NONPARAMETRIC STATISTICAL METHODS FOR

THE RANDOMIZED COMPLETE BLOCK DESIGN

by 5 44

LI-CHUN TAO

B. S., Taiwan Provincial Chung-hsing University, 1954

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and

Computer Science

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1968

Approved by:

W J Conover

Major Professor

LD
2668
R4
1968
T26
c.2

TABLE OF CONTENTS

1. INTRODUCTION

As we know from the F-test in the analysis of variance, we usually compare the observed F-value which is computed from the data with the theoretical F-value in the conventional F table. The theoretical F-values in the table are taken from the F-distributions with the corresponding degrees of freedom and the specified probabilities. The F-distribution is obtained under the condition that the population is normal. So the population of the data in which the observed F-value is obtained should also be normal. This is the assumption we need in the analysis of variance.

In many situations the population of the data may not meet this condition. If the shape of the population distribution function is known, then we can use the proper transformation to make the data satisfy this essential condition. Otherwise, many nonparametric methods can be used. This report will deal mainly with the chi-square test in the randomized complete block design case. A large sample is necessary for using this method and the minimum sample size can be reached by a working rule stated in Section 2.

In the second section we state the difference between a two-way randomized complete block arrangement table and a two-way contingency table, with the binomial transformation using the pooled median changing the former to the latter one.

The third section discusses the test of independence between two attributes in χ^2-test, which is comparable to testing the interaction of two attributes in the analysis of variance case.

The fourth and fifth sections deal with the methods to compute various χ^2's concerned with different types of experimental data, in which, of however, the contingency table should be formed at first.

The sixth section contains the concepts about the expected frequencies of χ^2-test. In the seventh section appears a normal score transformation. This is introduced by Fisher and Yates (1943) and is used for the ranked data. If we transform the quantitative data into ranks at first, the numerical data can also be analyzed by this method. The last two sections compare the method of χ^2-test and F-test. The F-test is better for normal populations and the χ^2-test needs larger samples to have the same power as the F-test. Some comments arose about Wilson's χ^2-test from Sheffield and McNemar, who indicated that the χ^2-test has less power than F-test.

It is true that if the population of the data is normal, the F-test is better than any other method; otherwise, if the data is not drawn from the normal population then the F-test is no longer the better one. The non-parametric methods are like wearing loose suits made to cover most people but not giving them a good fit. The transformation is used in statistical methods to transform the data into a normal distribution to meet the test assumption. It seems to change people's weight to fit them into the proper suits. When all of these methods are used, we may certainly have something to gain and also something to lose. Therefore, if we can find the proper method of analysis for every kind of population, this is the best way to do our job.

2. Randomized Complete Block Arrangement and the Contingency Table.

2.1 The Difference Between the Randomized Complete Block Two-Way Table and the Contingency Table.

The data of a randomized complete block design is generally of two way classification with one observation in each cell or plot. The observations in the cell are usually numerical measurements.

This design is devised to compare t treatments in n plots, with each treatment replicated in b plots, so that bt equals n. The n plots are divided into b blocks, such that within any block the plots are as homogeneous as possible, and the variation among blocks is known. The t treatments are randomly allocated to the t plots in each block. With b replications, we require b separate randomizations. A two-way classification table of such an arrangement for a randomized complete block design is given in Table 2.2.1.

If the observations in the two-way table of the randomized complete block design are replaced by frequencies, that table becomes a two-way frequency table. The treatment and block are two classified attributes. This is generally called a two-way r x c contingency table, (Table 2.2.2).

2.2 The Change from Randomized Complete Block Two-Way Table into a Contingency Table.

A binomial transformation can be used to change the randomized complete block two-way table into a contingency table. Table 2.2.3 for example, is the transformed form of Table 2.2.1. The method of transformation is at first to find the median of each block and then replace each observation above its respective block median by 1 and below or equal to its block

Table 2.2.1

Two-way Classification Table of RCB Design

Treatment	Block						Total	Mean
	1	2	...	j	...	b		
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1b}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2b}	$y_{2.}$	$\bar{y}_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ib}	$y_{i.}$	$\bar{y}_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮
t	y_{t1}	y_{t2}	...	y_{tj}	...	y_{tb}	$y_{t.}$	$\bar{y}_{t.}$
Total	$y_{.1}$	$y_{.2}$	...	$\bar{y}_{.j}$	...	$y_{.b}$	$y_{..}$	
Mean	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.j}$	...	$\bar{y}_{.b}$		$\bar{y}_{..}$

Where y_{ij} is the observation of the i^{th} treatment and the j^{th} block

$y_{i.}$ is the sum of the i^{th} treatment

$\bar{y}_{i.}$ is the mean of the i^{th} treatment

$y_{.j}$ is the sum of the j^{th} block

$\bar{y}_{.j}$ is the mean of the j^{th} block

$y_{..}$ is the grand total of all n observations

$\bar{y}_{..}$ is the grand mean of all n observations

Table 2.2.2

r x c Contingency Table

Treatment	Block						Total
	1	2	...	j	...	c	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1b}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2b}	$n_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ib}	$n_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮
r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rb}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	$n_{..}$

Where n_{ij} is the number of observations of the i^{th} treatment and the j^{th} block

$n_{..}$ is the total frequencies, or the total number of the observations in the design

$n_{i.}$ is the sum of frequencies of the i^{th} treatment

$n_{.j}$ is the sum of frequencies of the j^{th} block median by 0, then the number of 1's for each treatment is considered to be the frequencies of the successes, af_i, and that of 0's is considered that of the failures, bf_i. Such a 2 x r contingency table is obtained from the data of randomized complete block design as in Table 2.2.3.

Table 2.2.3.

2 x r Contingency Table where 'a' Means Above The Median and 'b' Means Below or Equal to The Median

Variate	Classification						Total
	1	2	...	i	...	r	
a	af_1	af_2	...	af_i	...	af_r	n_a
b	bf_1	bf_2	...	bf_i	...	bf_r	n_b
Total	n_1	n_2	...	n_i	...	n_r	n

where 'a' means above the median

'b' means below or equal to the median

af_i is the number of observations above the median in the i^{th} treatment

bf_i is the number of observations below or equal to the median of the i^{th} treatment

n_a is the number of total observations of above the median

n_b is the number of total observations of below or equal to the median

n is the number of total observations of all observations

In the two-way table of a randomized complete block design, if the number of observations in each cell is more than one, then the method of transformation is slightly different from the preceding one. That is, the median used here is the pooled median, Md, which is obtained from all the n observations instead of from each block. The number of successes and failures for each treatment is determined by counting the number of observations above and the number below or equal to the pooled median, Md.

The binomial transformation for a randomized complete block experiment can be used only if both t and b are large enough to make all the expected frequencies greater than or equal to 5. That is, in binomial populations both np and n(1-p) or nq should be greater than or equal to 5. This is a working rule for making the transformation effectively.

3. Interaction and Independence.

An r x c two-way contingency table is usually constructed for the purpose of studying the relationship between two attributes. In particular, we may wish to test whether the two attributes are related and dependent. If the two attributes are not related to each other, this means they are independent. On the other hand, if the two-way table is numerical measurement data, independence indicates no interaction between these two attributes. Thus we test interaction between two attributes in numerical measurement data in the same sense as we test independence between two attributes in a r x c contingency table. The following simple 2 x 2 table of artificial data is a numerical example to illustrate no interaction between two attributes, A and B.

Table 3.1

2 x 2 Table of Artificial Data

A	B 1	B 2	Total
1	10	12	22
2	13	15	28
Total	23	27	50

To see this, we could check 10 - 12 = 13 - 15 and 10 - 13 = 12 - 15, this means that the difference between the observations corresponding to any two levels of A is the same for all levels of B, and the difference between the observations for two levels of B is the same for all levels of A. This means that there is no interaction between two attributes of A and B. On the other hand, if we consider a 2 x 2 contingency table and let A be the variate, in which A_1 is "success" and A_2 is "failure", then the data becomes a binomial form so that B_1 and B_2 are two binomial samples. Now we see that the two relative frequencies or two binomial sample means are approximately equal, or $10/23 \doteq 12/27 \doteq 22/50 \doteq 0.4$. Therefore, we would say that the two attributes A and B are independent or the two binomial sample means are approximately equal. The reason that they are not exactly equal is accounted for by the sampling variation. Nevertheless, from this point of view, we know that the purpose of testing hypotheses of interaction for numerical two-way data and that of independence for two-way contingency data is the same.

4. The χ^2-Test for Randomized Complete Block Designs.

4.1 One Observation Per Cell.

The chi-square test in nonparametric methods may be used in many cases like the analysis of variance in parametric methods to test the hypothesis that the r samples are drawn from the same population, or that the r population means are equal. The difference between them is that the χ^2 test deals with multinomial populations, while the analysis of variance deals with normal populations. Thus, for the non-parametric analysis of randomized complete block experimental data, we may at first transform the two-way table of numerical observations into a 2 x r two-way frequency contingency table, which is shown in the Table 2.2.3.

