RANDOMIZED COMPLETE BLOCK DESIGNS

Introduction to Blocking

Nuisance factor: A factor that probably has an effect on the response, but is not a factor that we are interested in.

Types of nuisance factors and how to deal with them in designing an experiment:

Characteristics	**Examples**	**How to treat**
Unknown, uncontrollable	Experimenter or subject bias, order of treatments	Randomization Blinding
Known, uncontrollable, measurable	IQ, weight, previous learning, temperature	Analysis of Covariance
Known, moderately controllable (by choosing rather than adjusting)	Temperature, location, time, batch, particular machine or operator, age, gender, order, IQ, weight	Blocking

Randomization can in principal be used to take into account factors that can be treated by blocking, but blocking usually results in smaller error variance, hence better estimates of effect. Thus blocking is sometimes referred to as a method of *variance reduction design.*

The intuitive idea: Run in parallel a bunch of experiments on groups (called *blocks*) of units that are fairly similar.

The simplest block design: The *randomized complete block design* (RCBD)

v treatments
(They could be treatment combinations.)

b blocks, each with v units
Blocks chosen so that units within a block are alike (or at least similar) and units in different blocks are substantially different. (Thus the total number of experimental units is n = bv.)

The v experimental units within each block are randomly assigned to the v treatments. (So each treatment is assigned one unit per block.)

Note that experimental units are assigned randomly only within each block, not overall. Thus this is sometimes called a *restricted randomization*.

Example: Five varieties of wheat are to be compared to see which gives the highest yield. Eight plots of farmland are available for the experiment. The experimenter divides each plot into five subplots. For each of the 8 plots, the varieties of wheat were randomly assigned to the subplots of that plot.

Treatment factor =

Response =

Blocking factor =

Blocks =

Experimental units =

v = b =

n = # exp units =

RCBD Model:

$$Y_{hi} = \mu + \theta_h + \tau_i + \varepsilon_{hi}$$
$$\varepsilon_{hi} \sim N(0,\sigma^2)$$
$$\varepsilon_{hi}\text{'s independent}$$

where

- Y_{hi} is the random variable representing the response for treatment i observed in block h
- μ is a constant (which may be thought of as the overall mean – see below)
- θ_h is the (additive) effect of the h^{th} block
  ($h = 1, 2, \ldots, b$)
- τ_i is the (additive) effect of the i^{th} treatment
  ($i = 1, 2, \ldots, v$)
- ε_{hi} is the random error for the i^{th} treatment in the h^{th} block.

(Why is there no subscript t for observation number?)

Note:

1. This model is also called the *block-treatment model*. Formally, it looks just like a two-way main effects model – *but* remember:

   - There is just *one* factor (plus one block).
   - The randomization is just *within each block*.
   - Thus we do *not* have the conditions for a two-way analysis of variance.

2. Like the main-effects model, this is an *additive* model:

   - It does not provide for any interaction between block and treatment level.
   - It assumes that treatments have the same effect in every block, and the only effect of the block is to shift the mean response up or down.
   - If interaction between block and factor is suspected, then *either* a transformation is needed to remove interaction before using this model, *or* a design with more than one observation per block-treatment combination must be used.
   - Trying to add an interaction term in the RCBD would create the same problem as is encountered in two-way ANOVA with one observation per cell: the degrees of freedom for the error are zero, so the method of analysis breaks down.

3. This is an over-specified model; the additional constraints $\sum_{h=1}^{b}\theta_h = 0$ and $\sum_{i=1}^{v}\tau_i = 0$, are typically added, so that the treatment and block effects are thought of as deviations from the overall mean.

4. There is an alternate *means model*:

$$Y_{hi} = \mu_{ih} + \varepsilon_{hi}, \text{ where } \mu_{ih} = \mu + \theta_h + \tau_i.$$

5. Note that the i^{th} treatment mean is

$$\mu_i = \frac{1}{b}\sum_{h=1}^{b}(\mu + \theta_h + \tau_i).$$

Assuming the constraint $\sum_{h=1}^{b}\theta_h = 0$, this gives

$$\mu_i = \mu + \tau_i$$

Estimating and Analysis:

