STAT 360
Dr. Straight
Quiz 3, Due Tuesday, March 21

Note: *You are on your honor to do your own work and not help others!*
Submit a PDF file of your R Notebook.

For this quiz, use the `Pitching` data frame from the `Layman` package. We are interested in classifying pitchers as "low ERA" or "high ERA."

INCUDE LOTS OF COMMENTS! Explain what you are doing and what the results mean.

First, "clean up" the data frame. We are interested in the years from 2010 to 2019. We're going to look only at American League pitchers, and only those who faced a minimum of 100 batters. (We'll treat each "stint" by a pitcher in a season as a separate instance). Select the columns of interest and fix their types, as needed. Note: you should get 2422 rows in your data frame.

1. The median ERA (earned run average) is around 4, so let's classify those pitchers with an ERA less than 4 as "low ERA". Add a column to your data frame with this classification.

2. Form the training and testing data frames. Use an 80% - 20% split.

3. The "null model" simply predicts that every pitcher is "high ERA." What is its accuracy?

4. Perhaps HR (home runs allowed) is a good predictor. Actually, we should normalize this; that is, use HR/BFP (home runs allowed divided by number of batters faced). Build a classifier with this variable as the only predictor; call it `model1`. How does it do (on the training set)? (Compute its accuracy and its confusion matrix.)

5. Determine a numeric variable whose mean value for the low-ERA pitchers is different than its mean value for the high-ERA pitchers (in the sense that the confidence interval for the difference does not contain 0). Using `tally()` or appropriate graphs, show that this variable may be a good predictor. Add this predictor to `model1` to obtain `model2`. How does it do? (In addition to computing its accuracy and confusion matrix, produce a summary using `broom::tidy(model2)`. Interpret it.)

6. Apply `model2` to the testing set. How does it do?

7. Extra Credit: Build a "better" classifier than Dr. Straight!