Jour. Ind. Soc. Ag. Statistics
Vol. XXXVIII, No. 2 (1986), pp. 133-140

ON THE ANALYSIS OF 3-WAY CLASSIFICATION WITH UNEQUAL CELL FREQUENCIES

B. SINGH*
Central Sheep & Wool Research Institute, Avikanagar via Jaipur-304501
(Received : March, 1980)

SUMMARY

The expressions for 2 factor interactions sums of squares and 3 factor interaction (when one factor is at two levels) sum of squares using weighted squares of means procedure in 3-way unbalanced classification are presented. These formulae may be used on desk calculators without requiring any matrix inversion. The conditions, when these formulae give the values of the interaction sum of squares closer to the exact values than the upper bound, suggested by Federer & Zelel [1], are obtained. A numerical example of 3 × 3 × 2 experiment is used to compare the suggested procedure with the upper bound and the exact method of weighted squares of means.

Keywords : Interactions, upper bound, weighted mean, method of least squares.

Introduction

Unequal cell frequencies always cause non-orthogonality among various effects under experiment and hence the addition theorem of sum of squares does not hold unless sub class frequencies are proportional. A general method of analysis for such data is provided by fitting constants, the method of least squares. However, when constants are fitted for all main effects and interactions, the computing formulae become unmanageable when the number of factors and/or levels become large. When all the subclasses are filled and constants have been fitted for all main effects and interactions, the weighted squares of means procedure of analysis is equivalent to the complete least squares analysis (Harvey, [2]). The general

*Present address : Scientist S-2 (Ag. Stat.), L.E.S. Division, I.V.R.I., Izatnagar 243122 (U. P.).

procedure of analysis, by weighted squares of means, has been derived by Zelen and Federer [8] and illustrated by Federer & Zelen [1]. The method involves lengthy calculations mainly the matrix inversion in case of interactions sums of squares.

In this paper, a procedure of analysis, particularly for interactions sums of squares, is suggested without any matrix inversion. A numerical example of 3 × 3 × 2 factorial experiment is used to compare the suggested procedure with exact and upper bounds given by Federer & Zelen [1].

Expressions for Sums of Squares

Let A_1, A_2 and A_3 be 3 factors with levels m_1, m_2 and m_3 ($m_1 \leqslant m_2 \leqslant m_3$) respectively. Suppose $\bar{y}_{ijk}$ is the mean of the ith level of A_1, the jth level of A_2 and the kth level of A_3, and n_{ijk} is the corresponding cell frequency. Here i varies from 1 to m_1, j varies from 1 to m_2 and k varies from 1 to m_3.

2.1 *Main Effects Sums of Squares*

The formulae for main effects sums of squares are known (see Yates [7], Harvey [2], etc.) but included here for completion. The sum of squares due to the main effect A_1 can be writ ten as

$$SS(A_1) = \sum_{i=1}^{m_1} w_{i}\ldots Y_{i}^2\ldots - \frac{\left(\sum\limits_{i=1}^{m_1} w_i\ldots Y_i\ldots\right)^2}{\sum\limits_{i=1}^{m_1} w_i\ldots}, \tag{2.1.1}$$

where

$$Y_i\ldots = \sum_{j=1}^{m_2}\sum_{k=1}^{m_3} \bar{Y}_{ijk} \text{ and } w_i\ldots = \left(\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} \frac{1}{n_{ijk}}\right)^{-1}.$$

The expressions for $SS(A_2)$ and $SS(A_3)$ can be defined similarly.

2.2 *Two Factor Interactions Sums of Squares*

Zelen and Federer [8] derived the general expressions for sums of squares due to two factor interactions. The computation procedure for these sums of squares has been given by Federer & Zelen [1]. Their computing formula for sum of squares due to a two factor interaction consists of two terms, the upper bound and the correction term. The upper bound is simple to compute but the correction term involves the matrix inversion. Here we modify this upper bound for two factor interaction sum of squares to a form that estimates the sum of squares almost same as the exact value and avoids completely the matrix inversion.

