THE RANK ANALYSIS OF TRIPLE COMPARISONS

by

Robert Nixon Pendergrass, B.S.Ed., M.Ed., M.A.

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute
in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS

APPROVED:

Chairman, Advisory Committee

August 1957

Blacksburg, Virginia

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	5
	1.1 Methods of Rank Analysis	5
	1.2 The Bradley-Terry Method of Paired Comparisons	11
	1.3 Extension to Triple Comparisons	14
II.	MATHEMATICAL MODELS	16
	2.1 Experimental Procedure and Notation . . .	16
	2.2 The Bradley-Terry Model for Triple Comparisons	17
	2.3 A New Model for Triple Comparisons	20
	2.4 The Impossibility of Complete Independence	24
	2.5 Models for Blocks of Size Four and Five .	25
	2.6 Comparison of Models	28
III.	LIKELIHOOD-RATIO TESTS AND ESTIMATION	33
	3.1 The Likelihood Function	33
	3.2 Estimation of Parameters	36
	3.2.1 Solution of the Maximum-Likelihood Equations, Case (i)	39
	3.3 Likelihood-Ratio Tests	41
	3.3.1 The Generation of Tables	43
	3.3.2 Dominant Subsets of Sums of Ranks .	45
	3.4 Large-Sample Tests	48
	3.5 The Combination of Experiments	52
	3.6 A Coefficient of Agreement	53

Chapter		Page
IV.	A TEST OF THE MODEL FOR TRIPLE COMPARISONS . . .	56
	4.1 A Goodness-of-Fit Test of the Appropriateness of the Model	56
	4.2 Goodness-of-Fit Test Associated With Expected Frequencies	60
V.	SOME LARGE-SAMPLE RESULTS ON ESTIMATION AND POWER	64
	5.1 Introduction	64
	5.2 Distribution of Estimators for Case (i) . .	64
	5.3 Distribution of the Estimator for Case (ii)	74
	5.4 Large-Sample Distribution of $T = -2 \ln \lambda_1$, Case (i).	76
	5.4.1 Series Expansion of T	77
	5.4.2 Limiting Distribution of $\sqrt{n}\, y_i$ . . .	80
	5.4.3 Limiting Distribution of T under H_a	86
VI.	COMPARISONS OF POWERS OF TESTS	90
	6.1 Comparison of the Test for Triple Comparisons with the Test for Paired Comparisons .	90
	6.2 Efficiency of the Triple Comparisons Test Relative to Analysis of Variance	94
	6.3 Triple Comparisons Compared with a Multi-Multinomial Test	97
VII.	NUMERICAL EXAMPLES	104
	7.1 Estimation and Test of Significance Illustrated	104
	7.2 Goodness-of-Fit Test.	107
	7.3 Estimation of Variances and Covariances . .	108
	7.4 A Consumer Acceptance Study of Food Variants	110

Chapter	Page
VIII. SUMMARY	113
BIBLIOGRAPHY	116
APPENDICES	121
A. The Derivation of the Maximum-Likelihood Equations	121
B. Limiting Distribution of $\sqrt{n}\ y_i (i=1,\ldots,t)$	124
ACKNOWLEDGEMENTS	127
VITA	128

I. INTRODUCTION

1.1 Methods of Rank Analysis

When quantitative measurement of treatment effects is not possible, or perhaps is not practical, methods of rank analysis are frequently used to analyze the results of experiments. In experiments involving rank analysis, information concerning treatment effects is usually obtained by having observers rank the objects under consideration according to some trait or quality. Problems of statistical analysis based on ranks, or sums of ranks, arise in connection with experimental designs paralleling those used in the analysis of variance. These problems have been treated rather thoroughly for complete block designs, but only in the case of paired comparisons has much attention been given to the rank analysis of incomplete block designs. In this section, we shall discuss briefly some of the procedures that are now used in rank analysis. In the two following sections, we shall describe an important method of paired comparisons and discuss the need for extensions to the analysis of rankings of groups of size three.

A method of rank analysis involving transformations of the observed ranks through the use of scores for ordinal (or ranked) data was proposed by R. A. Fisher and Frank Yates (1953).[1] To apply this method, each given rank is transformed into the expected value of an observation with the same rank in a sample of independent observations of the same size from a unit normal population. Standard methods of analysis of variance are then applied to the transformed ranks. Although the efficacy of this procedure has as yet not been completely determined, it has been used extensively. The works of C. I. Bliss and co-workers (1943), W. C. Galinat and H. L. Everett (1949), and M. L. Greerwood and R. Salerno (1949), are typical applications of this method.

The problem of analysing the rankings of n objects by m different observers was discussed by Milton Friedman (1937). M. G. Kendall and B. Babington Smith (1939) treated the same topic and proposed the well-known coefficient of concordance. According to Kendall (1955, p. 95), this coefficient "measures, in a sense, the communality of judgements for the m observers." An observed coefficient of concordance may therefore be used to test

1. Dates in parenthesesidentify references that are listed in the bibliography. When page numbers are needed, they follow the date in parentheses.

the hypothesis that the observers have no agreement of preferences.

Kendall's coefficient of concordance was defined by the following equation:[2]

(1.1) $W = 12S/m^2(n^3 - n)$, where

m is the number of rankings, n is the number of objects ranked, and S is the observed sum of squares of the deviations of sums of ranks from their mean value $m(n+1)/2$. J. Durbin (1951) generalized this coefficient of concordance for use with balanced incomplete block designs in which the n objects are considered in blocks of size k; and, in the experiment as a whole, each is ranked m times. Using λ to indicate the number of blocks in which a particular pair of objects occurs, he defined W as follows:

(1.2) $W = 12S/\lambda^2(n^3 - n)$.

This is the ratio of observed S to maximum S, as was also the case for W in (1.1); and S retains its definition as the observed sum of squares of the deviations from the mean. An approximate test of the significance

2. An equation number (c,b) refers to equation b in chapter c.

of W may be obtained by considering $\lambda(n^2-1)W/(k+1)$ to be distributed as chi-square with $(n-1)$ degrees of freedom. Large values of W indicate concordant rankings and distinguishability of treatments or items.

The analysis of experiments involving paired comparisons has received considerable attention since publication of a paper by L. L. Thurstone (1927) on this subject. Some of the subsequent studies that have been conducted have led to generalizations of the method of Thurstone, and other investigations have led to entirely different procedures. One important extension of the work of Thurstone was made by F. Mosteller (1951 a, b, c); and, as a result, many writers now refer to the procedures as the Thurstone-Mosteller Method.

In the Thurstone Model for paired comparisons, as given by Mosteller, the responses X_i $(i = 1,\ldots,t)$ for each of t stimuli to be compared are assumed to be jointly normally distributed with means S_i and with variances of the differences $(X_i - X_j)$ equal to σ_d^2, $i \neq j$. The probability that the response to stimulus i is greater than the response to stimulus j (stimulus i is preferred to stimulus j, $X_i > X_j$) is given by the relation

$$(1.3)\quad p_{ij} = P(X_i > X_j) = (2\pi)^{-\frac{1}{2}} \int_{(S_i - S_j)}^{\infty} \exp(-y^2/2)\,dy.$$

Let c_{ij} indicate the observed proportional preference for stimulus i to stimulus j. From the c_{ij}, Mosteller obtained least squares solutions S_i' for the means S_i. He then obtained estimates p_{ij}' for p_{ij}. He proposed a goodness-of-fit test, essentially a test of the appropriateness of the assumed model, by setting

$$o_{ij} = \text{arc sin } c_{ij},$$

$$o_{ij}' = \text{arc sin } p_{ij}' \qquad (i,j = 1,\ldots t;\ i \neq j),$$

and then calculating

$$(1.4)\quad \chi^2 = \sum_{i=1}^{t-1} \sum_{\substack{j=1 \\ j\neq i}}^{t} \frac{n(o_{ij}' - o_{ij})^2}{821}.$$

The statistic in (1.4) is approximately distributed as chi-square with $(t-1)(t-2)/2$ degrees of freedom for large sample sizes n.

A second method of analysis of paired comparisons, and one with a somewhat different objective, involving a combinatorial-type test was developed by Kendall and

Babington Smith (1940). They used the number of circular triads[3] to calculate a "coefficient of consistency," a measure of the consistency of an individual observer. They also developed a "coefficient of agreement," a form of the concordance coefficient, which is a measure of agreement among observers. If we let γ_{ij} be the number of comparisons in which T_i is preferred to T_j and

$$\Sigma = \sum_{i=1}^{t} \sum_{\substack{j=1 \\ i \neq j}}^{t} \binom{\gamma_{ij}}{2},$$ this coefficient of agreement may be written as

$$(1.5) \quad U = \frac{2\Sigma}{\binom{m}{2}\binom{t}{2}} - 1,$$ where m is

the number of judges. The statistic U has a maximum of unity and a minimum of negative one. The maximum is attained only when there is complete agreement among judges, and the minimum is attained only when $m = 2$ and the two judges disagree on each decision. Kendall and Babington Smith suggested tests for this coefficient of agreement and indicated how it can be used to determine

3. A circular triad is illustrated by the following series of judgments: T_i preferred to T_j, T_j preferred to T_n, and T_n preferred to T_i, where T_i indicates the ith treatment or item considered.

if the quality under investigation can be associated with an underlying ordering scale.

1.2 The Bradley-Terry Method of Paired Comparisons

Recent research by R. A. Bradley and M. E. Terry (1952a) resulted in the development of a new method of analysis of paired comparisons that has received considerable attention in the literature of statistics and has been used extensively in sensory difference tests. They formulated a mathematical model as a generalization of the binomial model and used likelihood-ratio statistics to test specified hypotheses. According to Bradley and Terry (1952a, p. 324), "Although these tests basically agree with those of Kendall and Babington Smith, they subdivide the possible results from an experiment of a given size into more distinct subclasses, thus perhaps indicating better sensitivity."

Bradley and Terry proposed a mathematical model for a paired comparisons experiment by first assuming the existence of true treatment ratings, or real parameters, $\pi_1,\ldots,\pi_t$ for the t treatments, or items, involved in the experiment. Then, they assumed that these parameters satisfy the following conditions:

(i) $\pi_i \geq 0 \quad (i=1,\ldots,t)$.

(ii) $\sum_{i=1}^{t} \pi_i = 1$.

(iii) When treatment i is compared with treatment j, the probability that treatment i obtains the preferred rating (or a rank of one) is $\pi_i/(\pi_i + \pi_j)$.

The problem of quantifying the results of a paired comparisons experiment has been discussed by Louis Guttman (1946) and others. Guttman proposed a least squares method to obtain representative scale values for each of the stimuli being compared. Inspection of the Bradley-Terry model for paired comparisons proposed in the preceding paragraph shows that the estimation of the parameters defined there provides a simple solution to this problem of quantification. Using p_i to represent the estimate of π_i, Bradley and Terry (1952a, p. 326) state that "If the estimates are converted to logarithms, the values log p_i occur on a linear scale and permit over-all comparisons of experimental treatments."

An interesting evaluation of statistical techniques currently used in the analysis of paired comparisons was recently made by J. Edward Jackson and Mary Fleckenstein (1957). They discussed the work of M. G. Kendall on paired comparisons and illustrated results for methods developed by the following authors:

(i) Thurstone - Mosteller

(ii) Bradley - Terry

(iii) Scheffe (1952), and

(iv) Morrissey (1955) - Gulliksen (1956).

The latter two methods have not been discussed above. Scheffe does not use ranking but rather an analysis of variance technique with subjective scores. Morrissey and Gulliksen independently developed a technique to provide least squares solutions of the method of Thurstone for the case in which not all comparisons are made.

Using each of the four methods listed above, Jackson and Fleckenstein computed a response scale, or a representative scale value for each of the stimuli compared. These four response scales were found to be almost identical for the experiment which they considered. Some advantages and disadvantages were noted for each method. One of the conclusions reached was that, with reference to their investigations, the methods of Bradley and Terry are the most effective of all for a complete analysis of a paired comparisons experiment.

1.3 Extension to Triple Comparisons

In contrast to the widespread attention given to methods of paired comparisons, very little appears to have been done toward the development of comparable methods of triple or multiple comparisons that involve the rank analysis of incomplete block designs for blocks of size three or more. Bradley and Terry indicated that the mathematical model and test procedures which they developed for paired comparisons could be extended to the rank analysis of incomplete block designs for larger block sizes. In fact, in a preliminary report (Bradley and Terry, 1952b), they developed some of the theory involved with a particular model for triple comparisons. The authors did not present this paper for publication; and, from a survey of statistical literature, it appears that very little has been published on the subject of the analysis of experiments involving incomplete blocks of size three (or more) when only ordering, or ranking, is considered (An exception is the work of Durbin discussed above.).

The main purpose of this study is to consider methods of triple comparisons, or the rank analysis of incomplete block designs for blocks of size three. General extensions of the probability model for paired comparisons developed by Bradley and Terry are considered first. One of the new mathematical models proposed is then used to develop methods of estimation and test procedures for triple comparisons analogous to those developed by Bradley and Terry (1952a) and later by Bradley (1954a, b, 1955) for paired comparisons. Finally, the local asymptotic power of triple comparisons is compared with that of paired comparisons and with the analysis of variance. The procedures developed are illustrated by using data from a consumer acceptance study of food variants.

II. MATHEMATICAL MODELS

2.1 Experimental Procedure and Notation

Throughout this paper an effort will be made to keep the notation consistent with that introduced by Bradley and Terry for paired comparisons. Where modification of their symbolism is necessary, the new notation will generally be a logical generalization of the notation used in their papers.

Let t represent the number of treatments, or items, under consideration in an experiment. A complete repetition is now defined to be the set of all possible distinct blocks of size m that can be formed with the t treatments. If $t \geq m$, there are $\binom{t}{m}$ blocks in one complete repetition. The number of complete repetitions involved in the experiment is indicated by n.

The experiment is performed by having the items in each block ranked within that block according to some trait. The data thus obtained consist of a series of rankings. In general, the rank given to treatment i will be indicated by r_i. Additional subscripts may be used as needed to indicate the block, or the block and repetition in which the ranking occurred. For example, $r_{ik:jh}$ will denote the rank of treatment i in the kth repetition of the block containing treatments i, j, and h. A rank of one will be assigned the treatment most preferred in a given block. In blocks of size two then, $r_{i;j} = 1$ if treatment i is preferred to treatment j and $r_{i;j} = 2$ if treatment j is preferred to treatment i. When more than two items are ranked in a block, ranks one, two, three,..., are assigned in order of preference from the most preferred to the least preferred item.

2.2 The Bradley-Terry Model for Triple Comparisons

Reference was made in Section 1.3 to a preliminary report by Bradley and Terry (1952b) concerning the rank analysis of experimental designs in blocks of size three. In this paper they proposed a mathematical model by again postulating the existence of true treatment parameters, $\pi_1,\ldots,\pi_t$, which satisfy requirements (i) and (ii) of Section 1.2. They further assumed that, in a block containing treatments i, j, and h, the probability that treatment i obtains top rating (or a rank of one) is $\pi_i/(\pi_i + \pi_j + \pi_h)$ and, when this has occurred, it is assumed that, in the competition between treatments h and j for second place, the probability that treatment j obtains rank two is $\pi_j/(\pi_j + \pi_h)$. In terms of symbols previously defined, the resulting probability statement is

$$(2.1) \quad P(r_i < r_j < r_h) = \frac{\pi_i\pi_j}{(\pi_i + \pi_j + \pi_h)(\pi_j + \pi_h)} \, .$$

The probability model given in (2.1) has the desirable property that the unconditional probability that treatment i outranks treatment j is the same as in paired comparisons. This may be verified by noting that from (2.1)

$$P(r_i<r_j) = P(r_i<r_j<r_h) + P(r_i<r_h<r_j) + P(r_h<r_i<r_j),$$

$$(2.2) \quad = \frac{\pi_i\pi_j}{(\pi_i+\pi_j+\pi_h)(\pi_j+\pi_h)} + \frac{\pi_i\pi_h}{(\pi_i+\pi_j+\pi_h)(\pi_h+\pi_j)} + \frac{\pi_h\pi_i}{(\pi_i+\pi_j+\pi_h)(\pi_i+\pi_j)} \, ,$$

$$= \pi_i/(\pi_i + \pi_j).$$

Bradley and Terry showed that (2.1) does provide a workable model with which estimation and test procedures analogous to those of paired comparisons can be developed. However, they did not develop extensively the techniques of rank analysis for blocks of size three, nor did they explore thoroughly the properties of tests based on their model. This was due, partially at least, to the fact that the model proposed was not reversible in the sense that the probabilities expressed in (2.1) do not retain their form when the direction of ranking is reversed. The direction of ranking is reversed if, instead of assigning ranks 1, 2, 3 in that order from "best" to "worst", the ranks are assigned by first giving a rank of unity to the "worst" item and then ranking the remaining two items so that the "best" one now has a rank of three. Let ρ_i indicate the rank of item i in a ranking that is the reverse of one in which ranks are indicated by r_i. A probability model is reversible in the sense used here, if

$$P(r_i<r_j<r_h) = P(\rho_h<\rho_j<\rho_i)$$

and the probabilities are of the same form. In other words, if

$$\begin{aligned} &P(r_i<r_j<r_h) = f(\pi_i,\pi_j,\pi_h), \\ (2.3)\quad &P(\rho_i<\rho_j<\rho_h) = f(\beta_i,\beta_j,\beta_h), \text{ and} \\ &f(\pi_i,\pi_j,\pi_h) = f(\beta_h,\beta_j,\beta_i), \quad (i,j,h=1,\ldots,t) \end{aligned}$$

then the model is reversible. In terms of the Bradley-Terry Model, we have

$$P(r_i<r_j<r_h) = \frac{\pi_i\pi_j}{(\pi_i+\pi_j+\pi_h)(\pi_j+\pi_h)} \quad (i,j,h=1,\ldots,t)$$

and for reversibility we need to find β's such that

$$P(\rho_h<\rho_j<\rho_i) = \frac{\beta_h\beta_j}{(\beta_h+\beta_j+\beta_i)(\beta_j+\beta_i)} \quad (i,j,h=1,\ldots,t).$$

This model is not reversible since there is, in general, no set of β's such that

$$\frac{\beta_h \beta_j}{(\beta_i+\beta_j+\beta_h)(\beta_j+\beta_i)} = \frac{\pi_i \pi_j}{(\pi_i+\pi_j+\pi_h)(\pi_j+\pi_h)} \quad (i,j,h=1,\ldots,t).$$

This can be seen by considering the case of $t = 3$ in which $\pi_1 = 1/2$, $\pi_2 = 1/3$, and $\pi_3 = 1/6$. There is no corresponding set of β's such that the reversibility conditions are satisfied.

2.3 A New Model for Triple Comparisons

In the paragraph preceding Equation (2.1), it is implied that in ranking three items using the model given there, the item to be assigned the rank one is first selected and then, from the other two, the item to be given rank two is chosen. Another method of selection is to choose the item to be given rank three first and then select the item to be assigned rank two. We would like to have a probability model for triple comparisons which is unaffected by the order of assignment of ranks. Let $P(r_i<r_j<r_h|1)$ represent the probability of a given ranking in which the rank one is assigned first and let $P(r_i<r_j<r_h|3)$ represent the probability of a given ranking in which the rank three is assigned first. In terms

of these probabilities, we would like a model such that

$$(2.4) \qquad P(r_i<r_j<r_h|1) = P(r_i<r_j<r_h|3).$$

In subjective rankings of three items, an observer may find it very easy in some comparisons to select the "best" item first while in other comparisons it is easier to select the "worst" item first. This indicates the desirability of permitting the observer to assign ranks in either order, depending on the given comparison. If this is to be the case, there is a probability, say P(1), that the rank one is assigned first and a probability, P(3), that the rank three is assigned first. The probability of a given ranking under these conditions is given by

$$(2.5) \qquad P(r_i<r_j<r_h) = P(1)P(r_i<r_j<r_h|1) + P(3)P(r_i<r_j<r_h|3).$$

Since $P(1) + P(3) = 1$, condition (2.4), namely,

$$P(r_i<r_j<r_h|1) = P(r_i<r_j<r_h|3),$$

leads to

$$(2.6) \qquad P(r_i<r_j<r_h) = P(r_i<r_j<r_h|1) = P(r_i<r_j<r_h|3).$$

This indicates that a model which satisfies (2.4) also provides the same probability for (2.5) as given in (2.6). Hence, as we develop such a model, we shall discontinue the notation that indicates a particular order of assignment of ranks.

