A PARTIAL ORDERING OF p POPULATIONS

By

LARRY JOE STEPHENS

Bachelor of Science
Memphis State University
Memphis, Tennessee
1963

Master of Arts
University of Arizona
Tucson, Arizona
1965

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1972

A PARTIAL ORDERING OF p POPULATIONS

Thesis Approved:

P. Larry Claypool
Thesis Adviser

D. Leroy Folks

John Jobe

James E. Shamblin

D. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Dr. P. Larry Claypool for suggesting the problem to me and for the many helpful hours of discussion which we had. I would also like to thank Dr. J. Leroy Folks for serving as chairman of my advisory committee and for his advice during my stay at Oklahoma State. I wish to also thank Dr. John M. Jobe and Dr. James E. Shamblin for serving as members of my committee.

An expression of gratitude is extended to Dr. Gerald Funk for his many helpful suggestions and discussions. I would also like to thank my typist, Mrs. Mary Bonner, for her excellent work and patience.

I would also like to say thank you to my Mother and Father for making it possible for me to obtain an undergraduate degree and for the many other influences they have had on my life.

Finally, I wish to thank my wife, Jean, for her patience and understanding during my years in graduate school.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
	Ideas and Examples which Motivated the Problem	1
	Theoretical Formulation of the Problem	3
	Other Aspects of the Problem	4
II.	BRIEF REVIEW OF LITERATURE	6
	Friedman's Test	7
	Coefficient of Concordance	9
	Slippage Tests	10
	Extreme Rank Sum Test	12
	Estimation of the True Ranking	13
	Pairwise Comparisons and Comparison with a Control	13
	Asymptotically Efficient Tests in Two-Way Designs where Ranks are Used	16
	Incomplete Blocks in Ranking Experiments	20
	Extension to Unbalanced Cases	22
III.	THE PARTIAL RANKING STATISTIC AND IT'S DISTRIBUTIONAL PROPERTIES	26
	Moments of the Cell Scores	26
	Description of the Test Statistic and a Discussion of It's Distribution	28
	Approximations to the Distribution of S	34
	Estimation of the True Ranking of the Objects . .	52
	Asymptotic Relative Efficiency Considerations . .	54
IV.	AN EXTREME RANK SUM TEST OR A TEST FOR SLIPPAGE	60
V.	PAIRWISE COMPARISONS AND COMPARISONS WITH A CONTROL	81
	Pairwise Comparisons	81
	Comparisons with a Control Population	87
VI.	A PARTIAL ORDERING USING LINEARLY WEIGHTED SCORES	92

Chapter		Page
VII.	SUMMARY AND EXTENSIONS	111
A SELECTED BIBLIOGRAPHY		113
APPENDIX		114

LIST OF TABLES

Table		Page
I.	Distribution of S: (P = Prob. (S $\geq$ x))	35
II.	Percent Errors for the Chi-Square Approximation in the Tail Region	40
III.	A. R. E. Comparisons	59
IV.	Extreme Rank Sum Test	72

CHAPTER I

INTRODUCTION

1. Ideas and Examples Which Motivated the Problem

The statistical tests proposed in this dissertation are motivated by investigations which require a statement of preference by each of several individuals where it is either unreasonable or not practical for the statement of preference to include all items under consideration. Thus, only a subset of the items containing either the "most preferred" or "least preferred" items is actually ranked.

When the number of items is small and it is practical to rank all the items, then the test developed by Friedman in [4] may be used. The term partial ordering will be used to describe the case where only a subset of the items is ranked. When the subset is all of the items, the term full ordering will be used. The partial ordering may be thought of as a generalization of Friedman's test. A review of the literature pertaining to Friedman's test will be given in Chapter II.

Consider now some examples which motivated this study.

Example 1.1: Suppose a large supermarket is interested in reducing the number of detergents which it stocks and it wants to discontinue some brands and continue some others. A survey is conducted in which a number of housewives are asked to rank the best 5 of 15 brands which are available. Based upon the rank sum totals which each brand receives, the management decides which brands to discontinue.

Example 1.2: Suppose a certain city has recently appointed a Board of Industrial Development which initially wishes to determine which factors are most important to a firm which is locating a new plant. The Board selects n firms at random and asks that a company executive from each firm rank what he considers to be the k most important factors (from a list of p factors) in determining the site of a new plant.

Example 1.3: Consider the weekly poll of the top twenty major college football teams in the country throughout the season. Consider in particular the one conducted by the Associated Press. Those participating in the poll are sports writers or sports editors from member newspapers, and television and radio sportscasters. The panel consists of 55 individuals. Members of the panel generally are changed from year to year. Each section of the country is allocated a certain number of electors, depending on the number of NCAA-member universities and the number of major athletic conferences in the area. Each member of the panel votes for 15 teams each week. Twenty points are alloted for a first place vote, 18 points for second and 16 - 14 - 12 - 10 - 9 - 8 - 7 - 6 - 5 - 4 - 3 - 2 - 1 points respectively, on votes for third through fifteenth place. The Associated Press lists the twenty top teams in order, the number of points each received and the number of first-place votes each received.

2. Theoretical Formulation of the Problem

Examples 1.1 and 1.2 suggest the following theoretical formulation of the problem. Suppose there are p populations or objects to be considered and each of n judges is asked to rank the k "most preferred" objects according to some criterion of interest. The constant score c (either < 1 or $> k$) is assigned to each of the remaining $p - k$ to indicate that they have been judged to be "inferior" to each of the k objects which are ranked. The symbol r_{ij} is used for the score corresponding to the i^{th} judge and j^{th} population where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, p$. Two basic assumptions are made. They are: (1) The n p-variate random variables $(r_{i1}, \ldots, r_{ip})$, $i = 1, 2, \ldots, n$ are mutually independent. (The results within one block do not influence the results within the other blocks), and (2) each judge is capable of ranking the k most preferred objects according to some criterion of interest.

The null hypothesis which is to be tested is:

H_0: There are no differences in preference for the p objects.

The alternative hypothesis is:

H_A: At least one object is more preferable than at least one other object.

The test statistic used to test this hypothesis is based on the sums

$$r_j = \sum_{i=1}^{n} r_{ij} \quad \text{for } j = 1, 2, \ldots, p \qquad (1.1)$$

where r_j is the total of the scores received by the j^{th} object. The average sum of the scores assigned to each object will be denoted as $\bar{r}$ and may be calculated as

$$\bar{r} = \frac{\sum_{j=1}^{p} r_j}{p} = \frac{n k(k+1) + 2n(p-k)c}{2p} . \tag{1.2}$$

The test statistic S which will be used to test H_0 against H_A is

$$S = \sum_{j=1}^{p} (r_j - \bar{r})^2 . \tag{1.3}$$

If H_0 is true then each of the r_j's should be "close" to $\bar{r}$ and S should be "small". If, however, some of the r_j's are "large" then S will tend to be "large", tending to make one doubt H_0.

To test H_0 against H_A, the distribution of S must be known or an approximation to the distribution must be known. Also, since the distribution of S must be tabulated, some decision as to the appropriate value for c must be made. The well-known Friedman statistic [4] becomes a special case of the statistic S for either $k = p$ or $k = p - 1$ when each judge assigns the constant $c = 0$ in addition to the ranks $1, 2, \ldots, k$. These problems are discussed in Chapter III.

3. Other Aspects of the Problem

Another problem of interest is to determine the "best" object according to some criterion of interest. To do this, the following alternative hypothesis is considered:

H_A: At least one population has slipped to the right (in the direction of greater preference) of the others.

The statistic used to test H_0 against H_A is r_{max}, the maximum of the column sums. This problem is discussed in Chapter IV.

If θ_i; $i = 1, 2, \ldots, p$ are location parameters corresponding to the p populations, then it may be desirable to make pair-wise comparisons in terms of the θ_i or comparisons with a control population. These ideas are developed in Chapter V.

The football poll example mentioned earlier suggests the following possible extension of the test. Each judge assigns the following ascending scores,

$$a_1 \cdot 1 + b_1 < a_1 \cdot 2 + b_1 < \ldots < a_1 \cdot k_1 + b_1 < a_2 \cdot 1 + b_2$$
$$< a_2 \cdot 2 + b_2 < \ldots < a_2 \cdot k_2 + b_2 ,$$

and $k = k_1 + k_2$. Again, the remaining $(p - k)$ objects would receive a score of $c < a_1 \cdot 1 + b_1$. The top k_2 objects are assigned scores from one set of linearly ordered ranks and the next k_1 are assigned scores from a second set of linearly ordered ranks. Chapter VI is devoted to extending the ideas of Chapter III, IV and V to this problem.

CHAPTER II

BRIEF REVIEW OF LITERATURE

The basic idea of using ranks in a two-way design was introduced by Friedman [4] in 1937. Essentially the same problem was discussed by Kendall and Babington Smith [5] in 1939. The essential difference in the two is the test statistic used.

Doornbos and Prins [2] discussed the slippage problem in 1958. Thompson and Willke [11] in 1963 discussed essentially the same problem, using the same techniques to locate outliers in a two-way design where ranks instead of the actual observations are used.

Kendall [6] discusses the problem of estimating the true ranking of the objects or populations based on the column totals in a two-way rank analysis.

Pairwise comparisons and comparisons with a control in a Friedman rank analysis of a two-way classification are given by Miller [7].

Sen [10] considers asymptotically efficient tests in a two-way design where ranks are used. In this paper, a case where Friedman's test is optimal, as well as one where a partial ordering with $k = 1$ is optimal, is given.

Durbin [3] introduced a modification of the Kendall-Babington Smith test in 1951. The modification he made was motivated by the difficulty of ranking a large number of objects.

Bernard and VanElteren [1] discuss ranking procedures in a two-way design where there may be any number of observations in a given cell.

Each of the above topics will now be discussed in some detail.

1. Friedman's Test

Let X_{ij} be the observation in the i^{th} block corresponding to the j^{th} population in a two-way design. If there are p populations and n blocks, the data may be arranged as follows:

		Population			
		1	2	. . .	p
Block	1	X_{11}	X_{12}	. . .	X_{1p}
	2				
	. . .			. . .	
	n	X_{n1}	X_{n2}	. . .	X_{np}

Let r_{ij} denote the rank of X_{ij} relative to the other observations within the i^{th} block. That is, for block i, r_{ij} is the rank of the j^{th} observation with respect to the other observations in block i. The smallest observation in block i has rank 1, the next smallest has rank 2, etc. The original two-way layout is then replaced by the following one.

Population

Block	1	2	...	p
1	r_{11}	r_{12}	...	r_{1p}
2				
⋮			⋮	
n	r_{n1}	r_{n2}	...	r_{np}
	r_1	r_2	...	r_p

Define column totals r_j by the following

$$r_j = \sum_{i=1}^{n} r_{ij} \quad \text{for } j = 1, 2, \ldots, p$$

and the mean of the column totals as $\bar{r}$ where

$$\bar{r} = \frac{\sum_{j=1}^{p} r_j}{p} = \frac{n(p+1)}{2} .$$

The following assumptions are made:

(1) The n p-variate random variables $(r_{i1}, \ldots, r_{ip})$, $i = 1, 2, \ldots, n$ are mutually independent and,

(2) Within each block the observations may be arranged in increasing order according to some criterion of interest.

The null hypothesis H_0 is

H_0: There are no differences in the p populations.

The alternative hypothesis H_A is

H_A: At least one of the populations tends to yield larger observed values than at least one other population.

The statistic used to test H_0 versus H_A is

$$S = \sum_{j=1}^{p} \left(r_j - \frac{n(p+1)}{2} \right)^2 .$$

Friedman [4] shows that χ_r^2 defined by

$$\chi_r^2 = (12/n\,p(p+1))S$$

has a limiting chi-square distribution with $(p - 1)$ degrees of freedom. Tabulations of the exact distribution of χ_r^2 may be found in [4], [5] and [8] for small n and p.

Friedman points out in his original paper [4] that his test procedure may be used to avoid normality assumptions and also that in certain instances the data may be given directly in rank form.

2. Coefficient of Concordance

Kendall and Babington Smith [5] rescale Friedman's statistic to test H_0 versus H_A. Define S the same as in the Friedman case. The maximum value of S is

$$S_{max} = n^2(p^3 - p)/12$$

The coefficient of concordance W is defined as

$$W = S / S_{max} = 12 S/n^2(p^3 - p) .$$

W and χ_r^2 are related by the following equation

$$\chi_r^2 = n(p-1)W .$$

The same exact tabulations used for χ_r^2 may also be used for W. The first two moments of W are equated to those of a beta Type I distribution and it is shown that the third and fourth moments for the two distributions are close. The beta approximation is used for W for non-tabulated values of the exact distribution. It is of interest to note that Kendall and Babington Smith looked at W as a measure of community judgment among the n judges.

3. Slippage Tests

Doornbos and Prins [2] discuss a general slippage test and then apply it to variates following various specified distributions as well as the rank analysis of a two-way design. A discussion of the test follows. Again, assume there are p populations and n blocks and that $r_{.j}$, $j = 1, \ldots, p$ are column totals corresponding to the populations. Let θ_i, $i = 1, 2, \ldots, p$ be location parameters corresponding to the populations. The null hypothesis is:

$$H_0 : \theta_1 = \ldots = \theta_p$$

The alternatives considered are:

H_{A_1} : At least one θ_i has slipped to the right, or

H_{A_2}: At least one θ_i has slipped to the left.

The same basic assumptions are made by Doornbos and Prins as Friedman made. Also, Doornbos and Prins assumed that if the population distributions differ, they differ only in location parameter (as did Thompson and Willke). To test H_0 against H_{A_1} at significance level α, the statistic r_{max} is used and to test H_0 against H_{A_2} at significance level α, r_{min} is used.

The critical value for testing against H_{A_1} is the smallest integer R_α satisfying

$$P[r_i \geq R_\alpha] \leq \alpha/p .$$

H_0 is rejected in favor of H_{A_1} if $r_{max} \geq R_\alpha$.

The critical value for testing against H_{A_2} is the largest integer S_α satisfying

$$P[r_i \leq S_\alpha] \leq \alpha/p .$$

H_0 is rejected in favor of H_{A_2} if $r_{min} \leq S_\alpha$.

The marginal distribution of r_j must be known in order for the above two inequalities to be useful. The marginal distribution of r_j under H_0 is:

$$P[r_j = M] = \sum_{x=0}^{\infty} I_{M-px-n} \binom{n}{x} \binom{M-px-1}{n-1} (-1)^x p^{-n} ,$$

$j = 1, 2, \ldots, p$, where

$$\begin{cases} I_y = 0 & \text{if } y < 0 \\ I_y = 1 & \text{if } y \geq 0 \end{cases}$$

A table of R_α and S_α values may be found in [2].

4. Extreme Rank Sum Test

Thompson and Willke [11] using the same techniques as Doornbos and Prins [2] arrive at the following statement for testing against slippage in either direction.

$$p \cdot P(\overline{A}_j) - \binom{p}{2} P(\overline{A}_j \overline{A}_{j'}) \leq P(\overline{A}) \leq p \cdot P(\overline{A}_j)$$

for $j \neq j'$, where

$$A_j = \{(r_1, \ldots, r_p) \mid n + R < r_j < np - R\}$$

for $j = 1, 2, \ldots, p$ and

$$\overline{A} = \{(r_1, \ldots, r_p) \mid r_{min} \leq n + R \text{ or } r_{max} \geq np - R\}$$

If $p \cdot P(\overline{A}_j)$ is set equal to α, then R may be determined. Tables of R for α levels of .01, .03 and .05 may be found in [11].

The same inference is made here as in the slippage case, i.e., at least one population has slipped from the others and one of them has been found.

To see how the work of Thompson and Willke relate to that of Doornbos and Prins, consider slippage only to the right in the above discussion. In that case A_j becomes

$$A_j = \{(r_1, \ldots, r_p) | r_j < np - R\}$$

and $\bar{A}$ becomes,

$$\bar{A} = \{(r_1, \ldots, r_p) | r_{max} \geq np - R\}$$

and it is apparent that $R_\alpha = np - R$ where R is the largest integer such that $p \cdot P(\bar{A}_j) \leq \alpha$.

Similarly if only slippage to the left is considered, then $S_\alpha = n + R$. The tables given in [11] are for two sided slippage and those given in [2] are for one-sided slippage.

5. Estimation of the True Ranking

Kendall [6] suggests the following estimation procedure. Let r_j ; $j = 1, 2, \ldots, p$ be the p rank sums and R_j ; $j = 1, 2, \ldots, p$ be the estimates of the true rankings of the populations. Kendall suggests the following procedure. For the largest r_j, take $R_j = p$, for the next largest r_j, take $R_j = p - 1$, etc.

This procedure is "best" in the following sense: Subject to the constraint that the same scores be used that were used by the individual judges, it maximizes the average rank correlation between the estimated and the observed rankings and it minimizes the sums of squares of difference between the actual scores r_j and what they would be nR_j if all the rankings were identical.

6. Pairwise Comparison and Comparison With a Control

A discussion of pairwise comparison and comparison with a control may be found in [7]. Let X_{ij} ; $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, p$ be the observation in the i^{th} block corresponding to the j^{th} population. Assume that X_{ij} is distributed according to the density

$$f_{X_{ij}}(x) = f(x - \theta_j - \beta_i) \quad i = 1, \ldots, n; \; j = 1, \ldots, p .$$

The density f is fixed but unknown. Again assume the observations X_{ij} are replaced by their appropriate ranks r_{ij}. Define the column means as

$$\bar{r}_j = \frac{\sum_{i=1}^{n} r_{ij}}{n} = \frac{r_{.j}}{n} \quad \text{for } j = 1, 2, \ldots, p .$$

Three cases will now be discussed. The results in each case are based upon asymptotic theory and it will therefore be necessary to assume n is large in each case. The significance level in each case is α.

<u>Case 1</u>: The null hypothesis is,

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_p$$

The alternative hypothesis H_A is

$$H_A : \theta_j \neq \theta_{j'} \quad \text{for some pairs of } (j, j'),\ j \neq j' .$$

The null hypothesis is accepted if

$$|\bar{r}_j - \bar{r}_{j'}| \leq q^{\alpha}_{p,\infty} \left(\frac{p(p+1)}{12n} \right)^{1/2} \quad \text{for every } j \neq j'$$

where j and j' take on the values $1, 2, \ldots, p$. $q^{\alpha}_{p,\infty}$ is the upper α percentile of the range of p independent unit normal random variables. Values for $q^{\alpha}_{p,\infty}$ may be found in Table I of Appendix B of [7].

If $|\bar{r}_j - \bar{r}_{j'}|$ exceeds the critical value for at least one pair (j, j') the null hypothesis is rejected. For every difference $|\bar{r}_j - \bar{r}_{j'}|$ which exceeds the critical value, it is concluded that $\theta_j \neq \theta_{j'}$.

Case 2: Suppose a control population is introduced and that it has location parameter $\theta_0 = 0$. The null hypothesis H_0 is

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_p = 0$$

The alternative hypothesis H_A is

$$H_A : \theta_j > 0, \quad \text{for at least one } j, \; j = 1, 2, \ldots, p.$$

The null hypothesis is accepted if

$$\bar{r}_j - \bar{r}_0 \leq m^{\alpha}_{p(1/2)} \left[\frac{(p+1)(p+2)}{6n}\right]^{1/2} \quad \text{for every } j,$$

where $m^{\alpha}_{p(1/2)}$ is the upper α percentile point of the maximum of p unit normal random variables with common correlation $\rho = 1/2$. Values of $m^{\alpha}_{p(1/2)}$ may be found in Table IV, Appendix B of [7].

If at least one $\bar{r}_j - \bar{r}_0$ exceeds the critical value, then H_0 is rejected. Furthermore, for every j for which $\bar{r}_j - \bar{r}_0$ exceeds the critical value it is inferred that $\theta_j > 0$.

Case 3: This case is the same as Case 2 except the alternative H_A is:

$$H_A : \theta_j \neq 0, \quad \text{for at least one } j, \quad j = 1, 2, \ldots, p .$$

The null hypothesis is accepted if

$$|\bar{r}_j - \bar{r}_0| \leq |m|^{\alpha}_{p(1/2)} \left[\frac{(p+1)(p+2)}{6n}\right]^{1/2} \quad \text{for every } j$$

where $|m|^{\alpha}_{p(1/2)}$ is the upper α percentile point of the maximum absolute value of p unit normal random variables with common correlation $\rho = 1/2$. The distribution of $|m|^{\alpha}_{p(1/2)}$ may be found in Table IV, Appendix B of [7].

The null hypothesis is rejected if $|\bar{r}_j - \bar{r}_0|$ exceeds the critical value for at least one j. When this occurs, it is concluded that $\theta_j \neq 0$.

7. Asymptotically Efficient Tests in Two-Way Designs where Ranks are Used

Sen [10] gives the following type of discussion. Consider random variables $(X_{i1}, \ldots, X_{ip})$ (which may or may not be observable) underlying the ranks $(r_{i1}, \ldots, r_{ip})$ for $i = 1, \ldots, n$. It is then assumed that $X_{i1}, \ldots, X_{ip}$ are independently distributed according to continuous cumulative distribution functions $F_{i1}(x), \ldots, F_{ip}(x)$, respectively, for $i = 1, \ldots, n$. The null hypothesis states that

$$H_0 : F_{i1} = \ldots = F_{ip} = F_i \quad \text{for all } i = 1, \ldots, n .$$

Consider the translation alternatives,

$$F_{ij}(x) = F_i(x - T_j) \quad (j = 1, \ldots, p ; \ i = 1, \ldots, n) ;$$

$$\sum_{j=1}^{p} T_j = 0\,, \quad T = (T_1, \ldots, T_p)$$

Sen made the following assumptions:

(i) n (the number of blocks or observers) is large ;

(ii) $F_i(x)$ is absolutely continuous having a continuous density function $f_i(x)$, where

$$\int_{-\infty}^{\infty} f_i^2(x)\,dx < \infty\,, \quad i = 1, \ldots, n\,.$$

(iii) $T = n^{-1/2}\theta$; $\theta = (\theta_1, \ldots, \theta_p)$ has real and finite elements.

Sen confines himself to the following class of rank tests. Let $\{J(r,p) : r = 1, \ldots, p\}$ be p real-valued functions, where $J(r,p)$ is a function of r and p $(r = 1, \ldots, p)$.

Define

$$\overline{J} = p^{-1} \sum_{r=1}^{p} J(r,p) \tag{2.1}$$

and

$$A^2(J) = (p-1)^{-1} \sum_{r=1}^{p} \{J(r,p) - \overline{J}\}^2\,. \tag{2.2}$$

It is assumed that for any finite p, the $J(r,p)$ are all finite and are not all equal.

Define a rank statistic (vector) $T_n = (T_{n1}, \ldots, T_{np})$ where

$$T_{nj} = n^{-1} \sum_{j=1}^{n} J(r_{ij}, p)\,, \quad j = 1, \ldots, p\,. \tag{2.3}$$

Tests are based on statistics of the form

$$S_n = n[A^{-2}(J)] \sum_{j=1}^{p} \{T_{nj} - \bar{J}\}^2 . \tag{2.4}$$

The main objective is to select $\{J(r, p)\}$ in such a manner that the corresponding S_n leads to an asymptotically efficient test with respect to the classical analysis of variance test. A brief outline of Sen's results now follows. Define $\beta^{(i)}_{s,p-2}$ as follows,

$$\beta^{(i)}_{s,p-2} = \binom{p-2}{s} \int_{-\infty}^{\infty} \{F_i(x)\}^s \{1 - F_i(x)\}^{p-2-s} f_i^2(x)\,dx , \quad (s = 0, \ldots, p-2)$$

and let

$$\beta^{(i)}_{-1,p-2} = \beta^{(i)}_{p-1,p-2} = 0 \quad \text{for} \quad i = 1, \ldots, n .$$

Then, define

$$\bar{\beta}^{(n)}_{s,p-2} = n^{-1} \sum_{i=1}^{n} \beta^{(i)}_{s,p-2} , \qquad s = -1, 0, \ldots, p-1 .$$

and

$$\lambda_{r,n} = \bar{\beta}^{(n)}_{r-1,p-2} - \bar{\beta}^{(n)}_{r-2,p-2} , \qquad r = 1, 2, \ldots, p .$$

S_n, under the given translation alternatives, is shown to have asymptotically a non-central chi-squared distribution with $(p - 1)$ degrees of freedom and the non-centrality parameter

$$\Delta_n(J) = p^2 [A^{-2}(J)] \left(\sum_{j=1}^{p} \theta_j^2 \right) \left\{ \sum_{r=1}^{p} J(r, p) \lambda_{r,n} \right\}^2$$

If F_i possesses finite variance σ_i^2 for $i = 1, 2, \ldots, n$, then $(p - 1)\ F_{(p-1),(n-1)(p-1)}$ (where $F_{(p-1),(n-1)(p-1)}$ is the classical analysis of variance test statistic) has asymptotically, under the given translation alternatives, a non-central chi-squared distribution with $(p - 1)$ degrees of freedom and non-centrality parameter

$$\Delta_n^* = \left(\sum_{j=1}^{p} \theta_j^2\right) \Big/ \bar{\sigma}_n^2 \quad \text{where} \quad \bar{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \sigma_i^2 .$$

Using $\Delta_n(J)$ and Δ_n^*, the asymptotic relative efficiency (A. R. E.) of the test based on S_n with respect to the classical analysis of variance test is deduced to be equal to

$$e_n(J) = \bar{\sigma}_n^2 p^2 (p - 1) \left\{ \sum_{r=1}^{p} J(r, p) \lambda_{r,n} \right\}^2 \Big/ \left[\sum_{r=1}^{p} \{J(r, p) - \bar{J}\}^2 \right] \qquad (2.5)$$

which may also be written as

$$\left\{ \bar{\sigma}_n^2 p^2 (p-1) \sum_{r=1}^{p} \lambda_{r,n}^2 \right\}$$
$$\left\{ \left(\sum_{r=1}^{p} \{J(r, p) - \bar{J}\} \lambda_{r,n} \right)^2 \Big/ \left(\sum_{r=1}^{p} \lambda_{r,n}^2 \right) \left(\sum_{r=1}^{p} \{J(r, p) - \bar{J}\}^2 \right) \right\}$$

where the second factor in the last equation is uniformly bounded by 1, and the first factor is independent of $\{J(r, p) : r = 1, 2, \ldots, p\}$. Now, e_n^* given by

$$e_n^* = \bar{\sigma}_n^2 p^2 (p-1) \sum_{r=1}^{p} \lambda_{r,n}^2 \qquad (2.6)$$

is defined to be the A. R. E. of the test based on S_n, since it is the maximum value of $e_n(J)$. Therefore, a particular S_n test is said to be asymptotically efficient for $(F_1, \ldots, F_n)$ if $e_n(J) = e_n^*$ for that $(F_1, \ldots, F_n)$. It may be shown that $e_n(J) = e_n^*$ if and only if $J(r, p) - \overline{J} = k\lambda_{r,n}$ for all $r = 1, \ldots, p$ and k being a constant. If $F_1 = F_2 = \ldots = F_n = F$, then in many cases $\lambda_{r,n}$ are quite simple and the corresponding S_n statistics are not difficult to compute. Three examples from Sen's paper will now be given. In each case the set of quantities $J(r, p)$ given satisfy the above optimality condition.

