Jour. Ind. Soc. Ag. Statistics
Vol. XXXIX, No. 2 (1987), pp. 167-170

OPTIMAL LEVEL OF SIGNIFICANCE FOR A CONDITIONALLY SPECIFIED ESTIMATOR

S. K. SAXENA and V. P. GUPTA
University of Rajasthan, Jaipur

(Received : February, 1986)

SUMMARY

Singh and Gupta [2] have discussed a conditionally specified estimator of the error variance in a three-way layout with random effects. In this paper optimal level of significance has been derived by minimising a suitably defined risk function.

Keywords : Error variance; Risk function; Efficiency.

Introduction

For a three-way layout in the random effects model, exact F-tests for testing the main effects are not available even under the normality assumptions unless one of the first order interactions is zero. Let V_1, V_2, V_3 be the mean squares based on n_1, n_2, n_3, d.f. respectively corresponding to the interactions ABC, AC and AB and let $E(V_i) = \sigma_A^2, (i = 1, 2, 3)$. If $\sigma_{AB}^2 > 0$ and $\sigma_{AC}^2 > 0$. Scheffe [1] suggests the use of $V_A = V_3 + V_2 - V_1$ as an error variance, which, though unbiased gives an approximate F-test for testing $H_0 : \sigma_A^2 = 0$ against $H_1 : \sigma_A^2 > 0$. Singh and Gupta [2] used a preliminary test of significance (PTS) to test the hypothesis $H_p : \sigma_{AB}^2 = 0$ against $H_p : \sigma_{AB}^2 > 0$ and suggest the following conditionally specified procedure to estimate the error variance to be used in testing of H_0.

$$V = \begin{cases} \text{If } V_3/V_1 < \beta, & \text{use } V = V_2, \\ \text{If } V_3/V_1 \geqslant \beta, & \text{use } V = V_4, \end{cases}$$

where $\beta = F(n_3, n_1; \alpha_1)$ is the upper 100 $\alpha\%$ point of the F-distribution with n_3, n_1 d.f. Singh *et al.* also derived expression for bias and mean square error of V and on the basis of an empirical study recommended the use of $\beta = 1$.

In an effort to choose optimal significance level of PTS, Toyoda and Wallace [3] discussed the optimal significance points in PTS for the estimator in the linear normal regression model and pooling of two variances respectively. The object of the present paper is to investigate the optimal value of β by minimising a suitably defined risk function for the above conditionally specified procedure.

From Singh *et al.* the mse of the estimator V as a fraction of σ_1^4 is given by

$$\frac{\text{MSE}(V)}{\sigma_1^4} = \frac{2}{n_1} + \frac{2}{n_2\theta_{12}^2} + \frac{2}{n_3\theta_{13}^2} - \frac{1}{\theta_{13}^2}(1 + 2/n_3)\, I_p(a_3 + 2, a_1)$$
$$+ \frac{2}{\theta_{13}} I_p(a_3 + 1, a_1 + 1) - (1 + 2/n_1)\, I_p(a_3, a_1 + 2)$$
$$- 2(\theta_{13}^{-1} - 1)(I_p(a_3, a_1 + 1) - \theta_{13}^{-1} I_p(a_3 + 1, a_1))$$

where

$$a_1 = \frac{1}{2} n_1, \;\; a_2 = \frac{1}{2} n_2, \;\; a_3 = \frac{1}{2} n_3, \;\; p = X_c$$

when $\beta = 0$, i.e. we never pool the mean squares, then we have

$$\frac{\text{M.S.E.}(V_A)}{\sigma_1^4} = \frac{2}{n_1} + \frac{2\theta_{21}^2}{n_2} + \frac{2\theta_{31}^2}{n_3}$$

and when $\beta \to \infty$, i.e. we always pool the estimators, we have

$$\frac{\text{M.S.E.}(V_2)}{\sigma_1^4} \to \frac{2\theta_{21}^2}{n_2} + (\theta_{31}^2 - 1)^2$$

2. Optimality Criterion

Assuming $\theta_{21} = \theta_{31} - \theta$, MSE $(V_A)/\sigma_1^4$ and MSE $(V_2)/\sigma_1^4$ always have two intersections with respect to θ for any values of β and degrees of freedom, n_1 and n_2, provided that $n_3 - 2 \neq 0$. The roots are

$$\theta_1 = [n_3 - n_3\sqrt{(2n_3 + 2n_1 - 4)/n_1n_3}]/(n_3 - 2)$$

and

$$\theta_2 = [n_3 + n_3\sqrt{(2n_3 + 2n_1 - 4)/n_1n_3}]/(n_3 - 2)$$

It is easily shown that there are two cases : (i) $0 < \theta_1 < 1, 1 < \theta_2$, and (ii) $0 < \theta_1 < 1, \theta_2 < 0$. In any case there exists one interaction whose root is $\theta_1 \varepsilon (0, 1)$ MSE $(V)/\sigma_1^4$ depend on two parameters, namely β and θ, for a given set of degrees of freedom.

