VARIANCES OF VARIANCE COMPONENTS: II. THE UNBALANCED SINGLE CLASSIFICATION[1]

By John W. Tukey

Princeton University

1. Summary. The variance of the usual estimate of between variance components in an unbalanced single classification has been found for arbitrary infinite populations by Hammersley [1], who found it necessary to use rather heavy algebra. The methods of polykays are here applied to a family of weighted estimates to obtain the variances and covariances of the estimates of between and within variance components. These apply to arbitrary finite populations.

Weighting column means equally seems to give a better estimate than the classical proportional weighting for the between variance component as soon as (i) the between component exceeds $\frac{1}{2}$ of the within component in a moderately unbalanced design, or (ii) the between component exceeds the within component in a substantially unbalanced design. Slight further gains come from intermediate weighting. Numerical examples are given.

While pooling mean squares instead of sums of squares across columns loses accuracy, notably for the within variance component, doing the same in calculating the between variance component seems to have a minor effect. If the within contributions are sufficiently non-normal, this effect will be favorable.

2. Introduction. This paper closely follows the method and concept of the first paper of this series [2], familiarity with the techniques and results of which is assumed. The present paper deals with the unbalanced single classification, where we have observations in the various columns. The actual observations are supposed to be representable in the form

$$\text{(observation)} = \text{(column contribution)} + \text{(cell contribution)},$$

where each class of contribution arises from a separate population, or populations, and some independence is assumed for the selection or sampling of contributions in the different classifications (this is not a serious element of unrealism for a *single* classification situation).

A wide variety of models can be constructed within this framework. The way in which the lack of balance arises may be very important. If the number of observations in a column is at all related to the value of the corresponding column contribution (as might be the case if items with potentially extreme values were preferentially lost), the situation becomes very complex, and may be outside

Received April 6, 1956.

[1] Prepared in connection with research sponsored by the Office of Naval Research and based on part of Memorandum Report 45, "Finite Sampling Simplified," Statistical Research Group, Princeton University, which was written while the author was a Fellow of the John Simon Guggenheim Foundation.

the scope of both the present paper and the literature known to this writer. We shall assume a fixed pattern of column sizes and a random arrangement of the column contributions among them.

There still remain various possibilities for the cell contributions. These could, for example, be drawn from a single population, or from a family of populations, one per column. In the interests of simplicity, we shall begin with the case where only one population is involved.

3. Types of analysis. Having specified the probability model, we are not, as an acquaintance with balanced designs alone might suggest, through with the specification of the problem. There are various possible analyses to make of the observations, as we shall shortly see. Let $\{x_{ij}\}$, with $i = 1, 2, \cdots, c$ and $j = 1, 2, \cdots, r_i$, be the observations, let $\{x_{i+}\}$ be the column totals, x_{++} be the grand total, and let R, the sum of the r_i, be the total number of observations.

If we are to have unbiased estimates of the variance components, they will be quadratic functions of the x_{ij} with coefficients depending on c and the $\{r_i\}$. In principle, we could start with a general quadratic function and then optimize its coefficients in some way. In practice, we select two quadratic functions by some scheme involving elements of intuition, find how their average values are expressed linearly in terms of the variance components, and then form two linear combinations of the original quadratics whose average values are the variance components. These linear combinations are then our estimates. Much flexibility is possible in this situation, but only a limited amount of flexibility seems to be customary.

Within each column, reasons of symmetry favor using

$$\sum_j \left(x_{ij} - \frac{x_{i+}}{r_i}\right)^2 = \sum_j x_{ij}^2 - \frac{x_{i+}}{r_i},$$

but we may as well be prepared for the use of arbitrary weights in combining these pieces. For the first quadratic then, we take

$$J = \sum_i u_i \sum_j \left(x_{ij} - \frac{x_{i+}}{r_i}\right)^2 = \sum_{ij} u_i x_{ij}^2 - \sum_i u_i \frac{x_{i+}^2}{r_i}.$$

The average value of J will be shown to be $\sum u_i(r_i - 1)$ times the within-variance component, so that the within estimate is immediately constructible.

The other quadratic is usually definable in terms of the column means x_i/r_i and some weighted grand mean

$$\frac{1}{W} \sum w_i \frac{x_{i+}}{r_i},$$

where W is the sum of the weights w_i. The usual expression is the weighted sum of squares of deviations

$$L = \sum w_i \left(\frac{x_{i+}}{r_i} - \frac{1}{W} \sum w_j \frac{x_{j+}}{r_j}\right)^2 = \sum w_i \frac{x_{i+}^2}{r_i^2} - \frac{1}{W} \left(\sum w_j \frac{x_{j+}}{r_j}\right)^2.$$

We shall confine our analyses to this family of cases, which is parametrized by the $\{u_i\}$ and the $\{w_i\}$.

