ANALYSIS OF PAIRED COMPARISON DESIGNS

WITH INCOMPLETE REPETITIONS

by

John W. Wilkinson
University of North Carolina

This research was jointly supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 18(600)-458, and the National Research Council of Canada

Institute of Statistics
Mimeograph Series No. 147
April, 1956

A C K N O W L E D G M E N T S

I am deeply indebted to Professor R. C. Bose at whose suggestion this research was initiated. His continued encouragement and advice have contributed substantially to its completion, and I welcome this opportunity to indicate my gratitude.

I wish to acknowledge my appreciation for the financial assistance received from the United States Air Force, the National Research Council of Canada, and the Institute of Statistics of the University of North Carolina.

In addition, I wish to thank my wife, Jean, whose patience, understanding, and assistance considerably aided my completion of this research. Also, I am sincerely appreciative of the excellent cooperation received from Miss Martha Jordan, Mrs. Mary Ann Taylor, and Mrs. Hilda Kattsoff in their careful typing of the manuscript.

TABLE OF CONTENTS

INTRODUCTION

When quantitative measurement of treatment effects is not possible, or is possible but is not practicable, a method of paired comparison of treatments is frequently used. Experiments involving paired comparisons have mainly concerned the situation where each observer compares all possible pairs of treatments or objects. In certain types of experiments, this may require an excessive number of comparisons to be made by any observer. For this reason, experimental designs, symmetrical with respect to both observers and to objects, which do not require each observer to make all possible comparisons will be useful.

Historically, various procedures for reducing the number of comparisons required by each observer have been proposed (these procedures usually being somewhat dependent on advance knowledge of the treatments) but apparently the first research on the construction of such designs was carried out by Kendall [9][1]. His procedure was to assign pairs of objects to observers in accordance with tours around a "preference" polygon whose t vertices depict the t objects considered in the experiment. Later, Bose [1] constructed three series of paired comparison designs by making certain correspondences with known balanced incomplete block designs. One of these series produced the same series of designs as the tours around the preference polygon considered by Kendall [9]. However, neither Bose nor Kendall proposed any procedures

[1]The numbers in square brackets refer to the bibliography.

of analysis in their respective papers.

The main purpose of this study is to consider the problem of analysis of the paired comparison designs developed by Bose and Kendall. In addition, further designs are obtained by constructing the complementary designs corresponding to the designs given by Bose [1].

The analysis of experiments involving paired comparisons, for the situation where each observer compares all possible pairs of treatments, has been considered by several authors. Thurstone [16] proposed a method of paired comparisons based on a model which is essentially dependent on the assumption that judgments concerning characteristics of an object are normally distributed. Mosteller [13] has summarized the notions underlying Thurstone's method and has extended its scope.

M. G. Kendall and Babington Smith [12], and Kendall [10], [11], in their consideration of a method of paired comparisons, introduced a coefficient of agreement and a coefficient of consistence, and proposed tests of certain null hypotheses based on these statistics.

Bradley and Terry [2] considered a method of paired comparisons based on a mathematical model which is a generalization of the binomial model. Using this model, they obtained maximum-likelihood estimates of the true object ratings and likelihood ratio tests for certain classes of hypotheses. They derived the distributions of certain of their proposed test statistics and provided tables (later extended by Bradley [3]) of these distributions when the number of objects is small. Additional description of the method is given by Terry, Bradley, and Davis [15], and Bradley [5] has discussed the appropriateness of this model for paired comparisons. In [4], Bradley investigated the reliability of the estimators of the true object ratings and the large-sample evaluation of the power functions of his original tests, comparing them with the analysis of variance and with a multi-binomial procedure.

In this study, an analysis of the paired comparison designs developed by Bose and Kendall is proposed. Using the Bradley-Terry model, likelihood ratio tests of several classes of hypotheses are proposed, and in certain situations it is shown that significance levels for these tests may be obtained from available tables. The construction and description of these tests are displayed in Chapter II, while, in Chapter III, large-sample properties of the tests are discussed and some examples are presented to clarify the test procedures. In addition, in Chapter IV, an analogue to Kendall's coefficient of agreement is given for the case of incomplete repetitions and under certain assumptions is shown to provide an alternative test for a specific class of hypotheses. Chapter I is devoted to describing the paired comparison designs and discussing the procedure of constructing the complementary designs which are listed in the Appendix.

CHAPTER I

DESCRIPTION OF THE PAIRED COMPARISON DESIGNS

1.1. Introduction.

Since a large portion of this investigation is devoted to constructing and describing tests usable in connection with certain paired comparison designs given by Bose [1], it is deemed advisable to present a clear description of these designs. The designs have a high degree of symmetry as they are balanced by numbers of comparisons, objects compared, and numbers of observers on given comparisons. Since it is this symmetry that enables much of the ensuing mathematical manipulation, the following detailed description is quite pertinent.

1.2. Definition of paired comparison designs.

Suppose we have t objects which we desire to compare according to a certain characteristic, and we have v judges available to perform the comparisons. The procedure of comparison will be as follows:

(i) Each judge compares r pairs of objects $(1 < r \leq t(t-1)/2)$, and, for each pair compared, expresses his preference for one or the other object of the pair. (It may be desirable to allow the judge to express no preference with respect to either of the objects forming the pair. However, this possibility will not be entertained for the main part of our discussion.)

(ii) The pairs compared by any judge are all different.

(iii) Among the r pairs compared by each judge, each object appears equally often, say α times.

(iv) Each pair is compared by k judges, $(1 < k \leq v)$.

(v) Given any two judges, there are exactly λ pairs which are compared by both judges.

Designs satisfying the above conditions are referred to by Bose as linked paired comparison designs.

1.3. Relationships among the parameters.

Using the usual design of experiments terminology, suppose we consider each judge to correspond to a treatment and each pair of objects to a block. Then if a pair is compared by a judge, the block corresponding to the pair may be considered to contain the treatment corresponding to the judge. Therefore, if a paired comparison design as defined in Section 1.2 exists, there must exist a corresponding balanced incomplete block design with v treatments, $b = t(t-1)/2$ blocks, such that each block contains k treatments, each treatment occurs in r blocks, and two given treatments occur together in λ blocks. However, the existence of the balanced incomplete block design does not in general ensure the existence of the corresponding linked paired comparison design.

From the relationship among the parameters of balanced incomplete block designs, and from the definition of the paired comparison designs given in Section 1.2, we can obtain the following relationships among the parameters of a paired comparison design:

(1.3.1) $$r = t\alpha/2,$$

(1.3.2) $$b = t(t-1)/2,$$

(1.3.3) $$bk = vr, \quad \lambda(v-1) = r(k-1),$$

(1.3.4) $$v\alpha = k(t-1),$$

(1.3.5) $$b \geq v \quad \text{or} \quad r \geq k,$$

(1.3.6) $$t\alpha \geq 2k,$$

(1.3.7) $$r \geq \lambda,$$

(1.3.8) $$r \geq \alpha(\alpha+1)/2, \quad \lambda \leq \alpha(\alpha+1)/2.$$

1.4. Three series of paired comparison designs.

Series 1.

Bose obtains one group of designs by considering the case $\alpha = 2$. When this is the case, the parameters of the paired comparison design can be written in terms of t -- the number of objects to be compared -- and, in fact, these parameters are

$$(1.4.1) \quad v = \frac{(t-1)(t-2)}{2},\ b = \frac{t(t-1)}{2},\ r = t,\ k = t-2,\ \lambda = 2,\ \alpha = 2.$$

The existence of (1.4.1) implies the existence of the balanced incomplete block design with parameters

$$(1.4.2) \quad v = \frac{(t-1)(t-2)}{2},\ b = \frac{t(t-1)}{2},\ r = t,\ k = t-2,\ \lambda = 2,$$

which can be derived by writing down a solution of the symmetrical balanced incomplete block design

$$(1.4.3) \quad v = b = \frac{t(t-1)}{2} + 1,\ \ r = k = t,\ \ \lambda = 2,$$

and then deleting one block and all the treatments in this block. Conversely, Bose has shown that we can obtain a solution for the paired comparison design (1.4.1) for any value of t for which a solution of the symmetrical balanced incomplete block design (1.4.3) is known. Thus, four designs with the following parameters have been obtained:

$$(1.4.4) \quad \begin{array}{llllllll} \text{(i)} & t = 4, & v = 3, & b = 6, & r = 4, & k = 2, & \lambda = 2, & \alpha = 2; \\ \text{(ii)} & t = 5, & v = 6, & b = 10, & r = 5, & k = 3, & \lambda = 2, & \alpha = 2; \\ \text{(iii)} & t = 6, & v = 10, & b = 15, & r = 6, & k = 4, & \lambda = 2, & \alpha = 2; \\ \text{(iv)} & t = 9, & v = 28, & b = 36, & r = 9, & k = 7, & \lambda = 2, & \alpha = 2. \end{array}$$

The designs for these have been listed in Table 1 of [1]. For facilitating reference at a later stage, designs (i) and (ii) of (1.4.4) will be listed here. If we denote the objects to be compared by the natural numbers 1, 2, ..., t, and denote the judges by J_1, J_2, ..., J_v, then we can present the designs as in Table 1.4.1.

TABLE 1.4.1

DESIGN NUMBER	PARAMETERS	DESIGN	
		JUDGE	PAIRS ASSIGNED TO A JUDGE
(i), (1.4.4)	$t = 4$, $v = 3$	J_1	(1,4), (1,3), (2,4), (2,3)
	$b = 6$, $r = 4$	J_2	(1,3), (2,4), (1,2), (3,4)
	$k = 2$, $\lambda = 2$	J_3	(1,4), (1,2), (2,3), (3,4)
	$\alpha = 2$		
(ii), (1.4.4)	$t = 5$, $v = 6$	J_1	(3,5), (2,4), (1,3), (1,4), (2,5)
	$b = 10$, $r = 5$	J_2	(2,3), (3,4), (1,4), (1,5), (2,5)
	$k = 3$, $\lambda = 2$	J_3	(2,3), (3,5), (1,2), (4,5), (1,4)
	$\alpha = 2$	J_4	(3,5), (1,2), (3,4), (2,4), (1,5)
		J_5	(1,2), (3,4), (4,5), (1,3), (2,5)
		J_6	(2,3), (4,5), (2,4), (1,3), (1,5)

Series 2.

Bose obtains another series of paired comparison designs for the situation where the number of objects t is even, say $t = 2z$. When this is the case, the $z(2z - 1)$ pairs can be divided into $2z - 1$ sets of z pairs each, such that each object occurs exactly once among the pairs of a set. Then, considering a balanced incomplete block design with v^* treatments, $b^* = 2z - 1$ blocks, r^* replications, block size k^*, and in which every pair of treatments occurs together in the same block λ^* times, where each block corresponds to one set and each treatment to one judge, a paired comparison design can be obtained by assigning to each judge the sets of pairs corresponding to all blocks in which the treatment corresponding to the judge occurs. This procedure yields paired cmmparison designs with parameters

$t = 2z,\ v = v^*,\ b = z(2z - 1),\ r = zr^*,\ k = k^*,\ \lambda = z\lambda^*,\ \alpha = r^*.$

Designs obtained in this manner and with the following parameters are listed in Table 2 of [1]:

$$
(1.4.5)\quad
\begin{array}{lllllllll}
\text{(i)} & t = 4, & v = 3, & b = 6, & r = 4, & k = 2, & \lambda = 2, & \alpha = 2; \\
\text{(ii)} & t = 6, & v = 5, & b = 15, & r = 12, & k = 4, & \lambda = 9, & \alpha = 4; \\
\text{(iii)} & t = 8, & v = 7, & b = 28, & r = 12, & k = 3, & \lambda = 4, & \alpha = 3; \\
\text{(iv)} & t = 8, & v = 7, & b = 28, & r = 16, & k = 4, & \lambda = 8, & \alpha = 4.
\end{array}
$$

It is of interest to note that design (i) of (1.4.5) is the same as design (i) of (1.4.4).

Series 3.

In a somewhat similar manner, another series of paired comparison designs is obtained with parameters

$$t = 2z + 1,\quad v = v',\ b = z(2z + 1),\ r = (2z + 1)r',\ k = k',$$
$$\lambda = (2z + 1)\lambda',\ \alpha = 2r',$$

where v', $b' = z$, r', k', λ' are the parameters of a known balanced incomplete block design. The design with parameters

$$t = 7,\quad v = 3,\quad b = 21,\quad r = 14,\quad k = 2,\quad \lambda = 7,\quad \alpha = 4,$$

is given in Table 2 of [1]. It should be noted that the paired comparison designs obtained in this manner are the same as those considered by Kendall [9].

1.5. Some further paired comparison designs.

For experimentation using paired comparison methods, the usefulness of these designs would increase with increasing number of objects t; hence, some merit would seem to be attached to the listing of further designs for larger values of t. Twenty-seven additional designs, with

$4 \leq t \leq 25$, are listed in Appendix A. To determine the pairs of objects, or sets of pairs of objects, to be assigned to each judge, field plans for corresponding balanced incomplete block designs displayed in Cochran and Cox [6], and in Fisher and Yates [8] were utilized.

Several of these new designs were obtained in the following manner: Consider a known paired comparison design D with parameters t, v, b, r, k, λ, α. Then from this design we can construct a complementary design $\bar{D}$ by taking as the pairs to be compared by judge u, $(u = 1, \ldots, v)$, the $b - r = t(t - 1)/2 - r$ pairs not compared by judge u in the original design D. If we consider D and $\bar{D}$ together, then each judge would compare b pairs and each pair would be compared by v judges; hence, in $\bar{D}$ each pair will be compared by $v - k$ judges. Since in the $b = t(t - 1)/2$ pairs, each object appears $(t - 1)$ times, then in the pairs compared by each judge in $\bar{D}$ each object appears $t - 1 - \alpha$ times. In D, any two judges compare exactly $r - \lambda$ different pairs and λ common pairs, and hence $b - 2(r - \lambda) - \lambda = b - 2r + \lambda$ pairs are not compared by either judge. Therefore, in $\bar{D}$ any two judges will compare exactly $b - 2r + \lambda$ common pairs. Thus the parameters of the complementary design $\bar{D}$ in terms of the original design parameters, are

$$t_1 = t,\ v_1 = v,\ b_1 = b,\ r_1 = b-r,\ k_1 = v - k,$$
$$\lambda_1 = b - 2r + \lambda,\ \alpha_1 = t - 1 - \alpha.$$

CHAPTER II

SOME TEST PROCEDURES FOR EXAMINING THE RELATIVE RATING OF t OBJECTS

2.1. Introduction.

In this chapter several tests are developed for examining the relative rating of objects when a method of paired comparisons is employed. The experimental procedure used is that prescribed by the paired comparison designs obtained by Bose [1]. The tests are constructed using the Bradley-Terry model [2] and the method of maximum likelihood.

2.2. Experimental procedure and notation.

Consider a set of t objects $(a_1, a_2, \ldots, a_t)$. We wish to examine these objects with respect to a prescribed characteristic x, and we desire to express opinions concerning the relevant status of the objects with respect to each other.

Suppose we have available v judges, and we require each of these judges to compare, according to a certain characteristic x, r pairs of the $t(t-1)/2$ possible pairs of the set of objects $(a_1, a_2, \ldots, a_t)$, $(1 < r \leq t(t-1)/2)$. The r pairs to be compared by each judge will be specified by the field plan of the appropriate paired comparison design employed. Presently available designs are outlined in Chapter I and are given in detail in [1] and in Appendix A. We shall refer to one set of r pairs compared by a specific judge as an incomplete repetition or a complete repetition, depending on whether or not $r < t(t-1)/2$ or $r = t(t-1)/2$. When $r = t(t-1)/2$, the situation becomes that which is considered in detail by Bradley and Terry [2], Bradley [3, 4, 5], and Terry, Bradley, and Davis [15]. For this reason, we will strive to employ similar notation to that used by the above authors.

For each pair of objects compared by judge u, $(u = 1, \ldots, v)$, the judge expresses a preference for one object or the other, assigning the rank one to the object judged superior according to the characteristic x, and the rank two to the object judged inferior.

