Weather and the NFL

Rory Houghton-Berry, Edwin Park, Kathy Pierce

Introduction

When the Seattle Seahawks traveled to Minnesota for the NFC Wild card game in January 2016, they found themselves playing in -6ºF weather. This made the game one of the three coldest games in NFL history. Although it was far colder than the visiting Seahawks were used to, they managed to come away with the win. This type of game could only really be played in the NFL since football is distinct from most other sports in that games will go ahead in all kinds of weather. Since the NFL season takes place from September all the way through to February across the country, as well as in multiple climates, the conditions for each game throughout the season can vary immensely. For instance, Seattle can travel to Minnesota and play in the freezing cold, then travel to the much milder North Carolina the next week, as they did in the 2016 playoffs. Seeing this wide variation of weather conditions possible for NFL games, we decided to evaluate what effect differences in temperature, precipitation or playing surface could have on the results of NFL games, taking into account the strengths and weaknesses of the teams playing.

To perform this evaluation, we needed to isolate each weather condition that we wanted to examine, as well as break each NFL team down into certain types, based on the relative strengths of their passing and rushing offense and defense. Doing this would allow us to perform regressions to assess the impact each variation in the weather had on the games being played, as a function of the types of teams playing. This would allow us to answer questions linking different types of offense and different conditions, such as whether passing offenses do better in colder or warmer temperatures. Answering these smaller questions would then help us apply this knowledge into real NFL situations. For instance assessing how much chance the Seahawks had of beating the Vikings in frigid conditions, or whether the Rams' return to Los Angeles will benefit them. It could also allow us to make predictions about the types of conditions each NFL team would play well in. This could prove especially useful in predicting playoff outcomes, as those games are played in January and February, when we see the greatest disparity in weather conditions across different NFL cities.

Related Work

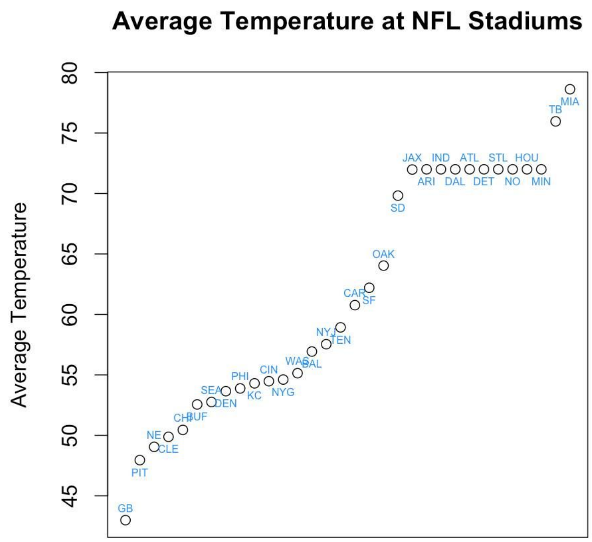
Many other people have attempted to evaluate the effect of weather on NFL games, especially in the lead up to the Seahawks' game in Minnesota. Most of these studies have focused on the effect of temperature alone, especially looking to see whether colder temperatures affect the amount of points scored.[1] Other studies, have chosen to focus on just one aspect of the game, for instance passing, and analyzed how different weather conditions, especially wind and temperature, have affected the effectiveness of the passing games of the teams involved.[2] The study that is the closest to what we are doing is this one posted on flex.io.[3] In this study, the authors analyzed a variety of weather conditions, including humidity and wind chill, but did not consider the playing surface, nor explicitly whether the game was played under a dome or not. From this they were able to ascertain the effect on overall point scoring of each variation in weather conditions. They did not break down the offense and defense into passing and rushing, though, and so only considered how weather affected overall game score, rather than how it affected each team's offense and defense.

