ARE BLOCKS DIFFERENT FROM RANDOM FACTORS?

by

Myra L. Samuels, George Casella and George P. McCabe
Purdue University and Cornell University *

Technical Report #88-26

Department of Statistics
Purdue University

December 1988

* This research was performed while George Casella was Visiting Professor, Department of Statistics, Purdue University.

★

ARE BLOCKS DIFFERENT FROM RANDOM FACTORS?

by

Myra L. Samuels, George Casella and George P. McCabe
Purdue University and Cornell University

Abstract

Can blocks be tested in a randomized blocks design? It is well–known that two different linear mixed models can be written which yield conflicting answers to this question. This paper examines the models from the point of view of statistical relevance. It is found that the question of testing blocks is not the same as the question of testing a random factor. Viewing blocking as a device to increase efficiency leads to the conclusion that blocks *can* be tested. Two tests for blocks are discussed and compared, and confidence intervals to assess the magnitude of the block effect are described.

KEY WORDS: Randomized blocks, Mixed model; Random factors; Analysis of variance; Linear models; Expected mean squares.

1. Introduction

In a two–way mixed model, how should one test the random factor if additivity is not assumed? Can blocks be tested in a randomized blocks design if blocks are regarded as random and additivity is not assumed? Although these questions have been around for several decades (Cornfield and Tukey, 1956; Wilk and Kempthorne, 1955, 1956), there is still some confusion as to the answers. The conflict is dramatically illustrated by the fact that the widely used SAS statistical package produces tables of expected mean squares (EMSs) which disagree with those given in most textbooks, and with another widely used statistical package, BMDP.

This paper will address these questions, with special attention to the important case of blocked designs. We consider not only randomized blocks designs but also nonrandomized observational designs which incorporate blocking.

To set the stage, consider the following examples of blocked designs. (Although our examples all involve human beings, our discussion is meant to be quite general.)

Example 1.1. Pairs of female twins are randomly selected from a source population to participate in a study of an anti–aging skin cream. One randomly chosen member of each pair receives the cream (Treatment 1) and the other receives placebo (Treatment 2). The observation Y is a measure of average skin thickness on the two forearms.

Example 1.2. The treatments are as in Example 1.1, but now the subjects are women (not twins) randomly selected from a source population. The two treatments are randomly allocated to the two arms of each subject, and skin thickness Y is measured on each arm.

Example 1.3. As in Example 1.2, the subjects are women randomly selected from a source population. The question of interest now is whether the density of neurons is greater in the right or the left hemisphere of the brain. Each measurement Y of neural density is taken from a CAT scan of the head.

Example 1.4. As in Example 1.1, pairs of female twins are observed, but now the twins are selected from a source population in which one member of the pair smokes cigarettes and the other does not. The observation Y is a measure of cardiac health.

Examples 1.1 and 1.2 are commonly called randomized blocks designs. In Example 1.1 the block is a pair of subjects, while in Example 1.2 each subject is a block. In the psychometric literature, Example 1.2 is also called a repeated measures design.

Examples 1.3 and 1.4 are blocked designs but are strictly observational: no randomized allocation is involved. Note that the primary purpose of blocking in Example 1.4 is to control confounding – that is, to remove the effect of covariates (such as age and genetic background) which might otherwise distort the observed relationship between smoking and cardiac health. By contrast, in Examples 1.1, 1.2 and 1.3 a valid study could be conducted without blocking, but blocking is expected to increase efficiency.

For designs such as the above, a standard statistical approach – a mixed–model analysis of variance (ANOVA) – leads to two conflicting answers as to whether the ANOVA table yields an F test for blocks. More generally, there are two conflicting answers as to which mean square should form the denominator of an F statistic for testing the random factor in a two–way mixed model. Several authors (Hartley and Searle 1969; Searle 1971a, Searle 1971b, pp. 401–404; Hocking 1973; Harville 1978; Hocking 1985, pp. 330-334) have discussed the conflict and have noted that in a certain sense the discrepancy can be resolved by suitable re–definition of the model parameters. This is an algebraic resolution, but it is not a statistical resolution because it sidesteps the question of which parameterization is preferable. We will discuss the problem from a statistical point of view and try to develop some statistical perspective to aid the experimenter who asks: Can I test blocks? Should I? And, if so, how?

In Section 2 the two conflicting answers are exhibited. In Sections 3, 4, and 5 the special case of two treatments is explored. In Section 3 it is argued, not only that blocks can be tested, but also that there are two ways to do so. In Section 4 two mixed models are defined, the paradox of the conflicting ANOVA tables is resolved, and the statistical implications of the resolution are discussed. In Section 5 the discussion is extended to the case of unequal variances. Extension to more than two treatments is considered in Section

6 and the case of replicated observations is treated in Section 7. The statistical resolution of the conflict is summarized and discussed in Section 8. Proofs and further details are contained in three appendices.

In discussing formal inference, we will assume that the random variables in the models are jointly normal. Much of our development, however, involves only first and second moments and is not dependent on normality assumptions.

2. The Issue

For simplicity, we will use the term "treatments" throughout our discussion, even though this usage is unnatural for an application like Example 1.3.

Let the data be represented by Y_{ijk}, where $i = 1, \ldots, I$ represents treatments, $j = 1, \ldots, J$ represents blocks and $k = 1, \ldots, K$ represents repeated measurements on the same treatment–block combination. (Although $K = 1$ in Examples 1.1—1.4 and in many blocked designs, we will find it helpful to also consider the case $K > 1$.) Table 2.1 shows the usual orthogonal decomposition of the total sum of squares, including terms for treatments, blocks, the treatment by block interaction and variation within a treatment–block combination.

Table 2.1 goes here

Corresponding to the decomposition of the sum of squares, it is natural to think in terms of an equation of the form

$$Y_{ijk} = \mu + T_i + B_j + G_{ij} + \varepsilon_{(ij)k} \tag{2.1}$$

where μ represents an overall mean, T_i the treatment effect, B_j the block effect, G_{ij} the treatment–block interaction and $\varepsilon_{(ij)k}$ the residual variation within a treatment–block combination. The equation (2.1) becomes a statistical model useful for guiding the analysis of data only when constraints and assumptions regarding the distributions of the various terms are defined. These issues will be addressed in detail in Section 4.

For general analysis of variance models, construction of F tests or quasi-F tests for testing hypotheses under normality assumptions is determined by a table of expected mean squares (EMSs). The controversy regarding blocks arises because, using models that appear similar on the surface, it is possible to calculate different expressions for the EMSs. The case where disagreement exists is the mixed model, in which treatments are regarded as fixed and blocks are regarded as random.

Routine application of the "EMS algorithm" (see, for example, Winer 1971 or Kirk 1982) to (2.1) with treatments fixed and the other terms random gives the first column (Version 1) of Table 2.2. Examination of theses EMSs suggests that the denominator for testing blocks should be MS(Within T*B). In the interesting and important case where $K = 1$, this term has no degrees of freedom, which seems to indicate that blocks cannot be tested. The Version 1 table is essentially what is presented in most textbooks (for instance, Scheffe, 1959; Winer, 1971; Snedecor and Cochran, 1980; Steel and Torrie, 1980; Hicks, 1982; Kirk, 1982; Montgomery, 1984) and by the computer program BMDP8V.

Table 2.2 goes here

On the other hand, the text by Searle (1971, p. 401) and the RANDOM statement in the SAS procedure GLM give the EMSs presented in the second column (Version 2) of Table 2.2. This EMS table can be obtained via the EMS algorithm from the following slight modification of (2.1):

$$Y_{ijk\ell} = \mu + T_i + B_j + G_{\ell(ij)} + \varepsilon_{(ij\ell)k}$$

where $\ell \equiv 1$. (See Appendix 1 for a further discussion of this model equation.) The Version 2 EMSs suggest that the denominator for testing blocks should be MS(T*B), and, in particular, that blocks <u>can</u> be tested when $K = 1$.

The EMS algorithm can be a convenient tool to guide an analysis, but we see that it can give different answers when applied to what appear to be very similar representations for the same set of data. Furthermore, statistical practice should not be determined by which software package is available to the user. We will try to clarify the situation by translating (2.1) into more fully specified models.

In the models we consider, the random variables are generated by sampling from (infinite) populations. Before proceeding, we note that other models are sometimes used for a randomized blocks analysis. For example, one might use the restriction error models of Anderson (Anderson 1970, Anderson and McLean 1974), or permutation models (Kempthorne 1952, pp. 135–151, White 1975). Such models are based on different assumptions than the models considered here, and can lead to different answers. (In the restriction error model, for example, blocks cannot be tested under any circumstances.) The models we consider, sometimes called population models, represent a common approach to modeling a statistical analysis. (Note that permutation models would be unnatural for Examples 1.3 and 1.4, since those examples do not involve randomization).