Let r_{iju} denote the rank assigned by judge u to the i-th object when it is compared with the j-th object. Let $\{a_i \longrightarrow a_j \mid u\}$ denote the preference indicated by judge u for object a_i over object a_j, $(i \neq j,\ i, j = 1, \ldots, t;\ u = 1, \ldots, v)$. Then

(2.2.1) $$r_{iju} = \begin{cases} 1, & \text{if } \{a_i \longrightarrow a_j \mid u\}, \\ 2, & \text{if } \{a_i \longleftarrow a_j \mid u\}, \end{cases}$$

if objects a_i and a_j are compared by judge u. From (2.2.1) it is readily observed that

(2.2.2) $$r_{iju} + r_{jiu} = 3, \qquad (i \neq j,\ i, j = 1, \ldots, t;\ u = 1, \ldots, v),$$

when objects a_i and a_j are compared by judge u. To handle the complication created by the fact that each judge does not compare all possible pairs of objects, we shall define

(2.2.3) $$n_{iju} = \begin{cases} 1, & \text{if } a_i \text{ and } a_j \text{ are compared by judge } u, \\ 0, & \text{if } a_i \text{ and } a_j \text{ are not compared by judge } u, \end{cases}$$

$(i \neq j,\ i, j = 1, \ldots, t;\ u = 1, \ldots, v)$, and we shall conventionally take

$$n_{iiu} = 0, \qquad (i = 1, \ldots, t;\ u = 1, \ldots, v).$$

Hence, for each judge u we can define an <u>incidence matrix</u> $N_u = (n_{iju})$ which is symmetric, has 1's or 0's for elements, and is specified by the field plan of the paired comparison design used. For example, for design (i) of (1.4.4) we have the following three incidence matrices:

$$N_1 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \qquad N_2 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix},$$

(2.2.4)

$$N_3 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

From the highly symmetrical nature of the original paired comparison designs, as indicated in Section 1.3, these incidence matrices possess the following interesting properties:

$$\sum_{i=1}^{t} n_{iju} = \sum_{j=1}^{t} n_{iju} = \alpha,$$

(2.2.5)
$$\sum_{i=1}^{t} \sum_{j=1}^{t} n_{iju} = t\alpha = 2r, \qquad (u = 1, \ldots, v),$$

$$\sum_{i=1}^{t} \sum_{j=1}^{t} \sum_{u=1}^{v} n_{iju} = 2rv = kt(t - 1).$$

If, in addition, we conventionally define

$$N_0 = kI,$$

where I is the (t x t) identity matrix, then

(2.2.6)
$$\sum_{u=0}^{v} N_u = kE,$$

where E is a (t x t) matrix, each element of which is unity.

To simplify notation, we will introduce the following conventions: $\sum_i$, $\sum_j$, $\sum_u$, $\sum_m$ and $\prod_i$, $\prod_j$, $\prod_u$, $\prod_m$ will indicate, respectively, single sums and products with respect to the depicted quantity over its range, where $i = 1, \ldots, t$; $j = 1, \ldots, t$; $u = 1, \ldots, v$; $m = 1, \ldots, g$. $\sum_j'$ will denote $\sum_{\substack{j=1 \\ j \neq i}}^{t}$ for some i, where this i will be clear from the context.

$\sum_{i<j}$ and $\prod_{i<j}$ will indicate, respectively, double sums and products, $i = 1, \ldots, j-1$; $j = 2, \ldots, t$. $\sum_{i \neq j}$ and $\prod_{i \neq j}$ will indicate, respectively, $\sum_{i=1}^{t} \sum_{\substack{j=1 \\ i \neq j}}^{t}$ and $\prod_{i=1}^{t} \prod_{\substack{j=1 \\ i \neq j}}^{t}$. Any departures from these conventions will be specified when the departure is incurred. In addition, log and ℓn will be used to denote common and natural logarithms, respectively.

2.3. Mathematical model (Bradley-Terry model).

Suppose there exist numbers $\pi_{1u}, \ldots, \pi_{tu}$ corresponding to judge u, $(u = 1, \ldots, v)$, where

$$\text{(2.3.1)} \qquad \begin{aligned} &\pi_{iu} \geq 0, && (i = 1, \ldots, t; \; u = 1, \ldots, v), \\ &\sum_i \pi_{iu} = 1, && (u = 1, \ldots, v), \end{aligned}$$

and such that

$$\text{(2.3.2)} \qquad \begin{aligned} P\left\{a_i \longrightarrow a_j \mid u\right\} &= \frac{\pi_{iu}}{\pi_{iu} + \pi_{ju}}, \\ P\left\{a_i \longleftarrow a_j \mid u\right\} &= \frac{\pi_{ju}}{\pi_{iu} + \pi_{ju}}, \end{aligned}$$

$(i \neq j, \; i, j = 1, \ldots, t; \; u = 1, \ldots, v)$.

The conditions in (2.3.1) are imposed partially for convenience and partially to ensure determinacy in certain systems of equations which will be encountered. These numbers $\pi_{1u}, \ldots, \pi_{tu}$, will be considered as true ratings (or preferences) of the t objects $a_1, \ldots, a_t$ corresponding to the judge u. Estimates of $\pi_{1u}, \ldots, \pi_{tu}$ will be denoted by $p_{1u}, \ldots, p_{tu}$, respectively.

2.4. The likelihood function.

From (2.3.2), the probability of judge u assigning rank r_{iju} to object a_i and rank $r_{jiu} = 3 - r_{iju}$ to object a_j upon comparison of the pair of objects is

$$(2.4.1) \quad n_{iju} \left(\frac{\pi_{iu}}{\pi_{iu} + \pi_{ju}} \right)^{2-r_{iju}} \left(\frac{\pi_{ju}}{\pi_{iu} + \pi_{ju}} \right)^{2-r_{jiu}}$$

$$= n_{iju} \frac{\pi_{iu}^{2-r_{iju}} \pi_{ju}^{2-r_{jiu}}}{\pi_{iu} + \pi_{ju}} \quad .$$

Provided that the pair of objects is compared by judge u, we observe that if $\{a_i \longrightarrow a_j | u\}$, then $r_{iju} = 1$, $r_{jiu} = 2$ and (2.4.1) becomes

$$n_{iju} \frac{\pi_{iu}}{\pi_{iu} + \pi_{ju}} = P \left\{ a_i \longrightarrow a_j | u \right\} \quad ,$$

and if $\{ a_i \longleftarrow a_j | u \}$, then $r_{iju} = 2$, $r_{jiu} = 1$ and (2.4.1) becomes

$$n_{iju} \frac{\pi_{ju}}{\pi_{iu} + \pi_{ju}} = P \left\{ a_i \longleftarrow a_j | u \right\} \quad .$$

If we assume probability independence between pairs of objects when compared by judge u, we obtain the likelihood function for the set of comparisons made by judge u,

$$L_u = \prod_R n_{iju} \frac{\pi_{iu}^{2-r_{iju}} \pi_{ju}^{2-r_{jiu}}}{\pi_{iu} + \pi_{ju}}$$

(2.4.2)
$$= \frac{\prod_i \pi_{iu}^{2\alpha - \sum_j n_{iju} r_{iju}}}{\prod_R n_{iju}(\pi_{iu} + \pi_{ju})}, \quad (u = 1, \ldots, v),$$

where R denotes the set of numbers $R = \{ i, j: i < j, i = 1, \ldots, j-1; j = 2, \ldots, t; n_{iju} = 1 \}$. If we assume probability independence between experiments performed by different judges, then from (2.4.2) we obtain the likelihood function for the complete set of comparisons made by the v judges,

(2.4.3)
$$L_{(1)} = \prod_u L_u,$$

and hence if we set

$$a_i^* = 2 \alpha v - \sum_j \sum_u n_{iju} r_{iju},$$

we have

(2.4.4)
$$\ln L_{(1)} = \sum_i a_i^* \ln \pi_{iu} - \sum_{i < j} \sum_u n_{iju} \ln(\pi_{iu} + \pi_{ju}).$$

2.5. Test I.

In this and succeeding sections, we will be examining hypotheses concerning the true ratings of the t objects, $a_1, \ldots, a_t$, under various assumptions. These assumptions will usually be depicted and taken care of by the alternative hypothesis considered. To start, we shall consider the situation where each of the v judges compares only one set of r pairs, the set being specified by the paired comparison design used. We desire information about the differences, if any, among the true ratings of the t objects. In this section, we are willing to make the assumption that the v judges are consistent as a group; that is, we desire to test the null hypothesis,

$$H_0: \ \pi_{iu} = 1/t,$$

against the alternative hypothesis,

(2.5.1)

$$H_1: \ \pi_{iu} = \pi_i, \qquad (i = 1, \ldots, t; \ u = 1, \ldots, v),$$

where the π_i's are not all equal.

When the alternative hypothesis H_1 is true, (2.4.4) becomes

$$\left\{ \ln L_1 \,|\, H_1 \right\} = \sum_i a_i^* \ln \pi_i - k \sum_{i < j} \ln(\pi_i + \pi_j), \tag{2.5.2}$$

where the simplification of the latter term is aided by the properties of the incidence matrices listed in (2.2.5) and (2.2.6).

We note that a_i^*, defined in (2.4.4), can be rewritten, using (1.3.4), as

$$a_i^* = 2k(t - 1) - \sum_j \sum_u n_{iju} \, r_{iju}.$$

When $r = t(t - 1)/2$, then $n_{iju} = 1$ for all i, j, u, $(i \neq j)$, k = v, and

$$a_i^* = 2v(t - 1) - \sum_j{}' \sum_u r_{iju},$$

and corresponds to a_i of [2] with n replaced by v.

Now maximizing (2.5.2), subject to the constraint $\sum_i \pi_i = 1$, yields the set of equations

(2.5.3)

$$\frac{a_i^*}{\pi_i} - k \sum_j{}' (\pi_i + \pi_j)^{-1} + \mu = 0, \qquad (i = 1, \ldots, t),$$

$$\sum_i \pi_i = 1.$$

Summing (2.5.3) with respect to i yields

$$(2\alpha t - 4r)v + \mu = 0,$$

which implies $\mu = 0$, since from (1.3.1), $r = t\alpha/2$. Hence the maximum-likelihood estimates $p_1, \ldots, p_t$ of $\pi_1, \ldots, \pi_t$ will be obtained from system of equations

$$\text{(2.5.4)} \qquad \begin{aligned} &\frac{a_i^*}{p_i} - k \sum_j{}' (p_i + p_j)^{-1} = 0, \qquad (i = 1, \ldots, t), \\ &\sum_i p_i = 1. \end{aligned}$$

When the null hypothesis H_0 is true, (2.4.4) becomes

$$\text{(2.5.5)} \qquad \begin{aligned} \{ \ell n L_1 | H_0 \} &= (2\alpha vt - 3rv) \ell n \frac{1}{t} - \frac{kt(t-1)}{2} \ell n \frac{2}{t} \\ &= - \frac{kt(t-1)}{2} \ell n\, 2 \\ &= - vr\, \ell n\, 2, \end{aligned}$$

using (2.2.2), (2.2.5), and (2.2.6).

Thus the likelihood ratio test of the null hypothesis H_0 against the alternative hypothesis H_1, as specified in (2.5.1), will be in terms of the likelihood ratio λ_1, where, if we define

$$\text{(2.5.6)} \qquad B^{(1)} = k \sum_{i<j} \log(p_i + p_j) - \sum_i a_i^* \log p_i,$$

where $p_1, \ldots, p_t$ are solutions of (2.5.4), then

$$\text{(2.5.7)} \qquad \ell n \lambda_1 = - \left\{ vr\, \ell n 2 - B^{(1)}\, \ell n\, 10 \right\} .$$

When $r = t(t-1)/2$, then $k = v$, and $B^{(1)}$ corresponds with B_1 of [2]. It will also be convenient at this stage to define the statistic

(2.5.8) $$T^{(1)} = -2 \ln \lambda_1 = 2\, vr \ln 2 - 2\, B^{(1)} \ln 10 .$$

For the likelihood ratio in (2.5.7) to be useful to test the hypothesis stated in (2.5.1), it would be desirable to have some knowledge of the distribution of $B^{(1)}$ -- at least when the null hypothesis H_0 is true.

It is possible to generate all combinations of the object sums of ranks. Then under the null hypothesis of equality of true object ratings for each judge, the probability of each combination of rank sums is attainable. This information would be sufficient to obtain the distribution of $B^{(1)}$ under H_0.

However, a direct computation of the probabilities of these various combinations of rank sums would be extremely tedious, even for small values of the design parameters. The fact that every permutation of the rank sums corresponding to each judge is not possible, since each judge does not compare all possible $t(t - 1)/2$ pairs, is the cause of the complication. However, we can utilize the symmetry of the designs to circumvent this particular difficulty.

Let A_i, $(i = 1, \ldots, t)$, denote the set of $v\alpha = k(t - 1)$ elements which are the $k(t - 1)$ ranks assigned by all the v judges to the object a_i when compared with the other $(t - 1)$ objects. That there are $v\alpha = k(t - 1)$ such ranks follows from the fact that each of the v judges compares a_i with exactly α of the remaining $(t - 1)$ objects. Also, from the design properties depicted by (1.3.1) - (1.3.4), (2.2.5) and (2.2.6), it follows that these $k(t - 1)$ ranks are those for a_i after comparison with each of the remaining $(t - 1)$ objects exactly k times. Hence, we can subdivide A_i into k disjoint subsets $A_{iu'}$, with each $A_{iu'}$

containing (t - 1) elements which are the ranks for a_i after comparison with each of the remaining (t - 1) objects exactly once. Thus if we denote the (t - 1) elements of $A_{iu'}$ by $r_{iju'}$, $(i \neq j;\ j = 1, \ldots, t)$, then the sum of elements of $A_{iu'}$ is

(2.5.9) $$S_{iu'} = \sum_j{}' r_{iju'},$$

and the sum of the elements of A_i is

(2.5.10) $$\sum_{u'=1}^{k} \sum_j{}' r_{iju'} = \sum_u \sum_j n_{iju} r_{iju}.$$

Hence the set of rank sums

$$\left(\sum_u \sum_j n_{1ju} r_{1ju}, \ldots, \sum_u \sum_j n_{tju} r_{tju}\right)$$

(2.5.11) $$= \left(\sum_{u'=1}^{k} \sum_j{}' r_{1ju'}, \ldots, \sum_{u'=1}^{k} \sum_j{}' r_{tju'}\right)$$

$$= \sum_{u'=1}^{k} (S_{1u'}, \ldots, S_{tu'}),$$

could be considered as being the set of rank sums obtained from k complete repetitions of all t(t - 1)/2 pairs. It is these rank sums which determine $B^{(1)}$, and hence its distribution can be determined from the distribution of the different combinations of the rank sums. The possibility of putting the rank sums in the form of (2.5.11) enables us to obtain the distribution of $B^{(1)}$, under the null hypothesis H_0 of equality of true object ratings, in a manner described by Bradley and Terry [2]. This procedure will be described briefly below, but for more detail the reference given should be consulted.

Consider the design (i) of (1.4.4) with parameters $t = 4$, $v = 3$, $b = 6$, $r = 4$, $k = 2$, $\lambda = 2$, $\alpha = 2$. The possible sets of revised rank sums, $(S_{1u'}, \ldots, S_{tu'})$, of (2.5.11) are

$$(2.5.12)\qquad \begin{array}{llll} \text{(i)} & (3, 4, 5, 6), & \text{(ii)} & (3, 5, 5, 5), \\ \text{(iii)} & (4, 4, 4, 6), & \text{(iv)} & (4, 4, 5, 5). \end{array}$$

Each of the twenty-four permutations of the elements of (i) has probability 1/64, each of the four permutations of the elements of (ii) and the four permutations of the elements of (iii) has probability 2/64, while each of the six permutations of the elements of (iv) has probability 4/64. The possible rank sums in (2.5.11) for $k = 2$ are obtained by adding the sets of elements of (i), (ii), (iii), (iv), in turn, to the corresponding elements of the sets of all possible permutations of the elements of these sets. The probability of a given permutation is obtained by multiplying the probability of the combination involved from (2.5.12) and the permutation used to produce the given permutation. To obtain the probability of a given new combination of rank sums, the probabilities for each permutation of the elements of the combination are added. By successive repetition of this process a desired number of times, the combinations of rank sums (2.5.11) and their corresponding probabilities under H_0 can be determined for any given k and t. Of course, for large k and t the procedure would be extremely laborious.

