

NONPARAMETRIC ANALOGUES OF ANALYSIS OF VARIANCE

by

RONALD K. LOHRDING

B. A. , Southwestern College, 1963

A MASTER 'S REPORT

submitted in partial fulfillment of the

requirements for the degree

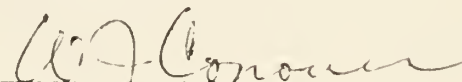
MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Approved by:


Major Professor

21)
2668
R4
1966
4833

CONTENTS

1. INTRODUCTION	1
2. COMPLETELY RANDOMIZED DESIGN	4
2.1. The Kruskal-Wallis One-way Analysis of Variance by Ranks.	4
2.2. Mosteller's k-Sample Slippage Test	7
2.3. Conover's k-Sample Slippage Test	8
2.4. A k-Sample Kolmogorov-Smirnov Test	11
2.5. The Contingency χ^2 Test.	13
3. RANDOMIZED COMPLETE BLOCK.	14
3.1. Cochran Q Test	15
3.2. Friedman's Two-way Analysis of Variance by Ranks .	16
4. PROCEDURES WHICH GENERALIZE TO SEVERAL DESIGNS .	17
4.1. Mood's Median Test.	18
4.2. Wilson's Median Test	21
4.3. The Randomized Rank-Sum Test	27
4.4. Rank Test for Paired-Comparison Experiments . . .	30
5. BALANCED INCOMPLETE BLOCK.	33
5.1. Bradley's Rank Analysis of Incomplete Block Designs	33
5.2. Durbin's Rank Analysis of Incomplete Block Designs	40
6. A 2×2 FACTORIAL EXPERIMENT.	44
7. PARTIALLY BALANCED INCOMPLETE BLOCK	47

8. OTHER RELATED TESTS 50
ACKNOWLEDGEMENTS 53
REFERENCES 54

NONPARAMETRIC ANALOGUES OF ANALYSIS OF VARIANCE

1. Introduction

J. V. Bradley (1960) proposes that the history of statistics can be divided into four major stages. The first or one parameter stage was when statistics were merely thought of as averages or in some instances ratios. The second or two parameter stage was when the concept of variance was introduced. Every population was thought to have a normal distribution which could be completely described with the mean and variance. In the third or multiparameter stage the discovery of many new distributions took place and with this the concept of using moments to identify the distribution. The fourth or no parameter stage is the area of this report. The first nonparametric statistics, sometimes called distribution free statistics, can be traced back as far as 1710. However, Savage (1953) places the "true beginning" of nonparametric statistics in 1936. From this point there has been rapid growth in this field.

There are several reasons for the rapid growth of nonparametric statistics. One of these reasons is the simplicity of derivation of most of the nonparametric tests in comparison to similar parametric tests. This has made the nonparametric test easier to understand for the average researcher.

In most cases, the ease and speed of calculation of the nonparametric tests make them very attractive to organizations not having a computer.

In general, the nonparametric tests are nearly as efficient as the parametric tests even under the assumptions of the parametric tests. When the assumptions of the parametric test are violated to a large degree, it is generally thought that the nonparametric tests are more powerful than the parametric tests (Bradley, 1960).

Siegel (1956) and Bradley (1960) have written books on nonparametric statistics that cover only the simplest analogues of analysis of variance. The purpose of this report is to provide nonparametric analogues for many basic parametric designs and new nonparametric procedures for the designs covered by Siegel (1956) and Bradley (1960).

In many areas of research, parametric analysis of variance designs are used extensively even when the assumptions are violated. This is usually done because the researcher is unfamiliar with the proper nonparametric procedure. Therefore, it is hoped that this paper will provide a reference for researchers interested in using nonparametric analysis of their data.

The first design discussed is the completely randomized design. The most popular and probably the most powerful of these analogues is the Kruskal-Wallis analysis of variance by ranks. There is also a discussion of two possible approaches to a k-sample slippage test. These two tests have the advantage of being quick and simple. The next analogue

of this section involves a k-sample Kolmogorov-Smirnov test that makes a test of significance on k cumulative distributions. Last is the well known contingency χ^2 test for k independent samples.

The second analogue is of the randomized complete block design. The first test considered is the Cochran Q test. This is used for only special types of data, usually of the + or - type. Next is the Friedman two-way analysis of variance by ranks. This is the most widely known test of this section (Siegel, 1956).

The third major area includes procedures that extend over several designs. These procedures are quite useful because an experimenter needs only to be acquainted with one process and then can generalize over several designs. A large part of this section is devoted to median tests which are analogues to several analysis of variance designs. The first of these procedures was developed by Mood (1950). Wilson (1956) revised Mood's method but Alluisi (1956) was responsible for developing the computational formulae which make this procedure very useable. Alluisi extends this procedure to three factors which is very similar to the three factor analysis of variance design. The last two tests of this section are recently developed tests involving rank analysis. These tests have procedures for completely randomized and randomized complete block analogues. These two tests appear to have considerable power and could be the most important part of this paper.

Section 5 includes two procedures for a balanced incomplete block analogue. These two tests are valuable when the data consist of small-scale comparisons.

The only nonparametric factorial experiment found was a 2×2 factorial with paired comparisons. This test is of little use due to the extreme complexity of computations.

The last major section is concerned with partially balanced incomplete block designs. This test by Giri (1965) gives procedures for a reasonable approximation to the randomization test using the F statistic.

The final section is a short summary of other related tests and papers that might be of considerable help to other persons doing research in this area.

2. Completely Randomized Design

For this design the experiment units or observations are randomly chosen for each treatment. It is not necessary for the treatments to have an equal number of observations; although, it usually simplifies computations. There are several nonparametric analogues of this design. Some of these will be discussed in this section and the remainder in section 4.

2.1 The Kruskal-Wallis One-way Analysis of Variance by Ranks.

The Kruskal-Wallis test is the most popular test of this section.

This popularity could be due to the fact that the Kruskal-Wallis test is nearly an exact analogue of the completely randomized design. Also, Bradley (1960) states that the asymptotic efficiency of the Kruskal-Wallis test compared to the completely randomized design is .955 or $3/\pi$.

The Kruskal-Wallis technique tests the null hypothesis that the k samples come from the same population or from identical populations against the alternative that they are from different populations.

The only assumptions required are that the random variables are from a continuous distribution and that the variables can at least be ordered (Siegel, 1956).

The scores from all of the k samples are placed in an ordered array. The observations are then ranked giving the rank of 1 to the smallest observation, to the next smallest 2, etc. The original observations are now identified only by their proper rank. These ranks are then replaced in their proper column and the columns or treatments are summed.

It has been shown by Kruskal-Wallis (1952) that if the k samples actually are from the same population, then H is distributed approximately as chi square with $k-1$ degrees of freedom (df) for sample sizes sufficiently large, where

$$(2.1.1) \quad H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1).$$

and

k = number of samples.

n_j = number of observations in the j^{th} sample.

$N = \sum_j n_j$ = number of observations in all samples combined.

R_j = sum of ranks in the j^{th} sample (column).

When all the k samples have $n \geq 5$, then a chi square table (Siegel, 1956) with $df = k - 1$ may be used to test significance. When $k = 3$ and $n \leq 5$ there are special tables available (Kruskal and Wallis, 1952).

When ties occur between two or more scores, each score is given the mean of the ranks for which it is tied. Since the value of H is somewhat influenced by ties, one may wish to correct for ties in computing H . To correct for the effect of ties, H is computed and then divided by

$$(2.1.2) \quad 1 - \frac{\sum_{j=1}^a T_j}{N^3 - N}$$

Where $T_j = t_j^3 - t_j$,

t_j = the number of tied observations in the j^{th} tied group of scores,

a = the number of different groups of tied observations.

If the number of ties is relatively small, the correction for ties makes very little difference in H .

2.2 Mosteller's k-sample Slippage Test.

A slippage test is concerned with the amount the location parameter of one distribution has slipped with respect to the other distributions.

Exact probabilities can be computed for Mosteller's test without difficulty and the test is easy to perform. However, it would appear that the power of this slippage test is low. The loss in power would be due to the fact that evenly spaced sample differences would not show significance; although, there could be high significance between the largest and smallest sample.

The null hypothesis is that all populations have the same locations against the alternative that one population has a larger location parameter than the rest.

It is assumed that the populations are continuous and are identically distributed. It is also assumed that the samples are independent and randomly chosen.

The slippage test is only concerned with the largest observations; therefore, this test is extremely sensitive to both shape and location of the upper tail of the distribution. This should be taken into consideration when conducting this test.

