AdoptIQ.sas: SAS Analysis of data from a two way factorial using contrasts and effects models

The example is the French nature/nurture IQ study. The data are in adoptiveIQ.xlsx. Background on the study is in the background .docx file.

Proc import can read data files in many different formats. We’ve used it to read .csv files. It can also read .xlsx files. Options allow you to read specific worksheets in the file and portions of a worksheet. Not needed here.

**Creating a variable with the combination of factor levels:**

To use contrasts, we need a variable that indicates each cell (combination of factor levels). This can be created “on the fly”, see below, or we can create a variable with those combinations. The easiest way to create a new variable is to concatenate two character strings. This has to be done in a data step. If you’re using a data step to read data, you can add code to compute new variables below the input statement. If you use proc import to create the data set, you need a separate data step.

Data adopt2; creates a new data set.

 Set adopt; by reading observations from a pre-existing data set.

ab = adoptive || '/' || biological; concatenates the value in adoptive, a / character, and the value in biological. || is the SAS ‘concatenate strings’ operator. You can see the result by printing the first few, here 5, observations in adopt2. Note the (obs=5) in the proc print.

**Fitting a one-way ANOVA:**

The basic code should be familiar by now. Proc glm. The variable identifying the groups goes in the class statement and on the right hand side of the model equation. The response variable is on the left hand side of the =. The estimate statements give you estimates, standard errors and tests for each quantity.

New things:

/clparm on the model statement. Requests confidence intervals, 95% by default, for all parameters. This includes regression coefficients and (what we want) results of estimate statements.

Contrast statement. Syntax looks like an estimate statement. Returns the SS for the contrast. You can specify multiple contrasts separated by commas and SAS will test the hypothesis that all contrasts = 0. For example, the last contrast (3df for groups) tests the hypothesis that each estimate = 0. This is equivalent to the 3 df test of ab in the one-way ANOVA. Because the data set is unbalanced, you can not add the ss for each component contrast because they are no longer orthogonal.

The next proc glm demonstrates creating all combinations of adoptive and biological “on the fly”, without using a data step to create a variable. Adoptive\*biological means all combinations. When that term follows the main effects in a model, it represents an interaction. When it is by itself, it is the equivalent of the ab variable.

**Fitting an effects model, for a 2 way ANOVA, and getting estimates:**

We want terms for the main effect of adoptive, the main effect of biological, and their interaction.

The class statement names both factors, but you don’t need/want the interaction here. Just the variables.

The model statement includes the main effects (the individual factor variable names) and their interaction, specified using \*.

There is a short cut: the | character stands for the specified combinations and all constituent main effects. I prefer to write out the terms so I can put them in the order I want, but that is just my preference. And, this only makes a difference when there are 3 or more factors in the model.

The lsmeans statement will give you marginal means (e.g. adoptive or biological) or cell means (adoptive\*biological). You have to request what you want.

The nice thing about the effects model specification is that you can write contrasts in terms of the marginal means, e.g. adoptive 1 -1, or the cell means, e.g. adoptive\*biological 1 -1 -1 1. In all cases, the order of coefficients matches the order of groups used by SAS.

Make sure to check the ordering in the class statement (for main effects) or the lsmeans statement (for main effects or combinations of factors).

**Estimating simple effects:**

This is hard in proc glm; because of the way simple effects relate to the parameters in the model. The easiest method is to use a lsmestimate statement, which is only available in proc mixed or proc glimmix.

The proc mixed code for a two way factorial looks just like the proc glm code. Since there is no random statement, all effects are considered fixed effects. The code file illustrates other minor differences between proc glm and proc mixed.

The lsmestimate statement allows you to specify contrasts in terms of cell means (or marginal means) instead of model parameters. For many quantities of interest, these two forms are equivalent. They are not for simple effects.

The order of things in an lsmestimate statement is different from that in an estimate statement. The model term being considered comes first. For simple effects, you need to look at each cell mean, so that is adoptive\*biological. Then you give a label and then the coefficients. If you want to compute multiple estimates using the same quantities, all you need to do is provide another label and set of coefficients.

**Order of statements in proc glm, mixed or glimmix:**

The class statement needs to come **before** the model statement.

Lsmeans, estimate and contrast statements need to come **after** the model statement.