After the 2 x r contingency table is obtained, we can compute the statistic χ^2 as follows:

$$\chi^2 = \sum_{i=1}^{r} \left[\frac{\left(af_i - \frac{n_i n_a}{n}\right)^2}{\frac{n_i n_a}{n}} + \frac{\left(bf_i - \frac{n_i n_b}{n}\right)^2}{\frac{n_i n_b}{n}} \right]$$

$$= \frac{\frac{(af_1)^2}{n_1} + \frac{(af_2)^2}{n_2} + \dots + \frac{(af_i)^2}{n_i} + \dots + \frac{(af_r)^2}{n_r} - \frac{(n_a)^2}{n}}{\frac{n_a}{n} \cdot \frac{n_b}{n}} \qquad (4.1.1)$$

which is approximately chi-square with $(r - 1)$ degrees of freedom, where all the notations in the formula (4.1.1) are the same as in Table 2.2.3.

As we mentioned in the previous section, the binomial transformation for a randomized complete block experiment in many cases can be used only if both r and b are large. If r, the number of treatments, is small, the χ^2-value needs to be corrected. The corrected value is

$$\chi_c^2 = \left(\frac{r - 1}{r}\right) \chi^2 . \tag{4.1.2}$$

This correction term originated from the relation between the chi-square test of independence and the analysis of variance.

4.1.1 Friedman's χ^2-Test.

For the case of one observation per cell in randomized complete block design, Friedman (1937) suggested that a quick method to test the same hypothesis that r population means are equal is at first to rank the observations in each block from 1 to b. Let $R_{i.}$ be the sum of the ranks of the observations from the i^{th} treatment, we may compute

$$\chi_F^2 = \frac{12}{br(r + 1)} \sum_{i=1}^{r} (R_{i.})^2 - 3b(r + 1) \tag{4.1.1.1}$$

Where b is the number of blocks or replicates

r is the number of treatments

$R_{i.}$ is the sum of ranks in the i^{th} treatment.

Under the null hypothesis, this statistic, χ_F^2, is distributed approximately as χ^2 distribution with (r - 1) degrees of freedom.

The integers 12 and 3 in the formula are constants, not dependent on the size of the experiment. This approximation is poor for small values of r and b. Friedman has prepared tables (Siegel 1956) of the exact distribution of χ_F^2 for some pairs of small values of r and b.

4.1.2 Cochran's Q-Test

Another method for the same case contributed by Cochran (Siegel 1956) is the Q-test. This test is particularly suitable when the data are in a nominal or dichotomized ordinal scale, such as 'yes' or 'no'; 'alive' or 'dead', 'success' or 'failure', and so on. This test determines whether the r related samples come from the same population with respect to the frequency of successes in the various samples.

The steps for this test are at first in the two-way table, to assign a '1' to each 'success' and a '0' to each 'failure', and then to determine the statistic Q by substituting the observed values into the following formula;

$$Q = \frac{(r-1)\left[r\sum_{i=1}^{r} G_i^2 - (\sum_{i=1}^{r} G_i)^2\right]}{r\sum_{j=1}^{b} L_j - \sum_{j=1}^{b} L_j^2} . \qquad (4.1.2.1)$$

where G_i is the total number of 'successes' in the i^{th} treatment

L_j is the total number of 'successes' in the j^{th} block

r is the number of treatments

b is the number of blocks (replications).

Under the hypothesis that the r population means are equal this Q-value is distributed approximately as chi-square distribution with $(r - 1)$ degrees of freedom.

The significance of the observed value of Q may be determined by reference to an ordinary χ^2-table.

4.2 More Observations Per Cell

Suppose that there are r rows, c columns, and h observations per cell. The observations are denoted by y_{ijk} with $i = 1,2,\ldots,r$; $j = 1,2,\ldots,c$; and $k = 1,2,\ldots,h$. The two-way table can be transformed into a 2 x r x c contingency table, (Table 4.2.1) by using the pooled median, Md. This table can also be written as Table 4.2.2.

From Table 4.2.2 the total χ^2-value can be calculated in general as

$$\chi_T^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left[\frac{\left(af_{ij} - \frac{n_{ij}\, n_a}{n}\right)^2}{\frac{n_{ij}\, n_a}{n}} + \frac{\left(bf_{ij} - \frac{n_{ij}\, n_b}{n}\right)^2}{\frac{n_{ij}\, n_b}{n}} \right] \tag{4.2.1}$$

with $(rc - 1)$ degrees of freedom.

The hypothesis tested for this case is that the main effects and interaction effects produce no change in the distribution of the data population. If the number of observations for each cell of the r x c table, $n_{ij} = af_{ij} + bf_{ij}$, are all equal, and if $n_a = n_b = \frac{n}{2}$, then χ_T^2 can be written as

$$\chi_T^2 = \frac{4rc}{n} \sum_{i=1}^{r} \sum_{j=1}^{c} \left(af_{ij} - \frac{n}{2rc}\right)^2 \tag{4.2.2}$$

and also if $n_a \neq n_b$, but all n_{ij} are equal, then χ_T^2 can be expressed as

$$\chi_T^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left[\frac{\left(af_{ij} - \frac{n_a}{rc}\right)^2}{\frac{n_a}{rc}} + \frac{\left(bf_{ij} - \frac{n_b}{rc}\right)^2}{\frac{n_b}{rc}} \right] . \tag{4.2.3}$$

For computing row or treatment χ_R^2 and column or block χ_C^2, we could change Table 4.2.1 into the form of Table 4.2.3 and Table 4.2.4, or namely 2 x r and 2 x c contingency tables respectively, then the two statistics are in general

Table 4.2.1

$2 \times r \times c$ Contingency Table with "a" Means Above and "b" Means Below or Equal to the Median, Md.

		1	2	...	j	...	c	Totals	
1	a	af_{11}	af_{12}	...	af_{1j}	...	af_{1c}	$an_{1.}$	$n_{1.}$
	b	bf_{11}	bf_{12}	...	bf_{1j}	...	bf_{1c}	$bn_{1.}$	
2	a	af_{21}	af_{22}	...	af_{2j}	...	af_{2c}	$an_{2.}$	$n_{2.}$
	b	bf_{21}	bf_{22}	...	bf_{2j}	...	bf_{2c}	$bn_{2.}$	
⋮		⋮	⋮		⋮		⋮	⋮	⋮
i	a	af_{i1}	af_{i2}	...	af_{ij}	...	af_{ic}	$an_{i.}$	$n_{i.}$
	b	bf_{i1}	bf_{i2}	...	bf_{ij}	...	bf_{ic}	$bn_{i.}$	
⋮		⋮	⋮		⋮		⋮	⋮	⋮
r	a	af_{r1}	af_{r2}	...	af_{rj}	...	af_{rc}	$an_{r.}$	$n_{r.}$
	b	bf_{r1}	bf_{r2}	...	bf_{rj}	...	bf_{rc}	$bn_{r.}$	
Totals	a	$an_{.1}$	$an_{.2}$	...	$an_{.j}$	...	$an_{.c}$	n_a	
	b	$bn_{.1}$	$bn_{.2}$	...	$bn_{.j}$	...	$bn_{.c}$	n_b	
		$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$		n

Table 4.2.2

2 x rc Contingency Table

	11	12	...	1c	21	...	31	...	r1	...	rc	Total
a	af_{11}	af_{12}	...	af_{1c}	af_{21}	...	af_{31}	...	af_{r1}	...	af_{rc}	n_a
b	bf_{11}	bf_{12}	...	bf_{1c}	bf_{21}	...	bf_{31}	...	bf_{r1}	...	bf_{rc}	n_b
Total	n_{11}	n_{12}	...	n_{1c}	n_{21}	...	n_{31}	...	n_{r1}	...	n_{rc}	n

Table 4.2.3

2 x r Contingency Table

	1	2	...	i	...	r	Total
a	$af_{1.}$	$af_{2.}$	...	$af_{i.}$	...	$af_{r.}$	n_a
b	$bf_{1.}$	$bf_{2.}$	...	$bf_{i.}$	...	$bf_{r.}$	n_b
Total	$n_{1.}$	$n_{2.}$	...	$n_{i.}$	...	$n_{r.}$	n

Table 4.2.4

2 x c Contingency Table

	1	2	...	j	...	c	Total
a	$af_{.1}$	$af_{.2}$	...	$af_{.j}$	...	$af_{.c}$	n_a
b	$bf_{.1}$	$bf_{.2}$	...	$bf_{.j}$	...	$bf_{.c}$	n_b
Total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	n

$$\chi_R^2 = \sum_{i=1}^{r}\left[\frac{(af_{i.} - \frac{n_{i.}n_a}{n})^2}{\frac{n_{i.}n_a}{n}} + \frac{(bf_{i.} - \frac{n_{i.}n_b}{n})^2}{\frac{n_{i.}n_b}{n}}\right] \tag{4.2.4}$$

with $(r - 1)$ degrees of freedom, where $n_{i.} = \sum_{j=1}^{c} n_{ij}$, and

$$\chi_C^2 = \sum_{j=1}^{c}\left[\frac{(af_{.j} - \frac{n_{.j}n_a}{n})^2}{\frac{n_{.j}n_a}{n}} + \frac{(bf_{.j} - \frac{n_{.j}n_b}{n})^2}{\frac{n_{.j}n_b}{n}}\right] \tag{4.2.5}$$

with $(c - 1)$ degrees of freedom, where $n_{.j} = \sum_{i=1}^{r} n_{ij}$.

