Journal of Statistics & Applied Science, Volume 1, Issue 2
ISSN 2718-0999

Inference in "One-Way" Random Designs - Discussing Sub-D and ANOVA based Estimators

Adilson Silva*[1,2]

[1] Faculty of Science and Technology, University of Cape Verde, Praia, Cape Verde

[2]CMA, Faculty of Science and Technology, New University of Lisbon, Caparica, Portugal

*Correspondence Author, e-mail: adilson.dasilva@docente.unicv.edu.cv

Abstract

Recently it was shown through simulations studies that Sub-D produces estimates with unbiased and lower variance-covariance estimates than the ANOVA-based estimator, except in case of random "one-way" balanced designs. In this designs the simulations studies suggested that they have the same variance-covariance estimates. This paper aims to compare the common ANOVA-based estimator to Sub-D in random "one-way" designs with two groups of treatment and in random "one-way" balanced designs. The comparison will be conducted through theoretical results and corroborated with simulation studies. It will be proved that the ANOVA-base estimator and Sub-D have exactly the same variance-covariance estimates in both above referred designs. The proof will be given firstly for random "one-way" designs with two groups of treatment and then for random "one-way" balanced designs.

Keywords: Sub-D, ANOVA, Variance Components, One-way Designs.

1. Introduction

Due to necessity of incorporate the amount of variations caused by certain uncontrollable sources of variations in statistical designs with fixed effects, for example the amount of variations within and/or between groups of treatments for that the experimenters are not able to control and those whose the levels must be randomly selected, in research field such as genetic, agriculture, animal breeding, and quality control and improvement, in early 1960 several designs with both fixed and random effects terms were introduced and widely investigated (see Khuri [4] and Silva [11]).

Among those designs we highlight the well known and widely discussed random "one-way" designs:

$$z_{ij} = \mu + \alpha_i + \epsilon_{ij},\ i = 1, \dots, k;\ j = 1, \dots, n_i, \tag{1.1}$$

$$\text{where} \begin{cases} k \ \text{ is the the number of groups of treatment;} \\ n_i \text{ is the number of observations within the } i\text{th group of treatment;} \\ \mu \ \text{ is the general mean (the fixed effect);} \\ \alpha_i \text{ is the random effect due to the } i\text{th group of treatment;} \\ \epsilon_{ij} \text{ is the random error due to the } j\text{th observation within the } i\text{th} \\ \quad \text{group of treatment.} \end{cases}$$

It is assumed that:

$$\begin{cases} \alpha_i \sim (0,\ \gamma_\alpha)\,, \ \text{ that is } \alpha_i\text{'s are i.i.d. with mean zero and variance } \gamma_\alpha; \\ \epsilon_{ij} \ \sim (0,\ \gamma_\epsilon)\,, \text{ that is } \epsilon_{ij}\text{'s are i.i.d. with mean zero and variance } \gamma_\epsilon; \\ cov(\alpha_i,\ \epsilon_{ij}) = 0,\ i = 1, \dots, k \text{ and } j = 1, \dots, n_i. \end{cases}$$

When all groups of treatment have the same number of observations, that is $n_i = n$, the model 1.1 is called random "one-way" balanced design. Otherwise it is called random "one-way" designs.

Random "one-way" designs are useful tools for modeling repeated measured data and, in particular, small sample and longitudinal data (see Wallace [15] and Khuri et al. [5]). For this designs several techniques and tools focussing on variance components estimation has been developed. Among than the most popular are those based on likelihood and ANOVA (see Demidenko [1] and Pinheiro and Bates [7], for instance). Recently, while doing research for his PhD Thesis, Silva (2017) developed a new estimator for variance components named Sub-D (see Silva [11], [12], [13] and Ferreira et al. [2]). On its approach Silva constructed and applied a finite sequence of orthogonal transformations (which he called sub-diagonalizations) to the covariance structure of the restricted design producing a set of sub-models which he used to create pooled estimators for the variance components.

Through simulations it was Shown that Sub-D produces very realistic estimative in random "one-way" balanced and unbalanced designs (see Silva [12]); in nested and crossed "two-way" unbalanced designs (see Silva [11]); and in nested "three-way" unbalanced designs (See Silva et al.[13]). In fact, the numerical simulation show that Sub-D produces reasonable and comparable estimates, sometimes slightly better than those obtained with REML and mostly better than tose obtained with Anova. However, due to the correlation between the sub-models on it's foundation, the variability of estimates produced with Sub-D is slightly greater then tose obtained with REML except in random "one-way" balanced designs. But, when compared with Anova, Sub-D produces estimates with unbiased and lower variance estimates than Anova-based estimator except in case of random "one-way" balanced designs. In this case, simulations studies suggested that Sub-D and Anova-based estimator has the same variance. Thus, this work aims to prove through theoretical results that for this designs Anova-based estimator and Sub-D have exactly the same variance. Moreover, this work also aims to propose a correction for a result in the deduction of one of the Sub-D's estimators for variance components estimators given in Silva [12].

First section is devoted to the introduction, and the second one to the background. Thirty section is reserved to prove that Anova-based estimator and Sub-D has exactly the same variance-covariance in random "one-way" balanced designs. Forth section is reserved to simulations studies, and the last one for the discussions.

