QÜESTIIÓ, Vol 12, n° 2, pp. 149-157, 1988

A SIMPLIFIED TREATMENT OF THE RESTRICTED ANALYSYS OF A SLIGHTLY DISPROPORTIONATE FACTORIAL EXPERIMENT

JOSÉ M. PRIETO* and JOSÉ M. CARIDAD**

University of Córdoba

This paper consider a procedure to obtain effect estimators in the least squares analysis of a slightly disproportionate factorial design when a sample survey is made of the results of an extensive experiment. Explicit formulae have been found for the restricted estimators and their variances, when the constraints normally imposed upon a proportional model are used. In addition, an approximate analysis of the original problem is used to perform that estimation, and an approximate analysis of variance table is proposed.

Keywords: Factorial designs, disproportionate subclass frequencies, constraints.

1. INTRODUCTION

In many experimental investigations, the resulting data can be displayed according to several factors in a complete "factorial" design. As a framework for presentation, let us initially considerer a two-day design. The corresponding fixed-effect linear model in usual notation is written:

$$(1)\quad y_{ijr} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijr} \quad (i = 1, ..., a; j = 1, ..., b; r = 1, ..., n_{ij})$$

where it is assumed that the errors e_{ijr} are independently distributed as $N(o, \sigma^2)$

*José M. Prieto - Universitat de Còrdova. E.T.S.I.A Dept. Matemàtica Aplicada - Aptat. 3048 - 14080 Còrdova

**José M. Caridad - Universitat de Còrdova. E.T.S.I.A Dept. d'Estadística - Aptat. 3048 - 14080 Còrdova

-Article rebut el abril de 1988

Since the above model is overparameterized, a set of identiafiability constraints of the form:

$$\sum_i v_i\alpha_i = 0 \ , \ \sum_j w_j\beta_j = 0 \ , \ \sum_j w_j\alpha\beta_{ij} = 0 \ (i = 1, \dots, a) \, ,$$
$$\sum_i v_i\alpha\beta_{ij} = 0 \ (j = 1, \dots, b)$$

must be imposed on the original factorial effects to ensure their estimability and interpretation. Considerable interest as well as a fair amount of controversy has centered on the choice of constraints to be used in unbalanced designs (/1/), the form of the constraints depending on the weights assigned to the cells or subclasses of the classification. The choice of the weights should be definable before establishing the sampling scheme, irrespective on whether the sampling is balanced or unbalanced (/2/, p. 97).

In this paper we will be concerned with investigation in which we make a sample survey of the results of an "extensive" experiment (/5/, p. 119). In this experimental context, the subclass numbers reflect the stratum sizes of the corresponding experimetal population, and the use of observed frequencies as weights is then interpretable.

It is a common practice in experimentation to call initially for a balanced design (/3/), althoug the completed experiment frequently has unequal subclass numbers, the usual thing being to analize this data with "usual" constraints ($v_i = w_i = 1$). If lack of balance is not great, this inappropriate analysis produces effect estimates which are not much different from the proper estimates (/4/, p. 131).

It is also common in experimentation to call initially for a design with "proportional" frequencies (/6/, p. 286), althoug the completed experiment has disproportionate subclass numbers. It is expected, however, that the data show slight deviations from proportionality. We assume such a slightly disproportionate classification.

For such an extensive experiment with a slightly disproportionate classification it would be meaningful to use constraints associated with the weights $v_i = n_{i.}$ and $w_j = n_{.j}$, normally imposed upon a proportional model (/7/), which we call "marginal" constraints.

As a progress report on the well-known "quasi balanced" design (previously mentioned), our interest is centered on a "quasi-proportional" design. On the other hand, for models with a large number of parameters, althoug the problem of the estimation of the parameters may be solved numerically by a computer (/8/, /9/), it is a particular advantage to find formulae for the estimators (/2/).

The purpose of this paper is then to examine for a disproportionate model with marginal constraints on the following question: Might the restricted estimators and their variances be expressed by explicit formulae? An accesory question is: is it possible to simplify the above non-orthogonal analysis and to obtain an approximate analysis of variance (anova) table?

