Math 189 Homework 5
Due Feb 17th, 2023

Problem 1

The baseball dataset collected the statistics of 263 players in Major League Baseball in the season 1986-1987. This dataset (baseball_5.csv) contains 5 variables selected from the original baseball dataset. The variable names and descriptions are listed in the table below.

Variable	Description
Salary	1987 annual salary on opening day in thousands of dollars
Hits	Number of hits in 1986
Walks	Number of walks in 1986
PutOuts	Number of put outs in 1986
CHits	Number of hits during his career

In this homework, we use Salary as the response variable and the rest four variables predictors/covariates.

1. Draw a scatter plot between Hits and Salary. Consider a simple linear regression using Hits as predictor. Estimate the regression coefficients and their standard errors. Add a line to the scatter plot according to the predicted linear curve. Do you think this line fits the data well? Calculate Residual Sum of Squares (RSS) and R^2.

2. Consider a multivariate linear model using Hits, Walks, PutOuts and CHits as predictors. Report the estimated regression coefficients and their standard errors. Calculate Residual Sum of Squares (RSS) and R^2. Test the marginal effects of each coefficient.

3. Compare the model fitted in 2 and 1 by their RSS and R^2. Test the model adequacy by letting the simple linear model as the null model and the multivariate linear model in 2 as the alternative model. What can you conclude?