Chapter 13

Complete Block Designs

We now begin the study of *variance reduction design*. Experimental error makes inference difficult. As the variance of experimental error (σ^2) increases, confidence intervals get longer and test power decreases. All other things being equal, we would thus prefer to conduct our experiments with units that are homogeneous so that σ^2 will be small. Unfortunately, all other things are rarely equal. For example, there may be few units available, and we must simply take what we can get. Or we might be able to find homogeneous units, but using the homogeneous units would restrict our inference to a subset of the population of interest. Variance reduction designs can give us many of the benefits of small σ^2, without necessarily restricting us to a subset of the population of units.

Variance reduction design

13.1 Blocking

Variance reduction design deals almost exclusively with a technique called *blocking*. A *block* of units is a set of units that are homogeneous in some sense. Perhaps they are field plots located in the same general area, or are samples analyzed at about the same time, or are units that came from a single supplier. These similarities in the units themselves lead us to anticipate that units within a block may also have similar responses. So when constructing blocks, we try to achieve homogeneity of the units within blocks, but units in different blocks may be dissimilar.

A block is a set of homogeneous units

Blocking designs are not completely randomized designs. The Randomized Complete Block design described in the next section is the first design we study that uses some kind of restricted randomization. When we design

Blocking restricts randomization

an experiment, we know the design we choose to use and thus the randomization that is used. When we look at an experiment designed by someone else, we can determine the design from the way the randomization was done, that is, from the kinds of restrictions that were placed on the randomization, not on the actual outcome of which units got which treatments.

There are many, many blocking designs, and we will only cover some of the more widely used designs. This chapter deals with *complete block designs* in which every treatment is used in every block; later chapters deal with *incomplete block designs* (not every treatment is used in every block) and some special block designs for treatments with factorial structure.

Complete blocks include every treatment

13.2 The Randomized Complete Block Design

The Randomized Complete Block design (RCB) is the basic blocking design. There are g treatments, and each treatment will be assigned to r units for a total of $N = gr$ units. We partition the N units into r groups of g units each; these r groups are our blocks. We make this partition into blocks in such a way that the units within a block are somehow alike; we anticipate that these alike units will have similar responses. In the first block, we randomly assign the g treatments to the g units; we do an independent randomization, assigning treatments to units in each of the other blocks. This is the RCB design.

RCB has r blocks of g units each

Block for homogeneity

Blocks exist at the time of the randomization of treatments to units. We cannot impose blocking structure on a completely randomized design after the fact; either the randomization was blocked or it was not.

Example 13.1 **Mealybugs on cycads**

Modern zoos try to reproduce natural habitats in their exhibits as much as possible. They therefore use appropriate plants, but these plants can be infested with inappropriate insects. Zoos need to take great care with pesticides, because the variety of species in a zoo makes it more likely that a sensitive species is present.

Cycads (plants that look vaguely like palms) can be infested with mealybug, and the zoo wishes to test three treatments: water (a control), horticultural oil (a standard no-mammalian-toxicity pesticide), and fungal spores in water (*Beauveria bassiana*, a fungus that grows exclusively on insects). Five infested cycads are removed to a testing area. Three branches are randomly chosen on each cycad, and two 3 cm by 3 cm patches are marked on each branch; the number of mealybugs in these patches is noted. The three

Table 13.1: Changes in mealybug counts on cycads after treatment. Treatments are water, *Beauveria bassiana* spores, and horticultural oil.

	Plant				
	1	2	3	4	5
Water	-9	18	10	9	-6
	-6	5	9	0	13
Spores	-4	29	4	-2	11
	7	10	-1	6	-1
Oil	4	29	14	14	7
	11	36	16	18	15

branches on each cycad are randomly assigned to the three treatments. After three days, the patches are counted again, and the response is the change in the number of mealybugs (before − after). Data for this experiment are given in Table 13.1 (data from Scott Smith).

How can we decode the experimental design from the description just given? *Follow the randomization!* Looking at the randomization, we see that the treatments were applied to the branches (or pairs of patches). Thus the branches (or pairs) must be experimental units. Furthermore, the randomization was done so that each treatment was applied once on each cycad. There was no possibility of two branches from the same plant receiving the same treatment. This is a restriction on the randomization, with cycads acting as blocks. The patches are measurement units. When we analyze these data, we can take the average or sum of the two patches on each branch as the response for the branch. To recap, there were $g = 3$ treatments applied to $N = 15$ units arranged in $r = 5$ blocks of size 3 according to an RCB design; there were two measurement units per experimental unit.

