Jour. Ind. Soc. Ag. Statistics
Vol. XL, No. 1 (1988), pp. 9-14

ON THE ANALYSIS OF DATA FROM TWO-WAY CLASSIFICATION WITH UNEQUAL SUB-CLASS FREQUENCIES

B. SINGH and C. B. TIWARI
Indian Veterinary Research Institute, Izatnagar-243122
(Received : November, 1985)

SUMMARY

Expression for improved upper bound for interaction sum of squares in two-way unbalanced classification is derived and compared numerically with the upper bound given by Federer and Zelen [1] and the exact value by the least squares procedure. An indirect approach for the analysis of variance is also suggested without any matrix inversion.

Keywords : Unbalanced classification; Upper bound; Method of least squares.

Introduction

In case of non-orthogonality caused by unequal cell frequencies the addition theorem of sums of squares does not hold except where sub-class frequencies are proportional. In such situations, when all the sub-classes are filled, a method of analysis using weighted squares of means was first suggested by Yates [7]. He, however, did not derive the expressions for the interaction sum of squares except in the cases of two factors at least one of which was only at level two. The general procedure of analysis by the weighted squares of means was given by Federer and Zelen [1] who also gave the exact expression for the interaction sum of squares. Their formula for computing the interaction sum of squares involved two terms, an upper bound for the sum of squares and a correction term. Their upper bound is simple to calculate but the correction term involves the solution of least square equations or matrix inversion which may be involved in many situations.

In this paper, an improved upper bound has been obtained for the interaction sum of squares in the case of two way unbalanced classification which has been numerically compared with the upper bound of Federer and Zelen and with the exact value obtained by the method of least squares. An indirect approach for analysis of two-way unbalanced classification is also suggested by using the upper bound or the improved upper bound without going for the solution of least squares equations or matrix inversion.

2. Expressions for Upper Bound and Improved Upper Bound for Interaction Sum of Squares

Let A and B be the two factors with levels a and b, respectively. Suppose Y_{ijk}, the kth observation of jth level of B and ith level of A, is represented by

$$Y_{ijk} = m + a_i + b_j + (ab)_{ij} + e_{ijk}, \quad (i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b;\ k = 1, 2, \ldots, n_{ij}) \tag{2.1}$$

where, m is the general mean, a_i the effect of the ith level of the factor A, b_j the effect of the jth level of B, $(ab)_{ij}$ the effect of (ij)th level of the interaction AB, e_{ijk} the error variable and n_{ij} (> 1) the (ij)th cell frequency.

The upper bound for interaction, derived by Federer and Zelen, is:

$$SS_{ABU} = \sum_{i=1}^{a} \left[\sum_{j=1}^{b} n_{ij} D_{ij}^2 - \frac{\left(\sum\limits_{j=1}^{b} n_{ij} D_{ij} \right)^2}{\sum\limits_{j=1}^{b} n_{ij}} \right], \tag{2.2}$$

where

$$D_{ij} = \bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

$$\bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}, \quad \bar{y}_{i.} = \frac{1}{b} \sum_{j=1}^{b} \bar{y}_{ij}$$

$$\bar{y}_{.j} = \frac{1}{a} \sum_{i=1}^{a} \bar{y}_{ij} \quad \text{and} \quad \bar{y}_{..} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \bar{y}_{ij}.$$

The expression (2.2) can be written as :

$$SS_{ABU} = \sum_{i=1}^{a} \left[\sum_{j=1}^{b} n_{ij} D_{ij}^{*2} - \frac{\left(\sum_{j=1}^{b} n_{ij} D_{ij}^{*} \right)^2}{\sum_{j=1}^{b} n_{ij}} \right], \tag{2.3}$$

where,

$$D_{ij}^{*} = \bar{y}_{ij} - \bar{y}_{.j} \tag{2.4}$$

For improved upper bound, we replace unweighted mean $\bar{y}_{.j}$ by a weighted mean $\bar{y}_{w.j}$ and D_{ij}^{*} by

$$d_{ij} = \bar{y}_{ij} - \bar{y}_{w.j}, \tag{2.5}$$

where,

$$\bar{y}_{w.j} = \frac{\sum_{i=1}^{a} w_i \bar{y}_{ij}}{\sum_{i=1}^{a} w_i} \quad \text{and} \quad w_i^{-1} = \sum_{j=1}^{b} \frac{1}{n_{ij}}$$