If $n_a = n_b = n/2$, and all n_{ij} are equal, the following two expressions can be used

$$\chi_R^2 = (\frac{4r}{n}) \sum_{i=1}^{r} (af_{i.} - \frac{n}{2r})^2 \tag{4.2.6}$$

where $af_{i.} = \sum_{j=1}^{c} af_{ij}$;

$$\chi_C^2 = (\frac{4c}{n}) \sum_{j=1}^{c} (bf_{.j} - \frac{n}{2c})^2 \tag{4.2.7}$$

where $bf_{.j} = \sum_{i=1}^{r} bf_{ij}$.

Also, if $n_a \neq n_b$ but all n_{ij} are equal, the following two formulas may be used.

$$\chi_R^2 = \sum_{i=1}^{r} \left[\frac{(af_{i.} - \frac{n_a}{r})^2}{\frac{n_a}{r}} + \frac{(bf_{i.} - \frac{n_b}{r})^2}{\frac{n_b}{r}} \right] \tag{4.2.8}$$

$$\chi_C^2 = \sum_{j=1}^{c} \left[\frac{(af_{.j} - \frac{n_a}{c})^2}{\frac{n_a}{c}} + \frac{(bf_{.j} - \frac{n_b}{c})^2}{\frac{n_b}{c}} \right] . \tag{4.2.9}$$

To detect the interaction effect of row and column we can compute χ_I^2 by subtracting, as is done in analysis of variance. That is

$$\chi_I^2 = \chi_T^2 - \chi_R^2 - \chi_C^2 \tag{4.2.10}$$

with $(r - 1)(c - 1)$ degrees of freedom.

The general expression for χ_I^2 is fairly complex and is given by Rao (1952).

5. Extension of Randomized Complete Block Design.

5.1 Randomized Complete Block Design with Two Treatments with One Observation Per Cell.

If only two treatments and b blocks are contained in the experimental data, the sign test may be used, and the computing method for this case is that a plus or minus sign is given to each difference of the b blocks, depending on whether the observation of the first treatment is greater or less than the observation of the second treatment. If there is no difference between the two treatments, plus and minus signs occur with equal probability. If the effect of the first treatment is greater than that of the second treatment one can expect an excess of plus signs, otherwise a deficit in plus signs. Therefore, the hypothesis that two treatment effects are equal is the same as that the probability of a plus sign is equal to 0.5, or $p = 0.5$.

Here again, a nonparametric method is essentially the binomial transformation. To test the hypothesis that $p = 0.5$, a χ^2-test may be used, provided that the number of blocks is greater than or equal to 10, by the working rule $bp \geq 5.0$.

Strictly speaking, the sign test is applicable only to the case in which all the b signs are either positive or negative. But in practice the two observations of a block are sometimes equal. When this occurs, such a block may be excluded from the test.

The χ^2 - value of the sign test is exactly the corrected chi-square χ_c^2 (4.1.2) for the randomized complete block experiment with 2 treatments and b blocks. This relation can be shown algebraically. The median of a block is the average of the two observations in that block. A plus sign implies that the first observation is greater than the second one in that block. Therefore, the number of observations greater than their block medians for the first treatment equals the number of observations less than their block medians for the second treatment. Therefore, the 2 x 2 contingency table is as follows:

	treat 1	treat 2	totals
no. of +'s	T	b - T	b
no. of -'s	b - T	T	b
totals	b	b	2b

The letter T in the above table is the number of plus signs. By the sign test

$$\chi^2 = \frac{(T - \frac{b}{2})^2}{b(0.5)(0.5)} = \frac{(2T - b)^2}{b} \quad . \tag{5.1.1}$$

By the method for randomized complete block experiment and formula (4.1.2)

$$\chi_c^2 = \frac{2 - 1}{2} \cdot \left[\frac{\frac{T^2}{b} + \frac{(b - T)^2}{b} - \frac{b^2}{2b}}{\frac{1}{2} \cdot \frac{1}{2}} \right], \tag{5.1.2}$$

which can be reduced to the same expression given in formula (5.1.1).

Other methods of nonparametric analysis for two related samples may be found in Siegel (1956).

5.2 Randomized Complete Block Design with Two Factors and no Combination

If the treatment contains two forms, A and C, both at m levels and also if there are b blocks in the experiment, the two-way arrangement is as given in Table 5.2.1.

For this data we may find the difference between corresponding levels of factor A and factor C in the b blocks.

To find the interaction between the factors and the blocks, the method is to tabulate the differences between values at corresponding levels for these two factors under the blocks. Then the next step is to determine the ranks of the differences (Table 5.2.2).

The following χ^2 - value can be used to test the hypothesis that two factors have no interaction with blocks.

Table 5.2.1

Two-way Table of Two Factor with No Combination in RCB Design

Treatment		Block				Total
		1	2	...	b	
A	1	a_{11}	a_{12}	...	a_{1b}	$a_{1.}$
	2	a_{21}	a_{22}	...	a_{2b}	$a_{2.}$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
	m	a_{m1}	a_{m2}	...	a_{mb}	$a_{m.}$
C	1	c_{11}	c_{12}	...	c_{1b}	$c_{1.}$
	2	c_{21}	c_{22}	...	c_{2b}	$c_{2.}$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
	m	c_{m1}	c_{m2}	...	c_{mb}	$c_{m.}$

$$\chi^2 = \frac{12}{mb(b+1)} \sum_{j=1}^{b} r_j^2 - 3m(b+1) \qquad (5.2.1)$$

with (b - 1) degrees of freedom, where b is the number of blocks, m is the number of levels, and r_j is the sum of ranks in the j^{th} block.

Table 5.2.2

The Difference and Rank Table of A-C

Level	Difference in Block I	Rank	Difference in Block II	Rank	...	Difference in Block b	Rank
1	$a_{11} - c_{11}$		$a_{12} - c_{12}$		...	$a_{1b} - c_{1b}$	
2	$a_{21} - c_{21}$		$a_{22} - c_{22}$		...	$a_{2b} - c_{2b}$	
$\vdots$	$\vdots$		$\vdots$		...	$\vdots$	
m	a_{m1} c_{m1}		a_{m2} c_{m2}		...	a_{mb} c_{mb}	
Total		r_1		r_2	...		r_b

The resulting χ^2-value can be compared with that of the conventional χ^2-value with respective degrees of freedom.

5.3 Randomized Complete Block Design with Three Factors and No Combination

If three factors, A, B and C, are involved in the treatment for a randomized complete block design, then the χ^2_I is the sum of χ^2's. One is obtained by finding the difference, A - B as the same manner shown in Table 5.2.2 for different blocks as in the last section, and another χ^2 is obtained by finding A + B - 2C for all blocks.

5.4 Randomized Complete Block Design with Four Factors and No Combination

In this case, we can use a similar procedure to find three components of χ^2. That is, the first χ^2 is obtained by finding the difference of A - B, the second χ^2 is by finding A + B - 2C, and the third χ^2 is by finding A + B + C - 3D, and thus χ^2_I is the sum of them.

If more than four factors are involved in the treatment with no combination, the method is the extension of the previous ones.

5.5 Randomized Complete Block Design with Two Factors and With Each Cell Containing More Than One Observation.

If the randomized complete block design includes two factors, the first factor has r levels, and the second factor has c levels. Then there are rc treatment combinations. Each treatment combination is repeated in b plots, and each plot contains n_{ijk} observations. Then, by using the binomial transformation, a 2 x rcb frequency contingency table can be obtained as the following table.

Table 5.5.1

2 x rcb Contingency Table with 'a' and 'b' Means Above and Below or Equal to the Median, Md.

	111	112	...	ijk	...	rcb	Total
a	af_{111}	af_{112}	...	af_{ijk}	...	af_{rcb}	n_a
b	bf_{111}	bf_{112}	...	bf_{ijk}	...	bf_{rcb}	n_b
Total	n_{111}	n_{112}	...	n_{ijk}	...	n_{rcb}	n

af_{ijk} is the number of observations in the ijk^{th} cell which are greater than Md.

bf_{ijk} is the number of observations in the ijk^{th} cell which are less than or equal to Md.

From this table we can compute the total chi-square to test the hypothesis that the main effects and interaction effects make no difference in the population distribution of the data. This statistic can be expressed as

$$\chi_T^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{b} \left[\frac{\left(af_{ijk} - \frac{n_{ijk}n_a}{n}\right)^2}{\frac{n_{ijk}n_a}{n}} + \frac{\left(bf_{ijk} - \frac{n_{ijk}n_b}{n}\right)^2}{\frac{n_{ijk}n_b}{n}} \right] \quad (5.5.1)$$

with (rcb - 1) degrees of freedom, where $n_{ijk} = af_{ijk} + bf_{ijk}$.

