The Efficiency of Blocking: How to Use MS(Blocks)/MS(Error) Correctly

MARVIN LENTNER, JESSE C. ARNOLD, and KLAUS HINKELMANN*

Even though there are no valid tests of block effects in randomized complete block and Latin square experiments, it is noted that a commonly used measure of efficiency is monotonically related to the F ratios used inappropriately for testing the effectiveness of blocking. Because of this relationship one can give beginning students a useful interpretation of otherwise inappropriate F statistics without introducing concepts of relative efficiency.

KEY WORDS: Block effects; Latin square design; Randomized complete block design; Relative efficiency.

1. INTRODUCTION

In many experiments, the experimental units are arranged in homogeneous sets called blocks. The objective of blocking is to account for the effects of nuisance factors that characterize the experimental material. Although units within blocks are to be as homogeneous as possible, units in different blocks are expected to be heterogeneous. As these two conditions approach their "ideal" states, blocking becomes very effective and yields "better" inference with respect to treatment effects than one would get from an experimental design without blocking. On the other hand, if blocking has been used and units in different blocks are not more heterogeneous than within blocks, then inference is not as good as one would have obtained from a design without blocking. Thus the researcher may wish to know, after the experiment has been done, whether blocking of experimental units was worthwhile, especially if these or similar units might be used in future experiments.

Measures of relative efficiency should be used to compare an experimental design where blocking is present with a similar experiment without the blocking in question. The relative efficiency (RE) of design D_1 compared with design D_2 is defined as follows:

$$\text{RE } (D_1 \text{ compared with } D_2) = \frac{\text{efficiency } D_1}{\text{efficiency } D_2} = \frac{\text{variance } D_2}{\text{variance } D_1}, \qquad (1)$$

where variance D refers to the variance of a treatment comparison using design D.

Two widely used experimental designs having blocks are the randomized complete block design (RCB) and latin square design (LS). We shall show that the estimated relative efficiencies (ERE's) used to evaluate these designs are monotonically related to the "F" ratios, which are frequently and inappropriately used for testing the effectiveness of blocking. This provides a useful way of interpreting the "F" ratios correctly without deriving the expressions for the ERE's.

In Section 2 we give a brief review of the notion and properties of block designs and two-way classifications, mainly to contrast these two situations and point out why they are different. In Section 3 we use these results for the RCB and give an alternate expression for the ERE of an RCB compared with a completely randomized design (CRD). The same argument is used in Section 4 for the ERE of an LS compared with an RCB.

2. BASIC IDEAS OF TWO-WAY CLASSIFICATION AND BLOCK DESIGNS

In many textbooks the concepts of experimental design, linear models, and analysis of variance are used almost synonymously. For example, the RCB is often equated with a two-way classification. As a result, a common linear model and analysis of variance are used to analyze data arising in different ways.

To elaborate briefly on this latter point, in a randomized block design treatments are assigned randomly to experimental units within blocks. This implies immediately that the two factors treatments and blocks are not "interchangeable." In a two-way classification, however, the factors are symmetric or "interchangeable." Data for this situation come either from an observational study, where the items observed are classified according to two factors, or from an experimental study, where the treatments are defined by level combinations from two factors and then applied to experimental units in a CRD.

Although the commonly used mathematical (linear) models for both situations described previously are basically of the same form, their statistical properties are quite different (a fact that is often overlooked). The major difference is due to the different randomization process, that is, restricted versus unrestricted randomization. For purposes of this note we shall not go into details of the theoretical aspects but shall refer the reader to some pertinent results in the literature. Wilk (1955) and Wilk and Kempthorne (1956a) investigated a general class of block-treatment experimental designs from the finite population and randomization point of view. Wilk and Kempthorne (1955, 1956a,b) also investigated, along the same lines, a two-factor experiment in the context of a CRD. Under the usual assumptions for the RCB design (no block-treatment interaction, application of each treatment to one experimental unit in each block) and the CRD (no two-factor interaction and one replication for each treatment combination), their results reduce to those given in Table 1. Under the RCB model, σ_b^2 denotes "block

*Marvin Lentner, Jesse C. Arnold, and Klaus Hinkelmann are Professors, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

Table 1. Analysis of Variance for RCB and Two-way Classification With One Observation Per Cell

