AN EXACT TEST OF SIGNIFICANCE IN THE BALANCED INCOMPLETE BLOCK DESIGN WITH RECOVERY OF INTER-BLOCK INFORMATION

By

DAVID LEE WEEKS

Bachelor of Science

Oklahoma Agricultural and Mechanical College

Stillwater, Oklahoma

1952

Submitted to the faculty of the Graduate School of
the Oklahoma Agricultural and Mechanical College
in partial fulfillment of the requirements
for the degree of
MASTER OF SCIENCE
May, 1957

AN EXACT TEST OF SIGNIFICANCE IN THE BALANCED INCOMPLETE BLOCK DESIGN

WITH RECOVERY OF INTER-BLOCK INFORMATION

Thesis Approved:

Franklin A. Graybill
Thesis Adviser

R B Deal

Robert MacVicar
Dean of the Graduate School

PREFACE

In the balanced incomplete block design, we are frequently interested in testing the null hypothesis that all the treatment effects are equal. In the conventional analysis, assuming Eisenhart's Model I, this hypothesis is tested using Snededor's "F" statistic, that is forming the ratio of the mean square for treatments (eliminating blocks) and the mean square for the intra-block error.

This thesis shows that when we assume the block effects random variables, i.e., assume Eisenhart's Model III, and the number of blocks is greater than the number of treatments, then there exist two independent tests of the null hypothesis. A method for combining the two tests is also given.

The author is deeply indebted to Dr. Franklin A. Graybill for his assistance in the preparation of this thesis and also for suggesting the problem.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Consider a balanced incomplete block design having the mathematical model (Eisenhart's Model I)

$$y_{ijm} = \mu' + \beta_i + \alpha'_j + e_{ijm}$$

$$i = 1, 2, \ldots, b$$

$$j = 1, 2, \ldots, t$$

$$m = 0, 1, \ldots, n_{ij}$$

where t treatments are applied to b blocks with k plots per block and r replicates per treatment. The following conditions hold:

$$bk = tr$$

$$n_{ij} = 1 \text{ if the j-th treatment occurs in block i}$$

$$= 0 \text{ if the j-th treatment does not occur in block i}$$

$$\sum_i n_{ij} = N_{.j} = r \text{ for all } j$$

$$\sum_j n_{ij} = N_{i.} = k \text{ for all } i$$

$$\sum_i n_{ij}n_{ij'} = \lambda \text{ for all pairs } (jj') \text{ where } j \neq j'.$$

The errors are normally distributed with

$$E\,(e_{ijm}) = 0$$

$$E\,(e_{ijm}e_{pqs}) = \sigma^2 \text{ if } i = p,\ j = q \text{ and } m = s$$

$$= 0 \text{ otherwise.}$$

We will consider in this paper from the outset the reparameterized model

$$y_{ijm} = \mu + \beta_i + \alpha_j + e_{ijm}$$

where

$$\mu = \mu' + \alpha'_{.}$$

$$\alpha_j = \alpha'_j - \alpha'_{.}.$$

This gives

$$\sum_j \alpha_j = 0.$$

We will also be interested in the model (Eisenhart's Model III) where the β_i's are assumed to be normally distributed with

$$E(\beta_i) = 0$$

$$E(\beta_i \beta_p) = \sigma_b^2 \quad \text{if } i = p$$

$$= 0 \quad \text{if } i \neq p$$

$$E(\beta_i e_{pqs}) = 0 \quad \text{for all } i,\ p,\ q \text{ and } s.$$

CHAPTER I

THE GENERAL TWO-WAY CLASSIFICATION

Consider the general two-way classification model with unequal numbers and no interaction, i.e., the Eisenhart Model I:

$$y_{ijm} = \mu + \beta_i + \alpha_j + e_{ijm}$$

$$i = 1, 2, \ldots, b$$

$$j = 1, 2, \ldots, t$$

$$m = 0, 1, \ldots, n_{ij}$$

where

$$\mu = \text{general mean}$$

$$\beta_i = \text{effect of the i-th block}$$

$$\alpha_j = \text{effect of the j-th treatment}$$

$$e_{ijm} = \text{a random variable with the following characteristics}$$

$$E\,(e_{ijm}) = 0$$

$$E\,(e_{ijm}e_{uvp}) = \sigma^2 \quad i = u,\ j = v,\ m = p$$

$$= 0 \quad \text{otherwise.}$$

The object here will be to show by the method of least squares the estimate of $(\alpha_i - \alpha_o)$ considering the β_i as fixed parameters. Proceeding, we have

$$e_{ijm} = y_{ijm} - \mu - \beta_i - \alpha_j$$

and we wish to minimize

$$\sum_i \sum_j \sum_m (e_{ijm})^2 = \sum_i \sum_j \sum_m (y_{ijm} - \mu - \beta_i - \alpha_j)^2$$

with respect to each of the parameters in the model. If we let

$$Z = \sum_i \sum_j \sum_m (e_{ijm})^2$$

we have

$$\frac{\partial Z}{\partial \mu} = -2 \sum_i \sum_j \sum_m (y_{ijm} - \mu - \beta_i - \alpha_j)$$

$$\frac{\partial Z}{\partial \beta_i} = -2 \sum_j \sum_m (y_{ijm} - \mu - \beta_i - \alpha_j)$$

$$\frac{\partial Z}{\partial \alpha_j} = -2 \sum_i \sum_m (y_{ijm} - \mu - \beta_i - \alpha_j)$$

Setting these partial derivatives equal to zero and solving for the estimates of the parameters we obtain the normal equations, namely,

$$N_{\cdot\cdot}\hat{\mu} + \sum_i N_{i\cdot}\hat{\beta}_i + \sum_j N_{\cdot j}\hat{\alpha}_j = Y_{\cdot\cdot\cdot}$$

$$N_{i\cdot}\hat{\mu} + N_{i\cdot}\hat{\beta}_i + \sum_j n_{ij}\hat{\alpha}_j = Y_{i\cdot\cdot}$$

$$N_{\cdot j}\hat{\mu} + \sum_i n_{ij}\hat{\beta}_i + N_{\cdot j}\hat{\alpha}_j = Y_{\cdot j\cdot}$$

where the dot subscript denotes summation over that subscript.

