PROPERTIES OF MODEL II—TYPE ANALYSIS OF VARIANCE TESTS, A: OPTIMUM NATURE OF THE F-TEST FOR MODEL II IN THE BALANCED CASE[1, 2]

By Leon H. Herbach

College of Engineering, New York University[3]

1. Summary. A distribution analogous to the canonical distribution used in testing the general linear hypothesis is developed for Model II analysis of variance for balanced classifications. As in the case of Model I analysis of variance, this standard distribution exhibits the sums of squares going into the analysis of variance table. By use of the standard form it is also shown that (i) all exact F-tests used in testing hypotheses based on balanced multiple classifications determine uniformly most powerful (u.m.p.) similar regions although they are not likelihood ratio (L.R.) tests, but (ii) in the balanced one-way classification, for all practical purposes, the test is an L.R. test, and is u.m.p. invariant. An *exact* F-test exists when we have a sum of squares, S_1 distributed as $(k + \sigma_0^2)$ times a chi-square variate, where $k > 0$, independently of S_2, which is distributed as k times a chi-square variate. The test is then to reject the hypothesis that $\sigma_0^2 = 0$ whenever S_1/S_2 is greater than some suitably chosen number, c. As a corollary to property (i) it is shown that "of all invariant tests of $\sigma_0^2 = 0$ against $\sigma_0^2 > 0$ whose power is a function of $\sigma_0^2/(k + \sigma_0^2)$ only, the test $S_1/S_2 > c$ is most powerful, providing S_1 and S_2, as defined above can be found."

2. Notation and terminology. We use the notation $p_\theta(x)$ for the probability density function (p.d.f.) of the vector-valued random variable, X, which depends on the vector-valued parameter $\theta \varepsilon \Omega$, where Ω will always represent the unrestricted parameter space. This notation is generic so that p may not be the same density each time it appears. The difference in functional form is indicated by the change in variable. The actual form will always be clear from the context. This same generic notation will be used for constants; c will usually be a constant, not necessarily the same one each time it appears. It will be clear from the context when c is not a constant. The subspace of Ω specified by the hypothesis being tested will be denoted by ω. No confusion will be caused when dealing with the hypothesis $H: \theta \varepsilon \omega$ if we sometimes speak of ω rather than H as the hypothesis. By a test of an hypothesis we mean any measurable function $\varphi(x)$ with the property that $0 \leqq \varphi(x) \leqq 1$. When X is observed to take on the value x one rejects H with probability $\varphi(x)$.

3. Introduction. In Model II (components of variance model) analysis of

Received May 8, 1958; revised January 15, 1959.

[1] Part of the author's doctoral dissertation submitted to Columbia University [3].

[2] Work partially sponsored by the Office of Naval Research.

[3] Scientific Paper No. 6a, Engineering Statistics Laboratory.

variance, the following stochastic model is assumed in the case of a two-way classification with K observations per cell:

$$X_{ijk} = \mu + e_i^A + e_j^B + e_{ij}^{AB} + e_{ijk}, \tag{3.1}$$
$$i = 1, \cdots, I; \quad j = 1, \cdots, J; \quad k = 1, \cdots K,$$

where X_{ijk} is the kth measurement on the (i, j)th cell, μ the main effect is assumed a constant and the "components" e_i^A, e_j^B, e_{ij}^{AB}, e_{ijk} are normally and independently distributed (NID) with means zero and variances σ_a^2, σ_b^2, σ_{ab}^2, σ_e^2 respectively. These will be referred to as the Model II assumptions. If, as here, one has the same number of observations in each cell, the classification is called *balanced*, otherwise unbalanced.

The corresponding model for Model I (general linear hypothesis model) analysis of variance is given by (3.1) where it is now assumed that in addition to μ, e_i^A, e_j^B and e_{ij}^{AB} are also constants, and e_{ijk} are the only random variables, and these are NID$(0, \sigma^2)$. Furthermore it is usually assumed that $\sum_i e_i^A = \sum_j e_j^B = \sum_i e_{ij}^{AB} = \sum_j e_{ij}^{AB} = 0$. These equations for the effects may be assumed without loss of generality in Model I but would violate the assumed independence in Model II. The usual theoretical procedure in setting up any Model I hypothesis, say $H_0 : a_i = 0$ $(i = 1, \cdots, I)$, is to find the likelihood ratio test of the hypothesis. This gives the usual F-test. In addition to having the backing of the intuitive appeal of the likelihood ratio test, the resulting F-test has been shown by Hsu [4], [5], Wald [14], Wolfowitz [16] and others to have many optimum properties.

Analysis of variance, to many, also means a technique of calculating the analysis of variance table given in Table 3.1, where

$$\begin{aligned}
S_1 &= IJK\,(X_{...} - \mu)^2 \\
S_2 &= JK\sum_i (X_{i..} - X_{...})^2 \\
S_3 &= IK\sum_j (X_{.j.} - X_{...})^2 \\
S_4 &= K\sum_i \sum_j (X_{ij.} - X_{i..} - X_{.j.} + X_{...})^2 \\
S_5 &= \sum_i \sum_j \sum_k (X_{ijk} - X_{ij.})^2.
\end{aligned} \tag{3.2}$$

TABLE 3.1

Analysis of Variance Table for a Balanced Two-way Classification

Source	d.f.	S.S.	m.s.	E (mean square)
mean	$\nu_1 = 1$	S_1	S_1	λ_1
A effect	$\nu_2 = (I - 1)$	S_2	S_2/ν_2	λ_2
B effect	$\nu_3 = (J - 1)$	S_3	S_3/ν_3	λ_3
AB interaction	$\nu_4 = (I - 1)(J - 1)$	S_4	S_4/ν_4	λ_4
error	$\nu_5 = IJ(K - 1)$	S_5	S_5/ν_5	λ_5

The mean row and E (mean square) column do not always appear in the usual analysis of variance table and will be explained later. The statistic used in Model I to test H_0 is $(\nu_5 S_2)/(\nu_2 S_5)$, which is distributed as F with ν_2 and ν_5 degrees of freedom.

The procedure in Model II is to set up the analysis of variance table that is *used in Model I*, and then to add a column which gives the expected mean squares. One then notes that the five mean squares are always independently distributed and that $S_i/(\nu_i\lambda_i)$ is distributed as χ^2 with ν_i degrees of freedom. Using the fact that the expected mean squares are

$$\begin{aligned} \lambda_1 &= \sigma_e^2 + K\sigma_{ab}^2 + JK\sigma_a^2 + IK\sigma_b^2 \\ \lambda_2 &= \sigma_e^2 + K\sigma_{ab}^2 + JK\sigma_a^2 \\ \lambda_3 &= \sigma_e^2 + K\sigma_{ab}^2 + IK\sigma_b^2 \\ \lambda_4 &= \sigma_e^2 + K\sigma_{ab}^2 \\ \lambda_5 &= \sigma_e^2 \end{aligned} \tag{3.3}$$

we have, under the hypothesis of no Model II A effect $(H_0':\sigma_a^2 = 0)$, that $\lambda_2 = \lambda_4$ and $(\nu_4 S_2)/(\nu_2 S_4)$ is distributed as F with ν_2 and ν_4 degrees of freedom. This, in fact, is the F-statistic used in testing H_0'. All exact F-tests used in Model II are obtained by taking ratios of mean squares which have equal λ's under the hypothesis, *whenever there are equal* λ's. No attempt has been made previously to show that these tests are optimum or to even show they are likelihood ratio (L.R.) tests, which they sometimes are not. Two of the purposes of this paper are to derive optimum properties for some tests of Model II hypotheses and to show that in this model the analysis of variance table can be obtained without borrowing it from Model I.

4. Some useful lemmas for a certain matrix. The following $n \times n$ matrix plays an important role in what follows:

$$A = \begin{Bmatrix} a+b & a & a \cdots & a \\ a & a+b & a \cdots & a \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a & a & a \cdots & a+b \end{Bmatrix}, \tag{4.1}$$

where a and b are either *scalars or square matrices* of the same size. Since A is a function of $a + b$ and a only, the notation

$$A = (a + b\backslash a)^4 \tag{4.2}$$

shall be used.

[4] It may be noted that for a, b scalars, $A = b\mathcal{I} + a\mathcal{I}^*$, where $\mathcal{I}$ is the unit matrix and $\mathcal{I}^*$ is the matrix with all elements unity.

