BU-417-M

USING THE R( ) NOTATION FOR REDUCTIONS IN SUMS OF SQUARES WHEN FITTING LINEAR MODELS*

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, New York

BU-417-M March, 1972

Summary

$R(\underline{b})$ is defined as the reduction in sum of squares due to fitting the linear model $E(\underline{y}) = \underline{X}\underline{b}$. Thus $R(\underline{b}) = \underline{y}'\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'\underline{y}$. The difference between the reductions due to fitting a model $E(\underline{y}) = \underline{X}_1\underline{b}_1 + \underline{X}_2\underline{b}_2$ and a sub-model $E(\underline{y}) = \underline{X}_1\underline{b}_1$ is defined as

$$R(\underline{b}_2|\underline{b}_1) = R(\underline{b}_1,\underline{b}_2) - R(\underline{b}_1) .$$

This notation provides unequivocal description of sums of squares in the analysis of unbalanced (i.e., 'messy') data. Through a general result for the expected value of $E[R(\underline{b}_2|\underline{b}_1)]$ it also provides estimators of variance components (Henderson's Method 3). It does, however, contain potential pitfalls; for example, in fitting the model for the 2-way crossed classification,

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} ,$$

it must be appreciated that

$$R(\underline{\beta}|\underline{\mu},\underline{\alpha},\underline{\gamma}) \equiv 0 .$$

true only if the two-way classification reduces to a one-way classification but not under the two-way classification for which the $\beta_j \equiv 0$ for all j. That is rank of γ_{ij} remains $(a-1)(b-1)$ and does not change to $a(b-1)$ as for the nested case. W.T.F.

1. Definitions

1.1. Reductions in sums of squares

The R( )-notation is defined by denoting as $R(\underline{b})$ the reduction in sum of squares due to fitting the familiar linear model

---

* Paper prepared for the session on "Messy Data Methodology" at the Spring Regional Meetings of ENAR of the Biometric Society, held in Ames, Iowa, April 26-28, 1972.

$$E(\underline{y}) = \underline{X}\underline{b} \quad . \tag{1}$$

Thus, where $\underline{b}^{\circ}$ is <u>any</u> solution to the normal equations

$$\underline{X}'\underline{X}\underline{b}^{\circ} = \underline{X}'\underline{y} \ , \tag{2}$$

say

$$\underline{b}^{\circ} = (\underline{X}'\underline{X})^{-}\underline{X}'\underline{y} \ , \tag{3}$$

where $(\underline{X}'\underline{X})^{-}$ is a generalized inverse of $\underline{X}'\underline{X}$, meaning that it is any matrix satisfying $\underline{X}'\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'\underline{X} = \underline{X}'\underline{X}$, then

$$R(\underline{b}) = \underline{b}^{\circ\prime}\underline{X}'\underline{y} \ . \tag{4}$$

The right-hand side of (4) represents, in familiar manner, the sum of products of the elements of the solution vector $\underline{b}^{\circ}$ multiplied by the corresponding elements of the right-hand sides $\underline{X}'\underline{y}$ of the normal equations. In this form $R(\underline{b})$ is readily calculated. It can also be expressed as

$$R(\underline{b}) = \underline{y}'\underline{X}(\underline{X}'\underline{X})^{-}\underline{X}'\underline{y} \ . \tag{5}$$

We take (4) and (5) as our formal definition of $R(\underline{b})$.

Suppose $\underline{b}$ is partitioned into 2 vectors $\underline{b}_1$ and $\underline{b}_2$ so that the model is

$$E(\underline{y}) = \underline{X}_1\underline{b}_1 + \underline{X}_2\underline{b}_2 \ . \tag{6}$$

The reduction in sum of squares for fitting this is denoted by

$$R(\underline{b}_1,\underline{b}_2) = \underline{y}'(\underline{X}_1 \quad \underline{X}_2)\begin{bmatrix} \underline{X}_1'\underline{X}_1 & \underline{X}_1'\underline{X}_2 \\ \underline{X}_2'\underline{X}_1 & \underline{X}_2'\underline{X}_2 \end{bmatrix}^{-}\begin{bmatrix} \underline{X}_1' \\ \underline{X}_2' \end{bmatrix}\underline{y} \ , \tag{7}$$

this being the direct analogue of (5). In connection with (6) we might also consider the sub-model

$$E(\underline{y}) = \underline{X}_1\underline{b}_1 \tag{8}$$

for the fitting of which the reduction in sum of squares is

$$R(\underline{b}_1) = \underline{y}'\underline{X}_1(\underline{X}_1'\underline{X}_1)^{-}\underline{X}_1'\underline{y}\ . \tag{9}$$

1.2. Differences between reductions

Differences between reductions in sums of squares are also accommodated by the notation. For example, the difference between $R(\underline{b}_1,\underline{b}_2)$ of (7) and $R(\underline{b}_1)$ of (9) is denoted $R(\underline{b}_2|\underline{b}_1)$:

$$R(\underline{b}_2|\underline{b}_1) \equiv R(\underline{b}_1,\underline{b}_2) - R(\underline{b}_1)\ . \tag{10}$$

In combination with the models (6) and (8) the symbol $R(\underline{b}_2|\underline{b}_1)$ indicates exactly what it means: the reduction in sum of squares due to fitting $E(\underline{y}) = \underline{X}_1\underline{b}_1 + \underline{X}_2\underline{b}_2$ over and above that due to fitting $E(\underline{y}) = \underline{X}_1\underline{b}_1$. It can also be described more succinctly as the reduction in sum of squares due to fitting $\underline{b}_1$ and $\underline{b}_2$ over and above fitting $\underline{b}_1$; or as due to fitting $\underline{b}_2$ after $\underline{b}_1$, in this latter description taking care to understand that by "$\underline{b}_2$ after $\underline{b}_1$" we mean "$\underline{b}_1$ and $\underline{b}_2$ over and above $\underline{b}_1$". But in this manner the meaning of the symbol $R(\underline{b}_2|\underline{b}_1)$ is clarion clear.

The notation is quite general and can be used for regression models, for familiar linear models involving main effects and interactions, and for combinations of the two, namely covariance models.

