Stat 471/571: Split plot designs and their analysis: part 2

Examples of split plot studies from various fields

Engineering calls these designs "hard-to-change factor" designs. The idea is that some factors (hard-to-change) require extended time to change the level; other factors can be changed quickly. In the Ag Eng. example below, changing the machinery takes perhaps an hour. Changing the combine speed can be done nearly instantaneously. Randomly assigning all combinations of machinery and speed will require many changes of machinery. A natural way to reduce the total study time is to run multiple speeds with one set of machinery, then change the machinery and run all speeds with that.

Application area	Main plot / treatment	Split plot / treatment
Ag. Eng. (combines)	15 minute run machinery	3 minute run speed
Agronomy	Field Irrigation	row variety
Biochemistry	96 well plate incubation time	individual well dose of chemical
Nutrition	person ethnicity, gender	period diet
Horticulture	water bath root temperature	pot species
Meat science	10lb batch of meat rosemary oil	package of hot dogs radiation dose
Education	class teaching method	student gender

Why choose a split plot design?

- Rarely deliberately chosen. Usually done out of necessity because can't apply a treatment to the "small" size eu. E.g. too time consuming to change machinery frequently, can't irrigate one row, don't have enough water baths to use one per plot.
- Can be used to add treatments to ongoing study.
  Park Grass Experiment, Rothamsted England. Effects of fertilization on hay yield. Started in 1856; data collected annually since them. Fertilization changes the soil pH. Lime treatments

(two levels: add lime, don't) started in 1903. Want to continue with original treatments and plots.
Solution: divide each 1856 plot into two halves. One gets lime; the other does not.
Starting 1965, used four lime levels by dividing each 1856 plot into quarters

- Using two sizes of eu makes the study possible but complicates the analysis.

Expected Mean Squares

One way to think about tests and appropriate error terms in an ANOVA table.

Context: The data are random variables. So are any quantities computed from data.
Examples we've already seen:

- Treatment mean
- Difference between two treatment means
- T statistic testing difference between two treatment means
- F statistic testing equality of 2 or more treatment means

We've (so far) focused on the variability in that random variable, e.g.

- standard error of a mean
- standard error of a difference between two means
- d.f. for a T statistic
- error d.f. for an F statistic

A random variable has an expected value = theoretical mean = value for an infinite-sized sample. The expected value of a random variable is usually written as E random variable. Examples are:

- Treatment mean: E $\overline{Y}_i = \mu_i$
- Difference of two treatment means: E $(\overline{Y}_i - \overline{Y}_j) = \mu_i - \mu_j$
- E T statistic, when null hypothesis is true, = 0
- E F statistic, when null hypothesis is true, = 1

Each of the Mean Squares in the ANOVA table is computed from data. So each is a random variable and has an expected value. That Expected Mean Square (EMS) depends on values of fixed effects and variances of random effects.

Simple example: Two way factorial in a CRD. The model: $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$. The expected value of the error variance, MSE, is σ^2. That's what justifies using $s = \sqrt{MSE}$ in T statistics. The other terms in the ANOVA table also have EMS. The ANOVA table:

Source	df	EMS
Factor A	$(a-1)$	$\sigma^2 + Q(\alpha_i)$
Factor B	$(b-1)$	$\sigma^2 + Q(\beta_j)$
Interaction	$(a-1)(b-1)$	$\sigma^2 + Q(\alpha\beta_{ij})$
Error	$ab(n-1)$	σ^2

Q() indicates a quadratic function of the indicated quantities. All that matters is that Q(anything) = 0 when all the "anythings" are the same. So $Q(\alpha_i) = 0$ when the main effects of factor A are equal. $Q(\alpha\beta_{ij}) = 0$ when all interaction effects = 0.

The important thing about the EMS: when the null hypothesis is true, the Q() part = 0, so that (for this design) EMS = σ^2. The resulting F statistic = MS(effect) / MS(error) = 1. If that ratio of EMS doesn't = 1, then your analysis is using the wrong error term for that hypothesis.