We shall make use of the following two lemmas:

LEMMA 4.1: *If A is of the form* (4.1) *then the determinant of A satisfies*

$$| A | = | b + na || b |^{n-1}$$

where $| D |$ *means the determinant of the matrix D.*

The proof is exactly the same as that given by Wilks ([15], p. 109) for the case in which a and b are scalars.

LEMMA 4.2: *If a, b are real numbers with* $b(b + na) \neq 0$,

$$A^{-1} = \frac{1}{(b + na)b} ([b + (n - 1)a] \backslash -a). \tag{4.3}$$

5. Standard form for the balanced two-way classification. Consider the two-way Model II classification with K observations per cell, given by (3.1). Let the transpose of the observation vector X be

$$\begin{aligned} X' = (&X_{111}, X_{211}, X_{311}, \cdots, X_{I11}; X_{121}, X_{221}, \cdots, X_{I21}; \cdots; \\ &X_{1J1}, X_{2J1}, X_{3J1}, \cdots, X_{IJ1}; X_{112}, X_{212}, X_{312}, \cdots; X_{I12}; \\ &X_{122}, X_{222}, \cdots, X_{I22}; \cdots; X_{1JK}, X_{2JK}, X_{3JK}, \cdots, X_{IJK}), \end{aligned} \tag{5.1}$$

that is, the triplets i, j, k are ordered so that

$$\begin{aligned} &(i, j, k) \text{ precedes } (i', j, k) && \text{if } i < i', \\ &(i, j, k) \text{ precedes } (i', j', k) && \text{if } j < j', \\ &(i, j, k) \text{ precedes } (i', j', k') && \text{if } k < k'. \end{aligned}$$

Let Σ be the covariance matrix of X and $N = IJK$. Then it is known that there exists an orthogonal matrix D with the following properties: (a) its first row is $N^{-\frac{1}{2}}\delta'$, where

$$\delta' = (1, 1, \cdots, 1), \tag{5.2}$$

a $1 \times N$ row vector, (b) the covariance matrix of $Z = DX$ is $D \Sigma D' = \Lambda =$ diag. $(\lambda_1, \cdots, \lambda_N)$, $\lambda_i > 0$ and (c) the λ's are the roots of the characteristic equation $| \Sigma - \lambda \mathcal{I} | = 0$, where $\mathcal{I}$ is the identity matrix.

We shall now find these λ's in terms of σ_a^2, σ_b^2, σ_{ab}^2 and σ_e^2. Now

$$| \underset{K \times K}{\Sigma} - \lambda \mathcal{I} | = | \mathcal{A} \backslash \mathcal{B} |$$

where

$$\begin{aligned} \mathcal{A} &= (A_1 \backslash B_1) \\ \underset{J \times J}{\mathcal{B}} &= (A_2 \backslash B_2) \\ A_1 &= (\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \sigma_e^2 - \lambda \backslash \sigma_b^2) \\ B_1 &= (\sigma_a^2 \backslash 0) = B_2 \\ \underset{I \times I}{A_2} &= (\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 \backslash \sigma_b^2). \end{aligned} \tag{5.3}$$

It should be noted that $\mathfrak{T}$ is an $N \times N$ matrix of scalars, but is a $K \times K$ matrix when the elements are submatrices ($\mathfrak{A}$'s and $\mathfrak{B}$'s). Repeated use of Lemma 4.1 yields

$$\begin{aligned} | \mathfrak{T} - \lambda \mathfrak{g} | &= |(\mathfrak{A} - \mathfrak{B}) + K\mathfrak{B} \| \mathfrak{A} - \mathfrak{B} |^{K-1} \\ &= | A_1 + (K - 1)A_2 \backslash KB_2 | \cdot | A_1 - A_2 \backslash 0 |^{K-1} \\ &= | A_1 + (K - 1)A_2 + K(J - 1)B_2 | \cdot | A_1 + (K - 1)A_2 \\ &\quad - KB_2 |^{J-1} \, | A_1 - A_2 |^{J(K-1)} \\ &= D_1 \cdot D_2 \cdot D_3 \text{, say,} \end{aligned}$$

where

$$\begin{aligned} D_1 &= | K(\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2) + \sigma_e^2 + K(J - 1)\sigma_a^2 - \lambda \backslash K\sigma_b^2 | \\ &= | K\sigma_{ab}^2 + \sigma_e^2 + JK\sigma_a^2 + IK\sigma_b^2 - \lambda \| K\sigma_{ab}^2 + \sigma_e^2 + JK\sigma_a^2 - \lambda |^{I-1} \\ D_2 &= | K(\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2) + \sigma_e^2 - \lambda - K\sigma_a^2 \backslash K\sigma_b^2 |^{J-1} \\ &= | K\sigma_{ab}^2 + IK\sigma_b^2 + \sigma_e^2 - \lambda |^{J-1} \cdot | K\sigma_{ab}^2 + \sigma_e^2 - \lambda |^{(I-1)(J-1)} \\ D_3 &= | \sigma_e^2 - \lambda \backslash 0 |^{J(K-1)} = | \sigma_e^2 - \lambda |^{IJ(K-1)}. \end{aligned}$$

Therefore the values of the $N = IJK$ characteristic roots are

$$\begin{aligned} \sigma_e^2 + K\sigma_{ab}^2 + IK\sigma_b^2 + JK\sigma_a^2 &= \lambda_1 \text{, say with multiplicity } 1 \\ \sigma_e^2 + K\sigma_{ab}^2 + JK\sigma_a^2 &= \lambda_2 \text{, say with multiplicity } (I - 1) \\ \sigma_e^2 + K\sigma_{ab}^2 + IK\sigma_b^2 &= \lambda_3 \text{, say with multiplicity } (J - 1) \\ \sigma_e^2 + K\sigma_{ab}^2 &= \lambda_4 \text{, say with multiplicity } (I - 1)(J - 1) \\ \sigma_e^2 &= \lambda_5 \text{, say with multiplicity } IJ(K - 1). \end{aligned}$$

These λ's are the same as the ones defined by (3.3). The orthogonality of D, the property that the first row of D is $N^{-\frac{1}{2}}\delta'$ and the fact that $EX_{ijk} = \mu$ imply that

$$\begin{aligned} EZ_{111} &= \sqrt{N}\mu \\ EZ_{ijk} &= 0, \qquad \text{for } (i, j, k) \neq (1, 1, 1). \end{aligned} \tag{5.4}$$

After the dissertation [3] was defended but before this paper was prepared, Dr. Howard Levene called the author's attention to the work of Nelder [11] whose method for obtaining the latent roots of a special case of matrices of the form (4.1) can be generalized to find our eigenvalues. However, it is felt that the algorithm, using Lemma 4.1 is more convenient, especially when higher multiple classifications are treated.

Let $\zeta = EZ$, the vector given in (5.4). We have shown that the vector variable X defined by (3.1) which is distributed as $N(\mu\delta, \mathfrak{T})$ where $\mathfrak{T} = (\mathfrak{A} \backslash \mathfrak{B})$

and $\mathfrak{A}$, $\mathfrak{B}$ are defined by (5.3) when $\lambda = 0$, may be transformed by an orthogonal matrix to yield a variable Z which has the following density:

$$(5.5) \quad \frac{|\Lambda|^{-\frac{1}{2}}}{(2\pi)^{N/2}} e^{-\frac{1}{2}[(z-\zeta)'\Lambda^{-1}(z-\zeta)]} = \frac{|\Lambda|^{-\frac{1}{2}}}{(2\pi)^{IJK/2}} \exp\left\{-\frac{1}{2}\left(\frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2} + \frac{s_3}{\lambda_3} + \frac{s_4}{\lambda_4} + \frac{s_5}{\lambda_5}\right)\right\},$$

where $\lambda_1, \cdots, \lambda_5$ are given by (3.3) and

$$(5.6) \quad \begin{aligned} s_1 &= (z_{111} - \sqrt{N}\mu)^2 \\ s_2 &= \sum_{i=2}^{I} z_{i11}^2 \\ s_3 &= \sum_{j=2}^{J} z_{1j1}^2 \\ s_4 &= \sum_{i=2}^{I}\sum_{j=2}^{J} z_{ij1}^2 \\ s_5 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=2}^{K} z_{ijk}^2 \end{aligned}$$

The reader should note that s_1 is not a statistic, since it contains μ. This particular expression for s_1 is used because of the symmetry it gives to (5.5), which we shall refer to as the Model II standard form of the probability density for the case of a balanced two way classification with K observations per cell. Note that (3.2) implies that

$$(5.7) \quad \lambda_1 = \lambda_2 + \lambda_3 - \lambda_4,$$

a fact which will be used later.