Thus, the improved upper bound is given by

$$SS_{ABIU} = \sum_{i=1}^{a} \left[\sum_{j=1}^{b} n_{ij} \; d_{ij}^{2'} - \frac{\left(\sum_{j=1}^{b} n_{ij} d_{ij} \right)^2}{\sum_{j=1}^{b} n_{ij}} \right]. \tag{2.6}$$

Following Singh [5], the expression (2.6) would give the value of interaction SS more close to the exact value than that of (2.2) if,

$$\sum_{j=1}^{b} \left[\sum_{i=1}^{a} n_{ij} \left(1 - \frac{n_{ij}}{\sum_{j=1}^{b} n_{ij}} \right) \right] \left[\sum_{i=1}^{a} \frac{1}{n_{ij}} \left(\frac{1}{a^2} - \frac{w_i^2}{\left(\sum_{i=1}^{a} w_i \right)^2} \right) \right] > 0 \tag{2.7}$$

It is interesting to note that the expression (2.6) becomes, as shown below, equal to the expression for interaction, as given by Yates [7], when the factor B has only two levels :

From (2.5) we write

$$d_{ij} = \bar{y}_{ij} - \frac{\sum_{i=1}^{a} w_i \bar{y}_{ij}}{\sum_{i=1}^{a} w_i}, \qquad j = 1, 2$$

From (2.6), SS_{ABIU} is written as

$$SS_{ABIU} = \sum_{i=1}^{a} w_i \, (d_{i1} - d_{i2})^2, \tag{2.8}$$

where

$$d_{i1} - d_{i2} = \bar{y}_{i1} - \bar{y}_{i2} - \frac{\sum_{i=1}^{a} w_i \, (\bar{y}_{i1} - \bar{y}_{i2})}{\sum_{i=1}^{a} w_i} \tag{2.9}$$

If we write, $d_i = \bar{y}_{i_1} - \bar{y}_{i2}$, then (2.8) can be simplified to

$$SS_{ABIU} = \sum_{i=1}^{a} w_i \, d_i^2 - \frac{\left(\sum_{i=1}^{a} w_i \, d_i \right)^2}{\sum_{i=1}^{a} w_i}, \tag{2.10}$$

which is the expression obtained by Yates [7] as may be seen from Snedecor and Cochran [4] and Steel and Torri [6], but the same is not true for the Federer and Zelen's upper bound except for their exact expression for interaction SS.

The computation of exact interaction SS from either of the methods, the Federer and Zelen method of weighted squares of means or the method of fitting constants, needs the solution of least square equations or matrix inversion. The improved upper bound (2.6) or the upper bound (2.2), however, do not need any matrix inversion. As revealed from the numerical comparison (see Table below in Sec. 4) the improved upper bound (2.6) is a better approximation to the exact value of the interaction SS and hence may be used for routine analysis.

3. Indirect Approach of Analysis for Two Way Unbalanced Classification

When interaction effect is significant, the adjusted sums of squares due to main effects may be computed by the weighted squares of means procedure (Yates, [7] and Federer and Zelen, [1]). In case of non-significant interaction, these formulae are no longer applicable (Kramer, [3]). The general procedure of getting the exact values of main effect sums of squares, when interaction effect is non-significant, is the least squares procedure of fitting constants that requires the solution of least squares equations and hence matrix inversion. An indirect procedure of analysis under such situation is given below which does not need any matrix inversion.

