\Theta[Y^o \leq C < X^o]$$

$$= P[E_2 \leq (C - B) \leq E_1 + \theta]$$

$$= \int_{-\infty}^{\infty} [1 - F(c - \Theta)] - F(c)\, dG(c) \quad ,$$

and

$$P_{2L}(\theta) = P_\Theta[X^o \leq C < Y^o]$$

$$= \int_{-\infty}^{\infty} [1 - F(c)] - F(c - \Theta)\, dG(c) \quad ,$$

can be utilized in conjunction with the formulas for $\mu_{2n}(\theta)$, $\sigma^2_{2n}(\theta)$ given in Lemma 2.4.1(c).

CHAPTER FOUR

SIMULATION RESULTS AND CONCLUSIONS

4.1 Introduction and Summary

In the previous chapters a class of test statistics was proposed for the purpose of testing for location difference in censored matched pairs. This class of test statistics is the set of all standardized linear combinations of two conditionally independent test statistics proposed in Chapter Two. It was shown in Chapter Three that under certain assumptions each member of this class is conditionally distribution-free and has an asymptotically normal distribution.

In this chapter a simulation study is presented in order to compare the powers of some members of the proposed class and a test proposed by Woolson and Lachenbruch (1980). In Section 4.2 the statistics that are to be compared will be identified and in Section 4.3 the results of the simulation studies will be listed. Section 4.4 will contain the conclusions that are drawn based on the simulation study.

4.2 Test Statistics To Be Compared

In this section the six test statistics to be compared using a simulation study will be identified. Four of these statistics are members of the class proposed in Chapter Three, another statistic is based on the two statistics proposed in Chapter Two, while the remaining one is a member of the family of statistics proposed by Woolson and Lachenbruch (1980).

At this point some additional notation will be needed. Let $Z_{(i)}$ for $i=1,2,\ldots,N_1$ denote the ordered absolute values of the $D_{1i} \varepsilon C1$ and let $Z_{(0)} = 0$ and $Z_{(N_1+1)} = \infty$. Let m_i, $i=0,1,\ldots,N_1$ denote the number of $|D_{2j}|$, $j=1,2,\ldots,N_2$, where $D_{2j} \varepsilon C2$, contained in the interval $[Z_{(i)}, Z_{(i+1)})$ while p_i, $i=0,1,\ldots,N_1$ represents the number of these D_{2j} which are positive. Also, let

$$M_j = \sum_{k=j}^{N_1} (m_i + 1) \quad .$$

The Woolson and Lachenbruch (W-L) test statistic (1980) utilizing logistic scores can be written as

$$T_{WL} = T_u + T_c$$

where

$$T_u = \sum_{i=1}^{N_1} [2\,\Psi(D_{1i}) - 1][1 - \prod_{j=1}^{i} M_j \,/\, (M_j + 1)]$$

$$T_c = \tfrac{1}{2}(2p_o - m_o) + \sum_{i=1}^{N_1} (2p_i - m_i)[1 - \tfrac{1}{2} \prod_{j=1}^{i} M_j/(M_j+1)] \quad .$$

Under H_o, the variance of T_{WL}, conditional on the observed pattern of censoring, is given by

$$\sigma_{WL}^2 = \sum_{i=1}^{N_1}[1 - \prod_{j=1}^{i} M_j / (M_j + 1)]^2 + \sum_{i=1}^{N_1} m_i[1-\tfrac{1}{2}\prod_{j=1}^{i} M_j / (M_j + 1)]^2$$

$$+ 1/4\ m_o \quad .$$

It should be noted that if there is no censoring, that is if $N_1 = n$ with probability one, then T_{WL} is equivalent to the Wilcoxon signed rank statistic.

The second test statistic studied in Section 4.3 is TMAX where

$$TMAX = MAX\ \{\ T_{1n}^*(N_1),\ T_{2n}^*(N_2)\}\ .$$

It is clear that TMAX utilizes the test statistics T_{1n} and T_{2n} which are proposed in Chapter Two since $T_{1n}^*(N_1)$ and $T_{2n}^*(N_1)$ are the standardized versions of T_{1n} and T_{2n}. Since, according to Lemma 2.4.1(a), T_{1n} and T_{2n} are conditionally independent, an exact conditional test of H_o can be performed using TMAX.