For completeness the Tang canonical form [13] of the joint density of (3.1) when the Model I assumptions are made will now be given. Then

$$X_{ijk}:\text{NID}(\mu_{ij}, \sigma^2)$$

where

$$\mu_{ij} = \mu + e_i^A + e_j^B + e_{ij}^{AB}.$$

Tang showed that there exists an orthogonal $N \times N$ matrix D whose first row is $N^{-\frac{1}{2}}\delta'$ for which $Z = DX$ has the density,

$$(5.8) \quad \begin{aligned} \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\Bigg[-\frac{1}{2\sigma^2}\Bigg\{&(z_{111} - \sqrt{N}\mu)^2 \\ &+ \sum_{i=2}^{I}(z_{i11} - a_i^A)^2 + \sum_{j=2}^{J}(z_{1j1} - a_j^B)^2 \\ &+ \sum_{i=2}^{I}\sum_{j=2}^{J}(z_{ij1} - a_{ij}^{AB}) + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=2}^{K} z_{ijk}^2\Bigg\}\Bigg] \end{aligned}$$

where the $a_i^A(a_j^B, a_{ij}^{AB})$ are linear combinations of the $e_i^A(e_j^B, e_{ij}^{AB})$ such that $a_i^A(a_j^B, a_{ij}^{AB})$ are zero if and only if all $e_i^A(e_j^B, e_{ij}^{AB})$ are zero. It should be noted that in Model I one transforms to change the means while in Model II one does so to change the covariance matrix. It can be shown [3] that the same orthogonal transformation, D, could be used in both models, so that we are justified in using the same letter Z in (5.6) and (5.8). Also, Tang showed for Model I, that $S_1, \cdots, S_5$ as given in (3.2) and (5.6) are the same, except that (3.2) expresses the S's in terms of original variates while (5.6) does so in terms of transformed variates. Since the transformations are the same in both models this shows that the sums of squares in (3.2) and (5.6) are the same in Model II. From this one can argue that the standard distribution (5.5) can be obtained from the analysis of variance table above since it is known that all rows are independently distributed. This was not done because we do not yet know what the properties of the tests based on Table 3.1 are. We propose to get the table, tests and optimum properties of the tests from (5.5), the standard form, which is easier to handle than the density of X, although all tests of hypotheses based on X can be transformed to tests based on Z. It should be noted that (5.5) is the density of Z although it is written in terms of the s's. From (5.5) and (5.6) it is clear that $\bar{X} = N^{-\frac{1}{2}}Z_{111}$, S_2, S_3, S_4 and S_5 are independently distributed as a normal variate and four multiples of χ^2 with $(I - 1)$, $(J - 1)$, $(I - 1)(J - 1)$ and $IJ(K - 1)$ degrees of freedom respectively. In the sequel we shall use this latter joint density, namely,

$$(5.9) \quad p(\bar{x}, s_2, s_3, s_4, s_5) = \left(\frac{N}{2\pi\lambda_1}\right)^{\frac{1}{2}} \exp\left[-\frac{N(\bar{x} - \mu)^2}{2\lambda_1}\right] \prod_{i=2}^{5} \frac{s_i^{\frac{\nu_i - 2}{2}} \exp\left(-\frac{s_i}{2\lambda_i}\right)}{(2\lambda_i)^{\nu_i/2}\Gamma(\nu_i/2)}$$

Densities (5.5) and (5.8) or (5.9) and (5.8) show clearly that under the hypothesis of no A effect, H_0' and H_0 respectively, S_2/S_4 and S_2/S_5 respectively are distributed as a multiple of F with the degrees of freedom indicated by the number of standard variates in each S. These are the statistics indicated at the end of Section 3. All F-tests used to test the non-existence of certain effects can be obtained this way.

6. Uniformly most powerful similar test for testing non-existence of main effects in the balanced one or two-way classification. This section will be devoted to showing that the F-test is the u.m.p. similar test for testing $\omega:\sigma_a^2 = 0$ against $\Omega - \omega:\sigma_a^2 > 0$ when one has a balanced one or two way Model II classification. Although the hypothesis to be tested is actually $\sigma_a^2 = 0$, $\sigma_b^2 \geqq 0$, $\sigma_{ab}^2 \geqq 0$ and $\sigma_e^2 > 0$ we defer to the usual practice of not explicitly stating the other inequalities when no confusion will result. A similar statement can be made in regard to the alternative hypothesis. In the two-way classification

$$(6.1) \quad \Omega = \{\theta \mid -\infty < \mu < \infty; \quad 0 < \lambda_5 \leqq \lambda_4 \leqq \lambda_2 \leqq \lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 < \infty; \quad \lambda_4 \leqq \lambda_3 \leqq \lambda_1\}$$

$$(6.2) \quad \omega = \{\theta \mid -\infty < \mu < \infty; \quad 0 < \lambda_5 \leqq \lambda_4 = \lambda_2 \leqq \lambda_3 = \lambda_1 < \infty\}$$

where $\theta = (\mu, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$. Since $z_{111} = \sqrt{N}\bar{x}$ it can be seen from (5.5) that a sufficient statistic under ω, for the two-way classification is

$$T = (\bar{X}, S_3, S_5, U) \tag{6.3}$$

where

$$U = S_2 + S_4 \tag{6.4}$$

We first prove the following

THEOREM 6.1: *For the standard distribution of the two-way Model* II *classification* (5.5) *the statistic T defined by* (6.3) *is complete on* ω, *where* ω *is determined by the hypothesis* $\sigma_a^2 = 0$ *and is defined by* (6.2).

PROOF: By the definition of completeness [7] we need to show that

$$\underset{\theta \varepsilon \omega}{E_\theta f(T) = 0}$$

implies $f(t) = 0$, (a.e.). For $\theta \varepsilon \omega$, we have, using (5.9),

$$E_\theta f(T) = c(\theta) \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} f(t) g(t, \lambda_3) h(t, \theta)\, ds_5\, du\, ds_3\, d\bar{x} \tag{6.5}$$

where

$$g(t, \lambda_3) = \exp\left\{ \frac{-N\bar{x}^2}{2\lambda_3} \right\} s_3^{\frac{\nu_3 - 2}{2}} u^{\frac{\nu_2+\nu_4-2}{2}} s_5^{\frac{\nu_5-2}{2}} \tag{6.6}$$

and

$$h(t, \theta) = \exp\left\{ \frac{N\mu\bar{x}}{\lambda_3} - \frac{s_3}{2\lambda_3} - \frac{u}{2\lambda_4} - \frac{s_5}{2\lambda_5} \right\}. \tag{6.7}$$

Let $S_3^* = S_3 + N\bar{X}^2$ and $T^* = (\bar{X}, S_3^*, S_5, U)$. Changing the variable of integration in (6.5) to t^* one gets for $\theta \varepsilon \omega$,

$$\begin{aligned} E_\theta f(T) = c(\theta) \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} & f^*(t^*) g^*(t^*) \\ & \exp\left\{ \frac{N\mu\bar{x}}{\lambda_3} - \frac{s_3^*}{2\lambda_3} - \frac{u}{2\lambda_4} - \frac{s_5}{2\lambda_5} \right\} ds_5\, du\, ds_3^*\, d\bar{x} \end{aligned} \tag{6.8}$$

where

$$f^*(t^*) = \begin{cases} f(t) & \text{if } s_3^* > N\bar{x}^2 \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

and

$$g^*(t^*) = (s_3^* - N\bar{x}^2)^{\frac{\nu_3-2}{2}} u^{\frac{\nu_2+\nu_4-2}{2}} s_5^{\frac{\nu_5-2}{2}}. \tag{6.10}$$

By the unicity property of the quadruple Laplace transform, (6.8) is identically zero for θ in a non-degenerate interval only if

$$f(t^*) g^*(t^*) \equiv 0 \quad \text{(a.e.)}. \tag{6.11}$$

Now $g^*(t^*) \not\equiv 0$ (a.e.). Thus (6.11) and (6.9) imply that $f(t) \equiv 0$ (a.e.) and the theorem is proved.