The upper bound for the interaction sums of squares $(A_1 \times A_2)$ given by Federer & Zelen [1] can be written as

$$SS(A_1 \times A_2) = \sum_{i=1}^{m_1} \left[\sum_{j=1}^{m_2} w_{ij\cdot} D_{ij}'^2 - \frac{\left(\sum_{j=1}^{m_2} w_{ij} \cdot D'_{ij\cdot} \right)^2}{\sum_{j=1}^{m_2} w_{ij} \cdot} \right], \quad (2.2.1)$$

where

$$D'_{ij\cdot} = y_{ij} - \frac{1}{m_1} \sum_{i=1}^{m_1} y_{ij},$$

$$y_{ij} = \sum_{k=1}^{m_3} Y_{ijk} \text{ and } w_{ij} = \left(\sum_{k=1}^{m_3} \frac{1}{n_{ijk}} \right)^{-1}.$$

Define

$$D_{ij} = y_{ij} - \frac{\sum_{i=1}^{m_1} w_{i\cdot\cdot} Y_{ij}}{\sum_{i=1}^{m_1} w_{i\cdot\cdot}}.$$

Then, the suggested value of the sum of squares due to interaction $(A_1 \times A_2)$ is given by

$$SS(A_1 \times A_2) = \sum_{i=1}^{m_1} \left[\sum_{i=1}^{m_2} w_{ij} \cdot D_{ij\cdot}^2 - \frac{\left(\sum_{j=1}^{m_2} w_{ij\cdot} D_{ij\cdot} \right)^2}{\sum_{j=1}^{m_2} w_{ij} \cdot} \right], \quad (2.2.2)$$

which is obtained from (2.2.1) after replacing D'_{ij}. by D_{ij}.

In $D'_{ij\cdot}$ the deviations are taken from the simple mean while in $D_{ij\cdot\cdot}$, the deviations are taken from the weighted mean with weights $w_{i\cdot\cdot}$ to give more precise estimator of the interaction parameter under unbalanced situations.

To ensure that the expression (2.2.2) would give the value for $SS(A_1 \times A_2)$ more close to the exact value than the upper bound (2.2.1), the following inequality should hold.

$$\sum_{j=1}^{m_2} \left[\sum_{i=1}^{m_1} w_{ij\cdot} \left(1 - \frac{w_{ij\cdot}}{\sum_{j=1}^{m_2} w_{ij\cdot}} \right) \right] \left[\sum_{i=1}^{m_1} \frac{1}{w_{i\cdot\cdot}} \left(\frac{1}{m_1^2} - \frac{w_{i\cdot\cdot}^2}{\left(\sum_{i=1}^{m_1} w_{i\cdot\cdot} \right)^2} \right) \right] \geqslant 0. \quad (2.2.3)$$

The inequality (2.2.3) can be obtained as follows. Under the null hypothesis of no interaction ($A_1 \times A_2$) effect, it can be shown that

$$E\,(\text{Upper bound} \mid H_0) = (m_1 - 1)\,(m_2 - 1)\,\sigma^2$$

$$+ \left[\frac{1}{m_1^2}\sum_{j=1}^{m_2}\left(\sum_{i=1}^{m_1} w_{ij.}\left(1 - \frac{w_{ij.}}{\sum_{j=1}^{m_2} w_{ij.}}\right)\right)\sum_{i=1}^{m_1}\frac{1}{w_{ij.}} - (m_2 - 1)\right]\sigma^2$$

and

$$E[SS(A_1 + A_2) \mid H_0] = (m_1 - 1)\,(m_2 - 1)\,\sigma^2$$

$$+ \left[\left(\frac{1}{\sum_{i=1}^{m_1} w_{i..}}\right)^2 \sum_{j=1}^{m_2}\left(\sum_{i=1}^{m_1} w_{ij.}\left(1 - \frac{w_{ij.}}{\sum_{j=1}^{m_2} w_{ij.}}\right)\right)\right.$$

$$\left.\sum_{i=1}^{m_1}\frac{w_{i..}^2}{w_{ij.}} - (m_2 - 1)\right]\sigma^2, \qquad (2.2.4)$$

where σ^2 is the variance among observations,

The value of $SS(A_1 \times A_2)$, (2.2.2) would be less or equal to the upper bound (2.2.1), if

$$E\,[\text{upper bound} \mid H_0] - E[SS(A_1 \times A_2) \mid H_0] \geqslant 0. \qquad (2.2.5)$$

By using results (2.2.4) in (2.2.5) we can easily deduce the inequality (2.2.3). Similar results, (2.2.3), for sums of squares $SS(A_2 \times A_3)$ and $SS(A_1 \times A_3)$ can also be obtained easily.