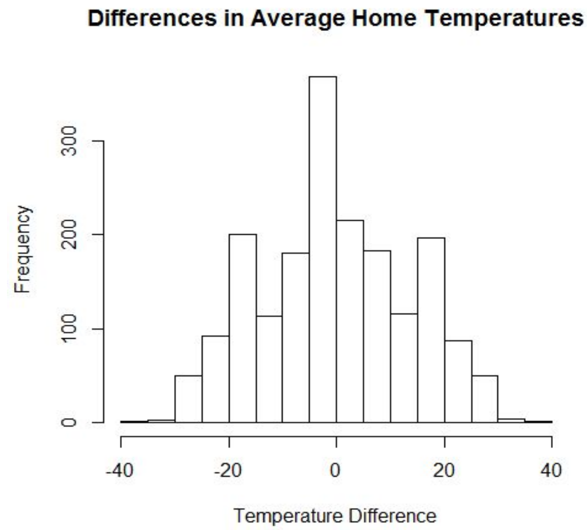
Data

We scraped weather data from *NFLweather.com*, which contains recorded weather data for each game from 2009 to 2015 seasons. Before performing our evaluations on the data, we assessed them to provide some insights on general weather trends in the NFL. What we found was that the temperature varied from the -6°F in Minnesota to 91°F in Jacksonville, the first week of the 2010 season, and the maximum wind speed reached 27 miles per hour in San Francisco in 2009. From the data we were also able to calculate the average home temperature for each NFL team, as displayed below.

---

[1] For instance *http://www.hawkblogger.com/2016/01/effect-of-cold-weather-on-nfl-games.html*
[2] For instance *http://archive.advancedfootballanalytics.com/2012/01/weather-effects-on-passing.html*
[3] *https://www.flex.io/blog/this-american-data-set-act-iii-weather-and-the-nfl/*





As the figure on the left shows, the average home temperatures are fairly evenly distributed, from the 42.6ºF of the Green Bay Packers, to the 78.9ºF of the Miami Dolphins. The multiple data points at 72ºF reflect the teams who play under roofs, since they always play in these standard conditions.

We were also able to assess the difference between the average home temperatures of the two teams in all the games in our data set. As this graph on the right shows, the differences in average home temperature are close to having a normal distribution. This is unsurprising given that the average temperature graph shown above, although quite evenly distributed, does show some clustering around certain temperatures, especially 72ºC. This means that there are multiple teams with little or no difference in their average home temperatures, and moreover, there are few teams at each extreme, meaning it is much more common for two teams with small differences to play than teams with large differences.

From this dataset, we aggregated these weather-related variables: { Temperature, Wind Speed, I(Fair), I(Rain), I(Snow), I(Indoor), I(Grass), I(Turf) }.[4] Since we had seven seasons of data from 2009 to 2015, we decided to use 2009-2014 season data as a training set, and the 2015 season data as a test set, and considered each team from different seasons (i.e. ‘14 Seahawks vs. ‘15 Seahawks) as different teams.

[4] Indicator variables are denoted by I().

Preliminary Methods

For our preliminary analysis, we were interested in first looking at the individual weather aspects and their effect on the score of the game. The weather types examined included temperature (degrees Fahrenheit), wind speed (mph), and indicators for rain, snow, and fair weather (absence of rain, snow, and heavy wind). We also looked at stadium conditions, which included indicators for indoor stadiums and for turf stadiums. We first ran linear regressions with each individual weather aspect as the only regression variable on the different score measurements. We checked the correlations between the weather variables and found that they were not too high to disqualify these analyses from providing preliminary information. We also performed a linear regression using all weather variables as regression variables as part of our preliminary analysis.

Preliminary Results

We regressed our weather variables on four different score measures, total Points Per Game (PPG), Score Differential between Home Team and Away Team, Home Score, and Away Score. We found that the weather variables with a statistically significant effect on PPG were wind speed (-0.26), fair weather (2.69), indoor stadiums (3.27), and turf stadiums (2.44). Wind speed has a negative impact on PPG since wind affects the passing games and field goal attempts of both teams. The significant positive impact of fair weather and indoor stadiums shows that good weather or stable conditions in an inside stadium allows more points in the game overall. Turf stadiums have a positive impact on PPG because it is believed turf allows players to run faster and has a positive impact on rushing yards in particular.

Next, we examined score differential between the home team and the away team to analyze factors that would give the home team an advantage or disadvantage over the away team. Temperature (-0.068) and fair weather (-1.51) have statistically significant negative impacts on the score differential. Although good weather allows more points overall, it is worse for the score differential for the home team. Rain, on the other hand, has a significant positive effect (3.72) on the score differential and provides an advantage for the home team. This advantage comes from the fact that poor weather at home is more likely to be poor weather that the home team is used to, and the home field advantage is able to pull the home team through the tough conditions.