To make this test, the sample containing the largest observation is determined and in it the experimenter counts the number, r , of observations which exceed all observations in all other samples. If n_i is the size of the i^{th} sample and N is the total number of observations in all

samples, then there are $n_i (n_i - 1) \dots (n_i - r + 1)$ or $n_i! / (n_i - r)!$ ways in which the r largest observations could have been placed in the i^{th} sample and $N(N - 1) \dots (N - r + 1)$ or $N! / (N - r)!$ ways in which they could have been located in different samples. The probability that the r largest observations will all be in a preselected sample is therefore

$\frac{n_i! / (n_i - r)!}{N! / (N - r)!}$, and the probability that they will all be in some one of the k samples is

$$(2.2.1) \quad \Pr(r) = \sum_{i=1}^k \frac{n_i! / (n_i - r)!}{N! / (N - r)!}$$

Since in the derivation, it was not required that the $(r - 1)^{\text{st}}$ largest observation be located in a different sample, the above probability is the probability that r or more of the largest observations will be located in a single sample (Bradley, 1960). This computation is not extremely complicated with tables of factorials (Siegel, 1956). However, exact tables have been published by Mosteller (1948) for the case of equal sized samples $n = 3, 5, 7, 10, 15, 20, 25$, with $2 \leq k \leq 6$ and $2 \leq r \leq 5$.

2.3 Conover's k -sample Slippage Test.

The main advantage of Conover's k -sample slippage test is the ease of computation. This test can be used where a quick and easy statistical test is desired to support the rejection of H_0 when it is obviously

wrong. Another advantage of this test is the ease of making multiple comparisons to find nonsignificant groupings of treatment means. It appears that Conover's k-sample slippage test is more versatile than Mosteller's k-sample slippage test and would be more powerful in most cases.

The data are arranged in k random samples with each containing n observations. The samples are ordered among themselves on the basis of the greatest value within each sample; these values being called the extreme values. The samples are then ranked according to their extreme values giving the sample with the largest extreme value the rank of one.

The null hypothesis is that all of the distribution functions of the samples are equal. This is to be tested against the alternative that at least two of the distribution functions differ by a shift in location parameter.

The distribution functions are all considered to be continuous and differ only by a location parameter.

The test statistic m equals the number of values from the sample of rank 1 exceeding the extreme value from the sample of rank $j = k$.

Using the tables by Conover (1966) for the appropriate values of k and n , and the desired level of significance, α , let $j = k$ and reject H_0 if the value of m is greater than or equal to the value in the table.

If significance is found, then multiple comparisons may be made. There are two possible procedures suggested by Conover (1966).

A possible third method, analogous to Duncan's new multiple range test would give all of the nonsignificant groupings of treatments in the

experiment. First, let m equal the number of observations from the sample of rank 1 that exceed the largest observation of the sample of rank 2 and enter the table as before except that now $j = 2$. If m does not exceed the critical value from the table, repeat the procedure for $j = 3, 4$, and so on until a significant reading is found. This would yield the first nonsignificant grouping.

Now take the sample of rank 2 and consider it to be the largest sample and repeat the process mentioned before with $k' = k - 1$. This procedure could produce a new group of nonsignificant treatments or just a subgroup of the first group. If a subgroup is produced, the procedure would again be followed for the sample of rank 3 and with a $k'' = k - 2$. This would continue until a second nonsignificant group of treatments is produced. This process would be continued for sample of rank 4, $\dots, k-1$. From this all of the nonsignificant groups of treatments could be found.

In case of ties, the method of determining m described previously will yield conservative results (Conover, 1966). If there is difficulty in determining the ranks of the samples because of ties in their extremes, it is suggested by Conover (1966) that these ties be resolved by comparing the next highest values from the samples in question, and assigning ranks correspondingly. This procedure may be continued until the tie is broken.

There is also a test for k samples of unequal size presented. However, this test is more of an analogue of a multiple comparison which does not require a previous analysis of variance.

The power of Conover's k-sample slippage test has not been computed but it seems possible that the power of this test must be greater than or equal to the power of Mosteller's test.

There are tables of significant values for $k \leq 20$ and for selected values of n given by Conover (1966).

2.4 A k-sample Kolmogorov-Smirnov Test.

The Kolmogorov-Smirnov test has one big advantage that is not present in most of the other nonparametric analogues of the completely randomized design. This advantage is that the scale of the distribution may be tested as well as the location.

The null hypothesis is that all of the samples come from the same distribution, with the alternative that at least some of the samples come from different distributions.

To make this test, the samples must first be ordered within themselves. Then the samples themselves are ordered according to their smallest observation. Now let Y_{ij} denote the observation where the number i is called the rank of Y_{ij} within the sample, and the number j is called the rank of the sample. S is the set of ordered samples from Y_{ij} . Conover (1965a) defines the empirical distribution function of the sample with rank j as:

$$\begin{aligned} S_j(y) &= 0 && \text{if } y < Y_{1j} \\ S_j(y) &= m/n && \text{if } Y_{mj} \leq y \leq Y_{m+1,j} \\ S_j(y) &= 1 && \text{if } Y_{nj} < y, \end{aligned}$$

where k is the number of equal random samples and n is the number of observations per sample.

To make this test, the experimenter may first graph all k samples on the same graph as discrete cumulative distribution functions or just look at the cumulative frequencies. Then the maximum vertical distance or the maximum difference is the test statistic

$$(2.4.1) \quad D_{j_1 j_2}^+(k, n) = \sup_y [S_{j_1}(y) - S_{j_2}(y)].$$

where $j_1 < j_2$ and $j_1 = 1, 2, \dots, k-1$; $j_2 = 2, 3, \dots, k$.

To test the null hypothesis, one could use $D_{1, k}^+(k, n)$. $D_{1, k}^+(k, n)$ will attain some value c/n . Thus $\Pr(D_{1, k}^+(k, n) \leq c/n)$ will give the probability of $D_{1, k}^+(k, n)$ being less than or equal to c/n under the null hypothesis so a decision may be reached, using

$$(2.4.2) \quad P(D_{j_1, j_2}^+(k, n) \leq \frac{c}{n}) = 1 - \frac{\binom{k-j_1}{j_2-j_1} \binom{2n-2}{n+c}}{\binom{k-j_1-(c+1)/n}{j_2-j_1} \binom{2n-2}{n-1}} +$$

$$\sum_{\alpha=0}^{j_2-j_1-1} \frac{(n-1)_c (c+1)_{j_2-j_1} (-1)^{j_2 j_1^{-\alpha}} (k-j_1-\alpha-1) (k-j_1-\alpha) (nk-nj_1-c-1-n\alpha)}{\alpha! (j_2-j_1-1-\alpha)! (nk-nj_1+n-1-n\alpha)_c}$$

where $(A)_c$ represents the falling factorial $A(A-1)(A-2)\dots(A-c+1)$.

The D^+ may be also used to make multiple comparisons of the treatments and Conover (1965a) gives a D^- which can be used to test for homogeneity of the scales of the distributions. These tests require long and difficult computations if done by hand; however, they could easily be put on a computer.

2.5 The Contingency χ^2 Test.

The contingency χ^2 is a well known test to any student of statistical methods. When frequencies in discrete categories constitute the data of research, the χ^2 test may be used to determine the significance of the differences among k independent groups.

To apply the χ^2 test, one first arranges the frequencies in a k by r table. The null hypothesis is that the k samples of frequencies have come from the same population or from identical populations. This hypothesis may be tested by

$$(2.5.1) \quad \chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where

O_{ij} = observed number of cases categorized in i^{th} row of j^{th} column,

E_{ij} = number of cases expected under H_0 to be categorized in the i^{th} row of the j^{th} column,

k = the number of columns,

r = the number of rows.

To find the expected frequency for each cell (E_{ij}), multiply the two marginal totals common to a particular cell, and then divide this product by the total number of observations. The χ^2 test requires that the expected frequencies in each cell should not be too small. When this requirement is violated, the results of the test are meaningless. It is recommended that fewer than 20 per cent of the cells should have an expected frequency of less than 5, and no cell should have an expected frequency of less than 1 (Cochran, 1954). If these requirements are not met by the data in the form in which they were originally collected, the researcher may combine adjacent categories so as to increase the E_{ij} 's in the various cells. Only after these conditions are met can one be sure of the level of significance. Tests of significance can be made by using the chi square distribution with $df = (k-1)(r-1)$.

3. Randomized Complete Block

The randomized complete block design has the experimental units divided into homogeneous groups called blocks. The treatments are then assigned randomly within each block. This is sometimes called the two-way analysis of variance. There are only two analogues in this section; however, there are several others in section 4.

3.1 Cochran Q Test

The Cochran Q test is a k sample extension of the McNemar test (Cochran, 1950). This extension provides a method for testing whether three or more matched sets of frequencies differ significantly among themselves. The Cochran test is particularly suitable for yes-no or pass-fail type of data. The exact power of this test is not known but it is considered to be less than that of Friedman's test (Siegel, 1956).

