INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeib Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 6HR
SMITH, Theodore MacDonald, 1948-
DISTRIBUTION-FREE TEST FOR THE EQUALITY
OF SEVERAL REGRESSION LINES.

The Ohio State University,
Ph.D., 1977
Statistics
DISTRIBUTION – FREE TEST FOR THE EQUALITY
OF SEVERAL REGRESSION LINES

DISSERTATION

Presented in Partial Fulfillment of the Requirement for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Theodore MacDonald Smith, B.S., M.S.

* * * * *

The Ohio State University
1977

Reading Committee: Approved By
Professor D. A. Wolfe
Professor M. A. Fligner
Professor G. E. Policello II


Department of Statistics
ACKNOWLEDGEMENTS

I want to express my appreciation to my adviser, Professor Douglas Wolfe, for suggesting this problem and for being extremely patient as I struggled to find a solution to the problem. I am deeply indebted to the people of the University Area Fellowship in Columbus, Ohio, and the people of the Jubilee Fellowship in Philadelphia, Pennsylvania, for the spiritual support, financial assistance and emotional support they so freely gave to myself and my family. I want to thank my father and my wife's family for their love and confidence they had continually shown to me throughout this endeavor. Finally I simply want to say thank you to my wife, Gee-Gee, and children, Aaron and Becca, for the many sacrifices they had to make so that I could finish this dissertation.
VITA

May 18, 1948 ....................... Born - Ft. Wayne, Indiana

1970 ......................................... B.S., Rose-Hulman Institute of Technology, Terre Haute, Indiana

1970-1972 ................................ Teaching Assistant, Department of Mathematics, The Ohio State University, Columbus, Ohio

1972 ......................................... M.S., The Ohio State University, Columbus, Ohio

1972-1975 ................................ Teaching Assistant, Department of Statistics, The Ohio State University, Columbus, Ohio

1973 ......................................... M.S., The Ohio State University, Columbus, Ohio

1975-1976 ................................ Consultant, Mershon Center, Dr. Saad Nagi, Program Director, The Ohio State University, Columbus, Ohio

1976-1977 ................................ Instructor, Department of Statistics, Temple University, Philadelphia, Pennsylvania

PUBLICATIONS

"Distribution - Free Tests for Equality of Several Regression Equations." Department of Statistics Technical Report No. 128, The Ohio State University, Columbus, Ohio.
FIELDS OF STUDY

Major Field: Statistics

Studies in Nonparametric Statistics. Professor D. A. Wolfe
Studies in Decision Theory. Professor E. Dudewicz
Studies in Analysis of Variance. Professor J. Singh
Studies in Multivariate Analysis. Professor P. Anderson
Studies in Statistical Inference. Professor D. A. Wolfe
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VITA</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>iii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>viii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>viii</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vi</td>
</tr>
</tbody>
</table>

## Chapter

### I. Introduction
- Literature Survey ........................................ 1
- Statement of Problem .................................... 16
- Motivation .................................................. 21

### II. DISTRIBUTION THEORY UNDER THE NULL HYPOTHESIS
- Introduction ............................................... 27
- Preliminary Results ..................................... 27
- A General Class of Rank-like Statistics ............ 34
- Estimator $\hat{\beta}$ .................................. 38

### III. LIKELIHOOD RATIO STATISTIC
- Introduction ............................................... 41
- Preliminary Notation .................................... 41
- Equivalent Form of $P_v$ ................................ 47

### IV. CONSISTENCY
- Introduction ............................................... 54
- Preliminary Definition and Results for $k = 2$ .... 54
- Consistency of Mann-Whitney-Wilcoxon ............... 58
- Consistency of Jonckheere Test. ........................ 67
- Consistency of $\hat{\beta}_n$ .............................. 71

### V. SMALL SAMPLE POWER STUDY
- Introduction ............................................... 77
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of Simulation Technique</td>
<td>78</td>
</tr>
<tr>
<td>Results</td>
<td>83</td>
</tr>
<tr>
<td>Discussion of Results</td>
<td>85</td>
</tr>
<tr>
<td>VI.</td>
<td>98</td>
</tr>
<tr>
<td>Summary</td>
<td>98</td>
</tr>
<tr>
<td>Extensions</td>
<td>99</td>
</tr>
<tr>
<td>New Research</td>
<td>101</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>103</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLE 1  Prob (type I error) verification ........................................ 86
TABLE 2  Power Curve \( n_1 = n_2 = 8 \) ................................. 87
Significance level = .05
TABLE 3  Power Curve \( n_1 = n_2 = 8 \) ................................. 89
Significance level = .01
TABLE 4  Power Comparison ................................................ 90
TABLE 5  Effect of the value of the intercept ...................... 91
TABLE 6  One-sided versus Two-sided tests ....................... 92
LIST OF FIGURES

FIGURE 1 ................................................................. 23
FIGURE 2 ................................................................. 25
Chapter One

1.1 Introduction

In terms of applications, one of the most widely used statistical models is the one-sample multilinear regression model:

\[(1.1) \quad Y_j = \alpha + \beta_1 x_{1j} + \ldots + \beta_p x_{pj} + Z_j; \quad j = 1, 2, \ldots, n\]

where \(Z_j\) is called the error term. In this model, the problems of interest are the classical problems of estimation of the parameters \(\alpha, \beta_1, \ldots, \beta_p\); tests of hypotheses that various subsets of the parameters are equal zero (or some other prespecified value); and construction of confidence intervals for the parameters. A closely related model is the k-sample multilinear regression model:

\[(1.2) \quad Y_{ij} = \alpha_i + \beta_{i1} x_{i1j} + \ldots + \beta_{ip} x_{i pj} + Z_{ij}; \quad 1 \leq i \leq k, 1 \leq j \leq n_i,\]

where \(Z_{ij}\) is called the error term. In this model, the main problems of interest are tests of hypotheses that various subsets of parameters are equal across the k samples. Two specific hypotheses are the following: for each \(k, 1 \leq k \leq p\)

\[(1.3) \quad H_0: \quad \beta_{i k} = \beta_k \text{ (unknown)} \quad \text{for all } i, \quad 1 \leq i \leq \frac{k}{p}\]
Expression (1.4) represents the general setting of the analysis of co-variance problem and (1.3) represents the general setting of the equality of regression problem. Solutions to all of these problems have been obtained under a variety of assumptions on the error term. It is the goal of this dissertation to present a class of distribution-free tests for the hypothesis (1.3) for the case $p = 1$ (that is, for the case of linear regression) using a general $k$-sample linear rank statistic based on the ranks of certain functions of the observations. Not only is each test statistic distribution-free, but there is also a subclass of statistics whose null distributions are equivalent to certain well-known (and well tabulated) two-sample (and $k$-sample) rank statistics. Now, in order to motivate both the assumptions placed on the model (1.2) and the proposed solution; a brief survey of other approaches to the problems associated with models (1.1) and (1.2) is presented first.

1.2 Literature Survey

For both models (1.1) and (1.2), where $p$ is arbitrary, the most restrictive assumptions usually placed on the error term are the following:
(A1) the error terms are independent, identically distributed random variables;

(B1) the distribution of each error term is normal with mean zero and variance \( \sigma^2 \); and

(C1) \( \{X_{j_k}: 1 \leq j \leq n; 1 \leq k \leq p \} \) (or in (1.2) \( \{X_{i,j_k}: 1 \leq i \leq k, 1 \leq j \leq n_1, 1 \leq k \leq p \}) \) are fixed known constants.

Under these assumptions, the method of least squares provides solutions to all the problems associated with both model (1.1) and (1.2). For model (1.1), the least squares estimators (LSE) are obtained either by minimizing the error sum of squares

\[
\sum_{j=1}^{n} (Y_j - \alpha - \sum_{m=1}^{p} \beta_m X_{jm})^2 = S(\alpha, \beta_1, \ldots, \beta_p)
\]

or by solving the \((p+1)\) normal equations

\[
\sum_{j=1}^{n} (Y_j - \alpha - \sum_{m=1}^{p} \beta_m X_{jm}) X_{j_k} = 0, \quad 1 \leq k \leq p
\]

(1.6) and

\[
\sum_{j=1}^{n} (Y_j - \alpha - \sum_{m=1}^{p} \beta_m X_{jm}) = 0.
\]

Since the LSE for each parameter has a normal distribution, an exact confidence interval for each parameter can be constructed. The test of hypothesis that a certain subset of the parameters equal zero is based on the ratio \((SSH-SSE)/SSE\), where
SSE = S(\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_p), \hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_p \text{ are the } LSE \text{ of } \alpha, \beta_1, \ldots, \beta_p, \text{ resp.},

and SSH is the sum of squares that is obtained when the null hypothesis \( H \) is true. The distribution of the test statistic (sometimes called the variance ratio test statistic) is a \( F \)-distribution. For the model (1.2) similar techniques can be used by considering the \( k \)-samples, each with \((p+1)\) parameters, as one combined sample with \( k(p+1) \) parameters (see Rao (1965), sec. 4g.4). Then a variance ratio type statistic, which also has a \( F \)-distribution, can be derived to test the hypothesis (1.3) and (and (1.4)).

Now, replace assumption (B1) with a less restrictive assumption:

(B2) the distribution of the error term is \( F(z) \) where \( F(*) \) is some known distribution function with finite variance.

Then under (A1), (B2), and (C1), the maximum likelihood technique is usually used to obtain estimators (MLE) of \( \alpha, \beta_1, \ldots, \beta_p \). The test of hypothesis that a certain subset of parameters in (1.1) equal zero or the test of (1.3) (or (1.4)) can be made with the likelihood ratio test (LRT). In general, only the asymptotic distributions are known for the MLE's and LRT's (Wald (1943)). Thus the effect of replacing (B1) with (B2) is that of replacing estimators and test statistics having known exact distributions with estimators and test statistics having only known asymptotic distributions. For the model (1.1) certain optimal properties of MLE and LRT have been proved by Wald (1943). For (1.2) the similar properties are shown to be still true by Sen (1969). Finally only approximate confidence intervals for
\( a, \beta_1, \ldots, \beta_p \) can be constructed based on the asymptotic distribution of the MLE's of \( a, \beta_1, \ldots, \beta_p \).

Now replace (B2) with the assumption:

(B3) The distribution of the error term is \( F(z) \)
where \( F(\cdot) \) is an absolutely continuous distribution function with unknown functional form and density function \( f(z) \).

A number of different solutions to problems in models (1.1) and (1.2) have been proposed under the assumptions (A1), (B3), and (C1). For model (1.1), with \( p = 1 \), one of the first solutions was proposed by Thiel (1950). He suggests using a test statistic based on Kendall's tau to test the hypothesis \( H: \beta = \beta_0 \) (known). The computation of the statistic uses the sequence

\[ \{(x_j, y_j - \beta_0 x_j), 1 \leq j \leq n\} \]

The null distribution of the statistic is the same as that of Kendall's tau; hence it is distribution-free under the null hypothesis. An exact confidence interval for \( \beta \) can be constructed based on the test statistic. Thiel also proposed as an estimator of \( \beta \),

\[ \beta^* = \text{median} \left\{ \frac{y_j - y_i}{x_j - x_i}, i \neq j, 1 \leq i < j \leq n \right\} \]

(Thiel assumed that \( x_i \neq x_j \) for \( i \neq j \). Sen (1968) relaxed this assumption and only used those pairs, \( (y_i, x_i) \) and \( (y_j, x_j) \), where \( x_i \neq x_j \).)
Again for $p = 1$, Hájek (1962) proposed a test of $H: \beta = 0$ based on a class of linear rank statistics. This test can be viewed as a generalization of the work done by Hoeffding (1950) and Terry (1951). Each test statistic has the form

$$T = \sum_{i=1}^{n} (x_i - \bar{x}) a_n(R_i), \quad \bar{x} = n^{-1} \sum_{i=1}^{n} x_i,$$

where $R_i$ is the rank of $Y_i$ among $Y_1, \ldots, Y_n$, where either

$$a_n(i) = E(\phi(U_n^{(i)}))$$

in which $U_n^{(i)}$ is the $i^{th}$ order statistic from a random sample of size $n$ from an uniform $[0,1]$ distribution and $\phi(u)$ is some particular square integrable score-generating function; or

$$a_n(i) = \phi(i/(n+1)).$$

The test based on $T$ is asymptotically locally most powerful for a given density function, $f(x)$, when

$$\phi(u) = f'(F^{-1}(u))/f(F^{-1}(u))$$

where $f'(x)$ is the derivative of $f(x)$, and $F^{-1}(u)$ is the inverse of the distribution function $F(u)$. (See Hájek and Šidák (1967), sec I.2.4 and III 3.1).

Adichie (1967a) proposed a class of rank test for $H: \alpha = \beta = 0$. Each test was based upon

$$T_1 = n^{-1} \sum_{j=1}^{n} \psi_n(R_j/(n+1)) \text{Sign} Y_j$$
\begin{align}
T_2 &= n^{-1} \sum_{j=1}^{n} x_j \psi_n(R_j/(n+1)) \text{ Sign } Y_j
\end{align}

where

\begin{align}
\psi_n(u) &= -g(G^{-1}((u/2)+(1/2))/g(G^{-1}((u/2)+(1/2))), 0 \leq u \leq 1
\end{align}

where $G^{-1}(u)$ is the inverse of $G(u)$, $G(u)$ is any known absolutely continuous symmetric distribution with density $g(u)$, $g'(u)$ is the derivative of $g(u)$, and finally $R_j$ equals the rank of $|Y_j|$ among $|Y_1|, \ldots, |Y_n|$. The Pitman efficiency of the proposed tests relative to the variance ratio test is shown to be the same as the efficiency of the corresponding rank score tests relative to the $t$-test in the two-sample location problem.

Extending the results of Hodges and Lehmann (1963) Adichie (1967b) proposed a class of estimators for $\alpha$ and $\beta$ based on (1.7) and (1.11). Instead of ranking $Y_1, \ldots, Y_n$, the idea is to rank $Y_1-a-bx_1, \ldots, Y_n-a-bx_n$ for each $a$ and $b$. Then the estimators are $\hat{\alpha}$ and $\hat{\beta}$ where

$$
\hat{\beta} = \sup \{ b : T(y-a-bx) > 0 \text{ for all } a \}
$$

$$
\hat{\beta} = \inf \{ b : T(y-a-bx) < 0 \text{ for all } a \}
$$

where $T$ is defined as in (1.7) and

\begin{align}
\hat{\alpha} &= (\hat{\beta} + \hat{\beta})/2;
\end{align}

$$
\alpha^* = \sup \{ a : T_1(y-a-\hat{\alpha} x) > 0 \}
$$

$$
\alpha^* = \inf \{ a : T_1(y-a-\hat{\alpha} x) < 0 \}
$$
where $T_1$ is defined as in (1.11) and

$$(1.16) \hat{\alpha} = (\alpha^* + \alpha^**) / 2.$$  

The joint distribution of $\hat{\alpha}$ and $\hat{\beta}$ is shown to be symmetric about the parameter point $(\alpha, \beta)$ and asymptotically multivariate normal. 

For the case of multilinear regression where $p$ is arbitrary, Jaeckel (1972) and Jurečková (1971) have each proposed a solution to the estimation problem. Both solutions can be viewed as a modification of the least-squares technique. Jaeckel modified the expression (1.5) by looking at the expression

$$(1.17) \sum_{j=1}^{n} \{a_n(R_j)(Y_j - \xi_1^p \beta_j x_{j1})\}$$

where $R_j$ is the rank of $(Y_j - \xi_1^p \beta_j x_{jm})$ among $(Y_1 - \xi_1^p \beta_1 x_{1m}), \ldots, (Y_n - \xi_1^p \beta_n x_{nm})$ and $a_n(i)$ is a sequence of a non-decreasing set of scores satisfying $\sum_{k=1}^{n} a_n(k) = 0$. Expression (1.17) can be viewed as a measure of dispersion of the residuals which is an alternative to the error sum of squares in (1.5). Jaeckel takes as his estimators of $\beta_1, \ldots, \beta_p$ those values which minimize (1.17). He treats $\alpha$ as a nuisance parameter and is not able to estimate it by his method. 

Jurečková proposed an analog to solving the normal equations in (1.6) by looking at

$$(1.18) n^{-1/2} \sum_{k=1}^{p} \xi_j a_n(R_j) \frac{a_n(R_j)}{\sum_{R_j}} = \xi_j S_{n\xi}$$
where $R_j$ is the rank of $(Y_j - \sum_{m=1}^{p} \beta_m x_{jm})$ among $(Y_j - \sum_{m=1}^{p} \beta_m x_{jm}), \ldots,$

$(Y_n - \sum_{m=1}^{p} \beta_m x_{nm})$; $a_n(i)$ is defined as in (1.8) or (1.9) (except that

$\phi(u)$ also is assumed to be non-decreasing); and $\bar{x}_j = \frac{1}{n} \sum_{j=1}^{n} x_{j}.$

Jurečková takes as her estimators of $\beta_1, \ldots, \beta_p$ those values which minimize (1.18). She is also able to obtain an estimate of the parameter $\alpha$. By a suitable choice of score functions, she shows that her estimators have the same limiting distribution as the MLE for a given density function, $f(x)$. Jaeckel showed that the partial derivative of (1.17) with respect to $\beta_k$ was $s_{nk}$, and that his estimators were asymptotically equivalent to Jurečková's estimators.

Hettmansperger and McKean (1976) proposed tests of hypotheses that certain subsets of parameters of the general linear model equal zero based on Jaeckel's measure of dispersion of residuals (1.17). The test statistic is based on the reduction in minimum dispersion due to the hypothesis and is shown to be asymptotically distribution-free. Since model (1.2) can be viewed as a general linear model, these proposed test statistics could be used to test the hypotheses (1.3) and (1.4).

Aside from Hettmansperger and McKean's work, the only other solutions for testing the hypothesis (1.3) under the general assumptions of (A1), (B3), and (C1) have been derived for the linear ($p=1$) regression case only. For $k=2$, Hollander (1970) proposed a distribution-free test of (1.3) based on the Wilcoxon signed-rank test. Assuming $n_1 = n_2 = 2n$, slope estimators of the
form

\[ U_{ij} = \frac{(Y_{i(j+n)} - Y_{ij})}{(x_{i(j+n)} - x_{ij})} \]

for \(1 \leq j \leq n, i = 1, 2\) are computed. Each estimator from the first sample is then randomly paired with exactly one estimator from the second sample. The differences of the from \(Z = U_{ij} - U_{2j}\) are used as the "observations" with the signed rank test statistic. Potthoff (1974) proposed a conservative test of (1.3) which neither required equal sample sizes nor involved an irrelevant randomization. Using all possible slope estimators of the form

\[ U_{jj'} = \frac{(Y_{ij} - Y_{ij'})}{(x_{ij} - x_{ij'})} \]

where \(i = 1, 2\), and \(1 \leq j, j' \leq n_i\); the test statistic is based on all possible differences between an estimator from the first sample and one from the second sample. The test statistic is computed using an upper bound for the variance, an upper bound which does not depend on the true underlying distribution.

For an arbitrary \(k\), both Sen (1969) and Adichie (1974) have each proposed a class of rank-score tests for (1.3), with \(p=1\), which are asymptotically distribution-free. Both classes of test statistics are based on the following definitions. For \(1 \leq i \leq k\), let

\[ \lambda_i = \frac{C_i^2}{C^2} \]

and also

\[ (1.19) \quad \bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}, C_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, C^2 = \sum_{i=1}^{k} C_i^2, \]

Let \(\phi(u)\) be an absolutely continuous non-decreasing function of \(u\), \(0 \leq u \leq 1\), such that \(\int_0^1 \phi^2(u) du < \infty\). Let \(a_n(i)\) be defined either as in (1.8) or (1.9) and let
(1.20) \[ A^2 = \int_0^1 \phi^2(u)du - [\int_0^1 \phi(u)du]^2. \]

Then for each value of \( b \), Sen defined for \( 1 \leq i \leq k \)

(1.21) \[ T_i(b) = \left[ \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) a_{n_i} (R_{ij}) \right] / A \cdot C_i, \]

where \( R_{ij} \) is the within sample rank of \( Y_{ij} - b \ x_{ij} \) among

\( Y_{i1} - b \ x_{i1}, \ldots, Y_{in_i} - b \ x_{in_i} \).

Using

(1.22) \[ T(b) = \sum_{i=1}^{k} C_i T_i(b) / C, \]

a pooled sample estimate of \( \beta \), the common unknown value of

\( \beta_1, \beta_2, \ldots, \beta_k \), under (1.3), is found in the same manner as Adichie (1967b) did in the one-sample case. The test statistic for (1.3) is

(1.23) \[ L = \sum_{i=1}^{k} T_i^2(\hat{\beta}) \]

where \( \hat{\beta} \) is the estimator of \( \beta \) based on (1.22) and \( T_i(b) \) is defined in (1.21). Sen proved that the null distribution is asymptotically chi-square with \( (k-1) \) degrees of freedom. He proved that the asymptotic relative efficiency of (1.23) relative to the variance ratio test (based on the least-squares technique) is the same as that of the corresponding rank test relative to the t-test for the two-sample location problem, namely

(1.24) \[ \frac{\sigma^2_F}{\sigma^2_F} \left\{ \int [d(\phi(F(z))/dz) dF(z)]^2 \right\} / A^2, \]

where \( \sigma^2_F \) is the variance of the underlying distribution \( F(z) \).
The asymptotic relative efficiency of (1.23) (with score generating function \( \phi(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \)) relative to the log likelihood ratio test is

\[
(1.25) \quad \left\{ \int \left[ \frac{d(\phi(F(z))/dz)}{dF(z)} \right] dz \right\}^2 / \Lambda^2 \text{ I}(F)
\]

where \( \text{I}(F) \) is the Fisher information of \( F(z) \). In the same article, Sen also proposed a test of (1.3) for the more general model where assumption (B3) is replaced by

(B3') the error term \( Z_{ij} \) has for its distribution function \( F_i(z) \), where each \( F_i(z) \) is absolutely continuous with finite Fisher information.