One interesting property of (2.2.2) is that it reduces to the exact expression given in many statistical books (Snedecor & Cochran [4], Steel & Torrie [5], etc.) when one factor has two levels, which is not true for the upper bound expression (2.2.1). For example, let $m_2 = 2$, then (2.2.2) can be shown to be

$$SS(A_1 \times A_2) = \sum_{i=1}^{m_1} w_{i..}\, D_{i..}^2 - \frac{\left(\sum_{i=1}^{m_1} w_{i..}\, D_{i..}\right)^2}{\sum_{i=1}^{m_1} w_{i..}}, \qquad (2.2.6)$$

where $D_{i..} = Y_{i1.} - Y_{i2.}$.

In case the inequality (2.2.3) is not satisfied, the expression (2.2.1) may be used as an approximation for $SS(A_1 \times A_2)$ except the case when one factor has two levels, in that case formula (2.2.6) may be used.

2.3 *Three Factor Interaction Sum of Squares*

When one of the factors has 2 levels (say $m_3 = 2$), the upper bound (Federer & Zelen, [1]) for 3-factor interaction sum of squares becomes

$$S_0(A_1 \times A_2 \times A_3) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} w_{ij\cdot} \left(D^*_{ij\cdot} - \frac{1}{m_2} \sum_{j=1}^{m_2} D^*_{ij\cdot} \right)^2, \tag{2.3.1}$$

where

$$D^*_{ij\cdot} = \overline{Y}_{ij1} - \overline{Y}_{ij2} - \frac{1}{m_1} \sum_{i=1}^{m_1} (\overline{Y}_{ij1} - \overline{Y}_{ij2}).$$

By the theory of least squares it can be shown that

$$\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} w_{ij\cdot} \left(D^*_{ij\cdot} - \frac{\sum_{j=1}^{m_2} w_{ij\cdot} \, D^*_{ij\cdot}}{\sum_{j=1}^{m_2} w_{ij\cdot}} \right)^2 \leqslant S_0(A_1 \times A_2 \times A_3). \tag{2.3.2}$$

Hence, on the basis of (2.3.2), the sum of squares for 3-factor interaction when one factor has two levels ($m_3 = 2$) is proposed as

$$S(A_1 \times A_2 \times A_3) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} w_{ij\cdot} \left(D^{**}_{ij\cdot} - \frac{\sum_{j=1}^{m_2} w_{ij\cdot} \, D^{**}_{ij\cdot}}{\sum_{j=1}^{m_2} w_{ij\cdot}} \right)^2$$

$$= \sum_{i=1}^{m_1} \left[\sum_{j=1}^{m_2} w_{ij\cdot} \, D^{**2}_{ij\cdot} - \frac{\left(\sum_{j=1}^{m_2} w_{ij\cdot} \, D^{**}_{ij\cdot} \right)^2}{\sum_{j=1}^{m_2} w_{ij\cdot}} \right], \tag{2.3.3}$$

where

$$D^{**}_{ij\cdot} = \overline{Y}_{ij1} - \overline{Y}_{ij2} - \sum_{i=1}^{m_1} w_{i\cdot\cdot} \, (\overline{Y}_{ij1} - \overline{Y}_{ij2}) \Big/ \sum_{i=1}^{m_1} w_{i\cdot\cdot},$$

It is noticed here that in $D^{**}_{ij\cdot}$ the deviations are taken from weighted mean with weights $w_{i\cdot\cdot}$ to have the more precise estimator for the interaction parameter as in two-factor interaction case.

The condition, when (2.3.3) given the value more close to the exact value than the L.H.S. of (2.3.2) and hence the upper bound, comes out to be the same as for 2-factor interaction sum of squares, i.e. the inequality (2.3.3).