If the data are arranged in a two-way table consisting of N rows and k columns, where the rows are the blocks, it is possible to test the null hypothesis that the proportion or frequency of responses of a particular kind is the same in each column, except for chance differences. Cochran (1950) has shown that if the number of rows is not too small, then

$$(3.1.1) \quad Q = \frac{(k - 1) \left[k \sum_{j=1}^k G_j^2 - \left(\sum_{j=1}^k G_j \right)^2 \right]}{N \left[k \sum_{i=1}^k L_i - \sum_{i=1}^k L_i^2 \right]}$$

where

G_j = total number of "successes" in j^{th} column,

L_i = total number of "successes" in i^{th} row.

Q is approximately distributed as a chi square random variable with $df = k - 1$. Therefore, tests for significance can be made from any chi square tables.

3.2 Friedman's Two-way Analysis of Variance by Ranks.

The Friedman test is one of the most popular two-way tests. This popularity is due to its high asymptotic efficiency of .91 when compared to analysis of variance and also due to the fact that the Friedman two-way analysis of variance by ranks is a very good analogue to the parametric randomized complete block design of analysis of variance (Bradley, 1960). The data of the tests are ranks. The scores in each row are ranked separately. That is, with k conditions being studied, the ranks in any row or block range from 1 to k . The Friedman test determines whether it is likely that the different columns of ranks came from the same population. If the treatment scores were independent of the conditions, the set of ranks in each column would represent a random sample and the rank totals for the various columns would be about equal. If the treatment scores were dependent on the conditions, then the rank totals would vary from one column to another.

The Friedman test determines whether the rank totals differ significantly. To make this test, we compute the value of a statistic which Friedman denotes χ_r^2 .

Friedman (1937) has shown that when the number of rows and columns is not too small, χ_r^2 is distributed approximately as chi square random variable with $df = k - 1$, where

$$(3.2.1) \quad \chi_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^k (R_j^2) - 3N(k+1),$$

and

N = the number of rows,

k = number of columns,

R_j = sum of ranks in j^{th} column.

When N and k are smaller than 4 the exact probability tables are available (Siegel, 1956).

4. Procedures Which Generalize to Several Designs

All of these procedures include analogues of completely randomized and randomized complete block. However, the median test is extended to a three factor experiment with k observations per cell. A three factor experiment has three main effects instead of the two main effects of the randomized complete block. A good example would be an experiment involving varieties of wheat, kinds of fertilizer and amounts of irrigations. In a three factor experiment the experimenter could test the main effects of wheat, fertilizer and irrigation; the two-way interaction of wheat with fertilizer, wheat with irrigation and fertilizer with irrigation; and the three-way interaction of wheat with fertilizer with irrigation.

4.1 Mood's Median Test.

The median test is a very useable test because of the way it can be used for several designs. However, the median test only has an asymptotic efficiency of .65 when compared to similar parametric tests.

4.1.1 The Median Test Analogue of the Completely Randomized Design.

The assumptions under H_0 for the median test procedures are that the data are continuous and that each of the C samples has the same distribution.

To apply the median test, one must first determine the median score for the combined C samples of scores. Then replace each score by a plus if the score is larger than the common median and by a minus if it is smaller than or equal to the common median.

The resulting sets of scores may be put into a C by 2 table. The first row table represents the frequencies of pluses in each of the C groups and the second row represents the minuses.

E_{11} 0_{11}	E_{12} 0_{12}	E_{13} 0_{13}	...	E_{1C} 0_{1C}
E_{21} 0_{21}	E_{22} 0_{22}	E_{23} 0_{23}	...	E_{2C} 0_{2C}

To test the null hypothesis that the C samples have come from the same population with respect to medians, the value of χ^2 is computed, using

$$(4.1.1) \quad X^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}} .$$

where

O_{ij} = observed number of cases categorized in the i^{th} row and the j^{th} column,

E_{ij} = number of cases expected under H_0 to be categorized in the i^{th} row of the j^{th} column,

R = number of rows in the table which will always be 2 in this procedure,

C = the number of different samples being tested.

It has been shown by Mood (1950) that X^2 is an approximate chi square random variable with $df = (C - 1)(R - 1)$. However, in this design R is always 2, so the df will always be $(C - 1)$.

Once the data have been categorized as plus and minus with respect to the common median and the resulting frequencies have been cast in a C by 2 table, the computation procedures for this test are exactly the same as those for the χ^2 test for C independent samples, presented previously.

4.1.2 The Median Test Analogue of the Randomized Complete Block Design.

The observations are denoted by X_{ij} with $i = 1, 2, \dots, R$ and $j = 1, 2, \dots, C$. Where R is the number of rows and C is the number of columns.

The null hypothesis is that the row effects are zero. This test is made by finding the median of each column and then by putting a plus sign if the observation is above the median of that column and a minus sign if it is below. This produces an R by C table of plus and minus signs. Let a_i be the number of plus signs in the i^{th} row. If there is no row effect, we would expect each row to have $C/2$ plus signs. The null hypothesis is therefore tested by testing whether the signs are divided evenly in rows. In fact, we may construct a $2 \times R$ contingency table.

Table (4.1.2)

				Totals
a_1	a_2	. . .	a_R	C_a
$C - a_1$	$C - a_2$. . .	$C - a_R$	$C(R - a)$

Where $a = R/2$ if R is even or $(R - 1)/2$ if r is odd. It turns out that the a_i do not have the ordinary contingency-table distribution (Mood, 1950). However, this X^2 table is approximately distributed as a chi square random variable with $R - 1$ degrees of freedom for large samples. So this

table may be tested like an ordinary contingency table with all marginal totals fixed.

The null hypothesis that the row effects are zero may therefore be tested by

$$(4.1.2) \quad X^2 = \frac{R(R-1)}{Ca(R-a)} \sum_{i=1}^R \left(a_i - \frac{Ca}{R} \right)^2 .$$

For practical purposes, the large-sample distribution is satisfactory if C is as large as 5 provided $RC \geq 20$. For smaller values, the exact probability can be computed by a complicated formula in Mood (1950). To test column effects, one would, of course, simply reverse the roles of rows and columns in the above test.

Mood (1950) has extended this to an analogue of the randomized complete block design with k observations per cell. However, Wilson has changed Mood's procedure into a much more useable procedure. Therefore, Wilson's procedure will be discussed.

4.2 Wilson's Median Test.

Wilson's median test procedure is as near to an exact analogue of the analysis of variance as one can get. Even the formulae are similar although the χ^2 tables are used to make the test instead of the F tables.

4.2.1 The Median Test Analogue to the Randomized Complete Block Design with K Observations Per Cell.

Assume that data have been collected in a study that used a two-factor analysis of variance design. There were R rows, C columns, and in all RC cells with K independent observations per cell. Now assume that the obtained data violated some of the assumptions of analysis of variance. Then Wilson (1954) suggests a nonparametric analysis of the data with the chi square method.

First, divide the distribution of scores into upper and lower halves. In one group there will be the $KRC/2$ highest scores in the distribution, and in the other group there will be the $KRC/2$ lowest scores in the total distribution. If ties occur with the median, Alluisi (1956) suggests that a coin be tossed to determine the placement of the tied scores.

Second, tabulate the number (frequency) of the $KCR/2$ upper-half scores that occurs in each of the RC cells, and prepare a summary matrix listing these frequencies and their sums.

Table (4. 2. 1)

Frequency Matrix of Upper -Half Scores								
Rows	Columns							Row Sums
	1	2	3	...	c	...	<u>C</u>	
1	f_{11}	f_{12}	f_{13}	...	f_{1c}	...	f_{1C}	$f_{1\cdot}$
2	f_{21}	f_{22}	f_{23}	...	f_{2c}	...	f_{2C}	$f_{2\cdot}$
3	f_{31}	f_{32}	f_{33}	...	f_{3c}	...	f_{3C}	$f_{3\cdot}$
.
.
.
r	f_{r1}	f_{r2}	f_{r3}	...	f_{rc}	...	f_{rC}	$f_{r\cdot}$
.
.
.
<u>R</u>	f_{R1}	f_{R2}	f_{R3}	...	f_{Rc}	...	f_{RC}	$f_{R\cdot}$
Column Sums	$f_{\cdot 1}$	$f_{\cdot 2}$	$f_{\cdot 3}$...	$f_{\cdot c}$...	$f_{\cdot C}$	$f_{\cdot\cdot}$

The chi square values for the separate components of variation may be computed from the data of the summary matrix. The formulae are presented in the following table.

