COMPONENTS OF VARIANCE ANALYSIS FOR PROPORTIONAL FREQUENCIES

By J. D. Bankier and R. E. Walpole

McMaster University and Virginia Polytechnic Institute

1. Summary. With the exception of papers by G. W. Snedecor, G. M. Cox, and H. F. Smith ([8], [9], [10]), there seems to be little about proportional frequencies in the literature. In this paper we consider two-way crossed classifications and two-way nested classifications. The expected values of the sums of squares are obtained in a form which is applicable to a variety of components of variance models. The tests of several hypotheses are considered.

2. The Type I model for two-way crossed classifications. We consider an experiment in which p treatments are applied to q blocks. The ith treatment is applied to the jth block n_{ij} times. The n_{ij}'s having been displayed in a matrix with n_{ij} in the ith row and jth column, we assume that the n_{ij}'s in a given row are proportional to the n_{ij}'s in any other row. This implies that

$$(1) \qquad n_{ij} = \frac{n_{i.}n_{.j}}{N},$$

where

$$n_{i.} = \sum_{j=1}^{q} n_{ij}, \qquad n_{.j} = \sum_{i=1}^{p} n_{ij}, N = \sum_{i=1}^{p} n_{i.}.$$

Consider the model

$$(2) \quad Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk_{ij}}, \quad i = 1, 2, \cdots, p; j = 1, 2, \cdots, q;$$

$k_{ij} = 1, 2, \cdots, n_{ij}$, where the $\epsilon_{ijk_{ij}}$'s are NID $(0, \sigma^2)$ and the parameters are subject to the conditions

$$(3) \qquad \sum_{i=1}^{p} n_{i.}\tau_i = \sum_{j=1}^{q} n_{.j}\beta_j = \sum_{i=1}^{p} n_{i.}(\tau\beta)_{ij} = \sum_{j=1}^{q} n_{.j}(\tau\beta)_{ij} = 0.$$

If we denote $E(Y_{ijk_{ij}})$ by ξ_{ij}, the above conditions are equivalent to defining

$$\mu = \bar{\xi}_{..}, \qquad \tau_i = \bar{\xi}_{i.} - \bar{\xi}_{..}, \qquad \beta_j = \bar{\xi}_{.j} - \bar{\xi}_{..}, \qquad (\tau\beta)_{ij} = \bar{\xi}_{ij} - \bar{\xi}_{i.} - \bar{\xi}_{.j} + \bar{\xi}_{..},$$

where

$$\bar{\xi}_{i.} = \frac{1}{n_{i.}} \sum_{j=1}^{q} n_{ij}\xi_{ij}, \qquad \bar{\xi}_{.j} = \frac{1}{n_{.j}} \sum_{i=1}^{p} n_{ij}\xi_{ij}, \qquad \bar{\xi}_{..} = \frac{1}{N} \sum_{i,j}^{p,q} n_{ij}\xi_{ij}.$$

A more realistic model, such as is considered by Anderson and Bancroft [1], will be studied in a later section.

We now rewrite Eqs. (2) in a form where the theory given by Anderson and

Received August 14, 1956, revised February 25, 1957.

Bancroft [1] may be applied. They may be put in the form

$$(4) \qquad Y_{ijk_{ij}} = \mu + \sum_{i'=1}^{p} U_{i'i}\tau_{i'} + \sum_{j'=1}^{q} V_{j'j}\beta_{j'} + \sum_{i',j'}^{p,q} W_{i'j'ij}(\tau\beta)_{i'j'} + \epsilon_{ijk_{ij}},$$

where

$$U_{i'i} = \delta_{i'i}, \qquad V_{j'j} = \delta_{j'j}, \qquad W_{i'j'ij} = \delta_{i'i}\delta_{j'j},$$

δ_{ij} being the Kronecker δ. If we order the $Y_{ijk_{ij}}$, calling them $Y_\alpha(\alpha = 1, 2, \cdots, N)$, we may write Eqs. (4) in the vector form

$$(5) \qquad Y = \mu + \sum_{i=1}^{p} U_i\tau_i + \sum_{j=1}^{q} V_j\beta_j + \sum_{i,j}^{p,q} W_{ij}(\tau\beta)_{ij} + \epsilon.$$

