SHOULD BLOCKS BE FIXED OR RANDOM?

Philip Dixon

Iowa State University, USA, pdixon@iastate.edu

Recommended Citation

Should blocks be fixed or random?

Philip M. Dixon
Department of Statistics,
Iowa State University,
Ames IA, 50011

Abstract:
Many studies include some form of blocking in the study design. Block effects are rarely of intrinsic interest; instead they are included in a model so that that model reflects restrictions on randomization that accompany the study design. I consider the question of how these block effects should be modeled: as fixed effects or as random effects. I discuss the consequences of the choice, including the recovery of inter-block information when available, give a simple example to illustrate the connection between recovery of inter-block information and pooling two estimators of a treatment effect, and give an example where fitting a model with random block effects can lead to the wrong answer. I suggest that block effects should be modeled as fixed effects unless there are compelling reasons to do otherwise.

1 Introduction

The randomized complete block design is probably the most commonly used experimental design in agronomy and related fields (Cochran and Cox, 1957, p. 107). Randomly assigning treatments within groups of otherwise similar units provides a more precise estimate of treatment differences while still providing valid estimates of the error variance. Blocking has a long history. "Local control", i.e., blocking, is one of R.A. Fisher's three principles of experimental design (attributed to Fisher (1925) by Yates 1965). However, field experimenters understood the value of blocking before Fisher (Yates 1965).

Although there are many different sorts of block designs, the most common is to have one and only one instance of each treatment within each block. So, for example, a study comparing 5 treatments would have blocks of 5 experimental units. The standard model to analyze data from such an experiment is:

$$Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}, \tag{1}$$

where μ is a quantity shared by all observations (often the overall average), β_i is an effect specific to block i, τ_j is an effect specific to treatment j, and ϵ_{ij} is an effect specific to each experimental unit (i.e., combination of block and treatment). The mean difference between treatments k and l is $\tau_k - \tau_l$, which is estimated by $\sum_i Y_{ik}/b - \sum_i Y_{il}/b$, where b is the number of blocks. The reasons for blocking is more clearly shown by the equivalent expression for the estimator: $\sum_i (Y_{ik} - Y_{il})/b$. The estimated difference in treatment means is the average of the within-block differences. Because the experimental units within each block are intended

to be more similar to each other, the variability of the within-block differences is expected to be smaller than the variability between two units randomly chosen from all units in the population.

Much of the subsequent discussion of block designs focuses on this commonly-used version, which I will call the standard RCBD. Many, but not all, of the principles illustrated using the standard RCBD apply to other variations of block designs. I will clearly indicate when I am discussing a variation from the standard RCBD.

Model (1) can be augmented by adding statements about one or more of the effects. For example, if the values of β_i are considered to be random variables drawn from a distribution, (1) can be augmented to specify what is known or assumed about those random variables. If all that is known is that the β_i are independent, identically distributed (iid) with a mean of 0 and a variance of σ_b^2, (1) becomes the model:

$$\begin{aligned} Y_{ij} &= \mu + \beta_i + \tau_j + \epsilon_{ij}, \\ \beta_i &\sim \text{iid}(0, \sigma_b^2). \end{aligned} \tag{2}$$

The random variables could be given a specific distribution, e.g., normal, which gives the model:

$$\begin{aligned} Y_{ij} &= \mu + \beta_i + \tau_j + \epsilon_{ij}, \\ \beta_i &\sim \text{iid N}(0, \sigma_b^2). \end{aligned} \tag{3}$$

In models (2) and (3), the block term is called a **random effect**, because values of β_i are modeled as values of a random variable with some specified properties. When that random specification is missing, as in (1), the block term is called a **fixed effect**.

A related but not identical way to define fixed and random effects focuses on "the manner in which they [the set of effects] were gathered and the environment from which they came ... Thus when inferences will be made about a distribution of effects from which those in the data are considered to be a random sample, the effects are considered as random; and when inferences are going to be confined to the effects in the model, the effects are considered fixed." (McCulloch, Searle, and Newhaus, 2008, p. 17). Similar ideas are given by Searle (1971, p. 383), Kreft and de Leeuw (1998, p. 11) and Toutenberg (2002, p. 111-112). This collection of definitions is closely connected to the distinction between Model I and Model II ANOVA in Eisenhart (1947).

The concept of an inference space (McLean, Sanders and Stroup 1991) provides additional insight into the definition of fixed and random effects. McLean, Sanders and Stroup define a narrow inference space as "inference to specific levels of random effects" and a broad inference space as "inference to the entire population of random effects." The parallel to McCulloch et al.'s definitions of fixed and random effects implies that one could consider all effects as random effects, then decide the appropriate inference space. Gelman (2005) explores the consequences of treating all effects as random effects. Although narrow-sense inference can

be obtained by carefully specifying coefficient matrices, it is identical to treating effects as fixed effects, at least for balanced data (McLean, Sanders and Stroup. 1991). Because it is easier to specify an effect as fixed or random than to specify the appropriate coefficients for narrow-sense inference, we will phrase the choice as fixed or random.