Table (4.2.2)

Summary of a Chi-Square Components Analysis		
Components	<u>df</u>	Chi-square
Columns	C-1	$\frac{2 \left[C \sum_{c=1}^C (f_{.c})^2 - (f_{..})^2 \right]}{f_{..}}$
Rows	R-1	$\frac{2 \left[R \sum_{r=1}^R (f_{r.})^2 - (f_{..})^2 \right]}{f_{..}}$
Interaction	(R-1)(C-1)	Total - Rows - Columns
Total	RC-1	$\frac{2 \left[RC \sum_{r=1}^R \sum_{c=1}^C f_{rc}^2 - (f_{..})^2 \right]}{f_{..}}$

In the above table $f_{r.}$ means the sum of all of the column frequencies for

a fixed row R or $f_{r.} = \sum_{c=1}^C f_{rc}$ and then $\sum_{r=1}^R (f_{r.})^2 = \sum_{r=1}^R \left(\sum_{c=1}^C f_{rc} \right)^2$.

The significance levels of the obtained values of chi square are determined by referring to tabled values of the chi square statistic with the df indicated in the figure. Interpretation of the results is analogous

to the interpretation given an analysis of variance, except that in the present method the number of measurements above the median is dealt with rather than the amount of deviation from the mean.

In building the summary frequency matrix, the $KRC/2$ lower-half scores could have been tabulated instead of the upper-half scores without loss of validity of the method. In either case, the obtained values of chi square are identical. The tabulation of the upper-half scores of the distribution is recommended because of the simplicity of interpretation of the results after the computation of the chi square values. A direct interpretation is possible with the matrix of upper-half scores. A high entry indicates a large number of scores above the median, and therefore, analogous to a high mean in analysis of variance.

The chi square value for the interaction component was computed indirectly by subtraction in the formula given. It could have been computed directly in the usual way that chi square is computed for a test of independence in a contingency table. When it is computed directly from either the summary frequency matrix of upper-half scores or from the matrix of lower-half scores, the obtained value must be multiplied by two to yield a chi square accounting for both halves of the total distribution of scores.

4.2.2 The Median Test Analogue of the Three-Factor Analysis of Variance.

The median test by Wilson has been extended to a three-factor

analysis by Alluisi (1956). There are R rows, C columns, and B blocks in this generalized three-factor experimental design; in all, there are RCB cells with K independent observations per cell. The three-factor summary frequency matrix contains tabulations of the $KRCB/2$ highest or lowest scores in the total distribution of scores. Then, the computational formulae for the several component chi square are as follows:

Table (4.2.3)

Summary of a Chi-square Components Analysis		
Components	df	Chi-square
Total	$RCB-1$	$\frac{2 \left[RCB \sum_{r=1}^R \sum_{c=1}^C \sum_{b=1}^B f_{rcb}^2 - f_{\dots}^2 \right]}{f_{\dots}} = \chi_T^2$
Row	$R-1$	$\frac{2 \left[R \sum_{r=1}^R (f_{r\dots})^2 - f_{\dots}^2 \right]}{f_{\dots}} = \chi_R^2$
Column	$C-1$	$\frac{2 \left[C \sum_{c=1}^C (f_{\dots c})^2 - f_{\dots}^2 \right]}{f_{\dots}} = \chi_C^2$
Block	$B-1$	$\frac{2 \left[B \sum_{b=1}^B (f_{\dots b})^2 - f_{\dots}^2 \right]}{f_{\dots}} = \chi_B^2$

Row by Column	$(R-1)(C-1)$	$\frac{2 \left[RC \sum_{r=1}^R \sum_{c=1}^C (f_{rc})^2 - f_{\dots}^2 \right]}{f_{\dots}} - \chi_R^2 - \chi_C^2 = \chi_{RC}^2$
Row by Block	$(R-1)(B-1)$	$\frac{2 \left[RB \sum_{r=1}^R \sum_{b=1}^B (f_{r \cdot b})^2 - f_{\dots}^2 \right]}{f_{\dots}} - \chi_R^2 - \chi_B^2 = \chi_{RB}^2$
Column by Block	$(C-1)(B-1)$	$\frac{2 \left[CB \sum_{c=1}^C \sum_{b=1}^B (f_{\cdot cb})^2 - f_{\dots}^2 \right]}{f_{\dots}} - \chi_C^2 - \chi_B^2 = \chi_{CB}^2$
Row by Column by Block	$(R-1)(C-1)(B-1)$	$\chi_T^2 - \chi_R^2 - \chi_C^2 - \chi_B^2 - \chi_{RC}^2 - \chi_{RB}^2 - \chi_{CB}^2 = \chi_{RCB}^2$

The above notation is the same as that used in figure (4.2.2).

The interpretation is the same as it would be for a three-factor analysis of variance.

4.3 The Randomized Rank-Sum Test.

The randomized rank-sum test is a very useable procedure. It has been shown by Bell and Doksum (1965) that for a two-sample test the rank-sum test has an asymptotic efficiency of one when compared with the t test for normally distributed data. Bell and Doksum (1965) have also

shown that when the assumptions for that test are not met, the rank-sum test has an asymptotic efficiency of greater than one. From this it would be expected that the K sample test would have a very high asymptotic efficiency when compared to the analysis of variance.

4.3.1 The Randomized Rank-Sum Analogue to the Completely Randomized Design.

There are K independent random samples of size n_1, n_2, \dots, n_K . Under the null hypothesis each sample has the same distribution. It is assumed that the distribution are continuous. The data are ranked as one

sample of size $N = \sum_{j=1}^K n_j$. Each observation, X_{ij} , has a specific rank,

$R(X_{ij})$.

Now take a random sample of size N from a normal population with mean 0 and variance 1, $(N(0, 1))$. This sample is then ranked and denoted by $Z(R(X_{ij}))$. Then replace each X_{ij} by $Z(R(X_{ij}))$, i.e., the Z value with the same rank in the Z sample as X_{ij} has in the X sample.

Then the test statistic is

$$(4.3.1) \quad V(H) = \sum_{i=1}^K n_i (\bar{Z}_{i.} - \bar{Z}_{..})^2$$

where

$$\bar{Z}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z(R(X_{ij}))$$

$$\bar{Z}_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^N Z(i) = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} Z(R(X_{ij}))$$

Then $V(H)$ is distributed as a chi square random variable with $K-1$ df (Bell and Doksum, 1965).

4.3.2 The Randomized Rank-Sum Analogue to the Randomized Complete Block Design.

There are C columns or treatments and R rows or blocks. The null hypothesis is that there are no effects due to treatments. It is assumed that there are row effects and that the columns have equal continuous distributions.

The X_{ij} 's are ranked within each row and denoted $R_i(X_{ij})$. Then R random samples of size C are taken from a $N(0, 1)$ population and each of the R samples are ranked within the samples and denoted $Z_i(R_i(X_{ij}))$. Then replace each X_{ij} by $Z_i(R_i(X_{ij}))$, i.e., replace X_{ij} by the Z value of the i^{th} sample of Z 's that has the same row-rank as X_{ij} . The test statistic is

$$(4.3.2) \quad Q(H) = R \sum_{j=1}^C (\bar{Z}_{\cdot j} - \bar{Z}_{\cdot\cdot})^2$$

where

$$\bar{Z}_{\cdot j} = \frac{1}{R} \sum_{i=1}^R Z_i(R_i(X_{ij})),$$

$$\bar{Z}_{..} = \frac{1}{RC} \sum_{j=1}^C \sum_{i=1}^r Z_i(R_i(X_{ij})).$$

$Q(H)$ is distributed as a chi square random variable with $C-1$ df (Bell and Doksum, 1965).

4.4 Rank Test for Paired-Comparison Experiments.

The rank test for paired comparisons involves more than just a generalization of the sign test (Mehra, 1964). This test also takes into account the magnitudes of the observed comparison difference. This test has an asymptotic efficiency of $3/\pi$ when compared with parametric analysis of variance designs (Mehra, 1964).

4.4.1 Rank Test for Paired-Comparisons that is an Analogue of the Completely Randomized Design.

Consider a paired-comparison experiment involving K treatments and suppose that each of the N_{ij} comparisons for a pair (i, j) of treatments ($1 \leq i < j \leq k$) provides an observed comparison difference Z_{ijg} , where $g = 1, 2, \dots, N_{ij}$. Let $G_{ij}(z)$ be a c.d.f. denoting the common distribution of the comparison differences and assume that $G_{ij}(z)$ is continuous. The hypothesis of no difference among the treatments can be expressed as $H_0 : G_{ij}(z) = G_{i'j'}(z)$ for any two pairs (i, j) and (i', j') .

To perform the test rank all the $N = \sum_{i=1}^k \sum_{j>i} N_{ij}$ absolute values

of the observed comparison differences $Z_{ijg}, (1 \leq i < j \leq k; g = 1, \dots, N_{ij})$.

