Stat 471/571: Split plot designs and their analysis

Q: What's the right error?

Example study: randomized experiment evaluating a new physical activity program

Goal: Increase kid's physical activity
Question: Does a new school-level program increase kid's physical activity?
Data: made up, based on some recent studies I've helped with

- Two treatments: Active (the new intervention), Control (no intervention)
- Schools are paired by county and school size. 10 blocks, 2 schools in each
  - Randomly assign intervention to one of the two schools in each block
- Measure kids physical activity after 1 year of the intervention
- Response is %time in moderate-vigorous physical activity (mvpa)
- In the subset we start with, there are measurements for 10 kids per school.
- Inspection of the data indicates severely unequal variances and skewed distributions
  - most (75%) kids have $< 7.1\%$ mvpa, 5% have $> 12.8\%$, maximum is 39.6%
- So analyze log(mvpa)
  - Approximately equal variances, more symmetrical distributions

Review:

- what is the experimental unit (eu)?
- what is the observational unit (ou)?
- do we have subsampling?

A model to assess the effect of the intervention is:

$$Y_{ijk} = \mu + \beta_i + \alpha_j + \gamma_{ij} + \varepsilon_{ijk}, \quad (1)$$

where:
i is the block and β_i is the block effect,
j is the treatment and α_j is the treatment effect,

ij identifies each school, and γ_{ij} is the variability between schools,
k identifies the student within a school, so ijk identifies each student
and ε_{ijk} is the variability between students within a school

Now, what if we classify students by gender? Not randomly assigned, but we're going to analyze the data as if it were.

Treat gender as if it were randomly assigned to a student.

- what is the experimental unit?
- what is the observational unit?

There are two sizes of eu:
school: assigned to intervention/control
student: assigned to gender

Vocabulary

The general name for a design with two sizes of eu is split plot design.

- Comes from RA Fisher's agricultural background
- Main plot: "larger" eu, e.g. school, that is then divided into multiple
- Split plots: "smaller" eu. e.g. student

Notice there are two different experimental design used here:

- How are treatments randomized to schools?: within blocks (pairs of schools)
  so RCBD at the main plot level
- How are treatments randomized to students within a school? randomly
  so CRD at the split plot level, within each school

This is an example of the most common split plot design:
blocking at the main plot level and a CRD at the split plot level
There are many, many other combinations of designs that you could use to do a study. Rather than talk about the model for "the" split plot design, we'll figure out how to construct a model for any sort of design with two sizes of eu.

Constructing a model for a split plot design

We will construct two models, one for the main plots (schools), the other for the split plots (students), then knit them together.

Main plots: ignore split plot treatment (gender), average students within a school. ou is now the school. 20 rows of data. Write out the design: 3 components: block, treatment, and error = block*treatment interaction. Here's the skeleton ANOVA table for the main plot part of the design.

Source	df
Block	9
Intervention	1
Error = Block*Intervention	9

If we only had main plots, the error would be included automatically. Here, it is specifically named in terms of other main plot characteristics because we will need that when we combine main and split plots.

Split plots: Think of each school as a "mini block" and write out the split part of the design. Here CRD with one treatment factor:

Source	df
School	19
Gender	1
Error	200 - 21

Now combine them. Notice that the main plot df sum to 19, the same df as mini-blocks. If they don't, there is a mistake somewhere.

Source	df
Block	9
Intervention	1
Error = Block*Intervention	9
Gender	1
Error	179

And add the treatment interactions. Interactions between main and split plot treatment factors only exist in the combined model. Interactions between main plot design factors (e.g. blocks) are usually not included. The result is the skeleton anova for the analysis.

Source	df	
Block	9	
Intervention	1	
Block*Intervention	9	Main plot error
Gender	1	
Gender*Intervention	1	
Error	178	Split plot error

In terms of an equation, we add the split plot treatments and interaction to equation (1).

$$\begin{aligned} Y_{ijkl} &= \mu + \beta_i + \alpha_j + \gamma_{ij} + \delta k + \alpha\delta_{jk} + \varepsilon_{ijkl}, \\ \gamma_{ij} &\sim N(0, \sigma^2_{school}, \\ \varepsilon_{ijkl} &\sim N(0, \sigma^2_{kid}, \end{aligned} \tag{2}$$

where, as before, below equation (1),
i is the block and β_i is the block effect,
j is the treatment and α_j is the treatment effect,
ij identifies each school, and γ_{ij} is the variability between schools.
And the new stuff at the split plot level is:
k is the gender of a student,
$\alpha\delta_{jk}$ is the interaction between treatment and gender,
l identifies the student within a school, so $ijkl$ identifies each student
and ε_{ijkl} is the variability between students within a school and gender