Jour. Ind. Soc. Ag. Statistics
55(1), 2002 : 47-62

Optimal Design for Point and Interval Estimation of Ratio of Variance Components in One Way ANOVA Model[1]

Himadri Ghosh
Indian Agricultural Statistics Research Institute, New Delhi 110 012

SUMMARY

In the present work, attempt has been made to estimate the ratio of variance components in one-way ANOVA model both in point estimation as well as interval estimation. Till now the variance components σ_g^2 and σ_e^2 in the linear model have been estimated separately from the expectation of the mean square due to different sources of variation. Under the asumption of normality of both the effects of different levels and the error components, the statistical independence between the sum of squares due to levels and errors has been established which helps to find out the unbiased estimator of $\frac{\sigma_g^2}{\sigma_e^2}$. In case of interval estimation the sum of squares due to levels is being represented as a linear combination of χ^2 distribution which will give a set of confidence intervals of $\frac{\sigma_g^2}{\sigma_e^2}$. Finally optimal design has been obtained which will minimize the large sample variance of the estimator and the maximum expected confidence length.

Key words : Variance components, Interval estimation, Large sample variance, Simulation, Optimal design.

1. Introduction

In many practical situations of statistical investigations to which a model applies, it is the deciding factor in determining whether effects are to be considered as random or fixed. In case of numerous real life problems generated from the branches of animal breeding, medication and clinics, manufacturing

---

[1] The paper was presented in the session on ISAS Young Scientist Award during the 55th Annual Conference of the Indian Society of Agricultural Statistics held at CIAE and IISS, Bhopal on 16th January, 2002 and adjudged best for 'ISAS Young Scientist Award, 2001'.

process, it is reasonable to assume that the level/effects of the factor come from a probability distribution and the underlying population of effects are usually considered to have infinite size. For example, animal breeders are interested in using breeding the daughter cows from crossing their cows with a sample of young bulls that are considered to be a random sample from the population of bulls, thereby help to increase the production of economically important products, the variance components σ_e^2 (error variance) and σ_g^2 (variance of the effects) are of much interest. They are needed, for example, for the ratio $h^2 = \frac{4\sigma_g^2}{\left(\sigma_g^2 + \sigma_e^2\right)}$, which is a parameter called heritability that is of great interest in genetics, not only in the breeding of animals but of plants too. A landmark paper dealing with the problem of estimating variance components is Henderson [3]. The present paper is motivated in the direction of estimating the ratio of variance components both in the case of point estimation as well as interval estimation than the estimation of σ_g^2 and σ_e^2 separately in one way ANOVA model where the effects are considered to be random. Due to the nonlinear form of the unbiased estimator of ratio of variance components which is a function of the quadratic forms, large sample variance of the estimator of ratio of variance components has been considered and its performance with respect to its exact variance has been studied by simulation technique. Finally, optimal one way designs have been obtained in the sense that the design minimize the variance and the maximum expected length of the confidence intervals.

2. *Expectation and Sampling Variance of the Sums of Squares*

2.1 Expectation of Sums of Squares

We can represent our model in matrix notation as

$$\mathbf{y} = \mu \mathbf{1} + D_1' \mathbf{g} + \mathbf{e} \tag{2.1}$$

where $\mathbf{y}$ is the vector of n observations, $\mathbf{g}$ is the $p \times 1$ vector of effects of different levels with E(**g**) = 0 and Var (**g**) = $\sigma_g^2 I$, **e** is the error vector with E(**e**) = 0 and Var (**e**) = $\sigma_e^2 I$, and $D_1 = \left(d_{uv}^{(1)}\right)$ is the $p \times n$ level versus observation matrix with $d_{uv}^{(1)} = 1$ if v^{th} observation receives u^{th} level and $d_{uv}^{(1)} = 0$ otherwise. Here **1** represents a column vector of all ones and I denotes an identity matrix. We assume that D_1 has full row rank. Equivalently

$$Y = X\begin{pmatrix} \mu \\ g \end{pmatrix} + \mathbf{e} \text{ where } X = \begin{pmatrix} 1 & D_1' \end{pmatrix}$$

In general, for a matrix X = (X_1 X_2), we have an identity among the matrices of quadratic forms given by

$$X(X'X)^{-}X' = X_1(X_1'X_1)^{-}X_1' + M_1X_2(X_2'M_1X_2)^{-}X_2'M_1 \qquad (2.2)$$

where T^{-} is a g–inverse of a matrix T and $M_1 = I - X_1(X_1'X_1)^{-}X_1'$ is idempotent matrix. From this identity we can obtain three quadratic forms, that is, the total corrected sum of squares (SST), the sum of squares due to levels (SSL) and the sum of squares due to error (SSE). Partitioning SST into SSL and SSE based on Henderson's Method III, we have

$$SST = SSL + SSE$$

Now

$$SST = \mathbf{y}' M \,\mathbf{y} \qquad (2.3)$$

$$SSL = \mathbf{y}' \left[MD_1' \left(D_1MD_1'\right)^{-} D_1M\right] \mathbf{y} \qquad (2.4)$$

$$SSE = \mathbf{y}' M_0\mathbf{y} \qquad (2.5)$$

where $M = I - \frac{1}{n}\mathbf{1}\mathbf{1}'$ and $M_0 = I - (1 \;\; D_1')[(1 \;\; D_1')'(1 \;\; D_1')]^{-}(1 \;\; D_1')'$