Let $r_{ijg} = \text{rank of } |Z_{ijg}| \text{ if } Z_{ijg} > 0$, otherwise let $r_{ijg} = 0$. Let $S_{ijg} = \text{rank of } |Z_{ijg}| \text{ if } Z_{ijg} \leq 0$, otherwise let $S_{ijg} = 0$.

$$(4.4.1) \quad R_N^{(i,j)} = \sum_{j=1}^{N_{ij}} r_{ijg} \quad \text{and} \quad S_N^{(i,j)} = \sum_{g=1}^{N_{ij}} s_{ijg}$$

are respectively the sums of the ranks of the positive and the negative Z 's corresponding to the $(i, j)^{\text{th}}$ pair. The test statistic is

$$(4.4.2) \quad L = \frac{6}{(N+1)(2N+1)} \frac{K}{K} \sum_{i=1}^K \sum_{j \neq i} \frac{V_N^{(i,j)}{}^2}{N_{ij}}$$

where

N = the total number of comparisons in the experiment,

and

$$V_N^{(i,j)} = R_N^{(i,j)} - S_N^{(i,j)}.$$

When H_0 is true, L is asymptotically distributed as a χ^2 variable, as $N \rightarrow \infty$, with $K-1$ df (Mehra, 1964).

4.4.2 Rank Test for Paired-Comparisons that is an Analogue of the Randomized Complete Block Design.

Let B denote the total number of blocks and N_{ijt} be the number

of comparisons for the $(i, j)^{\text{th}}$ pair in the t^{th} block ($1 \leq i < j \leq k$;

$t = 1, 2, \dots, B$). Let $N_t = \sum_{i=1}^K \sum_{j>i} N_{ijt}$ and $N = \sum_{t=1}^B N_t$. Let X_{it} be

the random variable denoting the characteristic of the i^{th} treatment in the t^{th} block and let $F_{it}(x)$ denote the distribution function corresponding to it. If $G_{ijt}(x)$, a continuous c.d.f., denotes the distribution function of the difference random variable $Z_{ijt} = X_{it} - X_{jt}$, then the null hypothesis is $G_{ijt}(x) = G_{i'j't'}(x)$. For any two treatment pairs (i, j) and (i', j') and the same t ($1 \leq i < j \leq k$; $t = 1, 2, \dots, B$), let R_{ijt} and S_{ijt} denote the sum of ranks of positive and negative Z 's respectively for the combination (i, j, t) and let $V_{ijt} = R_{ijt} - S_{ijt}$.

If $N_{ijt} = n$, for all combinations (i, j, t) , then to test the null hypothesis use

$$(4.4.3) \quad L^* = \frac{6}{n(N'+1)(2N'+1)K\mu} \sum_{i=1}^K \sum_t \sum_{j \neq i} V_{ijt}^2$$

where

λ = the number of treatments compared in each block,

μ = the number of blocks in which each pair (i, j) ,

$(1 \leq i < j \leq K)$ of treatments is compared,

h = the number of blocks in which each individual treatment i appears ($i = 1, 2, \dots, K$),

$\mu = \lambda(\lambda - 1)B/K(K - 1)$,

$$h = (K - 1) \mu / (\lambda - 1),$$

$$N' = 1/2 \lambda (\lambda - 1)n.$$

L^* is distributed approximately as a chi square random variable with $K - 1$ df (Mehra, 1964).

5. Balanced Incomplete Block.

In an incomplete block design, all of the treatments do not occur in every block. When every pair of treatments occurs together an equal number of times, the design is called balanced.

There are five conditions that determine a unique design. There are t treatments in b incomplete blocks of size k with n or r replicates on each treatment. In a balanced design, each pair of treatments occur together in the incomplete blocks in equal number, λ , times. By choosing values for three of these, one can solve for the other two by using the following relationships:

$$(5.1.1) \quad tr = kb$$

$$(5.1.2) \quad (t - 1)\lambda = r(k - 1)$$

5.1 Bradley's Rank Analysis of Incomplete Block Designs.

The rank analysis of incomplete block designs is a very practical procedure when one is concerned with ranking a large number of items or

when one wants to be sure of significance in ranking. It has been shown by Bradley (1955) that the method of rank analysis being compared to the analysis of variance has an asymptotic efficiency of from $1.5/\pi$ to $3/\pi$ depending upon the sample size. The smaller the number of treatments the better the asymptotic efficiency.

An experiment with t treatments involving paired comparisons will be considered. Assume that the treatment t_i has a true rating π_i , where π_i is a number that helps rate treatment i , ($\pi_i > 0, \sum \pi_i = 1$) and that the data are continuous. It is also assumed that when treatment i appears with treatment j and is ranked with the k observations in a block, the probability that treatment i has the higher rank is $\pi_i / (\pi_i + \pi_j)$.

The observations within each block are ranked relative to each other, and r_{ijk} will designate the rank of the i^{th} treatment in the k^{th} repetition of the block in which treatment i appears with treatment j . Estimates of π_1, \dots, π_t will be denoted by p_1, \dots, p_t respectively, and n will be reserved to denote the number of repetitions of the design where a repetition is defined to be a set of all pairs of treatments.

The likelihood function may be obtained by assuming statistical independence between blocks or pairs of treatments (Bradley and Terry, 1952). The likelihood function in the general form will be:

$$(5.1.3) \quad L = \prod_i \pi_i^{2n(t-1) - \sum_k \sum_j r_{ijk}} \prod_{i < j} (\pi_i + \pi_j)^{-n}.$$

There are two ways to analyze this experiment. The first way is the pooled analysis. This is when the experimenter is willing to assume that the treatments are homogeneous for every repetitions. This pooled analysis is very similar to pooled analysis in a χ^2 experiment and the same things are lost when performing the pooled analysis. The pooled analysis has the null hypothesis that $\pi_i = \frac{1}{t}$ for all i with the alternative that $\pi_i \neq \frac{1}{t}$ for some i . The π_i will be estimated by p_i and the likelihood ratio test depends on the statistic

$$(5.1.4) \quad B_1 = n \sum_{i < j} \log_{10}(p_i + p_j) - \sum_i [2n(t-1) - \sum_{k=1}^n \sum_{j \neq i} r_{ijk}] \log_{10} p_i.$$

Where p_i is found by solving the equations

$$(5.1.5) \quad \frac{a_i}{p_i} - n \sum_{j \neq i} (p_i + p_j)^{-1} = 0 \quad i, j = 1, \dots, t$$

and

$$(5.1.6) \quad \sum p_i = 1.$$

where

$$(5.1.7) \quad a_i = 2n(t-1) - \sum_{j \neq i} \sum_{k=1}^n r_{ijg}.$$

For small values of n the B_1 may be found in tables by Bradley and Terry (1952) and Bradley, R. A. (1954). With these tables, it is not

necessary to compute (5.1.4). The B_1 and the significant level can be read directly by looking at the proper values of p_i and $\sum_{i=1}^n r_i$ which is the sum of the ranks of treatment i . The tables will give values for a design as large as $t = 5$, $\lambda = 1$, $b = 10$, $r = n = 4$ and $k = 2$. For designs larger than this, a chi square approximation must be used. Bradley and Terry (1952) have shown that $-2 \ln \lambda_1$ is an approximate chi square random variable with $t - 1$ df, where

$$(5.1.8) \quad -2 \ln \lambda_1 = nt(t - 1) \ln 2 - 2B_1 \ln 10.$$

The second way to analyze this data is with a combined analysis. In many cases the assumption that the same true ratings exist for all groups is not realistic. If the detection of treatment differences is the main concern of the experimenter, a pooled analysis may be inappropriate and even give a non-significant result, while each alone exhibits significant treatment differences. This is particularly likely to happen where judge preferences may prohibit the setting up of uniform ranking criteria.

The combined analysis procedure is very similar to what Fryer (1966) calls the sum of χ^2 procedure, except the $\sum_{u=1}^n B_1^u$ is used instead of the $\sum \chi^2$. Where the n repetitions are subdivided into g groups of repetitions with n_u repetitions in the u^{th} group, $u = 1, \dots, g$.

The null hypothesis for this test is that $\pi_{iu} = \pi_i$ for all i and u , where u stands for the u^{th} repetition group, against the alternative $\phi_{iu} \neq \phi_i$ for some i and u . The B_1^u corresponds to B_1 of the pooled analysis for the n homogeneous repetitions of the u^{th} group and

$$(5.1.9) \quad B_1^c = \sum_{u=1}^n B_1^u.$$

The exact probability of B_1^c can be read from Bradley and Terry (1952). For large samples $-2 \ln \lambda_1^c$ is distributed as an approximate chi square random variable with $n(t - 1)$ df, where

$$(5.1.10) \quad -2 \ln \lambda_1^c = nt(t - 1) \ln 2 - 2B_1^c \ln 10.$$

Significance would be regarded as treatment difference although some of the judges may have disagreed on the treatments.