We must now solve the normal equations for the $\hat{a}_j$'s. To accomplish this we will determine each $\hat{\mu} + \hat{\beta}_i$ in terms of the observations. The equations for the β_i are

$$N_{i\cdot}\hat{\mu} + N_{i\cdot}\hat{\beta}_i + \sum_j n_{ij}\hat{\alpha}_j = Y_{i\cdot\cdot}$$

Then when

$i = 1$ $$(\hat{\mu} + \hat{\beta}_i) = \frac{1}{N_{1\cdot}}(Y_{1\cdot\cdot} - \sum_{j'} n_{1j'}\hat{\alpha}_{j'})$$

$$\vdots$$

$i = b$ $$(\hat{\mu} + \hat{\beta}_b) = \frac{1}{N_{b\cdot}}(Y_{b\cdot\cdot} - \sum_{j'} n_{bj'}\hat{\alpha}_{j'})$$

The first α_j equation is (for α_1)

$$Y_{.1.} = N_{.1}\hat{\mu} + \sum_i n_{i1}\hat{\beta}_i + N_{.1}\hat{\alpha}_1$$

Expanding, we have

$$Y_{.1.} = n_{11}(\hat{\mu} + \hat{\beta}_1) + n_{21}(\hat{\mu} + \hat{\beta}_2) + \cdot \quad \cdot \quad \cdot + n_{b1}(\hat{\mu} + \hat{\beta}_b) + N_{.1}\hat{\alpha}_1$$

Substituting in this equation for $(\hat{\mu} + \hat{\beta}_i)$ we obtain

$$n_{11}\left[\frac{1}{N_{1.}}(Y_{1..} - \sum_{j'} n_{1j'}\hat{\alpha}_{j'})\right] + n_{21}\left[\frac{1}{N_{2.}}(Y_{2..} - \sum_{j'} n_{2j'}\hat{\alpha}_{j'})\right] + \cdot \quad \cdot \quad \cdot$$
$$+ n_{b1}\left[\frac{1}{N_{b.}}(Y_{b..} - \sum_{j'} n_{bj'}\hat{\alpha}_{j'})\right] + N_{.1}\hat{\alpha}_1$$

Simplifying, we obtain

$$\begin{aligned} Y_{.1.} &= \sum_i n_{i1}\left[\frac{1}{N_{i.}}(Y_{i..} - \sum_{j'} n_{ij'}\hat{\alpha}_{j'})\right] + N_{.1}\hat{\alpha}_1 \\ &= \sum_i n_{i1}\left[\frac{1}{N_{i.}}(Y_{i..} - \sum_{j' \neq 1} n_{ij'}\hat{\alpha}_{j'} - n_{i1}\hat{\alpha}_1)\right] + N_{.1}\hat{\alpha}_1 \\ &= \sum_i n_{i1}\left[\frac{1}{N_{i.}}(Y_{i..} - \sum_{j' \neq 1} n_{ij'}\hat{\alpha}_{j'})\right] + N_{.1}\hat{\alpha}_1 - \sum_i \frac{n_{i1}^2}{N_{i.}}\hat{\alpha}_1 \end{aligned}$$

This is the first of the α_j equations. We may write the general α_j equation as

$$(1) \qquad (Y_{.j.} - \sum_i \frac{n_{ij}Y_{i..}}{N_{i.}}) = (N_{.j} - \sum_i \frac{n_{ij}^2}{N_{i.}})\hat{\alpha}_j - \sum_{\substack{j' \\ j' \neq j}} (\sum_i \frac{n_{ij}n_{ij'}}{N_{i.}})\hat{\alpha}_{j'}$$

Imposing the conditions of the balanced incomplete block design on equation (1) simplifies it greatly and thereby we are able to obtain estimates of the treatment effects easily as we shall see later.

The quantity to the left of the equality in (1) is usually denoted by Q_j and henceforth will be referred to as such.

CHAPTER II

THE BALANCED INCOMPLETE BLOCK DESIGN

The balanced incomplete block design is defined as a design in which there are t treatments applied to b blocks where there are $k < t$ plots per block and each treatment is replicated r times with any pair of different treatments occuring in all blocks λ times. The conditions enumerated in the introduction now hold.

There are two sources of information used to estimate the α_j effects using the estimation techniques for this design. The purpose here will be to derive a method for obtaining an exact test of significance that the treatment effects are equal in designs where the number of blocks is greater than the number of treatments, i.e., $b > t$, and where the block effects are assumed to be a normally distributed random variable.

We will first use the equation (1) to estimate α_j in the balanced incomplete block design. The incomplete block design imposes the following conditions on the two-way classification:

$$N_{\circ j} = r \quad \text{constant for all } j$$

$$N_{i\circ} = k \quad \text{constant for all } i$$

$$n_{ij}^2 = n_{ij}$$

$$\sum_i n_{ij}n_{ij'} = \lambda \text{ for all } j \neq j'.$$

Imposing the above conditions on equation (1), we obtain

$$(r - \frac{r}{k})\hat{\alpha}_j - \frac{\lambda}{k} \sum_{\substack{j' \\ j \neq j'}} \hat{\alpha}_{j'} = Q_j$$

or

$$(r - \frac{r}{k} + \frac{\lambda}{k})\hat{\alpha}_j - \frac{\lambda}{k}\sum_j \hat{\alpha}_j = Q_j$$

The Q_j becomes

$$Q_j = Y_{.j.} - \frac{\sum_i n_{ij} Y_{i..}}{N_{i.}} = V_j - \frac{T_j}{k}$$

where

V_j = total yield of treatment j

T_j = total of all blocks containing treatment j.

The estimate of the j-th treatment effect is then

$$\hat{\alpha}_j = \frac{Q_j}{(r - \frac{r}{k} + \frac{\lambda}{k})}$$

The additional source of treatment comparisons is given by the block totals when Model III is assumed, i.e., considering the model

$$B_i = k\mu + \sum_j n_{ij}\alpha_j + k\beta_i'$$

where the β_i' are considered a random variable distributed normally with

$$E(\beta_i') = 0$$

$$E(\beta_i'\beta_p') = \sigma^2 + k\sigma_b^2 \text{ if } i = p$$

$$= 0 \text{ if } i \neq p$$

In order to find estimates of the treatment effects under this model we minimize

$$\sum_i \beta_i'^2 = \sum_i \left(\frac{B_i}{k} - \mu - \frac{\sum_j n_{ij}\alpha_j}{k} \right)^2$$

with respect to the t + 1 parameters in the model.