We now obtain the expected values of SSL and SSE using results given in Searle *et al.* [7]. Let $R = D_1D_1' = \text{Diag}(r_1 \ldots r_p)$ and $\mathbf{r} = (r_1 \ldots r_p)' = D_1\mathbf{1}$. Using the definition of D_1 it can be verified, r_i gives the replication of level i, for $i = 1(1)p$ and the vector **r** of order $p \times 1$ is explained similarly. Also since we assume Rank (D_1) = p and tr R = n, where for a square matrix A, tr (A) stands for trace, we obtain

$$\begin{aligned}
E[SSL] &= \sigma_g^2 \,\text{tr}\left[MD_1' D_1\right] + \sigma_e^2 \;(\text{Rank}\;(1 \;\; D_1') - \text{Rank}(\mathbf{1}))\\
&= \sigma_g^2 \,\text{tr}\left[\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)D_1'D_1\right] + \sigma_e^2\,(p-1)\\
&= \sigma_g^2 \,\text{tr}\left[D_1D_1' - \frac{1}{n}(D_1 \;\; \mathbf{1})(D_1 \;\; \mathbf{1})'\right] + \sigma_e^2(p-1)\\
&= \sigma_g^2 \,\text{tr}\left[R - \frac{1}{n}\mathbf{r}\mathbf{r}'\right] + (p-1)
\end{aligned}$$

and $\quad E[SSE] = (n-p)\,\sigma_e^2$

Let $C_0 = R - \frac{1}{n}\mathbf{r}\mathbf{r}'$ and $W = D_1' - \frac{1}{n}\mathbf{1}\mathbf{r}'$

Then we can write

$$\left.\begin{aligned}
E[SSL] &= \sigma_g^2 \;\text{tr}\,C_0 + \sigma_e^2\,(p-1)\\
E[SSE] &= (n-p)\sigma_e^2
\end{aligned}\right\} \qquad (2.6)$$

2.2 Dispersion Matrix of $\begin{pmatrix} SSL \\ SSE \end{pmatrix}$

Now $SSL = \mathbf{y}'(MD'_1(D_1MD'_1)^- D_1M)\mathbf{y} = \mathbf{y}' WC_0^- W'\mathbf{y}$ (2.7)

since $$MD'_1 = D'_1 - \frac{1}{n}\mathbf{1}\mathbf{r}' = W \quad (2.8)$$

and $$D_1MD'_1 = D_1\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)D'_1 = D_1D'_1 - \frac{1}{n}\mathbf{r}\mathbf{r}' = C_0 \quad (2.9)$$

Note that $D_1W = C_0$ and $W'W = C_0$

Let $A_1 = WC_0^-W'$ then using the results given in Searle *et al.* [7] on variance and covariance of quadratic forms, under normality, we obtain

$$\begin{aligned} Var(SSL) &= Var(\mathbf{y}' WC_0^- W'\mathbf{y}) \\ &= Var(\mathbf{y}' A_1 \mathbf{y}) \\ &= 2\,tr\,(A_1V)^2,\ \text{where } V = Var(\mathbf{y}) = \sigma_g^2 D'_1D_1 + \sigma_e^2 I \\ &= 2\,tr\,\left[WC_0^-W'\left(\sigma_g^2 D'_1D_1 + \sigma_e^2 I\right)\right]^2 \\ &= 2\,tr\left[\sigma_g^4 WC_0^- W'D'_1D_1WC_0^- W'D'_1D_1\right. \\ &\quad + \sigma_g^2\sigma_e^2 WC_0^- W'D'_1D_1WC_0^- W' \\ &\quad \left. + \sigma_g^2\sigma_e^2 WC_0^- W'WC_0^- W'D'_1D_1 + \sigma_e^4 WC_0^- W'WC_0^- W'\right] \end{aligned}$$

Now evaluating each of the above four terms, we have (Ghosh and Das [2]).

$$tr\left[\sigma_g^4 WC_0^- W'D'_1D_1WC_0^- W'D'_1D_1\right] = \sigma_g^4\,tr\,C_0^2$$

$$tr\left[\sigma_g^2\sigma_e^2 WC_0^- W'D'_1D_1WC_0^- W'\right]$$

$$= tr\left[\sigma_g^2\sigma_e^2 WC_0^- W'WC_0^- W'D'_1D_1\right] = \sigma_g^2\sigma_e^2\,tr\,C_0$$

and $$tr\left[\sigma_e^4 WC_0^- W'WC_0^- W'\right] = (p-1)\sigma_e^4$$

Therefore, $Var(SSL) = 2\{\sigma_g^4\,tr\,C_0^2 + 2\,\sigma_e^2\sigma_g^2\,tr\,C_0 + (p-1)\,\sigma_e^4\}$