For the uniform distribution, the scores are,

$$J(r,p) = \begin{cases} 1, & r = 1 \\ 0, & 2 \leq r \leq p-1, \quad p \geq 4 \\ -1, & r = p \end{cases}$$

For the exponential distribution, the scores are

$$J(r,p) = \begin{cases} 1, & r = 1 \\ & \\ 0, & r = 2, \ldots, p \end{cases}, \quad p \geq 4$$

and for the logistic distribution, the scores are

$$J(r,p) = r \quad \text{for} \quad 1 \leq r \leq p. \quad , \quad p \geq 4$$

8. Incomplete Blocks in Ranking Experiments

If p, the number of objects to be ranked is large, then in a pure ranking situation it is desirable to reduce the number of objects each judge is to rank, since it is difficult to rank a large number of objects. One solution to this problem was suggested by Durbin [3]. In his scheme, each judge is presented a subset of the p objects and asked to rank the objects in that subset. The null and alternative hypothesis and the assumptions underlying the test are the same as in Friedman's case. In addition, the following two conditions are imposed on the scheme. (a) Each object should occur an equal number of times in the experiment as a whole, and (b) The number of times two particular objects occur together should be the same for all possible pairs of objects. Designs in which these conditions are fulfilled are called "balanced incomplete block designs."

A general description of the test statistic follows. Suppose p objects are presented in blocks of size k, and that each object is ranked n times. The number of blocks will be $\frac{np}{k}$. Within each block there are $\frac{k}{2}(k-1)$ comparisons between pairs. Consequently, the total number of comparisons is $\frac{np}{2}(k-1)$. Let λ be the number of blocks in which a particular pair of objects occurs. Then $\lambda p(p-1)/2 = np(k-1)/2$ so that

$$\lambda = \frac{n(k-1)}{p-1}$$

Let r_{ij} denote the rank assigned to the j^{th} object in the i^{th} block, where only k of the p objects are ranked, and define the statistic

$$S = \sum_{j=1}^{p} (r_j - \bar{r})^2$$

where

$$r_j = \sum_{i=1}^{np/k} r_{ij} \quad \text{for } j = 1, 2, \ldots, p .$$

and $\bar{r}$ is the average of the column totals. S attains its maximum value when there is perfect concordance among the rankings. The column totals are then some permutation of the integers $n, n+\lambda, \ldots, n+(p-1)\lambda$. Thus, the maximum value of S is $\lambda^2 p(p^2-1)/12$ and the coefficient of concordance is

$$W = \frac{12S}{\lambda^2 p(p^2-1)} .$$

Durbin suggests the beta distribution as an approximation to W. He finds the first two moments of W and equates them to the first two moments of the beta distribution and suggests that the third and fourth moments will also be close, but he doesn't really show that they will be. He also points out that

$$\chi_r^2 = \frac{\lambda(p^2-1)}{k+1} W$$

has, asymptotically, a chi-square distribution with $(p-1)$ degrees of freedom.

9. Extension to Unbalanced Cases

The Durbin test has been generalized to the case where some experimental units may contain several observations by Bernard and VanElteren [1]. The same basic assumptions are made that are made in Friedman's test. A brief discussion of Bernard and VanElteren's results will now be given.

Suppose there are np random variables X_{ij}, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, p$. The number of observations of X_{ij} is k_{ij}. Define

$k_i = \sum_{j=1}^{p} k_{ij}$ to be the number of observations in the i^{th} block. The k_i observations in each block are ranked. The ranks of observations of X_{ij} are said to belong to cell (i, j). In each ranking ties are allowed. The number of ranks in a tie is called the size of that tie. Denote by $t_{i\gamma}$ the number of ties of size γ and by g_i the size of the greatest tie in the i^{th} ranking.

Now, for each i, derive from the ranks in the i^{th} block the reduced ranks by subtracting $1/2(k_i + 1)$, the mean of the ranks in the i^{th} ranking. The sum of the reduced ranks in cell (i, j) is denoted by $\widetilde{\mu}_{ij}$. If $k_{ij} = 0$, set $\widetilde{\mu}_{ij} = 0$. Define column totals $\widetilde{\mu}_j$ as

$$\widetilde{\mu}_j = \sum_{i=1}^{n} \widetilde{\mu}_{ij}, \qquad j = 1, 2, \ldots, p .$$

Define the quantities

$$\sigma_{jj'} = - \sum_{i=1}^{n} k_{ij} k_{ij'} K_i$$

and

$$\sigma_{jj} = \sum_{i=1}^{n} k_{ij} (k_i - k_{ij}) K_i ,$$

where

$$K_i = \frac{k_i^3 - \sum_{\gamma} \gamma^3 t_{i\gamma}}{12 k_i (k_i - 1)} .$$

Also define the matrices

$$V_\mu = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} & \tilde{\mu}_1 \\ & \vdots & & \\ \sigma_{p1} & \cdots & \sigma_{pp} & \tilde{\mu}_p \\ \tilde{\mu}_1 & \cdots & \tilde{\mu}_p & 0 \end{bmatrix}$$

and

$$V = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ & \vdots & \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix} .$$

The test statistic for H_0, to be described below, is defined as follows. Consider the matrix formed from V_μ by deleting an arbitrary row and an arbitrary column, excluding the last row and the last column, and compute its determinant Δ_μ. Consider also the matrix obtained by deleting an arbitrary row and an arbitrary column from V and compute its determinant Δ. Then define

$$\chi_r^2 = \frac{|\Delta_\mu|}{|\Delta|}$$

The null hypothesis H_0 is

H_0 : For each ranking all possible manners of dividing the given sets of ranks into cells of the prescribed sizes have the same probability.

Under H_0 and the appropriate regularity conditions given in [1], χ_r^2 has a limiting χ^2 distribution with $p - 1$ degrees of freedom.

CHAPTER III

THE PARTIAL RANKING STATISTIC AND ITS DISTRIBUTIONAL PROPERTIES

The basic formulation of the problem was given in Section 2 of Chapter I. It will be convenient to think of the data as forming an $n \times p$ matrix with elements r_{ij} or as being arranged in a two-way table. In either case, the i^{th} row contains the scores assigned by judge i and the j^{th} column contains the scores assigned to object j by the various judges.

In Section 1, the moments of r_{ij} will be found. A description of the test statistic and its distribution will be given in Section 2. Section 3 considers approximations to the distribution of the test statistic. Section 4 is concerned with estimation of the true rankings of the objects and Section 5 deals with asymptotic relative efficiency considerations.

1. Moments of the Cell Scores

The distribution of r_{ij} under H_0 is:

$$P(r_{ij} = \alpha) = \begin{cases} \dfrac{1}{p}, & \alpha = 1, 2, \ldots, k \\ \dfrac{p-k}{p}, & \alpha = c \end{cases}$$

The first three moments of r_{ij} will now be derived. The mean, μ, of r_{ij} is

$$\mu = E(r_{ij}) = \sum_{i=1}^{k} i \cdot \frac{1}{p} + c \cdot \frac{(p-k)}{p}$$

$$= \frac{k(k+1) + 2c(p-k)}{2p} . \tag{3.1}$$

The variance σ^2 of r_{ij} is derived as follows.

$$\sigma^2 = E(r_{ij}^2) - \mu^2$$

$$= \sum_{i=1}^{k} i^2 \cdot \frac{1}{p} + c^2 \cdot \frac{(p-k)}{p} - \mu^2$$

$$= \frac{(p-k)k}{p^2} c^2 - \frac{(p-k)k(k+1)}{p^2} c + \frac{k(k+1)}{p} \left[\frac{2k+1}{6} - \frac{k(k+1)}{4p} \right] . \tag{3.2}$$

Now consider the covariance of r_{ij} and $r_{ij'}$. Define d_{ij} as

$$d_{ij} = r_{ij} - \mu . \tag{3.3}$$

Then

$$\text{cov}(r_{ij}, r_{ij'}) = E(d_{ij} \cdot d_{ij'})$$

$$= E_{d_{ij'}} [d_{ij'} E(d_{ij} \mid d_{ij'})]$$

$$= E_{d_{ij'}} \left[- \frac{d_{ij'}}{(p-1)} d_{ij'} \right] ,$$

since $\sum_{i=1}^{p} d_{ij} = 0$, and

$$E_{d_{ij}}(d_{ij} \mid d_{ij'}) = \sum_{\substack{j=1 \\ j \neq j'}}^{p} d_{ij} \cdot \frac{1}{p-1} \quad .$$

Therefore,

$$\begin{aligned} \operatorname{cov}(r_{ij}, r_{ij'}) &= -\frac{1}{(p-1)} E(d_{ij'}^{2}) \\ &= -\frac{1}{(p-1)} \sigma^{2} \quad . \end{aligned} \qquad (3.4)$$

Due to the independence of the rows, the covariance of r_{ij} and $r_{i'j'}$, $i \neq i'$ is zero.

The third moment about the mean, μ_3, is derived as follows:

$$\begin{aligned} \mu_3 &= E(r_{ij} - \mu)^3 \\ &= \left(\sum_{i=1}^{k} (i-\mu)^3 + (p-k)(c-\mu)^3 \right) \cdot \frac{1}{p} \\ &= \frac{k(k+1)}{2p} \left(\frac{k(k+1)}{2} - (2k+1)\mu + 3\mu^2 \right) + \frac{(p-k)}{p} (c-\mu)^3 - \frac{k}{p} \mu^3 \quad . \end{aligned} \qquad (3.5)$$

2. Description of the Test Statistic and a Discussion of It's Distribution

The statistic to be used to test H_0 will now be described. The column totals, r_j, are defined as in Equation (1.1).

The mean of the column totals is

$$\begin{aligned} \bar{r} &= \sum_{j=1}^{p} r_j / p = n\mu \\ &= \frac{n}{p} \left[\frac{k(k+1) + 2c(p-k)}{2} \right] \quad . \end{aligned}$$

The test statistic which will be used to test H_0 versus H_A is

$$S = \sum_{j=1}^{p} (r_j - \bar{r})^2 .$$

Before the problem of determining the distribution of S can be solved, a decision must be made in regard to the value of c to be used. This decision is facilitated by first making the arbitrary choice to rank the k most preferred objects in ascending order of preference; that is, the single most preferred object receives a rank of k. Based on this convention any value $c < 1$ may be assigned to the $p - k$ unranked objects to indicate that they have been judged to be inferior to each of the k objects that were ranked. The value zero is selected as the assigned value of c.

Aside from the obvious intuitive appeal of this selection there are four advantages to be noted for making the selection $c = 0$. First of all, the moments computed in the previous section are simplified computationally. Secondly, the expression for σ^2 is a quadratic function of c and assumes its minimum value at $c = \frac{k+1}{2}$. However this value does not infer the inferiority of choice which c must exhibit. On the other hand $c = 0$ minimizes σ^2 subject to the constraints that the above criterion be satisfied and that c be an integer.

The third advantage is that the statistic S reduces to the Friedman statistic given in Section 1 of Chapter II, for $k = p - 1$ as well as for the obvious case $k = p$. Thus S has already been tabulated for these values of k. Finally, investigations into the Asymptotic Relative Efficiency (A. R. E.) show that there is no unique optimum value c. However c = 0 does occur as the optimum value corresponding to various values of n, p and k for certain assumed underlying distributions and becomes an advantageous choice in tabulating the distributions of S.

Based on the convention described above each judge will assign some permutation of the scores 0, with multiplicity p - k, and 1, 2, ..., k, each with multiplicity one, to the p objects. There are $\frac{p!}{(p-k)!}$ distinct permutations of these scores possible where each permutation may be represented as a p-vector of scores. Under the null hypothesis each of these permutations is equally likely. Since the vectors of scores assigned by the n judges are mutually independent, there are $[\frac{p!}{(p-k)!}]^n$ permutations of scores possible for fixed n, p and k. Again these permutations are equally likely.

Corresponding to each of these $N = [\frac{p!}{(p-k)!}]^n$ permutations of scores assigned by the n judges is a p-vector of column totals, $(r_1, r_2, \ldots, r_p)$, where r_j is defined as in (1.1). Hence the statistic S may be evaluated for each of the N vectors of column totals. Let N_s denote the number of permutations for which S = s, then $P(S = s) = \frac{N_s}{N}$.

For fixed p and k the distribution of S may be constructed sequentially for successive values of n. For example, once the $[\frac{p!}{(p-k)!}]^m$ column-total vectors are obtained for n = m the distribution for n = m + 1 is obtained by adding each of the $\frac{p!}{(p-k)!}$ scoring vectors to each of the column-total vectors obtained at n = m. This procedure is quite laborious and may be made computationally more efficient by selective editing.

The editing will be facilitated by first observing that two or more column-total vectors which represent different permutations of the same collection of p integers will yield identical values of the statistic S. Secondly observe that adding each of the scoring vectors to two or more such column-total vectors (obtained for $n = m$) results in identical sets of column-total vectors at $n = m + 1$. Thus for $n = 1$ each of the column-total vectors is some permutation of the p-vector $(k, k-1, \ldots, 2, 1, 0, \ldots, 0)$ and, hence, $S = k(k+1)\frac{[(p-k)(k+1) + pk]}{12p}$ with probability one. For purposes of calculating the distribution of S at $n = 2$ only the p-vector $(k, k-1, \ldots, 2, 1, 0, \ldots, 0)$ needs to be retained. It may be carried forward with a frequency of one instead of $\frac{p!}{(p-k)!}$ without loss of generality. The distribution at $n = 2$ may now be calculated by adding each permutation of this vector to itself and calculating the relative frequency for each distinct value of S generated.

In general the distribution at $n = m + 1$ may be generated from that at $n = m$ in the following manner. Tabulate the distinct column-total vectors obtained when the elements of each column-total vector generated at $n = m$ are ordered and note the frequency associated with each of the distinct vectors with ordered elements. Then add each scoring vector to each of these distinct column-total vectors. Let the i^{th} distinct column-total vector occur with frequency $f_{i,m}$ at $n = m$. Now the column-total vector generated by adding any scoring vector occurs with frequency $f_{i,m}$ in the distribution at $n = m + 1$ as a result of this single action. Other similar actions may generate the same vector in which the corresponding frequencies will be accumulated.

The following example is given to illustrate the previous discussion.

Example 3.1: Let $p = 3$, $k = 1$. The three basic scoring vectors which can occur are $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, and under H_0, they are all equally likely. Now, fix one vector for the first judge's scores and consider the possible scoring vectors for two judges. Let $(1, 0, 0)$ be the fixed vector. Then the possible column-total vectors are,

$$(1, 0, 0) + (1, 0, 0) = (2, 0, 0)$$

$$(1, 0, 0) + (0, 1, 0) = (1, 1, 0)$$

$$(1, 0, 0) + (0, 0, 1) = (1, 0, 1)$$

$\bar{r} = \frac{2}{3}$ for each column-total vector. Now, consider the value of S for each of these vectors.

$$S = (2 - \tfrac{2}{3})^2 + (0 - \tfrac{2}{3})^2 + (0 - \tfrac{2}{3})^2 = 2.667 \quad \text{for} \quad (2, 0, 0)$$

$$S = (1 - \tfrac{2}{3})^2 + (1 - \tfrac{2}{3})^2 + (0 - \tfrac{2}{3})^2 = 0.667 \quad \text{for} \quad (1, 1, 0)$$

$$S = (1 - \tfrac{2}{3})^2 + (0 - \tfrac{2}{3})^2 + (1 - \tfrac{2}{3})^2 = 0.667 \quad \text{for} \quad (1, 0, 1)$$

The distribution of S for $n = 2$ is:

x	0.667	2.667
$P(S = x)$	$\frac{2}{3}$	$\frac{1}{3}$

$(1, 1, 0)$ and $(1, 0, 1)$ are different arrangements of the same scores.

Therefore, consider (1, 1, 0) with a weight of 2 and (2, 0, 0) with a weight of 1 and add the basic scoring vectors to these two only.

The resulting possible column-total vectors with their appropriate weights and S values are

Vector	Weight	S
(2, 1, 0)	2	2
(1, 2, 0)	2	2
(1, 1, 1)	2	0
(3, 0, 0)	1	6
(2, 1, 0)	1	2
(2, 0, 1)	1	2

The distribution of S is:

x	0	2	6
P(S = x)	$\frac{2}{9}$	$\frac{6}{9}$	$\frac{1}{9}$

In finding the distribution of S for $n = 4$, the necessary column-total vectors and their weights are:

$$(2, 1, 0),\ 6\ ;\ (1, 1, 1),\ 2\ ;\ (3, 0, 0),\ 1\ .$$

The three basic scoring vectors are added to (2, 1, 0), (1, 1, 1) and (3, 0, 0) and the appropriate weights are attached to the nine resulting vectors.

The process may be continued in this sequential manner until the desired number of cycles is completed.

Using an IBM-360, mod 65 computer, the exact distribution for S for the following cases is found in Table I: $p = 3$, $k = 1$, $3 \leq n \leq 10$; $p = 4$, $k = 1$, $3 \leq n \leq 10$; $p = 4$, $k = 2$, $3 \leq n \leq 8$; $p = 5$, $k = 1$, $3 \leq n \leq 7$; $p = 5$, $k = 2$, $3 \leq n \leq 4$. Refer to tabulations of Friedman's statistic in [4], [5] and [8] for $p = 3$, $k = 2$; $p = 4$, $k = 3$, $k = 4$; $p = 5$, $k = 4$, $k = 5$ and for various values of n.

It is clear that for larger values of n, an approximation to the distribution of S must be found. Two approximations will be discussed, a chi-square approximation and a beta approximation.

3. Approximations to the Distribution of S

Sen [10] points out that S_n, defined in Equation (2.4), under H_0, converges in law to a chi-square distribution with $(p-1)$ degrees of freedom. The form which S_n takes for a partial ordering will now be investigated. The general discussion concerning S_n may be found in Section 7 of Chapter II.

The scores $J(r, p)$ are the scores $1, 2, \ldots, k$ and 0 with a multiplicity of $(p-k)$. $\overline{J}$, $A^2(J)$ and T_{nj} defined in Equations (2.1), (2.2) and (2.3) take the following special forms,

$$\overline{J} = \frac{1}{p} \sum_{r=1}^{p} J(r, p) = \mu ,$$

$$A^2(J) = (p-1)^{-1} \sum_{r=1}^{p} \{J(r, p) - \overline{J}\}^2 = \frac{p}{p-1} \sigma^2$$

TABLE I

DISTRIBUTION OF S: (P = Prob. (S $\geq$ x))

x	P
p = 3, k = 1	
n = 3	
0.000	1.000
2.000	0.778
6.000	0.111
n = 4	
0.667	1.000
2.667	0.556
4.667	0.333
10.667	0.037
n = 5	
0.667	1.000
2.667	0.630
4.667	0.383
8.667	0.136
16.667	0.012
n = 6	
0.000	1.000
2.000	0.878
6.000	0.383
8.000	0.177
14.000	0.053
24.000	$0.0^{2}41$
n = 7	
0.667	1.000
2.667	0.712
4.667	0.520
8.667	0.232
10.667	0.136
12.667	0.078
20.667	0.021
32.667	$0.0^{2}14$
n = 8	
0.667	1.000
2.667	0.744
4.667	0.552
8.667	0.296
10.667	0.142
12.667	0.110
16.667	0.059
18.667	0.033
28.667	$0.0^{2}78$
42.667	$0.0^{3}46$
n = 9	
0.000	1.000
2.000	0.915
6.000	0.531
8.000	0.319
14.000	0.166
18.000	0.050
24.000	0.025
26.000	0.014
38.000	$0.0^{2}29$
54.000	$0.0^{3}15$
n = 10	
0.667	1.000
2.667	0.787
4.667	0.627
8.667	0.371
10.667	0.242
12.667	0.178
16.667	0.093
18.667	0.080
20.667	0.059
24.667	0.022
32.667	0.010
34.667	$0.0^{2}56$
48.667	$0.0^{2}11$
66.667	$0.0^{4}51$
p = 4, k = 1	
n = 3	
0.750	1.000
2.750	0.625
6.750	0.063
n = 4	
0.000	1.000
2.000	0.906
4.000	0.344
6.000	0.203
12.000	0.016
n = 5	
0.750	1.000
2.750	0.766
4.750	0.414
6.750	0.180
10.750	0.063
18.750	$0.0^{2}39$
n = 6	
1.000	1.000
3.000	0.736
5.000	0.531
9.000	0.180
11.000	0.063
17.000	0.019
27.000	$0.0^{3}98$
n = 7	
0.750	1.000
2.750	0.846
4.750	0.539
6.750	0.385
8.750	0.231
12.750	0.077
14.750	0.052
16.750	0.021
24.750	$0.0^{2}54$
36.750	$0.0^{3}24$
n = 8	
0.000	1.000
2.000	0.962
4.000	0.654
6.000	0.551
8.000	0.295
10.000	0.218
12.000	0.116
14.000	0.095
16.000	0.034
18.000	0.027
22.000	0.017

TABLE I (Continued)

x	P
n = 8	
24.000	$0.0^{2}67$
34.000	$0.0^{2}15$
48.000	$0.0^{4}61$
n = 9	
0.750	1.000
2.750	0.885
4.750	0.654
6.750	0.481
8.750	0.340
10.750	0.225
12.750	0.155
14.750	0.092
18.750	0.046
20.750	0.038
24.750	$0.0^{2}92$
30.750	$0.0^{2}54$
32.750	$0.0^{2}21$
44.750	$0.0^{3}43$
60.750	$0.0^{4}15$
n = 10	
1.000	1.000
3.000	0.856
5.000	0.720
9.000	0.431
11.000	0.261
13.000	0.167
17.000	0.109
19.000	0.052
21.000	0.037
25.000	0.018
27.000	0.016
29.000	0.011
33.000	$0.0^{2}30$
41.000	$0.0^{2}17$
43.000	$0.0^{3}63$
57.000	$0.0^{3}12$
75.000	$0.0^{5}38$
p = 4, k = 2	
n = 3	
0.750	1.000
n = 3	
2.750	0.903
4.750	0.778
6.750	0.611
8.750	0.500
10.750	0.292
12.750	0.250
14.750	0.167
18.750	0.083
20.750	0.069
24.750	$0.0^{2}69$
n = 4	
0.000	1.000
2.000	0.990
4.000	0.844
6.000	0.771
8.000	0.590
10.000	0.535
12.000	0.424
14.000	0.385
16.000	0.228
18.000	0.214
20.000	0.138
22.000	0.098
24.000	0.084
26.000	0.071
30.000	0.034
34.000	0.020
36.000	0.013
38.000	$0.0^{2}75$
44.000	$0.0^{3}58$
n = 5	
0.750	1.000
2.750	0.957
4.750	0.861
6.750	0.770
8.750	0.676
10.750	0.546
12.750	0.506
14.750	0.421
16.750	0.325
18.750	0.294
20.750	0.246
22.750	0.195
24.750	0.162
n = 5	
26.750	0.124
28.750	0.091
30.750	0.071
32.750	0.064
34.750	0.041
36.750	0.036
38.750	0.028
42.750	0.020
44.750	0.013
50.750	$0.0^{2}66$
52.750	$0.0^{2}46$
54.750	$0.0^{2}32$
56.750	$0.0^{2}22$
60.750	$0.0^{3}77$
68.750	$0.0^{4}48$
n = 6	
1.000	1.000
3.000	0.952
5.000	0.898
9.000	0.758
11.000	0.630
13.000	0.542
17.000	0.466
19.000	0.359
21.000	0.316
25.000	0.240
27.000	0.206
29.000	0.175
33.000	0.120
35.000	0.094
37.000	0.073
41.000	0.063
43.000	0.039
45.000	0.035
49.000	0.024
51.000	0.019
53.000	0.015
57.000	0.010
59.000	$0.0^{2}82$
61.000	$0.0^{2}60$
65.000	$0.0^{2}43$
67.000	$0.0^{2}29$
69.000	$0.0^{2}24$
75.000	$0.0^{2}13$
77.000	$0.0^{2}12$