By this process, the combinations of rank sums, their corresponding probabilities under H_0, their corresponding values of $B^{(1)}$ and its probabilities under H_0, for $t = 4$, $k = 2$, have been determined and are tabulated in (2.5.13).

(2.5.13)

$\sum_{u'=1}^{2} (S_{1u'}, \ldots, S_{4u'})$	$P\left\{\sum_{u'=1}^{2} (S_{1u'}, \ldots, S_{4u'}) \mid H_0\right\}$	$B^{(1)}$	$P\left\{B^{(1)} \mid H_0\right\}$
6, 8, 10, 12	24/4096	0	24/4096
6, 8, 11, 11	24/4096		
6, 9, 9, 12	24/4096	0.602	72/4096
7, 7, 10, 12	24/4096		
7, 7, 11, 11	24/4096	1.204	24/4096
6, 9, 10, 11	144/4096	1.498	288/4096
7, 8, 9, 12	144/4096		
6, 10, 10, 10	40/4096	1.806	80/4096
8, 8, 8, 12	40/4096		
7, 8, 10, 11	432/4096	2.359	432/4096
7, 9, 9, 11	336/4096	2.631	336/4096
7, 9, 10, 10	528/4096	2.898	1056/4096
8, 8, 9, 11	528/4096		
8, 8, 10, 10	408/4096	3.158	408/4096
8, 9, 9, 10	1224/4096	3.389	1224/4096
9, 9, 9, 9	152/4096	3.612	152/4096

From the previous discussion, it then becomes clear that the tables of [2] and [3] may be used to provide us with the distribution of $B^{(1)}$ under the null hypothesis H_0. The tables are available for design parameters t and k in the following range: t = 3, k = 1, ..., 10; t = 4, k = 1, ..., 8; t = 5, k = 1, ..., 5. These tables also list the estimates, p_1, ..., p_t, corresponding to the rank sum combination (2.5.11).

Thus a test procedure for (2.5.1), for t and k in the indicated range, is as follows: Determine the rank sums (2.5.11). Then from tables in [2] and [3] obtain the corresponding values for p_1, ..., p_t, $B^{(1)}$, and the probability P that $B^{(1)}$ will not be exceeded if the null hypothesis H_0 is true.

If either t or k is outside the range for the tables, the above test procedure cannot be used. However, if only k is outside the indicated range, it is possible to use the available tables to obtain the estimates p_1, ..., p_t, or at least a good first approximation to them, depending on whether or not there exists an integer c which divides the rank sums and k evenly and which is such that k/c is within the above indicated range for k. This technique is discussed in detail in [3] and [15], and the example given in Section 3.7, involving its use, will describe it further.

If $t > 5$, the available tables will not be of assistance even in obtaining approximations to p_1, ..., p_t except in special cases. Hence, to obtain estimates, p_1, ..., p_t, in this case, it will be necessary to solve equations (2.5.4). Some methods aiding in the solution of equations of this form are suggested by Dykstra [7], and Bradley and Terry [2].

Once p_1, ..., p_t have been obtained, $B^{(1)}$ can be evaluated from (2.5.6), but the significance level of $B^{(1)}$ can only be approximately determined. A discussion of this will be given in Chapter III.

2.6. Test II.

For the situation and assumptions used in Section 2.5, we now desire to test the null hypothesis H_0 against the alternative that the

π_i's under H_1 are split into two groups, the elements within each group being equal. That is, we desire to test the null hypothesis,

$$H_0: \quad \pi_{iu} = 1/t \ , \quad (i = 1,\ldots,t;\ u = 1,\ldots,v),$$

against the alternative hypothesis,

$$\text{(2.6.1)} \qquad H_2: \quad \pi_{iu} = \begin{cases} \pi, & (i = 1,\ldots,s) \\ \dfrac{1-s\pi}{t-s}, & (i = s+1,\ldots,t),(u=1,\ldots,v). \end{cases}$$

If we let p denote an estimate of π, then the equations(2.5.4) become

$$\text{(2.6.2a)} \qquad \frac{a_i^*}{p} - k\left\{ \frac{s-1}{2p} + \frac{(t-s)^2}{1+p(t-2s)} \right\} = 0, \qquad (i = 1,\ldots,s),$$

$$\text{(2.6.2b)} \qquad \frac{a_i^*(t-s)}{1-sp} - k\left\{ \frac{s(t-s)}{1+p(t-2s)} + \frac{(t-s)(t-s-1)}{2(1-sp)} \right\} = 0, \quad (i=s+1,\ldots,t).$$

Summing the equations in (2.6.2a), with respect to i from 1 to s, yields

$$\text{(2.6.3)} \qquad \frac{1}{p}\sum_{i=1}^{s} a_i^* - ks\left\{ \frac{s-1}{2p} + \frac{(t-s)^2}{1+p(t-2s)} \right\} = 0,$$

from which we obtain

$$p = \frac{2\sum_{i=1}^{s} a_i^* - ks(s-1)}{2ks(t-s)^2 - (t-2s)\left\{ 2\sum_{i=1}^{s} a_i^* - ks(s-1)\right\}} .$$

Upon substituting for a_i^* from (2.4.4), and using the parameter relationships of (1.3.1) - (1.3.4), we obtain

(2.6.4) $$p = \frac{ks(4t - s - 3) - 2 \sum_{i=1}^{s} \sum_{j} \sum_{u} n_{iju}\, r_{iju}}{ks(5st - 2t^2 - 6s + 3t) - 2(2s - t) \sum_{i=1}^{s} \sum_{j} \sum_{u} n_{iju}\, r_{iju}} .$$

Summing equations in (2.6.2b), with respect to i from s + 1 to t, yields the same result, (2.6.4) -- as is expected.

Now, by proceeding in a manner analagous to that given by Bradley in [4], we will obtain a test for (2.6.1). Let X denote the number of times an object of the first group of s objects ranks above an object of the second group of (t - s) objects. Then,

(2.6.5) $$\sum_{i=1}^{s} \sum_{j} \sum_{u} n_{iju}\, r_{iju} = \frac{3ks(s-1)}{2} + 2ks(t-s) - X,$$

and upon substitution in (2.6.4) we obtain

(2.6.6) $$p = \frac{X}{ks(t-s)^2 + (2s-t)X} .$$

From the model of Section 2.3, we obtain the probability that any object of the first group is ranked above any object of the second group,

$$\frac{\pi}{\pi + \frac{1-s\pi}{t-s}} = \frac{\pi(t-s)}{1 + (t-2s)\pi} ,$$

which under the null hypothesis H_0 equals 1/2. Now, from this discussion and from the observation that there are ks(t - s) comparisons of objects of the first group with objects of the second group, it becomes clear that the binomial or sign test will provide an appropriate test procedure for (2.6.1). The example given in Section 3.7 will further clarify the situation.

2.7. Test III.

Suppose we have g groups with v judges in each group. We are interested in investigating the equality of the true treatment ratings of the t objects under the assumption of within-group judge consistency, but not necessarily assuming between-group judge consistency. It should be observed that these g groups could contain the same v judges, but due to some additional feature, such as significant time lag between repetitions, or training gained from continued experimentation, we are unwilling to assume between-group judge consistency.

Let $\pi_{1u}^{m}, \ldots, \pi_{tu}^{m}$ represent the true ratings of the t objects corresponding to the u-th judge in the m-th group, $(u = 1, \ldots, v;$ $m = 1, \ldots, g)$. Then, we wish to test the null hypothesis,

$$H_0: \quad \pi_{iu}^{m} = 1/t,$$

(2.7.1) against the alternative hypothesis,

$$H_3: \quad \pi_{iu}^{m} = \pi_{i}^{m}, \quad (i = 1, \ldots, t; u ' 1, \ldots, v; m = 1, \ldots, g).$$

Let $n_{iju}^{m} = 1$ or 0 depending on whether or not the i-th and j-th objects are compared by the u-th judge of the m-th group, $(i \neq j;$ $i, j = 1, \ldots, t;$ $u = 1, \ldots, v;$ $m = 1, \ldots, g)$, and conventionally take $n_{iiu}^{m} = 0$, $(i = 1, \ldots, t;$ $u = 1, \ldots, v;$ $m = 1, \ldots, g)$. It is clear that the corresponding incidence matrices

(2.7.2) $$(n_{iju}^{m}), \quad (m = 1, \ldots, g; \quad u = 1, \ldots, v),$$

will have analagous properties to those given in (2.2.5) and (2.2.6).

Let r_{iju}^{m} be the rank assigned to the i-th object if compared with the j-th object by the u-th judge of the m-th group. Then, corresponding to (2.2.2), we have

(2.7.3) $$r^m_{iju} + r^m_{jiu} = 3, \quad (m = 1, \ldots, g; \ u = 1, \ldots, v),$$

when the i-th and j-th objects are compared by the u-th judge of the m-th group.

Now, under the assumption of probability independence between pairs of objects, judges, and groups of judges, and by defining

(2.7.4) $$a^*_{im} = 2\,\alpha\, v - \sum_j \sum_u n^m_{iju}\, r^m_{iju},$$

we obtain, by procedures used in Sections 2.4 and 2.5,

(2.7.5) $$\left\{\ln L_{(3)} \middle| H_3\right\} = \sum_m \left\{ \sum_i a^*_{im} \ln \pi^m_i - k \sum_{i<j} \ln(\pi^m_i + \pi^m_j) \right\},$$

(2.7.6) $$\left\{\ln L_{(3)} \middle| H_0\right\} = -v\, r\, g \ln 2,$$

where $L_{(3)}$ is the likelihood function, and $\left\{\ln L_{(3)} \middle| H_i\right\}$ denotes the natural logarithm of the likelihood function when the hypothesis H_i is true, $(i = 0, 3)$. Then, if we denote the likelihood ratio by λ_3 and define our test statistic to be $T^{(3)} = -2 \ln \lambda_3$, and define

(2.7.7) $$B^{(1)}_m = k \sum_{i<j} \log(p^m_i + p^m_j) - \sum_i a^*_{im} \log p_i,$$

where $p^m_1, \ldots, p^m_t$, the maximum likelihood estimates of $\pi^m_1, \ldots, \pi^m_t$, are solutions of the system of equations

(2.7.8) $$\frac{a^*_{im}}{p^m_i} - k \sum_j{}' (p^m_i + p^m_j)^{-1} = 0, \quad (i = 1, \ldots, t),$$
$$\sum_i p^m_i = 1, \quad (m = 1, \ldots, g),$$

(2.7.9) $$T^{(1)}_m = 2\, v\, r \ln 2 - 2\, B^{(1)}_m \ln 10,$$

we obtain, by the methods of Section 2.5,

(2.7.10) $$T^{(3)} = \sum_m T_m^{(1)} \quad .$$

Define

(2.7.11) $$B^{(3)} = \sum_m B_m^{(1)} \quad .$$

For $B^{(3)}$ or $T^{(3)}$ to provide a test for (2.7.1), some knowledge about the distribution of $B^{(3)}$ under the null hypothesis H_0 is needed. Clearly, the distribution of $B_m^{(1)}$ under H_0 is the same as the distribution of $B^{(1)}$ under H_0. Hence, the probability of a specified value B of $B^{(3)}$ can be obtained by determining the probabilities of the joint occurrences of all combinations of $B_1^{(1)}, \ldots, B_g^{(1)}$ such that $\sum_m B_m^{(1)} = B$, and then summing these probabilities. Thus the distribution of $B^{(3)}$ under H_0 can be obtained from the distribution of $B^{(1)}$ under H_0.

For example, consider the case $t = 4$, $k = 2$, $g = 2$. In this situation $B^{(3)} = B_1^{(1)} + B_2^{(1)}$, and the distributions of $B_1^{(1)}$ and $B_2^{(1)}$ under H_0 are given in (2.5.13) and in [2]. One possible value of $B^{(3)}$ is 1.204. This value can be obtained in the following ways:

(i) $B_1^{(1)} = 0.602$, $B_2^{(1)} = 0.602$;

(ii) $B_1^{(1)} = 0$, $B_2^{(1)} = 1.204$;

(iii) $B_1^{(1)} = 1.204$, $B_2^{(1)} = 0$.

In (i), three different sets of rank sums (6, 8, 11, 11), (6, 9, 9, 12), (7, 7, 10, 12), each with probability of occurrence 24/4096 when H_0 is true, can produce the value 0.602 for $B_1^{(1)}$ or $B_2^{(1)}$. Hence the probability of occurrence of event (i) under H_0 is $3^2(24/4096)(24/4096) = 5184/(4096)^2$. In (ii), $B_1^{(1)}$ is zero only when the set of rank sums

(6, 8, 10, 12) occurs, $B_2^{(1)} = 1.204$ only when the set of rank sums (7, 7, 11, 11) occurs. Hence the probability of occurrence of event (ii) under H_0 is $576/(4096)^2$. Similarly, the probability of occurrence of event (iii) under H_0 is $576/(4096)^2$. Thus,

$$P\left\{ B^{(3)} = 1.204 \,|\, H_0 \right\} = (5184 + 576 + 576)/(4096)^2 = 0.0^3 3777.$$

By proceeding in this manner, the distribution of $B^{(3)}$ under H_0 can be obtained.

For reference, the distribution of $B^{(3)}$ under H_0, for the above example, has been tabulated in Appendix B. For additional discussion and tables for $t = 3$, $g, k = 2, \ldots, 5$, consult [2].

We are now able to introduce the following test procedure for (2.7.1): Compute the rank sums

$$\sum_j \sum_u n_{iju}^m \, r_{iju}^m, \qquad (i = 1, \ldots, t),$$

corresponding to the m-th group of judges, $(m = 1, \ldots, g)$. Then, if the design parameters are within the range $t = 3$, $k = 1, \ldots, 10$; $t = 4$, $k = 1, \ldots, 8$; $t = 5$, $k = 1, \ldots, 5$, determine the corresponding value $B_m^{(1)}$ from the available tables in [2] and [3]. Sum these $B_m^{(1)}$'s with respect to m, thus obtaining $B^{(3)}$. If $t = 3$, $g, k = 2, \ldots, 5$; $t = 4$, $g, k = 2$, the exact significance level of $B^{(3)}$ is given in [2] or in Appendix B. Otherwise, without further extension of these tables, an exact test of (2.7.1) using $B^{(3)}$ will not be available. However, the significance level of $B^{(3)}$ can be approximately determined and will be discussed in Chapter III.

For large t, g, k, it will be necessary to solve the equations in (2.7.8) for $p_1^m, \ldots, p_t^m$, $(m = 1, \ldots, g)$, and then to evaluate $B^{(3)}$ using (2.7.7) and (2.7.11).

It should be noted that estimates of the true object ratings based on all the gv judges will not be available. It will only be possible to obtain estimates of the true object ratings corresponding to a specific group of v judges. This is entirely reasonable, though, since we were not originally willing to make the assumption of equality of the true ratings from group to group.