Chi-squares for three main effects, namely the two factor effects and the block effect, are computed in the same manner.

$$\chi_R^2 = \sum_{i=1}^{r}\left[\frac{\left(af_{i..} - \frac{n_{i..}n_a}{n}\right)^2}{\frac{n_{i..}n_a}{n}} + \frac{\left(bf_{i..} - \frac{n_{i..}n_b}{n}\right)^2}{\frac{n_{i..}n_b}{n}}\right] \tag{5.5.2}$$

$$\chi_C^2 = \sum_{j=1}^{c}\left[\frac{\left(af_{.j.} - \frac{n_{.j.}n_a}{n}\right)^2}{\frac{n_{.j.}n_a}{n}} + \frac{\left(bf_{.j.} - \frac{n_{.j.}n_b}{n}\right)^2}{\frac{n_{.j.}n_b}{n}}\right] \tag{5.5.3}$$

$$\chi_B^2 = \sum_{k=1}^{b}\left[\frac{\left(af_{..k} - \frac{n_{..k}n_a}{n}\right)^2}{\frac{n_{..k}n_a}{n}} + \frac{\left(bf_{..k} - \frac{n_{..k}n_b}{n}\right)^2}{\frac{n_{..k}n_b}{n}}\right] \tag{5.5.4}$$

where $n_{..k} = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ijk}$

$$bf_{..k} = \sum_{i=1}^{r}\sum_{j=1}^{c} bf_{ijk}\ .$$

These three Chi-squares, χ_R^2, χ_C^2, and χ_B^2 are distributed as χ^2 random variables with $(r - 1)$, $(c - 1)$, and $(b - 1)$ degrees of freedom.

The hypothesis tested is that the population means of different levels for all three main factors are identical.

The total interaction χ^2 can be computed by subtracting from χ_T^2 .

$$\chi_I^2 = \chi_T^2 - \chi_R^2 - \chi_C^2 - \chi_B^2\ . \tag{5.5.5}$$

This statistic is distributed approximately as χ^2 - distribution with $rcb - r - c - b + 2$ degrees of freedom

If χ_I^2 is significant, then we may make 2 x b x c, 2 x r x b, and

2 x r x c contingency tables across rows, columns, and blocks respectively. For each of these tables we can compute a χ^2_T as $RC\chi^2_T$, $RB\chi^2_T$, and $CB\chi^2_T$, so that the interactions for each pair of two main factors are

$$RC\chi^2_I = RC\chi^2_T - \chi^2_R - \chi^2_C \tag{5.5.6}$$

$$RB\chi^2_I = RB\chi^2_T - \chi^2_R - \chi^2_C \tag{5.5.7}$$

$$CB\chi^2_I = CB\chi^2_T - \chi^2_R - \chi^2_C \; . \tag{5.5.8}$$

These three statistics are distributed approximately as χ^2 distribution with $(r - 1)(c - 1)$, $(r - 1)(b - 1)$, and $(c - 1)(b - 1)$ degrees of freedom respectively.

Finally, the triple interaction χ^2 of row, column, and block is expressed as

$$\begin{aligned} RBC\chi^2_I &= \chi^2_T - \chi^2_R - \chi^2_C - \chi^2_B - RC\chi^2_I - RB\chi^2_I - CB\chi^2_I \\ &= \chi^2_I - RC\chi^2_I - RB\chi^2_I - CB\chi^2_I \end{aligned} \tag{5.5.9}$$

which is approximately distributed as a χ^2 random variable with $(r - 1)(c - 1)(b - 1)$ degrees of freedom.

To test the significance of all the χ^2 statistics of the main effects and interactions above, we may compare the observed χ^2-values with the conventional χ^2 table with the corresponding degrees of freedom.

6. The Expected Frequencies

6.1 Two-Way Classification

6.1.1 'i' and 'j' are Both 'variates'.

In the two way classification, if we suppose that the row and column are referred to as treatment and block respectively, the expected frequencies can be obtained under the hypothesis. If we let p_{ij} be the probability that an individual selected at random from the population is a member of all observations in the i^{th} row and j^{th} column of the r x c contingency table, and let $p_{i.}$ be the probability that an individual is a member of the i^{th} row, and also let $p_{.j}$ be the probability that an individual is a member of the j^{th} column (In this case n is fixed from sample to sample), then an r x c probability table is indicated as the following table, Table 6.1.1, which is formed from Table 2.2.2,

Table 6.1.1

r x c Probability Table

Row	Column						Total
	1	2	...	j	...	c	
1	p_{11}	p_{12}	...	p_{1j}	...	p_{1c}	$p_{1.}$
2	p_{21}	p_{22}	...	p_{2j}	...	p_{2c}	$p_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮
i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{ic}	$p_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮
r	p_{r1}	p_{r2}	...	p_{rj}	...	p_{rc}	$p_{r.}$
Total	$p_{.1}$	$p_{.2}$	...	$p_{.j}$	...	$p_{.c}$	1

where $$p_{ij} = \frac{n_{ij}}{n} = \frac{af_{ij} + bf_{ij}}{n} .$$

The hypothesis that the row and column or two attributes are independent may be written in the form

$$H_o: p_{ij} = p_{i.}p_{.j}$$

against $H_a: p_{ij} \neq p_{i.}p_{.j}$ $(i = 1,2, \ldots, r$ and $j = 1,2, \ldots, c)$. from some i and j,

If a sample of size n is selected and n_{ij} individuals of them are in the cell of the i^{th} row and j^{th} column, then the chi-square is conventionally computed as

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - np_{ij})^2}{np_{ij}} \tag{6.1.1.1}$$

with $(r - 1)(c - 1)$ degrees of freedom. Under the hypothesis, this expression may be written as

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - np_{i.}p_{.j})^2}{np_{i.}p_{.j}} \ . \tag{6.1.1.2}$$

Since the $p_{i.}$ and $p_{.j}$ are unknown, it is necessary to estimate them from the sample.

By the property of χ^2, the χ^2-test can be used if the estimates are maximum likelihood estimates, with one degree of freedom for each parameter estimated. Since $\sum_{i=1}^{r} p_{i.} = 1$ and $\sum_{j=1}^{c} p_{.j} = 1$, there are $r - 1 + c - 1 = r + c - 2$ parameters to be estimated; hence the proper number of degrees of freedom for testing the independence of two attributes in the $r \times c$ contingency table is $df = rc - 1 - (r + c - 2) = (r - 1)(c - 1)$.

To find the maximum likelihood estimates of the $p_{i.}$ and $p_{.j}$ we let $n_{i.}$ denote the sum of the frequencies in the i^{th} row and let $n_{.j}$ denote the sum of the frequencies in the j^{th} column. Since the frequencies n_{ij}

are discrete, the likelihood function of the sample is the probability of obtaining the sample in the order occured. Thus, using the same reasoning as that used to arrive at $p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}$, the likelihood function of the sample will be given by

$$L = \frac{n!}{\prod\limits_{i,j} n_{ij}!} \prod_{i=1}^{r} \prod_{j=1}^{c} p_{ij}^{n_{ij}} . \tag{6.1.1.3}$$

But because of H_o: $p_{ij} = p_{i.}p_{.j}$ and the definition of $n_{i.}$ and $n_{.j}$, this likelihood function reduces to

$$L = \frac{n!}{\prod\limits_{i,j} n_{ij}!} \prod_{i=1}^{r} \prod_{j=1}^{c} (p_{i.}p_{.j})^{n_{ij}}$$

$$= \frac{n!}{\prod\limits_{i,j} n_{ij}!} \prod_{i=1}^{r} \prod_{j=1}^{c} p_{i.}^{n_{ij}} \prod_{i=1}^{r} \prod_{j=1}^{c} p_{.j}^{n_{ij}}$$

$$= \frac{n!}{\prod\limits_{i,j} n_{ij}!} \prod_{i=1}^{r} p_{i.}^{\sum_{j=1}^{c} n_{ij}} \prod_{j=1}^{c} p_{.j}^{\sum_{i=1}^{r} n_{ij}}$$

$$= \frac{n!}{\prod\limits_{i,j} n_{ij}!} \prod_{i=1}^{r} p_{i.}^{n_{i.}} \prod_{j=1}^{c} p_{.j}^{n_{.j}} . \tag{6.1.1.3}$$

Now, let $p_{r.} = 1 - \sum\limits_{i=1}^{r-1} p_{i.}$, then

$$L = \frac{!}{\prod\limits_{i,j} n_{ij}!} (1 - \sum_{1}^{r-1} p_{i.})^{n_{r.}} \prod_{i=1}^{r-1} p_{i.}^{n_{i.}} \prod_{j=1}^{c} p_{.j}^{n_{.j}} \tag{6.1.1.4}$$

and $$\log L = n_{r.} \log(1 - \sum_{1}^{r-1} p_{i.}) + \sum_{i=1}^{r-1} n_{i.} \log p_{i.} + K \qquad (6.1.1.5)$$

where K does not involve the variable $p_{i.}$. Now, differentiating with respect to $p_{i.}$ and setting the derivative equal to zero to find a maximum,

$$\frac{\partial \log L}{\partial p_{i.}} = - \frac{n_{r.}}{1 - \sum_{1}^{r-1} p_{i.}} + \frac{n_{i.}}{p_{i.}} = 0 . \qquad (6.1.1.6)$$

Since $1 - \sum_{i=1}^{r-1} p_{i.} = p_{r.}$, this equation is equivalent to

$$p_{i.} = \frac{p_{r.}}{n_{r.}} n_{i.} = \lambda n_{i.} \qquad (6.1.1.7)$$

where λ does not depend upon the index i. Since this must hold for $i = 1, 2, \ldots, r$ and since

$$1 = \sum_{1}^{r} p_{i.} = \lambda \sum_{1}^{r} n_{i.} = \lambda n, \qquad (6.1.1.8)$$

it follows that $\lambda = 1/n$, and hence that the maximum likelihood estimate of $p_{i.}$ is

$$\tilde{p}_{i.} = \frac{n_{i.}}{n} . \qquad (6.1.1.9)$$

By symmetry, the maximum likelihood estimate of $p_{.j}$ is

$$\tilde{p}_{.j} = \frac{n_{.j}}{n} . \qquad (6.1.1.10)$$

If $p_{i.}$ and $p_{.j}$ in the formula (6.1.1.2) are replaced by their maximum likelihood estimates, the χ^2 will become

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2}{\frac{n_{i.} n_{.j}}{n}} \qquad (6.1.1.11)$$

with $(r - 1)(c - 1)$ degrees of freedom, but we should notice that this statistic is distributed as a χ^2 distribution provided that n is sufficiently large and H_o is true.