When the study goal is to estimate differences among treatments, the treatment effects, τ_j, are modeled as fixed effects. The experimental-unit-specific effects, ϵ_{ij}, are universally considered as random effects. It is less clear how the block effects, the β_i, should be modeled. Should the block effects be considered fixed or random? This paper provides perspectives on this question. Specifically, I discuss some of the opinions on this choice, review the consequences of this choice, illustrate recovery of between-block information, and highlight a potential concern with computing software. I end with my evaluation of this evidence, which is offered as a starting point for discussion.

2   Opinions on blocks as fixed or random effects

Commonly, blocks are considered to be random effects. Two textbook quotes are:

> "We take the pragmatic view that blocks consist of collections of experimental units. The object of the experiment is to make inferences, predictions of what will happen if the treatments are applied to new units from the population. In this sense the experimenter wants blocks to be random." (Giesbrecht and Gumpertz, 2004, p. 86)

> "By definition, blocks are always a random factor since not all levels of interest can be in the experiment." (Casella and Berger, 1990, p. 530)

In contrast, Paul Allison argues that blocks should always be considered as fixed effects (e.g., Allison 2009). His discussion and argument focus on observational studies, where block effects are not orthogonal to treatment effects, so the block means can provide information about differences in treatment means. I discuss Allison's view in section 6.3.

The diversity of opinion about the appropriate treatment of block effects is not restricted to the United States. Members of the German Region of the International Biometrics Society discussed the issue in 2006. There were proponents of both views (Moll, Piepho and Richter 2006; Piepho 2006) and no consensus.

Andrew Gelman (2005, section 6) identifies five published criteria to decide whether an effect is fixed or random. These are:

- Variation in the effect across subjects (Kreft and de Leeuw, 1998, section 1.3.3). Fixed effects are the same in all subjects; random effects vary.

- Focus of the study (various texts and papers by Searle, including McCulloch, Searle, and Neuhaus, 2008, section 1.6). Effects are fixed when the study focuses on those individual effects; effects are random when the study focuses on characteristics of an underlying population.

- Fraction of the population included in the study (Green and Tukey 1960). When the entire population of interest is included in the study, the effect is fixed; when a small fraction of the population is included, the effect is random.

- Is the effect a realization of a random variable (LaMotte 1983)? If it is, the effect is random; if not, the effect is fixed.

- How effect values are estimated. Values of fixed effects are estimated by least squares; Values of random effects are estimated with shrinkage (i.e., best linear unbiased predictors).

In my talk at the 2016 meeting, I surveyed opinions on fixed or random block effects. Participants who indicated a connection to ISU were excluded from the tabulation of results. I briefly described four designs and asked "Should block effects be considered fixed or random?". Participants were asked to choose from three possible answers: Fixed, Random, or Depends. If they chose Depends, they were further asked "On what?". The four designs were:

Study 1: Standard randomized complete block design, 10 blocks, 5 treatments , 50 eu's
Study 2: Standard randomized complete block design, 3 blocks, 5 treatments, 15 eu's
Study 3: resolvable incomplete block design (e.g. alpha lattice),
    36 incomplete blocks of size 4 forming 3 complete blocks, 48 treatments, 144 eu's
    Participants were asked about both the complete blocks and the incomplete blocks.
Study 4: quantitative treatments in an RCBD, 10 blocks, 5 levels, 50 eu's.

For all questions, the most common answer was Random (Table 1). More participants chose Fixed when the number of blocks was small (Study 2 with 3 blocks, compared to Study 1 with 10 blocks). Other aspects of the study had little effect on the choices. Relatively few participants provided additional information about their "Depends" answers. Among those who did, there were various answers, including "context, interest", "interaction", and "desire to make inference on blocks".

Clearly, both in the literature and among the conference participants, there are multiple opinions on how block effects should be modeled.

Study	Number of times chosen:		
	Fixed	Random	Depends
1: RCBD, 10 blocks	5	30	14
2: RCBD, 3 blocks	13	20	16
3: lattice design			
3a: complete blocks	8	24	13
3b: incomplete blocks	5	30	11
4: RCBD, quantitative	5	28	14

Table 1: Number of participants answering Fixed, Random, or Depends to questions about how block effects should be modeled. Some questions were not answered by everyone, so the total number of respondents varies by question.

3   Consequences of the choice

To better understand the choice of fixed or random, I review the consequences of that choice. The fixed block effect model is (1) with no distribution placed on the β_i. The random effect model is (2), with the mean and variance of the β_i specified.

3.1   Traditional randomized complete block design

A traditional randomized complete block design has one instance of each treatment in each block. For this design, almost all results are identical for both models (Table 2). The result that does depend on the choice is the standard error (se) of a treatment mean. When blocks are random, the block variance component contributes to the se of a mean. This corresponds to the view that inferences in a random block model are to new blocks in the study population, while inferences in a fixed block model are restricted to the blocks included in the study. Differences or linear contrasts among treatments are estimated within each block, so the variability among blocks cancels out. The variability among blocks does affect inferences about an individual treatment mean because there is no cancellation of the block effects.

