BIAS AND MEAN SQUARE ERROR OF A VARIANCE ESTIMATOR AFTER THREE PRELIMINARY TESTS OF SIGNIFICANCE IN A COMPONENT OF VARIANCE MODEL

by R. C. JAIN, *School of Studies in Statistics, Vikram University, Ujjain* 456010

(*Received* 11 *May* 1976; *after revision* 16 *December* 1976)

In an analysis of variance situation resulting from a balanced four-fold nested classification model, expressions for bias, mean square error of an error variance estimator to be used in subsequent inferences have been derived. A comparison between the mean square errors of the said estimator and the unbiased estimator have been made and recommendations regarding the use of the estimator for different levels of significance, degrees of freedom and population variance ratios have been given.

1. INTRODUCTION

Bancroft (1944) investigated the bias and mean square error of a variance estimator obtained after a preliminary test of equality of two variances. Mosteller (1948) studied the effect of using a preliminary test of significance in estimating the population mean. Bennett (1952) extended the studies of Mosteller by using the preliminary test of significance for equality of mean prior to estimating the mean or testing the hypothesis about mean. The effect of preliminary test on subsequent estimation has been studied by Asano (1960), Kitagawa (1963), Larson and Bancroft (1963), Gupta and Srivastava (1969), Srivastava and Gupta (1965), Singh (1971), and Srivastava (1972). The object of the present investigation is to study the bias, mean square error of a variance estimator obtained after three preliminary tests of significance in an ANOVA situation.

2. STATEMENT OF THE PROBLEM

Consider the following balanced four-fold nested classification experimental model (*see* Graybill 1961, p. 351),

$$y_{ijkmn} = \mu + a_i + b_{ij} + c_{ijk} + d_{ijkm} + e_{ijkmn} \qquad \text{...(1)}$$

where $i = 1, 2, ..., I;\ j = 1, 2, ..., J;\ k = 1, 2, ..., K;\ m = 1, 2, ..., M;$

$$n = 1, 2, ..., N.$$

All the random variables a_i, b_{ij}, c_{ijk}, d_{ijkm} and e_{ijkmn} are assumed to be independent and normally distributed with zero means and variances σ_a^2, σ_b^2, σ_c^2, σ_d^2, and σ_e^2 respectively.

The analysis of variance resulting from (1) is given in Table I.

TABLE I

Analysis of variance. Nested classification

[*All factors random.* $\sigma_b^2 > 0$, $\sigma_c^2 > 0$, $\sigma_d^2 > 0$.]

Source of variation	Degrees of freedom	Mean square	Expected mean square
Between A (treatments)	$I - 1 = n_5$	V_5	
Between B within A (true error)	$I(J-1) = n_4$	V_4	$\sigma_4^2 = \sigma_e^2 + N\sigma_d^2 + MN\sigma_c^2 + KMN\sigma_b^2$
Between C within B (doubtful error 3)	$IJ(K-1) = n_3$	V_3	$\sigma_3^2 = \sigma_e^2 + N\sigma_d^2 + MN\sigma_c^2$
Between D within C (doubtful error 2)	$IJK(M-1) = n_2$	V_2	$\sigma_2^2 = \sigma_e^2 + N\sigma_d^2$
Within D (doubtful error 1)	$IJKM(N-1) = n_1$	V_1	$\sigma_1^2 = \sigma_e^2$

The four mean squares V_i $(i = 1, 2, 3, 4)$ are independently distributed as $\chi_i^2 \sigma_i^2 / n_i$, where χ_i^2 is the central chi-square based on n_i degrees of freedom.

We are interested in the estimation of the true error variance σ_4^2 to be used in certain analysis, say, for testing the hypothesis of no treatment effects. From analysis of variance (Table I), we observe that one obvious estimate of σ_4^2 is V_4, which is unbiased and will have minimum variance if (1) is true. But when there is an uncertainty whether σ_b^2 and/or σ_c^2 and/or σ_d^2 are equal to zero, we construct other estimates of σ_4^2 whose mean square errors may be smaller than that of V_4. We construct below an estimator by adopting the following rule of procedure :

Step 1 : If $V_4/V_3 \geqslant F_1$, use $V = V_4$;

Step 2 : If $V_4/V_3 \leqslant F_1$, $V_{34}/V_2 \geqslant F_3$, use $V = V_{34}$;

Step 3 : If $V_4/V_3 \leqslant F_1$, $V_{34}/V_2 \leqslant F_3$, $V_{234}/V_1 \geqslant F_4$, use $V = V_{234}$;

Step 4 : If $V_4/V_3 \leqslant F_1$, $V_{34}/V_2 \leqslant F_3$, $V_{234}/V_1 \leqslant F_4$, use $V = V_{1234}$.

$$V_{34} = \frac{(n_3V_3 + n_4V_4)}{n_{34}}, \quad V_{234} = \frac{(n_2V_2 + n_3V_3 + n_4V_4)}{n_{234}},$$

$$V_{1234} = \frac{(n_1V_1 + n_2V_2 + n_3V_3 + n_4V_4)}{n_{1234}}; \quad n_{34} = n_3 + n_4, \quad n_{234} = n_2 + n_{34},$$

$n_{1234} = n_1 + n_{234}$; $F_1 = F(n_4, n_3; \alpha_1)$, $F_3 = F(n_{34}, n_2; \alpha_3)$ and $F_4 = F(n_{234}, n_1; \alpha_4)$.