In the next three sections we assume that $I = 2$ and $K = 1$, that is, there are only two treatments and one observation per treatment–block combination. This simplification accomplishes two things. First, we can explore the issue of testing blocks with reduced algebraic effort. Second, we can treat the case of unequal variances, which has some interesting consequences and is helpful in understanding the relationship between the block effect and the interaction effect.

3. To Block or Not to Block

In this section the testing and estimation of block effects will be approached from first principles, rather than from the viewpoint of a linear model like (2.1). In Section 4 the results will be related to two different versions of (2.1).

When an investigator considers the hypothesis that "There is no effect due to blocks," what exactly is meant? One interpretation, which we feel is the most reasonable, is that the study could just as well have been done without blocks, that is, as a completely randomized design (e.g., in Examples 1.1 and 1.2), or with independent sampling rather than paired sampling (e.g., in Examples 1.3 and 1.4). In other words, the hypothesis is interpreted to mean that nothing was accomplished by blocking.

What is blocking intended to accomplish? The answer can vary according to the nature of the study. When blocking serves the purpose of controlling confounding (as in Example 1.4), the conditions under which blocks can be ignored are complex. We will not consider this case, but rather will limit our attention to cases (as in Examples 1.1—1.3) where the purpose of blocking is to increase the efficiency of treatment comparisons. Thus, we consider only the situation in which a non–blocked design would be a valid alternative to the blocked design.

3.1 A General Model

For the case $I = 2$ and $K = 1$ we suppress the third subscript and let

$$\{(Y_{1j}, Y_{2j})';\ j = 1, \ldots, J\}$$

be independently and identically distributed (iid) random vectors each distributed as the vector $(Y_1, Y_2)'$. (Note that this formulation implicitly regards blocks as a random factor.)

The treatment difference to be estimated is $E(Y_1) - E(Y_2)$. To compare a blocked design with a non–blocked design it is appropriate to compare the variances of the differences $(Y_1 - Y_2)$ under the two designs.

For a non–blocked design, Y_1 and Y_2 are modeled as independent random variables. Thus, we have

$$\mathrm{Var}_N(Y_1 - Y_2) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) \qquad \text{(nonblocked)}$$

$$\mathrm{Var}_B(Y_1 - Y_2) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) - 2\ \mathrm{Cov}(Y_1, Y_2) \qquad \text{(blocked).}$$

A natural measure of the effect of blocking is the ratio λ of these two variances, which is called the relative efficiency: *

$$\lambda = \frac{\mathrm{Var}_N(Y_1 - Y_2)}{\mathrm{Var}_B(Y_1 - Y_2)} = \frac{\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2)}{\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) - 2\ \mathrm{Cov}(Y_1, Y_2)} \qquad (3.1)$$

* Strictly speaking, (3.1) is the asymptotic relative efficiency, since it does not account for the loss in degrees of freedom due to blocking.

To obtain the same variance as a blocked design with J observations on each treatment pair, λJ observations on each treatment would be required in a non–blocked design. Note that if the covariance between Y_1 and Y_2 is negative (a situation unlikely to be encountered in most applications, but possible in principle), then blocking results in a loss, rather than a gain, of efficiency.

If we let

$$\tilde{\rho} = \frac{\mathrm{Cov}(Y_1, Y_2)}{\frac{1}{2}[\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2)]}, \tag{3.2}$$

then we can write (3.1) as

$$\lambda = \frac{1}{1 - \tilde{\rho}}. \tag{3.3}$$

Note the similarity between $\tilde{\rho}$ and the Pearson correlation coefficient

$$\rho = \frac{\mathrm{Cov}(Y_1, Y_2)}{[\mathrm{Var}(Y_1)\mathrm{Var}(Y_2)]^{\frac{1}{2}}}.$$

The coefficients $\tilde{\rho}$ and ρ have the same sign. However, $|\tilde{\rho}| \leq |\rho|$, with equality if $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2)$. (The latter facts follow because the denominator of $\tilde{\rho}$ is the arithmetic mean of the variances, whereas the denominator of ρ is the geometric mean.)

From Equation (3.3) it is clear that in the context of this general model, issues regarding the effect of blocking can be addressed by consideration of the parameter $\tilde{\rho}$. In particular, the hypotheses of no block effect is expressed as

$$H_0 \colon \tilde{\rho} = 0$$

or equivalently,

$$H_0 \colon \mathrm{Cov}(Y_1, Y_2) = 0$$

and a measure of the effect of blocking is given by an estimate of $\tilde{\rho}$.

3.2 Testing for Blocks

Table 3.1 gives the ANOVA table with expected mean squares derived from the general formulation of Section 3.1. From Table 3.1 it follows that

$$\begin{aligned} \mathrm{EMS(Blocks)} &= \tfrac{1}{2}\mathrm{Var}(Y_1 + Y_2) = \tfrac{1}{2}[\mathrm{Var}\ (Y_1) + \mathrm{Var}\ (Y_2)] + \mathrm{Cov}(Y_1, Y_2) \\ \mathrm{EMS(T{*}B)} &= \tfrac{1}{2}\mathrm{Var}(Y_1 - Y_2) = \tfrac{1}{2}[\mathrm{Var}\ (Y_1) + \mathrm{Var}\ (Y_2)] - \mathrm{Cov}(Y_1, Y_2) \end{aligned} \tag{3.4}$$

Thus, under the hypothesis $H_0\colon \text{Cov}(Y_1, Y_2) = 0$, MS(Blocks) and MS(T*B) have the same expectation.

In the remainder of Section 3 we assume that the distribution of $(Y_1, Y_2)'$ is normal. Under this assumption, MS(Blocks) and MS(T*B) are each distributed as a scaled chi–squared random variable. Also, for now we impose the further assumption that $\text{Var}(Y_1) = \text{Var}(Y_2)$. Then, using the facts that

$$\text{Cov}(Y_1 + Y_2,\ Y_1 - Y_2) = \text{Var}(Y_1) - \text{Var}(Y_2) \tag{3.5}$$

and that MS(Blocks) is a function only of $(Y_1 + Y_2)$ and MS(T*B) is a function only of $(Y_1 - Y_2)$, the equality of variance assumption, together with normality, implies that MS(Blocks) and MS(T*B) are independent. Therefore, under the hypothesis $H_0\colon \text{Cov}(Y_1, Y_2) = 0$ we have

$$\frac{\text{MS(Blocks)}}{\text{MS(T*B)}} \sim F_{J-1,J-1}, \tag{3.6}$$

where $F_{m,n}$ is an F random variable with m and n degrees of freedom. Against a two–sided alternative the test based on (3.6) is the likelihood ratio test (LRT) (Mehta and Gurland 1969).

Referring back to Table 2.2, note that the F test based on (3.6) is suggested by the Version 2 EMSs but not by the Version 1 EMSs. We will return to this point in Section 4.

We have justified the F test for blocks as a test of $H_0\colon \text{Cov}(Y_1, Y_2) = 0$. We now might ask, why go through the gyrations of the development leading to (3.6), rather than directly testing the covariance. In fact, we can directly test the covariance using the Pearson product–moment correlation

$$r = \frac{\sum_j (Y_{1j} - \overline{Y}_{1\cdot})(Y_{2j} - \overline{Y}_{2\cdot})}{[(\sum_j (Y_{1j} - \overline{Y}_{1\cdot})^2 \sum_j (Y_{2j} - \overline{Y}_{2\cdot})^2)]^{\frac{1}{2}}}. \tag{3.7}$$

Under $H_0\colon \text{Cov}(Y_1, Y_2) = 0$, we have the familiar result that

$$\frac{\sqrt{J-2}\, r}{\sqrt{1 - r^2}} \sim t_{J-2}, \tag{3.8}$$

where t_n is a t random variable with n degrees of freedom.

Thus, we find that under the assumption of equal treatment variances there are two tests, based on (3.6) and (3.8), which are valid for testing $H_0\colon \mathrm{Cov}(Y_1, Y_2) = 0$. If in fact $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2)$, then the F test based on (3.6) is more powerful than the t test based on (3.8) (see Appendix 2); however, computations suggest that the difference in power is very small. The advantage of the t test is that it does not depend on the assumption of equal variances.

3.3 Estimation of the Block Effect

The investigator who is interested in testing blocks may also be interested in estimating the magnitude of the increase (or, perhaps, decrease) in efficiency due to blocking. Such an estimate would be helpful in planning future studies in similar settings. For example, if blocking is costly or inconvenient, and the anticipated gain in efficiency is small, then an experimenter might opt for a completely randomized design.