TABLE I (Continued)

x	P
n = 6	
81.000	$0.0^{3}50$
83.000	$0.0^{3}26$
89.000	$0.0^{4}76$
99.000	$0.0^{5}40$
n = 7	
0.750	1.000
2.750	0.974
4.750	0.904
6.750	0.844
8.750	0.772
10.750	0.684
12.750	0.641
14.750	0.571
16.750	0.480
18.750	0.455
20.750	0.402
22.750	0.339
24.750	0.305
26.750	0.263
28.750	0.226
30.750	0.209
32.750	0.187
34.750	0.146
36.750	0.123
38.750	0.111
40.750	0.093
42.750	0.089
44.750	0.072
46.750	0.057
48.750	0.050
50.750	0.039
52.750	0.033
54.750	0.029
56.750	0.025
58.750	0.021
60.750	0.019
62.750	0.017
64.750	0.011
66.750	$0.0^{2}98$
68.750	$0.0^{2}81$
72.750	$0.0^{2}59$
74.750	$0.0^{2}55$
76.750	$0.0^{2}37$
78.750	$0.0^{2}34$
80.750	$0.0^{2}26$
n = 7	
82.750	$0.0^{2}23$
84.750	$0.0^{2}20$
86.750	$0.0^{2}14$
88.750	$0.0^{2}12$
90.750	$0.0^{3}90$
92.750	$0.0^{3}76$
96.750	$0.0^{3}51$
98.750	$0.0^{3}44$
102.750	$0.0^{3}29$
104.750	$0.0^{3}22$
106.750	$0.0^{3}18$
108.750	$0.0^{3}11$
110.750	$0.0^{4}78$
112.750	$0.0^{4}43$
114.750	$0.0^{4}28$
122.750	$0.0^{5}74$
134.750	$0.0^{6}33$
n = 8	
0.000	1.000
2.000	0.994
4.000	0.933
6.000	0.906
8.000	0.809
10.000	0.766
12.000	0.690
14.000	0.668
16.000	0.550
18.000	0.537
20.000	0.468
22.000	0.427
24.000	0.392
26.000	0.360
30.000	0.277
32.000	0.234
34.000	0.225
36.000	0.193
38.000	0.175
40.000	0.138
42.000	0.127
44.000	0.109
46.000	0.101
48.000	0.087
50.000	0.085
52.000	0.066
54.000	0.061
n = 8	
56.000	0.045
58.000	0.038
62.000	0.035
64.000	0.027
66.000	0.026
68.000	0.020
70.000	0.017
72.000	0.015
74.000	0.014
76.000	$0.0^{2}94$
78.000	$0.0^{2}89$
80.000	$0.0^{2}79$
82.000	$0.0^{2}74$
84.000	$0.0^{2}64$
86.000	$0.0^{2}57$
88.000	$0.0^{2}42$
90.000	$0.0^{2}38$
94.000	$0.0^{2}27$
96.000	$0.0^{2}20$
98.000	$0.0^{2}18$
100.000	$0.0^{2}14$
102.000	$0.0^{2}14$
104.000	$0.0^{2}12$
106.000	$0.0^{3}98$
108.000	$0.0^{3}84$
110.000	$0.0^{3}76$
114.000	$0.0^{3}48$
116.000	$0.0^{3}37$
118.000	$0.0^{3}29$
120.000	$0.0^{3}27$
122.000	$0.0^{3}24$
126.000	$0.0^{3}14$
132.000	$0.0^{4}99$
134.000	$0.0^{4}90$
136.000	$0.0^{4}55$
138.000	$0.0^{4}43$
142.000	$0.0^{4}27$
144.000	$0.0^{4}18$
146.000	$0.0^{4}15$
150.000	$0.0^{5}78$
152.000	$0.0^{5}63$
162.000	$0.0^{5}42$
176.000	$0.0^{5}36$

TABLE I (Continued)

x	P
p = 5, k = 1	
n = 3	
1.200	1.000
3.200	0.520
7.200	0.040
n = 4	
0.800	1.000
2.800	0.808
4.800	0.232
6.800	0.136
12.800	$0.0^{2}80$
n = 5	
0.000	1.000
2.000	0.962
4.000	0.578
6.000	0.290
8.000	0.098
12.000	0.034
20.000	$0.0^{2}16$
n = 6	
0.800	1.000
2.800	0.885
4.800	0.539
6.800	0.328
10.800	0.098
12.800	0.027
18.800	$0.0^{2}80$
28.800	$0.0^{3}32$
n = 7	
1.200	1.000
3.200	0.839
5.200	0.624
7.200	0.301
9.200	0.220
11.200	0.113
15.200	0.032
17.200	0.023
19.200	$0.0^{2}72$
27.200	$0.0^{2}19$
39.200	$0.0^{4}64$
p = 5, k = 2	
n = 3	
0.800	1.000
2.800	0.955
4.800	0.865
6.800	0.760
8.800	0.580
10.800	0.400
12.800	0.295
14.800	0.183
16.800	0.137
18.800	0.092
22.800	0.048
24.800	0.033
28.800	$0.0^{2}25$
n = 4	
1.200	1.000
3.200	0.960
5.200	0.903
7.200	0.759
9.200	0.721
11.200	0.556
13.200	0.464
15.200	0.347
17.200	0.302
19.200	0.207
21.200	0.179
23.200	0.100
25.200	0.091
27.200	0.059
29.200	0.043
31.200	0.034
33.200	0.025
37.200	0.013
39.200	$0.0^{2}89$
41.200	$0.0^{2}81$
43.200	$0.0^{2}36$
45.200	$0.0^{2}21$
51.200	$0.0^{3}13$

and

$$T_{nj} = n^{-1} \sum_{i=1}^{n} J(r_{ij}, p) = \frac{r_j}{n} .$$

S_n, then may be written as

$$S_n = n \frac{p-1}{p\sigma^2} \sum_{j=1}^{n} \left(\frac{r_j}{n} - \mu \right)^2$$

$$= \frac{p-1}{np\sigma^2} \sum_{j=1}^{p} (r_j - n\mu)^2$$

$$= \frac{p-1}{np\sigma^2} \sum_{j=1}^{p} (r_j - \bar{r})^2$$

where

$$\sigma^2 = \frac{k(k+1)}{p} \left[\frac{2k+1}{6} - \frac{k(k+1)}{4p} \right] .$$

Instead of using S_n, for this statistic, the symbol χ_r^2 will be used as has been the custom of other authors defining similar statistics, i.e.,

$$\chi_r^2 = \frac{p-1}{np\sigma^2} \sum_{j=1}^{p} (r_j - \bar{r})^2 . \qquad (3.6)$$

The next question to arise is how well the chi-square distribution approximates the distribution of χ_r^2. Table II contains percent relative errors for the chi-square approximation in the tail area of the distribution. The x value is the value of S, P is the probability of S taking that value or a larger one. Let

$$\alpha = P\left[\frac{p-1}{np\sigma^2} S \geq \frac{p-1}{np\sigma^2} x \right] = P[S \geq x]$$

and

$$\hat{\alpha} = P\left[\chi_{(p-1)}^2 \geq \frac{p-1}{np\sigma^2} x \right]$$

TABLE II

PERCENT ERRORS FOR THE CHI-SQUARE APPROXIMATION IN THE TAIL REGION

x	P	Error
	p = 3, k = 1	
	n = 3	
6.000	.111	- 55
	n = 4	
4.667	.333	- 48
10.667	.037	- 51
	n = 5	
8.667	.136	- 45
16.667	.012	- 45
	n = 6	
8.000	.117	- 23
14.000	.053	- 44
24.000	$.0^{2}41$	- 40
	n = 7	
10.667	.136	- 25
12.667	.078	- 15
20.667	.021	- 42
32.667	$.0^{2}14$	- 33
	n = 8	
12.667	.110	- 16
16.667	.059	- 26
18.667	.033	- 10
28.667	$.0^{2}78$	- 40
42.667	$.0^{3}46$	- 27
	n = 9	
14.000	.166	- 41
18.000	.050	- 1
24.000	.025	- 26
26.000	.014	- 5
38.000	$.0^{2}29$	- 39
54.000	$.0^{3}15$	- 19
	n = 10	
12.667	.178	- 16
16.667	.093	- 12
18.667	.080	- 24
20.667	.059	- 23
	n = 10	
24.667	.022	10
32.667	.010	- 27
34.667	$.0^{2}56$	- 2
48.667	$.0^{2}11$	- 36
66.667	$.0^{4}51$	11
	p = 3, k = 2	
	n = 3	
14.000	.194	- 50
18.000	.028	79
	n = 4	
18.000	.125	- 16
24.000	.069	- 28
26.000	.042	- 7
32.000	$.0^{2}46$	>100
	n = 5	
24.000	.124	- 27
26.000	.093	- 20
32.000	.039	4
38.000	.024	- 6
42.000	$.0^{2}85$	77
50.000	$.0^{3}77$	>100
	n = 6	
26.000	.142	- 19
32.000	.072	- 4
38.000	.052	- 19
42.000	.029	4
50.000	.012	30
54.000	$.0^{2}81$	37
56.000	$.0^{2}55$	70
62.000	$.0^{2}17$	>100
72.000	$.0^{3}13$	>100
	n = 7	
32.000	.112	- 9
38.000	.085	- 22
42.000	.051	- 3
50.000	.027	3
54.000	.021	2

TABLE II (Continued)

x	P	Error
	n = 7	
56.000	.016	13
62.000	$.0^{2}84$	42
72.000	$.0^{2}36$	61
74.000	$.0^{2}27$	86
78.000	$.0^{2}12$	>100
86.000	$.0^{3}32$	>100
98.000	$.0^{4}21$	>100
	n = 8	
38.000	.120	- 22
42.000	.079	- 8
50.000	.047	- 6
54.000	.038	- 9
56.000	.030	- .4
62.000	.018	16
72.000	$.0^{2}99$	12
74.000	$.0^{2}80$	23
78.000	$.0^{2}48$	60
86.000	$.0^{2}24$	96
96.000	$.0^{2}11$	>100
98.000	$.0^{3}86$	>100
104.000	$.0^{3}26$	>100
114.000	$.0^{4}61$	>100
128.000	$.0^{5}36$	>100
	n = 9	
42.000	.107	- 9
50.000	.069	- 10
54.000	.057	- 13
56.000	.048	- 6
62.000	.031	4
72.000	.019	- 3
74.000	.016	4
78.000	.010	28
86.000	$.0^{2}60$	40
96.000	$.0^{2}35$	37
98.000	$.0^{2}29$	47
104.000	$.0^{2}13$	>100
114.000	$.0^{3}66$	>100
122.000	$.0^{3}35$	>100
126.000	$.0^{3}20$	>100
128.000	$.0^{4}97$	>100
134.000	$.0^{4}54$	>100
146.000	$.0^{4}11$	>100
162.000	$.0^{6}60$	>100

x	P	Error
	n = 10	
42.000	.135	- 9
50.000	.092	- 11
54.000	.078	- 14
56.000	.066	- 9
62.000	.046	- 1
72.000	.030	- 10
74.000	.026	- 4
78.000	.018	13
86.000	.012	18
96.000	$.0^{2}75$	10
98.000	$.0^{2}63$	17
104.000	$.0^{2}34$	64
114.000	$.0^{2}20$	68
122.000	$.0^{2}13$	79
126.000	$.0^{3}83$	>100
128.000	$.0^{3}51$	>100
134.000	$.0^{3}37$	>100
146.000	$.0^{3}18$	>100
150.000	$.0^{3}11$	>100
152.000	$.0^{4}85$	>100
158.000	$.0^{4}44$	>100
162.000	$.0^{4}20$	>100
168.000	$.0^{4}11$	>100
182.000	$.0^{4}21$	>100
200.000	$.0^{7}99$	>100
	p = 4, k = 1	
	n = 3	
2.750	.625	- 52
6.750	.063	- 53
	n = 4	
6.000	.203	- 45
12.000	.016	- 53
	n = 5	
6.750	.180	- 19
10.750	.063	- 44
18.750	$.0^{2}39$	- 53
	n = 6	
9.000	.180	- 38
11.000	.063	- 1
17.000	.019	- 46
27.000	$.0^{3}98$	- 55

TABLE II (Continued)

x	P	Error
	n = 7	
8.750	.231	- 26
12.750	.077	- 18
14.750	.052	- 26
16.750	.021	9
24.750	$.0^{2}54$	- 49
36.750	$.0^{3}24$	- 57
	n = 8	
12.000	.116	- 3
14.000	.095	- 24
16.000	.034	37
18.000	.027	8
22.000	.017	- 31
24.000	$.0^{2}67$	11
34.000	$.0^{2}15$	- 54
48.000	$.0^{4}61$	- 59
	n = 9	
12.000	.155	- 17
14.750	.092	- 5
18.750	.046	- 13
20.750	.038	- 30
24.750	$.0^{2}92$	27
30.750	$.0^{2}54$	- 37
32.750	$.0^{2}21$	8
44.750	$.0^{3}43$	- 58
60.750	$.0^{4}15$	- 61
	n = 10	
17.000	.109	- 28
19.000	.052	7
21.000	.037	4
25.000	.018	4
27.000	.016	- 22
29.000	.011	- 21
33.000	$.0^{2}30$	39
41.000	$.0^{2}17$	- 43
43.000	$.0^{3}63$	2
57.000	$.0^{3}12$	- 62
75.000	$.0^{5}38$	- 63
	p = 4, k = 2	
	n = 3	
14.750	.167	- 12
18.750	.083	- 6
20.750	.069	- 19
24.750	$.0^{2}69$	>100
	n = 4	
20.000	.138	3
22.000	.098	13
24.000	.084	4
26.000	.071	- 2
30.000	.034	26
34.000	.020	32
36.000	.012	58
38.000	$.0^{2}75$	>100
44.000	$.0^{3}58$	>100
	n = 5	
26.750	.124	- 3
28.750	.091	9
30.750	.071	15
32.750	.064	5
34.750	.041	34
36.750	.036	25
38.750	.028	35
42.750	.020	26
44.750	.013	55
50.750	$.0^{2}66$	73
52.750	$.0^{2}46$	>100
54.750	$.0^{2}32$	>100
56.750	$.0^{2}22$	>100
60.750	$.0^{3}77$	>100
68.750	$.0^{4}48$	>100
	n = 6	
33.000	.120	- 7
35.000	.094	1
37.000	.074	12
41.000	.063	- 7
43.000	.039	29
45.000	.035	23
49.000	.024	28
51.000	.019	37
53.000	.015	49
57.000	.010	51
59.000	$.0^{2}82$	61
61.000	$.0^{2}60$	87
65.000	$.0^{2}43$	85
67.000	$.0^{2}29$	>100

TABLE II (Continued)

x	P	Error
	n = 6	
69.000	$.0^{2}24$	>100
75.000	$.0^{2}13$	>100
77.000	$.0^{2}12$	>100
81.000	$.0^{3}50$	>100
83.000	$.0^{3}26$	>100
89.000	$.0^{4}76$	>100
99.000	$.0^{5}40$	>100
	n = 7	
38.750	.111	-1
40.750	.093	3
42.750	.089	-6
44.750	.072	1
46.750	.057	10
48.750	.050	9
50.750	.039	23
52.750	.033	25
54.750	.029	23
56.750	.025	24
58.750	.021	30
60.750	.019	27
62.750	.017	24
64.750	.011	56
66.750	$.0^{2}98$	57
68.750	$.0^{2}81$	65
72.750	$.0^{2}59$	71
74.750	$.0^{2}55$	57
76.750	$.0^{2}37$	>100
78.750	$.0^{2}34$	93
80.750	$.0^{2}26$	>100
82.750	$.0^{2}23$	>100
84.750	$.0^{2}20$	>100
86.750	$.0^{2}14$	>100
88.750	$.0^{2}12$	>100
90.750	$.0^{3}90$	>100
92.750	$.0^{3}76$	>100
96.750	$.0^{3}51$	>100
98.750	$.0^{3}44$	>100
102.750	$.0^{3}29$	>100
104.750	$.0^{3}22$	>100
106.750	$.0^{3}18$	>100
108.750	$.0^{3}11$	>100
110.750	$.0^{4}78$	>100
112.750	$.0^{4}43$	>100
114.750	$.0^{3}28$	>100

x	P	Error
	n = 7	
122.750	$.0^{5}74$	>100
134.750	$.0^{6}33$	>100
	n = 8	
46.000	.101	-2
48.000	.087	1
50.000	.085	-8
52.000	.066	5
54.000	.061	-1
56.000	.045	20
58.000	.038	25
62.000	.035	6
64.000	.027	22
66.000	.026	11
68.000	.020	29
70.000	.017	33
72.000	.015	34
74.000	.014	31
76.000	$.0^{2}94$	67
78.000	$.0^{2}89$	56
80.000	$.0^{2}79$	55
82.000	$.0^{2}74$	45
84.000	$.0^{2}64$	49
86.000	$.0^{2}57$	46
88.000	$.0^{2}42$	77
90.000	$.0^{2}38$	71
94.000	$.0^{2}27$	87
96.000	$.0^{2}20$	>100
98.000	$.0^{2}18$	>100
100.000	$.0^{2}14$	>100
102.000	$.0^{2}14$	>100
104.000	$.0^{2}12$	>100
106.000	$.0^{3}98$	>100
108.000	$.0^{3}84$	>100
110.000	$.0^{3}76$	>100
114.000	$.0^{3}48$	>100
116.000	$.0^{3}37$	>100
118.000	$.0^{3}29$	>100
120.000	$.0^{3}27$	>100
122.000	$.0^{3}24$	>100
126.000	$.0^{3}14$	>100
132.000	$.0^{4}99$	>100
134.000	$.0^{4}90$	>100
136.000	$.0^{4}55$	>100
138.000	$.0^{4}43$	>100

TABLE II (Continued)

x	P	Error
	n = 8	
142.000	$.0^{4}27$	>100
144.000	$.0^{4}18$	>100
146.000	$.0^{4}15$	>100
150.000	$.0^{5}78$	>100
152.000	$.0^{5}63$	>100
162.000	$.0^{5}42$	>100
176.000	$.0^{5}36$	>100
	<u>p = 5, k = 1</u>	
	n = 3	
3.200	.520	- 51
7.200	.040	- 57
	n = 4	
6.800	.136	- 45
12.800	$.0^{2}80$	- 62
	n = 5	
6.000	.290	- 31
8.000	.098	- 6
12.000	.034	- 48
20.000	$.0^{2}16$	- 69
	n = 6	
6.800	.328	- 31
10.800	.098	- 37
12.800	.027	12
18.800	$.0^{2}80$	- 56
28.800	$.0^{3}32$	- 75
	n = 7	
11.200	.113	- 19
15.200	.032	- 13
17.200	.023	- 34
19.200	$.0^{2}72$	14
27.200	$.0^{2}19$	- 65
39.200	$.0^{4}64$	- 80
	<u>p = 5, k = 2</u>	
	n = 3	
16.800	.137	- 1
18.800	.092	6
22.800	.048	5
24.800	.033	8
28.800	$.0^{2}25$	>100
	n = 4	
23.200	.100	24
25.200	.091	6
27.200	.059	27
29.200	.043	34
31.200	.034	31
33.200	.025	36
37.200	.013	52
39.200	$.0^{2}89$	76
41.200	$.0^{2}81$	46
43.200	$.0^{2}36$	>100
45.200	$.0^{2}21$	>100
51.200	$.0^{3}13$	>100

Then the percent relative error is equal to $100(\hat{\alpha} - \alpha)/\alpha$. The $\hat{\alpha}$ values were obtained by use of the CDTR subroutine of the IBM scientific subroutine package. It will be noted that in the region, $.01 \leq P \leq .1$, the percent relative error is reasonably small. As in the usual case, the approximation becomes poorer the further out in the tail region one goes.

It is interesting to compare the first three moments of χ_r^2 and χ^2 as this gives some idea of how good the approximation is.

It is well known that the first three moments of $\chi^2_{(p-1)}$ are

$$E[\chi^2_{(p-1)}] = p-1$$

$$E[\chi^2_{(p-1)} - (p-1)]^2 = 2(p-1)$$

$$E[\chi^2_{(p-1)} - (p-1)]^3 = 8(p-1) .$$

The first three moments of χ_r^2 will now be derived. First of all, χ_r^2 will be simplified as follows.

$$\chi_r^2 = \frac{p-1}{n p \sigma^2} \cdot S$$

$$= \frac{p-1}{n p \sigma^2} \sum_{j=1}^{p} (r_j - n\mu)^2$$

$$= \frac{p-1}{n p \sigma^2} \sum_{j=1}^{p} \left(\sum_{i=1}^{n} (r_{ij} - \mu) \right)^2$$

$$= \frac{p-1}{n p \sigma^2} \sum_{j=1}^{p} \left(\sum_{i=1}^{n} d_{ij} \right)^2$$

where, $d_{ij} = r_{ij} - \mu$

$$\chi_r^2 = \frac{p-1}{n p \sigma^2} \sum_{j=1}^{p} \left(\sum_{i=1}^{n} d_{ij}^2 + 2 \sum_{i \neq i'} \sum d_{ij} d_{i'j} \right)$$

$$= \frac{p-1}{n p \sigma^2} \left(\sum_{i=1}^{n} \sum_{j=1}^{p} d_{ij}^2 + 2 \sum_{i \neq i'} \sum \sum_{j=1}^{p} d_{ij} d_{i'j} \right)$$

$$= \frac{p-1}{n p \sigma^2}\left(n p \sigma^2 + 2 \sum_{i \neq i'}\sum \sum_{j=1}^{p} d_{ij} d_{i'j}\right)$$

$$= (p-1) + \frac{2(p-1)}{n p \sigma^2} \sum_{i \neq i'}\sum \left(\sum_{j=1}^{p} d_{ij} d_{i'j}\right) ,$$

where $\sum_{j=1}^{p} d_{ij}^2 = p\sigma^2$. Now, define $R_{ii'}$ as follows:

$$R_{ii'} = \left(\sum_{j=1}^{p} d_{ij} d_{i'j}\right) .$$

Then

$$\chi_r^2 = (p-1) + \frac{2(p-1)}{n p \sigma^2} \sum_{i \neq i'}\sum R_{ii'} . \qquad (3.8)$$

Now,

$$E[R_{ii'}] = E\left(\sum_{j=1}^{p} d_{ij} d_{i'j}\right)$$

$$= \sum_{j=1}^{p} E(d_{ij} d_{i'j})$$

$$= \sum_{j=1}^{p} E(d_{ij}) E(d_{i'j})$$

$$= 0$$

since there is independence from row to row and $E(d_{ij}) = 0$. It then follows from (3.8) that

$$E(\chi_r^2) = p - 1 .$$

Now consider the variance of χ_r^2.

$$E[\chi_r^2 - (p-1)]^2 = \left(\frac{2(p-1)}{np\sigma^2}\right)^2 E\left(\sum_{i \neq i'}\sum R_{ii'}\right)^2 . \qquad (3.9)$$

The term $\left(\sum_{i \neq i'}\sum R_{ii'}\right)^2$ when expanded contains terms of the form $R_{ii'}^2$, $R_{ii'}R_{i''i'''}$ and $R_{ii'}R_{ii''}$. Consider each of these terms.