From now on, the following *notations* will be used without any additional comments:

- $P_{R(X)}$ denotes the projection matrix onto the subspace spanned by the columns of a matrix X and $P_{R(X)^\top}$ the projection matrix onto the orthogonal complement of the subspace spanned by the columns of X;
- $\Sigma(x)$ denotes the variance-covariance matrix of a random vector x, i.e
$\Sigma(x) = Eb \quad \top$
$\mathbf{0}_{n,m}$ denotes an $n \times m$ matrix, while $\mathbf{0}_n$ denotes a null vector of dimension n; $\mathbf{1}_n$ denotes a vector of ones having both dimension n;
$\mathbf{J}_n$ denotes a $n \times n$ matrix of ones;
$z \sim (w, \Sigma)$ denotes a random vector z with mean w, and variance-covariance matrix Σ;
$z \sim N(w, \Sigma)$ denotes a random vector z with a normal distribution with mean w, and variance-covariance matrix Σ;
$r(A)$ denotes the rank of a matrix A;
$tr(A)$ denotes the trace of a matrix A
$\sum_{i \neq j}^{n}$ denotes $\sum_{i=1}^{n} \sum_{j=1}^{n}$ for $i \neq j$.

2. The Estimators: Anova and Sub-D

In this section we introduce and briefly discuss Sub-D and ANOVA-based estimators on design 1.1. Their MSE will be discussed. We will focus on case when the design has two groups of treatment, i.e $k = 2$, as well as the case when the design is balanced, that is $n_i = n, i = 1, \ldots, k$.

2.1. ANOVA-based Estimator

The analysis of variance (ANOVA) method of estimating the variance components γ_α and γ_ϵ in model 1.1 consists of equating observed values of the between group mean squares (MS_B) and within group mean square (MS_W) to their expected values, and solving the resulting equations for γ_α and γ_ϵ. This method produces unbiased estimators of γ_α and γ_ϵ. Such estimators are respectively given as

$$\begin{aligned}
\widehat{\gamma_\alpha}^A &= \frac{1}{n_o}(MS_B - MS_W) \\
&= \frac{1}{n_o}\left[\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{z_{i\bullet}} - \bar{z_{\bullet\bullet}})^2 - \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \bar{z_{i\bullet}})^2\right] \text{ and} \\
\widehat{\gamma_\epsilon}^A &= MS_W = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \bar{z_{i\bullet}})^2, \qquad (2.1)
\end{aligned}$$

where $n_o = \frac{N^2 - \sum_{i=1}^{k} n_i^2}{N(k-1)}$, $N = \sum_{i=1}^{k} n_i$, $\bar{z_{i\bullet}} = \sum_{j=1}^{n_i} \frac{z_{ij}}{n_i}$ and $\bar{z_{\bullet\bullet}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{z_{ij}}{N}$.

Following Searle [8], [9] (see Sahai and Ojeda [3]) the variance of ANOVA estimators $\widehat{\gamma_\alpha}^A$ and $\widehat{\gamma_\epsilon}^A$, are respectively given as

$$\begin{aligned}
\Sigma\left(\widehat{\gamma_\alpha}^A\right) &= \frac{2\gamma_\alpha^2}{\left(N^2 - \sum_{i=1}^{k} n_i^2\right)^2}\left[N^2\sum_{i=1}^{k} n_i^2 + \left(\sum_{i=1}^{k} n_i^2\right)^2 - 2N\sum_{i=1}^{k} n_i^3\right] \\
&\quad + \frac{4N\gamma_\alpha\gamma_\epsilon}{\left(N^2 - \sum_{i=1}^{k} n_i^2\right)} + \frac{2\gamma_\epsilon^2 N^2(N-1)(k-1)}{\left(N^2 - \sum_{i=1}^{k} n_i^2\right)^2 (N-k)} \text{ and} \\
\Sigma\left(\widehat{\gamma_\epsilon}^A\right) &= \frac{2\gamma_\epsilon^2}{N-k}. \qquad (2.2)
\end{aligned}$$

Numerical studies carried out by Singh [14] and Caro et al. [6] for different configurations of γ_α and γ_ϵ suggested that the unbalancedness of the data results in an increase of variance-covariance of $\Sigma\left(\widehat{\gamma_\alpha}^A\right)$ and $\Sigma\left(\widehat{\gamma_\epsilon}^A\right)$. Khuri et al. [5] proved that $\Sigma\left(\widehat{\gamma_\alpha}^A\right)$ attains its minimum for all γ_α and γ_ϵ when the data are balanced.

2.2. Sub-D

Lets take the matrix formulation of design (1.1):

$$z = \mu 1_N + Z\beta + \epsilon, \qquad (2.3)$$

where

$$Z = \begin{bmatrix} 1_{n_1} & 0_{n_1} & 0_{n_1} & \ldots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & 0_{n_2} & \ldots & 0_{n_2} \\ 0_{n_3} & 0_{n_2} & 1_{n_3} & \ldots & 0_{n_3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{n_k} & 0_{n_k} & 0_{n_k} & \ldots & 1_{n_k} \end{bmatrix}, \ \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} \text{ and } \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{kn_k} \end{bmatrix}, \qquad (2.4)$$

with $\beta \sim (0_k, \gamma_\alpha I_k)$, $\epsilon \sim (0_N, \gamma_\epsilon I_N)$ and β and ϵ mutually independent. Thus the model (2.3) may be rewritten as follow

$$z \sim \left(\mu 1_N,\ \gamma_\alpha Z Z^\top + \gamma_\epsilon I_N\right). \tag{2.5}$$

Let B be the $N \times (N-1)$ matrix whose columns are the $N-1$ orthonormal eigenvectors associated to the null eigenvalue of $\frac{1}{N} J_N$, where J_N denotes an $N \times N$ matrix of 1's. Using B it is possible to define (see Silva [12]) a new design (a restricted one) by projecting the design (2.5) onto the orthogonal complement of the vectorial subspace spanned by $\mu 1_N$, as follow

$$y = B^\top z \sim \left(\mu 0_{N-1},\ \gamma_\alpha M + \gamma_\epsilon I_{N-1}\right), \text{where } M = B^\top Z Z^\top B. \tag{2.6}$$