2.8. Test IV.

It is conceivable that the g groups of v judges considered in Section 2.7 could have had the same judges in each group. Then we would have v judges each making g incomplete repetitions of r pairs, the specific r pairs included in each repetition being specified by (2.7.2) which depends on the paired comparison design used. We desire to test for the equality of the true object ratings. The situation where we are unwilling to assume group consistency of judges from one repetition to the next is handled in Section 2.7, with the hypotheses being stated in (2.7.1). If we are willing to assume consistency of judges from one repetition to the next, we may wish to test the null hypothesis,

$$H_0:\ \pi_{iu}^{m} = 1/t,$$

(2.8.1) against the alternative hypothesis,

$$H_4:\ \pi_{iu}^{m} = \pi_i,\ (i = 1,\ldots,t;\ u = 1,\ldots,v;\ m = 1,\ \ldots,\ g).$$

If we define

$$a_i' = \sum_m a_{im}^{*} = \sum_m \left\{ 2\,\alpha\,v - \sum_j \sum_u n_{iju}^{m}\, r_{iju}^{m} \right\}, \tag{2.8.2}$$

then, under the assumption of probability independence between pairs of objects, judges, and groups of judges, and by procedures used in

Sections 2.4 and 2.5, we obtain

$$(2.8.3)\qquad \left\{\ln L_{(4)}|H_4\right\} = \sum_i a_i' \ln \pi_i - gk \sum_{i<j} \ln(\pi_i + \pi_j)$$

$$\left\{\ln L_{(4)}|H_0\right\} = - v\, r\, g \ln 2,$$

as in (2.7.6), the simplifications being enabled by (2.7.2), (2.2.5), and (1.3.1) - (1.3.4). Define

$$(2.8.4)\qquad B^{(4)} = g\, k \sum_{i<j} \log (p_i + p_j) - \sum_i a_i' \log p_i,$$

where a_i' is defined in (2.8.2), and where $p_1, \ldots, p_t$, the maximum-likelihood estimates of $\pi_1, \ldots, \pi_t$, are solutions of the system of equations

$$(2.8.5)\qquad \begin{aligned} &\frac{a_i'}{p_i} - g\, k \sum_j{}' (p_i + p_j)^{-1} = 0, \quad (i = 1, \ldots, t), \\ &\sum_i p_i = 1, \end{aligned}$$

which has been obtained by maximizing (2.8.3) subject to the constraint $\sum_i \pi_i = 1$. Then, if we denote the likelihood ratio by λ_4, and if we define the test statistic to be $T^{(4)} = -2 \ln \lambda_4$, we obtain from (2.7.6), (2.8.3), and (2.8.4),

$$(2.8.6)\qquad T^{(4)} = 2\, v\, r\, g \ln 2 - 2\, B^{(4)} \ln 10.$$

Upon examination of (2.5.2) and (2.5.6), we note the similarity between a_i' and a_i^*, and between $B^{(4)}$ and $B^{(1)}$, with k replaced by gk. Hence, for t and gk, within the range of the tables in [2] and [3], a test for (2.8.1) is provided by computing the rank sums

$$\sum_m \sum_j \sum_u n_{iju}^m\, r_{iju}^m, \qquad (i = 1, \ldots, t),$$

and, corresponding to this set of rank sums, determining $p_1, \ldots, p_t$,

$B^{(4)}$, and its significance level from these available tables.

If either t or gk is outside the range of the presently available tables, the above test procedure cannot be used. For this situation, we can determine $p_1, \ldots, p_t$ by direct solution of equations (2.8.5), or, if $t \leq 5$, we can use the method mentioned in Section 2.5. We can then evaluate $B^{(4)}$ and $T^{(4)}$ from (2.8.4) and (2.8.6), respectively, but the significance level can only be approximately determined. This will be discussed in Chapter III.

It should be noted that there are many ways in which g sets of r pairs could be assigned to each of the judges. No matter how the sets of pairs are assigned, the statistic $B^{(4)}$ will have the same distribution under the null hypothesis H_0 of (2.8.1). This would suggest that simplicity of assignment of the g sets to the v judges would be the popular criterion. However, it is felt that it would be better to have each judge compare as many different sets of r pairs as possible, as this may provide greater connectivity. The investigation of an exact criterion is desirable, but will not be considered at this time.

2.9. Test V.

Consider the situation of Section 2.7 where we had g groups of v judges, and where we assumed within-group judge consistency. This time we wish to test for judge consistency between groups; that is, we wish to test the null hypothesis,

$$H_4: \quad \pi_{iu}^{m} = \pi_i,$$

(2.9.1) against the alternative hypothesis,

$$H_3: \quad \pi_{iu}^{m} = \pi_i^{m}, \quad (i = 1, \ldots, t; \; u = 1, \ldots, v; \; m = 1, \ldots, g).$$

Under the assumptions of probability independence of Section 2.7, we have $\{\ell n\ L_{(3)} | H_3\}$, given in (2.7.5), and $\{\ell n\ L_{(3)} | H_4\}$, given in (2.8.3). Hence, if we denote the likelihood ratio by λ_5, and if we define our test statistic to be $T^{(5)} = -2\ \ell n\ \lambda_5$, we obtain

(2.9.2) $$T^{(5)} = 2(B^{(4)} - B^{(3)})\ell n\ 10,$$

where $B^{(3)}$ is defined in (2.7.11), and $B^{(4)}$ in (2.8.4). From the discussion of $B^{(3)}$ and $B^{(4)}$ in Sections 2.7 and 2.8, and hence from the correspondence we can make of $B^{(3)}$ and $B^{(4)}$ with statistics which are tabulated in [2] and [3], we can evaluate $T^{(5)}$ by using these tables, provided that our parameters lie within the range of the tables. When the parameters are outside this range, we must first solve equations (2.7.8) and (2.8.5), and then we must calculate $B^{(3)}$ and $B^{(4)}$ from (2.7.11) and (2.8.4), respectively.

However, the distribution of $T^{(4)}$ under the null hypothesis H_4 of (2.9.1) will depend on $\pi_1, \ldots, \pi_t$, and, hence, will not give us an exact parameter-free test. An approximate test will be discussed in Chapter III.

2.10. Test VI.

Suppose we have v judges each comparing g sets of r pairs of objects, $a_1, \ldots, a_t$, where the sets compared by each judge are specified by (2.7.2). Under the assumption of judge consistency from one repetition to the next, we wish to test for equality of the true object ratings; that is, we desire to test the null hypothesis,

H_0: $\pi_{iu} = 1/t$,

(2.10.1) against the alternative hypothesis,

H_6: No π_{iu} is assumed equal to any π_{jw},

$(i \neq j,\ u \neq w;\ \ i,j = 1, \ldots, t;\ u,w = 1, \ldots, v)$.

Then, under the assumptions of probability independence of Section 2.8, if we define

(2.10.2) $$b_{iu} = 2\,\alpha\, g - \sum_m \sum_j n^m_{iju}\, r^m_{iju},$$

we obtain

(2.10.3) $$\left\{\ln L_{(6)} \,|H_6\right\} = \sum_u \left\{ \sum_i b_{iu} \ln \pi_{iu} - \sum_m \sum_{i<j} n^m_{iju} \ln(\pi_{iu} + \pi_{ju}) \right\},$$

and

$$\left\{\ln L_{(6)} \,|H_0\right\} = - v\, r\, g \ln 2.$$

Then, if we denote the likelihood ratio by λ_6, and if we define the test statistic to be $T^{(6)} = -2 \ln \lambda_6$, we obtain

(2.10.4) $$T^{(6)} = \sum_u \left\{ 2\, g\, r \ln 2 - 2\, B_u \ln 10 \right\},$$

where

(2.10.5) $$B_u = \sum_m \sum_{i<j} n^m_{iju} \log(p_{iu} + p_{ju}) - \sum_i b_{iu} \log p_{iu},$$

and where $p_{1u}, \ldots, p_{tu}$, the maximum-likelihood estimates of $\pi_{1u}, \ldots, \pi_{tu}$, are solutions of the equations

(2.10.6)
$$\frac{b_{iu}}{p_{iu}} - \sum_m \sum_j{}' n^m_{iju} (p_{iu} + p_{ju})^{-1} = 0, \quad (i = 1, \ldots, t),$$

$$\sum_i p_{iu} = 1,$$

which were obtained by maximizing (2.10.3) subject to the constraint $\sum_i \pi_{iu} = 1$, $(u = 1, \ldots, v)$.

Observe that if $g = cv$, $c \geq 1$, such that each judge compares each of all the possible v different sets of r pairs c times, the situation reduces to the equivalent of each judge making the comparison of c k sets of all possible $t(t - 1)/2$ pairs. This situation is dealt with in [2]. Thus, if this method of repetition is employed, for $g = cv$, $T^{(6)}$ will provide an approximate test for (2.10.1) as described in Section 10 of [2] and Section 7.1 of [4].

It would be possible to provide an exact test for (2.10.1) by using the statistic $\sum_u B_u$ defined in (2.10.5). The distribution of B_u and $\sum_u B_u$ under the null hypothesis H_0 of (2.10.1) could be obtained in a manner similar to that for $B^{(1)}$. However, the distribution would depend on the paired comparison design being used, the number g of repetitions made by each judge, and the criterion for assigning the g sets of r pairs to each judge. The assignment of these sets is, in itself, a design problem, and it is conceivable that with the assignment of these sets to satisfy certain additional symmetry conditions, tabulation for small values of t would be practicable.

2.11. Test VII.

Consider the situation of Section 2.10 where v judges compare g sets of r pairs of objects $a_1, \ldots, a_t$, and where we assume judge consistency from one repetition to the next. This time we wish to test for agreement among the judges; that is, we desire to test the null hypothesis,

$$H_1: \pi_{iu} = \pi_i,$$

(2.11.1) against the alternative hypothesis,

$$H_6: \text{ No } \pi_{iu} \text{ is assumed equal to any } \pi_{jw},$$

$$(i \neq j,\ u \neq w;\ i,j = 1,\ldots,t;\ u,w = 1,\ldots,v).$$

If we denote the likelihood ratio by λ_7, and if we define the test statistic to be $T^{(7)} = -2\ln \lambda_7$, then, under the probability independence assumptions of Section 2.10, and from (2.10.3), (2.10.5), (2.8.3), and (2.8.4), we obtain

(2.11.2) $$T^{(7)} = 2(B^{(4)} - \sum_u B_u)\ln 10.$$

As in Section 2.10, it is observed that if $g = cv$, $c \geq 1$, such that each judge compares each of all the possible v different sets of r pairs c times, the situation becomes equivalent to each judge making the comparison of ck sets of all possible $t(t-1)/2$ pairs. For this situation, which is dealt with in [2] and [4], $T^{(7)}$ will provide an approximate test for (2.11.1). Any exact test for (2.11.1) provided by $T^{(7)}$ will not be a parameter-free test, since, under the null hypothesis H_1, it will depend on $\pi_1, \ldots, \pi_t$.

For the situations encountered here and in Section 2.10, it would be of considerable interest to investigate how good the approximate tests provided by $T^{(7)}$ and $T^{(6)}$ actually are for small g, when g is not some multiple of v but is comprised of as many different sets of r pairs as possible.

2.12. Extreme sets of rank sums.

For the situation of Section 2.5, the extreme values which the sum of ranks corresponding to the i-th object may have are

$$\sum_j \sum_u n_{iju} r_{iju} = 2\alpha v \text{ or } \alpha v, \qquad (i = 1, \ldots, t).$$

When either of these cases arise for any of the objects, equations in (2.5.4), which provide our maximum-likelihood estimates for the parameters, become somewhat valueless. In order for this rank sum to be

$2\alpha v$ (or αv), it is necessary that the i-th object be judged inferior (or superior) in all comparisons made by all judges. Thus, when these situations arise, it would seem reasonable to estimate π_i by 0 (or 1); then, drop this object from the analysis and consider only the remaining t - 1 objects.

Let us consider the case where the t-th object has the extreme sum of ranks

$$\sum_j \sum_u n_{tju} r_{tju} = 2\alpha v .$$

Clearly, there is no loss in generality in selecting the t-th object for convenience in notation. Omitting this object from consideration, we will have a reduced set of (t - 1) rank sums, namely,

$$\sum_j \sum_u n_{1ju} r_{1ju} - k, \ \ldots, \ \sum_j \sum_u n_{t-1,ju} r_{t-1,ju} - k. \tag{2.12.1}$$

Now, from the symmetry of the paired comparison designs as displayed in (2.2.5) and (2.2.6), we obtain the logarithm of the likelihood function $L_{(1)}$ under the alternative hypothesis H_1 of (2.5.1) to be

$$\begin{aligned} \left\{\ln L_{(1)} | H_1\right\} &= \sum_{i=1}^{t-1} \left\{ 2(\alpha v - k) - \sum_{j=1}^{t-1} \sum_u n_{iju} r_{iju} \right\} \ln \pi_i \\ &\quad - k \sum_{i<j}^{t-1} \ln(\pi_i + \pi_j) \\ &= \sum_{i=1}^{t-1} \left\{ 2k(t-2) - \sum_{j=1}^{t-1} \sum_u n_{iju} r_{iju} \right\} \ln \pi_i \\ &\quad - k \sum_{i<j}^{t-1} \ln(\pi_i + \pi_j) . \end{aligned} \tag{2.12.2}$$

It also follows easily that the system of equations, which provides maximum-likelihood estimates of $\pi_1, \ldots, \pi_{t-1}$, and the test statistic

$B^{(1)}$ will be the same as in (2.5.4) and (2.5.6), respectively, with t replaced by t - 1.

Hence, the test procedure for (2.5.1) will be the same as that outlined at the end of Section 2.5 with t replaced by t - 1, and with the original set of rank sums replaced by the reduced set (2.12.1).

For the case where the t-th object has the extreme rank sum

$$\sum_j \sum_u n_{tju} \, r_{tju} = \alpha \, v,$$

the omission of this object from consideration will result in the reduced set of (t - 1) rank sums

$$\sum_j \sum_u n_{1ju} \, r_{1ju} - 2k, \; \ldots, \; \sum_j \sum_u n_{t-1,ju} \, r_{t-1,ju} - 2k. \tag{2.12.3}$$

Similarly, the test procedure for (2.5.1), in this case, is the same as that outlined in Section 2.5, with t replaced by (t - 1), and with the original set of rank sums replaced by the reduced set (2.12.3).

In Test IV of Section 3.8, for the case where the t-th object has the extreme rank sum

$$\sum_m \sum_j \sum_u n^m_{tju} \, r^m_{tju} = 2 \, \alpha \, g \, v \text{ or } \alpha \, g \, v, \tag{2.12.4}$$

the test procedure for (2.8.1) is the same as described in Section 2.8, with t replaced by (t - 1), and with the original set of rank sums replaced by

$$\sum_m \sum_j \sum_u n^m_{1ju} \, r^m_{1ju} - gk, \; \ldots, \; \sum_m \sum_j \sum_u n^m_{t-1,ju} \, r^m_{t-1,ju} - gk,$$

(2.12.5) or

$$\sum_m \sum_j \sum_u n^m_{1ju} \, r^m_{1ju} - 2gk, \; \ldots, \; \sum_m \sum_j \sum_u n^m_{t-1,ju} \, r^m_{t-1,ju} - 2\, gk,$$

depending on whether or not the t-th object rank sum is $2 \, \alpha \, g \, v$ or $\alpha \, g \, v$, respectively.

For Tests III and V, of Sections 2.7 and 2.9, respectively, if an extreme rank sum occurs for the t-th object of the m-th group, namely,

$$\sum_j \sum_u n^m_{tju}\, r^m_{tju} = 2\,\alpha\, v \text{ or } \alpha\, v,$$

the test procedures described in the respective sections can be employed with t replaced by (t - 1), and with the original set of rank sums replaced by the reduced set of rank sums,

$$\sum_j \sum_u n^m_{1ju}\, r^m_{1ju} - k,\ \ldots,\ \sum_j \sum_u n^m_{t-1,ju}\, r^m_{t-1,ju} - k,$$

or

$$\sum_j \sum_u n^m_{1ju}\, r^m_{1ju} - 2k,\ \ldots,\ \sum_j \sum_u n^m_{t-1,ju}\, r^m_{t-1,ju} - 2k.$$

Such an alteration to the test procedure can be extended when more than one of the g groups has an extreme rank sum. For Test V, if the t-th object in all g groups has an extreme rank sum as in (2.12.4), in order to evaluate $B^{(4)}$ we will need to follow the procedure of Section 2.8, with t replaced by t - 1, and to employ the reduced set of rank sums of (2.12.5).

