Homework 2

MLRM 2009

Instructions. These questions are based on material in G&H , Chapter 4, and lecture and reading on linear regression, transformations, and diagnostics. Answer the following questions. Show your R code, your input, and your output. Feel free to ask me for hints if you get stumped.

1. In discussing transformation of variables, Weisberg mentions two rule, the *log rule* and the *range rule*.

   - *The log rule.* If the values of a variable range over more than one order of magnitude and the variable is strictly positive, then replacing the variable by its logarithm is likely to be helpful.
   - *The range rule.* If the range of a variable is considerably less than one order of magnitude, then any transformation of that variable is unlikely to be helpful.

   Cook and Weisberg discuss applying Box-Cox transformations two either the y or x variable, or both. They mention two additional easy-to-remember rules that can make manipulating the value of λ more straightforward. Their rules are:

   - To spread the small values of a variable, make the power λ smaller.
   - To spread the large values of a variable, make the power λ larger.

   Here is a famous example that is discussed in many regression textbooks. I want you to try it. Load the `alr3` library (You will have to download it and install it on your system if it is not already there.)

   ```
   > library(alr3)
   ```

   Then load the brains data.

   ```
   > data(brains)
   > attach(brains)
   ```

   Examine the structure of the file, the names, etc. Note these data are strictly positive.

(a) Do these data satisfy the log rule? Do they satisfy the range rule?

(b) Draw the scatterplot for predicting `BrainWt` from `BodyWt`. Do the small values of `BodyWt` need spreading? Do the small values of `BrainWt` need spreading? How about the large values?

(c) Try transforming both variables until you achieve a good linear fit.

(d) Test your resulting fit for linearity and equal variance using the test discussed in lecture.

2. Exercise 4.1 in Gelman and Hill, i.e., the exercise that begins: Logarithmic transformation and regression: consider the following regression equation

$$\log(\text{weight}) = -3.5 + 2.0\log(\text{height}) + \text{error}$$

Hints: Assume that the regression model as fitted actually satisfies the assumptions of the linear regression model. In part (a), assume that "factor" means "multiplicative factor." In part (b), consider using simulated data to help you draw the scatterplot.

3. G&H Exercise 4.3 *Hint.* Recall how to use the `curve` function in R to add plots of curves to a graph.

4. G&H Exercise 4.4 *Hint.* The data file *pollution.dta* is a Stata file. To read the file in R, put the file in your working directory, then use the following command sequence:

```
> library(foreign)
> pollution.data ← read.dta("pollution.dta")
> attach(pollution.data)
```

5. G&H Exercise 4.5

6. G&H Exercise 4.6

7. Weisberg, Exercise 1.2 *Hint*: You can manipulate the aspect ratio directly with the `asp` option in the `plot` command.

8. Weisberg, Exercise 1.3

9. Weisberg, Exercise 2.1

10. Weisberg, Exercise 2.4