

STAT 215 Lab 7: Indicators and Interactions

Colin Reimer Dawson

Last Revised October 27, 2017

This lab is a slight modification of Exercise 3.36 in the text. Write up and turn in your solutions as an RMarkdown document. Key R commands are listed at the end.

The dataset `MetabolicRate` in the `Stat2Data` package contains data about caterpillars. The metabolic rate of caterpillars is related to their body size. We would like to examine whether the relationship between size and metabolism changes throughout the caterpillar's life stages ("instars"). Between each instar, the caterpillar molts (sheds the outer layer of its skin). It does this five times before forming a cocoon or chrysalis in which it will finally transform into a moth or butterfly.

- (1) Fit simple linear regression models predicting metabolism `Mrate` using body size `BodySize`. Consider log transforms of either or both variables. What combination appears to give a model that best satisfies the regression conditions?
- (2) Plot the variables in your model using a separate plotting symbol for each `Instar`. You can add the argument `groups = factor(Instar)` to the `xyplot()` function to get this. Does the relationship between body size and metabolism appear to be different at different stages of life?
- (3) Fit a multiple regression model to predict `LogMrate` using both `LogBodySize` and `Instar` as predictors (treat `Instar` as categorical using `factor()`). R will automatically create multiple indicator variables out of the factor. Check the conditions of the model, and describe any violations you see (if any).
- (4) Notice that there are five levels of `Instar` but only four indicator variables included in the model. Why?
- (5) Interpret the role that each coefficient plays in the model.

- (6) Fit a new model that adds interaction terms (you can simply ask for the interaction between `LogBodySize` and `factor(Instar)`, and all relevant terms will be generated.) Interpret each coefficient.
- (7) Perform three nested F -tests to examine whether the non-parallel lines model is a significant improvement over the parallel lines model, whether the parallel lines model is a significant improvement over a single line model, and whether the non-parallel lines model is a significant improvement over the single line model.

Key R Commands

Usage patterns for some of the most important commands for multiple regression are given below. All-caps elements are placeholders.

```
### To fit a model with multiple predictors (no interaction terms)
MyModel1 <- lm(Response ~ Explanatory1 + Explanatory2 + ETCETERA,
               data = DataSetName)

### Fit a model with two predictors where Explanatory2 is categorical
MyModel2 <- lm(Response ~ Explanatory1 + factor(Explanatory2),
               data = DataSetName)

### Fit a model with an interaction between Explanatory 1 and 2
### where Explanatory 2 is categorical
MyModel3 <- lm(Response ~ Explanatory1 + factor(Explanatory2) +
               Explanatory1:factor(Explanatory2), data = DataSetName)

### Plot (1) residuals against fitted values, (2) Q-Q plot of residuals
### and (4) Cook's distance plot
plot(MyModel, which = 1)
plot(MyModel, which = 2)
plot(MyModel, which = 4)

### Histogram of residuals (could use dotPlot instead)
histogram(~residuals(MyModel), VARIOUS_OPTIONS)

### Nested F-test comparing Model1, which is nested in Model2
anova(Model1, Model2)
```