$F(n_i, n_j; \alpha_k)$ refers to the upper $100\,\alpha_k\%$ point of the F-distribution with n_i, n_j degrees of freedom.

3. Expected Value of the Variance Estimator V

The joint density of four independent mean squares V_1, V_2, V_3 and V_4 is given by

$$g(V_1, V_2, V_3, V_4) = C\left[\prod_{i=1}^{4} V_i^{(1/2)n_i - 1}\right]\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{4}\frac{n_iV_i}{\sigma_i^2}\right\}$$

where C is a constant. Let us make the transformation

$$u_1 = \frac{n_2V_2}{n_1V_1\theta_{21}}, \quad u_2 = \frac{n_3V_3}{n_2V_2\theta_{32}}, \quad u_3 = \frac{n_4V_4}{n_3V_3\theta_{43}}, \quad z = \frac{n_1V_1}{n_4},$$

where $$\theta_{21} = \frac{\sigma_2^2}{\sigma_1^2}, \quad \theta_{32} = \frac{\sigma_3^2}{\sigma_2^2}, \quad \theta_{43} = \frac{\sigma_4^2}{\sigma_3^2}.$$

The density of the new variates u_1, u_2, u_3, and z is given by

$$h(u_1, u_2, u_3, z) = K\, z^{a_{1234}-1}\, u_1^{a_{234}-1}\, u_2^{a_{34}-1}\, u_3^{a_4-1} \times \exp\left\{-\frac{a_4 z}{\sigma_1^2}(1 + u_1 + u_1u_2 + u_1u_2u_3)\right\} \qquad \ldots(2)$$

where $$K = \frac{a_4^{a_{1234}}}{[(\sigma_1^2)^{a_{1234}}.\Gamma a_1.\Gamma a_2.\Gamma a_3.\Gamma a_4]} \qquad \ldots(3)$$

$a_i = \frac{1}{2}n_i$, $a_{34} = a_3 + a_4$, $a_{234} = a_2 + a_{34}$, $a_{1234} = a_1 + a_{234}$.

The expected value of V is obtained by adding the conditional expectations which correspond to steps (1) - (4) of the test procedure. Now we derive the expressions for each one of them separately.

Expected Value of V for Step (1)

Let E_1 denote the expected value of the estimator

$$V_4 = zu_1u_2u_3\theta_{21}\theta_{32}\theta_{43} \text{ under step (1).}$$

Then,

$$E_1 = \frac{\theta_{21}\theta_{32}\theta_{43}}{P(D_1)} \int_{u_3=a}^{\infty} \int_{u_2=0}^{\infty} \int_{u_1=0}^{\infty} \int_{z=0}^{\infty} zu_1u_2u_3 \times h(u_1, u_2, u_3, z)\, dz\, du_1\, du_2\, du_3 \quad \ldots(4)$$

where D_1 denotes the domain $\{u_3 \geqslant a\}$ and

$$a = \frac{n_4F_1}{n_3\theta_{43}}.$$

If we integrate (4) and simplify the results, we obtain

$$\frac{E_1}{\sigma_4^2} = \frac{I_{x1}(a_3, a_{4+1})}{P(D_1)}; \quad \ldots(5)$$

where

$$x_1 = \frac{1}{(1 + a)}; \quad I_x(p, q) = \frac{B_x(p, q)}{B(p, q)}$$

and

$$B_x(p, q) = \int_0^x t^{p-1}(1 - t)^{q-1}\, dt.$$

Expected Value of V for Step (2)

Let E_2 denote the expected value of the estimator

$$V_{34} = n_4\theta_{21}\theta_{32}zu_1u_2 \frac{(1 + u_3\theta_{43})}{n_{34}}$$

under Step (2). Then,

$$E_2 = \frac{n_4\theta_{21}\theta_{32}}{n_{34}P(D_2)} \int_{u_3=0}^{a} \int_{u_2=\lambda_1}^{\infty} \int_{u_1=0}^{\infty} \int_{z=0}^{\infty} zu_1u_2(1 + u_3\theta_{43}) \times h(u_1, u_2, u_3, z)\, dz\, du_1\, du_2\, du_3 \quad \ldots(6)$$

where

$$\lambda_1 = \frac{c}{(1 + u_3\theta_{43})}, \quad c = \frac{n_{34}F_3}{n_2\theta_{32}}$$

and D_2 denotes the domain $\{u_3 \leqslant a,\ u_2 \geqslant \lambda_1\}$.

