Listy Biometryczne - Biometrical Letters
Vol. 33 (1996), No. 1, 1-23

On the existence of BLUEs under a randomization model for the randomized block design

Tadeusz Caliński[1]

Department of Mathematical and Statistical Methods,
Agricultural University of Poznań,
Wojska Polskiego 28, 60-637 Poznań, Poland

SUMMARY

The paper reconsiders the existence of the BLUEs under a randomization model for the randomized block design, questioned by Kaiser (1989; *J. Statist. Plann. Inference* **22**, 63-69). It is shown how two different randomization models for this design imply the existence, under certain conditions, of the BLUEs of the usual linear parametric functions of interest. The role of the double randomization, within and between blocks, is indicated. Finally, the reason for the negative conclusion drawn in Kaiser's paper is discussed.

KEY WORDS: best linear unbiased estimation, randomization models, randomization block design.

1. Introduction

In a paper by Kaiser (1989) it has been claimed that under a simple randomization model for the randomized block design the usual estimator of a contrast of treatment parameters is not the best linear unbiased and, moreover, that no best linear unbiased estimator (BLUE) exists for any such contrast. The aim of the present paper is to reconsider the problem of existence of the BLUEs, under two different randomization models, and to discuss the negative conclusion drawn by Kaiser (1989).

When describing the randomized block design, authors of the classical experimental design books pointed out the necessity of a random assignment of treatments to units (plots) within blocks (see, e.g., Fisher, 1935, Section 22; Kempthorne, 1952, Section 9.1; Cochran and Cox, 1957, Section 4.2; Cox, 1958, Chapter 5; Scheffé, 1959, Section 9.1; Finney, 1960, Section 3.1). However, in formulating mathematical models

---

[1] This paper was delivered as a lecture at the XXXVII Session of the Polish Biometric Society, Poznań, February 2-3, 1996.

for the analysis of experimental results obtained under the randomized block design, the role of randomization has not been understood in the same way by various authors. In most of the literature, the models discussed are based on the assumption that the residuals, caused by the apparent lack of homogeneity of experimental units, are uncorrelated and distributed around zero with equal variance. Such a model, called the assumed linear model, does not appreciably take into account the randomization implemented in designing and carrying out the experiment. Different models have been obtained when really recognizing, in the process of mathematical derivations, the randomization techniques employed. Among these more realistic approaches to modelling randomized experiments, two lines are of particular interest. One, which is confined to the intra-block randomization, stems back to the early works by Neyman (1923, 1935) and has been continued by Kempthorne (1952, 1955) and his followers (see, e.g., Zyskind, 1975). A model derived from this intra-block randomization approach and its implications for estimating linear parametric functions are considered in Section 2. Another line of formulating a randomization model was initiated by Nelder (1954), under some influence of Anscombe (1948), and extended by Nelder (1965a). A general exposition of this approach has been given by Bailey (1981). For the randomized block design case, an essential feature of this approach is that not only the intra-block but also the inter-block randomization is taken into account in formulating the model. A model so derived and implications following from it for linear parametric estimation are subjects of Section 3. Results presented in Sections 2 and 3 are in contradiction to the conclusion of Kaiser (1989). The reasons for this are discussed in Section 4. In an appendix (Section 5) a useful criterion for assessing whether a linear statistic is or is not the BLUE of its expectation is recalled.

2. A model derived from the intra-block randomization

As in the paper by Kaiser (1989), the model considered here is for the analysis of results of an experiment in the randomized (complete) block design with v treatments (or varieties) compared in b blocks of v units each. Under the usual unit-treatment additivity assumption, also adopted in Kaiser's paper, it is permissible to represent the conceptual true response to treatment j on unit k in block i by $\mu_{ik} + \tau_j$, where μ_{ik} and τ_j are some unknown parameters. It may also be justified to add a technical error component, as some authors do (e.g., Neyman, 1935; Kempthorne, 1952, pp. 132 and 151; Scheffé, 1959, p.293), but this will be avoided here for purely comparative reasons, to keep to the model used by Kaiser (1989, Section 2). A basic assumption in formulating the model for the observed responses is that the treatments are assigned to units (plots) at random in each block, independently from block to block. It follows from the generally accepted rule, described, e.g., by Finney (1960, p.23):

Before randomization

Block 1				Block 2				Block 3				Block 4				Block 5			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

After randomization

Block 1				Block 2				Block 3				Block 4				Block 5			
3	*4*	*2*	*1*	*4*	*3*	*1*	*2*	*2*	*1*	*4*	*3*	*1*	*2*	*4*	*3*	*1*	*3*	*2*	*4*

Fig.1. Randomization for an experiment with four treatments and five blocks. Here the unit labels are randomly permuted within blocks taken in the order given by their original labels.

"The procedure for any block is to select one plot at random for the first treatment, another at random for the second treatment, and so on, and to use a new random order in each block." This rule can be illustrated as shown in Figure 1.

Under the above assumptions of additivity and intra-block randomization, let y_{ij} denote the observed response to treatment j in block i. To use the matrix notation, let the observations be represented by

$$\mathbf{y} = [y_{11}, ..., y_{1v}, ..., y_{b1}, ..., y_{bv}]',$$

and the parameters by

$$\boldsymbol{\mu} = [\mu_{11}, ..., \mu_{1v}, ..., \mu_{b1}, ..., \mu_{bv}]' \quad \text{and} \quad \boldsymbol{\tau} = [\tau_1, ..., \tau_v]'.$$

Then the model can be written as

$$\mathbf{y} = (\mathbf{1}_b \otimes \mathbf{I}_v)\boldsymbol{\tau} + \boldsymbol{\rho}_r \boldsymbol{\mu} \tag{2.1}$$

[formula (2.1) of Kaiser, 1989], where $\boldsymbol{\rho}_r$ is a permutation matrix determined by the randomization (hence the subscript). Let the design random variables δ^j_{ik} (introduced by Kempthorne, 1952, Section 8.2) be defined as

$$\delta^j_{ik} = \begin{cases} 1 & \text{if a unit in block } i \text{ originally labelled } k \text{ receives by} \\ & \text{the randomization the label } j \text{ (to get the treatment } j\text{),} \\ 0 & \text{otherwise,} \end{cases}$$

for $j, k = 1, ..., v$ and $i = 1, ..., b$. Then the matrix $\boldsymbol{\rho}_r$ can be written explicitly as

$$\boldsymbol{\rho}_r = \text{diag}\ [\boldsymbol{\Delta}_1 : \ldots : \boldsymbol{\Delta}_b], \quad \text{where } \boldsymbol{\Delta}_i = [\delta_{ik}^j]$$

($j = 1, \ldots, v$, for rows, and $k = 1, \ldots, v$, for columns of $\boldsymbol{\Delta}_i$, $i = 1, \ldots, b$).

From the known distribution properties of the variables δ_{ik}^j (see, e.g., Scheffé, 1959, Section 9.1), the random vector $\mathbf{y}$ of the model (2.1) can be shown to have the properties

$$\text{E}(\mathbf{y}) = (\mathbf{1}_b \otimes \mathbf{I}_v)\boldsymbol{\tau} + (\mathbf{I}_b \otimes \mathbf{1}_v)\boldsymbol{\mu}_{(\cdot)} \tag{2.2}$$

and

$$\text{Cov}(\mathbf{y}) = \text{diag}[\sigma_{U,1}^2, \ldots, \sigma_{U,b}^2] \otimes (\mathbf{I}_v - v^{-1}\mathbf{1}_v\mathbf{1}_v'), \tag{2.3}$$

where $\boldsymbol{\mu}_{(\cdot)} = [\mu_{1.}, \ldots, \mu_{b.}]'$ and

$$\sigma_{U,i}^2 = (v-1)^{-1}\sum_{k=1}^{v}(\mu_{ik} - \mu_{i.})^2, \tag{2.4}$$

with $\mu_{i.} = v^{-1}\sum_{k=1}^{v}\mu_{ik}$, for $i = 1, \ldots, b$. (This dot notation for averaging will be used throughout the paper.)