6.1.2 'i' is a 'Way of Classification' and 'j' is a 'Variate'.

If we consider the row a way of classification, then the r x c probability table can be changed, so that $\sum_{j=1}^{c} p_{ij} = p_{i.} = 1$ and $n_{i.}$ is fixed.

So for such a row the likelihood function is

$$\frac{n_{i.}!}{\prod_{j=1}^{c} n_{ij}!} \prod_{j=1}^{c} p_{ij}^{n_{ij}} . \qquad (6.1.2.1)$$

Now we have r independent sets of sizes $n_{1.}, n_{2.}, \ldots, n_{r.}$ of independent observations such that $n_{i.}$ $(i = 1, 2, \ldots, r)$ is fixed from sample to sample. Under the hypothesis that p_{ij} for any column, is independent of row, or in other words,

$$H_o: p_{ij} = q_{.j} \text{ (say)}$$

against $H_a \neq H_o$,

where $q_{.j}$'s are arbitrary positive parameters such that

$\sum_{j=1}^{c} q_{.j} = \sum_{j=1}^{c} p_{ij} = p_{i.} = 1$, we have, therefore,

$$L = \prod_{i=1}^{r} \left[\frac{n_{i.}!}{\prod_{j=1}^{c} n_{ij}!} \prod_{j=1}^{c} p_{ij}^{n_{ij}} \right]$$

$$= \prod_{i=1}^{r} \frac{n_{i.}!}{\prod_{j=1}^{c} n_{ij}!} \prod_{j=1}^{c} p_{ij}^{\sum_{i=1}^{r} n_{ij}}$$

$$= \frac{\prod_{i=1}^{r} n_{i.}!}{\prod_{i,j}^{c} n_{ij}!} \prod_{j=1}^{c} q_{.j}^{n_{.j}} . \qquad (6.1.2.2)$$

Maximizing log L with respect to $q_{.j}$'s subject to $\sum_{j=1}^{c} q_{.j} = 1$ we obtain the maximum likelihood solutions: $q_{.j} = \frac{n_{.j}}{n}$. The number of independent parameters estimated from the data is c-1, and hence the test here is to be based on a statistic which has the χ^2-distribution with degrees of freedom $r(c - 1) - (c - 1) = (r - 1)(c - 1)$ and whose form is

$$\chi^2 = \sum_{i}^{r} \sum_{j}^{c} \frac{(n_{ij} - n_{i.} \frac{n_{.j}}{n})^2}{\frac{n_{i.} n_{.j}}{n}} . \qquad (6.1.2.3)$$

The result of the case of 'i' being variate and 'j' a way of classification may be obtained as the same manner as that above.

6.1.3 'i' and 'j' are Both 'Ways of Classification'.

The row and column of the contingency table are both ways of classification. If we suppose $n_{i.}$ and $n_{.j}$ in the r x c contingency table are both fixed from sample to sample, then both row and column marginal probabilities are all equal to 1, that is

$$\sum_{i=1}^{r} p_{ij} = \sum_{j=1}^{c} p_{ij} = 1 \quad \text{or } p_{i.} = p_{.j} = 1 . \qquad (6.1.3.1)$$

In this case the chi-square will be

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}} \qquad (6.1.3.2)$$

with $rc - (r + c - 1) = (r - 1)(c - 1)$ degrees of freedom.

6.2 Three-way Classification.

6.2.1 'i', 'j' and 'k' Are all 'Variates'.

Suppose we have a sample of independent observations such that p_{ijk} is the probability of an observation in the $(ijk)^{th}$ cell and n is fixed from sample to sample and if we let

$$\sum_{i=1}^{r} p_{ijk} = p_{.jk}, \quad \sum_{j=1}^{c} p_{ijk} = p_{i.k}, \quad \sum_{k=1}^{b} p_{ijk} = p_{ij.}$$

$$\sum_{i,j}^{r\ c} \sum p_{ijk} = p_{..k}, \quad \sum_{i,k}^{r\ b} \sum p_{ijk} = p_{.j.}, \quad \sum_{j,k}^{c\ b} \sum p_{ijk} = p_{i..} \qquad (6.2.1.1)$$

$$\sum_{i,j,k}^{r\ c\ b} \sum \sum p_{ijk} = p_{...} = 1$$

then the likelihood function is given by

$$L = \frac{n!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{\,n_{ijk}} \qquad (6.2.1.2)$$

under the hypothesis of independence between 'i' and 'j' for fixed 'k'.

$$H_o: \frac{p_{ijk}}{p_{..k}} = \frac{p_{i.k}\,p_{.jk}}{p_{..k}\,p_{..k}}$$

or

$$H_o: p_{ijk} = \frac{p_{i.k}\,p_{.jk}}{p_{..k}} ,$$

against $H_a \neq H_o$ $(i = 1,\ldots,r;\ \ j = 1,\ldots,c;\ \ k = 1,\ldots,b)$.

We then have

$$L \propto \prod_{ijk} \left(\frac{p_{i.k}p_{.jk}}{p_{..k}}\right)^{n_{ijk}} . \tag{6.2.1.3}$$

Maximizing log L with respect to the $p_{i.k}$'s, $p_{.jk}$'s and $p_{..k}$'s

subject to $\sum_{i=1}^{r} p_{i.k} = \sum_{j=1}^{c} p_{.jk} = p_{..k}$ and $\sum_{k=1}^{b} p_{..k} = 1$,

gives maximum-likelihood solutions

$$\tilde{p}_{i.k} = \frac{n_{i.k}}{n}$$

$$\tilde{p}_{.jk} = \frac{n_{.jk}}{n} \tag{6.2.1.4}$$

$$\tilde{p}_{..k} = \frac{n_{..k}}{n} .$$

The number of these estimated parameters is $(r - 1)b + (c - 1)b + (b - 1)$. The χ^2 used to test the hypothesis here is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{b} \frac{\left(n_{ijk} - \frac{n_{i.k}\,n_{.jk}}{n_{..k}}\right)^2}{\frac{n_{i.k}\,n_{.jk}}{n_{..k}}} \tag{6.2.1.5}$$

with $rcb - 1 - b(r - 1) - b(c - 1) - (b - 1) = b(r - 1)(c - 1)$ degrees of freedom.

Under the hypothesis that $p_{i.k} = p_{i..}p_{..k}$ and $p_{.jk} = p_{.j.}p_{..k}$ we can test the independence of 'i' and 'k' and 'j' and 'k'. Also if we let $p_{ijk} = p_{i..}p_{.j.}p_{..k}$, then we have

$$L = \prod_{ijk} (p_{i..}p_{.j.}p_{..k})^{n_{ijk}} . \tag{6.2.1.6}$$

To test the hypothesis we maximize log L with respect to

$p_{i..}$'s, $p_{.j.}$'s and $p_{..k}$'s

subject to

$$\sum_{i=1}^{r} p_{i..} = \sum_{j=1}^{c} p_{.j.} = \sum_{k=1}^{b} p_{..k} = 1$$

and obtain the solutions of maximum likelihood as:

$$\tilde{p}_{i..} = \frac{n_{i..}}{n}$$

$$\tilde{p}_{.j.} = \frac{n_{.j.}}{n} \tag{6.2.1.7}$$

$$\tilde{p}_{..k} = \frac{n_{..k}}{n} .$$

The number of independent parameters estimated from the data is $(r + c + b - 3)$, and hence the χ^2 used to test the hypothesis here will be

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{b} \frac{\left(n_{ijk} - \frac{n_{i..}n_{.j.}n_{..k}}{n^2}\right)^2}{\frac{n_{i..}n_{.j.}n_{..k}}{n^2}} \tag{6.2.1.8}$$

with $rcb - 1 - (r + c + b - 3) = rcb - r - c - b + 2$ degrees of freedom.