From equation (3.3) we saw that the relative efficiency is simply related to the parameter $\tilde{\rho}$. For a researcher interested in estimating the relative efficiency, an estimate of $\tilde{\rho}$ is therefore needed.

The natural estimator of $\tilde{\rho}$ is

$$\tilde{r} = \frac{\sum_j (Y_{1j} - \overline{Y}_{1\cdot})(Y_{2j} - \overline{Y}_{2\cdot})}{\frac{1}{2}[\sum_j (Y_{1j} - \overline{Y}_{1\cdot})^2 + \sum_j (Y_{2j} - \overline{Y}_{2\cdot})^2]}. \tag{3.9}$$

This statistic is the maximum likelihood estimator of $\tilde{\rho}$, whether or not it is assumed that $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2)$ (Mehta and Gurland 1969). Furthermore, it is easy to show that

$$\tilde{r} = \frac{\mathrm{MS(Blocks)} - \mathrm{MS(T*B)}}{\mathrm{MS(Blocks)} + \mathrm{MS(T*B)}}. \tag{3.10}$$

so that $\tilde{r}$ is easily calculated from the mean squares in the ANOVA table. The statistic $\tilde{r}$ is sometimes called an intraclass correlation coefficient or a reliability coefficient; more often, however, these names are given to a somewhat different statistic (see Winer 1971, pp. 286–287 and Cochran 1980, pp. 243–244).

We turn now to the problem of setting confidence limits on $\tilde{\rho}$, and thereby on λ. If it is assumed that $\text{Var}(Y_1) = \text{Var}(Y_2)$, then limits can be derived from the fact that the quantity $[(1-\tilde{\rho})(1+\tilde{r})]/[(1+\tilde{\rho})(1-\tilde{r})]$ follows an F distribution (Kristof 1972). The resulting $100(1-\alpha)\%$ confidence limits for $\tilde{\rho}$ are given by

$$\frac{1+\tilde{r}-(1-\tilde{r})F_{1-\frac{\alpha}{2};J-1,J-1}}{1+\tilde{r}+(1-\tilde{r})F_{1-\frac{\alpha}{2};J-1,J-1}}\ ,\quad \frac{1+\tilde{r}-(1-\tilde{r})F_{\frac{\alpha}{2};J-1,J-1}}{1+\tilde{r}+(1-\tilde{r})F_{\frac{\alpha}{2};J-1,J-1}} \tag{3.11}$$

where $F_{p;m,n}$ represents the pth percentile of an $F_{m,n}$ distribution. Note that the interval (3.11) is different from the interval more commonly given for intraclass correlation (as, for instance, implicitly by Snedecor and Cochran (1980, pp. 245–246)) because it is based on a different experimental design.

4. Two Models and Three Hypotheses

In this section we formulate two different linear mixed models for the blocked design with two treatments. In the context of these models we consider three hypotheses, each of which asserts, in a different sense, the absence of a block effect. Both models are special cases of the general model presented in Section 3.1. Recall that we assumed $\{(Y_{1j}, Y_{2j})' : j = 1, \ldots, J\}$ to be iid random vectors. Let

$$E(Y_{ij}) = \mu + \tau_i$$

where $\Sigma\tau_i = 0$, and let

$$\phi_\tau = \Sigma\tau_i^2.$$

Thus, ϕ_τ is the usual noncentrality parameter for treatment effects (the definition for general values of I is given in Section 6).

As a first step, we decompose the Y_{ij} as follows:

$$Y_{ij} = W_{ij} + \varepsilon_{ij} \tag{4.1}$$

where the ε_{ij} are iid random variables with mean zero and variance σ_ε^2. The W_{ij} are random variables that represent the mean value of Y_{ij} that would be obtained from a large number of observations of treatment i in block j, while the ε_{ij} represent variation about these means. For instance, in Example 1.3 W_{1j} and W_{2j} would be the actual values of neural density in the right and left hemispheres, while ε_{ij} would represent measurement error. We assume that the W_{ij} and the ε_{ij} are independent.

4.1 Two Models

We now formulate two models for the W_{ij}. The following model is a special case of Scheffe's (1956, 1959, pp. 261ff) model.

Model 1:

$$W_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} \tag{4.2}$$

where the β_j and γ_{ij} are random variables which are iid as j varies and for which

$$\left.\begin{aligned} &\text{(a)}\quad E(\beta_j) = 0 \\ &\text{(b)}\quad \operatorname{Var}(\beta_j) = \sigma_\beta^2 \\ &\text{(c)}\quad E(\gamma_{ij}) = 0 \\ &\text{(d)}\quad \operatorname{Var}(\gamma_{ij}) = \tfrac{1}{2}\sigma_\gamma^2 \\ &\text{(e)}\quad \gamma_{1j} + \gamma_{2j} = 0 \\ &\text{(f)}\quad \operatorname{Cov}(\beta_j, \gamma_{ij}) = 0 \end{aligned}\right\} \tag{4.3}$$

A more general version of Model 1 can be obtained by replacing (f) by

$$(f') \quad \operatorname{Cov}(\beta_j, \gamma_{ij}) \text{ constrained only by (4.3e)} \tag{4.4}$$

We will see that this more general version (which is Scheffe's model) is necessary to accommodate unequal variances. The notation in (4.3d) may seem unnatural, but it is used in most textbooks because it agrees with the results of the EMS algorithm. (The definition for general values of I is given in Section 6.) Some authors (e.g., Steel and Torrie 1980) choose a more natural notation.

A second model, given, for example, by Searle (1971, pp. 400–401) is the following.

Model 2:

$$W_{ij} = \mu + \tau_i + \tilde{\beta}_j + \tilde{\gamma}_{ij} \tag{4.5}$$

where the $\tilde{\beta}_j$ and $\tilde{\gamma}_{ij}$ are random variables which are iid as j varies and for which

$$\left.\begin{array}{ll} \text{(a)} & E(\tilde{\beta}_j) = 0 \\ \text{(b)} & \mathrm{Var}(\tilde{\beta}_j) = \sigma^2_{\tilde{\beta}} \\ \text{(c)} & E(\tilde{\gamma}_{ij}) = 0 \\ \text{(d)} & \mathrm{Var}(\tilde{\gamma}_{ij}) = \sigma^2_{\tilde{\gamma}} \\ \text{(e)} & \mathrm{Cov}(\tilde{\gamma}_{1j}, \tilde{\gamma}_{2j}) = 0 \\ \text{(f)} & \mathrm{Cov}(\tilde{\beta}_j, \tilde{\gamma}_{ij}) = 0 \end{array}\right\} \tag{4.6}$$

The key difference in the two models is in assumption (e) of (4.3) and (4.6). In Model 2, the terms $\tilde{\gamma}_{1j}$ and $\tilde{\gamma}_{2j}$ are uncorrelated while the analogous terms in Model 1 are perfectly negatively correlated because $\gamma_{1j} = -\gamma_{2j}$. Note that Model 1 allows the generalization (4.4), whereas replacement of (f) in (4.6) for Model 2 would require a change in assumption (e) also.

4.2 EMS Tables

For simplicity in the following developments, we will drop the subscript j and write the models as follows:

General:

$$\left.\begin{array}{l} Y_1 = W_1 + \epsilon_1 \\ Y_2 = W_2 + \epsilon_2 \end{array}\right\} \tag{4.7}$$

Model 1:

$$\left.\begin{array}{l} W_1 = \mu + \tau_1 + \beta + \gamma_1 \\ W_2 = \mu + \tau_2 + \beta + \gamma_2 \end{array}\right\} \tag{4.8}$$

Model 2:

$$\left.\begin{array}{l} W_1 = \mu + \tau_1 + \tilde{\beta} + \tilde{\gamma}_1 \\ W_2 = \mu + \tau_2 + \tilde{\beta} + \tilde{\gamma}_2 \end{array}\right\} \tag{4.9}$$

where we assume that $\{\beta, \gamma_1, \gamma_2\}$ satisfy the conditions (4.3) and that $\{\tilde{\beta}, \tilde{\gamma}_1, \tilde{\gamma}_2\}$ satisfy (4.6).

We can now express the EMSs of Table 3.1 in terms of the model parameters. From (4.7), (4.8) and (4.9) we note that for Model 1,

$$\left.\begin{aligned} Y_1 - Y_2 &= (\tau_1 - \tau_2) + (\gamma_1 - \gamma_2) + (\epsilon_1 - \epsilon_2) \\ Y_1 + Y_2 &= 2\mu + 2\beta + (\epsilon_1 + \epsilon_2) \end{aligned}\right\} \tag{4.10}$$

and for Model 2,

$$\left.\begin{aligned} Y_1 - Y_2 &= (\tau_1 - \tau_2) + (\tilde{\gamma}_1 - \tilde{\gamma}_2) + (\epsilon_1 - \epsilon_2) \\ Y_1 + Y_2 &= 2\mu + 2\tilde{\beta} + (\tilde{\gamma}_1 + \tilde{\gamma}_2) + (\epsilon_1 + \epsilon_2) \end{aligned}\right\} \tag{4.11}$$

Note that the key difference in these expressions is the absence of γ_i terms in the expression for $Y_1 + Y_2$ under Model 1. This is a consequence of the constraint in part (e) of (4.3). Applying the assumptions of each model to (4.10) and (4.11) yields immediately the EMSs given in Table 4.1. Note that these EMSs agree with the two versions in Table 2.2 for the case $I = 2,\ K = 1$.