Source	df	MS	E(MS)
	RCB		
Block	$b-1$	B^*	$\sigma^2 + t\sigma_b^2$
Treatment	$t-1$	T^*	$\sigma^2 + \sigma_u^2 + b\sigma_t^2$
Error	$(b-1)(t-1)$	E^*	$\sigma^2 + \sigma_u^2$
	Two-way classification		
F_A	$a-1$	A^*	$\sigma^2 + \sigma_{u^*}^2 + c\sigma_A^2$
F_C	$c-1$	C^*	$\sigma^2 + \sigma_{u^*}^2 + a\sigma_C^2$
Error	$(a-1)(c-1)$	E^{**}	$\sigma^2 + \sigma_{u^*}^2$

variability,'' σ_t^2 denotes ''treatment variability,'' σ_u^2 denotes variability among units within blocks, and σ^2 denotes variability due to all other extraneous components, and under the two-way classification model, σ_A^2 denotes ''factor A variability,'' σ_B^2 denotes ''factor B variability,'' and $\sigma_{u^*}^2$ denotes variability among experimental units. From Table 1, we observe that (a) blocks and treatments are ''asymmetric,'' whereas factors A and C are ''symmetric,'' and (b) there exists no test for equality of block effects (i.e., H_0: $\sigma_b^2 = 0$) except for the very unlikely event that $\sigma_u^2 = 0$.

3. RANDOMIZED COMPLETE BLOCK DESIGN

As we pointed out in the previous section there does not exist a valid test for block effects (e.g., Anderson and McLean 1974; Kempthorne 1952; Lentner and Bishop 1986). It may, however, be important to assess the usefulness of blocking, not for the current experiment but for future experiments using the same or similar experimental material. For this purpose Yates (1935), using the construct of a uniformity trial, introduced the notion of relative efficiency [see (1)]. Using the notation from Table 1, the ERE for comparing the RCB with the CRD is given by

$$\text{ERE(RCB compared with CRD)} = \frac{(b-1)B^* + b(t-1)E^*}{(bt-1)E^*}. \quad (2)$$

We note that Kempthorne (1952, 1955) derived (2) by using only randomization arguments, that is, restricted versus unrestricted randomization.

Rewriting (2), we observe that

$$\text{ERE(RCB compared with CRD)} = \alpha + (1-\alpha)H, \quad (3)$$

where $\alpha = b(t-1)/(bt-1)$ and $H = B^*/E^*$. Examination of the ERE in (3) reveals that

$$\begin{aligned} \text{ERE} &< 1 \quad \text{iff } H < 1 \\ &= 1 \quad \text{iff } H = 1 \\ &> 1 \quad \text{iff } H > 1. \end{aligned} \quad (4)$$

It should be noted that H is the ratio of block and error mean squares (MS), which many textbooks present, incorrectly, as an F statistic for testing block effects. It is clear from (3) that the commonly used efficiency measure, ERE, is a one-to-one monotone function of H. Thus, although H is not a valid F statistic for testing block effects, it may be used equivalently for judging the value of blocking in an RCB design: ERE > 1 or, alternatively, $H > 1$ indicates that the RCB is more effective than the CRD with the same number of replications for each treatment. A CRD would require b(ERE) replications to achieve the same efficiency as an RCB with b blocks. Whereas a value of H greater (smaller) than 1 implies a greater (lesser) efficiency, the value of H by itself does not provide full information about the efficiency measure; the degrees of freedom are important also, as can be seen in (3).

4. LATIN SQUARE DESIGNS

For a basic LS design, the experimental units are arranged in a square array according to row and column blockings. The t treatments are randomized to the units in such a way that each appears once in every ''row block'' and once in every ''column block.'' This makes row and column blockings orthogonal and, therefore, represents a straightforward extension of the basic RCB design. As with the RCB design, a valid test of equal treatment means exists (under appropriate assumptions), but valid tests of row and column blockings do not exist.

The assessment of one or both blocking factors in an LS design may be of interest. If only one blocking factor had been used, the experiment would have been conducted as an RCB design. For these, the relative efficiencies are

ERE_1 (LS compared with RCB, no row blocking)

$$= \frac{R^* + (t-1)E^*}{tE^*} \quad (5)$$

and

ERE_2 (LS compared with RCB, no column blocking)

$$= \frac{C^* + (t-1)E^*}{tE^*}, \quad (6)$$

where R^*, C^*, and E^* are row, column, and error MS, respectively, for the LS design. These measures may be rewritten as

$$\text{ERE}_1 = c + (1-c)H_R \quad (7)$$

and

$$\text{ERE}_2 = c + (1-c)H_C, \quad (8)$$

where $c = (t-1)/t$, $H_R = R^*/E^*$, and $H_C = C^*/E^*$. Again, the ratios, H_R and H_C, are given incorrectly in many textbooks as F statistics for testing row and column blockings, respectively. As seen in (7) and (8), the efficiency measures, ERE_1 and ERE_2, have the same format as the ERE given in (3). Thus, although H_R and H_C are not valid F statistics for testing row and column blockings, these ratios may be used equivalently for assessing the value of the respective blocking factor, as was indicated in (4).