Under assumptions (A1), (B3), (C1) and the restriction that \( \alpha_1 = \alpha_2 = \ldots = \alpha_k = \alpha \) (unknown), Adichie proposed a test of (1.3) based on the combined sample rankings of all \( k \) samples simultaneously. He defined for each \( i, 1 \leq i \leq k \)

\[
C_{ij}^{(i)} = \begin{cases} 
\lambda_i (x_{sj} - \bar{x}_s), & s = 1, \ldots, i-1, i+1, \ldots, k \\
(\lambda_i - 1) (x_{sj} - \bar{x}_s), & s = i.
\end{cases}
\]

Similar to Sen, a pooled sample estimator of \( \beta \), the common value under (1.3), is made based on the statistic

\[
(1.26) \quad S(\beta) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) \alpha_n (R_{ij})
\]

where, now, \( R_{ij} \) equals the rank of \( Y_{ij} - b x_{ij} \) in the combined ranking of all \( n = \sum_{i=1}^{k} n_i \) observations, and \( \alpha_n (i) \) is also defined as in (1.8) or (1.9). For \( 1 \leq i \leq k \), define
(1.27) \( T_i(\hat{\beta}) = \sum_{s=1}^{k} \sum_{j=1}^{n_s} c_{sj}^{(i)} n_s(\hat{R}_{sj}) \)

where \( \hat{\beta} \) is the estimator of \( \beta \) based on (1.26) and \( \hat{R}_{sj} \) equals the rank of \( Y_{sj} - \hat{\beta} x_{sj} \) in the combined sample. The test statistic is

(1.28) \( L = \sum_{i=1}^{k} \left( \frac{T_i(\hat{\beta})}{AC_i} \right)^2 \)

The null distribution of (1.28) is asymptotically chi-square with

\((k-1)\) degrees of freedom. The asymptotic relative efficiency results are the same as those of Sen (1969).

Adichie (1975a) dropped the restriction of equal intercept parameters and instead provided a class of rank score tests which

are asymptotically distribution-free for the hypothesis

(1.29) \( H: \alpha_i = \alpha \) (unknown), \( \beta_i = \beta \) (unknown)

requiring only a pooled sample estimator for \( \beta \). However the procedure is limited to designs where the group means \( \bar{x}_i \) are equal.

Define for each \( i, 1 \leq i \leq k \) and \( j = 1, ..., n_i \),

\[
d_{ij}^{(i)} = 0 \text{ if } s \neq i \]
\[
= 1 \text{ if } s = i.
\]

Use (1.26) to obtain a pooled estimate of \( \beta \), except let \( R_{ij} \) equal

the rank of \( Y_{ij} - b(x_{ij} - \bar{x}_i) \). For each \( i, 1 \leq i \leq k \), let

(1.30) \( T_{\alpha_i}(\hat{\beta}) = \sum_{s=1}^{k} \sum_{j=1}^{n_s} \left( d_{sj}^{(i)} - \bar{d}^{(i)} \right) n_s(\hat{R}_{sj}) \)

(1.31) \( T_{\beta_i}(\hat{\beta}) = \sum_{s=1}^{k} \sum_{j=1}^{n_s} c_{sj}^{(i)} n_s(\hat{R}_{sj}) \)

where again \( \hat{R}_{sj} \) is the rank of \( Y_{ij} - \hat{\beta}(x_{ij} - \bar{x}_i) \) in the combined ranking.
of all \( n \) residuals, \( \hat{\beta} \) is the pooled estimator based on (1.26),

\( a_n(i) \) is defined as in (1.8) or (1.9), and \( d_{sj}^{(i)}, c_{sj}^{(i)} \) are defined as before. Also for each \( i \), let

\[
v_i = n^{-1/2} \left( T_{\alpha}^{(i)}(\hat{\beta})/A \right), \quad u_i = \left( T_{\beta}^{(i)}(\hat{\beta})/C_{\beta}^{(i)} \right)
\]

Then, the test statistic for (1.29) is

\[
M = \sum_{i=1}^{k} (v_i^2 + u_i^2)
\]

The null distribution is asymptotically chi-square with \( 2k-2 \) degrees of freedom. The asymptotic relative efficiency results are also the same as in Sen (1969).

Tests for (1.3) against ordered alternatives were developed by Adichie (1976b). He developed both a likelihood ratio test (LRT) and a test that depends on a suitable linear combination of one group statistic as well as two rank analogs of the above two tests. The paper represents an application of the general theory of ordered alternatives to the regression problem. The rank analog of the LRT is

\[
\overline{x}_k^2(\phi) = \sum_{i=1}^{k} C_i^2 \left( \overline{T}_{ni} - \overline{T}_n \right)^2/A^2
\]

where \( \overline{T}_n = \sum \lambda_i \overline{T}_{ni}, \overline{T}_{ni} = A.T_i(\beta), \overline{T}_i, A_i, \lambda_i, \) and \( C_i^2 \) are defined in (1.21), \( \overline{T}_{ni} \) is the isotonic regression of \( \overline{T}_{ni} \), \( \phi(u) \) is a score-generating function which is expressible as a sum of two monotone square integrable functions, and the scores are defined as in (1.9). The statistic \( \overline{x}_k^2(\phi) \) has the same limiting distribution as the corresponding likelihood ratio rest used in the analysis of variance.
against ordered alternatives. The asymptotic relative efficiencies of $\frac{\chi^2_k}{k} (\phi(u))$ relative to the LRT is the same as the two-sample location test relative to the $t$-test.

Finally, as a note of interest, four papers are reviewed that deal with the testing of (1.4), the analysis of covariance. In Puri and Sen (1969) (which is a generalization of Quade (1967)) the independent (or concomitant) variable $X = (X_1, \ldots, X_p)$ is assumed to be a random variable whose (joint) distribution is identical across the $k$-samples. The conditional distribution of the dependent variable $Y$ given the concomitant variable $X$ has the form,

$$F_i(y|X=x) = F(y-a_i| X=x), \quad i = 1, \ldots, k.$$  

Then (1.4) is equivalent to the hypothesis of equal conditional distributions across the $k$-samples. The test statistic is based upon both ranks of the dependent variable and of the concomitant variables, and is conditionally permutationally distribution-free.

In 1972, Sen proposed a class of asymptotically distribution-free test statistics for (1.4) under the assumption that the independent variables are non-random treating the slope parameters, $\beta_i$ (which depend on the treatment), as nuisance parameters. The statistics are weighted combinations of several one-sample rank tests. Adichie (1975b) proposed a class of rank tests for (1.4) under the assumption of a common slope parameter, $\beta$. Each statistic is based upon the simultaneous ranking of all observations. In both Sen and Adichie, the asymptotic properties are the same as in Sen (1969).
1.3. Statement of Problem

In order to get distribution-free tests, the problem has been formulated in the following manner:

(D1) For $i = 1, \ldots, k$, let $\{(Y_{ij}, X_{ij}), 1 \leq j \leq n_i\}$ be $k$ independent bivariate random samples, where the $i^{th}$ sample is from a distribution which is absolutely continuous with respect to Lebesgue measure on the plane. Let $F_i(y, x)$ and $f_i(y, x)$ be the joint distribution function and joint density function, respectively. Let $\sum_{i=1}^{k} n_i = n$.

(D2) For each $i$, $1 \leq i \leq k$, assume that

$$f_i(y, x) - f(y - \alpha_i - \beta_i x, x)$$

where $f(y, x)$ is an unknown density function.

In addition to the joint distribution being absolutely continuous, it is also assumed that

(1.34) Marginal distribution of both $X$ and $Y$ is absolutely continuous with respect to Lebesgue measure.

(1.35) $m(Y|X = 0) = 0$

where $m(Y|X = x)$ is the median of the conditional distribution of $Y|X = x$,
where \((Y,X)\) have joint density function \(f(y,x)\).

The null hypothesis \(H\) and the sequence of alternative hypotheses \(K\) that will be considered are:

\[
(1.36) \quad H: \quad \beta_1 = \ldots = \beta_k = \beta \text{ (unknown)}
\]

\[
(1.37) \quad K: \quad \beta_i = \beta + \theta_i \quad \text{where } |\theta_i| < M
\]

for \(1 \leq i \leq k\).

Let \(\hat{\beta}_n = \hat{\beta}_n((Y_{11},X_{11}),\ldots,(Y_{1n},X_{1n}),\ldots,(Y_{k1},X_{k1}),\ldots,(Y_{kn},X_{kn}))\) be any combined sample estimator of the common value which has the following properties:

\[
(1.38) \quad \text{the value of } \hat{\beta}_n \text{ is invariant under any}
\]

permutation of the indices in the combined sample.

\[
(1.39) \quad \hat{\beta}_n \overset{D}{=} \beta \text{ under both } K \text{ and } H.
\]

The class of test statistics is defined in the following manner: Let \(\phi(u)\) be a non-decreasing score-generating function such that

\[
(1.40) \quad A^2 = \int_0^1 \phi^2(u) \, du - [\int_0^1 \phi(u) \, du]^2 < \infty
\]

Let \(a_n(i)\) be the scores, where either

\[
(1.41) \quad a_n(i) = E(\phi(U_n^{(i)}))
\]

where \(U_n^{(i)}\) is the \(i^{th}\) order statistic from a random sample of size
n from an uniform \([0,1]\) distribution, or,

\[
(1.42) \quad a_n(i) = \phi(i/(n+1)).
\]

Let \(\hat{R}_{ij}\) be the rank of the signed residuals \((Y_{ij} - \hat{\beta}_i X_{ij})\text{sgn}(X_{ij})\) among the combined sample \(\{(Y_{ij} - \hat{\beta}_i X_{ij})\text{sgn}(X_{ij}), 1 \leq i \leq k, 1 \leq j \leq n_i\}\), where

\[
\text{sgn}(u) = \begin{cases} 
1 & \text{if } u \geq 0 \\
-1 & \text{if } u < 0
\end{cases}
\]

Then for \(k=2\), define \(c_{1j} = 1, 1 \leq j \leq n_1, c_{2j} = 0, 1 \leq j \leq n_2\), and

\[
(1.43) \quad \hat{S}_n = \sum_{i=1}^{n_1} \sum_{j=1}^{n_i} c_{ij} a_n(\hat{R}_{ij}) = \sum_{j=1}^{n_1} a_n(\hat{R}_{1j}).
\]

For \(k > 2\), the statistic will depend on the alternative of interest. If the alternative is \(\beta_i \neq \beta_j\) for some \(i, j\) then define

\[
(1.44a) \quad \hat{S}_n = 12n^{-1}(n + 1)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\sum_{j=1}^{n_i} \hat{R}_{ij})^2 - 3(n + 1).
\]

If the alternative of interest is \(\beta_1 \leq \beta_2 \leq \ldots \leq \beta_k\) with at least one strict inequality, define for \(1 \leq i < i' \leq k\),

\[
S_{ii'} = \text{number of pairs } (Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}) \text{ where }
\]

\[
(Y_{ij} - \hat{\beta}_n X_{ij})\text{sgn}(X_{ij}) < (Y_{i'j'} - \hat{\beta}_n X_{i'j'})\text{sgn}(X_{i'j'}),
\]

and

\[
(1.44b) \quad \hat{S}_n = \sum_{1 \leq i < i' \leq k} \hat{S}_{ii'}.
\]

**Remark 1.1.a.** Conditions (D1) and (D2) are a natural generalization of the conditions (A1), (B3) and (C1), in which the "fixed" constants of (C1) are now assumed to be random variables. This generalization is a very realistic one because in many practical applications the
independent variable (i.e. the X-variable) is not easily controlled and is in fact a realization of a random variable. This is the situation for any bivariate data collected by a survey method in which neither variable is controlled. An example is a medical study in which the subject is first chosen in some random manner and then the bivariate measurements are taken. However, as is shown in chapter two, the situation in which the independent variable is actually a fixed controllable constant can not be viewed as a degenerate case of (D1) and (D2). Consequently, (1.43) and (1.44) can not really be viewed as competitors to or as replacements for the test statistics proposed in the literature, except in situations where the existing, fixed X's procedures are improperly used. Instead, these test statistics ((1.43) and (1.44)) represent the only available strictly distribution-free solution to the regression problem of testing (1.36) under the assumption that the independent variable is random.

1.1.b. Expression (1.35) is needed only to allow the usual interpretation of the parameter \( \alpha \) as the "intercept" in the regression model.

1.1.c. From a practical viewpoint, the only really critical assumptions imposed on this model are implied in expression (1.33). The first assumption is one of equal "intercepts" across the k-samples. Without this assumption, (1.43) and (1.44) will only detect departures from the joint hypothesis (1.29). Consequently, (1.43) and (1.44) can be used only to test for the equality among "regression lines" and not for parallellism among "regression lines". The second assumption is one of equal distribution of the \( X_{ij} \)'s across the k-samples. This is necessary to allow ranking within the combined sample, instead within individual samples only. More comments about these two assumptions will be made in chapters five and six.
Remark 1.2.a. Property (1.38) is not at all restrictive since most reasonable estimators will satisfy it. This property is necessary only to establish the distribution-free property of (1.43) and (1.44).

2b. Likewise (1.39) is not that restrictive. This property is necessary only to establish the consistency properties of (1.43) and (1.44b).

Remark 1.3.a. One method of constructing a score-generating function $\phi(u)$, as in (1.13), is to let

\begin{equation}
\phi(u) = -g'(G^{-1}(u))/g(G^{-1}(u))
\end{equation}

where $g(x)$ is a strongly unimodal density function (i.e., $\ln g(x)$ is convex). [See Hájek-Šidák (1967), sec 1.2.4]

3b. If $g(x) = \phi'(x)$, where $\phi'(x)$ is the standard normal density function, then the scores (1.41) are called the normal scores and the scores (1.42) are called the van der Waerden scores. Then under (1.36), (1.43) is equivalent to the normal score (Fisher-Yates-Terry-Hoeffding) two-sample location test statistic with (1.41) or to the van der Waerden two-sample location test statistic with (1.42).

3c. If $g(x)$ is the logistic density, then both (1.41) and (1.42) are called the Mann-Whitney-Wilcoxon scores. Then under (1.36), (1.43) is equivalent to the two-sample Mann-Whitney-Wilcoxon location test statistic.

3d. The statistic 1.44a is equivalent under (1.36) to the Kruskal-Wallis $k$-sample statistic, whereas 1.44b is equivalent to the Jonckheere $k$-sample statistic for ordered alternatives.
[In remark 3b, 3c and 3d, equivalence refers to the concept that both test statistics have the same null probability distribution.]

**Remark 1.4.** In accordance with the idea established by Moses (1963) and supported by Fligner, Hogg, and Killeen (1976), the test statistics (1.43) and (1.44) will be referred to as *rank-like* statistics to distinguish them from the usual rank statistics. This is done to emphasize the fact that the ranks $\hat{R}_{ij}$ are based upon certain functions of the observations and not upon the observations themselves (which in (D1) have no well-defined ranking in the first place.)

In chapter two the null distribution is derived for (1.43) and (1.44). Examples of density functions satisfying (1.33) are developed. In chapter three, the likelihood ratio test statistic is derived for densities satisfying (1.33). In chapter four, the consistency class for (1.43) with the Mann-Whitney-Wilcoxon scores and for the Jonckheere (1.44b) are derived. In chapter five, results of a Monte Carlo simulation power study are presented. In chapter six is a discussion of these results, extensions to other problems, and areas of new research.

**1.4. Motivation**

As stated before, the goal of this dissertation is to develop a class of tests for (1.36) which are truly distribution-free, not just asymptotically so. The key to this distribution-free property is contained in the following example.
Example 1.1. For \( k=2 \) and \( \alpha=0 \), Fig. 1 is representative of a regression setting in which \( \beta_1 \neq \beta_2 \) and \( P(X>0) = 1 \). The areas marked I and II represent a "likely" range of observations for sample 1 and sample 2, respectively. The line \( y=\hat{\beta}x \) represents a common regression line in which \( \hat{\beta} \) is any reasonable slope estimate based on the combined sample. The quantities \( Y_{1j} - \hat{\beta} X_{1j} \) and \( Y_{2k} - \hat{\beta} X_{2k} (\text{sgn} X_{1j} = 1) \) represent the signed residuals of arbitrary observations from sample 1 and sample 2, respectively. It is clear from Fig. (1.1) that the difference in \( \beta_1 \) and \( \beta_2 \) could be detected by considering the ranks of \( Y_{1j} - \hat{\beta} X_{1j} \) in the combined sample since the larger ranks are associated with sample 1 and the smaller ranks are associated with sample 2. In fact, in this case a test of (1.36) is essentially just a test of a difference in location between the quantities \( (Y_{1j} - \hat{\beta} X_{1j}) \) and \( (Y_{2k} - \hat{\beta} X_{2k}) \) (as long as the two samples lie only in the first and/or fourth quadrants, see remark 1.c.). Consequently for a test based on the ranks of \( \text{sgn}(X_{1j})(Y_{ij} - \hat{\beta} X_{1j}) \) in the combined sample to be distribution-free, it is sufficient that every arrangement of these ranks be equally likely. To guarantee this, it seems very reasonable (if not necessary) to require that the value of \( \hat{\beta} \) be invariant under all permutations of the \((Y_{1j}, X_{1j})\) pairs. (The corresponding distribution-free property is not true under the assumption that the \( \{X_{1j}\} \) are fixed known constants.) Now this condition is just a restatement of (1.38) and the distribution-free property will hold if the variables \((Y_{1j}, X_{1j})\) are jointly continuously distributed (i.e., if (D1) is true). The proof that such a condition is sufficient
to guarantee that every arrangement of the ranks is equally likely is given in Chapter 2.

**Example 1.2.** Again for \( k=2 \) and \( \alpha=0 \), Fig. 2 is representative of a regression setting in which \( \beta_1 \neq \beta_2 \), except now the distribution of the \( X \)'s includes both positive and negative values. The quantities \( \{Y_{1j}-\hat{\beta}X_{1j}, \ Y_{1j}, \ Y_{1j} \} \) and \( \{Y_{2k}-\hat{\beta}X_{2k}, \ Y_{2k}, \ Y_{2k} \} \)

represent two unsigned residuals of two arbitrary observations from sample 1 and sample 2, respectively. It is clear from Fig. 2 that any two-sample location statistic based on the ranks of the **unsigned** residuals will not detect that \( \beta_1 \neq \beta_2 \) since these ranks will (roughly) always be evenly distributed among the two samples. But this lack of consistency does not exist if the ranks of the **signed** residuals are used instead of the unsigned residuals. The larger ranks of the signed residuals are then associated with sample 1 and the smaller ranks with sample 2.

Although this next example does not constitute a proof in itself, it indicates very clear how the assumption that the \( \{X_{ij}\} \) are fixed constants can prevent the quantities \( \{Y_{ij}-\hat{\beta}X_{ij}\} \) from having every arrangement of ranks being equally likely.

**Example 1.3.** Let \( (Y_1,x_1), (Y_2,x_2), (Y_3,x_3) \) be a random sample of size 3 where \( x_1 < x_2 < x_3 \) and \( x_1 \) is a fixed known constant. Let

\[ \hat{\beta} = \text{median}\{(Y_2-Y_1)/(x_2-x_1), (Y_3-Y_1)/(x_3-x_1), (Y_3-Y_2)/(x_3-x_2)\} \]

(\( \hat{\beta} \) does satisfy the invariance condition in (1.38)). Since

\[ (Y_3-Y_1)/(x_3-x_1) = a(Y_2-Y_1)/(x_2-x_1) + (1-a)(Y_3-Y_2)/(x_3-x_2) \]
where $a = (x_2 - x_1)/(x_3 - x_1)$, and $0 < a < 1$, it is clear that

$\hat{\beta} = (Y_3 - Y_1)/(x_3 - x_1)$. Consider now the random variables

\[
Y_1 - x_1 \hat{\beta} = Y_1(x_3/(x_3 - x_1)) - Y_3(x_1/(x_3 - x_4))
\]

\[
Y_2 - x_2 \hat{\beta} = Y_2 + Y_1(x_2/(x_3 - x_4)) - Y_3(x_2/(x_3 - x_1))
\]

\[
Y_3 - x_3 \hat{\beta} = Y_1(x_3/(x_3 - x_1)) - Y_3(x_1/(x_3 - x_1)) = Y_1 - x_1 \hat{\beta}.
\]

If $R_i$ equals the rank of $Y_i - x_i \hat{\beta}$, $1 \leq i \leq 3$, then clearly with probability one there will be a tie among these random variables and hence, not every arrangement of the ranks can be equally likely.
CHAPTER TWO

Distribution Theory Under the Null Hypothesis

2.1 Introduction

A general class of distribution-free rank-like statistics is derived, which includes statistics of the form (1.43) and (1.44). In addition, several properties and examples of distribution functions satisfying condition (D2) are discussed. Finally, several estimators satisfying condition (1.38) are presented.

2.2 Preliminary Results

Proposition 2.1. Assumption (1.33) implies that for $i = 1, \ldots, k$

\begin{equation}
(2.1) \quad h_i(x) = h(x)
\end{equation}

where $h_i(x)(h(x))$ is the marginal density function associated with the joint density function $f_i(y,x) (f(y,x))$; and

\begin{equation}
(2.2) \quad L_i(y|X = x) - L(y - \alpha - \beta_i x|X = x)
\end{equation}

when $L_i(y|X = x)(L(y|X = x))$ is the conditional distribution function associated with the joint distribution function $F_i(y,x)(F(y,x))$.  