Taking the partial derivatives with respect to each of the parameters, setting the result equal to zero and solving for α_j, we obtain the least squares estimates of the α_j, namely,

$$\hat{\alpha}_j = \frac{T_j}{(r - \lambda)} \text{ .}$$

We now have two independent estimates of the treatment effects, i.e.,

$$(2) \quad \hat{\alpha}_j = \frac{V_j - \frac{T_j}{k}}{(r - \frac{r}{k} + \frac{\lambda}{k})} \qquad \text{and} \qquad (3) \quad \hat{\alpha}_j = \frac{T_j}{(r - \lambda)}$$

The analysis of variance for the balanced incomplete block design under the assumption of Model III is presented in TABLE I.

TABLE I

ANALYSIS OF VARIANCE UNDER MODEL III

Source	d. f.	Sum of Squares
Total	$bk - 1$	$\sum_i \sum_j \sum_m (y_{ijm} - \frac{Y_{...}}{bk})^2$
Blocks (ignoring treatments)	$b - 1$	$\frac{1}{k} \sum_i (B_i - \frac{B_.}{b})^2$
Treatment component	$t - 1$	$\frac{1}{k(r - \lambda)} \sum_j (T_j - \frac{T_.}{t})^2$
Remainder	$b - t$	Subtraction
Treatments (eliminating blocks)	$t - 1$	$\frac{1}{rE} \sum_j Q_j^2$
Intra-block error	w	Subtraction

where

$$\lambda = \frac{r(k - 1)}{(t - 1)}, \qquad E = \frac{t(k - 1)}{k(t - 1)} \quad \text{and} \quad w = bk - t - b + 1$$

We wish now to investigate in detail the quantity termed Blocks (ignoring treatments). Let B_i denote the i-th block total. Then in terms of the model,

$$B_i = k\mu + k\beta_i + \sum_j n_{ij}\alpha_j + \sum_j \sum_m e_{ijm}$$

We will now prove the following Lemma.

Lemma I.

<u>The B_i are distributed normally and independently with mean $k\mu + \sum_j n_{ij}\alpha_j$ and variance $k(\sigma^2 + k\sigma_b^2)$.</u>

Proof.

Mean of B_i:

$$E(B_i) = E\left(k\mu + k\beta_i + \sum_j n_{ij}\alpha_j + \sum_j \sum_m e_{ijm} \right)$$

$$= k\mu + \sum_j n_{ij}\alpha_j$$

Variance of B_i:

$$E(B_i - EB_i)^2 = E\left(k\mu + k\beta_i + \sum_j n_{ij}\alpha_j + \sum_j \sum_m e_{ijm} - k\mu - \sum_j n_{ij}\alpha_j\right)^2$$

$$= E\left(k\beta_i + \sum_j \sum_m e_{ijm}\right)^2 = E\left[k^2\beta_i^2 + \left(\sum_j \sum_m e_{ijm}\right)^2\right]$$

$$= k(\sigma^2 + k\sigma_b^2)$$

Covariance of B_iB_p $(i \neq p)$:

$$\underset{i\neq p}{E}\left(B_iB_p - EB_iEB_p \right) = E\left(k\mu + k\beta_i + \sum_j n_{ij}\alpha_j + \sum_j \sum_m e_{ijm}\right)\left(k\mu + k\beta_p\right.$$

$$\left. + \sum_{j'} n_{pj'}\alpha_{j'} + \sum_{j'} \sum_m e_{pj'm}\right) - \left(k\mu + \sum_j n_{ij}\alpha_j\right)\left(k\mu + \sum_{j'} n_{pj'}\alpha_{j'}\right)$$

$$= 0$$

and the Lemma is proved.

Now, consider the following identity in the B_i:

$$(4)\quad \left(\frac{B_i}{k} - \frac{B_.}{bk}\right) = \frac{1}{(r-\lambda)} \sum_p \sum_j n_{ij} n_{pj} \left(\frac{B_p}{k} - \frac{B_.}{bk}\right) + \left(\frac{B_i}{k} - \frac{B_.}{bk}\right) - \frac{1}{(r-\lambda)} \sum_p \sum_j n_{ij} n_{pj} \left(\frac{B_p}{k} - \frac{B_.}{bk}\right)$$

$$i = 1, 2, \ldots, b.$$

Before proceeding we define the following matrices:

$$\underset{b\times 1}{B} = \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ \cdot \\ B_b \end{bmatrix} \qquad \underset{b\times 1}{B_.} = \begin{bmatrix} a'B \\ a'B \\ \cdot \\ \cdot \\ \cdot \\ a'B \end{bmatrix} \qquad \text{where } \underset{1\times b}{a'} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix}.$$

We will denote (4) in matrix form as

$$\underset{b\times 1}{T} = \underset{b\times 1}{F} + \underset{b\times 1}{S}$$

Squaring both sides of (4) and summing on i, j and m we may write the resulting expression in matrix form as

$$\underset{1\times 1}{T^*} = \underset{1\times 1}{F^*} + \underset{1\times 1}{R}$$

where $T^* = c_1 T'T$, $F^* = c_2 F'F$ and R is a matrix composed of all remaining factors resulting from the summing and squaring.

Since T^* is a quadratic form in the B_i's, i.e.,

$$\underset{1\times 1}{T^*} = \sum_i \sum_j \sum_m \left(\frac{B_i}{k} - \frac{B_.}{bk}\right)^2 = k \sum_i \left(\frac{B_i}{k} - \frac{B_.}{bk}\right)^2 = \frac{1}{k} \sum_i \left(B_i - \frac{B_.}{bk}\right)^2,$$

we will denote T^* in matrix form as

$$\frac{1}{k} B'X_o'X_o B = T^*.$$

We must now find a matrix X_o such that

$$(5) \qquad \frac{1}{k}(B'X_o')(X_oB) = \frac{1}{k}\sum_i \left(B_i - \frac{B}{b} \right)^2.$$

Consider the matrix

$$\underset{b\times b}{X_o} = \begin{bmatrix} 1-\frac{1}{b} & -\frac{1}{b} & \cdot & \cdot & \cdot & -\frac{1}{b} \\ -\frac{1}{b} & 1-\frac{1}{b} & \cdot & \cdot & \cdot & -\frac{1}{b} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ -\frac{1}{b} & -\frac{1}{b} & \cdot & \cdot & \cdot & 1-\frac{1}{b} \end{bmatrix}.$$

Then using this X_o, the equation (5) holds. The matrix X_o is symmetric idempotent, i.e., $X_oX_o = X_o$. Therefore

$$\frac{1}{k} B'X_o'X_oB = \frac{1}{k} B'X_oB.$$

We wish now to find the distribution of this quadratic form. Before proceeding we state the following theorem, proved in reference (5).