If we integrate (6) and simplify the result, we get

$$\frac{E_2}{\sigma_4^2} = \frac{1}{n_{34}P(D_2)} \Big[n_4I_{x_2}(a_4 + 1, a_3) + \frac{n_3}{\theta_{43}} I_{x_2}(a_4, a_3 + 1) - K_1n_{234}cS_{np} \times$$

(equation continued on p. 5)

$$\times \left\{ \frac{1}{(a_{34} + n + 1)} \frac{1}{(1 + c)^{a_3+n-p} (c + \theta_{43})^{a_4+p}} \right.$$

$$\left. \times \left(\frac{B_{x_3}(a_4 + p, a_3 + n - p + 1)}{\theta_{43}(1 + c)} + \frac{B_{x_3}(a_4 + p + 1, a_3 + n - p)}{(c + \theta_{43})} \right) \right\} \Bigg],$$

where $x_2 = 1 - x_1$, $x_3 = \dfrac{a(c + \theta_{43})}{[1 + c + a(c + \theta_{43})]}$, $K_1 = \dfrac{\Gamma a_{234}}{\Gamma a_2 . \Gamma a_3 . \Gamma a_4}$ and

$$S_{np} = \sum_{n=0}^{a_2-1} (-1)^n \binom{a_2 - 1}{n} c^{a_{34}+n} \left\{ \sum_{p=0}^{n} \binom{n}{p} \right\}.$$

Expected Value of V for Step (3)

Let E_3 denote the expected value of the estimator

$$V_{234} = n_4 \theta_{21} z u_1 \frac{(1 + u_2\theta_{32} + u_2 u_3 \theta_{32}\theta_{43})}{n_{234}}$$

under step (3). Then,

$$E_3 = \frac{n_4\theta_{21}}{n_{234} P(D_3)} \int_{u_3=0}^{a} \int_{u_2=0}^{\lambda_1} \int_{u_1=\lambda_3}^{\infty} \int_{z=0}^{\infty} z u_1 (1 + u_2\theta_{32} + u_2 u_3 \theta_{32}\theta_{43})$$

$$\times h(u_1, u_2, u_3, z)\, dz\, du_1\, du_2\, du_3 \qquad \text{...(7)}$$

where

$$\lambda_3 = \frac{d}{1 + u_2\theta_{32} + u_2 u_3 \theta_{32}\theta_{43}}; \quad d = \frac{n_{234}F_4}{n_1\theta_{21}}$$

and D_3 denotes the domain $\{u_3 \leqslant a,\ u_2 \leqslant \lambda_1,\ u_3 \geqslant \lambda_3\}$.

If we integrate (7) and simplify, we get

$$\frac{E_3}{\sigma_4^2} = \frac{1}{P(D_3)} \left[K_1 S_{gh} \frac{B_{x_3}(a_4 + h, a_3 + g - h)}{\theta_{32}\theta_{43}^{(a_{34}+g)} (1 + c)^{a_3+g-h} (c + \theta_{43})^{a_4+h}} \right.$$

$$+ c K_1 S_{np} \left[\frac{1}{(a_{34} + n + 1)} \frac{1}{(1 + c)^{a_3+n-p} (c + \theta_{43})^{a_4+p}} \right.$$

$$\left. \times \left\{ \frac{B_{x_3}(a_4 + p, a_3 + n - p + 1)}{\theta_{43}(1 + c)} + \frac{B_{x_3}(a_4 + p + 1, a_3 + n - p)}{(c + \theta_{43})} \right\} \right]$$

$$+ dKS_{ij} \left[\frac{n_{1234}}{n_{234}(a_{234} + i + 1)} \right.$$

$$\times \left\{ S_{qlr} \frac{(d + \theta_{32})^{q-r} (d + \theta_{32}\theta_{43})^r B_{x_4}(a_4 + l + r, a_3 + j + q - l - r)}{\theta_{32}\theta_{43}(1 + d)(a_{34} + j + q) H_1^{a_3+j+q-l-r} H_2^{a_4+l+r}} + \right.$$

(*equation continued on p.* 6)

$$+ \frac{c}{(1+d)} S_{klm} \frac{(d+\theta_{32}\theta_{43})^m (d+\theta_{32})^{k-m}}{(a_{34}+j+k+1) H_1^{a_3+j+k-l-m} H_2^{a_4+l+m}}$$

$$\times \left(\frac{B_{x_4}(a_4+l+m, a_3+j+k-l-m+1)}{\theta_{43} H_1} \right.$$

$$\left. + \frac{B_{x_4}(a_4+l+m+1, a_3+j+k-l-m)}{H_2} \right) \Big\} \Big] \Big],$$

where

$$H_1 = 1 + d + cd + c\theta_{32}, \quad H_2 = \theta_{43} + d\theta_{43} + cd + c\theta_{32}\theta_{43};$$

$$x_4 = \frac{aH_2}{aH_2 + H_1};$$

$$S_{gh} = \sum_{g=0}^{a_2} (-1)^g \binom{a_2}{g} c^{a_{34}+g} \left\{ \sum_{h=0}^{g} \binom{g}{h} \right\}, \quad S_{ij} = \sum_{i=0}^{a_1-1} (-1^i)$$

$$\times \binom{a_1-1}{i} d^{a_{234}+i} \left\{ \sum_{j=0}^{i} \binom{i}{j} \frac{1}{(1+d)^{a_2+i-j}} \right\},$$

$$S_{qlr} = \sum_{q=0}^{a_2+i-j} (-1)^q \binom{a_2+i-j}{q} c^{a_{34}+j+q} \left[\sum_{l=0}^{j} \binom{j}{l} \left\{ \sum_{r=0}^{q} \binom{q}{r} \right\} \right]$$

and

$$S_{klm} = \sum_{k=0}^{a_2+i-j-1} (-1)^k \binom{a_2+i-j-1}{k} c^{a_{34}+j+k} \left[\sum_{l=0}^{j} \binom{j}{l} \left\{ \sum_{m=0}^{k} \binom{k}{m} \right\} \right].$$