Why did the experimenter block? Experience and intuition lead the experimenter to believe that branches on the same cycad will tend to be more alike than branches on different cycads—genetically, environmentally, and perhaps in other ways. Thus blocking by plant may be advantageous.

It is important to realize that tables like Table 13.1 hide the randomization that has occurred. The table makes it appear as though the first unit in every block received the water treatment, the second unit the spores, and so on. This is not true. The table ignores the randomization for the convenience of a readable display. The water treatment may have been applied to any of the three units in the block, chosen at random.

You cannot determine the design used in an experiment just by looking at a table of results, you have to know the randomization. There may be many

Follow the randomization to determine design

different designs that could produce the same data, and you will not know the correct analysis for those data without knowing the design. *Follow the randomization to determine the design.*

General treatment structure

An important feature to note about the RCB is that we have placed no restrictions on the treatments. The treatments could simply be g treatments, or they could be the factor-level combinations of two or more factors. These factors could be fixed or random, crossed or nested. All of these treatment structures can be incorporated when we use blocking designs to achieve variance reduction.

Example 13.2

Protein/amino acid effects on growing rats

Male albino laboratory rats (Sprague-Dawley strain) are used routinely in many kinds of experiments. Proper nutrition for the rats is important. This experiment was conducted to determine the requirements for protein and the amino acid threonine. Specifically, this experiment will examine the factorial combinations of the amount of protein in diet and the amount of threonine in diet. The general protein in the diet is threonine deficient. There are eight levels of threonine (.2 through .9% of diet) and five levels of protein (8.68, 12, 15, 18, and 21% of diet), for a total of 40 treatments.

Two-hundred weanling rats were acclimated to cages. On the second day after arrival, all rats were weighed, and the rats were separated into five groups of 40 to provide groupings of approximately uniform weight. The 40 rats in each group were randomly assigned to the 40 treatments. Body weight and food consumption were measured twice weekly, and the response we consider is average daily weight gain over 21 days.

This is a randomized complete block design. Initial body weight is a good predictor of body weight in 3 weeks, so the rats were blocked by initial weight in an attempt to find homogeneous groups of units. There are 40 treatments, which have an eight by five factorial structure.

13.2.1 Why and when to use the RCB

Block when you can identify a source of variation

We use an RCB to increase the power and precision of an experiment by decreasing the error variance. This decrease in error variance is achieved by finding groups of units that are homogeneous (blocks) and, in effect, repeating the experiment independently in the different blocks. The RCB is an effective design when there is a single source of extraneous variation in the responses that we can identify ahead of time and use to partition the units into blocks. Blocking is done at the time of randomization; you can't construct blocks after the experiment has been run.

There is an almost infinite number of ways in which units can be grouped into blocks, but a few examples may suffice to get the ideas across. We would like to group into blocks on the basis of homogeneity of the responses, but that is not possible. Instead, we must group into blocks on the basis of other similarities that we think may be associated with responses.

Some blocking is fairly obvious. For example, you need milk to make cheese, and you get a new milk supply every day. Each batch of milk makes slightly different cheese. If your batches are such that you can make several types of cheese per batch, then blocking on batch of raw material is a natural.

Block on batch

Units may be grouped spatially. For example, some units may be located in one city, and other units in a second city. Or, some units may be in cages on the top shelf, and others in cages on the bottom shelf. It is common for units close in space to have more similar responses, so spatial blocking is also common.

Block spatially

Units may be grouped temporally. That is, some units may be treated or measured at one time, and other units at another time. For example, you may only be able to make four measurements a day, and the instrument may need to be recalibrated every day. As with spatial grouping, units close in time may tend to have similar responses, so temporal blocking is common.

Block temporally

Age and gender blocking are common for animal subjects. Sometimes units have a "history." The number of previous pregnancies could be a blocking factor. In general, any source of variation that you think may influence the response and which can be identified prior to the experiment is a candidate for blocking.

Age, gender, and history blocks

13.2.2 Analysis for the RCB

Now all the hard work in the earlier chapters studying analysis methods pays off. The design of an RCB is new, but there is nothing new in the analysis of an RCB. Once we have the correct model, we do point estimates, confidence intervals, multiple comparisons, testing, residual analysis, and so on, in the same way as for the CRD.

