Analysis of Variance

The purpose of this TP is to carry out ANOVA. We will apply 2-way ANOVA to two problems with data in the **HSAUR3** package, one a balance design and the other unbalanced.

As usual (!), make sure that you read the **help** for any new data or functions that you use.

Weight gain data

Before carrying out our data analysis, first load the data, look at it and get an idea of its structure using the command **str**. Then, perform a small data exploration by calculating means and standard deviations of **weightgain** for each combination of **source** and **type**, and making a plot of the design.

```
data("weightgain", package = "HSAUR3")
weightgain
tapply(weightgain$weightgain, list(weightgain$source, weightgain$type), mean)
tapply(weightgain$weightgain, list(weightgain$source, weightgain$type), sd)
plot.design(weightgain)
```

Make sure that you understand how to interpret this output.

Does the assumption of homoscedasticity appear to hold for these data? Do there appear to be differences in mean weight gain for these two factors? Explain.

Now fit an anova model to the data. Here, we will fit the full model, including interactions, and produce the anova table. Before proceeding, write out the mathematical model that you are fitting.

```
wg.aov <- aov(weightgain ~ source*type, data=weightgain)
summary(wg.aov)
```

and extract the estimated model coefficient via:

```
coef(wg.aov)
```

By default, the model identifiability constraint is *treatment constrasts*, which you can see in **R**:

```
options("contrasts")
```

Treatment contrasts fit with a model where the coefficient for the first level of a model is set to 0. Thus, the interpretation of a given coefficient is that it estimates the *difference* between the effect of the first level and the level corresponding to the coefficient. (It is possible to change this if it makes more practical sense to use a different constraint for your problem \- see the help for **contrasts**. Here, treatment contrasts are fine.)

To get the levels for each factor (and their order) apply the command **levels** to each factor (**weightgain$source** and **weightgain$type**). By default, factor levels are in alphanumeric order, but if you prefer to change the order you can do that as well (google 'reorder factor levels R' to find several ways to do this). We don't need to do that here.

What are the interpretations of the model coefficients? [*Hint*: Your interpretation should correspond to the design plot you made.

The interaction term is marginally significant. You can examine the interaction graphically by making an interaction plot:

```
interaction.plot(weightgain$type, weightgain$source, weightgain$weightgain)
```

Again, make sure that you understand how to interpret this plot. Does the plot support the anova results you obtained?

Now make some diagnostic plots to assess the validity of the model assumptions:

```
layout(matrix(1:4,ncol=2))
plot(wg.aov)
```

Summarize your findings for your analyses of the **weightgain** data.

Foster feeding data

Here, you will repeat the analyses you did above for the **foster** data (also in the **HSAUR3** package).

Unlike the **weightgain** data, these data are unbalanced. Make sure that you understand why **weightgain** is balance and **foster** is not.

Carry out the same exploration on this dataset that you did above.

We will again fit the full model, but because the data are unbalanced we want to see how much the order of the variables affects inference. So you should examine the anova tables for the following 2 models:

```
model.1 <- aov(weight ~ litgen * motgen, data = foster)
model.2 <- aov(weight ~ motgen * litgen, data = foster)
```

Also, verify that you get the same results for the **weightgain** data when you change the order of the variables.

Are there large differences between the anova results for the 2 models? Are there any significant effects? Which ones?

Using model 1 as the basis of *post hoc* multiple comparisons, compute the Tukey honest significant differences:

```
foster.hsd <- TukeyHSD(model.1, "motgen")
foster.hsd
```

Make sure that you understand the output (remember our friend **help**?). There is a plot method for this object:

```
plot(foster.hsd)
```

Which differences are significant?

Summarize your findings for your analyses of the **foster** data.