27
Proof: (a) Clearly by (1.33) for each $i$,

$$h_i(x) = \int \phi_i(y, x) dy = \int \phi(y - \alpha - \beta_i x, x) dy$$

$$= \int \phi(z, x) dz \quad \text{(where } z = y - \alpha - \beta_i x)$$

$$= h(x).$$

(b) Also for each $i$, it is clear that the conditional density function

$$\phi_i(y | X = x) = \frac{\phi_i(y, x)}{h_i(x)}$$

$$= \frac{\phi(y - \alpha - \beta_i x, x)}{h(x)}$$

$$= \phi(y - \alpha - \beta_i x | X = x).$$

where

$$\phi(y | X = x) = \frac{\phi(y, x)}{h(x)}.$$

and the proof is completed.

Remark 2.1.a. Result (2.1) is an analog of the common sense dictum to conduct the experiment under "similar conditions" across the $k$ populations. In many practical applications the experimenter should attempt to insure the validity of (2.1) approximately, if not exactly. This allows for differences among the $k$ populations to exist only as differences among the parameters $\beta_1, \ldots, \beta_k$. [See also lemma 2.1].

1b. If the distribution of the "independent" variable is allowed to be degenerate, then (2.1) requires that each observation in the $k$ samples be taken at the same value of $X$. Consequently, the usual linear regression problem with the independent variable being a fixed known constant is not a degenerate case of the model stipulated by conditions (D1) and (D2).
Remark 2.2.a. Result (2.2) indicated clearly that condition (1.33) is one possible generalization of the linear regression problem with fixed "independent" variables.

2b. The proof of (2.2) indicates how to completely determine $k$ joint density functions which satisfy (1.33). By specifying the marginal density function $h(x)$ and the conditional density function $f(y|X = x)$, the joint density function for the $i^{th}$ sample is

$$f_i(y,x) = f(y - \alpha - \beta_i x|X = x)h(x).$$

Remark 2.3. The following conditions are sufficient to insure that $f_i(y,x)$ will satisfy conditions (D1) and (D2). Let $f(y,x)$ be a joint density function such that for constant $M_0$

$$f(y,x) = f(y) \cdot h(x),$$

$$h(x), f(y) \text{ continuous functions},$$

$$\text{median of the distribution associated with } f(y) \text{ is zero.}$$

Then

$$f_i(y,x) = f(y - \alpha - \beta_i x) \cdot h(x).$$

Expression (2.7) represents the situation in which the dependency between $X_{ij}$ and $Y_{ij}$ is solely a function of the regression parameter $\beta_i$. And if $\beta_i = 0$, then $X_{ij}$ and $Y_{ij}$ are stochastically dependent. Whereas, in (2.3), the dependency between $X_{ij}$ and $Y_{ij}$ could very well be a function of other quantities in addition to $\beta_i$. Nevertheless, as is shown in Lemma 2.1, this dependency must be the same across the $k$-samples when (1.36) is true.

Example 2.1. Possible examples of (2.7) are the following. For $h(x)$, choose one of the following: the uniform density  

\[ h(x) = \begin{cases} \frac{1}{b-a} & , \quad a \leq x \leq b, \\ 0 & , \quad \text{elsewhere}; \end{cases} \]

or the beta density
\[ h(x) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \left(1 - \frac{x}{b}\right)^{a-1} \left(1 - \frac{x}{a}\right)^{b-1} \]

for \(-M_0 \leq x \leq M_0\)
\[ = 0 \]
elsewhere.

For \(g(y)\), choose one of the following: the normal density
\[ g(y) = (2\pi \sigma^2)^{-\frac{1}{2}} \exp \left[ -(\frac{y}{\sigma})^2 \right] \quad -\infty < y < +\infty; \]
the logistic density
\[ g(y) = \frac{e^{-y}}{1 + e^{-y}}^2 \quad -\infty < y < +\infty; \]
the double exponential density
\[ g(y) = (\frac{1}{2})e^{-|y|} \quad -\infty < y < +\infty; \]
the Cauchy density
\[ g(y) = \pi^{-\frac{1}{2}}(1 + y^2)^{-\frac{1}{2}} \quad -\infty < y < +\infty. \]

The next example indicates that the class of joint density functions which satisfy the conditions of (D2) include the bivariate normal distribution. It will also serve to illustrate more clearly how (1.35) is essential only for interpretation of the parameter \(a\). The bivariate normal distribution of the random pair \((T_1, T_2)\) will be denoted by \(BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\). The joint density function has the form
\[ f(t_1, t_2) = (2\pi \sigma_1 \sigma_2(1 - \rho^2)^{\frac{1}{2}})^{-1} \exp \left[ -\frac{(t_1 - \mu_1)^2}{2\sigma_1^2} - \frac{(t_2 - \mu_2)^2}{2\sigma_2^2} - \rho(t_1 - \mu_1)(t_2 - \mu_2) \right] \]
where
\[ q = (1 - \rho^2)^{-1}[\sigma_{11}^{-2}(t - \mu_1)^2 - 2\rho\sigma_{12}^{-1}(t_1 - \mu_1)(t_2 - \mu_2) + \sigma_{22}^{-2}(t_2 - \mu_2)^2]. \]

**Example 2.2.** Let \((Y, X)\) and \((Y', X')\) have respectively joint density functions \(f(y, x)\) and \(f'_i(y, x)\) where \(f'_i(y, x)\) and \(f(y, x)\) satisfy (1.33).

Let \(f(y, x)\) be the joint density function of a \(BVN(a, b, s^2, t^2, r)\), where \(a, b, s, t,\) and \(r\) are fixed constants. Then
\[ f'_i(y, x) = f(y - a - \beta_1 x, x) \]
\[ = (2\pi st(1 - r^2)^{k-1})^{-1}\exp[-q^*/2] \]
where
\[ q^* = (1 - r^2)^{-1}[s^{-2}(y - a - \beta_1(x - b) - \beta_1 b - a)^2 \]
\[ - 2rs^{-1}t^{-1}(y - a - \beta_1(x - b) - \beta_1 b - a)(x - b) \]
\[ + t^{-2}(x - b)^2] \]
\[ = (1 - r^2)^{-1}[s^{-2}(y - a - \beta_1 b - a)^2 \]
\[ - 2(y - a - \beta_1 b - a)(x - b)(\beta_1 s^{-2} + r(st)^{-1}) \]
\[ + (x - b)^2(\beta_1 s^{-2} + 2\beta_1 r(st)^{-1} + t^{-2})]. \]

It is sufficient to recognize \(f'_i(x, y)\) as the density function of a \(BVN(u_{11}, u_{12}, \sigma_{11}^2, \sigma_{12}^2, \rho_1)\) by making the following identifications.

(i) \[ \sigma_{11}^2\rho_1^2(1 - \rho_1^2) = s^2 t^2 (1 - r^2) \]
(ii) \[ \sigma_{11}^2(1 - \rho_1^2) = s^2 (1 - r^2) \]
(iii) \[ \rho_1((1 - \rho_1^2)\sigma_{11}\sigma_{12})^{-1} = (1 - r^2)^{-1}(\beta_1 s^{-2} + r(st)^{-1}) \]
(iv) \[ \sigma_{12}^2(1 - \rho_1^2) = (1 - r^2)(\beta_1^2 s^{-2} + 2\beta_1 r s^{-1} t^{-2})^{-1} \]

(v) \[ \mu_{11} = a + \beta_1 b + a. \]

(vi) \[ \mu_{12} = b. \]

It follows from (i) and (ii) that

(vii) \[ \sigma_{12}^2 = t^2, \]

from (vii) and (iv) that

(viii) \[ (1 - \rho_1^2) = (1 - r^2)(\beta_1^2 t^2 s^{-2} + 2\beta_1 r s^{-1} t^{-2} + 1)^{-1} \]

from (ii) and (viii) that

(ix) \[ \sigma_{11}^2 = \beta_1^2 t^2 + 2\beta_1 rst + s^2 \]

and from (viii) and (ix) that

(x) \[ \rho_1^2 = (\beta_1^2 t^2 + 2\beta_1 rst + r^2 s^2) \sigma_{11}^{-2} \]

\[ = (\beta_1 t + rs)^2 \sigma_{11}^{-2} \]

Consequently, from (v), (vi), (vii), (ix), and (x), \( f_1(y, x) \) can be viewed as a density function for

\( \text{BVN}(\alpha + \beta_1 b + a, \beta_1^2 t^2 + 2\beta_1 rst + s^2, t^2, \beta_1 t + rs)(\beta_1^2 t^2 + 2\beta_1 rst + s^2)^{-\frac{1}{2}} \).

The conditional mean of \( Y_1 \) given \( X_1 = x \) is

(xi) \[ \mu_{Y_1 | X_1=x} = \mu_{11} + \rho_1 \sigma_{11} \sigma_{12}^{-1}(x - \mu_{12}) \]

\[ = (\alpha + a - st^{-1} x) + x(\beta_1 + st^{-1} r) \]
To satisfy condition (1.35) it is sufficient to examine the conditional mean

\[ \mu_{y|x=x} = a + rst^{-1}(x - b) \]

For \( \mu_{y|x=0} = 0 \), either \( a = b = 0 \) or \( a = r = 0 \). With the requirement that \( a = r = 0 \), (xi) reduces to

\[ (xi)' \quad \mu_{y_1|x_1=x} = \alpha + x\beta_1 \]

Clearly from (xi)', the requirement \( a = r = 0 \) allows \( \alpha \) to be interpreted as the "intercept" and \( \beta_1 \) as the "slope" of the (conditional) mean line. Also it should be noted that the requirement \( r = 0 \) allows \( f(y, x) \) to be derived using expression (2.7). With the requirement \( a = b = 0 \), (xi) reduces to

\[ (xi)' \quad \mu_{y_1|x_1=x} = \alpha + x(\beta_1 + xt^{-1}r) \]

Again \( \alpha \) represents the actual "intercept," but \( \beta_1 \) differs from the actual "slope" by a fixed known amount (\( st^{-1}r \)). But still hypothesis (1.36) is equivalent to the hypothesis of equal "slopes."

The following lemma indicates that the difference between the joint distribution governing the \( i^{th} \) sample and the joint distribution governing the \( j^{th} \) sample can only be in the difference between \( \beta_i \) and \( \beta_j \).

**Lemma 2.1.** Let \((Y_1, X_1)\) and \((Y_j, X_j)\) have respectively the joint density function \( f(y, x) \) and \( f_j(y, x) \). Assume that both \( f(y, x) \) and \( f_j(y, x) \) satisfy (1.33). Then the joint distribution of \((Y_1, X_1)\) is the same as that of \((Y_j - (\beta_j - \beta_1)X_j, X_j)\).
Proof: Clearly

\[ f_1(y, x) = f(y - \alpha - \beta_i x, x) \]
\[ = f(y - \alpha + (\beta_j - \beta_i)x - \beta_j x, x) \]
\[ = f_j(y + (\beta_j - \beta_i)x, x) \]

Let \( U_j = X_j \), \( V_j = Y_j - (\beta_j - \beta_i)X_j \). The absolute value of the Jacobian of the inverse transformation is one and the joint density of \( (V_j, U_j) \) is

\[ g(v_j, u_j) = f_j(v_j + (\beta_j - \beta_i)u_j, u_j) = f_1(v_j, u_j) \]

and the proof is completed.

Remark 2.4. It follows from Lemma 2.1 (as well as directly from (1.33)) that an equivalent hypothesis to (1.36) is

(2.8) \[ H^i: F_1(y, x) = F_2(y, x) = \ldots = F_k(y, x) = F(y, x) \]

where \( F_1(y, x) \) is defined in (D1).

2.3 A General Class of Rank-Like Statistics

In order to discuss the general technique for obtaining distribution-free rank-like statistics, the idea of a function that is \( p \)-tuple symmetric in \( N \) vector arguments must be introduced. Let \( x_i \), \( i = 1, \ldots, N \), be \( N \) arbitrary \( p \)-tuples of real numbers.
Definition 2.1. Let \( g(\cdot) \) be a function from \( pN \) dimensional Euclidean space to \( r \) dimensional Euclidean space such that
\[
g(x_1, \ldots, x_N) = g(x_{d_1}, \ldots, x_{d_N})
\]
for every permutation \((d_1, \ldots, d_N)\) of \((1, \ldots, N)\) and every set of \( N \) \( p \)-tuples \((x_1, \ldots, x_N)\). Then say that \( g(\cdot) \) is \( p \)-tuple symmetric in \( N \) vector arguments.

Using this idea of \( p \)-tuple symmetric functions, the following generalization of a theorem of Fligner, Hogg, and Killeen (1976) is obtained.

Theorem 2.1. [Smith and Wolfe (1977)]. Let
\[
x_i = (x_{i1}, \ldots, x_{ip}), \quad i = 1, \ldots, N
\]
be a random sample from some \( p \)-variate continuous distribution and let \( h(\cdot) \) be any real-valued function on \((p + r) \) dimensional Euclidean space. Let \( g(\cdot) \) be any function from \( pN \) dimensional Euclidean space to \( r \) dimensional Euclidean space that is \( p \)-tuple symmetric in \( N \) vector arguments and define the random variables
\[
W_i = h(x_i, g(x_1, \ldots, x_N))
\]
\( i = 1, \ldots, N \). Then \( W_1, \ldots, W_N \) are exchangeable random variables.

Proof: Let \((d_1, \ldots, d_N)\) be any permutation of \((1, \ldots, N)\). From the \( p \)-tuple symmetry of \( g(\cdot) \) it follows that
\[
g(x_1, \ldots, x_N) = g(x_{d_1}, \ldots, x_{d_N})
\]
In addition, since $X_1, \ldots, X_N$ are a random sample, then

$$(X_1, \ldots, X_N) \overset{d}{=} (X_{d_1}, \ldots, X_{d_N}).$$

These two facts combine to give immediately that

$$(W_1, \ldots, W_N) = (h(X_1, g(X_1, \ldots, X_N)), \ldots, h(X_N, g(X_1, \ldots, X_N)), h(X_{d_1}, g(X_{d_1}, \ldots, X_{d_N})), \ldots, h(X_{d_N}, g(X_{d_1}, \ldots, X_{d_N}))) = (W_{d_1}, \ldots, W_{d_N})$$

and the proof is completed.

The following corollary can be used to produce distribution-free rank-like statistics for many problems.

**Corollary 2.1.** Let $W_1, \ldots, W_N$ be defined as in the Theorem 2.1. If the functions $g(\cdot)$ and $h(\cdot)$ are such that $P(W_i = W_j) = 0$ for every $i \neq j$ then for every permutation $(d_1, \ldots, d_N)$ of $(1, \ldots, N)$

$$P((R_1, \ldots, R_N) = (d_1, \ldots, d_N)) = 1/N!$$

where $R_i$ is the rank of $W_i$ among $W_1, \ldots, W_N$.

Although the restriction that $P(W_i = W_j) = 0$ for all $i \neq j$ is not serious with regard to applications, it is necessary. If $p = r = 1$, $g(X_1, \ldots, X_N) = \text{median } (X_1, \ldots, X_N)$, and $h(s, t) = |s - t|$, for example, the variables $W_1, \ldots, W_N$ are exchangeable but the rank property of
Corollary 2.1 does not hold if \( N \) is even, since then there is a tied rank with probability one.

To apply the Theorem 2.1 and Corollary 2.1 to the model stipulated by conditions (D1) and (D2), let \( p = 2, r = 1, \)

\[
N = \sum_{i=1}^{k} n_i
\]

\( \hat{\theta} = g(X_1, \ldots, X_N) \)

and

\[
h(X_1, \hat{\theta}) = (X_{11} - \hat{\theta} X_{12}) \text{sgn}(X_{12}) = W_1
\]

The restriction that \( P(W_1 = W_j) = 0 \) for all \( i \neq j \) is equivalent to

\[
(2.9) \quad P(\hat{\theta} = [X_{11} - (\text{sgn} X_{j2}/\text{sgn} X_{12}) X_{j1}]^2 [X_{12} - (\text{sgn} X_{j2}/\text{sgn} X_{12}) X_{j2}]) = 0
\]

for all \( i \neq j \). The condition (1.38) is equivalent to \( g(\cdot) \) being 2-tuple symmetric in \( N \) vector arguments. Then, the following theorem is a direct result of Theorem 2.1 and Corollary 2.1.

**Theorem 2.2.** For the model stipulated by (D1) and (D2), let \( \hat{\theta} \) be any estimator based on the combined sample which satisfies (1.38) and (2.9). Then under the null hypothesis (1.36) [or (2.8)] the distribution of \( \hat{S}_n \), as defined in (1.43) or as defined in (1.44), is independent of the underlying joint distribution \( F(y, x) \).

**Remark 2.5.** In reference to remarks 1.3.b and 1.3.c, the practical importance of Theorem 2.2 is the existence of certain test statistics whose null distributions are well-known and well tabulated.
Aside from the work of Hollander (1970), no other solution to the problem of testing for the equality of linear regression lines has this property.

Remark 2.6. The appropriate critical region for testing (1.36) depends on the value of \( k \), the score-generating function \( \phi(u) \), and the alternative of interest. For example, if \( k = 2 \), the alternative is \( \beta_1 > \beta_2 \), and \( \hat{S}_n \) is the Mann-Whitney-Wilcoxon statistic

\[
\hat{S}_n = \sum_{j=1}^{n_1} R_{ij}
\]

the test rejects (1.36) if \( \hat{S}_n \) is too large.

Remark 2.7. It should be noted that as a result of Corollary 2.1, any \( k \)-sample rank statistic [not just those of the form in (1.43) or (1.44)] will be distribution-free under the null hypothesis.

2.4 Estimator \( \hat{\beta} \)

The only property of any estimator of \( \beta \) which is crucial for the distribution-freeness of any rank statistic based upon the signed residuals is the 2-tuple symmetry in \( N \) vector arguments [or (1.38)]. The following are three examples of such estimators. In all of the examples, \( (Y_i, X_i) \) represents the \( i^{\text{th}} \) observation from the combined sample.
Example 2.3.a. Let \( \hat{\beta} \) have the form proposed by Theil (1950) [or Sen (1968)] where

\[
\hat{\beta} = \text{median} \{ (Y_i - Y_j)/(X_i - X_j) : X_i \neq X_j, 1 \leq i < j \leq n \}
\]

Some care must be exercised when using this estimator. For example, if \( P(X_i > 0) = 1 \) and the number of pairs \( (i, j) \) in which \( X_i \neq X_j \) is odd, then

\[
\hat{\beta} = (Y_{i_0} - Y_{j_0})/(X_{i_0} - X_{j_0}) \quad \text{for some pair } (i_0, j_0)
\]

and condition (2.9) is violated.

2.3.b. Let \( \hat{\beta} \) be the usual least squares estimator, that is, let

\[
\hat{\beta} = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}
\]

where \( \bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i \) and \( \bar{X} = n^{-1} \sum_{i=1}^{n} X_i \). The advantage of this estimator is that it is contained in many statistical packages for computers (as well as for hand-held calculators). This means that with minor modifications these packages can be used to generate at least the signed residuals, if not the value of the chosen test statistic.

2.3.c. Let \( \hat{\beta} \) be the Hodge-Lehmann type estimator used by both Sen (1969) and Adichie (1967). That is, define

\[
T(b) = \sum_{i=1}^{n} (X_i - \bar{X})a_n(R_i(b))
\]
where $R_1(b)$ is the rank of $Y_1 - bX_1$ among $(Y_1 - bX_1, \ldots, Y_n - bX_n)$ and $a_n(i)$ is defined as in (1.41) or (1.42). Then let

$$\beta^* = \sup \{b : T(b) > 0\}$$

$$\beta^{**} = \inf \{b : T(b) < 0\}$$

and

$$\hat{\beta} = (\beta^* + \beta^{**})/2.$$
CHAPTER THREE

Likelihood Ratio Test

3.1 Introduction

The likelihood ratio test (LRT) for testing the null hypothesis (1.36) against the alternative in Theorem 3.1 derived for a given joint density function which satisfies (1.33), (1.34), and (1.35) following the work of Wald (1943). The asymptotic distribution of the LRT is given both under the null hypothesis and the alternative hypotheses. Finally the non-centrality parameter is evaluated for those joint distributions discussed in examples 2.1 and 2.2.

3.2 Preliminary Notation

The following notation will be used throughout this chapter.

Denote \( \beta' = (\beta_1, \ldots, \beta_k, \alpha) \). Let \( N_i, 0 < \lambda_1 < \ldots < \lambda_k < 1 \) be given (and fixed) such that

\[
(3.1) \quad n_i = \lambda_i N \quad \text{and} \quad \sum_{i=1}^{k} n_i = N.
\]

Denote the random pair \((Y_{ij}, X_{ij})\) by \(Q_{ij}\), for \(1 \leq i \leq k, 1 \leq j \leq n_i\).

Let \(Q_i = (Q_{i1}, \ldots, Q_{in_i})\), for \(1 \leq i \leq k\) and

\[
(3.2) \quad Q' = (Q'_1, \ldots, Q'_k)
\]
Let \( Q_{ij} \) have a joint density function \( f_i(q_{ij} | \alpha_i, \beta_j) \), which satisfy (1.33), (1.34), and (1.35). Then \( Q \), as defined in (3.2), has the joint density function

\[
(3.3) \quad g(q | \beta) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} f_i(q_{ij} | \alpha, \beta)
\]

Define the constants

\[
(3.4) \quad c_{ij} = -E_{\beta}^{\alpha^2 \log(g(q | \beta))/\partial \beta_i \partial \beta_j}
\]

where \( 1 \leq i \leq k + 1, 1 \leq j \leq k + 1 \) and \( \beta_{k+1} = \alpha \), and \( E_{\beta}[\cdot] \) is expectation with respect to the parameter vector \( \beta \). Now, \( Q \) will be considered as a single "observation" in the k-sample model. A single "observation" consists of \( N \) pairs.