1.3. The 2-way classification

Adapting the notation to the 2-way classification we use as the reduction in sum of squares

$$R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) \text{ for fitting } E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} . \tag{11}$$

In this model $i = 1,2,\cdots,a$, $j = 1,2,\cdots,b$, $k = 1,2,\cdots,n_{ij}$ for $n_{ij} \neq 0$ for s cells, and γ_{ij} is the interaction effect between the i^{th} level of the α-factor and the j^{th} level of the β-factor. The $\underline{\alpha}$, $\underline{\beta}$ and $\underline{\gamma}$ in the symbol $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ represent the α-effects, the β-effects and the γ-effects respectively.

Similar to (11) we have

$$R(\mu) \quad \text{for fitting } E(y_{ijk}) = \mu \tag{12}$$

$$R(\mu,\underline{\alpha}) \quad \text{for fitting } E(y_{ijk}) = \mu + \alpha_i \tag{13}$$

$$R(\mu,\underline{\beta}) \quad \text{for fitting } E(y_{ijk}) = \mu + \beta_j \tag{14}$$

$$R(\mu,\underline{\alpha},\underline{\beta}) \quad \text{for fitting } E(y_{ijk}) = \mu + \alpha_i + \beta_j . \tag{15}$$

All of these R's are calculated in accord with (5) by appropriate definition of $\underline{X}$ used there. We are concerned here not with methods of calculation but with the clarity of description provided by the R( ) notation.

2. Partitioning and describing sums of squares

Together with

$$SSE = \underline{y}'\underline{y} - R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) ,$$

the reductions in sums of squares shown in (11)-(15) can, as is well known, be used for partitioning the total sum of squares $\underline{y}'\underline{y}$ in at least two different ways.

Table 1

		Sum of squares due to fitting:		
$R(\mu)$	$= R(\mu)$	μ		
$R(\underline{\alpha}\mid\mu)$	$= R(\mu,\underline{\alpha}) - R(\underline{\alpha})$	$\mu,\underline{\alpha}$	after	μ
$R(\underline{\beta}\mid\mu,\underline{\alpha})$	$= R(\mu,\underline{\alpha},\underline{\beta}) - R(\mu,\underline{\alpha})$	$\mu,\underline{\alpha},\underline{\beta}$	after	$\mu,\underline{\alpha}$
$R(\underline{\gamma}\mid\mu,\underline{\alpha},\underline{\beta})$	$= R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) - R(\mu,\underline{\alpha},\underline{\beta})$	$\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}$	after	$\mu,\underline{\alpha},\underline{\beta}$
SSE	$= \underline{y}'\underline{y} - R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$			
Total	$= \underline{y}'\underline{y}$			

Table 2

$R(\mu)$	$= R(\mu)$	μ		
$R(\underline{\beta}\mid\mu)$	$= R(\mu,\underline{\beta}) - R(\underline{\beta})$	$\mu,\underline{\beta}$	after	μ
$R(\underline{\alpha}\mid\mu,\underline{\beta})$	$= R(\mu,\underline{\alpha},\underline{\beta}) - R(\mu,\underline{\beta})$	$\mu,\underline{\alpha},\underline{\beta}$	after	$\mu,\underline{\beta}$
$R(\underline{\gamma}\mid\mu,\underline{\alpha},\underline{\beta})$	$= R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) - R(\mu,\underline{\alpha},\underline{\beta})$	$\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}$	after	$\mu,\underline{\alpha},\underline{\beta}$
SSE	$= \underline{y}'\underline{y} - R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$			
Total	$= \underline{y}'\underline{y}$			

2.1. A variety of descriptions

The descriptions given on the right of these tables are implicit in the R's given on the left, and they relate directly to the models shown in (11)-(15). For example, $R(\underline{\alpha}\mid\mu)$ is the reduction in sum of squares due to fitting $E(y_{ijk}) = \mu + \alpha_i$ of (13) over and above that due to fitting $E(y_{ijk}) = \mu$ of (12). An abbreviated form of the descriptions can also be used; e.g., the terms in Table 1 can be described as

$$\begin{aligned}
&R(\mu) && : \text{due to } \mu \\
&R(\underline{\alpha}|\mu) && : \text{due to } \underline{\alpha}, \text{ after } \mu \\
&R(\beta|\mu,\underline{\alpha}) && : \text{due to } \underline{\beta}, \text{ after } \mu \text{ and } \underline{\alpha} \\
&R(\underline{\gamma}|\mu,\underline{\alpha},\underline{\beta}) && : \text{due to } \underline{\gamma}, \text{ after } \mu, \underline{\alpha} \text{ and } \underline{\beta}\ .
\end{aligned} \tag{16}$$

Alternative descriptions sometimes used are

$$\begin{aligned}
&R(\mu) && : \text{due to } \mu, \text{ ignoring } \underline{\alpha},\underline{\beta},\underline{\gamma} \\
&R(\underline{\alpha}|\mu) && : \text{due to } \underline{\alpha}, \text{ adjusted for } \mu, \text{ ignoring } \underline{\beta},\underline{\gamma} \\
&R(\underline{\beta}|\mu,\underline{\alpha}) && : \text{due to } \underline{\beta}, \text{ adjusted for } \mu,\underline{\alpha}, \text{ ignoring } \underline{\gamma} \\
&R(\underline{\gamma}|\mu,\underline{\alpha},\underline{\beta}) && : \text{due to } \underline{\gamma}, \text{ adjusted for } \mu,\underline{\alpha},\underline{\beta}\ .
\end{aligned} \tag{17}$$

These are sometimes abbreviated through not mentioning μ because its presence is considered to be "obvious"; e.g.

$$\begin{aligned}
&R(\underline{\alpha}|\mu) && : \text{due to } \underline{\alpha}, \text{ ignoring } \underline{\beta},\underline{\gamma} \\
&R(\underline{\beta}|\mu,\underline{\alpha}) && : \text{due to } \underline{\beta}, \text{ adjusted for } \underline{\alpha}, \text{ ignoring } \underline{\gamma} \\
&R(\underline{\gamma}|\mu,\underline{\alpha},\underline{\beta}) && : \text{due to } \underline{\gamma}, \text{ adjusted for } \underline{\alpha},\underline{\beta}\ .
\end{aligned} \tag{18}$$

Of the 4 styles of description, that of Tables 1 and 2 or its abbreviated form (16) is preferred. In (17) the use of "ignoring $\underline{\beta}$ and $\underline{\gamma}$" in the description of $R(\underline{\alpha}|\mu)$ connects this sum of squares to the model involving $\underline{\mu}$, $\underline{\alpha}$, $\underline{\beta}$ and $\underline{\gamma}$; but there is no need for this. $R(\underline{\alpha}|\mu)$ is the difference between the sums of squares due to fitting the two models $E(y_{ijk}) = \mu + \alpha_i$ and $E(y_{ijk}) = \mu$, and this fact is quite unrelated to the model $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$. It is therefore irrelevant to describe $R(\underline{\alpha}|\mu)$ in terms of "ignoring $\underline{\beta}$ and $\underline{\gamma}$. Descriptions

(17) and their shorter form (18), which inaccurately omits reference to μ, are therefore not as appropriate as the descriptions in Table 1 or their shorter form (16); and even (16) demands cautious use to ensure that it is interpreted in the manner of Table 1.