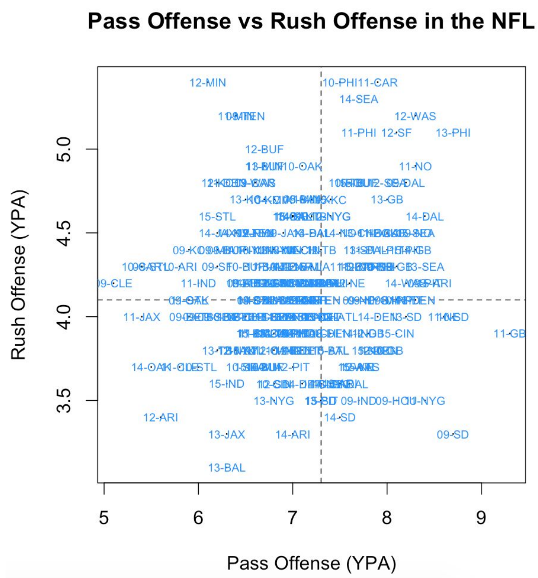
Home score had similar significant regressors as score differential. Away score was positively impacted by good weather in temperature (0.04) and indoor stadiums (1.23) and was negatively impacted by rain (-2.72) and wind (-0.14). Overall, our initial findings showed good weather had a positive impact on PPG and away score, but was a disadvantage for the home team. On the other hand, poor weather conditions like rain and lower temperatures had a positive impact on score differential for the home team and home team score. Wind negatively affected both teams, and indoor stadiums positively affected both teams' scores but provided a slightly larger positive impact on the home score.

We then performed a linear regression using all the weather variables as regression variables in our model. We performed this regression on PPG, Score Differential, Home Score, and Away Score. While there were a number of weather variables that were significant for each of the score measurements, we will only include those that had an impact greater than 1 point. For Score Differential, Rain (3.25) and Indoor stadiums (2.23) had significant positive impacts. Rain fits with our earlier findings that poor weather provides a significant advantage to the home team over the away team. The impact of Indoor stadiums could be due to the large amount of noise or other conditions in indoor stadiums that away teams may not be used to, or home field advantage could be falsely included in the coefficient for Indoor stadiums (discussed in limitations later on). Home score was also positively impacted by Indoor stadiums (1.77) and the same possible reasons discussed above could explain this effect as well.

Methods

For our more detailed analysis, we use regularized *normal* lasso regression model to predict the score differential for each game, and regularized *binomial* lasso regression model to predict whether home team would win, primarily focusing on the effects of weather on the characteristics of each team. For example, how well does a passing team (e.g. Patriots) play in a snowy condition? We use lasso regression, which is a type of regularization that shrinks some coefficients to exactly zero. It is similar to ridge regression, but it can also act as a feature selector, since it would shrink irrelevant coefficient estimates to zero. Since our model involves many predictors, lasso regression is an appropriate regularization to use.

Instead of using the identity of the team, we use strengths of pass offense, pass defense, rush offense, and rush defense. Strength of pass/rush offense is estimated by the average number of yards gained per pass/rush attempts, whereas strength of pass/rush defense is estimated by the average number of yards allowed per pass/rush attempt. To the right is a figure that shows the distribution of pass yards per attempt vs. rush yards per attempt in the NFL through the past six seasons.



To observe how a team would perform in a particular game, we consider both its offense and the opposing team's defense as follows, where PO = pass offense and RD = rush defense, etc:

$$Pass_H = (PO_H - \mu_{PO}) + (PD_A - \mu_{PD}) + \mu_{PO} \quad (1)$$

$$Pass_A = (PO_A - \mu_{PO}) + (PD_H - \mu_{PD}) + \mu_{PO} \quad (2)$$

$$Rush_H = (RO_H - \mu_{RO}) + (RD_A - \mu_{RD}) + \mu_{RO} \quad (3)$$

$$Rush_A = (RO_A - \mu_{RO}) + (RD_H - \mu_{RD}) + \mu_{RO} \quad (4)$$

When only using strengths of offense and defense to predict the score differential, our model is defined:

$$S_i = \alpha + \beta_P(Pass_H - Pass_A) + \beta_R(Run_H - Run_A) \quad (5)$$

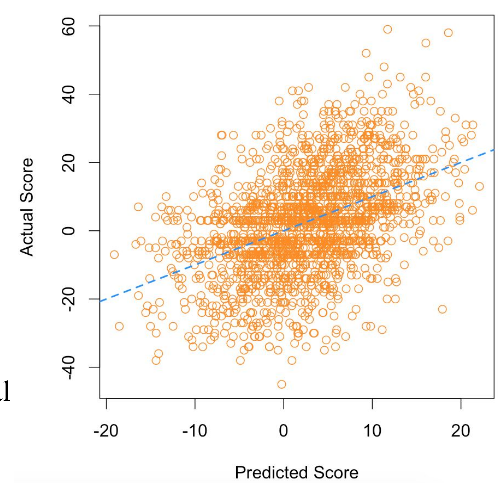
Since our project mainly focuses on the effects of weather on the type of plays, we integrate weather variables with type of plays by including *interaction* variables in our model. In other words, we expand the features in our model by multiplying each weather variable with expected yards gained per attempt. For simplicity, let P = $Pass_H + Pass_A$ and R = $Run_H + Rush_A$, then we have our final model:

$$\begin{aligned} S_i = \alpha &+ \beta_P\,(P) + \beta_R\,(R) \\ &+ \beta_{PT}\,Temp(P) + \beta_{PW}\,Wind(P) + \beta_{PG}\,Grass(P) + \beta_{PT}\,Turf(P) \\ &+ \beta_{PF}\,Fair(P) + \beta_{PR}\,Rain(P) + \beta_{PS}\,Snow(P) + \beta_{PI}\,Indoor(P) \\ &+ \beta_{RT}\,Temp(R) + \beta_{RW}\,Wind(R) + \beta_{RG}\,Grass(R) + \beta_{RT}\,Turf(R) \\ &+ \beta_{RF}\,Fair(R) + \beta_{RR}\,Rain(R) + \beta_{RS}\,Snow(R) + \beta_{RI}\,Indoor(R) \end{aligned}$$

When regularization is applied to this model, we seek to minimize the following equation, where ƛ is a regularization penalty, and cross-validation would be used to obtain the optimal ƛ: $\sum_{i=1}^{n}(y_i - S_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$

Results

After finding the optimal regularization penalty, we achieved the following coefficient estimates (shown below). Using this model, we were able to predict score differentials (favoring the home team) with mean-squared error (MSE) of 12.5 points. Using the normal model, we were able to predict the winner 67.3% of the time, whereas when using the binomial model, we were able to predict the winner 68.6% of the time.



$$\begin{aligned} \alpha &= 2.58 \\ \beta_P &= 3.98 \\ \beta_{PT} &= 0.005 \\ \beta_{PW} &= 0.032 \\ \beta_{PR} &= 0.553 \\ \beta_{PG} &= 0.0075 \\ \beta_{RT} &= 0.034 \\ \beta_{RW} &= -0.138 \\ \beta_{RI} &= -1.504 \end{aligned}$$

The coefficient estimates that are not shown are zeros, as a result of the shrinkage caused by the lasso regression. Our model overall suggests that passing favors passing for gaining score differential than rushing, with expected 4 point-increase per yard increase in expected passing YPA. According to these coefficient estimates, we see that for both passing and rushing (in particular), higher temperatures favor the home team. For example, in a 72-degree weather, a team with average of 5 yards per rush attempt should expect to score 12 more points than the opposing team, but in a 24-degree weather, the same team should expect to score only 4 more points than the opposing team. Furthermore, in windy or rainy weathers, passing is more preferred strategy for gaining score differentials than rushing. On grass fields, passing plays especially seem to be more dominant play than rushing for the home team. On the other hand, for indoors stadiums, home teams should focus more on passing, as rushing would give them a negative score differential, as our model predicts.

Conclusions

There are a number of limitations in our model. Detailed weather data were difficult to obtain and had to be scraped from a website. For example, our dataset had a number of missing data for wind speed and temperature, which we substituted by using the mean for each home field. Moreover, we had some skepticism about the data about inaccurate recordings. If there are multiple weather conditions (e.g. snow and fair) within a game, what weather is recorded? We were also unable to account for the home field advantage, which means that we may have overestimated the true effects of the weather variables and misattributed home field advantage to some of these weather variables' effects. We did not take into account the interaction effects between weather conditions, which may have decreased the accuracy of the model's predictions. Lastly, there may have been multicollinearity among the weather variables, so the true effects of the weather variables may have been misattributed to other weather variables.

Overall, weather factors that seem to be significant in affecting score differentials for home teams are temperature, wind speed, rain, and indoor stadiums. The fact that we were only using YPA as a measure for characteristics of a team, and their combination with weather variables, made it difficult to exactly predict the score differentials, leading to a relatively large MSE. Still, it is interesting to see the effect of weather on the outcome of NFL games, observing what helps and what doesn't. We hope that this project gives more insight about in-game strategies based on weather conditions and best location for relocating stadiums.