Statistics 2: Project 1

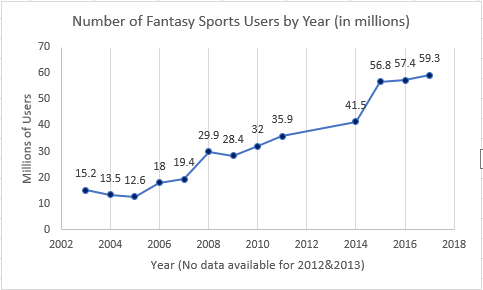
Sachin Chavan

Sean Kennedy

Kevin Thompson

# Introduction

Fantasy sports are a big business. Generating nearly 7 billion dollars annually (cite) and with approximately 60 million players in the US/Canada (cite), fantasy sports – particularly football – have become as big as the sports that they mimic. Players can compete head to head in leagues across a wide range of providers (Yahoo, CBS, ESPN, DraftKings, DraftStreet etc) – each with their own rulesets and stakes. Some are just friendly leagues set up between friends or coworkers, no real stakes other than bragging rights, others have significant monetary rewards for those that can get to the top of the leaderboard.



The explosion of weekly leagues over the course of the last few years has seen an already huge business get even larger. In the weekly cash leagues – each player is given a budget and drafts a completely new team every week. Each draft site uses a predictive model to set player salaries based on the number of points the model predicts that player to score. Budget constraints make it impossible to simply select the players that are predicted to score the most points – hence having a predictive model for which players will generate the best return on investment would be a huge advantage. In the following analysis we will attempt to build position specific models that can accurately predict the number of points a given NFL player is likely to score in a given week.

Data Description **Required**

**--Sean**

Exploratory Analysis **Required**

**--Sachin to create correlagram for target predictors (FantasyPoints)**

**--Sean/Sachin to make box/whisker plots of predictors**

Addressing Objective 1:

Restatement of Problem and the overall approach to solve it **Required**

**--Sean**

Model Selection **Required**

Type of Selection

**Options**: LASSO, RIDGE, ELASTIC NET,

Stepwise, Forward, Backward,

Manual / Intuition,

A mix of all of the above.

Checking Assumptions **Required**

**--TEAM / partition predictors**

Residual Plots

Influential point analysis (Cook’s D and Leverage)

Compare Competing Models **Optional (Helpful if using 2 model strategy)**

Via: Training and test set split or CV

**Possible Metrics**: (ASE, AIC, BIC, adj R2, etc)

Parameter Interpretation

Interpretation **Required**

Confidence Intervals **Required**

Final conclusions from the analyses of Objective 1 **Required**

In addition to overall conclusions, feel free to include additional insights or concerns gleaned from the analysis. What needs to be done next or how could we do it better next time?

Addressing Objective 2

State what route you are going to take 2way ANOVA or Time series and summarize the goal. **Required**

**-TS**

Main Analysis Content **Required**

This will depend on the route you take. I’m leaving it open here to see what you do.

**--ARIMA / lag analysis ACF/PACF**

Conclusion/Discussion **Required**

The conclusion should reprise the questions and conclusions of objective 2.

Appendix **Required**

Well commented SAS/R Code **Required**

Graphics and summary tables (Can be placed in the appendix or in the written report itself.)