The remaining test statistics are of the form

$$T_n(N_1,\ N_2) = (1 - L_n)^{\frac{1}{2}}\ T_{1n}^*(N_1) + L_n^{\frac{1}{2}}\ T_{2n}^*(N_2)\ . \qquad (4.2.1)$$

as defined below. The first test statistic of this type is

$$T\text{-}EQ = (.5)^{\frac{1}{2}}\ T_{1n}^*(N_1)\ +\ ).5)^{\frac{1}{2}} T_{2n}^*(N_2)$$

corresponding to $L_n = .5$ in (4.2.1). The second is

$$\text{T-SQR} = [N_1 / (N_1 + N_2)]^{\frac{1}{2}} T_{1n}^{*}(N_1) + [N_2 / (N_1 + N_2)]^{\frac{1}{2}} T_{2n}^{*}(N_2)$$

obtained by selecting L_n proportion to $N_2^{\frac{1}{2}}$. The third statistic sets L_n proportional to N_2^2 with the result

$$\text{T-SS} = [N_1 / (N_1^2 + N_2^2)^{\frac{1}{2}}] T_{1n}^{*}(N_1) + [N_2 / (N_1^2 + N_2^2)^{\frac{1}{2}}] T_{2n}^{*}(N_2) \quad .$$

The remaining test statistic to be studied is T-STD where

$$\text{T-STD} = \frac{T_{1n} + T_{2n} - N_1(N_1+1)/4}{\{[N_1(N_1 + 1)(2N_1 + 1)/24] + N_2\}^{\frac{1}{2}}} \, .$$

Note that T-STD is obtained by standardizing $T_{1n} + T_{2n}$ using its mean and standard deviation under H_o: $\Theta=0$.
Simple calculations can show that T-STD has the form (4.2.1) with

$$L_n = \sigma_{N_2}^2 / (\sigma_{N_1}^2 + \sigma_{N_2}^2)$$

where

$$\sigma_{N_1}^2 = N_1(N_1 + 1)(2N_1+1)/24$$

and

$$\sigma_{N_2}^2 = N_2 \, .$$

Therefore, weights given to $T_{1n}^{*}(N_1)$ and $T_{2n}^{*}(N_2)$ by T-STD are proportional to the null variances of T_{1n} and T_{2n}, respectively.

The statistics T-EQ, T-SQR, T-SS, and T-STD are all of the form (4.2.1) and are chosen to represent the class of all test statistics proposed in Chapter Three. In the next section some simulation results are given comparing these statistics among each other and with T_{WL} and TMAX.

4.3 Simulation Results

In this section, following Woolson and Lachenbruch (1980), the assumed model for X_i^o and Y_i^o, i=1,2, . . n, is the log linear model

$$\exp(X_i^o) = \phi V_{1i} W_i$$

$$\exp(Y_i^o) = V_{2i} W_i$$

where $\phi > 0$ is an unknown parameter, V_{1i} and V_{2i} are independent and identically distributed nonnegative random variables, and W_i is an independent nonnegative random variable for all i. The simulation results for each case are developed by generating 500 random samples of 50 observations of $(\log V_{1i}, \log V_{2i})$ and C_i where C_i is an independent observation of some censoring variable. It is clear that if $\theta = \log \phi$ then,

$$X_i^o - Y_i^o = \theta + (\log V_{1i} - \log V_{2i}) \text{ , } i=1,2, \ldots, n;$$

consequently, the distributional form of $X_i^o - Y_i^o$ depends on the distributional form of $\log V_{1i} - \log V_{2i}$.

Samples from four distributional forms for $X_i^o - Y_i^o$ were simulated in this study. The logistic distribution was simulated by generating V_{1i} and V_{2i} with independent exponential distributions. Another light-tailed distribution, the normal distribution, was simulated by generating $\log V_{1i}$ and $\log V_{2i}$ as independent standard normal variables. Two heavy-tailed distributions, the double exponential distribution and Ramberg-Schmeiser-Tukey (RST) Lambda Distribution (Randles and Wolfe, 1979, p. 416) were also studied. The double exponential distribution was arrived at by generating $\log V_{1i}$ and $\log V_{2i}$ with independent exponential distributions. The RST Lambda cumulative distribution function (c.d.f.) can not be expressed explicitly but its inverse c.d.f. can be given as follows:

$$F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2 \qquad 0 < u < 1.$$

As shown in Ramberg and Schmeiser (1972) when $\lambda_1 = 0$, $\lambda_3 = \lambda_4 = -1$ and $\lambda_2 = -3.0674$, the RST Lambda Distribution approximates the Cauchy distribution. This particular choice of RST Lambda Distribution was used to generate $\log V_{1i}$ and $\log V_{2i}$.

To generate the censoring random variable, the Uniform [0, B] distribution was utilized for the logistic and double exponential cases, while the natural logarithm of the Uniform [0, B] distribution was used in the normal and RST Lambda cases. The choice of B was made in each case in such a way so that, under H_o, the proportion of uncensored pairs was approximately 75% of the total sample size while approximately 20% of the total sample size consisted of pairs in which exactly one member of a pair was uncensored. Consequently, approximately 5% of the pairs were not utilized since they were pairs in which both members were censored. Since the results presented in this section only apply to the pattern of censoring described above, any conclusions drawn only apply to this form of censoring.