$$R_{ii'}^2 = \left(\sum_{j=1}^{p} d_{ij}d_{i'j}\right)^2$$

$$= \sum_{j=1}^{p} d_{ij}^2 d_{i'j}^2 + 2 \sum_{j > j'}\sum d_{ij}d_{i'j}d_{ij'}d_{i'j'} .$$

$$E(R_{ii'}^2) = \sum_{j=1}^{p} E(d_{ij}^2)E(d_{i'j}^2) + 2 \sum_{j > j'}\sum E(d_{ij}d_{ij'})E(d_{i'j}d_{i'j'})$$

$$= p(\sigma^2)^2 + p(p-1)\,\mathrm{cov}\,(r_{ij}, r_{ij'})^2$$

$$= p(\sigma^2)^2 + p(p-1) \cdot \frac{1}{(p-1)^2}(\sigma^2)^2 = \frac{p^2}{p-1}(\sigma^2)^2 . \qquad (3.10)$$

$$E(R_{ii'}R_{i''i'''}) = E\left[\sum_{j=1}^{p} d_{ij}d_{i'j} \sum_{j=1}^{p} d_{i''j}d_{i'''j}\right]$$

$$= 0 . \qquad (3.11)$$

$$E(R_{ii'}R_{ii''}) = E\left[\sum_{j=1}^{p} d_{ij}d_{i'j} \sum_{j=1}^{p} d_{ij}d_{i''j}\right]$$

$$= E\left[\sum_{j=1}^{p} d_{ij}^2 d_{i'j}d_{i''j} + \sum_{j \neq j'}\sum d_{ij}d_{ij'}d_{i'j}d_{i''j'}\right]$$

$$= 0 . \qquad (3.12)$$

Now from (3.9), (3.10), (3.11) and (3.12),

$$E[\chi_r^2 - (p-1)]^2 = \left(\frac{2(p-1)}{n p \sigma^2}\right)^2 \cdot \frac{n(n-1)}{2} \, E(R_{ii'}^2)$$

$$= 2 \cdot \frac{n-1}{n} (p-1) \quad .$$

The third moment of χ_r^2 about the mean will now be derived.

$$E[\chi_r^2 - (p-1)]^3 = \left(\frac{2(p-1)}{n p \sigma^2}\right)^3 E\left(\sum_{i \neq i'}\sum R_{ii'}\right)^3 \qquad (3.13)$$

The only terms in $\left(\sum_{i \neq i'}\sum R_{ii'}\right)^3$ with non-zero expectation are terms of the form $R_{ii'}^3$ and $R_{ii'} R_{i''i'} R_{i''i}$.

$$R_{ii'}^3 = \left(\sum_{j=1}^{p} d_{ij} d_{i'j}\right)^3$$

contains p terms of the form $d_{ij}^3 d_{i'j}^3$, $3p(p-1)$ terms of the form $d_{ij}^2 d_{i'j}^2 d_{ij'} d_{i'j'}$ and $p(p-1)(p-2)$ of the form $d_{ij} d_{i'j} d_{ij'} d_{i'j'} d_{ij''} d_{i'j''}$.

$$E(d_{ij}^3 d_{i'j}^3) = E(d_{ij}^3) E(d_{i'j}^3) = \mu_3^2$$

$$E(d_{ij}^2 d_{ij'}) = E_{d_{ij}} [d_{ij}^2 E_{d_{ij'}}(d_{ij'} | d_{ij}^2)]$$

$$= -\frac{1}{p-1} \mu_3 \quad ,$$

which implies that

$$E(d_{ij}^2 d_{i'j}^2 d_{ij'} d_{i'j'}) = E(d_{ij}^2 d_{ij'}) E(d_{i'j}^2 d_{i'j'})$$

$$= \frac{1}{(p-1)^2} \mu_3^2 \quad .$$

Now, consider $E(d_{ij} d_{ij'} d_{ij''})$.

$$E(d_{ij} d_{ij'} d_{ij''}) = E_{d_{ij}, d_{ij'}}[d_{ij} d_{ij'} E_{d_{ij''}}(d_{ij''} | d_{ij}, d_{ij'})]$$

$$= -\frac{1}{p-2} E_{d_{ij}, d_{ij'}}[d_{ij} d_{ij'}(d_{ij} + d_{ij'})]$$

$$= -\frac{1}{p-2} E_{d_{ij}, d_{ij'}}[d_{ij}^2 d_{ij'} + d_{ij} d_{ij'}^2]$$

$$= -\frac{2}{p-2} E(d_{ij}^2 d_{ij'})$$

$$= \frac{2}{(p-1)(p-2)} \mu_3 .$$

This implies that

$$E(d_{ij} d_{i'j} d_{ij'} d_{i'j'} d_{ij''} d_{i'j''}) = \frac{4}{(p-1)^2 (p-2)^2} \mu_3^2 .$$

It follows that

$$E(R_{ii'}^3) = p\mu_3^2 + 3p(p-1) \cdot \frac{1}{(p-1)^2} \mu_3^2 + p(p-1)(p-2) \cdot \frac{4}{(p-1)^2 (p-2)^2} \mu_3^2$$

$$= \mu_3^2 \left[p + \frac{3p}{p-1} + \frac{4p}{(p-1)(p-2)} \right]$$

$$= \frac{\mu_3^2}{(p-1)(p-2)} [p(p-1)(p-2) + 3p(p-2) + 4p]$$

$$= \frac{p^3}{(p-1)(p-2)} \mu_3^2 . \qquad (3.14)$$

In a similar fashion, it follows that

$$E(R_{ii'} R_{i''i'} R_{i''i}) = \left[E(d_{i'j}^2) - E(d_{i'j} d_{i'j'}) \right] E(R_{ii''}^2)$$

$$= \left[\sigma^2 + \frac{1}{p-1}\sigma^2\right] \frac{p^2}{p-1}(\sigma^2)^2$$

$$= \frac{p^3}{(p-1)^2}(\sigma^2)^3 \quad . \tag{3.15}$$

The third moment may now be found using (3.13), (3.14) and (3.15) to be

$$E[\chi_r^2 - (p-1)]^3 = \left(\frac{2(p-1)}{np\sigma^2}\right)^3 E\left(\sum_{i\neq i'}\sum R_{ii'}\right)^3 = \left(\frac{2(p-1)}{np\sigma^2}\right)^3$$

$$\cdot\left(\frac{n(n-1)}{2} \cdot \frac{p^3}{(p-1)(p-2)}\mu_3^2 + n(n-1)(n-2)\frac{p^3}{(p-1)^2}(\sigma^2)^3\right)$$

$$= \frac{4(p-1)^2(n-1)\mu_3^2}{n^2(p-2)(\sigma^2)^3} + \frac{8(n-1)(n-2)}{n^2}(p-1) \quad .$$

It may be noted that for n large, the first three moments of χ_r^2 are approximately those of $\chi^2_{(p-1)}$.

Following the approach of Kendall and Babington Smith [5], a statistic called the Coefficient of Concordance will be defined. Consider the maximum value of S, S_{max}. It occurs when the column totals r_j are some permutation of $n, 2n, \ldots, kn, \; 0 \; (p-k$ times).

$$S_{max} = \sum_{i=1}^{k}(ni - n\mu)^2 + (p-k)(0-n\mu)^2$$

$$= n^2 p\left\{\sum_{i=1}^{k}(i-\mu)^2 \cdot \frac{1}{p} + (0-\mu)^2 \cdot \frac{p-k}{p}\right\}$$

$$= n^2 p\sigma^2$$

The Coefficient of Concordance, W, then is defined to be

$$W = \frac{S}{n^2 p \sigma^2}, \quad 0 \leq W \leq 1 .$$

The exact tabulations of S provides exact tabulations of W since $P(S \geq x) = P\left(W \geq \frac{x}{n^2 p \sigma^2}\right)$. The distribution of W is approximated by the beta distribution. The beta distribution is given by

$$\beta(x; \ell, q) = \frac{1}{B(\ell, q)} x^{\ell - 1}(1 - x)^{q-1}, \quad (0 < x < 1; \ell > 0; q > 0)$$

The mean of the beta is $\frac{\ell}{\ell + q}$ and the variance is $\frac{\ell q}{(\ell + q)^2(\ell + q + 1)}$.

The mean and variance of W may be found by noting that

$$W = \frac{1}{n(p-1)} \chi_r^2 .$$

Then,

$$E(W) = \frac{1}{n(p-1)} (p-1) = \frac{1}{n}$$

$$\text{Var}(W) = \frac{1}{n^2(p-1)^2} \cdot 2 \frac{n-1}{n} (p-1) = 2 \frac{n-1}{n^3(p-1)} .$$

Now equating the means and variances of W and beta,

$$\frac{\ell}{\ell + q} = \frac{1}{n}$$

$$\frac{\ell q}{(\ell + q)^2(\ell + q + 1)} = 2 \frac{n-1}{n^3(p-1)}$$

one finds that

$$\ell = \frac{p-1}{2} - \frac{1}{n}$$

and

$$q = (n-1)\left(\frac{p-1}{2} - \frac{1}{n}\right) .$$

Kendall and Babington Smith point out in [5] that the third and fourth moments are also close in the full ordering. The fourth moment is not derived in this study for the partial ordering, but the third moments differ in the full and partial orderings and the fit will probably not be as good in the partial ordering since the third moment in the full ordering is $\frac{8(n-1)(n-2)}{n^5(p-1)^2}$ and the third moment in the partial ordering contains this term plus

$$\frac{4(n-1)\mu_3^2}{n^5(p-1)(p-2)(\sigma^2)^3} .$$

No computations were made to see how close the beta approximation would be to the distribution of W.

4. Estimation of the True Ranking of the Objects

Let R_j, $j = 1, 2, \ldots, p$ be the estimates of the true rankings of the populations. The same scores are used for the R_j's as each individual judge uses.

Consider first the average rank correlation, ρ_{ar}, between the estimated and the observed rankings. The correlation between the estimated rankings and the rankings given by the i^{th} judge is

$$\rho_i = \sum_{j=1}^{p} (R_j - \mu)(r_{ij} - \mu) \Bigg/ \sqrt{\sum_{j=1}^{p} (R_j - \mu)^2 \sum_{j=1}^{p} (r_{ij} - \mu)^2} \ .$$

The average of all n such correlations is

$$\rho_{ar} = \frac{1}{n} \sum_{i=1}^{n} \rho_i$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} (R_j - \mu)(r_{ij} - \mu) \Bigg/ \sqrt{\sum_{j=1}^{p} (R_j - \mu)^2 \sum_{j=1}^{p} (r_{ij} - \mu)^2}$$

$$= \text{constant} \cdot \left[\sum_{j=1}^{p} R_i r_j - n p \mu^2 \right] .$$

The average rank correlation, ρ_{ar}, is maximized if the R_j are chosen as follows. For the largest r_j, set $R_j = k$, for the next largest r_j, set $R_j = k - 1$, and continue in this fashion until the largest k of the r_j's have been considered. For the smallest $(p - k)$ of the r_j's, assign the value 0 to the corresponding R_j's.

As a second criteria, consider the sum of squares of differences between the actual column sums, r_j, and the sums, $n R_j$, which would occur if all the rankings were identical and equal to the estimated rankings. Call this sum of squares U. Then U is given by

$$U = \sum_{j=1}^{p} (r_j - n R_j)^2 = \text{constant} - 2n \sum_{j=1}^{p} r_j R_j \ .$$

U is minimized if $\sum_{j=1}^{p} r_j R_j$ is maximized. Therefore, U is minimized if the R_j are again chosen as suggested in the discussion concerning the average rank correlation.

In summary, this method of choosing the estimates, R_j, is "best" in the following sense. Subject to the constraint that the same scores be used that were used by the individual judges, the average rank correlation between the estimated and the observed rankings is maximized and the sum of squares of difference between the actual column sums and the sums, nR_j, which would occur if all rankings were identical and equal to the estimated rankings is minimized.

5. Asymptotic Relative Efficiency Considerations

In this section, an additional justification for the choice $c = 0$ will be given. It is shown in [10] that the A. R. E. of the test based on S_n, which is defined in Section 7 of Chapter II, with respect to the classical analysis of variance test is equal to

$$e_n(J) = \left\{ \bar{\sigma}_n^2 p^2 (p-1) \sum_{r=1}^{p} \lambda_{rn}^2 \right\} \cdot \left\{ \left(\sum_{r=1}^{p} \{J(r,p) - \bar{J}\} \lambda_{r,n} \right)^2 \Big/ \left(\sum_{r=1}^{p} \lambda_{r,n}^2 \right) \left(\sum_{r=1}^{p} \{J(r,p) - \bar{J}\}^2 \right) \right\}$$

where the second factor in the expression for $e_n(J)$ is uniformly bounded by 1 and the first factor is independent of $\bar{J}(r,p)$, $r = 1, \ldots, p$.

$$e_n^* = \bar{\sigma}_n^2 p^2 (p-1) \sum_{r=1}^{p} \lambda_{r,n}^2$$

is defined to be the A. R. E. of the ranking procedure since it is the maximum possible value of $e_n(J)$. A particular S_n test is said to be asymptotically efficient for $(F_1, \ldots, F_n)$ if $e_n(J) = e_n^*$ for that $(F_1, \ldots, F_n)$.

The main quantity of interest in the expression for $e_n(J)$ is the second factor which will be referred to as E, i.e.,

$$E = \left(\sum_{r=1}^{p} \{J(r,p) - \bar{J}\}\lambda_{r,n}\right)^2 \Big/ \left(\sum_{r=1}^{p} \lambda_{r,n}^2\right)\left(\sum_{r=1}^{p} \{J(r,p) - \bar{J}\}^2\right).$$

If $E = 1$ for a given set $\{J(r,p); r = 1, 2, \ldots, p\}$, and a given distribution $F(x)$, then that set of scores is optimum for the distribution. Taking the $J(r,p)$ values to be $1, 2, \ldots, k$ and c with a multiplicity of $(p-k)$, the value of c which maximizes E will be found, and the value of E at that c value, say E_c, will also be found. Then E will be evaluated with $c = 0$ and this E value will be called E_0. Then E_c and E_0 will be compared for various values of p and k.

It is noted that E is a rational function of c, in particular it is a ratio of two quadratic functions of c. The following theorem which is proved in the Appendix will be needed.

<u>Theorem 3.1</u>: If $f(x) = (a_1x^2 + a_2x + a_3)/(b_1x^2 + b_2x + b_3)$, $b_1x^2 + b_2x + b_3$ has no real zeros and $A_1x^2 + A_2x + A_3$ has two real zero's, where $A_1 > 0$ and

$$f'(x) = \frac{A_1x^2 + A_2x + A_3}{(b_1x^2 + b_2x + b_3)^2},$$

then $f(x)$ has an absolute maximum at

$$x = \frac{-A_2 - \sqrt{A_2^2 - 4A_1A_3}}{2A_1}$$

Consider the following three cases taken from [10].

Case 1: Uniform Distribution

$$\lambda_{r,n} = \begin{cases} 1, & r = 1 \\ 0, & (2 \leq r \leq p-1) \\ -1, & r = p \end{cases}$$

$E = \dfrac{c^2 - 2kc + k^2}{2p\sigma^2}$. If $k = 1$, $E = \dfrac{p}{4(p-1)}$, provided $c \neq 1$. The derivative of E, E' is

$$E' = \frac{1}{2}\,\frac{A_1c^2 + A_2c + A_3}{(p\sigma^2)^2}, \quad k \neq 1,$$

where

$$A_1 = (p-k)\,k(k-1)/p$$

$$A_2 = 2k(k+1)\,[(2k+1)/6 - k(k+1)/4p] - 2(p-k)k^3/p$$

$$A_3 = k^2(k+1)\,[(p-k)k/p - 2((2k+1)/6 - k(k+1)/4p)]$$

Case 2: Exponential Distribution

$$\lambda_{r,n} = \begin{cases} \dfrac{1}{p}, & r = 1 \\ -\dfrac{1}{p(p-1)}, & r > 1 \end{cases}$$

$$E = \frac{1}{(p-1)} \cdot \frac{\frac{(p-k)^2}{p^2} c^2 - 2\frac{(p-k)}{p}\left(k - \frac{k(k+1)}{2p}\right)c + \left(k - \frac{k(k+1)}{2p}\right)^2}{p\sigma^2} .$$

If $k = 1$, $E = 1$, provided $c \neq 1$. The derivative of E, E' is

$$E' = \frac{p}{(p-1)} \cdot \frac{A_1 c^2 + A_2 c + A_3}{(p\sigma^2)^2}$$

where

$$A_1 = \frac{(p-k)^2 k(k-1)}{p^2}$$

$$A_2 = 2\frac{(p-k)k}{p}\left\{\frac{(p-k)(k+1)}{p}\left(\frac{2k+1}{6} - \frac{k(k+1)}{4p}\right) - \left(k - \frac{k(k+1)}{2p}\right)^2\right\}$$

$$A_3 = \frac{(p-k)\,k(k+1)}{p}\left(k - \frac{k(k+1)}{2p}\right)\left(\left(k - \frac{k(k+1)}{2p}\right) - 2\left(\frac{2k+1}{6} - \frac{k(k+1)}{4p}\right)\right)$$

<u>Case 3</u>: Logistic Distribution

$$\lambda_{r,n} = \frac{2\{r - \frac{1}{2}(p+1)\}}{p(p^2-1)} , \qquad r = 1, 2, \ldots, p$$

$$E = \frac{12}{p(p+1)(p-1)}$$

$$\cdot \frac{\frac{k^2(p-k)^2}{4} c^2 + \frac{k^2(p-k)(k+1)(2k-3p+1)}{12} c + \frac{k^2(k+1)^2(2k-3p+1)^2}{144}}{p\sigma^2}$$

If $k = 1$, $E = 3/(p+1)$, provided $c \neq 1$. The derivative of E, E' is

$$E' = \frac{12}{p(p-1)(p+1)} \cdot \frac{A_1c^2 + A_2c + A_3}{(p\sigma^2)^2}$$

where

$$A_1 = \frac{(p-k)^2 k^3 (k+1)(k-1)}{12p}$$

$$A_2 = \frac{k^3(k+1)(p-k)}{2} \cdot \left\{ (p-k)\left(\frac{2k+1}{6} - \frac{k(k+1)}{4p} \right) - \frac{(k+1)}{36p}(2k-3p+1)^2 \right\}$$

$$A_3 = \frac{k^3(k+1)^2(p-k)(2k-3p+1)}{12} \left\{ \left(\frac{2k+1}{6} - \frac{k(k+1)}{4p} \right) + \frac{(k+1)(2k-3p+1)}{12p} \right\}$$

If $k = 1$, there is no optimum c value in the three cases and any c value except $c = 1$ is appropriate. The hypothesis of Theorem 3.1 are satisfied in all three cases for the following values of p and k: $3 \leq p \leq 10$, $2 \leq k \leq p-1$. The optimum c value, the value of E at that optimum c value, E_c, and the value of E at $c = 0$, E_0, are found in Table III for all three distributions. It is noted that the E_0 and E_c values are "close" and therefore the loss in A. R. E. is small in these three cases for these particular values of p and k when $c = 0$ is used instead of the optimum c value.

TABLE III

A. R. E. COMPARISONS

		Uniform			Exponential			Logistic		
p	k	c	E_c	E_0	c	E_c	E_0	c	E_c	E_0
3	2	0.00	1.00	1.00	1.00	1.00	0.75	0.00	1.00	1.00
4	2	0.50	0.75	0.73	1.00	1.00	0.76	-0.50	0.90	0.89
4	3	- 0.67	0.92	0.90	1.33	0.78	0.60	0.00	1.00	1.00
5	2	0.67	0.67	0.63	1.00	1.00	0.77	-1.00	0.80	0.78
5	3	0.33	0.67	0.66	1.33	0.79	0.60	-0.50	0.95	0.94
5	4	- 1.67	0.85	0.80	1.67	0.63	0.50	0.00	1.00	1.00
6	2	0.75	0.63	0.57	1.00	1.00	0.77	-1.50	0.71	0.69
6	3	0.67	0.58	0.56	1.33	0.80	0.60	-1.00	0.89	0.86
6	4	0.00	0.60	0.60	1.67	0.64	0.49	-0.50	0.97	0.96
6	5	- 3.00	0.80	0.71	2.00	0.52	0.43	0.00	1.00	1.00
7	2	0.80	0.60	0.54	1.00	1.00	0.78	-2.00	0.64	0.62
7	3	0.83	0.54	0.51	1.33	0.81	0.60	-1.50	0.82	0.79
7	4	0.56	0.52	0.51	1.67	0.65	0.49	-1.00	0.93	0.91
7	5	- 0.50	0.55	0.55	2.00	0.53	0.42	-0.50	0.98	0.98
7	6	- 4.67	0.76	0.64	2.33	0.44	0.38	0.00	1.00	1.00
8	2	0.83	0.58	0.52	1.00	1.00	0.78	-2.50	0.58	0.55
8	3	0.93	0.52	0.47	1.33	0.81	0.61	-2.00	0.76	0.72
8	4	0.83	0.48	0.46	1.67	0.66	0.49	-1.50	0.88	0.85
8	5	0.33	0.47	0.47	2.00	0.54	0.42	-1.00	0.95	0.94
8	6	- 1.17	0.51	0.50	2.33	0.46	0.36	-0.50	0.99	0.98
8	7	- 6.67	0.73	0.58	2.67	0.39	0.33	0.00	1.00	1.00
9	2	0.86	0.57	0.50	1.00	1.00	0.78	-3.00	0.53	0.50
9	3	1.00	0.50	0.45	1.33	0.81	0.61	-2.50	0.71	0.67
9	4	1.00	0.45	0.42	1.67	0.66	0.50	-2.00	0.83	0.79
9	5	0.75	0.43	0.42	2.00	0.55	0.42	-1.50	0.92	0.89
9	6	0.00	0.43	0.43	2.33	0.46	0.36	-1.00	0.97	0.95
9	7	- 2.00	0.48	0.46	2.67	0.40	0.32	-0.50	0.99	0.99
9	8	- 9.00	0.71	0.53	3.00	0.34	0.30	0.00	1.00	1.00
10	2	0.88	0.56	0.49	1.00	1.00	0.78	-3.50	0.49	0.46
10	3	1.05	0.49	0.43	1.33	0.81	0.62	-3.00	0.66	0.62
10	4	1.11	0.43	0.40	1.67	0.67	0.50	-2.50	0.79	0.74
10	5	1.00	0.40	0.38	2.00	0.56	0.42	-2.00	0.88	0.84
10	6	0.58	0.39	0.38	2.33	0.47	0.36	-1.50	0.94	0.91
10	7	- 0.44	0.40	0.40	2.67	0.40	0.32	-1.00	0.98	0.96
10	8	- 3.00	0.46	0.43	3.00	0.35	0.29	-0.50	0.99	0.99
10	9	-11.67	0.69	0.49	3.33	0.31	0.27	0.00	1.00	1.00

CHAPTER IV

AN EXTREME RANK SUM TEST OR A TEST FOR SLIPPAGE

In this chapter the null hypothesis of no difference in preference versus the alternative

$$H_A: \text{At least one population has slipped to the right,}$$

is considered. The test statistic, r_{max}, defined as

$$r_{max} = \max_{1 \le i \le p} \{r_i\}$$

or the maximum of the column sums, will be used to test H_0 against H_A. The basic test technique will be established in Theorem 4.2. Theorem 4.1 will contain a result needed in Theorem 4.2, and another result needed in the application of Theorem 4.1 will be given in Theorem 4.3. Finally, a table will be presented which is instrumental to the application of the results.

Before proving Theorem 4.1, consider the following lemmas.

Lemma 4.1: Suppose $a_1 \le a_2 \le \dots \le a_n$ and $b_1 \le b_2 \le \dots \le b_n$, then $\sum_{i=1}^{n} (a_i - \bar{a})(b_i - \bar{b}) \ge 0$ or equivalently,

$$n \sum_{j=1}^{n} a_j b_j \ge \left(\sum_{j=1}^{n} a_j \right) \left(\sum_{j=1}^{n} b_j \right) ,$$

where $\bar{a} = \sum_{j=1}^{n} \frac{a_j}{n}$ and $\bar{b} = \sum_{j=1}^{n} \frac{b_j}{n}$.

Proof: For $1 \leq k \leq n-1$, $0 \leq \sum_{j=k+1}^{n} (a_j - a_{j-k})(b_j - b_{j-k})$. Thus,

$$0 \leq \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} (a_j - a_{j-k})(b_j - b_{j-k})$$

or

$$0 \leq \left(\sum_{k=1}^{n-1} \sum_{j=k+1}^{n} a_j b_j + \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} a_{j-k} b_{j-k} \right) - \left(\sum_{k=1}^{n-1} \sum_{j=k+1}^{n} a_j b_{j-k} + \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} a_{j-k} b_j \right) .$$

Now, each $a_r b_r$ occurs $(n-1)$ times in the first grouped term and for each $s \neq t$, $a_s b_t$ occurs exactly once in the second grouped term. This means that the right hand side of the last inequality may be replaced by $n \sum_{j=1}^{n} a_j b_j - \sum_{k=1}^{n} \sum_{j=1}^{n} a_j b_k$ or

$$0 \leq n \sum_{j=1}^{n} a_j b_j - \sum_{k=1}^{n} \sum_{j=1}^{n} a_j b_k$$

$$= n \sum_{k=1}^{n} a_j b_j - \left(\sum_{j=1}^{n} a_j \right) \left(\sum_{j=1}^{n} b_j \right) . \qquad \triangle$$

Lemma 4.2: Let $p_i = P(x_i \leq g_i)$, $p_{ij} = P(x_i \leq g_i$ and $x_j \leq g_j)$, $(i \neq j)$, $q_i = P(x_i > g_i)$ and $q_{ij} = P(x_i > g_i$ and $x_j > g_j)$, $(i \neq j)$. Then $p_{ij} \leq p_i p_j$ is equivalent to $q_{ij} \leq q_i q_j$.