2. EXACT RESTRICTED ESTIMATION

We now tackle the proposed estimation problem for a sufficiently generic case, namely, the $a \times b \times c$ factorial model with the equation:

$$\begin{aligned} Y_{ijkr} &= \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + e_{ijkr} \\ &(i = 1, ..., a\,;\; j = 1, ..., b\,;\; k = 1, ..., c\,;\; r = 1, ..., n_{ijk}) \end{aligned} \tag{2}$$

with the same assumption as in (1). For this three-factor case, the marginal constraints we impose on the parameters are written:

$$\begin{array}{ll} \sum_i n_{i..}\, \alpha_i = 0 & \sum_i n_{.j.}\, \beta_j = 0 \\ \sum_k n_{..k}\, \gamma_k = 0 & \sum_j n_{.j.}\, \alpha\beta_{ij} = 0 \quad (i = 1, ..., a) \\ \sum_i n_{i..}\, \alpha\beta_{ij} = 0 \quad (j = 1, ..., b) & \sum_k n_{..k}\, \alpha\gamma_{ik} = 0 \quad (i = 1, ..., a) \\ \sum_i n_{i..}\, \alpha\gamma_{ik} = 0 \quad (k = 1, ..., c) & \sum_k n_{..k}\, \beta\gamma_{jk} = 0 \quad (j = 1, ..., b) \\ \sum_j n_{.j.}\, \beta\gamma_{jk} = 0 \quad (k = 1, ..., c) & \sum_k n_{..k}\, \alpha\beta\gamma_{ijk} = 0 \begin{pmatrix} i = 1, ..., a \\ j = 1, ..., b \end{pmatrix} \\ \sum_j n_{.j.}\, \alpha\beta\gamma_{ijk} = 0 \begin{pmatrix} i = 1, .., a \\ k = 1, ..., c \end{pmatrix} & \sum_i n_{i..}\, \alpha\beta\gamma_{ijk} = 0 \begin{pmatrix} j = 1, ..., b \\ k = 1, ..., c \end{pmatrix} \end{array} \tag{3}$$

The system of normal equations relative to a model with all the interactions, such as (2), may be written:

$$\begin{aligned} \bar{y}_{ijk.} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha\beta}_{ij} + \hat{\alpha\gamma}_{ik} + \hat{\beta\gamma}_{jk} + \hat{\alpha\beta\gamma}_{ijk} \\ &(i = 1, ..., a\,;\; j = 1, ..., b\,;\; k = 1, ..., c) \end{aligned} \tag{4}$$

where $\bar{y}_{ijk.}$ is the mean of a generic subclass. The solution of this system augmented with the above one (3) (with regard to the effect estimators) is unique. To express that solution concisely, we introduce the following notation:

$$n_{ijk}^{\simeq} = 1/n^2(n_{i..}\, n_{.j.}\, n_{..k}) \tag{5}$$

which implies that

$$n_{ij.}^{\simeq} = 1/n(n_{i..}\, n_{.j.}) \ , \ n_{i..}^{\simeq} = n_{i..} \ , \ n^{\simeq} = n \tag{6}$$

It is interesting to note that, for the disproportionate classification we are considering $n_{ijk}^{\simeq} \neq n_{ijk}$ and $n_{ij.}^{\simeq} \neq n_{ij.}$. Also, in terms of (5) and (6), we state the following additional notation:

$$\begin{aligned} y_{...}^{\simeq} &= 1/n \sum_i \sum_j \sum_k n_{ijk}^{\simeq}\, \bar{y}_{ijk.} \ , \ \bar{y}_{i...}^{\simeq} = 1/n_{i..} \sum_j \sum_k n_{ijk}^{\simeq}\, \bar{y}_{ijk.} \\ \bar{y}_{ijk.}^{\simeq} &= 1/n \sum_k n_{ijk}^{\simeq}\, \bar{y}_{ijk.} \ , \ \bar{y}_{ijk.}^{\simeq} = \bar{y}_{ijk.} \end{aligned} \tag{7}$$

where the means and marginal means of the original classification appear.

We are now in a position to state more formally the solution of above mentioned augmented normal equations.