In each case the hypothesis tested was H_o: $\theta=0$ against H_a: $\theta>0$, at level .05. Based on the asymptotic distribution of the test statistics, the critical value was chosen to be 1.645 in each case except for TMAX which has a critical value of 1.955. Tables 4.1, 4.2, 4.3, and 4.4 give the resulting power (the proportion of times the corresponding test statistic exceeds the critical value) of the six tests described in Section 4.2 for the logistic, normal, double exponential, and RST Lambda distributions, respectively.

TABLE 4.1

POWER OF THE TESTS FOR THE LOGISTIC DISTRIBUTION

	$\theta=0$	$\theta=.181$	$\theta=.453$	$\theta=.907$	$\theta=1.814$
T_{WL}	.050	.186	.532	.944	1.000
TMAX	.040	.128	.414	.900	1.000
T-EQ	.048	.172	.512	.890	.998
T-SQR	.054	.156	.474	.896	1.000
T-SS	.040	.152	.444	.894	1.000
T-STD	.040	.116	.248	.522	.922

TABLE 4.2

POWER OF THE TESTS FOR THE NORMAL DISTRIBUTION

	θ=0	θ=.141	θ=.354	θ=.707	θ=1.414
T_{WL}	.060	.180	.546	.960	1.000
TMAX	.046	.142	.344	.886	1.000
T-EQ	.052	.170	.514	.944	1.000
T-SQR	.054	.184	.514	.948	1.000
T-SS	.052	.156	.466	.936	1.000
T-STD	.050	.150	.418	.876	1.000

TABLE 4.3

POWER OF THE TESTS FOR THE DOUBLE EXPONENTIAL DISTRIBUTION

	θ=0	θ=.141	θ=.354	θ=.707	θ=1.414
T_{WL}	.064	.250	.728	.996	1.000
TMAX	.038	.188	.584	.970	1.000
T-EQ	.064	.262	.746	.998	1.000
T-SQR	.050	.264	.750	.998	1.000
T-SS	.056	.236	.742	.998	1.000
T-STD	.060	.224	.648	.974	1.000

TABLE 4.4

POWER OF THE TESTS FOR AN APPROXIMATE CAUCHY DISTRIBUTION

	θ=0	θ=.2	θ=.4	θ=.6	θ=.8	θ=1.0
T_{WL}	.050	.124	.254	.370	.528	.670
TMAX	.048	.102	.208	.304	.470	.596
T-EQ	.042	.126	.234	.350	.516	.652
T-SQR	.056	.132	.288	.394	.576	.720
T-SS	.060	.124	.272	.386	.594	.728
T-STD	.056	.126	.248	.356	.562	.694

Note that the values of Θ appearing in Tables 4.1, 4.2, and 4.3 are the values 0σ, $.1\sigma$, $.25\sigma$, $.5\sigma$, and 1σ where in Table 4.1 $\sigma^2 = \pi^2/3$, which is the variance of the logistic distribution, while in Tables 4.2 and 4.3 $\sigma^2 = 2$, which is the variance of the underlying normal and double exponential distributions, respectively. In Table 4.4 the value of Θ was chosen such that the corresponding percentiles of the RST Lambda Distribution are approximately equal to the corresponding percentiles determined by the values of Θ used in Tables 4.1, 4.2, and 4.3.

The simulation results displayed in this section show some differences in the powers of the test statistics considered. For example, TMAX appears to be less powerful than the other test statistics near the null hypothesis. However, the results indicate that the choice of critical value for TMAX is conservative since the .05 level of significance is not attained under the null hypothesis. Consequently, if the critical value is decreased until the .05 level of significance is attained, then TMAX cam be expected to perform better than it did in these studies. Nevertheless, all six test statistics are similar in that their powers not only incresse as Θ increases, but for large Θ they all have high powers.

In Table 4.1 the results indicate that for the logistic distribution T_{WL} and T-EQ have similar local (near the null hypothesis) powers with T-SQR and T-SS falling closely behind. Table 4.2 shows that T_{WL}, T-EQ, and T-SQR all have similar local power when the underlying distribution is normal. For the heavier tailed distributions the powers represented in Tables 4.3 and 4.4 show that T-SQR has the highest local power with the powers of T_{WL} and T-EQ only slightly lower.

4.4 Conclusions

The simulation study of Section 4.3 compared the large-sample powers for a test proposed by Woolson and Lachenbruch (1980) and some tests from the class of tests proposed in this dissertation. As noted, in the previous section no noticeable difference in large-sample powers was observed. Therefore, selection of a test procedure for testing H_o: $\Theta = 0$ is not clear based on the results given in Section 4.3. However, this selection should not only take into account the power of the test procedures, but also should be based on the availability of critical values and the level of difficulty involved in the implementation of the test procedures.