Expected Value of V for Step (4)

Let E_4 denote the expected value of the estimator

$$V_{1234} = n_4 z \frac{(1 + u_1\theta_{21} + u_1u_2\theta_{21}\theta_{32} + u_1u_2u_3\theta_{21}\theta_{32}\theta_{43})}{n_{1234}}$$

under step (4). Then,

$$E_4 = \frac{n_4}{n_{1234} P(D_4)} \int_{u_3=0}^{a} \int_{u_2=0}^{\lambda_1} \int_{u_1=0}^{\lambda_3} \int_{z=0}^{\infty} z(1 + u_1\theta_{21} + u_1u_2\theta_{21}\theta_{32}$$

$$+ u_1u_2u_3\theta_{21}\theta_{32}\theta_{43}) \, h(u_1, u_2, u_3, z) \, dz \, du_1 \, du_2 \, du_3 \qquad \ldots(8)$$

where D_4 denotes the domain $\{u_3 \leqslant a, \ u_2 \leqslant \lambda_1, \ u_1 \leqslant \lambda_3\}$.

Integrating (8) and simplifying the results, we obtain

$$\frac{E_4}{\sigma_4^2} = \frac{K}{P(D_4)} \Bigg[S^*_{ijklm} \frac{(d+\theta_{32})^{k-m} (d+\theta_{32}\theta_{43})^m B_{x_4}(a_4+l+m, a_3+j+k-l-m)}{\theta_{21}\theta_{32}\theta_{43} H_1^{a_3+j+k-l-m} H_2^{a_4+l+m}}$$

$$+ S_{ij} . S_{qlr} \frac{d(d+\theta_{32})^{q-r} (d+\theta_{32}\theta_{43})^r B_{x_4}(a_4+l+r, a_3+j+q-l-r)}{\theta_{32}\theta_{43}(1+d)(a_{234}+i+1)(a_{34}+j+q) H_1^{a_3+j+q-l-r} H_2^{a_4+l+r}}$$

$$+ S_{ij} . S_{klm} \frac{cd(d+\theta_{32})^{k-m} (d+\theta_{32}\theta_{43})^m}{(a_{234}+i+1)(a_{34}+j+k+1) H_1^{a_3+j+k-l-m} H_2^{a_4+l+m}}$$

$$\times \Bigg\{ \frac{B_{x_4}(a_4+l+m, a_3+j+k-l-m+1)}{H_1\theta_{43}} + \frac{B_{x_4}(a_4+l+m+1, a_3+j+k-l-m)}{H_2} \Bigg\} \Bigg],$$

where

$$S^*_{ijklm} = \sum_{i=0}^{a_1} (-1)^i \binom{a_1}{i} \frac{d^{a_{234}+i}}{(a_{234}+i)} \Bigg[\sum_{j=0}^{i} \binom{i}{j} \frac{1}{(1+d)^{a_2+i-j}}$$

$$\times \Bigg\{ \sum_{k=0}^{a_2+i-j-1} (-1)^k \binom{a_2+i-j-1}{k} \frac{c^{a_{34}+j+k}}{(a_{34}+j+k)} \Bigg\} \Bigg].$$

The expected value of the variance estimator V is obtained by combining the contributions of the four steps given by (5), (6), (7) and (8) respectively.

The amount of bias in the estimator V is given by

$$\text{Bias}(V) = E(V) - \sigma_4^2. \qquad \dots(9)$$

Partial Checks

(i) For $F_1 = F_3 = F_4 = 0$, the four mean squares are never pooled and, therefore, we have $a = c = d = 0$, which implies $x_2 = x_3 = x_4 = 0$ and $x_1 = 1.0$, and

$$E(V) = E(V_4) = \sigma_4^2, \ \text{Bias}(V) = \text{Bias}(V_4) = 0.$$

(ii) For $F_1 = F_3 = F_4 \to \infty$, the four mean squares are always pooled and, therefore, we have $a = c = d \to \infty$, which implies $x_2 = x_3 = x_4 = 1.0$ and $x_1 = 0$ and

$$E(V) = E(V_{1234}) = \frac{(n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2 + n_4\sigma_4^2)}{n_{1234}},$$

$$\text{Bias}(V) = \text{Bias}(V_{1234}) = \frac{n_1(\sigma_1^2 - \sigma_4^2) + n_2(\sigma_2^2 - \sigma_4^2) + n_3(\sigma_3^2 - \sigma_4^2)}{n_{1234}}.$$

4. Mean Square Error of the Variance Estimator V

The mean square error of the variance estimator V is obtained by the following relation

$$\text{M.S.E.}\ (V) = E(V^2) - 2\sigma_4^2\ E(V) + \sigma_4^4. \qquad \ldots(10)$$