Proof: $p_i = 1 - q_i$, $p_j = 1 - q_j$ imply that $p_i(1 - p_j) = q_j(1 - q_i)$. Also, $p_i - p_{ij} = q_j - q_{ij}$, since both equal to $P(x_i \leq g_j$ and $x_i > g_j)$.

It follows that $(p_i - p_i p_j) - (p_i - p_{ij}) = (q_j - q_i q_j) - (q_j - q_{ij})$ or $p_{ij} - p_i p_j = q_{ij} - q_i q_j$, which implies the result. △

This lemma is referred to on page 40 of [2].

Theorem 4.1: Under H_0, $P(r_i > a$ and $r_j > b) \leq P(r_i > a)\,P(r_j > b)$ where r_i and r_j are two column totals in the partial ordering scheme and a and b are integers between 0 and kn.

Proof: Using Lemma 4.2, the theorem will be proved by showing that $P(r_i \leq a$ and $r_j \leq b) \leq P(r_i \leq a)\,P(r_j \leq b)$. The proof will be by mathematical induction. The following fact will also be used when needed: If x and y are two integers and $0 < x < y$, then $\frac{(x-1)}{(y-1)} < \frac{x}{y}$.

Suppose $n = 1$ and $a = b = 0$, then

$$P(r_i \leq 0 \text{ and } r_j \leq 0) = \frac{(p-k)(p-k-1)}{p(p-1)}$$

$$P(r_i \leq 0) = P(r_j \leq 0) = \frac{p-k}{p}$$

and

$$P(r_i \leq 0 \text{ and } r_j \leq 0) < P(r_i \leq 0)\,P(r_j \leq 0)$$

since

$$\frac{(p-k-1)}{(p-1)} < \frac{(p-k)}{p} .$$

Suppose $a = b = \ell \neq 0$

$$P(r_i \leq \ell \text{ and } r_j \leq \ell) = \frac{[(p-k)+\ell)][(p-k)+\ell-1]}{p(p-1)}$$

$$P(r_i \leq \ell) = P(r_j \leq \ell) = \frac{(p-k)+\ell}{p}$$

Again, $P(r_i \leq \ell \text{ and } r_j \leq \ell) \leq P(r_i \leq \ell)\,P(r_j \leq \ell)$.

Suppose $b < a$, then

$$P(r_i \leq a \text{ and } r_j \leq b) = P(r_i \leq a \mid r_j \leq b) \cdot P(r_j \leq b)$$

$$= \frac{[a+(p-k)]-1}{p-1} \cdot \frac{b+(p-k)}{p}$$

$$P(r_i \leq a)\,P(r_j \leq b) = \frac{a+(p-k)}{p} \cdot \frac{b+(p-k)}{p}$$

Again $\frac{[a+(p-k)]-1}{p-1} < \frac{a+(p-k)}{p}$ implies the desired result. A similar argument holds for $a < b$.

Now suppose the desired result is true for $n = m$ judges, i.e., assume

$$P(r_i \leq a \text{ and } r_j \leq b \mid m) \leq P(r_i \leq a \mid m)\,P(r_j \leq b \mid m)$$

and show it follows for $n = m+1$ judges. Now,

$P(r_i \leq a \text{ and } r_j \leq b \mid m+1)$

$$= \sum_{e=0}^{k} \sum_{f=0}^{k} \cdot P(r_i \leq a \text{ and } r_j \leq b \mid r_{m+1,i} = e \text{ and } r_{m+1,j} = f)$$

$$\cdot P(r_{m+1,i} = e \text{ and } r_{m+1,j} = f)$$

$$= \sum_{e=0}^{k} \sum_{f=0}^{k} P(r_i \leq a - e \text{ and } r_j \leq b - f \mid m) \cdot P(r_{m+1,i} = e \text{ and } r_{m+1,j} = f)$$

$$= P(r_i \leq a \text{ and } r_j \leq b \mid m) \cdot \frac{(p-k)(p-k-1)}{p(p-1)}$$

$$+ \frac{p-k}{p(p-1)} \sum_{f=1}^{k} P(r_i \leq a \text{ and } r_j \leq b - f \mid m)$$

$$+ \frac{p-k}{p(p-1)} \sum_{e=1}^{k} P(r_i \leq a - e \text{ and } r_j \leq b \mid m)$$

$$+ \sum_{f=1}^{k} \sum_{\substack{e=1 \\ e \neq f}}^{k} P(r_i \leq a - e \text{ and } r_j \leq b - f \mid m) \cdot \frac{1}{p(p-1)}$$

$$\leq P(r_i \leq a \mid m) P(r_j \leq b \mid m) \cdot \frac{(p-k)(p-k-1)}{p(p-1)}$$

$$+ \frac{(p-k)}{p(p-1)} P(r_i \leq a \mid m) \sum_{f=1}^{k} P(r_j \leq b - f \mid m)$$

$$+ \frac{(p-k)}{p(p-1)} P(r_j \leq b \mid m) \sum_{e=1}^{k} P(r_i \leq a - e \mid m)$$

$$+ \sum_{e=1}^{k} \sum_{\substack{f=1 \\ e \neq f}}^{k} P(r_i \leq a - e \mid m) P(r_j \leq b - f \mid m) \cdot \frac{1}{p(p-1)} \tag{4.1}$$

by the induction assumption for $n = m$ judges. Now, $P(r_i \leq a \mid m+1) P(r_j \leq b \mid m+1)$ may be written as

$$P(r_i \leq a \mid m+1) P(r_j \leq b \mid m+1)$$

$$= \left[P(r_i \leq a \mid m) \frac{p-k}{p} + \sum_{e=1}^{k} P(r_i \leq a - e \mid m) \cdot \frac{1}{p} \right]$$

$$\cdot \left[P(r_j \leq b \mid m) \cdot \frac{p-k}{p} + \sum_{f=1}^{k} P(r_j \leq b - f \mid m) \cdot \frac{1}{p} \right]$$

$$= P(r_i \leq a|m) P(r_j \leq b|m) \cdot \frac{(p-k)^2}{p^2}$$

$$+ P(r_i \leq a|m) \sum_{f=1}^{k} P(r_j \leq b-f|m) \frac{(p-k)}{p^2}$$

$$+ P(r_j \leq b|m) \sum_{e=1}^{k} P(r_i \leq a-e|m) \cdot \frac{(p-k)}{p^2}$$

$$+ \sum_{e=1}^{k} \sum_{f=1}^{k} P(r_i \leq a-e|m) P(r_j \leq b-f|m) \cdot \frac{1}{p^2} .$$

If

$$P(r_i \leq a|m+1) P(r_j \leq b|m+1) - \left[P(r_i \leq a|m) P(r_j \leq b|m) \cdot \frac{(p-k)^2}{p^2} \right.$$

$$+ P(r_i \leq a|m) \sum_{f=1}^{k} P(r_j \leq b-f|m) \frac{(p-k)}{p^2}$$

$$+ P(r_j \leq b|m) \sum_{e=1}^{k} P(r_i \leq a-e|m) \cdot \frac{(p-k)}{p^2}$$

$$\left. + \sum_{e=1}^{k} \sum_{f=1}^{k} P(r_i \leq a-e|m) P(r_j \leq b-f|m) \cdot \frac{1}{p^2} \right] ,$$

which equals zero, is added to the right hand side of inequality (4. 1), then the inequality may be rewritten as

$$P(r_i \leq a \text{ and } r_j \leq b|m+1) \leq P(r_i \leq a|m+1) P(r_j \leq b|m+1)$$

$$- \frac{(p-k)k}{p^2(p-1)} P(r_i \leq a|m) P(r_j \leq b|m) + \frac{(p-k)}{p^2(p-1)} P(r_i \leq a|m)$$

$$\sum_{f=1}^{k} \cdot P(r_j \leq b-f|m) + \frac{(p-k)}{p^2(p-1)} P(r_j \leq b|m) \sum_{e=1}^{k} P(r_i \leq a-e|m)$$

$$+ \frac{1}{p^2(p-1)} \sum_{\substack{e=1 \\ e \neq f}}^{k} \sum_{f=1}^{k} P(r_i \le a - e | m) P(r_j \le b - f | m)$$

$$- \frac{1}{p^2} \sum_{e=1}^{k} P(r_i \le a - e | m) P(r_j \le b - e | m) . \tag{4.2}$$

Now, define A and B as follows,

$$A = \frac{\sum_{f=1}^{k} P(r_i \le a - f | m) + (p-k) P(r_i \le a | m)}{p}$$

$$B = \frac{\sum_{f=1}^{k} P(r_j \le b - f | m) + (p-k) P(r_j \le b | m)}{p}$$

Now, the inequality (4.2) may be written as

$$P(r_i \le a \text{ and } r_j \le b | m+1) \le P(r_i \le a | m+1) P(r_j \le b | m+1) - \frac{1}{p(p-1)}$$

$$\cdot \left\{ \sum_{e=1}^{k} [P(r_i \le a - e | m) - A][P(r_j \le b - e | m) - B] \right.$$

$$\left. + (p-k)[P(r_i \le a | m) - A][P(r_j \le b | m) - B] \right\} .$$

Lemma 4.1 implies

$$P(r_i \le a \text{ and } r_j \le b | m+1) \le P(r_i \le a | m+1) P(r_j \le b | m+1) .$$

Therefore, by induction

$$P(r_i > a \text{ and } r_j > b) \le P(r_i > a) P(r_j > b)$$

for any number of judges n. △

Theorem 4.2 will now be proved

Theorem 4.2: Let $S = \{\text{all possible permutations of column totals}\}$ and $\vec{r} = (r_1, \ldots, r_p)$ and $\alpha = p$. $P\{\vec{r} : \vec{r} \in S \text{ and } r_j > R\}$ for an arbitrary $j = 1, 2, \ldots, p$. Then under H_0,
$\alpha - \frac{1}{2}\alpha^2 \leq P\{\vec{r} : \vec{r} \in S \text{ and } r_{max} > R\} \leq \alpha$.

Proof: Let $A = \{\vec{r} : r_j \leq R \text{ for } j = 1, 2, \ldots, p \text{ and } \vec{r} \in S\}$. Let $A_j = \{\vec{r} : r_j \leq r \text{ and } \vec{r} \in S\}$. Now,

$$A = \bigcap_{j=1}^{p} A_j$$

$$\overline{A} = \overline{\bigcap_{j=1}^{p} A_j} = \bigcup_{j=1}^{p} \overline{A}_j = \bigcup_{j=1}^{p} \{\vec{r} : r_j > R ; \vec{r} \in S\}$$

$$= \{\vec{r} : r_{max} > R ; \vec{r} \in S\}$$

Now, Bonferroni's inequalities imply

$$p \cdot P(\overline{A}_j) - \binom{p}{2} \cdot P(\overline{A}_i \text{ and } \overline{A}_j) \leq P(\overline{A}) \leq p \cdot P(\overline{A}_j) .$$

Now Theorem 4.1 implies

$$P(\overline{A}_i \text{ and } \overline{A}_j) \leq P(\overline{A}_i) P(\overline{A}_j) = [P(\overline{A}_j)]^2 .$$

Then, one may write

$$p \cdot P(\overline{A}_j) - \frac{p(p-1)}{2} P(\overline{A}_j)^2 \leq P(\overline{A}) \leq p \cdot P(\overline{A}_j) .$$

It follows that

$$p \cdot P(\overline{A}_j) - \frac{p^2}{2} P(\overline{A}_j)^2 \leq P(\overline{A}) \leq p \cdot P(\overline{A}_j)$$

or

$$\alpha - \frac{1}{2}\alpha^2 \leq P\{\vec{r} : \vec{r} \in S \text{ and } r_{max} > R\} \leq \alpha$$

where

$$\alpha = p \cdot P\{\vec{r} : \vec{r} \in S \text{ and } r_j > R\} . \qquad \Delta$$

In order that Theorem 4.2 be of practical use, the marginal distribution of r_j under H_0, must be known. This marginal distribution will be given in Theorem 4.3.

<u>Theorem 4.3</u>: If r_j is the j^{th} column sum in the partial ordering scheme, then the marginal distribution of r_j under H_0 is given by

$$P(r_j = M) = \sum_{i=0}^{n-1} \binom{n}{i} \left(\frac{p-k}{p}\right)^i \left(\frac{k}{p}\right)^{n-1} \cdot \sum_{x=0}^{\infty} \frac{I_{(M-kx-n+i)} \binom{n-i}{x} \binom{M-kx-1}{n-i-1} (-1)^x}{k^{n-i}}$$

$$+ \left(\frac{p-k}{p}\right)^n J_M, \quad \text{where} \quad 0 \leq M \leq kn,$$

$$I_y = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases} \quad \text{and} \quad J_M = \begin{cases} 1, & M = 0 \\ 0, & M \neq 0 \end{cases}$$

<u>Proof</u>:

$$P(r_j = M) = \sum_{i=0}^{n} P(r_j = M \mid j^{th} \text{ object receives } i \text{ scores of } 0)$$

$$\cdot P(j^{th} \text{ object receives } i \text{ scores of } 0) .$$

Let E_i be the event

$$E_i = \{r_j = M | j^{th} \text{ object receives } i \text{ scores of } 0\} .$$

For $0 \leq i \leq n-1$, E_i corresponds to having $(n-i)$ cells, each of which could be filled with the integers 1 through k and having an integral total of M.

It is shown by Doornbos and Prins [2] that

$$P(E_i) = \sum_{x=0}^{\infty} \frac{I_{(M-kx-n+i)} \binom{n-i}{x} \binom{M-kx-1}{n-i-1} (-1)^x}{k^{n-i}}$$

for $i = 0, 1, \ldots, (n-1)$.

E_n is the event that $r_j = M$ given the j^{th} object received all 0's.

$$P(E_n) = 1 \text{ if } M = 0$$
$$= 0 \text{ if } M \neq 0 .$$

Let F_i be the event

$$F_i = \{j^{th} \text{ object receives } i \text{ scores of } 0\}$$

then

$$P(F_i) = \binom{n}{i} \left(\frac{p-k}{p}\right)^i \left(\frac{k}{p}\right)^{n-i} , \quad i = 0, 1, \ldots, n .$$

Then substituting for $P(E_i)$ and $P(F_i)$, the expression for $P(r_j = M)$ becomes

$$P(r_j = M) = \sum_{i=0}^{n-1} P(F_i)\,P(E_i) + P(E_n)\,P(F_n)$$

$$= \sum_{i=0}^{n-1} \binom{n}{i}\left(\frac{p-k}{p}\right)^i \left(\frac{k}{p}\right)^{n-i} \sum_{x=0}^{\infty} \frac{I_{(M-kx-n+i)}\binom{n-i}{x}\binom{M-kx-1}{n-i-1}(-1)^x}{k^{n-i}}$$

$$+ \left(\frac{p-k}{p}\right)^n \cdot J_M \, . \qquad \triangle$$

Using the results of Theorems 4.2 and 4.3, the critical values R may be found for a pre-set α level test. Table IV contains critical values of R for α levels of .01, .025 and .05 for $3 \leq p \leq 10$, $3 \leq n \leq 10$, $1 \leq k \leq p-1$. Table IV also contains a lower limit α_L and an upper limit α_U for α since in a discrete distribution, α cannot be obtained exactly. In the table α_L and α_U are defined as follows,

$$\alpha_L = \alpha - \frac{1}{2}\alpha^2$$

$$\alpha_U = \alpha$$

where α is defined in Theorem 4.2. If an r_{max} as large as R or larger is obtained H_0 is rejected and H_A is accepted.

The results given in Table IV were obtained by using a computer program and the IBM-360, Mod 65 at Oklahoma State University.

Tables for the case of the full ranking, i.e., k = p, and $3 \leq n \leq 10$ may be found in [2] and [11].

Consider now a comparison of the two basic test statistics developed thus far, i.e., S and r_{max}. By comparing Table II and IV it appears that the test based on r_{max} dominates the test based on S, i.e., any time one would reject H_0 based on r_{max}, then H_0

would also be rejected by using S. However, this is not always the case as the following example illustrates.

Example 4.1: Suppose $p = 4$, $k = 2$ and $n = 8$ and the following scores are assigned.

		Object			
		1	2	3	4
Judge	1	0	1	0	2
	2	0	1	0	2
	3	0	1	0	2
	4	0	1	0	2
	5	1	0	0	2
	6	1	0	0	2
	7	1	0	2	0
	8	1	0	2	0
	r_i	4	4	4	12

Now suppose H_0 is tested at $\alpha = .05$. $r_{max} = 12$, $S = (4-6)^2 + (4-6)^2 + (4-6)^2 + (12-6)^2 = 48$. $P(S \geq 48) = 0.087$ which means H_0 would not be rejected. The critical value for r_{max} is 12 and therefore H_0 would be rejected using r_{max}.

TABLE IV

EXTREME RANK SUM TEST

p	k	n	α = .01 R	α = .01 α_L	α = .01 α_U	α = .025 R	α = .025 α_L	α = .025 α_U	α = .05 R	α = .05 α_L	α = .05 α_U
3	1	3									
		4							4	.0363	.0371
		5				5	.0122	.0124	5	.0122	.0124
		6	6	.0041	.0042	6	.0041	.0042	6	.0041	.0042
		7	7	.0013	.0014	6	.0203	.0206	6	.0203	.0206
		8	7	.0077	.0078	7	.0077	.0078	7	.0077	.0078
		9	8	.0028	.0030	7	.0245	.0249	7	.0245	.0249
		10	9	.0010	.0011	8	.0101	.0103	8	.0101	.0103
3	2	3									
		4							8	.0363	.0371
		5				10	.0122	.0124	10	.0122	.0124
		6	12	.0041	.0042	12	.0041	.0042	11	.0283	.0289
		7	14	.0013	.0014	13	.0109	.0110	12	.0481	.0494
		8	15	.0041	.0042	14	.0203	.0206	14	.0203	.0206
		9	16	.0083	.0084	16	.0083	.0084	15	.0316	.0322
		10	18	.0033	.0034	17	.0139	.0141	16	.0442	.0453
4	1	3									
		4				4	.0155	.0157	4	.0155	.0157
		5	5	.0039	.0040	5	.0039	.0040	5	.0039	.0040
		6	6	.0009	.0010	5	.0183	.0186	5	.0183	.0186
		7	6	.0053	.0054	6	.0053	.0054	6	.0053	.0054
		8	7	.0015	.0016	6	.0167	.0170	6	.0167	.0170
		9	7	.0053	.0054	7	.0053	.0054	6	.0391	.0400
		10	8	.0016	.0017	7	.0139	.0141	7	.0139	.0141
4	2	3									
		4				8	.0155	.0157	8	.0155	.0157
		5	10	.0039	.0040	9	.0231	.0235	9	.0231	.0235
		6	11	.0068	.0069	11	.0068	.0069	10	.0326	.0333
		7	13	.0010	.0020	12	.0104	.0106	11	.0387	.0396
		8	14	.0032	.0033	13	.0134	.0135	12	.0440	.0452
		9	15	.0044	.0045	14	.0161	.0163	13	.0477	.0490
		10	16	.0056	.0057	15	.0183	.0186	15	.0183	.0186

TABLE IV (Continued)

p	k	n	α = .01 R	α = .01 α_L	α = .01 α_U	α = .025 R	α = .025 α_L	α = .025 α_U	α = .05 R	α = .05 α_L	α = .05 α_U
4	3	3									
		4				12	.0155	.0157	12	.0155	.0157
		5	15	.0039	.0040	14	.0231	.0235	14	.0231	.0235
		6	17	.0068	.0069	17	.0068	.0069	16	.0269	.0274
		7	19	.0087	.0088	19	.0087	.0088	18	.0288	.0294
		8	22	.0027	.0028	21	.0100	.0101	20	.0292	.0298
		9	24	.0033	.0034	23	.0107	.0108	22	.0287	.0292
		10	26	.0037	.0038	25	.0109	.0111	24	.0276	.0281
5	1	3							3	.0392	.0401
		4	4	.0079	.0081	4	.0079	.0081	4	.0079	.0081
		5	5	.0016	.0017	5	.0016	.0017	4	.0330	.0337
		6	5	.0079	.0081	5	.0079	.0081	5	.0079	.0081
		7	6	.0018	.0019	5	.0230	.0234	5	.0230	.0234
		8	6	.0061	.0062	6	.0061	.0062	6	.0061	.0062
		9	7	.0015	.0016	6	.0152	.0154	6	.0152	.0154
		10	7	.0043	.0044	7	.0043	.0044	6	.0313	.0319
5	2	3							6	.0392	.0401
		4	8	.0079	.0081	8	.0079	.0081	7	.0392	.0401
		5	9	.0095	.0097	9	.0095	.0097	8	.0483	.0497
		6	11	.0022	.0023	10	.0127	.0129	9	.0468	.0481
		7	12	.0031	.0033	11	.0134	.0136	10	.0468	.0481
		8	13	.0036	.0037	12	.0141	.0143	11	.0440	.0451
		9	14	.0040	.0041	13	.0139	.0141	12	.0414	.0424
		10	15	.0041	.0042	14	.0136	.0138	13	.0381	.0390
5	3	3							9	.0392	.0491
		4	12	.0079	.0081	12	.0079	.0081	11	.0392	.0401
		5	14	.0095	.0097	14	.0095	.0097	13	.0330	.0337
		6	16	.0089	.0090	16	.0089	.0090	15	.0283	.0289
		7	18	.0080	.0082	17	.0235	.0239	17	.0235	.0239
		8	20	.0070	.0071	19	.0190	.0193	18	.0454	.0465
		9	22	.0059	.0060	21	.0153	.0155	20	.0355	.0362
		10	24	.0049	.0050	23	.0122	.0124	22	.0277	.0282
5	4	3							12	.0392	.0401
		4	16	.0079	.0081	16	.0079	.0081	15	.0392	.0401
		5	19	.0095	.0097	19	.0095	.0097	18	.0330	.0337
		6	22	.0089	.0090	22	.0089	.0090	21	.0265	.0269
		7	25	.0076	.0077	24	.0209	.0212	24	.0209	.0212
		8	28	.0063	.0064	27	.0162	.0164	26	.0368	.0376
		9	31	.0050	.0052	30	.0125	.0126	29	.0276	.0281
		10	33	.0095	.0097	32	.0207	.0210	31	.0412	.0422

TABLE IV (Continued)

p	k	n	α = .01 R	α_L	α_U	α = .025 R	α_L	α_U	α = .05 R	α_L	α_U
6	1	3							3	.0273	.0278
		4	4	.0046	.0047	4	.0046	.0046	4	.0046	.0047
		5	5	.0007	.0008	4	.0198	.0201	4	.0198	.0201
		6	5	.0039	.0040	5	.0039	.0040	5	.0030	.0040
		7	6	.0007	.0008	5	.0119	.0121	5	.0119	.0121
		8	6	.0026	.0027	6	.0026	.0027	5	.0272	.0277
		9	6	.0067	.0069	6	.0067	.0069	6	.0067	.0069
		10	7	.0016	.0017	6	.0145	.0147	6	.0145	.0147
6	2	3							6	.0273	.0278
		4	8	.0046	.0047	7	.0228	.0232	7	.0228	.0232
		5	9	.0046	.0047	9	.0046	.0047	8	.0273	.0278
		6	10	.0059	.0060	9	.0236	.0240	9	.0236	.0240
		7	11	.0055	.0056	10	.0222	.0226	10	.0222	.0226
		8	12	.0054	.0056	11	.0191	.0194	11	.0191	.0194
		9	13	.0049	.0050	12	.0167	.0170	11	.0481	.0494
		10	14	.0045	.0046	13	.0142	.0144	12	.0401	.0411
6	3	3							9	.0273	.0278
		4	12	.0046	.0047	11	.0228	.0232	11	.0228	.0232
		5	14	.0046	.0047	13	.0160	.0163	13	.0160	.0163
		6	16	.0035	.0037	15	.0122	.0124	14	.0348	.0355
		7	17	.0089	.0091	16	.0238	.0242	16	.0238	.0242
		8	19	.0063	.0065	18	.0165	.0168	17	.0387	.0396
		9	21	.0045	.0046	20	.0114	.0116	19	.0264	.0268
		10	22	.0079	.0080	21	.0180	.0180	20	.0381	.0390
6	4	3							12	.0273	.0278
		4	16	.0046	.0047	15	.0228	.0232	15	.0228	.0232
		5	19	.0046	.0047	18	.0160	.0163	17	.0422	.0433
		6	22	.0035	.0037	21	.0107	.0109	20	.0273	.0278
		7	24	.0072	.0073	23	.0177	.0179	22	.0389	.0398
		8	27	.0047	.0049	26	.0117	.0117	25	.0248	.0252
		9	29	.0073	.0075	28	.0158	.0160	27	.0316	.0322
		10	32	.0047	.0048	30	.0202	.0205	29	.0379	.0388
6	5	3							15	.0273	.0278
		4	20	.0046	.0047	19	.0228	.0232	19	.0228	.0232
		5	24	.0046	.0047	23	.0160	.0163	22	.0422	.0433
		6	28	.0035	.0037	27	.0107	.0109	26	.0266	.0271
		7	31	.0070	.0071	30	.0168	.0170	29	.0359	.0367
		8	35	.0045	.0047	33	.0224	.0228	32	.0436	.0447
		9	38	.0067	.0068	37	.0140	.0142	36	.0273	.0278
		10	41	.0088	.0089	40	.0171	.0174	39	.0315	.0321

