Statistics 350
Assignment 6

1. A study was conducted (Duncan, 1950) used data (North and Hart, 1949) on occupational prestige to see if there was evidence of a relationship between prestige and education, income, and occupation type. The data can be found in *prestige.csv* on Canvas.
    a. Using the data in *prestige.csv* to see if there is evidence that education, income, and/or occupation type impacted the prestige of an application in 1950. Write out the appropriate hypothesis, test statistic, p-value and your conclusion.
    b. Looking the residuals, is there any evidence that there are problematic observations? If so, which professions are these?
    c. Fit a robust regression model to these data with Huber's psi function. Do the same variables appear significant when compared to the usual linear model fit?
    d. Find the weights associated with each of the observations (the rlm command in R returns these as an output named *w* and NOT weights). Plot the weights versus the observation number. Which professions have small weights and is this what you expected to see?

2. **Bootstrap and robust regression**. The data for this question can be found in *assignment6.*csv on Canvas. There is one predictor variable and one response variable. The purpose of this question is to construct a confidence interval using the bootstrap when using robust regression.
    a. Fit a linear model to the data. Report the fitted model.
    b. Looking at the residuals, would you feel comfortable constructing a confidence interval for the slope? Why or why not?
    c. Fit a robust regression model to these data with Huber's psi function. Report the fitted model.
    d. Plot the data (y versus x) and also the two fitted regression lines. With reference to the plot, compare the two fitted lines? Which one fits the data better in terms of the majority of the data.
    e. Since the residuals and errors are not normally distributed for the robust regression, confidence intervals are not readily available in the usual way. Instead, we are going to construct a bootstrap confidence interval for the slope. We are going to Bootstrap *cases* and not residuals
        i. Select a bootstrap sample (sample with replacement the pairs (x,y)) of size n = 1050 from your data.
        ii. Fit a robust regression as above.
        iii. Save the estimated slope.
        iv. Run steps i. – iii. 1000 times.

Construct a histogram of the 1000 recorded slopes (your bootstrap distribution for the slope).

f. The 2.5^{th} and 97.5^{th} percentiles of your bootstrap distribution can be used as an approximate 95% confidence interval. Report your estimated confidence interval.

g. Using the result in part f. test the hypothesis that the slope is different than zero (stat the hypothesis and how you came to your conclusion).