Table 4.1 goes here

4.3 Relationships among Model Terms and Parameters

From the Model 1 assumptions, it follows that

$$\left.\begin{aligned} \mathrm{Var}(W_1) = \mathrm{Var}(W_2) &= \sigma_\beta^2 + \tfrac{1}{2}\sigma_\gamma^2 \\ \mathrm{Cov}(W_1, W_2) &= \sigma_\beta^2 - \tfrac{1}{2}\sigma_\gamma^2. \end{aligned}\right\} \tag{4.12}$$

This implies

$$\left.\begin{aligned} \mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) &= \sigma_\beta^2 + \tfrac{1}{2}\sigma_\gamma^2 + \sigma_\varepsilon^2 \\ \mathrm{Cov}(Y_1, Y_2) &= \sigma_\beta^2 - \tfrac{1}{2}\sigma_\gamma^2. \end{aligned}\right\} \tag{4.13}$$

In contrast, Model 2 gives

$$\left.\begin{aligned} \mathrm{Var}(W_1) = \mathrm{Var}(W_2) &= \sigma_{\tilde{\beta}}^2 + \sigma_{\tilde{\gamma}}^2 \\ \mathrm{Cov}(W_1, W_2) &= \sigma_{\tilde{\beta}}^2. \end{aligned}\right\} \tag{4.14}$$

This implies

$$\left.\begin{aligned}\operatorname{Var}(Y_1) = \operatorname{Var}(Y_2) &= \sigma^2_{\tilde\beta} + \sigma^2_{\tilde\gamma} + \sigma^2_{\varepsilon}\\ \operatorname{Cov}(Y_1, Y_2) &= \sigma^2_{\tilde\beta}.\end{aligned}\right\} \tag{4.15}$$

Note that Model 2 imposes the constraint

$$\operatorname{Cov}(Y_1, Y_2) \geq 0. \tag{4.16}$$

We will see in Section 4.4 that Model 1 does not constrain the covariance. Under the constraint (4.16), we can relate the models in the following way:

$$\left.\begin{aligned}\beta &= \tilde\beta + \tfrac{1}{2}(\tilde\gamma_1 + \tilde\gamma_2)\\ \gamma_1 &= \tfrac{1}{2}(\tilde\gamma_1 - \tilde\gamma_2)\end{aligned}\right\} \tag{4.17}$$

and

$$\left.\begin{aligned}\sigma^2_{\beta} &= \sigma^2_{\tilde\beta} + \tfrac{1}{2}\sigma^2_{\tilde\gamma}\\ \sigma^2_{\gamma} &= \sigma^2_{\tilde\gamma}\end{aligned}\right\} \tag{4.18}$$

The relations (4.17) and (4.18) are given by Searle (1971, pp. 403–404) and Hocking (1973).

4.4 Representations

If the joint distribution of (W_1, W_2) is such that it cannot be represented by either Model 1 or Model 2, then the relationships (4.17) and (4.18) are nonsense. For this reason, and also in the interest of better understanding, we now ask whether and how a given joint distribution can be represented in each model.

Given any joint distribution of (W_1, W_2) such that

$$\operatorname{Var}(W_1) = \operatorname{Var}(W_2), \tag{4.19}$$

it is immediate from (4.2) and (4.7) that a representation of (W_1, W_2) in terms of Model 1 can always be found, namely

$$\left.\begin{aligned}\beta &= \tfrac{1}{2}(W_1^* + W_2^*)\\ \gamma_i &= W_i^* - \beta\end{aligned}\right\} \tag{4.20}$$

where

$$W_i^* = W_i - E(W_i).$$

(Furthermore, even if (4.19) is violated the relations (4.20) define a representation in terms of the generalized version of Model 1, with (4.3f) replaced by (4.4).)

For representation by Model 2, it is necessary that the joint distribution of (W_1, W_2) satisfy not only (4.19) but also

$$\mathrm{Cov}(W_1, W_2) \geq 0. \tag{4.21}$$

In this case, a representation in terms of Model 2 can be constructed by introducing a random variable Z which is independent of W_1 and W_2. The construction is

$$\left.\begin{aligned} \tilde{\beta} &= c_1(W_1^* + W_2^*) + c_2 Z + c_3 \\ \tilde{\gamma}_i &= W_i^* - \tilde{\beta} \end{aligned}\right\} \tag{4.22}$$

where c_1, c_2 and c_3 depend on the first and second moments of (W_1, W_2) and Z.

If Model 1 is additive ($\sigma_\gamma^2 = 0$), then Model 2 is also additive, and in this case the two models coincide and are unique. Otherwise, the construction (4.22) of Model 2 is not unique because the (nondegenerate) random variable Z is arbitrary.

4.5 The Hypotheses

Consider the following three hypotheses, each of which in some sense expresses the assertion that there is no difference between the blocks.

$$H_0^{(1)}\colon \sigma_\beta^2 = 0$$

$$H_0^{(2)}\colon \sigma_\beta^2 = 0,\ \sigma_\gamma^2 = 0$$

$$H_0^{(3)}\colon \sigma_{\tilde{\beta}}^2 = 0$$

We now ask whether (assuming normality of the random variables) the above hypotheses can be tested. Consideration of the likelihood function easily confirms what is suggested by Table 4.1 – that the parameters σ_β^2 and σ_γ^2 are non–identifiable (and so is any linear combination of them), so that neither $H_0^{(1)}$ nor $H_0^{(2)}$ can be tested. On the other hand, if the assumptions of Model 2 are satisfied then the statistic MS(Blocks)/MS(T*B) yields a valid F test of $H_0^{(3)}$.

The preceding paragraph and the relations (4.17) and (4.18) provide an algebraic resolution of the paradox of the conflicting EMSs. To guide the user in deciding which hypothesis is of interest in a given case, we proceed to discuss statistical interpretations of the hypotheses.

Consider first the interpretations of the Model 1 hypotheses, $H_0^{(1)}$ and $H_0^{(2)}$. To interpret $H_0^{(1)}$, let

$$\overline{W} = \tfrac{1}{2}(W_1 + W_2); \tag{4.23}$$

then, because of (4.20), $H_0^{(1)}$ asserts that $\overline{W}$ has variance zero. The stronger hypothesis $H_0^{(2)}$ asserts that the random variables W_1 and W_2 each have variance zero. The following example illustrates these interpretations.

<u>Example 3.1</u>. Consider Example 1.3, where Y is neural density. The hypothesis $H_0^{(2)}$ asserts that all women in the population have the same neural density in the left hemisphere, and also that all women in the population have the same neural density in the right hemisphere. The hypothesis $H_0^{(1)}$ asserts that the average neural density in the two hemispheres ($\overline{W}$) is the same for all women in the population.

The hypothesis $H_0^{(2)}$ is equivalent to its analog in Model 2, that is, to the hypothesis

$$H_0^{(4)}: \sigma^2_{\tilde{\beta}} = 0, \quad \sigma^2_{\tilde{\gamma}} = 0.$$

By contrast, $H_0^{(1)}$ cannot be expressed in terms of the parameters of Model 2. (Confusingly, however, the relation (4.18) seems to suggest that $H_0^{(1)}$ is equivalent to $H_0^{(4)}$; this is an illusion, for if $H_0^{(1)}$ is true and $\sigma^2_{\gamma} > 0$, then it follows from (4.12) that (4.21) is violated and Model 2 cannot be constructed, so that the parameters $\sigma^2_{\tilde{\beta}}$ and $\sigma^2_{\tilde{\gamma}}$ are meaningless.)

Turning now to the interpretation of $H_0^{(3)}$, we note from (4.15) that $H_0^{(3)}$ is equivalent to

$$H_0: \mathrm{Cov}(Y_1, Y_2) = 0.$$

Recall that in Section 3.1 it was argued that this hypothesis is a natural interpretation of the idea of no effect due to blocks. In terms of Model 2 we can rewrite the relative efficiency λ of equation (3.1) as

$$\lambda = \frac{\sigma^2_{\tilde{\beta}} + \sigma^2_{\tilde{\gamma}} + \sigma^2_{\varepsilon}}{\sigma^2_{\tilde{\gamma}} + \sigma^2_{\varepsilon}} \tag{4.24}$$

which explicitly indicates that within Model 2 the hypothesis of no efficiency gain due to blocking is $H_0\colon \sigma^2_{\tilde{\beta}} = 0$. Note, however, that within Model 2 the alternative hypothesis must be one–sided, that is, $H_A\colon \sigma^2_{\tilde{\beta}} > 0$.