To test if the judges have agreed or disagreed, something similar to a treatment with judge interaction may be tested. This test can be made by testing $-2 \ln \lambda_a$ which is an approximate chi square random variable with $(n - 1)(t - 1)$ df, where

$$(5.1.11) \quad \begin{aligned} -2 \ln \lambda_a &= -2(\ln \lambda_1^c - \ln \lambda_1) \\ &= 2(B_1 - B_1^c) \ln 10. \end{aligned}$$

If $-2 \ln \lambda_a$ is significant, it can be concluded that the judges were in disagreement among themselves, although each judge was consistent

throughout the experiment.

The procedure is very complicated to explain without an example. Therefore, an example given by Bradley and Terry will be used to explain the procedure. This experiment has three groups of pigs fed three rations and tests the taste of the resulting pork roasts.

Step 1. A competent panel of judges was selected and so instructed that all had experience with the experimental material.

Step 2. For each judge and for each repetition, six small containers were coded. Two samples from roasts from each of the three treatment groups of animals were placed in the containers and the three requisite pairs formed. Code numbers were recorded and the pairs presented to the judges in a random order together with score cards.

Step 3. For each pair a judge tasted each sample and recorded the value 1 for the sample preferred and 2 for the other sample.

Step 4. (Analysis). The experimenter collected and decoded the data for each judge and recorded the results as follows: C denotes the corn ration, Cp the corn plus peanut supplement ration, and CP the corn plus large peanut supplement ration. The treatment sums of ranks, $\sum r_i$, for C, Cp, CP are respectively 19, 13, 13, and 13, 15, and 17 for the two judges.

Table (5.1.1)

	1			2			3			4			5		
	C	C _p	CP	C	C _p	CP	C	C _p	CP	C	C _p	CP	C	C _p	CP
Pair		Judge 1													
C, C _p	2	1	-	2	1	-	2	1	-	2	1	-	2	1	-
C, CP	2	-	1	1	-	2	2	-	1	2	-	1	2	-	1
C _p , CP	-	2	1	-	1	2	-	2	1	-	1	2	-	2	1
		Judge 2													
C, C _p	2	1	-	2	1	-	1	2	-	1	2	-	1	2	-
C, CP	1	-	2	1	-	2	1	-	2	1	-	2	2	-	1
C _p , CP	-	1	2	-	1	2	-	2	1	-	2	1	-	1	2

Step 5. Since it was agreed that the results of the two judges should be combined, we enter the table A at $n = 5$ (Bradley and Terry, 1952, and Bradley, R. A., 1954). For judge 1, $\Pr(C) = 0.05$, $\Pr(C_p) = 0.47$, $\Pr(CP) = 0.47$, $B_1 = 2.917$, the significance level is 0.057; for judge 2 $\Pr(C) = 0.53$, $\Pr(C_p) = 0.30$, $\Pr(CP) = 0.17$, $B_1 = 4.034$ and the significance level is 0.404.

Step 6. The combined statistic B_1^C was obtained and has the value $2.917 + 4.034 = 6.951$. From the table B under the two equal groups, $n = 10$, the significance level for the combined test was found to be 0.069. It was concluded that it had not been demonstrated that ration difference was detectable by these judges.

Step 6a. If a decision to pool the data had been made, treatment sums of ranks added over the judges would have been 32, 28, 30, and the table A would have been used for $n = 10$. We would have found $\Pr(C) = 0.24$, $\Pr(C_p) = 0.43$, $\Pr(CP) = 0.32$, $B_1 = 8.797$ and the significance level would have been 0.630.

5.2 Durbin's Rank Analysis of Incomplete Block Design.

This procedure has some distinct advantages over Bradley's rank analysis of incomplete blocks. One of the biggest advantages is the use of the standard F tables instead of special tables. Durbin's method of computation is straightforward and easier to understand than Bradley's method. However, Bradley's method is quicker for small samples, once it is understood, and the test is more versatile.

For the balanced incomplete block experiment, the t objects or treatments are presented in blocks of k , and each treatment is ranked r times relative to the other treatments in that block. Then the number of blocks is b and the number of blocks in which a particular pair of objects occurs is λ . In each experiment, the treatments within each block are ranked. If there is no difference in the treatments, the expected value of the sum of ranks for each of the treatments would be equal. Thus, the null hypothesis is that the preferences are random, or that the treatment rank totals are not significantly different.

Let X_{ij} be the rank of the j^{th} object in the i^{th} replication and

$X_j = \sum_{i=1}^n X_{ij}$. Then Durbin (1951) shows that Kendall's coefficient of con-

cordance is

$$(5.2.1a) \quad W = \frac{12 \sum_{j=1}^t X_j^2 - 3tr^2(k+1)^2}{\lambda^2 t(t^2 - 1)} \quad \text{or}$$

$$(5.2.1b) \quad W = \frac{12 S}{\lambda^2 t(t^2 - 1)}$$

where

$S =$ the sum of squared deviations from the mean of the ranks.

This F^* is distributed approximately as an F random variable

where

$$(5.2.2) \quad F^* = \frac{\left(\frac{\lambda(t+1)}{k+1} - 1\right) W}{1 - W}$$

with

$$(5.2.3) \quad v_1 = \frac{rt \left(1 - \frac{k+1}{\lambda(t+1)}\right)}{\left(\frac{tr}{t-1} - \frac{k}{k-1}\right)} - \frac{2(k+1)}{\lambda(t+1)}$$

and

$$(5.2.4) \quad v_2 = \left(\frac{\lambda(t+1)}{k+1} - 1\right) v_1.$$

where v_1 and v_2 are the approximate degrees of freedom for the F distribution.

The F^* statistic is valid for any t and r , although the computations are rather long. For moderate t and large r the X^2 statistic provides an easier test than does the F^* statistic.

For moderate t and large r , the X^2 is approximately distributed as a χ^2 random variable with $t - 1$ degrees of freedom, where

$$(5.2.5) \quad X^2 = \frac{\lambda (t^2 - 1)}{k + 1} W.$$

Thus the null hypothesis can be accepted or rejected by looking in the χ^2 or F tables.

To make this procedure clear, an example from Durbin (1951) is given as follows.

Suppose it is desired to rank seven objects, A B C D E F G, in blocks of three. The Youden square layout shown below is ideal for this experiment (Federer, 1963).

Table (5.2.6)

A	B	C	D	E	F	G
B	C	D	E	F	G	A
D	E	F	G	A	B	C

Each column represents a block, and each row represents a single replication of the objects.

Suppose an ice cream manufacturer wants to test the taste preference of several people on his seven varieties of ice cream. If the above Youden square were repeated 3 times or if 3 similar Youden squares were used, the resulting 21 blocks could be assigned to 21 different tasters, where each taster samples three kinds of ice cream and ranks them from 1 to 3, giving the rank of 3 to his favorite flavor. The experiment was completed and the totals of the ranks for each treatment are as follows:

A	B	C	D	E	F	G
20	13	18	25	22	12	16

The mean of the rank totals is 18. Therefore, $S = 2^2 + 5^2 + 0^2 + 7^2 + 4^2 + 6^2 + 2^2 = 134$. From the layout, it is known that $t = 7$, $r = 9$, $k = 3$, $b = 21$ and $\lambda = 3$, these results give $W = .532$. Then the computed F^*

$$\frac{\left(\frac{3(8)}{4} - 1\right) (.532)}{1 - .532} = 5.684.$$

The degrees of freedom for the tabled F are $v_1 = 5.5 \approx 5$ and $v_2 = 27.5 \approx 27$, thus with $\alpha = .05$ the tabled value is $F(5, 27) = 2.57$. Therefore, reject the null hypothesis that the preferences were at random.

The example should clearly show that this test is straightforward and uncomplicated.

6. A 2×2 Factorial Experiment.

A 2×2 factorial experiment is designed to examine the effect of two factors. Each factor is applied at 2 levels, with 1 or more replications, so that a difference in effect can be observed.

Abelson and Bradley (1954) have extended the Bradley and Terry (1952) likelihood ratio test to a 2×2 factorial experiment. The Abelson and Bradley (1954) procedure is the only paper that gives the details for performing factorial experiments. However, Sheffield (1957) makes some comments on a possible extension of the median test to a 3×3 factorial experiment.

The Abelson and Bradley (1954) 2×2 factorial experimental has an asymptotic efficiency of $3/\pi$ when compared to the parametric 2×2 factorial experiment.

This procedure has one big disadvantage; its complicated computations. This test could be useable only with the help of a computer.

The Abelson and Bradley (1954) procedure considers factors A and B, at two levels each, with parameters representing the effects of the levels of these factors given by α_1 , α_2 , β_1 and β_2 respectively.