2.2. Possible confusions

The descriptions in Tables 1 and 2 show just what each reduction in sum of squares is; and this is implicit in the corresponding symbol of the R-notation. Furthermore, this notation precludes confusion such as can arise from other inaccurate and loose forms of description. For example, $R(\mu,\underline{\alpha})$, $R(\underline{\alpha}|\mu)$ and $R(\underline{\alpha}|\mu,\underline{\beta})$ are clearly distinct and their meanings are easily recognized. In contrast, the exact meaning of the oft-used and ill-defined phrase "the sum of squares due to the α-effects" is not clear, and there is no guarantee as to which of these terms is meant by it. The positive clarity provided by the R-notation is clearly advantageous.

2.3. Balanced and unbalanced data

Distinguishing between $R(\underline{\alpha}|\mu)$ and $R(\underline{\alpha}|\mu,\underline{\beta})$ is unimportant with balanced (equal subclass numbers) data, because these two sums of squares are then the same. But the distinction is vitally important with unbalanced (unequal subclass numbers) data, because $R(\underline{\alpha}|\mu)$ and $R(\underline{\alpha}|\mu,\underline{\beta})$ then represent two entirely different things; and the notation makes this clear. Furthermore, the notation succinctly identifies the distinction, whereas any careless or incomplete use of words readily clouds it. In addition, although the distinction is well known, many of us have learnt it only after much heartache. Perhaps wider use of the R( )-notation in teaching would lead to quicker understanding of the differences entailed, not only in the relatively simple case of Tables 1 and 2 but also in cases involving more than 2 factors. It would also lead to a better understanding of the relationships of unbalanced data to those of balanced data; e.g., that with balanced data $R(\underline{\alpha}|\mu)$ of Table 1 and $R(\underline{\alpha}|\mu,\underline{\beta})$ of Table 2 both become $bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2$.

Tables 1 and 2 are the basis of analysis of variance tables and ensuing F-statistics, which, under normality assumptions, can be used for testing certain hypotheses (see, e.g., Searle [1971], pp. 305-313). With unbalanced data these hypotheses take no simple form as they do with balanced data, and the R( )-notation is of little help in identifying the different hypotheses. However the notation does have advantages in the estimation of variance components.

3. Variance component estimation

3.1. A general result

It is shown in Searle [1968] that the expected value of $R(\underline{b}_2|\underline{b}_1)$ given in (10) is

$$E[R(\underline{b}_2|\underline{b}_1)] = tr\{\underline{X}_2'[\underline{I} - \underline{X}_1(\underline{X}_1'\underline{X}_1)^{-}\underline{X}_1']\underline{X}_2 E(\underline{b}_2\underline{b}_2')\} + \sigma_e^2[r(\underline{X}_1 \quad \underline{X}_2) - r(\underline{X}_1)] \ . \tag{19}$$

The advantage of this result is that the right-hand side does not involve $\underline{b}_1$; it is in terms of only $E(\underline{b}_2\underline{b}_2')$ and σ_e^2. This means that by judicious partitionings of any model $E(\underline{y}) = \underline{X}\underline{b}$ into various forms $E(\underline{y}) = \underline{X}_1\underline{b}_1 + \underline{X}_2\underline{b}_2$ a series of expressions can be developed from (19) for estimating variance components directly. For example, partitioning $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$ so that μ, $\underline{\alpha}$ and $\underline{\beta}$ constitute $\underline{b}_1$ and $\underline{\gamma}$ constitutes $\underline{b}_2$ makes

$$R(\underline{b}_2|\underline{b}_1) \equiv R(\underline{\gamma}|\mu,\underline{\alpha},\underline{\beta}) \tag{20}$$

and from (19) its expectation is a function of $E(\underline{\gamma}\underline{\gamma}')$ and σ_e^2. Since, in familiar variance components models with the γ-effects being random $E(\underline{\gamma}\underline{\gamma}') = \sigma_\gamma^2 I$, the expectation of (20) is, through (19), a linear combination of σ_γ^2 and σ_e^2. Similarly the expectations of $R(\underline{\beta},\underline{\gamma}|\mu,\underline{\alpha})$ and $R(\underline{\alpha},\underline{\beta},\underline{\gamma}|\mu)$ are linear combinations of σ_β^2, σ_γ^2, σ_e^2 and σ_α^2, σ_β^2, σ_γ^2, σ_e^2 respectively. By this means, together with SSE, estimators of the variance components are obtainable. This is, of course, Henderson's [1953] Method (3). Its whole basis rests upon (19), which is a useful algorithm for applying the method.

3.2. Computing methods

A useful comment on computing (19) is made in Mount and Searle [1972]. They point out that

$$\underline{M} = \underline{I} - \underline{X}_1(\underline{X}_1'\underline{X}_1)^-\underline{X}_1'$$

is symmetric and idempotent so that in (19) the term in $E(\underline{b}_2\underline{b}_2')$ can be expressed as

$$\begin{aligned} \mathrm{tr}\{\underline{X}_2'[\underline{I} - \underline{X}_1(\underline{X}_1'\underline{X}_1)^-\underline{X}_1']\underline{X}_2E(\underline{b}_2\underline{b}_2')\} &= \mathrm{tr}\{(\underline{X}_2'\underline{M}\underline{X}_2)E(\underline{b}_2\underline{b}_2')\} \\ &= \mathrm{tr}\{(\underline{X}_2'\underline{M})(\underline{X}_2'\underline{M})E(\underline{b}_2\underline{b}_2')\} \\ &= \mathrm{tr}\{(\underline{X}_2 - \hat{\underline{X}}_{2(1)})'(\underline{X}_2 - \hat{\underline{X}}_{2(1)})E(\underline{b}_2\underline{b}_2')\} \qquad (21) \end{aligned}$$

where

$$\hat{\underline{X}}_{2(1)} = \underline{X}_1(\underline{X}_1'\underline{X}_1)^-\underline{X}_1'\underline{X}_2$$

is the matrix of the usual least squares predicted values of the columns of $\underline{X}_2$, derived by regressing <u>each</u> column of $\underline{X}_2$ on <u>all</u> the columns of $\underline{X}_1$. For example, for (20), $E(\underline{b}_2\underline{b}_2')$ of (21) is $E(\underline{\gamma}\underline{\gamma}') = \sigma_\gamma^2 I_s$ when s sub-classes have data in them.