In order to derive a model for triple comparisons which satisfies (2.4), we again assume the existence of true treatment ratings $\pi_1, \ldots, \pi_t$ such that every $\pi_i \geq 0$ and $\sum_{i=1}^{t} {}_i = 1$. We also assume that after the selection of an extreme treatment in a given block, the comparison of the other two reverts to a paired comparison. That is,

$$(2.7) \quad P(r_i<r_j<r_h|1) = P(r_i=1)P(r_j<r_h|r_i=1),$$

$$(2.8) \quad P(r_i<r_j<r_h|3) = P(r_h=3)P(r_i<r_j|r_h=3), \text{ and}$$

$$(2.9) \quad P(r_i<r_j|r_h=3) = \pi_i/(\pi_i+\pi_j) = P(r_i<r_j|r_h=1).$$

To satisfy (2.4), we furthermore make an equivalent assumption that

$$(2.10) \quad P(r_i<r_j<r_h) = P(r_i=1)P(r_j<r_h|r_i=1)$$
$$= P(r_h=3)P(r_i<r_j|r_h=3).$$

A probability model satisfying these assumptions may be derived from a consideration first of

$$P(r_i<r_j<r_h) = P(r_i=1)P(r_j<r_h|r_i=1), \text{ or}$$

$$(2.11) \quad P(r_i<r_j<r_h) = [P(r_i=1)][\pi_i>(\pi_j+\pi_h)].$$

Then, using (2.9) and (2.10), it is easy to see that

$$(2.12) \quad P(r_i=1) = [P(r_h=3)][\pi_i/(\pi_i+\pi_j)]$$
$$+ [P(r_j=3)][\pi_i/(\pi_i+\pi_h)] \text{ and}$$

$$(2.13) \quad P(r_h=3) = [P(r_i=1)][\pi_j/(\pi_j+\pi_h)]$$
$$+ [P(r_j=1)][\pi_i/(\pi_i+\pi_j)].$$

Substituting for $P(r_h=3)$ and $P(r_j=3)$ in (2.7) and simplifying, we obtain

(2.14)

$$(\pi_j^2\pi_i+\pi_j^2\pi_h+\pi_h^2\pi_i+\pi_h^2\pi_j)P(r_i=1) = \pi_i^2(\pi_j+\pi_h)[P(r_j=1)+P(r_h=1)]$$
$$= \pi_i^2(\pi_j+\pi_h)[1-P(r_i=1)] \quad .$$

From (2.14), we get

$$(2.15) \quad P(r_i=1) = \frac{\pi_i^2(\pi_j+\pi_h)}{\pi_i^2(\pi_j+\pi_h) + \pi_j^2(\pi_i+\pi_h) + \pi_h^2(\pi_i+\pi_j)} \, .$$

Substitution of (2.15) in (2.11) yields

$$(2.16) \quad P(r_i<r_j<r_h) = \frac{\pi_i^2\pi_j}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)} \, .$$

The preceding derivation indicates that (2.16) is a general extension of the Bradley-Terry Model for paired comparisons to triple comparisons that satisfies the conditional independence of extreme rankings as assumed in (2.9). It is also independent of the order of assignment of ranks and may be used to represent the probability of the given ranking when the observer is permitted to assign either the rank one or three first.

The probability statement in (2.16) was first suggested by D. R. Cox (in correspondence with R. A. Bradley). He noted that the Bradley-Terry Model for triple comparisons was not reversible in the sense defined and proposed (2.16) as a reversible model. That this is true may be easily verified by dividing the numerator and denominator of the right member of (2.16) by $\pi_i\pi_j\pi_h$. This results in

$$(2.17)\quad P(r_i<r_j<r_h) = \frac{\frac{\pi_i}{\pi_h}}{\frac{\pi_i}{\pi_h}+\frac{\pi_i}{\pi_j} + \frac{\pi_j}{\pi_i}+\frac{\pi_j}{\pi_h} + \frac{\pi_h}{\pi_i}+\frac{\pi_h}{\pi_j}}.$$

Letting $\frac{\pi_i}{\pi_j} = \frac{\beta_j}{\beta_i}$, we find after substitution in (2.17) that

$$(2.18)\quad P(r_i<r_j<r_h) = \frac{\beta_h^2\beta_j}{\beta_h^2(\beta_j+\beta_i)+\beta_j^2(\beta_h+\beta_i)+\beta_i^2(\beta_j+\beta_h)}.$$

This indicates that if the π_i exist and are not zero so that the ratios used are meaningful, then the β_i exist as parameters which have the same relation to rankings in the reverse direction that the π_i have to the r_i of the original ranking.

2.4 The Impossibility of Complete Independence

The probability model proposed in (2.16) was derived by requiring the conditional independence of the rank of a pair of items when the third item in the block considered is an extreme. The question naturally arises as to the possibility of having the probability of a given ranking of a pair of items completely independent of the rank of the third item in the block. This would require

$$(2.19)\quad P(r_i<r_j|r_h=1)=P(r_i<r_j|r_h=2)=P(r_i<r_j|r_h=3), \text{ or}$$

$$(2.20)\quad \frac{P(r_i<r_j<r_h)}{P(r_j<r_i<r_h)} = \frac{P(r_i<r_h<r_j)}{P(r_j<r_h<r_i)} = \frac{P(r_h<r_i<r_j)}{P(r_h<r_j<r_i)} .$$

To investigate the possibility of a model satisfying (2.19), let $P(r_i<r_j<r_h) = a$, $P(r_j<r_i<r_h) = b$, $P(r_i<r_h<r_j) = c$, $P(r_j<r_h<r_i) = d$, $P(r_h<r_i<r_j) = e$, and $P(r_h<r_j<r_i) = f$. Now from the general relation in (2.20) we may write

$$(2.21)\quad \text{(i)}\quad a/b = c/d = e/f$$

$$\text{(ii)}\quad a/c = b/e = d/f$$

$$\text{(iii)}\quad a/f = b/d = c/e$$

Note that in (i) $a/b = c/d$, in (ii) $a/b = c/e$, and in (iii) $a/b = f/d$. It follows that $d = e$ and $f = c$. This permits us to write, from (i), $c/d = d/c$, which implies that $c = d$, since the letters represent probabilities. Extending this argument, we see that $c = d = f = e = a = b$. Consequently, the requirement of complete independence as stated in (2.19) can hold in general only if all probabilities are equal. This indicates that there can be no general model satisfying (2.19).

2.5 Models for Blocks of Size Four and Five

A second derivation of (2.16) may be found by noticing that the probability $P(r_i<r_j<r_h)$ involves three paired

comparisons in which $r_i<r_j,<r_i<r_h$, and $r_j<r_h$. Hence, it is reasonable to consider

$$(2.22)\quad P(r_i<r_j<r_h) = C\cdot P(r_i<r_j)P(r_i<r_h)P(r_j<r_h),$$

where $P(r_i<r_j) = \pi_i/(\pi_i+\pi_j)$. This gives

$$(2.23)\quad P(r_i<r_j<r_h) = C\,\frac{\pi_i^2\pi_j}{(\pi_i+\pi_j)(\pi_j+\pi_h)(\pi_i+\pi_h)}.$$

Since
$$P(r_i<r_j<r_h) + P(r_j<r_h<r_i) + P(r_h<r_i<r_j) + P(r_h<r_j<r_i) + P(r_j<r_i<r_h) + P(r_i<r_h<r_j) = 1,$$

we can easily determine that

$$C = \frac{(\pi_i+\pi_j)(\pi_j+\pi_h)(\pi_i+\pi_h)}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)}, \quad \text{and}$$

$$(2.24)\quad P(r_i<r_j<r_h) = \frac{\pi_i^2\pi_j}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)}$$

as in (2.16) .

A probability model for blocks of size four may be obtained by noticing that $P(r_i<r_j<r_h<r_m)$ involves six paired comparisons in which $r_i<r_j$, $r_i<r_h$, $r_i<r_m$, $r_j<r_h$, $r_j<r_m$, and $r_h<r_m$. Consequently, we may consider that

$$(2.25)\quad P(r_i<r_j<r_h<r_m) = C\,\frac{\pi_i}{\pi_i+\pi_j}\cdot\frac{\pi_i}{\pi_i+\pi_h}\cdot\frac{\pi_i}{\pi_i+\pi_m}\cdot\frac{\pi_j}{\pi_j+\pi_h}\cdot\frac{\pi_j}{\pi_j+\pi_m}\cdot\frac{\pi_h}{\pi_h+\pi_m}.$$

Since the sum of the probabilities over the 24 permutations of i,j,h, and m is equal to unity, we may determine C and find that

$$(2.26)\quad P(r_i<r_j<r_h<r_m) = \frac{\pi_i^3\pi_j^2\pi_h}{\pi_i^3D_{jhm}+\pi_j^3D_{ihm}+\pi_h^3D_{ijm}+\pi_m^3D_{ijh}}$$

where

$$(2.27)\ D_{ijh} = \pi_i^2(\pi_j+\pi_h) + \pi_j^2(\pi_i+\pi_h) + \pi_h^2(\pi_i+\pi_h).$$

It is interesting to note that the form of (2.26) is similar to that of (2.16) if we write the latter as

$$(2.28)\ P(r_i<r_j<r_h) = \frac{\pi_i^2\pi_j}{\pi_i^2 D_{jh}+\pi_j^2 D_{ih}+\pi_h^2 D_{ij}}\text{, where}$$

$$(2.29)\ D_{ij} = \pi_i + \pi_j .$$

This suggests that a comparable model for blocks of size five might well be

(2.30)

$$P(r_i<r_j<r_h<r_m<r_v) = \frac{\pi_i^4\pi_j^3\pi_h^2\pi_m}{\pi_i^4 D_{jhmv}+\pi_j^4 D_{ihmv}+\pi_h^4 D_{ijmv}+\pi_m^4 D_{ijhv}+\pi_v^4 D_{ijhm}},$$

where D_{ijhm} is the denominator in (2.26).

From this discussion, it appears that a probability model for blocks of any size may be obtained as an extension of the paired comparisons model of Bradley and Terry. However, with larger block sizes the expressions become increasingly complicated and difficult to handle. No further attempt will be made in this study to investigate methods of rank analysis using such probability models for blocks containing more than three treatments.

2.6 Comparison of Models

In the study of probability models for ranking blocks of size three, two other models that were considered are perhaps worthy of mention. The first of these may be written as

$$(2.31)\quad P(r_i<r_j<r_h) = \frac{\delta_i}{\delta_i+\delta_j+\delta_h} \cdot \frac{\pi_i}{(\pi_j+\pi_h)},$$ where

the π_i are parameters satisfying the requirements given in Section 1.2 for paired comparisons, and the δ_i are parameters such that $P(r_i=1) = \delta_i/(\delta_i+\delta_j+\delta_h)$. The complications due to the extra set of parameters resulted in this model being abandoned early in favor of (2.16).

Another model for triple comparisons was suggested by Professor John E. Freund as an extension of paired comparisons from a consideration of the different ways in which the three treatments could be presented to an observer to be ranked in pairs. In any triplet, if $r_i<r_j$ for the first pair and $r_j<r_h$ in the second pair, then r_i must be less than r_i in the third pair. To obtain this probability model, let us first consider the following listing of the six ways in which three treatments could be ranked in pairs, together with the paired comparisons probability of the ranking that results in $r_i<r_j<r_h$.

<u>Order of Presentation</u>

First	Second	Third	
T_i, T_j	T_j, T_h	T_h, T_i	$\frac{\pi_i}{\pi_i+\pi_j} \cdot \frac{\pi_j}{\pi_j+\pi_h}$
T_i, T_j	T_h, T_i	T_j, T_h	$\frac{\pi_i}{\pi_i+\pi_j} \cdot \frac{\pi_i}{\pi_i+\pi_h} \cdot \frac{\pi_j}{\pi_j+\pi_h}$
T_j, T_h	T_h, T_i	T_i, T_j	$\frac{\pi_j}{\pi_j+\pi_h} \cdot \frac{\pi_i}{\pi_i+\pi_h} \cdot \frac{\pi_i}{\pi_i+\pi_j}$
T_j, T_h	T_i, T_j	T_i, T_h	$\frac{\pi_j}{\pi_j+\pi_h} \cdot \frac{\pi_i}{\pi_i+\pi_j} \quad \pi$
T_h, T_i	T_i, T_j	T_j, T_h	$\frac{\pi_i}{\pi_i+\pi_h} \cdot \frac{\pi_i}{\pi_i+\pi_j} \cdot \frac{\pi_j}{\pi_j+\pi_h}$
T_h, T_i	T_j, T_h	T_i, T_j	$\frac{\pi_i}{\pi_i+\pi_j} \cdot \frac{\pi_j}{\pi_j+\pi_h} \cdot \frac{\pi_i}{\pi_i+\pi_j}$

Assuming that the order of presentation is suitably randomized so that the orders are equally likely to occur, we may obtain an expression for $P(r_i<r_j<r_h)$ by summing the probabilities listed above and dividing by six. The result is

$$(2.32) \quad P(r_i<r_j<r_h) = \frac{\pi_i^2\pi_j + \pi_i\pi_j\pi_h/3}{(\pi_i+\pi_j)(\pi_i+\pi_h)(\pi_j+\pi_h)}.$$

This model retains its form when the direction of ranking is reversed but it does not have the property of conditional independence as indicated in (2.9). Preliminary investigation of (2.32) indicated that the mathematics involved in later work with this model would be more difficult than with the one proposed in (2.16).

Throughout the remainder of this discussion, we will refer to the model indicated by (2.16) as Model I and the one proposed originally by Bradley and Terry as Model II. In this way we have

$$\text{I. } P(r_i<r_j<r_h) = \frac{\pi_i^2\pi_j}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)} ,$$

and

$$\text{II. } P(r_i<r_j<r_h) = \frac{\pi_i\pi_j}{(\pi_i+\pi_j+\pi_h)(\pi_j+\pi_h)} .$$

Note that from II

$$P(r_i<r_j<r_h) = \frac{\pi_i^2\pi_j}{\pi_i(\pi_i+\pi_j+\pi_h)(\pi_j+\pi_h)} ,$$

and, for this probability to be equal to that given by I, it is necessary for

$$2\pi_i\pi_j\pi_h = \pi_i\pi_h^2 + \pi_j\pi_h^2 , \quad \text{or}$$
$$2\pi_i\pi_j = (\pi_i+\pi_j)\pi_h$$

This last equation is strictly true only when $\pi_i = \pi_j = \pi_h$, or all parameters are equal and, in this case, both probabilities reduce to one-sixth. The equality will approximately hold when the π_i do not markedly differ from each other and $P(r_i<r_j<r_h)$ will be approximately the same under both models.

We have previously shown in (2.2) that for Model II

$$(2.33)\quad P(r_i<r_j) = \pi_i/(\pi_i+\pi_j) \ .$$

From Model I we get

$$P(r_i<r_j) = \frac{\pi_i^2\pi_j+\pi_i^2\pi_h+\pi_h^2\pi_i}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)} ,$$

or

$$(2.34)\quad P(r_i<r_j) = \frac{\pi_i}{\pi_i+\pi_j\left[1+\dfrac{\pi_i(\pi_j-\pi_h)}{\pi_i\pi_j+\pi_i\pi_h+\pi_h^2}\right]} .$$

This shows that the unconditional probability that treatment i outranks treatment j in Model I is not in general equal to the paired comparisons probability that $r_i<r_j$ as was the case with Model II. Again little difference will result between the models if the π_i are approximately equal.

All investigations that have been undertaken thus far have shown very little difference between Models I and II from the standpoint of probabilities except when there are large differences in parameter values. Model I is more desirable from the standpoint of reversibility. The symmetry of the denominator involved also results in this model being more suitable from the standpoint of ease of mathematical computations. The remainder of this paper, involving methods of estimation, likelihood-ratio test procedures, and power comparisons, will deal primarily with Model I. Additional comparisons will be made with Model II as these procedures are developed.

III. LIKELIHOOD-RATIO TESTS AND ESTIMATION

3.1 The Likelihood Function

In order to develop likelihood-ratio tests and maximum-likelihood methods for estimating parameters, an expression is needed for the likelihood function associated with a triple comparison experiment based on the model proposed in (2.16). This may be obtained by first considering the probability of a given ranking ($r_{ik:jh}$, $r_{jk:ih}$, $r_{hk:ij}$) in the kth repetition of the block containing treatments i, j, and h. The probability of such a set of ranks is

(3.1)

$$P(r_{ik:jh}, r_{jk:ih}, r_{hk:ij}) = \frac{\pi_i^{3-r_{ik:jh}} \pi_j^{3-r_{jk:ih}} \pi_h^{3-r_{hk:ij}}}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)} .$$

If we assume statistical independence between blocks, the probability of a given set of ranks for the kth complete repetition is given by[4]

$$(3.2)\quad \prod_{i<j<h} \frac{\pi_i^{3-r_{ik:jh}}\pi_j^{3-r_{jk:ih}}\pi_h^{3-r_{hk:ij}}}{\pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j)} .$$

This may be rewritten as

$$(3.3)\quad f(X_{(k)},\pi) = \frac{\prod_i \pi_i^{X_{i(k)}}}{\prod_{i<j<h} D_{ijh}} ,$$

where $X_{(k)}$ and π represent vectors $(X_{1(k)},\ldots,X_{t(k)})$ and $(\pi_1,\ldots,\pi_t)$, while

$$(3.4)\quad X_{i(k)} = \sum_{j<h}{}' (3-r_{ik:jh}) ,$$

and

$$(3.5)\quad D_{ijh} = \pi_i^2(\pi_j+\pi_h)+\pi_j^2(\pi_i+\pi_h)+\pi_h^2(\pi_i+\pi_j).$$

---

4. The symbols $\sum_i$ and $\prod_i$ will be used to indicate sums and products respectively with $i=1,\ldots,t$. $\sum_{i<j<h}$ and $\prod_{i<j<h}$ will represent sums and products respectively with $i<j<h$, $i,j,h=1,\ldots,t$. $\sum_{j<h}$ and $\sum_{j\neq h}$ will indicate summations in which $j<h$ and $j\neq h$ respectively with $j,h=1,\ldots,t$. $\sum'_{j<h}$ will be used to indicate sums in which $j\neq i$, $h\neq i$, $j<h$, $i,j,h=1,\ldots,t$, where i is the additional subscript in the expression following the summation sign. Departures from these conventions will be specified.

Note that $X_{i(k)}$ is twice the number of first place ranks for treatment i plus the number of second place ranks for treatment i in all comparisons involving threatment i in the k<u>th</u> repetition of triple comparisons.

Since the set of rankings for one complete repetition is, by assumption, independent in probability of the set of rankings for a different complete repetition, the likelihood function for n complete repetitions may be written as

$$(3.6)\quad L = \prod_{k=1}^{n} f(X_{(k)}, \pi),$$

or

$$(3.7)\quad L = \frac{\prod_{i} \pi_i^{a_i}}{\prod_{i<j<h} [D_{ijh}]^n},$$

where

$$(3.8)\quad a_i = \sum_{k=1}^{n} X_{i(k)}$$

$$= \sum_{k=1}^{n} \sum_{j<h}{}' (3 - r_{ik:jh})$$

$$= \frac{3n(t-1)(t-2)}{2} - \sum_{k=1}^{n} \sum_{j<h}{}' r_{ik:jh}.$$

The sum of ranks for treatment i for the complete experiment is $\sum_{k=1}^{n} \sum_{j<h}{}' r_{ik:jh}$.

3.2 Estimation of Parameters

Using the notation of Bradley and Terry (1952a, p. 326), we shall be interested in a general class of tests of the null hypothesis,

(3.9) $H_o: \pi_i = 1/t \quad (i = 1,\ldots,t)$

against alternative hypotheses

(3.10) $H_a: \pi_i = \pi_{(\alpha)} \quad (\alpha = 1,\ldots,m);$

$i = s_{\alpha-1}+1,\ldots,s_\alpha$, where $s_o = 1$, $s_m = t$,

and $\sum_{\alpha=1}^{m} (s_\alpha - s_{\alpha-1})\pi_{(\alpha)} = 1$. Let $N_\alpha = s_\alpha - s_{\alpha-1}$,

then $\sum_{\alpha=1}^{m} N_\alpha = t$ and $\sum_{\alpha=1}^{m} N_\alpha \pi(\alpha) = 1$.

The null hypothesis states that the true treatment ratings, or parameters, are identical, but the alternative specifies that the true treatment ratings are identical within each of m groups of treatments while the parameters may differ from group to group.

Estimators $p(\alpha)$ of $\pi(\alpha)$ will be obtained using maximum-likelihood procedures. For the special case when $m=t$, the estimator of the parameter π_i will be denoted by p_i.

The likelihood function under the alternative indicated in (3.10) is

$$(3.11)\quad L = 6^{-n\sum\limits_{\alpha=1}^{m}\binom{N_\alpha}{3}} \cdot 2^{-n\sum\limits_{\alpha\neq\beta}^{m}\binom{N_\alpha}{2}(N_\beta)} \cdot \prod_{\alpha=1}^{m}[\pi(\alpha)]^{a(\alpha)}$$

$$\cdot \prod_{\alpha\neq\beta}^{m}\left\{\pi^2(\alpha)[\pi(\alpha)+\pi(\beta)]+\pi^2(\beta)\pi(\alpha)\right\}^{-n\binom{N_\alpha}{2}(N_\beta)}$$

$$\cdot \prod_{\alpha<\beta<\gamma}^{m}\left\{\pi^2(\alpha)[\pi(\beta)+\pi(\gamma)]+\pi^2(\beta)[\pi(\alpha)+\pi(\gamma)]+\pi^2(\gamma)[\pi(\alpha)+\pi(\beta)]\right\}^{-nN_\alpha N_\beta N_\gamma},$$

where

$$(3.12)\quad a(\alpha) = \sum_{i=s_{\alpha-1}+1}^{s_\alpha}\ \sum_{k=1}^{n}\ \sum_{h>j}{}'(3-r_{ik:jh})-3n\binom{N_\alpha}{3}.$$

Through maximization of ln L, subject to the restraint

$$\sum_{\alpha=1}^{m} N_\alpha \pi(\alpha) = 1,$$

the maximum-likelihood equations[5] are found to be

$$(3.13)\quad \frac{a(\alpha)}{p(\alpha)} = n\Bigg[N_\alpha \sum_{\substack{\beta=1\\ \beta\neq\alpha}}^{m}\binom{N_\beta}{2}\frac{p^2(\beta)+2p(\beta)p(\alpha)}{p^2(\beta)[p(\beta)+p(\alpha)]+p(\beta)p^2(\alpha)}$$

$$+\binom{N_\alpha}{2}\sum_{\substack{\beta=1\\ \beta\neq\alpha}}^{m} N_\beta\,\frac{3p^2(\alpha)+2p(\alpha)p(\beta)+p^2(\beta)}{p^2(\alpha)[p(\alpha)+p(\beta)]+p(\alpha)p^2(\beta)}$$

$$+N_\alpha \sum_{\substack{\gamma>\beta\\ \gamma,\beta\neq\alpha}}^{m} N_\beta N_\gamma\,\frac{p^2(\beta)+2p(\alpha)[p(\beta)+p(\gamma)]+p^2(\gamma)}{p^2(\alpha)[p(\beta)+p(\gamma)]+p^2(\beta)[p(\alpha)+p(\gamma)]+p^2(\gamma)[p(\alpha)+p(\beta)]}\Bigg],$$

$$\alpha = 1,\ldots,m.$$

---

5. For a derivation of the likelihood equations, see Appendix A.

For determinancy, we need $\sum_{\alpha=1}^{m} N_\alpha p(\alpha)=1$.