If, in Test VI of Section 2.10, the t-th object has an extreme rank sum corresponding to the v-th judge, namely,

$$\sum_m \sum_j n^m_{tjv}\, r^m_{tjv} = 2\,\alpha\, g \text{ or } \alpha\, g,$$

then,

$$\begin{aligned}\left\{\ln L_{(6)}\,|H_6\right\} = {} & \sum_{u=1}^{v-1}\left\{\sum_i b_{iu} \ln\pi_{iu} - \sum_m \sum_{i<j} n^m_{iju} \ln(\pi_{iu} + \pi_{ju})\right\} \\ & + \sum_{i=1}^{t-1} \sum_m \left\{2(\alpha - n^m_{itv}) - \sum_{j=1}^{t-1} n^m_{ijv}\, r^m_{ijv}\right\} \ln \pi_{iv} \\ & - \sum_{i<j}^{t-1} \sum_m n^m_{ijv} \ln(\pi_{iv} + \pi_{jv}) \quad .\end{aligned} \tag{2.12.6}$$

Hence, the test statistic $T^{(6)}$ of (2.10.4) will become

$$T^{(6)} = 2\, gvr\, \ln 2 - 2\left(\sum_{u=1}^{v-1} B_u + B_v'\right)\ln 10, \tag{2.12.7}$$

where

$$B_v' = \sum_{i=1}^{t-1} \sum_m \left\{ 2(\alpha - n_{itv}^m) - \sum_{j=1}^{t-1} n_{ijv}^m\, r_{ijv}^m \right\} \log p_{iv} \tag{2.12.8}$$

$$- \sum_{i<j}^{t-1} \sum_m n_{ijv}^m \log (p_{iv} + p_{jv}),$$

where $p_{1v}, \ldots, p_{t-1,v}$ are solutions of

$$\frac{\sum_m \left\{ 2(\alpha - n_{itv}^m) - \sum_{j=1}^{t-1} n_{ijv}^m\, r_{ijv}^m \right\}}{p_{iv}} - \sum_{j=1}^{t-1} \sum_m n_{iju}^m (p_{iv}+p_{jv})^{-1}$$

$$= 0, \qquad (i = 1, \ldots, t-1), \tag{2.12.9}$$

$$\sum_{i=1}^{t-1} p_{iv} = 1.$$

The manner of extension of this procedure to the situation where other objects have extreme rank sums corresponding to certain judges is clear.

For Test VII of Section 2.11, the above procedure can also be employed if the t-th object has an extreme rank sum corresponding to the v-th judge. If, in addition, the t-th object has an extreme rank sum, as in (2.12.4), corresponding to each judge, the test statistic $T^{(7)}$ of (2.11.2) becomes

$$T^{(7)} = 2\left(B^{(4)\prime} - \sum_u B_u'\right)\ln 10, \tag{2.12.10}$$

where B_u' is defined in (2.12.8), and where $B^{(4)'}$ is obtained from $B^{(4)}$ of (2.8.4) by replacing t by (t - 1), and by using the reduced set of rank sums of (2.12.5).

CHAPTER III

SOME LARGE-SAMPLE PROPERTIES OF THE TESTS AND SOME EXAMPLES TO CLARIFY THE TEST PROCEDURES

3.1. Introduction.

The exact tests discussed in Chapter II are restricted by the limited range of available tables. In this chapter, approximate tests for these situations will be described. To obtain these approximate tests, we will redefine the parameters in such a way that we may make use of the results obtained by Bradley [4]. Also, since discussion with the use of incomplete repetitions may be somewhat obscure, some examples will be included to help clarify the test procedures.

3.2. Large-sample distribution of $T^{(1)}$ and $T^{(4)}$.

Let

$$y_i = t(p_i - \frac{1}{t}), \quad (i = 1, \ldots, t), \tag{3.2.1}$$

where $p_1, \ldots, p_t$ are the maximum-likelihood estimates of $\pi_1, \ldots, \pi_t$ given in (2.5.1). Now observe that

$$\sum_i \sum_j \sum_u n_{iju} r_{iju} = \frac{3kt(t-1)}{2} ;$$

hence,

$$\begin{aligned} \sum_i a_i^* &= 2\alpha vt - \frac{3kt(t-1)}{2} \\ &= \frac{kt(t-1)}{2} , \end{aligned} \tag{3.2.2}$$

where a_i^* is defined in (2.4.4). Now, upon substitution in (2.5.6) and (2.5.8), we obtain

$$(3.2.3)\qquad B^{(1)} = \frac{kt(t-1)}{2}\log 2 + k\sum_{i<j}\log\left\{1+\tfrac{1}{2}(y_i+y_j)\right\} - \sum_i a_i^*\log(1+y_i),$$

and

$$(3.2.4)\qquad T^{(1)} = 2k\sum_{i<j}\ell n\left\{1+\tfrac{1}{2}(y_i+y_j)\right\} - 2\sum_i a_i^*\,\ell n(1+y_i),$$

respectively. Substitution of (3.2.1) in (2.5.4) yields

$$a_i^* = \frac{k(1+y_i)}{2}\sum_j\left\{1+\tfrac{1}{2}(y_i+y_j)\right\}^{-1} ,$$

and hence, upon substitution for a_i^* in (3.2.4), we obtain

$$(3.2.5)\qquad T^{(1)} = k\sum_i(1+y_i)\,\ell n(1+y_i)\sum_j{}'\left\{1+\tfrac{1}{2}(y_i+y_j)\right\}^{-1}$$
$$+\,2k\sum_{i<j}\ell n\left\{1+\tfrac{1}{2}(y_i+y_j)\right\}^{-1}.$$

$T^{(1)}$, as expressed in (3.2.5), can be put in the form

$$(3.2.6)\qquad T^{(1)} = \tfrac{1}{4}kt\sum_i y_i^2 + R(y_i) ,$$

where $R(y_i)$ depends on higher powers of y_i than the second power. Then, in the manner followed by Bradley [4], if we redefine

$$(3.2.7)\qquad \pi_i = \frac{1}{t} + \frac{\delta_{ik}}{\sqrt{k}} ,\quad (i = 1, \ldots, t),$$

where δ_{ik} is a sequence of constants converging to δ_i as $k \to \infty$, it can be shown that $R(y_i)$ converges to zero in probability as $k \longrightarrow \infty$, and that $T^{(1)}$ has the same limiting distribution as $\frac{1}{4} k\, t \sum_i y_i^2$. This limiting distribution is, under H_1, a non-central χ^2-distribution with $(t - 1)$ degrees of freedom and parameter of non-centrality

$$\lambda^{(1)} = \frac{1}{4} t^3 \sum_i \delta_i^2 , \tag{3.2.8}$$

which, for large k, can be approximated by

$$\lambda^{(1)} = \frac{1}{4}k\, t^3 \sum_i (\pi_i - \frac{1}{t})^2 . \tag{3.2.9}$$

Under the null hypothesis H_0, $\pi_i = 1/t$, $\delta_{ik} = \delta_i = 0$, $(i = 1,\ldots,t)$. Therefore $\lambda^{(1)} = 0$ and thus, under H_0, $T^{(1)}$ has a limiting central χ^2-distribution with $(t - 1)$ degrees of freedom.

Similarly, $T^{(4)}$ has the same limiting distribution as $gk \longrightarrow \infty$.

Thus, approximate tests for (2.5.1) and (2.8.1) are provided by $T^{(1)}$ and $T^{(4)}$, respectively, with approximate significance levels being obtained from tables for the central χ^2-distribution with $(t - 1)$ degrees of freedom.

3.3. Large-sample distribution of $T^{(3)}$.

If, in (2.7.1), we redefine

$$\pi_i^m = \frac{1}{t} + \frac{\delta_{ik}^m}{\sqrt{k}} , \quad (i = 1, \ldots, t;\ \ m = 1, \ldots, t), \tag{3.3.1}$$

where δ_{ik}^m is a sequence of constants converging to δ_i^m as $k \longrightarrow \infty$, then, from Section 3.2 and from [4], it follows that $T_m^{(1)}$, which is defined in (2.7.9), has a limiting distribution, under H_3, which is a non-central χ^2-distribution with $(t - 1)$ degrees of freedom and parameter of non-centrality

$$\lambda_m^{(1)} = \tfrac{1}{4}t^3 \sum_i (\delta_i^m)^2 \ , \tag{3.3.2}$$

which, for large k, can be approximated by

$$\lambda_m^{(1)} = \tfrac{1}{4}k\ t^3 \sum_i (\pi_i^m - \tfrac{1}{t})^2 . \tag{3.3.3}$$

Since $T^{(3)} = \sum_m T_m^{(1)}$, it follows from the additivity property of χ^2 that $T^{(3)}$ has a limiting distribution, under H_3, which is a non-central χ^2-distribution with g(t-1) degrees of freedom and parameter of non-centrality

$$\lambda^{(3)} = \tfrac{1}{4}t^3 \sum_i \sum_m (\delta_i^m)^2 m \tag{3.3.4}$$

which, for large k, can be approximated by

$$\lambda^{(3)} = \tfrac{1}{4}\ k\ t^3 \sum_i \sum_m (\pi_i^m - \tfrac{1}{t})^2 \ . \tag{3.3.5}$$

Under the null hypothesis H_0, $\pi_i^m = \frac{1}{t}$, $\delta_{ik}^m = \delta_i^m = 0$, $(i = 1,\ldots,t;$ $m = 1,\ldots,g)$, and $\lambda^{(3)} = 0$. Hence, under H_0, $T^{(3)}$ has a limiting central χ^2-distribution with g(t-1) degrees of freedom and therefore provides an approximate test for (2.7.1).

3.4. Large-sample distribution of $T^{(5)}$.

From (2.9.2) we have

$$\begin{aligned} T^{(5)} &= 2(B^{(4)} - B^{(3)})\ \ell n\ 10 \\ &= 2vrg\ell n\ 2 - 2B^{(3)}\ell n\ 10 - \left\{2vrg\ell n\ 2 - 2B^{(4)}\ell n\ 10\right\} \qquad (3.4.1) \\ &= T^{(3)} - T^{(4)}. \end{aligned}$$

Then, with the parameters of (2.9.1) redefined as in (3.2.7) and (3.3.1), it follows, from Sections 3.2 and 3.3 and from [4], that $T^{(5)}$ has a limiting distribution, under H_3, which is a non-central χ^2-distribution with (g-1)(t-1) degrees of freedom and parameter of non-centrality

$$\lambda^{(5)} = \frac{1}{4} t^3 \sum_i \sum_m \left\{ (\delta_i^m)^2 - \delta_i^2 \right\} , \tag{3.4.2}$$

which, for large k, can be approximated by

$$\lambda^{(5)} = \frac{1}{4} k\ t^3 \sum_i \sum_m \left\{ (\pi_i^m)^2 - \pi_i^2 \right\} . \tag{3.4.3}$$

Under the null hypothesis H_4, $\pi_i^m = \pi_i$, $(i = 1, \ldots, t;\ m = 1, \ldots, g)$, $\lambda^{(5)} = 0$, and hence $T^{(5)}$ has a limiting central χ^2-distribution with (g-1)(t-1) degrees of freedom. Thus an approximate test for (2.9.1) is provided by $T^{(5)}$, the approximate significance level being obtainable from the table for the χ^2-distribution with (g-1)(t-1) degrees of freedom.

3.5. Large-sample distribution for $T^{(6)}$.

From (2.10.4) we have

$$\begin{aligned} T^{(6)} &= \sum_u (2gr \ln 2 - 2B_u \ln 10) \\ &= \sum_u T_u . \end{aligned} \tag{3.5.1}$$

If g = cv, $(c \geq 1)$, and the g incomplete repetitions are assigned to each judge in a manner such that he compares all possible v different sets of r pairs a total of c times, and if we redefine the parameters

(3.5.2) $$\pi_{iu} = \frac{1}{t} + \frac{\delta_{iuk}}{\sqrt{kc}}, \qquad (i = 1, \ldots, t; \;\; u = 1, \ldots, v),$$

where δ_{iuk} is a sequence of constants converging to δ_{iu} as $kc \longrightarrow \infty$, it follows, from Section 2.10, Section 3.2, and [4], that T_u, $(u = 1, \ldots, v)$, has a limiting distribution, under H_6, which is a non-central χ^2-distribution with (t-1) degrees of freedom and paramter of non-centrality

(3.5.3) $$\lambda_u^{(6)} = \frac{1}{4}t^3 \sum_i \delta_{iu}^2 ,$$

which, for large kc, can be approximated by

(3.5.4) $$\lambda_u^{(6)} = \frac{1}{4}\, kct^3 \sum_i (\pi_{iu} - \frac{1}{t})^2 .$$

Hence, from the additivity property of χ^2, it follows that $T^{(6)}$ has a limiting distribution, under H_6, which is a non-central χ^2-distribution with v(t-1) degrees of freedom and parameter of non-centrality

(3.5.5) $$\lambda^{(6)} = \frac{1}{4}t^3 \sum_i \sum_u \delta_{iu}^2 ,$$

which, for large kc, can be approximated by

(3.5.6) $$\lambda^{(6)} = \frac{1}{4}\, kct^3 \sum_i \sum_u (\pi_{iu} - \frac{1}{t})^2 .$$

Under the null hypothesis H_0, $\pi_{iu} = 1/t$, $(i = 1,\ldots,t; \; u = 1,\ldots,v)$, $\lambda^{(6)} = 0$, and hence $T^{(6)}$ has a limiting distribution, under H_0, which is a central χ^2-distribution with v(t-1) degrees of freedom. Hence an approximate test for (2.10.1) is provided by $T^{(6)}$, the approximate significance level being obtainable from a table for the χ^2-distribution with v(t-1) degrees of freedom.

3.6. Large-sample distribution for $T^{(7)}$.

From (2.11.2) we have

$$T(7) = 2(B^{(4)} - \sum_u B_u)\ln 10$$

$$(3.6.1) \qquad = \sum_u (2\,gr\ln 2 - 2B_u \ln 10) - (2grv\ln 2 - 2B^{(4)}\ln 10)$$

$$= T^{(6)} - T^{(4)} .$$

Hence, if $g = cv$, $(c \geq 1)$, and, if the repetitions are performed as prescribed in Section 3.5, then, from Sections 3.2 and 3.5, and from [4], it follows that $T^{(7)}$ has a limiting distribution, under H_6, which is a non-central χ^2-distribution with (v-1)(t-1) degrees of freedom and parameter of non-centrality

$$(3.6.2) \qquad \lambda^{(7)} = \frac{1}{4} t^3 \sum_i \sum_u (\delta_{iu}^2 - \delta_i^2),$$

which, for large kc, can be approximated by

$$(3.6.3) \qquad \lambda^{(7)} = \frac{1}{4} kc\, t^3 \sum_i \sum_u (\pi_{iu}^2 - \pi_i^2) .$$

Under the null hypothesis H_1, $\pi_{iu} = \pi_i$, $(i = 1, \ldots, t;\ u = 1, \ldots, v)$, $\lambda^{(7)} = 0$, and hence $T^{(7)}$ has a limiting distribution, under H_1, which is a central χ^2-distribution with (v-1)(t-1) degrees of freedom. Thus an approximate test for (2.11.1) is provided by $T^{(7)}$, the approximate significance level being obtainable from a table for the χ^2-distribution with (v-1)(t-1) degrees of freedom.

3.7. Examples of test procedures.[1]

Example 1.