Results for the traditional RCBD carry over to situations where treatments are unequally replicated but identically so in all blocks (e.g. treatment A occurs once, treatment B occurs 3 times, and treatment C occurs 4 times in every block). The technical requirement for the results in Table 2 is that block effects are orthogonal to treatment effects. This happens for both the traditional (each treatment once per block) and the unequally but identically replicated RCBD designs. The orthogonality condition is also met in the rare situation of proportional replication (e.g., in one block, treatments A and B each occur once, while in a second block, both treatments occur twice).

Result	Comparison	Block effects: Fixed	Random
Estimated treatment means	Same		
Estimated difference or contrast among treatments	Same		
Estimated error variance	Same		
Estimated se of treatment differences or contrasts	Same		
Inference (tests, intervals) for differences or contrasts	Same		
Standard error of a treatment mean	Different	$\sqrt{\frac{\sigma_e^2}{b}}$	$\sqrt{\frac{\sigma_e^2 + \sigma_b^2}{b}}$

Table 2: Comparison of results from fixed block effects and random block effects models for analyzing data from a traditional RCBD. σ_e^2 is the variance of the ϵ_{ij} in both models. σ_b^2 is the variance of the β_i in the random block effect model (2). b is the number of blocks.

3.2   Incomplete or unbalanced blocks

Consequences change considerably when the number of occurrences of a treatment is not the same in all blocks (unbalanced blocks) and extreme version of this when blocks do not include all treatments (incomplete blocks). In particular, the estimates of treatment differences change because the random block effect model "recovers inter-block information" (Giesebrecht and Gompertz, 2004, p. 203). The estimated treatment means also change to be consistent with the estimated differences.

To understand this effect, consider a study with 3 treatments conducted in 5 blocks, with unequal replication in some blocks (Table 3). Each treatment either occurs once in a block or 8 times in a block. The number of replications of each treatment in Blocks 4 and 5 are proportional, but those in Blocks 1, 2, and 3 are not. Using a random block effects model, the expected difference between the block means for Blocks 4 and 5 is

$$
\begin{aligned}
\text{E}\left[\bar{Y}_{4.} - \bar{Y}_{5.}\right] &= \mu + \beta_4 + (\tau_A + \tau_B + \tau_C)/3 - [\mu + \beta_5 + (\tau_A + \tau_B + \tau_C)/3] \\
&= \text{E}\ (\beta_4 - \beta_5) \\
&= 0.
\end{aligned}
$$

This difference in block means carries no information about treatment differences. Similar behavior happens for any pair of blocks with proportional treatment replication. The difference between Blocks 1 and 2, where the treatment replication is not proportional, is:

$$
\begin{aligned}
\text{E}\left[\bar{Y}_{1.} - \bar{Y}_{2.}\right] &= \mu + \beta_1 + (8\tau_A + \tau_B + \tau_C)/10 - [\mu + \beta_2 + (\tau_A + 8\tau_B + \tau_C)/10] \\
&= \text{E}\ [\beta_1 - \beta_2 + 7(\tau_A - \tau_B)/10] \\
&= 7(\tau_A - \tau_B)/10.
\end{aligned}
$$

The difference between these two block averages contains information about the difference between treatments A and B.

When blocks are treated as a fixed effect, treatment differences are estimated by averaging the within-block differences across blocks. The between-block, or inter-block, information is

block	# trt A	# trt B	# trt C
1	8	1	1
2	1	8	1
3	1	1	8
4	1	1	1
5	8	8	8

Table 3: Example of unequal replication. Table gives the number of times an experimental units assigned to each treatment in each of the five blocks.

ignored. When blocks are treated as a random effect, the estimated treatment differences combine the direct, i.e., within-block, comparison between two treatments and the inter-block information (Giesebrecht and Gompertz 2004, p. 204-205). The practical consequence is that switching how block effects are modeled, from fixed to random or vice-versa, will change the estimated treatment differences and hence the estimated treatment means.

4 Two illustrations of recovering between-block information

4.1 Made-up data with a Gaussian response

An example suggested by Dan Nettleton illustrates recovery of between-block information. The data in this study can be analyzed either by fitting a mixed model or by explicitly combining the within- and the between-block estimates of the treatment difference. This suggests a practical approach for a discrete-data situation where a generalized linear mixed model is not feasible.

The study compares two treatments in a total of 60 blocks. Twenty blocks are size 2 with one instance of treatment A and one instance of treatment B. Twenty more blocks are size 1 with only one instance of treatment A; the final 20 blocks are also size 1 with only one instance of treatment B. Treatment means for the three sets of blocks are given in Table 4.

		Average Y	
Blocks	With	trt A	trt B
1 - 20	both trt's	10.9938	10.1000
21 - 40	only trt A	10.5155	–
41 - 60	only trt B	–	9.4000

Table 4: Treatment means to illustrate combining within- and between-block information.