As in paired comparisons, we shall again consider the following two special cases of H_a:

(i) H_1: No π_i is assumed equal to any $\pi_j (i \neq j)$; that is, $m=t$.

(ii) H_2: $\pi_i = \pi \quad (i=1,\ldots,s)$
$= (1-s\pi)/(t-s) \quad (i=s+1,\ldots,t)$; that is, $m=2$.

The first case represents the most general of the alternative hypotheses indicated in (3.10), and the second case represents the least general.

The likelihood function for Case (i) is

$$L = \prod_i \pi_i^{a_i} \prod_{i<j<h} D_{ijh}^{-n},$$

as in (3.7). The maximum-likelihood equations may be obtained from (3.13), and are found to be

$$(3.14) \quad \frac{a_i}{p_i} = n \sum_{j<h}{}' \frac{2p_i(p_j+p_h)+p_j^2+p_h^2}{p_i^2(p_j+p_h)+p_j^2(p_i+p_h)+p_h^2(p_i+p_j)}$$
$$(i=1,\ldots,t), \text{ with } \sum_i p_i = 1.$$

For case (ii), the likelihood function may be written as follows:

$$(3.15) \quad L = \frac{6^{-n\left[\binom{s}{3}+\binom{t-s}{3}\right]} 2^{-n\left[(t-s)\binom{s}{2}+s\binom{t-s}{2}\right]} \pi^a \theta^b}{\left[\pi^2(\pi+\theta)+\theta^2\pi\right]^{n(t-s)\binom{s}{2}} \left[\pi^2\theta+\theta^2(\pi+\theta)\right]^{ns\binom{t-s}{2}}},$$

where

$$s\pi + (t-s)\theta = 1, \text{ or } \theta = (1-s\pi)/(t-s),$$

$$a = \sum_{i=1}^{s} \sum_{j<h}{}' \sum_{k=1}^{n} (3-r_{ik:jh}) - 3n\binom{s}{3},$$

$$b = \sum_{i=s+1}^{t} \sum_{j<h}{}' \sum_{k=1}^{n} (3-r_{ik:jh}) - 3n\binom{t-s}{3}.$$

The likelihood equation for this case may be obtained as a cubic equation in p, the estimator of π, from

$$\frac{a}{p} = n\left[\frac{(t-s)\binom{s}{2}[2p(p+\hat{\theta})+p^2+\hat{\theta}^2]}{[p^2(p+\hat{\theta})+\hat{\theta}^2p]} + \frac{s\binom{t-s}{2}[2p\hat{\theta}+\hat{\theta}^2]}{[p^2\hat{\theta}+\hat{\theta}^2(p+\hat{\theta})]}\right],$$

with

$$sp + (t-s)\hat{\theta} = 1 .$$

The p_i, as estimators of true treatment ratings, provide a means for quantifying the results of a triple comparisons experiment. Consequently, they permit over-all comparisons of treatments. They are also useful in the evaluation of large-sample test statistics to be proposed later.

3.2.1 Solution of the Maximum-Likelihood Equations, Case(i)

The system of equations given by (3.14) can be solved by iterative procedures. This involves finding, by any desirable means, first approximations to $p_1,\ldots,p_t$ and then getting second approximations from the first. Third approximations are obtained from the second approximations and this process of successive approximations is repeated until the desired precision is reached. For example, first estimates could be ${}^1p_i = 1/t$, $i = 1,\ldots,t$. Continuing the use of the superscript on the left side of the estimator to indicate the order of approximation, we next get

$$C({}^2p_i) = a_i\left[n\sum_{j<h}{}' \frac{2\,{}^1p_i({}^1p_j+{}^1p_h)+{}^1p_j^2+{}^1p_h^2}{{}^1p_i^2({}^1p_j+{}^1p_h)+{}^1p_j^2({}^1p_i+{}^1p_h)+{}^1p_h^2({}^1p_i+{}^1p_j)}\right]^{-1}.$$

The factor C is eliminated by using the relation $\sum_i p_i = 1$. From the second approximations, a third set could be obtained in the same way and the process continued.

Since the work involved in this iterative procedure is rather tedious and time-consuming, the best possible first estimates of the parameters should be used to start the iterative process. A method similar to one used by Otto Dykstra (1956) for paired comparisons provides a satisfactory set of first estimates here. First estimates are obtained by assuming that the estimators other than the ith one differ by very little and, as an approximation, we may set

$$(3.17)\quad p_j = \frac{1 - p_i}{t - 1}\quad (j = 1,\ldots, t,\ \ j \neq i).$$

Using these values in (3.14), we obtain

$$(3.18)\ \frac{a_i}{p_i} - \frac{n(t-1)(t-2)}{2} \cdot \frac{4p_i\left[\frac{1-p_i}{t-1}\right]+2\left[\frac{1-p_i}{t-1}\right]^2}{2p_i^2\left[\frac{1-p_i}{t-1}\right]+2\left[\frac{1-p_i}{t-1}\right]^2 \cdot \left[p_i+\frac{1-p_i}{t-1}\right]} = 0.$$

Some simplification of (3.18) yields the quadratic equations

$$(3.19)\ p_i^2[a_i(2t^2-6t+6)+n(-2^4+11t^3-22t^2+19t-6)]$$

$$+ p_i[a_i(2t-6)+n(-t^3+4t^2-5t+2)] + 2a_i = 0,$$

$$(i = 1,\ldots,t).$$

The positive roots of these equations have provided good approximations for almost all cases in which they have been used. The exceptional cases have been those in which there were great differences among the a_i and, consequently, large differences among the estimators p_i.

3.3 Likelihood-Ratio Tests

Let L(o) represent the likelihood function evaluated under the null hypothesis, and let L(a) denote the value of the likelihood function at its maximum under the alternative hypothesis. A general test statistic may be obtained as a function of the likelihood ratio

$$\lambda_a = \frac{L(o)}{L(a)} .$$

Exact tests may be found by using the likelihood ratio itself with small values of λ_a resulting in rejection of the null hypothesis. Exact tests may also be constructed from the function

(3.20)

$$B_a = n \sum_{\alpha<\beta<\gamma}^{m} N_\alpha N_\beta N_\gamma \ln\left[p_{(\alpha)}^2 (p_{(\beta)}+p_{(\gamma)}) + p_{(\beta)}^2 (p_{(\alpha)}+p_{(\gamma)}) + p_{(\gamma)}^2 (p_{(\alpha)}+p_{(\beta)})\right]$$

$$+ n \sum_{\beta\neq\alpha}^{m} \binom{N_\alpha}{2} N_\beta \ln\left[p_{(\alpha)}^2 (p_{(\alpha)}+p_{(\beta)}) + p_{(\beta)}^2 p_{(\alpha)}\right]$$

$$+ n \sum_{\alpha=1}^{m} \binom{N_\alpha}{3} \ln 6 + n \sum_{\alpha\neq\beta}^{m} \binom{N_\alpha}{2} N_\beta \ln 2 - \sum_{\alpha=1}^{m} a(\alpha) \ln p(\alpha).$$

B_a is a monotone function of the likelihood ratio chosen so that

(3.21) $\quad -2 \ln \lambda_a = 2n\binom{t}{3} \ln 6 - 2B_a.$

Small values of B_a constitute a critical region.

As yet there are no tables available for the exact distribution of the statistic B_a. Some investigation has been made of the possibility of constructing tables of exact distributions. These investigations have shown that there are two difficulties to overcome before tables of practical value can be constructed. The more serious of these is the difficulty encountered in doing the tremendous amount of computation involved to find exact distributions. The second difficulty is that of the length of the tables involved.

Some abbreviations would no doubt be possible, but if all possible entries of sums of ranks were included, such a table would be very large, even for small values of t and n. For example, under Case (i), if $t = 4$ and $n = 1$, there are 16 distinct sets of rank sums; while for $t = 4$ and $n = 2$, there are 76 distinct sets of rank sums.

3.3.1 The Generation of Tables

As indicated above, tables of exact distributions are tedious to construct. However, from a theoretical standpoint, they are very simple to obtain. To illustrate the method of generating such tables under conditions of H_0, we shall first consider Case (i) with three treatments. For one repetition, there is only one combination of rank sums, namely the set 1, 2, 3. The probability of a given ranking is 1/6 under the null hypothesis. For two repetitions, there are 36 equally likely ways in which ranks from the first repetition may be added to the corresponding ranks from the second repetition to obtain sums of ranks associated with each treatment. We can find the possible sets of sums of ranks for two repetitions by adding the numbers in the ordered set 1, 2, 3 in order with the numbers in the six permutations of 1, 2, 3 as indicated in Table 1. The other five columns that can be obtained by permuting the ordered set and then adding are the same as those obtained by using the ordered set 1, 2, 3. Consequently, we need to consider only this one ordered set.

TABLE 1. Generation of Sums of Ranks, n = 2, t = 3

Probability	Permutations			Sums with Ordered Set 1	2	3
1/6	1	2	3	2	4	6
1/6	1	3	2	2	5	5
1/6	2	1	3	3	3	6
1/6	2	3	1	3	5	4
1/6	3	1	2	4	3	5
1/6	3	2	1	4	4	4

The sets of ordered rank sums for two repetitions are thus found to be 2, 4, 6; 2, 5, 5; 3, 3, 6; 4, 4, 4; and 3, 4, 5. The probability associated with any one of the first four sets is 1/6, and the probability of the last set is 2/6. Use of these five sets of ordered rank sums with the permutations of the set 1, 2, 3 would now lead to the sets of rank sums and associated probabilities for three repetitions. A continuation of this process would lead to the sets of ordered sums of ranks for any given number of repetitions. A similar but more tedious procedure would produce corresponding results for any number of treatments.

To further illustrate the problems involved in constructing exact tables and to indicate some of the computational techniques, we have prepared the following table (Table 2) of sets of rank sums with their associated probabilities under H_o. Estimates of the parameters are indicated by p_i and the values under T_c are the results of using the p_i in an evaluation of $T = -2 \ln \lambda_1$. The subscript on λ_1 indicates that we are considering Case (i) of the alternative hypothesis.

3.3.2 Dominant Subsets of Sums of Ranks

A study of the contents of Table 2 gives some indication of results to be expected concerning the estimates of the parameters when extreme sets of rank sums, or when dominant subsets of rank sums, are a part of a set. A subset, say S_1, may be considered dominant, or be said to dominate the subset S_2, if the treatments that contribute to S_1 outrank those involved in S_2 in every block in which there is a comparison of treatments contributing to the two subsets. Examples in Table 2 are the extreme rank sums of 3 and 9. The sum 3 represents the smallest possible sum of ranks for a treatment and, in this case, the first treatment

TABLE 2. Distribution of $T = -2 \ln \lambda_1$, $t = 4$, $n = 1$

Rank Sums				Estimates of π_i				Distribution	
Σr_1	Σr_2	Σr_3	Σr_4	p_1	p_2	p_3	p_4	T_c	$P(T \leq T_c)$
3	5	7	9	1.00	____	____	____	14.33	1.00
3	6	6	9	1.00	____	____	____	11.56	0.98
4	4	7	9	0.50	0.50	____	____	11.56	0.98
4	4	8	8	0.50	0.50	____	____	8.79	0.94
4	5	6	9	0.53	0.31	0.16	____	7.66	0.93
3	6	7	8	1.00	____	____	____	7.66	0.93
3	5	8	8	1.00	____	____	____	7.02	0.78
3	7	7	7	1.00	____	____	____	6.59	0.76
5	5	5	9	0.33	0.33	0.33	____	6.59	0.76
4	5	7	8	0.52	0.31	0.11	0.06	4.04	0.72
4	6	6	8	0.53	0.20	0.20	0.07	3.49	0.61
5	5	6	8	0.35	0.35	0.22	0.08	2.46	0.54
4	6	7	7	0.54	0.20	0.13	0.13	2.46	0.54
5	5	7	7	0.35	0.35	0.15	0.15	1.59	0.31
5	6	6	7	0.36	0.24	0.24	0.16	0.78	0.22
6	6	6	6	0.25	0.25	0.25	0.25	0.00	0.02

dominates the others. The largest possible sum of ranks in Table 2 is 9 and, in this case, the first three treatments dominate the fourth. The set of rank sums 4, 4, 8, 8 illustrates the case in which two treatments dominate two others.

When a treatment, say the first one, receives the smallest possible sum of ranks, $n(t - 1)(t - 2)/2$, the set of estimates of the parameters is 1, 0,...,0. Substitution of these values in $-2 \ln \lambda_1 = 2[\ln L(1) - \ln L(o)]$ yields an indeterminate form for $\ln L(1)$. However, its limiting value may be found by assuming that the estimates are $1, b_1\delta, \ldots, b_{t-1}\delta$; and then, after substitution of these quantities for $p_1, \ldots, p_t$ into $-2 \ln \lambda_1$, taking the limit as δ approaches zero. After this has been done, $\ln L(1)$ is a function of $b_1, \ldots, b_{t-1}$. It has the appearance of the logarithm of a likelihood function involving both paired and triple comparisons. To evaluate it, maximum-likelihood estimates of $b_1, \ldots, b_{t-1}$ must be obtained. This can be rather tedious because of the form of the function involved.

In other cases of dominant subsets, a similar technique may be used to evaluate $-2 \ln \lambda_1$. The calculation of secondary estimates, such as the b_i above, provides some information about the relative values of the parameters

As an example, consider the set of rank sums 3, 6, 7, 8 from Table 2. If we substitute the values 1, 0, 0, 0 for p_1, p_2, p_3, p_4, in $-2 \ln \lambda_1$, we obtain

$$-2 \ln \lambda_1 = 8 \ln 6 + 2 \ln (0/0).$$

If we use the technique described above, we get

$$-2 \ln \lambda_1 = 8 \ln 6$$

$$+ 2 \ln \frac{b_1^3 b_2^2 b_3}{[b_1+b_2][b_1+b_3][b_2+b_3][b_1^2(b_2+b_3)+b_2^2(b_1+b_3)+b_3^2(b_1+b_2)]}.$$

The second term on the right is approximately twice the logarithm of the likelihood function for three treatments and two repetitions with rank sums 3, 4, 5. This function is $p_1^3\ p_2^2\ p_3/\hat{D}_{123}^2$. Estimates of $b_1 = .53$, $b_2 = .31$, and $b_3 = .16$ give some information concerning the relative value of the parameters of the last three treatments and permit evaluation of $-2 \ln \lambda_1$ as 7.66.

3.4 <u>Large-Sample Tests</u>

S. S. Wilks (1946) has indicated under very general conditions, that if λ is the likelihood ratio, $-2 \ln \lambda$ is asymptotically distributed as chi-square. A rather comprehensive treatment of this topic was also given by A. Wald (1943). We make use of this result in the following discussion in order to consider some approximate large-sample tests for triple comparisons.

Case (i). For the most general alternate hypothesis designated as H_1 above, we now have

(3.22)
$$-2 \ln \lambda_1 = 2n\binom{t}{3} \ln 6 + 2\sum_i a_i \ln p_i - 2n \sum_{i<j<h} \ln \hat{D}_{ijh},$$

where
$$\hat{D}_{ijh} = p_i^2(p_j+p_h) + p_j^2(p_i+p_h) + p_h^2(p_i+p_j).$$

Since there are no parameters to estimate under the null hypothesis and $(t - 1)$ independent parameters to estimate under this alternate hypothesis, $-2 \ln \lambda_1$ is approximately distributed as chi-square with $(t - 1)$ degrees of freedom when n, the number of repetitions, is large.

In an attempt to gain some insight into the rapidity of convergence of $-2 \ln \lambda_1$ to the chi-square distribution, a good deal of computation was carried out to determine first and second moments and the exact distribution of $-2 \ln \lambda_1$. The means and variances are given below (in Table 3) for the complete block case in which $t = 3$. These results indicate a fairly rapid convergence to the limiting values for means and variances, even for relatively samll values of n. It also seems reasonable to expect the convergence to be even more rapid for larger values of t since this is true in paired comparisons.

TABLE 3. Means and Variance of $-2 \ln \lambda_1$, $t = 3$

n	Mean	Variance
2	3.02	6.28
3	2.66	6.97
4	2.41	6.30
5	2.28	5.67
6	2.24	5.41
8	2.16	4.73
∞	2.00	4.00

Actual computation also showed a rapid convergence of the distribution of T to the limiting chi-square distribution in the right hand tails of the respective distributions. Some of these results, with chi-square comparisons, are shown below in Table 4 for the case where $t = 3$. T is used to indicate $-2 \ln \lambda_1$.

Case (ii). In this case, there are no parameters to estimate under the null hypothesis and only one independent parameter to estimate under H_2. As a result, when n is large, $-2 \ln \lambda_2$ is approximately distributed as chi-square with one degree of freedom. The statistic may be written as follows:

$$-2\ \ln \lambda_2 = 2n\left\{\binom{t}{3}-\left[\binom{s}{3}+\binom{t-s}{3}\right]\right\} \ln 6 - 2n\left[(t-s)\binom{s}{2}+s\binom{t-s}{2}\right]\ln 2$$

$$-2n\left\{(t-s)\binom{s}{2}\ln(p^2[p+\hat{\theta}]+p\theta^2)+s\binom{t-s}{2}\ln(p^2\hat{\theta}+\hat{\theta}^2[p+\hat{\theta}])\right\}$$

$$+2a\ \ln p + 2b\ \ln \hat{\theta}.$$

TABLE 4. Distribution of T and χ^2, t = 3

n = 6			n = 8		
T_c	$P(T\leq T_c)$	$P(\chi^2\leq T_c)$	T_c	$P(T\leq T_c)$	$P(\chi^2\leq T_c)$
2.48	.75	.71	4.34	.880	.880
3.26	.82	.80	5.13	.921	.919
4.55	.86	.90	5.76	.953	.942
4.85	.93	.91	7.01	.963	.970
6.38	.95	.96	7.70	.970	.977
7.58	.97	.98	8.07	.982	.981
8.60	.99	.98	9.07	.990	.989
10.87	.992	.995	10.85	.992	.995
13.18	.995	.998	11.42	.995	.996
13.87	.998	.999	13.91	.999	.999

3.5 The Combination of Experiments

In certain cases, the repetitions may be performed by different judges in subjective experiments, at different times, or at different locations. Under such circumstances as these, the over-all experiment may be thought of as consisting of g groups of repetitions, the uth group of size n_u, with $\sum_{u=1}^{g} n_u = n$. The treatment parameters are correspondingly indicated by $\pi_{1u},\ldots,\pi_{tu}$. If the experimenter is willing to assume that only one set of treatment parameters $\pi_1,\ldots,\pi_t$ exists as the alternative hypothesis for all groups of repetitions, the test procedure is the same as that already described. Total treatment sums of ranks are obtained by adding corresponding group treatment sums of ranks over the groups and the experiment is then treated as though one group of n repetitions, homogeneous in the indicated sense, had been made.

The pooled analysis described above is not appropriate for cases in which the alternative hypothesis that the same true treatment ratings exist for all groups is not realistic. Such an analysis may give results that are not significant, even though each group alone does show significant treatment differences. A more sensitive test for the detection of treatment differences is a combined analysis in which a test statistic is calculated for each group, and these results then combined for an overall test of the null hypothesis that there are no treatment differences. For this combined analysis, the alternative hypothesis may be specified as follows:

H_1^c: Within the uth of g groups, true ratings $\pi_{1u},\ldots,\pi_{tu}$, $\sum_{i=1}^{t} \pi_{iu} = 1$, exist; and these ratings may change from group to group.

Let λ_{1u} indicate the likelihood ratio for the uth group. For large n_u, $-2 \ln \lambda_{1u}$ is distributed approximately as chi-square with $(t - 1)$ degrees of freedom. Assuming that the group experiments are independent in probability, the appropriate test statistic is

$$(3.23) \quad -2 \ln \lambda_1^c = -2 \sum_{u=1}^{g} \ln \lambda_{1u} \; .$$

The statistic in (3.23) is approximately distributed as chi-square with $g(t - 1)$ degrees of freedom.

3.6 A Coefficient of Agreement

A measure of the consistency of ranking from group to group may be obtained by considering the following hypotheses:

H_o: $\pi_{iu} = \pi_i$, $(u = 1,\ldots,g;\; i = 1,\ldots,t)$;

H_1^c: π_{iu} is unrestricted by groups.

Letting L(o) denote the likelihood function under H_o, we see that

$$(3.24) \quad L(o) = \prod_i p_i^{a_i} \prod_{i<j<h} [\hat{D}_{ijh}]^{-n}$$

where a_i, p_i, and $\hat{D}_{ijh}$ have their original meanings as defined in (3.7) and (3.22). Similarly,

$$(3.25) \quad L^c_{(1)} = \prod_{u=1}^{g} \quad \prod_i p_{iu}^{a_{iu}} \prod_{i<j<h} [\hat{D}_{ijh}(p_{iu})]^{-n_u}$$

where p_{iu} = the maximum-likelihood estimator of π_{iu}, and a_{iu} is twice the number of first place ranks plus the number of second place ranks for treatment i in all comparisons involving treatment i in the uth group of repetitions.

Under H_o, a total of (t - 1) independent parameters must be estimated, while under H_1^c of this section, there are g(t - 1) independent parameters to be estimated. Consequently, when the number of repetitions is large,

$$(3.26) \quad -2 \ln \lambda_1^A = -2 \ln \frac{L(o)}{L^c(1)}$$

$$= 2 \sum_{u=1}^{g} [\sum_i a_{iu} \ln p_{iu} - n_u \sum_{i<j<h} \ln \hat{D}_{ijh}(p_{iu})]$$

$$- 2[\sum_i a_i \ln p_i - n \sum_{i<j<h} \ln \hat{D}_{ijh}]$$

is distributed approximately as chi-square with (g-1)(t-1) degrees of freedom. Large values of $-2 \ln \lambda_1^A$ in (3.26) imply discordant ranking from group to group, and hence, $-2 \ln \lambda_1^A$ is in fact a measure of group by treatment interaction.