In order to test the hypothesis that the independence between '(i,j)' and 'k' or

H_o: $p_{ijk} = p_{ij.} p_{..k}$

against $H_a \neq H_o$

$(i = 1,2,\ldots,r, \quad j = 1,2,\ldots,c, \quad k = 1,2,\ldots,b)$, we have

$$L \propto \prod_{i,j,k} (p_{ij.}\, p_{..k})^{n_{ijk}} \tag{6.2.1.9}$$

To test this hypothesis we maximize log L with respect to $p_{ij.}$'s and $p_{..k}$'s subject to $\sum_{i=1}^{r} \sum_{j=1}^{c} p_{ij.} = \sum_{k=1}^{b} p_{..k} = 1$ and obtain the maximum likelihood solutions as

$$\tilde{p}_{ij.} = \frac{n_{ij.}}{n}$$

$$\tilde{p}_{..k} = \frac{n_{..k}}{n} \, . \tag{6.2.1.10}$$

The number of independent parameters estimated from the data is $(rc - 1) + (b - 1)$ and hence the χ^2 used to test the hypothesis here will be

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{b} \frac{\left(n_{ijk} - \frac{n_{ij.}\, n_{..k}}{n}\right)^2}{\frac{n_{ij.}\, n_{..k}}{n}} \tag{6.2.1.11}$$

with $rcb - 1 - [(rc - 1) + (b - 1)] = (rc - 1)(b - 1)$ degrees of freedom.

The hypothesis of independence between 'i' and 'k' and between 'j' and 'k' is included under the hypothesis

H_o: $p_{i.k} = p_{i..}\, p_{..k}$ and $p_{.jk} = p_{.j.}\, p_{..k}$, and this implies the independence between 'ij' and 'k', and 'j' and 'k', as has been shown by Roy & Kastenbaum (1956).

6.2.2 'i' and 'j' are 'Variates' and 'k' is a 'Way of Classification'.

Suppose there are b of sizes $n_{..1},\ldots,n_{..b}$ of independent observations such that $n_{..k}$ $(k = 1,\ldots,b)$ is fixed from sample to sample and p_{ijk} is the probability of an observation in the $(ijk)^{th}$ cell, and $\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ijk} = p_{..k} = 1$. The likelihood function is given by

$$L = \prod_{k=1}\left[\frac{n_{..k}!}{\prod_{ij} n_{ijk}!}\prod_{ij} p_{ijk}^{\,n_{ijk}}\right]. \qquad (6.2.2.1)$$

Under the hypothesis of independence between 'i' and 'j' for each 'k' that is

$$H_o:\ p_{ijk} = p_{i.k}p_{.jk}$$

against $H_a \neq H_o$

$(i = 1,2,\ldots,r;\ j = 1,2,\ldots,c;\ k = 1,2,\ldots,b)$ we have

$$L \propto \prod_{ijk}(p_{i.k}p_{.jk})^{n_{ijk}}. \qquad (6.2.2.2)$$

We maximize log L with respect to the $p_{i.k}$'s and $p_{.jk}$'s subject to $\sum_{i=1}^{r} p_{i.k} = \sum_{j=1}^{c} p_{.jk} = p_{..k} = 1$, and obtain the maximum likelihood solutions

$$\tilde{p}_{i.k} = \frac{n_{i.k}}{n_{..k}},\quad \tilde{p}_{.jk} = \frac{n_{.jk}}{n_{..k}}.$$

The number of independent parameters estimated from the data is

$b(r - 1) + b(c - 1)$ and hence the χ^2 used to test the hypothesis here is

$$\chi^2 = \sum_{k=1}^{b} \left[\frac{\sum_{i=1}^{r} \sum_{j=1}^{c} (n_{ijk} - n_{..k} \frac{n_{i.k} n_{.jk}}{n_{..k}^2})^2}{(n_{..k} \frac{n_{i.k} n_{.jk}}{n_{..k}^2})} \right] \tag{6.2.2.3}$$

with $b(rc - 1) - b(r - 1) - b(c - 1) = b(r - 1)(c - 1)$ degrees of freedom.

For the hypothesis p_{ijk} independent of 'k', or

H_o: $p_{ijk} = q_{ij.}$ (say)

against $H_a \neq H_o$ (for all i,j and k), we have

$$L \propto \prod_{ijk} q_{ij.}^{n_{ijk}} \tag{6.2.2.4}$$

We maximize log L with respect to the $q_{ij.}$'s subject to $\sum_{ij} q_{ij.} = 1$, and obtain the maximum likelihood solutions:

$$\tilde{q}_{ij.} = \frac{n_{ij.}}{n} \quad . \tag{6.2.2.5}$$

The number of independent parameters to be estimated from the data is $(rc - 1)$ and hence the statistic χ^2 is

$$\chi^2 = \sum_{k=1}^{b} \left[\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ijk} - n_{..k} \frac{n_{ij.}}{n})^2}{(n_{..k} \frac{n_{ij.}}{n})} \right] \tag{6.2.2.6}$$

with $b(rc - 1) - (rc - 1) = (rc - 1)(b - 1)$ degrees of freedom.

6.2.3 'i' is a 'Variate' and 'j' and 'k' are 'Ways of Classification'.

Consider c x b independent sets of sizes $n_{.jk}$ of independent observations, such that $n_{.jk}$ ($j = 1,2,\ldots,c$, $k = 1,2,\ldots,b$) is fixed from sample to sample and p_{ijk} is the probability of an observation in the $(ijk)^{th}$ cell, and $\sum_{i=1}^{r} p_{ijk} = p_{.jk} = 1$. The likelihood function is

$$L = \prod_{j,k} \left[\frac{n_{.jk}!}{\prod_i n_{ijk}} \prod_i p_{ijk}^{\,n_{ijk}} \right] \tag{6.2.3.1}$$

under the hypothesis, that for any 'k', p_{ijk} is independent of 'j', that is

$$H_o:\ p_{ijk} = q_{i.k} \text{ (say)}$$

against $H_a \neq H_o$ (for all i,j and k), and

$$\sum_{i=1}^{r} q_{i.k} = \sum_{i=1}^{r} p_{ijk} = p_{.jk} = 1, \text{ we have}$$

$$L \propto \prod_{i,j,k} q_{i.k}^{\,n_{ijk}} \tag{6.2.3.2}$$

we maximize log L with respect to the $q_{i.k}$'s subject to $\sum_{i=1}^{r} q_{i.k} = 1$, and obtain the maximum likelihood solutions as

$$\tilde{q}_{i.k} = \frac{n_{i.k}}{n_{..k}} . \tag{6.2.3.3}$$

The number of independent parameters to be estimated from the data is $b(r - 1)$ and hence the statistic χ^2 used to test the hypothesis is

$$\lambda^2 = \sum_{j=1}^{c} \sum_{k=1}^{b} \left[\sum_{i=1}^{r} \frac{(n_{ijk} - n_{.jk} \frac{n_{i.k}}{n_{..k}})^2}{(n_{.jk} \frac{n_{i.k}}{n_{..k}})} \right] \qquad (6.2.3.4)$$

with $cb(r - 1) - b(r - 1) = b(r - 1)(c - 1)$ degrees of freedom.

Again for the hypothesis that for any 'j', p_{ijk} is independent of 'k' that is

$$H_o: \ p_{ijk} = q_{ij.} \ \text{(say)}$$

against $H_a \neq H_o$ (for all i, j, and k) where $\sum_{i=1}^{r} q_{ij.} = \sum_{i=1}^{r} p_{ijk} = p_{.ij} = 1.$

As mentioned above, the hypotheses

$$H_o: \ p_{ijk} = q_{i.k} \ \text{(say)}$$

together with

$$H_o: \ p_{ijk} = q_{ij.} \ \text{(say)},$$

implie that p_{ijk} is a pure function of 'i', i.e. that

$$p_{ijk} = q_{i..} \ \text{(say) (for all i, j and k)}.$$

If, in a one way classification in the usual analysis of variance, 'i' corresponds to the 'variate', 'j' to the 'concomitant variate' and 'k' to the 'way of classification', then it will be seen on a little reflection that

$$H_o: \ p_{ijk} = p_{i.k} \, p_{.jk}$$

against $H_a \neq H_o$ $(i = 1,\dots,r; \ j = 1,\dots,c; \ k=1,\dots,b)$

will be the analogue of the hypothesis of no regression, and

$$H_o: \; p_{ijk} = q_{i..} \text{ (say)},$$

against $H_a \neq H_o$ (for all i,j and k)

will be the analogue of the hypothesis of no covariance.

On the other hand, suppose we take 'j' and 'k' as just the two way classification, for example, if we take 'j' as, say, blocks and 'k' as, say, treatments in a randomized complete block experiment (with more than one and in general unequal number of replications in each cell). Then

$$H_o: \; p_{ijk} = q_{i.k} \text{ (say)}$$

against $H_a \neq H_o$ (for all i,j and k)

will be the analogue of no block effect for each treatment separately and

$$H_o: \; p_{ijk} = q_{ij.} \text{ (say)},$$

against $H_a \neq H_o$ (for all i,j and k)

will be the analogue of 'no treatment effect' for each block separately.

In other words, in the usual parlance of analysis of variance,

$$H_o: \; p_{ijk} = q_{i.k} \text{ (say)},$$

against $H_a \neq H_o$ (for all i, j and k)

combines the hypothesis of 'no main effect' and 'no interaction', while

$$H_o: \; p_{ijk} = q_{ij.} \text{ (say)},$$

against $H_a \neq H_o$ (for all i,j and k)

combines the hypotheses of another 'no main effect' and 'no interaction'.