I consider three estimators of the mean difference between treatments A and B. The first uses only the paired data in Blocks 1-20, the second uses only the unpaired data in Blocks 21-40,

and the third uses all the data (Table 5). The estimate from all the data is a weighted average of the paired- and unpaired- estimates, with weights that are the reciprocals of the estimated variances (Table 5). The estimate from both sets of data is $(1.1155 \times 7.0396 + 0.8938 \times 7.716)/(7.0396 + 7.716) = 0.9996$ with a standard error of $\sqrt{1/(7.0396 + 7.716)} = 0.2603$.

Data source	blocks	estimate	se	var	weight
paired obs.	1-20	0.8938	0.3600	0.1296	7.716
unpaired obs.	21-60	1.1155	0.3769	0.14201	7.0396
all	1-60	0.9996	0.2603		

Table 5: Three estimates of the difference between treatments A and B, with the inverse-variance weights used to pool the between- and within-block information.

When all the data are analyzed using a model with blocks treated as a fixed effect, the estimated treatment difference and its standard error are the same as the paired-data results. The unpaired data are ignored because the single observation in each block is used to estimate the block mean. When blocks are treated as a random effect, the estimated difference and its standard error are exactly the same as the results from combining the paired- and unpaired-data. The only difference is that the mixed model with random blocks provides an estimate of the variance between blocks, σ_b^2.

The design of this study is excellent for recovery of between-block information. The efficiency of the all-data estimator is $(0.3769/0.2603)^2 = 1.91$, which is extremely high. Most studies have lower efficiency because they have more treatments in each block, which reduces the efficiency of recovering between-block information, and blocking is more effective. In this example, blocking is not effective; the standard error of the paired-data estimate from 40 observations is only slightly smaller than the standard error from the unpaired-data estimate, also from 40 observations. The efficiency of recovering inter-block information is highest when blocking is least effective, i.e., when the block variance is small (Federer 1967, p. 268).

4.2 Example with discrete data

The concept of combining two sources of information can be used when it is not feasible to fit a mixed model. One practical situation is a study like that in section 4.1, except that the response is a Bernoulli random variable, e.g., yes or no for a trait of interest. We want to compare the probability of a yes response in the two treatments. Some observations are paired, so the Bernoulli response is observed for both treatments in a block. Other observations are not paired, so the response is observed for only one treatment.

I faced this issue when evaluating the performance of models to predict whether an introduced tree or shrub will escape from cultivation (Widrlechner et al. 2009). The Bernoulli response was whether the model prediction was incorrect (an error) or not. We needed to compare the performance of two models developed with slightly different data sets and applied to overlapping but not identical lists of species. For the species on both lists, we have paired

30

responses and the analysis follows the principles of a McNemar's test. We expect a strong positive correlation between the paired responses because some species are "hard" to predict correctly while other species are not. Those hard-to-predict-correctly species are likely to generate an error on both lists. The species on only one list are two independent samples, and the analysis follows the principles of a Chi-square test.

Trying to fit a generalized linear mixed model (GLMM) with species as blocks leads to serious numerical issues because of a dichotomous response with only 1 or 2 observations per block. The practical alternative is conditional logistic regression, which conditions on the block totals. The conditional logistic regression is asymptotically equivalent to McNemar's test (Santner and Duffy 1989, p. 268). Both analyses consider only the paired observations and ignore the unpaired part of the data set. Explicitly combining the two estimates provides answers that use all the data.

5   A false solution with a random block effects model

The final consequence involves details of fitting the random blocks model to Gaussian data. Under a fixed block effects model, there are closed-form solutions to the least-squares equations. In contrast, the recommended algorithm for the random block effects model, residual maximum likelihood (REML), is an iterative algorithm. The estimated block variance component, σ_b^2, is found by numeric maximization. It is easy to forgot that a log likelihood surface can be badly behaved, i.e., nearly flat in some parts of the parameter space or having more than one local maximum. In such situations, numeric maximization may provide incorrect estimates. It is especially easy to forget about numeric issues when the problem is widely regarded as "easy", such as fitting a model with random block effects. A data set, originally constructed to explore block effect models with a quantitative treatment, reminded me to never forget about numeric issues.

This study has 10 blocks, each with three experimental units. The treatment is a quantitative amount, e.g. of shading, of fertilizer, or of irrigation. The 10 blocks differ in the overall level of the quantitative treatment; they also differ in other unspecified ways that influence the mean response. The data are shown in Figure 1.