Denoting the elements of the vector U_i by $U_{i\alpha}$, we define

$$\bar{U}_i = \frac{1}{N}\sum_{\alpha=1}^{N} U_{i\alpha} = \frac{n_{i.}}{N}, \qquad u_{i\alpha} = U_{i\alpha} - \bar{U}_i$$

so that

$$0 = \sum_{\alpha=1}^{N} u_{i\alpha} = \sum_{i'=1}^{p} n_{i'.}u_{ii'}.$$

Similarly,

$$\bar{V}_j = \frac{n_{.j}}{N}, \qquad \bar{W}_{ij} = \frac{n_{ij}}{N}, \qquad \sum_{j'=1}^{q} n_{.j'}v_{jj'} = \sum_{i',j'}^{p,q} n_{i'j'}w_{iji'j'} = 0.$$

Changing our notation, we denote by $\bar{U}_i$, $\bar{V}_j$, $\bar{W}_{ij}$, the vectors $\bar{U}_iI$, $\bar{V}_jI$, $\bar{W}_{ij}I$, where I is a column vector all of those N elements are equal to unity. Then, if we set

$$u_i = U_i - \bar{U}_i, \qquad v_j = V_j - \bar{V}_j, \qquad w_{ij} = W_{ij} - \bar{W}_{ij},$$

we may write Eq. (5) in the form

$$(6) \qquad Y = \mu + \sum_{i=1}^{p} u_i\tau_i + \sum_{j=1}^{q} v_j\beta_j + \sum_{i,j}^{p,q} w_{ij}(\tau\beta)_{ij} + \epsilon.$$

It is necessary that the u_i's, v_j's, and w_{ij}'s form a linearly independent set of vectors. Since this is not the case, we use the conditions (3) to eliminate τ_p, β_q, $(\tau\beta)_{iq}(i = 1, 2, \cdots, p)$ and $(\tau\beta)_{pj}(j = 1, 2, \cdots, q-1)$, obtaining

$$(7) \qquad \begin{aligned} Y = \mu &+ \sum_{i=1}^{p-1}\left(u_i - \frac{n_{i.}}{n_{p.}}u_p\right)\tau_i + \sum_{j=1}^{q}\left(v_j - \frac{n_{.j}}{n_{.q}}v_q\right)\beta_j \\ &+ \sum_{i=1}^{p-1}\sum_{j=1}^{q-1} n_{ij}\left(\frac{w_{ij}}{n_{ij}} - \frac{w_{iq}}{n_{iq}} - \frac{w_{pj}}{n_{pj}} + \frac{w_{pq}}{n_{pq}}\right)(\tau\beta)_{ij} + \epsilon. \end{aligned}$$

We note that

$$u_i - \frac{n_{i.}}{n_{p.}}u_p = U_i - \frac{n_{i.}}{n_{p.}}U_p$$

and a similar statement may be made about the coefficient vectors of β_j and $(\tau\beta)_{ij}$. Making use of the relations

$$U_i.U_{i'} = n_{i.}\delta_{ii'}, \quad U_i.V_j = n_{ij}, \quad U_i.W_{i'j} = n_{ij}\delta_{ii'}, \quad V_j.V_{j'} = n_{.j}\delta_{jj'},$$
$$V_j.W_{ij'} = n_{ij'}\delta_{jj'}, \quad W_{ij}.W_{i'j'} = n_{ij}\delta_{ii'}\delta_{jj'},$$

it may be proved that the coefficient vectors of the τ's, β_j's, $(\tau\beta)_{ij}$'s form three sets of linearly independent vectors and a vector from any set is orthogonal to the vectors of the other two sets. Thus, when the three sets are combined, they form a set of linearly independent vectors.

We shall be interested in testing three hypotheses

$$H_1: (\tau\beta)_{ij} = 0, \qquad i = 1, 2, \cdots, p-1; j = 1, 2, \cdots, q-1,$$
$$H_2: \tau_i = 0, \qquad i = 1, 2, \cdots, p-1,$$
$$H_3: \beta_j = 0, \qquad j = 1, 2, \cdots, q-1.$$