Now let A_i be the matrix whose rows are the set of $g_i = r(A_i)$ orthonormal eigenvectors associated to the eigenvalue θ_i, $i = 1, \ldots, h$, of M; Let also $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\epsilon}^S$ denote the Sub-D estimator of γ_α and γ_ϵ, respectively. Thus, following Silva[5], we that

$$\begin{aligned} \widehat{\gamma_\alpha}^S &= \frac{1}{h^*} \sum_{i=1}^{h} \theta_i \left(h y^\top P_i y - \sum_{j=1}^{h} y^\top P_j y \right) \\ &= y^\top \Lambda_\alpha y, \end{aligned} \tag{2.7}$$

where $\Lambda_\alpha = \frac{1}{h^*} \sum_{i=1}^{h} \theta_i \left(h P_i - \sum_{j=1}^{h} P_j \right)$, $h^* = h \sum_{i=1}^{h} \theta_i^2 - \left(\sum_{i=1}^{h} \theta_i \right)^2$, and $P_i = \frac{A_i^\top A_i}{g_i}$, and

$$\begin{aligned} \widehat{\gamma_\epsilon}^S &= \frac{1}{h^*} \sum_{i=1}^{h} \theta_i \left(\theta_i \sum_{j=1}^{h} y^\top P_j y - \sum_{j=1}^{h} \theta_j y^\top P_j y \right) \\ &= y^\top \Lambda_\epsilon y, \end{aligned} \tag{2.8}$$

where $\Lambda_\epsilon = \frac{1}{h^*} \sum_{i=1}^{h} \theta_i \sum_{j=1}^{h} \left(\theta_i - \theta_j \right) P_j$.

2.2.1. The Correct Version of Sub-D. Unfortunately, it seems that the algebraic manipulation at the time of Sub-D's deduction did not work as well as Silva [12] wished since we found that his deduction of $\widehat{\gamma_\epsilon}^S$ is wrong. The correct one is the one we presented here at (2.8). It worth to remark that:

(1) The above elucidated error in the deduction of $\widehat{\gamma_\epsilon}^S$ at Silva[5] (Section 3) lies on (the wrong) computation of $(\Theta^\top \Theta)^{-1}$. Indeed, with $\Theta = \begin{bmatrix} \theta_1 & 1 \\ \vdots & \vdots \\ \theta_h & 1 \end{bmatrix}$, we found that

$$\Theta^\top \Theta = \begin{bmatrix} \sum_{i=1}^{h} \theta_i^2 & \sum_{i=1}^{h} \theta_i \\ \sum_{i=1}^{h} \theta_i & h \end{bmatrix} \text{ so that } (\Theta^\top \Theta)^{-1} = \frac{1}{h^*} \begin{bmatrix} h & -\sum_{i=1}^{h} \theta_i \\ -\sum_{i=1}^{h} \theta_i & \sum_{i=1}^{h} \theta_i^2 \end{bmatrix}, \tag{2.9}$$

but unfortunately a miscalculation led Silva[5] to find $\frac{1}{h^*} \begin{bmatrix} h & -\sum_{i=1}^{h} \theta_i \\ -\sum_{i=1}^{h} \theta_i & \sum_{i=1}^{h} \theta_i \end{bmatrix}$ for $(\Theta^\top \Theta)^{-1}$ instead of the equation at right side of (2.9), which on it's turn let to a wrong deduction of $\widehat{\gamma_\epsilon}^S$.

(2) The miscalculation in the deduction of $\widehat{\gamma_\epsilon}^S$ did not reflected in the section 'Numerical Example' of Silva[5], since the computation of $(\Theta^\top \Theta)^{-1}$ was done through a software (R).

From now on we refer to the correct version of $\widehat{\gamma_\epsilon}^S$ given in (2.8).

The next Theorem proposes the variance-covariance of both $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\epsilon}^S$.

Theorem 2.1. *Let* $\lambda_s = h^2 \sum_{i=1}^{h} \frac{\theta_i^s}{g_i} - 2h \sum_{i=1}^{h} \theta_i \sum_{i=1}^{h} \frac{\theta_i^{s-1}}{g_i} + \left(\sum_{i=1}^{h} \theta_i\right)^2 \sum_{i=1}^{h} \frac{\theta_i^{s-2}}{g_i}$, $s = 2, 3, 4.$ *Then:*

(a) $\Sigma\left(\widehat{\gamma_\alpha}^S\right) = \frac{2\gamma_\alpha^2}{h^{*2}}\lambda_4 + \frac{4\gamma_\alpha\gamma_\epsilon}{h^{*2}}\lambda_3 + \frac{2\gamma_\epsilon^2}{h^{*2}}\lambda_2;$

(b) $\Sigma\left(\widehat{\gamma_\epsilon}^S\right) = \frac{2}{h^{*2}} \sum_{j=1}^{h} \left[\frac{\left(\sum_{i=1}^{h} \theta_i(\theta_i - \theta_j)\right)^2}{g_j}\right] \left(\gamma_\alpha^2\theta_j^2 + 2\gamma_\alpha\gamma_\epsilon\theta_j + \gamma_\epsilon^2\right).$

