ISSN 0258-7122
Bangladesh J. Agril. Res. 34(3) : 395-401, September 2009

PROBLEMS OF USUAL WEIGHTED ANALYSIS OF VARIANCE (WANOVA) IN RANDOMIZED BLOCK DESIGN (RBD) WITH MORE THAN ONE OBSERVATIONS PER CELL WHEN ERROR VARIANCES VARY FROM CELL TO CELL

MD. LUTFOR RAHMAN[1] AND KALIPADA SEN[2]

Abstract

It is well known that classical analysis of variance (ANOVA) is not suitable for heteroscedastic layouts. Weighted analysis of variance (WANOVA) is the only way to deal with such situations. Problems of usual WANOVA in Randomized. Block Design (RBD) with more than one observations per cell with interaction when error variances vary from cell to cell are discussed in this paper.

Key Words : Weighted analysis of variance, error variances, cell to cell.

Introduction

Simple heteroscedasticity is discussed in detail by Sen (1984), Sen & Ponnuswamy (1991) and Rahaman and Sen (1995). In this paper, heteroscedasticity is considered in a RBD assuming error variances vary from cell to cell. There are more than one observations per cell in this layout. The two-way general heteroscedastic model is considered here with interaction. On the assumption that the unequal error variances are known, WANOVA are derived using usual weighted least square methd. This method has some arbitrariness and problems such as:

- solution of dependent equations,
- imposition of arbitrary constraints both on parameters and on estimators of parameters,
- the problem of non-testabitity, etc.

All such problems are analytically discussed in this paper.

General heteroscedastic model for RBD

Suppose there are p treatments and q blocks in RBD with multiple observations per cell which are included in this experiment. The observations in such an experiment may be arranged in a two-way table. Let there are r observations in each of the pq cells of the table.

[1] Associate Professor, Department of Statistics, Biostatistics & Informatics, University of Dhaka, [2]Professor, Department of Statistics, Biostatistics & Informatics, University of Dhaka, Dhaka, Bangladesh.

The fixed effect additive model with interaction for the above situation may be taken as follows :

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

for i = 1, 2 ...., p, j = 1, 2, ...q and k= 1, 2, ..., r where

y_{ijk}= the k-th observation of the (i,j) the cell

μ = the general mean effect

α_i = i-th treatment effect

β_j = j-th block effect

$(\alpha\beta)_{ij}$ = the interaction effect of the simultaneous occurrence of the i-th treatment and j-th block

e_{ijk}'s are normally and independently distributed with mean zero and variance α_{ij}^2 which varies from cell to cell.

Our main objective of this analysis is to test the following gypotheses:

H_{AB}: $(\alpha\beta)_{11} = (\alpha\beta)_{12} = ...= (\alpha\beta)_{pq} = 0$ against

H: all $(\alpha\beta)_{ij}$'s are not equal to zero (2)

HA: $\alpha_1 = \alpha_2 = ...= \alpha_p$ against

H: all α_i's are not equal (3)

HB: $\beta_1 = \beta_2 =...=\beta_q$ against

H: all β_i's are not equal (4)

Usual WANOVA and analytical problems

Assuming that the error variances (α_{ij}^2) are known with weight, $W_{ij}=\frac{1}{\alpha_{ij}^2}$ for i= 1, 2, ..., p; j= 1, 2, ..., q, the Weighted Least Square (WLS) estimators of the parameters are obtained by minimizing

$$L = \sum_i\sum_j\sum_k W_{ij}(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2$$

with respect to μ, α_i, β_j and $(\alpha\beta)_{ij}$, respectively, as follows;

$$\frac{\partial L}{\partial \mu} = 0 \Rightarrow -2\sum_i\sum_j\sum_k W_{ij}(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}) = 0$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow -2\sum_j \sum_k W_{ij}(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}) = 0$$

$$\frac{\partial L}{\partial \beta_j} = 0 \Rightarrow -2\sum_i \sum_k W_{ij}(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}) = 0$$

$$\frac{\partial L}{\partial (\alpha\beta)_{ij}} = 0 \Rightarrow -2\sum_k W_{ij}(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}) = 0$$