7. Normal Score Transformation.

R. A. Fisher (1943) designed a normal score transformation for the ranked data. Many characteristics of various objects can not be measured numerically, but they can be ranked in an orderly sequence, such kinds of data as in judging ice cream, bread, cake, candy, chocolate, all food tests, tea and coffee tests, and furthermore tests for clothing, sports, cars, courses, etc. We may not express our preference in a quantitative measure, but we can rank the different flavors, as 1, 2, 3 and so on. For this ranked data we can replace each rank by a normal score which can be found in the statistical table for Biological Agricultural and Medical Research of Fisher and Yates (1943). This table gives the average deviate of the r^{th} largest of samples of n observations drawn from a normal distribution which has a unit variance; that is, if $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ is an ordered sample from a standard normal distribution, the table gives $E(X_{(r)})$.

The application of this table is very simple. We now consider an example of the ranked and randomized complete block design. Four flavors of ice cream were evaluated by 10 judges. Each judge ranked the flavors, 1,2,3, or 4 with 1 being the most preferred, and with the results in the following Table 7.1.

Table 7.1

Ranked Data for Testing the Flavors of Four Ice Creams

Judge	Flavor			
	A	B	C	D
1	2	1	4	3
2	1	2	3	4
3	2	1	4	3
4	3	2	4	1
5	2	1	4	3
6	2	3	4	1
7	1	2	3	4
8	2	1	4	3
9	2	1	4	3
10	3	1	2	4

After we transform the ranks in the table into normal scores we may have the new two-way Table 7.2.

Table 7.2

The Normal Score Transformed Data from Table 7.1

Judge	Flavor				Total
	A	B	C	D	
1	0.30	1.03	-1.03	-0.30	0
2	1.03	0.30	-0.30	-1.03	0
3	0.30	1.03	-1.03	-0.30	0
4	-0.30	0.30	-1.03	1.03	0
5	0.30	1.03	-1.03	-0.30	0
6	0.30	-0.30	-1.03	1.03	0
7	1.03	0.30	-0.30	-1.03	0
8	0.30	1.03	-1.03	-0.30	0
9	0.30	1.03	-1.03	-0.30	0
10	-0.30	1.03	0.30	-1.03	0
Total	3.260	6.780	-7.510	-2.530	0

For this data we can consider the judges as blocks and flavors as treatments in a randomized complete block design and then do the conventional analysis of variance. The results obtained are shown in the following table 7.3.

Table 7.3

Analysis of Variance Table for Testing the Flavors of Four Ice creams

Source of Variation	DF	Sum of Squares	Mean Square	F-Value
Treatment	3	11.9397	3.9799	9.6990**
Error	27	11.0783	0.4103	
Total	30	23.0179		

And also if we use a 5% significant level, the multiple range test results are as follows.

Treatment	Mean
C	-0.7510
D	-0.2530
A	0.3260
B	0.6280

Here we should note that since the block totals are zero, we are not able to find differences among blocks. The block degrees of freedom should be subtracted from that of the total. The normal score transformation may apply not only on ranked data but also on quantitative data, and second numerical example shows the analysis of variance for the normal score transformed randomized complete block data. The data includes 5 treatments and

10 blocks. The transformed scores and the results of analysis of variance are shown in the following Table 7.5 and 7.6.

Table 7.4

Two-Way Table of Randomized Complete Block Design

Block	Treatment				
	1	2	3	4	5
1	46	50	69	48	44
2	48	46	47	60	40
3	32	50	46	54	59
4	42	48	65	47	44
5	39	37	49	50	55
6	48	58	59	68	50
7	49	50	42	58	47
8	30	44	63	46	71
9	48	40	47	46	43
10	34	39	47	37	55

Table 7.5

Normal Score Transformed Data from Table 7.4

Block	Treatment				
	1	2	3	4	5
1	-0.50	0.50	1.16	0.00	-1.16
2	0.50	-0.50	0.00	1.16	-1.16
3	-1.16	0.00	-0.50	0.50	1.16
4	-1.16	0.50	1.16	0.00	-0.50
5	-0.50	-1.16	0.00	0.50	1.16
6	-1.16	0.00	0.50	1.16	-0.50
7	0.00	0.50	-1.16	1.16	-0.50
8	-1.16	-0.50	0.50	0.00	1.16
9	1.16	-1.16	0.50	0.00	-0.50
10	-1.16	0.00	0.50	-0.50	1.16
Total	-5.14	-1.82	2.66	3.98	0.32

Table 7.6

Analysis of Variance Table for the Data in Table 7.5

Source of Variation	DF	Sum of Square	Mean of Square	F-value
Treatment	4	5.27540	1.3188	1.78
Error	36	26.63696	0.7399	
Total	40	31.91200		

Also, the grand total is equal to zero and all the block totals are equal to zero, so the component of blocks is completely eliminated. The total sum of squares is just $\sum_{i=1}^{t} \sum_{j=1}^{b} y^2$. Also the number of degrees of freedom for the total sum of square is reduced, because the component of blocks is eliminated.

In using the normal score transformation, ties are permitted. If two ranks or observations in the same block are identical, the average of the corresponding normal scores is used.

Furthermore, for the randomized complete block design, this transformation can be extended to two factors or more than two factorial experiments. In this case, each of the treatments can be divided into several levels. Then the experiment becomes the factorial type. After the transformation is made for these kinds of experiment as above, then the conventional analysis of variance or even regression can also be used.

For food test experiments, because it is not easy to rank more than 4 products effectively at a time, this method is limited. Fisher's normal score table can be applied for up to 50 treatments.

8. χ^2-Test and F-Test for Ordinary Binomial Population.

For convenience, we describe the relationship between the χ^2-test and the F-test in the complete randomized design case, when examining data from binomial populations, as in the test for equal medians.

As we know, the means of sample size n drawn from an ordinary binomial population with p and q which are not necessarily equal follow approximately the normal distribution with the population mean equal to p and variance equal to $p(1 - p)/n$. Then the sample means, $\bar{y}_i$'s may be considered a sample of t (number of treatment) observations drawn from a normal population with mean equal to p and variance equal to $p(1 - p)/n$. From this and by definition, the χ^2-statistic is given by

$$\chi^2 = \frac{\sum_{i=1}^{t} (\bar{y}_i - \bar{\bar{y}})^2}{\frac{p(1-p)}{n}}$$

$$= \frac{n \sum_{i=1}^{t} (\bar{y}_i - \bar{\bar{y}})^2}{p(1 - p)} \tag{8.1}$$

$$= \frac{\text{Among sample SS}}{p(1 - p)}$$

where $\bar{\bar{y}}$ is the mean of $\bar{y}_i$. This χ^2 will follow approximately the χ^2 distribution with $(t - 1)$ degrees of freedom. Since the variance of a binomial population is equal to $p(1 - p)$, the $\bar{\bar{y}}(1 - \bar{\bar{y}})$ may be used as pooled estimate of $p(1 - p)$, and then

$$\chi^2 = \frac{\sum_{i=1}^{t} n(\bar{y}_i - \bar{\bar{y}})^2}{\bar{\bar{y}}(1 - \bar{\bar{y}})} = \frac{\text{Among sample SS}}{\bar{\bar{y}}(1 - \bar{\bar{y}})} \tag{8.2}$$

is distributed approximately as a chi-square random variable with $(t - 1)$ degrees of freedom.

By hypothesis, if p is specified, then we can use pq to estimate $p(1 - p)$.

For the F-statistic we commonly use

$$F = \frac{\dfrac{\text{Among sample SS}}{t - 1}}{\dfrac{\text{Within sample SS}}{\sum n - t}} = \frac{\text{Among sample MS}}{\text{Within sample MS}} \tag{8.3}$$

with $(t - 1)$ and $(\sum n - t)$ degrees of freedom. This means that the two statistics χ^2 and F, are similar, because the χ^2 may be expressed in a way that resembles an F statistic,

$$F' = \frac{\chi^2}{t - 1} = \frac{\dfrac{\text{Among sample SS}}{t - 1}}{\bar{\bar{y}}(1 - \bar{\bar{y}})} = \frac{\text{Among sample MS}}{\bar{\bar{y}}(1 - \bar{\bar{y}})} \tag{8.4}$$

with $t - 1$ and ∞ degrees of freedom.

Notice that in this case the within sample mean square is replaced by $\bar{\bar{y}}(1 - \bar{\bar{y}})$. This is the difference between normal and binomial population cases. For normal population, σ^2 is directly estimated by s^2, the error mean squares, and for binomial population $\sigma^2 = p(1 - p)$. So in a basic sense, these two tests χ^2 and F', are similar.

Now, we can consider the term, $\bar{\bar{y}}(1 - \bar{\bar{y}})$, which is the total mean square, because in a binomial population, the observations y's are both 0's and 1's, the grand total is $\sum y = \sum 1 = G$ (say), the total SS is

$\sum y^2 - \frac{G^2}{\sum n} = G - \frac{G^2}{\sum n}$, and the total mean square is approximately equal to

$$\frac{1}{\sum n}\left[G - \frac{G^2}{\sum n}\right] = \frac{G}{\sum n} - \left(\frac{G}{\sum n}\right)^2$$

$$= \bar{\bar{y}} - \bar{\bar{y}}^2 = \bar{\bar{y}}(1 - \bar{\bar{y}}) \tag{8.5}$$

in which we just replaced the total degrees of freedom $\sum n - 1$ by $\sum n$. So, consequently, the total mean square is only slightly greater than $\bar{\bar{y}}(1 - \bar{\bar{y}})$. Furthermore, the total mean square is the weighted average of the among sample and within sample mean squares, with their number of degrees of freedom being the weights.

