STAT 313/513 Assignment 9
Due Tuesday, 6/1

Read the problem description, but answer the following questions rather than those listed in the text. When citing R output, feel free to refer to displays from earlier problem parts. You don’t need to display the same output over and over.

Finding the best treatment for cat fleas, p 229

1. Answer text questions (1) – (3).

2. Plot the density of fleas on the focal cat against each of the predictors in the dataset. (You should have 4 plots). For each plot, comment on any apparent relationships between the response and predictor, and any hints of transformations (in x or y) that might be needed, if any.

3. Fit the additive model: fleas = carpet + ncats + hairlength + treatment and display the model summary (i.e., coefficients etc. using summary). Verify the value of $R^2$ that you computed in part 1(2).

4. Display the following six diagnostic plots (using standardized residuals) for the model. For each plot, describe any patterns or indications of violations of the linear model requirements. Note that not all plots will indicate problems.
   a. Residuals versus fitted values (1 plot)
   b. Normal plot of residuals (1 plot)
   c. Residuals versus each of the four predictors (4 plots)

5. Refit the model in part 3, but transform the response using $\log_{10}$. Note that this model is different than the model in Box 11.10(c) which considered a $\ln$ transformation. Display the model summary.

6. Repeat the plots in part 4 using the new model with the log transformed response. Do they generally look better?

7. Should any interactions be in the model?
   a. Considering the log transformed response, build a model that includes all possible two-way interactions between the predictors. Display the drop1 table and comment on the significance of the interaction terms.
   b. In one test, examine the significance of all of the interaction terms by comparing the models of 7a and 5. Are any interactions significant? Do your results agree with part 7a?

8. Taking into consideration what you have learned above...
   a. Answer text question (4) and describe how you arrived at your final model
b. Use your model to predict the mean flea density for cats that live in homes with 3 cats, have 6mm hair and are allowed on the carpet. (Note, I didn’t specify the flea treatment…how you going to handle that?)

Dangers of Stepwise Regression

Type in the following R code (you don’t need to type the comments):

```
# make 100 random normal y observations
y = rnorm(100)
# make 50 x variables each of 100 random normal observations
x = matrix(rnorm(100*50), ncol = 50)
# put x and y into a dataframe for analysis
# The random data frame has garbage in it. It’s all random
# noise. There are no real relationships
random = data.frame(y, x)
# fit a model using all of the data (i.e., one y, and 50 predictor
# variables (x’s))
m1 = lm(y ~ ., data = random)
summary(m1)
```

1. Based on the summary output, is the model overall a good fit of the data? Are any of the predictors significant? Explain.

2. Now use stepwise regression to find a good model for this noise.

```
m2 = step(m1, dir = 'both')
summary(m2)
```

Is the resulting model significant? Are any predictors significant? How can this be with only noise (random data)?