Consider a random sample of size \( v \), \( Q_1, \ldots, Q_v \), \( 1 < v < \infty \), where \( Q^v = (Q^v_1, \ldots, Q^v_k) \) is the \( v^{th} \) "observation" of the form in (3.2). Each random sample represents a combined sample from the \( k \) populations in which there are \( v \cdot n_i \) pairs \((X_{ij}, Y_{ij})\) from the \( i^{th} \) population and \( v \cdot N \) pairs from all \( k \) populations. Denote \( \beta^*_{v1} \) to be the maximum likelihood estimate of \( \beta_i \), \( 0 \leq i \leq k \), based on the \( v \) "observations" \( Q_1, \ldots, Q_v \).

An equivalent hypothesis to (1.36) is

\[
(3.5) \quad H'': \quad \beta_1 - \beta_2 = \beta_1 - \beta_2 = \ldots = \beta_{k-1} - \beta_k = 0
\]

For \( 1 \leq i \leq k - 1 \), define \( \xi_i = \beta_i - \beta_{i+1} \) and define \( \xi_k = \beta_k \) and \( \xi_{k+1} = \beta_{k+1} = \alpha \). Then it is clear that

\[
\beta_i = \sum_{i=1}^{k} \xi_j \quad \text{and} \quad \beta_{k+1} = \xi_{k+1}
\]
Let the inverse Jacobian [(k + 1) \times (k \times 1) matrix] be

\[(3.6) \quad \left| \frac{\partial \theta_j}{\partial \xi_i} \right| = \left| \overline{d}_{ij} \right| \]

\[
\begin{bmatrix}
D_{11} & D_{12} \\
D_{21} & D_{22}
\end{bmatrix}
\]

where

\[D_{11}(k \times k) = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}, \quad D_{22} = 1
\]

\[D_{12}(k \times 1) = \begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix} = D_{21}^t
\]

Denote \(|\overline{c}_{ij}| = D^{'} \left| c_{ij} \right| D, \quad \left| c^*_{ij} \right| = \left| \overline{c}_{ij} \right|^{-1}\) and finally

\[\left| c^*_{pq} \right| = \left| \sigma^*_{pq} \right|^{-1}, \quad 1 \leq p \leq k-1, \quad 1 \leq q \leq k-1. \quad \text{[Note that \(\left| c^*_{pq} \right| \) is the inverse of the asymptotic covariance matrix of the maximum likelihood estimators for \(\xi_1, \ldots, \xi_{k-1}\) only.]} \]

Define the statistic

\[(3.7) \quad P_v = v \sum_{p=1}^{k} \sum_{q=1}^{k-1} (\beta^*_{vp} - \beta^*_{v(p+1)})(\beta^*_{vq} - \beta^*_{v(q+1)})c^*_{pq}
\]

\[= v \sum_{p=1}^{k-1} \sum_{q=1}^{k-1} \xi^*_{vp} \xi^*_{vq} c^*_{pq}
\]

where \(\xi^*_v = (\xi^*_v, \ldots, \xi^*_v)\).
Let $\Omega$ be the $(k + 1)$ dimensional parameter space of all possible values of the parameters $(\beta_1, \ldots, \beta_k, \alpha)$. Let $\omega$ represent the 2 dimensional subspace of $\Omega$ that is associated with the null hypothesis (1.36). Let

$$L_v = \sup_{v} \left[ \prod_{j=1}^{v} g(q_j | \theta) \right] / \sup_{\Omega} \left[ \prod_{j=1}^{v} g(q_j | \theta) \right]$$

where $g(q_j | \theta)$ is defined in (3.3). Then $L_v$ is called the likelihood ratio test (LRT) for (1.36).

Additional regularity assumptions about $g(q | \theta)$ are necessary to derive the asymptotic distribution of $L_v$ (and $P_v$). Since

$$g(q | \theta) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} f(y_{ij} - \alpha - \beta_i x_{ij}, x_{ij})$$

all assumptions will be given in terms of assumptions about the joint density function $f(y - \alpha - \beta x, x) = f(y - D_1 - D_2 x, x)$. Let $D = (D_1, D_2)$.

**E1.** Differentiation with respect to the parameters $D$ and integration with respect to the joint density $f(y - D_1 - D_2 x, x)$ can be interchanged. That is, for $1 \leq i, j \leq 2$

$$0 = \iint \partial f(y - D_1 - D_2 x, x) / \partial D_i \, dy \, dx$$

$$= \iint \partial^2 f(y - D_1 - D_2 x, x) / \partial D_i \, D_j \, dy \, dx$$

Denote $\partial f(u, x) / \partial u \big|_{u=y-D_1-D_2x}$ by $\partial f(u, x) / \partial u$ and $\partial^2 f(u, x) / \partial^2 u \big|_{u=y-D_1-D_2x}$ by $\partial^2 f(u, x) / \partial^2 u$.

**E2.** $\partial f(u, x) / \partial u$ and $\partial^2 f(u, x) / \partial^2 u$ are continuous functions of $D_1$ and $D_2$. 

The next assumption is necessary to insure that

\[
v^{-1} \sum_{m=1}^{m} \partial \xi_{n}(g_{m}|^{R})/\partial \beta \partial j F_{c_{ij}}
\]

Denote \( \psi_{lm}(y,x,D_{1},\delta) \) as the greatest lower bound and \( \phi_{lm}(y,x,D_{1},\delta) \) as the least upper bound of \( \partial^{2} \xi_{nf}(y - D_{1} - D_{2} x, x)/\partial D_{x} \partial D_{m} \), \( 1 \leq l, m \leq 2 \) with respect to the \( |D_{1} - D_{2}| \leq \varepsilon \), for \( 1 \leq l \leq 2 \).

E3. (a) For any sequence \{D_{1}^{v}\} \{D_{2}^{v}\} and \{\delta_{v}\} for which

\[
\lim_{v \to +\infty} D_{1}^{v} = \lim_{v \to +\infty} D_{2}^{v} = D_{1} = D_{2} = 12, \text{ and } \lim_{v \to +\infty} \delta_{v} = 0
\]

(as \( v \to +\infty \)), it follows that

\[
\lim_{v \to +\infty} E_{D_{1}}(\psi_{lm}(y,x,D_{1}^{v},\delta_{v})) = \lim_{v \to +\infty} E_{D_{1}}(\phi_{lm}(y,x,D_{2}^{v},\delta_{v}))
\]

\[
= E(\partial^{2} \xi_{nf}(y - D_{1} - D_{2} x, x)/\partial D_{x} \partial D_{m}) = -c_{lm}
\]

uniformly in \( D \).

(b) \( \exists \varepsilon > 0 \) such that \( E_{D_{1}}(\psi_{lm}(y,x,D_{2}^{v},\delta))^{2} \) and

\( E_{D_{1}}(\phi_{lm}(y,x,D_{2}^{v},\delta))^{2} \) are bounded functions of \( D_{1}, D_{2} \) and \( \delta \) in the domain \( D_{\varepsilon} \) defined by

\[
|D_{1}^{v} - D_{2}^{v}| < \varepsilon, \ |\delta| < \varepsilon.
\]

E4. The matrix \( ||c_{ij}|| \) is positive definite.
E5. Denote by $E_v$ the set of all sample points for which $\beta^*_v$ exists. Then \( \lim_{v \to +\infty} P(E_v | \beta) = 1 \) (uniformly in $\beta$) and \( \lim_{v \to +\infty} P( | \beta^* - \beta | < \varepsilon | \beta ) = 1 \) (uniformly in $\beta$), (where $| \beta^* - \beta | < \varepsilon$ denotes the vector of inequalities $| \beta^*_{v_i} - \beta_{v_i} | < \varepsilon$, $1 \leq i \leq k + 1$).

Denote the distribution of a non-central chi-square with $r$ degrees of freedom and non-centrality parameter $\Delta$ by $F(y; r, \Delta)$. The following theorem follows directly from Wald (1943) (pages 478 and 480).

Theorem 3.1. Under conditions E1 to E5, (1.33) and the sequence of alternatives $\beta_v$, $\beta_{v_i} = \beta + (vN)^{-\frac{1}{2}} \theta_1$, the following are true

(a) \( \lim_{v \to +\infty} P[ -2\ln L_v - \varepsilon < \theta_v < -2\ln L_v + \varepsilon | \beta_v ] = 1 \)

(b) \( \lim_{v \to +\infty} P[ -2\ln L_v < y | \beta_v ] - F(y; k - 1, \Delta_v(\beta_v) ) ] = 0 \)

uniformly in $y$, where

\[
\Delta_v(\beta_v) = v \sum_{p=1}^{k-1} \sum_{q=1}^{k-1} (\beta_{vp} - \beta_v(p+1))(\beta_{vq} - \beta_v(q+1))c^*_{pq}
\]

where $| |c^*_{pq} | |$ is defined in (3.6),

(c) when $\beta \in \omega$, the asymptotic distribution of $-2\ln L_v$ (and $P_v$) is central chi-square with $k - 1$ degrees of freedom.
3.3 An Equivalent Form of $P_v$

An algebraically equivalent form of $P_v$, which also gives a more useful form of the non-centrality parameter $\Delta_0(\beta_v)$, is

\[
(3.9) \quad P_v = c_2(f) \sum_{i=1}^{k} n_i (\beta_{v_i}^{*} - \beta_{v_0}^{*})^2
\]

where $\beta_{v_0}^{*} = \sum_{i=1}^{k} \lambda_{i} \beta_{v_i}^{*}$, $\beta_{v_i}^{*}$ is the MLE of $\beta_i$, $n_i$ and $\lambda_i$ are defined in (3.1), and $c_2(f)$ is a function of the joint density function $f(y,x)$.

Under the sequence of alternatives in Theorem 3.1 the non-centrality parameter is

\[
(3.10) \quad \Delta_v(\beta_v) = c_2(f) \sum_{i=1}^{k} \lambda_{i} (\theta_{i} - \sum_{i=1}^{k} \lambda_{i} \theta_{i})^2
\]

Before stating the proposition which establishes the validity of (3.9) [and (3.10)] the following definitions are made:

Since $f(y - \alpha - \beta x, x) = \lambda(y - \alpha - \beta x|x)h(x)$, it follows that

\[
(3.11) \quad \begin{align*}
(a) & \quad \int \left[ (1/\lambda(y - \alpha - \beta x, x)) \frac{\partial f(u,x)}{\partial u} \right]^2 f(y - \alpha - \beta x, x) dy dx \\
& = \int \left[ (1/\lambda(y - \alpha - \beta x|x)) \frac{\partial \lambda(y|x)}{\partial u} \right]^2 \lambda(y - \alpha - \beta x|x) dy \\
& \quad \cdot h(x) dx \\
& = \int \left[ (1/\lambda(y|x)) \frac{\partial \lambda(y|x)}{\partial u} \right]^2 \lambda(y|x) dy h(x) dx \\
& = \int I(L(y|x)) h(x) dx = c_0(f)
\end{align*}
\]

where $\partial f(u,x)/\partial u$ [and similarly $\partial \lambda(u|x)/\partial u$] is defined as in (E.2); and $I(L(y|x))$ is Fisher information of the conditional distribution function $L(y|x)$ associated with the joint density function $f(y,x)$ [in (1.33)];

\[
(b) \quad \int x \left[ (1/f(y - \alpha - \beta x, x)) \frac{\partial f(u,x)}{\partial u} \right]^2 f(y - \alpha - \beta x, x) dy dx \\
= \int x I(L(y|x)) h(x) dx = c_1(f)
\]
(c) $\int x^2 [(1/f(y - \alpha - \beta x, x)) \partial f(u, x)/\partial u]^2 f(y - \alpha - \beta x, x) dy dx$

$$= \int x^2 I(L(y | x) h(x) dx = c_2(f)$$

[It is implied by the notation used and is crucial to the above calculations that $f(y, x)$ does not depend on $\beta$ in any manner.]

**Proposition 3.1.** Under the assumptions (1.33) and (3.1)

$$V X \times C S J_{p}^{k-1} \times C S J_{q}^{k-1}$$

$$\sum_{p=1}^{k-1} \sum_{q=1}^{k-1} (\beta_{v p}^{*} - \beta_{v (p+1)}^{*}) (\beta_{v q}^{*} - \beta_{v (q+1)}^{*}) c_{pq}^{*}$$

$$= c_2(f) \sum_{p=1}^{k} \sum_{q=1}^{k} (\beta_{v p}^{*} - \beta_{v (p+1)}^{*}) (\beta_{v q}^{*} - \beta_{v (q+1)}^{*})$$

$$\sum_{i=1}^{a_{pq}} \sum_{i=b_{pq}}^{k} n_i/N$$

$$= c_2(f) \sum_{i=1}^{k} n_i (\beta_{v 0}^{*} - \beta_{v 0}^{*})^2$$

where $\beta_{v 0}^{*}$ is defined in (3.9), $c_{pq}^{*}$ is defined in (3.6), and (3.12)

$$a_{pq} = \min(p, q), \ b_{pq} = \max(p, q), \ 1 \leq p, q \leq k - 1.$$  

**Proof:** The first equality is proved by showing that

$$c_{pq}^{*} = c_2(f) \sum_{i=1}^{a_{pq}} \sum_{i=b_{pq}+1}^{k} n_i/N$$

Since

$$\ln g(q | \vec{\beta}) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln f(y_{ij} - \beta_{i} x_{ij} - \beta_{k+1} x_{ij})$$

it is clear that for $1 \leq i \leq k$,
(3.13) 

(a) \[- \frac{\partial^2 \ln g(q|\beta)}{\partial^2 \beta_i} \]
\[- \sum_{j=1}^{n_i} \left[ \frac{1}{f(y_{ij} - \beta_i^1 x_{ij} - \beta_{k+1}^1 x_{ij}^1)} \left( \frac{\partial^2 f(u,x_{ij})}{\partial^2 u} x_{ij}^2 \right) \right. \\
- \left. \left( x_{ij}^1 f(y_{ij} - \beta_i^1 x_{ij} - \beta_{k+1}^1 x_{ij}^1) \right) \left( \frac{\partial f(u,x_{ij})}{\partial u} \right) \right]^2 \]

(b) \[- \frac{\partial^2 \ln g(q|\beta)}{\partial \beta_i \partial \beta_{k+1}} \]
\[- \sum_{j=1}^{n_i} \left[ \frac{1}{f(y_{ij} - \beta_i^1 x_{ij} - \beta_{k+1}^1 x_{ij}^1)} \left( \frac{\partial^2 f(u,x_{ij})}{\partial^2 u} \right) \right. \\
- \left. \left( x_{ij}^1 f(y_{ij} - \beta_i^1 x_{ij} - \beta_{k+1}^1 x_{ij}^1) \right) \left( \frac{\partial f(u,x_{ij})}{\partial u} \right) \right]^2 \]

(c) \[- \frac{\partial^2 \ln g(q|\beta)}{\partial^2 \beta_{k+1}} \]
\[- \sum_{j=1}^{n_i} \left[ \frac{1}{f(y_{ij} - \beta_i^1 x_{ij} - \beta_{k+1}^1 x_{ij}^1)} \left( \frac{\partial^2 f(u,x_{ij})}{\partial^2 u} \right) \right. \\
+ \left. \frac{3}{2} \left( \frac{1}{f(y_{ij} - \beta_i^1 x_{ij} - \beta_{k+1}^1 x_{ij}^1)} \left( \frac{\partial f(u,x_{ij})}{\partial u} \right) \right) \right]^2 \]

(d) for \(1 < i \neq j < k\)
\[- \frac{\partial^2 \ln g(q|\beta)}{\partial \beta_i \partial \beta_j} = 0 \]

where \(\frac{\partial^2 f(u,x_{ij})}{\partial^2 u}\) and \(\frac{\partial f(u,x_{ij})}{\partial u}\) is defined as in (E.2). From (E.1), (1.33), and (3.11) it follows that for \(1 \leq i \leq k\)

(3.14) 

(a) \(c_{i1} = n_i c_2(f)\)

(b) \(c_{i(k+1)} = n_i c_1(f)\)

(c) \(c_{(k+1)(k+1)} = Nc_0(f)\)

(d) \(c_{ij} = 0, \quad 1 \leq i \neq j \leq k\)
Using (3.6) it follows that

\[(3.15) \quad ||c_{ij}|| = D'||c_{ij}||D\]

Then

\[E_{11} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}\]

\[E_{11}(k-1) \times (k-1) = ||e_{11}||; \quad e_{11}^i = c_2(f) \sum_{i=1}^{a_{pq}} n_i\]

\[E_{12} = ||e_{12}||, \quad e_{12}^i = c_2(f) \sum_{i=1}^{a_{pq}} n_i\]

\[e_{2q}^i = c_1(f) \sum_{i=1}^{a_{pq}} n_i, \quad 1 \leq q \leq k - 1\]

\[E_{21} = E_{12}; \text{ and}\]

\[E_{22} = \begin{bmatrix} c_2(f) & c_1(f) \\ c_1(f) & c_0(f) \end{bmatrix}\]

Then

\[||c_{ij}^*|| = \begin{bmatrix} E_{11.2}^{-1} & -E_{11}^{-1}E_{12}E_{22.1}^{-1} \\ -E_{22}^{-1}E_{21}E_{11.2}^{-1} & E_{22.1}^{-1} \end{bmatrix}\]

where \(E_{11.2} = E_{11} - E_{12}E_{22}E_{21}\), \((k-1) \times (k-1)\) matrix,

\[E_{22.1} = E_{22} - E_{21}E_{11}E_{12}\]

Finally

\[||c_{pq}|| = E_{11.2} = ||e_{pq}|| - ||f_{pq}||\]

where

\[||f_{pq}|| = E_{12}E_{22}^{-1}E_{21}; \quad f_{pq} = c_2(f) \sum_{i=1}^{a_{pq}} n_i \sum_{i=1}^{b_{pq}} n_i/N\]
Therefore
\[ c_{pq}^* = c_{pq} = c_{pq} = c_2(f) \sum_{i=1}^{a_{pq}} n_i (1 - (\sum_{i=1}^{b_{pq}} n_i / N)) \]
\[ = c_2(f) \sum_{i=1}^{a_{pq}} n_i \sum_{i=b_{pq}+1}^{k} n_i / N \] (since \( \sum_{i=1}^{k} n_i = N \)).

which proves the first equality.

The second equality follows by expanding \( \sum_{i=1}^{k} n_i (\beta_{v1}^* - \beta_{v0}^*)^2 \) and collecting terms. Since \( \lambda_j N = n_j \)

\[ \sum_{i=1}^{k} n_i (\beta_{v1}^* - \beta_{v0}^*)^2 = \sum_{i=1}^{k} \lambda_j (\beta_{v1}^* - \beta_{v0}^*)^2 \]

\[ = (1/N^2) \sum_{i=1}^{k} n_i (\beta_{v1}^* \sum_{j=1}^{k} n_j - \sum_{j=1}^{k} n_j \beta_{v0}^*)^2 \]

\[ = (1/N^2) \sum_{i=1}^{k} \sum_{j=1}^{k} n_j (\beta_{v1}^* - \beta_{v0}^*)^2. \]

For \( i < j \), \( \beta_{v1}^* - \beta_{v0}^* = \sum_{l=j+1}^{j-1} (\beta_{vl}^* - \beta_{v(l+1)}^*) \) and for \( i > j \),

\[ \beta_{v1}^* - \beta_{v0}^* = \sum_{l=j+1}^{i-1} (\beta_{vl}^* - \beta_{v(l+1)}^*) \] and (3.16) can be rewritten as

\[ \sum_{i=1}^{k} \sum_{j=1}^{k} n_j (\beta_{vl}^* - \beta_{v(l+1)}^*)^2 \]

\[ = (1/N^2) \sum_{i=1}^{k} \sum_{j=1}^{k} n_j (\beta_{vl}^* - \beta_{v(l+1)}^*)^2 = \]

\[ \sum_{l=j+1}^{i-1} \sum_{j=1}^{k} n_j (\beta_{vl}^* - \beta_{v(l+1)}^*)^2. \]
\[ = \left(\frac{1}{N^2}\right) \sum_{i=1}^{k} n_i \left( \sum_{l=1}^{k-1} h_{il} (\beta^*_v - \beta^*_v (l+1)) \right)^2 \]

where

\[ h_{il} = \begin{cases} 
\sum_{j=l+1}^{k} n_j & \text{if } l \geq i \\
- \sum_{j=1}^{l} n_j & \text{if } l < i 
\end{cases} \]

By expanding the squared term and interchanging the summations in (3.17), it becomes

\[ (3.18) \quad \left(\frac{1}{N^2}\right) \sum_{l=1}^{k-1} \sum_{l'=1}^{k-1} (\beta^*_v (l+1)) (\beta^*_v (l'+1)) (\sum_{i=1}^{k} n_i h_{il} h_{i'l'}) \]

Fix \( l \leq l' \), then

\[ \sum_{i=1}^{k} n_i h_{il} h_{i'l'} = \sum_{i=1}^{k} [n_i (\sum_{j=l+1}^{k} n_j) (\sum_{j=l'+1}^{k} n_j)] + \sum_{i=l+1}^{l'} [n_i (- \sum_{j=1}^{l} n_j) (\sum_{j=l'+1}^{k} n_j)] = (\sum_{i=1}^{k} n_i) (\sum_{i=l+1}^{k} n_i) - \sum_{i=l+1}^{l'} n_i + \sum_{i=l+1}^{k} n_i \]

\[ = (\sum_{i=1}^{k} n_i) (\sum_{i=l'+1}^{k} n_i) + \sum_{i=l+1}^{l'} n_i + \sum_{i=1}^{k} n_i \]

\[ a_{l,l'} \]

\[ = (\sum_{i=1}^{k} n_i) (\sum_{i=b_{l,l'+1}}^{k} n_i) N \]
Finally (3.18) can be written as
\[ \frac{1}{N} \sum_{\ell=1}^{k-1} \sum_{\ell'=1}^{k-1} (\beta_{\ell}^* - \beta_{\ell+1}^*) (\beta_{\ell'}^* - \beta_{\ell'+1}^*) \left( \sum_{i=1}^{a_{\ell\ell'}} n_i \right) \left( \sum_{i=b_{\ell\ell'+1}}^{k} n_i \right) \]
and the second equality is proved.