The large-sample tests proposed for Case (i) in this chapter may be summarized as in Table 5.

TABLE 5. Large-sample Analysis

Statistic*	Hypothesis	Limiting Distribution
$2n\binom{t}{3}\ln 6+2\sum_i a_i \ln p_i-2n\sum_{i<j<h}\ln \hat{D}_{ijh}$	$H_o:\pi_i=1/t$ $H_1:\pi_i$	χ^2_{t-1}
$2n\binom{t}{3}\ln 6+2\sum_{u=1}\left[\sum_i a_{iu}\ln p_{iu}-n_u\sum_{i<j<h}\ln\hat{D}_{ijh}(p_{iu})\right]$	$H_o:\pi_{iu}=1/t$ $H_1:\pi_{iu}$	$\chi^2_{g(t-1)}$
$2\sum_{u=1}^{g}\left[\sum a_{iu}\ln p_{iu}-n_u\sum_{i<j<h}\ln\hat{D}_{ijh}(p_{iu})\right]$ $-2\left[\sum_i a_i\ln p_i-n\sum_{i<j<h}\ln\hat{D}_{ijh}\right]$	$H_o:\pi_{iu}=\pi_i$ $H_1:\pi_{iu}$	$\chi^2_{(g-1)(t-1)}$

* $\hat{D}_{ijh} = p_i^2(p_j+p_h) + p_j^2(p_i+p_h) + p_h^2(p_i+p_j)$

$\hat{D}_{ijh}(p_{iu}) = p_{iu}^2(p_{ju}+p_{hu}) + p_{ju}^2(p_{iu}+p_{hu}) + p_{hu}^2(p_{iu}+p_{ju})$.

$$a_i = \frac{3n(t-1)(t-2)}{2} - \sum_{k=1}^{n}\sum_{j<h}{}' r_{ik:jh}$$

$a_{iu} = \dfrac{3n_u(t-1)(t-2)}{2}$ - (sum of ranks for treatment i in the $u\underline{th}$ group).

IV. A TEST OF THE MODEL FOR TRIPLE COMPARISONS

4.1 A Goodness-of-Fit Test of the Appropriateness of the Model

The statistical tests of the preceding chapter are appropriate only for data that conform to the assumptions underlying the given model for triple comparisons. In order to determine significant departures from the model, a goodness-of-fit test of the appropriateness of the model for triple comparisons will be developed in the following paragraphs.

The most general model available for a triple comparisons experiment is one in which each block is treated as a small independent multinomial-type experiment. In order to indicate this, let $\bar{\pi}_{ijh}$ be a parameter representing the probability that treatments i, j, and h are ranked 1, 2, and 3 respectively in a given block in which they are compared. Then, for each triplet, there are six such parameters, possibly totally unrelated for distinct triplets. Correspondingly, let f_{ijh} represent the number of times treatments i, j, and h are ranked 1, 2, and 3 respectively in an experiment of n complete repetitions. The likelihood function, collected over separated components for each triplet, may now be written as

(4.1)
$$\bar{L}(\bar{\pi}_{ijh}) = \prod_{i<j<h} [\bar{\pi}_{ijh}^{f_{ijh}} \bar{\pi}_{jhi}^{f_{jhi}} \bar{\pi}_{hij}^{f_{hij}} \bar{\pi}_{hji}^{f_{hji}} \bar{\pi}_{jih}^{f_{jih}} \bar{\pi}_{ihj}^{f_{ihj}}].$$

From their definition it is obvious that

$$(4.2) \quad f_{ijh}+f_{jhi}+f_{hij}+f_{hji}+f_{jih}+f_{ihj} = n$$

$$(i,\ j,\ h = 1,\ldots,t;\ i \neq j \neq h),\ \text{and}$$

$$(4.3) \quad \bar{\pi}_{ijh}+\bar{\pi}_{jhi}+\bar{\pi}_{hij}+\bar{\pi}_{hji}+\bar{\pi}_{jih}+\bar{\pi}_{ihj} = 1$$

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

In order to test the appropriateness, or goodness-of-fit, of the model for triple comparisons for homogeneous repetitions, we need a test of the null hypothesis

$$(4.4) \quad H_o\colon \bar{\pi}_{ijh} = \frac{\pi_i^2\pi_j^2}{\pi_i^2(\pi_j+\pi_h) + \pi_j^2(\pi_i+\pi_h) + \pi_h^2(\pi_i+\pi_j)}$$
$$= \pi_i^2\pi_j/D_{ijh}\ ,$$

against the more general alternative,

$$(4.5) \quad H_I\colon \bar{\pi}_{ijh} \neq \pi_i^2\pi_j/D_{ijh} \qquad \text{for some } i, j, \text{ and } h.$$

Note that in the alternative we attempt to represent the data in terms of $5\binom{t}{3}$ independent parameters. The null hypothesis is much more restrictive, since through it, we attempt to represent the data with only $(t-1)$ independent parameters. The alternative fits the observations exactly in the sense that the estimated expected frequencies of rankings are the same as the observed frequencies.

An approximate test of H_o against H_I is possible using likelihood-ratio procedures. The test statistic depends on the functions $\bar{L}(\bar{p}_{ijh}|H_o)$ and $\bar{L}(\bar{p}_{ijh}|H_I)$, where $\bar{L}$ is defined by (4.1), and these functions represent evaluations of $\bar{L}$ in terms of maximum-likelihood estimators $\bar{p}_{ijh}$ obtained under the assumptions of H_o and H_I respectively. Under the null hypothesis, the estimators $\bar{p}^o_{ijh}$ of $\bar{\pi}_{ijh}$ have already been found to be

$$(4.6)\quad \bar{p}^o_{ijh} = p_i^2 p_j / \hat{D}_{ijh},$$

where the p_i are the estimators of π_i discussed in Section 3.2. Under the assumptions of H_I, estimators $\bar{p}^I_{ijh}$ of $\bar{\pi}_{ijh}$ are easily seen to be

$$(4.7)\quad \bar{p}^I_{ijh} = f_{ijh}/n,$$

the usual maximum-likelihood estimators of multinomial parameters.

Letting $\lambda_I = \bar{L}(\bar{p}_{ijh}|H_o)/\bar{L}(\bar{p}_{ijh}|H_I)$, $-2 \ln \lambda_I$ is asymptotically distributed as chi-square. The degrees of freedom may be obtained by noting that under H_I, a total of $t(t-1)(t-2)$ parameters are estimated, subject to $t(t-1)(t-2)/6$ restraints, as indicated in (4.3); and, under

H_o, there are (t-1) independent parameters. Hence, there are $\frac{5t(t-1)(t-2)}{6}$ - (t-1) degrees of freedom for chi-square.

We may write the test statistic as follows:

$$(4.8)\quad -2 \ln \lambda_I = 2[\ln \bar{L}(\bar{p}_{ijh}|H_I) - \ln \bar{L}(\bar{p}_{ijh}|H_o)]$$

$$= 2 \sum_{i<j<h} \Sigma'' f_{ijh}[\ln \bar{p}^{I}_{ijh} - \ln \bar{p}^{o}_{ijh}],$$

where Σ'' indicates the sum over the six permutations of i, j, and h. Using (4.6) and (4.7) in (4.8), we see that

(4.9)

$$-2 \ln \lambda_I = 2 \sum_{i<j<h} \Sigma'' f_{ijh}[\ln f_{ijh} - \ln n - \ln(p_i^2 p_j) + \ln \hat{D}_{ijh}]$$

$$= 2[\sum_{i<j<h} \Sigma'' f_{ijh} \ln f_{ijh} - n\binom{t}{3}\ln n + n \sum_{i<j<h} \ln \hat{D}_{ijh} - \sum_i a_i \ln p_i],$$

where $a_i = \sum_{j<h}' \sum_{k=1}^{n} (3 - r_{ik:jh})$ as before.

When the repetitions are grouped into g groups, within-group tests of goodness-of-fit may be made for each group as described above. If we follow the notation of Section 3.5, the desired test is that of the null hypothesis,

$$(4.10)\quad H_o: \bar{\pi}_{ijhu} = \pi_{iu}^2 \pi_{ju}/D_{ijhu},$$

$$(u = 1,\ldots,g;\ i, j, h = 1,\ldots,t;\ i \neq j \neq h)$$

against the alternative,

$$(4.11)\quad H_{II}: \bar{\pi}_{ijhu} \neq \pi_{iu}^2 \pi_{ju}/D_{ijhu}$$

for some i, j, h, and fixed u, where

$$D_{ijhu} = \pi_{iu}^2(\pi_{ju}+\pi_{hu}) + \pi_{ju}^2(\pi_{iu}+\pi_{hu}) + \pi_{hu}^2(\pi_{iu}+\pi_{ju}).$$

Now the test statistic for the u*th* group may be denoted by $-2 \ln \lambda_{Iu}$, where the subscript u indicates that the statistic is obtained as in (4.8) but only the u*th* group is considered. Since the rankings are independent from group to group, the over-all goodness-of-fit statistic is

$$(4.12) \quad -2 \ln \lambda_{II} = -2 \sum_{u=1}^{g} \ln \lambda_{Iu}.$$

For large values of n_u, $u = 1,\ldots,g$, this statistic is approximately distributed as chi-square with $g[5\binom{t}{3}-(t-1)]$ degrees of freedom.

4.2 Goodness-of-Fit Test Associated with Expected Frequencies

In the discussion of Section 4.1 above, reference was made to $-2 \ln \lambda$ as being asymptotically distributed as chi-square. The quantity $-2 \ln \lambda$ may be associated with the more common chi-square goodness-of-fit test by the following approximations. Let

$$(4.13) \quad f_{ijh}^{o} = n\bar{p}_{ijh}^{o},$$

and note that

$$(4.14) \quad f_{ijh} = n\bar{p}_{ijh}^{I},$$

the latter from (4.7).

Substitution of these values into (4.8) gives

(4.15)

$$-2 \ln \lambda_I = 2 \sum_{i<j<h} \Sigma'' f_{ijh}[\ln(f_{ijh}/n) - \ln(f^o_{ijh}/n)]$$

$$= 2 \sum_{i<j<h} \Sigma'' f_{ijh} \ln(f_{ijh}/f^o_{ijh}).$$

Let

$$(4.16) \quad f_{ijh}/f^o_{ijh} = 1 + \epsilon_{ijh},$$

where ϵ_{ijh} may be either positive or negative. Substituting (4.16) into (4.15), we get

$$(4.17) \quad -2 \ln \lambda_{II} = 2 \sum_{i<j<h} \Sigma'' f^o_{ijh}(1+\epsilon_{ijh})\ln(1+\epsilon_{ijh}).$$

If, as an approximation, we use the first two terms in the series expansion of the logarithm, we get

$$(4.18) \quad -2 \ln \lambda_I \doteq 2 \sum_{i<j<h} \Sigma'' f^o_{ijh}(1+\epsilon_{ijh})(\epsilon_{ijh}-\epsilon^2_{ijh}/2)$$

$$\doteq 2 \sum_{i<j<h} \Sigma'' f^o_{ijh}[\epsilon_{ijh}+\epsilon^2_{ijh}/2-\epsilon^3_{ijh}/2].$$

From (4.16), $f_{ijh} = f^o_{ijh} + \epsilon_{ijh}f^o_{ijh}$ and hence

$$(4.19) \quad \sum_{i<j<h} \Sigma'' \epsilon_{ijh}f^o_{ijh} = \sum_{i<j<h} \Sigma''(f_{ijh} - f^o_{ijh})$$

$$= n\binom{t}{3} - n\binom{t}{3} = 0.$$

If we continue to ignore powers of ϵ_{ijh} greater than the second, we now get

$$(4.20) \quad -2 \ln \lambda_I \doteq \sum_{i<j<h} \Sigma'' f^o_{ijh}\epsilon^2_{ijh}.$$

Since $\epsilon_{ijh} = (f_{ijh} - f^{o}_{ijh})/f^{o}_{ijh}$, this last approximation may be written as follows:

$$(4.21) \quad -2 \ln \lambda_{I} \doteq \sum_{i<j<h} \Sigma'' (f_{ijh} - f^{o}_{ijh})^{2}/f^{o}_{ijh}.$$

This relation (4.21) gives the usual expression for the chi-square goodness-of-fit statistic. From its derivation, it is seen to be a good approximation to $-2 \ln \lambda_{I}$ if ϵ_{ijh} is small. It is meaningless if $\epsilon_{ijh} = -1$, which occurs when an observed frequency is zero. In practice, with reasonably large experiments, the approximation of (4.21) will generally agree closely with the form of (4.15) and this justifies taking the statistic (4.21) to have a chi-square distribution; an alternate justification could be obtained beginning with the form (4.21) without reference to its derivation included here.

When the repetitions are grouped into g groups, a slight modification of the notation (to indicate the association of frequencies with groups) permits us to write

$$(4.22) \quad -2 \ln \lambda_{II} \doteq \sum_{u=1}^{g} \sum_{i<j<h} \Sigma'' (f_{ijhu} - f^{o}_{ijhu})^{2}/f^{o}_{ijhu},$$

where f_{ijhu} represents the number of times treatments i, j, and h were ranked 1, 2, and 3 respectively in the <u>uth</u> group of repetitions, and f^{o}_{ijhu} is the corresponding expected value under H_{o}. This is the association with frequencies that is obtained from the test statistic in

(4.12) for the case when repetitions are grouped. Consequently, (4.22) may be tested using chi-square with $g[5\binom{t}{3} - (t-1)]$ degrees of freedom.

The developments of this chapter parallel those of Bradley (1954b), who dealt with the appropriateness of the model developed for paired comparisons. In this paper, he applied the goodness-of-fit test to paired comparisons data and showed that the model did appear to be appropriate in practice. J. W. Hopkins (1954) did an experimental methodological study for paired comparisons and also suggested that the model proposed was appropriate. When sufficient experimental evidence is available, we shall be able to assess the appropriateness of the model for triple comparisons proposed in this dissertation. Departures from the model will sometimes occur, but it is conjectured that they will be rare and that assignable causes may usually be found to explain them.

V. SOME LARGE-SAMPLE RESULTS ON ESTIMATION AND POWER

5.1 Introduction

In Section 3.3 some indication was given of the difficulties encountered in finding the exact distribution of $T = -2 \ln \lambda_1$ under the null hypothesis of treatment equality for Case (i). In that case, the null hypothesis that all parameters are equal is tested against the alternative that no π_i is assumed equal to π_j for all $j \neq i$. The difficulties are even more troublesome in the event that the null hypothesis is not true. Consequently, only a large-sample evaluation and comparison of power functions with other test procedures will be made here. Large-sample results concerning the distribution of the estimators p_i of π_i are first obtained. They give some indication of the reliability of the estimators and are useful in the investigation of the power of the test based on T.

5.2 Distribution of Estimators for Case (i)

To consider the distribution of estimators for Case (i), we first refer to some results already stated. In Equation (3.1), the probability of a given set of ranks for one complete repetition was written as

$$(5.1)\quad f(X, \pi) = \prod_i \pi_i^{X_i} \prod_{i<j<h} D_{ijh}^{-1},$$

where

$$X_i = \sum_{j<h}{}' (3 - r_{i:jh}),$$

and

$$D_{ijh} = \pi_i^2(\pi_j+\pi_h) + \pi_j^2(\pi_i+\pi_h) + \pi_h^2(\pi_i+\pi_j).$$

Here X and π represent vectors $(X_1,\ldots,X_t)$ and $\pi_1,\ldots,\pi_t)$. As noted in Section 3.1, X_i is twice the number of times treatment i ranks first plus the number of times treatment i ranks second in the complete repetition.

The parameters $\pi_1,\ldots,\pi_t$ are subject to the single restriction $\sum_i \pi_i = 1$. If we take $\pi_t = 1 - \sum_{i=1}^{t-1} \pi_i$, we shall then have (t - 1) independent parameters $\pi_1,\ldots,\pi_{t-1}$. Let $p_1,\ldots,p_{t-1}$ be the maximum-likelihood estimators of these parameters as obtained in Section 3.2. According to K. C. Chanda (1954), if $f(X,\pi)$ satisfies certain conditions, then p_i is a consistent estimator of π_i, and $\sqrt{n}(p_1 - \pi_1),\ldots,$ $\sqrt{n}(p_{t-1} - \pi_{t-1})$ have asymptotically a joint (t - 1)-variate normal distribution with zero means and dispersion matrix given by the inverse of $[\lambda'_{ij}]$, where

$$(5.2)\quad \lambda'_{ij} = E\left(\frac{\partial \ln f}{\partial \pi_i} \cdot \frac{\partial \ln f}{\partial \pi_j}\right).$$

The required conditions in terms of our problem are as follows:

(i) The point represented by the vector π lies in a $(t - 1)$-dimensional region Ω; for almost all X and for all $\pi \in \Omega$,

$$\frac{\partial \ln f}{\partial \pi_i}, \ \frac{\partial^2 \ln f}{\partial \pi_i \partial \pi_j}, \text{ and } \frac{\partial^3 \ln f}{\partial \pi_i \partial \pi_j \partial \pi_h}$$

exist for $i, j, h = 1, \ldots, (t - 1)$.

(ii) For almost all X and every point $\pi \in \Omega$,

$$\left|\frac{\partial f}{\partial \pi_i}\right| < F_i(X), \ \left|\frac{\partial^2 f}{\partial \pi_i \partial \pi_j}\right| < F_{ij}(X), \text{ and}$$

$$\left|\frac{\partial \ln f}{\partial \pi_i \partial \pi_j \partial \pi_h}\right| < H_{ijh}(X) \text{ while } E[H_{ijh}(X)] \text{ is}$$

finite for $i, j, h = 1, 2, \ldots, (t - 1)$.

(iii) For all $\pi \in \Omega$ the matrix

$\lambda' = [\lambda'_{ij}]$, where λ'_{ij} is defined in (5.2),

is positive definite and $|\lambda'_{ij}|$ is finite.

The first two conditions are readily verified from the form of $f(X,\pi)$ provided that each π_i has a positive lower bound, a situation which we will assume because otherwise treatments whose parameters are zero would have to be excluded from the experiment. Condition (iii) will be verified after obtaining some preliminary results.

Taking the logarithm of $f(X,\pi)$ and differentiating with respect to π_i $(i = 1,\ldots,[t-1])$ while regarding π_t as a function of π_i, we obtain

(5.3)

$$\frac{\partial \ln f}{\partial \pi_i} = \frac{X_i}{\pi_i} - \frac{X_t}{\pi_t} - \sum_{j<h}{}' \frac{2\pi_i(\pi_j+\pi_h)+\pi_j^2+\pi_h^2}{D_{ijh}} + \sum_{\substack{j<h\\ j,h\neq t}} \frac{2\pi_t(\pi_j+\pi_h)+\pi_j^2+\pi_h^2}{D_{tjh}}.$$

Now we note that

$$E(X_i) = \sum_{j<h}{}' E(3 - r_{i:jh})$$

$$= \sum_{j<h}{}' \left[2\frac{\pi_i^2(\pi_j+\pi_h)}{D_{ijh}} + \frac{\pi_j^2\pi_i}{D_{ijh}} + \frac{\pi_h^2\pi_i}{D_{ijh}}\right],$$

or

$$(5.4)\quad E(X_i) = \sum_{j<h}{}' \frac{2\pi_i^2(\pi_j+\pi_h)+\pi_j^2\pi_i+\pi_h^2\pi_i}{D_{ijh}},\ i = 1,\ldots,t.$$

It follows that

$$(5.5)\quad \frac{\partial \ln f}{\partial \pi_i} = \left(\frac{X_i}{\pi_i} - \frac{X_t}{\pi_t}\right) - E\left(\frac{X_i}{\pi_i} - \frac{X_t}{\pi_t}\right).$$

But from (5.5), we see that

$$\lambda'_{ij} = E\left(\frac{\partial \ln f}{\partial \pi_i}\cdot\frac{\partial \ln f}{\partial \pi_j}\right)$$

$$= \mathrm{Cov}\left[\left(\frac{X_i}{\pi_i} - \frac{X_t}{\pi_t}\right)\left(\frac{X_j}{\pi_j} - \frac{X_t}{\pi_t}\right)\right]$$

$$= \frac{\mathrm{Cov}(X_i,X_j)}{\pi_i\pi_j} - \frac{\mathrm{Cov}(X_i,X_t)}{\pi_i\pi_t} - \frac{\mathrm{Cov}(X_j,X_t)}{\pi_i\pi_t} + \frac{\mathrm{Var}(x_t)}{\pi_t^2},$$

or

$$(5.6)\quad \lambda'_{ij} = \lambda_{ij} - \lambda_{it} - \lambda_{jt} + \lambda_{tt},$$

where

$$(5.7) \quad \lambda_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\pi_i \pi_j}, \quad (i,\ j = 1, \ldots, t).$$

In order to express λ_{ij} in terms of the parameters π_i, let

$$(5.8) \quad x_{i:jh} = 3 - r_{i:jh} .$$

Since $X_i = \sum'_{j<h} x_{i:jh}$, and the $x_{i:jh}$ of this sum are statistically independent, it follows that

$$(5.9) \quad \mathrm{Var}(X_i) = \sum'_{j<h} \mathrm{Var}(x_{i:jh}),$$

and

$$(5.10) \quad \mathrm{Cov}(X_i, X_j) = \sum_{h \neq i,j} \mathrm{Cov}(x_{i:jh}, x_{j:ih}) .$$

We may find $\mathrm{Var}(x_{i:jh})$ as follows:

$$(5.11) \quad \mathrm{Var}(x_{i:jh}) = E(x_{i:jh})^2 - [E(x_{i:jh})]^2$$

$$= 4\,\frac{\pi_i^2(\pi_j+\pi_h)}{D_{ijh}} + \frac{\pi_j^2\pi_i}{D_{ijh}} + \frac{\pi_h^2\pi_i}{D_{ijh}} - \left[\frac{2\pi_i^2(\pi_j+\pi_h) + \pi_j^2\pi_i + \pi_h^2\pi_i}{D_{ijh}}\right]^2$$

$$= \frac{\pi_i}{D_{ijh}^2}\,[4\pi_i\pi_j\pi_h(\pi_j+\pi_h)^2 + (\pi_j+\pi_h)(\pi_j^2+\pi_h^2)(\pi_i^2+\pi_j\pi_h)].$$

In like manner,

$$(5.12) \quad \mathrm{Cov}(x_{i:jh}, x_{j:ih}) = E(x_{i:jh}x_{j:ih}) - E(x_{i:jh})E(x_{j:ih})$$

$$= \frac{-\pi_i\pi_j}{D_{ijh}^2}[2\pi_i\pi_j\pi_h(\pi_i+2\pi_h+\pi_j)+2\pi_h^3(\pi_i+\pi_j)-\pi_h^2(\pi_i^2+\pi_j^2)+\pi_i^2\pi_j^2+\pi_h^4] .$$

Using (5.11) and (5.12) as needed, we get

(5.13)

$$\lambda_{ij} = -\sum_{h\neq i,j} \frac{2\pi_i\pi_j\pi_h(\pi_i+2\pi_h+\pi_j)+2\pi_h^3(\pi_i+\pi_j)-\pi_h^2(\pi_i^2+\pi_j^2)+\pi_i^2\pi_j^2+\pi_h^4}{D_{ijh}^2}$$

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

$$\lambda_{ii} = \frac{1}{\pi_i} \sum_{j<h}{}' \frac{4\pi_i\pi_j\pi_h(\pi_j+\pi_h)^2+(\pi_j+\pi_h)(\pi_j^2+\pi_h^2)(\pi_i^2+\pi_j\pi_h)}{D_{ijh}^2}$$

$(i=1,\ldots,t)$.