The restrictions (3) imply that not only the parameters in a given hypothesis are zero but also all other parameters of the same kind. To test H_1, we first compute

$$SSE = \sum_{i,j,k_{ij}} [Y_{ijk_{ij}} - m - t_i - b_j - (tb)_{ij}]^2,$$

where m, t_i, b_j and $(tb)_{ij}$ are the least squares estimates of μ, τ_i, β_j, and $(\tau\beta)_{ij}$, respectively, and SSE is the minimized value of the residual sum of squares. Next, we compute SSE_1, the corresponding minimum obtained under the assumption that H_1 holds. Then

$$R = \sum_{\alpha=1}^{N} y_\alpha^2 - SSE,$$

where $y_\alpha = Y_\alpha - \bar{Y}$, is the reduction in the sum of squares when all the parameters are used while

$$R_1 = \sum_{\alpha=1}^{N} y_\alpha^2 - SSE_1$$

is the reduction due to the parameters left when H_1 is true. The additional reduction in the sum of squares due to the $(\tau\beta)_{ij}$'s is

$$SS(TB) = R - R_1 = SSE_1 - SSE.$$

In the same way, SSE_2 and SSE_3 denote the minima obtained subject to H_2 and H_3, respectively, and the reductions in the sum of squares due to the τ_i's and the β_j's are

$$SST = SSE_2 - SSE \quad \text{and} \quad SSB = SSE_3 - SSE,$$

respectively. Anderson and Bancroft [1] show that

$$\sum_{\alpha=1}^{N} y_\alpha^2 = SST + SSB + SS(TB) + SSE,$$

and that, subject to the corresponding hypotheses, SST, SSB, $SS(TB)$ and SSE are independently distributed as $\chi^2\sigma^2$ with $p - 1$, $q - 1$, $(p - 1)(q - 1)$, and $N - pq$ degrees of freedom. The hypotheses H_1, H_2, and H_3 are tested by the statistics

$$F_1 = \frac{MS(TB)}{MSE}, \qquad F_2 = \frac{MST}{MSE}, \qquad F_3 = \frac{MSB}{MSE},$$

respectively, where MSE, for example, is SSE divided by the corresponding number of degrees of freedom.

3. The sums of squares. The following theorem, a slight generalization of one stated by Mann [5], will be used in computing the sums of squares.

THEOREM A. *If*

$$E(Y) = \mu I + \sum_{k=1}^{p} X_k \tau_k$$

and

(1) $I = \sum_{k=1}^{s} X_k, \qquad s \leqq p,$

(2) $X_1, X_2, \cdots, X_s$ *form a mutually orthogonal set of vectors,*

(3) $\sum_{k=1}^{s} n_k \tau_k = 0, \qquad \sum_{k=1}^{s} n_k \neq 0,$

(4) *any number of other conditions hold for* $\tau_{s+1}, \tau_{s+2}, \cdots, \tau_p$, *such that the method of Lagrange multipliers may be used,*

then condition (3) *may be ignored in the minimizing of*

$$SSE = \left(Y - \mu I - \sum_{k=1}^{p} X_k \tau_k\right)^2.$$

Our estimates of μ, τ_i, β_j, $(\tau\beta)_{ij}$ are m, t_i, b_j, $(tb)_{ij}$, respectively, where these values minimize SSE subject to the conditions (3). By Theorem A, we may ignore the conditions on the τ_i's and the β_j's. The conditions on the $(\tau\beta)_{ij}$'s will have to be considered in the computation of SSE_2 and SSE_3 but in the computation of SSE they can be avoided by expressing SSE in a different form. We have

$$SSE = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \xi_{ij})^2$$

Taking partial derivatives, we find our estimate of ξ_{ij} is $\hat{\xi}_{ij} = \bar{Y}_{ij.}$, where this notation indicates an average over the missing subscript. Then, by the invariance property of such estimators,

$$m = \bar{Y}_{...}, \qquad t_i = \bar{Y}_{i..} - \bar{Y}_{...}, \qquad bj = \bar{Y}_{.j.} - \bar{Y}_{...},$$
$$(tb)_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{..},$$

and

$$SSE = \sum_{i,j,k_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{ij.})^2 = \sum_{i,j,k_{ij}} Y_{ijk_{ij}}^2 - \sum_{i,j}^{p,q} \frac{Y_{ij.}^2}{n_{ij}}$$

where $Y_{ij.}$ is the sum of all the observations on the ith treatment in the jth block.