To find the mean square error of V, it is necessary to find $E(V^2)$. This is done by the method of evaluation of integrals, as in Section 3. The final expression for $E(V^2)$ expressed as a fraction of σ_4^4 is as follows :

$$\frac{E(V^2)}{\sigma_4^4} = 1 + \frac{2}{n_4} + R_3 I_{x_2}(a_4, a_3 + 2) + R_4 I_{x_2}(a_4 + 1, a_3 + 1) + R_5 I_{x_2}(a_4 + 2, a_3)$$

$$+ K_1' \Bigg[S_{np} \frac{R_6}{(a_{34} + n + 2)(c + \theta_{43})^{a_4+p}(1 + c)^{a_3+n-p}}$$

$$\times \left\{ \frac{B_{x_3}(a_4 + p, a_3 + n - p + 2)}{\theta_{43}^2 (1 + c)^2} + \frac{2B_{x_3}(a_4 + p + 1, a_3 + n - p + 1)}{\theta_{43}(1 + c)(c + \theta_{43})} + \frac{B_{x_3}(a_4 + p + 2, a_3 + n - p)}{(c + \theta_{43})^2} \right\} + S_{gh} \frac{2c}{(a_{34} + g + 1)}$$

$$\times \frac{1}{(1 + c)^{a_3+g-h}(c + \theta_{43})^{a_4+h}\,\theta_{32}\theta_{43}}$$

$$\times \left\{ \frac{B_{x_3}(a_4 + h, a_3 + g - h + 1)}{\theta_{43}(1 + c)} + \frac{B_{x_3}(a_4 + h + 1, a_3 + g - h)}{(c + \theta_{43})} \right\}$$

$$+ S_{gh}^* \frac{B_{x_3}(a_4 + h, a_3 + g - h)}{(a_{34} + g)\,\theta_{32}^2\,\theta_{43}^2\,(1 + c)^{a_3+g-h}(c + \theta_{43})^{a_4+h}} \Bigg]$$

$$+ K' \Bigg[S_{ij} \frac{R_7}{(a_{234} + i + 2)} \Big\{ S_q^*\ S_l S_r^*$$

$$\times \frac{(d + \theta_{32})^{q-r}(d + \theta_{32}\theta_{43})^r B_{x_4}(a_4 + l + r, a_3 + j + q - l - r)}{(1 + d)^2 (a_{34} + j + q)\,\theta_{32}^2\,\theta_{43}^2\, H_1^{a_3+j+q-l-r} H_2^{a_4+l+r}} + S_{qlr}$$

$$\times \frac{2c(d + \theta_{32})^{q-r}(d + \theta_{32}\theta_{43})^r}{(1 + d)(a_{34} + j + q + 1)\,\theta_{32}\theta_{43}\, H_1^{a_3+j+q-l-r} H_2^{a_4+l+r}}$$

$$\times \left(\frac{B_{x_4}(a_4 + l + r, a_3 + j + q - l - r + 1)}{H_1 \theta_{43}} + \frac{B_{x_4}(a_4 + l + r + 1, a_3 + j + q - l - r)}{H_2} \right)$$

$$+ S_{klm} \frac{c^2(d + \theta_{32})^{k-m}(d + \theta_{43}\theta_{32})^m}{(a_{34} + j + k + 1)\, H_1^{a_3+j+k-l-m} H_2^{a_4+l+m}} \times$$

(*equation continued on p.* 9)

$$\times \left\{ \frac{B_{x_4}(a_4 + l + m,\, a_3 + j + k - l - m + 2)}{H_1^2\, \theta_{43}^2} \right.$$

$$+ \frac{2B_{x_4}(a_4 + l + m + 1,\, a_3 + j + k - l - m + 1)}{H_1 H_2 \theta_{43}}$$

$$\left. \left. + \frac{B_{x_4}(a_4 + l + m + 2,\, a_3 + j + k - l - m)}{H_2^2} \right\} \right\}$$

$$+ S_{ij}^* \frac{2d}{(a_{234} + i + 1)\, \theta_{21}\theta_{32}\theta_{43}} \left\{ S_{klm}^* \frac{(d + \theta_{32})^{k-m}}{(a_{34} + j + k + 1)} \right.$$

$$\times \frac{(d + \theta_{32}\theta_{43})^m}{H_1^{a_3+j+k-l-m}\, H_2^{a_4+l+m}} \left(\frac{B_{x_4}(a_4 + l + m + 1,\, a_3 + j + k - l - m + 1)}{H_1 \theta_{43}} \right.$$

$$\left. + \frac{B_{x_4}(a_4 + l + m + 1,\, a_3 + j + k - l - m)}{H_2} \right)$$

$$\left. + S_{slt}^* \frac{(d + \theta_{32})^{s-t}\, (d + \theta_{32}\theta_{43})^t\, B_{x_4}(a_4 + l + t,\, a_3 + j + s - l - t)}{(a_{34} + j + s)\, (1 + d)\, \theta_{32}\theta_{43}\, H_1^{a_3+j+s-l-t}\, H_2^{a_4+l+t}} \right\}$$

$$\left. + S_{ijklm}^{**} \frac{(d + \theta_{32})^{k-m} (d + \theta_{32}\theta_{43})^m\, B_{x_4}(a_4 + l + m,\, a_3 + j + k - l - m)}{\theta_{21}^2\, \theta_{32}^2\, \theta_{43}^2\, H_1^{a_3+j+k-l-m}\, H_2^{a_4+l+m}} \right],$$

