HW 4 solutions

Stats and sports class

Fall 2019

```r
library(tidyverse)
library(Lahman)
Teams_1 <- Teams %>%
  filter(yearID >= 2000) %>%
  mutate(X1B = H - X2B - X3B - HR)
# Part II: Model assessment

fit_1 <- lm(R ~ X1B + X2B + X3B + HR, data = Teams_1)
fit_2 <- lm(R ~ X1B + X2B + X3B + HR + BB, data = Teams_1)
fit_3 <- lm(R ~ X1B + X2B + X3B + HR + BB + SO, data = Teams_1)
fit_4 <- lm(R ~ X1B + X2B + X3B + HR + BB + SO + CS, data = Teams_1)
fit_5 <- lm(R ~ X1B + X2B + X3B + HR + BB + SO + CS + lgID, data = Teams_1)
fit_6 <- lm(R ~ X1B + X2B + X3B + HR + BB + SO + CS + lgID + SB, data = Teams_1)

options(scipen=999)
```

Question 9

Using the AIC criteria, which of the six models would you recommend for measuring runs scored on a team-wide level? From a baseball perspective, what does your choice suggest about certain measurements as far as their link to runs scored?

Answer: The model with the lowest AIC is best.

In my code:

```r
AIC(fit_1); AIC(fit_2); AIC(fit_3); AIC(fit_4); AIC(fit_5); AIC(fit_6)
```

The lowest AIC is model 6 – although answers may vary depending on which years your database goes up to.

Question 10

One of the coefficients in `fit_5` and `fit._6` is `lgID`. Generate a table of the `lgID` in your data set. What does this variable refer to?

Answer This is the league identification, and depends on if a player played in the American (AL) or National (NL) leagues.

Question 11

Using the code below, the coefficient for `league = "NL"` is negative. Interpret this coefficient. What about baseball's rules make it important to consider which league each team played in? Note: you can google the differences between the American League and the National League to guide you.

```r
library(broom)
tidy(fit_5)
```

Answers

Teams that play in the National League score, on average, 11.4 fewer runs in a season that teams in the American League, holding all other variables constant (or adjusting for all other variables in the model).

Pitchers bat in the National League, so it is not surprising that there are fewer runs scored among these teams.