When there is only one experimental unit, all EMS for treatment effects have the form $\sigma^2 + Q(something)$ (as above), so it is appropriate to use the error MS as the denominator of an F statistics. Not so for a split plot study!

"Simple" split plot study: RCBD for main plots, CRD for split plots, equal sample sizes. b blocks, $m = 2$ main plot treatments (two treatments, one each per block), s split plot treatments, one each per main plot, and r replicates per combination of split plot treatment and main plot. Example: physical activity study has $b = 10$, $m = 2$, $s = 2$, and $r = 5$.

Source	df	Expected MS
Block	$b-1$	not relevant
Main plot treatment	$m-1$	$\sigma^2 + b\sigma^2_{main} + Q(main)$
Main plot error	$(b-1)(m-1)$	$\sigma^2 + b\sigma^2_{main}$
Split plot treatment	$s-1$	$\sigma^2 + Q(split)$
Split*Main	$(m-1)(s-1)$	$\sigma^2 + Q(interaction)$
Split plot error	bmsr - m(b+s-1)	σ^2

As an aside, you can estimate σ^2_{main}, the variance component for the main plots from the expected mean squares and observed mean squares for the Main plot and Split plot error lines. That's the ANOVA, also called method of moments, estimator of variance components.

Using EMS to determine appropriate F statistics Remember, from above, that when the null hypothesis is true, $Q()$ equals zero, and the ratio of EMS should = 1. If you have the expected mean squares, then these tell you the appropriate F statistic for any test.

Example: Test split× main interaction: Numerator is Split*Main with EMS: $\sigma^2 + Q(interaction)$. When $Q() = 0$, that is σ^2, so the appropriate denominator is a term with am EMS of σ^2. That's the error. So the appropriate F statistic to test no interaction is $F = MS(split * main)/MS(error)$.

Example: Test main plot treatment: Numerator is Main plot treatment with EMS: $\sigma^2 + b\sigma^2_{main} + Q(main)$. When $Q() = 0$, that is $\sigma^2 + b\sigma^2_{main}$, so the appropriate denominator is a term with am EMS of $\sigma^2 + b\sigma^2_{main}$. That's the main plot error line. So the appropriate F statistic to test no interaction is $F = MS(main)/MS(mainploterror)$.

If you argue "the other direction", you find out what hypothesis is being tested by a proposed F statistic. Consider F = MS(main) / MS(error). The ratio of EMS is: $\sigma^2 + b\sigma^2_{main} + Q(main)$ / σ^2. That = 1 only when $\sigma^2_{main} = 0$ and $Q(main) = 0$. So you can "reject" either when there are no differences between treatment means (the intended test) or when there is sufficient variability among main plots ($\sigma^2_{main} > 0$) or a combination of both. Only by using the correct error term do you get a test of the intended piece.

F tests for physical activity study

The ANOVA table:

Source	df	MS	F	p-value
Block	9	–	–	–
Intervention	1	7.093	14.07	0.0045
Block*Int	9	0.504	–	–
Gender	1	2.800	8.28	0.0045
Gender*Int	1	1.997	5.91	0.016
Error	178	0.338	–	–

Notice that the F statistics for Gender and Gender*Intervention (split plot effects) use Error as the denominator while the F statistic for Intervention (randomized to school) use the Block*Interaction, i.e., the main plot error as the denominator.

The estimated variance components are:

- Block*intervention = main plot error: 0.0166
- Residual = split plot error: 0.3380

Planning a split plot study

Two relevant aspects:

- The main plot MS is usually larger than the split plot error.
  Because of the "mini-blocks" created by the main plots.
- The main plot error has fewer df than the split plot error (lots fewer here!).
  So t quantiles for tests and confidence intervals will be larger

Consequences:

- Estimates of main plot effects have larger standard errors and wider confidence intervals than do estimates of split plot effects.
- Tests of main plot effects have lower power than do tests of split plot effects.