For the situation of Section 2.5, consider $t = 5$ specimens of handwriting compared pairwise by $v = 6$ judges, with each judge recording a preference for one object of each pair which he compares. We will refer to the factors which influence judgment as characteristic x. The r pairs compared by the u-th judge, $(u = 1, \ldots, 6)$, will be dictated by the field plan of design (ii) of (1.4.4) displayed in Table 1.4.1. The incidence matrices for this design, obtainable from Table 1.4.1, are:

(3.7.1)

$$N_1 = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix},$$

$$N_3 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \quad N_4 = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix},$$

1. The data used in these examples is a small portion of that collected for an experiment conducted at the Psychometric Laboratory of the University of North Carolina.

$$N_5 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}, \qquad N_6 = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix} .$$

From (3.7.1) it is easily seen that

$$\sum_i n_{iju} = \sum_j n_{iju} = 2, \qquad (u = 1, \ldots, 6),$$

$$3\, I(5 \times 5) + \sum_u N_u\, (5 \times 5) = 3\, E(5 \times 5).$$

Now, for judge u, $(u = 1, \ldots, v)$, we construct a $(t \times t)$ preference matrix $R_u = (r_{iju})$, where $r_{iju} = 1$ or 2 if $\{a_i \longrightarrow a_j | u\}$ or $\{a_i \longleftarrow a_j | u\}$, respectively. If objects a_i and a_j are not compared by judge u, we will conventionally take the elements r_{iju} and r_{jiu} of R_u to be zero. Thus, the preference matrix R_u will have 1's and 2's as elements corresponding to the unit elements of N_u, and zeros corresponding to the zero elements of N_u. From this definition of R_u, we observe that the sums of the rows of R_u yield a $(t \times 1)$ vector, the elements of which are the rank sums of the t objects corresponding to judge u. For example, the preference matrix R_1 and the rank sums vector $\sum_j (r_{1j1}, \ldots, r_{tj1})$ corresponding to judge J_1 for this experiment are

$$R_1 = \begin{pmatrix} 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 2 \\ 2 & 2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \end{pmatrix},$$

$$\sum_j (r_{ij1}) = R_1\underline{1} = \begin{pmatrix} 3 \\ 2 \\ 3 \\ 4 \\ 3 \end{pmatrix} \begin{pmatrix} \sum_j n_{1j1}\, r_{1j1} \\ \sum_j n_{2j1}\, r_{2j1} \\ \sum_j n_{3j1}\, r_{3j1} \\ \sum_j n_{4j1}\, r_{4j1} \\ \sum_j n_{5j1}\, r_{5j1} \end{pmatrix},$$

where $\underline{1}$ is a unit column vector of appropriate dimension. By this procedure we obtain for each judge the following rank sums vectors:

$$\begin{aligned} J_1&: (3, 2, 3, 4, 3) \\ J_2&: (2, 2, 4, 3, 4) \\ J_3&: (2, 4, 3, 4, 2) \\ J_4&: (2, 3, 3, 3, 4) \\ J_5&: (4, 2, 2, 4, 3) \\ J_6&: (2, 2, 4, 4, 3). \end{aligned} \tag{3.7.2}$$

Summing these with respect to judges yields

$$\left(\sum_j \sum_u n_{1ju}\, r_{1ju}, \ \ldots, \ \sum_j \sum_u n_{5ju}\, r_{5ju}\right) = (15, 15, 19, 22, 19). \tag{3.7.3}$$

Hence,

(3.7.4) $$(a_1^*, \ldots, a_5^*) = (9, 9, 5, 2, 5),$$

where a_i^* is defined in (2.4.4). The rank sums vector in (3.7.3) is what we require in order to use the tables of [3]. Entering these tables for t = 5 and for k = 3 complete repetitions, we obtain

$$(p_1, \ldots, p_5) = (.38, .38, .10, .03, .10), \; B^{(1)} = 6.686,$$

$$P\left\{ B^{(1)} \leq 6.686 \,|H_0 \right\} = 0.0404.$$

Hence, we would conclude, at the 0.0404 level of significance, that the five handwriting specimens are different with respect to the characteristic x under the assumption that the judges are consistent as a group.

It is of interest to note, from (2.5.8), that

$$T^{(1)} = 2 \, v \, r \, \ln 2 \; - 2 \, B^{(1)} \ln 10$$

$$= 10.80$$

$$\cong \chi^2_{(4)} \text{ from Section 3.2.}$$

From the large-sample properties of $T^{(1)}$, we obtain the approximate significance level to be

$$P\left\{ \chi^2_{(4)} \geq 10.80 \,|H_0 \right\} = 0.028.$$

Thus, although the approximate test will give too many significant results, the approximate significance level obtained is reasonably close to the exact significance level, even for k as small as 3.

Example 2.

We will now consider an example for the situation portrayed in Section 2.8. Suppose that the six judges of Example 1 make two incomplete repetitions with r pairs in each repetition. The r pairs compared by the u-th judge in the first repetition are those indicated in (3.7.1), and, hence, the incidence matrices for the first repetition are

$$(n^1_{iju}) = (n_{iju}), \qquad (u = 1, \ldots, 6).$$

The r pairs compared by the u-th judge in the second repetition are indicated by the incidence matrices

$$(n^2_{iju}) = (n_{ij,u+1}), \ (u = 1, \ldots, 5), \ (n^2_{ij6}) = (n_{ij1}).$$

The rank sums vectors corresponding to each judge for the first repetition are given in (3.7.2). By an analagous procedure to that used in Example 1, we compute the rank sums vectors corresponding to each judge on the second repetition:

(3.7.5)

$$\begin{aligned} J_1&: (2, 3, 2, 4, 4) \\ J_2&: (3, 2, 3, 3, 4) \\ J_3&: (2, 3, 3, 4, 3) \\ J_4&: (2, 3, 3, 3, 4) \\ J_5&: (3, 2, 3, 4, 3) \\ J_6&: (2, 2, 4, 4, 3). \end{aligned}$$

Summing these with respect to judges, we obtain

(3.7.6) $$\left(\sum_j \sum_u n^2_{1ju} r^2_{1ju}, \ldots, \sum_j \sum_u n^2_{5ju} r^2_{5ju}\right) = (14, 15, 18, 22, 21).$$

Hence,

(3.7.7) $$(a^*_{12}, \ldots, a^*_{52}) = (10, 9, 6, 2, 3),$$

where a^*_{12} is defined in (2.7.7). By summing the vectors in (3.7.6) and (3.7.3), we obtain

(3.7.8) $(\sum_m \sum_j \sum_u n^m_{1ju} r^m_{1ju}, \ldots, \sum_m \sum_j \sum_u n^m_{5ju} r^m_{5ju}) = (29, 30, 37, 44, 40)$.

Hence, we obtain

(3.7.9) $(a_{1}^{'}, \ldots, a_{5}^{'}) = (19, 18, 11, 4, 8)$,

where $a_{i.}^{'}$ is defined in (2.8.3). The result in (3.7.9) can be checked by adding vectors in (3.7.7) and (3.7.4). The elements of the vector (3.7.8) are the rank sums which enter into the statistic $B^{(4)}$ of (2.8.4). Here t = 5 and gk = 6 -- the corresponding number of complete repetitions -- are outside the range of available tables. Hence, we are unable to obtain either the value of $B^{(4)}$ or its significance level directly from the tables. However, since t = 5, we can use a procedure outlined in [3] and [15] as a means to assist us to evaluate $B^{(4)}$. Dividing gk and the elements of (3.7.8) by two, we obtain

(3.7.10) gk/2 = 3, (3.7.8)/2 = (14.5, 15, 18.5, 22, 20).

Then, for rank sums vectors

(14, 15, 19, 22, 20) and (15, 15, 18, 22, 20),

we enter the tables of [3] for t = 5 and for 3 complete repetitions and obtain corresponding estimates

(.51, .33, .08, .02, .05) and (.38, .38, .14, .03, .07),

respectively. Then, by linear interpolation, we obtain an approximation to the estimates, $p_1, \ldots, p_5$, corresponding to the elements of (3.7.10) to be

(.445, .355, .110, .025, .060),

which, after adjustment to add to 1, is

(.447, .357, .111, .025, .060).

Starting with these approximations and using the iterative formula given in [2], we obtain

(3.7.11) $(p_1, \ldots, p_5) = (.441, .361, .106, .029, .063)$.

Substituting these values in (2.8.4), we obtain

$$B^{(4)} = 6 \Big\{ \log .802 + \log .547 + \log .470 + \log .504 + \log .467 + \log .390 + \log .424 + \log .135 + \log .169 + \log .092 \Big\}$$
$$- 19 \log .441 - 18 \log .361 - 11 \log .106 - 4 \log .029 - 8 \log .063$$
$$= 12.556.$$

This value substituted in (2.8.6) yields $T^{(4)} = 25.353$. From Section 3.2, it follows that

$$P\left\{ T^{(4)} \geq 25.353 \,|\, H_0 \right\} < .0005,$$

approximately. Hence, under the assumption that the judges are consistent as a group and consistent from one repetition to the next, we would conclude, at less than the 0.0005 level of significance (approximately), that the five handwriting specimens are different with respect to the characteristic x.

Example 3.

Suppose that in the previous example we were unwilling to assume judge consistency from one repetition to the next. The situation then becomes that of Section 2.7. Here we have $g = 2$ groups of $v = 6$ judges comparing t objects according to a characteristic x, and according to this characteristic the judges in each group would not be considered the same.

Using the data of Examples 1 and 2 for each repetition separately, we obtain

$(p_1, \ldots, p_5) = (.38, .38, .10, .03, .10)$, $B_1^{(1)} = 6.686$, $T_1^{(1)} = 10.800$,

$(p_1, \ldots, p_5) = (.51, .33, .10, .02, .03)$, $B_2^{(1)} = 5.598$, $T_2^{(1)} = 15.809$.

Hence, from (2.7.10),

$$T^{(3)} = 10.800 + 15.809$$
$$= 26.609.$$

From Section 3.3, $T^{(3)}$ has, under H_0, approximately a χ^2-distribution with $g(t - 1) = 8$ degrees of freedom, and therefore an approximate significance level for Test (2.7.1) is

$$P\left\{T^{(3)} \geq 26.609 | H_0\right\} = 0.001.$$

Hence, under the assumption of within-group judge consistency only, we conclude, at approximately the 0.001 level of significance, that the five handwriting specimens are different with respect to the characteristic x.

Example 4.

Suppose for the experiment used in Examples 2 and 3 we now desire to test for the consistency from one group of repetitions to the next. Then, the test given in Section 2.9 is the one we need.

From (2.9.2) we have the test statistic

$$T^{(5)} = 2(B^{(4)} - B^{(3)})\ln 10$$
$$= T^{(3)} - T^{(4)} \quad \text{, from (3.4.1),}$$
$$= 26.609 - 25.353 \quad \text{, from Examples 3 and 2,}$$
$$= 1.256.$$

From Section 3.4, under H_4 of (2.9.1), $T^{(5)}$ has approximately a χ^2-distribution with $(g - 1)(t - 1) = 4$ degrees of freedom. Therefore, the approximate significance level for Test (2.9.1) in this case is

$$P\{T^{(5)} \geq 1.256 | H_4\} = 0.86.$$

Hence, under the assumption of within-group judge consistency, we have no evidence to doubt the hypothesis that the group judging criterion is the same from one repetition to the next.

Example 5.

Suppose that in Example 1 we were only interested in testing the null hypothesis -- that the true object ratings are equal -- against the alternative that thte true ratings are in two groups, within which they are equal, but between which they are not necessarily equal. A test for this situation is given in Section 2.6. For the data given in Example 1, we wish to test the null hypothesis,

$$H_0: \quad \pi_{iu} = 1/t,$$

against the alternative hypothesis,

$$H_2: \quad \pi_{iu} = \begin{cases} \pi, & (i = 1, \ldots, s), \\ \dfrac{1 - s\pi}{t - s}, & (i = s+1, \ldots, t; \ u = 1, \ldots, v). \end{cases}$$

In our example $s = 2$, $t = 3$. To determine X, the number of times an object of the first group of $s = 2$ objects is ranked above an object of the second group of $t - s = 3$ objects, recode the objects in such a way that $a_1, \ldots, a_s$ are in the first group and $a_{s+1}, \ldots, a_t$ are in the second group. Then for judge u, $(u = 1, \ldots, v)$, construct a new preference matrix corresponding to R_u of Example 1, but this time use the notation

$$r_{iju} = \begin{cases} 1 \quad , \quad \text{if } \{a_i \longrightarrow a_j | u\} \\ 0 \quad , \quad \text{if } \{a_i \longleftarrow a_j | u\} \end{cases} ,$$

if objects a_i and a_j are compared by judge u. Then X will be the sum of the elements of the v, $(s \times \overline{t - s})$ sub-matrices contained in the upper right-hand corner of the new t x t preference matrices corresponding to each judge. For the data of our example, the 6, (2 x 3) sub-matrices are:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} , \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} , \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} , \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} .$$

Summing the elements of these, we obtain X = 15. Hence, from (2.6.6), an estimate of π is

$$p = \frac{X}{ks(t - s)^2 + (2s - t)X} = 0.385 .$$

If we consider the preference of an object in the first group over an object in the second group as a success, we have X = 15 successes in ks(t - s) = 18 trials. The probability of such an occurrence, under the hypothesis that the probability is one-half that an object of the first group is ranked over an object of the second group, is

$$P(X = 15 | H_0) = 0.0031 .$$

In addition,

$$P(X \geq 15 | H_0) = 0.0096.$$

(These probabilities have been obtained from [14]). Hence, under the assumption that the judges are consistent as a group, we would reject the null hypothesis of equality of true object ratings in favor of the alternative hypothesis H_2, at approximately the 0.01 level of significance.

CHAPTER IV

AN ANALOGUE TO KENDALL'S COEFFICIENT OF AGREEMENT FOR THE CASE OF INCOMPLETE REPETITIONS

4.1. Introduction.

M. G. Kendall and Babington Smith [12], and Kendall [10, 11], introduced a coefficient of agreement for the case where v observers each compare all possible t(t - 1)/2 pairs. In this chapter, an analagous statistic for the case where each judge compares only r pairs is constructed, and some of its implications are discussed. Under certain assumptions, it is shown to provide an alternative test for the situation described in (2.5.1).

4.2. Some definitions and notation.

Corresponding to judge u, (u = 1, ..., v), we define a (t x t) matrix

$$P_u = (p_{iju}), \tag{4.2.1}$$

where

$$p_{iju} = \begin{cases} 1 \text{ ,if } \{a_i \longrightarrow a_j | u\}, \\ \\ 0 \text{ ,if } \{a_i \longleftarrow a_j | u\}, \end{cases}$$

if objects a_i and a_j are compared by judge u, and where p_{iju} is always 0 if a_i and a_j are not compared by judge u. From the design properties exemplified by the incidence matrices (2.2.3) corresponding to each judge, we observe that

$$p_{iju} = \begin{cases} 1 \text{, or } 0, & \text{if } n_{iju} = 1, \\ \\ 0 & \text{, if } n_{iju} = 0, \end{cases}$$

and, hence, r elements of P_u will be unity and the remaining $t^2 - r$ elements will be zero. The matrix P_u will be defined to be the preference matrix corresponding to judge u.

Now we shall define a combined preference matrix P as

$$(4.2.2) \qquad P(t \times t) = \sum_u P_u\,(t \times t).$$

If we let γ_{ij}, $(i,j = 1, \ldots, t)$, denote the element in cell (i,j) of P, then

$$(4.2.3) \qquad \gamma_{ii} = 0 \ , \qquad (i = 1, \ldots, t),$$

$$(4.2.4) \qquad \gamma_{ji} = k - \gamma_{ij} \ , \qquad (i \neq j,\ i,j = 1, \ldots, t).$$

These follow from the properties of the incidence matrices, with (4.2.4) specifically following from the fact that each pair of objects is compared by exactly k judges. Also,

$$(4.2.5) \qquad \sum_{i,j} \gamma_{ij} = v\,r = \frac{kt(t-1)}{2} \ .$$

We also observe that if all the judges agree perfectly on their allocation of preferences, P will have $t(t-1)/2$ elements k and $t(t+1)2$ elements 0. Poorest agreement among the judges will be reflected in P when

$$\gamma_{ij} = \frac{k}{2}, \qquad \text{if } k \text{ is even},$$

$$= \frac{k \pm 1}{2}\ , \text{ if } k \text{ is odd},$$

$(i \neq j\ ;\ i,j = 1, \ldots, t)$.

4.3. A coefficient of agreement.

Define

$$\Sigma = \sum_{i \neq j} \binom{\gamma_{ij}}{2} . \tag{4.3.1}$$

Physically, Σ can be interpreted as the sum of the number of agreements between pairs of judges. The maximum value of Σ is $\binom{t}{2}\binom{k}{2}$, and is attained when the judges are in perfect agreement in the allocation of their preferences. The minimum value for Σ is

$$t(t-1)\binom{k/2}{2} \qquad \text{, if k is even,}$$

$$\binom{t}{2}\left\{\binom{(k+1)/2}{2} + \binom{(k-1)/2}{2}\right\} \qquad \text{, if k is odd,}$$

and is attained when the judges are in poorest agreement. Now define

$$u = \frac{2\Sigma}{\binom{t}{2}\binom{k}{2}} - 1. \tag{4.3.2}$$

Then, from the above discussion, u has a maximum value of 1 and a minimum value of

$$-\frac{1}{k-1} \qquad \text{, if k is even,}$$

$$-\frac{1}{k} \qquad \text{, if k is odd.}$$

The magnitude of u gives a measure of the agreement among the judges. When each judge compares all $t(t-1)/2$ pairs of objects, the u of (4.3.2) is equivalent to Kendall's coefficient of agreement in [10].