The choice $w_i = r_i$ gives the customary analyses, which treat observations as important and columns as unimportant. This is appropriate when testing significance or when the column variance component is small compared with the within variance component.

The choice $w_i = 1$ gives the equally-weighted-column analyses, which treat columns as important and observations as unimportant. This is appropriate when the column variance component is large compared with the within variance component.

Some intermediate choice of weights may indeed be preferred.

Finally, as we shall see, it is sometimes possible to choose the weights so that, although the pattern is unbalanced, the analysis becomes a balanced analysis in the sense of Paper I [2]. This, we shall see, occurs when we try to make the estimate of the between variance component unbiased whenever the variances within the various columns differ.

We shall try to obtain as general answers as seem useful for this family of cases.

4. Model and elementary results. Our model is

$$x_{ij} = \mu + \eta_i + \omega_{ij}, \qquad i = 1, 2, \cdots, c, \qquad j = 1, 2, \cdots, r_i;$$

$$\eta\text{'s from } n, k_1, k_{11}, \cdots,$$

$$\omega\text{'s from } N, K_1, K_{11}, \cdots;$$

independently and randomly sampled and arranged.

It is easy to see that the values of the η do not affect the value of the within sum of squares J. Since J is quadratic in the η's and ω's and invariant under a common translation of all the values in either population, we must have

$$\text{ave}\{J\} = \phi K_2.$$

Since L is also quadratic and invariant, we must have

$$\text{ave}\{L\} = \zeta k_2 + \xi K_2.$$

The estimates of the variance components will then be

$$\text{within} = \frac{1}{\phi} J, \qquad \text{between} = \frac{1}{\zeta} L - \frac{\xi}{\phi\zeta} J.$$

Arguments similar to those just used, and entirely parallel to those used in Paper I for the balanced case, now determine the finite population corrections and the vanishing of certain coefficients in the expressions for variances and covariance. The results are

$$\text{var}\{\text{between}\} = \left(\alpha_1 - \frac{1}{n}\right) k_4 + \left(\beta_1 - \frac{2}{n-1}\right) k_{22} + \gamma_1 k_2 K_2 + \delta_1 K_4 + \epsilon_1 K_{22},$$

$$\text{var}\{\text{within}\} = \left(\delta_2 - \frac{1}{N}\right) K_4 + \left(\epsilon_2 - \frac{2}{N-1}\right) K_{22},$$

$$\text{cov}\{\text{between, within}\} = \delta_3 K_4 + \epsilon_3 K_{22}.$$

We are left with the task of determining ϕ, ζ, ξ, α_1, β_1, γ_1, δ_1, δ_2, δ_3, ϵ_1, ϵ_2, ϵ_3; we may do this by treating separately the cases of (i) single minimal unit populations, which together give us ϕ, ζ, ξ, α_1, δ_1, δ_2, δ_3, and (ii) normal theory, which gives us β_1, γ_1, ϵ_1, ϵ_2, ϵ_3. These are our next tasks. We shall find it helpful to calculate, as intermediate quantities, some of the coefficients in the formulas for the variances and covariances of J and L. These formulas are of the forms

$$\text{var}\{L\} = \left(\alpha_L - \frac{\zeta^2}{n}\right) k_4 + \left(\beta_L - \frac{2\zeta^2}{n-1}\right) k_{22} + \gamma_L k_2 K_2 + \left(\delta_L - \frac{\xi^2}{n}\right) K_4 + \left(\epsilon_L - \frac{2\xi^2}{n-1}\right) K_{22},$$

$$\text{var}\{J\} = \left(\delta_J - \frac{\phi^2}{n}\right) K_4 + \left(\epsilon_J - \frac{2\phi^2}{n-1}\right) K_{22},$$

$$\text{cov}\{J, L\} = \left(\delta_c - \frac{\phi\xi}{n}\right) K_4 + \left(\epsilon_c - \frac{2\phi\xi}{n-1}\right) K_{22}.$$

5. A single-unit column. If we take the special case where all ω's are zero ($K_2 = K_4 = K_{22} = 0$) and the η's are a minimal unit population (just enough to go around, all zero except one and that one equal to unity, $n = c$, $k_2 = k_4 = c/1$, $k_{22} = 0$), then we can make the first step. We will have $J = 0$, and if the unit η falls into the jth column (an event of probability $1/c$), we shall have

$$x_{j+} = r_j, \quad \text{other} \quad x_{i+} = 0, \qquad L = w_j - w_j^2/W = a_j,$$

which defines a_j. (If we write θ_j for the relative weight w_j/W, then $a_j = W\theta_j(1 - \theta_j)$.)

