**B. Brewer, D. Nguyen, A. Scheerer February 13, 2019**

**MSDS 6372 Group Project I**

**INTRODUCTION:**

For our group project, we decided to look at the professional football team New England Patriots. Our goal is twofold:

1) To determine which statistics explain the number of regular season wins for the Patriots since Tom Brady has been playing on the team (last 18 seasons).

2) \*\*\*whatever we are doing for our two-way ANOVA\*\*\*

**DATA DESCRIPTION:**

Our dataset includes all offensive and defensive stats (70 variables) for each of the Patriots’ past 18 seasons (2001 to 2018). Each season is one observation.

**EXPLORATORY ANALYSIS:**

We used SAS to complete our first objective and R to complete our second.

**OBJECTIVE 1:**

**Problem:**

Which variables correlate to regular season wins for Tom Brady’s Patriots?

**Overall Approach:**

We are going to determine which variables are predictors for regular season wins. Then, we are going to create a multiple regression model to predict regular season wins.

**Determining Predictors:**

First, we loaded the dataset into SAS.[1] Then, we performed exploratory data analysis on the data by creating multiple scatterplot matrices to see which if any variables showed a correlation with regular season wins.[2] The reason why we created multiple matrices is that there were 70 variables, and nothing would show if we created one single matrix. We determined that 16 variables showed a correlation with regular season wins. (None of the variables had to be transformed.) We created a scatterplot matrix with these 16 variables.[3]



**Checking Assumptions:**

multiple regression assumpitons:

resiguals are normally distributed (preidctors and response variables don’t have to be)

constant variance

observations are independent

multicollinearity (check thru VIFs)

outliers and leverage (check thru residual diagnostics)

**Model Selection:**

LARS

LASSO

stepwise

leave-one-out cross validation is K-fold cross validation taken to its logical extreme

**Parameter Interpretation:**

RegSeasonWins = 8.567177 + 0.023588(PointsDifferntial)

**Interpretation:**

If the New England Patriots finish the regular season with a 0 point differential, they will win 8.567177 games. 315.11 PD

**Confidence Intervals:**

95% confidence limits intercept 7.09350 10.04086

95% confidence limits pointsdifferential 0.01498 0.03220

**Final Conclusion from the Analyses of Objective 1:**

Bill Belichick is the real goat, not Tom Brady.

**OBJECTIVE 2:**

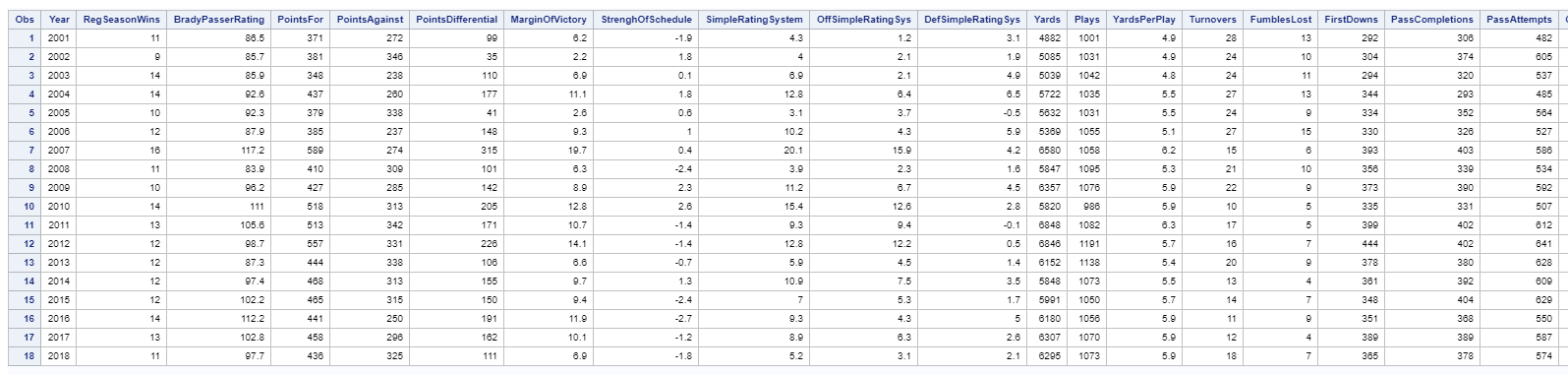
\*\*\*answer to objective 2 here\*\*\*

**APPENDIX:**

[1]

PROC IMPORT OUT= WORK.pats  
 DATAFILE= "/home/daveknockwin0/PatriotsYearlyStats.csv"  
 DBMS=CSV REPLACE;  
 GETNAMES=YES;  
 DATAROW=2;  
RUN;

proc print data=pats;  
run;



[2]