Since $\lambda'_{ij} = \mathrm{Cov}\ (\frac{X_i}{\pi_i} - \frac{X_t}{\pi_t}),(\frac{X_j}{\pi_j} - \frac{X_t}{\pi_t})$, the matrix $[\lambda'_{ij}]$ is a dispersion matrix and is therefore non-negative definite. It is furthermore positive definite since $X_1,\ldots,X_{t-1}$, and hence $\frac{\partial \ln f}{\partial\pi_1},\ldots,\frac{\partial \ln f}{\partial\pi_{t-1}}$, are free of linear restrictions. The preceding calculations of λ'_{ij} in terms of parameters indicate that $|\lambda'_{ij}|$ is finite provided that all π_i have a positive lower bound as was assumed earlier. As a result, and according to Chanda, we can conclude that the $p_i (i = 1,\ldots,[t - 1])$ are consistent estimators and that $\sqrt{n}(p_1,-\pi_1),\ldots,\sqrt{n}(p_{t-1}-\pi_{t-1})$ have asymptotically a joint (t-1)-variate normal distribution with zero means and dispersion matrix given by the inverse of $[\lambda'_{ij}]$.

The lack of symmetry introduced into the elements of the dispersion matrix $[\lambda'_{ij}]^{-1}$ by considering π_t to be a function of the other (t-1) parameters may be removed by the following method developed by Bradley (1955) for paired comparisons. This method is based on the fact that the $t \times t$ matrix $[\lambda_{ij}]$ for triple comparisons is singular, as was the corresponding matrix for paired comparisons. In both cases, we have

$$(5.14)\quad \pi_i\lambda_{ii} + \sum_j{}' \pi_j\lambda_{ij} = 0.$$

To verify this, we use (5.13) to obtain

$$(5.15)$$

$$\pi_i\lambda_{ii} = \sum_{j<h}{}' \frac{4\pi_i\pi_j\pi_h(\pi_j+\pi_h)^2+(\pi_j+\pi_h)(\pi_j^2+\pi_h^2)(\pi_i^2+\pi_j\pi_h)}{D_{ijh}^2},$$

$$= \sum_{j<h}{}' N_{ijh}/D_{ijh}^2,$$

and

$$(5.16)\quad \sum_j{}'\pi_j\lambda_{ij} = -\sum_j{}' \sum_{h\neq i,j}\left[\frac{2\pi_i\pi_j^2\pi_h(\pi_i+2\pi_h+\pi_j)}{D_{ijh}^2} + \frac{2\pi_j\pi_h^3(\pi_i+\pi_j)-\pi_j\pi_h^2(\pi_i^2+\pi_j^2)+\pi_i^2\pi_j^3+\pi_j\pi^4}{D_{ijh}^2}\right]$$

$$= -\sum_j{}' \sum_{h\neq i,j} M_{ijh}/D_{ijh}^2,$$

where the definitions of N_{ijh} and M_{ijh} are obvious. We now note that

$$\pi_i\lambda_{ii} + \sum_j{}'\pi_j\lambda_{ij} = \sum_{j<h}{}' \frac{N_{ijh} - (M_{ijh} + M_{ihj})}{D_{ijh}^2} = 0$$

since $N_{ijh} - (M_{ijh} + M_{ihj}) = 0$.

Transforming $[\lambda_{ij}]$ by subtracting the elements of the last row and then the elements of the last column from corresponding elements of the remaining rows and columns respectively, we obtain

$$(5.17)\qquad |\lambda_{ij}| = \begin{vmatrix} [\lambda'_{ij}] & [\lambda_{it} - \lambda_{tt}]' \\ [\lambda_{jt} - \lambda_{tt}] & \lambda_{tt} \end{vmatrix} = 0,$$

where $[\lambda_{it} - \lambda_{tt}]'$ and $[\lambda_{jt} - \lambda_{tt}]$ are respectively column and row vectors of $(t - 1)$ elements. From (5.17), we see that

$$(5.18)\qquad |\lambda'_{ij}| = \begin{vmatrix} [\lambda'_{ij}] & [\lambda_{it} - \lambda_{tt}]' \\ [\lambda_{jt} - \lambda_{tt}] & 1 + \lambda_{tt} \end{vmatrix} .$$

In the determinant of the right member of (5.18), adding the elements of the last row and then the elements of the last column to the corresponding elements of the remaining rows and columns respectively results in

$$(5.19)\qquad |\lambda'_{ij}| = |\, 1 + \lambda_{ij}| = \begin{vmatrix} [1 + \lambda_{ij}] & [1]' \\ [0] & 1 \end{vmatrix} ,$$

where $[0]$ is a row vector of $(t - 1)$ zero elements and $[1]'$ is a column vector of $(t - 1)$ unit elements. From (5.19), we obtain

$$(5.20)\qquad |\lambda'_{ij}| = -\begin{vmatrix} [\lambda_{ij}] & [1]' \\ [1] & 0 \end{vmatrix} .$$

Inspection of the cofactor of λ_{ij} in the right member of (5.20) indicates that it may be shown to be equal to the cofactor of λ'_{ij} in $|\lambda'_{ij}|$. This may be done by transforming the cofactor of λ_{ij} by subtracting λ_{rt} times the elements of the last row from corresponding elements of the rth row, and then subtracting $(\lambda_{tc} - \lambda_{tt})$ times the elements of the last column from the corresponding elements of the cth column. It follows that if σ_{ij} is the covariance of $\sqrt{n}(p_i - \pi_i)$ and $\sqrt{n}(p_j - \pi_j)$, $[i, j = 1,\ldots,(t-1)]$, then

$$(5.21)\quad \sigma_{ij} = \frac{\text{cofactor of } \lambda_{ij} \text{ in } \begin{vmatrix} [\lambda_{ij}] & [1]' \\ [1] & 0 \end{vmatrix}}{\begin{vmatrix} [\lambda_{ij}] & [1]' \\ [1] & 0 \end{vmatrix}} .$$

Since the model for triple comparisons is symmetric in the parameters $\pi_1,\ldots,\pi_t$ and estimators $p_1,\ldots,p_t$, (5.21) applies for all variances and covariances with $i, j=1,\ldots,t$ on the basis of its symmetry.

Inspection of (5.21) shows that the variances and covariances desired are simply the elements of the t-square principal minor of the inverse of the matrix of the determinant in the denominator. This inverse matrix may be evaluated by elementary methods if t is small. Since $|\lambda_{ij}| = 0$, the usual Doolittle methods of matrix inversion can not be used. Consequently, if t is large, it appears desirable to invert $[\lambda'_{ij}]$ by a Doolittle method in order to obtain the variances and covariances for i, j = 1,...,(t - 1). Since

$$(5.22)\quad \sqrt{n}(p_t - \pi_t) = -\sum_{i=1}^{t-1} \sqrt{n}(p_i - \pi_i),$$

the remaining variances and covariances associated with $\sqrt{n}(p_t - \pi_t)$ are obtained from

$$(5.23)\quad \sigma_{tt} = \sum_{i=1}^{t-1} \sum_{j=1}^{t-1} \sigma_{ij} \quad \text{and}$$

$$\sigma_{it} = \sum_{j=1}^{t-1} \sigma_{ij},\ i = 1,\ldots,(t - 1).$$

Since $\sqrt{n}(p_t - \pi_t)$ is a linear combination of variates having a multivariate normal distribution, $\sqrt{n}(p_1 - \pi_1),\ldots,$ $\sqrt{n}(p_t - \pi_t)$ have, for large values of n, a singular multivariate normal distribution with zero means and dispersion matrix $[\sigma_{ij}]$ given by (5.21). For large sample sizes then, we may take $p_1,\ldots,p_t$ to be jointly normally distributed with means $\pi_1,\ldots,\pi_t$ and dispersion matrix $[\sigma_{ij}/n]$. As a result, it is possible to consider, for large n, any linear function $\sum_i b_i p_i$ to be normally distributed with mean $\sum_i b_i \pi_i$ and variance $\sum_{i=1}^{t} \sum_{j=1}^{t} b_i b_j \sigma_{ij}/n$. Orthogonal linear comparisons may then be made as in the analysis of variance.

In actual practice, it will usually be necessary to estimate variances and covariances. Because of the consistency of maximum-likelihood estimators, it is possible to define estimators $\hat{\lambda}_{ij}$ of λ_{ij} as the same functions of $p_1,\ldots,p_t$ that the λ_{ij} are of $\pi_1,\ldots,\pi_t$ in (5.13). Then a consistent estimator $\hat{\sigma}_{ij}$ of σ_{ij} may be obtained by defining the sample dispersion matrix $[\hat{\sigma}_{ij}]$ to be the same function of the $\hat{\lambda}_{ij}$ as $[\sigma_{ij}]$ is of λ_{ij} in (5.21).

Under the null hypothesis that all parameters are equal, we find

(5.24) $\lambda_{ij} = -t^2(t - 2)/3, \quad i \neq j, i,j=1,\ldots,t;$

and

(5.25) $\lambda_{ii} = t^2(t - 1)(t - 2)/3, \quad i=1,\ldots,t.$

The matrix inversion indicated in (5.21) may now be performed algebraically and the result is

(5.26) $\sigma_{ij} = \dfrac{-3}{t^4(t - 2)}, \quad i \neq j, i,j=1,\ldots,t;$

and

(5.27) $\sigma_{ii} = \dfrac{3(t - 1)}{t^4(t - 2)}, \quad i=1,\ldots,t.$

5.3 Distribution of the Estimator for Case (ii)

The distribution problem for Case (ii) is similar to that of Section 5.2. The only difference is that, under Case (ii), there is only one independent parameter to estimate. In this case, the alternate hypothesis is $\pi_i = \pi$ for $i = 1,\ldots,s$, and $\pi_i = \theta$ for $i = s + 1,\ldots,t$, where $s\pi + (t - s)\theta = 1$. The probability for a given set of rankings for a complete repetition may be written as

$$(5.28)\quad f(X,\pi) = \frac{6^{-\left[\binom{s}{3}+\binom{t-s}{3}\right]}\, 2^{-\left[(t-s)\binom{s}{2}+s\binom{t-s}{2}\right]}\, \pi^{X_a}\, \theta^{X_b}}{\left[\pi^2(\pi+\theta)+\theta^2\pi\right]^{(t-s)\binom{s}{2}}\left[\pi^2\theta+\theta^2(\pi+\theta)\right]^{s\binom{t-s}{2}}}$$

where

$$X_a = \sum_{i=1}^{s} \sum_{j<h}{}' (3 - r_{i:jh}) - 3\binom{s}{3},$$

and

$$X_b = \sum_{i=s+1}^{t} \sum_{j<h}{}' (3 - r_{i:jh}) - 3\binom{t-s}{3} .$$

Now it may be shown that $\sqrt{n}(p - \pi)$, where p is the estimator of π, is normally distributed with mean zero and variance c^2, where

$$(5.29)\quad 1/c^2 = E\left(\frac{\partial \ln f}{\partial \pi}\right)^2 .$$

The details of the proof will be omitted, but we shall indicate the results that are obtained in calculating this variance.

Proceeding as in Section 5.2, we find

$$(5.30)\quad \frac{\partial \ln f}{\partial \pi} = \frac{X_a}{\pi} - \frac{sX_b}{(t-s)\theta} - E\left[\frac{X_a}{\pi} - \frac{sX_b}{(t-s)\theta}\right],$$

and it follows that

$$(5.31)\quad E\left[\frac{\partial \ln f}{\partial \pi}\right]^2 = \frac{\mathrm{Var}(X_a)}{\pi^2} + \frac{s^2\,\mathrm{Var}(X_b)}{(t-s)^2\,\theta^2} - \frac{2s\,\mathrm{Cov}(X_a,X_b)}{(t-s)\pi\theta}.$$

Substituting appropriate expressions for the variances and covariance in (5.31) and simplifying the result, we obtain

(5.32)

$$E\left[\frac{\partial \ln f}{\partial \pi}\right]^2 = \frac{s(t-s)(t-2)}{2} \cdot \frac{\pi^3\theta+4\pi^2\theta^2+\pi\theta^3}{(\pi^2+\pi\theta+\theta^2)^2}\left[\frac{1}{\pi}+\frac{s}{(t-s)\theta}\right]^2 .$$

For any given value of π, the variance of $\sqrt{n}(p - \pi)$ can be found as the reciprocal of (5.32). Under the null hypothesis that $\pi = 1/t$, we obtain

$$(5.33) \quad \mathrm{Var}\left[\sqrt{n}(p - \pi)\right] = \frac{3(t - s)}{t^4 s(t - 2)} .$$

5.4 Large-Sample Distribution of $T = -2 \ln \lambda_1$, Case (i)

In Section 3.4 we indicated that, under the null hypothesis of Case (i), $T = -2 \ln \lambda_1$ is asymptotically distributed as chi-square with $(t - 1)$ degrees of freedom. We shall now verify this while investigating the power of the test procedure based on T. Relative efficiencies of the triple comparisons test will then be obtained in comparison with other test procedures in the following chapter. In order to begin this development, we first consider the expansion of T in terms of powers of $(p_i - 1/t)$, $i=1,\ldots,t$.

5.4.1 Series Expansion of T

To obtain the required series expansion of T, we rewrite T as

$$(5.34)\quad T = 2n\binom{t}{3} \ln 6 + 2\sum_i a_i \ln p_i - 2n \sum_{i<j<h} \ln \hat{D}_{ijh},$$

where p_i is the estimator of $_i$,

$$a_i = \sum_{k=1}^{n} \sum_{j<h}{}' (3 - r_{ik:jh}),$$

and

$$\hat{D}_{ijh} = p_i^2(p_j + p_h) + p_j^2(p_i + p_h) + p_h^2(p_i + p_j).$$

Let

$$(5.35)\quad y_i = t(p_i - 1/t),\ i = 1,\ldots,t.$$

Then

$$p_i = (1 + y_i)/t,$$

and

$$(5.36)\quad T = 2n\binom{t}{3} \ln 6 + 2\sum_i a_i[\ln(1+y_i) - \ln t]$$
$$- 2n \sum_{i<j<h} [\ln(1+u_{ijh}) + \ln 6 - 3 \ln t],$$

where

$$(5.37)$$
$$u_{ijh} = \tfrac{1}{6}[6(y_i+y_j+y_h)+4(y_iy_j+y_iy_h+y_jy_h)+2(y_i^2+y_j^2+y_h^2)$$
$$+ y_i^2(y_j+y_h)+y_j^2(y_i+y_h)+y_h^2(y_i+y_j)].$$

By use of the result that

$$(5.38)\quad \sum_i a_i = 3n\binom{t}{3},$$

it is easily seen that

$$(5.39)\quad T = 2\sum_i a_i \ln(1+y_i) - 2n \sum_{i<j<h} \ln(1+u_{ijh}).$$

By rearranging (3.9) we obtain

$$a_i = n \sum_{j<h}{}' [2p_i^2(p_j+p_h)+p_i p_j^2+p_i p_h^2][\hat{D}_{ijh}]^{-1}.$$

It follows then from (5.35) and (5.37) that

$$(5.40)\quad a_i = \frac{n(1+y_i)}{6} \sum_{j<h}{}' [w_{ijh}][1+u_{ijh}]^{-1}, \text{ where}$$

$$(5.41)\quad w_{ijh} = 6 + 4(y_i+y_j+y_h)+2(y_i y_j+y_i y_h)+y_j^2+y_h^2.$$

Now T may be written

$$(5.42)\quad T = \frac{n}{3} \sum_i (1+y_i)\ln(1+y_i) \sum_{j<h}{}' (w_{ijh})(1+u_{ijh})^{-1}$$

$$-2n \sum_{i<j<h} \ln(1+u_{ijh}).$$

From the definition of y_i, it is obvious that

$$\sum_i y_i = 0 \quad \text{and} \quad 2\sum_{i<j} y_i y_j = -\sum_i y_i^2.$$

By making use of these relations and the identity $(1 + X)^{-1} = 1 - X + X^2/(1 + X)$, as well as the usual Taylor expansions for the logarithmic functions, we obtain

$$(5.42)\quad T = \frac{nt(t-2)}{3} \sum_i y_i^2 + R(y),$$

where R(y) involves powers of the variates higher than the second.

We next consider the function R(y) which may be written as follows:

$$(5.43)\quad R(y) = \frac{n}{3}\sum_i \Bigg\{\frac{-(t-2)(t-3)}{2}\, y_i^3$$

$$+[3(t-1)(t-2)-(t-2)(t-3)y_i]\left[\frac{y_i^3+y_i^4}{3(1+\theta_i y_i)^3} - \frac{y_i^3}{2}\right]$$

$$+\left[y_i + \frac{y_i^2-y_i^3}{2} + \frac{y_i^3+y_i^4}{3(1+\theta_i y_i)^3}\right][(t-2)(\sum_j{}' y_j^2 - 2y_i^2)$$

$$-\sum_{j<h}{}'[w_{ijh} - 6]w_{ijh} + \frac{w_{ijh}u_{ijh}^2}{1+u_{ijh}})]\Bigg\}$$

$$-\frac{n}{3}\sum_{i<j<h}\Bigg\{y_i^3\frac{\partial^3 F_{ijh}}{\partial y_i^3} + y_j^3\frac{\partial^3 F_{ijh}}{\partial y_j^3} + y_h^3\frac{\partial^3 F_{ijh}}{\partial y_h^3}$$

$$+3y_i^2y_j\frac{\partial^3 F_{ijh}}{\partial y_i^2\partial y_j} + 3y_j^2y_i\frac{\partial^3 F_{ijh}}{\partial y_j^2\partial y_i} + 3y_i^2y_h\frac{\partial^3 F_{ijh}}{\partial y_i^2\partial y_h}$$

$$+3y_j^2y_h\frac{\partial^3 F_{ijh}}{\partial y_j^2\partial y_h} + 3y_h^2y_j\frac{\partial^3 F_{ijh}}{\partial y_h^2\partial y_j} + 3y_h^2y_i\frac{\partial^3 F_{ijh}}{\partial y_h^2\partial y_i}$$

$$+6y_iy_jy_h\frac{\partial^3 F_{ijh}}{\partial y_i\partial y_j\partial y_h}\Bigg\},$$

where $0<\theta_i<1, F_{ijh} = \ln(1+u_{ijh})$, and the derivatives of F_{ijh} are evaluated at some point $(\theta_{ijh}y_i, \theta_{ijh}y_j, \theta_{ijh}y_h)$ for $0<\theta_{ijh}<1$. We note that each of the derivatives involved is a rational function of y_i, y_j, and y_h.

Now we make the assumption, which we shall examine in the next subsection, that $\sqrt{n}\, y_i$ $(i=1,\ldots,t)$ have a joint limiting distribution. Assuming this result, we can easily see that $(\sqrt{n}\, y_i)^c n^{-\epsilon}$, for any $\epsilon > 0$ and positive integer c, converges in probability to zero. The function R(y) of (5.43) is a rational function of terms of the form $(\sqrt{n}\, y_i)^c n^{-\epsilon}$ in which c is a positive integer and $\epsilon \geq 1/2$. We may conclude then from Slutsky's Theorem (Cramér, 1946, p. 255) that, if the $\sqrt{n}\, y_i$ $(i=1,\ldots,t)$ have a limiting distribution, R(y) converges stochastically to zero. It follows now from a convergence theorem of Cramér (1946, p. 254) that T has the same limiting distribution as $\frac{nt(t-2)}{3} \sum_i y_i^2$.