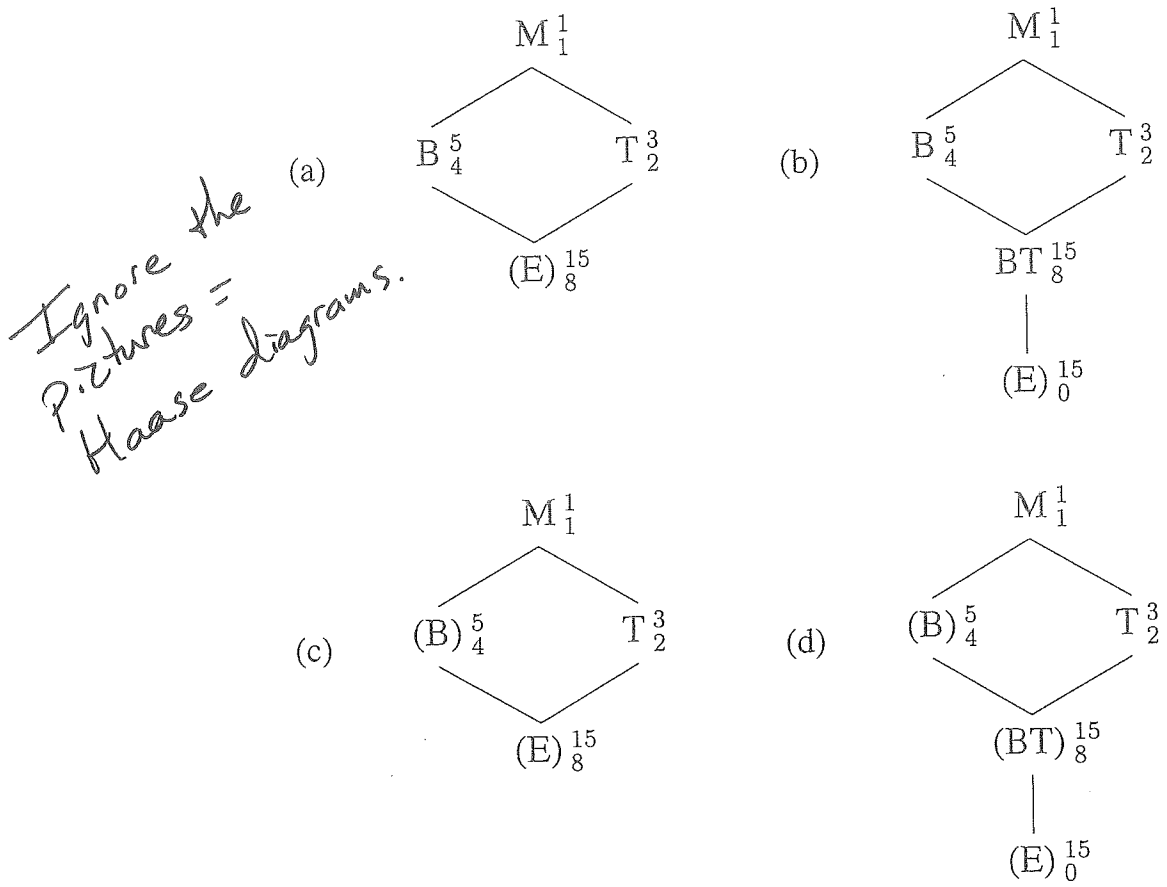
Nothing new in analysis of RCB

Let y_{ij} be the response for the ith treatment in the jth block. The standard model for an RCB has a grand mean, a treatment effect, a block effect, and experimental error, as in

Blocks usually assumed additive

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \ .$$

This standard model says that treatments and blocks are additive, so that treatments have the same effect in every block and blocks only serve to shift



Ignore the pictures = Haase diagrams.

Figure 13.1: Models for a Randomized Complete Block.

the mean response up or down. Hasse diagrams (a) or (c) in Figure 13.1 correspond to this standard model.

All reasonable models for RCB use the same analysis

To complete the model, we must decide which terms are random and which are fixed; we must also decide whether to use the standard additive model given above or to allow for the possibility that treatments and blocks interact. Fortunately, all variations lead to the same operational analysis procedure for the RCB design. Figure 13.1 shows Hasse diagrams for four different sets of assumptions for the RCB. Panels (a) and (b) assume the blocks are fixed, and panels (c) and (d) assume the blocks are random. Panels (a) and (c) assume that blocks do not interact with treatments (as in the standard model above), and panels (b) and (d) include an interaction between blocks and treatments. In all four cases, we will use the $(r-1)(g-1)$ degree of freedom term below treatments as the denominator for treatments. This is true whether we think that the treatments are fixed or random; what differs is how this denominator term is interpreted.