In summary, the block effect expressed by the parameter $\sigma^2_{\tilde{\beta}}$ of Model 2 and tested by the F statistic MS(Blocks)/MS(T∗B) corresponds to the approach presented in Section 3.

5. Unequal Variances

In many ANOVA settings, the assumption of homogeneity of variance might be regarded as a necessary evil. However, in the present context of a mixed model with $I = 2$, it is not necessary, and the data analyst could (and, we will argue, sometimes should) consider dropping it. Note that the generalized version of Model 1, using (4.4), can accommodate unequal variances.

5.1 Test for Treatment Difference

Thus far we have ignored the test of treatment difference. The usual F test of $H_0\colon \tau_1 = \tau_2$, based on the ratio of MS(Treatments) to MS(T∗B), is the square of the t statistic for paired samples. Thus, the test of treatment difference is valid even if the variances of Y_1 and Y_2 are not equal.

5.2 Test for Blocks

If $\mathrm{Var}(Y_1) \neq \mathrm{Var}(Y_2)$ then the F test for blocks discussed in Sections 3 and 4 is not valid. The argument leading to (3.6) fails because, although MS(Blocks) and MS(T∗B) have the same expectation under the null hypothesis (from (3.4)), without the equal variance assumption these mean squares are no longer independent (from (3.5)). Thus, generalizing Model 2 to accommodate unequal variances would be undesirable, because the model would then suggest an invalid F test for blocks.

On the other hand, when $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$ are unconstrained, the test for blocks based on the Pearson correlation given by (3.8) is valid (and is the LRT) for testing H_0: $\mathrm{Cov}(Y_1, Y_2) = 0$.

5.3 Confidence Interval for $\tilde{\rho}$

Recall from section 3.1 that the relative efficiency of the blocked design compared to the nonblocked design is given by

$$\lambda = \frac{1}{1-\tilde{\rho}}$$

where

$$\tilde{\rho} = \frac{\mathrm{Cov}(Y_1, Y_2)}{\frac{1}{2}[\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2)]}$$

The problem of constructing a confidence interval for $\tilde{\rho}$ when $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$ are unconstrained was solved by Kristof (1972), who showed that

$$A \, \frac{\tilde{r} - \tilde{\rho}}{\sqrt{1-\tilde{\rho}^2}} \sim t_{J-2} \; ,$$

where

$$A = \frac{r}{\tilde{r}}\sqrt{\frac{J-2}{1-r^2}}$$

and $\tilde{r}$ is given by (3.9) and r by (3.7). It follows that $100(1-\alpha)\%$ confidence limits for $\tilde{\rho}$ are given by

$$\frac{\tilde{r} \pm B\sqrt{1-\tilde{r}^2+B^2}}{1+B^2} \; , \qquad (5.1)$$

where $B = A^{-1}t_{1-\alpha/2;J-2}$ and $t_{p;n}$ is the pth percentile of a t_n distribution. The confidence limits for λ are therefore $1/(1-L)$ and $1/(1-U)$ where L and U are the lower and upper confidence limits for $\tilde{\rho}$.

5.4 Examples

The assumption of equal variances is not as innocuous as may at first appear. In the two–way mixed model, the assumption has substantive force because it stringently limits the kinds of interactions that are permitted. Specifically, it follows from the definition of Model 1 ((4.3) and (4.4)) that $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2)$ if and only if

$$\mathrm{Cov}(\gamma_1, \beta) = 0 \qquad (5.2)$$

This condition says that the interaction term is uncorrelated with the term representing the main effect of blocks. The condition is perhaps more intuitively meaningful if it is expressed as

$$\text{Cov}(W_1 - W_2, \overline{W}) = 0, \tag{5.3}$$

where $\overline{W} = \frac{1}{2}(W_1 + W_2)$. This condition asserts that the difference between responses is uncorrelated with the average response.

The following two examples illustrate the limitations of the equal variance constraint.

Example 5.1. As in Example 1.2, suppose Y_1 and Y_2 are measured skin thickness (and W_1 and W_2 are "true" skin thickness) in the two differently treated arms of each subject. In the source population, assume that women differ in their overall skin thickness, as follows:

50% have "thick" skin, defined as $\overline{W} = 150$

50% have "thin" skin, defined as $\overline{W} = 100$

Assume further that women differ in their "sensitivity" to the difference between Treatment 1 and Treatment 2:

50% are "sensitive", defined as $W_1 - W_2 = 40$

50% are "insensitive", defined as $W_1 - W_2 = 0$

Will such a population satisfy the assumption of equal variances? The answer is Yes if sensitivity is distributed independently of overall thickness, and No if it is not. Table 5.1 shows the joint distribution of sensitivity and overall thickness for two hypothetical populations, A and B. Population A would satisfy (5.3) but Population B would not.

A common source of violation of the equal–variance assumption is a multiplicative treatment effect, as in the following example.

Example 5.2. As in Example 1.3, suppose Y_1 and Y_2 are measured neural density in the right and left hemispheres, and assume that the true neural density is 10% higher in the right than in the left hemisphere, so that

$$W_1 = 1.1W_2$$

Clearly, (5.3) would be violated in this situation.

The foregoing examples show that the case of unequal variances can arise quite naturally in practice, and may be necessary to adequately model many situations. If a flexible model is required, where variances and, hence, interactions are not restricted, then the t test based on (3.8) is the only way to test the hypothesis of no block effect, $H_0\colon \mathrm{Cov}(Y_1, Y_2) = 0$, and (5.1) should be used if an interval estimate of the intrablock correlation $\tilde{\rho}$ is desired.

5.5 Testing for Equality of Variances

The hypothesis

$$H_0\colon \mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) \tag{5.4}$$

is of interest in a blocked design for two reasons. First, if this hypothesis is false then blocks should not be tested with the ANOVA test (3.6), but rather with the Pearson correlation test (3.8). Second, as illustrated in Examples 5.1 and 5.2, falsity of (5.4) may indicate the presence of a certain kind of interaction – namely, interaction correlated with the block effect.

Using equation (3.5), the homogeneity of variance hypothesis (5.4) is equivalent to

$$H_0\colon \mathrm{Cov}(Y_1 + Y_2,\ Y_1 - Y_2) = 0.$$

Under this null hypothesis,

$$r'\sqrt{\frac{J-2}{1-r'^2}} \sim t_{J-2}\ , \tag{5.5}$$

where r' is the observed Pearson correlation between $Y_1 + Y_2$ and $Y_1 - Y_2$.

The test based on (5.5) was suggested by Pitman (1939) and also by Morgan (1939), who showed that it is the LRT of (5.4) when the alternative hypothesis is two-sided. Pitman also showed how to obtain a confidence interval for the ratio $\mathrm{Var}(Y_1)/\mathrm{Var}(Y_2)$ (see also Kendall and Stuart, 1979, Chap. 20).

6. More than Two Treatments

The discussion in Sections 3, 4 and 5 was limited to the case of two treatments ($I = 2$). The models discussed in Section 4 extend in a natural way to the case of arbitrary I. The decomposition (4.1) extends immediately and the noncentrality parameter is

$$\phi_\tau = \frac{1}{I-1}\Sigma\tau_i^2.$$

Models 1 and 2 can be defined by (4.2)—(4.6) for general I, with conditions (4.3)(d) and (e) replaced by

$$\text{Var}(\gamma_{ij}) = \left(\frac{I-1}{I}\right)\sigma_\gamma^2$$

and

$$\Sigma_i\gamma_{ij} = 0$$

respectively.

For these generalized models, the same discrepancy occurs as for $I = 2$; that is, the EMSs for blocks disagree because the parameter $\sigma_{\tilde{\beta}}^2$ is not the same as the parameter σ_β^2.