Theorem 1. The restricted estimators for a disproportionate $a \times b \times c$ factorial model with marginal constraints are expressed by the following explicit formulae:

$$\begin{aligned} \hat{\mu} &= \bar{y}_{....}^{\simeq} \\ \hat{\alpha}_i &= \bar{y}_{i...}^{\simeq} - \bar{y}_{....}^{\simeq} \\ \hat{\alpha\beta}_{ij} &= \bar{y}_{ij..}^{\simeq} - \bar{y}_{i..}^{\simeq} - \bar{y}_{.j..}^{\simeq} - \bar{y}_{....}^{\simeq} \\ \hat{\alpha\beta\gamma}_{ijk} &= \bar{y}_{ijk.} - \bar{y}_{ij..}^{\simeq} - \bar{y}_{i.k.}^{\simeq} - \bar{y}_{.jk.}^{\simeq} + \bar{y}_{i...}^{\simeq} + \bar{y}_{.j..}^{\simeq} + \bar{y}_{..k.}^{\simeq} - \bar{y}_{....}^{\simeq} \end{aligned} \tag{8}$$

The variances of these estimators are:

(9)

$$\mathrm{Var}\ (\hat{\mu}) = \sigma^2/n^2 \sum_i \sum_j \sum_k \left(n_{ijk}^{\simeq}\right)^2 / n_{ijk}$$

$$\mathrm{Var}\ (\hat{\alpha}_i) = \sigma^2((1/n_{i..} - 1/n)^2 \sum_j \sum_k \left(n_{ijk}^{\simeq}\right)^2 / n_{ijk} + 1/n^2 \sum_{i'} \sum_j \sum_k \left(n_{i'jk}^{\simeq}\right)^2 / n_{i'jk}\Big)$$

$$\mathrm{Var}\ (\hat{\alpha\beta}_{ij}) = \sigma^2 \Big((1/n_{ij.}^{\simeq} - 1/n_{i..} - 1/n_{.j.} + 1/n\Big)^2 \sum_k \left(n_{ijk}^{\simeq}\right)^2 / n_{ijk} + (1/n_{i..} - 1/n)^2 \sum_{j'} \sum_k \left(n_{ij'k}^{\simeq}\right)^2 / n_{ij'k} + (1/n_{.j.} - 1/n)^2 \sum_{i'} \sum_k \left(n_{i'jk}^{\simeq}\right)^2 / n_{i'jk} + 1/n^2 \sum_{i'} \sum_{j'} \sum_k \left(n_{i'j'k}^{\simeq}\right)^2 / n_{i'j'k}\Big)$$

$$\mathrm{Var}\ (\alpha\hat{\beta}\gamma_{ijk}) = \sigma^2 \Big((1/n_{ijk}^{\simeq} - 1/n_{ij.}^{\simeq} - 1/n_{i.k.}^{\simeq} - 1/n_{.jk}^{\simeq} + 1/n_{i..} + 1/n_{.j.} + 1/n_{..k} - 1/n)^2 \times \left(n_{ijk}^{\simeq}\right)^2 / n_{ijk} + \Big(1/n_{ij.}^{\simeq} - 1/n_{i..} - 1/n_{.j.} + 1/n)^2 \sum_{k'} \left(n_{ijk'}^{\simeq}\right)^2 / n_{ijk'}$$

$$+ \Big(1/n_{i.k}^{\simeq} - 1/n_{i..} - 1/n_{.k.} + 1/n\Big)^2 \sum_{j'} \left(n_{ij'k}^{\simeq}\right)^2 / n_{ij'k}$$

$$+ \Big(1/n_{.jk}^{\simeq} - 1/n_{.j.} - 1/n_{..k} + 1/n\Big)^2 \sum_{i'} \left(n_{i'jk}^{\simeq}\right)^2 / n_{i'jk}$$

$$+ (1/n_{i..} - 1/n)^2 \sum_{j'} \sum_{k'} \left(n_{ij'k'}^{\simeq}\right)^2 / n_{ij'k'} + (1/n_{.j.} - 1/n)^2 \sum_{i'} \sum_{k'} \left(n_{i'jk'}^{\simeq}\right)^2 / n_{i'j'k'}$$

$$+ (1/n_{..k} - 1/n)^2 \sum_{i'} \sum_{j'} \left(n_{i'j'k}^{\simeq}\right)^2 / n_{i'j'k} + 1/n^2 \sum_{i'} \sum_{j'} \sum_{k'} \left(n_{i'j'k'}^{\simeq}\right)^2 / n_{i'j'k'}$$