Thus,

$$A = \sum a_j = W - \frac{1}{W}\sum w_j^2,$$

and hence

$$\frac{A}{c} = \text{ave}\{L\} = \zeta k_2 + \xi K_2 = \zeta\frac{1}{c}$$

so that $\zeta = A$. Correspondingly,

$$\text{var}\{L\} = \frac{1}{c}\sum a_j^2 - \left(\frac{A}{c}\right)^2 = \left(\alpha_L - \frac{\zeta^2}{c}\right)\frac{1}{c}.$$

so that

$$\alpha_L = \sum a_j^2 = W^2 \sum \left(\frac{w_j}{W}\right)^2 \left(1 - \frac{w_j}{W}\right)^2$$

6. A single-unit observation. Take, now, the special case where the ω's are a minimal unit population ($N = R$, $K_2 = K_4 = 1/R$, $K_{22} = 0$) and all the η's vanish ($k_2 = k_4 = k_{22} = 0$). If the single non-zero ω falls in the jth column, an event whose probability is now r_j/R and not $1/c$, then we have

$$J = u_j\left(1 - \frac{1}{r_j}\right), \qquad L = \frac{w_j}{r_j^2} - \frac{1}{W}\left(\frac{w_j}{r_j}\right)^2 = b_j.$$

(Note that $a_j = r_j^2 b_j$.)

The average value of J is

$$\text{ave } \{J\} \sum \frac{u_j r_j}{R}\left(1 - \frac{1}{r_j}\right) = \phi K_2 = \frac{\phi}{R}$$

whence $\phi = \sum u_j(r_j - 1)$. The variance of J is

$$\text{var } \{J\} = \sum \frac{r_j}{R} u_j^2 \left(1 - \frac{1}{r_j}\right)^2 - \left(\frac{\phi}{R}\right)^2 = \left(\delta_J - \frac{\phi^2}{R}\right)\frac{1}{R},$$

so that

$$\delta_J \sum u_j^2 r_j \left(1 - \frac{1}{r_j}\right)^2 = \sum u_j^2 \left(r_j - 2 + \frac{1}{r_j}\right).$$

The average value of L is

$$\sum \frac{r_j}{R} b_j = \text{ave } \{L\} = \zeta k_2 + \xi K_2 = \frac{\xi}{R},$$

so that

$$\xi = \sum r_j b_j = \sum \frac{w_j}{\gamma_j} - \frac{1}{W} \sum \frac{w_j^2}{r_j}.$$

For the two special choices of the w_j, this reduces to

$$\xi = c - 1 \qquad (\text{for } w_j \equiv r_j),$$

$$\xi = \left(1 - \frac{1}{c}\right) \sum \frac{1}{r_j} \qquad (\text{for } w_j \equiv 1).$$

The variance of L is

$$\text{var } \{L\} = \sum \frac{r_j}{R} b_j^2 - \left(\frac{\xi}{R}\right)^2 = \left(\delta_L - \frac{\xi^2}{R}\right)\frac{1}{R},$$

so that

$$\delta_L = \sum r_j b_j^2.$$

The covariance of J and L is

$$\text{cov}\ \{J, L\} = \sum \frac{r_j}{R} b_j u_j \left(1 - \frac{1}{r_j}\right) - \left(\frac{\phi}{R}\right)\left(\frac{\xi}{R}\right) = \left(\delta_c - \frac{\phi\xi}{R}\right)\frac{1}{R},$$

whence

$$\delta_c = \sum u_j b_j (r_j - 1).$$

7. Normal theory. We now need to calculate variances and covariances of J and L on normal theory (where $k_4 = K_4 = 0$, $K_{22} = K_2^2$, $n = N = \infty$). The sums of squares within each separate column are distributed as multiples of chi square, independently of each other and of L, so that we have

$$\text{var}\ \{J\} = \sum u_j^2 \frac{2(r_j - 1)^2 K_2}{r_j - 1} = \epsilon_J K_{22}$$

whence

$$\epsilon_J = 2 \sum u_j^2 (r_j - 1),$$

and

$$\text{cov}\ \{J, L\} = 0 = \epsilon_c K_{22},$$

whence

$$\epsilon_c = 0.$$

In calculating var $\{L\}$, it will be convenient to assume that all means are zero, so that

$$\text{var}\ \left\{\left(\sum c_{ij} x_{ij}\right)^2\right\} = 2\ \left(\text{var}\ \left\{\sum c_{ij} x_{ij}\right\}\right)^2,$$

and to write

$$x_{i\cdot} = \frac{x_{i+}}{x_i}, \qquad x_{-} = \sum w_i x_{i\cdot},$$

when

$$\text{var}\ \{x_i\} = k_2 + \frac{1}{r_i} K_2,$$

$$\text{var}\ \{x_-\} = \left(\sum w_i^2 k_2 + \left(\sum w_i^2/r_i\right) K_2\right),$$