<u>THEOREM I.</u>

If a vector $\underset{p\times 1}{Y}$ is distributed as the p-variate normal, mean vector μ_1, variance covariance matrix $\sigma_1^2 I$, then $Y'AY$ is distributed as the non-central chi-square distribution with parameters f and λ if and only if $A(\sigma^2 I)$ is idempotent, where f is the rank of $A(\sigma^2 I)$ and $\lambda = \frac{1}{2}\mu_1'A\mu_1$.

Applying this to the problem of finding the distribution of $\frac{1}{k} B'X_oB$ we have

$$\underset{b\times 1}{B} \sim N_b \left[m + \pi, \; k(\sigma^2 + k\sigma_b^2)I \right]$$

where

$$\underset{b\times 1}{m} = \begin{bmatrix} k\mu \\ k\mu \\ \cdot \\ \cdot \\ \cdot \\ k\mu \end{bmatrix} \quad \text{and} \quad \underset{b\times 1}{\pi} = \begin{bmatrix} \sum_j n_{1j}\alpha_j \\ \sum_j n_{2j}\alpha_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_j n_{bj}\alpha_j \end{bmatrix} .$$

Then $B'AB$ is distributed as the non-central chi-square with parameters f_o and λ_o, where

$$A = \frac{X_o}{k(\sigma^2 + k\sigma_b^2)} \qquad f_o = (b - 1)$$

$$\lambda_o = \frac{(m' + \pi')X_o(m + \pi)}{2k(\sigma^2 + k\sigma_b^2)}$$

for

$$A(\sigma^2 I) = \frac{X_o}{k(\sigma^2 + k\sigma_b^2)} k(\sigma^2 + k\sigma_b^2)I$$

$$= X_o I = X_o$$

which has been found to be symmetric idempotent of rank $(b - 1)$.

The non-centrality parameter λ_o must be evaluated in order to completely specify the distribution. Proceeding, we have

$$2k(\sigma^2 + k\sigma_b^2)\lambda_o = m'X_o m + m'X_o\pi + \pi'X_o m + \pi'X_o\pi.$$

It may be readily verified that the first three terms to the right of the equality are zero. The remaining term $\pi'X_o\pi$ then equals

$$\sum_i \left(\sum_j n_{ij}\alpha_j \right)^2 = \sum_i \sum_j \sum_{j'} n_{ij}n_{ij'}\alpha_j\alpha_{j'} = \sum_i \sum_j n_{ij}\alpha_j^2 + \sum_i \sum_j \sum_{\substack{j' \\ j\neq j'}} n_{ij}n_{ij'}\alpha_j\alpha_{j'}$$

$$= r \sum_j \alpha_j^2 + \lambda \sum_j \sum_{\substack{j' \\ j\neq j'}} \alpha_j\alpha_{j'} = r \sum_j \alpha_j^2 - \lambda \sum_j \alpha_j^2 = (r - \lambda) \sum_j \alpha_j^2 .$$

Therefore the non-centrality λ_o is

$$(6) \qquad \frac{(r - \lambda)}{2k(\sigma^2 + k\sigma_b^2)} \sum_j \alpha_j^2 .$$

We now consider the term

$$F^* = \sum_i \sum_j \sum_m \frac{1}{(r - \lambda)} \sum_p \sum_j n_{ij}n_{pj}\left(\frac{B_p}{k} - \frac{B_.}{bk} \right)^2$$

and we will show that F^* is equal to

$$(7) \qquad \frac{1}{k(r - \lambda)} \sum_j \left(\sum_i n_{ij}B_i - \frac{k}{t}B_. \right)^2$$

and also

$$(8) \qquad \frac{1}{k(r - \lambda)} \sum_j \left(T_j - \frac{1}{t} T_. \right)^2$$

where T_j is the total of all blocks containing treatment j and $T_. = \sum_j T_j$.

Denoting the quantity $\left(\frac{B_i}{k} - \frac{B_.}{bk} \right)$ by $\bar{\bar{B}}_i$, we have

$$F^* = \sum_i \sum_j \sum_m \left[\frac{1}{(r - \lambda)} \sum_p \sum_j n_{ij}n_{pj}\bar{\bar{B}}_p \right]^2 = \frac{1}{(r - \lambda)^2} \sum_i \left[\sum_j n_{ij}\left(\sum_p n_{pj}\bar{\bar{B}}_p \right) \right]^2$$

$$= \frac{k}{(r - \lambda)^2} \sum_i \left[\sum_j n_{ij}\left(\sum_p n_{pj}\bar{\bar{B}}_p \right)^2 + \sum_p \sum_q \sum_j \sum_{\substack{j' \\ j\neq j'}} n_{ij}n_{pj}n_{qj'}\bar{\bar{B}}_p\bar{\bar{B}}_q \right]$$

$$= \frac{k}{(r - \lambda)^2} \left[r \sum_j \left(\sum_p n_{pj}\bar{\bar{B}}_p \right)^2 + \lambda \sum_p \sum_q \sum_j n_{pj}(k - n_{qj})\bar{\bar{B}}_p\bar{\bar{B}}_q \right]$$

$$= \frac{k}{(r - \lambda)^2} \left[r \sum_j \left(\sum_p n_{pj}\bar{\bar{B}}_p \right)^2 - \lambda \sum_j \left(\sum_p n_{pj}\bar{\bar{B}}_p \right)^2 \right]$$

$$= \frac{k}{(r-\lambda)^2}(r-\lambda)\sum_j\left(\sum_p n_{pj}\bar{B}_p\right)^2 = \frac{k}{(r-\lambda)}\sum_j\left(\frac{1}{k}\sum_p n_{pj}B_p - \frac{r}{bk}B_{\circ}\right)^2$$

$$= \frac{1}{k(r-\lambda)}\sum_j\left(\sum_i n_{ij}B_i - \frac{k}{t}B_{\circ}\right)^2 = \frac{1}{k(r-\lambda)}\sum_j\left(T_j - \frac{1}{t}T_{\circ}\right)^2$$

which are the forms (7) and (8).