As has been noted in this dissertation, critical values for the proposed test statistics, $T_n(N_1,N_2)$, when the sample size is small, are easily obtained. This is not the case for the Woolson and Lachenbruch (1980) statistic, T_{WL}, since the critical values need to be derived from its permutation distribution conditional on the observed scores. This can be time consuming and impractical since the critical values of T_{WL} depend on the observed sample and, thus, can not be tabulated. Tables of critical values for $T_n(N_1,N_2)$ are easily obtained as noted in the Appendix. As for large sample sizes, Theorem 3.3.4 shows that the critical values for $T_n(N_1,N_2)$ can be found by using the standard normal distribution. The conditional distribution of the statistic T_{WL} under H_o is shown by Woolson and Lachenbruch (1980) to have an asymptotic normal distribution as $\min(n_1,n_2) \to \infty$, but it is not clear if this holds unconditionally as the overall sample size, n, tends to infinity.

Computation of the values of the statistics $T_n(N_1,N_2)$ is not difficult since it only involves the calculation of the values of a Wilcoxon signed rank statistic and a binomial statistic. The value of the statistic T_{WL} is more difficult to find and for larger sample sizes is only easily obtainable through the use of a computer. Therefore, for moderate and large sample sizes the lack of computing facilities makes the use of T_{WL} somewhat impractical.

The simulation study in Section 4.3 does not include small-sample power results since these results are not easily obtained. This is due to the difficulty encountered when trying to find critical values for T_{WL}, since these values depend on the particular sample observed. Consequently, the selection of the test procedure can now be made based upon the availability of critical values, the level of difficulty of calculating the corresponding test statistic, and the large-sample power results of the previous section. Based on these criteria T-EQ is the recommended statistic due to its ease of computation, the availability of tables of critical values as in the Appendix, and since the large-sample power results contained in Section 4.3 shows its performance is close to the best of the statistics considered in each case. However, if the form of the underlying distributions is known, then it might be possible to apply the method of Section 3.5 to find the test statistic $T_n(N_1,N_2)$ which maximizes the conditional Pitman Asymptotic Relative Efficiency. In that case, the resulting statistic would be the recommended choice.

APPENDIX

TABLES OF CRITICAL VALUES FOR TESTING FOR DIFFERENCE IN LOCATION

The tables in this appendix list the critical values of

$$T_n(n_1,n_2) = (.5)^{\frac{1}{2}}T_{1n}^*(n_1) + (.5)^{\frac{1}{2}}T_{2n}^*(n_2)$$

at the α equal to .01, .025, .05, and .10 levels of significance for $n_1=1,2,\ldots,10$ and $n_2=1,2,\ldots,15$. The critical values for larger n_1 or n_2 can be approximated by the critical values obtained from the standard normal distribution. When $n_1=0$ the critical values can be obtained from the critical values of the standardized Binomial distribution with parameters n_2 and $p = .5$. Similarly, when $n_2 = 0$ the critical values can be obtained from the critical values of the standardized Wilcoxon signed rank statistic based on n_1 observations. The critical values of this test statistic are derived for each n_1 and n_2 by convoluting the standardized Wilcoxon signed rank distribution based on n_1 observations with the standardized Binomial distribution with n_2 observations. Since both distributions are discrete, exact .01, .025, .05, and .10 level critical values do not always exist, these levels are only approximated. In addition to the critical values, the attained significance level of each critical value (i.e., the probability that $T_n(n_1,n_2)$ is greater than or equal to the critical value) is given in the parentheses.

$\underline{n_1=1}$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	---	---	---
α = .025	---	---	---
α = .05	---	---	---
α = .10	---	---	1.932 (.0625)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	---	---	2.439 (.0078)
α = .025	---	2.288 (.0156)	2.439 (.0078)
α = .05	2.121 (.0313)	2.288 (.0156)	1.862 (.0547)
α = .10	2.121 (.0313)	1.656 (.0938)	1.862 (.0547)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.578 (.0039)	2.707 (.0020)	2.357 (.0098)
α = .025	2.578 (.0039)	2.207 (.0176)	2.357 (.0098)
α = .05	2.043 (.0313)	2.207 (.0176)	1.886 (.0449)
α = .10	2.043 (.0313)	1.293 (.0742)	1.886 (.0449)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.496 (.0054)	2.626 (.0029)	2.340 (.0096)
α = .025	2.049 (.0273)	2.200 (.0164)	2.340 (.0096)
α = .05	2.049 (.0273)	1.773 (.0566)	1.742 (.0366)
α = .10	1.529 (.0864)	1.638 (.0569)	1.334 (.0985)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.472 (.0056)	2.597 (.0032)	2.032 (.0088)
α = .025	1.842 (.0231)	2.219 (.0143)	1.985 (.0296)
α = .05	1.688 (.0668)	1.561 (.0453)	1.666 (.0299)
α = .10	1.450 (.0676)	1.463 (.1064)	1.301 (.0773)