In obtaining SSE_1, all of the conditions (3) may be ignored, but, to determine SSE_2 and SSE_3, the method of Lagrange multipliers must be used. As a result of these calculations, we find that

$$SS(TB) = \sum_{i,j}^{p,q} n_{ij}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2,$$

$$SST = \sum_{i=1}^{p} n_{i.}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = \sum_{i=1}^{p} \frac{Y_{i..}^2}{n_{i.}} - \frac{Y_{...}^2}{N},$$

and

$$SSB = \sum_{j=1}^{q} n_{.j}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 = \sum_{j=1}^{q} \frac{Y_{.j.}^2}{n_{.j}} - \frac{Y_{...}^2}{N}.$$

4. Other models. We still assume that

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk_{ij}}.$$

For the Type II model we assume that the τ_i's, β_j's $(\tau\beta)_{ij}$'s and $\epsilon_{ijk_{ij}}$'s are NID with zero means and variances σ_τ^2, σ_β^2, $\sigma_{\tau\beta}^2$, and σ^2, respectively. For the Type III model we assume that the τ_i's, β_j's, and $(\tau\beta)_{ij}$'s come from finite independent populations of size $P > p$, $Q > q$, and PQ, respectively, with zero means and variances

$$\sigma_\tau^2 = \frac{\sum_{i=1}^{P} \tau_i^2}{P-1}, \qquad \sigma_\beta^2 = \frac{\sum_{j=1}^{Q} \beta_j^2}{Q-1}, \qquad \sigma_{\tau\beta}^2 = \frac{\sum_{i,j}^{P,Q} (\tau\beta)_{ij}^2}{(P-1)(Q-1)}.$$

The assumption of zero means implies that

$$\sum_{i=1}^{P} \tau_i = 0, \qquad \sum_{j=1}^{Q} \beta_j = 0, \qquad \sum_{i,j}^{P,Q} (\tau\beta)_{ij} = 0,$$

and, in addition, we assume that

$$\sum_{i=1}^{P} (\tau\beta)_{ij} = \sum_{j=1}^{Q} (\tau\beta)_{ij} = 0.$$

For the mixed model we may assume that the τ_i's, β_j's, and $(\tau\beta)_{ij}$'s are of any of the types described above. In addition when the τ_i's, say, are of Type I and the β_i's of Type II, Anderson and Kempthorne ([1], [2]) have shown that it is desirable to assume that, corresponding to each β_j, there exists a population of $(\tau\beta)_{ij}$'s consisting of p elements such that

$$\sum_{i=1}^{p} (\tau\beta)_{ij} = 0, \qquad \sigma_{\tau\beta}^2 = \frac{\sum_{i=1}^{p} (\tau\beta)_{ij}^2}{p-1}.$$

If the τ_i's came from a Type III population, we would replace p by P in the above definitions, and if the roles of the τ_i's and β_j's were interchanged, we would interchange i and j and replace p by q. We always assume the $\epsilon_{ijk_{ij}}$'s are NID $(0, \sigma^2)$.

5. The expected values of the sums of squares. In every case we shall arbitrarily begin with the sums of squares obtained for the Type I model. To determine their expected values, we shall make use of the following theorem which is a slight generalization of one stated by Tukey [11].

THEOREM B.

If $y_1, y_2, \cdots, y_p$ *have means* $\mu_1, \mu_2, \cdots, u_p$, *variances* $\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2$, *and every pair has the same covariance*, λ, *then*

$$E\left\{\sum_{i=1}^{p} n_i(y_i - \bar{y}.)^2\right\} = \sum_{i=1}^{p} n_i(\mu_i - \bar{\mu}.)^2 + \sum_{i=1}^{p} n_i\left(1 - \frac{n_i}{N}\right)(\sigma_i^2 - \lambda)$$

where

$$\bar{y}. = \frac{\sum_{i=1}^{p} n_i y_i}{N}, \qquad \bar{\mu}. = \frac{\sum_{i=1}^{p} n_i \mu_i}{N}, \qquad N = \sum_{i=1}^{p} n_i .$$

We find that

$$SST = \sum_{i=1}^{p} n_{i.}(w_i - \bar{w}.)^2, \qquad SSB = \sum_{j=1}^{q} n_{.j}(y_j - \bar{y}.)^2,$$

$$SS(TB) = \sum_{i,j}^{p,q} n_{ij}(z_i - \bar{z}.)^2, \qquad SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\epsilon_{ijk_{ij}} - \bar{\epsilon}_{ij.})^2,$$

where

$$w_i = \tau_i + (\overline{\tau\beta})_{i.} + \bar{\epsilon}_{i..}, \qquad y_j = \beta_j + (\overline{\tau\beta})_{.j} + \bar{\epsilon}_{.j.},$$