This is a quadratic form in the B_i and we shall write it in matrix notation as follows:

$$\underset{t\times 1}{Z} = \begin{bmatrix} n_{11} & n_{21} & n_{31} & \cdot & \cdot & \cdot & n_{b1} \\ n_{12} & n_{22} & n_{32} & \cdot & \cdot & \cdot & n_{b2} \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & & & & \\ n_{1t} & n_{2t} & n_{3t} & \cdot & \cdot & \cdot & n_{bt} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ \cdot \\ B_b \end{bmatrix} - \frac{k}{t}\begin{bmatrix} B_{\circ} \\ B_{\circ} \\ \cdot \\ \cdot \\ \cdot \\ B_{\circ} \end{bmatrix}$$

Then, using the notation used previously and letting N denote the coefficient matrix of B, we have

(9) $$Z = NB - \frac{k}{t}B^*_{\circ}$$

where

$$\underset{1\times t}{B^{*\prime}_{\circ}} = \begin{bmatrix} B_{\circ} & B_{\circ} & \cdot & \cdot & \cdot & B_{\circ} \end{bmatrix}$$

Then the form $\frac{1}{k(r-\lambda)}Z'Z$ is equivalent to (7) and (8). Using (9) we have

$$\frac{1}{k(r-\lambda)}Z'Z = \frac{1}{k(r-\lambda)}\left(B'N' - \frac{k}{t}B^{*\prime}_{\circ}\right)\left(NB - \frac{k}{t}B^*_{\circ}\right).$$

Letting A be a txb matrix of ones, we may write $B^* = AB$ and

$$\frac{1}{k(r-\lambda)}Z'Z = \frac{1}{k(r-\lambda)}\left(B'N'NB - \frac{k}{t}B'N'AB - \frac{k}{t}B'A'N + \frac{k^2}{t^2}A'A\right)$$

which may be written as a quadratic form in the B_i's as

(10) $$\frac{1}{k(r-\lambda)} B'(N'N - \frac{k}{t} N'A - \frac{k}{t} A'N + \frac{k^2}{t^2} A'A)B.$$

Examining the product of $A'_{bxt}N_{txb}$ we find that the result is a matrix of order bxb with every element equal to k. Denote this matrix by K*. Similarly the product of A'A is a bxb matrix of all t's. Denote A'A by T*. If we now let a bxb matrix of all ones be denoted by δ, we may write (10) as

$$\frac{1}{k(r-\lambda)} B'(N'N - \frac{2k}{t} K^* + \frac{k^2}{t^2} T^*)B$$

or

$$\frac{1}{k(r-\lambda)} B'(N'N - \frac{k^2}{t} \delta)B.$$

We now desire to ascertain whether the form

$$U = \frac{1}{k(r-\lambda)} (N'N - \frac{k^2}{t} \delta)$$

is idempotent. The p-p'th element of UU is

(11) $$\frac{1}{k^2(r-\lambda)^2} \sum_i (\sum_{j'} n_{p'j'}n_{ij'} - \frac{k^2}{t})(\sum_j n_{ij}n_{pj} - \frac{k^2}{t}),$$

and the p-p'th element of U is

(12) $$\frac{1}{k(r-\lambda)} (\sum_j n_{pj}n_{p'j} - \frac{k^2}{t}).$$

We will now expand (11) and find the relationship to (12). Expanding (11) we obtain

(13) $$\frac{1}{k^2(r-\lambda)^2}(\sum_i \sum_j \sum_{j'} n_{p'j'}n_{ij'}n_{pj} - \frac{k^2}{t} \sum_i \sum_{j'} n_{p'j'} - \frac{k^2}{t} \sum_i \sum_j n_{ij}n_{pj} + \sum_i \frac{k^4}{t^2})$$

Evaluating (13) term by term we have

(a) $$\frac{1}{k^2(r-\lambda)^2} (\sum_i \sum_j \sum_{j'} n_{p'j'}n_{ij'}n_{pj})$$

$$= \frac{1}{k^2(r-\lambda)^2}\left(\sum_i \sum_j n_{p'j} n_{ij} n_{pj} + \sum_i \sum_j \sum_{\substack{j' \\ j \neq j'}} n_{p'j'} n_{ij'} n_{pj}\right)$$

$$= \frac{1}{k^2(r-\lambda)^2}\left(r \sum_j n_{p'j} n_{pj} + \lambda k \sum_j n_{pj} - \lambda \sum_j n_{p'j} n_{pj}\right)$$

$$= \frac{1}{k^2(r-\lambda)^2}\left[(r-\lambda)\sum_j n_{pj} n_{p'j} + \lambda k^2\right]$$

(b) $\frac{-1}{k^2(r-\lambda)^2}\left(\frac{k^2}{t}\right)\left(\sum_i \sum_{j'} n_{p'j'} n_{ij'}\right)$

$$= \frac{-rk^3}{tk^2(r-\lambda)^2} = \frac{-rk}{t(r-\lambda)^2}$$

(c) $\frac{-1}{k^2(r-\lambda)^2}\left(\frac{k^2}{t}\right)\left(\sum_i \sum_j n_{ij} n_{pj}\right)$

$$= \frac{-rk}{t(r-\lambda)^2}$$

(d) $\frac{1}{k^2(r-\lambda)^2} \sum_i \frac{k^4}{t^2} = \frac{bk^2}{t^2(r-\lambda)^2}$.

Combining the results of (a), (b), (c) and (d) we have as the p-p'th element of UU

$$\frac{1}{k^2(r-\lambda)} \sum_j n_{pj} n_{p'j} + \frac{\lambda}{(r-\lambda)^2} - \frac{2rk}{t(r-\lambda)^2} + \frac{bk^2}{t^2(r-\lambda)^2}.$$

Simplifying the above result we obtain

$$\frac{1}{k^2(r-\lambda)}\left(\sum_j n_{pj} n_{p'j} - \frac{k^2}{t}\right). \tag{14}$$

This result differs from the p-p'th element of U (12) by the factor $\frac{1}{k}$. Therefore since $UU = \frac{1}{k} U$, this implies that kU is idempotent, i.e.,

$$(kU)(kU) = k^2 UU = k^2 \frac{1}{k} U = kU.$$

Since F* is a quadratic form in the B_i's, then F* = B'AB for some matrix A. Letting U be denoted by X_1 and applying Theorem I we have

$$(15) \qquad B'AB = \frac{1}{k(\sigma^2 + k\sigma_b^2)} B'(kX_1)(kX_1)B \sim \chi'^2(f_1, \lambda_1)$$

for if

$$A = \frac{kX_1}{k(\sigma^2 + k\sigma_b^2)}, \text{ then } A(\sigma_1^2 I) = \frac{kX_1}{k(\sigma^2 + k\sigma_b^2)} k(\sigma^2 + k\sigma_b^2)$$

$$= kX_1$$

which we have shown to be idempotent.