5. CONCLUDING REMARKS

From the point of view of teaching elementary or applied statistics courses, the relationships between the relative efficiencies and the respective ratios of MS enables instructors to give students a method of assessing general gains (or losses) due to blocking without the need for introducing efficiency concepts. Not only is the relationship monotone one to one, but the efficiency is unity when the ratio of MS is unity (the point of no gain due to blocking). For either efficiency or ratio of MS, judgment of significant gains (or losses) due to blocking is subjective and must also be considered from a practical standpoint.

[*Received May 1987. Revised November 1988.*]

REFERENCES

Anderson, V. L., and McLean, R. A. (1974), *Design of Experiments*, New York: Marcel Dekker.

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

——— (1955), "The Randomization Theory of Experimental Inference," *Journal of the American Statistical Association*, 50, 946–967.

Lentner, M., and Bishop, T. (1986), *Experimental Design and Analysis*, Blacksburg, VA: Valley Book Co.

Wilk, M. B. (1955), "The Randomization Analysis of a Generalized Randomized Block Design," *Biometrika*, 42, 70–79.

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

——— (1956a), "Analysis of Variance: Preliminary Tests, Pooling, and Linear Models," WADC Technical Report 55-244, Wright-Patterson Air Force Base, Ohio.

——— (1956b), "Some Aspects of the Analysis of Factorial Experiments in a Completely Randomized Design," *The Annals of Mathematical Statistics*, 27, 950–985.

Yates, F. (1935), "Complex Experiments," *Journal of the Royal Statistical Society* (supplement), 2, 181–247.

Exploratory Plots for Paired Data

PAUL R. ROSENBAUM*

1. PLOTTING MATCHED–PAIR DIFFERENCES

Paired data consist of bivariate measurements, (a_i, b_i), $i = 1, \ldots, I$, where the a_i's and b_i's are on a common scale. The pair (a_i, b_i) may be measurements on a single subject i at two times, or measurements on two distinct subjects who are similar in some way.

The paired t statistic, the Wilcoxon signed rank statistic, and many other conventional statistical procedures for paired data focus on the matched-pair differences, $b_i - a_i$. It is natural, then, to plot the differences. Still, the differences tell only part of the story—they omit information about the marginal distributions of the a_i's and the b_i's and the dependence between the a_i's and b_i's. Moreover, an extreme or outlying difference, $b_i - a_i$, may result from an extreme a_i or an extreme b_i or, when the a_i's and b_i's are strongly related, from an exceptional pair in which the relationship does not hold. Examination of the differences alone cannot distinguish these three cases. An exploratory display should tell more of the story.

The example discussed here is based on data from Morton et al. (1982), a comparison of blood lead levels in $I = 33$ pairs of children. The parent of one child in each pair worked in a factory in Oklahoma in which lead is used in the manufacture of batteries. The parents of the second child in each pair, the control child, had not worked in an industry using lead for five years. The control child was matched to the treated child on the basis of age and neighborhood of residence. The thought was that parents who are exposed to lead while working might bring lead home in their clothes and hair, thereby exposing their children as well. Morton et al. compared blood lead levels (μg/dl of whole blood) for the pairs of exposed (b_i) and control (a_i) children using the signed rank statistic, presenting the data in tabular form, and plotting the marginal distributions in each group using histograms that ignore the pairing.

Figure 1 is a "sliding square plot." The center of the square is a scatterplot of the (a_i, b_i) pairs in which the horizontal and vertical axes have the same 0–80 μg/dl scale. Points project to the east to the boxplot (Tukey 1977) of the marginal distribution of the lead levels for exposed children, the b_i's. Points project to the north to the boxplot of the marginal distribution of the lead levels from control children, the a_i's. Points project to the southwest along the diagonal to the boxplot of the marginal distribution of the matched-pair differences, the $b_i - a_i$'s. Recall that in a boxplot, the center line is the median, the box ends at the quartiles, and extreme observations are indicated individually. Multiple projections of a single scatterplot have been used for a different purpose in Rosenbaum (1981). Multiple projections of a scatterplot allow us to locate a single point in several marginal distributions.

Figure 1 leads to the following observations. Since the lower quartile of the differences is 4, more than three quarters of the exposed children had higher lead levels than their matched controls, and the typical difference was 15 μg/dl. The lead levels for exposed children are not only higher but also more variable than for controls, as is seen from the width of the horizontal and vertical boxplots. Matching children on the basis of age and neighborhood did not produce a strong dependence within pairs; arguably, the matching was not very successful. The extreme difference of 60 μg/dl was due to an exposed child with an extremely high lead level; however, the level for this child's matched control is not unusual, being at the lower quartile of the controls.

*Paul R. Rosenbaum is Associate Professor, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104-6302.