$$z_i = (\tau\beta)_{ij} - (\overline{\tau\beta})_{i.} + \bar{\epsilon}_{ij.} - \bar{\epsilon}_{i..}$$

and

$$\bar{\tau}. = \frac{\sum_{i=1}^{p} n_{i.}\tau_i}{N}, \qquad \bar{\beta}. = \frac{\sum_{j=1}^{q} n_{.j}\beta_j}{N}, \qquad (\overline{\tau\beta})_{i.} = \frac{\sum_{j=1}^{q} n_{.j}(\tau\beta)_{ij}}{N},$$

$$(\overline{\tau\beta})_{.j} = \frac{\sum_{i=1}^{p} n_{i.}(\tau\beta)_{ij}}{N}, \qquad (\overline{\tau\beta})_{..} = \frac{\sum_{i,j}^{p,q} n_{ij}(\tau\beta)_{ij}}{N}.$$

In order to apply Theorem B, we need the variances and covariances of the w_i, y_j, and z_i in a form that does not depend on the form of the model. By using the methods employed by Bennett and Franklin [3], we find for the Type III model that

$$\mu_i = E(w_i) = (1 - \delta_\tau)\tau_i, \qquad \bar{\mu}. = 0,$$

$$\sigma_i^2 = \delta_\tau\left(1 - \frac{1}{P}\right)\sigma_\tau^2 + \delta_{\tau\beta}\left(1 - \frac{1}{P}\right)\frac{1}{N^2}\left(\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q}\right)\sigma_{\tau\beta}^2 + \frac{\sigma^2}{n_{i.}}.$$

$$\lambda = -\delta_\tau \frac{\sigma_\tau^2}{P} - \delta_{\tau\beta} \frac{1}{PN^2} \left(\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q} \right) \sigma_{\tau\beta}^2 ,$$

where $\delta_\tau = 0$ if the τ_i's come from a Type I population, $\delta_\tau = 1$ otherwise, and a similar definition holds for $\delta_{\tau\beta}$.

Application of Theorem B enables us to find $E(SST)$ and division by $p - 1$ gives us

$$E(MST) = \sigma^2 + \frac{\delta_{\tau\beta} a}{(p-1)N^2} \left(\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q} \right) \sigma_{\tau\beta}^2 + \frac{\delta_\tau a}{p-1} \sigma_\tau^2 + \frac{(1-\delta_\tau)}{p-1} \sum_{i=1}^{p} n_{i.}\tau_i^2 ,$$

where

$$a = N - \frac{1}{N} \sum_{i=1}^{p} n_{i..}^2 .$$

Similarly

$$E(MSB) = \sigma^2 + \frac{\delta_{\tau\beta} b}{(q-1)N^2} \left(\sum_{i=1}^{p} n_{i.}^2 - \frac{N^2}{P} \right) \sigma_{\tau\beta}^2 + \frac{\delta_\beta b}{q-1} \sigma_\beta^2 + \frac{(1-\delta_\beta)}{q-1} \sum_{i=1}^{q} n_{.j}\beta_j^2 ,$$

where

$$b = N - \frac{1}{N} \sum_{j=1}^{q} n_{.j}^2 ,$$

and

$$E[MS(TB)] = \sigma^2 + \frac{\delta_{\tau\beta} ab}{(p-1)(q-1)N} \sigma_{\tau\beta}^2 + \frac{(1-\delta_{\tau\beta}) \sum_{i,j}^{p,q} n_{ij}(\tau\beta)^2_{ij}}{(p-1)(q-1)}$$

Finally, by the theory for the Type I model, we know that SSE is distributed as $\chi^2\sigma^2$ with $N - pq$ degrees of freedom and hence $E(MSE) = \sigma^2$. We also note that, if all the $n_{ij} = 1$, $SSE = 0$ and it is impossible to carry out any of the F tests which involve division by this quantity.