Then, on writing the model (11) as

$$\underline{y} = \mu\underline{1} + \underline{X}_\alpha\underline{\alpha} + \underline{X}_\beta\underline{\beta} + \underline{X}_\gamma\underline{\gamma} + \underline{e} \tag{22}$$

the expected value of (20) is, through (19) and (21),

$$\begin{aligned} E[R(\underline{\gamma}|\mu,\underline{\alpha},\underline{\beta})] &= tr\{[\underline{X}_\gamma - \hat{\underline{X}}_{\gamma(\mu,\alpha,\beta)}]'[\underline{X}_\gamma - \hat{\underline{X}}_{\gamma(\mu,\alpha,\beta)}]\sigma^2_\gamma I_s \\ &\qquad + \sigma^2_e\{r[\underline{1}\ \ \underline{X}_\alpha\ \ \underline{X}_\beta\ \ \underline{X}_\gamma] - r[\underline{1}\ \ \underline{X}_\alpha\ \ \underline{X}_\beta]\} \\ &= \sigma^2_\gamma\ tr[\underline{\Delta}'_{\gamma(\mu,\alpha,\beta)}\underline{\Delta}_{\gamma(\mu,\alpha,\beta)}] + \sigma^2_e(s - a - b + 1) \end{aligned} \tag{23}$$

where

$$\begin{aligned} \underline{\Delta}_{\gamma(\mu,\alpha,\beta)} &= \underline{X}_\gamma - \hat{\underline{X}}_{\gamma(\mu,\alpha,\beta)} \\ &= \{\underline{\delta}_{k:\gamma(\mu,\alpha,\beta)}\} \quad \text{for} \quad k = 1,2,\cdots,s \end{aligned} \tag{24}$$

is the N $\times$ s matrix whose columns (denoted by $\underline{\delta}$) are columns of deviations of observed values from predicted values (i.e., columns of estimated residuals) derived by regressing each of the s columns of $\underline{X}_\gamma$ on all the columns of $[\underline{1}\ \ \underline{X}_\alpha\ \ \underline{X}_\beta]$. Substituting (24) into (23) gives

$$E[R(\underline{\gamma}|\mu,\underline{\alpha},\underline{\beta})] = \sigma^2_\gamma \sum_{k=1}^{s}\left(\sum_{\ell=1}^{N}\delta^2_{\ell k:\gamma(\mu,\alpha,\beta)}\right) + \sigma^2_e(s - a - b + 1)\ . \tag{25}$$

where $\delta_{\ell k:\gamma(\mu,\alpha,\beta)}$ is the estimated residual corresponding to the ℓ^{th} element in the k^{th} column of $\underline{X}_\gamma$ after regressing the columns of $\underline{X}_\gamma$ on $[\underline{1}\ \ \underline{X}_\alpha\ \ \underline{X}_\beta]$. Thus the coefficient of σ^2_γ in (25) is a sum of sums of squares of estimated residuals.

The characteristic just developed in (25), that of the coefficient of a σ^2 in (19) being a sum of sums of squares of estimated residuals is true quite generally whenever $E(\underset{\sim}{b}_2\underset{\sim}{b}_2')$ in (19), and equivalently (22), has the form of a diagonal matrix having matrices $\sigma^2\underset{\sim}{I}$ for its sub-matrices. For example, consider deriving $E[R(\underset{\sim}{\alpha},\underset{\sim}{\gamma}|\mu,\underset{\sim}{\beta})]$ for the model (22). For (19) we have $\underset{\sim}{b}_1' = [\mu \quad \underset{\sim}{\beta}']$ and $\underset{\sim}{b}_2' = [\underset{\sim}{\gamma}' \quad \underset{\sim}{\gamma}']$ with $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ having corresponding values. Furthermore, $E(\underset{\sim}{b}_2\underset{\sim}{b}_2') = \begin{bmatrix} \sigma_\alpha^2\underset{\sim}{I}_a & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma_\gamma^2\underset{\sim}{I}_s \end{bmatrix}$.

Hence, from (19)

$$E[R(\underset{\sim}{\alpha},\underset{\sim}{\gamma}|\mu,\underset{\sim}{\beta})] = \mathrm{tr}\left\{\begin{bmatrix} \underset{\sim}{X}_\alpha' \\ \underset{\sim}{X}_\gamma' \end{bmatrix}[\underset{\sim}{I} - \underset{\sim}{W}][\underset{\sim}{X}_\alpha \quad \underset{\sim}{X}_\gamma]\begin{bmatrix} \sigma_\alpha^2\underset{\sim}{I}_a & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma_\gamma^2\underset{\sim}{I}_s \end{bmatrix}\right\}$$

$$+ \sigma_e^2\{r[\underset{\sim}{1} \quad \underset{\sim}{X}_\alpha \quad \underset{\sim}{X}_\beta \quad \underset{\sim}{X}_\gamma] - r[\underset{\sim}{1} \quad \underset{\sim}{X}_\beta]\}$$

where

$$\underset{\sim}{W} = \begin{bmatrix} \underset{\sim}{1}' \\ \underset{\sim}{X}_\beta' \end{bmatrix}\begin{bmatrix} \underset{\sim}{1}'\underset{\sim}{1} & \underset{\sim}{1}'\underset{\sim}{X}_\beta \\ \underset{\sim}{X}_\beta'\underset{\sim}{1} & \underset{\sim}{X}_\beta'\underset{\sim}{X}_\beta \end{bmatrix}^{-}[\underset{\sim}{1} \quad \underset{\sim}{X}_\beta] \,. \tag{26}$$