5.4.2 Limiting Distribution of $\sqrt{n}\, y_i$

In order to determine the joint limiting distribution of $\sqrt{n}\, y_i$ $(i=1,\ldots,t)$, we make use of the following relation:

(5.44)

$$\frac{a_i}{n} = \frac{1}{6}(1+y_i) \sum_{j<h}{}' \frac{6+4(y_i+y_j+y_h)+2(y_iy_j+y_iy_h)+y_j^2+y_h^2}{1+u_{ijh}}$$

$$= \frac{1}{6}(1+y_i) \sum_{j<h}{}' w_{ijh}(1+u_{ijh})^{-1}.$$

From this we obtain

$$(5.45)\quad \frac{a_i}{n} = \frac{1}{6}(1 + y_i) \sum_{j<h}{}' w_{ijh}\left(1 - u_{ijh} + \frac{u_{ijh}^2}{1+u_{ijh}}\right).$$

If we again make use of the relation $\sum_i y_i = 0$ to obtain $\sum_j{}' y_j = -y_i$, we find

$$(5.46)\quad \frac{a_i}{n} = \binom{t-1}{2} + \frac{t(t-2)}{3} y_i + R_2(y),$$

where

$$(5.47)\quad R_2(y) = \frac{1}{6} \sum_{j<h}{}' \Big\{(y_j^2+y_h^2+2y_iy_j+2y_iy_h)$$

$$-(w_{ijh}u_{ijh} - 6[y_i+y_j+y_h])+y_i(w_{ijh}-6)$$

$$+(1+y_i)\frac{w_{ijh}u_{ijh}^2}{1+u_{ijh}} - y_iw_{ijh}u_{ijh}\Big\}.$$

Rearranging (5.46) and multiplying through by $\sqrt{n}$, we find

$$(5.48)\quad \frac{3\sqrt{n}}{t(t-2)}\left[\frac{a_i}{n} - \binom{t-1}{2}\right] = \sqrt{n}\, y_i + \frac{3\sqrt{n}}{t(t-2)} R_2(y).$$

In Equation (5.48), we again have $\sqrt{n}\, y_i$ occurring with a remainder that converges stochastically to zero provided $\sqrt{n}\, y_i$ $(i=1,\dots,t)$ have a limiting distribution. The problem of proving that $\sqrt{n}\, y_i$ $(i=1,\dots,t)$ have a limiting distribution appears to be very complex and has not been solved in its entirety. Some aspects of this problem will be discussed in Appendix B. We shall find the form of the limiting distribution of the $\sqrt{n}\, y_i$ here by noting that, if $\sqrt{n}\, y_i$ have a limiting distribution, then $\sqrt{n}\, R_2(y)$ converges stochastically to zero, and the limiting distribution of $\sqrt{n}\, y_i$ $(i=1,\ldots,t)$ is the same as that of

$$\frac{3\sqrt{n}}{t(t-2)}\left[\frac{a_i}{n} - \binom{t-1}{2}\right], \quad (i = 1,\ldots,t) .$$

We shall determine this distribution by finding the limiting characteristic function in the following paragraphs.

In order to obtain the joint limiting distribution of $\sqrt{n}\, y_i$ $(i=1,\ldots,t)$, we first consider the joint distribution of z_i $(i=1,\ldots,t)$, where

$$(5.49) \qquad z_i = \sqrt{n}\left[\frac{a_i}{n} - \frac{(t-1)(t-2)}{2}\right] .$$

In (5.8) we defined $x_{ik:jh} = 3 - r_{ik:jh}$. In terms of these variables, we can write

$$(5.50) \qquad z_i = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{j<h}{}' (x_{ik:jh} - 1) .$$

Now the characteristic function of $z_1,\ldots,z_t$ may be written as

$$(5.51) \qquad \phi_n(\tau) = E\left[\exp(\sqrt{-1} \sum_i z_i \tau_i)\right]$$

$$= E\left[\exp\left(\frac{\sqrt{-1}}{\sqrt{n}} \sum_{k=1}^{n} \sum_i \sum_{j<h}{}' [x_{ik:jh} - 1]\tau_i\right)\right]$$

From the independence of the $x_{ik:jh}$ for the different repetitions $(k = 1,\ldots,n)$, we have

$$(5.52)\quad \phi_n(\tau) = \left[E\left\{\exp(\sqrt{-1}/\sqrt{n}\ \sum_i \sum_{j<h}{}' [x_{i:jh}1]\tau_i)\right\}\right]^n,$$

where $x_{i:jh}$ is representative of the set of random variables $x_{ik:jh}$ $(k = 1,\ldots,n)$. Expanding the exponential function, we obtain

$$(5.53)\quad \phi_n(\tau) = \left[E\left\{1 - \frac{\sqrt{-1}}{\sqrt{n}} \sum_i \sum_{j<h}{}' (x_{i:jh}1)\tau_i\right.\right.$$

$$\left.\left. - \frac{1}{2n}(\sum_i \sum_{j<h}{}' [x_{i:jh}1]\tau_i)^2 + \frac{1}{n^{3/2}} R_1'\right\}\right]^n$$

It is now necessary to redefine the parameters under the alternative hypothesis of Case (i) so that

$$(5.54)\quad \pi_i = 1/t + \delta_{in}/\sqrt{n} \qquad (i=1,\ldots,t),$$

where δ_{in} represents a sequence of constants converging to δ_i as $n \rightarrow \infty$. This refefinition of parameters permits us to find the limiting distribution of T and the power of the test procedure for alternatives in the locality of the parameter point determined by the null hypothesis. This is the technique that is commonly used in investigations of the limiting power functions of test procedures. The results obtained in this way are useful as approximations to the power functions of the tests involved. They are also useful in obtaining the local asymptotic relative efficiency of two different tests.

In taking the expectations indicated in (5.53), let us first consider

$$(5.55)\quad E\sum_{j<h}{}'(x_{i:jh}1) = \sum_{j<h}{}'\left[P(r_{i:jh}1) - P(r_{i:jh}3)\right]$$

$$= \sum_{j<h}{}'\left[\pi_i^2(\pi_j+\pi_h)-(\pi_j^2\pi_h+\pi_h^2\pi_j)\right] D_{ijh}^{-1}.$$

Substitution of $1/t + \delta_{in}/\sqrt{n}$ for π_i results in

$$E\sum_{j<h}{}'(x_{i:jh}1) = \sum_{j<h}\frac{t(2\delta_{in}-\delta_{jn}-\delta_{hn})}{3\sqrt{n}} + \frac{R_2'}{\sqrt{n}},$$

where $\lim\limits_{n\to\infty}\sqrt{n}\, R_2' = M_2$. Since $\sum\limits_i \pi_i = 1$, it is clear from the definition of δ_{in} that $\sum\limits_i \delta_{in} = 0$. It follows that

$$\sum_{j<h}{}'(2\delta_{in}-\delta_{jn}-\delta_{hn}) = (t-1)(t-2)\delta_{in} - (t-2)\sum_j{}'\delta_{jn}$$

$$= (t-1)(t-2)\delta_{in}+(t-2)\delta_{in},$$

and

$$(5.57)\quad E\sum_{j<h}(x_{i:jh}1) = \frac{t^2(t-2)\delta_{in}}{3\sqrt{n}} + \frac{R_2'}{\sqrt{n}}.$$

A similar but somewhat more tedious procedure produces the result that

$$(5.58)\quad E\Big[\sum_i \sum_{j<h}{}' (x_{i:jh}1)\tau_i\Big]^2 = \sum_i \frac{(t-1)(t-2)\tau_i^2}{3} - \sum_{i\neq j} \frac{(t-2)\tau_i\tau_j}{3} + R_3'$$

where $\lim_{n\to\infty} \sqrt{n}\, R_3' = M_3$. As a result of these computations, we may write

$$(5.59)\quad \phi_n(\tau) = \left[1 + \frac{\sqrt{-1}}{3n}\sum_i t^2(t-2)\tau_i\delta_{in} - \frac{1}{2n}\left\{\sum_i \frac{(t-1)(t-2)\tau_i}{3} - \sum_{i\neq j}\frac{(t-2)\tau_i\tau_j}{3}\right\} + \frac{R}{n}\right]^n,$$

where $R \to 0$ as $n \to \infty$. Letting $n \to \infty$, we obtain, from (5.59),

$$(5.60)\quad \phi(\tau) = \exp\left\{\frac{\sqrt{-1}}{3}\sum t^2(t-2)\delta_i\tau_i - \frac{1}{2}\left[\frac{(t-1)(t-2)}{3}\sum_i \tau_i^2 - \frac{(t-2)}{3}\sum_{i\neq j}\tau_i\tau_j\right]\right\},$$

where we have used $\lim_{n\to\infty} \phi_n(\tau) = \phi(\tau)$.

We note now that ϕ is the moment generating function of a multivariate normal distribution with means $t^2(t-2)\delta_i/3$ $(i=1,\ldots,t)$, and with variances and covariances equal to $(t-1)(t-2)/3$ and $-(t-2)/3$ respectively. According to Cramér (1946, p. 102), this is sufficient to conclude that

$z_1,\ldots,z_t$ have a joint limiting multivariate normal distribution with the means, variances, and covariances given above. It follows that

$$\frac{3\sqrt{n}}{t(t-2)}\left[\frac{a_i}{n} - \binom{t-1}{2}\right] = \frac{3z_i}{t(t-2)} \quad (i=1,\ldots,t)$$

have a joint limiting multivariate normal distribution with means $t\delta_i$ $(i=1,\ldots,t)$, and with variances and covariances equal to $3(t-1)/t^2(t-2)$ and $-3/t^2(t-2)$ respectively.

5.4.3 Limiting Distribution of T under H_a

We now use the results of Subsection 5.4.2 to determine the limiting distribution of T which, as was shown, must be the same as that of $\frac{nt(t-2)}{3}\sum_i y_i^2$. Let

$$(5.61) \quad u_i = \sqrt{nt(t-2)/3}\; y_i \quad (i=1,\ldots,t).$$

It follows that T has the same limiting distribution as $\sum_i u_i^2$. As functions of $\sqrt{n}\, y_1,\ldots,\sqrt{n}\, y_t$, it is easy to see that $u_1,\ldots,u_t$ have a joint limiting distribution that is multivariate normal with

$$(5.62) \quad E(u_i) = \sqrt{t(t-2)/3}\; E(\sqrt{n}\, y_i)$$

$$= t\sqrt{t(t-2)}\,\delta_i ,$$

$$(5.63)\quad \mathrm{Var}(u_i) = \frac{t(t-2)}{3}\,\mathrm{Var}(\sqrt{n}\; y_i)$$

$$= (t-1)/t \;,$$

and

$$(5.64)\quad \mathrm{Cov}(u_i, u_j) = -1/t \;.$$

Now we let

$$Z' = Gu' \;,$$

where Z' and u' are column vectors with components $Z_1, \ldots, Z_t$ and $u_1, \ldots, u_t$ while G is the Helmert matrix

$$G = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \ldots & 0 & \ldots & 0 \\ \vdots & & & & & \\ \frac{1}{\sqrt{i(i+1)}} & \frac{1}{\sqrt{i(i+1)}} & \ldots & \frac{-i}{\sqrt{i(i+1)}} & 0 \ldots & 0 \\ \vdots & & & & & \\ \frac{1}{\sqrt{t}} & \frac{1}{\sqrt{t}} & \ldots & \frac{1}{\sqrt{t}} & \ldots & \frac{1}{\sqrt{t}} \end{bmatrix}$$

From this transformation, we obtain $Z_t = 0$. Consequently, T has the same limiting distribution as $\sum_{i=1}^{t-1} Z_i^2$. Since

$$Z_i = \frac{1}{\sqrt{i(i+1)}} \sum_{j=1}^{i} u_j - \frac{i}{\sqrt{i(i+1)}}\, u_{i+1} \qquad (i=1,\ldots,[t-1]),$$

we find

$$\mathrm{Var}(Z_i) = 1 \quad (i=1,\ldots, t-1)$$

and

$$\mathrm{Cov}(Z_i, Z_j) = 0 \quad (i\neq j,\ i,\ j=1,\ldots, t-1) .$$

We also know that

$$\sum_{i=1}^{t-1} \left[E(Z_i)\right]^2 = \sum_{i=1}^{t} \left[E(u_i)\right]^2 = \frac{t^3(t-2)}{3} \sum_{i=1}^{t} \delta_i^2 .$$

This is sufficient to conclude that, under the alternate hypothesis that $\pi_i = t^{-1} + n^{-\frac{1}{2}}\delta_{in}$, $\sum_{i=1}^{t-1} Z_i^2$, and hence T, has a limiting non-central chi-square distribution with (t-1) degrees of freedom and parameter of non-centrality

$$\lambda = \frac{t^3(t-2)}{3} \Sigma\delta_i^2 .$$

The power of the triple comparisons test based on T is asymptotically given by

$$(5.66) \quad \beta(\lambda|\alpha,(t-1),\infty) = \int_{\chi^2_{\alpha,(t-1)}}^{\infty} f(T)\, dT ,$$

where $\chi^2_{\alpha,(t-1)}$ is the α-level significance value of a central chi-square distribution with (t-1) degrees of freedom and

$$(5.67) \quad f(T) = \frac{\exp\left[-\frac{1}{2}(T+\lambda)\right]}{2^{\frac{1}{2}(t-1)}} \sum_{h=0}^{\infty} \frac{T^{\frac{1}{2}(t-1)+h-1}\ \lambda^h}{\Gamma\left[\frac{1}{2}(t-1)+h\right] 2^{2h} h!} .$$

In (5.66), we have followed Bradley (1955, p. 459) in his modification of the notation $\beta(\lambda\ \alpha, \nu_1, \nu_2)$ as used by E. S. Pearson and H. O. Hartley (1951). In this paper, Pearson and Hartley provided charts for the non-central F distribution for given degrees of freedom and $\alpha = 0.01$ and 0.05. They plotted β against ϕ where $\phi = \sqrt{\lambda/t}$. Tables which may be used to evaluate (5.66) have been provided by P. C. Tang (1938) and E. Fix (1949).

An approximation to the power of the triple comparisons test for large n may be obtained by assuming the non-central chi-square distribution and setting

$$\frac{\delta_i}{\sqrt{n}} = \pi_i - 1/t \ .$$

Then

$$(5.68) \quad \lambda = \frac{nt^3(t-2)}{3} \sum_i (\pi_i - 1/t)^2 \ .$$

In Section 3.4, we indicated that under the null hypothesis T is distributed approximately with a central chi-square distribution for large n. This follows directly from the proceding discussion since $\pi_i = 1/t$ and $\delta_{in} = \delta_i = 0$ $(i=1,\ldots,t)$ under H_0. As a result, $\lambda = 0$ and the limiting distribution of T is that of the central chi-square distribution with $(t - 1)$ degrees of freedom.

VI. COMPARISONS OF POWERS OF TESTS

6.1 Comparison of the Test for Triple Comparisons with the Test for Paired Comparisons

The investigation of the power of a statistical test is very useful as a means of assessing the merits of the test in contrast to alternative procedures. An important alternative of the method of triple comparisons is the Bradley-Terry method of paired comparisons. We shall first compare these two methods for alternative hypothesis (i). Then the triple comparisons test based on T will be compared with two other possible tests.

The developments in Chapter V have paralleled similar results that were obtained by Bradley (1955) in his investigation of the power of the test for paired comparisons. He found that the test statistic for paired comparisons, which corresponds to the statistic T for triple comparisons, has a limiting non-central chi-square distribution with (t - 1) degrees of freedom. The parameter of non-centrality for the paired comparisons test is given by the equation

$$(6.1) \quad \lambda_2' = t^3 \sum_i (\delta_i')^2/4 ,$$

where the Case (i) alternative hypothesis is

$$(6.2) \quad \pi_i = 1/t + \delta_{in'}'/\sqrt{n'} \text{ (with } \lim_{n' \to \infty} \delta_{in'}' = \delta_i').$$

To compare the power of the triple comparisons test with the power of the paired comparisons test, we need an association of the δ_i' of paired comparisons with the δ_i of (5.54) for triple comparisons. In order to obtain the association needed, we note that for paired comparisons

$$(6.3) \quad P(r_i<r_j) = \pi_i/(\pi_i+\pi_j) \ .$$

Substituting from (6.2) for the parameters in (6.3), we find

$$(6.4) \quad P(r_i<r_j) = \frac{1}{2} + \frac{t}{4}\left(\frac{\delta_{in'}' - \delta_{jn'}'}{\sqrt{n'}}\right) + R_1 \ ,$$

where R_1 involves terms in $(n')^{-\frac{1}{2}m}$, $m \geq 2$, and n' is the number of complete repetitions in the paired comparisons experiment.

From the probability model for triple comparisons, we obtain

$$(6.5) \quad P(r_i<r_j) = P(r_i<r_j<r_h)+P(r_i<r_h<r_j)+P(r_h<r_i<r_j)$$

$$= (\pi_i^2\pi_j+\pi_i^2\pi_h+\pi_h^2\pi_i)/D_{ijh} \ .$$

Under the alternative hypothesis considered in the discussion of the power of the triple comparisons test,

$$(6.6) \quad \pi_i = 1/t + \delta_{in}/\sqrt{n} \ ,$$

where n is the number of complete repetitions in the triple comparisons experiment, and $\lim_{n\to\infty} \delta_{in} = \delta_i$. Substitution for π_i in (6.5) results in

$$(6.7)\quad P(r_i < r_j) = \frac{1}{2} + \frac{t}{3}\left(\frac{\delta_{in} - \delta_{jn}}{\sqrt{n}}\right) + R_2 ,$$

where R_2 involves terms in $[n^{-\frac{1}{2}}]^m$, $m \geq 2$.

The alternative hypotheses indicated by (6.2) and (6.6) provide essentially equal probabilities for (6.4) and (6.7) when

$$(6.8)\quad \frac{t}{3}\left(\frac{\delta_{in} - \delta_{jn}}{\sqrt{n}}\right) = \frac{t}{4}\left(\frac{\delta'_{in'} - \delta'_{jn'}}{\sqrt{n'}}\right) (i \neq j;\ i,j = 1,\ldots,t).$$

The n and n' of (6.8) have been defined as the number of complete repetitions respectively in triple and paired comparisons experiments. Since power comparisons of tests should be made in terms of comparable numbers of replications, we define n_2 and n_3 to be the number of replications respectively in a paired comparisons and triple comparisons experiment. It follows that

$$(6.9)\quad n_2 = n'(t-1) \quad \text{and}$$
$$n_3 = n(t-1)(t-2)/2 .$$

Using the relations in (6.9) to substitute for n' and n in (6.8), we obtain

$$(6.10)\quad \frac{\sqrt{(t-2)}\,(\delta_{in} - \delta_{jn})}{3\sqrt{2n_3}} = \frac{\delta'_{in'} - \delta'_{jn'}}{4\sqrt{n_2}} \ (i,j,\ldots t; i \neq j).$$

For (6.10) to hold for all i and j, we must have

$$(6.11)\quad \frac{\sqrt{t-2}\ \delta_{in}}{3\sqrt{2n_3}} = \frac{\delta'_{in'}}{4\sqrt{n_2}} .$$

As a result,

$$(6.12)\quad \frac{(t-2)\sum\limits_i \delta_{in}^2}{9n_3} = \frac{\sum\limits_i (\delta'_{in})^2}{8n_2} .$$

We are interested in the relative efficiency of the two test as given by the limit of the ratio n_2/n_3 as the number of replications increases while the powers of the two tests remain equal. To find this ratio, we let λ_3 represent the parameter of non-centrality for triple comparisons; that is,

$$(6.13)\quad \lambda_3 = \frac{t^3(t-2)}{3} \sum_i \delta_i^2 .$$

If we now set $\lambda'_2 = \lambda_3$, thus requiring that the two tests have equal powers, we find

$$(6.14)\quad \frac{t^3}{4}\sum_i (\delta'_i)^2 = \frac{t^3(t-2)}{3}\sum_i \delta_i^2 , \text{ or}$$

$$\frac{n_2}{4}\sum_i \frac{(\delta'_i)^2}{n_2} = \frac{n_3(t-2)}{3}\sum_i \frac{\delta_i^2}{n_3}$$

It follows that

$$(6.15)\quad \frac{n_2}{n_3} = \frac{4(t-2)}{3}\,\frac{\sum\limits_i \delta_i^2/n_3}{\sum\limits_i (\delta'_i)^2/n_2}$$

From (6.12) and the requirement for equal alternatives, we now get

$$(6.16)\quad \lim_{n_2,n_3\to\infty} n_2/n_3 = [4(t-2)/3]\,[9/8(t-2)] = 3/2 .$$

Equation (6.16) gives the relative efficiency of triple comparisons to paired comparisons as 3 to 2 or 150%. This indicates that $1\frac{1}{2}$ times as many replications are required for the method of paired comparisons as for the method of triple comparisons for tests of equal powers. In terms of complete repetitions this means that 3(t-2) complete repetitions of a paired comparisons experiment are required to provide a test with power equal to that of a test based on four complete repetitions of a triple comparisons experiment.

6.2 Efficiency of the Triple Comparisons Test Relative to Analysis of Variance

Since the analysis of variance is used extensively, it is of interest to compare the triple comparisons test based on T with analysis of variance. In order to make this comparison, let n'' be the number of replications of a randomized block design while n_2 and n_3 represent the number of replications for paired and triple comparisons experiments respectively. Furthermore, let $\delta''_{in''}$, corresponding to $\delta'_{in'}$ for paired comparisons, be a sequence of constants associated with the analysis of variance such that $\lim_{n''\to\infty} \delta''_{in''} = \delta''_i$. For treatment differences in a paired comparisons test and an analysis of variance test to be comparable with the ratio n''/n' fixed, Bradley (1955) showed that

$$(6.18) \qquad \lim_{n',n''\to\infty}\left[\sum_i \frac{(\delta''_{in''})^2}{n''} \Big/ \sum_i \frac{(\delta'_{in'})^2}{n'}\right] = 1 .$$

He also showed that the parameter of non-centrality, λ''_r, for the analysis of variance test can be written as

$$(6.19) \qquad \lambda''_r = \frac{1}{4}\pi t^2(t-1)\sum_i(\delta''_i)^2 .$$

According to O. Kempthorne (1952, p. 533), the parameter of non-centrality, λ'', for incomplete block designs may be obtained from (6.19) by multiplying by the efficiency factor, $(k-1)t/k(t-1)$, to adjust for the efficiency of incomplete block designs relative to randomized block designs. The symbol k in the efficiency factor represents the block size and t represents the number of treatments. It follows that for blocks of size three

$$(6.20) \qquad \lambda'' = \frac{\pi t^3}{6}\sum(\delta''_i)^2 .$$

To obtain the relative efficiency of triple comparisons with respect to the analysis of variance, we must find the limit of the ratio n''/n_3 as the number of replications increases while the powers of the two tests remain equal. We do this by setting $\lambda'' = \lambda_3$ and solving for n''/n_3. We get $\frac{t^3(t-2)}{3}\Sigma\delta_i^2 = \frac{\pi t^3}{6}\Sigma(\delta_i'')^2$, or

$$(6.21)\quad \frac{n''}{n_3} = \frac{[t^3(t-2)/3]\,\Sigma\,\delta_i^2/n_3}{(\pi t^3/6)\Sigma(\delta_i'')^2/n''}$$

$$= \frac{2(t-2)}{\pi}\left[\frac{\Sigma\delta_i^2}{n_3}\Big/\frac{\Sigma(\delta_i'')^2}{n''}\right].$$

From (6.18) and the requirement of equal alternatives, we may rewrite this as follows:

$$(6.22)\quad \frac{n''}{n_3} = \frac{2(t-2)}{\pi}\left[\frac{\Sigma\delta_i^2}{n_3}\Big/\frac{\Sigma(\delta_i')^2}{n_2}\right].$$

Now, if we make use of (6.12), we find the limiting ratio

$$(6.23)\quad \frac{n''}{n_3} = \frac{2(t-2)}{\pi}\cdot\frac{9}{8(t-2)}$$

$$= \frac{9}{4\pi}.$$

The asymptotic relative efficiency of triple comparisons to the analysis of variance is 9 to 4π or 71.6%. The comparison was made though when the analysis of variance would be the appropriate method. The triple comparisons test was formulated largely for subjective tests as an extension of paired comparisons and may be useful in many cases where the analysis of variance is not applicable.