We consider three models for these data:

$$\text{No block effects: } Y_{ij} \;=\; \mu + \tau\, X_{ij} + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0, \sigma_e^2) \tag{4}$$
$$\text{Fixed block effects: } Y_{ij} \;=\; \mu + \beta_i + \tau\, X_{ij} + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0, \sigma_e^2) \tag{5}$$
$$\text{Random block effects: } Y_{ij} \;=\; \mu + \beta_i + \tau\, X_{ij} + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0, \sigma_e^2),\ \beta_i \sim N(0, \sigma_b^2) \tag{6}$$

I used SAS proc mixed with default settings to fit the fixed block model (5) and the random block model (6). The results are very different. The fitted lines from the fixed block effects

Annual Conference on Applied Statistics in Agriculture
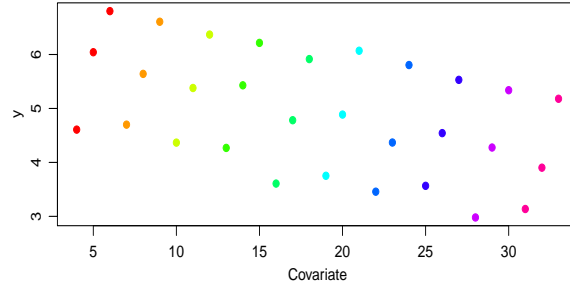Kansas State University

Figure 1: Data for the example with a quantitative variable. Blocks are indicated by colors.

model (5) are clearly based on the trend within each block (Figure 3). The fitted line from the random block effect model is clearly based on the trend across the blocks (Figure 3) and is very similar to the fitted line from the no block model that ignores block effects.

Because the blocks have very different mean covariate values, the block effects are not orthogonal to the regression slope so the estimated slope, $\hat{\tau}$ depends on $\hat{\sigma}_b^2$ (Figure 2). When $\hat{\sigma}_b^2$ is close to 0, the estimated regression slope is negative; when $\hat{\sigma}_b^2$ is large, the estimated regression slope is close to that from the fixed block effects model (Figure 2).
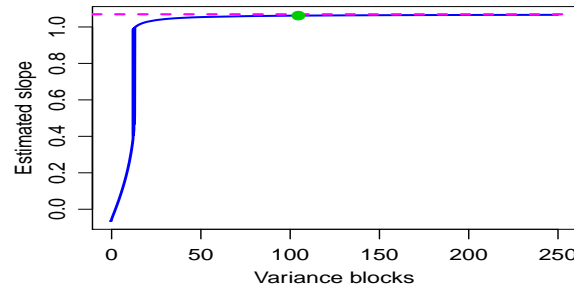

Figure 2: Estimated regression slope as a function of the block variance component. The dashed line is the estimated slope from the fixed block model.

The details of fitting the random block effects model are extremely important because the estimated slope depends on $\hat{\sigma}_b^2$. One issue is whether the "variance" component for blocks is bounded below by 0. The random block line shown in Figure 3 is for the unbounded REML estimate of σ_b^2 (SAS proc mixed nobound). That estimate is $\hat{\sigma}_b^2 = -0.42$. Although variances must be non-negative, σ_b^2 has another interpretation as the covariance between two observations in the same block, which could be positive or negative. When the nobound option is omitted, $\hat{\sigma}_b^2 = 0$ and the fitted line is the same as that from the no blocks model.

The second issue with fitting the random block effects model is that the default estimates of $\hat{\sigma}_b^2 = -0.42$ or $\hat{\sigma}_b^2 = 0$ are local maximums of the REML log likelihood function. Neither one is the correct REML estimate. This problem with the default REML fit is demonstrated by plotting the profile REML deviance as a function of σ_b^2 (Figure 4). When evaluated for values of σ_b^2 from -0.7 to 10, there is a minimum at approximately -0.4. A numerical minimization
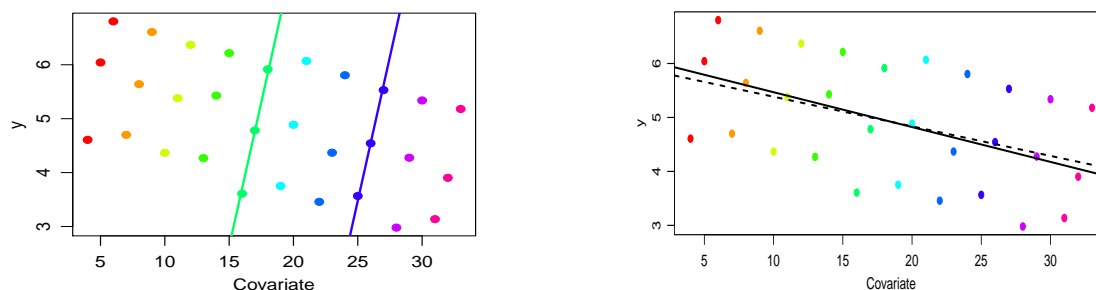
32

Figure 3: Fitted regressions from the fixed blocks model (left) and the random blocks (right, solid) and no blocks models (right, dashed). Only 2 of the 10 block-specific lines are shown for the fixed-block model.