Proof. (See Shayle et al. [10] for variance-covariance of a quadratic form) Part (a):

$$\begin{aligned}
\Sigma\left(\widehat{\gamma_\alpha}^S\right) &= 2tr\left(y^\top \Lambda_\alpha y\right) = 2tr\left[(\Lambda_\alpha(\gamma_\alpha M + \gamma_\epsilon))^2\right] \\
&= 2\gamma_\alpha^2 tr\left[(\Lambda_\alpha M)^2\right] + 4\gamma_\alpha\gamma_\epsilon tr\left[\Lambda_\alpha M \Lambda_\alpha\right] + 2\gamma_\epsilon^2 tr\left[\Lambda_\alpha^2\right] \\
&= \frac{2\gamma_\alpha^2}{(h^*)^2}\left[h^2 \sum_{i=1}^{h} \frac{\theta_i^4}{g_i} - 2h \sum_{i=1}^{h} \theta_i \sum_{i=1}^{h} \frac{\theta_i^3}{g_i} + \left(\sum_{i=1}^{h} \theta_i\right)^2 \sum_{i=1}^{h} \frac{\theta_i^2}{g_i}\right] \\
&\quad + \frac{4\gamma_\alpha\gamma_\alpha}{(h^*)^2}\left[h^2 \sum_{i=1}^{h} \frac{\theta_i^3}{g_i} - 2h \sum_{i=1}^{h} \theta_i \sum_{i=1}^{h} \frac{\theta_i^2}{g_i} + \left(\sum_{i=1}^{h} \theta_i\right)^2 \sum_{i=1}^{h} \frac{\theta_i}{g_i}\right] \\
&\quad + \frac{2\gamma_\alpha^2}{(h^*)^2}\left[h^2 \sum_{i=1}^{h} \frac{\theta_i^2}{g_i} - 2h \sum_{i=1}^{h} \theta_i \sum_{i=1}^{h} \frac{\theta_i}{g_i} + \left(\sum_{i=1}^{h} \theta_i\right)^2 \sum_{i=1}^{h} \frac{1}{g_i}\right] \\
&= \frac{2}{(h^*)^2}\left(\lambda_4\gamma_\alpha^2 + 2\lambda_3\gamma_\alpha\gamma_\epsilon + \lambda_2\gamma_\epsilon^2\right). \qquad (2.10)
\end{aligned}$$

Part (b):

$$\begin{aligned}
\Sigma\left(\widehat{\gamma_\epsilon}^S\right) &= 2tr\left(y^\top \Lambda_\epsilon y\right) \\
&= 2\gamma_\alpha^2 tr\left[(\Lambda_\epsilon M)^2\right] + 4\gamma_\alpha\gamma_\epsilon tr\left[\Lambda_\epsilon M \Lambda_\epsilon\right] + 2\gamma_\epsilon^2 tr\left[\Lambda_\epsilon^2\right] \\
&= \frac{2\gamma_\alpha^2}{(h^*)^2} \sum_{j=1}^{h} \frac{\theta_j^2}{g_j}\left(\sum_{i=1}^{h} \theta_i(\theta_i - \theta_j)\right)^2 + \frac{4\gamma_\alpha\gamma_\epsilon}{(h^*)^2} \sum_{j=1}^{h} \frac{\theta_j}{g_j}\left(\sum_{i=1}^{h} \theta_i(\theta_i - \theta_j)\right)^2 \\
&\quad + \frac{\gamma_\epsilon^2}{(h^*)^2} \sum_{j=1}^{h} \frac{1}{g_j}\left(\sum_{i=1}^{h} \theta_i(\theta_i - \theta_j)\right)^2 \\
&= \frac{2}{(h^*)^2} \sum_{j=1}^{h} \left[\frac{\left(\sum_{i=1}^{h} \theta_i(\theta_i - \theta_j)\right)^2}{g_j}\right] \left(\gamma_\alpha^2\theta_j^2 + 2\gamma_\alpha\gamma_\epsilon\theta_j + \gamma_\epsilon^2\right). \qquad (2.11)
\end{aligned}$$

□

3. Estimation in Designs with two groups of treatments

It is not so evident a strict comparison between the variance-covariance of Sub-D and Anova-based estimators, but when the design has a fixed $k = 2$ groups of treatment, no matter the number of observation for each group, it seems that they are somehow comparable.

When $k = 2$ it follows that $N = n_1 + n_2$ and $n_0 = \frac{N^2-(n_1^2+n_2^2)}{N^2}$, and so the ANOVA-based estimators reduce to

$$\begin{aligned}\widehat{\gamma_\alpha}^A &= \frac{1}{n_0}\left[n_1(\bar{z_{1\bullet}} - \bar{z_{\bullet\bullet}})^2 + n_2(\bar{z_{2\bullet}} - \bar{z_{\bullet\bullet}})^2)\right] \\ &\quad -\frac{1}{n_o(N-2)}\left[\sum_{j=1}^{n_1}\left(z_{1j} - \bar{z_{1\bullet}}\right)^2 + \sum_{j=1}^{n_2}\left(z_{2j} - \bar{z_{2\bullet}}\right)^2\right] \text{ and} \\ \widehat{\gamma_\epsilon}^A &= \frac{1}{N-2}\left[\sum_{j=1}^{n_1}\left(z_{1j} - \bar{z_{1\bullet}}\right)^2 + \sum_{j=1}^{n_2}\left(z_{2j} - \bar{z_{2\bullet}}\right)^2\right].\end{aligned}$$

As we may easily conclude, their respective variance-covariance will be given as

$$\begin{aligned}\Sigma\left(\widehat{\gamma_\alpha}^A\right) &= 2\gamma_\alpha^2 + \left(\frac{2N}{n_1n_2}\right)\gamma_\alpha\gamma_\epsilon + \frac{N^2(N-1)}{2(n_1n_2)^2(N-2)}\gamma_\epsilon^2 \text{ and} \\ \Sigma\left(\widehat{\gamma_\epsilon}^A\right) &= \frac{2\gamma_\epsilon^2}{N-2}. \qquad (3.1)\end{aligned}$$

When $k = 2$, it follows that $h = 2$, that is M will only have two eigenvalues, θ_1 and θ_2, and since $r(M) = k - 1 = 1$ it follows that $\theta_2 = 0$. Under these conditions we have that

$$\Lambda_\alpha = \frac{P_1 - P_2}{\theta_1} \text{ and } \Lambda_\epsilon = P_2, \qquad (3.2)$$

and therefore the estimators boils down to

$$\widehat{\gamma_\alpha}^S = y^\top\left(\frac{P_1 - P_2}{\theta_1}\right)y \text{ and } \widehat{\gamma_\epsilon}^S = y^\top P_2 y.$$

The results for their respective variance-covariance follow as a consequente of Theorem 2.1.