Least squares fits: Since the model is formally the same as the main-effects model, the process of finding least squares estimates is the same, yielding estimates (with notation appropriately changed)

$$\mu\hat{} = \bar{y}_{\bullet\bullet}$$

$$\theta_h\hat{} = \bar{y}_{h\bullet} - \bar{y}_{\bullet\bullet}$$

$$\tau_i\hat{} = \bar{y}_{\bullet i} - \bar{y}_{\bullet\bullet}$$

$$y_{hi}\hat{} = \mu\hat{} + \theta_h\hat{} + \tau\hat{} =$$

$$= \bar{y}_{h\bullet} + \bar{y}_{\bullet i} - \bar{y}_{\bullet\bullet}$$

Thus the error sum of squares for this model is

ssE =

As with the two-way main effects model,

$$MSE = SSE/(b-1)(v-1)$$

is an unbiased estimator of σ^2.

Note: Since $n = bv$,

$$(b-1)(v-1) = bv - b - v + 1 = n - b - v + 1$$

Model checking: Important as always. Look especially for potential problems with:

- the normality assumption
- unequal error variance by block or treatment
- treatment-block interaction

To check for possible block-treatment interaction, form an "interaction plot" by plotting each y_{hi} against each treatment level i and connecting points for each block h. If corresponding line segments are parallel, this suggests both no interaction and small error variability

Note: Since there is just one observation per (block, treatment level) combination, there is no way to check the equal variance assumption at that fine a level.

Hypothesis test and Analysis of Variance Table:

We are interested in testing equality of treatment means. Thus we wish to test the null hypothesis

H_0: $\mu_1 = \mu_2 = \ldots = \mu_v$

against the alternate

H_a: $\mu_i \neq \mu_j$ for at least one pair i,j.

Note: Since $\mu_i = \mu + \tau_i$, we can restate the hypotheses as

H_0: $\tau_1 = \tau_2 = \ldots = \tau_v = 0$

and

H_a: at least one $\tau_i \neq 0$.

We can construct an F-test in the usual manner:

Consider the submodel corresponding to the null hypothesis, namely

$Y_{hi} = \mu + \theta_h + \varepsilon_{hi}$

This has least squares fits

$(y_{hi}\hat{})_0 = (\bar{y}_{\bullet\bullet}) + (\bar{y}_{h\bullet} - \bar{y}_{\bullet\bullet})$

$= \bar{y}_{h\bullet}$

and hence error sum of squares

$ssE_0 =$

The difference $ssT = ssE_0 - ssE$ is called the *sum of squares for treatment*. Our test statistic for H_0 is

$$\frac{ssT/(v-1)}{ssE/(b-1)(v-1)}.$$

As usual, the numerator is denoted msT (with v-1 degrees of freedom) and the denominator msE (with (b-1)(v-1) degrees of freedom, as mentioned above.

The test statistic has an F distribution with v-1 degrees of freedom in the numerator, (b-1)(v-1) in the denominator.

Note:

- The above test is the same as the F-test for the treatment factor we would get by two-way ANOVA considering treatment and block as two factors in a main effects model. Thus we can test our hypothesis by using a two-way ANOVA main-effects software routine. *But* we only look at the test for T.

- We can define ssB and msB (using b-1 degrees of freedom), but we *don't* get a legitimate F-test for the null hypothesis "No block effect," since the conditions for proving that the would-be test statistic has an F-distribution are not met, because the blocks are chosen, not randomly assigned.

- Nonetheless, the ratio msB/msE can be considered as an informal measure of the effect of the blocking factor – if the ratio is large, that suggests that the blocking "factor" has a large effect, and that the variance reduction obtained by blocking was probably helpful in by improving the precision in the comparison of treatment means.

- The algebra works out to show that

$$ssTot = ssB + ssT + ssE,$$

and the degrees of freedom add accordingly.

Contrasts:

In the RCBD, all contrasts (with coefficient sum zero) in the treatment effects τ_i are estimable, and the techniques of Chapter 4 still apply, with the following observed:

- The estimate of τ_i is $\hat{\tau}_i = \bar{y}_{\bullet i} - \bar{y}_{\bullet\bullet}$
- Since in a contrast $\sum c_i \tau_i$, we have $\sum c_i = 0$, the estimate of the contrast is $\sum c_i \bar{y}_{\bullet i}$
- The number of replicates is equal to the number b of blocks.
- The error degrees of freedom are (b-1)(v-1).
- The msE used is the one obtained by the block design analysis. (Thus the Minitab automatic procedures will not work for a block design.)