We turn now to the question of testing blocks when $I > 2$ and $K = 1$. Suppose first that the covariance matrix of $(Y_1, \ldots, Y_I)$ is compound symmetric – that is, that the Y_i have the same variance and the same pairwise correlation

$$\tilde{\rho} = \rho = \text{corr}(Y_i, Y_{i'}) \quad i \neq i' \tag{6.1}$$

(The assumptions of Model 2 imply not only compound symmetry but also $\tilde{\rho} \geq 0$). Then the results given in Sections 3 and 4 for the case $I = 2$ extend naturally. The F test for treatments is valid. The hypothesis $H_0\colon \sigma_\beta^2 = 0$ cannot be tested. What can be tested by an F test is the hypothesis $H_0\colon \sigma_{\tilde{\beta}}^2 = 0$ which is equivalent to $H_0\colon \tilde{\rho} = 0$. The efficiency gain due to blocking is

$$\lambda = \frac{1}{1-\tilde{\rho}}$$

and the maximum likelihood estimator of $\tilde{\rho}$ is

$$\tilde{r} = \frac{\sum\limits_{i\neq i'}\sum\limits_{j}(Y_{ij} - \overline{Y}_{i\cdot})(Y_{i'j} - \overline{Y}_{i'\cdot})}{(I-1)\sum\limits_{i}\sum\limits_{j}(Y_{ij} - \overline{Y}_{i\cdot})^2}. \tag{6.2}$$

Note that $\tilde{r}$, which is the direct extension of (3.9), can be interpreted as the average observed covariance divided by the average observed variance. In Appendix 3 a computational formula for $\tilde{r}$ and a confidence interval for $\tilde{\rho}$ are given.

If the assumption of compound symmetry is dropped, the considerations of the proceeding sections do not immediately extend. In the case of an arbitrary covariance matrix with $I > 2$, the validity of the F test for treatments becomes questionable. We will not pursue this case any further here.

7. The Case of Replication ($K > 1$)

The case $K > 1$ represents replicate measurements within the (i,j)th block–treatment combination. For instance, in Example 1.2 one might make K independent measurements on each forearm of each subject. Another example is the generalized randomized blocks design in which each block contains IK experimental units which are randomly allocated to the I treatments.

As in Section 2, we let Y_{ijk} represent the kth observation on the ith treatment in the jth block; we decompose Y_{ijk} as

$$Y_{ijk} = W_{ij} + \varepsilon_{ijk} \tag{7.1}$$

where the random vectors $(W_{1j}, \ldots, W_{Ij})$ are iid as j varies, and the ε_{ijk} are iid random variables with mean 0 and variance σ_ε^2 which are uncorrelated with the W_{ij}.

The models for W_{ij} discussed in the preceding sections carry over unchanged to the present case. As in the case $K = 1$, there are two seemingly similar hypotheses which actually address different questions. The hypothesis $H_0\colon \sigma_{\tilde{\beta}}^2 = 0$ asserts that blocking has no effect on the efficiency of treatment comparisons, whereas the hypothesis $H_0\colon \sigma_{\bar{\beta}}^2 = 0$ asserts that the "true" response W, averaged over treatments, is the same for all blocks.

The entire discussion in Sections 3–6 of testing, estimation, and interpretation of $\sigma_{\tilde{\beta}}^2$ and $\tilde{\rho}$ can be readily carried over to the present case by identifying Y_{ij} of the previous discussion with $\overline{Y}_{ij\cdot}$ of the present case.

In contrast with the previous discussion, as suggested by the Version 1 EMSs in Table 2.2 the hypothesis $H_0: \sigma^2_\beta = 0$ can be tested when $K > 1$. The F statistic for this test is the ratio of MS(Blocks) to MS(Within T*B), and is, of course, not the same as that used to test $H_0: \sigma^2_{\tilde{\beta}} = 0$.

8. Summary and Discussion

We have seen that, even when a blocked design is modeled as a mixed model, the question of testing blocks is not identical to the question of testing the random factor. The difference arises because "testing the random factor" has two different meanings corresponding to two different parametrizations of the model.

Model 1, which agrees with the mixed–model ANOVAs given in most textbooks, expresses the hypothesis of no effect of the random factor as

$$H_0: \sigma^2_\beta = 0 \tag{8.1}$$

The hypothesis (8.1), which is the direct analog of the hypothesis of no main effect in a fixed–effects ANOVA, can be tested by the F ratio MS(Blocks)/MS(Within T*B); the test requires within–block replication ($K > 1$). In a discussion of mixed models, Kempthorne (1975) writes:

> Is there a case for testing the main effect of the random factor? I think it is hard to make one ...

We agree with this statement in reference to the hypothesis (8.1), which is what Kempthorne had in mind. Cases like Example 1.3, where (8.1) is a natural hypothesis, probably are rather rare. [The parameter σ^2_β may, however, be of interest in animal genetics, where mixed models are used which incorporate interaction between environment (fixed) and genotype (random); see, for instance, Muir (1985). There appears to be some controversy concerning the choice of models in this context (Fernando et al., 1984; Yamada and Sugimoto, 1988); because the choice depends on specific genetic considerations, we do not consider it here.]

On the other hand, in the setting of a blocked design, where the blocks are of interest only insofar as they can enhance the comparison of the treatments, it is natural to express the hypothesis of no block effect, not as (8.1), but rather as

$$H_0\colon \tilde{\rho} = 0 \tag{8.2}$$

where $\tilde{\rho}$ is the intrablock correlation. The hypothesis (8.2) implies that blocking has no effect on the variance of inter–treatment contrasts; the efficiency gain due to blocking is equal to $1/(1-\tilde{\rho})$. For instance, in the case considered by Kempthorne (1975) of locations (random) and varieties (fixed) of corn, the quantity $1/(1-\tilde{\rho})$ expresses the effect of blocking by location on the efficiency of comparisons among varieties. (In a field setting, the effect of locations is usually large, but in a greenhouse it may be rather small.)

Within Model 2, the hypothesis (8.2) can be expressed, in analogy to (8.1), as

$$H_0\colon \sigma^2_{\tilde{\beta}} = 0. \tag{8.3}$$

However, (8.2) is of interest not only within Model 2 but also in two cases where the constraints implied by Model 2 — compound symmetry of the covariance matrix and nonnegative $\tilde{\rho}$ — are violated. First, if compound symmetry holds then the efficiency gain due to blocking is equal to $1/(1-\tilde{\rho})$ even if $\tilde{\rho}$ is negative. Second, in the case of two treatments ($I = 2$) the efficiency gain due to blocking is equal to $1/(1-\tilde{\rho})$ even if $\tilde{\rho}$ is negative or compound symmetry fails (that is, $\mathrm{Var}(Y_1) \neq \mathrm{Var}(Y_2)$).

We have considered two tests of (8.2) under the assumption of normality. First, if the covariance matrix is compound symmetric, then the F test of (8.2) based on MS(Blocks)/MS(T∗B) is valid. The second test arises in the case $I = 2$; then, even if the treatment variances are unequal, the hypothesis (8.2) can be tested by a t test based on the Pearson correlation coefficient. Corresponding to the two tests of (8.2) in the case $I = 2$, two different confidence intervals for $\tilde{\rho}$ can be constructed.

The question of whether $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2)$ has played a nontrivial role in our development for the case $I = 2$, and we have indicated in Section 5.4 how to test for equality of variances. From a practical point of view, it should be noted that inequality of variances may also indicate that the basic linear model (4.1) is inappropriate, and/or that the usual t test for a treatment difference would lack power. In some situations a logarithmic transformation, corresponding to a multiplicative model, may yield improved power.

Many discussions of the randomized blocks design assume that the treatment–block interaction is zero. Such an additivity assumption implies that the conditions of Model 2 are satisfied and that the hypotheses (8.1) and (8.2) are equivalent. In our opinion, the additivity assumption is unnecessarily restrictive for investigations in which blocks are regarded as random. Indeed, the notion that a treatment can be represented by the average response over a conceptual population of people, plots of ground, or other units is fundamental to statistical thinking. For this reason it appears to us to be useful to develop a view of blocked designs which does not depend on the additivity assumption.

We have not dealt with the case where an experimenter prefers to regard blocks as a fixed factor rather than as a random factor. In this case, the EMS table is unambiguous: the denominator for the F test on blocks must be MS(Within T*B) unless additivity of treatments and blocks is assumed, in which case the pooled value of MS(T*B) and MS(Within T*B) can be used instead.

How do our conclusions about testing blocks reflect on the merits of Models 1 and 2? The strengths of Model 1 are its explicitness (see Section 4.4) and its great generality; however, Model 1 is heuristically misleading for a blocked design because it suggests an inappropriate F ratio for blocks. On the other hand, while Model 2 suggests the appropriate F ratio, this model has serious limitations. First, as discussed above, Model 2 implies equal treatment variances, a constraint which is unnecessary for $I = 2$. Second, Model 2 constrains intrablock correlations to be nonnegative, a constraint which is irrelevant to the question of testing blocks. If the correlation is indeed negative, then blocking decreases, rather than increases, the efficiency of treatment comparisons. This may sometimes happen in practice; for instance, Snedecor and Cochran (1980, p. 244) note that competition between animals in a pen may produce negative intrablock correlations. Hocking (1985, p. 334) suggests weakening Model 2 to allow negative correlation; however, this modification would vitiate the heuristic value of the model, because the term $\tilde{\beta}_j$ appearing in the model equation would have a negative variance! From the foregoing, it would appear that neither Model 1 nor Model 2 is entirely satisfactory for representing blocked designs in the presence of interaction. For mixed models in which the random factor does *not* represent blocks, the choice between models may depend on the context of the investigation. Note, however, that in some applications what is wanted is a variance components model, which (as indicated in Appendix 1) is different from either Model 1 or Model 2.