TABLE IV (Continued)

p	k	n	α = .01 R	α_L	α_U	α = .025 R	α_L	α_U	α = .05 R	α_L	α_U
7	1	3				3	.0202	.0205	3	.0202	.0205
		4	4	.0029	.0030	4	.0029	.0030	4	.0029	.0030
		5	5	.0004	.0005	4	.0128	.0130	4	.0128	.0130
		6	5	.0022	.0023	5	.0022	.0023	4	.0337	.0344
		7	5	.0067	.0068	5	.0067	.0068	5	.0067	.0068
		8	6	.0012	.0013	5	.0158	.0160	5	.0158	.0160
		9	6	.0033	.0034	6	.0033	.0034	5	.0312	.0318
		10	6	.0074	.0075	6	.0074	.0075	6	.0074	.0075
7	2	3				6	.0202	.0205	6	.0202	.0205
		4	8	.0029	.0030	7	.0144	.0146	7	.0144	.0146
		5	9	.0025	.0026	8	.0169	.0171	8	.0169	.0171
		6	10	.0030	.0031	9	.0131	.0133	9	.0131	.0133
		7	11	.0026	.0027	10	.0117	.0119	9	.0394	.0403
		8	11	.0092	.0094	11	.0092	.0094	10	.0317	.0324
		9	12	.0076	.0078	11	.0242	.0246	11	.0242	.0246
		10	13	.0060	.0061	12	.0189	.0192	12	.0189	.0192
7	3	3				9	.0202	.0205	9	.0202	.0205
		4	12	.0029	.0030	11	.0144	.0146	10	.0427	.0438
		5	13	.0087	.0088	13	.0087	.0088	12	.0291	.0296
		6	15	.0060	.0061	15	.0183	.0186	13	.0463	.0475
		7	17	.0039	.0040	16	.0111	.0113	15	.0286	.0291
		8	18	.0069	.0071	17	.0176	.0178	16	.0400	.0409
		9	20	.0043	.0044	18	.0245	.0249	18	.0245	.0249
		10	21	.0066	.0067	20	.0150	.0153	19	.0321	.0327
7	4	3				12	.0202	.0205	12	.0202	.0205
		4	16	.0029	.0030	15	.0144	.0146	14	.0427	.0438
		5	18	.0087	.0088	17	.0230	.0234	17	.0230	.0234
		6	21	.0049	.0051	20	.0131	.0133	19	.0309	.0315
		7	23	.0074	.0076	22	.0172	.0175	21	.0363	.0371
		8	25	.0096	.0098	24	.0203	.0206	23	.0400	.0410
		9	28	.0054	.0055	26	.0226	.0230	25	.0424	.0434
		10	30	.0064	.0065	28	.0242	.0246	27	.0436	.0447
7	5	3				15	.0202	.0205	15	.0202	.0205
		4	20	.0029	.0030	19	.0144	.0146	18	.0427	.0438
		5	23	.0087	.0088	22	.0230	.0234	22	.0230	.0234
		6	27	.0049	.0051	26	.0124	.0125	25	.0274	.0279
		7	30	.0067	.0068	29	.0148	.0150	28	.0299	.0304
		8	33	.0080	.0081	32	.0162	.0164	31	.0306	.0312
		9	36	.0088	.0089	35	.0167	.0170	34	.0302	.0308
		10	39	.0092	.0093	38	.0167	.0170	36	.0486	.0500

TABLE IV (Continued)

p	k	n	α = .01 R	α_L	α_U	α = .025 R	α_L	α_U	α = .05 R	α_L	α_U
7	6	3				18	.0202	.0205	18	.0202	.0205
		4	24	.0029	.0030	23	.0144	.0146	22	.0427	.0438
		5	28	.0087	.0088	27	.0230	.0234	27	.0230	.0234
		6	33	.0049	.0051	32	.0124	.0125	31	.0271	.0275
		7	37	.0067	.0068	36	.0144	.0146	35	.0286	.0292
		8	41	.0077	.0079	40	.0154	.0156	39	.0286	.0291
		9	45	.0083	.0084	44	.0155	.0157	42	.0467	.0479
		10	49	.0085	.0085	48	.0151	.0153	46	.0428	.0438
8	1	3				3	.0155	.0157	3	.0155	.0157
		4	4	.0019	.0020	4	.0019	.0020	4	.0019	.0020
		5	4	.0087	.0088	4	.0087	.0088	4	.0087	.0088
		6	5	.0013	.0014	4	.0234	.0238	4	.0234	.0238
		7	5	.0041	.0042	5	.0041	.0042	4	.0486	.0500
		8	5	.0097	.0099	5	.0097	.0099	5	.0097	.0099
		9	6	.0018	.0019	5	.0196	.0199	5	.0196	.0199
		10	6	.0040	.0041	6	.0040	.0041	5	.0350	.0357
8	2	3				6	.0155	.0157	6	.0155	.0157
		4	7	.0097	.0098	7	.0097	.0098	7	.0097	.0098
		5	9	.0014	.0015	8	.0111	.0113	7	.0420	.0430
		6	9	.0078	.0079	9	.0078	.0079	8	.0351	.0359
		7	10	.0067	.0068	9	.0241	.0245	9	.0241	.0245
		8	11	.0049	.0050	10	.0185	.0188	10	.0185	.0188
		9	12	.0038	.0039	11	.0131	.0133	10	.0411	.0421
		10	12	.0097	.0099	12	.0097	.0099	11	.0290	.0295
8	3	3				9	.0155	.0157	9	.0155	.0157
		4	11	.0097	.0098	11	.0097	.0098	11	.0097	.0098
		5	13	.0051	.0052	12	.0183	.0186	12	.0183	.0186
		6	15	.0032	.0034	14	.0105	.0107	13	.0278	.0283
		7	16	.0057	.0059	15	.0157	.0159	14	.0385	.0394
		8	17	.0088	.0089	16	.0212	.0215	15	.0476	.0488
		9	19	.0049	.0050	18	.0118	.0120	17	.0268	.0273
		10	20	.0066	.0067	19	.0150	.0152	18	.0320	.0326
8	4	3				12	.0155	.0157	12	.0155	.0157
		4	15	.0097	.0098	15	.0097	.0098	14	.0288	.0294
		5	18	.0051	.0052	17	.0135	.0137	16	.0338	.0345
		6	20	.0069	.0070	19	.0170	.0173	18	.0377	.0385
		7	22	.0085	.0086	21	.0186	.0189	20	.0381	.0389
		8	24	.0093	.0095	23	.0193	.0196	22	.0376	.0384
		9	26	.0098	.0099	25	.0192	.0195	24	.0359	.0367
		10	28	.0098	.0100	27	.0186	.0189	26	.0338	.0345

TABLE IV (Continued)

p	k	n	α = .01 R	α_L	α_U	α = .025 R	α_L	α_U	α = .05 R	α_L	α_U
8	5	3				15	.0155	.0157	15	.0155	.0157
		4	19	.0097	.0098	19	.0097	.0098	18	.0288	.0294
		5	23	.0051	.0052	22	.0135	.0137	21	.0302	.0308
		6	26	.0063	.0065	25	.0143	.0145	24	.0297	.0303
		7	29	.0068	.0069	28	.0142	.0144	27	.0276	.0281
		8	32	.0068	.0069	31	.0133	.0135	29	.0435	.0446
		9	35	.0064	.0066	33	.0217	.0220	32	.0372	.0380
		10	38	.0059	.0060	36	.0188	.0190	35	.0315	.0322
8	6	3				18	.0155	.0157	18	.0155	.0157
		4	23	.0097	.0098	23	.0097	.0098	22	.0288	.0294
		5	28	.0051	.0052	27	.0135	.0137	26	.0302	.0308
		6	32	.0063	.0065	31	.0140	.0142	30	.0279	.0284
		7	36	.0065	.0066	35	.0131	.0133	33	.0439	.0450
		8	40	.0062	.0063	38	.0212	.0215	37	.0364	.0372
		9	44	.0056	.0057	42	.0179	.0181	40	.0482	.0496
		10	47	.0087	.0088	45	.0244	.0248	44	.0388	.0397
8	7	3				21	.0155	.0157	21	.0155	.0157
		4	27	.0097	.0098	27	.0097	.0098	26	.0288	.0294
		5	33	.0051	.0052	32	.0135	.0137	31	.0302	.0308
		6	38	.0063	.0065	37	.0140	.0142	36	.0278	.0283
		7	43	.0065	.0066	41	.0242	.0246	40	.0424	.0435
		8	48	.0061	.0062	46	.0204	.0207	45	.0347	.0355
		9	52	.0098	.0099	51	.0169	.0172	49	.0440	.0461
		10	57	.0082	.0084	55	.0226	.0230	54	.0356	.0364
9	1	3				3	.0122	.0124	3	.0122	.0124
		4	4	.0013	.0014	4	.0013	.0014	3	.0442	.0453
		5	4	.0062	.0063	4	.0062	.0063	4	.0062	.0063
		6	5	.0008	.0009	4	.0169	.0171	4	.0169	.0171
		7	5	.0026	.0027	5	.0026	.0027	4	.0356	.0364
		8	5	.0063	.0064	5	.0063	.0064	5	.0063	.0064
		9	6	.0010	.0011	5	.0129	.0131	5	.0129	.0131
		10	6	.0023	.0024	5	.0234	.0238	5	.0234	.0238
9	2	3				6	.0122	.0124	5	.0481	.0494
		4	7	.0068	.0069	7	.0068	.0069	7	.0068	.0069
		5	8	.0077	.0078	8	.0077	.0078	7	.0301	.0307
		6	9	.0049	.0050	8	.0244	.0248	8	.0244	.0248
		7	10	.0041	.0042	9	.0155	.0158	9	.0155	.0158
		8	11	.0028	.0029	10	.0114	.0116	9	.0368	.0377
		9	11	.0076	.0077	11	.0076	.0077	10	.0258	.0262
		10	12	.0053	.0055	11	.0171	.0173	11	.0171	.0173

TABLE IV (Continued)

p	k	n	α = .01 R	α_L	α_U	α = .025 R	α_L	α_U	α = .05 R	α_L	α_U
9	3	3				9	.0122	.0124	8	.0481	.0494
		4	11	.0068	.0079	10	.0203	.0206	10	.0203	.0206
		5	13	.0032	.0033	12	.0122	.0124	11	.0367	.0375
		6	14	.0064	.0066	13	.0176	.0178	12	.0464	.0477
		7	15	.0092	.0093	14	.0239	.0243	14	.0239	.0243
		8	17	.0047	.0049	16	.0120	.0122	15	.0285	.0290
		9	18	.0062	.0063	17	.0148	.0150	16	.0327	.0333
		10	19	.0076	.0077	18	.0171	.0174	17	.0363	.0371
9	4	3				12	.0122	.0124	11	.0481	.0494
		4	15	.0068	.0069	14	.0203	.0206	13	.0468	.0481
		5	17	.0085	.0086	16	.0220	.0223	16	.0220	.0223
		6	20	.0039	.0040	18	.0232	.0235	17	.0477	.0490
		7	22	.0045	.0047	20	.0219	.0223	19	.0439	.0451
		8	24	.0047	.0048	22	.0205	.0208	21	.0391	.0400
		9	25	.0095	.0096	24	.0185	.0187	23	.0344	.0351
		10	27	.0087	.0088	26	.0165	.0167	25	.0299	.0305
9	5	3				15	.0122	.0124	14	.0481	.0494
		4	19	.0068	.0069	18	.0203	.0206	17	.0468	.0481
		5	22	.0085	.0086	21	.0190	.0193	20	.0398	.0408
		6	25	.0081	.0082	24	.0172	.0174	23	.0341	.0348
		7	28	.0074	.0075	27	.0148	.0150	26	.0278	.0283
		8	31	.0064	.0065	29	.0223	.0226	28	.0389	.0398
		9	34	.0054	.0055	32	.0177	.0180	31	.0303	.0309
		10	36	.0080	.0081	34	.0235	.0239	33	.0384	.0303
9	6	3				18	.0122	.0124	17	.0481	.0494
		4	23	.0068	.0069	22	.0203	.0206	21	.0468	.0481
		5	27	.0085	.0086	26	.0190	.0193	25	.0376	.0385
		6	31	.0077	.0079	30	.0157	.0159	29	.0297	.0302
		7	35	.0066	.0067	33	.0229	.0233	32	.0397	.0406
		8	38	.0099	.0100	37	.0174	.0177	35	.0479	.0492
		9	42	.0076	.0077	40	.0218	.0221	39	.0350	.0238
		10	45	.0098	.0099	44	.0161	.0163	42	.0397	.0406
9	7	3				21	.0122	.0124	20	.0481	.0494
		4	27	.0068	.0069	26	.0203	.0206	25	.0468	.0481
		5	32	.0085	.0086	31	.0190	.0193	30	.0376	.0385
		6	37	.0077	.0079	36	.0155	.0157	35	.0287	.0292
		7	42	.0064	.0065	40	.0215	.0218	39	.0365	.0373
		8	46	.0092	.0093	45	.0159	.0161	43	.0423	.0433
		9	51	.0069	.0070	49	.0191	.0193	47	.0463	.0475
		10	55	.0085	.0086	53	.0315	.0218	52	.0328	.0335

TABLE IV (Continued)

			α = .01			a = .025			α = .05		
p	k	n	R	α_L	α_U	R	α_L	α_U	R	α_L	α_U
9	8	3				24	.0122	.0124	23	.0481	.0494
		4	31	.0068	.0069	30	.0203	.0206	29	.0468	.0481
		5	37	.0085	.0086	36	.0190	.0193	35	.0376	.0385
		6	43	.0077	.0079	42	.0155	.0157	41	.0286	.0291
		7	49	.0064	.0064	47	.0212	.0216	46	.0358	.0365
		8	54	.0090	.0092	53	.0156	.0158	51	.0408	.0418
		9	60	.0067	.0068	58	.0183	.0186	56	.0440	.0451
		10	65	.0082	.0083	63	.0204	.0207	61	.0458	.0470
10	1	3				3	.0099	.0101	3	.0099	.0101
		4	4	.0010	.0011	4	.0010	.0011	3	.0363	.0371
		5	4	.0045	.0047	4	.0045	.0047	4	.0045	.0047
		6	5	.0005	.0006	4	.0126	.0128	4	.0126	.0128
		7	5	.0017	.0018	5	.0017	.0018	4	.0269	.0273
		8	5	.0043	.0044	5	.0043	.0044	5	.0043	.0044
		9	5	.0088	.0090	5	.0088	.0090	5	.0088	.0090
		10	6	.0014	.0015	5	.0162	.0164	5	.0162	.0162
10	2	3				6	.0099	.0101	5	.0392	.0401
		4	7	.0049	.0051	7	.0049	.0051	6	.0420	.0431
		5	8	.0055	.0057	7	.0223	.0227	7	.0223	.0227
		6	9	.0032	.0034	8	.0176	.0179	8	.0418	.0428
		7	10	.0026	.0027	9	.0104	.0106	8	.0418	.0428
		8	10	.0074	.0075	10	.0074	.0075	9	.0252	.0256
		9	11	.0046	.0047	10	.0169	.0171	10	.0169	.0171
		10	12	.0031	.0032	11	.0105	.0107	10	.0334	.0341
10	3	3				9	.0099	.0101	8	.0392	.0401
		4	11	.0049	.0051	10	.0148	.0151	10	.0148	.0151
		5	12	.0085	.0087	12	.0085	.0087	11	.0267	.0272
		6	14	.0041	.0043	13	.0116	.0118	12	.0325	.0332
		7	15	.0057	.0058	14	.0155	.0158	13	.0369	.0377
		8	16	.0072	.0073	15	.0179	.0182	14	.0414	.0424
		9	17	.0086	.0088	16	.0199	.0202	15	.0435	.0446
		10	18	.0097	.0098	17	.0216	.0219	16	.0449	.0461
10	4	3				12	.0099	.0101	11	.0392	.0401
		4	15	.0049	.0051	14	.0148	.0151	13	.0343	.0351
		5	17	.0055	.0057	16	.0149	.0152	15	.0365	.0373
		6	19	.0063	.0064	18	.0150	.0152	17	.0317	.0323
		7	21	.0061	.0062	20	.0133	.0135	19	.0278	.0283
		8	23	.0056	.0057	21	.0235	.0238	20	.0443	.0454
		9	25	.0050	.0051	23	.0196	.0199	22	.0363	.0371
		10	26	.0086	.0087	25	.0162	.0164	24	.0294	.0299

TABLE IV (Continued)

			$\alpha = .01$			$\alpha = .025$			$\alpha = .05$		
p	k	n	R	α_L	α_U	R	α_L	α_U	R	α_L	α_U
10	5	3				15	.0099	.0101	14	.0392	.0401
		4	19	.0049	.0051	18	.0148	.0151	17	.0343	.0351
		5	22	.0055	.0057	21	.0125	.0127	20	.0268	.0273
		6	25	.0048	.0049	23	.0215	.0218	22	.0409	.0418
		7	27	.0085	.0086	26	.0163	.0165	25	.0299	.0305
		8	30	.0065	.0066	28	.0220	.0223	27	.0380	.0388
		9	32	.0091	.0092	31	.0160	.0162	29	.0450	.0462
		10	35	.0067	.0068	33	.0196	.0199	32	.0322	.0328
10	6	3				18	.0099	.0101	17	.0392	.0401
		4	23	.0049	.0051	22	.0148	.0151	21	.0343	.0351
		5	27	.0055	.0057	26	.0125	.0127	24	.0465	.0478
		6	30	.0093	.0095	29	.0180	.0182	28	.0328	.0334
		7	34	.0069	.0070	32	.0227	.0231	31	.0384	.0393
		8	37	.0090	.0091	36	.0155	.0158	34	.0419	.0429
		9	41	.0062	.0063	39	.0175	.0178	37	.0439	.0450
		10	44	.0072	.0074	42	.0189	.0192	40	.0446	.0457
10	7	3				21	.0099	.0101	20	.0392	.0401
		4	27	.0049	.0041	26	.0148	.0151	25	.0343	.0351
		5	32	.0055	.0057	31	.0125	.0127	29	.0451	.0463
		6	36	.0092	.0093	35	.0171	.0173	34	.0302	.0307
		7	41	.0065	.0066	39	.0201	.0204	38	.0333	.0340
		8	45	.0079	.0080	43	.0218	.0221	42	.0344	.0351
		9	49	.0088	.0089	47	.0224	.0227	46	.0342	.0349
		10	53	.0093	.0094	51	.0222	.0226	49	.0484	.0497
10	8	3				24	.0099	.0101	23	.0392	.0401
		4	31	.0049	.0051	30	.0148	.0151	29	.0343	.0351
		5	37	.0055	.0057	36	.0125	.0127	34	.0451	.0463
		6	42	.0092	.0093	41	.0170	.0172	40	.0296	.0301
		7	48	.0064	.0065	46	.0194	.0196	45	.0316	.0322
		8	53	.0075	.0077	51	.0203	.0206	49	.0481	.0494
		9	58	.0082	.0083	56	.0203	.0206	54	.0453	.0465
		10	63	.0084	.0085	61	.0196	.0199	59	.0419	.0429
10	9	3				27	.0099	.0101	26	.0392	.0401
		4	35	.0049	.0051	34	.0148	.0151	33	.0343	.0351
		5	42	.0055	.0057	41	.0125	.0127	39	.0451	.0463
		6	48	.0092	.0093	47	.0170	.0172	46	.0295	.0301
		7	55	.0064	.0065	53	.0192	.0195	52	.0312	.0318
		8	61	.0075	.0076	59	.0200	.0203	57	.0468	.0481
		9	67	.0080	.0082	65	.0197	.0200	63	.0436	.0447
		10	73	.0082	.0083	71	.0189	.0192	69	.0399	.0409

CHAPTER V

PAIRWISE COMPARISONS AND COMPARISONS WITH A CONTROL

In this chapter, the topic of multiple comparisons will be considered. All of the results established will be based upon asymptotic theory, that is, will be based upon the assumption that n is large.

1. Pairwise Comparisons

Let θ_j, $j = 1, 2, \ldots, p$ be p location parameters corresponding to the p populations under consideration. It is assumed that if the population distributions differ, they differ only in location parameters. The null hypothesis is

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_p .$$

The first alternative to be considered is

$$H_{A_1} : \theta_j \neq \theta_{j'} \text{ for at least one set of } (j, j'), \; j \neq j' .$$

The following two theorems will be needed to establish some of the results of this chapter. The first theorem may be found on page 108 of C. R. Rao's book [9]. The second theorem may be found on page 90 of Tucker's text [11].

Theorem 5.1: Consider independent identically distributed k-dimensional variables

$$\vec{U}_n' = (U_{1n}, \ldots, U_{kn}), \quad n = 1, 2, \ldots$$

which have first and second order moments,

$$E(\vec{U}_n) = \mu, \quad D(\vec{U}_n) = \Sigma .$$

Define the sequence of random variables

$$\vec{\overline{U}}_n = (\vec{\overline{U}}_{1n}, \ldots, \vec{\overline{U}}_{kn}), \quad n = 1, 2, \ldots$$

where $\overline{U}_{in} = \frac{1}{n} \sum_{j=1}^{n} U_{ij}$. Then the asymptotic distribution of $\sqrt{n}\,(\vec{\overline{U}}_n - \mu)$ is $N_k(0, \Sigma)$.

Theorem 5.2: Let $T : E^{(k)} \rightarrow E^{(m)}$ be a continuous mapping, and let $X^{(1)}, X^{(2)}, \ldots$ be a sequence of k-dimensional random variables such that $F_{X^{(n)}} \xrightarrow{\mathfrak{L}} F_X$, then $F_{TX^{(n)}} \xrightarrow{\mathfrak{L}} F_{TX}$.

Lemma 5.1: $f(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n)$ is a continuous transformation.

Proof:

$$\begin{aligned} |f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| &= |\max\{x_1, \ldots, x_n\} - \max\{y_1, \ldots, y_n\}| \\ &= |x_i - y_j| \\ &\leq \max\{|x_i - y_i|, |x_j - y_j|\} \\ &\leq \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \end{aligned}$$

$$= \|(x_1, \ldots, x_n) - (y_1, \ldots, y_n)\| .$$

If $\|(x_1, \ldots, x_n) - (y_1, \ldots, y_n)\| < \delta$, then $|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| < \epsilon$ when $\epsilon = \delta$. △

Now, define column means in the partial ordering scheme as follows,

$$\bar{r}_j = \frac{r_{.j}}{n} = \frac{\sum_{i=1}^{n} r_{ij}}{n} .$$

Then,

$$\operatorname{var}(\bar{r}_j) = \frac{\operatorname{var}(r_{ij})}{n} = \frac{\sigma^2}{n}$$

$$\operatorname{cov}(\bar{r}_j, \bar{r}_{j'}) = -\frac{1}{n(p-1)} \operatorname{var}(r_{ij})$$

$$\operatorname{cor}(\bar{r}_j, \bar{r}_{j'}) = -\frac{1}{p-1} .$$

Now, consider the set of $\frac{p(p-1)}{2}$ differences $(\bar{r}_j - \bar{r}_{j'})$.

$$E(\bar{r}_j - \bar{r}_{j'}) = 0, \quad j \neq j', \quad j = 1, 2, \ldots, p$$

$$\operatorname{var}(\bar{r}_j - \bar{r}_{j'}) = \operatorname{var}(\bar{r}_j) + \operatorname{var}(\bar{r}_{j'}) - 2 \operatorname{cov}(\bar{r}_j, r_{j'})$$

$$= \frac{2}{n} \operatorname{var}(r_{ij}) - 2\left(-\frac{1}{n(p-1)} \operatorname{var}(r_{ij})\right)$$

$$= \frac{2p}{n(p-1)} \operatorname{var}(r_{ij}), \quad j \neq j'; \; j = 1, 2, \ldots, p .$$

Next, consider the covariance matrix of the differences $(\overline{r}_j - \overline{r}_{j'})$, $j \neq j'$, i.e., look at terms of the following form,

$$\mathrm{cov}\,(\overline{r}_j - \overline{r}_{j'}, \overline{r}_{j''} - \overline{r}_{j'''}), \quad j \neq j', \quad j'' \neq j''' .$$

$j \neq j'$ and $j'' \neq j'''$ implies the following cases.