Let V be defined by the following 1:1 transformation

$$U = S_2 + S_4 \qquad\qquad S_2 = UV$$
$$\text{(6.12)} \qquad\qquad \text{or}$$
$$V = S_2/(S_2 + S_4) \qquad\qquad S_4 = U(1 - V).$$

Since, as can be seen from (5.5), $(\bar{X}, S_2, S_3, S_4, S_5)$ is sufficient under Ω then $W = (T, V)$ is also. Using (6.12) and the density of (S_2, S_4) given by (5.9) we have for the density of (U, V),

$$\text{(6.13)} \qquad p_\theta(u, v) = c(\theta)u^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)} e^{-\frac{u}{2\lambda_4}}, \qquad \theta \in \Omega.$$

But under ω, $\lambda_2 = \lambda_4$ and (6.13) becomes for $\theta_0 \in \omega$

$$\text{(6.14)} \qquad p_{\theta_0}(u, v) = c(\theta_0)u^{\frac{\nu_2+\nu_4-2}{2}} e^{-\frac{u}{2\lambda_4}} v^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}},$$

which shows that U and V are independent under ω. Since (5.9) and (6.4) show clearly that $(\bar{X}, S_3, S_5)$ and (U, V) are always independent this means that T and V are independent under ω and we have

$$\text{(6.15)} \qquad p_\theta(v \mid t) = p_\theta(v) = cv^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}}, \qquad \text{for} \quad \theta \in \omega.$$

Now we are in a position to prove

THEOREM 6.2: *The F-test, which rejects the hypothesis when V is greater than some constant, determines a uniformly most powerful similar region for testing* $\omega{:}\sigma_a^2 = 0$ *against* $\Omega - \omega{:}\sigma_a^2 > 0$.

PROOF: We make use of the fact [7, p. 317] that if T is a sufficient statistic for $\theta \in \omega$, and if T is complete on ω then *all* similar tests of size α,

$$E_\theta \varphi(W) \equiv \alpha, \quad \theta \in \omega, \quad W = (T, V)$$

have Neyman structure [12] with respect to T, i.e. satisfy

$$\text{(6.16)} \qquad \int \varphi(t, v)p_\theta(v \mid t)\, dv \underset{t}{\equiv} \alpha, \qquad \text{(a.e.) for all} \quad \theta \in \omega.$$

Subject to this we wish to maximize the power at a *particular* alternative $\theta_1 \in \Omega - \omega$; that is we desire

$$\int \left\{ \int \varphi(t, v)p_{\theta_1}(v \mid t)\, dv \right\} p_{\theta_1}(t)\, dt = \text{maximum}.$$

Using (6.14) and (6.15) these conditions become

$$\text{(6.17)} \qquad c\int_0^1 \varphi(t, v)v^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}} dv = \alpha$$

and

$$c \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \left\{ \int_0^1 \varphi(t, v) p_{\theta_1}(v \mid t)\, dv \right\} p_{\theta_1}(t)\, dt = \max$$

respectively, where

$$p_{\theta_1}(v \mid t) = \frac{u^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)-\frac{u}{2\lambda_4}}}{p_{\theta_1}(u)}.$$

This will be achieved if for each value of t we have

$$(6.18) \quad c \int_0^1 \varphi(t, v) u^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)-\frac{u}{2\lambda_4}}\, dv = \max,$$

where we recall $t = (\bar{x}, s_3, s_5, u)$. But, finding for fixed t (and *a fortiori* for fixed u) a test $\varphi(t, v)$ satisfying (6.17) and (6.18) is a problem whose solution is given at once by the fundamental Neyman-Pearson lemma to be $\varphi(t, v) = 1$ when

$$cu^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)-\frac{u}{2\lambda_4}} > cv^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}}$$

or

$$e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)} > c(\theta_1, t)$$

or

$$e^{kv} > c(\theta_1, t), \qquad k > 0.$$

or

$$v > c(\theta_1, t).$$

The "constant", $c = c(\theta_1, t)$ is determined by (6.17) or

$$\frac{1}{B\left(\frac{\nu_2}{2}, \frac{\nu_4}{2}\right)} \int_{v>c} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}}\, dv = \alpha.$$

Consequently c is independent of both θ_1 and t,

$$\varphi(t, v) = 1 \quad \text{when} \quad v = \frac{s_2}{s_2 + s_4} > c$$

and the usual F-test is u.m.p. similar.[5]

Of course, Theorem 6.2 was proved only for the balanced two-way classification, but using the standard form in the next section it can be proved in the same way for the balanced one-way classification.

To show where the proof breaks down when applied to testing the hypothesis

[5] In commenting on an earlier draft of this paper, Dr. Werner Gautschi pointed out that in testing $\omega: \sigma_a^2 + \sigma_b^2 = 0$, the T corresponding to (6.3) namely $T = (\bar{X}, S_2 + S_3 + S_4, S_5)$ is complete on ω, but the method of Theorem 6.2 does not seem to help one show that the test based on $(S_2 + S_3)/(S_2 + S_3 + S_4)$ is u.m.p. similar.

of $\omega:\sigma_{ab}^2 = 0$ against $\Omega - \omega:\sigma_{ab}^2 > 0$ we try to prove the analogue of Theorem 6.1. The region Ω is still given by (6.1), but ω is now given by

$$\omega = \{\theta \mid -\infty < \mu < \infty; \quad 0 < \lambda_5 = \lambda_4 \leqq \lambda_2 \leqq \lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 < \infty; \quad \lambda_4 \leqq \lambda_3 \leqq \lambda_1 < \infty\}.$$

Clearly a sufficient statistic under ω is $T = (\bar{X}, S_2, S_3, U')$, where

$$U' = (S_4 + S_5).$$

Now

$$E_\theta f(T) = c(\lambda_2 + \lambda_3 - \lambda_4) \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} f(t) g(t, \lambda_2 + \lambda_3 - \lambda_4) h(t, \theta) \, du' \, ds_2 \, ds_3 \, d\bar{x}$$

where

$$g(t, \lambda_2 + \lambda_3 - \lambda_4) = \exp\left\{ -\frac{N\bar{x}^2}{2(\lambda_2 + \lambda_3 - \lambda_4)} \right\} s_2^{\frac{\nu_2 - 2}{2}} s_3^{\frac{\nu_3 - 2}{2}} (u')^{\frac{\nu_4 + \nu_5 - 2}{2}}$$

and

$$h(t, \theta) = \exp\left\{ \frac{N\mu\bar{x}}{\lambda_2 + \lambda_3 - \lambda_4} - \frac{s_2}{2\lambda_2} - \frac{s_3}{2\lambda_3} - \frac{u'}{2\lambda_4} \right\}.$$

The proof of Theorem 6.1 made use of the fact that

$$\exp\left\{ -\frac{N\bar{x}^2}{2\lambda_3} \right\} \exp\left\{ -\frac{s_3}{2\lambda_3} \right\} = \exp\left\{ -\frac{s_3^*}{2\lambda_3} \right\}$$

for $S_3^* = S_3 + N\bar{X}^2$. This method will not work here because the λ_i associated with the mean, viz. $\lambda_1 \equiv \lambda_2 + \lambda_3 - \lambda_4$, does not equal any other λ_i. However, a lemma due to Gautschi[6] [17] and appearing in this issue of the *Annals* can be used to show completeness under this ω and thus that the F-test of $\sigma_{ab}^2 = 0$ is u.m.p. similar.

7. Likelihood ratio test for the balanced one way classification. We shall show that for the *balanced*, *one-way* classification the likelihood ratio (L.R.) test is not the F-test, but for purposes of significance testing we can act as if it were. Let us consider I populations where the jth measurement on the ith population is given by

$$X_{ij} = \mu + e_i^A + e_{ij}, \tag{7.1}$$

$$i = 1, 2, \cdots, I; \quad j = 1, 2, \cdots, J; \quad IJ = N. \tag{7.2}$$

The usual Model II assumptions are made, namely, that μ is a constant and e_i^A, e_{ij} are normally and independently distributed with zero means and vari-

[6] Dr. Gautschi independently derived the standard form which proved so useful in this work.

ances σ_a^2, σ_e^2 respectively. Let D be the usual orthogonal transformation that transform X_{ij}, suitably ordered, to Z_{ij} which have the standard distribution

$$\frac{|\Lambda|^{-\frac{1}{2}}}{(2\pi)^{N/2}} \exp\left\{ -\frac{1}{2}\left[\frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2} + \frac{s_3}{\lambda_3}\right]\right\}, \tag{7.3}$$

where $S_1 = (Z_{11} - \sqrt{N}\mu)^2$, $S_2 = \sum_{i=2}^{I} Z_{i1}^2$, $S_3 = \sum_{j=2}^{J} \sum_{i=1}^{I} Z_{ij}^2$ and

$$\lambda_1 = \lambda_2 = \sigma_e^2 + J\sigma_a^2 \tag{7.4}$$

$$\lambda_3 = \sigma_e^2 \tag{7.5}$$

$$|\Lambda| = \lambda_1 \lambda_2^{I-1} \lambda_3^{I(J-1)}. \tag{7.6}$$

Clearly

$$\lambda_2 \geqq \lambda_3 > 0. \tag{7.7}$$

To test $H_0 : \sigma_a^2 = 0$ (or $\lambda_2 = \lambda_3$) it is well known that the usual F-test is equivalent to rejecting H_0 if $G > C$ where $G = S_2/S_3$ and C is a constant which depends on the level of significance.