In our case, we used the pooled median as a cutting point to transform the data into the binomial form, and to test the hypothesis that the t treatment populations have the same median, that is $p = 1 - p = q = 0.5$. Under this case we may replace the term $\bar{\bar{y}}(1 - \bar{\bar{y}})$ by $p(1 - p) = pq = 1/4$.

From the discussion above we see the χ^2-test is equivalent to the analysis of variance, if we use the total mean square as the error term. That is to say, the χ^2-test and the analysis of variance usually yield the same conclusion in testing the hypothesis that t population means are equal.

9. Comments and Discussion.

9.1 Basic Technique.

The basic technique of the non-parametric methods in this report is to make a contingency table based on a pooled median. If the dimensions of the table are 2 x 2 the data may be interpreted as two samples drawn from two binomial populations. If the dimensions are 2 x t (or 2 x r), the data may be interpreted as t samples drawn from t binomial populations, and also as 2 samples drawn from t-attributes multinomial populations. If the dimensions are r x c (or r x c), the data may be either interpreted as r random samples drawn from c attributes multinomial populations or c random samples drawn from r attributes multinomial populations. Either interpretation may yield the same result.

9.2 Application of Mean or Median.

As we know that the normal population is symmetric, the mean and the median are equal. So all of the discussion concerning the mean also pertains to the median. The test of the hypothesis that the t population means are equal is the same test as for t population medians being equal. For the binomial population in this report all the discussion about tests of hypotheses is about the median instead of the mean. The median has an important property; that is, the median is transformable. For example, for the 5 observations 14, 15, 26, 100, 125, the median is 26 and the mean is 56. Suppose we use the square root transformation, then the corresponding transformed values are 3.74, 3.87, 5.10, 10.00, 11.18, where the transformed median is 5.10, which is the square root of the original median 26, but the new mean is 6.78, which is no longer the square root of the original mean 56. For any transformation this is true, so when we use a transformation with the analysis of variance, we are actually making comparisons among the medians on the original scale. In this report for cases in which the population is not normal, the mean and median may not be the same, so the median is used directly for the transformation.

In section 8, we see that the χ^2-test is similar to the F'test. The χ^2-value seems to be larger than F', but the corresponding F-value in the table is also larger than that of F', because the degrees of freedom for the denominator of F' is larger. The χ^2 test seems to have a slightly higher probability of committing a Type II error than has the analysis of variance. However, the F-test is also not beyond reproach, because the populations are binomial and not normal. If the population is not normal, the analysis of variance tends to reject the true hypothesis more frequently than the significance level specified. Therefore, the F-test seems to have a higher probability of committing a Type I error than that of χ^2-test.

9.3 Individual Degree of Freedom.

The individual degree of freedom can be used on any contingency table except that of 2 x 2 in which case the number of degrees of freedom is already equal to 1. The basic technique of the individual degree of freedom is to reduce the dimension of the contingency table to 2 x 2 out of the r x c contingency table. The purpose of the individual degree of freedom is to increase the power of the test.

9.4 Sheffield's Comments.

Sheffield (1957) reinterpreted Wilson's method in a similar manner. He considered that the hypothesis in Wilson's method is that each observation in a cell has 50% chance of falling above the pooled median. If n is the number of observations per cell, then the range of the possible frequencies above the medi n is from 0 to n, and the mean is equal to n/2. The variance of a frequency is npq or n(0.5)(0.5) = n/4, since the hypothesis is that $p = q = 0.5$. He repeated the example with the 3 x 3 factorial experiment including 16 replicates in each cell. The range of observation in each cell is from 0 to 16. The mean of each cell is $16/2 = 8$, and variance of cell is $16/4 = 4$. The obtained frequency table is

Table 9.4.1

The Fictitious 3 x 3 Factorial Experimental Data

Dials	Illumination			Total
	1	2	3	
A	14	12	11	37
B	9	7	8	24
C	6	3	2	11
Total	29	22	21	72

and the analysis of variance is as follows.

Table 9.4.2

Analysis of Variance Table for the Data in Table 9.4.1

Source of Variation	DF	SS	MS	F	P	Wilson's χ^2	P
Dials	2	112.67	56.34	14.08	<0.01	28.168	$<0.1\%$
Illumination	2	12.67	6.34	1.58	>0.05	3.188	10%
Interaction	4	2.67	0.67	0.17	--	0.664	--
Total	8	128.00	16.00	4.00	<0.01		

Sheffield indicated that the F for illumination is not at all significant by the nonparametric test but would be well within the 5% level if tested in the conventional way, and he also mentioned that in a typical 3 x 3 factorial design experiment with only one observation per cell, there is no within cell error because of the lack of replications. The only error term available in such a case is the interaction of the two marginal variables. If the parametric approach or F-test is applied, the F-value for illumination against interaction is $6.34/0.67 = 9.5$, which is well beyond the 6.94 needed at the 5% level for 2 and 4 degrees of freedom. The corresponding nonparametric test ($F = 1.58$) does not even reach the 20% level of confidence.

Sheffield concluded with the comment that Wilson's test involves two parts: first the procedure for creating approximately normal data from the original nonnormal data with cutting by a pooled median; second the procedure for testing obtained variance, npq. Only the second part of the method is the distribution-free part.

9.5 McNemar's Comments.

McNemar (1957) contrasted the results of Wilson's test and the F - test for some data of two-way classification which are published in other textbooks. From the levels of significance reached by way of F - test and Wilson's test, most of them, for row effects, column effects, and for interaction effects, indicated that the probabilities of reaching the significance needed for the F - test is smaller than that of Wilson's test, so the power of Wilson's test is much lower than that of F - test.

ACKNOWLEDGMENT

I am deeply indebted to Dr. W. G. Conover for his advice and guidance in the preparation of this report.

[illegible]

[illegible]. [illegible]. McGraw-Hill [illegible] Co., Inc. New York.

[illegible] (1960). [illegible] – [illegible], WADC [illegible].

[illegible] (1966). The Power of Chi-square Tests for [illegible]. J. of the Amer. Stat. Assoc. 61 965-975.

[illegible]. [illegible] Designs in Industry. John Wiley & Sons, Inc. New York.

Cramer, [illegible] (1946). Mathematical Methods of Statistics. Princeton Univ. Press.

Fisher, R. A. and Yates, F. (1943). Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd Ltd. London.

Fisher, [illegible] (1950). Contribution to Mathematical Statistics. John Wiley & Sons, Inc. New York.

Fisher, R. [illegible] (19[illegible]). The Design of Experiments. Oliver & Boyd, London.

Goodman, L. A. (1964). Simple Methods for Analyzing Three Factors Interaction in Contingency Tables. J. of the Amer. Stat. Assoc. 59 319 -352.

Hoel, P. G. (1954). Introduction to Mathematical Statistics. John Wiley & Sons, Inc. New York.

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

[illegible] (1964). Statistical Inference I. Edwards Brothers, Inc. [illegible].

[illegible] (19[illegible]). Statistical Analysis in Biology. Methuen and Co. Ltd. London.

NONPARAMETRIC STATISTICAL METHODS FOR

THE RANDOMIZED COMPLETE BLOCK DESIGN

by

LI-CHUN TAO

B. S., Taiwan Provincial Chung-hsing University, 1954

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and

Computer Science

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1968

The purpose of this report is to introduce the application of the chi-square test as a nonparametric test on the randomized complete block design. This test is one of the large sample methods. So before we can make use of this method, we should have large samples. The minimum sample size can be obtained from the working rule given in Section 2.

In order to use the chi-square test for the randomized complete block design, we first of all need to change the randomized complete block two way table into a two way contingency table. In other words, we have to transform the continuous data into discrete multinomial data with the median as a cutting point. A multinomial data set is a set of observations which can be classified into r categories. If r = 2 the multinomial data become binomial data. The method for this transformation is called the binomial transformation and is stated in the second section.

In the third section we stated that the test of independence between two attributes in χ^2-test, is comparable to the test of interaction between two attributes in the analysis of variance case.

The fourth and fifth sections deal with the methods to compute various χ^2's concerned with different types of experimental data to test the hypotheses that the treatment population means are the same, in which, of course, the contingency table should be formed at first. In the discussion we started with one observation and then more observations per cell data. An extension of the methods applies to factorial experiments on the randomized complete block design, in which both no combination and combinations among levels of factors are discussed. The various χ^2's are computed to test the hypotheses about the significance of the different main effects and interaction effects.

The sixth section contains the concepts of the expected frequencies of χ^2-test in both two and three way classification. The method of the derivation of the expected frequencies used is the maximum likelihood method.

In the seventh section appears a normal score transformation. This is introduced by Fisher and Yates (1943) and is used for the analysis of the ranked data. If we transform the quantitative data into ranks at first, the numerical data can also be analyzed by this method. After the normal score transformation has been made all the methods used in normal populations can be used in the ranked data.

The last two sections compared the χ^2-test and the F-test, and the situations of using mean and median. The F-test is better for normal populations and the χ^2-test needs larger samples to have the same power as the F-test. Since the normal distribution is symmetrical, the mean and median are tested in normally distributed data while only the median is compared in binomially distributed data.