Randomized complete block design

Recall 1-way AOV: "completely randomized design with single factor". There are n_{T} experimental units (book is now using notation N), randomly assigned among t treatments; objective of study is to obtain inferences about the treatment means μ_1, μ_2, ..., μ_t. Reparameterization will be helpful for extending the AOV model to more complicated designs:

$$\begin{aligned} \alpha_1 &= \mu_1 - \mu \\ \alpha_2 &= \mu_2 - \mu \\ &\vdots \\ \alpha_t &= \mu_t - \mu \end{aligned}$$

The α_i's are the **treatment effects**; here μ is the overall mean (the mean of the μ_i's) and $\alpha_1 + \alpha_2 + \cdots + \alpha_t = 0$.

The statistical model becomes

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where $\epsilon_{ij} \sim$ normal$(0,\, \sigma^2)$, or equivalently,

$$Y_{ij} \sim \text{normal}\big(\mu + \alpha_i,\, \sigma^2\big)$$

The **RCBD design** incorporates an additional categorical variable, called a **blocking variable** or **block**, to account for possible heterogeneity in experimental circumstances. For instance, a typical block in an agricultural experiment is a field— fields differ substantially in soil quality, etc., and the same experimental treatment might produce different means in different fields.

Formally, the design is as follows: within each of b blocks, assign 1 experimental unit at random to each of t treatments. Thus, all treatments appear within each block, and each block-treatment combination receives 1 experimental unit, which produces the observed response y_{ij}:

	block				
treatment	1	2		b	
1	y_{11}	y_{12}	$\cdots$	y_{1b}	$\overline{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2b}	$\overline{y}_{2\cdot}$
$\vdots$			$\vdots$		$\vdots$
t	y_{t1}	y_{t2}	$\cdots$	y_{tb}	$\overline{y}_{t\cdot}$
mean	$\overline{y}_{\cdot 1}$	$\overline{y}_{\cdot 2}$		$\overline{y}_{\cdot b}$	

The statistical model is

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

where $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, and $\epsilon_{ij} \sim$ normal(0, σ^2). Here β_j is the effect of the jth block on the mean of Y_{ij}.

Note that β_j is the same for every treatment. It is not possible to estimate an **interaction** between block and treatment, because each block/trt combination is assigned only 1 experimental unit.

ML estimates of model parameters:

$$\widehat{\mu} = \overline{y}_{..} = \sum_i \sum_j y_{ij}/N$$
$$\widehat{\alpha}_i = \overline{y}_{i.} - \overline{y}_{..}$$
$$\widehat{\beta}_j = \overline{y}_{.j} - \overline{y}_{..}$$

$$\widehat{\sigma}^2 = \left(\frac{1}{(b-1)(t-1)}\right)\sum_{i=1}^{t}\sum_{j=1}^{b}\left[y_{ij} - \left(\widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j\right)\right]^2$$

Note: $N = bt$, and
$N - (b-1) - (t-1) - 1 = (b-1)(t-1)$

Sums of squares:

$$\text{SS(block)} = t\sum_{j=1}^{b}\left(\overline{y}_{.j} - \overline{y}_{..}\right)^2 = t\sum_j \widehat{\beta}_j^{\,2}$$

$$\text{SS(treatment)} = b\sum_{i=1}^{t}(\overline{y}_{i.} - \overline{y}_{..})^2 = b\sum_i \widehat{\alpha}_i^{\,2}$$

$$\text{SS(error)} = \sum_{i=1}^{t}\sum_{j=1}^{b}\left[y_{ij} - \left(\widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j\right)\right]^2 = \sum_{i,j} e_{ij}^2$$

And as always, $\text{SS(total)} = \sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{..})^2$

The AOV table for the RCBD design:

Source	SS	df	MS	F
block	SS(block)	$b-1$	$\frac{\text{SS(block)}}{b-1}$	$\frac{\text{MS(block)}}{\text{MS(error)}}$
trt	SS(trt)	$t-1$	$\frac{\text{SS(trt)}}{t-1}$	$\frac{\text{MS(trt)}}{\text{MS(error)}}$
error	SS(error)	$(b-1)(t-1)$	$\frac{\text{SS(error)}}{(b-1)(t-1)}$	
total	SS(total)	$bt-1$		

Hypothesis tests

Block effect

$$\begin{aligned} &\text{H}_0\text{: } \beta_1 = \beta_2 = \cdots = \beta_b = 0 \\ &\text{H}_a\text{: } \quad\quad \neq \end{aligned}$$

Test statistic is $f = \frac{\text{MS(block)}}{\text{MS(error)}}$ using F$(b-1, (b-1)(t-1))$ distribution.

Treatment effect

$$\begin{aligned} &\text{H}_0\text{: } \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0 \\ &\text{H}_a\text{: } \quad\quad \neq \end{aligned}$$

Test statistic is $f = \frac{\text{MS(trt)}}{\text{MS(error)}}$ using F$(t-1, (b-1)(t-1))$ distribution.

RCBD remarks

1. Missing observation(s) produces unbalanced design; block and treatment are no longer orthogonal (& hand-calculating formulas are no longer available). Parameters can still be estimated, and effects tested, in PROC GLM. Use type III sums of squares:

```
MODEL Y=BLOCK TRT / SS3;
```

Type III SS tests the effect last, after the other variables are entered in the model.

2. If blocks have been randomly chosed from some defined population of blocks (block might be a person in a study, say, or a plot of land picked from farms in southern Idaho), one might consider a model in which

$$\beta_j \sim \text{normal}\left(0, {\sigma_\beta}^2\right)$$

This is called a **random effects** model (or, more precisely, a mixed effects model) & the experiment is called a **random blocks** experiment (rather than randomized block). We will look at a few of these types of models toward the end of the course. The ordinary RCBD model is a **fixed effects** model.