Often, one factor is more important than the rest. When possible, **put the more important factor at the split plot level**.

If possible to run a study as a 2 way factorial with one size of eu or as a split plot, you're better off doing the 2 way factorial. The split plot has a slightly higher power for split plot effects (because of "mini-blocks"). This is more than offset by a vastly smaller power for main plot effects.

Estimates of means and differences of means.

Goal for this block of material: If you look at output for a split plot analysis, you will notice all sorts of unusual things, including:

- Different standard errors for different comparisons of means, even when sample sizes the same.
- Fractional degrees of freedom (e.g., 24.6)

My goal here is explain why those happen and why they should be expected in a split plot analysis. There are lots of equations; you are not expected to manipulate any of these equations. They are included to provide the reasons that support my claims.

Standard errors of treatment effects (differences or contrasts among means):
Depend on whether main plot or split plot level, coefficents here are for differences.

Notation, as before:
m # of main plot levels,
s # of split plot levels,
b # of blocks,
r # of replicates of each split plot level

In the PA study, $m = 2$, $s = 2$, $b = 10$ and $r = 5$.

Quantity	se of mean	se of difference
Main treatment (e.g., Intervention)	$\frac{\sigma^2}{sbr} + \frac{\sigma^2_{main}}{sr}$	$\sqrt{2}\left(\frac{\sigma^2}{sbr} + \frac{\sigma^2_{main}}{sr}\right)$
Split treatment (e.g., Gender)	$\frac{\sigma^2}{brm} + \frac{\sigma^2_{main}}{bm}$	$\sqrt{2}\frac{\sigma^2}{brm}$
Cell mean	$\frac{\sigma^2}{br} + \frac{\sigma^2_{main}}{b}$	–

Two things to notice:

- the se of a main plot treatment mean =

$$\begin{aligned} se &= \frac{\sigma^2}{sbr} + \frac{\sigma^2_{main}}{sr} \\ &= \frac{\sigma^2 + b\sigma^2_{main}}{sbr} \\ &= \frac{MS(\text{main plot error})}{sbr} \end{aligned}$$

  Since the se of a main plot treatment mean is proportional to the Main plot error, we know the df for a T test or F test. It's the df for the main plot error.

- The se of split plot treatment means can not be simplified like that. Not clear what the df is.

Differences of cell means come in two versions, with different se's. Here is the conceptual explanation without formulae.

- Difference between two split plot treatments in the same main plot, e.g., Male/Intervention - Female/Intervention. This difference is an average of differences within each school (mini-block). So this only involves the error variance.
- Difference between two main plot treatments in the same or different split plots, e.g. Female/Control - Female/Intervention, or Female/Control - Male/Intervention. This involves different schools so this se involves both the error and main plot variances.

Degrees of freedom for inconvenient standard errors

The se for a split plot mean can be written as $\frac{1}{brm}(\sigma^2 + r\sigma^2_{main})$. There is no term in the ANOVA table with an expected mean square of $\sigma^2 + r\sigma^2_{main}$, so we can't do the same thing that we did with the main plot treatment mean. However, $\sigma^2 + r\sigma^2_{main}$ can be written as a linear combination of the EMS(Error)= σ^2 and EMS(main plot error)= $\sigma^2 + b\sigma^2_{main}$. That is:

$$\sigma^2 + r\sigma^2_{main} = \left(\frac{b-r}{b}\right) MS(\text{error}) + \left(\frac{r}{b}\right) MS(\text{main plot error})$$

For this study design, that is $0.5MS(\text{Error}) + 0.5MS(\text{main plot error})$. The question is what is an appropriate df for this linear combination?

