STAT 420: Homework #7

Spring 2015, Dalpiaz

Due: March 20, 3:00PM

You may use `R` for all problems on this homework. Please circle or highlight relevant output where possible. Perform any tests at an $\alpha = 0.05$ significance level.

Problem 1

The data set `mammals` from the `MASS` package contains the average body weight in kilograms (x) and the average brain weight in grams (y) for 62 species of land mammals.

```
library(MASS)
data(mammals)
```

```
View(mammals)
?mammals
```

(a) What are the smallest and largest body weights in the dataset?

(b) What are the smallest and largest brain weights in the dataset?

(c) Plot average brain weight (y) vs. the average body weight (x).

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.

(d) Since the body weights do range over more than one order of magnitude and are strictly positive, we will use log(body weight) as our predictor. Use the Box-Cox method to verify that log(brain weight) is a "recommended" transformation of the response variable. That is, verify that $\lambda = 0$ is among the "recommended" values of λ when considering,

$$g_\lambda(y) = \beta_0 + \beta_1 \log(\text{body weight}) + \epsilon$$

Please include the relevant plot in your results.

(e) Fit the model justified in part (d). That is fit a model with log(brain weight) as the response and log(body weight) as a predictor. Plot log(brain weight) versus log(body weight) and add the regression line to the plot. Does a linear relationship seem to be appropriate here?

(f) Use a QQ-plot to check the normality of the errors for the model fit in part (e).

(g) Use the model from part (e) to predict the average brain weight of a Hippopotamus. (average body weight 2300 kilograms) Construct a 95% prediction interval.

Problem 2

The data set `longley` from the `faraway` package contains macroeconomic data for predicting employment.

```
library(faraway)
data(longley)
```

```
View(longley)
?longley
```

(a) Find the correlation between each of the variables in the dataset.

(b) Fit a model with `Employed` as the response. Calculate the variance inflation factor for each of the predictors. Do any of the VIFs suggest multicollinearity?

(c) What proportion of observed variation in `Population` is explained by a linear relationship with the other predictors?

(d) Calculate the partial correlation coefficient for `Population` and `Employed` with the effects of the other predictors removed.

(e) Fit a new model with `Employed` as the response and the predictors from the model in (b) which were significant. Calculate the variance inflation factor for each of the predictors. Do any of the VIFs suggest multicollinearity?

(f) Use an F-test to compare the models in parts(b) and (e). Which do you prefer?