In panels (a) and (c), where we assume that blocks and treatments are additive, the $(r-1)(g-1)$ degree of freedom term is the usual error and the only random term below treatments. In panel (d), this term is the block by treatment interaction and is again the natural denominator for treatments. In panel (b), the correct denominator for treatments is "error," but "error" cannot be estimated because we have 0 degrees of freedom for error (only one observation for each treatment in each block). Instead, we must use the block by treatment interaction as a surrogate for error and recognize that this surrogate error may be too large if interaction is indeed present. Thus we will arrive at the same inferences regardless of our assumptions on randomness of blocks and interaction between treatments and blocks.

Denominator for treatments is $(r-1)(g-1)$ degree of freedom interaction or error

The computation of estimated effects, sums of squares, contrasts, and so on is done exactly as for a two-way factorial. In this the *model* we are using to analyze an RCB is just the same as a two-way factorial with replication $n = 1$, even though the *design* of an RCB is not the same.

One difference between an RCB and a factorial is that we do not try to make inferences about blocks, even though the machinery of our model allows us to do so. The reason for this goes back to thinking of F-tests as approximations to randomization tests. Under the RCB randomization, units are assigned at random to treatments, but units always stay in the same block. Thus the block effects and sums of squares are not random, and there is no test for blocks; blocks simply exist. More pragmatically, we blocked because we believed that the units within blocks were more similar, so finding a block effect is not a major revelation.

Do not test blocks—they were not randomized

Mealybugs, continued

Example 13.3

We take as our response the mean of the two measurements for each branch from Table 13.1. The ANOVA table follows:

	DF	SS	MS	F-stat	p-value
Blocks	4	686.4	171.60		
Treatments	2	432.03	216.02	12.2	.0037
Error	8	141.8	17.725		

There is fairly strong evidence for differences in mealybugs between the treatments, and there is no evidence that assumptions were violated.

Looking more closely, we can use pairwise comparisons to examine the differences. We compute the pairwise comparisons (HSD's or LSD's or whatever) exactly as for ordinary factorial data. The underline diagram below shows the HSD at the 5% level:

Water	Spores	Oil
-4.57	-2.97	7.53

Here we see that spores treatment cannot be distinguished from the control (water) treatment, but both can be distinguished from the oil treatment.

Standard model has blocks additive

The usual assumption made for an RCB model is that blocks and treatments do not interact. To some degree this assumption is forced on us, because as we saw from the Hasse diagrams, there is little we can do besides assume additivity. When the treatments have a factorial structure, we could have a model with blocks random and interacting with the various factors. In such a model, the error for factor A would be the A by block interaction, the error for factor B would be the B by block interaction, and so on. However, the standard model allows treatment factors to interact, whereas blocks are still additive.

Transform for additivity

Assuming that blocks and treatments are additive does not make them so. One thing we can do with potential interaction in the RCB is investigate transformable nonadditivity using Tukey one-degree-of-freedom procedures. When there is transformable nonadditivity, reexpressing the data on the appropriate scale can make the data more additive. When the data are more additive, the term that we use as error contains less interaction and is a better surrogate for error.

13.2.3 How well did the blocking work?

Gain in variance, lose in degrees of freedom

The gain from using an RCB instead of a CRD is a decrease in error variance, and the loss is a decrease in error degrees of freedom by $(r-1)$. This loss is only severe for small experiments. How can we quantify our gain or loss from an RCB? As discussed above, the "F-test" for blocks does not correspond to a valid randomization test for blocks. Even if it did, knowing simply that the blocks are not all the same does not tell us what we need to know: how much have we saved by using blocks? We need something other than the F-test to measure that gain.

Relative efficiency measures sample size savings

Suppose that we have an RCB and a CRD to test the same treatments; both designs have the same total size N, and both use the same population of units. The efficiency of the RCB relative to the CRD is the factor by which the sample size of the CRD would need to be increased to have the same information as the RCB. (Information is a technical term; think of two designs with the same information as having approximately the same power or yielding approximately the same length of confidence intervals.) For example, if an RCB with fifteen units has relative efficiency 2, then a CRD using the

same population of units would need 30 units to obtain the same information. Units almost always translate to time or money, so reducing N by blocking is one good way to save money.