Remark 3.1. Under conditions of remark 2.3, it is clear that
\[ c_2(f) = I(L(y)) \int x^2 h(x) dx \]
\[ = I(L(y)) (\sigma_x^2 + \mu_x^2) \]
where \( L(y) \) is the distribution function associated with the density \( \ell(y) \) and \( \sigma_x^2 \) and \( \mu_x \) are the variance and mean respectively associated with the density \( h(x) \). For example 2.2, with \( a = r - 0, b, s^2, t^2 \) known constants
\[ c_2(f) = s^{-2} (t^2 + b^2). \]

Remark 3.2. Since \( f(y - a - \beta x, x) = \ell(y - a - \beta x | x) h(x) \)
the MLE's of \( a \) and \( \beta \) have the same form as the MLE's in the case where
the independent variables are fixed known constants.
Chapter Four

Consistency

4.1. Introduction

In this chapter the consistency classes of the two-sample rank-like statistic using the Mann-Whitney-Wilcoxon scores and of the k-sample rank-like statistic using the Jonckheere scores are derived. In addition some general conditions are given which insure that the least-squares estimator $\beta_n$ of $\beta$ (example 2.3.b) is a consistent estimator under the alternative $\beta_1 \geq \beta_2$.

4.2. Preliminary Definitions and Results for $k = 2$.

The following assumptions are made:

(4.1) Let $f(u,v)$ be an arbitrary joint density function for a distribution which is absolutely continuous with respect to the Lebesgue measure on the plane and whose marginal distributions are also absolutely continuous with respect to Lebesgue measure.

For $i = 1,2$, let $\{(Y_{ij},X_{ij}): 1 \leq j \leq n_i\}$ be two bivariate samples, each from a population with joint density function

$\begin{equation}
(4.2) f_i(y,x) = f(y - \alpha - \beta_i x, x).
\end{equation}$

For each $b$, $i = 1,2$, and $1 \leq j \leq n_i$, define
(4.3) \[ Z_{ij}(b) = (Y_{ij} - bX_{ij}) \text{ sgn } X_{ij} \]

(4.4) \[ R_{ij}(b) = \text{rank}(Z_{ij}(b)) \]

where the rank is computed from the combined sample, and

(4.5) \[ G^b_1(z) = P(Z_{ij}(b) \leq z). \]

For \( 1 \leq k \leq n_1, 1 \leq m \leq n_2 \), let

(4.6) \[ D_{km}(b) = 1 \text{ if } Z_{2m}(b) < Z_{1k}(b) \]
\[ = 0 \text{ if } Z_{2m}(b) > Z_{1k}(b). \]

Then the Mann-Whitney form of the Mann-Whitney-Wilcoxon rank statistic is defined as

(4.7) \[ S_n(b) = \sum_{k=1}^{n_1} \sum_{m=1}^{n_2} D_{km}(b). \]

Assume that the alternative is of the form

(4.8) \[ \beta_1 = \beta + \theta_1 > \beta_2 = \beta + \theta_2. \]

Let \( \hat{\beta}_n \) be any estimator of \( \beta \) in (4.8) based on the combined sample such that

(4.9) \[ \hat{\beta}_n \overset{P}{\rightarrow} \beta. \]

Then

\[ S_n(\hat{\beta}_n) = S_n(1.43). \]

The next lemma demonstrates that for any \( b \), \( Z_{2m}(b) \) is stochastically smaller than \( Z_{1k}(b) \) when \( \beta_1 > \beta_2 \).

Lemma 4.1. Under the conditions (4.1), (4.2), (4.3) and (4.8), for any \( b \)

\[ G^b_1(z) \leq G^b_2(z). \]
with strict inequality for at least one $z$.

Proof: It is very clear that if $(Y, X)$ has a joint density function \( f(y - \alpha - \beta x, x) \), then \( (Y - \beta X, X) = (U, V) \) has a joint density function \( f(u - \alpha, v) \). Using this fact, it follows that

\[
Z_{1j}(b) \overset{D}{=} (U + (\beta_1 - b)V) \operatorname{sgn} V
\]
\[
Z_{2j}(b) \overset{D}{=} (U + (\beta_2 - b)V) \operatorname{sgn} V
\]

where "\( D \)" means equality of distributions and \((U, V)\) has a joint density function \( f(u - \alpha, v) \). Because \( \beta_1 > \beta_2 \) and \( V \operatorname{sgn} V = |V| \)

\[
G_1^b(z) = P(Z_{1j}(b) \leq z) = P((U + (\beta_1 - b)V) \operatorname{sgn} V \leq z)
\]
\[
\leq P((U + (\beta_2 - b)V) \operatorname{sgn} V \leq z) = P(Z_{2j}(b) \leq z) = G_2^b(z)
\]

and the proof is completed.

The next lemma gives the expected value of \( S_n(b) \) and develops the necessary properties for consistency.

Lemma 4.2. Under the conditions (4.1), (4.2), (4.3), (4.5), and (4.7), for any value of \( b \)

(a) \( E(S_n(b)) = n_1 n_2 \pi(b) = n_1 n_2 \int G_2^b(z)g_1^b(z) \, dz \)

in which \( g_1^b(z) = dG_1^b(z)/dz \).

(b) If \( G_2^b(z) = G_1^b(z) \), then \( \pi(b) = 1/2 \).

(c) If \( G_2^b(z) \geq G_1^b(z) \), then \( \pi(b) > 1/2 \).

Proof: (a) For any \((k, m)\) pair,

\[
E(D_{km}(b)) = P(Z_{2m}(b) < Z_{1k}(b))
\]
\[
= \int G_2^b(z)g_1^b(z) \, dz = \pi(b).
\]

Thus

\[
E(S_n(b)) = \sum_{k=1}^{n_1} \sum_{m=1}^{n_2} E(D_{km}(b)) = n_1 n_2 \pi(b).
\]
Part (b) and (c) are immediate and the proof is completed.

The parameter $\pi(b)$ can be written as a sum in the following manner.

For any pair $(k,m)$, $1 \leq k \leq n_1$, $1 \leq m \leq n_2$, define

$$\pi_1(b) = P(Z_{2m}(b) < Z_{1k}(b) \text{ and } |X_{2m}| < |X_{1k}|)$$
$$\pi_2(b) = P(Z_{2m}(b) < Z_{1k}(b) \text{ and } |X_{2m}| > |X_{1k}|).$$

Clearly then

$$\pi(b) = \pi_1(b) + \pi_2(b)$$

because the marginal distribution of $X$ is absolutely continuous.

Although $S_n(b)$ is not monotonic in $b$, the next lemma shows that $S_n(b)$ is the sum of two monotonic functions, one non-increasing and the other non-decreasing. First define

$$Q_{km} = 1 \quad \text{if } |X_{2m}| < |X_{1k}|$$
$$= 0 \quad \text{if } |X_{2m}| > |X_{1k}|.$$

Define

$$(4.11) \quad T_1(b) = \sum_{k=1}^{n_1} \sum_{m=1}^{n_2} D_{km}(b)Q_{km}$$
$$(4.11) \quad T_2(b) = \sum_{k=1}^{n_1} \sum_{m=1}^{n_2} D_{km}(b)(1 - Q_{km}).$$

Clearly

$$S_n(b) = T_1(b) + T_2(b).$$

**Lemma 4.3.** Under conditions (4.1)-(4.3), (4.5)-(4.8), and (4.11) and for $a,b$ any values such that $a < b$ it follows that

(a) $T_1(a) \geq T_1(b)$
(b) $T_2(a) \leq T_2(b)$. 
Proof: (a) $T_1(a) \geq T_1(b)$ if for each pair $(k,m)$ $D_{km}(b)Q_{km} = 1$ implies that $D_{km}(a)Q_{km} = 1$. Now $D_{km}(b)Q_{km} = 1$ implies that both

\begin{equation}
|x_{2m}| < |x_{1k}| \quad \text{and} \quad \text{sgn } x_{2m} < (Y_{1k} - bX_{1k}) \quad \text{sgn } x_{1k},
\end{equation}

(4.12)

It follows from (4.12) and (4.13) that because $x \text{ sgn } x = |x|$,

\begin{equation}
Y_{2m} \text{ sgn } x_{2m} < Y_{1k} \text{ sgn } x_{1k} - b(|x_{1k}| - |x_{2m}|) < Y_{1k} \text{ sgn } x_{1k} - a(|x_{1k}| - |x_{2m}|).
\end{equation}

(4.14)

From (4.14) it follows that

\begin{equation}
(Y_{2m} - ax_{2m}) \text{ sgn } x_{2m} < (Y_{1k} - ax_{1k}) \text{ sgn } x_{1k}
\end{equation}

and that

\begin{equation}
D_{km}(a)Q_{km} = 1.
\end{equation}

(b) In a similar manner it can be shown that $D_{km}(a)(1 - Q_{km}) = 1$ implies that $D_{km}(b)(1 - Q_{km}) = 1$ and the proof is completed.

It follows from the last lemma that if $a < b < c$ then

\begin{equation}
T_1(c) + T_2(a) \leq S_n(b) \leq T_1(a) + T_2(c).
\end{equation}

Finally it should be noted that

\begin{equation}
E(T_1(b)) = n_1n_2\varpi_1(b)
\end{equation}

\begin{equation}
E(T_2(b)) = n_1n_2\varpi_2(b).
\end{equation}

4.3. Consistency of the Mann-Whitney-Wilcoxon

The first lemma, which is adapted from Fligner (1975), gives sufficient conditions in which a statistic based on an estimator of a parameter will have the same stochastic limit as the same statistic based on the true value of the parameter. The second and third lemmas are needed
to show that \((n_1 n_2)^{-1} S_n(\hat{\beta})\) and the same stochastic limit as
\((n_1 n_2)^{-1} S_n(\beta)\) which clearly is \(\pi(\beta)\).

For the first lemma let \(W_n(b) = W_n(Y_{11}, \ldots, Y_{12n}, X_{2n}; b)\) and \(V_n(c) = V_n(Y_{11}, \ldots, Y_{12n}, X_{2n}; c)\) be arbitrary statistics.

**Lemma 4.4.** Let \(\hat{\beta}_n\) be an estimator of \(\beta\) such that
(a) \(\hat{\beta}_n \Rightarrow \beta\).
If for any positive constant \(c\),
(b) \(\sup_{|v| \leq c} |W_n(\beta - v) - m(\beta)| \leq V_n(c)\)
(c) \(V_n(c) \leq M(c)\)
(d) \(\lim_{m \to \infty} M(c_m) = 0\), where \(\{c_m\}\) is any decreasing sequence of
positive numbers such that \(c_m \to 0\) as \(m \to \infty\).

Then \(W_n(\beta) \Rightarrow m(\beta)\).

**Proof:** For any \(c > 0\), \(\epsilon > 0\) there exist \(N_1(c, \epsilon)\) such that
\(P(A_n(c, \epsilon)) = P(|\beta_n - \beta| < c) > 1 - \epsilon\) for \(n > N_1(c, \epsilon)\).
In addition for any \(\delta, \delta > 0\), there exists \(N_2(c, \epsilon, \delta)\) such that
\(P(B_n(c, \epsilon, \delta)) = P(V_n(c) \leq M(c) + \delta) > 1 - \epsilon\)
for \(n > N_2(c, \epsilon, \delta)\). Pick a specific decreasing sequence \(\{c_m\}\). Then there
exists \(m^*\) such that for \(c^* = c_{m^*}\), \(M(c^*) < \delta\).
For \(n > \max(N_1(c^*, c), N_2(c^*, c, \delta))\)
\(P(A_n(c^*, \epsilon) \cap B_n(c^*, \epsilon, \delta)) > 1 - 2\epsilon\)
and for each element in \(A_n(c^*, \epsilon) \cap B_n(c^*, \epsilon, \delta)\) there exist a \(v\) so that
\(|v| \leq c^*, \hat{\beta}_n = \beta - v\), and
\[ |W_n(\hat{\alpha}_n) - m(\beta)| = |W_n(\beta - \nu) - m(\beta)| \]
\[ \leq \sup_{|\nu| \leq c} \ |W(\beta - \nu) - m(\beta)| \leq V_n(c^*) \leq 2\delta. \]

Hence \( W_n(\hat{\alpha}) \not\equiv m(\beta) \) and the proof is completed.

It follows from (4.15) that for \( |\nu| \leq c \),
\[ T_1(\beta + c) + T_2(\beta - c) \leq S_n(\beta - \nu) \leq T_1(\beta - c) + T_2(\beta + c) \]
and that because \( \pi(\beta) = \pi_1(\beta) + \pi_2(\beta) \)
\[ (4.17) \quad \sup_{|\nu| \leq c} |(n_1n_2)^{-1}s_n(\beta - \nu) - \pi(\beta)| \leq \]
\[ |(n_1n_2)^{-1}(T_1(\beta - c) + T_2(\beta + c)) - \pi(\beta)| + \]
\[ |(n_1n_2)^{-1}(T_1(\beta + c) + T_2(\beta - c)) - \pi(\beta)| \]
\[ \leq |(n_1n_2)^{-1}T_1(\beta - c) - \pi_1(\beta)| + |(n_1n_2)^{-1}T_2(\beta + c) - \pi_2(\beta)| + \]
\[ |(n_1n_2)^{-1}T_1(\beta + c) - \pi_1(\beta)| + |(n_1n_2)^{-1}T_2(\beta - c) - \pi_2(\beta)| \]
\[ \leq |(n_1n_2)^{-1}T_1(\beta - c) - \pi_1(\beta)| + |(n_1n_2)^{-1}T_2(\beta + c) - \pi_2(\beta + c)| + \]
\[ |(n_1n_2)^{-1}T_1(\beta + c) - \pi_1(\beta + c)| + |(n_1n_2)^{-1}T_2(\beta - c) - \pi_2(\beta - c)| + \]
\[ |\pi_1(\beta - c) - \pi_1(\beta)| + |\pi_2(\beta + c) - \pi_2(\beta)| + \]
\[ |\pi_1(\beta + c) - \pi_1(\beta)| + |\pi_2(\beta - c) - \pi_2(\beta)|. \]

Thus, to satisfy conditions (c) and (d) of Lemma 4.4 with
\[ V_n(c) = |(n_1n_2)^{-1}(T_1(\beta - c) + T_2(\beta + c)) - \pi(\beta)| \]
\[ + |(n_1n_2)^{-1}(T_1(\beta + c) + T_2(\beta - c)) - \pi(\beta)| \]
and \( M(c) = |\pi_1(\beta - c) - \pi_1(\beta)| + |\pi_2(\beta + c) - \pi_2(\beta)| \]
\[ + |\pi_1(\beta + c) - \pi_1(\beta)| + |\pi_2(\beta - c) - \pi_2(\beta)|, \]
it suffices to show, for $c > 0$ and any value of $b$, that

$$(n_1n_2)^{-1}T_1(b) \overset{P}{\rightarrow} \pi_1(b)$$

$$\lim_{c \to 0} \pi_1(b \pm c) = \pi_1(b).$$

The first result is proved in lemma 4.5 and the second result is proved in lemma 4.6.

**Lemma 4.5.** Under conditions (4.1)-(4.8), (4.10), (4.11), and $\min(n^i, n_2) \to \infty$, for $i = 1, 2$, and any value of $b$

$$(n_1n_2)^{-1}T_1(b) \overset{P}{\rightarrow} \pi_1(b)$$

**Proof:** Only the proof for $i = 1$ is given because the proof for $i = 2$ is very analogous. Because

$$E((n_1n_2)^{-1}T_1(b)) = \pi_1(b)$$

it suffices to show that

$$\text{var}((n_1n_2)^{-1}T_1(b)) \to 0 \quad \text{as} \quad \min(n_1, n_2) \to \infty.$$
where

\[ \pi_{11}(b) = P(Z_{2m}(b) < Z_{1k}(b) \text{ and } Z_{2q}(b) < Z_{1k}(b) \text{ and } |X_{2m}| < |X_{1k}| \text{ and } |X_{2q}| < |X_{1k}|) \]

\[ \pi_{12}(b) = P(Z_{2m}(b) < Z_{1k}(b) \text{ and } Z_{2m}(b) < Z_{1p}(b) \text{ and } |X_{2m}| < |X_{1k}| \text{ and } |X_{2m}| < |X_{1p}|) \].

Thus

\[ \text{var}(T_1(b)) = n_1n_2\pi_1(b)(1 - \pi_1(b)) \]

\[ + n_1n_2(n_2 - 1)(\pi_{11}(b) - \pi_1^2(b)) \]

\[ + n_2n_1(n_1 - 1)(\pi_{12}(b) - \pi_1^2(b)) \]

\[ = n_1n_2(\pi_1(b) - \pi_1^2(b))(n_1 + n_2 - 1) \]

\[ + (n_2 - 1)\pi_{11}(b) + (n_1 - 1)\pi_{12}(b)) \]

Clearly then \( \text{var}(T_1(b)/n_1n_2) \to 0 \) as \( \min(n_1, n_2) \to \infty \) and the proof is complete.

For each value of \( b \), define the sets

\[ A(b) = \{(Y_1, X_1, Y_2, X_2) : (Y_2 - bX_2) \ sgn \ X_2 < (Y_1 - bX_1) \ sgn \ X_1 \]

and \( |X_2| < |X_1| \} \]

\[ B(b) = \{(Y_1, X_1, Y_2, X_2) : (Y_2 - bX_2) \ sgn \ X_2 < (Y_1 - bX_1) \ sgn \ X_1 \]

and \( |X_2| > |X_1| \} \).
The next lemma shows that if \( b_1 < b_2 \), then
\[
A(b_1) \supseteq A(b_2) \\
B(b_1) \subseteq B(b_2).
\]

**Lemma 4.6.** Let \( b_1 < b_2 \) and assume that the joint distribution of \((Y_1, X_1, Y_2, X_2)\) is absolutely continuous with respect to the Lebesque measure on four dimensional Euclidean space. Then

(a) \( A(b_1) \supseteq A(b_2) \)

(b) If \( \{c_n\} \) is any monotonically decreasing sequence of positive numbers so that \( \lim_{n \to \infty} c_n = 0 \), then for any value of \( b \)
\[
\lim_{n \to \infty} P(A(b \pm c_n)) = P(A(b)) \\
\lim_{n \to \infty} P(B(b \pm c_n)) = P(B(b))
\]

**Proof:** (a) Let \((Y_1, X_1, Y_2, X_2) \in A(b_2)\). Then \( |X_2| < |X_1| \) and

\[
Y_2 \sgn X_2 < Y_1 \sgn X_1 - b_2(|X_1| - |X_2|) \\
< Y_1 \sgn X_1 - b_1(|X_1| - |X_2|)
\]

and thus \((Y_1, X_1, Y_2, X_2) \in A(b_1)\). In a very analogous manner
\( B(b_1) \subseteq B(b_2) \).
(b) From (a) it follows that

\[ A(b + c_n) \subseteq A(b + c_{n+1}) \quad \text{and} \]

\[ \lim_{n \to \infty} P(A(b + c_n)) = P(\bigcup_{n=1}^{\infty} A(b + c_n)). \]

Clearly \( A(b) \supseteq \bigcup_{n=1}^{\infty} A(b + c_n) \). Let \((Y_1, X_1, Y_2, X_2)\) be an arbitrary element in \(A(b)\). Then

\[ Y_2 \sgn X_2 < Y_1 \sgn X_1 - b(|X_1| - |X_2|). \tag{4.18} \]

Expression (4.18) can be rewritten as

\[ b < (Y_1 \sgn X_1 - Y_2 \sgn X_2)/(|X_1| - |X_2|). \]

Let \( c_n^* \) be any element in the sequence \( \{c_n\} \) so that

\[ b < b + c_n^* < (Y_1 \sgn X_1 - Y_2 \sgn X_2)/(|X_1| - |X_2|). \tag{4.19} \]

Multiplying by \(|X_1| - |X_2|\), the expression in (4.19) becomes

\[ Y_2 \sgn X_2 - (b + c_n^*)|X_2| < Y_1 \sgn X_1 - (b + c_n^*)|X_1| \]

and consequently

\[ (Y_1, X_1, Y_2, X_2) \in A(b + c_n^*) \subseteq \bigcup_{n=1}^{\infty} A(b + c_n). \]

Thus

\[ A(b) = \bigcup_{n=1}^{\infty} A(b + c_n) \]
and

\[ \lim_{n \to \infty} P(A(b + c_n)) = P(A(b)). \]

Also from (a) it follows that

\[ A(b - c_n) \supset A(b - c_{n+1}) \]

and that

\[ \lim_{n \to \infty} P(A(b - c_n)) = P(\bigcap_{n=1}^{\infty} A(b - c_n)). \]

Let \((Y_1, X_1, Y_2, X_2) \in \bigcap_{n=1}^{\infty} A(b - c_n)\). Then for every \(n\),

\[ Y_2 \sgn X_2 < Y_1 \sgn X_1 - \frac{(b - c_n)(|X_1| - |X_2|)}{n}\]

\[ < Y_1 \sgn X_1 - b(|X_1| - |X_2|) + c_n(|X_1| - |X_2|). \]

Since \(c_n \geq 0\), \(|X_1| - |X_2| > 0\), \(\lim_{n \to \infty} c_n = 0\), it follows that

\[ Y_2 \sgn X_2 \leq Y_1 \sgn X_1 - b(|X_1| - |X_2|). \]

This implies that \((Y_1, X_1, Y_2, X_2) \in A(b) \cup \{(Y_1, X_1, Y_2, X_2): \]

\[ Y_2 \sgn X_2 = Y_1 \sgn X_1 + b(|X_1| - |X_2|) \quad \text{and} \quad |X_1| > |X_2|) \]

\[ = A(b) \cup E(b). \]

Consequently \(\bigcap_{n=1}^{\infty} A(b - c_n) \subseteq A(b) \cup E(b)\). Very clearly \(A(b) \cup E(b) \subseteq \)

\[ \bigcap_{n=1}^{\infty} A(b - c_n) \quad \text{and thus} \]

\[(4.20) \quad \bigcap_{n=1}^{\infty} A(b - c_n) = A(b) \cup E(b). \]
Since the underlying joint distribution is absolutely continuous,

\[ P(\bigcap_{n=1}^{\infty} A(b - c_n)) = P(A(b)) + P(E(b)) = P(A(b)). \]

Exactly the same technique will show that

\[ \lim_{n \to \infty} P(B(b \pm c_n)) = P(B(b)) \]

and the proof is complete.