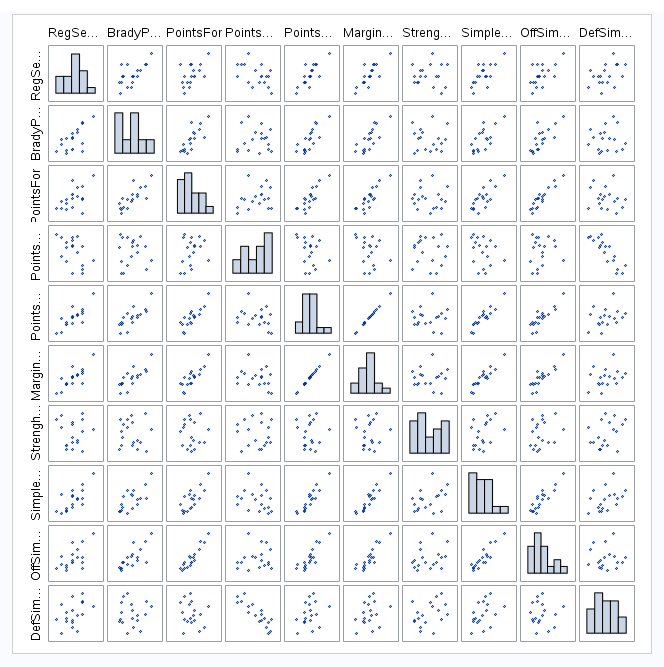
proc sgscatter data=pats;

matrix RegSeasonWins BradyPasserRating PointsFor PointsAgainst PointsDifferential MarginOfVictory StrenghOfSchedule

SimpleRatingSystem OffSimpleRatingSys DefSimpleRatingSys

/ diagonal=(histogram);

run;



proc sgscatter data=pats;

matrix RegSeasonWins Yards Plays YardsPerPlay Turnovers FumblesLost FirstDowns PassCompletions PassAttempts PassYards

PassTouchdowns PassInterceptions NetYardsPerPass PassFirstDowns RushAttempts

/ diagonal=(histogram);

run;



proc sgscatter data=pats;

matrix RegSeasonWins RushYards RushTouchdowns RushYardsPerAttempt RushFirstDowns Penalties PenaltyYards PenaltyFirstDowns

NumberDrives DriveScorePercent DriveTurnoverPercent AvgStartingPosition

/ diagonal=(histogram);

run;



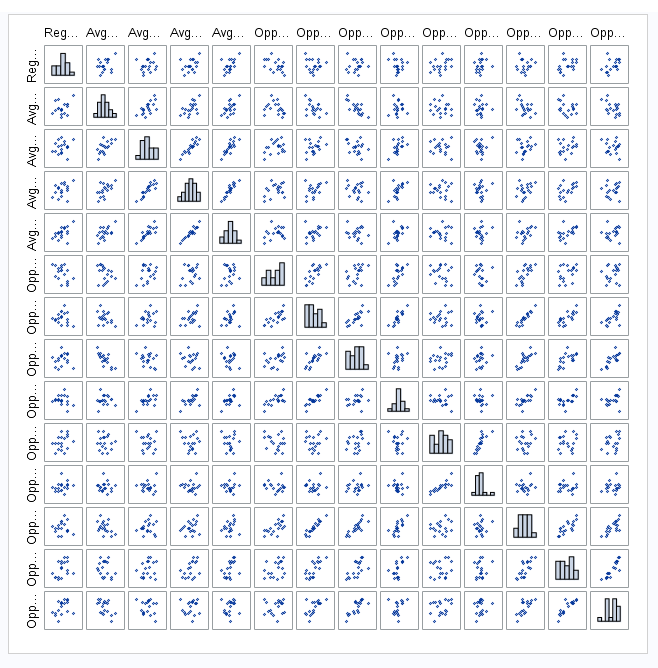
proc sgscatter data=pats;

matrix RegSeasonWins AvgDriveTime AvgDrivePlays AvgDriveYards AvgDrivePoints OppPointsFor OppYards OppPlays OppYardsPerPlay

OppTurnovers OppFumblesLost OppFirstDowns OppPassCompletions OppPassAttempts

/ diagonal=(histogram);

run;



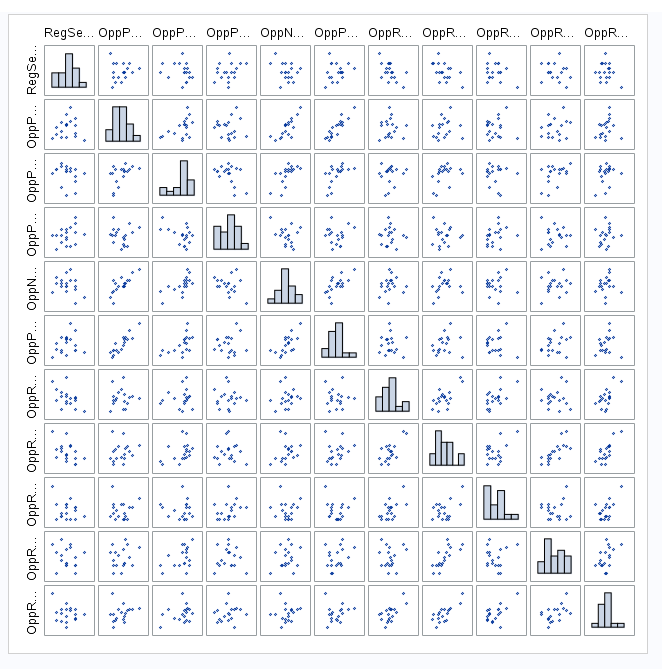
proc sgscatter data=pats;

