PSTAT 126 Homework #2

P1) Explain why the coefficient of determination R^2 satisfies $0 \leq R^2 \leq 1$. Is the same true of the adjusted coefficient of determination R_a^2? Give a reason.

P2) Show, in the context of Simple Linear Regression, that the residuals e_n, $n = 1,\ldots,N$, are normally-distributed if the noise/error terms ϵ_n,$n=1,\ldots,N$,are. Give your reasoning.

P3) For any given, fixed number $N$of samples, explain why and in what sense each $\hat{\beta}_m$, m = 0, ...,M, is an optimal estimator for β_m, m = 0,...,M, respectively.

P4) In R, use the lm( ) function to generate a summary output report with Temp as the response against the predictors Ozone, Wind, and Month, using the "built-in" airquality dataset. Is the regression model overall a significant one that adds insight beyond a simple "intercept-only" model? How do you know? Which of the three predictor variables are deemed important for the regression model? How do you know this? Based on the summary report, what do the residuals say about the validity of the regression model?

P5) With the built-in mtcars dataset, taking mpg as the Y-variable and disp and hp as the x-variables, use R to solve the resulting least squares multiple regression problem for the vector $\hat{\mathbb{B}}$ (see course slides 31-32) by direct computation of the matrix product

$$(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y},$$

assuming invertibility of the matrix $\mathbb{X}^T\mathbb{X}$. To do this, use the R code:

```
n = nrow(mtcars)
p = length(coef(mtcars))
X = cbind(rep(1, n), mtcars$disp, mtcars$hp)
y = mtcars$mpg

(beta_hat = solve(t(X) %*% X) %*% t(X) %*% y)
```

Do you know, using R, another way to do this computation and generate the vector $\hat{\mathbb{B}}$? Do do the computation to produce this vector a different way. Include your code in your answer, and compare the results with those of the other method.

P6) Do Problem 2.13 in the Weisberg (2014) text. Note that "regression of dheight on mheight" in the first part means that mheight is the predictor variable.

P7) Do Problem 2.15 in the Weisberg (2014) text.

P8) Do Problem 2.17.1 in the Weisberg (2014) text. With respect to Problem 2.17.2, simply compute $\hat{\beta}_1$, but you can ignore the rest of it (Hint: Models are fit in R without the intercept by adding a -1 to the formula. Also, you do not need to do Problem 2.17.3.)