Now it will be interesting to see whether the model (2.1) gives rise to the same BLUEs as those obtainable under the usual assumed linear model in which the unit errors are uncorrelated and identically distributed with zero mean and constant variance [i.e. the model (2.2) of Kaiser, 1989]. Note that, with regard to this, a candidate for a BLUE is a linear function $b^{-1}(\mathbf{1}_b' \otimes \mathbf{c}')\mathbf{y}$, where $\mathbf{c}$ is a $v \times 1$ vector of some constant coefficients. Its expectation is $\mathbf{c}'\boldsymbol{\tau} + \mathbf{c}'\mathbf{1}_v\mu_{..}$, where $\mu_{..} = b^{-1}\mathbf{1}_b'\boldsymbol{\mu}_{(\cdot)}$. So, the function $b^{-1}(\mathbf{1}_b' \otimes \mathbf{c}')\mathbf{y}$ is unbiased for $\mathbf{c}'\boldsymbol{\tau} + \mathbf{c}'\mathbf{1}_v\mu_{..}$. But is it the BLUE? To answer this question the well known criterion of Zyskind (1967, Theorem 3), recalled in the Appendix, can be applied. According to it, a necessary and sufficient condition for $b^{-1}(\mathbf{1}_b' \otimes \mathbf{c})\mathbf{y}$ to be the BLUE of its expectation, is the equality

$$(\mathbf{I}_{bv} - \mathbf{P})\text{Cov}(\mathbf{y})(\mathbf{1}_b \otimes \mathbf{c}) = \mathbf{0}, \tag{2.5}$$

where $\mathbf{P}$ is the orthogonal projector on the column space of the matrix $[\mathbf{1}_b \otimes \mathbf{I}_v : \mathbf{I}_b \otimes \mathbf{1}_v]$. Since

$$\mathbf{I}_{bv} - \mathbf{P} = (\mathbf{I}_b - b^{-1}\mathbf{1}_b\mathbf{1}_b') \otimes (\mathbf{I}_v - v^{-1}\mathbf{1}_v\mathbf{1}_v'),$$

and $\text{Cov}(\mathbf{y})$ is of the form (2.3), the condition (2.5) can equivalently be written as

$$[\sigma_{U,1}^2 - \sigma_{U,.}^2, \ldots, \sigma_{U,b}^2 - \sigma_{U,.}^2]' \otimes (\mathbf{I}_v - v^{-1}\mathbf{1}_v\mathbf{1}_v')\mathbf{c} = \mathbf{0}, \tag{2.6}$$

where

$$\sigma_{U,.}^2 = b^{-1}\sum_{i=1}^{b}\sigma_{U,i}^2. \tag{2.7}$$

Evidently, (2.6) is satisfied if and only if either $\mathbf{c}$ is proportional to the vector $\mathbf{1}_v$ or the equality

$$\sigma^2_{U,1} = ... = \sigma^2_{U,b} \qquad (= \sigma^2_{U,.}) \tag{2.8}$$

holds.

Thus, it has been shown that

(i) the parametric function $\tau_. + \mu_{..}$, where $\tau_. = v^{-1}\mathbf{1}'_v\boldsymbol{\tau}$, i.e. the overall true response averaged over the v treatments, has the BLUE in the form

$$\widehat{\tau_. + \mu_{..}} = (vb)^{-1}\mathbf{1}'_{bv}\mathbf{y},$$

i.e. equal to the general observed mean;

(ii) a parametric function $\mathbf{c}'\boldsymbol{\tau}$, where $\mathbf{c}'\mathbf{1}_v = 0$, i.e. a contrast of treatment parameters, has the BLUE in the form

$$\widehat{\mathbf{c}'\boldsymbol{\tau}} = b^{-1}(\mathbf{1}'_b \otimes \mathbf{c}')\mathbf{y} \tag{2.9}$$

if and only if the equality (2.8) holds, i.e., the unit error variances (2.4) are constant for all the b blocks.

Certainly, the condition required in (ii) is a well known result. Zyskind (1975, p.654), e.g., writes: "In the case of the complete randomized block design homogeneity of the various intrablock variances is required".

3. A model derived from the intra-block and inter-block randomization

Suppose that, in addition to the assumptions made in Section 2, it is assumed that not only the units are randomized within the blocks but also the blocks are independently randomized between themselves. This assumption follows from the rule given by Nelder (1954, Section 2): "choose a block at random and reorder its members at random (...); repeat the procedure with one of the remaining blocks chosen at random (...), and so on". This rule can be illustrated as shown in Figure 2. Note the difference in the randomizations between Figures 1 and 2.

Following, as before, the unit-treatment additivity approach (also accepted by Nelder, 1965b, Section 3) and using the same definitions of the vectors $\mathbf{y}$, $\boldsymbol{\mu}$, and $\boldsymbol{\tau}$, as in Section 2, one is justified in writing the model in the form

$$\mathbf{y} = (\mathbf{1}_b \otimes \mathbf{I}_v)\boldsymbol{\tau} + \boldsymbol{\Pi}_r\boldsymbol{\mu}, \tag{3.1}$$

where the permutation matrix $\boldsymbol{\Pi}_r$ determined by the randomization differs now from $\boldsymbol{\rho}_r$ in (2.1) by the fact that not only the labels of units within blocks but also the labels of blocks are randomized. Using, in accordance with this, additional design

Before randomization

Block 1				Block 2				Block 3				Block 4				Block 5			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

After randomization

Block *3*				Block *1*				Block *5*				Block *4*				Block *2*			
2	*1*	*4*	*3*	*3*	*4*	*2*	*1*	*1*	*3*	*2*	*4*	*1*	*2*	*4*	*3*	*4*	*3*	*1*	*2*

Fig.2. Randomization for an experiment with four treatments and five blocks. Here the unit labels are randomly permuted within blocks taken in the order given by their randomly permuted labels 3, 1, 5, 4, 2. (For the sake of comparison, it is assumed that the random permutations within blocks have been obtained exactly as in Fig. 1.)

random variables γ_i^h , defined as

$$\gamma_i^h = \begin{cases} 1 & \text{if a block originally labelled } i \text{ receives} \\ & \text{by the randomization the label } h, \\ 0 & \text{otherwise,} \end{cases}$$

for $h, i = 1, ..., b$, the matrix $\mathbf{\Pi}_r$ can be written explicitly as

$$\mathbf{\Pi}_r = (\mathbf{\Gamma} \otimes \mathbf{I}_v)\boldsymbol{\rho}_r, \quad \text{where} \quad \mathbf{\Gamma} = [\gamma_i^h]$$

($h = 1, ..., b$, for rows, and $i = 1, ..., b$, for columns of $\mathbf{\Gamma}$).

Due to the distribution properties of the variables γ_i^h, similar to those of δ_{ik}^j's , the random vector $\mathbf{y}$ of the model (3.1) has the properties

$$\mathrm{E}(\mathbf{y}) = (\mathbf{1}_b \otimes \mathbf{I}_v)\boldsymbol{\tau} + \mathbf{1}_{bv}\mu_{..} \tag{3.2}$$

and

$$\mathrm{Cov}(\mathbf{y}) = \sigma_B^2(\mathbf{I}_b - b^{-1}\mathbf{1}_b\mathbf{1}_b') \otimes \mathbf{1}_v\mathbf{1}_v' + \sigma_U^2\mathbf{I}_b \otimes (\mathbf{I}_v - v^{-1}\mathbf{1}_v\mathbf{1}_v'), \tag{3.3}$$

where $\sigma_U^2 = \sigma_{U,\cdot}^2$, i.e. the average defined in (2.7).