We must now evaluate the two parameters of the non-central chi-square distribution in (15). We will evaluate the non-centrality parameter λ_1 first. We know that $\lambda_1 = \frac{1}{2}\mu_1' A\mu_1$. Substituting for μ_1 and A we have

$$2k(r - \lambda)(\sigma^2 + k\sigma_b^2)\lambda_1 = (m' + \pi')kX_1(m + \pi).$$

Letting the coefficient of λ_1 be c_1 and substituting for kX_1 we obtain

$$c_1\lambda_1 = (m' + \pi')(N'N - \frac{k^2}{t})(m + \pi).$$

Expanding we obtain

$$c_1\lambda_1 = m'N'Nm + 2m'N'N\pi - \frac{2k^2}{t}m'\delta m + \pi'N'N\pi - \frac{k^2}{t}\pi'\delta\pi.$$

Simplifying we find that λ_1 equals

$$\frac{(r - \lambda)}{2k(\sigma^2 + k\sigma_b^2)} \sum_j \alpha_j^2.$$

We may now apply the fact that the rank of an idempotent matrix is equal to the trace to find f_1 in (15). Thus f_1 may be found to be $t - 1$, and we have the distribution in (15) completely specified.

We now have

$$B'X_oB = B'(kX_1)B + R$$

and let us denote R by the quadratic form in the B_i's as $B'X_2B$. We may then write

$$B'X_oB = B'(kX_1)B + B'X_2B$$

or

$$B'X_2B = B'(X_o - kX_1)B$$ and this implies the relation $X_2 = (X_o - kX_1)$ and we may write $X_o = kX_1 + (X_o - kX_1)$.

Let us examine the product of the two matrices kX_1 and X_2. We have then

$$(kX_1)(X_o - kX_1) = kX_1X_o - k^2X_1X_1 = kX_1X_o - kX_1 = kX_1(X_o - I).$$

Now $(X_o - I)$ is a matrix where every element is $-\frac{1}{b}$. In element form $kX_1(X_o - I)$ is

$$\begin{bmatrix} \sum_j n_{1j}n_{1j} - \frac{k^2}{t} & \sum_j n_{1j}n_{2j} - \frac{k^2}{t} & \cdots & \sum_j n_{1j}n_{bj} - \frac{k^2}{t} \\ \sum_j n_{2j}n_{1j} - \frac{k^2}{t} & \sum_j n_{2j}n_{2j} - \frac{k^2}{t} & \cdots & \sum_j n_{2j}n_{bj} - \frac{k^2}{t} \\ \vdots & \vdots & & \vdots \\ \sum_j n_{bj}n_{1j} - \frac{k^2}{t} & \sum_j n_{bj}n_{2j} - \frac{k^2}{t} & \cdots & \sum_j n_{bj}n_{bj} - \frac{k^2}{t} \end{bmatrix} \begin{bmatrix} -\frac{1}{b} & \cdots & -\frac{1}{b} \\ -\frac{1}{b} & \cdots & -\frac{1}{b} \\ \vdots & & \vdots \\ -\frac{1}{b} & \cdots & -\frac{1}{b} \end{bmatrix}$$

and the p-p'th element of the product is

$$\frac{1}{k(r-\lambda)} \sum_i \left(\sum_j n_{pj}n_{p'j} - \frac{k^2}{t} \right) = \frac{1}{k(r-\lambda)} \left(rk - \frac{bk^2}{t} \right) = 0.$$

Using this result we have

(16) $$X_o = kX_1 + (X_o - kX_1).$$

Squaring both sides we obtain

$$X_o X_o = k^2 X_1 X_1 + 2kX_1(X_o - kX_1) + (X_o - kX_1)(X_o - kX_1)$$

or
$$X_o = kX_1 + \varphi + (X_o - kX_1)(X_o - kX_1)$$

or
$$X_o - kX_1 = (X_o - kX_1)(X_o - kX_1).$$

Since $X_2 = (X_o - kX_1)$ the above expression becomes $X_2 = X_2 X_2$ and we have shown that X_2 is idempotent.

We now have $X_o = kX_1 + X_2$ and we have found the ranks of X_o and X_1 to be $(b - 1)$ and $(t - 1)$ respectively. It is well known that the sum of the ranks of two matrices is greater than or equal to the rank of the sum. Applying this to (16), we have if we denote the unknown rank of X_2 as q,

$$b - 1 \leqq t - 1 + q$$

or
$$b - t \leqq q$$

or
$$b - t = q - c^2$$

or
$$q = b - t + c^2.$$

Then
$$b - 1 = t - 1 + b - t + c^2$$

or
$$c^2 = 0 = c.$$

Therefore the rank of X_2 is $(b - t)$.

We have yet to find the distribution of $B'X_2B$. Applying Theorem I we have

$$B'AB \sim \chi'^2 (b - t, \lambda_2)$$

for if we let

$$A = \frac{X_2}{k^2(\sigma^2 + k\sigma_b^2)} \text{, then } A(\sigma_1^2 I) = \frac{X_2}{k(\sigma^2 + k\sigma_b^2)} k(\sigma^2 + k\sigma_b^2)$$

or $A(\sigma_1^2 I) = X_2$, which we have shown to be idempotent of rank $(b - t)$. λ_2 may be found from the relation

$$X_o = kX_1 + X_2.$$

Multiplying on the right and left by $(m + \pi) = \mu_o$ (say), we have

$$\mu_o' X_o \mu_o = k\mu_o' X_1 \mu_o + \mu_o' X_2 \mu_o.$$

If we divide this equation through by the quantity $2k(\sigma^2 + k\sigma_b^2)$ we will have an equation in the non-centralities of the quadratic forms T*, F* and R. The non-centralities of T* and F* have been shown to be equal and therefore the non-centrality of R is zero. Then R is distributed as the central chi-square with $(b - t)$ degrees of freedom.

The non-centrality of F* is a function of $\sum_j \alpha_j^2 = \sum_j (\alpha_j' - \alpha_o')^2$.