Corollary 3.1. *Consider the conditions of Theorem 2.1, and let Let $k = 2$. Then,*

(a) $\Sigma\left(\widehat{\gamma_\alpha}^S\right) = 2\gamma_\alpha^2 + \frac{4}{\theta_1}\gamma_\alpha\gamma_\epsilon + 2\left(\frac{g_2+1}{g_2\theta_1^2}\right)\gamma\epsilon^2$;

(b) $\Sigma\left(\widehat{\gamma_\epsilon}^S\right) = \frac{2\gamma_\epsilon^2}{g_2}$.

Proof. Nothing that $h = 2$, and so $g_1 = 1$ and $\theta_2 = 0$, and applying Theorem 2.1 the results follow. □

It worth to notice that since both Sub-D and Anova-based estimators are unbiased their respective mean square error (MSE) are equal to their respective variance-covariance. This remark allows us to infer about the quality of these estimators.

Remark 3.1. *With $MSE(\hat{q})$ denoting the MSE of an estimator $\hat{q}$ of a parameter q, we notice the following:*

- *Sub-D: $MSE\left(\widehat{\gamma_\alpha}^S\right) = \Sigma\left(\widehat{\gamma_\alpha}^S\right)$ and $MSE\left(\widehat{\gamma_\epsilon}^S\right) = \Sigma\left(\widehat{\gamma_\epsilon}^S\right)$;*
- *Anova: $MSE\left(\widehat{\gamma_\alpha}^A\right) = \Sigma\left(\widehat{\gamma_\alpha}^A\right)$ and $MSE\left(\widehat{\gamma_\epsilon}^A\right) = \Sigma\left(\widehat{\gamma_\epsilon}^A\right)$.*

The next result gives a comparative framework of the estimators in design with two groups of treatment.

Proposition 3.1. *Let $k = 2$. Then:*

(a) $MSE\left(\widehat{\gamma_\epsilon}^S\right) = MSE\left(\widehat{\gamma_\epsilon}^A\right)$;

(b) $MSE\left(\widehat{\gamma_\alpha}^S\right) = MSE\left(\widehat{\gamma_\alpha}^A\right)$, *if* $\theta_1 = \frac{2n_1n_2}{n_1+n_2}$.

Proof. These results are consequences of Corollary 3.1. Indeed, since $r(M) = 1$ we have that $g_1 = k - 1 = 1$ and $g_2 = N - K = N - 2$, so that

$$\frac{4}{\theta_1} = \frac{2(n_1+n_2)}{n_1 n_2} = \frac{2N}{n_1 n_2} \text{ and } \frac{2(g_2+1)}{g_2\theta_1^2} = \frac{N^2(N-1)}{2\left(n_1 n_2\right)^2 (N-2)}, \tag{3.3}$$

provide $\theta_1 = \frac{2n_1 n_2}{n_1+n_2}$. $\square$

The condition $\theta_1 = \frac{2n_1n_2}{n_1+n_2}$ for which $MSE\left(\widehat{\gamma_\alpha}^S\right) = MSE\left(\widehat{\gamma_\alpha}^A\right)$ imposed in Proposition 3.1 consists in a measure to compare the quality of estimators, in the sense that if $\theta_1 < \frac{2n_1n_2}{n_1+n_2}$ it holds that Sub-D is better than Anova-based estimator for γ_α and Anova based estimator is better if $\theta_1 > \frac{2n_1n_2}{n_1+n_2}$. In fact, as we may see through simulations studies (see tables 1 and 2),

$$\theta_1 = \frac{2n_1n_2}{n_1+n_2}$$

whatever the values of n_1 and n_2, and so $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\alpha}^A$ have exactly the same MSE.

For some combinations of parameters γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$ we simulated $s = 10000$ repeated designs, using $\beta \sim \mathcal{N}(0,\ \gamma_\alpha)$ and $e \sim \mathcal{N}(0,\ \gamma_\epsilon)$, and $n_1 = 101$ and $n_2 = 20$. For each simulated design both estimators was applied and the parameters γ_α and γ_ϵ was estimated. Next, the average of the estimated values for the parameters was computed as well as the standard deviations of the respective estimated values. See the results in Tables 1 and 2 and an R function for simulating both estimators in tables 3. As we may see, independently of the configuration for the parameters γ_α and γ_ϵ as well as the configuration for the number of elements in each groups of treatment, the estimates and the respective standard deviations found are the same for both estimators.

Table 1. Simulations for different values of γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, with $n_1 = 101$, $n_2 = 20$ and $s = 10000$. **Actual value** denotes the actual values of the parameters; **Estimate** denotes the estimated values of the parameters; **Stand. Dev.** denotes the standard deviations of the estimated values.