The results (3.2) and (3.3) coincide with those given by Nelder (1954), since his constants σ^2, ρ_1 and ρ_2 can be defined as follows:

$$\sigma^2 = \frac{b-1}{b}\sigma_B^2 + \frac{v-1}{v}\sigma_U^2, \quad \rho_1 = \left(\frac{b-1}{b}\sigma_B^2 - \frac{1}{v}\sigma_U^2\right)/\sigma^2 \quad \text{and} \quad \rho_2 = -\frac{1}{b}\sigma_B^2/\sigma^2.$$

Now, it is interesting to establish for which linear functions of the expectation vector (3.2) the BLUEs exist. Applying the criterion of Zyskind (1967) (see Appendix), a necessary and sufficient condition for a linear function $\mathbf{w}'\mathbf{y}$ to be the BLUE of $\mathrm{E}(\mathbf{w}'\mathbf{y}) = \mathbf{w}'[(\mathbf{1}_b \otimes \mathbf{I}_v)\boldsymbol{\tau} + \mathbf{1}_{bv}\mu_{..}]$ can be found. It is of the form

$$[(\mathbf{I}_b - b^{-1}\mathbf{1}_b\mathbf{1}_b') \otimes \mathbf{I}_v][\sigma_B^2(\mathbf{I}_b - b^{-1}\mathbf{1}_b\mathbf{1}_b') \otimes \mathbf{1}_v\mathbf{1}_v' + \sigma_U^2\mathbf{I}_b \otimes (\mathbf{I}_v - v^{-1}\mathbf{1}_v\mathbf{1}_v')]\mathbf{w} = \mathbf{0},$$

and is satisfied for any values of σ_B^2 and σ_U^2 if and only if the equalities

$$[(\mathbf{I}_b - b^{-1}\mathbf{1}_b\mathbf{1}_b') \otimes \mathbf{1}_v\mathbf{1}_v']\mathbf{w} = \mathbf{0}$$

and

$$[(\mathbf{I}_b - b^{-1}\mathbf{1}_b\mathbf{1}_b') \otimes \mathbf{I}_v]\mathbf{w} = \mathbf{0}$$

hold simultaneously. Evidently, this is satisfied if and only if the vector $\mathbf{w}$ belongs to the column space of the matrix $(\mathbf{1}_b \otimes \mathbf{I}_v)$.

Thus, for any $v \times 1$ vector $\mathbf{c}$ and any scalar s, the function $(s\mathbf{1}_b' \otimes \mathbf{c}')\mathbf{y}$ is the BLUE of its expectation, $sb(\mathbf{c}'\boldsymbol{\tau} + \mathbf{c}'\mathbf{1}_v\mu_{..})$. In particular, taking $s = b^{-1}$, the function $(b^{-1}\mathbf{1}_b' \otimes \mathbf{c}')\mathbf{y}$ is for any $v \times 1$ vector $\mathbf{c}$ the BLUE of $\mathbf{c}'\boldsymbol{\tau} + \mathbf{c}'\mathbf{1}_v\mu_{..}$. From the definition of the vector $\mathbf{y}$ in (3.1), the estimator can be written as

$$\mathbf{c}'(\widehat{\boldsymbol{\tau} + \mathbf{1}_v\mu_{..}}) = \mathbf{c}'(b^{-1}\sum_{h=1}^{b}\mathbf{y}_h),$$

where $\mathbf{y}_h = [y_{h1}, ..., y_{hv}]'$, for $h = 1, ..., b$, which is the same as (2.9) if $\mathbf{c}'\mathbf{1}_v = 0$. Since the result holds for any $\mathbf{c}$, it follows that the BLUE exists also directly for the vector $\boldsymbol{\tau} + \mathbf{1}_v\mu_{..}$, in the form

$$\widehat{\boldsymbol{\tau} + \mathbf{1}_v\mu_{..}} = b^{-1}\sum_{h=1}^{b}\mathbf{y}_h.$$

So, it can be concluded that the average true response to any treatment is best linearly and unbiasedly estimated by the corresponding observed mean response, taken over the b units to which the treatment has been applied according to the complete randomized block design, provided that the randomization is performed independently not only within each block but also between the blocks. Due to this double randomization, the covariance matrix of $\mathbf{y}$ is modified from (2.3) to (3.3), and hence the equality (2.8) is not required any more.

It should be mentioned that the model (3.1) is a particular case of that used by Rao (1959) for incomplete block experiments, and that the estimation results presented here are in full agreement with the more general results given by Kala (1991) and by Caliński and Kageyama (1991, 1996).

4. Discussion on Kaiser's result

In the paper by Kaiser (1989) the known result given in Section 2 of the present paper, under (ii), is accompanied by the following statement. "If (2.4) is constant for all i then (2.9) is best among unbiased estimators of the form $\mathbf{w}'\mathbf{y}$ with $\mathbf{w}$ (being) a vector of *fixed* constants". Then, however, it is argued that: "In a randomized block design (...) the data is not only the vector of responses, $\mathbf{y}$, it is also known which units received each treatment". Because of this, it is claimed (Kaiser, 1989, p. 65) that the search for the best estimator should be made "among unbiased estimators, linear in $\mathbf{y}$, but where the coefficient of y_{ij} may depend on the randomization". By such extension of the class of estimators a theorem (Theorem 1) is proved, which exhibits an estimator that is unbiased, linear (in the above sense) and has zero variance at any preselected values of the parameters. From this theorem it is concluded that (2.9) cannot be the BLUE of a contrast $\mathbf{c}'\boldsymbol{\tau}$, even if the condition (2.8) is satisfied.

To see whether this conclusion is justified, let the estimator used in that theorem be written in the notation of the present paper, i.e. as

$$\widetilde{\mathbf{c}'\boldsymbol{\tau}} = \widehat{\mathbf{c}'\boldsymbol{\tau}} - b^{-1}(\mathbf{1}_b' \otimes \mathbf{c}')\boldsymbol{\rho}_r\boldsymbol{\mu}^0\mathbf{1}_{bv}'\mathbf{y}/(b\mathbf{1}_v'\boldsymbol{\tau}^0 + \mathbf{1}_{bv}'\boldsymbol{\mu}^0), \tag{4.1}$$

where $\boldsymbol{\mu}^0$ and $\boldsymbol{\tau}^0$ are some preselected arbitrary vectors (of order $bv \times 1$ and $v \times 1$, respectively), and $\mathbf{c}$ defines a contrast, i.e. $\mathbf{c}'\mathbf{1}_v = 0$, for which $\widehat{\mathbf{c}'\boldsymbol{\tau}}$ is as in (2.9). In fact, under the model (2.1), the estimator (4.1) is unbiased for $\mathbf{c}'\boldsymbol{\tau}$. Also, at $\boldsymbol{\mu} = \boldsymbol{\mu}^0$ and $\boldsymbol{\tau} = \boldsymbol{\tau}^0$ the estimator becomes equal to $\mathbf{c}'\boldsymbol{\tau}^0 = \mathbf{c}'\boldsymbol{\tau}$, and thus has zero variance. But can it be called a linear estimator in the sense used in the general Gauss-Markov theorem? To answer this question write the estimator (4.1), equivalently, as

$$\widetilde{\mathbf{c}'\boldsymbol{\tau}} = b^{-1}(\mathbf{1}_b' \otimes \mathbf{c}')[\mathbf{I}_{bv} - (b\mathbf{1}_v'\boldsymbol{\tau}^0 + \mathbf{1}_{bv}'\boldsymbol{\mu}^0)^{-1}\boldsymbol{\rho}_r\boldsymbol{\mu}^0\mathbf{1}_{bv}']\mathbf{y},$$

i.e. in the form $\mathbf{w}'\mathbf{y}$, to note that the vector $\mathbf{w}$ here is not of (known) constant coefficients. In fact all of them are functions of the δ_{ik}^j's and so are random variables. Therefore, the estimator $\widetilde{\mathbf{c}'\boldsymbol{\tau}}$ cannot be named linear in the usual sense (see, e.g., Zyskind, 1975, Section 3; Harville, 1976, Section 2), and so cannot compete with $\widehat{\mathbf{c}'\boldsymbol{\tau}}$ in the class of linear unbiased estimators of $\mathbf{c}'\boldsymbol{\tau}$.