matrix RegSeasonWins OppPassYards OppPassTouchdowns OppPassInterceptions OppNetYardsPerPass OppPassFirstDowns OppRushAttempts

OppRushYards OppRushTouchdowns OppRushYardsPerAttempt OppRushFirstDowns

/ diagonal=(histogram);

run;



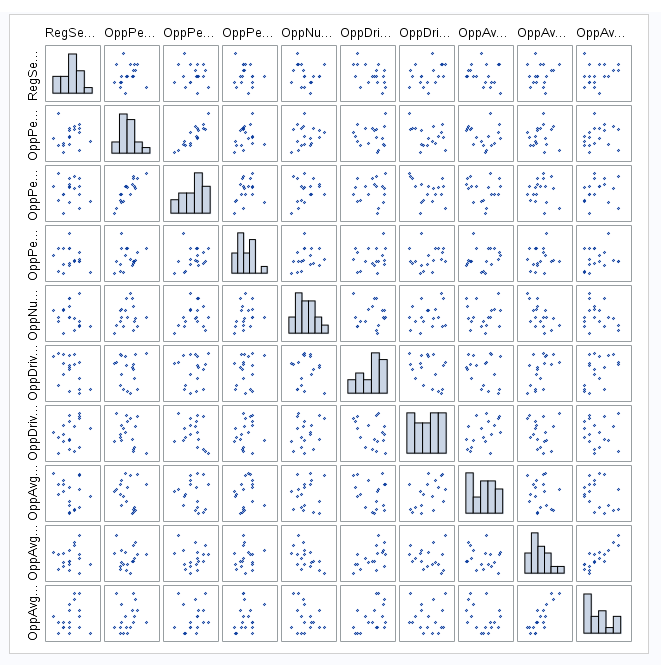
proc sgscatter data=pats;

matrix RegSeasonWins OppPenalties OppPenaltyYards OppPenaltyFirstDowns OppNumberDrives OppDriveScorePercent OppDriveTurnoverPerent

OppAvgStartingPosition OppAvgDriveTime OppAvgDrivePlays

/ diagonal=(histogram);

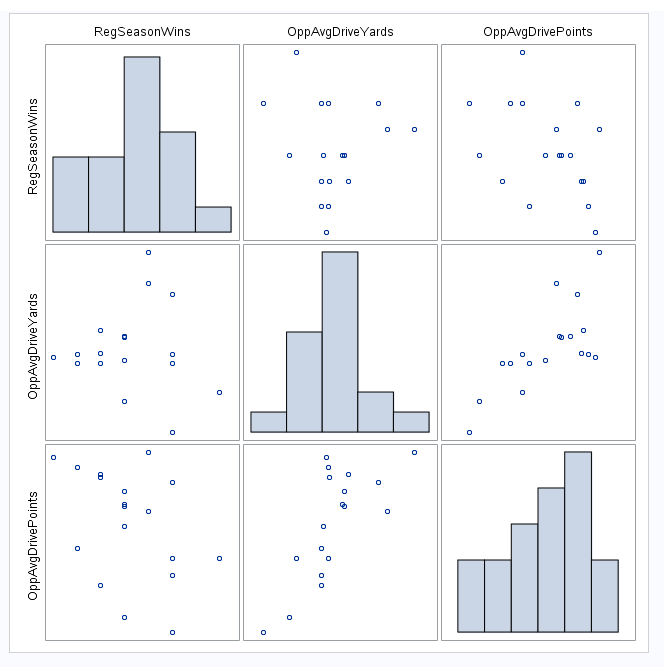
run;



proc sgscatter data=pats;

matrix RegSeasonWins OppAvgDriveYards OppAvgDrivePoints/ diagonal=(histogram);

run;



[3]

proc sgscatter data=pats;  
matrix RegSeasonWins BradyPasserRating PointsFor PointsAgainst PointsDifferential MarginOfVictory SimpleRatingSystem Turnovers FirstDowns   
PassTouchdowns PassInterceptions NetYardsPerPass DriveScorePercent DriveTurnoverPercent AvgDriveTime AvgDrivePoints / diagonal=(histogram);  
run;