Sub-D	γ_α	γ_ϵ	**ANOVA**	γ_α	γ_ϵ
Actual value	**0.5**	**1**	**AV**	**0.5**	**1**
Estimate	0.50129	0.99912	**Estimate**	0.50129	0.99912
Stand. Dev.	0.75534	0.12768	**Stand. Dev.**	0.75534	0.12768
Actual value	**1**	**0.5**	**AV**	**1**	**0.5**
Estimate	0.99809	0.50001	**Estimate**	0.9980	0.50001
Stand. Dev.	1.42519	0.06520	**Stand. Dev.**	1.42519	0.06520
Actual value	**0.75**	**0.5**	**AV**	**0.75**	**0.5**
Estimate	0.756304	0.50095	**Estimate**	0.75630	0.50095
Stand. Dev.	1.08095	0.06576	**Stand. Dev.**	1.08095	0.06576
Actual value	**0.5**	**0.75**	**AV**	**0.5**	**0.75**
Estimate	0.50144	0.74989	**Estimate**	0.50144	0.74989
Stand. Dev.	0.75143	0.09628	**Stand. Dev.**	0.75143	0.09628
Actual value	**0.5**	**0.1**.	**AV**	**0.5**	**0.1**
Estimate	0.49695	0.10004	**Estimate**	0.49695	0.10004
Stand. Dev.	0.71919	0.01301	**Stand. Dev.**	0.71919	0.01301
Actual value	**0.1**	**0.5**.	**AV**	**0.1**	**0.5**
Estimate	0.10171	0.50099	**Estimate**	0.10171	0.50099
Stand. Dev.	0.16582	0.06458	**Stand. Dev.**	0.16582	0.06458

Table 2. Simulations for different values of γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, with $n_1 = 20$, $n_2 = 101$ and $s = 10000$. **Actual value** denotes the actual values of the parameters; **Estimate** denotes the estimated values of the parameters; **Stand. Dev.** denotes the standard deviations of the estimated values.

Sub-D	γ_α	γ_ϵ	**ANOVA**	γ_α	γ_ϵ
Actual value	**0.5**	**1**	**AV**	**0.5**	**1**
Estimate	0.50746	1.00002	**Estimate**	0.50746	1.00002
Stand. Dev.	0.75025	0.12910	**Stand. Dev.**	0.75025	0.12910
Actual value	**1**	**0.5**	**AV**	**1**	**0.5**
Estimate	1.00721	0.50095	**Estimate**	1.00721	0.50095
Stand. Dev.	1.44209	0.06572	**Stand. Dev.**	1.44209	0.06572
Actual value	**0.75**	**0.5**	**AV**	**0.75**	**0.5**
Estimate	0.74427	0.50020	**Estimate**	0.74427	0.50020
Stand. Dev.	1.07430	0.06574	**Stand. Dev.**	1.07430	0.06574
Actual value	**0.5**	**0.75**	**AV**	**0.5**	**0.75**
Estimate	0.50204	0.75085	**Estimate**	0.50204	0.75085
Stand. Dev.	0.74323	0.09775	**Stand. Dev.**	0.74323	0.09775
Actual value	**0.5**	**0.1**.	**AV**	**0.5**	**0.1**
Estimate	0.50690	0.09980	**Estimate**	0.50690	0.09980
Stand. Dev.	0.71661	0.01299	**Stand. Dev.**	0.71661	0.01299
Actual value	**0.1**	**0.5**.	**AV**	**0.1**	**0.5**
Estimate	0.10217	0.49945	**Estimate**	0.10217	0.49945
Stand. Dev.	0.16710	0.06501	**Stand. Dev.**	0.16710	0.0650

Table 3. The R Code applied to Simulate and test Sub-D and ANOVA-based estimators in an unbalanced "one-way" random with two groups of treatments. Tables 1 and 2 show some examples.

With regard to the optimality of design (1.1) Sub-D allows to set theoretical and consistent results. Optimality designs provide accurate statistical inference by choosing the number of groups of treatments and number of observations at each group in oder to minimize the variance of estimating interested parameters, such as $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\alpha}^A$, which is our case.

According with Corollary 3.1,

$$\Sigma\left(\widehat{\gamma_\alpha}^S\right) = 2\gamma_\alpha^2 + \frac{4}{\theta_1}\gamma_\alpha\gamma_\epsilon + \frac{2(N-1)}{(N-2)\theta_1^2}\gamma\epsilon^2 \text{ and}$$
$$\Sigma\left(\widehat{\gamma_\epsilon}^S\right) = \frac{2\gamma_\epsilon^2}{N-2}, \tag{3.4}$$

recalling $g_2 = N-2$ and $\theta_1 = \frac{2n_1n_2}{n_1+n_2}$

Noting that θ_1 depends on N through n_1 and n_1, and $\frac{N-1}{N-2} \approx 1$ providing N is a large natural number, results in (3.4) allow us to remark that the bigger is θ_1 the smaller are $\Sigma\left(\widehat{\gamma_\alpha}^S\right)$ and $\Sigma\left(\widehat{\gamma_\epsilon}^S\right)$. More over, it can be proved that θ_1 is not greater than the maximum of n_1 and n_2.

Proposition 3.2. *Whatever n_1 and n_2,*

$$\theta_1 \leq max\{n_1, n_2\}.$$

Proof. Firstly, lets suppose $n_1 = n_2$. Then $\theta_1 = \frac{2n_1^2}{n_1} = n_1$.