Thus the normal equations are as follows:

$$\mu : r\hat{\mu}W.. + r\sum_i W_i.\hat{\alpha}_i + r\sum_j W._j\hat{\beta}_j + r\sum_i\sum_j W_{ij}\left(\widehat{\alpha\beta}\right)_{ij} = \sum_i\sum_j\sum_k W_{ij}y_{ijk}$$

$$\alpha_i : r\hat{\mu}W_i + rW_{i.}\hat{\alpha}_i + r\sum_j W_{.j}\hat{\beta}_j + r\sum_j W_{ij}\left(\widehat{\alpha\beta}\right)_{ij} = \sum_j\sum_k W_{ij}y_{ijk}$$

$$\beta_j : r\hat{\mu}W_{.j} + r\sum_j W_{ij}\hat{\alpha}_i + rW_{.j}\beta_j + r\sum_i W_{ij}\left(\widehat{\alpha\beta}\right)_{ij} = \sum_i\sum_k W_{ij}y_{ijk}$$

$$(\alpha\beta) : r\hat{\mu} + r\hat{\alpha}_i + r\hat{\beta}_j + r\left(\widehat{\alpha\beta}\right)_{ij} = \sum_k y_{ijk}$$

$$\text{where}\, w.. = \sum_i\sum_j W_{ij},\ W_i. = \sum_j^k W_{ij} \text{ and } W_{.j} = \sum_i W_{ij}$$

Since the equations are dependent there are infinite se of solutions of the above normal equations. Under the following constraints:

$$\sum_i W_{i.}\hat{\alpha}_i = \sum_j W_{.j}\hat{\beta}_j = \sum_i W_{ij}\left(\widehat{\alpha\beta}\right)_{ij} = \sum_j Wij\left\langle\widehat{\alpha\beta}\right\rangle_{ij} = 0 \qquad (5)$$

for all i= 1, 2, 3, ...p and j= 1, 2, 3, ..., q, we have the following standard

$$\hat{\mu} = \frac{\sum_i\sum_j\sum_k W_{ij}y_{ijk}}{rW..} = \bar{y}\cdots$$

$$\hat{\alpha}_i = \frac{\sum_j\sum_k W_{ij}y_{ijk}}{rW_{.j}} - \hat{\mu} = \bar{y}_{i..} - \bar{y}...$$

$$\hat{\beta}_j = \frac{\sum_j \sum_k W_{ij} y_{ijk}}{rW_{.j}} - \hat{\mu} = \overline{y}_{.j.} - \overline{y}...$$

$$\left(\widehat{\alpha\beta}\right)_{ij} = \frac{\sum_k y_{ijk}}{r} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j = \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}...$$

Here the WLS estimators of α_i's and β_j's are not unbiased. In addition to their linear constants are not often unbiased and hence hypotheses are not always testable. Under the linear model (1), the sample weighted residual sum of squares (SS) is

$$\text{SSE(W)} = \min \sum_i \sum_j \sum_k W_{ij}\left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}\right)^2$$

$$= \sum_i \sum_j \sum_k W_{ij}\left(y_{ijk} - \overline{y}_{ij.}\right)^2 \text{ with } pq(r-1)\text{d.f.} \tag{7}$$

The weighted residual SS under H_{AB} is

$$S_1^2(W) = \min \sum_i \sum_j \sum_k W_{ij}\left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}\right)^2 \text{ subject to } H_{AB}$$

$$= \sum_i \sum_j \sum_k W_{ij}\left(y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}...\right)^2 \text{ with } (pqr - p - q + 1) \text{ d.f.} \tag{8}$$

Thus SS due to H_{AB} is

$$\text{SS AB (W)} = S_2^2(W)\text{-SSE (W)}$$

$$= r\sum_i \sum_j W_{ij}\left(\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}._{j.} + \overline{y}_{...}\right)^2 \text{ with } (p-1)(q-1) \text{ d.f.} \tag{9}$$

The weighted residual SS under H_A is

$$S_2^2 = \min \sum_i \sum_j \sum_k W_{ij}\left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}\right)^2 \text{ subject to } H_A$$

$$= \sum_i \sum_j \sum_k W_{ij}\left(y_{ijk} - \overline{y}_{ij} + \overline{y}_{i..} - \overline{y}_{...}\right)^2 \text{ with } (pqr - pq + p - 1) \text{ d.f.} \tag{10}$$

Thus SS due to H_A is

$$\text{SSA (W)} = S_2^2\text{-SSE (W)}$$

$$= r\sum_{i} W_{i.}\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2 \text{ with (p - 1) d.f.} \tag{11}$$

Similarly, it can be shown that SS due to H_B is

$$\text{SSB (W)} = r\sum_{j} W_{.j}\left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2 \text{ with (q-1) d.f.} \tag{12}$$

It is interesting to note that desirable expected values of different SS and their desirable sampling distributions are not feasible until one impose further restrictions on the linear parameters of the model and this is shown in the next section.