$(i' = 1, \ldots, i-1, i+1, \ldots, a;\ j' = 1, \ldots, j-1, j+1, \ldots, b;\ k' = 1, \ldots, k-1, k+1, \ldots, c)$

where the error variance is estimated by:

$$(10) \qquad \hat{\sigma}^2 = \sum_i \sum_j \sum_k \sum_r \left(y_{ijkr} - \bar{y}_{ijk.}\right)^2 / (n - a \times b \times c)$$

Proof. If we introduce (5) to (7) in (8), the resulting expressions for the estimators satisfy (3) and (4) (as may be checked). On the other hand, formulae (9) may be verified by including (7) in (8) and finding the variances of the resulting expressions, taking into account the model assumptionm mentioned below expression (1).

The estimation carried out in this Section is valid in the experimental context described at the end os Section 1.

Remark. Because of the non-orthogonality of the original designal data, it is not possible to find formulae for the statistics of the F-tests.

3. AN APPROXIMATE ANALYSIS

3.1. APPROXIMATE RESTRICTED ESTIMATION

Next, we propose an approximate analysis in which the cell means $\bar{y}_{ijk.}$ are treated as thoug they were averages of $n_{ijk}^{\simeq}$ observations, where $n_{ijk}^{\simeq}$ is given in (5) in terms of the original data. We can imagine a hypothetical classification associated to the original one that possesses (5) and (6) as its corresponding frequencies and (7) as its means. Such a classification turns out to be "proportional" (as may be verified).

The following is an analogous result to that stated in Section 2.

Theorem 2. Approximate estimators for a model with a slightly disproportionate classification are given by the proper formulae (8). The variances of these estimators in this case are:

(11)

$$
\begin{aligned}
\mathrm{Var}(\hat{\mu}) &= \sigma^2/n\\
\mathrm{Var}(\hat{\alpha}_i) &= \sigma^2/n^3(n(n-n_{i..})^2/n_{i..}+n^2(n-n_{i..})/n)=\sigma^2(n-n_{i..})/nn_{i..}\\
\mathrm{Var}(\hat{\alpha\beta}_{ij}) &= \sigma^2/n^3\left((n(n-(n_{i..}+n_{.j.})+n^{\simeq}_{ij.})^2/n^{\simeq}_{ij.}\right)+(n-n_{.j.})(n-n_{i..})^2/n_{i..}\\
&\quad +(n-n_{i..})(n-n_{ij.})^2/n_{.j.}+(n-(n_{i..}+n_{ij.})+n^{\simeq}_{ij.})/n)\\
\mathrm{Var}(\hat{\alpha\beta\gamma}_{ijk}) &= \sigma^2/n^3\left(n(n-(n_{i..}+n_{.j.}+n_{..k})+(n^{\simeq}_{ij.}+n^{\simeq}_{i.k.}+n^{\simeq}_{.jk})-n^{\simeq}_{ijk}\right)^2/n^{\simeq}_{ijk}\\
&\quad +(n-n_{..k})\left(n-(n_{i..}+n_{.j.})+n^{\simeq}_{ij.}\right)^2/n^{\simeq}_{ij.}\\
&\quad +(n-n_{.j.})\left(n-(n_{i..}+n_{..k})+n^{\simeq}_{i.k}\right)^2/n^{\simeq}_{i.k}\\
&\quad +(n-n_{i..})\left(n-(n_{.j.}+n_{..k})+n^{\simeq}_{.jk}\right)^2/n^{\simeq}_{.jk}\\
&\quad +(n-(n_{.j.}+n_{..k})+n^{\simeq}_{.jk}(n-n_{i..})^2/n_{i..}\\
&\quad +(n-(n_{i..}+n_{..k})+n^{\simeq}_{i.k}(n-n_{.j.})^2/n_{.j.}\\
&\quad +(n-(n_{i..}+n_{.j.})+n^{\simeq}_{ij.}(n-n_{..k})^2/n_{..k}\\
&\quad \left.+(n-(n_{i..}+n_{.j.}+n_{..k})+(n^{\simeq}_{ij.}+n^{\simeq}_{i.k}+n^{\simeq}_{.jk})-n^{\simeq}_{ijk})/n\right)
\end{aligned}
$$

where the error variance is estimated as in the exact analysis (10).