The maximum likelihood (M.L.) estimates, $\hat{\mu}_\Omega$, $\hat{\lambda}_{2\Omega}$, $\hat{\lambda}_{3\Omega}$, are values which maximize the likelihood (7.3) subject to the condition that (7.7) is satisfied by the estimates, i.e.

$$\hat{\lambda}_{2\Omega} \geqq \hat{\lambda}_{3\Omega}. \tag{7.8}$$

Equating to zero the derivatives of the likelihood with respect to μ, λ_2, λ_3 one gets as solutions

$$\hat{\mu}_\Omega = z_{11}/N \tag{7.9}$$

$$\hat{\lambda}'_{2\Omega} = S_2/I \tag{7.10}$$

$$\hat{\lambda}'_{3\Omega} = S_3/[I(J-1)]. \tag{7.11}$$

Since differentiation *may* give, as solutions, values which do not satisfy condition (7.8), these estimates have primes to distinguish them from the "correct" M.L. estimates, which do satisfy (7.8) and are unprimed. That is, if

$$\hat{\lambda}'_{2\Omega} \geqq \hat{\lambda}'_{3\Omega},$$

then (7.10) and (7.11) are the correct M.L. estimates. Because $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$ is equivalent to $G < (J-1)^{-1}$ it remains only to see what the estimates are when (7.10) and (7.11) do not give the "correct" M.L. estimates, i.e., when $G < (J-1)^{-1}$. Since L, the logarithm of the likelihood may be written as a function of λ_2 plus a function of λ_3 it is clear that the values of λ_2 that maximize L, considered as a mathematical function defined for all positive λ_2 and λ_3, rather than as a likelihood (i.e. disregarding the restriction $\lambda_2 \geqq \lambda_3$), for fixed λ_3 is the same λ_2 as is given by (7.10) and similarly for the value of λ_3 that maximizes L for fixed λ_2. Also $\partial L/\partial\lambda_2 \lesseqgtr 0$ or $\partial L/\partial\lambda_3 \lesseqgtr 0$ according as $\lambda_2 \gtrless S_2/I$ or $\lambda_3 \gtrless S_3/[I(J-1)]$. This means that for any fixed λ_2, L decreases as λ_3 moves away from $\hat{\lambda}'_{3\Omega}$ in either direction and similarly for λ_2 and $\hat{\lambda}'_{2\Omega}$ when λ_3 is fixed. Now, by (7.8), the point $(\hat{\lambda}_{2\Omega}, \hat{\lambda}_{3\Omega})$ in the λ_2, λ_3 plane cannot lie above the line $\lambda_3 = \lambda_2$. Suppose it were (strictly) below this line and $(\hat{\lambda}'_{2\Omega}, \hat{\lambda}'_{3\Omega})$ were above the line, i.e. $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$. If $\hat{\lambda}_{3\Omega} < \hat{\lambda}'_{3\Omega}$, then one can increase L by increasing $\hat{\lambda}_{3\Omega}$; if $\hat{\lambda}_{3\Omega} \geqq \hat{\lambda}'_{3\Omega}$, L can be increased by decreasing $\hat{\lambda}_{2\Omega}$. In both of these cases the assumption that L is maximized at $(\hat{\lambda}_{2\Omega}, \hat{\lambda}_{3\Omega})$ is violated. Hence, whenever $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$, the "correct" maximum likelihood estimates are on the line, $\lambda_3 = \lambda_2$, which is the ω region. Thus maximum likelihood corrects negative estimates by making them zero. The maximum likelihood estimates are *then*

$$\hat{\mu}_\omega = \frac{Z_{11}}{N} \tag{7.12}$$

$$\hat{\lambda}_{2\omega} = \hat{\lambda}_{3\omega} = \frac{S_2 + S_3}{IJ}. \tag{7.13}$$

From (7.3) it can be seen that the square of the likelihood ratio is

$$R^2 = \frac{|\hat{\Lambda}_\Omega|}{|\hat{\Lambda}_\omega|} \exp\{z'\hat{\Lambda}_\Omega^{-1}z - z'\hat{\Lambda}_\omega^{-1}z\},$$

where $z' = \{z_{11} - \sqrt{N}\hat{\mu}, z_{21}, z_{31}, \cdots, z_{I1}; z_{12} \cdots z_{IJ}\}$. The subscripts are ordered as in (5.2) and $|{}_\Omega\hat{\Lambda}|$ and $|\hat{\Lambda}_\omega|$ are the maximum likelihood estimates of $|\Lambda|$. Since both $z'\hat{\Lambda}_\Omega^{-1}z$ and $z'\hat{\Lambda}_\omega^{-1}z$ can be shown to equal N, by a procedure given in the next section, $R^2 = |\hat{\Lambda}_\Omega| / |\hat{\Lambda}_\omega|$. By (7.6), this becomes

$$R^2 = \frac{\hat{\lambda}_{2\Omega}^I \hat{\lambda}_{3\Omega}^{I(J-1)}}{\hat{\lambda}_{3\omega}^{IJ}} \tag{7.14}$$

which is unity when $G < (J-1)^{-1} = G_0$, say. For $G \geqq G_0$, (7.9) to (7.13) imply that

$$R^2 = \frac{J^{IJ}}{(J-1)^{I(J-1)}} \frac{S_2^I S_3^{I(J-1)}}{(S_2 + S_3)^{IJ}}, \tag{7.15}$$

whence

$$R^{2/I} = KG\left(\frac{1}{G+1}\right)^J, \qquad K = J^J/[(J-1)^{J-1}] > 0. \tag{7.16}$$

For values of R below one and values of G above $(J-1)^{-1}$, the L.R. test and the G or F test will now be shown to be equivalent. Since low values of λ are significant, to show the equivalence of the two tests for this range of G it is only necessary to show that $R^{2/I}$ is a decreasing function of G or that

$$\frac{d}{dG}\left[G\left(\frac{1}{1+G}\right)^J\right] = \frac{1 + G - JG}{(1+G)^{J+1}} \tag{7.17}$$

is negative. Clearly, (7.17) is negative when $1 + G - JG < 0$ which is equivalent to $G > (J-1)^{-1} = G_0$. Also if $G = G_0$ in (7.16), $R = 1$. We have already seen that $R = 1$ when $G < G_0$. Now, let $\alpha_0 = \Pr\{G > G_0\}$, which is the probability that $R < 1$. Hence $1 - \alpha = \Pr\{R = 1\}$. Thus the atomic positive probability mass at $R = 1$ means that there are no L.R. tests of $\sigma_a^2 = 0$, for the balanced, one-way classification, with level of significance greater than α_0 but less than 1. However, when an L.R. significance test does exist it is the F-test. Since $F = I(J - 1)G/(I - 1)$, $G > G_0$ is equivalent to $F > F_0$, where $F_0 = I/(I - 1)$. For all significance levels up to and including the 25 per cent level [9] the percentage points of F with $(I - 1)$ and $I(J - 1)$ degrees of freedom for finite values of I and J greater than 1 are greater than F_0 while the 50 percentage points are less than F_0 for all these values of I and J. Inasmuch as it is unlikely that one wishes to use a significance level between 25 and 50 per cent, for all *practical* purposes, the F-test and L.R. test are equivalent in the case of the *balanced-one-way* classification.

Although the F-tests of no population effect are the same under Models I and II, this quirk of the L.R. test does not exist in Model I. It is known that then the L.R. test is precisely the F-test. In Model I, the L.R. and the F-statistic are strictly decreasing functions of one another *and* there is no positive probability mass at $R = 1$.