4.4. Tests based on Σ.

A calculated value of Σ, without any knowledge of the implications of this value, would be somewhat meaningless. We are interested in the question: If we are willing to use the Bradley-Terry model, would Σ provide a useful test for the situation outlined in (2.11.1)? In pursuing an answer to this question, we insert the following discussion:

From (4.2.4) we observe that the contribution to Σ from the pair of cells (i,j), (j,i) of P is

$$\binom{\gamma_{ij}}{2} + \binom{k - \gamma_{ij}}{2} .$$

The probability that the contribution from this pair of cells to Σ is

$$\binom{\gamma}{2} + \binom{k - \gamma}{2}$$

is equal to the probability that $\gamma_{ij} = \gamma$, which, under the null hypothesis H_1: $\pi_{iu} = \pi_i$, $(i = 1, \ldots, t;\ u = 1, \ldots, v)$, and under the assumption of probability independence between judges, is

$$(4.4.1)\quad \binom{k}{\gamma}\left(\frac{\pi_i}{\pi_i+\pi_j}\right)^{\gamma}\left(\frac{\pi_j}{\pi_i+\pi_j}\right)^{k-\gamma} = \binom{k}{\gamma}\frac{\pi_i^{\gamma}\,\pi_j^{k-\gamma}}{(\pi_i+\pi_j)^k} .$$

Hence, the probability generating function for the contribution to Σ from the above pair of cells is

$$(4.4.2)\quad f_{ij}(t) = \frac{1}{(\pi_i+\pi_j)^k}\sum_{\gamma=0}^{k}\binom{k}{k-\gamma}\pi_i^{k-\gamma}\,\pi_j^{\gamma}\,t^{\binom{k-\gamma}{2}+\binom{\gamma}{2}} .$$

Therefore, under the assumption of probability independence between pairs of objects, the probability generating function for Σ becomes

$$(4.4.3) \qquad f(t) = \prod_{i<j} f_{ij}(t) .$$

From (4.4.3) we can obtain

$$P\left\{\Sigma = x\right\} = \text{the coefficient of } t^x,$$

and

$$P\left\{\Sigma \geq x\right\} = \sum_{i \geq x} (\text{the coefficients of } t^i) .$$

Thus, the distribution of Σ under the null hypothesis H_1 depends on $\pi_1, \ldots, \pi_t$; hence, Σ will not provide an exact parameter-free test for (2.11.1).

However, as will be indicated, Σ is useful to test the null hypothesis H_0 that the preferences are allocated at random. Under H_0, the probability generating function for the contribution to Σ from the pair of cells (i,j), (j,i), which is given in (4.4.2), becomes

$$(4.4.4) \qquad g(t) = \frac{1}{2^k} \sum_{\gamma=0}^{k} \binom{k}{k-\gamma} t^{\binom{k-\gamma}{2} + \binom{\gamma}{2}} .$$

The probability generating function for Σ, under H_0, becomes

$$(4.4.5) \qquad h(t) = \left[g(t) \right]^{k(k-1)/2} .$$

The probability generating function given in (4.4.5) is identical with that for the situation where the combined preference matrix is constructed by summing the k preference matrices corresponding to k judges who compare all possible $t(t-1)/2$ pairs. This is the case considered by Kendall. Thus we may avail ourselves of his tables of the probability that a certain value of Σ will be attained or exceeded. These tables are available for parameters in the following range: $k = 3,\ t = 2, \ldots, 8$; $k = 4,\ t = 2, \ldots, 6$; $k = 5,\ t = 2, \ldots, 5$; $k = 6,\ t = 2, \ldots, 4$. Further tables, useful for some of the presently available paired comparison designs, are included in Appendix C.

An exact test ϕ of H_0 for parameters in the above range can, therefore, be defined as

$$(4.4.6) \qquad \phi(\Sigma) = \begin{cases} 1 \quad , & \text{if } \Sigma > c_\alpha \\ a \quad , & \text{if } \Sigma = c_\alpha \\ 0 \quad , & \text{if } \Sigma < c_\alpha \end{cases} ,$$

where a and c_α are such that

$$E\left\{ \phi(\Sigma) \,|H_0 \right\} = \alpha,$$

where $0 < \alpha < 1$.

For larger values of k and t, Kendall has indicated that

$$\left\{ \Sigma - \frac{1}{2}\binom{t}{2}\binom{k}{2}\frac{(k-3)}{(k-2)} \right\} \frac{4}{k-2}$$

has approximately a χ^2-distribution with

$$\nu = \frac{\binom{t}{2} k(k-1)}{(k-2)^2}$$

degrees of freedom, this approximation being reasonably good for parameters outside the above indicated range.

If we assume the v judges to be consistent as a group, (4.4.6) could be considered as a test for the equality of true object ratings as outlined in (2.5.1). It is conjectured that $\phi(\Sigma)$ is less powerful than the test based on $B^{(1)}$, say $\psi(B^{(1)})$, given in Section 2.5. A proof of this for the general case is somewhat intractable, but is supported by cases for $v = 3$, $t = 2$, and $v = 3$, $t = 3$ as indicated below:

<u>Case (i)</u> $v = 3,\ t = 2.$

Consider

$$\phi(\Sigma) = \begin{cases} 1 & , \quad \Sigma \geq 3 \\ 0 & , \quad \Sigma < 3 \end{cases}, \qquad \psi(B^{(1)}) = \begin{cases} 1 & , B^{(1)} < .602 \\ 1/6 & , B^{(1)} = .602 \\ 0 & , B^{(1)} > .602 \end{cases},$$

where $E\left\{ \phi(\Sigma) \mid H_0 \right\} = E\left\{ \psi(B^{(1)}) \mid H_0 \right\} = 0.125.$

Let $A_1 = (\pi_1 + \pi_2)(\pi_1 + \pi_3)(\pi_2 + \pi_3)$,

$$A_2 = \frac{1}{3}\left\{ \pi_2\pi_3(\pi_1^2 - \pi_2\pi_3)^2 + \pi_1\pi_2(\pi_3^2 - \pi_1\pi_2)^2 + \pi_1\pi_3(\pi_2^2 - \pi_1\pi_3)^2 \right\}:$$

then

$$E\left\{ \psi(B^{(1)}) \mid H_1: \pi_1, \pi_2, \pi_3 \right\} = \frac{1}{A_1^2}\left\{ \pi_1^4(\pi_2^2 + \pi_3^2) + \pi_2^4(\pi_1^2 + \pi_3^2) \right.$$

$$\left. + \pi_3^4(\pi_1^2 + \pi_2^2) + \frac{1}{6}\left[2\pi_1\pi_2\pi_3(\pi_1^3 + \pi_2^3 + \pi_3^3) + 2(\pi_1^3\pi_2^3 + \pi_1^3\pi_3^3 + \pi_2^3\pi_3^3) \right] \right\}$$

$$= E\left\{ \phi(\Sigma) \mid H_1: \pi_1, \pi_2, \pi_3 \right\} + A_2/A_1^2$$

$$\geq E\left\{ \phi(\Sigma) \mid H_1: \pi_1, \pi_2, \pi_3 \right\} ,$$

since $A_2 \geq 0$. Similarly we can show that, for these values of v and t, the above property obtains for tests $\phi(\Sigma)$ and $\psi(B^{(1)})$ of any equal size.

<u>Case (ii)</u> $v = 3,\ t = 3.$

Consider

$$\phi(\Sigma) = \begin{cases} 1 & , \quad \Sigma > 9 \\ 3/4 & , \quad \Sigma = 9 \\ 0 & , \quad \Sigma < 9 \end{cases} , \qquad \psi(B^{(1)}) = \begin{cases} 1 & , \ B^{(1)} \leq 0 \\ 0 & , \ B^{(1)} > 0 \end{cases} ,$$

where $E \left\{ \phi(\Sigma) \mid H_0 \right\} = E \left\{ \psi(B^{(1)}) \mid H_0 \right\} = 0.0117$. Let

$$A_3 = \frac{1}{4} \left\{ \pi_1^3(\pi_2^3 - \pi_3^3)^2 + \pi_2^3(\pi_1^3 - \pi_3^3)^2 + \pi_3^3(\pi_1^3 - \pi_2^3)^2 \right\} ;$$

then,

$$E \left\{ \psi(B^{(1)}) \mid H_1 : \pi_1, \pi_2, \pi_3 \right\} = \frac{1}{A_1^3} \left\{ \pi_1^6(\pi_2^3 + \pi_3^3) + \pi_2^6(\pi_1^3 + \pi_3^3) + \pi_3^6(\pi_1^3 + \pi_2^3) \right\}$$

$$= E \left\{ \phi(\Sigma) \mid H_1 : \pi_1, \pi_2, \pi_3 \right\} + A_3/A_1^3$$

$$\geq E \left\{ \phi(\Sigma) \mid H_1 : \pi_1, \pi_2, \pi_3 \right\} ,$$

since $A_3 \geq 0$.

APPENDIX A

FURTHER PAIRED COMPARISON DESIGNS FOR $4 \leq t \leq 25$

The following tables list the pairs of objects to be assigned to each judge for twenty-seven new paired comparison designs. For Series 2 and Series 3 designs, the pairs assigned to each judge are assigned in sets. For Series 2, $t = 2z$, and we divide the $z(2z-1)$ pairs into $2z-1$ sets of z pairs each, such that each object occurs exactly once among the pairs of a set. The $2z-1$ sets are obtained by developing, mod $(2z-1)$, the initial set

$$(1,\ 2z-2),\ (2,\ 2z-3),\ \ldots,\ (z-1,\ z),\ (0,\ \infty),$$

where $0, 1, 2, \ldots, 2z-1, \infty$ denote the $2z$ objects, and the object ∞ remains unchanged in the development. The initial sets for even values of t from 4 through 24 are tabulated in Table A1.

For Series 3 designs, $t = 2z+1$, and we divide the $z(2z+1)$ pairs into z sets of $(2z+1)$ pairs each, such that each object occurs exactly twice among the pairs of a set. The z sets are obtained by developing, mod $(2z+1)$, the second elements of each pair of the initial set

$$(0,\ 1),\ (1,\ 2),\ (2,\ 3),\ \ldots,\ (2z-1,\ 2z),\ (2z,\ 0),$$

leaving the first elements of each pair unchanged. The initial sets for odd values of t from 5 through 19 are tabulated in Table A2.

For certain Series 2 and eries 3 designs, the sets of pairs to be assigned to each judge can be obtained by developing an initial set with respect to a certain modulus. These designs are listed in Table A3, in which Roman numerals are used to denote the sets of pairs obtained from the initial set I. Series 2 and Series 3 designs which cannot be listed in this way are presented in Table A4, the sets of pairs for each judge being completely indicated. In Table A5, two Series 1 designs, which are complements of designs given in Table 1 of [1], are listed. In this case, the actual pairs of objects to be compared by each judge are presented.

TABLE A1

t	Initial Sets of Pairs	
4	(1,2),(0,∞)	mod 3
6	(1,4),(2,3),(0,∞)	mod 5
8	(1,6),(2,5),(3,4),(0,∞)	mod 7
10	(1,8),(2,7),(3,6),(4,5),(0,∞)	mod 9
12	(1,10),(2,9),(3,8),(4,7),(5,6),(0,∞)	mod 11
14	(1,12),(2,11),(3,10),(4,9),(5,8),(6,7),(0,∞)	mod 13
16	(1,14),(2,13),(3,12),(4,11),(5,10),(6,9),(7,8),(0,∞)	mod 15
18	(1,16),(2,15),(3,14),(4,13),(5,12),(6,11),(7,10),(8,9),(0,∞)	mod 17
20	(1,18),(2,17),(3,16),(4,15),(5,14),(6,13),(7,12),(8,11),(9,10),(0,∞)	mod 19
22	(1,20),(2,19),(3,18),(4,17),(5,16),(6,15),(7,14),(8,13),(9,12),(10,11),(0,∞)	mod 21
24	(1,22),(2,21),(3,20),(4,19),(5,18),(6,17),(7,16),(8,15),(9,14),(10,13),(11,12),(0,∞)	mod 23

TABLE A2

t	Initial Sets of Pairs	
5	(0,1),(1,2),(2,3),(3,4),(4,0)	mod 5
7	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,0)	mod 7
9	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,0)	mod 9
11	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,9), (9,10),(10,0)	mod 11
13	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,9), (9,10),(10,11),(11,12),(12,0)	mod 13
15	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,9), (9,10),(10,11),(11,12),(12,13),(13,14),(14,0)	mod 15
17	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,9), (9,10),(10,11),(11,12),(12,13),(13,14),(14,15),(15,16), (16,0)	mod 17
19	(0,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7),(7,8),(8,9), (9,10),(10,11),(11,12),(12,13),(13,14),(14,15),(15,16), (16,17),(17,18),(18,0)	mod 19

TABLE A3

Design number	Parameters	Sets of pairs	Sets of pairs assigned to initial judge	
1.	$t = 8$, $v = 7$ $b = 28$, $r = 24$ $k = 6$, $\lambda = 20$ $\alpha = 6$	t(8)	I,II,III,IV,V,VI	mod 7
2.	$t = 9$, $v = 4$ $b = 36$, $r = 27$ $k = 3$, $\lambda = 18$ $\alpha = 6$	t(9)	I,II,III	mod 4
3.	$t = 10$, $v = 3$ $b = 45$, $r = 30$ $k = 2$, $\lambda = 15$ $\alpha = 6$	t(10)	I,II,IV,V,VII,VIII	mod 9
4.	$t = 10$, $v = 9$ $b = 45$, $r = 40$ $k = 8$, $\lambda = 35$ $\alpha = 8$	t(10)	I,II,III,IV,V,VI,VII,VIII	mod 9
5.	$t = 11$, $v = 5$ $b = 55$, $r = 44$ $k = 4$, $\lambda = 33$ $\alpha = 8$	t(11)	I,II,III,IV	mod 5
6.	$t = 12$, $v = 11$ $b = 66$, $r = 30$ $k = 5$, $\lambda = 12$ $\alpha = 5$	t(12)	I,IV,VI,VII,VIII	mod 11
7.	$t = 12$, $v = 11$ $b = 66$, $r = 36$ $k = 6$, $\lambda = 18$ $\alpha = 6$	t(12)	II,III,V,IX,X,XI (Complement of number 6)	mod 11

TABLE A3 (continued)

Design number	Parameters	Sets of pairs	Sets of pairs assigned to initial judge	
8.	$t = 13$, $v = 3$ $b = 78$, $r = 52$ $k = 2$, $\lambda = 26$ $\alpha = 8$	t(13)	I,II,IV,V	mod 6
9.	$t = 14$, $v = 13$ $b = 91$, $r = 28$ $k = 4$, $\lambda = 7$ $\alpha = 4$	t(14)	I,V,XI,XIII	mod 13
10.	$t = 14$, $v = 13$ $b = 91$, $r = 63$ $k = 9$, $\lambda = 42$ $\alpha = 9$	t(14)	II,III,IV,VI,VII,VIII,IX,X,XII (Complement of number 9)	mod 13
11.	$t = 15$, $v = 7$ $b = 105$, $r = 45$ $k = 3$, $\lambda = 15$ $\alpha = 6$	t(15)	I,V,VII	mod 7
12.	$t = 15$, $v = 7$ $b = 105$, $r = 60$ $k = 4$, $\lambda = 30$ $\alpha = 8$	t(15)	II,III,IV,VI (Complement of number 11)	mod 7
13.	$t = 16$, $v = 15$ $b = 120$, $r = 56$ $k = 7$, $\lambda = 24$ $\alpha = 7$	t(16)	I,VI,VIII,XI,XII,XIV,XV	mod 15

TABLE A3 (continued)

Design number	Parameters	Sets of pairs	Sets of pairs assigned to initial judge
14.	t = 16, v = 15 b = 120, r = 64 k = 8, λ = 32 α = 8	t(16)	II,III,IV,V,VII,IX,X,XIII mod 15 (Complement of number 13)
15.	t = 20, v = 19 b = 190, r = 90 k = 9, λ = 40 α = 9	t(20)	I,VI,VII,X,XIII,XIV,XV,XVI, XVIII mod 19
16.	t = 20, v = 19 b = 190, r = 100 k = 10, λ = 50 α = 10	t(20)	II,III,IV,V,VIII,IX,XI,XII, XVII,XIX mod 19 (Complement of number 15)

TABLE A4

Design number	Parameters	Sets of pairs	Judge	Sets of pairs assigned to each judge
1.	t=13, v=4 b=78, r=39 k=2, λ=13 α=6	t(13)	J_1	I,III,V
			J_2	I,IV,VI
			J_3	II,III,VI
			J_4	II,IV,V

TABLE A4 (Continued)

