Explanation of SAS code in blockex.sas: Analyzing data from an RCBD

This document explains the code in blockex.sas. This analyzes a small study using an RCBD with 3 treatments and 10 blocks. The treatments are labeled T1, T2 (the two treatments) and C (control). The blocks are numbered 1, 2, ... 10.

First, we read the data from the .csv file. Nothing new here.

**Blocks as fixed effects:**

When blocks are fixed effects, we can use proc glm to analyze the data. Both the block and treatment variables define groups, so both are named in the class statement. The model statement looks like the model statement for a 1 way ANOVA, but it has both block and treatment on the right hand side. Because SAS produces partial (Type III) sums of squares, the order of terms doesn’t matter.

The proc glm statement includes the option plots=diagnostics, which produces a panel with 9 diagnostic plots, including the plot of residuals vs predicted values. That table is in the middle of the output.

The output includes three tables that look like ANOVA tables. These can be understood by remembering that every F test corresponds to a comparison between two models.

The first one (with model and error) looks the most like an ANOVA table. This compares the entire model (block + trt) to an intercept only model. This question is rarely of interest in any study, and especially not in a block design, because you expect there to be differences among blocks. The most useful number in this table is the error MS, which is the pooled error variance. The error SD = sqrt(Error MS) is reported as Root MSE in the table immediately below the Overall ANOVA table.

The third table (2nd that looks like an ANOVA table) has a column labelled Type I SS. This gives results for sequential tests. You start with just an intercept. The first line compares the model with that term to the intercept only model. Here, that is blocks compared to intercept. The second line compares the model with that term added to the model without it. Here, that is blocks + trt to blocks. That is the test you care about, because that is the test of differences among treatments after adjusting for any differences among blocks.

The last table (3rd that looks like an ANOVA table) has a column labelled Type III SS. This gives results for partial tests. These compare the model with all terms to the model with in the indicated term. Here, the block line compares block + trt to trt. The trt line compares block + trt to block. You see that the last line in both the type I and type III tables is comparing the same pair of models, but the first line is not.

When the data are balanced, i.e. the same sample size for each combination of block and treatment, the type I tests and type III tests are identical for all terms. I use this as a quick check for balance.

You can add any ‘after the ANOVA’ operations after the model statement. No change here from what you did with a 1 way ANOVA.

It's almost always a good idea to check assumptions. My quick check is a plot of X=predicted values vs Y = residuals, which you can get two different ways. Adding plots=diagnostics to the proc glm statement produces a panel of 9 plots. The top left one is residual vs predicted value. If you only want that plot and not the other 8, then you can store residuals and predicted values using an output statement, then plot those values.

You want a "flat fat sausage". A trumpet shape or an unusually large residual are signs of trouble (unequal variances or outlier). This plot looks fine to me.

The syntax of the output statement is out= <<dataset name>> followed by keyword = <<variable name>> pairs. You provide data set and variable names but you must use the keywords for the quantities you want. The keyword for saving residuals is r; the keyword for saving predicted values is p. There are lots of other quantities you can store. The help page for the proc glm output statement lists them all.

**Blocks as random effects:**

We combine ideas from 1 way ANOVA and variance components. The model is fit using proc mixed. Both variables are in the class statement. As with last week, the fixed effects go in the model statement and the random effects go in a random statement. I add /ddfm=kr (or equivalently for simple models /ddfm=satterth) to the model statement so SAS computes appropriate error degrees of freedom. The model statement is followed by whatever “after the ANOVA” analyses you need.

Differences between proc glm and proc mixed, other than the random statement.

1) Don’t put the same variable in the model and random statements. You’ll really confuse SAS.

2) The default residual\*predicted value plot is plots=residualpanel in proc mixed but plots=diagnostics in glm.

3) There is no output statement in proc mixed. You can save residuals and predicted values by adding the outp=<<dataset name>> option to the model statement. This creates a data set with a Resid variable, a Pred variable, and some others (e.g. se pred, ci for pred). You choose the data set name, but you can’t choose the variable names.

The residual vs predicted value plots for the fixed effects and random effects analyses don’t look the same. That’s because block effects are calculated differently in the two models. Usually, it doesn’t make any difference to the interpretation of the plots.