An approximate value for 3-factor interaction in general 3-way classification may be obtained from the following expression. This expression is an approximation similar to the upper bound (2.3.3) when one factor has two levels.

$$S(A_1 \times A_2 \times A_3) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \left[\sum_{k=1}^{m_3} n_{ijk} \; D^2_{ijk} - \frac{\left(\sum\limits_{k=1}^{m_3} n_{ijk} D_{ijk} \right)^2}{\sum\limits_{k=1}^{m_3} n_{ijk}} \right], \tag{2.3.4}$$

where

$$D_{ijk} = \overline{Y}_{ijk} - \overline{Y}_{i.k} - \frac{\sum\limits_{i=1}^{m_1} w_{i..} \; (\overline{Y}_{ijk} - \overline{Y}_{i.k})}{\sum\limits_{i=1}^{m_1} w_{i..}}$$

with

$$\overline{Y}_{i.k} = \frac{1}{m_2} \sum_{j=1}^{m_2} \overline{Y}_{ijk}.$$

In general n-way classification, the n_{ijk} are replaced by w_{ijk} in (2.3.4) for computing 3-factor interaction sum of squares. The weights w_{ijk} can be defined similarly as $w_{ij.}$ or $w_{i..}$.

3. An Example

Data used for the comparison pertain to the averages for body weights at weaning age of 3 breeds of sheep born during years 1975, 1976 and 1977 (Table 1). One factor is year (A_1, $m_1 = 3$), 2nd factor is breed (A_2, $m_2 = 3$) and the 3rd factor is sex (A_3, $m_3 = 2$). The values of sums of squares for interactions obtained by 3 procedures, the exact, the upper bound and the suggested one, are given in Table 2. The values of interactions sums of squares when one factor has 2 levels ($A_1 \times A_3$, $A_2 \times A_3$) are the same for exact and proposed procedure and hence the upper bounds for these sums of squares have not been calculated. It is clear from Table 2 that the values computed by the suggested procedure for 2 factor interaction ($A_1 \times A_2$) and 3-factor interaction ($A_1 \times A_2 \times A_3$) are closer to the exact value than the corresponding upper bounds. These results were also supported numerically by the inequality (2.2.3).

TABLE 1

Breed A_2	$A_2(1)$		$A_2(2)$		$A_2(3)$	
Year A_1	$A_3(1)$	$A_3(2)$	$A_3(1)$	$A_3(2)$	$A_3(1)$	$A_3(2)$
Mean $A_1(1)$	7.1	6.62	7.45	7.2327	7.6883	6.8692
No. of observation	6	5	49	55	30	13
Mean $A_1(2)$	8.0625	7.8619	8.4303	9.2176	9.2006	8.9208
No. of observation	16	21	33	34	39	24
Mean $A_1(3)$	5.7333	6.7833	6.92	5.9444	7.83	6.9666
No. of observation.	18	12	15	9	10	9

$A_3(1)$—Male, $A_3(2)$—Female.

TABLE 2—*SS* FOR INTERACTIONS COMPUTED BY DIFFERENT PROCEDURES

Source	*Exact value*	*Suggested value*	*Upper bound*
$A_1 \times A_3$	4.5328	4.5328	—
$A_2 \times A_3$	6.30205	6.30205	—
$A_1 \times A_2$	10.38638	11.23538	12.4002
$A_1 \times A_2 \times A_3$	19.317618	19.688586	21.901616

Discussion

Kramer [3] claimed that weighted squares of means procedure seems to be optimum when interactions come out to be significant, and proposed a new method for computing the main effects sums of squares when interaction effects are non-significant. In the situation when interactions seem to be significant the method presented in this paper could be recommended for routine use, as it estimates the interactions sums of squares (eliminating all other effects) very close to the exact value.

The procedure of analysis discussed in this paper is most practicable for animal breeding experiments as sex is always at two levels and season too is generally at two levels, as the condition for 3-factor interaction sum of squares is that one factor should have two levels.

ACKNOWLEDGEMENT

Author is grateful to referees for their critical comments and valuable suggestions for improving the manuscript.

REFERENCES

[1] Federer, W. T. and Zelen, M. (1966) : Analysis of multifactor classification with un-equal 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 numbers, *Biometrics*, 11 : 441-452.

[4] Snedecor, G. W. and Cochran, W. G. (1967) : *Statistical Methods*, 6th Ed., Oxford & IBH Publishing Co. : 484.

[5] Steel, G. D. and Torrie, J. H. (1960) : *Principles and Procedures of Statistics*, McGraw Hill : 252.

[6] Yates, F. (1933) : The principles of orthogonality and confounding in replicated experiments, *Jour. of Agric. Sciences*, 23 : 108-145.

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

[8] Zelen, M. and Federer, W. T. (1965) : III analysis of factorials with unequal number of observations, *Sankhya*, 25 : 383-400.