Summary of Lab Week 12

Factorial ANOVA

When there are more than one factor, we have what's called a factorial design, and we run a factorial ANOVA. An example dataset is in "lab12-1.txt". Look at the whole dataset to see how it's arranged. There are three columns (variables): two factors and the dependent variable. Each row represents a different subject, showing which level (group) that subject is in for each factor as well as the subject's score on the dependent variable. The subjects could be ordered in any way we choose, but they're ordered here so that you can see how the factors relate to each other. For all three levels of Factor1, we have a set of subjects at both levels of Factor2. This makes for $3 \cdot 2 = 6$ groups of subjects.

There are three hypotheses we can test using this dataset. First, we can ask whether there's an effect of `Factor1`, meaning whether the means for the three levels of this factor are the same. Second, we can ask whether there's an effect of `Factor2`. Third, we can ask whether the two factors interact, meaning whether the effect of `Factor2` depends on the level of `Factor1` (or vice versa). All of these questions can be answered at once using the `anova()` command.

The first step is to tell R what model we want to test. Once again, the model is a way of expressing what variable we want to explain (`X`) and what kinds of explanations we want to use (`Factor1`, `Factor2`, and their interaction). So, we use the `lm()` function with `X` on the left and the two factors on the right. Each factor is converted to a nominal variable using the `factor()` function. The two factors are separated by a `*` to indicate we want to include the interaction (to leave out the interaction, we would use `+`).

```
> model = lm(d$X ~ factor(d$Factor1)*factor(d$Factor2))
```

Now input this model to `anova()`. You get the same kind of table as before, but with more lines. This time, there are three ways of explaining variability in the outcome, and each gets its own line showing the sum of squares, mean square, F, and p-value. Each line works in the same as in a simple ANOVA; the only difference is that the sums of squares are more complicated to compute (especially for the interaction).

Interactions

The test of the interaction in the previous example shows that the effect of `Factor2` depends on the level of `Factor1` (and vice versa). Let's look at this interaction graphically. First, make a matrix of the means of all six groups. The rows of the matrix will indicate the level of `Factor1`, and the columns will indicate the level of `Factor2`. (Remember that you'll have to create the variable for your matrix before you can start defining its components. You can use `M = matrix(nrow=3,ncol=2)` to do this.)

```
> M[1,1] = mean(X[Factor1==1 & Factor2==1])  #etc.
```

Once you've made the matrix, you can plot the means. First plot a set of red points showing the mean outcome as a function of `Factor1`, for the groups that have `Factor2 = 1`. Then add a line in blue for the groups that have `Factor2 = 2`.

```
> plot(1:3,M[,1],col='red')
> points(1:3,M[,2],col='blue')
```

The difference between the two lines shows the effect of `Factor2`. When the red line is above the blue line, the average outcome is greater when `Factor2 = 1` than when `Factor2 = 2`. When the blue line is above the red line, the effect of `Factor2` is reversed. Now you should be able to see why there's an interaction. When `Factor1` equals `1` or `2`, the effect of `Factor2` goes in one direction, but when `Factor1` equals `3`, the effect of `Factor2` goes in the opposite direction. In other words, the effect of `Factor2` depends on the level of `Factor1`. Therefore, to explain the differences among all six groups, we can't just talk in terms of effects of the two factors separately (i.e., main effects). Instead, we have to think about the combined effect of the two factors together, i.e. their interaction.