$$\text{cov}\ \{x_i, x_-\} = \left(w_i k_2 + \frac{w_i}{r_i} K_2\right),$$

and since

$$L = \sum w_i x_{i\cdot}^2 - \frac{1}{W} x_-^2,$$

we have

$$\begin{aligned}\text{var}\ \{L\} &= 2\sum w_i^2\left(k_2 + \frac{1}{r_i}K_2\right)^2\\ &\quad + \frac{2}{W^2}\left(\sum w_i^2k_2 + \left(\sum w_i^2/r_i\right)K_2\right)^2\\ &\quad - \frac{4}{W}\sum w_i\left(w_ik_2 + \frac{w_i}{r_i}K_2\right)\\ &= \beta_L k_2^2 + \gamma_L k_2K_2 + \epsilon_L K_2^2,\end{aligned}$$

whence

$$\begin{aligned}\beta_L &= 2\left(\sum w_i^2 - \frac{2}{W}\sum w_i^3 + \frac{1}{W^2}\left(\sum w_i^2\right)\right)\\ &= 2\left(\sum a_i^2 + \frac{1}{W^2}\left(\left(\sum w_i^2\right)^2 - \left(\sum w_i^4\right)\right)\right),\\ \gamma_L &= 4\left(\sum \frac{w_i}{r_i} - \frac{2}{W}\sum \frac{w_i^3}{r_i} + \frac{1}{W^2}\sum{}_i^2\sum \frac{w_i^2}{r_i}\right)\\ &= 4\left(\sum \frac{a_j^2}{r_j} + \frac{1}{W^2}\left[\left(\sum w_j^2\right)\left(\sum \frac{w_j^2}{r_j}\right) - \left(\sum \frac{w_j^4}{r_j}\right)\right]\right),\\ \epsilon_L &= 2\left(\sum \frac{w_i^2}{r_i^2} - \frac{2}{W}\sum \frac{w_i^3}{r_i^2} + \frac{1}{W^2}\left(\sum \frac{w_i^2}{r_i}\right)^2\right)\\ &= 2\left(\sum \frac{a_j^2}{r_j^2} + \frac{1}{W^2}\left(\left(\sum \frac{w_i^2}{r_i}\right)^2 - \sum \frac{w_i^4}{r_i^2}\right)\right).\end{aligned}$$

8. Combined results. Combining all of the results above, using such relations as

$$\alpha_1 = \frac{1}{\zeta^2}\alpha_L,$$

$$\delta_1 = \frac{1}{\zeta^2}\delta_L - \frac{2\xi}{\phi\zeta^2}\delta_C + \frac{\xi^2}{\phi^2\zeta^2}\delta_J,$$

$$\delta_3 = \frac{1}{\zeta\phi}\delta_C - \frac{\xi}{2}\delta_J,$$

and writing $\psi = \xi/\phi$, we find

$$\alpha_1 = \frac{1}{A^2}\sum a_j^2,$$

$$\beta_1 = \frac{2}{A^2}\left(\sum a_j^2 + \frac{1}{W^2}\left(\left(\sum w_i^2\right)^2 - \left(\sum w_i^4\right)\right)\right),$$

$$\gamma_1 = \frac{4}{A^2}\left(\sum \frac{a_j^2}{r_j} + \frac{1}{W^2}\left(\left(\sum w_j^2\right)\left(\sum \frac{w_j^2}{r_j}\right) - \left(\sum \frac{w_j^4}{r_j}\right)\right)\right),$$

$$\delta_1 = \frac{1}{A^2}\left(\sum r_j b_j^2 - 2\psi \sum u_j b_j (r_j - 1) + \psi^2 \sum u_j^2\left(r_j - 2 + \frac{1}{r_j}\right)\right),$$

$$\epsilon_1 = \frac{2}{A^2}\left(\sum \frac{a_j^2}{r_j^2} + \frac{1}{W^2}\left(\left(\sum \frac{w_i^2}{r_i}\right)^2 - \sum \frac{w_i^4}{r_i^2}\right) + \psi^2 \sum u_j^2 (r_j - 1)\right)$$

$$= \frac{2}{A^2}\left(\sum \frac{a_j^2}{r_j^2} + \frac{1}{W^2}\left(\left(\sum \frac{w_i^2}{r_i}\right)^2 - \left(\sum \frac{w_i^4}{r_i^2}\right)\right)\right) + \frac{\left(\sum r_j b_j\right)^2}{A^2}\,\epsilon_2,$$

$$\delta_2 = \frac{1}{\left(\sum u_j (r_j - 1)\right)^2} \sum u_j^2\left(r_j - 2 + \frac{1}{r_j}\right),$$

$$\epsilon_2 = \frac{1}{\left(\sum u_j (r_j - 1)\right)^2} \sum u_j^2 (r_j - 1),$$

$$\delta_3 = \frac{1}{A\left(\sum u_j (r_j - 1)\right)}\left(\sum u_j b_j (r_j - 1) - \psi \sum u_j^2\left(r_j - 2 + \frac{1}{r_j}\right)\right)$$