References

ADDELMAN, SIDNEY (1970), "Variability of Treatments and Experimental Units in the Design and Analysis of Experiments," *Journal of the American Statistical Association*, **65**, 1095–1108.

ANDERSON, VIRGIL L. (1970), "Restriction Errors for Linear Models (An Aid to Develop Models for Designed Experiments)," *Biometrics*, **26**, 255–268.

ANDERSON, VIRGIL L. and McLEAN, ROBERT A. (1974), *Design of Experiments, A Realistic Approach*, New York: Marcel Dekker.

CORNFIELD, JEROME and TUKEY, JOHN W. (1956), "Average Values of Mean Squares in Factorials," *Annals of Mathematical Statistics*, **27**, 907–949.

FERNANDO, R. L., KNIGHTS, S. A. and GIANOLA, D. (1984), "On a Method of Estimating the Genetic Correlation between Characters Measured in Different Experimental Units," *Theoretical and Applied Genetics*, **67**, 175–178.

HARTLEY, H. O. and SEARLE, S. R. (1969), "A Discontinuity in Mixed Model Analysis," *Biometrics*, **25**, 573–576.

HARVILLE, DAVID A. (1978), "Alternative Formulations and Procedures for the Two–Way Mixed Model," *Biometrics*, **34**, 441–453.

HICKS, CHARLES R. (1982), *Fundamental Concepts in the Design of Experiments* (3rd ed.), New York: Holt, Rinehart and Winston.

HOCKING, R. R. (1973), "A Discussion of the Two–Way Mixed Model," *American Statistician*, **27**, 148–154.

HOCKING, R. R. (1985), *The Analysis of Linear Models*, Monterey, California: Brooks/Cole.

KEMPTHORNE, OSCAR (1952), *The Design and Analysis of Experiments*, New York: John Wiley.

KEMPTHORNE, OSCAR (1975), "Fixed and Mixed Models in the Analysis of Variance," *Biometrics*, **31**, 473–486.

KENDALL, M. and STUART, A. (1979), *The Advanced Theory of Statistics, Volume II*, London: Chapman and Hall.

KIRK, ROGER E. (1982), *Experimental Design: Procedures for the Behavioral Sciences* (2nd ed.), Monterey, California: Brooks/Cole.

KRISTOF, WALTER (1963), "The Statistical Theory of Stepped–up Reliability Coefficients When a Test Has Been Divided into Several Equivalent Parts," *Psychometrika*, **28**, 221–238.

KRISTOF, WALTER (1972), "On a Statistic Arising in Testing Correlation," *Psychometrika*, **37**, 377–384.

MEHTA, J. S. and GURLAND, JOHN (1969), "Some Properties and an Application of a Statistic Arising in Testing Correlation," *Annals of Mathematical Statistics*, **40**, 1736–1745.

MONTGOMERY, DOUGLAS C. (1984), *Design and Analysis of Experiments* (2nd ed.), New York: John Wiley.

MORGAN, W. A. (1939), "A Test for the Significance of the Difference between the Two Variances in a Sample from a Normal Bivariate Population," *Biometrika*, **31**, 13–19.

MUIR, WILLIAM M. (1985), "Relative Efficiency of Selection for Performance of Birds Housed in Colony Cages Based on Production in Single Bird Cages," *Poultry Science*, **64**, 2239–2247.

PITMAN, E. J. G. (1939), "A Note on Normal Correlation," *Biometrika*, **31**, 9–12.

SCHEFFE, HENRY (1956). "A 'Mixed Model' for the Analysis of Variance," *Annals of Mathematical Statistics* **27**, 23–36.

SCHEFFE, HENRY (1959). *The Analysis of Variance*, New York: John Wiley.

SEARLE, S. R. (1971a), "Topics in Variance Component Estimation," *Biometrics*, **27**, 1–76.

SEARLE, S. R. (1971b), *Linear Models*, New York: John Wiley.

SNEDECOR, GEORGE W. and COCHRAN, WILLIAM G. (1980). *Statistical Methods* (7th ed.), Ames, Iowa: The Iowa State University Press.

STEEL, ROBERT G. D. and TORRIE, JAMES H. (1980), *Principles and Procedures of Statistics* (2nd ed.), New York: McGraw–Hill.

WHITE, ROBERT F. (1975), "Randomization and the Analysis of Variance," *Biometrics* **31**, 555–572.

WILK, M. B. and KEMPTHORNE, O. (1955), "Fixed, Mixed, and Random Models," *Journal of the American Statistical Association*, **50**, 1144–1167.

WILK, M. B. and KEMPTHORNE, O. (1956), "Some Aspects of the Analysis of Factorial Experiments in a Completely Randomized Design," *Annals of Mathematical Statistics*, **27**, 950–985.

WINER, B. J. (1971), *Statistical Principles in Experimental Design* (2nd ed.), New York: McGraw–Hill.

YAMADA, YUKIO and SUGIMOTO, ISAO (1988), "Parametric Relationship between GE Interaction and Genetic Correlation," *Theoretical and Applied Genetics*, to appear.

APPENDIX 1. A COMPONENTS OF VARIANCE MODEL

In Section 2 it was stated that if the EMS algorithm is applied to the model

$$Y_{ijkl} = \mu + T_i + B_j + G_{l(ij)} + e_{(ijl)k} \tag{A1}$$

then setting $l = 1$ yields the Version 2 EMSs of Table 2.2. However, as we now show, the model (A1), as usually applied, is quite different from those considered in the body of the paper. Our concern has been to model interaction between factors T and B. By contrast, in the usual application of (A1) the G term represents, not interaction, but a variance component. The following is a typical application from animal breeding.

Example A1. In equation A1, define

$$\mu = \text{grand mean}$$

$$T_i = \text{effect of herd } i$$

$$B_j = \text{effect of year } j$$

$$G_{l(ij)} = \text{effect of sire } l$$

$$Y_{ijkl} = \text{milk yield of } k\text{th daughter cow of } l\text{th sire in } i\text{th herd in the } j\text{th year.}$$

Suppose the herd effect T_i is regarded as fixed, the year effect B_j is regarded as random, and the sire effect $G_{l(ij)}$ is regarded as random.

In an application like Example A1, the G term represents a variance component. The difference between a variance component and an interaction can easily be appreciated heuristically as follows. If the model were modified so that years were regarded as fixed rather than as random, then the G term, which models variability inherent in sires, would still be random. In other words, the G term is random *regardless* of whether the B term is random. By contrast, in interaction models such as those considered in this paper, the G term is random *because* the B term in random.

More precisely, in the variance components model the sire effects $G_{l(ij)}$ are random variables which are not only mutually independent and independent of the B_j, but furthermore are *conditionally* independent given the year. The situation is quite different for the models we have considered: conditional on blocks (that is, on W_{ij}), the γ terms are constant in our Model 1 and the $\tilde{\gamma}$ terms are perfectly correlated (because of (4.22)) in our Model 2.

Note that, with the above variance components interpretation, the model (A1) is *additive* with respect to the effects T and B, in the sense that, for herd i and year j, the average yield of many offspring of many sires would be the sum of a component depending only on i and a component depending only on j. In fact, the model could be generalized by adding another random term, say H_{ij}, to represent the interaction between herd and year. The component $(\mu + T_i + B_j + H_{ij})$ would then be analogous to our W_{ij}, and could be represented by either our Model 1 or Model 2.

APPENDIX 2. THEORETICAL CONSIDERATIONS

The properties of the tests of variances and covariances which are discussed in Sections 3, 4 and 5 are closely linked because, for bivariate normal samples, a duality exists between tests of variance and of covariance. Suppose that (X_{1j}, X_{2j}), $j = 1, \ldots, J$, are a random sample from a bivariate normal population with parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_X)$. Define

$$Y_{1j} = X_{1j} + X_{2j}$$
$$Y_{2j} = X_{1j} - X_{2j}.$$

Then the pairs (Y_{1j}, Y_{2j}), $j = 1, \ldots, J$, are also a random sample from a bivariate normal population with parameters $(\mu_1', \mu_2', \sigma_1'^2, \sigma_2'^2, \rho_Y)$. Furthermore, since

$$\begin{aligned}\text{Cov}(Y_{1j}, Y_{2j}) &= \text{Var}\,(X_{1j}) - \text{Var}\,(X_{2j}) \\ &= \sigma_1^2 - \sigma_2^2,\end{aligned}$$

it is clear that the hypotheses

$$H_0\colon \sigma_1^2 = \sigma_2^2 \quad \text{and} \quad H_0'\colon \rho_Y = 0$$

are equivalent, that is, H_0 is true if, and only if, H_0' is true. Thus, testing equality of variance is equivalent to testing lack of correlation in normal samples. (This is one of the facts exploited by Pitman, 1939).