Thus

$$E[R(\underset{\sim}{\alpha},\underset{\sim}{\gamma}|\mu,\underset{\sim}{\beta})] = \mathrm{tr}\begin{Bmatrix} \underset{\sim}{X}_\alpha'(\underset{\sim}{I} - \underset{\sim}{W})\underset{\sim}{X}_\alpha\sigma_\alpha^2 & \underset{\sim}{X}_\alpha'(\underset{\sim}{I} - \underset{\sim}{W})\underset{\sim}{X}_\gamma\sigma_\gamma^2 \\ \underset{\sim}{X}_\gamma'(\underset{\sim}{I} - \underset{\sim}{W})\underset{\sim}{X}_\alpha\sigma_\alpha^2 & \underset{\sim}{X}_\gamma'(\underset{\sim}{I} - \underset{\sim}{W})\underset{\sim}{X}_\gamma\sigma_\gamma^2 \end{Bmatrix} + \sigma_e^2(s - a)$$

$$= \sigma_\alpha^2 \,\mathrm{tr}[\underset{\sim}{X}_\alpha'(\underset{\sim}{I} - \underset{\sim}{W})\underset{\sim}{X}_\alpha] + \sigma_\gamma^2 \,\mathrm{tr}[\underset{\sim}{X}_\gamma'(\underset{\sim}{I} - \underset{\sim}{W})\underset{\sim}{X}_\gamma] + \sigma_e^2(s - a) \,, \tag{27}$$

and from the form of $\underset{\sim}{W}$ in (26) this is, following the $\underset{\sim}{\Delta}$-notation of (23)

$$E[R(\underset{\sim}{\alpha},\underset{\sim}{\gamma}|\mu,\underset{\sim}{\beta})] = \sigma_\alpha^2 \,\mathrm{tr}[\underset{\sim}{\Delta}_{\alpha(\mu,\beta)}'\underset{\sim}{\Delta}_{\alpha(\mu,\beta)}] + \sigma_\gamma^2 \,\mathrm{tr}[\underset{\sim}{\Delta}_{\gamma(\mu,\beta)}'\underset{\sim}{\Delta}_{\gamma(\mu,\beta)}] + \sigma_e^2(s - a) \,.$$

So we see that the coefficients of σ^2_α and σ^2_γ are sums of sums of squares of estimated residuals just as is that of σ^2_γ in (25).

3.3. Mixed models

A particular advantage of the general result (19) is in mixed models. So long as $\underline{b}_1$ includes all the fixed effects of a mixed model, (19) leads to variance components estimators that are unencumbered by the fixed effects in the model. For example, suppose that in the model (11) and (22) the β's are fixed effects. Then (25) and (27), together with SSE, provide estimators of σ^2_α, σ^2_γ and σ^2_e. This is, of course, one of the merits of Henderson's Method 3 – that it yields estimators of the variance components of a mixed model without interference from the fixed effects.

3.4. A warning

One aspect of (19) must not be overlooked. The form of $R(\underline{b}_2|\underline{b}_1)$ whose expectation is given in (19) is such that $\underline{b}_1,\underline{b}_2$ constitute between them the full model under which expectation is being taken. In other words $R(\underline{b}_2|\underline{b}_1) = R(\underline{b}_1,\underline{b}_2) - R(\underline{b}_1)$ is the difference in reductions in sums of squares due to fitting the full model and some sub-model thereof. This is the case with all the terms considered in sections 3.1 and 3.2. But it is <u>not</u> the case, for example, with $R(\underline{\alpha}\,|\mu,\underline{\beta}) = R(\mu,\underline{\alpha},\underline{\beta}) - R(\mu,\underline{\beta})$. The expected value in (19) does not apply to $R(\underline{\alpha}|\mu,\underline{\beta})$ for the model (11) and (22) because $\mu,\underline{\alpha},\underline{\beta}$ do not constitute the elements of that model, under which expectation is being taken. Of course, the expectation of $R(\underline{\alpha}|\mu,\underline{\beta})$ could be obtained from (19) under the model $y = \mu\underline{1} + \underline{X}_\alpha\underline{\alpha} + \underline{X}_\beta\underline{\beta} + \underline{\epsilon}$, but not under the model (11) and (22).

4. Conditions on the $\underline{b}$'s of $R(\underline{b}_2|\underline{b}_1)$

4.1. Conditions

On the face of it, $R(\underline{b}_2|\underline{b}_1)$ of (10) is a valid expression for any partitioning $\underline{b}' = [\underline{b}_1' \;\; \underline{b}_2']$ of $\underline{b}$ in the model $E(\underline{y}) = \underline{X}\underline{b}$. It goes without saying, however, that in this partitioning all effects corresponding to the same factor should be in either $\underline{b}_1$ or $\underline{b}_2$, this being necessary for purposes of interpretation.

Another condition on $\underline{b}_1$ and $\underline{b}_2$ must also be upheld, in order for $R(\underline{b}_2|\underline{b}_1)$ to be other than identically zero: $\underline{b}_1$ cannot contain the effects of an interaction factor unless all of that factor's lower order interactions and main effects are also in $\underline{b}_1$. The same must also be true for $\underline{b}_1, \underline{b}_2$ taken together.

We illustrate this second condition with an example from the 2-way classification model of (11) and (22).

4.2. Illustration

In denoting $R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma})$ by $R(\underline{b}_2|\underline{b}_1)$ we have $\underline{b}_1' = [\mu \;\; \underline{\alpha}' \;\; \underline{\gamma}']$, which contains interaction effects γ_{ij} but not the main effects β_j involved in those interactions. Because of this, $\underline{b}_1$ does not satisfy the second condition of section 4.1, and $R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma})$ is therefore identically zero. This we now show.

Formally we have

$$R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma}) = R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) - R(\mu,\underline{\alpha},\underline{\gamma}) \,. \tag{28}$$

As in (11), $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ is the reduction in sum of squares due to fitting

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} \,, \tag{29}$$

and so, as is well-known,

$$R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2/n_{ij} \quad \text{for } n_{ij} \neq 0 \,. \tag{30}$$

Similarly $R(\mu,\underline{\alpha},\underline{\gamma})$ is, by definition, the reduction in sum of squares due to fitting the model

$$E(y_{ijk}) = \mu + \alpha_i + \gamma_{ij} \,.$$

This is indistinguishable from the model for a 2-way nested (hierarchical) classification, for which the reduction in sum of squares is familiar as $\sum_{i=1}^{a} \sum_{j=1}^{b} \bar{y}_{ij.}^2/n_{ij.}$; i.e.

$$R(\mu,\underline{\alpha},\underline{\gamma}) = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2/n_{ij} \quad \text{for } n_{ij} \neq 0 \,. \tag{31}$$

Hence

$$R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) \equiv R(\mu,\underline{\alpha},\underline{\gamma}) = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2/n_{ij} \quad \text{for } n_{ij} \neq 0$$

and so, in (28)

$$R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma}) \equiv 0 \,. \tag{32}$$

To emphasize this result we demonstrate it with a small example. The example is then further used to demonstrate a procedure for calculating a non-zero value that can, quite erroneously, be used in place of (32).