6.3 Triple Comparisons Compared with a Multi-Multinomial Test

The triple comparisons analogue of the multi-binomial test (Bradley, 1955, p. 460) for the rank analysis of designs having a block size of two is a multi-multinomial test for use with blocks of size three. This latter test is suggested by the most general model available for triple comparisons, which was discussed early in Chapter IV. Following the notation introduced there, we have

(6.24) π_{ijh} = probability that treatments i, j, and h rank 1, 2, and 3 respectively in a given block,

f_{ijh} = number of times treatments i, j, and h rank 1, 2, 3 respectively in n complete repetitions.

Under the hypothesis of no treatment differences $\pi_{ijh} = 1/6$ and the expected value of f_{ijh} is n/6. The test involving treatments i, j, and h that is usually made is based on the following statistic:

$$(6.25)\quad S_{ijh} = \frac{(f_{ijh}-n/6)^2 + (f_{ihj}-n/6)^2 + (f_{hji}-n/6)^2}{n/6}$$

$$+ \frac{(f_{hij}-n/6)^2 + (f_{jih}-n/6)^2 + (f_{jhi}-n/6)^2}{n/6}$$

$$= 6n\Sigma''(p_{ijh} - 1/6)^2 ,$$

where Σ'' indicates that the summation is over the six permutations of i, j, and h; and

(6.26) $p_{ijh} = f_{ijh}/n$.

For large n, the appropriate test is that of chi-square with 5 degrees of freedom.

A test statistic for n complete repetitions may be obtained by regarding the repetitions as $t(t-1)(t-2)/6$ unrelated multinomial tests, each based on n trials. The usual statistic for n complete repetitions is

(6.27) $S = \sum_{i<j<h} S_{ijh}$,

which has a limiting chi-square distribution with $5t(t-1)(t-2)/6$ degrees of freedom.

In order to compare this test with that developed for triple comparisons, it is necessary to assume that the latter is applicable; that is,

(6.28) $\pi_{ijh} = \pi_i^2\pi_j/D_{ijh}$.

The alternative hypothesis for the test based on T was stated $\pi_i = 1/t + \delta_{in}/\sqrt{n}$ and the hypothesis analogous to that for the statistic S is

(6.29) $\pi_{ijh} = 1/6 + \mu_{ijhn}/\sqrt{n}$, where $\lim_{n\to\infty} \mu_{ijhn} = \mu_{ijh}$.

Since (6.28) is assumed appropriate, we may replace π_i with $1/t + \delta_{in}/\sqrt{n}$ and find

$$(6.30)\quad \pi_{ijh} = 1/6 + \frac{t(\delta_{in} - \delta_{hn})}{6\sqrt{n}} + R,$$

where R involves terms in $(n^{-\frac{1}{2}})^m$ with $m \geq 2$. It follows that, for large values of n, we may identify μ_{ijhn} with $t(\delta_{in} - \delta_{hn})/6$ or

$$(6.31)\quad \mu_{ijh} = t(\delta_i - \delta_h)/6$$

This relation of parameters will permit a comparison of the powers of the tests based on S and T.

Let $y_{ijh} = \sqrt{6n}\,(p_{ijh} - 1/6)$. By arguments similar to those in Chapter V, under the hypothesis of (6.29),

$$E(y_{ijh}) = \sqrt{6}\,\mu_{ijh} = t(\delta_i - \delta_h)/\sqrt{6}$$

Var $(y_{ijh}) = 5/6$, and Cov $(y_{ijh}, y_{\alpha\beta\gamma}) = -1/6$, where α, β, and γ are values of i, j, and h so that $y_{ijh} \neq y_{\alpha\beta\gamma}$. For large n, the y's have a singular multivariate normal distribution.

We now define $u_1 = y_{ijh}$, $u_2 = y_{ihj}$, $u_3 = y_{hji}$, $u_4 = y_{hij}$, $u_5 = y_{jih}$, and $u_6 = y_{jhi}$. With this notation, the transformation $z' = \Gamma u'$, where Γ is a Helmert matrix, results in $z_1, \ldots, z_5$ being asymptotically normally and independently distributed. We obtain

$$z_i = [i(i+1)]^{-\frac{1}{2}} \sum_{j=1}^{i} u_j - i[i(i+1)]^{-\frac{1}{2}}\, u_{i+1} \text{ for } i=1,\ldots,5;$$

$$z_6 = 6^{-\frac{1}{2}} \sum_{i=1}^{6} u_i = 0; \text{ and } \mathrm{Var}(z_i) = 1 \quad (i=1,\ldots,5).$$

It follows that $S_{ijh} = \sum_{i=1}^{5} z_i^2$ has a limiting chi-square distribution with five degrees of freedom and parameter of non-centrality

$$(6.32)\quad \lambda_{ijh} = \sum_{i=1}^{5} [E(z_i)]^2 = \Sigma''[E(y_{ijh})]^2$$

$$= t^2\ \Sigma''(\delta_i - \delta_h)^2/6$$

$$= \frac{2t^2}{3}\ (\delta_i^2+\delta_j^2+\delta_h^2) - (\delta_i\delta_j+\delta_i\delta_h+\delta_j\delta_h)\ .$$

From the independence of the S_{ijh} and the additive nature of chi-square it followsthat for the n complete repetitions

$$S = \sum_{i<j<h} S_{ijh}$$

has a limiting chi-square distribution with $5t(t-1)(t-2)/6$ degrees of freedom and parameter of non-centrality

$$(6.33)\quad \lambda_s = \sum_{i<j<h} \lambda_{ijh} = \frac{2t^2}{3}\left[\frac{(t-1)(t-2)}{2}\Sigma\delta_i^2 + \frac{(t-2)}{2}\Sigma\delta_i^2\right]$$

$$= \frac{t^3(t-2)}{3}\Sigma\delta_i^2$$

The parameter of non-centrality in (6.33) is the same as that found for the regular triple comparisons method based on T. Differences in power will be a result of the differences in the degrees of freedom for the two tests.

In order to compare the test procedure based on S with the triple comparisons method based on T, the relative efficiency may be obtained in terms of the efficacies of the tests according to the method of E. J. G. Pitman and G. Noether (1955). For a given statistic T_n, based on a sample of size n, Noether defines

$$(6.34)\quad E(T_n) = \psi_n(\theta, T_n)$$

$$(6.35)\quad Var(T_n) = \sigma_n^2(\theta, T_n).$$

The efficacy of the test of $H_o: \theta = \theta_o$ against the alternative $H: \theta \quad \theta_o$ is

$$(6.36)\quad R_n^{\frac{1}{m\delta}}(\theta_o, T_n) = [\psi_n^{(m)}(\theta_o, T_n)/\sigma_n(\theta_o, T_n)]^{\frac{1}{m\delta}}$$

where $\theta_n = \theta_o + k/n^{\delta}$, and m is the order of the first non-zero derivative of $\psi_n(\theta, T_n)$ evaluated at $\theta = \theta_o$.

For n complete repetitions, let T_n and S_n represent respectively the T and S statistics previously discussed for triple comparisons and the multi-multinomial test. If we define

$$(6.37)\quad \theta_n = \sum_i \delta_{in}^2/n$$

then T_n and S_n have asymptotically non-central chi-square distributions with the same parameter of non-centrality

$$\lambda = \lim_{n \to \infty} \frac{nt^3(t-2)}{3} \theta_n.$$

The degrees of freedom for tests based on T and S are (t-1) and 5t(t-1)(t-2)/6 respectively.

According to Noether, we may use

$\Psi_n(\theta, T_n) = (t-1)+nt^3(t-2)\theta/3,$

$\Psi_n(\theta, S_n) = 5t(t-1)(t-2)/6+nt^3(t-2)\theta/3,$

$\sigma_n^2(0, T_n) = 2(t-1)$, and

$\sigma_n^2(0, S_n) = 5t(t-1)(t-2)/3.$

Differentiating Ψ_n with respect to θ, we find, since $\theta_o = 0$

$$(6.38) \quad \Psi_n'(0, T_n) = nt^3(t-2)/3 = \Psi_n'(0, S_n).$$

This indicates that m = 1 and, from the definition of θ_n, $\delta = 1$. It follows that

$R_n(0, T_n) = nt^3(t-2)/3\sqrt{2(t-1)}$, and

$R_n(0, S_n) = nt^3(t-2)/\sqrt{15(t-1)(t-2)}.$

The relative efficiency of S to T is

$$(6.39) \quad \lim_{n \to \infty} \frac{R_n(0, S_n)}{R_n(0, T_n)} = \sqrt{\frac{6}{5t(t-2)}} .$$

For t = 3, the relative efficiency of S to T is approximately 63%. As t gets larger the relative efficiency of S to T obviously gets smaller. On the basis of this power comparison, it is apparent that, when it is applicable, the method of triple comparisons is more efficient than the multi-multinomial test.

VII. NUMERICAL EXAMPLES

7.1 Estimation and Test of Significance Illustrated

Since these test procedures for triple comparisons have not been used much in experiments, we do not have sufficient data from experiments to illustrate properly all of the techniques discussed. Some procedures will of necessity be illustrated by using hypothetical data, while others will be illustrated by using data from a consumer acceptance study of food variants.

We shall first consider a hypothetical experiment involving four treatments and 40 complete repetitions in which it is desired to test the null hypothesis

$$(7.1)\quad H_o:\pi_i = 1/4 \qquad (i=1,2,3,4)$$

against the alternative hypothesis

$$H_1:\text{no } \pi_i \text{ is assumed equal to } \pi_j \ (i\neq j;\ i,j=1,2,3,4).$$

The experimental results are given below in Table 6 in which f_{ijh} indicates the number of times the ranks of treatments i, j, and h were respectively 1, 2, and 3 when these three treatments were compared. The numbers in parentheses are the corresponding estimates of expected frequencies.

TABLE 6. Frequency (Expected Frequency) of Rankings in an Experiment with $t = 4$, $n = 40$.

$f_{123}=10$	(9)	$f_{124}=12$	(11)	$f_{134}=10$	(11)	$f_{234}=8$	(9)
$f_{132}=8$	(8)	$f_{142}=8$	(8)	$f_{143}=8$	(8)	$f_{243}=6$	(7)
$f_{213}=8$	(7)	$f_{214}=8$	(9)	$f_{341}=8$	(4)	$f_{342}=6$	(6)
$f_{231}=6$	(5)	$f_{241}=6$	(5)	$f_{314}=8$	(8)	$f_{324}=8$	(8)
$f_{312}=4$	(6)	$f_{412}=4$	(4)	$f_{413}=4$	(5)	$f_{423}=6$	(5)
$f_{321}=4$	(5)	$f_{421}=2$	(3)	$f_{431}=2$	(4)	$f_{432}=6$	(5)

Since a_i is equal to twice the number of times treatment i ranks first plus the number of times treatment i ranks second in the experiment, we find from Table 6 that

$$(7.2)\quad a_1 = 148,\ a_2 = 126,\ a_3 = 116, \text{ and } a_4 = 90.$$

For first approximations of the p_i, we substitute $t = 4$ and $n = 40$ into (3.19) to obtain

$$(7.3)\quad (7a_i-900)p_i^2 + (a_i-80)p_i + a_i = 0.$$

Substituting the values of a_i as given in (7.2) and solving, we obtain (after adjusting so that $\sum_i p_i = 1$) first approximations of

$$(7.4)\quad p_1 = .33,\ p_2 = .27,\ p_3 = .24, \text{ and } p_4 = .16.$$

A second set of approximations is obtained by using these first approximations in the right member of

$$(7.5)\quad p_i = a_i\left[n \sum_{j<h} \frac{2p_i(p_j+p_h)+p_j^2+p_h^2}{\hat{D}_{ijh}}\right]^{-1} \quad (i=1,\ldots,4),$$

where

$$\hat{D}_{ijh} = p_i^2(p_j+p_h) + p_j^2(p_i+p_h) + p_h^2(p_i+p_j).$$

This iterative procedure may be repeated until the desired precision is reached. Table 7 gives some indication of the speed of convergence of the estimators. The column under T shows how the value of the test statistic changes with changes in the estimators p_i. We shall use the values of the estimators correct to three decimal places as given in the fourth approximation.

TABLE 7. Successive Approximations of p_i and Corresponding Values of T

Approximation	p_1	p_2	p_3	p_4	T
First	.33	.27	.24	.16	14.54
Second	.32	.26	.23	.18	16.47
Third	.323	.259	.235	.183	16.61
Fourth	.322	.259	.236	.183	16.70
Fifth	.3216	.2594	.2358	.1832	16.71

The values of T in Table 7 were computed by substituting the a_i given in (7.2) and the p_i in the table into formula (3.22) to find $-2 \ln \lambda_1$. For a test of significance, we compare the computed T-value (16.70) with chi-square tables with three degrees of freedom. The result here indicates that the null hypothesis of equality of parameters should be rejected.

7.2 Goodness-of-Fit Test

We now use the data from the preceding section to illustrate the two methods of computing the statistic for the goodness-of-fit test. First we use the relation, from (4.9),

$$-2 \ln \lambda_I = 2\left[\sum_{i<j<h} \Sigma' f_{ijh} \ln f_{ijh} - n \binom{t}{3} \ln n + n \sum_{i<j<h} \ln \hat{D}_{ijh} - \sum_i a_i \ln p_i\right],$$

where Σ'' indicates the sum over the six permutations of i, j, and h. Substituting the observed frequencies from Table 6 and the values of the estimators p_i from line four of Table 7 into this equation, we find

(7.6) $-2 \ln \lambda_I = 7.94$

The approximation of $-2 \ln \lambda_I$ in terms of frequencies requires the calculation of expected frquencies, f^o_{ijh}, for the given model. They are obtained from

$$f^o_{ijh} = n\, \bar{p}^o_{ijh} = n\, p_i^2 p_j / \hat{D}_{ijh}.$$

These values have been computed for the data given in Section (7.1) and have been inserted in parentheses beside the observed frequencies in Table 6. Substituting the observed and expected frequencies from Table 6 into (4.21), that is,

$$-2 \ln \lambda_I = \sum_{i<j<h} \Sigma'' (f_{ijh} - f_{ijh}^o)^2 / f_{ijh}^o ,$$

we obtain

$$(7.7) \quad -2 \ln \lambda_I = 8.00$$

This result agrees very well with (7.6).

After computing $-2 \ln \lambda_I$, the goodness-of-fit test is completed by comparing the test statistic with chi-square tables with $\frac{5t(t-1)(t-2)}{6} - (t-1)$ degrees of freedom. Large values of the statistic $-2 \ln \lambda_I$ imply that the model does not fit the data. For the data that we have used with $t = 4$, we have 17 degrees of freedom and the value of the test statistic is 7.94, which indicates a very good fit of the model to the data in Table 6.

7.3 Estimation of Variances and Covariances

In order to illustrate the estimation of variances and covariances of estimators, we continue to use the data of Section (7.1) to obtain the variances and covariances of $\sqrt{n}(p_1-\pi_1)$, $\sqrt{n}(p_2-\pi_2)$, $\sqrt{n}(p_3-\pi_3)$, and $\sqrt{n}(p_4-\pi_4)$. From

Table 7, we obtain the estimates $p_1 = .322$, $p_2 = .259$, $p_3 = .236$, and $p_4 = .183$. These values of $p_i(i=1,2,3,4)$ are substituted for π_i in (5.13) to obtain $\hat{\lambda}_{ij}(i,j=1,2,3,4)$. From this computation, we obtain

$$(7.8)\quad \left[\hat{\lambda}_{ij}\right] = \begin{bmatrix} 17.93 & -7.80 & -8.44 & -9.69 \\ -7.80 & 29.01 & -10.94 & -13.27 \\ -8.44 & -10.94 & 36.12 & -15.98 \\ -9.69 & -13.27 & -15.98 & 56.82 \end{bmatrix}.$$

It follows from (5.6) that

$$(7.9)\quad \left[\hat{\lambda}_{ij}\right] = \begin{bmatrix} 94.13 & 71.98 & 74.05 \\ 71.98 & 112.37 & 75.13 \\ 74.05 & 75.13 & 124.90 \end{bmatrix}.$$

Inverting $\left[\hat{\lambda}'_{ij}\right]$, we obtain

$$(7.10)\quad \left[\hat{\lambda}'_{ij}\right]^{-1} = \begin{bmatrix} .025627 & -.010467 & -.008898 \\ -.010467 & .019161 & -.005322 \\ -.008898 & -.005322 & .016482 \end{bmatrix}.$$

If we now substitute values from (7.10) into (5.23) to find $\hat{\sigma}_{4j}$ (j=1,2,3,4), we get

$$(7.11)\quad \left[\hat{\sigma}_{ij}\right] = \begin{bmatrix} .025627 & -.010467 & -.008898 & -.006262 \\ -.010467 & .019161 & -.005322 & -.003372 \\ -.008898 & -.005322 & .016482 & -.002262 \\ -.006262 & -.003372 & -.002262 & .011896 \end{bmatrix}.$$

7.4 A Consumer Acceptance Study of Food Variants

As a final numerical example, we shall consider the results of an experiment (conducted by a firm which we are not at liberty to identify) concerning the relative merits of three kinds of orange juice. The objective of the experiment was to determine which of the three juices was most likely to be preferred by consumers. In an effort to determine this, the experiment was divided into two groups of 150 complete repetitions each. Since there were only three treatments, a complete repetition consisted of one ranking of the three juices. The results of these 300 repetitions are given below in Table 8 in which f_{ijh} refers to the ranking 1, 2, and 3 for juices i, j, and h respectively.

TABLE 8. Orange Juice Rankings

	Group One	Group Two	Total
f_{123}	62	88	150
f_{132}	17	21	38
f_{213}	29	17	46
f_{312}	6	2	8
f_{231}	15	7	22
f_{321}	8	2	10
Not ranked	13	13	26

The "not ranked" frequencies in Table 8 include two failures to express any preference and 24 rankings in which no preferences were expressed between two of the juices. In the analysis of the data, we shall ignore these frequencies and consider $n_1 = 137$, $n_2 = 137$, and $n = 274$ with $t = 3$. Some of the results are summarized in Table 9 in terms of the notation of Section 3.5 for the grouped data.

TABLE 9. Results for Orange Juice Rankings

Group One		Group Two		Total	
$a_{11}=193$	$p_{11}=.484$	$a_{12}=237$	$p_{12}=.691$	$a_1=430$	$p_1=.572$
$a_{21}=158$	$p_{21}=.360$	$a_{22}=138$	$p_{22}=.233$	$a_2=296$	$p_2=.307$
$a_{31}=60$	$p_{31}=.156$	$a_{32}=36$	$p_{32}=.076$	$a_3=96$	$p_3=.121$
$\hat{D}_{123}(p_{i1})=.1936$		$\hat{D}_{123}(p_{i2})=.2244$		$\hat{D}_{123}=.2182$	
$-2 \ln \lambda_{11}=74.38$		$-2 \ln \lambda_{12}=176.07$		$-2 \ln \lambda_1=230.92$	

To test the hypothesis that there are no differences in the juices, we use the combined analysis and calculate the test statistic from (3.23)

$$(7.12) \quad -2 \ln \lambda_1^c = -2[\ln \lambda_{11} + \ln \lambda_{12}]$$

$$= 250.45.$$

Comparing this value with chi-square tables with $g(t-1) = 4$ degrees of freedom, we reject the hypothesis of no differences in the juices.

We now illustrate the computation involved in the calculation of a coefficient of agreement to measure the consistency of the ranking of the two groups. From (3.26), we have

$$(7.13)\quad -2 \ln \lambda_1^A = 2\sum_{u=1}^{g}\left[\sum_i a_{iu} \ln p_{iu} - n_u \sum_{i<j<h} \ln \hat{D}_{ijh}(p_{iu})\right]$$

$$-2\left[\sum_i a_i \ln p_i - n \sum_{i<j<h} \ln \hat{D}_{ijh}\right]$$

$$= 19.52.$$

If this value is compared with chi-square tables for two (in general $[g-1][t-1]$) degrees of freedom, we see that the hypothesis of consistency must be rejected.

The examples that have been given here should illustrate the techniques involved in the method of triple comparisons. The most tedious and time-consuming computations involve the calculation of estimates of the treatment parameters. After this has been done, the remaining computations are very easy to do.

VIII. SUMMARY

A mathematical model for triple comparisons is developed as an extension of the Bradley-Terry model for paired comparisons. With the model that is proposed for triple comparisons, either the item to be given the rank of one or the item to be given the rank of three may be selected first. After the assignment of either of these extreme ranks, the comparison of the other two items in the given block reverts to a paired comparison.