It is understood that $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $i = 1, 2$ and the restraints

$$(6.1) \quad \alpha_1 + \alpha_2 = 1 \quad \text{and} \quad \beta_1 + \beta_2 = 1$$

The four treatment combinations are designated by $T(\alpha_1, \beta_1)$, $T(\alpha_2, \beta_1)$,

$T(\alpha_1, \beta_2)$, and $T(\alpha_2, \beta_2)$. Let $\Sigma r(\alpha_1)$ represent the total sum of ranks for all comparisons and all n repetitions of the two treatment combinations containing factor A at level one. Similar definitions follow for the other sums. The likelihood function is

$$(6.2) \quad L(\alpha, \beta) = \alpha_1^{c_1} \alpha_2^{c_2} \beta_1^{d_1} \beta_2^{d_2} [(\alpha_1 \beta_1 + \alpha_2 \beta_2)(\alpha_1 \beta_2 + \alpha_2 \beta_1)]^{-n} [(\alpha_1 + \alpha_2)(\beta_1 + \beta_2)]^{-2n},$$

where

$$c_1 = 11n - \Sigma r(\alpha_1)$$

$$c_2 = 11n - \Sigma r(\alpha_2)$$

$$d_1 = 11n - \Sigma r(\beta_1)$$

$$d_2 = 11n - \Sigma r(\beta_2)$$

n = the number of replications.

The maximum likelihood estimates of $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are

$$(6.3) \quad \hat{\beta}_1 = d_1/4n \qquad \hat{\beta}_2 = d_2/4n$$

$$\hat{\alpha}_1 = c_1/4n \qquad \hat{\alpha}_2 = c_2/4n$$

A summary of the test procedures will be given with an explanation of the test statistic to follow the table, where the π_{ij} are the same as those defined in section 5.1.

Table (6.1)

Test Procedures for the 2 x 2 Factorial			
Test	Hypotheses	Statistic	df
Treatment effects assuming no interaction	$H_0 = H_1: \alpha_1 = \alpha_2, \beta_1 = \beta_2.$ $H_a = H_3: \pi_{ij} = \alpha_i \beta_j; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.	$-2 \ln \lambda_{13}$	2
A-factor	$H_0 = H_2: \alpha_1 = \alpha_2, \beta_1$ not assumed equal to $\beta_2.$ $H_a = H_3: \pi_{ij} = \alpha_i \beta_j; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.	$-2 \ln \lambda_{23} =$ $-2(\ln \lambda_{13} - \ln \lambda_{12}).$	1
B-factor assuming no A-factor effect	$H_0 = H_1: \alpha_1 = \alpha_2, \beta_1 = \beta_2.$ $H_a = H_2: \alpha_1 = \alpha_2, \beta_1 \neq \beta_2.$	$-2 \ln \lambda_{12}$	1
B-factor	$H_0 = H_{2'}: \beta_1 = \beta_2, \alpha_1$ not assumed equal to $\alpha_2.$ $H_a = H_3: \pi_{ij} = \alpha_i \beta_j; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.	$-2 \ln \lambda_{2'3} =$ $-2(\ln \lambda_{13} - \ln \lambda_{12'}).$	1
A-factor assuming no B-factor effect	$H_0 = H_1: \alpha_1 = \alpha_2, \beta_1 = \beta_2.$ $H_a = H_{2'}: \alpha_1 \neq \alpha_2, \beta_1 = \beta_2.$	$-2 \ln \lambda_{12'}$	1
A X B interaction	$H_0 = H_3: \pi_{ij} = \alpha_i \beta_j; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both. $H_a = H_4: \pi_{ij} \neq \alpha_i \beta_j$ for some i and $j.$	$-2 \ln \lambda_{34} =$ $-2(\ln \lambda_{14} - \ln \lambda_{13}).$	1
Treatment effects admitting interaction	$H_0 = H_1: \pi_{ij} = 1/4.$ $H_a = H_4: \pi_{ij}$ possibly unequal to $\alpha_i \beta_j$ for some i and $j.$	$-2 \ln \lambda_{14}$	3

The formulae for the above test statistics are as follows:

$$(6.4) \quad -2 \ln \lambda_{13} = 12 n \ln 2 + 2 c_1 \ln \hat{\alpha}_1 + 2 c_2 \ln \hat{\alpha}_2 + 2 d_1 \ln \hat{\beta}_1 \\ + 2 d_2 \ln \hat{\beta}_2 - 2 n \ln (\hat{\alpha}_1 \hat{\beta}_1 + \hat{\alpha}_2 \hat{\beta}_2) - 2 n \ln (\hat{\alpha}_1 \hat{\beta}_2 + \hat{\alpha}_2 \hat{\beta}_1)$$

$$(6.5) \quad -2 \ln \lambda_{12} = 2 d_1 \ln d_1 + 2 d_2 \ln d_2 - 8 n \ln n - 8 n \ln 2$$

$$(6.6) \quad -2 \ln \lambda_{12'} = 2 c_1 \ln c_1 + 2 c_2 \ln c_2 - 8 n \ln n - 8 n \ln 2$$

$$(6.7) \quad -2 \ln \lambda_{14} = 12 n \ln 2 + 2 \sum_{i,j} [6 n - \sum_r (\alpha_i, \beta_j)] \ln p_{ij} \\ - 2 n \sum_{\substack{i \leq g, j \leq h \\ (i, j \neq g, h)}} \ln (p_{ij} + p_{gh}) = B_1.$$

B_1 can be found in the tables by Bradley and Terry (1952).

A rank analysis for 2×2 factorial treatments in the method of paired comparisons has been presented. The theory could be developed for general factorial sets of treatments but the solutions of estimating equations present formidable difficulties (Abelson and Bradley, 1954). The 2×2 factorial could be used in taste testing experiments, where paired comparisons are particularly appropriate, and there may be other applications.

7. Partially Balanced Incomplete Block

The partially balanced incomplete block (PBIB) design is a very complex analysis of variance. There are times when such a design would save considerable time and money in conducting an experiment, but

generally this design has more theoretical than practical use. In the balanced incomplete block, all pairs of treatments occur λ_i times; this sometimes requires a very large number of repetitions. The partially balanced incomplete block experiments have a set of two or more λ_i 's which could change the number of replications required.

In many practical situations, the proper assumptions for the analysis of variance of the PBIB do not hold. Therefore, Giri (1965) has shown that a reasonable approximation of the randomization test based on the F statistic is equivalent to modifying the normal theory test by multiplying both of the numbers of df of the F distribution by a factor depending on the heterogeneity of variance of the blocks.

The discussion is limited to 2 associate PBIB designs with $\lambda_1 = 0$, where one PBIB is associated with λ_1 and the other with λ_2 . $\lambda_1 = 0$ was selected to make the computations less involved. For two associated PBIB there must be two λ 's, λ_1 and λ_2 , and related to the λ 's are n_1 and n_2 , where n_i is the number of treatment pairs that occur together in the same block λ_i times.

The null hypothesis is that there are no significant treatment effects. Giri (1965) states that an approximation to the randomization test of the null hypothesis based on the statistic F is equivalent to modifying the normal theory test by multiplying both the numbers of df by the factor ϕ , using

$$(7.1) \quad \phi = \frac{2 [b(k-1) - (t-1)] k b^2 (k-1)^2}{[(t-1)(\lambda_2 - 1) + 2(r-1)(t-1)(k-1) + 2n_1 \lambda_2] \left(1 - \frac{V}{b-1}\right) - 2}$$

Where the design parameters are the same as those defined for the BIB design. When the block variances are homogeneous, $V = 0$. When they are not homogeneous, $V = b(k-1)$. Giri (1965) says in practice V will rarely be zero. However, this gives a lower bound to values of ϕ . In general ϕ will deviate from unity and the adjusted test will be more sensitive than the unadjusted normal theory test. As an example, table (7.2) will present the unadjusted and the adjusted values of F for some two associated PBIB's.

Table (7.2)

Parameters of PBIBD (First Kind)	Upper 5 per cent F (Unadjusted) Value	Upper 5 per cent F (Adjusted) Value	
		$V = 0$	$V = 1/2(b-1)$
$t = 14, b = 16, r = 8$ $k = 7, n_1 = 1, n_2 = 12$ $\lambda_1 = 0, \lambda_2 = 4$	1.78	1.78	1.50
$t = 15, b = 50, r = 10$ $k = 3, n_1 = 4, n_2 = 10$ $\lambda_1 = 0, \lambda_2 = 2$	1.91	1.93	1.57
$t = b = 9, r = k = 3$ $n_1 = 2, n_2 = 6, \lambda_1 = 0$ $\lambda_2 = 1$	3.07	3.58	2.47

This discussion of the nonparametric PBIB designs by no means gives all of the necessary information for performing a PBIB experiment. One must be completely familiar with the parametric PBIB to be able to do a nonparametric PBIB. This is because the computations are the same with the only difference being in the value of F used for rejection. For information on the analysis of the PBIB design, see Federer (1963).

8. Other Related Tests

8.1. The Randomization Test

The randomization tests not only have practical value, but also have value in showing the general nature of nonparametric tests.