Then, under the null hypothesis H_o: $\alpha_1' = \alpha_2' = \ldots = \alpha_t'$, the non-centrality of F* is zero and F* is distributed as the central chi-square with $(t - 1)$ degrees of freedom. The foregoing results are summarized in Table II.

TABLE II

SUB-ANALYSIS OF VARIANCE UNDER MODEL III

Source	d. f.	Sum of Squares
Blocks (ignoring treatments)	$b - 1$	$\frac{1}{k} B' X_o B = Q$
Treatment component	$t - 1$	$B' X_1 B = Q_1$
Remainder	$b - t$	$\frac{1}{k} B' X_2 B = Q_2$

Then, under the null hypothesis, the ratio

$$\frac{Q_1(b - t)}{Q_2(t - 1)}$$

is distributed as Snedecor's "F" with $(t - 1)$ and $(b - t)$ degrees of freedom.

We now have two independent tests of the null hypothesis for the treatment effects assuming Eisenhart's Model III. The next chapter will be concerned with combining these two tests of significance into a single test of significance.

CHAPTER III

COMBINING INDEPENDENT TESTS OF SIGNIFICANCE

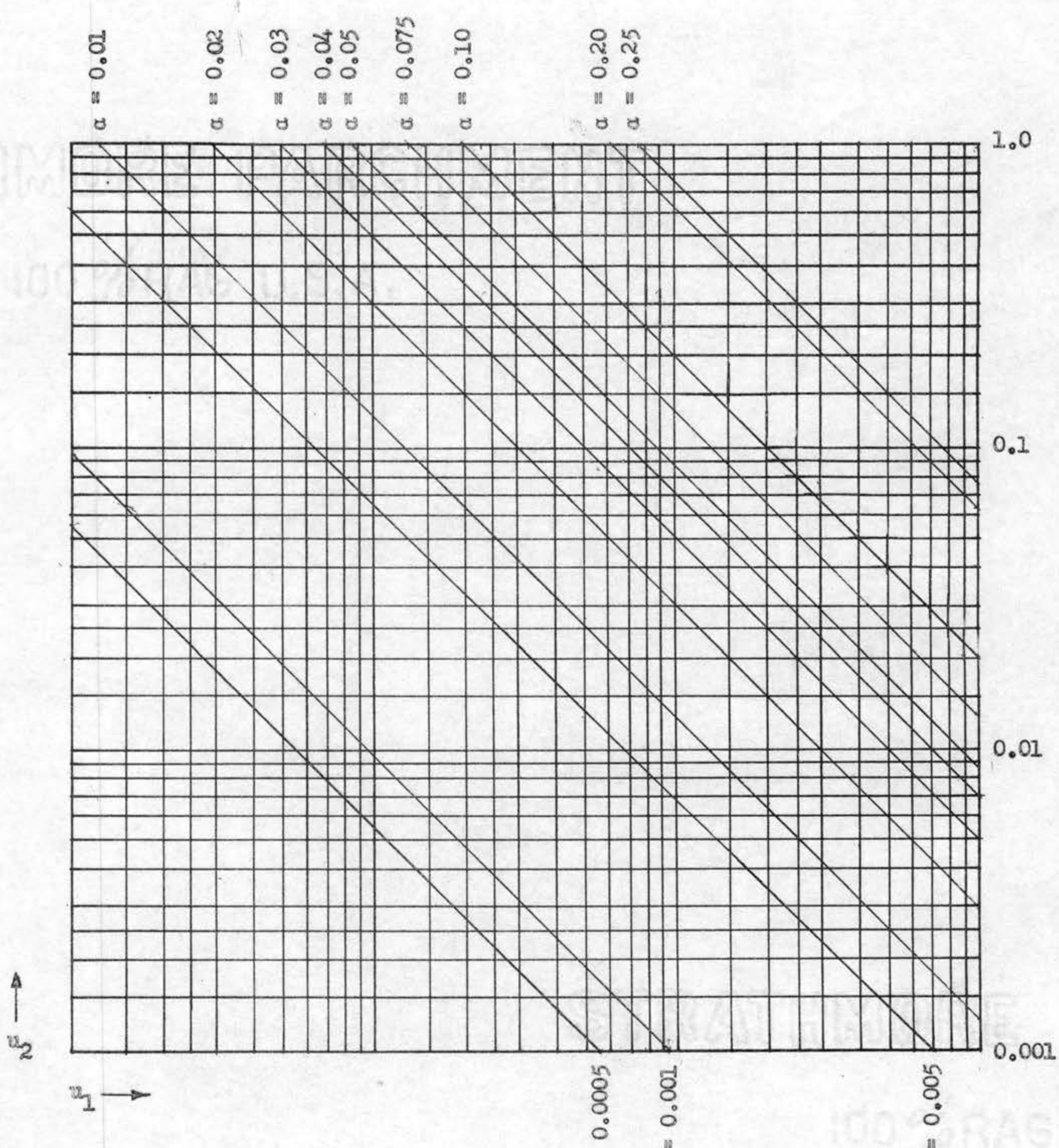
We have shown that two independent tests of significance of H_o: $\alpha_1' = \alpha_2' = . . . = \alpha_t'$ can be obtained in a balanced incomplete block design when b is greater than t and when Eisenhart's Model III is assumed. The purpose of this chapter will be to give a method of combining these two independent tests.

There exist many criteria for combining independent tests and the one which will be considered here is by R. A. Fisher. His criteria consists of rejecting H_o if and only if $u_1 u_2 \leqq c$ where u_1 and u_2 are the significance levels of the two independent tests and c is a predetermined constant corresponding to the desired significance level.

It has been shown that $-2 \log_e u_1 u_2$ is distributed as the chi-square variate with 4 degrees of freedom when H_o is true. Then if x is such that Probability $[\chi^2(4) \leqq x] = \alpha$, where $1 - \alpha$ is the desired significance level, and setting $-2 \log_e c = x$, we find that $\log_e c = -x/2$ and c may be computed from chi-square and $\log_e$ tables.

In the table which follows, c has been computed for several values of α and the function $u_1 u_2 = c$ plotted on log paper so that the curve is represented by a straight line. To find the significance level of the combined independent tests, merely find the significance level of each "F" and find the point in the $u_1 u_2$ plane. A point falling in the area between say the α_p and α_q levels of significance ($\alpha_p < \alpha_q$) is significant at the α_q but not at the α_p level.

