Advances and Applications in Statistics
© 2017 Pushpa Publishing House, Allahabad, India
http://www.pphmj.com
http://dx.doi.org/10.17654/AS051040277
Volume 51, Number 4, 2017, Pages 277-282 **ISSN: 0972-3617**

A METHOD TO MINIMIZE THE SUM OF THE VARIANCES OF THE ESTIMATORS OF THE VARIANCE COMPONENTS IN STAIR NESTED DESIGNS

C. Fernandes and P. Ramos

Área Departamental de Matemática
Instituto Superior de Engenharia de Lisboa
Portugal and Centro de Matemática e Aplicações (CMA), FCT, UNL
Portugal

Abstract

Stair nested designs may be a good alternative to balanced nested designs since we can work with fewer observations and the amount of information for the different factors is more evenly distributed. In stair nested designs the number of treatments is the sum of the "actice" factor levels so these designs lead to a great economy. A method will be used to minimize the sum of the variances of the estimators of the variance components.

1. Introduction

There are many practical situations where it is realistic to suppose that observations $y_1, ..., y_n$ are realizations of random variables $Y_1, ..., Y_n$ which

Received: June 13, 2017; Accepted: August 24, 2017
2010 Mathematics Subject Classification: 62J10.
Keywords and phrases: stair nested designs, variance components, minimization.

have the same mean μ and whose variance-covariance matrix is defined by a hierarchy of f nested random factors. There is a variance component associated with each factor, and it is desired to estimate these variance components.

In general, suppose that the factors $F_1, ..., F_f$ have random effects with variance components $\sigma_1^2, ..., \sigma_f^2$, and that factor F_j is nested in factor F_{j-1} for $j = 2, ..., f$. Assume that $f \geq 2$. Let m_1 be the maximum number of levels of factor F_1 that it is feasible to use, and, for $j = 2, ..., f$, let m_j be the maximum number of levels of factor F_j that it is feasible to use within each level of factor F_{j-1}. In industrial settings m_1 may be large, but there are many other practical situations in which $m_1 \leq m_j$ for $2 < j \leq f$.

Traditional designs to estimate variance components used so-called balanced nesting (e.g., Khuri et al. [5]). Within each combination of levels of factors $F_1, ..., F_j$, the same number of levels of factor F_{j+1} is used. These designs are orthogonal designs, and the estimators of the variance components are independent. However, they need a large number of observations, and the number of degrees of freedom for estimating the variance components are very different. Stair nested designs have been introduced for the estimation of variance components and to overcome these disadvantages of the balanced nested designs. They are reviewed in Section 2.

A method to minimize the sum of the estimated variances of the estimators of the variance components is presented in Section 3.

2. Stair Nested Designs

Stair nested designs, which were introduced by Cox and Solomon [1] and presented by Fernandes et al. [2-4], may be a good alternative to balanced nested designs since we can work with fewer observations and the amount of information for the different factors is more evenly distributed. There is effectively one component for each factor. For $j = 1, \ldots, f$, factor F_j has a_j "active" levels in component j and only a single level in each other component. If we write "+" between components then the stair nested design is written as

$$a_1/1/\cdots/1 \quad + \quad 1/a_2/1/\cdots/1 \quad + \quad \cdots \quad + \quad 1/\cdots/1/a_f$$

and has $n = \sum_{r=1}^{f} a_r$. For $j = 1, \ldots, f$, factor F_j has v_j levels, where $v_j = \sum_{r=1}^{j} a_r + f - j$. Figure 1 shows a stair nested design with $f = 3$, $a_1 = 3, a_2 = 2$ and $a_3 = 4$. Thus there are $n = 9$ observations.

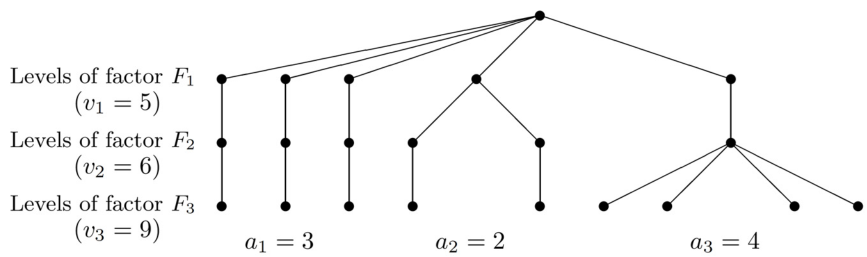

Figure 1. The stair nested design "3/1/1 + 1/2/1 + 1/1/4".

Factor F_1 has v_1 levels, where $v_1 = a_1 + f - 1$. The first a_1 levels are the "active" levels for this factor; each of them is combined with a single level of each of the remaining factors. The first observation sub-vector $\mathbf{Y}_1$ is constituted by the observations $y_1, \ldots, y_{a_1}$; it has variance-covariance matrix $\mathbf{V}_1$ given by $\mathbf{V}_1 = \gamma_1 \mathbf{I}_{a_1}$, with $\gamma_1 = \sum_{r=1}^{f} \sigma_r^2$, where $\mathbf{I}_k$ is the identity matrix of order k.

When $1 < j \leq f$, for level $a_1 + j - 1$ of the first factor we take a single level of each of the factors $F_2, \ldots, F_{j-1}$ and a_j "active" levels of factor F_j. For each of these we take only one level of the remaining factors. The sub-vector $\mathbf{Y}_j$ is constituted by the observations $y_{s_{j-1}+1}, \ldots, y_{s_j}$, where $s_j =$

$\sum_{r=1}^{j} a_r$. It has variance-covariance matrix $\mathbf{V}_j$ given by $\mathbf{V}_j = \omega_j \mathbf{J}_{a_j} + \gamma_j \mathbf{I}_{a_j}$, with $\omega_j = \sum_{r=1}^{j-1} \sigma_r^2$ and $\gamma_j = \sum_{r=j}^{f} \sigma_r^2$, where $\mathbf{J}_k$ is the $k \times k$ matrix with all entries equal to 1.