4.3. Example

Suppose for 2 rows and 3 columns, i.e., $a = 2$ and $b = 3$ in (11), the numbers of observations in the 6 cells are those of Table 3.

Table 3. n_{ij} values.

i	$j = 1$	$j = 2$	$j = 3$	Totals $n_{i.}$
1	3	2	1	6
2	2	0	2	4
Totals $n_{.j}$	5	2	3	$n_{..} = 10$

The normal equations corresponding to the model (29) for these n_{ij}-values are

$$\begin{array}{l} \underline{\text{Equation number}} \\ \text{i} : \\ \text{ii} : \\ \text{iii} : \\ \text{iv} : \\ \text{v} : \\ \text{vi} : \\ \text{vii} : \\ \text{viii} : \\ \text{ix} : \\ \text{x} : \\ \text{xi} : \end{array} \quad \left[\begin{array}{c:cc:ccc:ccccc} 10 & 6 & 4 & 5 & 2 & 3 & 3 & 2 & 1 & 2 & 2 \\ \hdashline 6 & 6 & \cdot & 3 & 2 & 1 & 3 & 2 & 1 & \cdot & \cdot \\ 4 & \cdot & 4 & 2 & 0 & 2 & \cdot & \cdot & \cdot & 2 & 2 \\ \hdashline 5 & 3 & 2 & 5 & \cdot & \cdot & 3 & \cdot & \cdot & 2 & \cdot \\ 2 & 2 & 0 & \cdot & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot \\ 3 & 1 & 2 & \cdot & \cdot & 3 & \cdot & \cdot & 1 & \cdot & 2 \\ \hdashline 3 & 3 & \cdot & 3 & \cdot & \cdot & 3 & \cdot & \cdot & \cdot & \cdot \\ 2 & 2 & \cdot & \cdot & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot \\ 2 & \cdot & 2 & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot \\ 2 & \cdot & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & 2 \end{array}\right] \left[\begin{array}{c} \mu^o \\ \hdashline \alpha_1^o \\ \alpha_2^o \\ \hdashline \beta_1^o \\ \beta_2^o \\ \beta_3^o \\ \hdashline \gamma_{11}^o \\ \gamma_{12}^o \\ \gamma_{13}^o \\ \gamma_{21}^o \\ \gamma_{23}^o \end{array}\right] = \left[\begin{array}{c} y_{...} \\ \hdashline y_{1..} \\ y_{2..} \\ \hdashline y_{.1.} \\ y_{.2.} \\ y_{.3.} \\ \hdashline y_{11.} \\ y_{12.} \\ y_{13.} \\ y_{21.} \\ y_{23.} \end{array}\right] \qquad (33)$$

The equation numbers i,ii,$\cdots$ are for ease of reference; the dots in the matrix represent zeros; and $\mu^o, \alpha_1^o, \cdots$ are the elements of a solution vector $\underline{b}^o$

for the normal equations $\underline{X}'\underline{X}\underline{b}^{o} = \underline{X}'\underline{y}$ of (2).

Equations (33) are 11 equations in 11 unknowns; but they have rank 5. This is because there are 6 linearly independent relationships among the equations: ii and iii sum to i; iv, v and vi sum to i; vii, viii and ix sum to ii; ix and x sum to iii; vii and x sum to iv; viii and v are the same; and ix and xi sum to vi. Of these 7 relationships the last (and others) is a consequence of the preceding 6, and so they constitute 6 linearly independent relationships. A solution of (33) can therefore be obtained by putting 6 elements of the solution vector equal to zero, crossing out the corresponding equations and solving what remains. The simplest set of 6 elements to put equal to zero is

$$\mu^{o} = 0 = \alpha_1^{o} = \alpha_2^{o} = \beta_1^{o} = \beta_2^{o} = \beta_3^{o} \,. \tag{34}$$

Doing this, and crossing out the corresponding equations, i.e., numbers i through vi, leaves

$$\begin{aligned} 3\gamma_{11}^{o} &= y_{11.} \\ 2\gamma_{12}^{o} &= y_{12.} \\ \gamma_{13}^{o} &= y_{13.} \\ 2\gamma_{21}^{o} &= y_{21.} \\ 2\gamma_{23}^{o} &= y_{23.} \end{aligned} \tag{35}$$

with the familiar solution

$$\gamma_{ij}^{o} = \bar{y}_{ij.} \quad \text{for } n_{ij} \neq 0 \,. \tag{36}$$

Using $R(\underline{b})$ of (4) then gives

$$R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) = \sum_{i=1}^{2} \sum_{j=1}^{3} y_{ij.}^2/n_{ij} \quad \text{for } n_{ij} \neq 0 \tag{37}$$

as in (30).

Now the model for calculating $R(\mu,\underline{\alpha},\underline{\gamma})$ is

$$E(y_{ijk}) = \mu + \alpha_i + \gamma_{ij} \; . \tag{38}$$

The normal equations for this corresponding to the n_{ij}'s of Table 3 are

$$\begin{array}{l} \underline{\text{Equation number}} \\ \text{i}' \; : \\ \text{ii}' \; : \\ \text{iii}' : \\ \text{iv}' \; : \\ \text{v}' \; : \\ \text{vi}' \; : \\ \text{vii}' : \\ \text{viii}' : \end{array} \quad \left[\begin{array}{c:cc:ccccc} 10 & 6 & 4 & 3 & 2 & 1 & 2 & 2 \\ \hdashline 6 & 6 & \cdot & 3 & 2 & 1 & \cdot & \cdot \\ 4 & \cdot & 4 & \cdot & \cdot & \cdot & 2 & 2 \\ \hdashline 3 & 3 & \cdot & 3 & \cdot & \cdot & 2 & \cdot \\ 2 & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & 1 & \cdot & 2 \\ 2 & \cdot & 2 & \cdot & \cdot & \cdot & 2 & \cdot \\ 2 & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & 2 \end{array}\right] \left[\begin{array}{c} \mu^{\circ\circ} \\ \hdashline \alpha_1^{\circ\circ} \\ \alpha_2^{\circ\circ} \\ \hdashline \gamma_{11}^{\circ\circ} \\ \gamma_{12}^{\circ\circ} \\ \gamma_{13}^{\circ\circ} \\ \gamma_{21}^{\circ\circ} \\ \gamma_{23}^{\circ\circ} \end{array}\right] = \left[\begin{array}{c} y_{\cdot\cdot\cdot} \\ \hdashline y_{1\cdot\cdot} \\ y_{2\cdot\cdot} \\ \hdashline y_{11\cdot} \\ y_{12\cdot} \\ y_{13\cdot} \\ y_{21\cdot} \\ y_{23\cdot} \end{array}\right] \tag{39}$$