Let $A_2 = M_0$. Now using the fact $D_1M_0 = 0$ and $A_1M_0 = 0$, we have

$A_1VA_2 = A_1(\sigma_g^2\ D'_1D_1 + \sigma_e^2 I)\,M_0 = A_1(\sigma_g^2\ D'_1D_1M_0 + \sigma_e^2 M_0) = 0$

Therefore, $Cov(SSL, SSE) = Cov(\mathbf{y}' A_1\mathbf{y}, \mathbf{y}'A_2\mathbf{y}) = 2\,tr[A_1VA_2V] = 0$

Finally, $V(SSE) = 2(n-p)\sigma_e^4$

Then we have

$$\text{Disp}\begin{pmatrix} SSL \\ SSE \end{pmatrix} = \begin{pmatrix} 2\{\sigma_g^4 \text{ tr } C_0^2 + 2\sigma_e^2\sigma_g^2 + (p-1)\sigma_e^4\} & 0 \\ 0 & 2(n-p)\sigma_e^4 \end{pmatrix} \quad (2.10)$$

3. Unbiased Estimation of Ratio of Variance Components $\frac{\sigma_g^2}{\sigma_e^2}$ *Under Normality*

Before obtaining the unbiased estimator of the ratio of variance components we shall state and prove the following results and theorems which will be used subsequently.

Result 3.1. Let $\mathbf{y}^{n\times 1}$ follows $N(\mu, V)$. Then

$\mathbf{y}'A\mathbf{y} \sim \chi'^2\ (\vartheta = R(A), \delta^2 = \mu' A\mu)$ iff AV is idempotent

Proof. Since $\mathbf{y} \sim N(\mu, V)$, then

$$V^{-1/2}\mathbf{y} = \mathbf{z} \sim N(V^{-1/2}\mu, I) \text{ where } V = V^{1/2}V^{1/2}$$

Now, $\mathbf{y}' A \mathbf{y} = \mathbf{z}' V^{1/2} A V^{1/2}\mathbf{z}$ where by Fisher - Cochran theorem

$$\mathbf{z}' V^{1/2} A V^{1/2}\mathbf{z} \sim \chi'^2\ (\vartheta, \delta^2)$$

iff $V^{1/2} A V^{1/2} V^{1/2} A V^{1/2} = V^{1/2} A V^{1/2}$

$\Leftrightarrow AVAV^{1/2} = AV^{1/2}$

$\Leftrightarrow AVAV = AV$ i.e AV is idempotent

Hence the proof.

Result 3.2. Let $\mathbf{y}^{nx1} \sim N(\mu, V)$, $V > 0$, $A, B \geq 0$. Then $\mathbf{y}' A \mathbf{y}$ and $\mathbf{y}' B \mathbf{y}$ are independently distributed iff $A V B = 0$.

Proof. Since A and B are positive semidefinite, we can write

$$A = A_1^{n\times r_1} A_1' \text{ and } B = B_1^{n\times r_2} B_1'$$

where r_1 and r_2 are the ranks of A and B.

Then $\mathbf{y}' A \mathbf{y} = \mathbf{y}' A_1 A_1' \mathbf{y}$ and $\mathbf{y}' B \mathbf{y} = \mathbf{y}' B_1 B_1' \mathbf{y}$

Now, $\text{Cov}\ (A_1' Y, B_1' Y) = A_1' V B_1 = 0$

since $AVB = 0$

$\Leftrightarrow (A_1' A_1) A_1' V B_1' (B_1' B_1) = 0$

$\Leftrightarrow A_1' V B_1 = 0$

Based on the above results we have the following theorems.

Theorem 3.1. $\frac{M_0}{\sigma_e^2}\left(\sigma_g^2 D_1' D_1 + \sigma_e^2 I_n\right)$ is idempotent.

Proof. Follows from the fact that $M_0(1\ D_1') = 0$ and M_0 is idempotent, since $X(X'X)^- X'$ is idempotent.

Theorem 3.2. $\mu\ 1' \frac{M_0}{\sigma_e^2} \mu 1 = 0$

Proof. Follows from the first part of Theorem 3.1.

Theorem 3.3. $\frac{Y' M_0 Y}{\sigma_e^2} \sim \chi'^2 (R(M_0) = n - p, \delta^2 = 0)$

Proof. Since $EY = \mu 1$, the proof follows from Result 3.1, Theorem 3.1, Theorem 3.2 and assuming $R(X) = p$.

Theorem 3.4. SSL and SSE are independently distributed.

