IEOR 4574: Business Analytics for OR Professor Jacob Leshno
Spring 2016 Professor Adam Elmachtoub

Assignment 1: Linear Regression

Due date: February 4, 11:59pm

**Attention: Please prepare two files for each homework assignment: a .pdf file
for your answers including relevant figurs, and a .R file for your relevant R
scripts. File names should be `Last_First_hw.pdf` and `Last_First_hw.R`, e.g.,
`Obama_Barack_1.pdf` and `Obama_Barack_1.R`. Your submissions must be based on
your own original work. Late submissions will not be accepted.**

In this problem you will need to analyze the EggProduction dataset. The data includes
daily information about egg production (number of eggs laid per chicken), feed (amount of
feed per chicken) and daily temperature (in degrees Celsius). Several of the questions ask
you to interpret the results, you might find it useful to go through the entire analysis before
coming back to these questions.

1. Download `EggProudction.csv` to your computer. You can use `setwd()` to set the
   current working directory. Load the data and print a summary of the variables.

2. Run a regression of eggs on feed and interpret the result. Does the result make sense
   to you?

3. Now include temperature in the model. Run a regression of eggs on feed and temper-
   ature and interpret the result. Does this result make sense to you?

4. Plot feed against temperature to understand how the temperature affects the amount
   of feed the chickens receive. From the graph you should be able to see that the
   amount of feed changes when temperature takes certain values. Create a new bi-
   nary/discrete/categorical variable from the temperature column that captures this
   phenomenon. Print a summary of the new variable.

5. Regress eggs on feed, temperature, and the new variable you created. Interpret the
   results.

6. Randomly and evenly divide the data into a training and test dataset. What is the
   best model you would use to predict egg production based on this dataset? Why?

7. For your best model, what is a 99% confidence interval for the regression coefficients.
   Interpret the results.

8. For your best model, what is a 90% prediction interval if the feed was 25 and the
   temperature was -1. Interpret the results.

9. Provide an intuitive explanation for what changed between the first two regressions
   and the last regression.