$$= \delta' - \frac{\sum r_j b_j}{A}\,\delta_2,$$

$$\epsilon_3 = \frac{-2\psi}{A\left(\sum u_j (r_j - 1)\right)^2} \sum u_j^2 (r_j - 1) = \frac{\sum r_j b_j}{A}\,\epsilon_2,$$

where

$$a_j = w_j - \frac{1}{W} w_j^2, \qquad A = \sum a_j,$$

$$b_j = \frac{a_j}{r_j^2}, \qquad \psi = \frac{\sum r_j b_j}{\sum u_j (r_j - 1)},$$

$$\delta' = \frac{1}{A} \frac{\sum u_j b_j (r_j - 1)}{\sum u_j (r_j - 1)}.$$

These results are not easy to digest, but, computationally, they are quite manageable. If we introduce $g_j = u_j(r_j - 1)$ and put $f_j = a_j/A$, so that $b_j/A = f_j/r_j^2$, and rearrange the order of the equations, we may write them in the following way:

$$\alpha_1 = \sum f_j^2,$$

$$\beta_1 = 2\alpha_1 + \frac{2}{A^2 W^2}\left(\left(\sum w_j^2\right)^2 - \left(\sum w_j^4\right)\right),$$

$$\gamma_1 = 4\sum (f_j^2/r_j) + \frac{4}{A^2 W^2}\left[\left(\sum w_j^2\right)\left(\sum (w_j^2/r_j)\right) - \sum (w_j^4/r_j)\right],$$

$$\delta_2 = \sum u_j g_j \left(1 - \frac{1}{r_j}\right) \Big/ \left(\sum g_j\right)^2,$$

$$\begin{aligned}
\epsilon_2 &= 2\sum u_j g_j/(\sum g_j)^2, \\
\epsilon_3 &= -[\sum (f_j/r_j)]\epsilon_2, \\
\delta' &= \sum(f_j/r_j^2)g_j/\sum g_j, \\
\delta_3 &= \delta' - [\sum (f_j/r_j)]\delta_2, \\
\delta_1 &= \sum(f_j^2/r_j^3) - 2[\sum(f_j/r_j)]\delta' + [\sum(f_j/r_j)]^2\delta_2, \\
\epsilon_1 &= 2\sum(f_j^2/r_j^2) + \frac{2}{A^2W^2}[(\sum(w_j^2/r_j))^2 - \sum(w_j^4/r_j^2)] + (\sum f_i/r_j)^2\epsilon_2.
\end{aligned}$$

(The precise form of these equations has been chosen with computation in mind; it takes account of the fact that the w_i are likely to be small integers.)

9. Special cases. The quantities appearing in the formulas for α_1 to ϵ_5 fall naturally into several groups according to which of the $\{r_i\}$, $\{w_i\}$, and $\{u_i\}$ they

TABLE 1

Quantities depending on w_i and r_i but not on u_i

Quantity	General Form	Form for $w_i = r_i$	Form for $w_i = 1$
W	$\sum w_i$	R	c
a_j	$w_j - \frac{1}{W} w_j^2$	$r_1 - \frac{1}{R} r_i^2$	$1 - \frac{1}{c}$
$\sum a_j^2$	$\sum \left(w_j - \frac{1}{W} w_j^2\right)^2$	$\sum \left(r_i - \frac{1}{R} r_i^2\right)^2$	$c\left(1 - \frac{1}{c}\right)^2$
$A = \sum a_j$	$W - \frac{1}{W}\sum w_j^2$	$R - \frac{1}{R}\sum r_j^2$	$c - 1$
α_1	—	—	$\frac{1}{c}$
—	$(\sum w_i^2) - (\sum w_i^4)$	$(\sum r_i^2)^2 - (\sum r_i^4)$	$(c-1) = AW$
$\sum \frac{a_j^2}{r_j}$	$\sum \frac{1}{r_j}\left(w_j - \frac{1}{W} w_j^2\right)^2$	$\sum r_j \left(1 - \frac{1}{R} r_j\right)^2$	$\left(1 - \frac{1}{c}\right)^2 \sum \frac{1}{r_j}$
—	$(\sum w_j^2)\left(\sum \frac{w_j^2}{r_j} - \sum \frac{w_j^4}{r_j}\right)$	$(\sum r_j)R - \sum r_j^3$	$(c-1)\sum \frac{1}{r_j}$
$\sum \frac{a_j^2}{r_j^2}$	$\sum \frac{1}{r_j^2}\left(w_j - \frac{1}{W} w_j^2\right)^2$	$\sum \left(1 - \frac{1}{R} r_j\right)^2$	$\left(1 - \frac{1}{c}\right)\sum \frac{1}{r_j^2}$
—	$\left(\sum \frac{w_j^2}{r_j}\right)^2 - \sum \frac{w_j^4}{r_j^2}$	$R^2 - \sum r_j^2 = AW$	$\left(\sum \frac{1}{r_j}\right)^2 - \left(\sum \frac{1}{r_j^2}\right)$
$\sum r_j b_j$	$\sum \frac{1}{r_j}\left(w_j - \frac{1}{W} w_j^2\right)$	$c - 1$	$\left(1 - \frac{1}{c}\right)\sum \frac{1}{r_j}$
$\sum r_j b_j^2$	$\sum \frac{a_j^2}{r_j^3} = \sum \frac{1}{r_j^3}\left(w_j - \frac{1}{W} w_j^2\right)^2$	$\sum \frac{1}{r_j}\left(1 - \frac{r_j}{R}\right)^2$	$\left(1 - \frac{1}{c}\right)^2 \sum \frac{1}{r_j^3}$
AW	$W^2 - \sum w_k^2$.	$R^2 - \sum r_j^2$	$c(c-1)$