Proof. The variance and covariance matrix of **y** is $\sigma_g^2\ D_1' D_1 + \sigma_e^2 I_n$

$$\text{Now } A_1 V A_2 = \left[M D_1' (D_1 M D_1')^- D_1 M\right] \left[\sigma_g^2 D_1' D_1 + \sigma_e^2 I_n\right] M_0$$

$$= \sigma_e^2 \left[M D_1' (D_1 M D_1')^- D_1' M\right] M_0 \text{ since } D_1 M_0 = 0$$

$$= 0 \text{ since} \begin{pmatrix} 1' \\ D_1 \end{pmatrix} M_0 = 0 \text{ and } (1\ D_1')[(1\ D_1')'(1\ D_1')]^- (1\ D_1')' M_0 = 0$$

hence the Theorem follows from Result 3.2.

$$\text{Now, } E\left[\frac{SSL}{SSE}\right] = E[SSL] E\left[\frac{1}{SSE}\right], \text{ (by Theorem 3.4)}$$

$$= E[SSL] \frac{1}{\sigma_e^2} E\left[\frac{1}{\frac{SSE}{\sigma_e^2}}\right]$$

$$= E[SSL] \frac{1}{\sigma_e^2} E\left[\frac{1}{\chi_{n-p}^2}\right], \text{ (by result 3.2 and result 3.3)}$$

$$= \left[\sigma_g^2 trC_0 + \sigma_e^2 (p-1)\right] \frac{1}{\sigma_e^2} k(n-p)$$

where $$k(n-p) = E\left[\frac{1}{\chi_{n-p}^2}\right] = \frac{1}{n-p-2}, n-p > 2$$

$$= \left[\frac{\sigma_g^2}{\sigma_e^2} \operatorname{tr}C_0 + (p-1) \right] k(n-p)$$

$$\text{Hence, } E \frac{\frac{SSL}{SSE} - (p-1)\,k(n-p)}{(\operatorname{tr}C_0)\,k(n-p)} = \frac{\sigma_g^2}{\sigma_e^2} \tag{3.1}$$

which gives

$\mathbf{R}_{est} = \dfrac{\frac{SSL}{SSE} - (p-1)\,k(n-p)}{(\operatorname{tr}C_0)\,k(n-p)}$ as an unbiased estimator of ratio of variance components $\dfrac{\sigma_g^2}{\sigma_e^2}$.

4. Large Sample Variance of the Unbiased Estimator of Ratio of Variance Components and Optimal Design

4.1 Large Sample Variance

It is to be noted from the expression of the unbiased estimator that it is a function of the ratio of SSL and SSE. Due to the random effects of the levels in the model, SSL does not follow a multiple of χ^2 distribution for general unbalanced data related to one way ANOVA model with n observations and p levels. SSL follows as a linear combination of independent χ^2 random variables with $p_h - 1$ degrees of freedom where p_h is the number of levels in the h^{th} group having each of the levels in the h^{th} group being equal replication r_h, h = 1(1)c (see section 5) plus c − 1 χ^2 random each having one degree of freedom and the coefficient as a function of r_h, σ_g^2 and σ_e^2 (La Motte [4]).

Keeping in mind the difficulties of deriving the exact variance expression of the unbiased estimator, we shall derive the large sample variance of the unbiased estimator by using the results given in C.R. Rao [6]. A simulation study for the comparison of the large sample variance expression with the variance of the unbiased estimator based on simulated observation with different parametric values of σ_g^2 and σ_e^2 has been carried out in Section 4.3.

Let $g(SSL, SSE) = \dfrac{SSL}{SSE}$

Then $$V(\mathbf{R}_{est}) = V\left[\frac{\frac{SSL}{SSE} - (p-1)\,K(n-p)}{(\text{tr}\ C_0)\,K(n-p)}\right]$$

$$= V\left[\frac{(n-p-2)}{\text{tr}\,C_0}\ \frac{SSL}{SSE}\right]$$

$$= \frac{1}{s^2}\,V\left[\frac{SSL}{(\text{tr}\,C_0/s)}\Big/(SSE/n-p-2)\right]$$

where s is the replication value.

$$= \frac{1}{p}\frac{1}{s^2}\left[\left(\frac{\partial g(SSL^*, SSE^*)}{\partial SSE}\right)^2 \Big/_{ESSL^*,\,ESSE^*} V(\sqrt{p}\ SSL^*)\right.$$

$$+ 2\,\frac{\partial g(SSL^*, SSE^*)}{\partial SSL^*}\,\frac{\partial g(SSL^*, SSE^*)}{\partial SSE^*}\Big/_{ESSL^*,\,ESSE^*} \text{Cov}(SSL^*, SSE^*)$$

$$\left. + \left[\left(\frac{\partial g(SSL^*, SSE^*)}{\partial SSE^*}\right)^2 \Big/_{ESSL^*,\,ESSE^*} V(\sqrt{p}\ SSE^*)\right.\right.$$

$$= \frac{2}{ps^2}\left[\frac{1}{(ESSE^*)^2}\,\frac{ps^2}{\text{tr}^2\,C_0}\left\{\sigma_g^4\ \text{tr}\ C_0^2 + 2\ \sigma_e^2\,\sigma_g^2\ \text{tr}\ C_0 + (p-1)\,\sigma_e^4\right\}\right.$$

