MAT453 - Assignment 4

Spring 2025

Assignment Due (by 11:59 P.M.): Sunday, Feb 9

Directions: You may discuss the exercises with other students and with the instructor, but the work you turn in must be your own. You will need the to submit your R code and answers to the questions below in **one** word or pdf file.

Data: The dataset is called *Hitters*. You can run:

```
> library(ISLR)
> data(Hitters)
```

It contains Major League Baseball Stats from the 1986 and 1987 seasons. Each sample represents a player. We are interested in the following variables:

- Salary: 1987 annual salary on opening day in thousands of dollars
- AtBat: Number of times at bat in 1986
- Hits: Number of hits in 1986
- HmRun: Number of home runs in 1986
- Runs: Number of runs in 1986

Exercises: (10 points total) Answer the following questions:

1. Fit a multiple linear regression predicting *Salary* from *AtBat*, *Hits*, *HmRun*, and *Runs*.
2. Refer to the R output in Q1. What is the correct interpretation of the p-value of *Runs* (i.e., 0.830517)? (Not what would you call it, or what would you conclude, or what would you do with it, but what does the numeric value of this number actually mean?)
3. Refer to the R output in Q1. What is the total sum of squares (i.e., SSTotal) of the response *Salary*?
4. Calculate the Bayesian information criterion (BIC) of the model.
5. For a new sample with
$$X_h = [1, 380, 96, 8, 48]^T$$
for the intercept, *AtBat*, *Hits*, *HmRun*, and *Runs*, respectively, calculate the 95% prediction interval.