Here we have 8 equations in 8 unknowns; and they too have rank 5 because there are 3 linearly independent relationships among them: ii' and iii' sum to i'; iv', v' and vi' sum to ii'; and vii' and viii' sum to iii'. A solution is therefore obtained by putting 3 elements of the solution vector equal to zero, the easiest being

$$\mu^{\circ\circ} = 0 = \alpha_1^{\circ\circ} = \alpha_2^{\circ\circ} \; .$$

Using this, and crossing out corresponding equations from (39), namely i' through iii', leaves

$$\begin{aligned} 3\gamma_{11}^{oo} &= y_{11\cdot} \\ 2\gamma_{12}^{oo} &= y_{12\cdot} \\ \gamma_{13}^{oo} &= y_{13\cdot} \\ 2\gamma_{21}^{oo} &= y_{21\cdot} \\ 2\gamma_{23}^{oo} &= y_{23\cdot} \end{aligned} \tag{40}$$

which are exactly the same as (35), with solution

$$\gamma_{ij}^{oo} = \bar{y}_{ij\cdot} \quad \text{for } n_{ij} \neq 0 \ . \tag{41}$$

Hence $R(\underline{b})$ of (4) gives

$$R(\mu,\underline{\alpha},\underline{\gamma}) = \sum_{i=1}^{2}\sum_{j=1}^{3} y_{ij\cdot}^2/n_{ij} \quad \text{for } n_{ij} \neq 0 \tag{42}$$

as in (31); and so $R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma}) = 0$ as in (32). All this is quite straightforward and well-known.

4.3. Other solutions of normal equations

Although the most easily obtainable solution to (33) is that shown in (36) derived from using (34), it is not an uncommon practice to use in place of (34) such expressions as

$$\begin{aligned} \alpha_1^o + \alpha_2^o &= 0 \\ \beta_1^o + \beta_2^o + \beta_3^o &= 0 \\ \gamma_{11}^o + \gamma_{12}^o + \gamma_{13}^o &= 0 \\ \gamma_{21}^o + \gamma_{23}^o &= 0 \\ \gamma_{11}^o + \gamma_{21}^o &= 0 \\ \gamma_{13}^o + \gamma_{23}^o &= 0 \end{aligned} \tag{43}$$

or, perhaps

$$\alpha_2^o = 0 = \beta_3^o = \gamma_{13}^o = \gamma_{23}^o = \gamma_{21}^o = \gamma_{12}^o \,. \tag{44}$$

The first of these, (43), is analogous to the procedure frequently used with balanced data of having effects sum to zero; and (44) is that of putting certain 'last' effects equal to zero. When using (43), they and equations (33) are solved simultaneously. With (44) corresponding equations of (33) are crossed out. In neither case will the solution vector be the same as (36) but, through the well-known invariance property of reductions in sums of squares, $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ will in both cases be exactly as in (37). For example, using (44) in (33) and crossing out equations from (33) that correspond to (44), leaves the equations

$$\begin{bmatrix} 10 & 6 & 5 & 2 & 3 \\ 6 & 6 & 3 & 2 & 3 \\ 5 & 3 & 5 & \cdot & 3 \\ 2 & 2 & \cdot & 2 & \cdot \\ 3 & 3 & 3 & \cdot & 3 \end{bmatrix} \begin{bmatrix} \mu^o \\ \alpha_1^o \\ \beta_1^o \\ \beta_2^o \\ \gamma_{11}^o \end{bmatrix} = \begin{bmatrix} y_{\cdot\cdot\cdot} \\ y_{1\cdot\cdot} \\ y_{\cdot 1 \cdot} \\ y_{\cdot 2 \cdot} \\ y_{11\cdot} \end{bmatrix} \,. \tag{45}$$

The solution to these is

$$\begin{bmatrix} \mu^o \\ \alpha_1^o \\ \beta_1^o \\ \beta_2^o \\ \gamma_{11}^o \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -1 & 3 & 1 & -2 & -3 \\ -1 & 1 & 2 & 0 & -2 \\ 0 & -2 & 0 & 3 & 2 \\ 1 & -3 & -2 & 2 & 4\frac{2}{3} \end{bmatrix} \begin{bmatrix} y_{\cdot\cdot\cdot} \\ y_{1\cdot\cdot} \\ y_{\cdot 1 \cdot} \\ y_{\cdot 2 \cdot} \\ y_{11\cdot} \end{bmatrix} \tag{46}$$

from which the reduction in sum of squares is, from (4),

$$R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) = \mu^{o}y_{...} + \alpha_1^{o}y_{1..} + \beta_1^{o}y_{.1.} + \beta_2^{o}y_{.2.} + \gamma_{11}^{o}y_{11.} \quad . \tag{47}$$

Tedious algebra reduces this to (37).

5. Erroneous computing methods

5.1. Computing procedure

The calculation of $R(\underline{b})$ as outlined in (1), (2) and (4), and typified more specifically in the example just given is as follows.

Table 4. Computing Procedure for $R(\underline{b})$

(a): Write the model as $E(\underline{y}) = \underline{X}\underline{b}$ of (1);

(b): Write the normal equations as $\underline{X}'\underline{X}\underline{b}^{o} = \underline{X}'\underline{y}$ of (2);

(c): Amend the equations to be of full rank, usually by setting some elements of the solution equal to zero, as typified in (34), (40) and (44);

(d): Obtain a solution $\underline{b}^{o}$ from the amended equations;

(e): Calculate $R(\underline{b})$ as $\underline{b}^{o\prime}\underline{X}'\underline{y}$ of (4).