6. Models with no interaction. In this case, for the Type I model,

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_j + \epsilon_{ijk_{ij}} .$$

We find, as in Sec. 2, that

$$m = \bar{Y}_{...}, \qquad t_i = \bar{Y}_{i..} - \bar{Y}_{...}, \qquad b_j = \bar{Y}_{.j.} - \bar{Y}_{...} .$$

and

$$SSE_1 = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (Y_{ijk_{ij}} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{..})^2$$

of that section plays the role of SSE. We also saw in Sec. 2 that

$$SSE_1 = SSE + SS(TB)$$

so, if we had accepted

$$H_1: (\tau\beta)_{ij} = 0,$$

and decided to change models in midstream, all that would be necessary to obtain SSE_1 would be to pool the interaction and error sums of squares. We find the same expressions for SST and SSB as in Sec. 3 and that the degrees of freedom associated with SSE_1 are those obtained by pooling the degrees of freedom associated with $SS(TB)$ and SSE. To obtain the expected values of MST and MSB one need only omit the terms involving the $(\tau\beta)_{ij}$, and it is easily verified that $E(MSE_1) = \sigma^2$. A discussion as to when pooling is desirable is to be found in Bechhofer's thesis [2] and in a paper by Bozivich, Bancroft and Hartley [4].

7. Distributions of the sums of squares. Corresponding to the hypotheses

$$H_1: (\tau\beta)_{ij} = 0, \qquad H_2: \tau_i = 0, \qquad H_3: \beta_j = 0,$$

we have the hypotheses

$$\sigma_{\tau\beta} = 0, \qquad \sigma_\tau = 0, \qquad \sigma_\beta = 0,$$

if the corresponding variables are from other than a Type I population. Then' since the populations have zero means, it follows that the corresponding variables are equal to zero. We have already referred to the tests for the above hypotheses for the Type I model at the end of Sec. 2. If there is no interaction term, subject to H_2 and H_3, the sums of squares SST and SSB reduce to the corresponding expressions for the Type I model and the tests of Sec. 6 apply no matter which model we may be considering. If there is an interaction term, the same argument shows that the Type I test can be used for H_1. Thus our problem is reduced to testing H_2 and H_3 when there is interaction and we are not dealing with a Type I model.

We first consider the Type II model where the parameters are NID with zero means and variances σ_τ^2, σ_β^2, and $\sigma_{\tau\beta}^2$. If all the n_{ij}'s are equal to n, say, using methods similar to those of Mood [6], it can be shown that SST, SSB, $SS(TB)$ and SSE are independently distributed as $\chi^2 E(MST)$, $\chi^2 E(MSB)$, $\chi^2 E[MS(TB)]$, and $\chi^2\sigma^2$ with $p - 1$, $q - 1$, $(p - 1)(q - 1)$, and $N - pq$ degrees of freedom, respectively. These results hold independent of the validity of H_1, H_2, and H_3. Since some of the details differ from those given by Mood we shall outline the proof of the above results.

The theory for the Type I model shows that

$$t_i = \bar{Y}_{i..} - \bar{Y}_{...}, \qquad bj = \bar{Y}_{.j.} - \bar{Y}_{...}, \qquad (tb)_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...},$$

$i = 1, 2, \cdots, p - 1; j = 1, 2, \cdots, q - 1$, are distributed independently of

$$SSE = \sum_{i,j,k_{ij}}^{p,q,n_{ij}} (\epsilon_{ijk_{ij}} - \bar{\epsilon}_{ij.})^2.$$

Therefore any function of these statistics is distributed independently of SSE, and, in particular, this holds for t_p, b_q, $(tb)_{pj}$ and $(tb)_{iq}$. These results hold for the particular case where $Y_{ijk} = \epsilon_{ijk}$. Hence

$$\bar{\epsilon}_{i..} - \bar{\epsilon}_{...}, \qquad \bar{\epsilon}_{.j.} - \bar{\epsilon}_{...}, \qquad \bar{\epsilon}_{ij.} - \bar{\epsilon}_{i..} - \bar{\epsilon}_{.j.} + \bar{\epsilon}_{...},$$

$i = 1, 2, \cdots, p; j = 1, 2, \cdots, q$, are distributed independently of SSE. It may be shown that any variable of the above three types is independent of any variable of the other two types by computing the appropriate covariances. We know that

$$SST = qn \sum_{i=1}^{p} (w_i - \bar{w}_.)^2$$

where

$$w_i = \tau_i + (\overline{\tau\beta})_{i.} + \bar{\epsilon}_{i..}, \qquad E(w_i) = 0, \qquad \text{var}(w_i) = \sigma_\tau^2 + \frac{\sigma_{\tau\beta}^2}{q} + \frac{\sigma^2}{qn},$$

$$\text{cov}(w_{i'}, w_i) = 0, \qquad E(MST) = \sigma^2 + n\sigma_{\tau\beta}^2 + qn\sigma_\tau^2.$$

It follows that $SST / E(MST)$ has a χ^2 distribution with $p - 1$ degrees of freedom. Similarly SSB is distributed as $\chi^2 E(MSB)$ with $q - 1$ degrees of freedom.