Efficiency is denoted by E with a subscript to identify the designs being compared. The relative efficiency of an RCB to a CRD is given in the following formula:

Relative efficiency is the ratio of variances times a degrees of freedom adjustment

$$E_{\text{RCB:CRD}} = \frac{(\nu_{rcb}+1)(\nu_{crd}+3)}{(\nu_{rcb}+3)(\nu_{crd}+1)} \frac{\sigma^2_{crd}}{\sigma^2_{rcb}} ,$$

where σ^2_{crd} and σ^2_{rcb} are the error variances for the CRD and RCB, $\nu_{rcb} = (r-1)(g-1)$ is the error degrees of freedom for the RCB design, and $\nu_{crd} = (r-1)g$ is the error degrees of freedom for the CRD of the same size. The first part is a degrees of freedom adjustment; variances must be estimated and we get better estimates with more degrees of freedom. The second part is the ratio of the error variances for the two different designs. The efficiency is determined primarily by this ratio of variances; the degrees of freedom adjustment is usually a smaller effect.

We will never know the actual variances σ^2_{crd} or σ^2_{rcb}; we must estimate them. Suppose that we have conducted an RCB experiment. We can estimate σ^2_{rcb} using MS_E for the RCB design. We estimate σ^2_{crd} via

Estimate σ^2_{crd} with a weighted average of MS_E and MS_{Blocks}

$$\widehat{\sigma}^2_{crd} = \frac{(r-1)MS_{\text{Blocks}} + ((g-1)+(r-1)(g-1))MS_E}{(r-1)+(g-1)+(r-1)(g-1)}$$

This is the weighted average of MS_{Blocks} and MS_E with MS_{Blocks} having weight equal to the degrees of freedom for blocks and MS_E having weight equal to the sum of the degrees of freedom for treatment and error. This is *not* the result of simply pooling sums of squares and degrees of freedom for blocks and error in the RCB.

Mealybugs, continued

Example 13.4

For the mealybug experiment, we have $g = 3$, $r = 5$, $\nu_{rcb} = (r-1)(g-1) = 8$, $\nu_{crd} = r(g-1) = 12$, $MS_{\text{Blocks}} = 171.6$, and $MS_E = 17.725$, so we get

$$\begin{aligned}
\widehat{\sigma}^2_{crd} &= \frac{4 \times 171.6 + (2+8) \times 17.725}{4+2+8} = 61.69 , \\
\frac{(\nu_{rcb}+1)(\nu_{crd}+3)}{(\nu_{rcb}+3)(\nu_{crd}+1)} &= \frac{9 \times 15}{11 \times 13} = .944 , \\
\widehat{E}_{\text{RCB:CRD}} &= \frac{(\nu_{rcb}+1)(\nu_{crd}+3)}{(\nu_{rcb}+3)(\nu_{crd}+1)} \frac{\widehat{\sigma}^2_{crd}}{MS_E} ,
\end{aligned}$$

$$= .944 \times \frac{61.69}{17.725} = 3.29 \ .$$

We had five units for each treatment, so an equivalent CRD would have needed $5 \times 3.29 = 16.45$, call it seventeen units per treatment. This blocking was rather successful. Observe that even in this fairly small experiment, the loss from degrees of freedom was rather minor.

13.2.4 Balance and missing data

Balance makes inference easier

The standard RCB is balanced, in the sense that each treatment occurs once in each block. Balance was helpful in factorials, and it is helpful in randomized complete blocks for the same reason: it makes the calculations and inference easier. When the data are balanced, simple formulae can be used, exactly as for balanced factorials. When the data are balanced, adding 1 million to all the responses in a given block does not change any contrast between treatment means.

Missing data in an RCB destroy balance. The approach to inference is to look at treatment effects adjusted for blocks. If the treatments are themselves factorial, we can compute whatever type of sum of squares we feel is appropriate, but we always adjust for blocks prior to treatments. The reason is that we believed, before any experimentation, that blocks affected the response. We thus allow blocks to account for any variability they can before examining any additional variability that can be explained by treatments. This "ordering" for sums of squares and testing does not affect the final estimated effects for either treatments or blocks.

Treatments adjusted for blocks

13.3 Latin Squares and Related Row/Column Designs

Randomized Complete Block designs allow us to block on a single source of variation in the responses. There are experimental situations with more than one source of extraneous variation, and we need designs for these situations.

Example 13.5 **Addled goose eggs**

The Canada goose (*Branta canadensis*) is a magnificent bird, but it can be a nuisance in urban areas when present in large numbers. One population control method is to addle eggs in nests to prevent them from hatching. This method may be harmful to the adult females, because the females fast while incubating and tend to incubate as long as they can if the eggs are unhatched.