The 5 steps in this computing procedure, and their sequence, are important. The starting point is (a), the model, from which are derived the normal equations (b). Since any solution of these suffices to yield $R(\underline{b})$ we amend the equations as in (c), to derive a solution $\underline{b}^{o}$ in (d) and use it in (e) to calculate $R(\underline{b})$. The expressions already used in this procedure for calculating $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ and $R(\mu,\underline{\alpha},\underline{\gamma})$ for the 2-way classification are summarized in Table 5.

Table 5. Examples of using the Computing Procedure for $R(\underline{b})$ of Table 4

Procedure step	For $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ Case 1	For $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ Case 2	For $R(\mu,\underline{\alpha},\underline{\gamma})$
(a): model	(29)	(29)	(38)
(b) normal equations	(33)	(33)	(39)
(c): amended equations	(35)	(45)	(40)
(d): solution	(36)	(46)	(41)
(e): calculation	(37)	(47)	(42)

As seen in (37) and (42), $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ and $R(\mu,\underline{\alpha},\underline{\gamma})$ are identically equal and so $R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma})$, their difference, is identically zero as previously discussed.

5.2. A wrong use of normal equations

The model corresponding to $R(\mu,\underline{\alpha},\underline{\gamma})$ is the same as that corresponding to $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ after reducing it by omitting the β's. It is important to observe at what point in the computing procedure this reduction of the model has to take place. It is at point (a), the writing down of the model. This is so because (a) is the foundation of the computing procedure. Thus the derivation of $R(\mu,\underline{\alpha},\underline{\gamma})$ from the model corresponding to $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ starts from reducing the model (29) by omitting the β's and so getting the model (38). Derivation of each R( ) then proceeds in accord with Table 4, as shown in Table 5.

An error that is sometimes made is to implement the reduction not at step (a) but after step (c). In the case of the amended equations (35) the reduction of omitting the β's is of no consequence, since the β's are already gone from (35) through being put equal to zero in (34), the precursor of (35). But this inconsequential effect is not universal.

Suppose the amended equations (45) had been used. The calculation of $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma})$ is unaffected, as noted after (47). But if the reduction of omitting β's is now implemented, in anticipation of deriving $R(\mu,\underline{\alpha},\underline{\gamma})$ from (45), the result is in fact, <u>not</u> $R(\mu,\underline{\alpha},\underline{\gamma})$. For, omitting β's from (45) gives the equations

$$\begin{bmatrix} 10 & 6 & 3 \\ 6 & 6 & 3 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} \mu^o \\ \alpha_1^o \\ \gamma_{11}^o \end{bmatrix} = \begin{bmatrix} y_{...} \\ y_{1..} \\ y_{11.} \end{bmatrix} \tag{48}$$

with solution

$$\begin{bmatrix} \mu^o \\ \gamma_1^o \\ \gamma_{11}^o \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 3 & -3 & 0 \\ -3 & 7 & -4 \\ 0 & -4 & 8 \end{bmatrix} \begin{bmatrix} y_{...} \\ y_{1..} \\ y_{11.} \end{bmatrix} . \tag{49}$$

Computing $\underline{b}^{o\prime}\underline{X}'\underline{y}$ from this, as in (4), gives

$$\underline{b}^{o\prime}\underline{X}'\underline{y} = \mu^o y_{...} + \alpha_1^o y_{1..} + \gamma_{11}^o y_{11.}$$

$$= \frac{1}{12}(3y_{...}^2 + 7y_{1..}^2 + 8y_{11.}^2 - 6y_{...}y_{1..} - 8y_{1..}y_{.1.}) \tag{50}$$

and no amount of algebra, tedious or otherwise, will reduce this to $R(\mu,\underline{\alpha},\underline{\gamma})$ of (42).

The clue to the fact that (48) does not lead to $R(\mu,\underline{\alpha},\underline{\gamma})$ is that equations (48) have rank 3 whereas the correct normal equations corresponding to $R(\mu,\underline{\alpha},\underline{\gamma})$, namely (39), have rank 5. Taking account of reducing the model after step (c) of the computing procedure in Table 6 is therefore wrong; the correct place is at its foundation, the model, at step (a).

We will refer to the model $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$ as the <u>full model</u> and the model obtained by omitting the β_j's, namely $E(y_{ijk}) = \mu + \alpha_i + \gamma_{ij}$ as the <u>reduced model</u>. In brief, the computing procedure of Tables 4 and 5 for $R(\mu,\alpha,\beta,\gamma)$ is the sequence

(a) full model

(b) normal equations for full model

(c) amended equations for full model

(d) solution

(e) calculation.

And the sequence for $R(\mu,\underline{\alpha},\underline{\gamma})$ is

(a) $\begin{cases}\text{full model}\\ \text{reduce the model}\end{cases}$

(b) normal equations for reduced model

(c) amended equations for reduced model

(d) solution

(e) calculation

The sequence is <u>not</u>

(a) full model

(b) normal equations for the full model

(c) $\begin{cases}\text{amended equations for the full model}\\ \text{reduce the amended equations}\end{cases}$

(d) solution

(e) calculation

Implementation of the reduction from the full model comes at (a), in the model, and not at (c), in the amended equations.

5.3. Consequences for variance component estimation

If erroneous computing for $R(\mu,\underline{\alpha},\underline{\gamma})$ is used in $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) - R(\mu,\underline{\alpha},\underline{\beta})$ for $R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma})$ the latter will not be zero, as it should be. Use of (19) for evaluating the expected value of $R(\underline{\beta}|\mu,\underline{\alpha},\underline{\gamma})$ and equating it to the erroneous non-zero computed value will therefore give a wrong equation for estimating variance components. True it is that the wrongly computed expression for $R(\mu,\underline{\alpha},\underline{\gamma})$ is just a quadratic form in the vector of observations $\underline{y}$ and could be used for variance component estimation; so it could. But its expected value would not be in accord with what (19) gives for the expected value of $R(\mu,\underline{\alpha},\underline{\beta},\underline{\gamma}) - R(\mu,\underline{\alpha},\underline{\gamma})$.

References

Henderson, C. R. [1953]. Estimation of variance and covariance components. Biometrics 9:226-252.

Mount, T. D. and Searle, S. R. [1972]. Estimating variance components in covariance models. Paper BU-403-M in the Biometrics Unit, Cornell University, Ithaca, New York. Submitted to Econometrica.

Searle, S. R. [1968]. Another look at Henderson's methods of estimating variance components. Biometrics 24:749-788.

Searle, S. R. [1971]. Linear Models. Wiley, New York.