1. (j, j', j'', j''') all j's are unique.
2. (j, j', j, j'')
3. (j, j', j'', j)
4. (j', j, j, j'')
5. (j', j, j'', j)
6. (j, j', j, j')
7. (j, j', j', j)

Using the fact that

$$\mathrm{cov}\,(X_1 - X_2, X_3 - X_4) = \mathrm{cov}\,(X_1, X_3) - \mathrm{cov}\,(X_2, X_3) - \mathrm{cov}\,(X_1, X_4) + \mathrm{cov}\,(X_2, X_4) ,$$

it may be verified that

$$\mathrm{cov}\,(\overline{r}_j - \overline{r}_{j'}, \overline{r}_{j''} - \overline{r}_{j'''}) = 0$$

$$\mathrm{cov}\,(\overline{r}_j - \overline{r}_{j'}, \overline{r}_j - \overline{r}_{j''}) = \frac{p}{n(p-1)}\,\mathrm{var}\,(r_{ij})$$

$$\mathrm{cov}\,(\overline{r}_j - \overline{r}_{j'}, \overline{r}_{j''} - \overline{r}_j) = -\frac{p}{n(p-1)}\,\mathrm{var}\,(r_{ij})$$

$$\mathrm{cov}\,(\overline{r}_{j'} - \overline{r}_j, \overline{r}_j - \overline{r}_{j''}) = -\frac{p}{n(p-1)}\,\mathrm{var}\,(r_{ij})$$

$$\text{cov}(\bar{r}_{j'} - \bar{r}_j, \bar{r}_{j''} - \bar{r}_j) = \frac{p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(\bar{r}_j - \bar{r}_{j'}, \bar{r}_j - \bar{r}_{j'}) = \frac{2p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(\bar{r}_j - \bar{r}_{j'}, \bar{r}_{j'} - \bar{r}_j) = -\frac{2p}{n(p-1)} \text{var}(r_{ij})$$

Now, let $Z_1, \ldots, Z_p$ be independent normal random variables with mean zero and variance $\frac{p}{n(p-1)} \text{var}(r_{ij})$. Then the covariance matrix of the differences $Z_j - Z_{j'}$, $j \neq j'$, $j = 1, \ldots, p$ is the same as that of the differences $(\bar{r}_j - \bar{r}_{j'})$, $j = 1, \ldots, p$, $j \neq j'$, i.e.,

$$\text{cov}(Z_j - Z_{j'}, Z_{j''} - Z_{j'''}) = 0$$

$$\text{cov}(Z_j - Z_{j'}, Z_j - Z_{j''}) = \frac{p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(Z_j - Z_{j'}, Z_{j''} - Z_j) = -\frac{p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(Z_{j'} - Z_j, Z_j - Z_{j''}) = -\frac{p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(Z_{j'} - Z_j, Z_{j''} - Z_j) = \frac{p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(Z_j - Z_{j'}, Z_j - Z_{j'}) = \frac{2p}{n(p-1)} \text{var}(r_{ij})$$

$$\text{cov}(Z_j - Z_{j'}, Z_{j'} - Z_j) = -\frac{2p}{n(p-1)} \text{var}(r_{ij})$$

<u>Theorem 5.3</u>: The asymptotic distribution of

$$\max_{j,j'} \frac{\{|\bar{r}_j - \bar{r}_{j'}|\}}{\sqrt{\frac{p}{n(p-1)} \text{var}(r_{ij})}}$$

coincides with the distribution of

$$\max_{j,j'} \frac{\{|Z_j - Z_{j'}|\}}{\sqrt{\frac{p}{n(p-1)} \operatorname{var}(r_{ij})}} .$$

<u>Proof:</u> In Theorem 5.1, let the elements of $\vec{U}_i'$ be of the form $(r_{ij} - r_{ij'})$. The elements of $\vec{\overline{U}}_i$ are of the form $(\bar{r}_j - \bar{r}_{j'})$. Then Theorem 5.1 implies $\sqrt{n}\ (\vec{\overline{U}}_n - 0)$ has an asymptotic distribution that is $N_q(0, \Sigma)$, where $q = \frac{1}{2} p(p-1)$. Let Z be the vector of differences $Z_j - Z_{j'}$. Then $\sqrt{n}\ \overline{U}_n$ and $\sqrt{n}\ Z$ for n large, both have a multivariate normal distribution with the same mean vector and covariance matrix. The distribution of

$$\max_{j,j'} \frac{\{|\bar{r}_j - \bar{r}_{j'}|\}}{\sqrt{\frac{p}{n(p-1)} \operatorname{var}(r_{ij})}}$$

coincides with the distribution of

$$\max_{j,j'} \frac{\{|Z_j - Z_{j'}|\}}{\sqrt{\frac{p}{n(p-1)} \operatorname{var}(r_{ij})}}$$

since a continuous transformation has been made. Lemma 5.1 and Theorem 5.2 give the desired results. △

The distribution of

$$\max_{j,j'} \frac{\{|Z_j - Z_{j'}|\}}{\sqrt{\frac{p}{n(p-1)} \operatorname{var}(r_{ij})}}$$

is the distribution of the range of p independent unit normal random variables. The symbol used in the literature to represent this random variable is $Q_{p,\infty}$. A discussion of $Q_{p,\infty}$ may be found in Chapter II of [7]. The critical points $q^{\alpha}_{p,\infty}$ may be found in Table I of Appendix B of [7].

The following probability statement holds approximately for large n,

$$P\left\{\max_{j,j'} \frac{\{|\bar{r}_j - \bar{r}_{j'}|\}}{\sqrt{\frac{p}{n(p-1)} \operatorname{var}(r_{ij})}} \leq q^{\alpha}_{p,\infty}\right\} = 1 - \alpha .$$

This then implies the following inequalities hold with probability $1-\alpha$.

$$|\bar{r}_j - \bar{r}_{j'}| \leq \sqrt{\frac{p}{n(p-1)} \operatorname{var}(r_{ij})} \quad q^{\alpha}_{p,\infty} \tag{5.1}$$

or

$$|\bar{r}_j - \bar{r}_{j'}| \leq \sqrt{\frac{k(k+1)}{n(p-1)}\left(\frac{2k+1}{6} - \frac{k(k+1)}{4p}\right)} \cdot q^{\alpha}_{p,\infty} .$$

Any difference $|\bar{r}_j - \bar{r}_{j'}|$ that exceeds the critical value is taken to indicate that $\theta_j \neq \theta_{j'}$.

2. Comparisons With a Control Population

Suppose now that a control population is introduced so that there is a total of $p+1$ populations. Each of the n judges is asked to rank the best $k+1$ and assign the score zero to the remaining $(p-k)$. Let θ_0 be the location parameter in the control population and without loss of generality suppose $\theta_0 = 0$. The null hypothesis is

$$H_0 : \theta_1 = \ldots = \theta_p = 0.$$

The first alternative to be considered is

$$H_{A_1} : \theta_j > 0 \quad \text{for at least one } j, \quad j = 1, 2, \ldots, p .$$

Consider the vector of differences $\bar{r}_j - \bar{r}_0$ where $\bar{r}_0$ is the column mean for the control population and $\bar{r}_j$ is the column mean for the j^{th} population.

Taking the expressions for $\text{var}(\bar{r}_j - \bar{r}_{j'})$ and $\text{cov}(\bar{r}_j - \bar{r}_{j'}, \bar{r}_{j''} - \bar{r}_{j'})$ and replacing p by $p+1$ and k by $k+1$ one finds,

$$\text{var}(\bar{r}_j - \bar{r}_0) = \frac{2(p+1)}{np} \text{var}(r_{ij})$$

$$\text{cov}(\bar{r}_j - \bar{r}_0, \bar{r}_{j'} - \bar{r}_0) = \frac{p+1}{np} \text{var}(r_{ij})$$

where

$$\text{var}(r_{ij}) = \frac{(k+1)(k+2)}{p+1} \left[\frac{2k+3}{6} - \frac{(k+1)(k+2)}{4(p+1)} \right] .$$

Let $Z_0, Z_1, \ldots, Z_p$ be $p+1$ identically distributed normal random variables with mean zero and variance $\frac{p+1}{np} \text{var}(r_{ij})$. Then consider the vector of differences $Z_j - Z_0$, $j = 1, \ldots, p$.

$$\text{var}(Z_j - Z_0) = \frac{2(p+1)}{np} \text{var}(r_{ij}) \quad j = 1, 2, \ldots, p$$

$$\text{cov}(Z_j - Z_0, Z_{j'} - Z_0) = \frac{p+1}{np}\,\text{var}(r_{ij}) \qquad j \neq j' .$$

The correlation between $Z_j - Z_0$ and $Z_{j'} - Z_0$ and the correlation between $\bar{r}_j - \bar{r}_0$ and $\bar{r}_{j'} - \bar{r}_0$ are of interest also.

$$\text{cor}(Z_j - Z_0, Z_{j'} - Z_0) = \text{cor}(\bar{r}_j - \bar{r}_0, \bar{r}_{j'} - \bar{r}_0) = \frac{1}{2} .$$

By using similar reasoning as that given in Theorem 5.3, it may be shown that the asymptotic distribution of

$$\max_j \frac{\{\bar{r}_j - \bar{r}_0\}}{\sqrt{\frac{2(p+1)}{np}\,\text{var}(r_{ij})}}$$

coincides with the distribution of

$$\max_j \frac{\{Z_j - Z_0\}}{\sqrt{\frac{2(p+1)}{np}\,\text{var}(r_{ij})}} .$$

Let

$$M_{p(\frac{1}{2})} = \max_j \frac{\{Z_j - Z_0\}}{\sqrt{\frac{2(p+1)}{np}\,\text{var}(r_{ij})}} .$$

$M_{p(\frac{1}{2})}$ is then distributed as the maximum of p equally correlated $(\rho = \frac{1}{2})$ unit normal random variables. The critical points $m^{\alpha}_{p(\frac{1}{2})}$ may be found in Table IV of Appendix B of Miller's book [7]. The following probability statement holds approximately for large n.

$$P\left\{\max_j \frac{\{\bar{r}_j - \bar{r}_0\}}{\sqrt{\frac{2(p+1)}{np}\,\mathrm{var}\,(r_{ij})}} \le m^{\alpha}_{p(\frac{1}{2})}\right\} = 1 - \alpha\ .$$

This implies the following inequalities hold with approximately probability $1 - \alpha$,

$$\bar{r}_j - \bar{r}_0 \le \sqrt{\frac{2(p+1)}{np}\,\mathrm{var}\,(r_{ij})}\quad m^{\alpha}_{p(\frac{1}{2})} \tag{5.2}$$

or

$$\bar{r}_j - \bar{r}_0 \le \sqrt{\frac{2(k+1)(k+2)}{np}\left(\frac{2k+3}{6} - \frac{(k+1)(k+2)}{4(p+1)}\right)} \cdot m^{\alpha}_{p(\frac{1}{2})}\ .$$

Now, if any $\bar{r}_j - \bar{r}_0$ exceeds the right hand side this is taken to indicate that $\theta_j > 0$.

Instead of alternative H_{A_1}, suppose the alternative is H_{A_2} where

$$H_{A_2} : \theta_j \ne 0 \qquad j = 1, 2, \ldots, p\ .$$

Again, consider $p+1$ identically distributed normal random variables $Z_0, \ldots, Z_p$ each with mean 0 and variance $\frac{p+1}{np}\,\mathrm{var}\,(r_{ij})$. Then the asymptotic distribution of

$$\max_j \frac{\{|\bar{r}_j - \bar{r}_0|\}}{\sqrt{\frac{2(p+1)\,\mathrm{var}\,(r_{ij})}{np}}}$$

coincides with the distribution of

$$\max_{j} \frac{\{|Z_j - Z_0|\}}{\sqrt{\frac{2(p+1)}{np} \operatorname{var}(r_{ij})}} = |M|_{p(\frac{1}{2})} ,$$

the absolute maximum of p equally correlated $(\rho = \frac{1}{2})$ unit normal random variables. The critical points $|m|^{\alpha}_{p(\frac{1}{2})}$ may be found in Table IV of Appendix B of Miller's book [7]. The following inequalities hold with probability approximately $1 - \alpha$.

$$|\bar{r}_j - \bar{r}_o| \leq \sqrt{\frac{2(p+1)}{np} \operatorname{var}(r_{ij})} \tag{5.3}$$

or

$$|\bar{r}_j - \bar{r}_0| \leq \sqrt{\frac{2(k+1)(k+2)}{np} \left(\frac{2k+3}{6} - \frac{(k+1)(k+2)}{4(p+1)}\right)} \cdot |m|^{\alpha}_{p(\frac{1}{2})}$$

$$j = 1, 2, \ldots, p .$$

Any difference, $|\bar{r}_j - \bar{r}_0|$ which exceeds the right hand side is taken to indicate that $\theta_j \neq 0$.

CHAPTER VI

A PARTIAL ORDERING USING LINEARLY WEIGHTED SCORES

This chapter is devoted to extending the results in Chapters III, IV and V to problems such as that illustrated by Example 1.3. Suppose each of n judges is presented p objects of which the k most preferred objects are to be ranked in ascending order of preference, according to some criterion of interest, with the remaining objects assigned a "rank" of zero. It will be assumed that each of the assumptions stated in Section 1.2 is valid. In addition, it is assumed that each rank r assigned to a given object is replaced by the score s_r where

$$\begin{aligned} s_r &= 0 \quad \text{if} \quad r = 0 \\ &= a_1 \cdot r + b_1 \quad \text{if} \quad r = 1, 2, \ldots, k_1 \\ &= a_2 \cdot (r - k_1) + b_2 \quad \text{if} \quad r = k_1 + 1, \ldots, k . \end{aligned} \tag{6.1}$$

It is further assumed that the fixed integers a_1, a_2, b_1 and b_2 are such that a_1 and a_2 are positive and $s_0 < s_1 < s_2 < \ldots < s_k$. It will be computationally convenient to define k_2 as $k - k_1$.

The column totals are given by

$$r_j = \sum_{i=1}^{n} r_{ij} \quad \text{for} \quad j = 1, 2, \ldots, p$$

where r_{ij} is the score from (6.1) which is assigned to the j^{th} object by the i^{th} judge. The average of the column sums is given by

$$\bar{r} = \sum_{j=1}^{p} \frac{r_j}{p} = \frac{n}{p}\left[\frac{1}{2} a_1 k_1(k_1+1) + b_1 k_1 + \frac{1}{2} a_2 k_2(k_2+1) + b_2 k_2\right] .$$

The statistic

$$S = \sum_{j=1}^{p} (r_j - \bar{r})^2$$

may be used to test the null hypothesis that there is no difference in preferences for the p objects against the alternative that at least one object is preferred to at least one other object. Under H_0, the distribution of r_{ij} is given by

$$p(r_{ij} = \alpha) = \frac{1}{p} \quad \text{for} \quad \alpha = s_r, \quad r = 1, 2, \ldots, k$$

$$= \frac{p-k}{p} \quad \text{for} \quad \alpha = 0 .$$

The mean μ_r is given by

$$\mu_r = E(r_{ij}) = \sum_{\ell=1}^{k_1} (a_1 \ell + b_1) \cdot \frac{1}{p} + \sum_{\ell=1}^{k_2} (a_2 \ell + b_2) \cdot \frac{1}{p} + 0 \cdot \frac{p-k}{p}$$

$$= \left(\frac{a_1 k_1(k_1+1)}{2} + b_1 k_1\right) \cdot \frac{1}{p} + \left(\frac{a_2 k_2(k_2+1)}{2} + b_2 k_2\right) \cdot \frac{1}{p} . \qquad (6.2)$$

The variance σ_r^2 is given by

$$\sigma_r^2 = E(r_{ij}^2) - \mu_r^2$$

$$= \sum_{\ell=1}^{k_1} (a_1\ell + b_1)^2 \cdot \frac{1}{p} + \sum_{\ell=1}^{k_2} (a_2\ell + b_2)^2 \cdot \frac{1}{p} - \mu_r^2$$

$$= \frac{a_1^2 k_1(k_1+1)(2k_1+1)}{6p} + \frac{a_1 b_1 k_1(k_1+1)}{p} + \frac{k_1 b_1^2}{p}$$

$$+ \frac{a_2^2 k_2(k_2+1)(2k_2+1)}{6p} + \frac{a_2 b_2 k_2(k_2+1)}{p} + \frac{k_2 b_2^2}{p}$$

$$- \left[\left(\frac{a_1 k_1(k_1+1)}{2} + b_1 k_1\right) \cdot \frac{1}{p} + \left(\frac{a_2 k_2(k_2+1)}{2} + b_2 k_2\right) \cdot \frac{1}{p}\right]^2 . \qquad (6.3)$$

For fixed values of the parameters p, k, k_1, a_1, a_2, b_1 and b_2 the exact distribution of the statistic S may be constructed corresponding to small values of n. The procedure is identical to that described in Section 3.2 where the $\frac{p!}{(p-k)!}$ vectors of scores available to each judge are now the permutations of the p-vector $(s_k, \ldots, s_2, s_1, 0, \ldots, 0)$. Observe that this statistic reduces to the one given in Chapter III for $a_1 = a_2 = 1$, $b_1 = 0$ and $b_2 = k_1$.

Since the statistic S is a special case of the statistic S_n defined by Sen [10] and given as equation (2.4), it follows that

$$\chi_r^2 = \frac{p-1}{np\sigma_r^2} S$$

has a limiting χ^2 distribution with $(p-1)$ degrees of freedom.

Also, if desired, a coefficient of concordance, W, could be defined as

$$W = \frac{S}{S_{max}}$$

where S_{max} is the maximum value of S occuring when the r_j's are permutations of the numbers $n(a_1 + b_1)$, $n(2a_1 + b_1), \ldots, n(k_2 a_2 + b_2)$, and 0 (with a multiplicity of $(p - k)$) .

The procedure for extending the extreme rank sum test of Chapter IV to the case of linearly weighted scores will be based upon more general versions of Theorems 4.1 and 4.3. Theorem 4.2 and Lemmas 4.1 and 4.2 are directly applicable, however, some additional notation will be necessary. Define the sets G, G_1 and G_2 as

$$G = \{ s_r | r = 0, 1, \ldots, k\}$$

$$G_1 = \{ s_r | = 0, 1, \ldots, k_1\}$$

and

$$G_2 = \{ s_r | r = k_1 + 1, \ldots, k\} \ .$$

Also define $N(s)$ as the multiplicity of scores available to each judge having a value less than or equal to s for s an element of G. That is, $N(0) = p - k$, $N(s_1) = (p - k) + 1, \ldots, N(s_k) = p$.

<u>Theorem 6.1</u>: Under H_0, $P(r_i > a \text{ and } r_j > b) \leq P(r_i > a) P(r_j > b)$ where r_i and r_j are two column totals in the partial ordering scheme using linearly weighted scores and a and b are admissable values for the column sums.

<u>Proof</u>: Using Lemma 4.2, the theorem will be proved by showing that $P(r_i \leq a \text{ and } r_j \leq b) \leq P(r_i \leq a) P(r_j \leq b)$. The proof will be by mathematical induction.

For $n = 1$ judge, the discussion is as follows. Suppose $a = b = 0$, then

$$P(r_i \leq 0 \text{ and } r_j \leq 0) = \frac{(p-k)(p-k-1)}{p(p-1)}$$

$$P(r_i \leq 0) = P(r_j \leq 0) = \frac{p-k}{p}$$

and

$$\frac{(p-k-1)}{(p-1)} < \frac{(p-k)}{p}$$

implies that $P(r_i \leq 0 \text{ and } r_j \leq 0) < P(r_i \leq 0)\,P(r_j \leq 0)$. Suppose $a = b = \ell \neq 0$

$$P(r_i \leq \ell \text{ and } r_j \leq \ell) = \frac{N(\ell)\,[N(\ell) - 1]}{p(p-1)}$$

$$P(r_i \leq \ell) = P(r_j \leq \ell) = \frac{N(\ell)}{p}$$

and again $P(r_i \leq \ell \text{ and } r_j \leq \ell) < P(r_i \leq \ell)\,P(r_j \leq \ell)$. Suppose $b < a$, then

$$P(r_i \leq a \text{ and } r_j \leq b) = P(r_i \leq a \mid r_j \leq b) \cdot P(r_j \leq b)$$

$$= \frac{N(a) - 1}{p - 1} \cdot \frac{N(b)}{p}$$

and

$$P(r_i \leq a) \cdot P(r_j \leq b) = \frac{N(a)}{p} \cdot \frac{N(b)}{p}$$

and again,

$$\frac{N(a) - 1}{p - 1} < \frac{N(a)}{p}$$

implies that $P(r_i \leq a \text{ and } r_j \leq b) < P(r_i \leq a)\, P(r_j \leq b)$. A similar argument holds for $a < b$.

Now, suppose the desired result is true for $n = m$ judges, i.e., assume

$$P(r_i \leq a \text{ and } r_j \leq b \mid m) \leq P(r_i \leq a \mid m)\, P(r_j \leq b \mid m)$$

and show that it follows for $n = m+1$ judges.

$$P(r_i \leq a \text{ and } r_j \leq b \mid m+1) =$$

$$\sum_{e=0}^{k} \sum_{f=0}^{k} \cdot P(r_i \leq a \text{ and } r_j \leq b \mid r_{m+1,i} = s_e \text{ and } r_{m+1,j} = s_f)$$

$$\cdot P(r_{m+1,i} = s_e \text{ and } r_{m+1,j} = s_f)$$

$$= \sum_{e=0}^{k} \sum_{f=0}^{k} P(r_i \leq a - s_e \text{ and } r_j \leq b - s_f \mid m)$$

$$\cdot P(r_{m+1,i} = s_e \text{ and } r_{m+1,j} = s_f)$$

$$= P(r_i \leq a \text{ and } r_j \leq b \mid m) \frac{(p-k)(p-k-1)}{p(p-1)}$$

$$+ \frac{(p-k)}{p(p-1)} \sum_{f=1}^{k} P(r_i \leq a \text{ and } r_j \leq b - s_f \mid m)$$

$$+ \frac{(p-k)}{p(p-1)} \sum_{e=1}^{k} P(r_i \leq a - s_e \text{ and } r_j \leq b \mid m)$$

$$+ \sum_{f=1}^{k} \sum_{\substack{e=1 \\ e \neq f}}^{k} P(r_i \leq a - s_e \text{ and } r_j \leq b - s_f \mid m) \cdot \frac{1}{p(p-1)}$$

$$\leq P(r_i \leq a \mid m)\, P(r_j \leq b \mid m) \cdot \frac{(p-k)(p-k-1)}{p(p-1)}$$

$$+ \frac{(p-k)}{p(p-1)} P(r_i \le a|m) \sum_{f=1}^{k} P(r_j \le b - s_f|m)$$

$$+ \frac{(p-k)}{p(p-1)} P(r_j \le b|m) \sum_{e=1}^{k} P(r_i \le a - s_e|m)$$

$$+ \sum_{\substack{e=1 \\ e \ne f}}^{k} \sum_{f=1}^{k} P(r_i \le a - s_e|m) P(r_j \le b - s_f|m) \cdot \frac{1}{p(p-1)} \qquad (6.4)$$

by the induction assumption for $n = m$ judges. Now, $P(r_i \le a|m+1) P(r_j \le b|m+1)$ may be written as,

$$P(r_i \le a|m+1) P(r_j \le b|m+1)$$

$$= \left[P(r_i \le a|m) \cdot \frac{p-k}{p} + \sum_{e=1}^{k} P(r_i \le a - s_e|m) \cdot \frac{1}{p}\right]$$

$$\cdot \left[P(r_j \le b|m) \cdot \frac{p-k}{p} + \sum_{f=1}^{k} P(r_j \le b - s_f|m) \cdot \frac{1}{p}\right]$$

$$= P(r_i \le a|m) P(r_j \le b|m) \cdot \frac{(p-k)^2}{p^2}$$

$$+ P(r_i \le a|m) \sum_{f=1}^{k} P(r_j \le b - s_f|m) \cdot \frac{(p-k)}{p^2}$$

$$+ P(r_j \le b|m) \sum_{e=1}^{k} P(r_i \le a - s_e|m) \cdot \frac{(p-k)}{p^2}$$

$$+ \sum_{e=1}^{k} \sum_{f=1}^{k} P(r_i \le a - s_e|m) P(r_j \le b - s_f|m) \cdot \frac{1}{p^2} \quad .$$

If

$$P(r_i \le a|m+1)P(r_j \le b|m+1) - \Bigg[P(r_i \le a|m)P(r_j \le b|m)\cdot\frac{(p-k)^2}{p^2}$$

$$+ P(r_i \le a|m)\sum_{f=1}^{k} P(r_j \le b - s_f|m)\cdot\frac{(p-k)}{p^2}$$

$$+ P(r_j \le b|m)\sum_{e=1}^{k} P(r_i \le a - s_e|m)\cdot\frac{(p-k)}{p^2}$$

$$+ \sum_{e=1}^{k}\sum_{f=1}^{k} P(r_i \le a - s_e|m)P(r_j \le b - s_f|m)\cdot\frac{1}{p^2}\Bigg] ,$$

which equals zero, is added to the right hand side of inequality (6.4), then the inequality may be rewritten as,

$$P(r_i \le a \text{ and } r_j \le b|m+1) \le P(r_i \le a|m+1)P(r_j \le b|m+1)$$

$$- \frac{(p-k)k}{p^2(p-1)} P(r_i \le a|m)P(r_j \le b|m)$$

$$+ \frac{(p-k)}{p^2(p-1)} P(r_i \le a|m)\sum_{f=1}^{k} P(r_j \le b - s_f|m)$$

$$+ \frac{(p-k)}{p^2(p-1)} P(r_j \le b|m)\sum_{e=1}^{k} P(r_i \le a - s_e|m)$$

$$+ \frac{1}{p^2(p-1)} \sum_{e=1}^{k}\sum_{\substack{f=1\\ e\neq f}}^{k} P(r_i \le a - s_e|m)P(r_j \le b - s_f|m)$$

$$- \frac{1}{p^2}\sum_{e=1}^{k} P(r_i \le a - s_e|m)P(r_j \le b - s_e|m) \quad . \qquad (6.5)$$

Define A and B as follows,

$$A = \frac{\sum_{f=1}^{k} P(r_i \le a - s_f|m) + (p-k)P(r_i \le a|m)}{p}$$

$$B = \frac{\sum_{f=1}^{k} P(r_j \le b - s_f \,|\, m) + (p-k)\, P(r_j \le b \,|\, m)}{p} .$$

Now, the inequality (6.5) may be written as

$$P(r_i \le a \text{ and } r_j \le b \,|\, m+1) \le P(r_i \le a \,|\, m+1)\, P(r_j \le b \,|\, m+1)$$
$$- \frac{1}{p(p-1)} \left\{ \sum_{e=1}^{k} \left[P(r_i \le a - s_e \,|\, m) - A \right] \right.$$
$$\cdot \left[P(r_j \le b - s_e \,|\, m) - B \right]$$
$$\left. + (p-k) \left[P(r_i \le a \,|\, m - A \right] \left[P(r_j \le b \,|\, m) - B \right] \right\} .$$

Applying Lemma 4.1 to the quantities

$$P(r_i \le a - s_e \,|\, m), \quad e = 1, 2, \ldots, k, \quad P(r_i \le a \,|\, m) \text{ (p - k multiplicity)}$$

and the quantities,

$$P(r_j \le b - s_e \,|\, m), \quad e = 1, 2, \ldots, k, \quad P(r_j \le b \,|\, m) \text{ (p - k multiplicity)}$$

yields the result that

$$P(r_i \le a \text{ and } r_j \le b \,|\, m+1) \le P(r_i \le a \,|\, m+1)\, P(r_j \le b \,|\, m+1) .$$

Therefore by induction and Lemma 4.2,

$$P(r_i > a \text{ and } r_j > b) \le P(r_i > a)\, P(r_j > b)$$

for any number of judges n. △

The marginal distribution of a column sum, r_j, under H_0, will now be determined. Some lemmas will be needed before Theorem 6.2 is proved.