Proof. In fact, equations (8) are reminiscent of the weel-known formulae that perform the estimation with marginal constraints in a model with a proportional classification, such as the hipothetical one mentioned earlier. On the other hand, if we replace in (9) the original frequencies n_{ijk}, $n_{ij.}$,..., bye their associated ones $n^{\simeq}_{ijk.}$, $n^{\simeq}_{ij.}$, ..., we obtain (11).

3.2. Aproximate analysis of variance table

For a classification with slight deviations from balance, the usual constrained analysis ($v_i = w_j = 1$) leads to an aproximate anova table, that dos not give misleading results, provided the cell sample sizes are not too unequal (/10/, p. 124).

Next, as an extension of the just mentioned result, we propose another table for a disproportionate classification in terms of (5) to (8).

Table 1. Approximate anova table for a three-way disporportionate factorial classifcation.

Source of Variation	Degrees of Freedom	Sum of Squares
A main effects	$a - 1$	$\sum_i n_{i..} (\hat{\alpha}_i)^2$
B main effects	$b - 1$	$\sum_j n_{.j.} (\hat{\beta}_j))^2$
C main effects	$c - 1$	$\sum_k n_{..k} (\hat{\gamma}_k)^2$
AB interactions	$(a - 1)(b - 1)$	$\sum_i \sum_j n^{\simeq}_{ij.} (\hat{\alpha\beta}_{ij})^2$
AC interactions	$(a - 1)(c - 1)$	$\sum_i \sum_k n^{\simeq}_{i.k} (\hat{\alpha\gamma}_{ik})^2$
BC interactions	$(b - 1)(c - 1)$	$\sum_j \sum_k n^{\simeq}_{.jk} (\hat{\beta\gamma}_{jk})^2$
ABC interactions	$(a - 1)(b - 1)(c - 1)$	$\sum_i \sum_k \sum_k n^{\simeq}_{ijk} (\alpha\hat{\beta}\gamma_{ijk})^2$
errors	$n - a\ b\ c$	$\sum_i \sum_j \sum_k \sum_r (y_{ijkr} - \bar{y}_{ijk.})^2$

The sums of squares in this Table do not add up exactly to the total sum of squares, but the discrepancy is not great, provided the sample sizes are only slightly disproportional. If so, the proposed aproximate anova table yields reasonable approximations.

4. REFERENCES

[1] **Nelder, J.A.**: "A reformulation of Linear Models (with Discusions)" Journal of the Royal Statistical Society, 1977. Ser. A, 140, Part 1, pp. 48-76.

[2] **Arnold, S.F.**: "The Theory of Linear Models and Multivariate Analysis". Willey, New York, 1981.

[3] **Winer, B.J.**: "Statistical Principles in Experimentals Design", McGraw-Hill Book Co., New York, 1971.

[4] **Bargmann, R.E.**: "Computational Aids for the Analysis of Balanced an Unbalanced Multiple Factor-Factorial Desgins", Proceedings of the Statistical Computign Section, American Satatistical Association 1975, pp. 128-133.

[5] **Kendall, M.G.** and **Stuart, A.**: "The Advanced Theory os Statistics", Vol. 3: Design and Analysis, and Time Series, Griffin, London, 1976.

[6] **Bock, R.D.**: "Multivariate Statistical Methods in Behavioral Research", McGraw-Hill Book Co., New York, 1975.

[7] **Scheffe, H.**: "The Analysis of Variance", Willey, New York, 1959.

[8] **Prieto, J.M.**: "Métodos algebraicos en el Análisis de Modelos Factoriales Desequilibrados", Ph. D. Thesis, Tx.: Author, Córdoba, 1986.

[9] **Prieto, J.M.** and **Caridad, J.M.**: "Restricted Estimation in Umbalanced Factorial Models: An APL Programs Package", Qüestiió, 1988, vol. 12, n.1, pp. 77-96.

[10] **Miller, R.G.**: "Beyond Anova, Basics of Applied Statistics", Willey, New York, 1986.