STATS 500 - Homework 2

Due in class Wednesday, September 27

1. Based on Chapter 2, problems 1 and 2 (page 30) The dataset `uswages` is drawn as a sample from the Current Population Survey in 1988.

1. Fit a regression model with weekly wages as the response and years of education and experience as predictors. Present the output.
2. What percentage of variation in the response is explained by these predictors?
3. Which observation has the largest (positive) residual? Give the case number.
4. Compute the mean and median of the residuals. Explain what the difference between the mean and the median indicates.
5. Compute the correlation of the residuals with the fitted values. Plot residuals against fitted values.
6. For two people with the same education and one year difference in experience, what would be the difference in predicted weekly wages?
7. Fit the same model but with log(weekly wages) as the response and interpret the regression coefficient for experience. Which model has a more natural interpretation?

Hints: Useful R functions for the homework: library(), data(), lm(), summary(), residuals(), fitted(), which.max(), mean(), median(), and cor(). Note that the experience variable has some negative values which most likely indicate missing data. Those observations should be removed from the analysis.

Solutions to this problem should be about 2-3 pages long.

2. Suppose have the multiple linear regression model as given in class, i.e.,

$$y = X\beta + \epsilon$$

where β is $(p+1) \times 1$, y and ϵ are $n \times 1$, and X is $n \times (p+1)$. Show that the critical values of β relative to the minimization problem

$$\min_{\beta} \|y - X\beta\|^2$$

exactly corresponds to the solution β of the normal equations, i.e.,

$$X^\top X\beta = X^\top y$$

The critical values are the values of the vector β that ensure that all partial derivatives are equal to 0, i.e.,

$$\frac{\partial \|y - X\beta\|^2}{\partial \beta_j} = 0, \qquad j = 0, \ldots, p$$