TABLE 2

Quantities depending on $\{u_i\}$ and $\{r_i\}$ but not on $\{w_i\}$

Quantity	General Form	Form for $u_i \equiv 1$	Form for $u_i \equiv \frac{1}{r_i - 1}$
—	$\sum u_j\,(r_j - 1)$	$R - c$	c
—	$\sum u_j^2\,(r_j - 1)$	$R - c$	$\sum \frac{1}{r_j - 1}$
—	$\sum u_j^2\left(r_j - 2 + \frac{1}{r_j}\right)$	$R - 2c + \sum \frac{1}{r_j}$	$\sum \frac{1}{r_j}$
δ_2		$\frac{1}{R} - \frac{c^2 - \sum \frac{1}{r_j}}{(R - c)^2}$	$\frac{1}{c^2} \sum \frac{1}{r_j}$
			$= \frac{1}{R} + \frac{1}{c^2} \sum \left(\frac{1}{r_j} - \frac{c}{R}\right)$
ϵ_2		$\frac{2}{R - c}$	$\frac{2}{c^2} \sum \frac{1}{r_j - 1}$

involve. The various quantities and their special values are given in Tables 1 and 2 with the exception of δ', which is the only quantity essentially involving both $\{w_i\}$ and $\{u_i\}$. Its special values are:

$$\frac{1}{A}\frac{1}{R - c}\left(c - 1 + \frac{c}{R} - \sum \frac{1}{r_j}\right) \qquad (w_i \equiv r_i, \qquad u_i \equiv 1),$$

$$\frac{1}{A}\frac{1}{c}\sum \frac{1}{r_j} - \frac{1}{R} \qquad \left(w_i \equiv r_i, \qquad u_i \equiv \frac{1}{r_i - 1}\right),$$

$$\frac{1}{c(r - c)}\left(\sum \frac{1}{r_j} - \sum \frac{1}{r_j^2}\right) \qquad (w_i \equiv 1, \qquad u_i \equiv 1),$$

$$\frac{1}{c^2}\sum \frac{1}{r_j^2} \qquad \left(w_i \equiv 1, \qquad u_i \equiv \frac{1}{r_i - 1}\right).$$

It is now quite clear that the result is not likely ever to become algebraically simple, whatever the values of $\{w_i\}$ and $\{u_i\}$.

One of the simplest cases arises when $w_i \equiv r_i$ and $u_i \equiv 1$. Here we have

$$\begin{aligned} \text{var}\,\{\text{between}\} = {} & \left\{\frac{\sum(r_j(R - r_j))^2}{(\sum r_j(R - r_j))^2} - \frac{1}{n}\right\} k_4 \\ & + \left[\frac{R^2 \sum r_j^2 - 2R \sum r_j^3 + (\sum r_j^2)^2}{S^2} - \frac{1}{n - 1}\right] k_{22} \\ & + \frac{4R}{S} k_2 K_2 + \frac{R^2(R - 1)^2}{S^2(R - c)^2}\left[\sum \left(\frac{1}{r_i} - \frac{c}{R}\right)\right] K_4 \\ & + \frac{2(c - 1)(R - 1)R^2}{(r - c)S^2} K_{22}, \end{aligned}$$

$$\text{var}\,\{\text{within}\} = \left(\frac{1}{(R-c)^2}\left\{\sum \frac{1}{r_j} - \frac{c^2}{R}\right\} + \left\{\frac{1}{R} - \frac{1}{N}\right\}\right) K_4 + \left(\frac{2}{R-c} - \frac{2}{N-1}\right) K_{22},$$

$$\text{cov}\,\{\text{between, within}\} = -\frac{R(R-1)}{S(r-c)^2}\left\{\sum \frac{1}{r_j} - \frac{c^2}{R}\right\} K_4 - \frac{2R(c-1)}{S(R-c)} K_{22},$$

where $R = \sum r_i$ and $S = \sum r_i(R - r_i) = R^2 - \sum r_i^2$.