$$\left. + \frac{(ESSL^*)^2}{(ESSE^*)^4}\,\frac{1}{(s-1)(n-p-2)}\left\{(n-p)\,\sigma_e^4\right\}\right]$$

$$= \frac{2(n-p-2)^2}{p(n-p)^2\,\sigma_e^4}\left[p\sigma_g^4\,\frac{\text{tr}C_0^2}{\text{tr}^2C_0} + 2\sigma_e^2\,\sigma_g^2\left(p + \frac{(p-1)(n-p-2)}{(n-p)(s-1)}\right)\frac{1}{\text{tr}C_0}\right.$$

$$\left. + (p-1)\sigma_e^4\left(p + \frac{(p-1)(n-p-2)}{(n-p)(s-1)}\right)\frac{1}{\text{tr}^2C_0} + \sigma_g^4\,\frac{(n-p-2)}{(n-p)(s-1)}\right] \qquad (4.1)$$

4.2 Optimal Design

In section 4.1, we have explicitly obtained the variance of $\frac{\hat{\sigma}_g^2}{\sigma_e^2}$. Let D (p, n) be the class of design related to one way classified data with p lines and n observations. We need the following well known result, see, for example, Cheng [1].

Lemma 4.1. For given positive integers ϑ and t, the minimum of $n_1^2 + n_2^2 + .. + n_\vartheta^2$ subject to $n_1 + n_2 \ldots + n_\vartheta = t$, where n_i's are non negative integers, is obtained when $t - \vartheta[t/\vartheta]$ of n_i's are equal to $[t\ \vartheta] + 1$ and $\vartheta - t + \vartheta[t/\vartheta]$ are equal to $[t/\vartheta]$, where [z] denotes the largest integer not exceeding z.

Lemma 4.2. Consider a real symmetric matrix A of order m having rank r. Then $\frac{\text{tr}\, A^2}{(\text{tr}\, A)^2} \geq \frac{1}{r}$

and the equality is attained when the non zero eigen values of A are equal.

Proof. See Ghosh and Das [2].

In order to minimise $V(\mathbf{R}_{est})$ within the class of design D(p, n), it is sufficient to minimise $\frac{\text{tr}\, C_0^2}{(\text{tr}\, C_0)^2}$ and $\frac{1}{(\text{tr}\, C_0)}$.

Lemma 4.3. For any design $d \in D(p, n)$

$$\text{tr}\,(C_{0d}) \leq \frac{n(p-1)}{p}$$

Equality holds iff $r_{di} = n/p = r$ for $i = 1(1)p$

Proof. For any design $d \in D(p, n)$

$$\text{tr}\,(C_{0d}) = \sum_{i=1}^{p} r_{di} - \frac{1}{n}\sum_{i=1}^{p} r_{di}^2$$

$$= n - \frac{1}{n}\sum_{i=1}^{p} r_{di}^2$$

Now, since $\sum_{i=1}^{p} r_{di} = n$ and $n/p = r$, using Lemma 4.1 we get $\sum_{i=1}^{p} r_{di}^2 \geq \frac{n^2}{p}$

Hence, $\text{tr}(C_{0d}) \leq n - \frac{n}{p} = \frac{n(p-1)}{p}$. By Lemma 4.1, equality above is attained iff $r_{di} = n/p = r$ for $i = 1(1)p$.

Making an appeal to the results of Lemma 4.2 and 4.3, we establish the following theorem.

Theorem 4.1. Let $d_0^* \in D(p, n)$ be a design of one way classified data and suppose $C_0 d_0^*$ satisfies

(i) $\text{tr}(C_0 \tau_{ij}) = \frac{n(p-1)}{p}$

(ii) $C_0 d_0^*$ is completely symmetric in the sense that $C_0 d_0^*$ has all its diagonal elements equal and all its off diagonal elements equal

Then d_0^* is optimal in D(p, n).

4.3 Simulation Study

The optimality study of the estimator of the ratio of variance components from its large sample variance expression has been validated by the exhaustive simulation of the observations from normal population with its covariance structure depending upon the design matrix D_1 and the variance components σ_g^2 and σ_e^2. The optimal design for estimating the ratio of variance components $\frac{\sigma_g^2}{\sigma_e^2}$ in the class of design with 20 observations and 5 levels is found out to be d_0^* which is of 20 observations and each of the five levels having replication number 4. The estimator has been computed based on the observation vector of dimension 20 in each of the iteration of the simulation with varying number of iteration ranging from 30, 35, ..., 50 and 100. The variance of the iterated values of the estimator is then computed and compared with numerical value of the large sample variance in (4.1). The observations from normal population have been generated by IML of Random Normal Number in SAS and by Box-Muller transformation of uniformly distributed random variables given in William *et al.* [8].