Now, without lost of generality, let $n_1 > n_2$. Then there exists a natural number b holding $0 < b \leq n_1$ such that $n_2 = n_1 - b$. Thus,

$$\theta_1 = \frac{2n_1^2 - 2bn_1}{2n_1 - b}. \tag{3.5}$$

By contradiction, suppose $\theta_1 > n_1$, i.e.

$$\frac{2n_1^2 - 2bn_1}{2n_1 - b} > n_1 \leftrightarrow -bn_1 > 0,$$

which is an absurd since by definition $n_1 > 0$ and $b > 0$. Therefore, θ cannot be greater than n_1. For the case when $n_2 > n_1$ we proceed identically. □

The proof of Proposition 3.2 provide a robust tool to discuss the optimality of design (1.1) with respect to Sub-D. In fact, supposing (with out lost of generality) that $n_1 \geq n_2$ and so $n_2 = n_1 - b$ and $\theta = \frac{2n_1n_2}{n_1+n_2}$, for some natural b, we easily prove that

$$\theta_1 \longrightarrow n_1 \text{ as } b \longrightarrow 0. \tag{3.6}$$

In practice, this means the "more balanced" the model is the smaller the variances of $\widehat{\gamma_\alpha}^S$ and $\widehat{\gamma_\epsilon}^S$ are. In order to do that lets consider the real function $t(b) = \theta_1 = \frac{2n_1^2-2bn_1}{2n_1-b}$, $0 \leq b \geq n_1$. Thus, we found the following: since $t'(b) = \frac{-2n_1^2}{(2n_1-b)^2} < 0$ (meaning that θ_1 is a decreasing function of b) and $t''(b) = \frac{-4n_1^2}{(2n_1-b)^3} < 0$ (meaning that θ_1 is a face-down concavity function of b), results (3.6) follows. $t'(b)$ and $t''(b)$ denote the first and second derivate of function $t(b)$ at b, respectively.

4. Estimation in Balanced "One-Way" Designs

For random "one-way" balanced designs, that is the case when $n_i = n$, $i =, 1 \ldots, k$, the ANOVA estimators for variance components γ_1 and γ_2, are given as (see Sahai and Ojeda [3]).

$$\begin{aligned} \widehat{\gamma_\alpha}^{Ab} &= \frac{1}{n}\left[\left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n(\bar{z_{i\bullet}} - \bar{z_{\bullet\bullet}})^2 - \left(\frac{1}{k(n-1)}\right)\sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z_{i\bullet}})^2\right] \\ \widehat{\gamma_\epsilon}^{Ab} &= \left(\frac{1}{k(n-1)}\right)\sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z_{i\bullet}})^2, \end{aligned} \tag{4.1}$$

with $n = n_i$, $\bar{z_{i\bullet}} = \frac{1}{n}\sum_{j=1}^{n} z_{ij}$ and $\bar{z_{\bullet\bullet}} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}$. The variance of the ANOVA estimators $\widehat{\gamma_\alpha}^{Ab}$ and $\widehat{\gamma_\epsilon}^{Ab}$ are respectively given as

$$\begin{aligned} \Sigma\left(\widehat{\gamma_\alpha}^{Ab}\right) &= \frac{2\gamma_1^2}{k-1} + \frac{4\gamma_1\gamma_2}{n(k-1)} + \frac{2k(n-1)\gamma_2^2}{kn^2(n-1)(k-1)} \text{ and} \\ \Sigma\left(\widehat{\gamma_\epsilon}^{Ab}\right) &= \frac{2\gamma_\epsilon^2}{k(n-1)}. \end{aligned} \tag{4.2}$$

When discussing Sub-D for such a design, we found that M has only two eigenvalues: $\theta_1 = n$ with multiplicity $g_1 = k - 1$, and $\theta_2 = 0$ with multiplicity $g_2 = N - k = k(n-1)$.

In this case the respective Sub-D estimators for variance components γ_α and γ_ϵ, become:

$$\widehat{\gamma_\alpha}^{Sb} = y^\top (\Lambda_{\alpha b})\, y \text{ and } \widehat{\gamma_\epsilon}^{Sb} = y^\top (\Lambda_{\epsilon b})\, y, \tag{4.3}$$

where

$$\Lambda_{\alpha b} = \frac{A_1^\top A_1}{n(k-1)} - \frac{A_2^\top A_2}{nk(n-1)} \text{ and } \Lambda_{\epsilon b} = \frac{A_2^\top A_2}{k(n-1)}. \tag{4.4}$$

As a consequence of Proposition 2.1 we find that:

$$\begin{aligned} \Sigma\left(\widehat{\gamma_\alpha}^{Sb}\right) &= \frac{2}{k-1}\gamma_\alpha^2 + \frac{4}{n(k-1)}\gamma_\alpha^2\gamma_\epsilon^2 + \frac{2(kn-1)}{kn^2(n-1)(n-1)}\gamma_\epsilon^2 = \Sigma\left(\widehat{\gamma_\alpha}^{Ab}\right), \text{ and} \\ \Sigma\left(\widehat{\gamma_\epsilon}^{Sb}\right) &= \frac{2\gamma_\epsilon^2}{k(n-1)} = \Sigma\left(\widehat{\gamma_\epsilon}^{Ab}\right), \end{aligned} \tag{4.5}$$

and so, consequently, we have the following corollary.

Corollary 4.1. *Let $n_i = n$, $i = 1, \ldots, k$. Then:*

(a) $MSE\left(\widehat{\gamma_\epsilon}^S\right) = MSE\left(\widehat{\gamma_\epsilon}^A\right)$;
(b) $MSE\left(\widehat{\gamma_\alpha}^S\right) = MSE\left(\widehat{\gamma_\alpha}^A\right)$.

For simulation purpose, we toke the same combinations of the parameters γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, and simulated $s = 10000$ repeated designs, using $\beta \sim \mathcal{N}(0,\ \gamma_\alpha)$ and $e \sim \mathcal{N}(0,\ \gamma_\epsilon)$ and $k = 10$ and $n = 23$. For each simulated design, both estimators are applied and the parameters γ_α and γ_ϵ was estimated. Then the average of the estimated values for the parameters was computed as well as the standard deviations of the respective estimated values. The number of groups of treatments and number of observations for each groups was respectively chosen as $k = 10$ and $n = 23$. These values was chosen with no reason other then the simulation purpose. As shown through theoretical results the estimates for both estimators will be equal no matter the number of groups and number of observations for each groups are taken. The results are in Table 4 and an R function to simulate both estimators in Table 5. As we may see, independently of the configuration for the parameters γ_α and γ_ϵ, the estimates and the respective standard deviations found are the same for both estimators.

