STAT 632, HW 3

Reading: Chapter 5, Sections 5.1 and 5.2, from *A Modern Approach to Regression.*
Optional Reading: Chapter 3, pp. 71–92, from *An Introduction to Statistical Learning.*

Exercise 1. For this question use the `Auto` data set from the `ISLR` package. To access this data set first install the package using `install.packages("ISLR")` (this only needs to be done once). Then load the package into R using the command `library(ISLR)`. You can read about this data set in the help menu by entering the command `help(Auto)`.

(a) Make a scatter plot with `mpg` on the y-axis, and `horsepower` on the x-axis.

(b) Use the `lm()` function to estimate a second degree (quadratic) polynomial regression model. That is, fit the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$, where $Y =$ `mpg` and $x =$ `horsepower`. Use the `summary()` function to print the results.

(c) Use the fitted regression model to make a prediction and 95% prediction interval for the `mpg` of a vehicle that has `horsepower` $= 150$.

(d) Add the fitted second degree polynomial regression curve to the scatter plot of `mpg` versus `horsepower`. You may use either the base-R or `ggplot2` approach.

(e) Make a plot of the residuals versus fitted values, and a QQ plot of the standardized residuals. Comment on whether or not there are any violations of the assumptions for regression modeling.

Exercise 2:[1] For this question use the `Carseats` data set from the `ISLR` package.

(a) Fit a multiple linear regression model to predict `Sales` using `Price`, `Urban`, and `US`.

(b) Provide an interpretation of each coefficient in the model. Note that some of the variables are qualitative.

(c) Write our the equation for the fitted model.

(d) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$?

(e) On the basis of the your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

(f) How well do the models in (a) and (e) fit the data?

(g) Using the model from (e), obtain 95% confidence intervals for the coefficients.

[1] From *An Introduction to Statistical Learning*, Exercise 10, with slight modifications