With the randomization test, the exact probability under H_0 can be obtained without making any assumptions about normality or homogeneity of variance. Under most conditions, randomization tests are the most powerful of the nonparametric test (Fisher, 1935).

The randomization test can be extended to most models of analysis of variance. However, there is one major drawback. Except for small samples, the computations required for the rejection region are extremely long.

Therefore, this procedure will not be discussed in detail. However, with the use of a computer, this procedure might become useable.

8.2. Rank Methods in Analysis of Variance.

Hodges and Lehmann (1962) have given a procedure using ranks that is a simplification of the randomization test. They have explained the 2 sample blocked test and say it can be extended to a randomized complete block design. There is an example of an analysis of two latin squares. There is also an analysis of a balanced incomplete block experiment presented. The use of ranks make this test much easier to compute than the randomization test. However, as the experiment increases in size, the computations become complicated. The test is much more useable than the randomization test although it would require the services of a computer for large experiments.

This procedure has an asymptotic efficiency of $\pi/3$, when compared to similar normal tests.

8.3. Multiple Comparisons

Used in conjunction with the parametric analysis of variance are multiple comparisons. These multiple comparisons are very useful for completing the analysis of the data.

Nemenyi (1963) has prepared his Ph.D. dissertation on distribution-free multiple comparisons. Some of these multiple comparisons procedures require a previous analogue to an analysis of variance; others do not. This fact makes Nemenyi's (1963) dissertation and this report compliments to

each other for an experimenter wishing to perform nonparametric analysis of his data.

ACKNOWLEDGEMENTS

I am very much indebted to Dr. W. J. Conover for his advice and guidance in the preparation of this report. I am also indebted to Dr. K. S. Banerjee and Marvin Lenter for their help in the interpretation of certain papers and to Dr. Leslie Marcus for his guidance in finding references.

REFERENCES

- Abelson, R. M. and Bradley, R. A. (1954). A 2 X 2 Factorial with Paired Comparisons. Biometrics 10 487-502.
- Alluisi, E. A. (1956). Computational Formulae for a Distribution Free Test of Analysis-of-Variance Hypothesis. WADD Tech. Report No. 56-339.
- Bell, C. B. and Doksum, K. A. (1965). Some New Distribution-Free Statistics. The Ann. of Math. Stat. 36 203-214.
- Bradley, J. V. (1960). Distribution Free Statistical Tests. WADD Tech. Report. 60-661.
- Bradley, R. A. and Terry, M. E. (1952). The Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons. Biometrika 39 324-335.
- Bradley, R. A. (1954). The Rank Analysis of Incomplete Block Designs. II Additional Tables for the Method of Paired Comparisons. Biometrika 41 502-537.
- Bradley, R. A. (1955). Rank Analysis of Incomplete Block Designs. III. Biometrika 42 450-470.
- Cochran, W. G. (1950). The Comparison of Percentages in Matched Samples. Biometrika. 37 256-262.
- Cochran, W. G. (1954). Some Methods for Strengthening the Common χ^2 Tests. Biometrics. 10 417-451.
- Conover, W. J. (1965a). Several k-Sample Kolmogorov-Smirnov Tests. The Ann. of Math. Stat. 36 1019-1026.
- Conover, W. J. (1965b). A k-Sample Model in Order Statistics. The Ann. of Math. Stat. 36 1223-1235.
- Conover, W. J. (1966). Two k-Sample Slippage Tests. An unpublished manuscript.
- Durbin, J. (1951). Incomplete Blocks in Ranking Experiments. British J. of Psychology, Statistical Section. 4 85-90.

- Federer, W. T. (1963). Experimental Design. The MacMillan Co., New York.
- Fisher, R. A. (1935). The Design of Experiments. Oliver and Boyd, Edinburgh.
- Friedman, Milton. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. J. of the Amer. Stat. Assoc. 32 675-701.
- Fryer, H. C. (1966). Concepts and Methods of Experimental Statistics. Allyn and Bacon Inc., Boston.
- Gabriel, K. R. (1963). Analysis of Variance of Proportions with Unequal Frequencies. J. of the Amer. Stat. Assoc. 58 1133-1157.
- Giri, N. (1965). On the F-Test in the Intrablock Analysis of a Class of Two Associate PBIB Designs. J. of the Amer. Stat. Assoc. 60 285-294.
- Goodman, L. A. (1964). Simple Methods for Analysing Three-Factor Interaction in Contingency Tables. J. of the Amer. Stat. Assoc. 59 319-353.
- Hodges, J. L., Jr. and Lehmann, E. L. (1962). Rank Methods for Combination of Independent Experiments in the Analysis of Variance. The Ann. of Math. Stat. 33 482-497.
- Hopkins, J. W. (1954). Incomplete Block Rank Analysis: Some Taste Test Results. Biometrics. 10 391-402.
- Krishna, Iyer, P. V. (1951). A Nonparametric Method of Testing k Samples. Nature 167 33.
- Lancaster, H. O. (1951). Complex Contingency Tables Treated by the Partition of χ^2 . J. of the Royal Stat. Soc. 13 242-249.
- Mehra, K. L. (1964). Rank Tests for Paired Comparisons. The Ann. of Math. Stat. 35 122-137.
- Mosteller, Frederick. (1948). A k-Sample Slippage Test for an Extreme Population. Ann. of Math. Stat. 19 58-65.
- Mood, A. M. (1950). Introduction to the Theory of Statistics. McGraw-Hill Book Co., New York.

- Nemenyi, P. (1963). Distribution-Free Multiple Comparisons. Ph.D. Dissertation. State University of New York Downstate Medical Center.
- Pitman, E. J. G. (1937). Significance Tests Which May be Applied to Samples from Any Populations. Supplement to the J. of the Royal Stat. Soc. 4 119-30.
- Sheffield, F. D. (1957). Comments on a Distribution Free Factorial Design Analysis. Psychological Bulletin 54 426-428.
- Savage, I. R. (1953). Bibliography of Nonparametric Statistics and Related Topics. J. of the Amer. Stat. Assoc. 48 844-906.
- Siegel, S. (1956). Nonparametric Statistics. McGraw-Hill Book Company, New York.
- Sutcliff, J. P. (1957). A General Method of Analysis of Frequency Data for Multiple Classification Designs. Psychological Bulletin. 54 134-137.
- Van Elteren, P. and Noether, G. E. (1959). The Asymptotic Efficiency of the χ^2 Test for a Balanced Incomplete Block Design. Biometrika 46 475-476.
- Wilson, K. V. (1956). A Distribution-Free Test of Analysis of Variance Hypothesis. Psychological Bull. 53 98-101.

NONPARAMETRIC ANALOGUES OF ANALYSIS OF VARIANCE

by

RONALD K. LOHRDING

B. A. , Southwestern College, 1963

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

ABSTRACT

The analysis of variance is one of the most important statistical tools of research. However, many times proper assumptions do not hold for the analysis of variance. This paper, nonparametric analogues of analysis of variance, provides procedures which minimize the necessary assumptions.

The purpose of this report is to provide a useful and practical paper on the nonparametric analogues of the analysis of variance.

The first design discussed is the completely randomized design. The most popular and probably the most powerful of these analogues is the Kruskal-Wallis analysis of variance by ranks. There is also a discussion of two possible approaches to a k-sample slippage test. These two tests have the advantage of being quick and simple. The next analogue of this section involves a k-sample Kolmogorov-Smirnov test that makes a test of significance on k cumulative distributions. Last is the well known contingency χ^2 test for k independent samples.

The second analogue is of the randomized complete block design. The first test considered is the Cochran Q test. This is used for only special types of data, usually of the + or - type. Next is the Friedman two-way analysis of variance by ranks. This is the most widely known test of this section.

The third major area includes procedures that extend over several designs. These procedures are quite useful because an experimenter needs only to be acquainted with one process and then can generalize over several designs. A large part of this section is devoted to median tests which are analogues to several analysis of variance designs. The first of these procedures was developed by Mood (1950). Wilson (1956) revised Mood's method but Alluisi (1956) was responsible for developing the computational formulae which make this procedure very useable. Alluisi extends this procedure to three factors which is very similar to the three-factor analysis of variance design. The last two tests of this section are recently developed tests involving rank analysis. These tests have procedures for completely randomized and randomized complete block analogues. These two tests appear to have considerable power and could be the most important part of this paper.

Section 5 includes two procedures for a balanced incomplete block analogue. These two tests are very valuable when an experimenter is interested in having judges rank several items accurately.

The only nonparametric factorial experiment found was a 2×2 factorial with paired comparisons. This test is of little use due to the extreme complexity of computations.

The last major section is concerned with partially balanced incomplete block design. This test by Giri (1965) gives procedures for a reasonable approximation to the randomization test using the F statistic.

The final section is a short summary of other related tests and papers that might be of considerable help to other persons doing research in this area.