Thus, Theorem 1 of Kaiser (1989) does not show that $\widehat{\mathbf{c}'\boldsymbol{\tau}}$, defined in (2.9), is not the BLUE of the contrast $\mathbf{c}'\boldsymbol{\tau}$ under the model (2.1). Also, it cannot be used as an argument for the nonexistence of the BLUE of any contrast, either under the randomization model (2.1) or under that in (3.1), if the notion of BLUE is to be understood in the sense used in the general Gauss-Markov theorem.

Furthermore, in the case of a randomized experiment it does not seem appropriate to demand an extension of the linear estimation under a randomization model by using estimators linear in $\mathbf{y}$ but with coefficients depending on the randomization. Once the randomization has been applied to the units of the experiment and the treatments have been assigned to the units accordingly, the knowledge of these facts is to be included into the model, by incorporating it in the appropriate permutation matrix, as indicated in Sections 2 and 3 (see also Zyskind, 1975, Section 2). Then it is sufficient to base the analysis on the derived linear model directly, since, as distinct from finite population sampling, no further randomization of the observed responses is needed for estimating the linear parametric functions in the best linear unbiased way.

5. Appendix

THEOREM (Zyskind, 1967). *Under the assumptions* $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ *and* $Cov(\mathbf{y}) = \mathbf{V}$, *a known linear function* $\mathbf{w}'\mathbf{y}$ *is the best linear unbiased estimator (BLUE) of its expectation,* $\mathbf{w}'\mathbf{X}\boldsymbol{\beta}$, *if and only if the condition* $(\mathbf{I} - \mathbf{P_X})\mathbf{V}\mathbf{w} = \mathbf{0}$ *holds, where* $\mathbf{P_X}$ *denotes the orthogonal projector on* $\mathcal{C}(\mathbf{X})$, *the column space of* $\mathbf{X}$, *i.e., if and only if the vector* $\mathbf{V}\mathbf{w}$ *belongs to* $\mathcal{C}(\mathbf{X})$.

Proof. Let $\mathbf{u}'\mathbf{y}$ be any known linear function different from $\mathbf{w}'\mathbf{y}$ but such that $\mathrm{E}(\mathbf{u}'\mathbf{y}) = \mathbf{w}'\mathbf{X}\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$. Then $\mathbf{u}'\mathbf{y} = \mathbf{w}'\mathbf{y} + (\mathbf{u}'\mathbf{y} - \mathbf{w}'\mathbf{y}) = \mathbf{w}'\mathbf{y} + \mathbf{f}'\mathbf{y} = (\mathbf{w} + \mathbf{f})'\mathbf{y}$, where $\mathbf{f} = \mathbf{u} - \mathbf{w}$ is such that $\mathbf{f}'\mathbf{X} = \mathbf{0}$, i.e., that $\mathbf{f} = (\mathbf{I} - \mathbf{P_X})\mathbf{a}$ for some nonzero vector $\mathbf{a}$ [which means that $\mathbf{f} \in \mathcal{C}^{\perp}(\mathbf{X})$, the orthogonal complement of $\mathcal{C}(\mathbf{X})$]. Hence, $\mathrm{Var}(\mathbf{u}'\mathbf{y}) = (\mathbf{w} + \mathbf{f})'\mathbf{V}(\mathbf{w} + \mathbf{f}) = \mathbf{w}'\mathbf{V}\mathbf{w} + 2\mathbf{f}'\mathbf{V}\mathbf{w} + \mathbf{f}'\mathbf{V}\mathbf{f}$. This shows that $\mathrm{Var}(\mathbf{u}'\mathbf{y}) < \mathbf{w}'\mathbf{V}\mathbf{w}$ if and only if $2\mathbf{f}'\mathbf{V}\mathbf{w} + \mathbf{f}'\mathbf{V}\mathbf{f} < 0$, and this inequality holds for any vector $\mathbf{f}$ such that either $\mathbf{f}'\mathbf{V}\mathbf{f} = 0$ and $\mathbf{f}'\mathbf{V}\mathbf{w} < 0$, or $\mathbf{f}'\mathbf{V}\mathbf{f} > 0$ and $\mathbf{f}'\mathbf{V}\mathbf{w}/\mathbf{f}'\mathbf{V}\mathbf{f} < -1/2$, unless $\mathbf{f}'\mathbf{V}\mathbf{w} = 0$. But the latter equality holds for any $\mathbf{f} \in \mathcal{C}^{\perp}(\mathbf{X})$ if and only if $(\mathbf{I} - \mathbf{P_X})\mathbf{V}\mathbf{w} = \mathbf{0}$. This proves the necessity of the condition. To see its sufficiency, note that if it holds, then $\mathrm{Var}(\mathbf{u}'\mathbf{y}) \geq \mathrm{Var}(\mathbf{w}'\mathbf{y})$ for any vector $\mathbf{f} = \mathbf{u} - \mathbf{w} \in \mathcal{C}^{\perp}(\mathbf{X})$, the equality holding if and only if (in addition) $\mathbf{f}'\mathbf{V}\mathbf{f} = 0$. However, the latter holds (together with $\mathbf{f}'\mathbf{V}\mathbf{w} = 0$) if and only if either $\mathbf{u} = \mathbf{w}$ or $\mathbf{u}'\mathbf{V}\mathbf{w}/\mathbf{w}'\mathbf{V}\mathbf{w} = \mathbf{u}'\mathbf{V}\mathbf{w}/\mathbf{u}'\mathbf{V}\mathbf{u} = 1$, i.e. the correlation between $\mathbf{u}'\mathbf{y}$ and $\mathbf{w}'\mathbf{y}$ is equal to 1, which means that these estimators are "almost surely" identical. □

This proof is a modified version of that given originally by Zyskind (1967, Theorem 3). Also note that the theorem follows from a known more general result (see Rao, 1973, Section 5a.2).

Acknowledgements

This work was supported by KBN Grant No. 2 1129 91 02. The author is grateful to Radosław Kala for his careful review of an earlier version of the paper.

References

Anscombe, F.J. (1948). Contribution to the discussion on D.G. Champernowne's "Sampling theory applied to autoregressive sequences". *J. Roy. Statist. Soc. Ser. B* **10**, 239.

Bailey, R.A. (1981). A unified approach to design of experiments. *J. Roy. Statist. Soc. Ser. A* **144**, 214-223.

Caliński, T. and Kageyama, S. (1991). On the randomization theory of intra-block and inter-block analysis. *Listy Biometryczne – Biometrical Letters* **28**, 97-122.

Caliński, T. and Kageyama, S. (1996). The randomization model for experiments in block designs and the recovery of inter-block information. *J. Statist. Plann. Inference* **52**, 359-374.

Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*, 2nd ed. John Wiley, New York.

Cox, D.R. (1958). *Planning of Experiments.* John Wiley, New York.

Finney, D.J. (1960). *An Introduction to the Theory of Experimental Design.* University of Chicago Press, Chicago.

Fisher, R.A. (1935). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Statist.* **4**, 384-395.

Kaiser, L.D. (1989). Nonexistence of BLUE's under a randomization model for the randomized block design. *J. Statist. Plann. Inference* **22**, 63-69.

Kala, R. (1991). Elementy teorii randomizacji. III. Randomizacja w doświadczeniach blokowych (Elements of the randomization theory. III. Randomization in block experiments). *Listy Biometryczne – Biometrical Letters* **28**, 3-23.

Kempthorne, O. (1952). *The Design and Analysis of Experiments.* John Wiley, New York.

Kempthorne, O. (1955). The randomization theory of experimental inference. *J. Amer. Statist. Assoc.* **50**, 946-967.

Nelder, J. A. (1954). The interpretation of negative components of variance. *Biometrika* **41**, 544-548.

Nelder, J.A. (1965a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. (London) A* **283**, 147-162.

Nelder, J.A. (1965b). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proc. Roy. Soc. (London) A* **283**, 163-178.

Neyman, J. (1923). Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych (Sur les applications de la théorie des probabilités aux expériences agricoles. Essay de principes). *Roczniki Nauk Rolniczych* **10**, 1-51. [A partial English translation by D.M. Dąbrowska and T.P. Speed (1990). *Statistical Science* **5**, 465-472.]