Consider the three sets of variables

$$(\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{..}, \qquad (\overline{\tau\beta})_{.j} - (\overline{\tau\beta})_{..}, \qquad (\overline{\tau\beta})_{ij} - (\overline{\tau\beta})_{i.} - (\overline{\tau\beta})_{.j} + (\overline{\tau\beta})_{..}.$$

As with the ϵ_{ijk}'s, it may be shown that any variable of the above three types is independent of any variable of the other two types. Then it follows that the three sets of variables

$$w_i - \bar{w}_., \qquad y_j - \bar{y}_., \qquad z_i - \bar{z}.$$

are independently distributed and hence so are SST, SSB, and $SS(TB)$.

If, in the results for the Type I model, we set μ, the τ_i's and the β_j's equal to zero and assume the $(\tau\beta)_{ij}$'s are NID $(0, \sigma_{\tau\beta}^2)$,

$$Y_{ijk} = (\tau\beta)_{ij} + \epsilon_{ijk}, \qquad \bar{Y}_{ij.} = (\tau\beta)_{ij} + \bar{\epsilon}_{ij.},$$

$$E(\bar{Y}_{ij.}) = 0, \qquad \text{var}(\bar{Y}_{ij.}) = \sigma_{\tau\beta}^2 + \frac{\sigma^2}{n},$$

and the $\bar{Y}_{ij.}$'s are independent. We carry out an analysis of variance on the $\bar{Y}_{ij.}$'s according to the model of Sec. 6, where there is no interaction, with the N of *that* section equal to pq and $n_{ij} = 1$, to obtain

$$SSE_1 = SSE + SS(TB) = \sum_{i,j}^{p,q} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

since, under these conditions, the Y_{ijk} of that section is equal to $\bar{Y}_{ij.}$. The theory of Sec. 6, when we replace σ^2 by $\sigma_{\tau\beta}^2 + \sigma^2/n$, tells us that SSE_1 is distributed as $\chi^2(n\sigma_{\tau\beta}^2 + \sigma^2)$ and hence

$$SS(TB) = \sum_{i,j}^{p,q} n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{...} + \bar{Y}_{...})^2$$

is distributed as $\chi^2(n\sigma_{\tau\beta}^2 + \sigma^2)$ with $pq - p - q + 1 = (p-1)(q-1)$ degrees of freedom. It then follows that the appropriate tests for H_1, H_2, H_3 are given by the statistics

$$F_1 = \frac{MS(TB)}{MSE}, \qquad F_2 = \frac{MST}{MS(TB)}, \qquad F_3 = \frac{MSB}{MS(TB)},$$

respectively. A proof of these results is also outlined by Anderson and Bancroft [1].

The above results are for the case where $n_{ij} = n$. If this condition does not hold, we can no longer say that F_2 and F_3 have the F distribution. This may be shown by considering the special case where $p = 3$, $q = 2$, $n_{11} = n_{12} = 1$, $n_{21} = n_{22} = 2$, $n_{31} = n_{32} = 3$ and $N = 12$. Then the moment generating function of SST is

$$[1 - 4(11x + 3\sigma^2)t/3 + 4(36x^2 + 22x\sigma^2 + 3\sigma^4)t^2/3]^{-1/2}$$

where $x = \sigma_\tau^2 + \sigma_{\tau\beta}^2/2$. This is not the moment generating function of a variable of the form $c\chi^2$ unless $x = 0$. Thus there is no hope of F_2 having an F distribution and a similar argument holds for F_3.

For a Type III model with interaction we cannot expect to obtain the distributions necessary for F tests of H_2 and H_3 since the $(\tau\beta)_{ij}$'s are not normally distributed. An approach similar to the one given above could be used in the case of the mixed model.