Since the first of these checks with Hammersley's result [1] for the between variance, we can have reasonable confidence that the result is right, since the algebra involved here is somewhat different from his.

10. Numerical examples. In order to learn what these formulas imply, it seems necessary to carry out at least a few numerical examples. The coefficients for several special choices of $\{w_i\}$ and $\{u_i\}$ and each of the following sets of $\{r_i\}$ are given in Tables 3, 4, and 5:

(Set I) $\{r_i\}$ = 10, 10, 10, 5, 5.

(Set II) $\{r_i\}$ = 10, 10, 6, 6, 2, 2.

(Set III) $\{r_i\}$ = 4, 4, 4, 4, 2, 2, 2, 2, 2, 2.

TABLE 3

Coefficients in variance and covariance of variance components for an unbalanced design of structure 5^2, 10^3

w_i, for $r_i = 5$	5	2	1	5	2	1
w_i, for $r_i = 10$	10	3	1	10	3	1
u_i, for $r_i = 5$	1	1	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
u_i, for $r_i = 10$	1	1	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
α_1	.21200	.20429	.20000	.21200	.20428	.20000
β_1	.54080	.51437	.50000	.54080	.51437	.50000
γ_1	.12800	.12996	.14000	.12800	.12996	.14000
δ_1	.00010	.00018	.00031	.00000	.00002	.00008
ϵ_1	.00913	.00997	.01182	.00929	.01014	.01201
δ_2	.02506	.02506	.02506	.02800	.02800	.02800
ϵ_2	.05714	.05714	.05714	.06667	.06667	.06667
δ_3	−.00008	−.00010	−.00014	.00010	.00025	.00048
ϵ_3	−.00731	−.00759	−.00800	−.00853	−.00886	−.00933
Variance of between variance component for $k_{22} = \frac{1}{4}K_{22}$, $k_2K_2 = \sqrt{k_{22}K_{22}}$, and						
$k_4 = -k_{22}$, $K_4 = -K_{22}$	$.621k_{22}$	$.609k_{22}$	$.626k_{22}$	$.622k_{22}$	$.619k_{22}$	$.628k_{22}$
$k_4 = K_4 = 0$	$.833k_{22}$	$.814k_{22}$	$.827k_{22}$	$.834k_{22}$	$.815k_{22}$	$.828k_{22}$
$k_4 = 4k_{22}$, $K_4 = 4K_{22}$	$1.683k_{22}$	$1.634k_{22}$	$1.632k_{22}$	$1.682k_{22}$	$1.632k_{22}$	$1.629k_{22}$

Both common sense and an examination of the tables of coefficients show us that if the between component is much larger than the within component, we will do better, in calculating the between component, to weight the column means equally. Similarly, if the between component is very small, we will do best to weight the column means in proportion to the number of entries. The big question is, Where does the crossover take place? Tables 3, 4, and 5 also give the variance of the between component when the between component is $\frac{1}{4}$ the within component for various degrees of non-normality. When we examine these values, we see that changes in k_4/k_{22} and K_4/K_{22} can have effects which are large compared to the weighting system. We see further that in Set I (two columns of 5 entries and 3 of 10) it is already better to use equal weights when (between) = $\frac{1}{4}$ (within) than to use proportional ones, although a slight further gain can be had from an intermediate weighting system. In Set II (two columns each of 10, 6, and 2) equal weighting is not yet as good as proportional to number. However, a weighting procedure which weights columns of 10 and 6 both twice as much as a column of 2 is better than either for most sorts of non-normality. For this case, where the ratio of extreme column sizes is $10/2 = 5$, equal weighting is better

TABLE 4

Coefficients in variances and covariance of variance components for an unbalanced design of structure $10^2, 6^2, 2^2$

