STAT 4130: Homework 3

Due: 2023-06-25

Question 1

Prove that for SLR: $y_i = \beta_0 + \beta_1 x_i$, where $i = 1, 2, \ldots, n$

The leverage :

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^{n} (x_k - \bar{x})^2}, \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Question 2

We will work on a dataset with 28 observations on National Football League 1976 Team Performance. This data contains the following columns:

- y: Games won in a 14 game season
- x1: Rushing yards
- x2: Passing yards
- x3: Punting average (yards/punt)
- x4: Field Goal Percentage (FGs made/FGs attempted)
- x5: Turnover differential (turnovers acquired - turnovers lost)
- x6: Penalty yards
- x7: Percent rushing (rushing plays/total plays)
- x8: Opponents' rushing yards
- x9: Opponents' passing yards

```
#To load the data
data = read.table('hw3.txt')
```

2.a Fit a simple linear regression model relating games won y to yards gained rushing by opponents x8.

2.b Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?

2.c Construct the analysis-of-variance table and test for significance of regression.

2.d Find a 95% CI on the slope.

2.e What percent of the total variability in y is explained by this model?

2.f Find a 95% CI on the mean number of games won if opponents' yards rushing is limited to 2000 yards. How about 95% prediction interval?

2.g Construct and interpret a plot of the residuals versus the predicted response.

2.h Plot the residuals versus the team passing yardage, x2. Does this plot indicate that the model will be improved by adding x2 to the model?

2.i Try to build your "best" model with three predictors, and use Residual Plus Component Plot to interpret your work. (Note: this question is open-ended, but your choice must come with some theoretical basis. Don't choose predictors arbitrarily)