$\underline{n_1=2}$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	---	---	---
α = .025	---	---	---
α = .05	---	---	2.173 (.0313)
α = .10	---	1.949 (.0625)	1.541 (.0625)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	---	2.530 (.0078)	2.681 (.0039)
α = .025	2.363 (.0156)	2.530 (.0078)	2.103 (.0273)
α = .05	1.730 (.0313)	1.897 (.0547)	2.048 (.0313)
α = .10	1.656 (.0938)	1.897 (.0547)	1.526 (.0898)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.820 (.0020)	2.316 (.0098)	2.438 (.0054)
α = .025	2.187 (.0176)	2.316 (.0098)	1.966 (.0273)
α = .05	1.750 (.0586)	1.684 (.0459)	1.656 (.0688)
α = .10	1.555 (.0742)	1.449 (.1006)	1.334 (.0908)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.290 (.0139)	2.029 (.0098)	2.357 (.0056)
α = .025	1.920 (.0166)	2.015 (.0299)	1.949 (.0231)
α = .05	1.658 (.0569)	1.603 (.0380)	1.765 (.0533)
α = .10	1.396 (.1106)	1.382 (.0985)	1.919 (.1159)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.321 (.0120)	2.206 (.0088)	2.325 (.0053)
α = .025	2.081 (.0144)	1.952 (.0243)	1.959 (.0193)
α = .05	1.601 (.0453)	1.705 (.0604)	1.594 (.0535)
α = .10	1.297 (.1088)	1.327 (.1229)	1.327 (.0952)

$\underline{n_1=3}$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	---	---	---
α = .025	---	2.134 (.0313)	2.359 (.0156)
α = .05	---	1.756 (.0625)	1.603 (.0469)
α = .10	1.841 (.0625)	1.378 (.0938)	1.542 (.0938)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.548 (.0078)	2.337 (.0078)	2.488 (.0039)
α = .025	2.170 (.0156)	2.083 (.0273)	1.911 (.0293)
α = .05	1.792 (.0547)	1.705 (.0508)	1.711 (.0625)
α = .10	1.414 (.1016)	1.450 (.0977)	1.333 (.1055)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.249 (.0098)	2.256 (.0093)	2.312 (.0139)
α = .025	2.092 (.0166)	2.000 (.0239)	1.934 (.0254)
α = .05	1.714 (.0459)	1.756 (.0415)	1.650 (.0505)
α = .10	1.401 (.1016)	1.378 (.0908)	1.370 (.1106)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.167 (.0098)	2.248 (.0090)	2.359 (.0120)
α = .025	2.028 (.0299)	1.870 (.0239)	1.981 (.0215)
α = .05	1.650 (.0526)	1.773 (.0541)	1.603 (.0477)
α = .10	1.272 (.1052)	1.396 (.0917)	1.225 (.1028)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.143 (.0090)	2.268 (.0124)	2.386 (.0101)
α = .025	1.779 (.0245)	1.890 (.0250)	2.034 (.0269)
α = .05	1.737 (.0422)	1.512 (.0494)	1.682 (.0485)
α = .10	1.408 (.0125)	1.134 (.0988)	1.304 (.1128)

$n_1=4$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	---	2.291 (.0156)	2.516 (.0078)
α = .025	1.998 (.0313)	2.033 (.0313)	1.999 (.0234)
α = .05	1.740 (.0625)	1.775 (.0469)	1.741 (.0391)
α = .10	1.481 (.0938)	1.291 (.1094)	1.441 (.1016)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.189 (.0117)	2.356 (.0059)	2.248 (.0107)
α = .025	1.998 (.0273)	1.981 (.0293)	1.929 (.0244)
α = .05	1.740 (.0508)	1.723 (.0430)	1.671 (.0527)
α = .10	1.291 (.1055)	1.349 (.1055)	1.352 (.0957)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.369 (.0093)	2.291 (.0120)	2.379 (.0085)
α = .025	1.871 (.0249)	2.016 (.0251)	1.998 (.0256)
α = .05	1.595 (.0498)	1.758 (.0500)	1.695 (.0515)
α = .10	1.318 (.1055)	1.291 (.0994)	1.268 (.1061)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.236 (.0091)	2.357 (.0103)	2.299 (.0078)
α = .025	1.927 (.0278)	2.009 (.0211)	1.933 (.0245)
α = .05	1.720 (.0475)	1.672 (.0504)	1.699 (.0532)
α = .10	1.291 (.1099)	1.312 (.0994)	1.333 (.0991)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.281 (.0081)	2.286 (.0101)	2.222 (.0095)
α = .025	1.889 (.0272)	2.028 (.0252)	1.946 (.0257)
α = .05	1.631 (.0482)	1.669 (.0518)	1.643 (.0502)
α = .10	1.249 (.0987)	1.291 (.1035)	1.322 (.1004)

