

NRES 798 — Lab 11

Multiple comparisons, etc.

1 ANOVA, means, comparisons

1.1 Plant growth

Load the `PlantGrowth` data set: `data(PlantGrowth)`. Examine it (`?PlantGrowth`, `summary`, etc.).

Do an ANOVA to test the effect of treatment. Write the model equation:

$$Y_{ij} = \mu +$$

(what are i and j ?). Get the ANOVA table. Is the effect statistically significant?

Estimate the expected weight for each treatment (a) by computing the mean for each, and (b) using the parameter estimates from `lm`:

Control	Trtmt. 1	Trtmt. 2

Use Tukey's HSD test to compare the three treatment levels (`tukeyHSD`). Understand the output.

`plot` the result from `tukeyHSD`. What does it mean?

1.2 Crabs

Load the `crabs` data: `data(crabs, package='MASS')`. Inspect.

Add to the data frame a variable `size`, the product of carapace length, width and depth in cubic centimetres.

Do the same as with the plant growth data before, but now looking at the effects of species and sex on size. Include an interaction term. The table of estimated sizes should look like

	Female	Male
Blue		
Orange		

Visualize with `interaction.plot`.

2 Inside *R*

Wonder why plotting categorical data gives boxplots and plotting numerical data gives scattergrams? And `summary` for `aov` gives an ANOVA table but `summary` of an `lm` or of a data frame give something else? These are called *generic functions*, and they simply look at the type of the argument (object) and call a specialized function that knows how to process that object. This is a characteristic of the *object oriented programming* approach. In *R* the specialized functions are typically named as the generic function name followed by a dot and the object type (or “class”): `summary.lm`, `summary.aov`, etc. There are many of these generic functions.

That can help in understanding the *R Help*. Look at `?summary`, `?plot`, `?lm`, in particular the section *See Also*. Look over a few more help pages, and see if you spot more instances of this.

3 What else is there?

Load the *R Commander*: `library(Rcmdr)`. Explore the menus, from left to right. In *Data* see the data sets available in various packages. Select one as “active” to play with, perhaps `crabs`. Watch the commands typed by *R Commander* and the results in the main *RStudio* console.

Under *Statistics*, there are summary statistics and hypothesis tests for normality, comparing means, proportions, variances. Also analysis of contingency tables (categorical = f (categoricals)), and “dimensional analysis” (various multivariate analysis and classification methods, some of it black magic used in plant ecology “ordination” and in psychology). *Fit models* includes regression, ANOVA/ANCOVA, and other related methods. Define a linear model to play with, perhaps the crabs ANOVA from before; it becomes the active model, button on upper-right.

The *Models* tab includes many of the things that we have seen, plus a few others. For any unfamiliar item, click on it, see the command in the console and look at its help page. Note that some are functions specific to the *R Commander*, mostly simple convenience wrappers; type the name and *Enter* to display and see what lies behind.

Distributions is just for the *p*-, *q*- and *r*- functions of the various distributions. There are also lots of *Graphs*, and other menu-driven command generation conveniences.

Explore and experiment. This is the tip of the iceberg, but it covers the methods most commonly used, plus quite a few more esoteric ones.