w_i, for $r_i = 2$	2	1	1	2	1	1
w_i, for $r_i = 6$	6	2	1	6	2	1
w_i, for $r_i = 10$	10	2	1	10	3	1
u_i, for $r_i = 2$	1	1	1	1	1	1
u_i, for $r_i = 6$	1	1	1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
u_i, for $r_i = 10$	1	1	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
α_1	.20271	.17638	.16667	.20271	.17638	.16667
β_1	.51349	.42950	.40000	.51349	.42950	.40000
γ_1	.14173	.15728	.20444	.14173	.15728	.20444
δ_1	.00091	.00269	.00639	.00003	.00043	.00187
ϵ_1	.01465	.02328	.04028	.01713	.02689	.04544
δ_2	.02837	.02837	.02837	.04259	.04259	.04259
ϵ_2	.06667	.06667	.06667	.14568	.14568	.14568
δ_3	$-.00074$	$-.00126$	$-.00193$	.00052	.00250	.00510
ϵ_3	$-.01181$	$-.01425$	$-.01704$	$-.02581$	$-.03115$	$-.03723$
Variance of between variance component for $k_{22} = \frac{1}{4}K_{22}$, $k_2K_2 = \sqrt{k_{22}K_{22}}$, and						
$k_4 = -k_{22}$, $K_4 = -K_{22}$	$.649k_{22}$	$.650k_{22}$	$.778k_{22}$	$.663k_{22}$	$.674k_{22}$	$.816k_{22}$
$k_4 = K_4 = 0$	$.856k_{22}$	$.837k_{22}$	$.970k_{22}$	$.866k_{22}$	$.852k_{22}$	$.991k_{22}$
$k_4 = 4K_{22}$, $K_4 = 4K_{22}$	$1.681k_{22}$	$1.586k_{22}$	$1.739k_{22}$	$1.676k_{22}$	$1.564k_{22}$	$1.687k_{22}$

TABLE 5

Coefficients in variances and covariance of variance components for an unbalanced design of structure 4^5, 2^5

w_i, for $r_i = 2$	2	1	2	1
w_i, for $r_i = 4$	4	1	4	1
u_i, for $r_i = 2$	1	1	1	1
u_i, for $r_i = 4$	1	1	$\frac{1}{3}$	$\frac{1}{3}$
α_1	.10900	.10000	.10900	.10000
β_1	.24500	.22222	.24500	.22222
γ_1	.15000	.16667	.15000	.16667
δ_1	.00123	.00366	.00001	.00059
ϵ_1	.03670	.04840	.04050	.05308
δ_2	.03438	.03438	.03750	.03750
ϵ_2	.10000	.10000	.13333	.13333
δ_3	−.00113	−.00195	.00016	.00156
ϵ_3	−.03375	−.03750	−.04500	−.05000
Variance of between variance component for $k_{22} = \frac{1}{4}K_{22}$, $k_2K_2 = \sqrt{k_{22}K_{22}}$, and				
$k_4 = -k_{22}$, $K_4 = -K_{22}$	$.578k_{22}$	$.634k_{22}$	$.598k_{22}$	$.666k_{22}$
$k_4 = K_4 = 0$	$.692k_{22}$	$.749k_{22}$	$.707k_{22}$	$.768k_{22}$
$k_4 = 4k_{22}$, $K_4 = 4K_{22}$	$1.147k_{22}$	$1.208k_{22}$	$1.143k_{22}$	$1.177k_{22}$

TABLE 6

Comparative variance of the between variance component in unbalanced and balanced designs*

Pattern of Columns r_i	R	Variance for $k_4 = K_4 = 0$, $k_{22} = \frac{1}{4}K_{22}$, and $k_2K_2 = \sqrt{k_{22}K_{22}} = \frac{1}{2}K_{22}$
8, 8, 8, 8, 8	40	$.785k_{22}$
5, 5, 10, 10, 10	40	$.814k_{22}$
7, 7, 7, 7, 7	35	$.832k_{22}$
5, 5, 5, 5, 5, 5	30	$.797k_{22}$
2, 2, 6, 6, 10, 10	36	$.837k_{22}$
4, 4, 4, 4, 4, 4	24	$.928k_{22}$
3, 3, 3, 3, 3, 3, 3, 3, 3, 3	30	$.662k_{22}$
2, 2, 2, 2, 2, 4, 4, 4, 4, 4	30	$.692k_{22}$
2, 2, 2, 2, 2, 2, 2, 2, 2, 2	20	$1.089k_{22}$

* With weights chosen from those in Tables 3, 4, and 5 to minimize this variance.

than proportional weighting for k_{22} near, but somewhat smaller, than K_{22}. In Set III (five columns each of 2 and 4) we have not computed an intermediate weighting system. Here proportional weighting is preferred until k_{22} rises to somewhat above K_{22}.

If we examine the effect of changing the $\{u_i\}$, we see that the case $u_i = 1/(r_i - 1)$, which corresponds to pooling mean squares, rather than sums of squares, across columns, is slightly less favorable in each set unless $K_4 = 4K_{22}$, when the reverse holds.

Finally, it is interesting to compare the variances of the between component in the unbalanced designs with those in balanced cases. This is done for one case in Table 6. The loss in effective number of observations for a ratio of 2 to 1 in column sizes is rather small, being perhaps 3 observations in the first and last cases. The loss for the middle case is larger, but not as large as might have been expected from the $10/2 = 5$ ratio of extreme column sizes.

REFERENCES

[1] J. M. Hammersley, "The unbiased estimate and standard error of the interclass variance," *Metron*, Vol. 15 (1949), 189–205.

[2] John W. Tukey, "Variances of variance components: I. Balanced Designs," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 722–736.