Explanation of split plot code in PA.sas

Really, the only new thing this week is the combination of two old things: specifying random effects (week 3) and specifying factorial treatment structures (weeks 6 and 7). The computer handles all the ugly details. I illustrate this using the PA study.

These data are simulated based on characteristics of a real study. These data are similar to, but not identical with, the example used in class.

Reminder about the study design:

Schools grouped into blocks, each with 2 similar schools. Treatments randomly assigned to schools within block: either intervention or control. PA measured on kids within schools. Second "treatment" factor is gender: boy or girl.

There are two data sets: pa2.xlsx has the balanced subset. This has 5 boys and 5 girls in each school. pa.csv has the full data set. Unequal numbers of boys and girls in each school.

We use proc mixed to analyze data with both fixed and random effects. The plots=residualpanel option (no / ) requests a panel of diagnostic plots. The residuals are called conditional residuals because they are based on the predicted values of the random effects. Here, those the differences among schools (block\*school in the model) that are the main plot errors. The conditional residuals assess the characteristics of the residual errors. Here, that’s the errors for each kid.

The code should look familiar and interpretable:

 All factors are named in the class statement.

 Fixed effects go in the model statement.

 Random effects go in the random statement.

 All “after the ANOVA” analyses, e.g. means, estimates, pairwise comparisons, go last.

The /ddfm=kr in the model statement tells SAS how to compute Denominator Degrees of FreedoM. kr is the Kenward-Roger method. /ddfm=satterth is the Satterthwaite method. They give the same results for a split plot model. When they don’t, the KR method is more reliable. Just make sure you specify one of them. The default is a method that SAS calls containment, which uses the model specification to figure out appropriate error terms. Both KR and Satterthwaite use the structure of the data, specifically how the random effect levels change relative to the fixed effects to figure out the appropriate error terms.

Proc mixed allows all the “after the ANOVA” analyses that you’ve used with proc glm. Two differences from proc glm;

 Lsmeans gives you standard errors by default, so /stderr is not needed.

 You get all pairwise differences by /diff and the output is organized a lot more conveniently.

Proc mixed provides multiple methods to estimate variances of the random effects. The default is a method called REML that is the generally recommended method. If you want to get expected mean squares, add method=type3 to the proc mixed statement. The second invocation of proc mixed demonstrates this. You get a ANOVA table that includes the expected mean squares and names the error term used for each F test.

For data from a split plot design when all estimated variances are positive, i.e. > 0, REML and using EMS, give the same results. When a variance is estimated as 0, look back at the notes for week 3 (subsampling) to see what happens when some estimated variances are 0. That is a serious concern and needs to be addressed. If there isn’t some other issue, e.g., outliers or unequal variances, there are two easy ways to get a valid analysis:

 Use method=type3, which can estimate a negative variance

 Add nobound to the proc mixed statement. This uses REML but allows a negative variance

When the main plot design is CRD:

You still need to indicate the main plot error. This requires a variable that indicates each school. I’ll demonstrate using the PA data, i.e. ignoring blocks. Bs is the variable that uniquely identifies each school. Remove block from the model statement and put bs in the random statement instead of block\*school. If you use ddfm=kr or ddfm=satterth, SAS uses the structure of the data to figure out that bs is nested in school. You can also explicitly indicate that by putting bs(school) in the random statement. If you use the default containment method for df, you must indicate the nesting explicitly.