algorithm that starts at a value between -0.7 and 10 will converge to either $\hat{\sigma}_b^2 = -0.42$ when unbounded or $\hat{\sigma}_b^2 = 0$ if bounded below at 0. The corresponding regression lines are the random block effects solution and the no blocks solution shown in Figure 3.
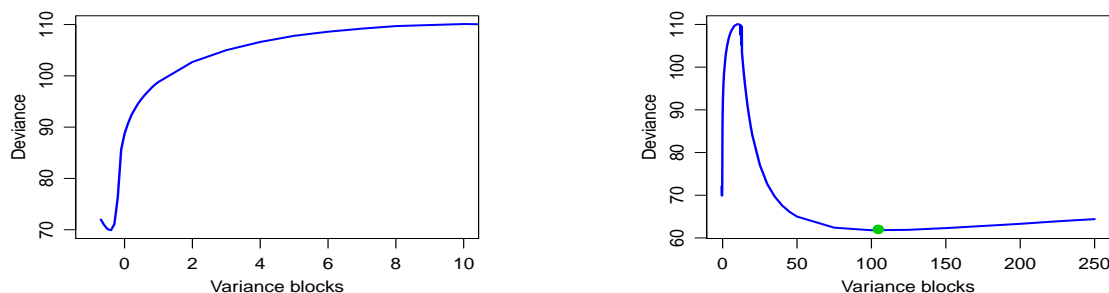

Figure 4: Profile deviance as a function of σ_b^2 for values of σ_b^2 from -0.7 to 10 (left) and from -0.7 to 250 (right).

When the profile deviance is plotted for larger values of σ_b^2 (Figure 4, right), the profile deviance, and hence the REML log likelihood function, is seen to be multi-modal. The true minimum deviance is at $\sigma_b^2 = 105$ (Figure 4, right). The correct REML estimate of the slope is approximately 1.06, which is essentially the same as the estimated slope from the fixed block effects model.

The false convergence is not unique to SAS proc mixed. SAS proc glimmix, the lme() and lmer() function in the R nlme and lme4 packages, and ASREML give the same default results. It appears that all start the iterative optimization at $\sigma_b^2 = 0$. When SAS proc mixed is started at $\sigma_b^2 = 20$, it converges to the true minimum.

Using ANOVA estimates of the variance components provides even worse results, at least as implemented in SAS as proc mixed method = type3. The estimated slope is $\hat{\beta} = 0.63$ and estimated block variance component is $\hat{\sigma}_b^2 = 1.068$. The method=type3 algorithm is not iterative, so the estimated slope is not consistent with the estimated block variance component. Conditional on $\sigma_b^2 = 1.068$, the generalized least squares estimate of the regression slope is $\hat{\beta}_{gls} = -0.027$, which is very different from the reported $\hat{\beta}_{gls} = 0.63$.

33

6   Should block effects be fixed or random?

Putting together everything about the consequences of the choice, should block effects be modeled as a fixed effect or a random effect? My answer is that blocks should be a fixed effect, unless you have compelling reasons to treat them as a random effect. My argument for fixed block effects has four parts, depending on the nature of both the study and the blocks.

6.1   When block effects are orthogonal to treatment effects

This situation includes the traditional RCBD with each treatment occurring once in a block and other designs with equal or proportional replication of treatments within each block. There is no inter-block information to recover and the choice of fixed or random block effects is irrelevant for most results (Table 2). My preference for a fixed block effects analysis in this situation is driven by my desire for a compact summary of the differences between treatments. A commonly used graphical summary is a plot of the treatment means with information about their uncertainty, e.g. bars indicating $\pm$ 1 standard error (Figure 5) or indicating 95% confidence intervals for the mean. Such a plot indirectly provides information about each pairwise difference of treatment means, without having to show each specific pair of differences. When the blocks are modeled as fixed effects, the interpretation of this plot is exactly the same as the interpretation in a completely randomized design. Using the traditional RCBD with b blocks to simplify the notation, the standard error of a difference of two treatment means is $\sqrt{2\sigma_e^2/b}$, which is larger than the standard error shown on the plot, $\sqrt{\sigma_e^2/b}$, by a constant factor of $\sqrt{2}$.

When blocks are modeled as a random effect, the additional variance component in the treatment mean (Table 2) complicates the relationship between the two standard errors. The result can be a plot of treatment means like Figure 5, right, where the se's for the means are large but the differences are precisely estimated. The discrepancy between the left and right panels in Figure 5 depends on the size of σ_b^2. When blocking is extremely effective, σ_b^2 is large and the plot of treatment means is most misleading as a display of comparisons among treatments.