$\underline{n_1=5}$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	---	2.430 (.0078)	2.274 (.0117)
α = .025	1.947 (.0313)	2.049 (.0234)	1.892 (.0273)
α = .05	1.756 (.0469)	1.667 (.0547)	1.701 (.0508)
α = .10	1.375 (.1094)	1.430 (.0938)	1.320 (.0977)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.272 (.0098)	2.379 (.0098)	2.394 (.0093)
α = .025	2.082 (.0215)	1.997 (.0244)	2.008 (.0283)
α = .05	1.700 (.0488)	1.676 (.0498)	1.631 (.0532)
α = .10	1.319 (.1035)	1.295 (.0996)	1.255 (.0967)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.348 (.0076)	2.286 (.0099)	2.317 (.0090)
α = .025	2.004 (.0254)	1.976 (.0220)	1.947 (.0279)
α = .05	1.660 (.0559)	1.667 (.0530)	1.666 (.0486)
α = .10	1.316 (.0950)	1.333 (.0988)	1.285 (.1013)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.325 (.0104)	2.351 (.0086)	2.274 (.0099)
α = .025	1.943 (.0252)	1.969 (.0240)	1.919 (.0244)
α = .05	1.687 (.0514)	1.643 (.0543)	1.675 (.0520)
α = .10	1.312 (.0940)	1.307 (.1027)	1.293 (.1036)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.242 (.0099)	2.360 (.0095)	2.327 (.0102)
α = .025	2.019 (.0256)	1.992 (.0249)	1.962 (.0262)
α = .05	1.648 (.0501)	1.617 (.0493)	1.613 (.0526)
α = .10	1.267 (.1028)	1.240 (.1002)	1.308 (.0983)

$n_1=6$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	2.264 (.0078)	2.408 (.0078)	2.337 (.0098)
α = .025	1.967 (.0234)	1.964 (.0273)	1.965 (.0254)
α = .05	1.671 (.0547)	1.667 (.0547)	1.668 (.0527)
α = .10	1.374 (.1094)	1.371 (.1016)	1.299 (.1035)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.264 (.0107)	2.357 (.0098)	2.267 (.0107)
α = .025	1.967 (.0254)	1.952 (.0229)	1.970 (.0273)
α = .05	1.671 (.0488)	1.655 (.0498)	1.674 (.0525)
α = .10	1.340 (.1025)	1.319 (.0986)	1.260 (.1021)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.358 (.0090)	2.315 (.0096)	2.290 (.0102)
α = .025	1.945 (.0236)	1.964 (.0261)	1.967 (.0258)
α = .05	1.648 (.0504)	1.667 (.0499)	1.671 (.0518)
α = .10	1.321 (.1010)	1.315 (.1017)	1.321 (.1018)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.305 (.0099)	2.289 (.0101)	2.337 (.0098)
α = .025	2.006 (.0235)	1.975 (.0244)	1.931 (.0245)
α = .05	1.566 (.0503)	1.678 (.0490)	1.632 (.0534)
α = .10	1.265 (.1043)	1.307 (.1043)	1.299 (.1023)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.284 (.0100)	2.313 (.0103)	2.321 (.0096)
α = .025	1.987 (.0244)	1.949 (.0249)	1.956 (.0252)
α = .05	1.648 (.0534)	1.653 (.0500)	1.660 (.0505)
α = .10	1.308 (.1000)	1.289 (.0989)	1.294 (.1011)

$\underline{n_1=7}$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	2.261 (.0078)	2.315 (.0098)	2.300 (.0098)
α = .025	1.902 (.0273)	1.956 (.0273)	1.962 (.0244)
α = .05	1.663 (.0547)	1.673 (.0508)	1.703 (.0508)
α = .10	1.305 (.1171)	1.315 (.1055)	1.344 (.0996)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.261 (.0107)	2.298 (.0095)	2.251 (.0106)
α = .025	2.012 (.0244)	1.940 (.0254)	1.971 (.0258)
α = .05	1.663 (.0518)	1.666 (.0513)	1.653 (.0510)
α = .10	1.305 (.1064)	1.307 (.1011)	1.294 (.1053)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.292 (.0108)	2.315 (.0103)	2.261 (.0105)
α = .025	1.941 (.0250)	1.956 (.0265)	1.909 (.0260)
α = .05	1.695 (.0478)	1.641 (.0501)	1.663 (.0498)
α = .10	1.287 (.0991)	1.315 (.1025)	1.305 (.1033)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.298 (.0103)	2.313 (.0102)	2.300 (.0103)
α = .025	1.970 (.0252)	1.954 (.0257)	1.962 (.0249)
α = .05	1.642 (.0522)	1.647 (.0516)	1.653 (.0526)
α = .10	1.284 (.1056)	1.305 (.1000)	1.295 (.1034)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.295 (.0102)	2.329 (.0096)	2.241 (.0103)
α = .025	1.937 (.0263)	1.971 (.0249)	1.982 (.0254)
α = .05	1.664 (.0503)	1.651 (.0499)	1.637 (.0494)
α = .10	1.305 (.1010)	1.312 (.1000)	1.278 (.0982)