Methods of analysis of triple comparisons based upon ranks and using the proposed model are developed. A general class of tests of the null hypothesis that treatment ratings are equal is discussed and maximum-likelihood estimators of treatment ratings are obtained. Test statistics are derived as functions of the likelihood ratio. The null hypothesis that treatment ratings are equal is tested against two special alternatives. The alternative hypothesis (i) makes no assumptions of equality of treatment ratings and (ii) makes the assumption that there are only two groups of treatment ratings and within group ratings do not differ but the two groups themselves may differ. Large-sample distributions of the statistics are discussed and

approximate tests by means of the chi-square distribution are recommended.

The procedures that are developed appear to be applicable in most problems where qualitative measurements are reliable. They should be useful in problems involving subjective ranking by a panel of judges for the detection of differences in items or treatments. A method of combining the results of several judges is given which permits an over-all test of significance without the usual assumption that members of a panel agree upon the nature of the differences to be detected. The consistency of judges, or of panels of judges, can be tested by means of a coefficient of agreement.

A goodness-of-fit test of the appropriateness of the model for triple comparisons is developed. By means of this test, it is possible to determine if the data under consideration conform to the assumptions of the model. By an association with expected frequencies, the proposed test is shown to be approximately equivalent to the usual chi-square test of goodness of fit.

The asymptotic distribution of the estimators of the treatment ratings is obtained. The variances and covariances thus obtained provide some information concerning the reliability of the estimators. It is shown that when the number of repetitions of the experiment is large, orthogonal linear comparisons of the estimators can be made as in the analysis of variance.

For the most general alternative hypothesis, the proposed test statistic is shown to have a limiting non-central chi-square distribution. It is found that in comparison with the analysis of variance, the relative efficiency of this method of triple comparisons is 72%. However, the comparison is made when the analysis of variance would be the appropriate method. The method of triple comparisons can be used for subjective tests for which the analysis of variance does not appear to be appropriate. When compared with the method of paired comparisons, the relative efficiency of the method of triple comparisons is 150%.

Many of the techniques proposed for the analysis of triple comparisons are illustrated through numerical examples. Estimators of treatment ratings and the test statistic for the general alternative are computed from a given set of data. Estimated variances and covariances of the estimators of the example are also computed. The same data is used to illustrate the test of the appropriateness of the model. The results of a consumer acceptance study of food variants are used to illustrate the combined analysis and the test of consistency.

BIBLIOGRAPHY

Bliss, C. I., Anderson, E. O., and Marland, R. E. (1943) "A Technique for Testing Consumer Preferences with Special References to the Constituents of Ice Cream." Storrs Agricultural Experiment Station Bulletin 251, 1-20.

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

Bradley, R. A. and Terry, M. E. (1952b) "The Rank Analysis of Incomplete Block Designs. II. The Method for Blocks of Size Three." Bi-annual Report No. 4, Virginia Agricultural Experiment Station. Appendix A.

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

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

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

Chanda, K. C. (1954) "A Note on the Consistency and Maxima of the Roots of Likelihood Equations." Biometrika, 41, 56-60.

Cramér, H. (1946) Mathematical Methods of Statistics. Princeton University Press, Princeton.

Durbin, J. (1951) "Incomplete Blocks in Ranking Experiments." British Journal of Psychology, 4, 85-90.

Dykstra, Otto (1956) "A Note on the Rank Analysis of Incomplete Block Designs - Applications Beyond the Scope of Existing Tables." Biometrics, 12, 301-306.

Fisher, R. A. and Yates, F. (1953) Statistical Tables for Biological, Agricultural and Medical Research. Fourth Edition, Oliver and Boyd, Edinburgh, 25, 76.

Fix, Evelyn (1949) "Tables of Non-central χ^2." University of California Publications in Statistics, 1(2), 15-21.

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

Galinat, W. C. and Everett, H. L. (1949) "A Technique for Testing Flavor in Sweet Corn." Agronomy Journal, 41, 443-445.

Greenwood, M. L. and Salerno, R. (1949) "Palatability of Kale in Relation to Cooking." Food Research, 14, 314-319.

Gridgeman, N. T. (1955) "The Bradley-Terry Probability Model and Preference Tasting." Biometrics, 11, 335-343.

Gulliksen, Harold (1956) "A Least Squares Solution for Paired Comparisons with Incomplete Data." Psychometrika, 21, 125-134.

Guttman, Louis (1946) "An Approach for Quantifying Paired Comparisons and Rank Order." Annals of Mathematical Statistics, 17, 144-163.

Hopkins, J. W. (1954) "Incomplete Block Rank Analysis: Some Taste Test Results." Biometrics, 10, 391-398.

Jackson, J. E. and Fleckenstein, Mary (1957) "An Evaluation of Some Statistical Techniques Used in the Analysis of Paired Comparisons Data." Biometrics, 13, 51-64.

Kempthorne, O. (1952) Design and Analysis of Experiments, John Wiley and Sons, Inc., New York.

Kendall, M. G. and Babington Smith, B. (1939) "The Problem of M Rankings." Annals of Mathematical Statistics, 10, 275-293.

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

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

Kendall, M. G. (1955b) Rank Correlation Methods. Hafner Publishing Company, New York.

Morrissey, J. H. (1955) "New Method for Assignment of Psychometric Scale Values from Incomplete Paired Comparisons." Journal of the Optical Society of America, 45, 373-378.

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

Mosteller, Frederick (1951b) "Remarks on the Method of Paired Comparisons: II. The Effect of Aberrant Standard Deviation when Equal Standard Deviations and Equal Correlations Are Assumed." Psychometrika, 16, 203-206.

Mosteller, Frederick (1951c) "Remarks on the Method of Paired Comparisons: III. A Test of Significance for Paired Comparisons when Equal Standard Deviations and Equal Correlations Are Assumed." Psychometrika, 16, 207-218.

Noether, G. (1955) "On a Theorem of Pitman." Annals of Mathematical Statistics, 26, 64-68.

Pearson, E. S. and Hartley, H. O. (1951) "Charts for the Power Function of Analysis of Variance Tests, Revised from the Non-central F-Distribution." Biometrika, 38, 112-116.

Scheffé, H. (1952) "An Analysis of Variance for Paired Comparisons." Journal of the American Statistical Association, 47, 381-400.

Tang, P. C. (1938) "The Power Function of Analysis of Variance Tests with Tables and Illustrations of Their Use." Statistical Research Memoirs, 2, 126-157.

Thurstone, L. L. (1927) "Psychophysical Analysis." American Journal of Psychology, 38, 368-389.

Wald, A. (1943) "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations Is Large." Transactions of American Mathematical Society, 54, 426-482.

Wilks, S. S. (1946) Mathematical Statistics, Princeton University Press, Princeton.

APPENDIX A

A.1 Derivation of the Maximum-Likelihood Equations

In order to find the maximum-likelihood equations, we let

$$\text{(A.1)}\quad F = \ln L + \mu \sum_{\alpha=1}^{m} N_\alpha \pi_{(\alpha)}$$

$$= K + \sum_{\alpha=1}^{m} a_{(\alpha)} \ln \pi_{(\alpha)} - n\Bigg\{ \sum_{\alpha=1}^{m} \sum_{\substack{\beta=1 \\ \alpha\neq\beta}}^{m} \binom{N_\alpha}{2} N_\beta \ln\left(\pi^2_{(\alpha)}[\pi_{(\alpha)}+\pi_{(\beta)}] + \pi^2_{(\beta)}\pi_{(\alpha)}\right)$$

$$+ \sum_{\alpha<\beta<\gamma}^{m} N_\alpha N_\beta N_\gamma \ln\left(\pi^2_{(\alpha)}[\pi_{(\gamma)}+\pi_{(\beta)}] + \pi^2_{(\beta)}[\pi_{(\alpha)}+\pi_{(\gamma)}] + \pi^2_{\gamma}[\pi_{(\alpha)}+\pi_{(\beta)}]\right)\Bigg\}$$

$$+ \mu \sum_{\alpha=1}^{m} N_\alpha \pi_{(\alpha)}$$

where L is the likelihood function of (3.11), K is a constant, and μ is a Lagrange multiplier. Differentiation of F with respect to $\pi_{(\alpha)}$ results in

$$\text{(A.2)}\quad \frac{\partial F}{\partial \pi_{(\alpha)}} = \frac{a_{(\alpha)}}{\pi_{(\alpha)}} - n\Bigg\{ N_\alpha \sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{m} \binom{N_\beta}{2} \frac{\pi^2_{(\beta)} + 2\pi_{(\beta)}\pi_{(\alpha)}}{\pi^2_{(\beta)}[\pi_{(\beta)}+\pi_{(\alpha)}] + \pi^2_{(\alpha)}\pi_{(\beta)}}$$

$$+ \binom{N_\alpha}{2} \sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{m} N_\beta \frac{3\pi^2_{(\alpha)} + 2\pi_{(\alpha)}\pi_{(\beta)} + \pi^2_{(\beta)}}{\pi^2_{(\alpha)}[\pi_{(\alpha)}+\pi_{(\beta)}] + \pi^2_{(\beta)}\pi_{(\alpha)}}$$

$$+ N_\alpha \sum_{\substack{\beta<\gamma \\ \beta,\gamma\neq\alpha}}^{m} N_\beta N_\gamma \frac{\pi^2_{(\beta)} + 2\pi_{(\alpha)}[\pi_{(\beta)}+\pi_{(\gamma)}] + \pi^2_{(\gamma)}}{\pi^2_{(\alpha)}[\pi_{(\beta)}+\pi_{(\gamma)}] + \pi^2_{(\beta)}[\pi_{(\alpha)}+\pi_{(\gamma)}] + \pi^2_{(\gamma)}[\pi_{(\alpha)}+\pi_{(\beta)}]}\Bigg\}$$

$$+ \mu N_\alpha, \quad (\alpha = 1,\ldots,m).$$

Setting the partial derivatives in (A.2) equal to zero, multiplying through by $\pi_{(\alpha)}$, and then summing over α, we obtain

(A.3) $$\sum_{\alpha=1}^{m} a_{(\alpha)} - n\left\{\sum_{\alpha=1}^{m}\sum_{\beta\neq\alpha}^{m} N_\alpha \binom{N_\beta}{2} \frac{\pi_{(\alpha)}[\pi^2_{(\beta)}+2\pi_{(\beta)}\pi_{(\alpha)}]}{\pi^2_{(\beta)}[\pi_{(\beta)}+\pi_{(\alpha)}]+\pi^2_{(\alpha)}\pi_{(\beta)}}\right.$$

$$+\sum_{\alpha=1}^{m}\sum_{\beta\neq\alpha}^{m}\binom{N_\alpha}{2}N_\beta \frac{\pi_{(\alpha)}[3\pi^2_{(\alpha)}+2\pi_{(\alpha)}\pi_{(\beta)}+\pi^2_{(\beta)}]}{\pi^2_{(\alpha)}[\pi_{(\alpha)}+\pi_{(\beta)}]+\pi^2_{(\beta)}\pi_{(\alpha)}}$$

$$\left.+\sum_{\alpha=1}^{m}\sum_{\substack{\beta<\gamma \\ \beta,\gamma\neq\alpha}}^{m} N_\alpha N_\beta N_\gamma \frac{\pi_{(\alpha)}[\pi^2_{(\beta)}+2\pi_{(\alpha)}(\pi_{(\beta)}+\pi_{(\gamma)})+\pi^2_{(\gamma)}]}{\pi^2_{(\alpha)}[\pi_{(\gamma)}+\pi_{(\beta)}]+\pi^2_{(\beta)}[\pi_{(\alpha)}+\pi_{(\gamma)}]+\pi^2_{(\gamma)}[\pi_{(\alpha)}+\pi_{(\beta)}]}\right\}$$

$$+\mu\sum_{\alpha=1}^{m} N_\alpha \pi_{(\alpha)} = 0 .$$

The first two sums within the braces combine easily into

(A.4)

$$\sum_{\alpha=1}^{m}\sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{m} N_\alpha \binom{N_\beta}{2} \frac{[\pi_{(\alpha)}\pi^2_{(\beta)}+2\pi^2_{(\alpha)}\pi_{(\beta)}]+[3\pi^3_{(\beta)}+2\pi^2_{(\beta)}\pi_{(\alpha)}+\pi^2_{(\alpha)}\pi_{(\beta)}]}{\pi^3_{(\beta)}+\pi^2_{(\beta)}\pi_{(\alpha)}+\pi^2_{(\alpha)}\pi_{(\beta)}}$$

$$= 3\sum_{\alpha=1}^{m}\sum_{\substack{\beta=1 \\ \beta\neq\alpha}}^{m} N_\alpha \binom{N_\beta}{2} .$$

The last sum within the braces simplifies, by combining terms of like denominators, into the quantity $3\sum_{\alpha<\beta<\gamma} N_\alpha N_\beta N_\gamma$.

From (3.12), we find

$$(A.5)\quad \sum_{\alpha=1}^{m} a_{(\alpha)} = \sum_{\alpha=1}^{m} \sum_{i=s_{\alpha-1}+1}^{s_\alpha} \sum_{j<h}{}' \sum_{k=1}^{n} (3 - r_{ik:jh}) - 3n \sum_{\alpha=1}^{m} \binom{N_\alpha}{3}$$

$$= 3n\binom{t}{3} - 3n \sum_{\alpha=1}^{m} \binom{N_\alpha}{3} .$$

Into (A.3) we now substitute the appropriate values for the summations involving the π's to obtain, since $\Sigma N_\alpha \pi_{(\alpha)} = 1$,

$$(A.6)\quad 3n\left\{\binom{t}{3} - \left[\sum_{\alpha=1}^{m} \binom{N_\alpha}{3} + \sum_{\alpha\neq\beta}^{m} N_\alpha \binom{N_\beta}{2} + \sum_{\alpha\ \beta\ \gamma}^{m} N_\alpha N_\beta N_\gamma\right]\right\} + \mu = 0.$$

In order to evaluate the expression within the braces, consider the identity

$$(A.7)\quad (1 + X)^t = \prod_{\alpha=1} (1 + X)^{N_\alpha} .$$

The coefficient of X^3 in the left member of (A.7) must be equal to the coefficient of X^3 in the right member. Hence

$$(A.8)\quad \binom{t}{3} = \sum_{\alpha=1}^{m} \binom{N_\alpha}{3} + \sum_{\alpha\neq\beta}^{m} N_\alpha \binom{N_\beta}{2} + \sum_{\alpha<\beta<\gamma} N_\alpha N_\beta N_\gamma .$$

It follows that $\mu = 0$, and the likelihood equations may be written as in (3.13) where the $\pi_{(\alpha)}$ are replaced by $p_{(\alpha)}$ representing the maximum-likelihood estimators satisfying those equations.

APPENDIX B

B.1 Limiting Distribution of $\sqrt{n}\, y_i$ $(i = 1,\ldots,t)$

In order to examine more fully the assumption made in Subsection 5.4.2 that $\sqrt{n}\, y_i$ $(i = 1,\ldots,t)$ have a joint limiting distribution, we rewrite some of the relations given in Chapter V as follows:

$$\text{(B.1)} \quad \sqrt{n}\, y_i = \sqrt{n}\, t(p_i - 1/t), \ (i = 1,\ldots,t);$$

$$\text{(B.2)} \quad \frac{3\sqrt{n}}{t(t-2)}\left[\frac{a_i}{n} - \binom{t-1}{2}\right] = \sqrt{n}\, y_i + \frac{3\sqrt{n}}{t(t-2)} R_2(y);$$

and

$$\text{(B.3)} \quad R_2(y) = \frac{1}{6}\sum_{j\ h}{}' \Big[(y_j^2 + y_h^2 + 2y_i y_j + 2y_i y_h) + (w_{ijh}u_{ijh} - 6[y_i + y_j + y_h]) + y_i(w_{ijh} - 6) + (1 + y_i)\frac{w_{ijh}u_{ijh}^2}{1 + u_{ijh}} - y_i w_{ijh} u_{ijh} \Big].$$

The hypothesis to be tested is that of

$$\text{(B.4)} \quad H_0 : \pi_i = t^{-1} \ (i = 1,\ldots,t)$$

against the alternative hypothesis

$$\text{(B.5)} \quad H_1 : \pi_i = t^{-1} + n^{-\frac{1}{2}}\delta_{in} \ (i = 1,\ldots,t),$$

where $\lim_{n \to \infty} \delta_{in} = \delta_i$.

We have made several attempts to prove directly that $\sqrt{n}\, y_i$ $(i = 1,\ldots,t)$ have a joint limiting distribution. Various methods have been tried, including an examination of the weaker assumption that $(n^{\frac{1}{2}+\epsilon})y_i$ $(i = 1,\ldots,t)$ have a joint limiting distribution. By using the results of Subsection 5.4.2, which showed that $\dfrac{3\sqrt{n}}{t(t-2)}\left[\dfrac{a_i}{n} - \binom{t-1}{2}\right]$ for $i = 1,\ldots,t$ have a joint limiting multivariate normal distribution, it can be shown from (B.2) that ϵ must be zero under the assumption that $(n^{\frac{1}{2}+\epsilon})y_i$ $(i = 1,\ldots,t)$ have a limiting distribution. The weaker assumption thus leads to the conclusion that $\sqrt{n}\, y_i$ $(i = 1,\ldots,t)$ have a limiting distribution, but we have been unable to prove that this assumption holds. Another approach to the problem involved trying to find bounds of $\sqrt{n}\, y_i$ in terms of functions of $\dfrac{3\sqrt{n}}{t(t-2)}\left[\dfrac{a_i}{n} - \binom{t-1}{2}\right]$, but we have not been able to find such functions.

It was shown in Section 5.2 that $\sqrt{n}(p_1-\pi_1),\ldots,\sqrt{n}(p_t-\pi_t)$ are asymptotically distributed as the multivariate normal distribution with zero means and a known dispersion matrix which is a function of $\pi_1,\ldots,\pi_t$. Since $\sqrt{n}\, y_i$ is a function of $\sqrt{n}(p_i - \pi_i)$ as given by the equation

(B.6) $\sqrt{n}\, y_i = \sqrt{n}\, t\left[(p_i - \pi_i)+(\pi_i - t^{-1})\right]$, the results of Section 5.2 provide some justification for the assumption that $\sqrt{n}\ y_i$ $(i = 1,\ldots,t)$ have a limiting distribution. The special alternative hypothesis given in (B.5) complicates the calculation of means, variances, and covariances from the functional relation given in (B.6). However, by merely assuming that $\sqrt{n}\ y_i$ $(i = 1,\ldots,t)$ have some form of joint limiting distribution, the form of the distribution was completely determined in Subsection 5.4.2 from the relation given here as (b.2). This procedure permitted us to find the distribution of $\sqrt{n}\ y_i$ $(i = 1,\ldots,t)$ on the basis of an assumption that seems to be valid.

ACKNOWLEDGEMENTS

The author is deeply grateful to Dr. Ralph A. Bradley for his assistance and counsel at all times and for the privilege of working under his guidance. The assistance and advice of Dr. John E. Freund and Dr. Leonard D. McFadden are appreciated. The author also wishes to thank Dr. Boyd Harshbarger for the encouragement and assistance given him during his period of study at the Virginia Polytechnic Institute.

The author acknowledges with thanks the help derived from The Southern Fellowships Fund. The Fund's grant made it possible to undertake the research which led to this dissertation. The author appreciates the assistance given by the Office of Naval Research which resulted in the completion of this research.

The author is grateful to and for the care with which they have typed the final manuscript.

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2

ABSTRACT

THE RANK ANALYSIS OF TRIPLE COMPARISONS

by

Robert Nixon Pendergrass, B.S. Ed., M. Ed., M.A.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS

August, 1957

General extensions of the probability model for paired comparisons, which was developed by R. A. Bradley and M. E. Terry, are considered. Four generalizations to triple comparisons are discussed. One of these models is used to develop methods of analysis of data obtained from the ranks of items compared in groups of size three.

The mathematical model recommended for triple comparisons is based upon the assumption of the existence of true treatment ratings, or parameters. A general class of tests of the null hypothesis that treatment ratings are equal is discussed, and maximum-likelihood estimators of the treatment parameters are obtained. These estimators are used in the evaluation of test statistics that are derived as functions of the likelihood ratio. The null hypothesis that treatment ratings are equal is tested against two special alternatives. The alternative hypothesis (i) makes no assumptions of equality of treatment ratings. Alternative hypothesis (ii) makes the assumption that there are only two groups of treatments and within group treatments do not differ in ratings but the two groups themselves may have different ratings. The asymptotic distributions of test statistics are discussed and approximate tests by means of the chi-square distribution are recommended.

The application of any statistical technique to numerical data is strictly appropriate only when the data conform to the assumptions upon which that statistical method is based. A test of the goodness of fit of the model for triple comparisons is developed in order to test the appropriateness of the model. By an association with expected frequencies, the proposed test is shown to be approximately equivalent to the usual chi-square test of goodness of fit.

The procedures that are developed for the analysis of triple comparisons should be useful in problems involving subjective ranking by a panel, or panels, of judges for the detection of differences in items or treatments. A method of combining the results of several judges is given which permits an over-all test of significance without requiring the usual assumption that members of a panel agree upon the nature of the differences to be detected. A coefficient of agreement is obtained to measure the consistency of judges, or of panels of judges.

The estimators of the treatment ratings are shown to be asymptotically normally distributed. Their variances and covariances are obtained to provide information concerning the reliability of the estimators. When the number of repetitions of the experiment is large, orthogonal linear comparisons of the estimators can be made as in the analysis of variance.

The test statistic proposed for the most general alternative hypothesis is shown to have a limiting non-central chi-square distribution. The results of this power study are used to obtain local asymptotic relative efficiencies of the triple comparisons test with other test procedures. When compared with the analysis of variance, the relative efficiency of the method of triple comparisons is 72%. When compared with paired comparisons, the relative efficiency of triple comparisons is 150%.

Many of the techniques proposed for the analysis of triple comparisons are illustrated through numerical examples. The results of a consumer acceptance study of food variants are used in the illustrations to indicate some of the problems encountered in actual experimentation.