Satterthwaite, a quantitative psychologist, worked this out and published his solution in 1941 (Psychometrika). His approximation provides an approximate df for linear combinations of mean squares. The computation uses the values and df for all MS in the linear combination. In case you want to know the equation, it is: d.f. for $c_1\text{MS}_1 + c_2\text{MS}_2$:

$$\text{df} = \frac{(c_1\text{MS}_1 + c_2\text{MS}_2)^2}{\frac{(c_1\text{MS}_1)^2}{\text{df}_1} + \frac{(c_2\text{MS}_2)^2}{\text{df}_2}}$$

The result is almost always a non-integer df. For this study, the df for a split plot marginal mean is 24.6. The df for a cell mean is also 24.6. Computers can figure out the appropriate tail quantiles for tests of confidence intervals.

More recently, Kenward and Roger (1997, Biometrics) worked out an approximation that works in more situations than Satterthwaite considered. Their df is the result of a matrix computation that provides no insight (at least to me). For balanced split plot designs, the two approximations are identical, so it doesn't matter which you use. For unbalanced split plot designs, the two approximations are very close, so essentially no practical difference. There are other situations, e.g. some repeated measures analyses (covered in a few weeks), where the differences matter and KR is preferred.

And, the complicate the issue, there are multiple Kenward-Roger adjustments, e.g., first-order, second-order, or "improved", based on Kenward, Roger (2009, Comp. Stat Data An.).

Default computing methods

Software	Default or recommended df
JMP	Kenward Roger, 1st order
SAS	/ddfm=kr: Kenward-Roger, 2nd order, but can specify lots of others
R (emmeans)	Kenward-Roger, not clear which version

My suggestion: don't worry about the details. Definitely report what software you used.

Estimates for the PA study, using SAS, KR:

Type	Quantity	Estimate	se	df
Marginal mean	Main plot (Intervention)	1.507	0.071	9
	Split plot (Female)	1.200	0.065	24.6
Cell mean	Female, Intervention	1.289	0.092	24.6
Diff. of marginal means	Main plot	0.377	0.100	9
	Split plot	0.237	0.082	178
Interaction	Int x Gender	0.400	0.164	178
Difference of cell means	M/I - F/I	0.436	0.116	178
	F/I - F/C	0.177	0.164	24.6

Things to notice / confirm general patterns:

- Marginal means and differences have integer df, but small because only 10 blocks
- Every thing else has non-integer df because of Satterthwaite/Kenward-Rogers approximations
- SE for difference of main plot means = $\sqrt{2}\times$ se for main plot mean
- SE for difference of split plot means $< \sqrt{2}\times$ se for split plot mean
  Not much less, because main plot variance component is close to 0 and much less than residual error

- Two different se's and df for comparison of cell means

What if data are not balanced (i.e., unequal sample sizes)?

The actual PA study had very different numbers of boys and girls at each school. No problem analyzing unbalanced split plot data, but you lose even the few "nice" things for balanced data.

To illustrate this, here are is the ANOVA table with EMS and error terms (F statistic denominators) for the full data set (591 kids).

The ANOVA table:

Source	df	E MS	F denominator
Block	9	–	–
Intervention	1	$\sigma^2 + 28.682\sigma^2_{main} + Q()$	0.0092 MS(Error) + 0.9908 MS(block*Int)
Block*Int	9	$\sigma^2 + 28.947\sigma^2_{main}$	–
Gender	1	$\sigma^2 + Q()$	MS(Error)
Gender*Int	1	$\sigma^2 + Q()$	MS(error)
Error	569	–	–

Main plot term is "constructed" so its df is not an integer.

Summary of the "what you need to know", without the why about a split plot design

- The analysis is a lot more complicated, but the computer handles all of that
- Remember to make the main plot error a random effect
- Check that the main plot variance component > 0.
  Big problems (and no clear guidance) when $= 0$.
- Estimates / tests of main plot treatment means have larger se's, lower df, and hence lower power than estimates / tests of split plot treatment means.
- Don't be surprised by non-integer error df, especially with unbalanced data.
- If you look at pairwise comparisons between cell means, don't be surprised by different se and df for different combinations of cells, even in balanced data.