It follows directly from the last lemma that

\[ \pi_1(\beta \pm c) + \pi_1(\beta) \quad \text{as} \quad c \to 0. \]

The next theorem contains the consistency class for different alternatives. The proof of the theorem is contained in the last six lemmas.

**Theorem 4.1.** For conditions (4.1)-(4.3),(4.5)-(4.9) and \( \min(n_1,n_2) \to \infty \) the consistency class of each type of test (as specified by a specific critical region) is given in the following table. The expression in the parenthesis is the alternative in terms of \( \beta_1 \) and \( \beta_2 \). The values \( k_\gamma(n_1,n_2) \) are such that

\[
P(S \leq k_\gamma(n_1,n_2)) = \gamma
\]

where \( S \) has the null distribution of the Mann-Whitney form of the Mann-Whitney-Wilcoxon statistic.

<table>
<thead>
<tr>
<th>Consistency Class</th>
<th>Critical Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_{1}(z) \leq G_{2}(z) \quad (\beta_1 &gt; \beta_2) )</td>
<td>( \hat{S}<em>n &gt; k</em>{1-a}(n_1,n_2) )</td>
</tr>
<tr>
<td>( G_{1}(z) \geq G_{2}(z) \quad (\beta_1 &lt; \beta_2) )</td>
<td>( \hat{S}_n \leq k_a(n_1,n_2) )</td>
</tr>
<tr>
<td>( G_{1}(z) \neq G_{2}(z) \quad (\beta_1 \neq \beta_2) )</td>
<td>Either ( \hat{S}<em>n &gt; k</em>{1-a}(n_1,n_2) ) or ( \hat{S}<em>n \leq k</em>{a_2}(n_1,n_2) ) where ( a = a_1 + a_2 ).</td>
</tr>
</tbody>
</table>
4.4. Consistency of the Jonckheere Test

For an arbitrary \( k \), let \( \{(Y_{ij}, X_{ij}) : 1 \leq i \leq k, 1 \leq j \leq n_i\} \) be \( k \) bivariate samples each from a population with joint density function as defined in (4.1) and (4.2). Assume \( \lim n_i/(n_1 + \ldots + n_k) = \lambda_1 \).

Let \( Z_{ij}(b), R_{ij}(b) \) and \( G^b(z) \) be defined as in (4.3), (4.4), and (4.5) respectively. Assume that the alternative of interest is the ordered alternative

\[
(4.19) \quad K: \beta_1 \leq \beta_2 \leq \ldots \leq \beta_k
\]

with at least one strict inequality. It follows from lemma 4.1 that a more general alternative that includes (4.19) is for any \( b \)

\[
(4.20) \quad G^b_1(z) \geq G^b_2(z) \geq \ldots \geq G^b_k(z)
\]

with at least one strict inequality for at least one value of \( z \). For each pair \( (i, i') \) so that \( i < i' \), define for \( 1 \leq k \leq n_i \), \( 1 \leq m \leq n_i \)

\[
(4.21) \quad D_{km}(b) = 1 \quad \text{if} \quad Z_{im}(b) < Z_{i'k}(b)
\]

\[
= 0 \quad \text{if} \quad Z_{im}(b) > Z_{i'k}(b).
\]

Let

\[
S_{ii'}(b) = \sum_{k=1}^{n_i} \sum_{m=1}^{n_i} D_{km}(b).
\]

Finally define the Jonckheere statistic as

\[
S_n(b) = \sum_{1 \leq i < i' \leq k} S_{ii'}(b).
\]
where \( n = \sum_{i=1}^{k} n_i \). If for each \( i \), \( \beta_i = \beta + \theta_i \), then let \( \hat{\beta}_n \) be any estimator of \( \beta \) based on the combined sample. Let

\[
\hat{S}_n = S_n(\hat{\beta}).
\]

In order to apply the results from the last section, these additional definitions are made. Define for each pair \((i,i')\), \( i < i' \) and \((k,m)\), \( 1 \leq k \leq n_i \), \( 1 \leq m \leq n_i \)

\[
\begin{align*}
ii',Q_{km} = 1 & \quad \text{if } |X_{im}| < |X_{i'k}| \\
& = 0 \quad \text{if } |X_{im}| > |X_{i'k}|
\end{align*}
\]

Define

\[
\begin{align*}
ii',T_1(b) = & \sum_{k=1}^{n_i} \sum_{m=1}^{n_i} ii',D_{km}(b) ii',Q_{km} \\
ii',T_2(b) = & \sum_{k=1}^{n_i} \sum_{m=1}^{n_i} ii',D_{km}(b)(1 - ii',Q_{km}).
\end{align*}
\]

Clearly then

\[
S_n(b) = \sum_{1 \leq i < i' \leq k} \sum_{1 \leq i < i' \leq k} ii',T_1(b) + \sum_{1 \leq i < i' \leq k} ii',T_2(b).
\]

In the same manner define

\[
\begin{align*}
ii',\pi_1(b) = & \mathbb{P}(Z_{im}(b) < Z_{i'k}(b) \text{ and } |X_{im}| < |X_{i'k}|) \\
ii',\pi_2(b) = & \mathbb{P}(Z_{im}(b) < Z_{i'k}(b) \text{ and } |X_{im}| > |X_{i'k}|) \\
ii',\pi(b) = & \mathbb{P}(Z_{im}(b) < Z_{i'k}(b))
\end{align*}
\]
Clearly then

\[ E(S_n(b)) = \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi} n_i n_{i'} i i' \pi(b) \]

\[ = \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi_1} n_i n_{i'} i i' \pi_1(b) + \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi_2} n_i n_{i'} i i' \pi_2(b). \]

It follows from lemma 4.3 that for \( a < b < c \) and

\[ \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi} T_1(c) + \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi} T_2(a) \]

\[ \leq S_n(b) \leq \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi} T_1(a) + \sum_{1 \leq i < i' \leq k} \sum_{i i' \pi} T_2(c). \]

It follows from lemma 4.5 that for \( h = 1, 2 \)

\[ n^{-2} \sum_{i i' \pi} T_h(b) \leq \sum_{i i' \pi} T_h(b) \lambda_i \lambda_i'. \]

From lemma 4.6 it follows that for \( h = 1, 2 \) and any sequence \( \{c_n\} \) of positive numbers such that \( \lim c_n = 0 \)

\[ \sum_{i i' \pi} J_h(b \pm c) \to \sum_{i i' \pi} J_h(b) \quad (as \ n \to \infty). \]

Finally, for any \( c > 0 \),

\[ \sup_{|v| \leq c} \left| n^{-2} S_n(\beta - v) - \sum_{i < i'} \lambda_i \lambda_i' \sum_{i i' \pi} (\beta) \right| \]

\[ \leq \sum_{h=1}^2 \left( \sum_{i < i'} \sum_{n^{-2}} T_h(\beta - c) - \sum_{i i' \pi} \sum_{i i' \pi} (\beta - c) \right) \]

\[ + \sum_{h=1}^2 \left( \sum_{i < i'} \sum_{n^{-2}} T_h(\beta + c) - \sum_{i i' \pi} \sum_{i i' \pi} (\beta + c) \right) \]
It follows from (4.23), (4.24), (4.25) and lemma 4.4 that if \( \hat{\beta}_n \approx \beta \) then

\[
\hat{S}_n \approx \sum_{i < i'} \sum_{h=1}^{2} \lambda_{i} \lambda_{i'} \left( \pi_h(\beta - c) - \pi_h(\beta) \right)
\]

Consequently the consistency class of \( \hat{S}_n \) is the same as that of the usual Jonckheere statistic. These results are summed up in the following theorem.

**Theorem 4.2.** Under the conditions (4.19), \( \hat{\beta}_n \) a consistent estimator of \( \beta \), and \( \lim n_1/n = \lambda_1 \), the test which consists of rejecting \( H_0 \) for large values of \( \hat{S}_n \) is consistent if and only if

\[
\sum_{i < i'} \sum_{h=1}^{2} \lambda_{i} \lambda_{i'} \left( \int G_i(z) g_i(z) dz - 1/2 \right) > 0
\]

under the alternative.

**Remark 4.1.** More work is needed to show the consistency class for the Kruskal-Wallis k-sample test.
4.5. Consistency of $\hat{\beta}_n$

In addition to conditions (4.1) and (4.2), assume that $f(u,v)$ satisfies the conditions

(4.26) $f(u,v) = \ell(u) \cdot h(v)$.

(4.27) $\mu_1 = \int u \ell(u) \, du$ finite, $\mu_2 = \int v h(v) \, dv$ finite,

$\sigma_1^2 = \int (u - \mu_1)^2 \ell(u) \, du < +\infty$, $\sigma_2^2 = \int (v - \mu_2)^2 h(v) \, dv < +\infty$.

The next lemma gives the means and variance-covariance structure of $(Y_{ij}, X_{ij})$ in terms of $\alpha, \beta_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$.

Lemma 4.7. Let $(Y, X)$ have a joint density function $f(y, x) = \ell(y - \alpha - \beta x) h(x)$, where $f(y, x)$ satisfies conditions (4.1), (4.2), (4.26) and (4.27). Then

(a) $E(Y|X) = \mu_1 + \alpha + \beta x$, $\text{var}(Y|X) = \sigma_1^2$.

(b) $\mu_y = E(Y) = \mu_1 + \alpha + \beta \mu_2$, $\mu_x = E(X) = \mu_2$.

(c) $\text{cov}(X, Y) = \sigma_{xy} = \beta \sigma_2^2$.

(d) $\sigma_y^2 = \text{var}(Y) = \sigma_1^2 + \beta^2 \sigma_2^2 < +\infty$,

$\sigma_x^2 = \text{var}(X) = \sigma_2^2 < +\infty$.

Proof: (a) The conditional density of $Y|X$ (form 2.2) is $\ell(y - \alpha - \beta x)$. Then
\[ E(Y|X) = \int y \ell(y - \alpha - \beta x) \, dy \]
\[ = \int (y + \alpha + \beta x) \ell(y) \, dy \]
\[ = \alpha + \beta x + \mu_1. \]

\[ \text{var}(Y|X) = \int (y - \alpha - \beta x - \mu_1)^2 \ell(y - \alpha - \beta x) \, dy \]
\[ = \int (y - \mu_1)^2 \ell(y) \, dy \]
\[ = \sigma_1^2. \]

(b) \[ \mu_Y = E(E(Y|X)) = E(\alpha + \mu_1 + \beta x) \]
\[ = \alpha + \mu_1 + \beta \mu_2. \]

From proposition 2.1 it follows that the density of \( X \) is \( h(x) \). Therefore \( \mu_X = E(X) = \mu_1 \) and \( \text{var}(X) = \sigma_2^2. \)

(c) \[ \text{cov}(X,Y) = \iint (y - (\mu_1 + \alpha + \beta \mu_2))(x - \mu_2) \ell(y - \alpha - \beta x) h(x) \, dydx \]
\[ = \iint (y - \mu_1 + \beta(x - \mu_2))(x - \mu_2) \ell(y) h(x) \, dydx \]
\[ = \iint (y - \mu_1)(x - \mu_2) \ell(y) h(x) \, dydx \]
\[ + \iint \beta(x - \mu_2)^2 \ell(y) h(x) \, dydx \]
\[ = 0 + \beta \sigma_2^2. \]

(d) \[ \text{var}(Y) = \text{var}(E(Y|X)) + E(\text{var}(Y|X)) \]
\[ = \beta^2 \sigma_2^2 + \sigma_1^2. \]
and the proof is completed.

**Remark 4.2.** The expression for $E(Y|X)$ in lemma 4.7 can always be replaced by the median regression line

$$(4.28) \quad m(Y|X) = \alpha + \beta x + m_1,$$

where $m_1$ is the median of the distribution associated with $\xi(v)$.

Obviously if $\xi(v)$ is symmetric, (4.28) and (a) are exactly the same. But if $\mu_1$ does not exist (such as if $\xi(v)$ is the Cauchy density) (a) has no meaning whereas (4.28) still makes sense.

Let the least-squares estimator of $\beta$ be

$$\hat{\beta}_n = 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \frac{(X_{ij} - \bar{X})(Y_{ij} - \bar{Y})}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (X_{ij} - \bar{X})^2}$$

where $\bar{X} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij}/(n_1 + n_2)$, $\bar{Y} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} Y_{ij}/(n_1 + n_2)$.

**Theorem 4.3.** Under conditions (4.1), (4.2), (4.8), (4.26), (4.27),

$n_1$

and for $i=1,2$, $E((\sum_{i=1}^{n_1} (X_i - \bar{X}_i)^2)^{-1}) = 0$ as $n_1 \rightarrow \infty$ where $\bar{X}_i = \sum_{i=1}^{n_1} X_i$

$$\hat{\beta}_n \overset{P}{\rightarrow} \beta = \lambda_1 \beta_1 + \lambda_2 \beta_2,$$

where $\lambda_1 = \lim n_1/(n_1 + n_2)$ as $\min(n_1, n_2) \rightarrow \infty$.

**Proof:** The numerator of $\hat{\beta}_n$ can be rewritten as

$$\sum_{j=1}^{n_1} (X_{1j} - \bar{X})(Y_{1j} - \bar{Y}) + \sum_{j=1}^{n_2} (X_{2j} - \bar{X})(Y_{2j} - \bar{Y})$$
For \( i = 1, 2 \), define

\[
S_i = \left( \frac{\sum (X_{ij} - \bar{X}_1)^2/n_i}{\frac{2}{n_1} \sum (X_{ij} - \bar{X}_1)^2/(n_1 + n_2)} \right).
\]

Then

\[
\hat{\beta}_n = \left( \frac{n_1}{(n_1 + n_2)} \right) \left( \frac{\sum (X_{ij} - \bar{X}_1)(Y_{ij} - \bar{Y}_1)}{\sum (X_{ij} - \bar{X}_1)^2} \right) S_1
\]

\[
+ \left( \frac{n_2}{(n_1 + n_2)} \right) \left( \frac{\sum (X_{ij} - \bar{X}_2)(Y_{ij} - \bar{Y}_2)}{\sum (X_{ij} - \bar{X}_2)^2} \right) S_2
\]

\[
+ \left( \frac{n_1}{(n_1 + n_2)} \right) (\bar{X}_1 - \bar{X})(\bar{Y}_1 - \bar{Y}) \left( \frac{2}{n_1} \sum (X_{ij} - \bar{X})^2/(n_1 + n_2) \right)
\]

\[
+ \left( \frac{n_2}{(n_1 + n_2)} \right) (\bar{X}_2 - \bar{X})(\bar{Y}_2 - \bar{Y}) \left( \frac{2}{n_1} \sum (X_{ij} - \bar{X})^2/(n_1 + n_2) \right).
\]

Because \( \sigma_x^2 < \infty \) and \( \sigma_y^2 < \infty \), it follows that, as \( \min(n_1, n_2) \to \infty \),

\[
S_1 \xrightarrow{p} 1
\]

\[
(\bar{X}_1 - \bar{X}) \xrightarrow{p} 0
\]

\[
(\bar{Y}_1 - \bar{Y}) \xrightarrow{p} (\beta_1 - (\lambda_1 \beta_1 + \lambda_2 \beta_2)) \mu_2
\]
Now, let \( \hat{\beta}_1 \) be the least squares estimator of \( \beta_1 \) based on the 1\(^{st} \) sample only. Then

\[
E(\hat{\beta}_1 \mid (X_{11}, \ldots, X_{in_1})) = \beta_1, \quad \text{and}
\]

\[
\text{var}(\hat{\beta}_1 \mid (X_{11}, \ldots, X_{in_1})) = \sigma_1^2/\sigma_{x_1}^2.
\]

Also

\[
\text{var}(\hat{\beta}_1) = E[\text{var}(\hat{\beta}_1 \mid (X_{11}, \ldots, X_{in_1}))]
\]

\[
+ \text{var}(E(\hat{\beta}_1 \mid (X_{11}, \ldots, X_{in_1})))
\]

\[
= E(\sigma_1^2/\sigma_{x_1}^2) + \text{var}(\hat{\beta}_1)
\]

\[
= E(\sigma_1^2/\sigma_{x_1}^2).
\]

Since \( \sigma_{x_1}^2 = \sum_{i=1}^{n_1} (X_i - \overline{X}_1)^2 \), it follows that as \( \min(n_1, n_2) \to \infty \),

\[\text{var}(\hat{\beta}_1) \to 0\]

and hence \( \hat{\beta}_1 \xrightarrow{p} \beta \). Consequently, \( \hat{\beta}_n \xrightarrow{p} \lambda_1 \beta_1 + \lambda_2 \beta_2 \), and the proof is completed.

Remark 4.3. The same technique will show that \( \hat{\beta}_n \) is a consistent estimator of \( \sum_{i=1}^{k} \lambda_i \beta_i \), for \( k \) arbitrary and \( \lambda_i = \lim n_i/(n_1 + \ldots + n_k) \).
Remark 4.4. More work is needed to show whether the Hodge-Lehmann type estimator or the median type estimator (example 2.3.c and example 2.3.a respectively) are consistent when the X-variables are allowed to be random. It should be noted that even if they are consistent, their stochastic limits need not be $\lambda_1 \beta_1 + \lambda_2 \beta_2$.

Remark 4.5. It is not known whether $\hat{\beta}_n$ is consistent if $\sigma_1^2$ does not exist as with the Cauchy distribution.
Chapter Five

Small Sample Power Study

5.1. Introduction

A Monte Carlo simulation study was performed under a variety of possible underlying joint distributions, of possible sample sizes, and of possible alternatives to compare the power of the two-sample rank-like statistic $S_n$ (1.43) using the Mann-Whitney-Wilcoxon scores with the power of the least squares statistic. Five specific areas were studied. First, under the null hypothesis, the 5% (and 1%) critical values both from the exact null distribution and the approximate null distribution of the Mann-Whitney-Wilcoxon rank statistic were verified to be the 5% (and 1%) critical values for $S_n$. (See Table 1, Table 3). Second, for a fixed sample size ($n_1 = n_2 = 8$) the (estimated) power curves of both $S_n$ and the least squares statistic are presented in Table 2 and Table 3. Third, for a fixed alternative ($\beta_1 - \beta_2 = 1$) the power of $S_n$ is compared with that of the least squares statistic for a variety of sample sizes (8-8, 16-16, 26-16, 16-26, 26-26). (See Table 4) Fourth, some indication of how the value of the "intercept" $\alpha$ affects the power of $S_n$ is presented in Table 5. Fifth, a comparison of the power of the least squares statistic, which is designed to detect two-sided alternatives only, with the power of $S_n$ using only the appropriate one-sided critical region is presented in Table 6.
5.2. Description of the Simulation Technique

Using the idea of remark 2.3, the following lemma indicates how to generate a sample from an underlying bivariate distribution satisfying conditions (1.33) through (1.35).

Lemma 5.1. Let \( U, V, X, \text{ and } Y \) be univariate random variables. If \( U \) and \( V \) are independent and have density functions \( h(u) \) and \( \ell(v) \), respectively, then the joint density function of \( (Y,X) \), where

\[
(5.1) \quad Y = V + \beta \cdot U + \alpha \quad \text{and} \quad X = U
\]
is

\[
(5.2) \quad f(y,x) = \ell(y - \alpha - \beta x) \cdot h(x)
\]

Proof. The joint density function of \( (U,V) \) is \( h(u) \cdot \ell(v) \). The inverse transformation is

\[
U = X \quad \text{and} \quad V = Y - \beta X - \alpha.
\]
The Jacobian is 1 and if \( f(y,x) \) is the joint density of \( (Y,X) \), then

\[
 f(y,x) = \ell(y - \beta x - \alpha) \cdot h(x)
\]
and the proof is completed.