Both reviewers were concerned about the use and interpretation of displays like Figure 5. Although commonly used, they are not statistically accurate displays of significant treatment differences. An accurate presentation of significant differences requires presenting information about each of the $k(k-1)/2$ treatment differences. I believe figures like Figure 5 are compact approximate displays of treatment differences. Their value is their compactness. Showing the fixed block effect standard error is more consistent with the interpretation of differences. This standard error does not estimate the precision of a treatment mean when blocks a simple random sample from some larger population.
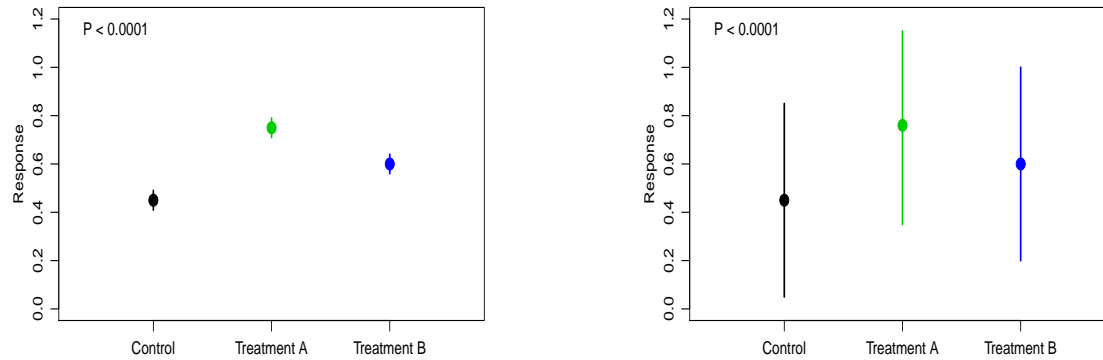
Figure 5: Plot of treatment means with vertical bars indicating $\pm$ 1 standard error. Plot on the left is for the fixed block effects model; plot on the right is for the random block effects model. The p-value for the 2 df ANOVA test of equal means is given in each panel.

6.2 When blocks are not a simple random sample from some larger population

The random block effect model assumes that the blocks used in the study are a simple random sample from some larger population of potential blocks. This assumption is crucial to the computation of the standard error in the random blocks model.

My experience is that blocks are rarely a random sample from some larger population. They are usually happenstance collections of experimental units. For example, consider a typical field experiment. Experience is that variation in a field is spatially structured, so the field is commonly divided into blocks to provide local control of the variability. For descriptive purposes, let us imagine there are six blocks, each with 5plots. In my opinion, those six blocks are not a sample from some larger population of blocks. If they are a sample from anything, that sample is a complete enumeration from the finite population of those six blocks. When the blocks used in a study are a complete enumeration of all possible blocks in the population, the correct standard error for a random block effects model is σ_e^2/n, because the only randomization is that of treatments to plots within a block.

Another example where the random blocks standard error of a mean is artificial because the blocks are not a random sample from some larger population is blocks that are constructed as weight groups. Imagine a study of pigs, where blocks are formed by initial weight. One block is those pigs from 50 kg to 55 kg, a second block is pigs from 55 kg to 60 kg, a third block is pigs from 60 kg to 65 kg, and so on up to a block with pigs from 85 kg to 90 kg. Each block has 6 pigs. For simplicity of exposition, imagine a response that is highly correlated with initial weight, so blocking is very effective. Although pigs may be randomly sampled from an available population, the variance between the blocks does not match the variance of initial weights in the population, because of the selection of 6 pigs per weight group (Figure 6). If very light pigs (50-55 kg) and very heavy pigs (85-90 kg) are less common in the population, the variance component for blocks is larger than the variance in the population.

35

Again, the standard error of a treatment mean under a random block effects model is not an appropriate estimate of the variability in the larger population of pigs.
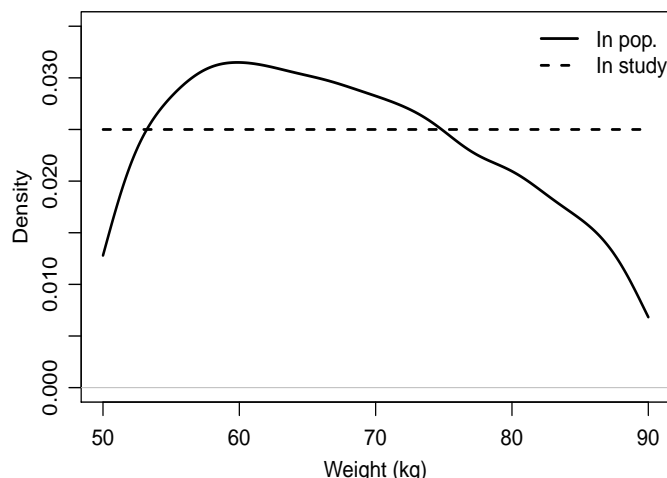

Figure 6: Density of pig initial weights in the available population of pigs (solid) and after grouping and selection into blocks defined by initial weight (dashed).

There are situations where blocks are actually a sample from a larger population, e.g., when a block is an individual animal or human being and those individuals are randomly sampled from the population of interest. The random sample is key, because then the variance component for blocks does estimate the variance between individuals in the population. In this case, the compelling reason to model block effects as a random factor is that the standard error of the mean is an appropriate estimator of the variability in a random sample in the population.

6.3   In an observational study with inter-block information

The context for my discussion is an observational study in which individuals are blocks and the "treatment" of interest varies both within and between individuals. Treatment differences could be estimated solely by within-individual information, solely by between-individual information, or by the combination of the two. Treating blocks as a fixed effect only uses within-individual information; treating blocks as a random effect uses the combination of within- and between-individual information.

Paul Allison (2009) argues that the between-individual information is likely to be biased by omitted confounding variables. The within-block information is less biased (or unbiased) because the effects of block-specific confounding variables cancel out of the estimated treatment differences. There is a tradeoff between bias and efficiency. Treating blocks as a random effect may provide a more efficient estimator of the treatment effect but that estimator is likely to be biased. The fixed block effects analysis is less efficient but less biased. Hence, he argues strongly for using the fixed block effects model.

