STAT 213: One-Way ANOVA as Multiple Regression

Colin Reimer Dawson

Last Revised April 4, 2016

This lab is a modification of Exercises 5.37 and 7.34.

Sea slugs live on vaucherian seawed, but the larvae from these sea slugs need to locate this type of seaweed to survive. A study was done to try to determine whether chemicals that leach out of the seaweed attract the larvae. Seawater was collected over a patch of this kind of seaweed at 5-minute intervals as the tide was coming in and, presumably, mixing with the chemicals. The idea was that as more seawater came in, the concentration of the chemicals was reduced. Each sample of water was divided into 6 parts. Larvae were then introduced to this seawater to see what percentage metamorphosed. Is there a difference in this percentage over the 5 time periods? The data can be found in the `SeaSlugs` dataset from the `Stat2Data` package. The `Time` variable measures the number of minutes since the tide came in, and the `Percent` variable measures the proportion of a sample of 15 sea slug larvae (ranging from 0 to 1, not 0 to 100) that metamorphose during the time period in question.

(a) First, fit a simple linear regression model, using `Time` as a predictor of `Percent`. Write down the predicted `Percent` for `Time = 10`. Do you have any concerns with the model? Explain.

(b) If we treat `Time` as a categorical variable, we can fit an ANOVA (multiple means) model. Write down the form of this model and then fit it in R (remember that you will need to use `factor()` to tell R to treat `Time` as categorical). Plot the means (using `xyplot(..., type = c("p","a"))`) and write down the predicted `Percent` for `Time = 10`. Check the ANOVA conditions, test whether there is evidence that the means differ in the population, and describe your conclusions.

(c) Again treating `Time` as categorical, write down and fit a multiple regression model with indicator variables for the different time periods (you can let `R` create the indicators for you by simply doing `lm(Percent ~ factor(Time), ...)`). What is the meaning of each coefficient? What do the t-tests in the `summary()` output tell you? Write down the predicted `Percent` for `Time = 10`.

(d) What are your overall conclusions?