Neyman, J. (1935), with co-operation of K. Iwaszkiewicz and St. Kołodziejczyk. Statistical problems in agricultural experimentation (with discussion). *J. Roy. Statist. Soc. Suppl.* **2**, 107-180.

Rao, C.R. (1959). Expected values of mean squares in the analysis of incomplete block experiments and some comments based on them. *Sankhyā* **21**, 327-336.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. John Wiley, New York.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley, New York.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38**, 1092-1109.

Zyskind, G. (1975). Error structures, projections and conditional inverses in linear model theory. In: J.N. Srivastava, Ed., *A Survey of Statistical Design and Linear Models.* North-Holland, Amsterdam, 647-663.

Received 22 February 1996

O istnieniu najlepszych liniowych estymatorów nieobciążonych w modelu randomizacyjnym dla układu bloków losowanych

STRESZCZENIE

W pracy rozważa się na nowo sprawę istnienia najlepszych liniowych estymatorów nieobciążonych w modelu randomizacyjnym, w związku z zakwestionowaniem tego istnienia przez Kaisera (1989). Pokazuje się, jak dwa różne modele randomizacyjne implikują istnienie, przy pewnych warunkach, najlepszych estymatorów liniowych nieobciążonych dla zwykle interesujących liniowych funkcji parametrycznych. Zwraca się uwagę na rolę randomizacji podwójnej, wewnątrz i między blokami. W końcu, wskazuje się na przyczynę negatywnego wniosku, do którego doszedł Kaiser.

SŁOWA KLUCZOWE: najlepszy liniowy estymator nieobciążony, modele randomizacyjne, układ bloków losowanych.

Comments on T. Caliński's paper

Radosław Kala[1], *Department of Mathematical and Statistical Methods, Agriculture University of Poznań, Poland*

First I would like to express my thanks to Prof. Caliński for focusing our attention on the paper by Lee Kaiser (1989). This paper is really controversial. It was published in JSPI, a journal of high quality, and was recommended by Oscar Kempthorne, probably the first author who fully elaborated a model taking into account the randomization processes. On the other hand, this paper causes certain problems presented on today's session.

It is a pity that Lee Kaiser is not among us, since his point would be most interesting. Anyway, I have spent several hours studying the paper and searching for the seeds of the controversy. As a result I would like to present the following remarks.

Models. Adopting Kaiser's notation we can write the randomization model in the form

$$\mathbf{y} = \boldsymbol{\Delta}\boldsymbol{\tau} + \boldsymbol{\rho}_r\boldsymbol{\mu}, \tag{1}$$

where $\mathbf{y} = (y_{11}, ..., y_{1v}, ..., y_{b1}, ..., y_{bv})'$ is a vector of bv observations, $\boldsymbol{\Delta} = (\mathbf{1}_b \otimes \mathbf{I}_v)$, $\boldsymbol{\tau} = (\tau_1, ..., \tau_v)'$ is a vector of treatment effects, $\boldsymbol{\rho}_r$ is a permutation matrix corresponding to the randomization process within blocks, while $\boldsymbol{\mu} = (\mu_{11}, ..., \mu_{1v}, ..., \mu_{b1}, ..., \mu_{bv})'$ is a vector of unit effects. The usual linear model corresponding to a block design can be expressed as

$$\mathbf{y} = \boldsymbol{\Delta}\boldsymbol{\tau} + \mathbf{D}\boldsymbol{\beta} + \mathbf{e}, \tag{2}$$

where $\boldsymbol{\Delta}$ and $\boldsymbol{\tau}$ are as in (1), $\mathbf{D} = (\mathbf{I}_b \otimes \mathbf{1}_v)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_b)'$ is a vector of block effects, while $\mathbf{e}$ is a vector of random errors.

Comparing the models (1) and (2) Kaiser claims (p.64, bottom) that a fundamental difference between these two models is that in the randomization model *the data includes* $\boldsymbol{\rho}_r$. What it means, is difficult to guess.

Let us assume that $\boldsymbol{\rho}_r$ is a fixed matrix reflecting the result of randomization performed in the experiment. This interpretation corresponds to Kaiser's statement on p. 67: *In the usual linear model the data is the vector of responses while in the randomization model the data also includes which treatment was applied to each unit.* In such a case the randomization model is highly overparameterized. It contains v parameters for treatment effects and $N = bv$ parameters for the unit effects, where b is the number of blocks. Since we have only N responses, the estimation subspace coincides with the space of observations, and hence the model has no statistical sense.

So, let us assume that $\boldsymbol{\rho}_r$ is a random matrix reflecting the randomization process. In consequence, the expectation of $\mathbf{y}$ in both models has exactly the same

[1] Presented after the lecture at the XXXVII Session of the Polish Biometric Society, Poznań, February 2-3, 1996

representation. It means that not only the estimation subspaces in both models are the same, but also the bases for these subspaces are the same. The difference between models (1) and (2) is shifted to the dispersion matrix of $\mathbf{y}$, which is a point discussed by Caliński in details. Note, that in this case the phrase: *the data includes* $\boldsymbol{\rho}_r$, means only that during the course of the experiment some randomization was performed.

It should be observed that even if $\boldsymbol{\rho}_r$ is a random matrix then the model (1) has more parameters than the model (2). In the former model we have $v+b$ fixed effects and a separate variance component for each block, while in the latter model we have $v+b$ fixed effects and only one error variance. Of course, the larger number of unknown variance components causes problems in estimation. They have been precisely shown by Caliński. Note, however, that the process of randomization has reduced the set of parameters from the initial number $v+N(N=bv)$ to $v+2b$. In my opinion this reduction of parameters is the main profit resulting from the randomization. That this operation is useful, has been also demonstrated today, particularly when showing that the randomization of blocks, in addition to randomization within blocks, reduces the number of variance components from b to 2. In result, more functions of fixed parameters admit the BLUE.

Estimation procedure. The inclusion of the matrix $\boldsymbol{\rho}_r$ to the data set, proposed by Kaiser, results in a new class of estimators. They are linear in $\mathbf{y}$, but the coefficients of observations can be dependent on $\boldsymbol{\rho}_r$. If $\boldsymbol{\rho}_r$ is fixed, the estimator developed by Kaiser in his Theorem 1 is operational, in the sense that it is possible to preselect the vectors $\boldsymbol{\tau}^0$ and $\boldsymbol{\mu}^0$ such that $\mathbf{y}=\boldsymbol{\Delta}\boldsymbol{\tau}^0+\boldsymbol{\rho}_r\boldsymbol{\mu}^0$. But it is not unbiased. If $\boldsymbol{\rho}_r$ is a random matrix then Kaiser's estimator is unbiased, but not operational, as is admitted by the author in his remark after Theorem 1.

Note, moreover, that if we allow to extend the class of estimators by adding to the usual class the nonoperational procedures as well as functions with random coefficients, then we can formulate more general theorem.

Let $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}$ *be a simple linear model, let* δ *be any random variable, independent on* $\mathbf{y}$, *and with zero expectation, and let* $\mathbf{A}$ *be an arbitrary matrix. Then the estimator*

$$\bar{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}+\delta\mathbf{A}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^0) \tag{3}$$

is (i) *linear in* $\mathbf{y}$, (ii) *unbiased for* $\boldsymbol{\beta}$, *and* (iii) $\bar{\boldsymbol{\beta}}=\boldsymbol{\beta}^0$ *if* $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}^0$.

The estimation procedure following from (3) belongs to the class of randomized decisions, and formally is correct. But it is useless. The probability of the event that $\mathbf{y}$ belongs to the estimation space, i.e. that $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}^0$, is zero. Thus we can say that the estimator (3) can improve the result of estimation in linear models only if the experimenter co-operates with a perfect prophet.

Randomization and finite population inference. Kaiser has emphasized, in his Section 5, that the randomization model is similar to a finite population sampling model, since in both models there are more data than the vector of responses and there are more parameters than responses. However, if we look at the paper by Godambe (1955), referred to by Kaiser, then we see much more dissimilarities than similarities. To be precise, recall from Godambe (1955) the main problem of sampling from a finite population.

