STA 207 HW-8
Due Date: 12/5 by 10:20AM

Problem: For the MLBStandings2016 data, do the following:

a. [35 points] Regress WinPct on ERA and League and report the fitted model. Interpret the regression coefficient for the League predictor. Make a plot of ERA vs WinPct with separate lines for the two leagues. Is League a significant predictor of WinPct in the presence of ERA. Show hypothesis test for the League predictor.
b. [25 points] For the model regressing WinPct on the predictors ERA and League include the interaction term. Report the fitted model and interpret the coefficients. Make a plot for the two leagues with different intercepts and slopes. Check if the interaction term is statistically significant.
c. [25 points] Regress WinPct on the predictors ERA and Runs including their interaction term. Report the fitted model and interpret the coefficients. Check if the interaction term is statistically significant.
d. [15 points] Perform ANOVA with FM as model with ERA, Runs, and their interaction term and RM as the ERA and Runs model.

Due Date: 12/9/23 by 10 PM
Start Early and Get Help!!!
Be Creative!
Data Analysis Project (One per group)

Rember that the goal of this project is to apply the regression analysis techniques we have covered over the course of semester to your project and to learn communicating your findings in a meaningful way, that is, interpretations of findings in the context of the data. These techniques include:

- Simple Linear Regression and statistical inference
- Multiple linear regression and statistical inferences
- Regression Diagnostics (LINE)
- Goodness of Fit
- Transformations
- Hypotheses Tests (all three)
- Confidence Interval Estimation
- Multi-collinearity
- Unusual Observations: Leverage, Outliers, and Influential Points
- ANOVA
- Regression for Qualitative Variable: Dummy Variables

Do the followins:

I) **Background and Research Question [10 points]:** Share the problem you are working on? Why is this problem interesting and important. State the research goals of the analysis in terms of regression model. Keeping it short and interesting.

II) **Multiple Linear Regression [40 points]**

- Start with a summary of the model you reported in HW-7.

- If some of the assumptions were not met for your MLR (HW-7), show if transformations we have learned can help or not. Your goal is to improve the model in terms of meeting the LINE conditions.
- Detect if there are any points with high leverage, outliers, and influential in the original model. Show all steps in the analysis.
- Remove influential points and re-fit the model to the remaining data Discuss the goodness of fit and LINE conditions as compared to the original model.
- Report ANOVA of model from above and test overall fit of your regression model.
- Discuss if all predictors are statistically significant or not (t-test).
- Fit a reduced model with the highly correlated and statistically insignificant predictors removed. Compare ANOVA for the original model as full model and reduced model. Show all steps and which model would you keep (General F-test).

III) Dummy Variables [40 points]

- Explain your categorical predictor variable and different categories.
- Show how the scatterplots of Y vs X (numeric) varies for different categories in scatterplots.
- Include categorical variable in the MLR model selected above with intercepts different for different categories. Show fitted models for each category, plot of fitted model for each category. Interpret the coefficients in context of your goals.
- Include interaction term in the above model so that you have different intercepts and slopes for different categories. Show fitted models for each category, plot of fitted model for each category. Interpret the coefficients in context of your goals.
- Include all numerical and categorical predictors (interaction term as well). Show fitted models for each category, plot of fitted model for each category. Interpret the coefficients in context of your goals.

Conclusion [10 points]: Discuss the main findings from your analysis (in context of the problem) and how the analysis you showed addresses the research goals.