Jour. Ind. Soc. Ag. Statistics
Vol. XLI, No. 3 (1989), pp. 279-282

ANALYSIS OF LSD WITH A MISSING VALUE

SHEELA S. DEO
University of Poona, Pune-7
(Received : October, 1987)

SUMMARY

In a linear model with a missing observation, one can substitute an algebraic quantity and then minimize the error sum of squares for the augmented model. This gives the correct error sum of squares, but does not produce the correct hypothesis sum of squares for testing a linear hypothesis about the parameters. The sum of squares obtained is biased but practitioners still use it. The distribution of this biased sum of squares is derived in this paper for a Latin Square Design and the consequences of using this biased sum of squares on the level of significance of the test and critical points are examined.

Keywords : Missing value, Bias, Linear hypothesis.

Introduction

In a Latin square design (LSD) if one observation is missing, an estimate is used in place of it for analysis. This gives error s.s., SSE, correctly but the hypothesis sum of squares is biased. The bias is well known (Kshirsagar [3]). The distribution of biased hypothesis s.s. for a LSD is worked out here, following the lines of Deo and Kharshikar [1].

2. Analysis of LSD

Without loss of generality, let us assume that the observation pertaining to treatment 1 in the first column and first row of a Latin Square design of order ν is missing.

An estimate of the missing value, α is substituted in place of it, where α is given by

$$\alpha = [\nu (R_1 + C_1 + T_1) - 2G]/(\nu-1)(\nu-2) \quad (1)$$

Here R_1, C_1 and T_1 denote the sum of available observations on first row, first column and first treatment and G is the grand total of v^2-1 observations.

If the hypothesis is of testing equality of row effects, *i.e.* $H_{01} : \alpha_1 = \alpha_2 = \ldots = \alpha_v$ then the substitute for a missing observation when hypothesis is true is given by

$$\beta = \{ v (C_1 + T_1) - G\} / (v-1)^2 \tag{2}$$

We note here that we can divide the row s.s. into $v-1$ orthogonal row contrasts one of which can be considered as

$$\alpha_v + \alpha_{v-1} + \ldots + \alpha_2 - (v-1) \alpha_1 \tag{3}$$

The best linear unbiased estimator (BLUE) of it is

$$[R_v + R_{v-1} + \ldots + R_2 - (v-1) (R_1 + \alpha)]/v \tag{4}$$

where R_i denotes the sum of observations in ith row, $i = 1, \ldots, v$.

This is also equal to $(\beta - \alpha) (v-1)^2 /v$. (5)

When the hypothesis H_{01} is true, we get

$$E (\beta - \alpha) = 0 \tag{6}$$

$$\text{var} (\beta - \alpha) = v^2\sigma^2/\{(v - 1)^2 (v - 2)\} \tag{7}$$

Hence for testing

$$H_{v-1} : \alpha_v + \alpha_{v-1} + \ldots + \alpha_2 - (v-1) \alpha_1 = 0$$

the statistics $(\beta - \alpha)^2/\text{var} (\beta - \alpha)$ has a chi-square distribution with one d.f.

The s.s. for other $(v-2)$ orthogonal contrasts have $\sigma^2\chi^2_1$ distribution independent of each other and they are not affected by a missing observation (cf. Deo and Kharshikar [1]). (8)

The row s.s. is positively biased and the bias is given by

$$(v-1)^2 (\beta - \alpha)^2/v^2 \tag{9}$$

Thus, the exact s.s. for H_{v-1} and bias together gives

$$(\beta-\alpha)^2 \frac{(v-1)^2 (v-2)}{v^2} \left[1 + \frac{1}{v-2} \right] \tag{10}$$

which is distributed as $\frac{\nu-1}{\nu-2}\ \sigma^2\chi_1^2$ (11)

Hence biased row s.s. SSR ($\propto$) has the distribution,

$$\sigma^2\chi^2_{\nu-2} + \frac{\nu-1}{\nu-2}\ \sigma^2\chi_1^2 \qquad (12)$$

We note that the distribution of SSR ($\propto$) is same whatever may be a missing observation.

Since there is a symmetry in LSD about row, column and treatment, the distribution of SST ($\propto$) and SSB ($\propto$) are also same as above when there is a missing observation. The hypothesis for testing equality of column effects $H_{02} : \beta_1 = \beta_2 \ldots = \beta_\nu$ and that for testing equality of treatment effects is $H_{03} : \tau_1 = \tau_2 = \ldots = \tau_\nu$.

The substitute for missing observation when the above hypotheses are true is given as

$$\{\nu\,(R_1 + T_1) - G\} / (\nu-1)^2 \qquad (13)$$

and $\{\nu\,(R_1 + C_1) - G\} / (\nu-1)^2$ respectively.

Johnson and Kotz [2] proved that if Z is a mixture of chi-squares with $E\,(z) = a_1$ and Var $(z) = 2a_2$ then za_1/a_2 is a better χ^2 with d.f. a_1^2/a_2. (14)

When H_{01} is true

$$E\left[\frac{SSR\,(\propto)}{\sigma^2}\right] = \nu-2 + (\nu-1)/(\nu-2) \qquad (15)$$

and

$$\tfrac{1}{2}\,\text{Var}\left[\frac{SSR\,(\propto)}{\sigma^2}\right] = \nu-2 + (\nu-1)^2/(\nu-2)^2 \qquad (16)$$

then

$$N^* = (\text{fe}\ SSR\,(\propto))\ /\ (a_1\ SSE\,(\propto))$$

has 'F' distribution with a_1^2/a_2 and fe d.f. Here

$$\frac{a_1^2}{a_2}\ \{ (\nu-2)^2 + (\nu-1) \}^2 / \{ (\nu-2)^3 + (\nu-1)^2\}$$

and fe $= \nu^3 - 3\nu + 1$. (17)

Practioners use SSR ($\propto$) for testing equality of row effects. They generally use the statistic

$$N = \text{fe}\,(SSR\,(\propto))/\,[(\nu-1)\,SSE\,(\propto)] \tag{18}$$

Then the test usually used is

reject H_{01}, if $N \geqslant F_0$

where F_0 is the $(1 - \alpha)$ cumulative point for F distribution with $((\nu-1), \text{fe})$ d.f.

For such a test the actual size is equal to

$$P\,[N > F_0 \mid H_o\,] = P\,[N^* > F_0\,(\nu-1)/a_1 \mid H_0\,]$$

which is larger than the intended size $\propto$. The actual sizes for these tests when intended size is 0.05 and 0.01 is given in Table 1 of Deo and Kharshikar [1].

To get a test of level $\propto$ one should use N^* statistics and $F^* = F_o\,(\nu-1)/a_1$ as a critical point of F distribution with $(a_1^2/a_2$, fe) d.f.

REFERENCES

[1] Deo, S. S. and Kharshikar, A. V. (1985) : On consequences of one missing observation in a balanced design. *Commun. Statist.*, **A14** : 3091-3100.

[2] Johnson, N. L. and Kotz, S. (1970) : *Conutinous Univariate distributions* —2, Houghton Miffiin Boston.

[3] Kshirsagar, A. M. (1971). Bias due to missing plots. *Amer. Statist.*, **25** : 47-50.