Lemma 6.1: Suppose there are n cells and each cell may contain any one of the integers from 1 to k. Let T be the sum of the integers in the n cells. The number of partitions that give the sum T is given by

$$\sum_{x=0}^{\infty} I_{T-kx-n} \binom{n}{x} \binom{T-kx-1}{n-1} (-1)^x , \quad n \leq T \leq nk$$

where

$$I_y = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases} .$$

The proof of this result may be found in page 439 of [2].

Lemma 6.2: Suppose there are n cells and each cell may contain any one of the weighted scores $a+b, 2a+b, \ldots, ka+b$. Let T be the sum of the integers in the n cells. The number of partitions that give the sum T is given by

$$\sum_{x=0}^{\infty} I_{\left[\frac{T-nb}{a}-kx-n\right]} \binom{n}{x} \binom{\frac{T-nb}{a}-kx-1}{n-1} (-1)^x$$

where the admissible values of T lie between $n(a+b)$ and $n(ak+b)$, inclusive, and

$$I_y = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 . \end{cases}$$

The proof is an immediate consequence of Lemma 6.1.

<u>Lemma 6.3</u>: Suppose there are p objects and n judges and that each judge assigns the following weighted scores: $a+b, 2a+b, \ldots, ka+b$. The other $(p-k)$ objects are assigned the value zero. The marginal distribution of r_j, the j^{th} column total, under H_0, is given by the following

$$P(r_j = M) = \sum_{i=0}^{n-1} \binom{n}{i} \left(\frac{p-k}{p}\right)^i \left(\frac{k}{p}\right)^{n-i} \cdot \sum_{x=0}^{\infty} I_{[\frac{M-(n-i)b}{a} - kx - n + i]} \binom{n-i}{x}$$

$$\cdot \binom{\frac{M-(n-i)b}{a} - kx - 1}{n-i-1} (-1)^x \Big/ k^{n-i} + \left(\frac{p-k}{p}\right)^n J_M$$

where

$$I_y = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases}, \quad J_M = \begin{cases} 1, & M = 0 \\ 0, & M \neq 0 \end{cases}.$$

The proof of Lemma 6.3 is similar to the proof of Theorem 4.3.

<u>Theorem 6.2</u>: The distribution of r_j for a partial ordering using the linearly weighted scores in (6.1), under H_0, is as follows

$$P(r_j = M) = P(r_j = M \mid \beta = 0) \left(\frac{p-k_2}{p}\right)^n +$$

$$+ \sum_{\ell=1}^{n-1} \left(\sum_{g=\ell}^{\ell k_2} P(r_{.j} = M \mid \beta = \ell ,\ S_{II} = a_2 g + \ell b_2) \right)$$

$$\cdot\ P(S_{II} = a_2 g + \ell b_2 \mid \beta = \ell) \cdot \binom{n}{\ell} \left(\frac{k_2}{p}\right)^{\ell} \left(\frac{p-k_2}{p}\right)^{n-\ell}$$

$$+ P(r_{.j} = M \mid \beta = n) \left(\frac{k_2}{p}\right)^{n}$$

where β is the number of elements from G_2 in the partition of $r_{.j} = M$, and S_{II} is the sum of those elements and,

(i)

$$P(r_{.j} = M \mid \beta = 0) = \sum_{i=0}^{n-1} \binom{n}{i} \left(\frac{p-k_1-k_2}{p-k_2}\right)^{i} \left(\frac{k_1}{p-k_2}\right)^{n-i}$$

$$\cdot \sum_{x=0}^{\infty} I_{\left[\frac{M-(n-i)b_1}{a_1} - k_1 x - n + i\right]}$$

$$\cdot \binom{n-i}{x} \binom{\frac{M-(n-i)b_1}{a_1} - k_1 x - 1}{n-i-1} (-1)^{x} \Big/ k_1^{\,n-i}$$

$$+ \left(\frac{p-k_1-k_2}{p-k_2}\right)^{n} J_M$$

and

$$I_y = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases} , \qquad J_M = \begin{cases} 1, & M = 0 \\ 0, & M \neq 0 \end{cases} .$$

(ii)

$$P(r_j = M \mid \beta = \ell, S_{II} = a_2 g + \ell b_2) = \sum_{i=0}^{n-\ell-1} \binom{n-\ell}{i} \left(\frac{p-k_1-k_2}{p-k_2}\right)^i \left(\frac{k_1}{p-k_2}\right)^{n-\ell-i}$$

$$\cdot \sum_{x=0}^{\infty} I_{\left[\frac{M-a_2 g-\ell b_2-(n-\ell-i)b_1}{a_1} - k_1 x-(n-\ell)+i\right]} \binom{n-\ell-i}{x}$$

$$\cdot \binom{\frac{M-a_2 g-\ell b_2-(n-\ell-i)b_1}{a_1} - k_1 x - 1}{n-\ell-i-1} (-1)^x \Big/ k_1^{n-\ell-i}$$

$$+ \left(\frac{p-k_1-k_2}{p-k_2}\right)^{n-\ell} J_{(M-a_2 g-\ell b_2)},$$

$$J_{(M-a_2 g-\ell b_2)} = \begin{cases} 1, & M-a_2 g-\ell b_2 = 0 \\ 0, & M-a_2 g-\ell b_2 \neq 0 \end{cases}$$

(iii)

$$P(S_{II} = a_2 g + \ell b_2 \mid \beta = \ell) = \frac{\sum_{x=0}^{\infty} I_{[g-k_2 x-\ell]} \binom{\ell}{x} \binom{g-k_2 x-1}{\ell-1} (-1)^x}{k_2^{\ell}}$$

and

(iv)

$$P(r_j = M \mid \beta = n) = \sum_{x=0}^{\infty} I_{\left[\frac{M-nb_2}{a_2} - k_2 x - n\right]} \binom{n}{x} \binom{\frac{M-nb_2}{a_2} - k_2 x - 1}{n-1} (-1)^x \Big/ k_2^n$$

Proof: There are n cells to be filled with the elements from G_1 and G_2. The element 0 from G_1 will be $(p-k)$ times as likely as any of the other elements from G_1 or G_2. Let β be the number of elements from G_2 that occur in a given partition of $r_j = M$, and let S_{II} be the sum of these β elements from G_2. If one conditions on the number of elements from G_2, then the following probability statement follows

$$\begin{aligned} P(r_j = M) &= \sum_{\ell=0}^{n} P(r_j = M \mid \beta = \ell)\, P(\beta = \ell) \\ &= P(r_j = M \mid \beta = 0)\, P(\beta = 0) + \sum_{\ell=1}^{n-1} P(r_j = M \mid \beta = \ell)\, P(\beta = \ell) \\ &\quad + P(r_j = M \mid \beta = n)\, P(\beta = n)\ . \end{aligned}$$

Now condition on the sum of the terms from G_2 in the statement $P(r_j = M \mid \beta = \ell)\, P(\beta = \ell)$, i.e.,

$$P(r_j = M \mid \beta = \ell) = \sum_{D \in Z} P(r_j = M \mid \beta = \ell\ ,\ S_{II} = D)\, P(\beta = \ell\ ,\ S_{II} = D)$$

where Z is the set of values which S_{II} can take on when ℓ elements are taken from G_2, i.e., D takes on values between $\ell(a_2 + b_2)$ and $\ell(a_2 k_2 + b_2)$ but not necessarily all the integral values between $\ell(a_2 + b_2)$ and $\ell(a_2 k_2 + b_2)$. Then $P(r_j = M)$ may be written as

$$\begin{aligned} P(r_j = M) &= P(r_j = M \mid \beta = 0) \\ &\quad + \sum_{\ell=1}^{n-1} \left[\sum_{D \in Z} P(r_j = M \mid \beta = \ell\ ,\ S_{II} = D)\, P(\beta = \ell\ ,\ S_{II} = D) \right] \\ &\quad + P(r_j = M \mid \beta = n)\, P(\beta = n) \end{aligned}$$

Now $P(\beta = \ell, S_{II} = D)$ may be replaced by $P(S_{II} = D \mid \beta = \ell) P(\beta = \ell)$. Therefore, $P(r_j = M)$ may be written as follows,

$$P(r_j = M) = P(r_j = M \mid \beta = 0) = (\beta = 0)$$

$$+ \sum_{\ell=1}^{n-1} \left(\sum_{D=\ell(a_2+b_2)}^{\ell(a_2k_2+b_2)} P(r_j = M \mid \beta = \ell, S_{II} = D) P(S_{II} = D \mid \beta = \ell) P(\beta = \ell) \right)$$

$$+ P(r_j = M \mid \beta = n) = P(r_j = M \mid \beta = 0) \left(\frac{p - k_2}{p} \right)^n$$

$$+ \sum_{\ell=1}^{n-1} \left(\sum_{D=\ell(a_2+b_2)}^{\ell(a_2k_2+b_2)} P(r_j = M \mid \beta = \ell, S_{II} = D) P(S_{II} = D \mid \beta = \ell) \binom{n}{\ell} \left(\frac{k_2}{p} \right)^{\ell} \left(\frac{p-k_2}{p} \right)^{n-\ell} \right)$$

$$+ P(r_j = M \mid \beta = n) \left(\frac{k_2}{p} \right)^n .$$

The expression for $P(r_j = M)$ may be simplified by considering the term

$$\sum_{D=\ell(a_2+b_2)}^{\ell(a_2k_2+b_2)} P(r_j = M \mid \beta = \ell, S_{II} = D) P(S_{II} = D \mid \beta = \ell) .$$

Certain of the probabilities $P(S_{II} = D \mid \beta = \ell)$ may be zero. This may be rectified by making the transformation,

$$g = \frac{D - \ell b_2}{a_2} .$$

Then

$$\sum_{D=\ell(a_2+b_2)}^{\ell(a_2k_2+b_2)} P(r_i = M \mid \beta = \ell\,,\ S_{II} = D)\, P(S_{II} = D \mid \beta = \ell)$$

$$= \sum_{g=\ell}^{\ell k_2} P(r_j = M \mid \beta = \ell\,,\ S_{II} = a_2 g + \ell b_2)\, P(S_{II} = a_2 g + \ell b_2 \mid \beta = \ell)$$

where now, S_{II} takes on values for each value of g, so there are no "gaps" in the sum. Now, $P(r_j = M)$ may be written as

$$P(r_j = M) = P(r_j = M \mid \beta = 0)\left(\frac{p-k_2}{p}\right)^n$$

$$+ \sum_{\ell=1}^{n-1}\left(\sum_{g=\ell}^{\ell k_2} P(r_j = M \mid \beta = \ell\,,\ S_{II} = a_2 g + b_2)\, P(S_{II} = a_2 g + \ell b_2 \mid \beta = \ell)\right.$$

$$\left.\cdot\binom{n}{\ell}\left(\frac{k_2}{p}\right)^{\ell}\left(\frac{p-k_2}{p}\right)^{n-\ell}\right) + P(r_j = M \mid \beta = n)\left(\frac{k_2}{p}\right)^n \quad .$$

Now, each of the separate probabilities given in (i), (ii), (iii) and (iv) in the statement of the theorem will be considered.

The probability, $P(r_j = M \mid \beta = 0)$ corresponds to the event that there are n cells to be filled from the set G_1. Lemma 6.3 gives this probability directly as,

$$P(r_j = M \mid \beta = 0) = \sum_{i=0}^{n-1}\binom{n}{i}\left(\frac{p-k_1-k_2}{p-k_2}\right)^i\left(\frac{k_1}{p-k_2}\right)^{n-i}$$

$$\cdot \sum_{x=0}^{\infty} I_{\left[\frac{M-(n-i)b_1}{a_1} - k_1 x - n + i\right]}\binom{n-i}{x}\binom{\frac{M-(n-i)b_1}{a_1} - k_1 x - 1}{n-i-1}(-1)^x \Bigg/ k_1^{n-i}$$

$$+\left(\frac{p-k_1-k_2}{p-k_2}\right) J_M ,$$

where

$$I_y = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases} , \qquad J_M = \begin{cases} 1, & M = 0 \\ 0, & M \neq 0 \end{cases} .$$

The probability, $P(r_j = M \mid \beta = \ell , S_{II} = a_2 g + \ell b_2)$ corresponds to the event that there are $n - \ell$ cells to be filled from G_1 and the sum of the $n - \ell$ cells is equal to $M - a_2 g - \ell b_2$. Lemma 6.3 gives this probability as

$$P(r_j = M \mid \beta = \ell , S_{II} = a_2 g + \ell b_2) = \sum_{i=0}^{n-\ell-1} \binom{n-\ell}{i} \left(\frac{p-k_1-k_2}{p-k_2}\right)^i \left(\frac{k_1}{p-k_2}\right)^{n-\ell-i}$$

$$\cdot \sum_{x=0}^{\infty} I_{\left[\frac{M-a_2 g-\ell b_2-(n-\ell-i)b_1}{a_1} - k_1 x - (n-\ell)+i\right]} \binom{n-\ell-i}{x}$$

$$\cdot \binom{\frac{M-a_2 g-\ell b_2-(n-\ell-i)b_1}{a_1} - k_1 x - 1}{n-\ell-i-1} (-1)^x \Big/ k_1^{n-\ell-i}$$

$$+ \left(\frac{p-k_1-k_2}{p-k_2}\right)^{n-\ell} J_{M-a_2 g-\ell b_2}$$

where I_y is as defined previously, and

$$J_{M-a_2 g-\ell b_2} = \begin{cases} 1, & M - a_2 g - \ell b_2 = 0 \\ 0, & M - a_2 g - \ell b_2 \neq 0 \end{cases} .$$

The probability, $P(S_{II} = a_2 g + \ell b_2 \mid \beta = \ell)$ corresponds to the event of ℓ cells being filled from the set G_2 and having sum $a_2 g + \ell b_2$. Lemma 6.2 gives this probability as

$$P(S_{II} = a_2 g + \ell b_2 \mid \beta = \ell) = \sum_{x=0}^{\infty} I_{\left[\frac{a_2 g + \ell b_2 - \ell b_2}{a_2} - k_2 x - \ell\right]} \binom{\ell}{x} \binom{\frac{a_2 g + \ell b_2 - \ell b_2}{a_2} - k_2 x - 1}{\ell - 1} (-1)^x \Big/ k_2^{\ell}$$

$$= \sum_{x=0}^{\infty} I_{[g - k_2 x - \ell]} \binom{\ell}{x} \binom{g - k_2 x - 1}{\ell - 1} (-1)^x \Big/ k_2^{\ell} .$$

The probability, $P(r_j = M \mid \beta = n)$ corresponds to the event that there are n cells to be filled from G_2 and the sum of the n cells is M. Lemma 6.2 gives this probability as

$$P(r_j = M \mid \beta = n) = \sum_{x=0}^{\infty} I_{\left[\frac{M - nb_2}{a_2} - k_2 x - n\right]} \binom{n}{x} \binom{\frac{M - nb_2}{a_2} - k_2 x - 1}{n - 1} (-1)^x \Big/ k_2^{2} . \qquad \triangle$$

If values are assigned to a_1, b_1, a_2, b_2, k_1 and k_2, then by using Theorems 6.1, 6.2 and 4.2, a table similar to Table IV could be constructed for the Extreme Rank sum test.

As concerns pairwise comparisons and comparisons with a control, the same theory as developed in Chapter V is directly applicable to the case of a partial ordering using linearly weighted scores. The basic results given by Equations (5.1), (5.2) and (5.3) are exactly the same with the exception that $var(r_{ij})$ is now σ_r^2 given in Equation (6.3).

CHAPTER VII

SUMMARY AND EXTENSIONS

The purpose of the research described in this dissertation was to develop statistical tests to deal with situations where a group of objects or populations are partially ordered, i.e., where each of n judges rank the k most preferred of p objects or populations according to some criterion of interest. It is assumed that the judges are capable of ordering the k most preferred objects and that they do so independently of one another. Also, the $p-k$ unranked objects are assigned some appropriate score of c by each judge to indicate their inferiority.

To test the null hypothesis of no difference in preference among the objects versus the general alternative that at least one object was more preferable than at least one other object, the test statistic S, where S is the sum of squares of deviations of the column rank sums for the objects from their mean, was used. The exact distribution of S was found for small values of n and p. Since the tabulation of the distribution of S is rather laborious, a chi-square approximation to it's distribution was found. Numerical results concerning the chi-square approximation are presented. A beta approximation to the ratio of S to its maximum value is also discussed.

An Extreme Rank Sum Test based on r_{max}, the maximum column rank sum, is presented and a table for its use is given. The purpose of the test is to determine if certain objects have slipped to the right of other objects in the scheme. Also, it is inferred that the object slipping the most has been found when r_{max} is significant.

A chapter dealing with multiple comparisons is presented. Finally, a chapter is presented where the ranks given by the judges are replaced by linearly weighted scores. Results given in the case where the ranks are not weighted are extended to the weighted scores case.

Some possible extensions of the results will now be discussed. An investigation of how well the beta distribution approximates the coefficient of concordance is one area that could be investigated. Some study to determine the proper weighting on the ranks in the linearly weighted scores case is a possible area of further investigation. Power studies for the Extreme Rank Sum Test would also be of interest.

The following problem would also be of interest. It would be desireable to choose some subset of the p objects and to be able to state with some probability, P^*, that the "best" object or "best" ℓ objects would be included in this subset.

The following variation of the problem is also reasonable. Suppose it is not possible for each judge to rank the same number of objects. Suppose judge i is capable of ranking only k_i objects. The same topics discussed where each judge ranked an equal number of objects are of interest here also.

A SELECTED BIBLIOGRAPHY

(1) Bernard, A. and Van Elteren, ph. "A Generalization of the Method of m Rankings." Indagationes Mathematicae. Vol. 15 (1953) 358-369.

(2) Doornbos, R. and Prins, H. J., "On Slippage Tests I, II, III." Indagationes Mathematicae. Vol. 20 (1958) 38-46, 47-55, 438-447.

(3) Durbin, J., "Incomplete Blocks in Ranking Experiments." The British Journal of Psychology, Statistical Section. Vol. IV, Part II (1951) 85-90.

(4) Friedman, M., "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." Journal of the American Statistical Association. Vol. 32 (1937) 675-701.

(5) Kendall, M. G. and Babington Smith, B., "The Problem of m Rankings." Annals of Mathematical Statistics. Vol. 10 (1939) 275-287.

(6) Kendall, M. G. Rank Correlation Methods, Hafner Book Co., New York, 2nd Edition, 1955.

(7) Miller, Rupert G. Simultaneous Statistical Inference, McGraw-Hill Book Co., Inc., New York, 1966.

(8) Owen, D. B. Handbook of Statistical Tables, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962.

(9) Rao, C. R. Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., New York, 1965.

(10) Sen, P. K. "Asymptotically Efficient Tests by the Method of n Rankings." Journal of the Royal Statistical Society. Vol. B30 (1968) 312-317.

(11) Thompson, W. A., Jr. and Willke, T. A. "On an Extreme Rank Sum Test for Outliers." Biometrika. Vol. 50 (1963) 375-383.

(12) Tucker, Howard G. A Graduate Course in Probability, Academic Press, New York and London, 1967.

APPENDIX

The purpose of this appendix is to provide a proof of Theorem 3.1 in Section 5 of Chapter III.

Theorem: If

$$f(x) = \frac{a_1x^2 + a_2x + a_3}{b_1x^2 + b_2x + b_3} ,$$

$b_1x^2 + b_2x + b_3$ has no real zeros and $A_1x^2 + A_2x + A_3$ has two real zeros, where $A_1 > 0$ and

$$f'(x) = \frac{A_1x^2 + A_2x + A_3}{(b_1x^2 + b_2x + b_3)^2} ,$$

then $f(x)$ has an absolute maximum at

$$x = \frac{-A_2 - \sqrt{A_2^2 - 4A_1A_3}}{2A_1} .$$

Proof: $f(x)$ is continuous for every x. The first derivative of $f(x)$ is

$$f'(x) = \frac{(2a_1x + a_2)(b_1x^2 + b_2x + b_3) - (a_1x^2 + x_2x + a_3)(2b_1x + b_2)}{(b_1x^2 + b_2x + b_3)^2}$$

$$= \frac{A_1x^2 + A_2x + A_3}{(b_1x^2 + b_2x + b_3)^2} .$$

Let

$$R_1 = \frac{-A_2 - \sqrt{A_2^2 - 4A_1A_3}}{2A_1}$$

and

$$R_2 = \frac{-A_2 + \sqrt{A_2^2 - 4A_1A_3}}{2A_1}$$

Then,

$$f'(x) > 0, \quad \text{for} \quad x \in (-\infty, R_1)$$

$$f'(x) < 0, \quad \text{for} \quad x \in (R_1, R_2)$$

$$f'(x) > 0, \quad \text{for} \quad x \in (R_2, \infty) .$$

Clearly, $f(x)$ has a relative maximum at R_1. Also $f(x)$ approaches $y = \frac{a_1}{b_1}$ as an asymptote as $x \to \infty$ or as $x \to -\infty$.

The fact that $f(x)$ increases to R_1, decreases between R_1 and R_2 and increases past R_2, and approaches $y = a_1 | b_1$ as an asymptote for large positive x and large negative x implies R_1 is the absolute maximum for $f(x)$. Δ

VITA

Larry Joe Stephens

Candidate for the Degree of

Doctor of Philosophy

Thesis: A PARTIAL ORDERING OF p POPULATIONS

Major Field: Statistics

Biographical:

Personal Data: Born in Pillowville, Tennessee, October 26, 1941, the son of Mr. and Mrs. Johnie Stephens.

Education: Graduated from Millington Central High School, Millington, Tennessee, May, 1959; received the Bachelor of Science degree from Memphis State University in May, 1963, with a double major in mathematics and physics; received the Master of Arts from the University of Arizona in May, 1965, with a major in mathematics; attended Yale University in the summer of 1965; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1972.

Professional Experience: Graduate Teaching Assistant, University of Arizona, 1963-1965; Research Assistant, Lawrence Radiation Laboratory, Livermore, California, Summer 1964; Instructor at Christian Brothers College, Memphis, Tennessee, 1965-1966; Instructor at Gonzaga University, Spokane, Washington, 1966-1967; Graduate Teaching Assistant, Oklahoma State University, 1967-1972; Aero-Space Technologist, NASA, Huntsville, Alabama, Summer 1968; Research Assistant, Los Alamos Scientific Laboratory, Los Alamos, New Mexico, Summers of 1969 and 1970.