Assume \( \alpha = 0, \beta_1 = \beta \text{ and } \beta_2 = 0 \). For any two density functions \( \ell(v), h(u) \), two independent random samples \( V_1, \ldots, V_N \) and \( U_1, \ldots, U_N \), where \( n_1 + n_2 = N \), can be generated. Then define the first sample as follows: for \( 1 \leq j \leq n_1 \)

\[
(5.3) \quad Y_{1j} = V_j + \beta U_j \quad \text{and} \quad X_{1j} = U_j.
\]
By lemma 5.1, the joint density of \( (Y_{1j}, X_{1j}) \) is

\[
(5.4) \quad f_1(y,x) = f(y - \beta x, x) = \ell(y - \beta x) \cdot h(x).
\]
In a similar fashion, the second sample is defined: for \( 1 \leq j \leq n_2 \)
The joint density of \((Y_{2j}, X_{2j})\) is
\[
f_2(y, x) = f(y - 0\cdot x, x) = f(y) \cdot h(x).
\]

The estimator for the common value of the slope was the usual least squares estimator based on the combined sample. An informal preliminary study was made using also the median estimator discussed in example 2.3.a. Because the two estimators appeared to perform very similarly and because the median estimator was a very inefficient estimator in terms of computer time, only the least squares estimator was used in the final computations.

The rank-like statistic \(\hat{S}_n\) is the Mann-Whitney-Wilcoxon rank statistic applied to the ranks of the signed residuals
\[
(Y_{ij} - \hat{\beta}X_{ij}) \text{ sgn } X_{ij}.
\]
That is
\[
\hat{S}_n = \sum_{j=1}^{n_1} \hat{R}_{ij}, \quad \text{where } 1 \leq j \leq n_1
\]
\[
\hat{R}_{ij} = \text{rank}((Y_{ij} - \hat{\beta}X_{ij}) \text{ sgn } X_{ij})
\]
in the combined sample. Because the least square statistic is designed to detect two-sided alternatives only, a two-sided symmetric critical region was used with \(\hat{S}_n\) (except for table 6).

The least squares statistic, \(T\), is of the form
\[
T = ((\text{SSH} - \text{SSE})/1)/((\text{SSE}/N-3))
\]
where \(\text{SSE}\) is the minimum of
\[
2 \sum_{i=1}^{n_1} \sum_{j=1}^{13} \left(Y_{ij} - \alpha - \beta_iX_{ij}\right)^2,
\]
and where SSH is the minimum of

\[ 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (Y_{ij} - \alpha - \beta X_{ij})^2. \]

Because the actual calculation of T does not depend on whether the "independent" variables are fixed constants or random variables, the actual form of SSE (and of SSH) can be found in Draper-Smith (1966), problem 2.D.b. For the sake of convenience,

\[
\text{SSE} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} Y_{ij}^2 - \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} b_{ij}^2 X_{ij}^2 + a^2 N - 2a \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} Y_{ij}^2;
\]

\[
a = \frac{(\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} Y_{ij} - \sum_{j=1}^{n_1} X_{ij} (\sum_{j=1}^{n_1} X_{ij} Y_{ij} / \sum_{j=1}^{n_1} X_{ij}^2))}{(N - \sum_{j=1}^{n_1} ((\sum_{j=1}^{n_1} X_{ij})^2 / \sum_{j=1}^{n_1} X_{ij}^2))};
\]

\[
b_i = \frac{(\sum_{j=1}^{n_1} X_{ij} Y_{ij} - a \sum_{j=1}^{n_1} X_{ij}) / \sum_{j=1}^{n_1} X_{ij}^2;}
\]

\[
\text{SSH} = (\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} Y_{ij}^2 - (\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} Y_{ij}/N)^2) - b(\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij} Y_{ij} - (\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij} Y_{ij}/N));
\]

\[
b = (\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij} Y_{ij} - (\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij} E Y_{ij}/N))/
\]

\[
(\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij}^2 - ((\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} X_{ij})^2 / N)).
\]
When the "independent" variables are fixed known constants and the
"dependent" variables are normally distributed, the null distribution
of \( T \) is an \( F \)-distribution with degrees of freedom \((1,N-3)\). Because
the null distribution does not depend on the actual values of the "in-
dependent" variables, the null distribution of \( T \) is still an \( F \)-dis-
tribution as long as the distribution of \( Y|X \) is normal. In any case,
the critical values of an \( F \)-distribution with \((1,N-3)\) degrees of free-
dom were used for each possible underlying joint distribution.

The different underlying joint distributions were specified in the
following manner using lemma 5.1. For \( 1 \leq i \leq 2 \), and for the nine
pairs \((k,m)\), \( 1 \leq k,m \leq 3 \), define

\[
\xi_{i}^{km}(y,x) = \xi_{m}(y - \beta_{i}x) \cdot h_{k}(x)
\]

where \( \beta_{1} = \beta \) and \( \beta_{2} = 0 \) and \( h_{k}(u) \) and \( \xi_{m}(v) \) are defined so as
to satisfy the conditions (2.4) through (2.7) of remark 2.3. The fol-
lowing is a list of \( h_{k}(u) \) and \( \xi_{m}(v) \).

For \( k = 1 \) (truncated normal)

\[
h_{1}(u) = \begin{cases} 
\frac{(\phi(3) - \phi(-3))^{-1}(2\pi)^{-1/2}\exp(-u^2/2)}{0} & -3 \leq u \leq 3 \\
0 & \text{otherwise}
\end{cases}
\]

where \( \phi(x) \) is the standard normal distribution function.

For \( k = 2 \) (truncated exponential)

\[
h_{2}(u) = \begin{cases} 
(E(3.7))^{-1}\exp(-u) & 0 \leq u \leq 3.7 \\
0 & \text{otherwise}
\end{cases}
\]

where \( E(x) = 1 - \exp(-x) \), the exponential distribution function.
For \( k = 3 \) (uniform \((-2,2))

\[
h_3(u) = \begin{cases} 
1/4 & \text{if } -2 < u < 2 \\
0 & \text{otherwise.}
\end{cases}
\]

For \( m = 1 \) (normal distribution)

\[
x_1(v) = (2\pi)^{-1/2} \exp(-v^2/2) \quad -\infty < v < +\infty.
\]

For \( m = 2 \) (Cauchy distribution)

\[
x_2(v) = \pi^{-1}(1 + v^2)^{-1} \quad -\infty < v < +\infty.
\]

For \( m = 3 \) (shifted exponential)

\[
x_3(v) = \begin{cases} 
\exp(-(v + .6931471)) & \text{if } -.6931471 < v < +\infty \\
0 & \text{otherwise.}
\end{cases}
\]

Note that the median of the distribution with density \( x_3(v) \) is zero since the median of a standard exponential is .6931471. Using example 2.2, for \( k = m = 1 \), \( x_1(y,x) \) represents a truncated bivariate normal distribution.

The last item to describe is how the two independent random samples used in (5.3) and (5.5) \( U_1, \ldots, U_N \) and \( V_1, \ldots, V_N \) were generated for each possible pair \((k,m)\). First a sequence of uniform random deviates between zero and one, \( W_1, \ldots, W_n \), were generated either by the Temple University CDC 6400 function RANF(.) (for Table 1, Table 4, Table 6) or by The Ohio State University IBM 370 function UNI(.) (for Table and Table 3). Then the following transformations were used (Fishman (1973)). If \( W_1, W_2 \) are uniform random deviates, then \( V_1, V_2 \) are standard normal deviates where
\[
V_1 = (-2 \log W_1)^{1/2} \sin(2\pi W_2)
\]
\[
V_2 = (-2 \log W_1)^{1/2} \cos(2\pi W_2).
\]

For a standard exponential distribution

\[ V = -\log W \]

has an exponential distribution. For a Cauchy distribution,

\[ V = \tan(\pi(W - .5)) \]

has a Cauchy distribution. For a truncated normal (truncated exponential) distribution, use only those \( V \)'s so that \( |V| \leq 3 \) \((0 \leq |V| \leq 3.7)\). For a shifted exponential distribution,

\[ V = -\log W - .6931471 \]

has an exponential distribution with median equal zero. For an uniform distribution over the interval \((-2,2)\), let

\[ V = 4W - 2. \]

5.3. Results

Table 1 presents a verification that the 5% two-sided critical values for the usual Mann-Whitney-Wilcoxon and the usual least squares statistic whose null distribution is an F-distribution with \((1,N-3)\) degrees of freedom give the same level of significance for \( \hat{S}_n \) and \( T \). For sample size 8-8, the exact Mann-Whitney-Wilcoxon null distribution was used (Hollander-Wolfe). For all other sample sizes the normal approximation to the null distribution was used. Across the top of the table, the values of \( m = 1,2,3 \) correspond to the normal distribution \( \mathcal{N}_1(v) \), the Cauchy distribution \( \mathcal{C}_2(v) \) and the shifted
exponential distribution \((L_3(v))\), respectively. Down the side of the table, the values of \(k = 1, 2, 3\) correspond to the truncated normal distribution \((h_1(u))\), the truncated exponential distribution \((h_2(u))\) and the uniform \((-2, 2)\) distribution \((h_3(u))\), respectively. The body of the table contains the estimated level of significance for both \(T\) and \(\hat{S}_n\) for three different sample sizes.

Table 2 presents a comparison of the power curve of \(\hat{S}_n\) with the power curve of \(T\) for a single sample size \(n_1 = n_2 = 8\) and for all nine underlying joint densities. Across the top of the table are the different values of the difference \(\beta = \beta_1 - \beta_2\). Down the side of the table are the different combinations of densities. In the body of the table are the estimates of the power of \(T\), of \(\hat{S}_n\), and of the ratio (power of \(\hat{S}_n\)/power of \(T\))(\((\hat{S}_n/T)\)).

Table 3 presents both a verification of the 1% critical values for \(\hat{S}_n\) and \(T\) for \(n_1 = n_2 = 8\) (0.00 column) and a comparison of the power of \(\hat{S}_n\) and of \(T\) for several different alternatives. The format of the table is exactly the same as that of Table 2.

Table 4 presents a comparison of power of \(\hat{S}_n\) and of \(T\) for several different sample sizes at a single alternative \(\beta_1 - \beta_2 = 1.00\) and the level of significance equal to .05. The format of the table is the same as that of Table 1.

Table 5 presents the comparison of the power of \(T\) and of \(\hat{S}_n\) when the "intercept" \(\alpha\) is zero and when the "intercept" \(\alpha\) is .6931471 (median of the standard exponential distribution). The "dependent" variable was allowed only to have either the shifted exponential
distribution ("intercept" equal zero) or the standard exponential distribution ("intercept" equal .6931471). The "independent" variable was allowed to have all three possible distributions, namely $h_1(u)$, $h_2(u)$, and $h_3(u)$. Only a sample size of $n_1 = n_2 = 8$ and level of significance equal to .05 was used.

Table 6 presents a comparison of the power of $T$ and of $\hat{S}_n^*$, where $\hat{S}_n^*$ is the rank-like statistics geared to detect the one-sided alternative $\beta_1 > \beta_2$. The level of significance was .05, and only the sample sizes 8-8, 16-16, and 26-26 were used, as well as only the alternative $\beta_1 - \beta_2 = 1.00$.

5.4. Discussion of Results

Three observations can be made about the results in Table 1. First, for sample size 8-8, the level of significance for $\hat{S}_n$ was guaranteed to be 5% by Theorem 2.2 which stated that the null distribution of $\hat{S}_n$ is the same as the null distribution of the Mann-Whitney-Wilcoxon. The only case in which the estimated level differed significantly from .05 is for $k = 2$, $m = 1$. Secondly, for the other sample sizes, it is clear that the normal approximation to the Mann-Whitney-Wilcoxon can also be used for $\hat{S}_n$. This approximation appeared to fare the worst for $k = 2$, $m = 1$, $n_1 = n_2 = 16$, $n_1 = n_2 = 26$ and for $k = 2$, $m = 3$, $n_1 = n_2 = 16$. Thirdly, for the cases when the "dependent" variable was non-normally distributed ($m = 2$ and $m = 3$), the critical values from the $F$-distribution were still fairly accurate. As would be expected, the greatest errors occurred for $m = 2$ (the Cauchy distribution).
## TABLE 1.

Prob (type I error) verification

Level of significance = .05\(^a\)

<table>
<thead>
<tr>
<th>(k)</th>
<th>(n_1 - n_2)</th>
<th>1 (NORMAL)</th>
<th>2 (CAUCHY)</th>
<th>3 (EXPONENTIAL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(T) (\hat{S}_n)</td>
<td>(T) (\hat{S}_n)</td>
<td>(T) (\hat{S}_n)</td>
<td>(T) (\hat{S}_n)</td>
</tr>
<tr>
<td>1</td>
<td>8-8</td>
<td>.049 .047</td>
<td>.057 .049</td>
<td>.053 .057</td>
</tr>
<tr>
<td></td>
<td>16-16</td>
<td>.051 .048</td>
<td>.054 .048</td>
<td>.055 .047</td>
</tr>
<tr>
<td></td>
<td>26-26(^b)</td>
<td>.044 .046</td>
<td>.062 .049</td>
<td>.051 .041</td>
</tr>
<tr>
<td>2</td>
<td>8-8</td>
<td>.064 .075(^c)</td>
<td>.064 .051</td>
<td>.056 .047</td>
</tr>
<tr>
<td></td>
<td>16-16</td>
<td>.074 .089(^c)</td>
<td>.060 .047</td>
<td>.047 .038(^c)</td>
</tr>
<tr>
<td></td>
<td>26-26(^b)</td>
<td>.067 .083(^c)</td>
<td>.064 .048</td>
<td>.053 .054</td>
</tr>
<tr>
<td>3</td>
<td>8-8</td>
<td>.050 .052</td>
<td>.045 .048</td>
<td>.046 .048</td>
</tr>
<tr>
<td></td>
<td>16-16</td>
<td>.050 .045</td>
<td>.038 .045</td>
<td>.050 .047</td>
</tr>
<tr>
<td></td>
<td>26-26(^b)</td>
<td>.047 .048</td>
<td>.034 .046</td>
<td>.050 .048</td>
</tr>
</tbody>
</table>

\(^a\) Except for sample sizes denoted by (b), all estimates are based upon 8000 iterations. Maximum error is .005 (using a 95% confidence interval).

\(^b\) Estimates for this sample size are based upon 2000 iterations. Maximum error is .01

\(^c\) Indicates a variation from a nominal 5% level of significance.
TABLE 2

Power Curve $n_1 = n_2 = 8$

Significance level = .05  8000 iterations

<table>
<thead>
<tr>
<th>$\beta_1 - \beta_2$</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-m</td>
<td>T</td>
<td>$\hat{S}_n$</td>
<td>$\hat{S}_n/T$</td>
<td>T</td>
</tr>
<tr>
<td>1-1</td>
<td>.070</td>
<td>.065</td>
<td>.92</td>
<td>.131</td>
</tr>
<tr>
<td>1-2</td>
<td>.063</td>
<td>.055</td>
<td>.87</td>
<td>.070</td>
</tr>
<tr>
<td>1-3</td>
<td>.080</td>
<td>.075</td>
<td>.94</td>
<td>.164</td>
</tr>
<tr>
<td>2-1</td>
<td>.117</td>
<td>.112</td>
<td>.96</td>
<td>.218</td>
</tr>
<tr>
<td>2-2</td>
<td>.067</td>
<td>.053</td>
<td>.80</td>
<td>.074</td>
</tr>
<tr>
<td>2-3</td>
<td>.099</td>
<td>.087</td>
<td>.88</td>
<td>.203</td>
</tr>
<tr>
<td>3-1</td>
<td>.078</td>
<td>.073</td>
<td>.93</td>
<td>.174</td>
</tr>
<tr>
<td>3-2</td>
<td>.049</td>
<td>.057</td>
<td>1.15</td>
<td>.055</td>
</tr>
<tr>
<td>3-3</td>
<td>.094</td>
<td>.093</td>
<td>.99</td>
<td>.213</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta_1 - \beta_2$</th>
<th>1.25</th>
<th>1.50</th>
<th>2.00</th>
<th>2.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-m</td>
<td>T</td>
<td>$\hat{S}_n$</td>
<td>$\hat{S}_n/T$</td>
<td>T</td>
</tr>
<tr>
<td>1-1</td>
<td>.522</td>
<td>.352</td>
<td>.68</td>
<td>.655</td>
</tr>
<tr>
<td>1-2</td>
<td>.111</td>
<td>.126</td>
<td>1.14</td>
<td>.125</td>
</tr>
<tr>
<td>1-3</td>
<td>.592</td>
<td>.489</td>
<td>.83</td>
<td>.703</td>
</tr>
<tr>
<td>2-1</td>
<td>.682</td>
<td>.445</td>
<td>.65</td>
<td>.790</td>
</tr>
<tr>
<td>2-2</td>
<td>.138</td>
<td>.134</td>
<td>.98</td>
<td>.157</td>
</tr>
<tr>
<td>2-3</td>
<td>.683</td>
<td>.501</td>
<td>.73</td>
<td>.784</td>
</tr>
<tr>
<td>3-1</td>
<td>.674</td>
<td>.527</td>
<td>.78</td>
<td>.807</td>
</tr>
<tr>
<td>3-2</td>
<td>.120</td>
<td>.177</td>
<td>1.48</td>
<td>.154</td>
</tr>
<tr>
<td>3-3</td>
<td>.729</td>
<td>.655</td>
<td>.90</td>
<td>.834</td>
</tr>
</tbody>
</table>
TABLE 2' (CONTINUED)

<table>
<thead>
<tr>
<th>$\beta_1 - \beta_2$</th>
<th>3.00</th>
<th>4.00</th>
<th>5.00</th>
<th>10.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-m</td>
<td>$T$</td>
<td>$\hat{S}_n$</td>
<td>$\hat{S}_n / T$</td>
<td>$T$</td>
</tr>
<tr>
<td>1-1</td>
<td>.982</td>
<td>.871</td>
<td>.89</td>
<td>.996</td>
</tr>
<tr>
<td>1-2</td>
<td>.290</td>
<td>.350</td>
<td>1.21</td>
<td>.370</td>
</tr>
<tr>
<td>1-3</td>
<td>.972</td>
<td>.903</td>
<td>.93</td>
<td>.994</td>
</tr>
<tr>
<td>2-1</td>
<td>.985</td>
<td>.845</td>
<td>.86</td>
<td>.996</td>
</tr>
<tr>
<td>2-2</td>
<td>.339</td>
<td>.330</td>
<td>.97</td>
<td>.448</td>
</tr>
<tr>
<td>2-3</td>
<td>.978</td>
<td>.867</td>
<td>.89</td>
<td>.996</td>
</tr>
<tr>
<td>3-1</td>
<td>.998</td>
<td>.975</td>
<td>.98</td>
<td>1.00</td>
</tr>
<tr>
<td>3-2</td>
<td>.347</td>
<td>.470</td>
<td>1.36</td>
<td>.461</td>
</tr>
<tr>
<td>3-3</td>
<td>.993</td>
<td>.970</td>
<td>.98</td>
<td>.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta_1 - \beta_2$</th>
<th>15.00</th>
<th>20.00</th>
<th>25.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-m</td>
<td>$T$</td>
<td>$\hat{S}_n$</td>
<td>$\hat{S}_n / T$</td>
</tr>
<tr>
<td>1-2</td>
<td>.789</td>
<td>.881</td>
<td>1.12</td>
</tr>
<tr>
<td>2-2</td>
<td>.820</td>
<td>.836</td>
<td>1.02</td>
</tr>
<tr>
<td>3-2</td>
<td>.824</td>
<td>.923</td>
<td>1.12</td>
</tr>
</tbody>
</table>

^a Maximum error is .012.
TABLE 3

Power Curve $n_1 = n_2 = 8$

Significance level = .01 8000 iterations

<table>
<thead>
<tr>
<th>$\beta_1 - \beta_2$</th>
<th>0.00</th>
<th>.50</th>
<th>1.00</th>
<th>1.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-m T S_n S_n/T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-1</td>
<td>.010</td>
<td>.039</td>
<td>.159</td>
<td>.393</td>
</tr>
<tr>
<td>1-2</td>
<td>.014</td>
<td>.020</td>
<td>.027</td>
<td>.044</td>
</tr>
<tr>
<td>1-3</td>
<td>.013</td>
<td>.057</td>
<td>.234</td>
<td>.484</td>
</tr>
<tr>
<td>2-1</td>
<td>.014</td>
<td>.081</td>
<td>.295</td>
<td>.586</td>
</tr>
<tr>
<td>2-2</td>
<td>.026</td>
<td>.028</td>
<td>.041</td>
<td>.066</td>
</tr>
<tr>
<td>2-3</td>
<td>.015</td>
<td>.086</td>
<td>.317</td>
<td>.663</td>
</tr>
<tr>
<td>3-1</td>
<td>.010</td>
<td>.052</td>
<td>.237</td>
<td>.570</td>
</tr>
<tr>
<td>3-2</td>
<td>.008</td>
<td>.018</td>
<td>.023</td>
<td>.054</td>
</tr>
<tr>
<td>3-3</td>
<td>.012</td>
<td>.079</td>
<td>.332</td>
<td>.644</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta_1 - \beta_2$</th>
<th>2.00</th>
<th>2.50</th>
<th>3.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-m T S_n S_n/T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-1</td>
<td>.649</td>
<td>.820</td>
<td>.923</td>
</tr>
<tr>
<td>1-2</td>
<td>.074</td>
<td>.100</td>
<td>.148</td>
</tr>
<tr>
<td>1-3</td>
<td>.699</td>
<td>.831</td>
<td>.919</td>
</tr>
<tr>
<td>2-1</td>
<td>.802</td>
<td>.907</td>
<td>.956</td>
</tr>
<tr>
<td>2-2</td>
<td>.103</td>
<td>.152</td>
<td>.206</td>
</tr>
<tr>
<td>2-3</td>
<td>.786</td>
<td>.893</td>
<td>.942</td>
</tr>
<tr>
<td>3-1</td>
<td>.892</td>
<td>.946</td>
<td>.988</td>
</tr>
<tr>
<td>3-2</td>
<td>.091</td>
<td>.144</td>
<td>.198</td>
</tr>
<tr>
<td>3-3</td>
<td>.840</td>
<td>.936</td>
<td>.975</td>
</tr>
</tbody>
</table>