The correspondence between the hypotheses carries over to the likelihood ratio tests (LRTs) for the hypotheses, whether or not the other parameters are restricted. Consider the pair of hypotheses

$$\begin{array}{lcl} H_0\colon \sigma_1^2 = \sigma_2^2, \rho_X = 0 & & H_0'\colon \rho_Y = 0,\ \sigma_1'^2 = \sigma_2'^2 \\ & \text{and} & \\ H_1\colon \sigma_1^2 \neq \sigma_2^2, \rho_X = 0 & & H_1'\colon \rho_Y \neq 0,\ \sigma_1'^2 = \sigma_2'^2 \end{array}$$

The LRT of H_0 vs. H_1 is based on the ratio of sums of squares. Using the relationship

$$\frac{\sum(X_{1j} - \overline{X}_1)^2}{\sum(X_{2j} - \overline{X}_2)^2} = \frac{\sum(Y_{1j} - \overline{Y}_1)^2 + \sum(Y_{2j} - \overline{Y}_{2\cdot})^2 + 2\sum(Y_{1j} - \overline{Y}_{1\cdot})(Y_{2j} - \overline{Y}_{2\cdot})}{\sum(Y_{1j} - \overline{Y}_1)^2 + \sum(Y_{2j} - \overline{Y}_{2\cdot})^2 - 2\sum(Y_{1j} - \overline{Y}_{1\cdot})(Y_{2j} - \overline{Y}_{2\cdot})},$$

and the fact that the LRT of H_0' vs. H_1' is based on the righthand side of the above equality, we see that the tests are equivalent.

There is also a duality between the more general pair of hypotheses

$$H_0: \sigma_1^2 = \sigma_2^2 \qquad H_0': \rho_Y = 0$$
and
$$H_1: \sigma_1^2 \neq \sigma_2^2 \qquad H_1': \rho_Y \neq 0$$

where any unmentioned parameter is left unspecified. The LRT of H_0 vs. H_1, with μ_1, μ_2 and ρ_X unspecified, was first given by Morgan (1939). An argument similar to the one above shows that this LRT is equivalent to the LRT for testing H_0' vs. H_1' with the variances unspecified.

In terms of an ANOVA, the usual F test applied to blocks, based on (3.6), is the LRT of $H_0: \sigma_1^2 = \sigma_2^2$ under the constraint $\rho_X = 0$. The correlation t test based on (3.8) is the LRT of $H_0': \rho_Y = 0$ with $\sigma_1'^2, \sigma_2'^2$ unspecified. Thus, the ANOVA F test can be regarded as a restricted version of the correlation t test.

It can be shown that if the restriction holds ($\rho_X = 0$), then the F test is more powerful; the argument is based on the fact that the correlation t test is not based on a minimal sufficient statistic if $\rho_X = 0$. If the restriction is not satisfied, then the ANOVA F test is not valid and no comparison can be made.

APPENDIX 3. DETAILS ON $\tilde{r}$

The coefficient $\tilde{r}$ defined in (6.2) can be written in terms of the ANOVA mean squares as (see Hocking 1985, p. 325)

$$\tilde{r} = \frac{\text{MS(Blocks)} - \text{MS(T}*\text{B)}}{\text{MS(Blocks)} + (I-1)\text{MS(T}*\text{B)}},$$

an extension of (3.10).

In Model 2, $\tilde{r}$ can be interpreted as

$$\tilde{r} = \frac{\hat{\sigma}^2_{\tilde{\beta}}}{\hat{\sigma}^2_{\tilde{\beta}} + \hat{\sigma}^2_{\tilde{\gamma}} + \hat{\sigma}^2_{\varepsilon}},$$

where $\hat{\sigma}^2_{\tilde{\beta}}$ and $\hat{\sigma}^2_{\tilde{\gamma}} + \hat{\sigma}^2_{\varepsilon}$ are the unbiased estimates of $\sigma^2_{\tilde{\beta}}$ and $\sigma^2_{\tilde{\gamma}} + \sigma^2_{\varepsilon}$ derived from the EMSs.

Using results of Kristof (1963), under the compound symmetry assumption $100(1-\alpha)\%$ confidence limits for $\tilde{\rho}$ can be written as

$$\frac{1+(I-1)\tilde{r}-(1-\tilde{r})F_{1-\alpha/2;J-1,(I-1)(J-1)}}{1+(I-1)\tilde{r}+(I-1)(1-\tilde{r})F_{1-\alpha/2;J-1,(I-1)(J-1)}},$$

$$\frac{1+(I-1)\tilde{r}-(1-\tilde{r})F_{\alpha/2;J-1,(I-1)(J-1)}}{1+(I-1)\tilde{r}+(I-1)(1-\tilde{r})F_{\alpha/2;J-1,(I-1)(J-1)}}$$

This interval is an extension of (3.11) and, as noted in Section 3, is different from the interval more commonly given for intraclass correlation because it is based on a different experimental design.

Table 2.1. Orthogonal Decomposition of the Total Sum of Squares

Source	df	SS
Treatments	$I-1$	$JK\sum_i(\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot\cdot\cdot})^2$
Blocks	$J-1$	$IK\sum_j(\overline{Y}_{\cdot j\cdot}-\overline{Y}_{\cdot\cdot\cdot})^2$
T*B	$(I-1)(J-1)$	$K\sum_i\sum_j(\overline{Y}_{ij\cdot}-\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot j\cdot}+\overline{Y}_{\cdot\cdot\cdot})^2$
Within T*B	$IJ(K-1)$	$\sum_i\sum_j\sum_k(Y_{ijk}-\overline{Y}_{ij\cdot})^2$
Total	$IJK-1$	$\sum_i\sum_j\sum_k(Y_{ijk}-\overline{Y}_{\cdot\cdot\cdot})^2$

Table 2.2: Two Versions of Expected Mean Squares for Mixed Model with Treatments Fixed, Blocks Random

	Expected Mean Squares	
	Version 1	Version 2
Treatments	$JK\phi_T + K\sigma_G^2 + \sigma_e^2$	$JK\phi_T + K\sigma_G^2 + \sigma_e^2$
Blocks	$IK\sigma_B^2 + \sigma_e^2$	$IK\sigma_B^2 + K\sigma_G^2 + \sigma_e^2$
T*B	$K\sigma_G^2 + \sigma_e^2$	$K\sigma_G^2 + \sigma_e^2$
Within T*B	σ_e^2	σ_e^2

Table 3.1. ANOVA Table for $I = 2$ and $K = 1$

Source	SS	df	EMS
Treatments	$\frac{1}{2}J(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot})^2$	1	$J\phi_\tau + \frac{1}{2}\mathrm{Var}(Y_1 - Y_2)$
Blocks	$2\sum_j(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})^2$	$J-1$	$\frac{1}{2}\mathrm{Var}(Y_1 + Y_2)$
T*B	$\frac{1}{2}\sum_j[(Y_{1j} - Y_{2j}) - (\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot})]^2$	$J-1$	$\frac{1}{2}\mathrm{Var}(Y_1 - Y_2)$

Table 4.1. EMSs for Model 1 and Model 2 when $I = 2$ and $K = 1$

Source	EMS	
	Model 1	Model 2
Treatments	$J\phi_\tau + \sigma_\gamma^2 + \sigma_\epsilon^2$	$J\phi_\tau + \sigma_{\tilde{\gamma}}^2 + \sigma_\epsilon^2$
Blocks	$2\sigma_\beta^2 + \sigma_\epsilon^2$	$2\sigma_{\tilde{\beta}}^2 + \sigma_{\tilde{\gamma}}^2 + \sigma_\epsilon^2$
T∗B	$\sigma_\gamma^2 + \sigma_\epsilon^2$	$\sigma_{\tilde{\gamma}}^2 + \sigma_\epsilon^2$

Table 5.1. Two Populations for Example 5.1.

Overall thickness	Sensitivity	W_1	W_2	Relative Frequency Pop. A	Relative Frequency Pop. B
Thin	Insensitive	100	100	.5	.1
Thick	Insensitive	150	150	.5	.4
Thin	Sensitive	120	80	.5	.4
Thick	Sensitive	170	130	.5	.1