Table 4.1. $\sigma_g^2 = 1.50, \sigma_e^2 = 0.30$, Large Sample Variance = 10.44575

Iteration Number	Exact Variance by SAS Random Number	Exact Variance by Box-Muller Transformation
30	9.9398163	9.0213099
35	10.9393419	9.0213224
40	9.9389777	9.0213306
45	10.9384685	11.0213362
50	11.9377258	11.0213402
100	9.9405069	10.0213530

Table 4.2. $\sigma_g^2 = 1.25, \sigma_e^2 = 0.25$, Large Sample Variance = 10.45426

Iteration Number	Exact Variance by SAS Random Number	Exasct Variance by Box-Muller Transformation
30	10.3237824	10.3180805
35	9.3237965	10.3180913
40	9.3238056	10.3180982
45	10.3238119	11.3181029
50	10.3238163	10.3181063
100	10.3238306	10.3181172

5. Interval Estimation of the Ratio of Variance Components

5.1 Scalar Notation for the One Way Random Model

In the random one way model, several observations are taken at each of several levels of a treatment factor (or classification). The observation y_{ij}, where i denotes the level and j the observation with level i, are presumed to follow the model

$$Y_{ij} = \mu + g_i + e_{ij},\ i = 1(1)p \tag{5.1}$$

where all g_i, e_{ij} are uncorrelated random variables with zero means; the g_i's have variance $\sigma_g^2 \geq 0$ and the e_{ij}'s have variance $\sigma_e^2 > 0$, and μ, σ_e^2 and σ_g^2 are unknown parameters.

For the developments to follow it is convenient to consider groups of levels such that each level in a group has the same number of observations. Let c be the number of distinct r_h's , say, $r_1 < r_2 < \ldots < r_c$. Let p_h, h = 1, 2, ..., c denote the number of levels having exactly r_h observations. Then the total number of levels is $p = \sum_h p_h$ and the total number of observations is $n = \sum_{h=1} p_h r_h$. Let the j^{th} observation in the i^{th} level in group h be denoted by y_{hij}. Let y_h be the vector of $p_h\ r_h$ observation from p_h levels in group h.

$$\text{Let } K_h = I_{p_h} \otimes \mathbf{1}_{r_h} \tag{5.2}$$

the Kronecker product of an $p_h \times p_h$ identity matrix and an r_h vector of ones.

Then

$$E(\mathbf{y}_h) = \mu \mathbf{1}_{p_h x r_h}, \text{Cov}\left(y_h / \sigma_g^2, \sigma_e^2\right) = \sigma_g^2 K_h K_h' + \sigma_e^2 I, \ h = 1(1)c \tag{5.3}$$

Let $\mathbf{y}' = \left(y_1' \dots y_c'\right)$ and

$$\text{Then } E(\mathbf{y}) = \mu \mathbf{1}_n, \ \text{Cov}(\mathbf{y}/\sigma_g^2, \sigma_e^2) = \sigma_g^2 KK' + \sigma_e^2 I \tag{5.5}$$

The balanced model has c = 1. An unbalanced model with three level having three, five and three observations each, respectively, would have c = 2, $p_1 = 2$, $r_1 = 3$, $p_2 = 1$, $r_2 = 5$.

5.2 μ – Invariant Quadratics Among Level/Groups, Among Groups and within Level

If A is an $n \times n$ real symmetric matrix, the quadratic $Y' A Y$ is called μ invariant, if $(\mathbf{y} + \alpha \mathbf{1})' A(\mathbf{y} + \alpha \mathbf{1}) = \mathbf{y}' A\mathbf{y}$ for all real scaler α. Equivalently $A\mathbf{1} = 0$ or there exists an $n \times n$ symmetric matrix C such that (see La Motte [4], Appendix A1).

$$A = \left(I - \frac{1}{n}\mathbf{11}'\right) C \left(I - \frac{1}{n}\mathbf{11}'\right) \tag{5.6}$$

Let H be an $n \times (n - 1)$ matrix with i^{th} column

$$H_i = \frac{\left(\mathbf{1}_i', -i, 0\right)'}{\sqrt{i(i+1)}} \tag{5.7}$$

The columns of H form an orthonormal base for the subspace of vectors orthogonal to **1**.

$$H'H = I_{n-1}, \ HH' = I - \frac{1}{n}\mathbf{11}' \tag{5.8}$$

Then if $\mathbf{y}' A\mathbf{y}$ is μ invariant

$$\mathbf{y}' A\mathbf{y} = \mathbf{y}' HH' CHH' \mathbf{y} = \mathbf{z}' H' CH\mathbf{z} \tag{5.9}$$

$$\text{with } \mathbf{z} = H' \mathbf{y} \sim \mathbf{N}_{n-1}(0, \ \sigma_g^2 H' KK' H + \sigma_e^2 \ I) \tag{5.10}$$

5.3 Reduction to Q

Following Olsen, *et al.* [5], let $0 = \lambda_0 < \lambda_1 < \dots < \lambda_k$ be the k + 1 distinct eigen values of $H'KK'H$, with multipliers $c_0, c_1, \dots, c_k$ respectively. Let P_i, $i = 0(1)k$ be an $(n - 1) \times c_i$ matrix whose column are orthogonal eigen vectors of $H'KK'H$ corresponding to λ_I. Then the exponent in the density function of **z** becomes

$$\mathbf{z}'(\sigma_g^2 H' KK' H + \sigma_e^2 I)^{-1}\mathbf{z} = \sum_{i=0}^{k} (\sigma_g^2 \lambda_i + \sigma_e^2)^{-1}\mathbf{z}' P_i P_i' \mathbf{z} \qquad (5.11)$$

so that $Q_i = \mathbf{z}' P_i P_i' \mathbf{z}$, i = 0 (1)k are independent and

$$(\sigma_g^2 \lambda_i + \sigma_e^2)^{-1} Q_i \qquad (5.12)$$

follows χ^2 distribution with c_i degrees of freedom, i = 0(1)k.