Calculation of expectation of Different SS

In order to find the expected values of different SS, parametric constraints are imposed similar to (5) as follows:

$$\sum_{i} W_{i.}\alpha_i = \sum_{j} W_{.j}\beta_j = \sum_{i} W_{ij}(\alpha\beta)_{ij} = \sum_{j} W_{ij}(\alpha\beta)_{ij} = 0$$

Under above constraints,

$$\text{SSA (W)} \quad = r\sum_{i} W_{i.}\left(\bar{y}_{i...} - \bar{y}_{...}\right)^2$$

$$\text{Where, } \bar{y}_{...} = \frac{\sum_{i}\sum_{j}\sum_{k} W_{ij} y_{ijk}}{rW_{..}}$$

$$= \frac{\sum_{i}\sum_{j}\sum_{k} W_{ij}\left(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}\right)}{rW_{..}}$$

$$= \mu + \bar{e}_{...}$$

Here, $\bar{e}_{...}$ = the weighted average of error values e_{ijk}

$$= \frac{\sum_{i}\sum_{j}\sum_{k} W_{ij} e_{ijk}}{rW_{...}}$$

$$\bar{y}_i = \frac{\sum_{j}\sum_{k} W_{ij} y_{ijk}}{rW_{i.}}$$

$$= \frac{\sum_j \sum_k W_{ij}\left(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}\right)}{rW_{i.}}$$

$$= \mu + \alpha_i + \bar{e}_{i..}$$

so that SSA (W)] $= r\sum_i W_{i.}\left(\alpha_i + \bar{e}_{i..} - \bar{e}_{...}\right)^2$

Thus E[SSA (W) $= r\sum_i W_{i.}\alpha_i^2 + rE[\sum_i W_{i.}\left(\bar{e}_{i..} - \bar{e}_{...}\right)^2]$

$$= r\sum_i W_{i.}\alpha_i^2 + r(p-1)$$

Similarly it can be shown that,

E[SSB (W)] $= r\sum_j W_{.j}\beta_j^2 + r(q-1)$

E[SSAB (W)] $= r\sum_i \sum_j W_{ij}(\alpha\beta)_{ij}^2 + (p-1)(q-1)$

E[SSE (W)] = pq (r-1)

For testing hypothesis of the above sections, it is essential to impose the above parametric restrictions. In section 3, it has been shown that to estimate the parameters it needs to impose restriction on the estimators.

Table 1. WANAVA for RBD

Sources of Variations	d.f.	SS	E (SS)
A	(p-1)	SSA (W)	$(p-1) + r\sum_i W_{i.}\alpha_i^2$
B	(q-1)	SSB (W)	$(q-1) + r\sum_j W_{.j}\beta_{j.}^2$
AB	(p-1) (q-1)	SSAB (W)	$(p-1)(q-1) + r\sum_i \sum_j W_{ij}(\alpha\beta)_{ij}^2$
Error	(pq (r-1)	SSE (W)	pq (r-1)
Total	(pqr-1)	SST (W)	

A means ....

B means ....

A×B means ....

The corresponding WANOVA table is shown as Table 1. By the general theorem of Sen (1984). it can be stated that SSE (W) and SSA (W) under H_A are independently distributed as central χ^2 with pq (r-1) and (p-1) d.f. respectively. Also SSE (W) and SSB (W) under H_B are independently distributed as central χ^2's with pq (r-1) and (q-1) d.f. respectively. Hence, WANOVA χ^2 can be performed for H_A and H_B, especially when error variations are known.

Conclusion

Usual weighted least squares method provides a number of arbitrariness and analytical problems in designing models. These are discussed analytically in this paper with general heteroscedastic model in RBD. New method can be developed in order to avoid all such analytical problems.

References

Rahaman, L. and Sen, K. 1995. On a new technique of WANAVA with heteroscedastic 3-way model having interactions. *Dhaka University Journal of Science* **43**(1): 133-140.

Sen, K. 1984. Some contributions to hetroscedastic analysis of variance (HANOVA), *PhD Thesis. Department of Statistics, University of Madras, India.*

Sen, K. 1991 On a general theorem of WANOVA with the general heteroscedastic model not of full rank. *Journal of Bangladesh Academy of Sciences* **15**(2): 149-151.

Sen K. and K.N. Ponnuswamy. 1991. Comparitive studies on the adequacy of some new test procedures for testing equality of means when population variances are unequal and unknown. *Journal of Statistical Research* **25**(1&2): 59-69.