$n_1=8$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	2.192 (.0098)	2.287 (.0098)	2.314 (.0093)
α = .025	1.895 (.0273)	1.990 (.0244)	1.992 (.0239)
α = .05	1.697 (.0488)	1.693 (.0498)	1.695 (.0488)
α = .10	1.400 (.0947)	1.297 (.1064)	1.324 (.0972)
	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.305 (.0098)	2.274 (.0103)	2.326 (.0096)
α = .025	1.994 (.0242)	1.977 (.0240)	1.947 (.0269)
α = .05	1.683 (.0505)	1.642 (.0532)	1.667 (.0497)
α = .10	1.301 (.1040)	1.306 (.1027)	1.287 (.0997)
	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.287 (.0107)	2.287 (.0106)	2.291 (.0100)
α = .025	1.951 (.0253)	1.985 (.0256)	1.971 (.0251)
α = .05	1.653 (.0504)	1.688 (.0498)	1.674 (.0498)
α = .10	1.297 (.1032)	1.297 (.0984)	1.301 (.1023)
	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.284 (.0097)	2.315 (.0097)	2.314 (.0097)
α = .025	1.936 (.0260)	1.957 (.0256)	1.930 (.0251)
α = .05	1.639 (.0510)	1.660 (.0500)	1.633 (.0499)
α = .10	1.294 (.0968)	1.302 (.1004)	1.299 (.1032)
	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.272 (.0100)	2.304 (.0102)	2.299 (.0103)
α = .025	1.971 (.0249)	1.944 (.0259)	1.965 (.0250)
α = .05	1.674 (.0499)	1.647 (.0518)	1.643 (.0498)
α = .10	1.285 (.0990)	1.287 (.1007)	1.303 (.1001)

$n_1=9$

	$n_2=1$	$n_2=2$	$n_2=3$
$\alpha = .01$	2.173 (.0098)	2.298 (.0093)	2.272 (.0110)
$\alpha = .025$	1.922 (.0244)	1.963 (.0254)	1.958 (.0239)
$\alpha = .05$	1.670 (.0508)	1.634 (.0498)	1.685 (.0496)
$\alpha = .10$	1.335 (.1064)	1.298 (.1001)	1.309 (.0972)

	$n_2=4$	$n_2=5$	$n_2=6$
$\alpha = .01$	2.294 (.0103)	2.293 (.0099)	2.286 (.0107)
$\alpha = .025$	1.959 (.0248)	1.958 (.0248)	1.951 (.0262)
$\alpha = .05$	1.670 (.0505)	1.661 (.0508)	1.634 (.0499)
$\alpha = .10$	1.298 (.1012)	1.288 (.1007)	1.289 (.1038)

	$n_2=7$	$n_2=8$	$n_2=9$
$\alpha = .01$	2.300 (.0100)	2.298 (.0101)	2.309 (.0101)
$\alpha = .025$	1.965 (.0248)	1.963 (.0251)	1.953 (.0250)
$\alpha = .05$	1.649 (.0508)	1.634 (.0485)	1.649 (.0496)
$\alpha = .10$	1.294 (.1017)	1.296 (.1012)	1.304 (.0997)

	$n_2=10$	$n_2=11$	$n_2=12$
$\alpha = .01$	2.305 (.0100)	2.288 (.0100)	2.282 (.0103)
$\alpha = .025$	1.943 (.0244)	1.953 (.0242)	1.947 (.0255)
$\alpha = .05$	1.662 (.0499)	1.672 (.0500)	1.634 (.0502)
$\alpha = .10$	1.300 (.0987)	1.283 (.0980)	1.288 (.1017)

	$n_2=13$	$n_2=14$	$n_2=15$
$\alpha = .01$	2.283 (.0100)	2.306 (.0100)	2.302 (.0097)
$\alpha = .025$	1.948 (.0247)	1.932 (.0251)	1.960 (.0254)
$\alpha = .05$	1.639 (.0500)	1.638 (.0500)	1.655 (.0498)
$\alpha = .10$	1.300 (.1008)	1.300 (.1001)	1.290 (.1017)

$\underline{n_1=10}$

	$n_2=1$	$n_2=2$	$n_2=3$
α = .01	2.185 (.0093)	2.261 (.0105)	2.270 (.0107)
α = .025	1.968 (.0210)	1.973 (.0247)	1.958 (.0253)
α = .05	1.680 (.0483)	1.685 (.0503)	1.670 (.0505)
α = .10	1.320 (.1079)	1.324 (.1030)	1.309 (.1030)