where

$$S_{ijklm}^{**} = \sum_{i=0}^{a_1+1} (-1)^i \binom{a_1 + 1}{i} \frac{d^{a_{234}+i}}{(a_{234} + i)} \left[\sum_{j=0}^{i} \binom{i}{j} \frac{1}{(1 + d)^{a_2+i-j}} \right.$$

$$\times \left\{ \sum_{k=0}^{a_2+i-j-1} (-1)^k \binom{a_2 + i - j - 1}{k} \frac{c^{a_{34}+j+k}}{(a_{34} + j + k)} \right.$$

$$\left. \left. \times \left\{ \sum_{l=0}^{j} \binom{j}{l} \left(\sum_{m=0}^{k} \binom{k}{m} \right) \right\} \right\} \right],$$

$$S_{ij} S_q^*\, S_l S_r^* = S_{ij} \left[\sum_{q=0}^{a_2+i-j+1} (-1)^q \binom{a_2 + i - j + 1}{q} c^{a_{34}+j+q} \right.$$

$$\left. \times \left\{ \sum_{l=0}^{j} \binom{j}{l} \left(\sum_{r=0}^{q} \binom{q}{r} \right) \right\} \right],$$

$$S^*_{ij}\,S^*_{slt} = \sum_{i=0}^{a_1} (-1)^i \binom{a_1}{i} d^{a_{234}+i}\left[\sum_{j=0}^{i}\binom{i}{j}\frac{1}{(1+d)^{a_2+i-j}}\right.$$

$$\times\left\{\sum_{s=0}^{a_2+i-j}(-1)^s\binom{a_2+i-j}{s}c^{a_{34}+j+s}\right.$$

$$\left.\left.\times\left\{\sum_{l=0}^{j}\binom{j}{l}\left(\sum_{t=0}^{s}\binom{s}{t}\right)\right\}\right\}\right],$$

$$S^*_{gh} = \sum_{g=0}^{a_2+1}(-1)^g\binom{a_2+1}{g}c^{a_{34}+g}\left[\sum_{h=0}^{g}\binom{g}{h}\right];\; R_3 = \frac{n_3(n_3+2)}{n_{34}^2\,\theta_{43}^2},$$

$$R_4 = \frac{2n_3n_4}{\theta_{43}n_{34}^2},\; R_5 = \frac{(n_4+2)\,(n_4^2 - n_{34}^2)}{n_4 n_{34}^2},$$

$$R_6 = \frac{c^2(n_{34}^2 - n_{234}^2)}{n_{34}^2},\; R_7 = \frac{d^2(n_{234}^2 - n_{1234}^2)}{n_{234}^2};$$

$$K_1' = K_1\left(1 + \frac{2}{n_{234}}\right) \text{ and } K' = K\left(1 + \frac{2}{n_{1234}}\right).$$

The relative efficiency of V to V_4 expressed in percentage is

$$\text{R.E.} = \frac{\dfrac{2\sigma_4^4}{n_4}}{\text{M.S.E.}(V)} \times 100\%.$$

Partial Checks

(i) For $F_1 = F_3 = F_4 = 0$, which implies that $a = c = d = 0$

and $x_1 = 1.0,\; x_2 = x_3 = x_4 = 0,$

and $E(V^2) = \sigma_4^4\left(1 + \dfrac{2}{n_4}\right)$, M.S.E.$(V)$ = M.S.E.$(V_4) = \dfrac{2\sigma_4^4}{n_4}$.

(ii) For $F_1 = F_3 = F_4 \rightarrow \infty$, which implies that $a = c = d \rightarrow \infty$ and

$x_1 = 0,\; x_2 = x_3 = x_4 = 1.0$, and

$$E(V^2) = \frac{2(n_1\sigma_1^4 + n_2\sigma_2^4 + n_3\sigma_3^4 + n_4\sigma_4^4) + (n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2 + n_4\sigma_4^2)^2}{n_{1234}^2},$$

$$\text{M.S.E.}(V) = \text{M.S.E.}(V_{1234}) = \frac{2(n_1\sigma_1^4 + n_2\sigma_2^4 + n_3\sigma_3^4 + n_4\sigma_4^2)}{n_{1234}^2} + [\text{Bias }(V_{1234})]^2.$$

5. Discussion

Here, we discuss the results of bias and the mean square error of the variance estimator V on the basis of the theoretical results derived in Sections (3 and 4) and the numerical results assembled in the Appendix (Tables 1 - 4). From (9) and (10), we note that the bias and mean square error of V are functions of 10 parameters, viz., 4 degrees of freedom n_1, n_2, n_3, n_4; 3 population variance ratios, θ_{21}, θ_{32}, θ_{43} and 3 levels of significance $\alpha_1, \alpha_3, \alpha_4$. According to Bozivich, Bancroft and Hartley (1956), the degrees of freedom are determined and fixed by the ANOVA table and the variance ratios are, in general, unknown to the experimenter. In order to control the bias and to minimize the M.S.E. of V, the three levels of singificance are only at the disposal of the experimenter. In the absence of any *a priori* justification for taking them different, we have taken them equal to α.

The empirical study of bias and relative efficiency has been done for the two sets of degrees of freedom, viz., (i) $n_1 = 8, n_2 = 6, n_3 = 4, n_4 = 2$, (ii) $n_1 = n_2 = n_3 = n_4 = 6$, at $\alpha = 0$, 0.05, 0.10, 0.25, 1.00 corresponding to a set of values of θ_{21}, θ_{32}, θ_{43}.