By least squares procedures the sum of squares due to interaction is :

$$SS_{AB} = R\{m, a_i, b_j, (ab)_{ij}\} - R(m, a_i, b_j), \tag{3.1}$$

where $R\{m, a_i, b_j, (ab)_{ij}\}$ is the total reduction in sum of squares due to the constants fitted in the interaction model and is equal to $\sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \bar{y}_{ij}^2$, the between cell sum of squares, since all the variability among AB sub-class is accounted for by the main effects and interaction and $R(m, a_i, b_j)$ is the total reduction in the sum of squares due to the constants fitted in the model without interaction and may be written as follows :

$$R(m, a_i, b_j) = \sum_{i=1}^{a} n_{i\cdot} \bar{y}_{0i\cdot}^2 + S_{B/A}^2 \tag{3.2}$$

$$= \sum_{j=1}^{b} n_{\cdot j} \bar{y}_{0\cdot j}^2 + S_{A/B}^2 , \tag{3.3}$$

where

$$n_{i\cdot} = \sum_{j=1}^{b} n_{ij}, \qquad n_{\cdot j} = \sum_{i=1}^{a} n_{ij} \quad \text{and}$$

$$\bar{y}_{0i\cdot} = \frac{1}{n_{i\cdot}} \sum_{j=1}^{b} n_{ij} \bar{y}_{ij}, \quad \bar{y}_{0\cdot j} = \frac{1}{n_{ij}} \sum_{i=1}^{a} n_{ij} \bar{y}_{ij},$$

and $S_{A/B}^2$ and $S_{B/A}^2$ are the adjusted sums of squares due to A adjusted for B and B adjusted for A, respectively.

Our problem is to obtain $S_{B/A}^2$ and $S_{A/B}^2$. If we know $R(m, a_i, b_j)$ the estimates for $S_{B/A}^2$ and $S_{A/B}^2$ can be obtained from (3.2) and (3.3), respectively, after substituting the value of $R(m, a_i, b_j)$. The value of $R(m, a_i, b_j)$ may be obtained from (3.1) after substituting the value of $R\{m, a_i, b_j,$

$(ab)_{ij}\}$, as between cell sum of squares, and the interactions sum of squares. The value of interaction sum of squares may be obtained as an approximation from upper bound or the improved upper bound as mentioned above. The improved upper bound is the better approximation than the upper bound, so it may be used in place of interaction sum of squares in (3.1) to obtain $R(m, a_i, b_j)$ and then $S^2_{B/A}$ and $S^2_{A/B}$. In this way the matrix inversion or the solution of least squares equations in analysis of two way unbalanced classification is avoided.

4. Numerical Examples

The following Table shows the comparative picture of the values of interaction sum of squares obtained by least squares procedure, upper bound (Federer and Zelen) and improved upper bound from four sets of data. All the computations were rounded up to 4 decimal places.

Method	*Sets of data*			
	1	*2*	*3*	*4*
Upper bound	1.5132	7.9719	125.5709	114.4755
Improved upper bound	1.5097	6.9835	123.2708	116.6128
Exact (least squares)	1.4896	6.4048	117.4965	113.3571

ACKNOWLEDGEMENT

The authors are grateful to Dr. D. D. Joshi, Professor of Statistics and Director, Institute of Social Sciences, Agra for his valuable comments and improvements in presentation.

REFERENCES

[1] Federer, W. T. and Zelen, M. (1966): Analysis of multifactor classification with unequal number of observations, *Biometrics*, **22** : 525-552.

[2] Harvey, W. R. (1966) : Least squares analysis of data with unequal sub-class numbers. ARS-USDA, Maryland-20705.

[3] Kramer, C. Y. (1955) : On the analysis of variance of a two-way classification with unequal sub-class number, *Biometrics*, **11** : 441-452.

[4] Snedecor, G. W. and Cochran, W. G. (1967) : *Statistical Methods*, 6th Ed., Oxford and IBH Publishing Co.

[5] Singh, B. (1986) : On the analysis of 3-way classification with unequal cell frequencies, *J. Indian Soc. Agril. Statist.*, **38** : 133-140.

[6] Steel, G. B. and Torrie, J. H. (1960) : *Principles of Statistics*, McGraw Hill.

[7] Yates, F. (1934) : The analysis of multiple classification with unequal numbers in the different classes, *J. Amer. Statist. Assoc.*, **29** : 51-66.