6.4   In an experimental study with inter-block information

In an experimental study, the within-block information provides an unbiased estimate of treatment effects, because of the random assignment of treatments to units within the block. Whether or not the between-block information provides unbiased estimates of the treatment differences depends on the details of the study. If groups of treatments are randomly assigned to incomplete blocks, then the between-block information about treatment differences is unbiased, because of that randomization.

As the example in Section 4.1 illustrates, recovering the inter-block information is equivalent to a weighted average of two estimates, where the weights depend on σ_b^2 and σ_e^2. Because those variance components are estimated, the weights are random variables, which has consequences for the statistical properties of the weighted average. Specifically, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimates of θ, the weighted average $w_1\hat{\theta}_1 + w_2\hat{\theta}_2$ is unbiased when w_1 and w_2 are fixed constants. When the weights are estimated, the weighted average $\hat{w}_1\hat{\theta}_1 + \hat{w}_2\hat{\theta}_2$ is unbiased only if the estimated weights are uncorrelated with the estimated parameters (Claeskens 2016).

Federer (1967, p. 266) recommends to not recover inter-block information unless the block degrees of freedom block is larger than 12 to 14. His reason is that the block variance component is poorly estimated when the block d.f. is small.

7   Summary

The choice of model for blocks, as a fixed effect or as a random effect, has consequences for at least some results of an analysis. For some designs, where there is inter-block information about the treatments, the choice can have a large effect. I definitely agree that there are situations where treating blocks as a random effect is justifiable for the study at hand. But, I argue that treating blocks as a fixed effect is usually the more appropriate choice. At a minimum, I recommend that a data analyst be aware of the consequences of the choice.

8   Acknowledgements

References

Allison, P.D. (2009) *Fixed Effects Regression Models* Sage Publications, Los Angeles.

Casella, G. & Berger, R.L. (1990) *Statistical Inference* Duxbury Press, Belmont, CA.

Claeskens, G., Magnus, J.R., Vasnev, A.L., & Wang, W. (2016) The forecast combination puzzle: a simple theoretical explanation. International Journal of Forecasting 32:754-762.

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

Eisenhart, C. (1947) The assumptions underlying the analysis of variance. Biometrics 3:1-21.

Federer, W.T. (1967) *Experimental Design: Theory and Application* Indian edition. Oxford and IBH Publishing, Calcutta.

Fisher, R.A. (1925) *Statistical Methods for Research Workers* Oliver and Boyd, Edinburgh.

Gelman, A. (2005) Analysis of variance: why it is more important than ever. *Annals of Statistics* 33:1-31.

Giesebrecht, F.G. & Gompertz, M.L. (2004) *Planning, Construction, and Statistical Analysis of Comparative Experiments.* Wiley, Hoboken, NJ.

Green, B.F. & Tukey, J.W. (1960) Complex analyses of variance: general problems. *Psychometrika* 25:127-152.

Kreft, I. & deLeeuw, J. (1998) *Introducing Multilevel Modeling* Sage Publications, London.

LaMotte, L.R. (1983). Fixed-, random-, and mixed-effects models. *Encyclopedia of Statistical Science* 3:137-141.

McCulloch, C.E. Searle, S.R. & Newhaus, J.M. (2008) *Generalized, Linear, and Mixed Models*, 2nd. ed. Wiley, New York.

McLean, R.A., Sanders, W.L., & Stroup, W.W. (1991) A unified approach to mixed linear models. *American Statistician* 45(1):54-64.

Moll, E., Piepho, H-P & Richter, C. 2006. Blocks fix oder zuf allig? Available at `http://www.biometrische-gesellschaft.de/fileadmin/AG_Daten/Landwirtschaft/tagungsberichte_pdf/06_Blocks_MollPiephoRichter.pdf`. Accessed 10 May 2017.

Piepho, H-P. 2006. Anmerkungen zur Frage "Bl ocke fix oder zuf allig"? Available at `http://www.biometrische-gesellschaft.de/fileadmin/AG_Daten/Landwirtschaft/ tagungsberichte_pdf/08_Blocks_Piepho.pdf`. Accessed 10 May 2017.

Santner, T.J. and Duffy, D.E. 1989. *The Statistical Analysis of Discrete Data* Springer-Verlag, New York.

Searle, S. (1971) *Linear Models* Wiley, New York.

Toutenberg, H. (2002) *Statistical Analysis of Designed Experiments*, 2nd. ed. Springer, New York.

Widrlechner, M.P., Thompson, J.R., Kapler, E.J., Kordecki, K., Dixon, P.M., & Gates, G. (2009) A test of four models to predict the risk of naturalization on non-native woody plants in the Chicago region. Journal of Environmental Horticulture 27(4):241-250.

Yates, F. (1965) A fresh look at the basic principles of the design and analysis of experiments. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 4:777-790.