First, we consider the bias in V. Tables 1 and 2 give the bias for the two sets expressed as a fraction of σ_4^2. It increases numerically as we increase the level of singnificance, except at $\alpha = 1.0$, where it reduces to zero. This is in accordance with the results obtained by Bancroft (1944), Singh (1971) and Srivastava (1972). For both the sets of degrees of freedom and $\alpha \leqslant 0.10$, the bias first decreases and then increases as θ_{43} increases for fixed values of θ_{32} and θ_{21}. It decreases continuously for $\alpha = 0.25$. For $1.0 \leqslant (\theta_{32}, \theta_{21}) \leqslant 5.0$ and $1.0 < \theta_{43} \leqslant 5.0$, $\alpha > 0.10$ and $\alpha \geqslant 0.25$ give an adequate control of bias for $n_1 = 8$, $n_2 = 6$, $n_3 = 4$, $n_4 = 2$ and $n_1 = n_2 = n_3 = n_4 = 6$ respectively.

For small n_4, the estimator V is more efficient or equally efficient as the unbiased estimator V_4, but for large n_4 and θ_{43} small, V is more efficient over V_4. The relative efficiency is maximum at $\alpha = 0$ and for small values of θ_{43}, it decreases monotonically as α increases. It also decreases monotonically as θ_{43} and/or θ_{32} and/or θ_{21} increase. It appears from Tables 1–4 that θ_{21} does not have a major contribution, while θ_{32} has some contribution and θ_{43} has the maximum contribution towards the bias and mean square error of V.

Indiscriminate use of V_4 is not desirable, because it sometimes fails to take advantage of other estimates of σ_4^2, which have less mean square error than that of V_4. When $\theta_{43} > 3.0$, $\alpha \geqslant 0.25$ is recommended and for $\theta_{43} \leqslant 3.0$, $\alpha = 0.05$ is recommended as the preliminary level of significance.

ACKNOWLEDGEMENT

The author is grateful to Dr. V. P. Gupta for suggesting the problem.

REFERENCES

Asano, C. (1960). Estimation after preliminary tests of significance and their applications to biometrical researches. *Bull. math. Statist. Fukuoha*, **9**, 1-24.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. math. Statist.*, **15**, 190-204.

Bennet, B. M. (1952). Estimation of means on the basis of preliminary tests of significance. *Ann. Inst. statist., Math. Tokyo, Japan* 31-43.

Bozivich, H., Bancroft, T. A., and Hartley, H. O. (1956). Power of analysis of variance test procedures for certain incompletely specified models I. *Ann. math. Statist.*, **27**, 1017-43.

Graybill, F. A. (1961). An Introduction to Linear Statistical Models, Vol. I. McGraw-Hill Book Co., New York.

Gupta, V. P., and Srivastava, S. R. (1969). Bias and mean square error of an estimation procedure after two preliminary tests of significance in ANOVA model I. *Sankhya*, Ser. B, **31** (Pt. 3), 319-32.

Kitagawa, T. (1963). Estimation after preliminary tests of significance. *Univ. Calif. Publs. Statist.*, **3**, 147-86.

Larson, H. J., and Bancroft, T. A. (1963). Biases in prediction by regression for certain incompletely specified models. *Biometrika*, **50**, 391-402.

Mosteller, F. (1948). On pooling data. *J. Am. statist. Ass.*, **43**, 231-42.

Singh, J. (1971). Pooling mean squares. *J. Am. statist. Ass.*, **66**, 82-85.

Srivastava, S. R., and Gupta, V. P. (1965). Estimation after preliminary testing in ANOVA model I. *Biometrics*, **21** (3), 752-58.

Srivastava, S. R. (1972). Pooling mean squares in ANOVA model II. *J. Am. statist. Ass.*, **67**, 676-79.

(*For Appendix see pages* 13–16)

APPENDIX

TABLE 1

Bias/σ_4^2 for $n_1 = 8,\ n_2 = 6,\ n_3 = 4,\ n_4 = 2.$

θ_{21}	θ_{32}	θ_{43}	α, Level of significance				
			0	0.05	0.10	0.25	1.0
1.0	1.0	1.0	0	0.1323	0.2268	0.3846	0
		3.0	− 0.6000	− 0.2084	− 0.0811	0.0422	0
		5.0	− 0.7200	− 0.1683	− 0.0652	0.0134	0
	5.0	1.0	− 0.5600	− 0.0148	0.1000	0.2475	0
		3.0	− 0.7867	− 0.1894	− 0.0810	0.0239	0
		5.0	− 0.8320	− 0.1466	− 0.0603	0.0066	0
3.0	1.0	1.0	− 0.2667	0.0365	0.1604	0.3384	0
		3.0	− 0.6889	− 0.2183	− 0.0885	0.0357	0
		5.0	− 0.7733	− 0.1698	− 0.0672	0.0109	0
	5.0	1.0	− 0.6133	− 0.0007	0.1048	0.2474	0
		3.0	− 0.8044	− 0.1866	− 0.0802	0.0239	0
		5.0	− 0.8427	− 0.1455	− 0.0600	0.0066	0
5.0	1.0	1.0	− 0.3200	0.0786	0.1829	0.3407	0
		3.0	− 0.7067	− 0.2065	− 0.0835	0.0361	0
		5.0	− 0.7840	− 0.1643	− 0.0651	0.0111	0
	5.0	1.0	− 0.6240	0.0019	0.1055	0.2475	0
		3.0	− 0.8080	− 0.1862	− 0.0801	0.0239	0
		5.0	− 0.8448	− 0.1453	− 0.0600	0.0066	0

TABLE 2

Bias/σ_4^2 for $n_1 = n_2 = n_3 = n_4 = 6$.