Table 4. Simulations for different values of γ_α and γ_ϵ ranging over $\{0.1, 0.5, 0.75, 1.0\}$, with $k = 10$, $n = 23$ and $s = 10000$. **Actual value** denotes the actual values of the parameters; **Estimate** denotes the estimated values of the parameters; **Stand. Dev.** denotes the standard deviations of the estimated values.

Sub-D	γ_α	γ_ϵ	ANOVA	γ_α	γ_ϵ
Actual value	**0.5**	**1**	**AV**	**0.5**	**1**
Estimate	0.49668	0.99980	**Estimate**	0.49668	0.99980
Stand. Dev.	0.25322	0.09545	**Stand. Dev.**	0.25322	0.09545
Actual value	**1**	**0.5**	**AV**	**1**	**0.5**
Estimate	0.99885	0.50044	**Estimate**	0.99885	0.50044
Stand. Dev.	0.47794	0.04757	**Stand. Dev.**	0.47794	0.04757
Actual value	**0.75**	**0.5**	**AV**	**0.75**	**0.5**
Estimate	0.75206	0.49970	**Estimate**	0.75206	0.49970
Stand. Dev.	0.36972	0.04706	**Stand. Dev.**	0.36972	0.04706
Actual value	**0.5**	**0.75**	**AV**	**0.5**	**0.75**
Estimate	0.50078	0.74935	**Estimate**	0.50078	0.74935
Stand. Dev.	0.25221	0.07101	**Stand. Dev.**	0.25221	0.07101
Actual value	**0.5**	**0.1.**	**AV**	**0.5**	**0.1**
Estimate	0.50214	0.09996	**Estimate**	0.50214	0.09996
Stand. Dev.	0.23539	0.00954	**Stand. Dev.**	0.23539	0.00954
Actual value	**0.1**	**0.5.**	**AV**	**0.1**	**0.5**
Estimate	0.10025	0.49985	**Estimate**	0.10025	0.49985
Stand. Dev.	0.05769	0.04768	**Stand. Dev.**	0.05769	0.04768

Table 5. The R Code applied to Simulate and test Sub-D and ANOVA-based estimators in random "one-way" balanced designs. Table 4 shows an example.

5. Discussion

As we may see in Silva et al. [2], Silva [12], Silva [11] and Silva et al.[13], through simulations studies, Sub-D has proven its value. When compared to Anova-based estimator it was shown that Sub-D produces estimates with unbiased and lower standard deviations, except in case of random "one-way" balanced designs. In this sense we tough convenient to investigate the performance of both estimators in such a designs; and we found that not only they have the same performance in random "one-way" balanced designs but also in random "one-way" designs with two groups of treatments. In fact this was proven through theoretical results (see Proposition 3 and Corollary 4.1), corroborated with simulations studies (see tables 1, 2, and 3, regarding the random "one-way" designs with two groups of treatments, and tables 4 and 5, regarding the random "one-way" balanced designs.)

Acknowledgements

References

[1] Demidenko, E., Mixed Models: Theory and Applications (Wiley Series in Probability and Statistics), Wiley-Interscience, 2004.

[2] Ferreira, D., Ferreira, S., Nunes, C., Fonseca, M., Silva, A. and Mexia, J., *Estimation and incommutativity in mixed models*, Journal of Multivariate Analysis **161**, 58-67, 2017.

[3] Hardeo, S. and Ojeda, M., *Analysis of Variance for Random Models. Unbalanced Data, Theory, Methods, Applications, and Data Analysis*, 2003.

[4] Khuri, A., *Design for variance components estimation: past and present*, International Statistical Review, Vol. 68 (3), 311–322, 2000.

[5] Khuri, A., Mathew, T. and Sinha, B., 'Statistical tests for mixed linear models', *John Wiley & Sons, Inc., New York*, 1998.

[6] Mulder, M. B., Caro, T. M., Chrisholm, J. S., Dumont, J. P., Hall, R. L., Hinde, R. A., and Ohtsuka, R., *The use of quantitative observational techniques in anthropology [and comments and replies*, Current Anthropology, **26** (3), 323-335, 1985.

[7] Pinheiro, J. and Bates, D., *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing Book Series (SCO), 2000.

[8] Searle, S. R., *Topics in variance components estimation*, Biometrics, **27**, 1-76, 1971.

[9] Searle, S .R., *Linear Models*, Wiley, New York, 1971.

[10] Searle, S. R., Casella, G. and McCulloch, C. E., *Variance Components*, Wiley Series in Probabliity and Statistics, 1992.

[11] Silva, A., *Variance components estimation in mixed linear models*, Ph.D Thesis, New University of Lisbon, 2017.

[12] Silva, A. and Fonseca, M., *The method sub-d for variance components estimation in random one-way designs*, International Journal of Multivariate Data Analysis, **1** (4), 327-336, 2018.

[13] Silva, A., Moreira, A. and Fonseca, M., *Inference in Mixed Linear Models with four Variance Components - Sub-D and Sub-DI*, Int. J. of Data Analysis Techniques and Strategies, **12** (4), 2020.

[14] Singh, B., *On effect of unbalancedness and heteroscedasticity on the ANOVA estimator of group variance component in one-way random model*, Biometrical J., **34**, 91-96, 1992.

[15] Wallace, D. D., *An approximate f statistic for testing population effects in longitudinal studies via mixed models*, Ph.D Thesis, University of North Carolina at Chapel Hill, 1993.