M119 Exam Chapters 3-5 Do-At-Home Part
P Staley Baseball Stats Analysis

Batting average is a measure of how well a team is hitting. Earned run average is a measure of how well a team is pitching (There is a table of baseball term/abbreviations on the DAH Data page). Each student has been assigned a year to use for the following analysis of baseball statistics (see www.staley-classes.org website under the stats column Exam 2 Do-At-Home Data).

1. **Compare with Given Distribution**. If the long term density curve for major league team-batting-averages is a normal distribution with mean .267 and standard deviation .01 then

a. What batting average would we expect San Diego needs to be in the top 40% of major league teams? (Based on our hypothetical normal distribution)
b. What team batting average would you expect to be Q1? Q3?
c. What percent of teams (over many years) would we expect to have a higher batting average than San Diego had in your assigned year?
d. Same question as above for Boston.

For your assigned year, do the following:

2. **Collect Data**. Make a table with the following columns: team, batting average, earned run average, and win percentage. You will need to use more than one column to compute the win percentage. Print out the table along with your explanation of how you computed the win percentage. Be sure to check your data table for accuracy.

3. **Standardized Scores**. Use your data table to compute means and standard deviations and then use those to compute z-scores for team-batting-averages, team-earned-run-averages and win percentages. Add the z-scores as columns to your table and print out the result.

4. **Predicting Wins from Batting Average**. Using the data for your year, make a scatterplot with batting average as the explanatory variable and win percentage as the response variable. Print out your scatterplot.

a. What is the correlation of batting average to win percentage?
b. Find the linear regression of win percentage to batting average.
c. Based on your regression analysis—if a team had been able to increase their team-batting-average by .005 (one half of one percent) how much would we expect their win percentage have increased?

5. **Predicting Wins from ERA**. Using the data for your year, make a scatterplot with team-earned-run-average as the explanatory variable and win percentage as the response variable. Print out your scatterplot

a. What is the correlation of team-earned-run-average to win percentage?
b. Find the linear regression of win percentage to team-earned-run-average.
c. Bases on your regression analysis—if a team had been able to decrease their team-earned-run-average by .100 how much would we expect their win percentage have increased?

6. **Conclusions?** Considering your analysis in questions 4 and 5, which is better at explaining the variation in win percentage—team-earned-run-average or team-batting-average? How much better? (i.e. quantify your answer in terms of variation).