Let $Y = \{y_\lambda, \lambda = 1, ..., N\}$ be a set of N individuals characterized by the value of some variable y. The experimenter is interested in estimating the total $T = y_1 + ... + y_N$ by making n successive random drawings from Y. Observe that if $n = N$, i.e. the sampling is exhaustive, the problem does not exist. It is the first dissimilarity, since the randomization process adopted by Kaiser takes into account all units. The second difference is contained in the formulation of the problem. In the randomization model the knowledge of individual unit effects is more interesting, than the knowledge of their sum T. The third dissimilarity follows from the method used. When sampling from a finite population, the values $y_1, ..., y_N$ are considered as fixed numbers and the experimenter is searching for the sampling design leading to a sample s_n of n elements ($n << N$) that is most convenient for inference about T. The permissible designs cover any scheme of sampling: stratified and unstratified, with and without replacement, independent and dependent. In result, the estimator, linear in s_n, in its most general form has the coefficients determined by the sampling scheme. So, in the finite population sampling model the estimator can have random coefficients in a natural way. It should be stressed that such estimators are operational provided being not dependent on any preselected value of unknowns.

Thus there are more dissimilarities than similarities between randomization models and finite population sampling models. It does not mean, however, that there is no common base. This is explained in an approach to modelling the experiments presented in a sequence of papers by Kala (1989, 1990, 1991).

References

Godambe (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. Ser. B* **17**, 269-278.

Kala, R. (1989). Elementy teorii randomizacji. I. Próba zrandomizowana. *Listy Biometryczne - Biometrical Letters* **26**, 41-55.

Kala, R. (1990). Elementy teorii randomizacji. II. Modelowanie doświadczeń prostych. *Listy Biometryczne - Biometrical Letters* **27**, 31-45.

Kala, R. (1991). Elementy teorii randomizacji. III. Randomizacja w doświadczeniach blokowych. *Listy Biometryczne - Biometrical Letters* **28**, 3-23.

Sanpei Kageyama, *Department of Mathematics, Faculty of School Education, Hiroshima University, Japan*

The basic principles of experimental design, "replication", "randomization" and "local control", were formulated by Fisher (1925, 1926). Beside the replication and the local control, the need for randomization was recognized as a necessary condition for obtaining a valid estimate of error and, consequently, for a valid use of the test of significance provided by the analysis of variance.

Block designs were originally used in agricultural experiments with the aim of allowing all treatments to be compared within similar conditions. A block is a compact set of experimental units possibly uniform in their conditions and in a number equal to the number of treatments and, since the treatments are assigned to units within such blocks at random, the resulting experiment is said to be designed in randomized blocks, called the *randomized block design.* This is reserved solely for the classical design composed of complete blocks, i.e. such which allows all the treatments to be allocated in each block. Recall that the adjective "randomized" is used only for naming the classical design of randomized blocks. When describing the randomized block design, most of the books on design of experiment point out the necessity of a random assignment of treatments to units within blocks. However, in formulating mathematical models for the analysis of experimental results obtained under the randomized block design, the role of randomization has not been understood in the same way by various research theorists and practitioners.

Glance at some history. On the statistical analysis of results obtainable from experiments conducted in block designs vast literature is available, starting from the early works of Yates (1936). Most of the literature makes references to the so-called intra-block analysis only. This analysis, however, does not provide full information on treatment differences, unless the within-block homogeneity is completely achieved by a successful choice of the experimental units and their block structure. Such ideal situation is seldom met in practice and, therefore, a more realistic model is to be considered as a basis for the statistical analysis of experimental data. Thus the usual linear additivity model does not appreciably take into account the randomization implemented in designing and carrying out the experiment. In deriving an appropriate model, the randomization procedures involved in laying out the experiment should definitely be taken into account, particularly as the main purpose of the randomization is to allow the experimental data to be considered as observations on random variables of certain homogeneous distributions. Several authors have tried to build models that fully recognize the randomization techniques employed. Among the various approaches of particular interest in this paper are two lines of development, along which the randomization is incorporated into the model building. One originates from Neyman (1923, 1935) and was extended by Kempthorne (1952, 1955). Most of the references in this line can be found in White (1975) and Kempthorne (1977), here called the intra-block randomization. Another was initiated by Nelder (1954), under some influence of Anscombe (1948), here called the intra-block and inter-block randomization. The essential references in this line have been given by Bailey (1981) and by Bailey and Rowley (1987). For the randomized block design case, an essential feature of this approach is that not only the intra-block but also the inter-block randomization is taken into account in formulating the model.

With regard to block designs, the randomization model leads to an analytical procedure which takes into account also the information on treatment differences that are partially or totally confounded with block differences. After Yates (1939, 1940), this procedure is called the recovery on inter-block information. It is presented here for a randomized block design in Section 3 with the simple combinability of the intra-block and inter-block information, the earlier Section 2 being related to the intra-block analysis, according to the commonly adopted attitude today towards the theory of block designs. The main attraction of this paper is that the results presented in Sections 2 and 3 are in contradiction to the conclusion of Kaiser (1989) regarding the existence of the BLUEs under a randomization model for the randomized block design. This is discussed in Section 4 under the above mentioned two different randomization models by showing the existence of the BLUEs of the usual linear parametric functions of interest, under certain conditions, along with the clear understanding of the role of the double randomization within and between blocks.

In Section 2, to compare with the Kaiser paper, the usual unit-treatment additivity and the intra-block randomization are assumed. Here the treatments are assigned to units at random in each block, independently from block to block. It is shown that the parametric function of the overall true response averaged over all treatments has the BLUE, and that a parametric function through a treatment contrast has the BLUE if and only if the unit error variances are constant for all the blocks. The last condition is known due to Zyskind (1975).

In Section 3, in addition to the assumptions in Section 2, it is further assumed that not only the units are randomized within the blocks but also the blocks are independently randomized between themselves. It is shown that the average true response to any treatment is best linearly and unbiasedly estimated by the corresponding observed mean response, taken over the units (in the number of blocks) to which the treatment has been applied according to the randomized block design, under the double randomization. The estimation results here coincide with the more general results given by Kala (1991) and by Caliński and Kageyama (1991, 1996).

Section 4 is a highlight of this paper on the existence of BLUEs under the setup above. What is the validity of the definition of a class of *linear* estimators in the vector of responses? This is understood differently between Kaiser (1989) and the present author. As far as the notion of BLUE is explained in the sense used in the general Gauss-Markov theorem, I agree with the claim by the author who points out the contradiction to the conclusion of Kaiser (1989). The statements described in the last paragraph of the paper are also instructive.

It is an interesting paper. This type of contrary is often hard to be understood by some theoretical researchers who have different philosophy and other understanding about the randomization and the model building based on it. In this sense it would be the best if the discussion in this paper could be accompanied with some illustrations.

References

Bailey, R.A. and Rowley, C.A. (1987). Valid randomization. *Proc. Roy. Soc. London Ser. A* **410**, 105-124.

Fisher, R.A. (1925). *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

Fisher, R.A. (1926). The arrangement of field experiments. *J. Ministry Agriculture* **33**, 503-513.

Kempthorne, O. (1977). Why randomize? *J. Statist. Plann. Inference* **1**, 1-25.

White, R.F. (1975). Randomization in the analysis of variance. *Biometrics* **31**, 555-571.

Yates, F. (1939). The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugen.* **9**, 136-156.

Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen.* **10**, 317-325.

John A. Nelder, *Department of Mathematics, Imperial College, London, UK*

What Caliński's argument shows is that, for the randomized-block design, the randomization argument I gave in my 1965 papers establishes an isomorphism between the standard Gauss-Markov models based on sums of independent random effects and those derived from the symmetric covariance structures induced by randomization. The only difference is that in the latter it may be necessary to allow negative variance components. The fact that it is necessary to randomize the blocks becomes obvious once incomplete-block designs have to be modelled, but this is hidden in the particular case of complete randomized blocks. The whole point of the randomization argument is to justify the use of the symmetric likelihood implied by the standard model. It is, of course, possible that the data will contradict this assumption, by showing, for example, that the blocks are not exchangeable, or that there is a consistent trend within blocks that should be removed and the treatment effects adjusted accordingly. Alternatively one may seek a model for the plot effects based on spatial correlation patterns; there is now a huge literature on this.