It is of interest to note that there is a modified L.R. test which is equivalent to the F-test for the Model II, balanced, one-way classification. One can reason that if in (7.10) and (7.11) $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$, then the estimate of σ_a^2 as given by (7.4) and (7.5) is negative. Then one way to modify or "correct" the estimates so that the estimate of σ_a^2 is zero, is to use as estimates (although they are no longer M.L.),

$$(7.18) \qquad \hat{\lambda}_{2\Omega c} = \hat{\lambda}_{3\Omega c} = \frac{S_3}{I(J - 1)}.$$

If these are put in (7.14), and if K is a positive constant

$$(7.19) \qquad R^2 = \left(\frac{K}{1 + G}\right)^{IJ}$$

which is a strictly decreasing function of G. Hence, if the estimates given by (7.18) are used when $\hat{\lambda}'_{2\Omega} > \hat{\lambda}'_{3\Omega}$, this modified L.R. test is equivalent to the F-test. Little can be said for this procedure, since the information in S_2 is not used and when the estimate of σ_a^2 is negative one can argue almost as easily, by ignoring the information in S_3, that the corrected estimates should be

$$(7.20) \qquad \hat{\lambda}_{2\Omega c} = \hat{\lambda}_{3\Omega c} = S_2/I.$$

If these are used in (7.14),

$$(7.21) \qquad R^2 = \left(\frac{K}{1 + 1/G}\right)^{IJ}$$

where K is again a positive constant. This is a strictly *increasing* function of G

and the modified test using it is certainly not equivalent to an F-test with large values significant.

8. Likelihood ratio test for the balanced two-way classification. This section will be devoted to showing by means of a counter-example that when testing $\omega:\sigma_a^2 = 0$ in the two-way classification, not only is the L.R. test not the F-test, but (unlike the balanced one-way classification) is not even equivalent to it for small levels of significance. The L.R. test is a function of S_2, S_3 and S_4, while the F-test is a function of only S_2 and S_4. From (5.5) the logarithm of the likelihood for $\theta \varepsilon \Omega$, Ω given by (6.1), is

$$L_\theta = -\frac{N}{2}\log 2\pi - \frac{1}{2}\left\{\log|\Lambda| + \sum_{i=1}^{5}\frac{s_i}{\lambda_i}\right\} \tag{8.1}$$

where $|\Lambda| = \lambda_1\lambda_2^{\nu_2}\lambda_3^{\nu_3}\lambda_4^{\nu_4}\lambda_5^{\nu_5}$. The s_i are defined by (5.6) and the λ's by (3.3). Recall that for all $\theta \varepsilon \Omega$

$$\lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 \tag{8.2}$$

and

$$\lambda_1 \geqq \lambda_2 \geqq \lambda_4; \qquad \lambda_1 \geqq \lambda_3 \geqq \lambda_4; \qquad \lambda_4 \geqq \lambda_5 > 0. \tag{8.3}$$

Rather than maximize L_θ subject to (8.2) we shall use a more general side condition, use of which will be made below, namely to maximize L_θ subject to $\sum_1^5 b_i\lambda_i = 0$ and $\sum_1^5 c_i\lambda_i = 0$ by making use of Lagrange multipliers $\beta/2$ and $\gamma/2$. Let

$$M = L_\theta + \frac{\beta}{2}\sum_{i=1}^{5} b_i\lambda_i + \frac{\gamma}{2}\sum_{i=1}^{5} c_i\lambda_i.$$

Equating to zero the derivative of M with respect to β, γ, μ, λ_i, $(i = 1, 2, \cdots, 5)$ one obtains

$$\hat{\mu} = Z_{111}/\sqrt{N} \tag{8.4}$$

$$-\nu_i + \frac{S_i}{\hat{\lambda}_i} + \beta b_i\hat{\lambda}_i + \gamma c_i\hat{\lambda}_i = 0, \qquad i = 1, 2, \cdots, 5 \tag{8.5}$$

$$\sum_1^5 b_i\hat{\lambda}_i = 0, \qquad \sum_1^5 c_i\hat{\lambda}_i = 0, \tag{8.6}$$

where the carats indicate that these are the maximizing values. Adding the five equations in (8.5) and making use of (8.6) we obtain

$$\sum_{i=1}^{5}\frac{S_i}{\hat{\lambda}_i} = \sum_{i=1}^{5}\nu_i = N,$$

and the exponent in (5.5) is $-N/2$ when the maximizing values of the parameters are used. Thus the well-known result [18] when there is no condition on the λ's is also true if the λ's are linearly dependent.

Under Ω, $b_1 = b_4 = 1$, $b_2 = b_3 = -1$, $b_5 = 0$, $c_i = 0$, $(i = 1, 2, \cdots, 5)$ and (8.4)–(8.6) become

$$\hat{\mu}_\Omega = Z_{111}/\sqrt{N} \tag{8.7}$$

$$\begin{aligned} \hat{\lambda}_{1\Omega}^{-1} - \beta &= 0 \\ \nu_2\hat{\lambda}_{2\Omega}^{-1} - S_2\hat{\lambda}_{2\Omega}^{-2} + \beta &= 0 \\ \nu_3\hat{\lambda}_{3\Omega}^{-1} - S_3\hat{\lambda}_{3\Omega}^{-2} + \beta &= 0 \\ \nu_4\hat{\lambda}_{4\Omega}^{-1} - S_4\hat{\lambda}_{4\Omega}^{-2} - \beta &= 0 \end{aligned} \tag{8.8}$$

$$\hat{\lambda}_{5\Omega} = S_5/\nu_5 . \tag{8.9}$$

If the solutions of (8.2) and (8.8) satisfy (8.3), they are also M.L. estimates. If not then one would have to get the "correct" M.L. estimates by some procedure similar to the one used in the previous section. This will be unnecessary because we shall show that even when these solutions satisfy (8.3) the L.R. statistic is not a function of F alone. Hereafter we confine ourselves to the part of the z space where (8.3) is satisfied by the stationary values. Eliminating the Lagrange multiplier, and performing some simplifications one may write (8.2) and (8.8) as

$$\begin{aligned} \hat{\lambda}_{1\Omega} &= \hat{\lambda}_{2\Omega} + \hat{\lambda}_{3\Omega} - \hat{\lambda}_{4\Omega} \\ S_2\hat{\lambda}_{2\Omega}^{-1} &= \nu_2 + \hat{\lambda}_{2\Omega}\hat{\lambda}_{1\Omega}^{-1} \\ S_3\hat{\lambda}_{3\Omega}^{-1} &= \nu_3 + \hat{\lambda}_{3\Omega}\hat{\lambda}_{1\Omega}^{-1} \\ S_4\hat{\lambda}_{4\Omega}^{-1} &= \nu_4 - \hat{\lambda}_{4\Omega}\hat{\lambda}_{1\Omega}^{-1} \end{aligned} \tag{8.10}$$

Similarly, the logarithm of the likelihood under $\omega: \sigma_a^2 = 0$ or $\lambda_1 = \lambda_3$, $\lambda_2 = \lambda_4$ is given by (8.1) subject to

$$b_1 = b_4 = 1, \quad b_2 = b_3 = -1, \quad b_5 = 0$$

$$c_2 = 1, \quad c_4 = -1, \quad c_1 = c_3 = c_5 = 0.$$

Using this last condition (8.4)–(8.6) can be simplified to

$$\begin{aligned} \hat{\mu}_\omega &= Z_{111}/\sqrt{N} \\ \hat{\lambda}_{3\omega} &= S_3/I \\ \hat{\lambda}_{4\omega} &= (S_2 + S_4)/I(J - 1) \\ \hat{\lambda}_{5\omega} &= S_5/IJ(K - 1). \end{aligned} \tag{8.11}$$

As in the estimate under Ω we treated only the case when the stationary values satisfy $\hat{\lambda}_{3\omega} \geqq \hat{\lambda}_{4\omega} \geqq \hat{\lambda}_{5\omega} > 0$.