Preposition 5.1. The quadratic forms Q_i, i = 1(1)k are generated from the linear set in the column space of K which represent the contrast in the observations and the quadratic form Q_0 represents the sum of squares generated from the linear set which are orthogonal to column space of K.

Proof. $Q_i = \mathbf{z}' P_i P_i' \mathbf{z} = \mathbf{y}' H P_i P_i' H' \mathbf{y} = \mathbf{y}' (HP_i)(HP_i)' \mathbf{y}$

Since P_0 is the eigen matrix with order $(n-1) \times C_0$ of the matrix H′ KK′ H corresponding to the eigen value zero, we have

$$(H' KK' H)P_0 = 0$$

$$\Leftrightarrow \quad (H' K)(H' K)' P_0 = 0$$

$$\Leftrightarrow \quad (H' K)' P_0 = 0$$

since column space of (H′ K)(H′ K)′ is same as column space of H′ K

$$\Leftrightarrow \quad K\, HP_0 = 0$$

Hence the second part of the proposition 5.1 is established.

Also $(HP_i)'(HP_0) = P_i' H' HP_0 = P_i' P_0 = 0$, i = 1(1)k

and $(HP_i)'(HP_j) = P_i' H' HP_j = P_i' P_j = 0$, $1 \le i < j \le k$

Hence the first part of the preposition is established.

Further, $\sum_{h,i,j} (y_{hij} - \bar{y}_{...})^2 = \mathbf{y}'HH'\mathbf{y} = \mathbf{z}'\mathbf{z} = \mathbf{z}'\left(\sum_{i=0}^{k} P_i P_i'\right)\mathbf{z} = \sum_{i=0}^{k} Q_i$, so that the Q_i's partition the total SOS into k + 1 independent quadratics that is $Q_0, Q_1, \ldots, Q_k$ constitute an ANOVA table.

Result 5.1. The eigen values of H′ KK′ H are λ_0 with multiplicity n – p and for each $p_i > 1$, $\lambda = r_i$ is an eigen value with multiplicity $p_i - 1$.

The remaining eigen values are roots of

$$g(\lambda) = \sum_{i=1}^{c} p_i r_i \prod_{j \ne i} (\lambda - r_j)$$

Proof. see La Motte [4], Appendix.

The quadratics $Q_0, Q_1, \ldots, Q_k$ are described in Table 5.1.

Table 5.1

Source	d.f.	SOS
Among levels	$p-1$	$Q_1+Q_2+..+Q_k$
Among levels/groups	$p-c$	$Q_1+Q_2+..+Q_c$
Among levels/group1[1]	p_1-1	Q_1
Among levels/group2	p_2-1	Q_2
⋮	⋮	⋮
Among levels/group	p_c-1	Q_c
Contrast 1	1	Q_{c+1}
⋮	⋮	⋮
Contrast c – 1	1	Q_k
Within levels	$n-p$	Q_0

[1]Among levels /within groups h should be omitted if $p_h = 1$

5.3 Construction of Confidence Interval with Confidence Coefficient $1-\alpha$

From (5.12) we get $(\sigma_g^2 r_h+\sigma_e^2)^{-1}Q_h$ follows a χ^2 distribution with $p_h-1, (p_h > 1)$ d.f. independent of $\sigma_e^{-2} Q_0$ which will follow a χ^2 distribution with $n-p$ d.f.

Set $\alpha^* = \alpha/c$, where c is the number of distinct r_h's in the design.

Let $L_{1d} = F_{1-\alpha^*/2, p_1-1, n-p}$ and $U_{1d} = F_{\alpha^*/2, p_1-1, n-p}$

$$\text{Then, } P\left[L_{1d} \le \frac{(p_1-1)^{-1}(\sigma_e^2+r_1\sigma_g^2)^{-1}Q_1}{(n-p)^{-1}\sigma_e^{-2}Q_0} \le U_{1d}\right] = 1-\alpha^*$$

$$\Leftrightarrow \quad P\left[\frac{(n-p)Q_1}{r_1(p_1-1)U_{1d}Q_0} - \frac{1}{r_1} \le \frac{\sigma_g^2}{\sigma_e^2} \le \frac{(n-p)Q_1}{r_1(p_1-1)L_{1d}Q_0} - \frac{1}{r_1}\right] = 1-\alpha^* \quad (5.13)$$

gives a confidence interval of $\frac{\sigma_g^2}{\sigma_e^2}$ with confidence coefficient $1-\alpha$.