To me, Kaiser's paper illustrates an attitude towards statistics that I find very unsatisfactory. It is based purely on deductive arguments and seems to lack any contact with the important inferential problems that exist in analysing data from the real world. These inferential problems are not easy and cannot be solved by purely deductive mathematical arguments. They are, however, the problems that statisticians ought to be tackling. I hope that Caliński's paper will show that we need not be concerned by Kaiser's results; there are more important things to do.

Rosemary A. Bailey, *School of Mathematical Sciences, Queen Mary and Westfield College, London, UK*

I am grateful for the opportunity to comment on Caliński's rebuttal of Kaiser's paradox. I am afraid that I disagree with both of them about whether the statistical analyst should know which treatment was applied to which plot. Caliński seems to say that we should always ignore this information; Kaiser says that we know it only if the experiment has been randomized. I believe that we should always know it. When I design an experiment in blocks (whether complete or incomplete, whether randomized or not) I prepare a data sheet with four columns. The first column contains the block number (i); the second contains the number (k) of the plot in that block; the third contains the code (j) for the treatment applied to that plot, which is the unique j such that $\delta_{ik}^{j} = 1$; and the fourth contains the response.

If the second column is missing when the data arrive, I ask the experimenter why. It may be that the second column should simply consist of $1, \ldots, v$ in order b times, but I need to verify this, so that I can scan the data for spatial patterns caused by, say, a patch of disease in the crop or rain starting part-way through the harvest. If the second column is missing and the third column has the same order in every block then I am suspicious. This may mean that the experiment was not randomized (this affects the conclusions that I draw from the analysis). On the other hand, it may be that the experiment was randomized but the experimenter has written the data in treatment order in each block in the mistaken belief that this will help me. Now there are three potential problems: perhaps the experimenter harvested the plots in treatment order, not in field order, thus making the experiment systematic and destroying the blocking system; perhaps the experimenter has done everything in the correct order but made mistakes copying the data to a different order on the data sheet; and any information on spatial patterns has been lost.

In short, in practical terms it is nonsense to expect the statistician to be ignorant of which plot the datum came from. For the mathematical theory, it is impossible to go beyond very simple structures (consider a Latin square, for example) while using a hybrid notation like y_{ij}, where i is part of the plot information and j is the treatment information. The response we are discussing is the one on plot k in block i: it should be called y_{ik}. In my own development of randomization theory (Bailey, 1981, 1991) I always label the responses by the plots. The root of Kaiser's paradox is his inappropriate labelling of the responses.

A second problem with Kaiser's approach to randomization, which he inherits from Kempthorne and which Caliński mimics, is the assumption that the μ_{ik} are constants and so $\sum_i \sum_j y_{ij}$ is constant for any given set of treatments. Again, this is nonsense in practical terms. The randomness in the assumed response is our modelling of the non-repeatability of an experiment. For example, suppose that the whole of block 1 is low-lying, on clay soil, while block 7 is higher and better drained. In a wet year μ_{11} and μ_{12} will probably be lower that μ_{71} if the response is yield; in a dry year the opposite will happen. If we regard the μ_{ik} as random variables then we can capture such variation by specifying that μ_{11} and μ_{12} are more highly correlated with each other than either is with μ_{17}.

Even if we accept all Kaiser's assumptions, there is still a flaw in his argument. His estimator $\tilde{\phi}$ of τ_1 is

$$b^{-1}\sum_i y_{i1} - d\sum_i\sum_j y_{ij}/e,$$

where $d = b^{-1}\sum_i \mu_{i1}$ and $e = b\sum_j \tau_j + \sum_i\sum_j \mu_{ij}$. He claims that d and e are both constants. Caliński objects because the constant d involves the random variable δ^j_{ik} : he says that therefore $\tilde{\phi}$ cannot be called linear. This objection vanishes once the responses are labelled by plots.

My objection is that $\tilde{\phi}$ is either not unbiased or not linear. What happens when we replace treatment 1 by a new treatment whose effect is τ_1^*? If the value of e in $\tilde{\phi}$ is not changed then $\tilde{\phi}$ is biased, because τ_1 contributes to the value of e. On the other hand, if e is changed to $b(\tau_1^* + \sum_{j\geq 2}\tau_j) + \sum_i\sum_j \mu_{ij}$ then $\tilde{\phi}$ is not linear in the responses. In fact, in this case $\tilde{\phi}$ in not even an estimator at all, because it is no longer a function just of the responses and other known values.

It seems to me that Kaiser's paradox has nothing whatever to do with randomization. It arises from forgetting that an estimator must not depend on the quantity being estimated, and so it should be ignored.

However, Caliński's exposition here certainly does contribute to our understanding of randomization. It is a matter of taste whether one calculates with Kempthorne's indicator random-variables or with the orbits of the permutation group on pairs of plots (cf. Bardin and Azaïs, 1990; Bailey, 1991): both approaches give the covariance matrices in (2.3) and (3.3). In Section 3 Zyskind's criterion is equivalent to $\mathbf{w}$ belonging to the treatment space, no matter what the values of σ_B^2 and σ_U^2 are; it is not necessary to assume that the criterion holds for all values of σ_B^2 and σ_U^2. Caliński's conclusion that simple averages are BLUEs not only of treatment contrasts but also of the overall mean is a special case of the important result of Zyskind (1967) and Kruskal (1968) that all ordinary-least-squares estimators are BLUEs if $\mathbf{P_X}$ commutes with $\mathbf{V}$ for all values of the unknown parameters in $\mathbf{V}$. Designs satisfying this condition are precisely those which I (and many others) call *orthogonal.*

I disagree slightly with Caliński in Section 2. It is well known that if two unbiased estimators of the same quantity have different variances then the best linear combination of them is weighted inversely to those variances. So I do not think that $b^{-1}(\mathbf{1}_b' \otimes \mathbf{c}')\mathbf{y}$ can be a candidate for a BLUE. Suppose that $\mathbf{w}'\mathbf{y}$ is a BLUE, and put $z_{ij} = \sigma_{U,i}^2 w_{ij}$. Then Zyskind's criterion shows that $\mathbf{z}$ is the sum of a blocks vector and a treatments vector, so that there are constants g_i and h_j such that $w_{ij} = g_i + h_j/\sigma^2_{U,i}$. Taking $g_i = (bv)^{-1}$ and $h_j = 0$ gives the BLUE of $\tau_. + \mu_{..}$; taking $\sum_i g_i = 0$ and $h_j = 0$ gives the BLUE of the block contrast $v \sum_i g_i \mu_{i.}$; while taking $g_i = 0$ and $\sum_j h_j = 0$ gives the BLUE of $\left(\sum_i (\sigma^2_{U,i})^{-1}\right)\left(\sum_j h_j \tau_j\right)$. Thus, in general, BLUEs exist only if the ratios $\sigma^2_{U,1} :: \ldots :: \sigma^2_{U,b}$ are known. This is analogous to the result for the existence of BLUEs in non-orthogonal designed experiments with two or more strata (Houtman and Speed, 1983).

The clear discussion in Section 2 and 3 exposes much of the common inconsistent thinking about randomization and block designs. If there is no inter-block randomization then there are no BLUEs of treatment contrasts and there are no estimators of the variances of the BLUEs of block contrasts. Why, then, do so few people perform inter-block randomization ? If the blocks are complete then such randomization does not enlarge the class of possible layouts for experiments so there is no practical need for such randomization even though the theory assumes it. An unfortunate side effect is that many experimenters think in terms of randomization of treatments, not randomization of plots: when faced with a resolvable incomplete-block design they may either ruin the design by randomizing treatments separately in each replicate or omit the necessary inter-block randomization.

If inter-block randomization is performed then Section 3 shows that there are BLUEs of treatment effects. It also shows unequivocally that blocks are random effects. Why then do so many people analyse data from randomized complete-block designs as if the block effects were fixed and the intra-block variances constant ?