FIGURE 1



GRAPH OF FUNCTION $u_1 u_2 = c(\alpha)$

CHAPTER IV

APPLICATION OF TECHNIQUES

We will consider here an example and work through the analysis showing the techniques which may be used to best advantage when Model III is assumed for this design.

Consider the following layout in a balanced incomplete block with $b = 6$, $t = 4$, $k = 2$, $r = 3$ and $\lambda = 1$, and artificial data.

TABLE III

STATISTICAL LAYOUT

Treatment	1	2	3	4	Block Totals
Block 1	6	8			14
2			13	7	20
3	7		13		20
4		12		2	14
5	5			3	8
6		10	11		21
Treatment Totals	18	30	37	12	97 Grand Total

Letting V_j denote the j-th treatment total and T_j denote the total of all blocks containing treatment j, form the following quantities:

	V_j	T_j	$T_j/2$	$V_j - T_j/2$
j = 1	18	42	21.0	− 3.0
= 2	30	49	24.5	5.5
= 3	37	61	30.5	6.5
= 4	12	42	21.0	− 9.0
Totals	97 = GT	194 = 2GT	97 = GT	0.0

and also find the quantities $r\not{E} = 2$ and $k(r - \lambda) = 4$.

We are now ready to compute the Analysis of Variance for this layout. The sums of squares are obtained in the following manner:

Total: $6^2 + 7^2 + \ldots + 3^2 - \frac{97^2}{12} = 154.92$

Blocks (ignoring treatments): $\frac{14^2 + 20^2 + \ldots + 21^2}{2} - \frac{97^2}{12} = 64.42$

Treatment component: $\frac{1}{4}\left(42^2 + 49^2 + 61^2 + 42^2 - \frac{194^2}{4} \right) = 60.25$

Remainder: By subtraction, $64.42 - 60.25 = 4.17$

Treatments (eliminating blocks): $\frac{1}{2}\left[(-3)^2 + 5.5^2 + 6.5^2 + (-9.)^2 \right] = 81.25$

Intra-block error: By subtraction, $154.92 - 64.42 - 81.25 = 9.25$

The results of this particular layout are summarized in TABLE IV which is shown on the next page.

Going to FIGURE I and looking on the u_1 axis for $u_1 = 0.0975$ and on the u_2 axis for $u_2 = 0.0625$ we find that the resulting test is significant between the 3% and 4% level, and therefore we would reject the null hypothesis at the 4% level but not at the 3% level.

TABLE IV

ANALYSIS OF VARIANCE FOR STATISTICAL LAYOUT IN TABLE III

Source	d. f.	S.S.	M.S.	F
Total	11	154.92		
Blocks (ignoring treatments)	5	64.42		
Treatment component	3	60.25	20.08	9.65
Remainder	2	4.17	2.08	
Treatments (eliminating blocks)	3	81.25	27.08	8.79
Intra-block error	3	9.25	3.08	

and we find that

$$\text{Probability } [F_{tab}(3,2) > 9.65] = 0.0975 = u_1$$

and

$$\text{Probability } [F_{tab}(3,3) > 8.79] = 0.0625 = u_2 .$$

CONCLUSIONS

The results of this thesis may be stated in the form of a theorem namely,

THEOREM

When Eisenhart's Model III is assumed in a balanced incomplete block design and the number of blocks is greater than the number of treatments, then there exist two independent tests of the null hypothesis H_o: $\alpha_1' = \alpha_2' = \alpha_3' = . \quad . \quad . = \alpha_t'$, i.e., that all the treatment effects are equal.

If an exact method of combining independent tests is employed to combine the two tests of H_o, then the test of H_o under Model III is exact.

These results give rise to problems which could be investigated that are not solved here. A few of which are

(1) examining the power of the test of the null hypothesis under Model III,

(2) developing a criteria for combining independent confidence intervals on the same parameter of a distribution and

(3) developing a criteria for combining independent estimates of the same parameter of a distribution.

BIBLIOGRAPHY

(1) Anderson, R. L. and Bancroft, T. A. Statistical Theory in Research. New York: McGraw-Hill Book Company, Inc., 1952.

(2) Birnbaum, Allan. "Combining Independent Tests of Significance." Journal of the American Statistical Association, 49 (September, 1954), 559-574.

(3) Eisenhart, C. "The Assumptions Underlying the Analysis of Variance." Biometrics, 3, 1-21, 1947.

(4) Fisher, R. A. Statistical Methods for Research Workers. Fourth and later Editions, Edinburgh and London: Oliver and Boyd, 1932 and later, Section 21.1.

(5) Graybill, Franklin A. and Marsaglia, George. "Idempotent Matrices and Quadratic Forms in the General Linear Hypothesis." To be published in Annals of Mathematical Statistics, September, 1957.

(6) Kempthorne, Oscar. Design and Analysis of Experiments. New York: John Wiley and Sons, Inc., 1950.

(7) Pearson, E. S. "The Probability Integral Transformation for Testing Goodness of Fit and Combining Independent Test of Significance." Biometrika 30, 1/2, 134-148, 1938.

(8) Rao, C. Radhakrishna. Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc., 1952.

VITA

David Lee Weeks
candidate for the degree of
Master of Science

Thesis: AN EXACT TEST OF SIGNIFICANCE IN THE BALANCED INCOMPLETE BLOCK DESIGN WITH RECOVERY OF INTER-BLOCK INFORMATION

Major: Mathematics

Minor: Statistics

Biographical and Other Items:

Born: June 24, 1930 at Boone, Iowa

Undergraduate Study: O. A. M. C., 1948-52

Graduate Study: O. A. M. C., 1954-55, 1956-57

Experiences: Army Signal Corps, I US Corps in Korea 1952-54; Graduate Assistant in Mathematics Department 1954-55, 1956-57

Member of Phi Kappa Phi, Pi Mu Epsilon and The American Statistical Association

Date of Final Examination: May, 1957

THESIS TITLE: AN EXACT TEST OF SIGNIFICANCE IN THE BALANCED INCOMPLETE BLOCK DESIGN WITH RECOVERY OF INTER-BLOCK INFORMATION

AUTHOR: David Lee Weeks

THESIS ADVISER: Dr. Franklin Graybill

The content and form have been checked and approved by the author and thesis adviser. The Graduate School Office assumes no responsibility for errors either in form or content. The copies are sent to the bindery just as they are approved by the author and the faculty adviser.

TYPIST: David Lee Weeks