Since the exponent of the likelihood when the estimates under Ω *or* ω are inserted has been shown above to equal $-N/2$, the square of the likelihood ratio is

$$R^2 = \frac{|\hat{\Lambda}_\Omega|}{|\hat{\Lambda}_\omega|} = \frac{L'_\Omega}{L'_\omega} \tag{8.12}$$

where

$$L'_\Omega = (\hat{\lambda}_{2\Omega} + \hat{\lambda}_{3\Omega} - \hat{\lambda}_{4\Omega})\hat{\lambda}_{2\Omega}^{\nu_2}\hat{\lambda}_{3\Omega}^{\nu_3}\hat{\lambda}_{4\Omega}^{\nu_4} \tag{8.13}$$

$$L'_\omega = (S_3/I)^I[(S_2 + S_4)/I\nu_2]^{I\nu_2} \tag{8.14}$$

and $\hat{\lambda}_{i\Omega}$, $i = 2, 3, 4$ satisfy (8.10). We have seen that the F-test of $\omega:\sigma_a^2 = 0$ is a function of S_2 and S_4 alone and does not depend on S_3. It appears that R^2 *may* depend on S_3 since its denominator, L'_ω does. However L'_Ω *may* equal S_3^I times a factor independent of S_3, in which case R^2 will be independent of S_3. It was shown [3], by comparing the solutions, (8.12) for two examples differing only in values for s_3, that R^2 *does* depend on S_3.

In Section 10 it will be shown that both the L.R. and F-tests are invariant tests, but the F-test is to be preferred since it has an optimum property, namely, of being the u.m.p. similar test.

9. Uniformly most powerful invariant test in the balanced, one-way classification is the F-test. It will be shown that for the balanced, one-way classification, when the standard variable Z has distribution (7.3) the u.m.p. invariant test of $\omega:\sigma_a^2 = 0$ against $\Omega - \omega:\sigma_a^2 > 0$ (or using (7.4) and (7.5) of $\omega:\theta = 1$ against $\Omega - \omega:\theta > 1$ where $\theta = \lambda_2/\lambda_3$) is the F-test. We partition the Z vector as follows. Let $Z' = [Z_{(1)}, Z'_{(2)}, Z'_{(3)}]$ where $Z_{(1)} = Z_{11}$, $Z_{(2)}$ is the column vector whose elements are Z_{i1}, $i = 2, 3, \cdots, I$ and $Z_{(3)}$ is the column vector whose elements are Z_{ij}; $i = 1, 2, \cdots, I$; $j = 2, 3, \cdots, J$. The elements of $Z_{(2)}$ and $Z_{(3)}$ may be ordered in any way. Clearly the problem remains invariant under the following groups of transformations, each of which is a normal subgroup of the product group of the previous ones:

$$Z^*_{(1)} = Z_{(1)} + c, \qquad Z^*_{(\alpha)} = Z_{(\alpha)}, \quad \alpha = 2, 3. \tag{9.1}$$

$$Z^*_{(\alpha)} = D_{(\alpha)}Z_{(\alpha)}; \qquad D_{(\alpha)} \text{ orthogonal}, \quad \alpha = 1, 2, 3 \tag{9.2}$$

$$Z^*_{(\alpha)} = cZ_{(\alpha)}; \qquad \alpha = 1, 2, 3; \quad c \neq 0. \tag{9.3}$$

A maximal invariant [8] under the product of all three groups is

$$G = \left(\sum_{i=2}^I Z_{i1}^2\right)/\left(\sum_{j=2}^J \sum_{i=1}^I Z_{ij}^2\right) = \frac{S_2}{S_3}.$$

It may be pointed out that unlike Model I the group of orthogonal transformations is unnecessary if we agree to base all decisions on the *sufficient* statistic (Z_{11}, S_2, S_3) of Section 7. Starting with this statistic the first and third group of transformations (additive and multiplicative group) will lead to G as a maximal invariant in the class of sufficient statistics.

To show that the test which determines the critical region $G > c$ (or the equivalent F-test which rejects ω when $W = \nu_3 G/\nu_2 > c$) is the u.m.p. invariant test one need only show it is the u.m.p. test based on G. Under ω, W is distributed as F with ν_2 and ν_3 degrees of freedom, while, under $\Omega - \omega$, it is distributed as θ times F with ν_2 and ν_3 degrees of freedom, i.e. the probability density of G is

$$(9.4) \qquad p_\theta(g) = c\theta^{\frac{\nu_3}{2}} g^{\frac{\nu_2 - 2}{2}} (\theta + g)^{-\left(\frac{\nu_2+\nu_3}{2}\right)}, \qquad \theta \geqq 1,$$

where $\theta = \lambda_2/\lambda_3$. By the Neyman-Pearson lemma the most powerful test of $\theta = 1$ based on G against a particular alternative $\theta = \theta_0 > 1$ is given by $\varphi(g) = 1$ when

$$\theta_0^{\frac{\nu_3}{2}} \left(\frac{1+g}{\theta_0 + g}\right)^{\frac{\nu_2+\nu_3}{2}} > c$$

or

$$(9.5) \qquad \frac{1+g}{\theta_0 + g} > c.$$

The left member of (9.5) is an increasing function of g, since its derivative with respect to g is $(\theta_0 - 1)/(\theta_0 + g)^2$ which is positive. Hence this test is equivalent to $\varphi(g) = 1$ when $g > c$. Since the value of c is determined by integrating the upper tail of (9.4) for $\theta = 1$, it is not dependent on the particular alternative. Thus for the balanced one-way classification one may replace the class of similar tests by the somewhat more reasonable class of invariant tests and show that in this more reasonable class the usual F-test of $\sigma_a^2 = 0$ against $\sigma_a^2 > 0$ is also u.m.p.

10. Invariance in the balanced two-way classification. It will now be shown why there *may* not be *any* uniformly most powerful invariant test in the case of the balanced two-way classification. We are interested in the test of $\omega: \sigma_a^2 = 0$ against $\Omega - \omega: \sigma_a^2 > 0$ (or, if we let $\psi_1 = \lambda_2/\lambda_4$, of testing $\omega: \psi_1 = 1$ against $\Omega - \omega: \psi_1 > 1$) for the standard variate Z whose distribution is given by (5.5). The group of transformations analogous to those in the last section will be considered. As in that section we partition the Z vector thus:

$$Z' = [Z_{(1)}, Z'_{(2)}, Z'_{(3)}, Z'_{(4)}, Z'_{(5)}],$$

where $Z_{(1)} = Z_{111}$ and $Z_{(\alpha)}$, for $\alpha = 2, 3, 4, 5$, is the column vector of the Z's (in any order) appearing in the sums S_α of (5.6). The problem remains invariant under the same types of groups of transformations as in the preceding section, namely (9.1) for $\alpha = 2, 3, 4, 5$ and (9.2), (9.3) for $\alpha = 1, \cdots, 5$. A maximal invariant under the product group of the three groups, is U, V, W where

$$(10.1) \qquad U = S_2/S_4, \quad V = S_3/S_4, \quad W = S_4/S_5.$$

Any test based on U, V, W will have power based on the maximal invariant induced in the parameter space, namely,

$$(10.2) \qquad \psi = (\psi_1, \psi_2, \psi_3),$$

where

$$\psi_1 = \lambda_2/\lambda_4, \qquad \psi_2 = \lambda_3/\lambda_4, \qquad \psi_3 = \lambda_4/\lambda_5. \tag{10.3}$$

As in the balanced one-way classification (Section 9), the orthogonal group of transformations corresponding to (9.2) is unnecessary if we agree to base all decisions on the sufficient statistic $(Z_{111}, S_2, S_3, S_4, S_5)$ of Section 5.