θ_{21}	θ_{32}	θ_{43}	α, Level of significance				
			0	0.05	0.10	0.25	1.0
1.0	1.0	1.0	0	0.0655	0.1138	0.1988	0
		3.0	− 0.5000	− 0.1726	− 0.0838	− 0.0035	0
		5.0	− 0.6000	− 0.1048	− 0.0432	− 0.0033	0
	5.0	1.0	− 0.4000	− 0.0275	0.0403	0.1200	0
		3.0	− 0.6333	− 0.1328	− 0.0671	− 0.0054	0
		5.0	− 0.6800	− 0.0808	− 0.0350	− 0.0034	0
3.0	1.0	1.0	− 0.1667	− 0.0095	0.0636	0.1688	0
		3.0	− 0.5556	− 0.1716	− 0.0832	− 0.0046	0
		5.0	− 0.6333	− 0.1013	− 0.0421	− 0.0035	0
	5.0	1.0	− 0.4333	− 0.0155	0.0441	0.1202	0
		3.0	− 0.6444	− 0.1318	− 0.0669	− 0.0054	0
		5.0	− 0.6867	− 0.0805	− 0.0309	− 0.0034	0
5.0	1.0	1.0	− 0.2000	0.0192	0.0823	0.1716	0
		3.0	− 0.5667	− 0.1633	− 0.0802	− 0.0044	0
		5.0	− 0.6400	− 0.0985	− 0.0413	− 0.0034	0
	5.0	1.0	− 0.4400	− 0.0140	0.0443	0.1202	0
		3.0	− 0.6467	− 0.1317	− 0.0669	− 0.0054	0
		5.0	− 0.6880	− 0.0805	− 0.0349	− 0.0034	0

TABLE 3

Relative efficiency of V to V_4 expressed as percentage for

$n_1 = 8, \; n_2 = 6, \; n_3 = 4, \; n_4 = 2.$

θ_{21}	θ_{32}	θ_{43}	α, Level of significance				
			0	0.05	0.10	0.25	1.00
1.0	1.0	1.0	1000.00	193.50	149.20	113.60	100.00
		3.0	263.16	112.30	106.40	107.80	100.00
		5.0	187.97	95.80	97.70	102.80	100.00
	5.0	1.0	288.68	163.56	140.51	113.99	100.00
		3.0	158.38	111.98	105.39	105.13	100.00
		5.0	142.22	97.73	98.02	101.69	100.00
3.0	1.0	1.0	737.70	200.64	158.15	118.44	100.00
		3.0	203.82	111.09	105.82	107.12	100.00
		5.0	163.87	95.63	97.47	102.45	100.00
	5.0	1.0	245.37	169.03	142.17	114.03	100.00
		3.0	151.63	112.63	105.56	105.14	100.00
		5.0	138.71	97.94	98.09	101.69	100.00
5.0	1.0	1.0	609.76	214.59	164.28	119.19	100.00
		3.0	194.13	113.38	106.75	107.22	100.00
		5.0	159.56	96.51	97.83	102.48	100.00
	5.0	1.0	237.73	170.24	142.41	114.04	100.00
		3.0	150.32	112.71	105.59	105.14	100.00
		5.0	138.02	97.97	98.09	101.69	100.00

TABLE 4

Relative efficiency of V to V_4 expressed as percentage for $n_1 = n_2 = n_3 = n_4 = 6$.

θ_{21}	θ_{32}	θ_{43}	α, Level of significance				
			0	0.05	0.10	0.25	1.00
1.0	1.0	1.0	400.00	142.69	126.98	109.79	100.00
		3.0	120.00	88.39	92.34	101.35	100.00
		5.0	86.96	82.59	91.68	99.12	100.00
	5.0	1.0	163.93	136.17	131.54	116.92	100.00
		3.0	78.53	95.90	96.23	100.74	100.00
		5.0	68.85	88.44	90.88	99.24	100.00
3.0	1.0	1.0	360.00	153.75	141.30	120.82	100.00
		3.0	99.69	88.98	93.08	101.26	100.00
		5.0	78.67	83.56	91.10	99.24	100.00
	5.0	1.0	144.69	143.68	134.08	117 04	100.00
		3.0	76.01	96.37	96.31	100.74	100.00
		5.0	67.58	88.53	93.19	99.24	100.00
5.0	1.0	1.0	322.58	168.35	152.21	123.23	100.00
		3.0	96.15	91.68	94.24	101.38	100.00
		5.0	84.14	84.47	91.42	99.27	100.00
	5.0	1.0	141.16	144.99	134.30	117.04	100.00
		3.0	75.51	96.39	96.34	100.74	100.00
		5.0	67.33	88.53	93.19	99.24	100.00