For $j = 1, ..., f$, the sub-vectors $\mathbf{Y}_j$ are mutually independent. The sums of squares S_j and mean squares M_j are given by $S_j = \mathbf{Y}_j' \mathbf{K}_{a_j} \mathbf{Y}_j$ and $M_j = S_j / d_j$, where $\mathbf{K}_k = \mathbf{I}_k - k^{-1} \mathbf{J}_k$. Thus S_j and S_k are independent when $j \neq k$ and multivariate normality implies that the distribution of S_j / γ_j is χ^2 on d_j degrees of freedom, where $d_j = a_j - 1$. Then $E(S_j) = d_j \gamma_j$ so $\hat{\gamma}_j = M_j$ is an unbiased estimator for γ_j for $j = 1, ..., f$. Thus we have the following unbiased estimators for the variance components: $\hat{\sigma}_f^2 = \hat{\gamma}_f = M_f$, and $\hat{\sigma}_j^2 = M_j - M_{j+1} = \hat{\gamma}_j - \hat{\gamma}_{j+1}$ for $j = 1, ..., f - 1$. Hence $Var(\hat{\sigma}_f^2) = 2\gamma_f^2 / d_f$ and $Var(\hat{\sigma}_j^2) = 2(\gamma_j^2 / d_j + \gamma_{j+1}^2 / d_{j+1})$ when $1 \leq j < f$.

Finally, if $1 \leq j < f$ and $\sigma_j^2 = 0$, then M_j / M_{j+1} has the F distribution on d_j and d_{j+1} degrees of freedom. Thus this ratio of mean squares may be used to test the hypothesis that $\sigma_j^2 = 0$.

3. Minimizing the Sum of the Variances of the Estimators

With $d_j = a_j - 1$ for $j = 1, ..., f$, we intend to, through a proper choice of $d_1, ..., d_f$, minimize

$$\sum_{j=1}^{f} Var(\hat{\sigma}_j^2) = 2\sum_{j=1}^{f-1} \left(\frac{\gamma_j^2}{d_j} + \frac{\gamma_{j+1}^2}{d_{j+1}} \right) = 2\frac{\gamma_1^2}{d_1} + 4\sum_{j=2}^{f-1} \frac{\gamma_j^2}{d_j} + 2\frac{\gamma_f^2}{d_f}$$

subject to the constraint $\sum_{j=1}^{f} d_j = n,$ with n known. Using the Lagrange multipliers with the auxiliary function

$$L(d_1, \ldots, d_f, \lambda) = 2\frac{\gamma_1^2}{d_1} + 4\sum_{j=2}^{f-1}\frac{\gamma_j^2}{d_j} + 2\frac{\gamma_f^2}{d_f} + \lambda\left(\sum_{j=1}^{f} d_j - n\right),$$

we have to solve equation $\nabla L(d_1, \ldots, d_f, \lambda) = 0$ and this implies solving equations

$$\begin{cases} \dfrac{\partial L}{\partial d_1} = 0 \\ \dfrac{\partial L}{\partial d_j} = 0, \quad j = 2, \ldots, f-1 \\ \dfrac{\partial L}{\partial d_f} = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases} \Leftrightarrow \begin{cases} 2\dfrac{\gamma_1^2}{d_1^2} = \lambda \\ 4\dfrac{\gamma_j^2}{d_j^2} = \lambda, \quad j = 2, \ldots, f-1 \\ 2\dfrac{\gamma_f^2}{d_f^2} = \lambda \\ \sum_{j=1}^{f} d_j - n = 0 \end{cases}.$$

From the first three conditions we obtain $b_j^2 = 2\gamma_j^2/\gamma_1^2$ for $b_j = d_j/d_1,$ with $j = 2, \ldots, f-1$ and $b_f^2 = \gamma_f^2/\gamma_1^2$ for $b_f = d_f/d_1.$

From the last condition we have

$$d_1 = \frac{n}{1 + \sum_{j=2}^{f} b_j}.$$

For $j = 2, \ldots, f,$ we know that $d_1 = d_j/b_j,$ then using the previous results we get

$$d_j = \frac{nb_j}{1 + \sum_{r=2}^{f} b_r}.$$

Now, obtaining estimators for $\gamma_j,$ with $j = 1, \ldots, f,$ through pre-sampling and choosing the value of n, we can optimize the distribution of the number of levels for each factor, in order to minimize the sum of the variances of the estimators of the variance components.

Acknowledgements

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).

References

[1] D. Cox and P. Solomon, Components of Variance, Chapman and Hall, 2003.

[2] C. Fernandes, P. Ramos and J. Mexia, Balanced and step nesting designs - application for cladophylls of asparagus, JP J. Biostat. 4(3) (2010a), 279-287.

[3] C. Fernandes, P. Ramos and J. Mexia, Algebraic structure of step nesting designs, Discuss. Math. Probab. Stat. 30(2) (2010b), 221-235.

[4] C. Fernandes, P. Ramos and J. Mexia, Crossing balanced and stair nested designs, Electron. J. Linear Algebra 25 (2012), 22-48.

[5] A. Khuri, T. Mathew and B. Sinha, Statistical Tests for Mixed Linear Models, John Wiley and Sons, 1998.