By transforming the density of (S_2, S_3, S_4, S_5) as given in (5.9) to that of (S_5, U, V, W) and integrating out s_5 we obtain [3]

$$p_\psi(u, v, w) = \frac{\Gamma\left(\frac{\nu_2 + \nu_3 + \nu_4 + \nu_5}{2}\right)}{\psi_1^{\frac{\nu_2}{2}}\psi_2^{\frac{\nu_3}{2}}\psi_3^{\frac{\nu_2+\nu_3+\nu_4}{2}}} \frac{u^{\frac{\nu_2-2}{2}} v^{\frac{\nu_3-2}{2}} w^{\frac{\nu_2+\nu_3+\nu_4-2}{2}}}{\delta^{\frac{\nu_2+\nu_3+\nu_4+\nu_5}{2}}} \tag{10.4}$$

where

$$\delta = \delta(u, v, w; \psi) = \frac{uw}{\psi_1\psi_3} + \frac{w}{\psi_3} + \frac{vw}{\psi_2\psi_3} + 1 \tag{10.5}$$

This shows that the density of u, v, w is indeed dependent on λ only through the maximal invariant $\psi = (\psi_1, \psi_2, \psi_3)$. The Neyman-Pearson lemma gives as the most powerful test of $H_0:\lambda = \lambda^0$ against $H_1:\lambda = \lambda^1$ (where $\lambda^i = (\lambda_2^i, \lambda_3^i, \lambda_4^i, \lambda_5^i)$, $\lambda_2^i/\lambda_4^i = \psi_1^i$, $i = 0, 1$ and $\psi_1^0 = 1 < \psi_1^1$), based on (U, V, W) the one which rejects H_0 when

$$\frac{p_{\psi^1}(u, v, w)}{p_{\psi^0}(u, v, w)} = \left[\frac{\frac{w(u+1)}{\psi_3^0} + \frac{vw}{\psi_2^0\psi_3^0} + 1}{\frac{uw}{\psi_1^1\psi_3^1} + \frac{w}{\psi_3^1} + \frac{vw}{\psi_2^1\psi_3^1} + 1}\right]^{\frac{\nu_2+\nu_3+\nu_4+\nu_5}{2}} \cdot \left(\frac{\psi_1^0}{\psi_1^1}\right)^{\frac{\nu_2}{2}} \left(\frac{\psi_2^0}{\psi_2^1}\right)^{\frac{\nu_3}{2}} \left(\frac{\psi_3^0}{\psi_3^1}\right)^{\frac{\nu_2+\nu_3+\nu_4}{2}} > c. \tag{10.6}$$

Since the distribution of V and W depend on ψ_2^0 and ψ_3^0 under H_0 there seems to be little likelihood of obtaining a *uniformly* most powerful invariant test based on a statistic involving U, V and W from (10.6). It was not obvious from the fact that the maximal invariant was vector valued that no u.m.p. invariant test exists, since conceivably (10.6) might involve only one of the elements of the vector. For example if (10.6) were a function of U alone then once again the usual F-test would be uniformly most powerful. Although our probability ratio, (10.6) showed that there is no u.m.p. invariant test based on the given product group of transformations, there still may be one with respect to a larger group of transformations. For example, if in the last section we had stopped after the second group of transformations obtaining as a maximal invariant S_2, S_3 (rather than S_2/S_3) a situation analogous to (10.6) would have resulted. This may mean that another group of transformations, unknown to the author, may leave the problems invariant in the case of the balanced two-way classification and the maximal invariant under the product of the four groups is U.

Even if there are no further invariant transformations, an optimum test in this case can be obtained by *decreasing* the class of invariant tests. We have seen that a maximal invariant under G is the vector consisting of any three independent ratios of S_2, S_3, S_4, S_5 and that G induced a group $\bar{G}$ under which a maximal invariant in the parameter space is the vector composed of the corresponding three ratios of λ_2, λ_3, λ_4, λ_5. But $\psi_1 = \lambda_2/\lambda_4$ (or its reciprocal) seems to be the only one that is independent of nuisance parameters under ω. Also S_2/S_4 (or its reciprocal) appears to be the only part of the maximal invariant under G whose distribution is a function of ψ_1 only. Thus, *it seems reasonable* to restrict our class to S_2/S_4. Then we obtain a u.m.p. test as in the last section. We now show

THEOREM 10.1: *Of all invariant tests of* ω: $\sigma_a^2 = 0$ *against* $\Omega - \omega: \sigma_a^2 > 0$ *in the balanded two-way classification whose power is a function of* ψ_1 *only, the usual* F*-test is most powerful.*

PROOF: *If* it can be shown that S_2/S_4 is the only invariant statistic whose power is a function of ψ only, the above assertion is true. However we have already shown a stronger result in Section 6, which includes this result, namely, the usual F-test is the u.m.p. similar test. Similarity in this example means

$$E_\theta \varphi(X) = \alpha, \qquad \theta \in \omega \text{ (i.e. } \psi_1 = 1),$$

while we want our test to satisfy

$$\begin{aligned} E_\psi \varphi(X) &= \text{const} = \alpha, \text{ say} \quad \text{for } \psi \in \omega \text{ (i.e. } \psi_1 = 1) \\ &= f(\psi_1) \qquad \psi \in \Omega - \omega \text{ (i.e. } \psi_1 > 1) \\ X &= h(U, V, W). \end{aligned}$$

By $\psi \in \omega$ we mean that the components of ψ satisfy (6.2). There clearly is a similar test for every invariant test which is a function of ψ_1 only. Since the u.m.p. similar test is based on U, an invariant statistic, Theorem 10.1 is proved. Of course invariance added nothing in this case.

11. Balanced multi-way classifications. The procedure of Section 5 can be used to obtain the standard form for any balanced multi-way classification. The evaluation of $| \mathfrak{T} - \lambda \mathfrak{I} |$ just becomes a little more tedious as the number of factors increases. Of special interest is the case of the multi-fold, hierarchical or nested classification [6] model which is very useful in survey sampling theory [1]. The three-fold classification may be represented by

$$X_{ijkm} = \mu + e_i^A + e_{ij}^{AB} + e_{ijk}^{ABC} + e_{ijkm}$$

with μ a constant, e_i^A, e_{ij}^{AB}, e_{ijk}^{ABC}, e_{ijkm}, normally and independently distributed with means zero and variances σ_a^2, σ_{ab}^2, σ_{abc}^2, σ_e^2 and the range of subscripts as usual. In this special case the hypothesis that any variance component, except σ_e^2, equals zero can be tested by an F-test and the method of Section 6 can be used to show these tests are u.m.p. similar. Even in this special case the methods of Section 9 cannot be used to show u.m.p. invariance unless the multi-fold classification is one-fold, which is the same as the one-way case treated in Section 9. However, Gautschi's [17] lemma must be used to prove that the usual F-tests are u.m.p. similar in the *non-hierarchical* multi-way classifications, when there are more than two classifications.

12. Acknowledgements. The author is indebted to many people, especially to Drs. H. Scheffé, H. Raiffa, T. W. Anderson and Erich Lehmann for helpful suggestions and to Dr. H. Scheffé for suggesting the problem that led to the paper.

REFERENCES

[1] W. G. Cochran, *Sampling Techniques*, John Wiley and Sons, New York, 1953.
[2] S. L. Crump, "The estimation of components of variance in multiple classifications", unpublished Ph.D. thesis, Ames, Iowa, Iowa State College Library, 1947.
[3] L. H. Herbach, "Optimum properties of analysis of variance tests based on Model II and some generalizations of Model II", Scientific Paper No. 6, Engineering Statistics Laboratory, New York University, 1957 (Ph.D. Dissertation, Columbia University).
[4] P. L. Hsu, "Analysis of variance from the power function standpoint", *Biometrika*, Vol. 32 (1941), p. 62.
[5] P. L. Hsu, "On the power function of the E^2 test and the T^2 test", *Ann. Math. Stat.*, Vol. 16 (1945), p. 278.
[6] O. Kempthorne, *The Design and Analysis of Experiments*, John Wiley and Sons, New York, 1952.
[7] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation, Part I", *Sankhya*, Vol. 10 (1950), pp. 305–340.
[8] E. L. Lehmann, "Some principles of the theory of testing hypotheses", *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[9] M. Merrington and C. M. Thompson, "Tables of percentage points of the inverted beta (F) distribution", *Biometrika*, Vol. 33 (1943), pp. 73–88.
[10] A. M. Mood, *Introduction to the Theory of Statistics*, McGraw Hill, New York, 1950.
[11] J. A. Nelder, "The interpretation of negative components of variance", *Biometrika*, Vol. 41 (1954), pp. 544–548.
[12] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability", *Phila. Trans. Roy. Soc. London, Ser. A*, Vol. 236 (1937), pp. 333–380.
[13] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use", *Stat. Res. Memoirs*, Vol. 2 (1938), p. 126.
[14] A. Wald, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), p. 434.
[15] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, New Jersey, 1947.
[16] J. Wolfowitz, "The power of the classical tests associated with the normal distribution," *Ann. Math. Stat.*, Vol. 20 (1949), p. 540.
[17] W. Gautschi, "Some remarks on Herbach's paper, 'Optimum nature of the F-test for Model II in the balanced case,'" *Ann. Math. Stat.*, Vol. 30 (1959), pp. 960–963.
[18] T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, John Wiley, New York, 1958, p. 47.