	$n_2=4$	$n_2=5$	$n_2=6$
α = .01	2.315 (.0097)	2.282 (.0104)	2.273 (.0099)
α = .025	1.968 (.0244)	1.938 (.0253)	1.982 (.0249)
α = .05	1.666 (.0508)	1.650 (.0500)	1.624 (.0496)
α = .10	1.320 (.0997)	1.289 (.1025)	1.264 (.0997)

	$n_2=7$	$n_2=8$	$n_2=9$
α = .01	2.309 (.0097)	2.324 (.0098)	2.300 (.0103)
α = .025	1.961 (.0244)	1.968 (.0256)	1.968 (.0250)
α = .05	1.643 (.0501)	1.676 (.0498)	1.647 (.0507)
α = .10	1.300 (.1005)	1.320 (.1000)	1.287 (.1011)

	$n_2=10$	$n_2=11$	$n_2=12$
α = .01	2.315 (.0097)	2.321 (.0099)	2.318 (.0097)
α = .025	1.954 (.0250)	1.967 (.0251)	1.957 (.0253)
α = .05	1.651 (.0509)	1.673 (.0500)	1.645 (.0515)
α = .10	1.291 (.1013)	1.312 (.1004)	1.286 (.0993)

	$n_2=13$	$n_2=14$	$n_2=15$
α = .01	2.306 (.0100)	2.306 (.0101)	2.318 (.0100)
α = .025	1.954 (.0256)	1.961 (.0249)	1.958 (.0255)
α = .05	1.642 (.0501)	1.657 (.0501)	1.665 (.0504)
α = .10	1.297 (.1001)	1.296 (.1005)	1.305 (.1001)

BIBLIOGRAPHY

Anscombe, F.J. (1952). Large-Sample Theory of Sequential Estimation. Proc. Cambridge Philos. Soc., 48, 600-607.

Callaert, H. and Janssen, P. (1978). The Berry-Esseen Theorem for U-statistics. The Annals of Statistics, 6, 417-421.

Holt, J.D. and Prentice, R.L. (1974). Survival Analysis in Twin Studies and Matched Pairs Experiments. Biometrika, 61, 17-30.

Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal Likelihoods Based on Cox's Regression and Life Model. Biometrika, 60, 267-278.

Leavitt, S.S. and Olshen, R.A. (1974). The Insurance Claims Adjuster as Patient's Advocate: Quantitative Impact. Report for Insurance Technology Company, Berkeley, California.

Lehmann, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

Miller, R.G. (1981). Survival Analysis. Wiley, New York.

Prentice, R.L. (1978). Linear Rank Tests with Right Censored Data. Biometrika, 65, 167-179.

Ramberg, J.S., and Schmeiser, B.W. (1972). An Approximate Method for Generating Symmetric Random Variables. Communications of the Association for Computing Machinery, Inc., 15, 987-990.

Randles, R.H., and Wolfe, D.A. (1979). Introduction to the Theory of Nonparametric Statistics. Wiley, New York.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

Spoule, R.N. (1974). Asymptotic Properties of U-statistics. Trans. Am. Mathematical Soc., 199, 55-64.

van Eeden, C. (1963). The Relation Between Pitman's Asymptotic Relative Efficiency of Two Tests and the Correlation Coefficient Between Their Test Statistics. The Annals of Mathematical Statistics, 34, 1442-1451.

Woolson, R.F., and Lachenbruch, P.A. (1980). Rank Tests for Censored Matched Pairs. Biometrika, 67, 597-606.

BIOGRAPHICAL SKETCH

Edward Anthony Popovich was born in Detroit, Michigan, on March 2, 1957. He moved to San Diego, California, in 1961 and remained there until he moved to Satellite Beach, Florida, in 1964. After graduating from Satellite High School in 1975, he enrolled at the University of Florida. He received his Bachelor of Science degree in Mathematics in 1977 at which time he received the Four Year Scholar award as valedictorian at his commencement. He entered the Graduate School at the University of Florida in 1978, and later received his Master of Statistics degree in 1979. He expects to receive the degree of Doctor of Philosophy in December, 1983. He is a member of the American Statistical Association.

His professional career has included teaching various courses in both the Mathematics and Statistics Departments at the University of Florida and working two summers for NASA at the Kennedy Space Center in Florida. He has been the recipient of the Wentworth Scholarship, Graduate School fellowships and teaching assistantships during his academic career at the University of Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

P. V. Rao

Pejaver V. Rao, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gerald J. Elfenbein
Associate Professor of Immunology
and Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ronald H. Randles

Ronald H. Randles
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andrew J. Rosalsky

Andrew J. Rosalsky
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Dennis D. Wackerly

Dennis D. Wackerly
Associate Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School, and was accepted for partial fulfillment of the requirements of the degree of Doctor of Philosophy.

December, 1983

Dean for Graduate Studies
and Research