8. The two-way nested classifications. This model is discussed by Bennett and Franklin [3]. We assume that

$$Y_{ijk_{ij}} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{ijk_{ij}},$$

$$\sum_{i=1}^{p} n_{i.}\tau_i = 0, \qquad \sum_{j=1}^{q} n_{.j}\beta_{j(i)} = 0, \qquad i = 1, 2, \cdots, p.$$

We test two hypotheses,

$$H_1: \beta_{j(i)} = 0, \qquad i = 1, 2, \cdots, p; j = 1, 2, \cdots, q,$$

and

$$H_2: \tau_i = 0, \qquad i = 1, 2, \cdots, p,$$

using the statistics, for the Type I model,

$$F_1 = \frac{MSB}{MSE} \text{ (with } p(q-1) \text{ and } N - pq \text{ degrees of freedom)},$$

and

$$F_2 = \frac{MST}{MSE} \text{ (with } p - 1 \text{ and } N - pq \text{ degrees of freedom)},$$

where SST and SSE have the values given earlier and

$$SSB = \sum_{i,j}^{p,q} n_{ij}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2 = \sum_{i,j}^{p,q} \frac{Y_{ij.}^2}{n_{ij}} - \sum_{i=1}^{p} \frac{Y_{i..}^2}{n_{i.}}.$$

The Type II model is defined as before but, in the case of the Type III model, we assume that the τ_i's come from a finite population of size P, mean zero, and variance

$$\sigma_\tau^2 = \frac{\sum_{i=1}^{P} \tau_i^2}{P - 1}$$

while the $\beta_{j(i)}$'s come from P populations of size Q, corresponding to the different values of i, these populations being independent of each other and the population of τ_i's, with zero means and common variance

$$\sigma_\beta^2 = \frac{\sum_{j=1}^{Q} \beta_{j(i)}^2}{Q - 1}.$$

The expected values of the mean squares are

$$E(MST) = \sigma^2 + \frac{\delta_\beta a}{(p-1)N^2}\left[\sum_{j=1}^{q} n_{.j}^2 - \frac{N^2}{Q}\right]\sigma_\beta^2 + \frac{\delta_\tau a}{p-1}\sigma_\tau^2 + \frac{(1-\delta_\tau)}{p-1}\sum_{i=1}^{p} n_{i.}\tau_i^2,$$

$$E(MSB) = \sigma^2 + \frac{\delta_\beta b}{p(q-1)}\sigma_\beta^2 + \frac{(1-\delta_\beta)}{p(q-1)}\sum_{i,j}^{p,q} n_{ij}\beta_{j(i)}^2, \qquad E(MSE) = \sigma^2,$$

where the δ's have the meaning assigned in Sec. 5 and

$$a = N - \frac{\sum_{i=1}^{p} n_{i.}^2}{N}, \qquad b = N - \frac{\sum_{j=1}^{q} n_{.j}^2}{N}$$

Examination of MSB indicates that, no matter what model we may use, we may test the hypothesis H_1 by the statistic F_1 given earlier in this section. For a Type II model with $n_{ij} = n$, we test H_2 with the statistic

$$F = \frac{MST}{MSB}.$$

For other cases an approximate method must be used such as is given elsewhere in the literature [7].

REFERENCES

[1] R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research*, McGraw-Hill, New York, 1952, pp. 177–182, 234, 314–319, and 340.

[2] R. E. Bechhofer, "The effect of preliminary tests of significance on the size and power of certain tests of univariate linear hypotheses," Ph.D. Thesis, Columbia University Library.

[3] C. A. Bennett and N. L. Franklin, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley & Sons, New York, 1954, pp. 470–477.

[4] Helen Bozivich, T. A. Bancroft and H. O. Hartley, "Power of analysis of variance test procedures for certain incompletely specified models, I," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1017–1043.

[5] H. B. Mann, *Analysis and Design of Experiments*, Dover, New York, 1949, pp. 39–40.

[6] A. M. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill, New York, 1950, p. 344.

[7] F. E. Satterthwaite, "An approximate distribution of estimates of variance components," *Biometrics*, Vol. 2 (1946), pp. 110–114.

[8] H. F. Smith, "Analysis of variance with unequal but proportionate numbers of observations in the sub-classes of a two-way classification," *Biometrics*, Vol. 7 (1951), pp. 70–74.

[9] G. W. Snedecor and G. M. Cox, "Disproportionate Subclass Numbers in Tables of Multiple Classification," Iowa State College, Agr. Expt. Bull. 180 (1935).

[10] G. W. Snedecor, *Statistical Methods*, 4th ed., Iowa State College Press, Ames, Iowa, pp. 281–284.

[11] J. W. Tukey, "Dyadic anova, an analysis of variance for vectors," *Human Biology*, Vol. 21 (1949), pp. 65–110.

[12] O. Kempthorne, *The Design and Analysis of Experiments*, John Wiley & Sons, New York, 1952, p. 574.