STATS 415, Homework 2

Due Thursday Jan 25, 2018

Please write your name, uniqname, and section (number, time, or GSI name) on the front page of your homework.

Turn in a printout of your homework in the lecture or in your GSI's mailbox across room 305A West Hall, no later than 5pm on the due date.

This exercise relates to the `Carseats` data set in the `ISLR` package (or it can be found on the book's website). You may use `help(Carseats)` to learn more about the data set.

1. Fit a multiple regression model to predict `Sales` using all other variables in the model. Report the values of coefficients, and how well the model fits (using R^2). Include a plot of residuals and comment on any interesting features.

2. Which variables correspond to significant p-values? What is the null hypothesis the p-values are testing?

3. Drop all the variables that are not significant in the previous model. (Note: this is not the best way to do model selection; we will study better ways later). Fit the linear model with the remaining variables (but no interaction). It will include one categorical variable, `ShelveLoc`. Compare the fit of the model to the previous one using R^2.

4. Use the `anova()` to formally compare the two models and state your conclusion. Comment on the difference between their R^2 in light of your conclusion.

5. Write out the model from the previous question in equation form and interpret the coefficients. Be careful with the coefficients of the categorical variable.

6. Add an interaction term between the categorical variable `ShelveLoc` and the variable `Price`. Refit the model, report the estimated coefficients, and interpret the coefficients of the interaction term. Do the

p-values associated with them suggest the interaction term is necessary?

7. Use the `anova()` to formally compare model from Q3 to the model from Q6 and state your conclusion.

Please limit your solution to at most 7 pages.