* Maximum error is .012.
### TABLE 4

Power Comparison

Significance level = .05<sup>a</sup>

Different sample sizes. Single alternative: $\beta_1 = 1, \beta_2 = 0$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$n_1 - n_2$</th>
<th>$m$</th>
<th>$\hat{S}_n$</th>
<th>$\hat{S}_n/T$</th>
<th>$\hat{S}_n$</th>
<th>$\hat{S}_n/T$</th>
<th>$\hat{S}_n$</th>
<th>$\hat{S}_n/T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2 (CAUCHY)</td>
<td>3 (EXPONENTIAL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-8</td>
<td>.369</td>
<td>.258</td>
<td>.70</td>
<td>.088</td>
<td>.099</td>
<td>1.12</td>
<td>.450</td>
<td>.370</td>
</tr>
<tr>
<td>16-16</td>
<td>.692</td>
<td>.467</td>
<td>.68</td>
<td>.092</td>
<td>.150</td>
<td>1.63</td>
<td>.733</td>
<td>.689</td>
</tr>
<tr>
<td>26-16</td>
<td>.789</td>
<td>.595</td>
<td>.75</td>
<td>.093</td>
<td>.187</td>
<td>2.01</td>
<td>.821</td>
<td>.775</td>
</tr>
<tr>
<td>16-26</td>
<td>.797</td>
<td>.590</td>
<td>.74</td>
<td>.095</td>
<td>.187</td>
<td>1.96</td>
<td>.823</td>
<td>.774</td>
</tr>
<tr>
<td>26-26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.900</td>
<td>.722</td>
<td>.80</td>
<td>.115</td>
<td>.234</td>
<td>2.05</td>
<td>.911</td>
<td>.894</td>
</tr>
<tr>
<td>8-8</td>
<td>.532</td>
<td>.348</td>
<td>.65</td>
<td>.109</td>
<td>.104</td>
<td>.96</td>
<td>.553</td>
<td>.404</td>
</tr>
<tr>
<td>16-16</td>
<td>.870</td>
<td>.632</td>
<td>.73</td>
<td>.115</td>
<td>.158</td>
<td>1.37</td>
<td>.845</td>
<td>.733</td>
</tr>
<tr>
<td>26-16</td>
<td>.922</td>
<td>.727</td>
<td>.79</td>
<td>.107</td>
<td>.180</td>
<td>1.69</td>
<td>.906</td>
<td>.808</td>
</tr>
<tr>
<td>16-26</td>
<td>.918</td>
<td>.717</td>
<td>.78</td>
<td>.108</td>
<td>.189</td>
<td>1.75</td>
<td>.904</td>
<td>.830</td>
</tr>
<tr>
<td>26-26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.977</td>
<td>.860</td>
<td>.88</td>
<td>.115</td>
<td>.226</td>
<td>1.97</td>
<td>.962</td>
<td>.933</td>
</tr>
<tr>
<td>8-8</td>
<td>.502</td>
<td>.396</td>
<td>.79</td>
<td>.092</td>
<td>.133</td>
<td>1.44</td>
<td>.584</td>
<td>.538</td>
</tr>
<tr>
<td>16-16</td>
<td>.839</td>
<td>.686</td>
<td>.82</td>
<td>.099</td>
<td>.216</td>
<td>2.18</td>
<td>.854</td>
<td>.854</td>
</tr>
<tr>
<td>26-16</td>
<td>.915</td>
<td>.798</td>
<td>.87</td>
<td>.098</td>
<td>.269</td>
<td>2.75</td>
<td>.920</td>
<td>.924</td>
</tr>
<tr>
<td>16-26</td>
<td>.917</td>
<td>.794</td>
<td>.87</td>
<td>.099</td>
<td>.274</td>
<td>2.77</td>
<td>.918</td>
<td>.925</td>
</tr>
<tr>
<td>26-26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.972</td>
<td>.891</td>
<td>.92</td>
<td>.088</td>
<td>.340</td>
<td>3.86</td>
<td>.965</td>
<td>.969</td>
</tr>
</tbody>
</table>

---

<sup>a</sup> Except for sample sizes denoted by (b), all estimates are based upon 8000 iterations. Maximum error is .012 (using 95% confidence interval).

<sup>b</sup> Power estimates for this sample size are based upon 2000 iterations. Maximum error is .024.
TABLE 5.

Effect of the value of the intercept

\[ n_1 = n_2 = 8 \quad \beta_1 - \beta_2 = 1.00^a \]

<table>
<thead>
<tr>
<th>k</th>
<th>Intercept = 0</th>
<th>Intercept = .69</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>( \hat{S}_n )</td>
</tr>
<tr>
<td>1</td>
<td>.450</td>
<td>.370</td>
</tr>
<tr>
<td>2</td>
<td>.553</td>
<td>.404</td>
</tr>
<tr>
<td>3</td>
<td>.584</td>
<td>.538</td>
</tr>
</tbody>
</table>

\(^a\) All estimates are based upon 8000 iterations. Maximum error is .012 (using 95% confidence interval).
### TABLE 6

One-sided versus Two-sided tests

Significance level = .05^a \( \beta_1 - \beta_2 = 1.00 \)

<table>
<thead>
<tr>
<th>(m)</th>
<th>1 (NORMAL)</th>
<th>2 (CAUCHY)</th>
<th>3 (EXPONENTIAL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k)</td>
<td>( n_1 - n_2 )</td>
<td>( T )</td>
<td>( \hat{S}_n^* )</td>
</tr>
<tr>
<td>8-8</td>
<td>0.369</td>
<td>0.385</td>
<td>0.088</td>
</tr>
<tr>
<td>1</td>
<td>0.692</td>
<td>0.613+</td>
<td>0.092</td>
</tr>
<tr>
<td>16-16</td>
<td>0.900</td>
<td>0.822+</td>
<td>0.115</td>
</tr>
<tr>
<td>26-26b</td>
<td>0.932</td>
<td>0.874+</td>
<td>0.109</td>
</tr>
<tr>
<td>8-8</td>
<td>0.532</td>
<td>0.474+</td>
<td>0.109</td>
</tr>
<tr>
<td>2</td>
<td>0.870</td>
<td>0.753+</td>
<td>0.115</td>
</tr>
<tr>
<td>16-16</td>
<td>0.977</td>
<td>0.910+</td>
<td>0.115</td>
</tr>
<tr>
<td>26-26b</td>
<td>0.972</td>
<td>0.943+</td>
<td>0.088</td>
</tr>
<tr>
<td>8-8</td>
<td>0.502</td>
<td>0.537+</td>
<td>0.092</td>
</tr>
<tr>
<td>3</td>
<td>0.839</td>
<td>0.805+</td>
<td>0.099</td>
</tr>
<tr>
<td>16-16</td>
<td>0.972</td>
<td>0.943+</td>
<td>0.088</td>
</tr>
<tr>
<td>26-26b</td>
<td>0.972</td>
<td>0.943+</td>
<td>0.088</td>
</tr>
</tbody>
</table>

---

^a All estimates except those denoted by (b) are based upon 8000 iterations. Maximum error is .012 (using 95% confidence interval).

^b Power estimates for this sample size are based upon 2000 iterations. Maximum error is .024.

\(+\) Indicates an increase (decrease) in the power of \( \hat{S}_n^* \) over the power of \( T \).
A number of comments can be made concerning the results in Table 2. First, it appears that the greatest vertical distance between the power curve of $T$ and that of $\hat{S}_n$ occurs somewhere near the alternative $\beta_1 - \beta_2 = 1.25$, except for the case 1-2, and 2-2. At or near this alternative, $\hat{S}_n$ fared the worst when the underlying joint distribution was either bivariate normal (1-1) or normal-exponential (2-1). This is not at all surprising since $T$ is designed for underlying distributions (of the "dependent" variable) which are normal. $\hat{S}_n$ fared the best against the Cauchy distribution ($m = 2$) in which case it almost always out-performed $T$. Again this was to be some what expected since in general normal theory tests do not perform as well as rank tests when the underlying distribution is heavy-tailed, such as it is with the Cauchy distribution. When the underlying distribution for the dependent variable was exponential, $\hat{S}_n$ and $T$ were fairly comparable in power, although $T$ always had a slightly better power for any given alternative. Aside from the cases using the Cauchy distribution, $\hat{S}_n$ fared the best for the case 3-3 (exponential-uniform). In this case, the greatest loss in power appears to be approximately 10%. Finally it should be noted that for each value of $m$, $\hat{S}_n$ fared the best for $k = 3$. This seems to suggest that if the independent variables are really fixed, controllable constants within a specified range, not too much (if anything) is lost in terms of power if the values of these constants are first chosen according to an uniform distribution and then use $\hat{S}_n$ to test (1.36).
Very similar patterns are found in Table 3 (level of significance is .01) as were found in Table 1 and Table 2. The only errors in the significance level for $T$ were for when the "dependent" variable had a Cauchy distribution. The maximum vertical distance between the power curves appear to take place somewhere near the alternative $\beta_1 - \beta_2 = 1.50$. This time the magnitude of this distance appears to be somewhat larger than it was in Table 2. Again, the greatest loss of power occurs for the case 1-1 and 2-1. Aside from the Cauchy distribution, $\hat{S}_n$ fared the best again in the case 3-3, in which the greatest loss in power is 18% (approximately).

The same type of conclusions can be found in Table 4 as were found in Table 2 and Table 3. The only new observation that could be made is to notice that for $m = 2$, the power of $\hat{S}_n$ increases considerably faster as the sample sizes increase than the power of $T$ for all three values of $k$.

The results in Table 5 indicate very clearly that the performance of $\hat{S}_n$ is not invariant to the value of the "intercept" except in the case $k = 2$. In order to get some intuition about the effect of the "intercept", consider again example 1.2 in a very qualitative terms (not at all trying to be geometrically or statistically precise). In example 1.2 (for $\alpha = 0$), the use of signed residuals has the effect of rotating or "flipping" the second and third quadrants of the graph in Figure 2 about the x-axis, causing most of the signed residuals from the first sample to be positive and most from the second sample to be negative. Essentially, for $\alpha = 0$, this "flipping" guarantees that if $\beta_1 > \beta_2$,
then generally all the ranks of the signed residuals from the first sample will be larger than all the ranks of signed residuals from the second sample, and a two-sample rank statistic for location will detect that \( \beta_1 > \beta_2 \). But as the value of the "intercept" increases positively (negatively) more and more of the unsigned residuals \( Y_{1j} - \hat{\beta}X_{1j} \) will become positive (negative) regardless from which sample they come. This is because the residuals are computed as the vertical distance from the line \( y = \beta x \), and the line remains unchanged for any value of the intercept. The "flipping" will cause the signed residuals from the first sample as well as the signed residuals from the second sample to generally be both positive and negative. Now if \( \beta_1 > \beta_2 \), then among the positive signed residuals, the ranks of those from the first sample will be larger than those from the second sample; and among the negative signed residuals the ranks of those from the first sample will be larger than those from the second sample. But obviously the ranks of positive signed residuals from the second sample will be larger than the ranks of the negative signed residuals from the first sample. Consequently it does not seem unreasonable to expect that a two-sample rank statistic would not detect \( \beta_1 > \beta_2 \) for the value of the intercept large as often as it would for intercept = 0. In the case where the \( P(X > 0) = 1 \) (or \( P(X < 0) = 1 \)), the effect of the intercept should disappear since there is really no "flipping" taking place. And this seems to take place in the table for \( k = 2 \). Finally, as it would be expected, the power of \( T \) does not change with the value of the intercept, since the estimate of the intercept is included in the computation of the test statistic.
This dependence of the power of $S^*_n$ on the true value of the "intercept" is an unfortunate property of $S^*_n$ and more research is needed to study the extent of this dependence. One important question to consider is what is the greatest decrease in power as $|\alpha|$ increases without bound. This will enable a conservative estimate of power to be given which does not depend on $\alpha$. More about this effect of $\alpha$ will be mentioned in chapter six.

The surprising result in Table 6 is that when the "dependent" variable is normally distributed, $T$ still performed better than $S^*_n$ for one sided alternatives, although the difference in power decreased considerably from the difference found in Table 4. In practically all other cases, $S^*_n$ had a higher power than that of $T$. Although there are normal theory tests for ordered alternatives, they are more difficult to construct and use. Because $S^*_n$ can be used either as an one-sided test or a two-sided test, it has a slight advantage over $T$. This advantage should be even more pronounced when the number of samples is greater than two.

In summary, $S^*_n$ compares very well with $T$ when used as a one-sided test. When used as a two-sided test, $S^*_n$ compares very well when the underlying distribution of the "dependent" variable is heavier tailed than the normal distribution. When the "dependent" variable is normally distributed, $S^*_n$ compares unfavorably with $T$. The distribution of the "independent" variable has a definite effect on the power of $S^*_n$. A conjecture for measuring this effect of the "independent" variable is to consider the parameter
\[ \tau = \frac{\int |x|h(x) \, dx}{\int x^2 h(x) \, dx} \]

\[ = \frac{1}{1 + \left( \frac{\mu_{|x|}^2}{\sigma_{|x|}^2} \right)} \]

where \( \mu_{|x|} \) is the mean of \( |X| \) and \( \sigma_{|x|}^2 = \sigma_x^2 \) is the variance of \( |X| \). It is conjectured that the higher the value of \( \tau \), the higher the power of \( \hat{S}_n \) in comparison to \( T \). For the three distributions considered in the tables, the values of \( \tau \) are

- Normal (Truncated) \( .64 \)
- Exponential (Truncated) \( .52 \)
- Uniform \((-2,2)\) \( .75 \)

Finally more research is needed on removing the effect of the "intercept" \( \alpha \) on the power of \( \hat{S}_n \). The obvious way of handling this problem is to estimate \( \alpha \) from the combined sample and then base \( \hat{S}_n \) on the ranks of the "true" residuals,

\[ (Y_{ij} - \hat{\alpha} - \hat{\beta}X_{ij}) \, \text{sgn} \, X_{ij}. \]

More about this topic will be mentioned in chapter six.
It has been shown that a class of exactly distribution-free tests of the hypothesis of equal regression lines from k different samples (with equal intercepts) can be developed under the assumption that both the dependent and independent variables are jointly distributed according to (1.33). This assumption represents a different view philosophically of the linear regression setting in the sense that the case in which the independent variables are fixed known constants is not a special case of (1.33). Each rank-like test statistic is based upon the signed "residual" \( (y_{ij} - \hat{\beta} x_{ij}) \) \( \text{sgn} x_{ij} \), where \( \hat{\beta} \) is any estimator from the combined sample which is 2-tuple symmetric and consistent in the sense of (4.9). The exact null distribution of certain members of this class is equivalent certain well-known 2-sample (and k-sample) rank statistics for location such as the 2-sample Mann-Whitney-Wilcoxon (or the k-sample Kruskal-Wallis or k-sample Jonckheere). As would be expected because the rank-like statistic does not use the actual values of the "independent" variable, as does the least-squares statistic or the LRT, there is in general a loss of power when using the rank-like statistic. This loss in power may very well be the price for having distribution-free statistics in the regression setting.

It should be noted that the estimator \( \hat{\beta} \) plays a very minor role in both the exact distribution theory and the asymptotic theory. This is because the requirement that \( \hat{\beta} \) is 2-tuple symmetric (used in the exact distribution setting) is extremely mild, and in the asymptotic
setting \( \hat{\beta} \) can be replaced by \( \beta \) (the true, unknown value) which in turn can be assumed to be zero since the rank-like statistic is invariant to the value of \( \beta \). On the other hand, the parameter \( \alpha \) is, in the true sense of the expression, a nuisance parameter, as can be seen from Table 5.5.

6.2. Extensions

(a) The requirement of equal intercepts can be replaced by the (essential) requirement that all regression lines intersect at a single point, namely at \( X = x_0 \). Then (1.33) would be rewritten as

\[
(1.33)' \quad f(y,x) = f(y-\alpha-\beta_i(x-x_0),x).
\]

Then under (1.33)', the intercepts are for \( 1 \leq i \leq k \)

\[ \alpha_i = \alpha - \beta_i x_0. \]

If \( x_0 \) is known, the rank-like statistics would be based upon the "observations"

\[ \hat{Z}_{ij} = (Y_{ij} - \hat{\beta}(X_{ij} - x_0)) \text{sgn}(x_{ij} - x_0), \]

and the results in chapters 2 and 4 still are valid. If \( x_0 \) is unknown and if \( P(X > x_0) = 1 \), then both \( \alpha \) and \( x_0 \) can be ignored and

\[ Z_{ij} = Y_{ij} - \hat{\beta} x_{ij} \]

can be used to test the hypothesis of equal regression lines. In general, if \( x_0 \) is unknown, an obvious approach to take is to try to use

\[ Z_{ij} = (Y_{ij} - \hat{\beta}(X_{ij} - \hat{x}_0)) \text{sgn}(x_{ij} - \hat{x}_0) \]
for the "observations", where \( \hat{x}_0 \) is a suitable estimator of \( x_0 \).

Conditions (1.38) and (1.39) are the most plausible restrictions to place on \( \hat{x}_0 \).

(b) The effect of \( \alpha \) may be removed by considering the "true" estimated residuals (for equal intercept case)

\[
Z_{ij}^* = (Y_{ij} - \hat{a} - \hat{\beta}(X_{ij})) \text{ sgn } X_{ij}.
\]

Again suitable restrictions must be placed on the estimator \( \hat{\alpha} \). One possible estimator would be the Hodges-Lelmann type estimator developed by Aidchie (1967b) (see (1.16)).

(c) Finally (a) and (b) could be combined by considering

\[
Z_{ij}^* = (Y_{ij} - \hat{a} - \hat{\beta}(X_{ij} - \hat{x}_0)) \text{ sgn } (X_{ij} - \hat{x}_0).
\]

Clearly, in (a), (b), (c), if all estimators are 2-tuple symmetric, theorem 2.1 will still be valid.

(d) The results in chapter two can easily be extended to the multilinear regression setting (\( p > 1 \), see (1.2)), using the unsigned residuals

\[
Z_{ij}^* = Y_{ij} - \sum_{k=1}^{p} \beta_{ik} X_{ijk}.
\]

(See Smith-Wolfe (1977)). Unfortunately, the idea of "flipping", as discussed in chapter five, becomes very complicated in the multilinear setting. The requirement of equal intercepts in the case \( p = 1 \) becomes the requirement that the regression hyper planes

\[
y - \beta_{11}x_1 - \ldots - \beta_{1p}x_p = \alpha, \ 1 \leq i \leq k
\]

all intersect in a single \( p \)-dimensional hyperplane whose projection
onto the p-dimensional subspace \( \{ (x_1, \ldots, x_p) \} \) has the equation

\[
a_1 x_1 + a_2 x_2 + \ldots + a_p x_p = 0,
\]

where \( a_1, \ldots, a_p \) must be known. Then the location rank statistic

should be based upon the "observations"

\[
Z_{ij}^* = (Y_{ij} - \bar{E}) \hat{\beta}_{ik} x_{ij} \prod_{k=1}^{p} \text{sgn} (\sum_{l=1}^{p} a_l x_{ij}^l).
\]

The requirement that the \( a_i \)'s must be known places a severe restriction
upon the use this technique. Further study is needed to remove this
restriction.

6.3. New Research

(a) One very obvious area of research is the effect \( \hat{\beta} \) has
on the small sample power of \( \hat{S}_n \). It is hoped that certain

general criteria can be developed for choosing which estimator \( \hat{\beta} \)
should be used. Also it may be fruitful to try to develope criteria,
such as in adaptive inferences, for selecting which scores should
be used.

(b) Another area that is being studied at present is the calcu-
lation of asymptotic relative efficiencies (A.R.E.) of (1.43) or (1.44a)
relative to the LRT for general scores, \( a_n(i) \). The key result needed
in this area is that

\[
\frac{\hat{S}_n - E(S_n(\beta))}{\text{var}_H(S_n(\beta))} - \frac{S_n(\beta) - E(S_n(\beta))}{\text{var}_H S_n(\beta)} \xrightarrow{p} 0
\]
To show this result, the condition (1.39) will have to be strengthened to

\[ n^{1/2} |\hat{\beta} - \beta| \quad \text{bounded in probability.} \]

Consequently to find the A.R.E. it will also be necessary to find what conditions will insure that \( \hat{\beta} \) is bounded in probability.

(c) An intriguing technique for incorporating the case in which the independent variable is truly a fixed controllable constant is to randomly choose the constants according to some probability law, such as the uniform distribution. This technique corresponds to assuming that \( f(y,x) = \xi(y)h(x) \), except now \( h(x) \) is a known density function. And in a certain sense, the choosing of \( h(x) \) is analogous to choosing a particular experimental design. It may be profitable to try to transfer some of the concepts of optimal design to a concept of an optimal choice of \( h(x) \).

(d) It may be possible to apply the technique of theorem 2.1 to find a distribution-free test for the analysis of covariance problem (see (1.4)).

(e) For the case \( k > 2 \), the next question asked after rejecting \( H \) is where does the difference lie. Consequently a complementary problem to this testing problem is one of estimating contrasts among the \( \beta_i \)'s.
REFERENCES