Design number	Parameters	Sets of pairs	Judge	Sets of pairs assigned to each judge
2.	t=16,v=6	t(16)	J_1	I,IV,VII,X,XIII
	b=120,r=40		J_2	I,V,VIII,XI,XIV
	k=2,λ=8		J_3	II,IV,IX,XII,XIV
	α=5		J_4	II,VI,VII,XI,XV
			J_5	III,V,IX,X,XV
			J_6	III,VI,VIII,XII,XIII
3.	t=16,v=6	t(16)	J_1	II,III,V,VI,VIII,IX,XI,XII,XIV,XV
	b=120,r=80		J_2	II,III,IV,VI,VII,IX,X,XII,XIII,XV
	k=4,λ=48		J_3	I,III,V,VI,VII,VIII,X,XI,XIII,XV
	α=10		J_4	I,III,IV,V,VIII,IX,X,XII,XIII,XIV
			J_5	I,II,IV,VI,VII,VIII,XI,XII,XIII,XIV
			J_6	I,II,IV,V,VII,IX,X,XI,XIV,XV (Complement of number 2)
4.	t=16,v=10	t(16)	J_1	I,II,III,IV,V,VI
	b=120,r=48		J_2	I,II,VII,VIII,IX,X
	k=4,λ=16		J_3	I,III,VII,XI,XII,XIII
	α=6		J_4	I,IV,VIII,XIII,XIV,XV
			J_5	II,V,IX,XI,XIII,XIV
			J_6	II,VI,VII,XII,XIV,XV
			J_7	III,V,VIII,X,XII,XIV
			J_8	III,VI,IX,X,XIII,XV
			J_9	IV,V,VII,X,XI,XV
			J_{10}	IV,VI,VIII,IX,XI,XII

TABLE A4 (Continued)

Design number	Parameters	Sets of pairs	Judge	Sets of pairs assigned to each judge
5.	t=16,v=10 b=120,r=72 k=6,λ=40 α=9	t(16)	J_1	I,IV,VII,VIII,IX,XI,XII,XIV,XV
			J_2	I,III,IV,VI,VII,IX,X,XII,XIII
			J_3	V,VI,VIII,IX,X,XII,XIII,XIV,XV
			J_4	I,III,IV,V,VI,X,XI,XIV,XV
			J_5	I,II,III,VIII,X,XI,XII,XIII,XIV
			J_6	II,III,IV,VI,VII,VIII,XIII,XIV,XV
			J_7	II,IV,V,VII,X,XI,XII,XIII,XV
			J_8	I,II,IV,V,VI,VIII,IX,XI,XIII
			J_9	I,II,III,V,VII,VIII,IX,X,XV
			J_{10}	II,III,V,VI,VII,IX,XI,XII,XIV (Complement of number 4)
6.	t=22,v=7 b=231,r=66 k=2,λ=11 α=6	t(22)	J_1	I,II,III,IV,V,VI
			J_2	I,VII,VIII,IX,X,XI
			J_3	II,VII,XII,XIII,XIV,XV
			J_4	III,VIII,XII,XVI,XVII,XVIII
			J_5	IV,IX,XIII,XVI,XIX,XX
			J_6	V,X,XIV,XVII,XIX,XXI
			J_7	VI,XI,XV,XVIII,XX,XXI
7.	t=22,v=7 b=231,r=165 k=5,λ=110 α=15	t(22)		Complement of number 6

TABLE A4 (Continued)

Design number	Parameters	Sets of pairs	Judge	Sets of pairs assigned to each judge
8.	$t=22, v=21$ $b=231, r=55$ $k=5, \lambda=11$ $\alpha=5$	t(22)	J_1	I,V,IX,XIII,XVII
			J_2	I,VI,X,XIV,XVIII
			J_3	I,VII,XI,XV,XIX
			J_4	I,VIII,XII,XVI,XX
			J_5	II,V,X,XV,XX
			J_6	II,VI,IX,XVI,XIX
			J_7	II,VII,XII,XIII,XVIII
			J_8	II,VIII,XI,XIV,XVII
			J_9	III,V,XI,XVI,XVIII
			J_{10}	III,VI,XII,XV,XVII
			J_{11}	III,VII,IX,XIV,XX
			J_{12}	III,VIII,X,XIII,XIX
			J_{13}	IV,V,XII,XIV,XIX
			J_{14}	IV,VI,XI,XIII,XX
			J_{15}	IV,VII,X,XVI,XVII
			J_{16}	IV,VIII,IX,XV,XVIII
			J_{17}	I,II,III,IV,XXI
			J_{18}	V,VI,VII,VIII,XXI
			J_{19}	IX,X,XI,XII,XXI
			J_{20}	XIII,XIV,XV,XVI,XXI
			J_{21}	XVII,XVIII,XIX,XX,XXI

TABLE A4 (Continued)

Design Number	Parameters	Sets of pairs	Judge	Sets of pairs assigned to each judge
9.	t=22, v=21 b=231, r=176 k=16, λ=132 α=16	t(22)		Complement of number 8

TABLE A5

Design number	Parameters	Judge	Pairs assigned to each judge
1.	t=6, v=10	J_1	(1,4),(1,6),(2,6),(1,5),(2,4), (3,4),(2,5),(3,6),(3,5)
	b=15, r=9	J_2	(1,6),(4,6),(2,4),(1,3),(3,4), (2,5),(1,2),(3,5),(5,6)
	k=6, λ=5	J_3	(4,6),(1,6),(2,3),(1,5),(2,4), (3,6),(1,2),(4,5),(3,5)
	α=3	J_4	(4,6),(2,3),(1,5),(2,4),(1,3), (2,5),(3,6),(5,6),(1,4)
		J_5	(2,6),(1,5),(2,4),(1,3),(3,4), (3,6),(1,2),(4,5),(5,6)
		J_6	(1,4),(1,6),(2,3),(3,4),(2,5), (3,6),(1,2),(4,5),(5,6)
		J_7	(1,4),(1,6),(2,3),(2,6),(2,4), (1,3),(4,5),(3,5),(5,6)
		J_8	(1,4),(4,6),(2,6),(1,3),(2,5), (3,6),(1,2),(3,5),(4,5)
		J_9	(1,4),(4,6),(2,3),(2,6),(1,5), (3,4),(1,2),(3,5),(5,6)
		J_{10}	(4,6),(1,6),(2,3),(2,6),(1,5), (1,3),(3,4),(2,5),(4,5) Complement of number (iii) of (1.4.4)

TABLE A5 (Continued)

Design number	Parameters	Judge	Pairs assigned to each judge
2.	$t=9, v=28$ $b=36, r=27$ $k=21, \lambda=20$ $\alpha=6$		Complement of number (iv) of (1.4.4)

APPENDIX B

This table gives the distribution of $B^{(3)}$ defined in (2.7.11) for case t = 4, g = k = 2. P_1 denotes the probability of $B^{(3)}$ attaining the indicated value, whereas P_2 denotes the probability that the indicated value of $B^{(3)}$ will not be exceeded.

$B^{(3)}$	P_1	P_2	$B^{(3)}$	P_1	P_2
0	$.0^{4}3433$	$.0^{4}3433$	3.500	$.0^{2}9064$	.03936
0.602	$.0^{3}2060$	$.0^{3}2403$	3.563	$.0^{2}1236$	.04059
1.204	$.0^{3}3777$	$.0^{3}6180$	3.612	$.0^{3}8163$	.04141
1.498	$.0^{3}8240$	$.0^{2}1442$	3.760	$.0^{2}3502$	.04491
1.806	$.0^{3}4349$	$.0^{2}1877$	3.835	$.0^{3}9613$	.04587
2.100	$.0^{2}2472$	$.0^{2}4349$	3.857	.01483	.06070
2.359	$.0^{2}1236$	$.0^{2}5585$	3.991	.01051	.07121
2.408	$.0^{3}7210$	$.0^{2}6306$	4.102	$.0^{2}3021$	.07423
2.631	$.0^{3}9613$	$.0^{2}7267$	4.129	.01154	.08577
2.702	$.0^{3}8240$	$.0^{2}8091$	4.165	$.0^{2}4120$	.08989
2.898	$.0^{2}3021$	.01111	4.214	$.0^{2}1305$	.09119
2.961	$.0^{2}3708$	.01482	4.362	$.0^{2}1167$	.09236
2.996	$.0^{2}4944$	.01976	4.396	.03625	.1286
3.010	$.0^{3}2289$	.01999	4.437	$.0^{2}3204$	.1318
3.158	$.0^{2}1167$	.02116	4.593	$.0^{2}3502$	.1353
3.233	$.0^{2}2884$	.02404	4.656	.01401	.1493
3.304	$.0^{2}2747$	.02679	4.704	.01007	.1594
3.389	$.0^{2}3502$	.03029	4.718	.01112	.1705

$B^{(3)}$	P_1	P_2	$B^{(3)}$	P_1	P_2
4.816	$.0^{3}4349$	.1710	5.796	.06647	.5228
4.887	.04202	.2130	5.971	$.0^{2}7828$	.5306
4.964	$.0^{2}3891$	.2169	6.020	.04903	.5796
4.990	.01730	.2342	6.056	.05136	.6310
5.110	$.0^{2}5219$	.2394	6.243	$.0^{2}6088$	.6371
5.195	.01167	.2511	6.287	.1541	.7912
5.257	.05438	.3054	6.316	$.0^{2}9922$	.8011
5.262	$.0^{2}6729$	.3122	6.510	.01913	.8202
5.418	$.0^{2}1450$	.3136	6.547	.05953	.8798
5.517	.02101	.3346	6.770	$.0^{2}7393$	.8871
5.529	.04230	.3769	6.778	.08930	.9764
5.748	.06303	.4400	7.001	.02218	.9986
5.789	.01634	.4563	7.224	$.0^{2}1377$	1.0000

APPENDIX C

In the following table, the probability, P, that a value Σ will be attained or exceeded is tabulated for k = 2, t = 4, ..., 10; k = 4, t = 8.

k = 2, t = 4		k = 2, t = 5		k = 2, t = 6	
Σ	P	Σ	P	Σ	P
0	1.0000	0	1.0000	0	1.0000
1	.9844	1	.9990	1	.9999
2	.9063	2	.9893	2	.9995
3	.6565	3	.9453	3	.9963
4	.3438	4	.8281	4	.9824
5	.1094	5	.6230	5	.9408
6	.01563	6	.3770	6	.8491
		7	.1719	7	.6964
		8	.05469	8	.5000
		9	.01074	9	.3036
		10	$.0^{2}9766$	10	.1509
				11	.05923
				12	.01758
				13	$.0^{2}3693$
				14	$.0^{3}4883$
				15	$.0^{4}3052$

k = 2, t = 7		k = 2, t = 8		k = 2, t = 9			
Σ	P	Σ	P	Σ	P	Σ	P
3	.9998	5	.9999	8	.9998	23	.06625
4	.9993	6	.9995	9	.9994	24	.03262
5	.9964	7	.9981	10	.9980	25	.01441
6	.9867	8	.9937	11	.9943	26	$.0^{2}5666$
7	.9608	9	.9822	12	.9856	27	$.0^{2}1967$
8	.9054	10	.9564	13	.9674	28	$.0^{3}5966$
9	.8083	11	.9075	14	.9338	29	$.0^{3}1563$
10	.6682	12	.8275	15	.8785	30	$.0^{4}3480$
11	.5000	13	.7142	16	.7975	31	$.0^{5}6457$
12	.3318	14	.5747	17	.6911	32	$.0^{6}9708$
13	.1917	15	.4253	18	.5660	33	$.0^{6}1136$
14	.09462	16	.2858	19	.4340	34	$.0^{8}9706$
15	.03918	17	.1725	20	.3089	35	$.0^{9}5384$
16	.01330	18	.09247	21	.2025	36	$.0^{10}1455$
17	$.0^{2}3600$	19	.04358	22	.1215		
18	$.0^{3}7448$	20	.01785				
19	$.0^{3}1106$	21	$.0^{2}6270$				
20	$.0^{4}1049$	22	$.0^{2}1860$				
21	$.0^{6}4768$	23	$.0^{3}4561$				
		24	$.0^{4}9000$				
		25	$.0^{4}1372$				
		26	$.0^{5}1516$				
		27	$.0^{6}1080$				
		28	$.0^{8}3725$				

k = 2, t = 10				k = 4, t = 8			
Σ	P	Σ	P	Σ	P	Σ	P
11	.9999	29	.03623	94	.0780	110	$.0^{3}22$
12	.9996	30	.01785	95	.0599	111	$.0^{3}16$
13	.9988	31	$.0^{2}8047$	96	.0454	112	$.0^{3}10$
14	.9967	32	$.0^{2}3304$	97	.0339	113	$.0^{4}65$
15	.9920	33	$.0^{2}1229$	98	.0249	114	$.0^{4}40$
16	.9822	34	$.0^{3}4120$	99	.0181	115	$.0^{4}24$
17	.9638	35	$.0^{3}1235$	100	.0129	120	$.0^{5}16$
18	.9324	36	$.0^{4}3287$	101	.0090	125	$.0^{7}83$
19	.8837	37	$.0^{5}7687$	102	.0061	130	$.0^{8}29$
20	.8144	38	$.0^{5}1561$	103	.0041	135	$.0^{10}80$
21	.7243	39	$.0^{6}2709$	104	.0026	140	$.0^{11}15$
22	.6170	40	$.0^{7}3939$	105	.0016	145	$.0^{13}20$
23	.5000	41	$.0^{8}4667$	106	.0010	150	$.0^{15}16$
24	.3830	42	$.0^{9}4327$	107	$.0^{3}65$	155	$.0^{17}11$
25	.2757	43	$.0^{10}2944$	108	$.0^{3}44$	160	$.0^{21}97$
26	.1856	44	$.0^{11}1307$	109	$.0^{3}31$	168	$.0^{25}52$
27	.1163	45	$.0^{13}2842$				
28	.06758						

BIBLIOGRAPHY

[1] Bose, R. C., "Paired Comparison Designs for Testing Concordance Between Judges," to appear in Biometrika, 43 (1956).

[2] Bradley, R. A. and Terry, M. E., "The Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons," Biometrika, 39 (1952), 324-345.

[3] Bradley, R. A., "The Rank Analysis of Incomplete Block Designs. II. Additional Tables for the Method of Paired Comparisons," Biometrika, 41 (1954), 502-537.

[4] Bradley, R. A., "The Rank Analysis of Incomplete Block Designs. III. Some Large-Sample Results on Estimation and Power for a Method of Paired Comparisons," Biometrika, 42 (1955), 450-470.

[5] Bradley, R. A., "Incomplete Block Rank Analysis: On Appropriateness of the Model for a Method of Paired Comparisons," Biometrics, 10 (1954), 375-390.

[6] Cochran, W. G. and Cox, G. M., Experimental Designs, John Wiley and Sons, New York, 1950.

[7] Dykstra, Otto, Jr., "A Note on the Rank Analysis of Incomplete Block Designs; Applications Beyond the Scope of Existing Tables," Unpublished manuscript, 1956.

[8] Fisher, R. A. and Yates, F., Statistical Tables for Biological, Agricultural and Medical Research, Hafner Publishing Company Inc., New York, 1949.

[9] Kendall, M. G., "Further Contributions to the Theory of Paired Comparisons," Biometrics, 11 (1955), 43-62.

[10] Kendall, M. G., Advanced Theory of Statistics, Vol. I, Charles Griffin and Company Limited, London, 1952.

[11] Kendall, M. G., Rank Correlation Methods, Charles Griffin and Company Limited, London, 1948.

[12] Kendall, M. G. and Babington Smith, B., "On the Method of Paired Comparisons," Biometrika, 31 (1940), 324-345.

[13] Mosteller, Frederick, "Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations," Psychometrika, 16 (1951), 3-9.

[14] _Tables of the Binomial Probability Distribution_, Department of Commerce, National Bureau of Standards, Applied Mathematics Series 6, Washington, D. C., 1950.

[15] Terry, M. E., Bradley, R. A. and Davis, L. L., "New Designs and Techniques for Organoleptic Testing," _Food Technology_, 6 (1952), No. 7, 250-254.

[16] Thurstone, L. L., "Psychophysical Analysis," _American Journal of Psychology_, 38 (1927), 368-389.