References

Bailey, R.A. (1991). Strata for randomized experiments. *J. Roy. Statist. Soc. Ser. B* **53**, 27-28.

Bardin, A. and Azaïs, J.-M. (1990). Une hypothèse minimale pour une théorie des plans d'expériences randomisés. *Rev. Statist. Appl.* **38**, 5-20.

Houtman A.M. and Speed, T.P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.* **11**, 1069-1085.

Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Statist.* **39**, 70-75.

Lee Kaiser, *Statistics Department, Pharmaceutical Division, Bayer Corporation, West Haven, CT, USA*

I am grateful for the opportunity to comment on the paper by T. Caliński. I will use the notation in my JSPI paper. Caliński and I have shown the following results for the estimation of a contrast $\sum c_j \tau_j$:

1. $\sum c_j \bar{y}_j$ is minimum variance unbiased for $\sum c_j \tau_j$ in the class of estimators $\mathbf{a}'\mathbf{y}$ when variability among units within blocks is constant across blocks.

2. $\sum c_j \bar{y}_j$ is minimum variance unbiased for $\sum c_j \tau_j$ in the class of estimators $\mathbf{a}'\mathbf{y}$, when block labels are randomized and subsequently ignored.

3. $\sum c_j \bar{y}_j$ is not minimum variance unbiased for $\sum c_j \tau_j$ in the class of estimators $\mathbf{a}_r'\mathbf{y}$, i.e., where the coefficient of a response value can depend on the randomization.

4. No minimum variance unbiased estimator exists for $\sum c_j \tau_j$.

5. $\sum c_j \bar{y}_j$ is admissible as an estimator of $\sum c_j \tau_j$ in the class of estimators $\mathbf{a}_r'\mathbf{y}$.

I do not dispute items 1 and 2 and T. Caliński does not dispute items 3, 4 and 5. The sole question is whether an estimator of the form $\mathbf{a}_r'\mathbf{y}$ is "linear". It is clearly linear in $\mathbf{y}$ as I pointed out in my paper. I leave it to the reader to make his own decision. If he decides that $\mathbf{a}_r'\mathbf{y}$ is a linear estimator, then item 3 implies that $\sum c_j \bar{y}_j$ is not a BLUE.

Rejoinder

Tadeusz Caliński

I would like to thank all who took part in the discussion for their interesting and helpful comments.

In Prof. Kala's comments the important problem of including the permutation matrix $\boldsymbol{\rho}_r$ to the data set has been clarified. His theorem concerning the estimator $\bar{\boldsymbol{\beta}}$, in formula (3), clearly reveals the consequences of allowing the class of unbiased linear estimators to be extended by including functions of observations with random coefficients, as proposed by Kaiser (1989). I am also grateful to Radosław Kala for commenting on the inferential differences between randomization models and finite population sampling models, a point to which I have not paid enough attention in my paper.

Prof. Kageyama's extensive comments throw some light on the historical background of the problem of randomization in the design and analysis of experiments in randomized blocks. His comments apply not only to the complete block design, considered in the present paper, but also to the general block design, where the recovery of inter-block information is an essential issue. The practical relevance of the randomization model in this general case is discussed more thoroughly in a joint paper, just published (Caliński and Kageyama, 1996, Section 2.2). The problem of the existence of BLUEs under the randomization model for the general block design has already been considered in an earlier joint paper (Caliński and Kageyama, 1991, Section 2.2). The general existence conditions obtained there coincide with those given by Kala (1991, Theorem 3). The conditions considered in the present paper are special cases of those. In all these papers the same approach is adopted, in which the relevant permutation matrix is taken into account in deriving the model. Unlike in Kaiser's (1989) approach, however, linear unbiased estimators are understood there in the sense used in the general Gauss-Markov theorem.

I am very glad that Prof. Nelder has been so kind to comment on my present paper. Much of the work mentioned above has been stimulated by his early papers (Nelder, 1954, 1965a,b) and, therefore, his opinion is of great value for this research. I fully agree with his comments, and particularly I am pleased with his statement that the necessity of randomizing the blocks so obvious for incomplete-block designs *is hidden in the particular case of complete randomized blocks.* In practice this causes sometimes misunderstandings at the stage of laying out the experiment. I hope that the two figures included in the present paper make the point clear. Certainly I agree that there are also many other important problems related to block designs that call for considerations.

Similarly, I appreciate very much the comments made by Prof. Rosemary Bailey, which I regard not only as extremely interesting from the theoretical and practical point of view, but also as very helpful in clarifying the main points of the controversy. Again, as in the case of John Nelder's comments, I am particularly grateful for her comments upon the necessity of *inter-block randomization.* I wish I could say that nobody analyses the experimental data according to the unrealistic assumptions pointed out in her last interrogative sentence.

As to Prof. Bailey's disagreements with some extracts from my text, it seems that they result mainly from misunderstanding of what I wanted to say. I must then apologize for being not sufficiently clear in my writing. Let me, therefore, explain my position. First of all, there was not my intention to suggest that *we should always ignore* the *information* on *which treatment was applied to which plot.* When writing (at the end of Section 4) that *the knowledge of these facts is to be included into the model* I did not want to suggest that the information on the assignment of treatments to units is to be forgotten at the stage of the analysis. I could hardly imagine an educated experimenter who would not record the experimental data in the manner described by Rosemary Bailey in her comments, whether the analysis had to be based on a randomization or any other model. Secondly, it seemed to me justified to regard in the considered model the quantities μ_{ik} as constants, at least so long as one wants to infer from that particular single experiment. Such assumption on μ_{ik} has been made not only by Kaiser (and consequently in my discussion with him) but also by Nelder (1954, p. 544) and by Bailey (1981, p. 215). Certainly, if the experiment were to be analysed within a series of experiments repeated in space (an experimental area) or/and in time (seasons), then μ_{ik} would have to be regarded as a random variable. Then, however (as rightly noticed by Bailey, 1981), a more general theory would emerge, here unnecessarily complicating the discussion. Thirdly, in my reference to the properties of Kaiser's estimator $\tilde{\phi}$, i.e. estimator (4.1) in my paper, I took for granted that $\boldsymbol{\mu} = \boldsymbol{\mu}^0$ and $\boldsymbol{\tau} = \boldsymbol{\tau}^0$, and I considered, following his approach, the permutation matrix $\boldsymbol{\rho}_r$ as random. Certainly, if $\boldsymbol{\rho}_r$ were to be considered as a realization of that matrix, my objection to call $\tilde{\phi}$ linear would vanish. But then the estimator would not remain unbiased. Finally, when considering $b^{-1}(\mathbf{1}_b' \oplus \mathbf{c}')\mathbf{y}$ as *a candidate for a BLUE*, I was referring to *the usual assumed linear model.* Certainly, as shown, it is not a good candidate under the randomization model (2.1), unless the variances (2.4) are constant over all blocks. I fully agree that, in general, we would have a right candidate for a BLUE if we knew the ratios among those variances, but such knowledge is usually not available in practice.

The contribution to the discussion made by Dr Kaiser is of particular value. He rightly draws attention to five statements related to the problem under discussion, and then reduces the controversy to the question *whether an estimator of the form* $\mathbf{a}_r'\mathbf{y}$ *is "linear"*, if the coefficient vector $\mathbf{a}_r$ can depend on the randomization. He concludes that if we agree that this estimator is linear, then his statement in item 3 implies that the estimator considered in the discussed paper is not a BLUE. My answer is as follows:

(1) The function $\mathbf{a}_r'\mathbf{y}$ is formally linear, but in a broader sense than it is understood in the estimation theory based on the general Gauss-Markov theorem in which the term BLUE is used.

(2) There is no justification for regarding the estimators considered in the present paper as being of the form $\mathbf{a}_r'\mathbf{y}$, with $\mathbf{a}_r$ depending on the randomization. This has been pointed out at the end of Section 4 of my paper, and now is fully clarified by the exhaustive comments made by Rosemary Bailey and Radosław Kala.