After some simplification and using the result E $F_{n,m} = m/(m-2)$, the expected confidence length of the interval given in (5.13) comes out to be

$$\left(\frac{\sigma_g^2}{\sigma_e^2}+\frac{1}{r_1}\right)\left(\frac{n-p}{n-p-2}\right) \qquad (5.14)$$

by normalising the expected length with respect to the distance between L_{1d} and U_{1d} defined as $\left(\frac{1}{L_{1d}}-\frac{1}{U_{1d}}\right)$.

The other c − 1 confidence intervals of $\frac{\sigma_g^2}{\sigma_e^2}$ can be constructed as follows

$$\left(\frac{(n-p)Q_h}{r_h(p_h-1)U_{hd}\,Q_0}-\frac{1}{r_h},\ \frac{(n-p)Q_h}{r_h(p_h-1)L_{hd}\,Q_0}-\frac{1}{r_h}\right) \qquad (5.15)$$

with confidence coefficient $1-\alpha^*$. Thus the confidence length of the h^{th} confidence interval is

$$\left(\frac{(n-p)Q_h}{r_h(p_h-1)L_{hd}\,Q_0}-\frac{(n-p)Q_h}{r_h(p_h-1)U_{hd}\,Q_0}\right),\ h=1(1)c$$

and the expected length of the confidence interval being

$$\left(\frac{\sigma_g^2}{\sigma_e^2}+\frac{1}{r_h}\right)\left(\frac{n-p}{n-p-2}\right),\ h=1(1)c \qquad (5.16)$$

From (5.13) and (5.15) and by applying Bonferroni's inequality ,we get

$$P\left[\frac{\sigma_g^2}{\sigma_e^2}\in\bigcap_{h=1}^{c}\left(\frac{(n-p)Q_h}{r_h(p_h-1)U_{hd}}-\frac{1}{r_h},\ \frac{(n-p)Q_h}{r_h(p_h-1)L_{hd}}-\frac{1}{r_h}\right)\right]\geq 1-\alpha \qquad (5.17)$$

Now $\underset{1\le h\le c}{\text{Max}}\ E\left[\frac{(n-p)Q_h}{r_h(p_h-1)L_{hd}Q_0}-\frac{(n-p)Q_h}{r_h(p_h-1)U_{hd}Q_0}\right]$

$$=\left(\frac{n-p}{n-p-2}\right)\left(\frac{\sigma_g^2}{\sigma_e^2}+\frac{1}{r_{min}}\right) \qquad (5.18)$$

where $r_{min} = \underset{1\le h\le c}{\text{Minimum}}\ r_h$

Remark. The expression in equation (5.18) can be taken as the loss function for interval estimation of the ratio of variance components by virtue of the fact that the ultimate confidence interval with confidence coefficient $1-\alpha$ has been obtained based on the c number of different confidence intervals with confidence coefficient $1-\alpha^*$.

A design is said to be optimal if, among all designs in D with n observations and p lines, d minimizes $\left(\frac{\sigma_g^2}{\sigma_e^2}+\frac{1}{r_{min}}\right)$. So it is enough to maximize $r_{d,min}$.

Theorem 5.1. Given n/p is an integer, let $d_0^* \in D(p, n)$ be a one way classified design and suppose $r_{d_0^* h} = n/p$, $h = 1(1)p$, then d_0^* is optimal in D (p, n).

Proof. The proof of the theorem follows by contradiction.

Let d be a design where $r_{d,min} > r_{d_0^*, min} = n/p$. Then

$$\sum_{h=1}^{c_d} r_{dh} p_{dh} > \sum_{h=1}^{c_d} \frac{n}{p} p_{dh} = \frac{n}{p} \sum_{h=1}^{c_d} p_{dh} = n, \text{since} \sum_{h=1}^{c_d} p_{dh} = p$$

which is a contradiction.

Hence the proof follows.

REFERENCES

[1] Cheng, C.S. (1978). Optimality of certain asymmetrical experimental designs. *Ann. Statist.*, **6**, 1239-1261.

[2] Ghosh, H. and Das, A. (2001). Optimal diallel cross designs for estimation of heritability. To appear in *JSPI*.

[3] Henderson, C.R. (1953). Estimation of variance components and covariance components. *Biometrics*, **9**, 226-252.

[4] La Motte, R.L. (1976). Invariant quadratic estimators in the random, one-way ANOVA model. *Biometrics*, **32**, 793-804.

[5] Olsen, A., Seely, J. and Birkes, D. (1974). Invariant quadrative unbiased estimation for two variance components. Unpublished menuscript.

[6] Rao, C.R. (1973). *Linear Statistical Inference and its Applications.* 2nd ed., Wiley Eastern Limited.

[7] Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components.* Wiley Interscience, New York.

[8] William, H. Press, Saul, A., Teukolsky, William, T., Vetterling and Brian, P., Flannery (1993). *Numerical Recipes in C.* 1st ed., Cambridge University Press.