$$\text{Min}\,[\text{MSE}(V_A)/\sigma_1^4,\ \text{MSE}(V_3/\sigma_1^4)] = \begin{cases} \text{MSE}(V_A)/\sigma_1^4, & \text{if } 0 \leq \theta \leq \theta_1 \\ \text{MSE}(V_2)/\sigma_1^4, & \text{if } \theta_1 \leq \theta \leq 1 \end{cases}$$

First we define the efficiency of the estimator with respect to the always-pool and never pool estimators expressed as a fraction of σ_1^4 as

$$\text{Min}\,[\text{MSE}(V_A),\ \text{MSE}(V_2)] - \text{MSE}(V)$$

The decision criterion for choosing β is to maximise the efficiency over the whole range of θ; i.e.

$$\underset{\beta}{\text{Max}}\ G(\beta) = \underset{\beta}{\text{Max}} \int_0^1 \{\text{Min}\,[\text{MSE}(V_A),\ \text{MSE}(V_2)] - \text{MSE}(V)\}\, d_\theta$$

Note that this is equivalent to maximizing the average efficiency provided that the prior distribution of θ is diffused. $G(\beta)$ can be written as

$$G(\beta) = \int_0^{\theta_1} [\text{MSE}(V_A) - \text{MSE}(V)]\, d_\theta + \int_{\theta_1}^1 [\text{MSE}(V_2) - \text{MSE}(V)] d_\theta$$

$$G(\beta) = \int_{\theta_1}^1 c_1\theta^2 d_\theta - \int_{\theta_1}^1 2\theta d_\theta + \int_{\theta_1}^1 c_2 d_\theta + \int_0^1 c_3\theta^2 I_p\,(a_3, a_1)\, d_\theta$$
$$- \int_0^1 [2I_p\,(a_3 + 1, a_1 + 1) - 2I_p\,(a_3, a_1 + 1)$$
$$- I_p\,(a_3 + 1, a_1)]\, \theta d_\theta$$

where

$$p = (n_3\beta)/(n_1\theta + n_3\beta),$$
$$c_1 = 1 - 2/n_3,\ c_2 = 1 - 2/n_1 \text{ and } c_3 = 1 + 2/n_3.$$

Evaluating the integrals and simplyfying we get

$$\begin{aligned} G(\beta) = {} & \tfrac{1}{2}\, c_1(1 - \theta_1^3) - (1 - \theta_1^2) + c_2(1 - \theta_1) + \tfrac{1}{3}\, c_1\, I_p\,(a_3, a_1) \\ & + c_4 I_p\,(a_3, a_1 + 2) - I_p\,(a_3 + 1, a_1 + 1) - I_p\,(a_3, a_1 + 1) \\ & + \tfrac{1}{2}\, I_p\,(a_3 + 1, a_1) - \tfrac{1}{3}\, k^3 c_3\, [B_p(a_3 - 3, a_1 + 3) - B^{-1}\,(a_3, a_1)] \\ & + k^2\, B^{-1}\,(a_3 + 1, a_1 + 1)\, [B_p(a_3 - 1, a_1 + 3) - 1] \\ & - k\, B^{-1}\,(a_3, a_1 + 1)\, [k - 1 - k\, B_p(a_3 - 2, a_1 + 3) \\ & \qquad + B_p(a_3 - 1, a_1 + 2)] \\ & + \tfrac{1}{2}\, k^2\, B^{-1}\,(a_3 + 1, a_1)\, [1 - B_p\,(a_3 - 1, a_1 + 2)] \\ & + c_4 k\, B^{-1}\,(a_3, a_1 + 2)\, [1 - B_p\,(a_3 - 1, a_1 + 2)] \end{aligned}$$

where

$$k = \frac{n_3\beta}{n_1} \quad \text{and} \quad c_4 = 1 + \frac{2}{n_1}$$

Differentiating and simplyfying the equation

$$\frac{\partial G(\beta)}{\partial \beta} = 0$$

we find that the value of $\partial^2 G(\beta)/\partial\beta^2$ at $\beta = 1$ is positive.

This shows that $\beta = 1$ is the optimal significance level.

REFERENCES

[1] Scheffe, H. (1959) : *Analysis of Variance*, John Wiley and Sons, New York.

[2] Singh, Umesh and Gupta, V. P. (1976) : Bias and mean square error of an estimation procedure. After a preliminary test of significance for three way layout in ANOVA Model II, *Sankhya, A, Series B*, **38** (II).

[3] Toyoda, T. and Wallace, T. D. (1975) : Estimation of variance after preliminary Test of Homoginity and optimal levels of significance for the pre test, *Journal of Econometrics*, 395-404.