STATS 551 Final Project

Projecting Success of NFL Wide Receivers Based on Draft Position and Rookie Year Performance

Group #14
Anthony Paolillo
Seth Corbridge
Krittin Tangboriboonrat

Introduction

The NFL Draft spans over three days every April and are some of the most important days for a NFL front office. After months of intensively scouting hundreds of player's college careers and evaluating their performance at the combine, front offices attempt to best predict the careers of these players in hopes of improving their organization. However, any general manager will tell you that this is no perfect science. Players that ultimately don't live up to their expectations can set a franchise back significantly based on how much draft capital has been invested in them. However, the quicker a team can see that an upgrade is needed, the quicker that a team can move on and replace a specific player.

The motivation for our project is to better project NFL wide receivers by taking into account their draft position and rookie year statistics. With the average salary of this position skyrocketing, it has become increasingly more important to draft rookie receivers, as their contracts for the first 4-5 years are fixed and significantly cheaper. While all players develop at different rates, a year of games under a player's belt can give us a baseline understanding of the ability of a player, more so better than just his college statistics and pre-draft notes. We hope to not only update our beliefs about a player after a rookie season, but also give teams a better idea of when they should draft receivers for maximum future gain.

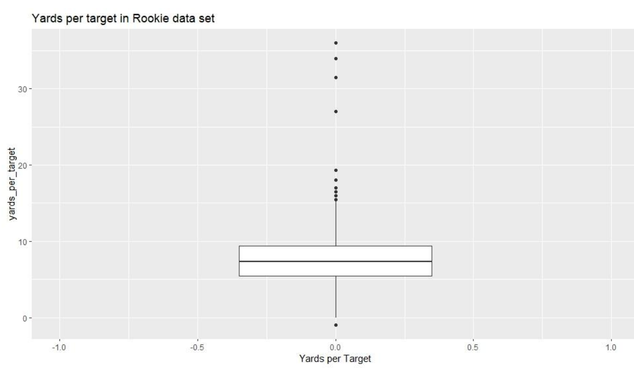
To quantify the success of a wide receiver, we will focus our analysis on his yards per target. Yards per target has a lot of desirable qualities. For example, yards per target is not dependent on time played in a game. There is a wide variance in time played between players, so using a statistic that does not depend on time played increases the generalizability of the results. Another benefit of using yards per target is that it is not as dependent on the team compared to other stats, such as EPA (Expected Points Added). A player that is thrown the ball every play will have no inherent advantage over a player thrown the ball once a game in this stat. A potential downside of yards per target is that players with few targets in a season will have a very large variance. For example, one player has 30 yards per target, but that is with only one target. We will deal with this at a later point. Wide receivers with higher yards per target capitalize more when they are thrown the ball, and such, can typically be considered better overall players.

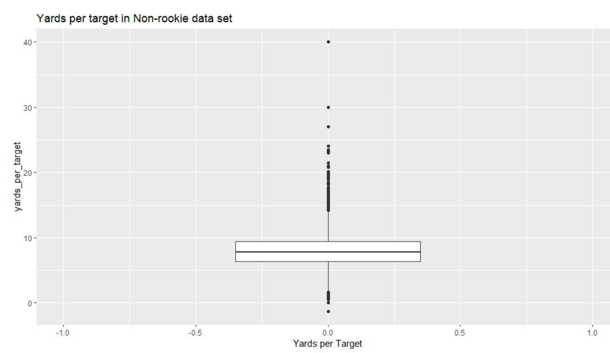
To model the receivers' success after their rookie year, we will construct a distribution for a wide receiver's yards per target that is influenced by prior beliefs, draft position, and season statistics. We will also build a predictive model aiming to estimate a player's yards per target in future seasons based on his rookie season. Within our modeling, we will implement a hierarchical structure to our model in order to compare receivers based on the round they were selected in. By sampling from our new distribution and comparing coefficients, we will more accurately be able to predict a player's future success after his rookie season and compare performances across rounds to determine what rounds are best to select receivers in.

Exploratory Data Analysis

In an effort to have data reflective of the NFL today, we decided to span our analysis over the past 20 seasons. Our dataset includes every season for all wide receivers that were drafted from 2003 to 2022. The statistics for our receivers comes from *Pro Football Reference,* which are official league statistics. These were loaded using the *nflfastR* package. To get draft information, we used an NFL draft dataset provided by Lee Sharpe, and then merged the two datasets together. Once we had all seasons of data combined with each player's draft information, we separated our data into two sections: Rookie and Non-Rookie. In each, a singular row contained a season for a particular player, along with his yards per target, round selected, pick number, yards, touchdowns, and other stats. For simplicity, we omitted any players who did not record any receiving statistics in a given season; It would be unfair to give a 0 to players if they were injured or did not get the ball thrown to them.

With the data collected and organized sufficiently, we needed to understand our yards per target variable in both groups before beginning to understand the relationship between a player's rookie season and non-rookie statistics. Shown below are the boxplots for yards per target in rookie seasons and in non-rookie seasons.

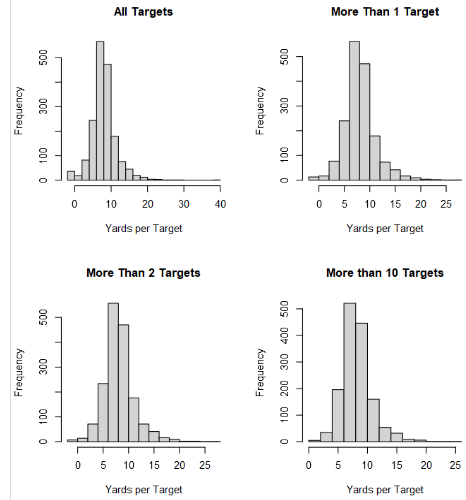




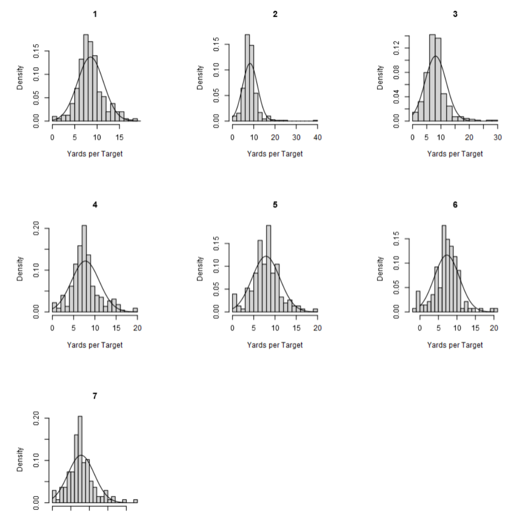
Starting with the rookie data, the boxplot shows that the median value is at 7.323. There are many outliers in the plot, with a maximum value of 36. This occurs due to the number of targets that the players achieved was low. Since the low number of the denominator, the target, caused the extreme value of yards per target, it appears that we will have to create a target threshold in order to best represent the entirety of our data.

Similarly, the boxplot of the yards per target in the non-rookie data shows that the median value is at 7.814, which is slightly higher than the rookie data. We do see, however, a lower IQR in the non-rookie data, which shows that the variance may be higher across the rookie dataset. There are also many outliers in this data, with a maximum value of 40, due to the lower number of targets that these players achieved.

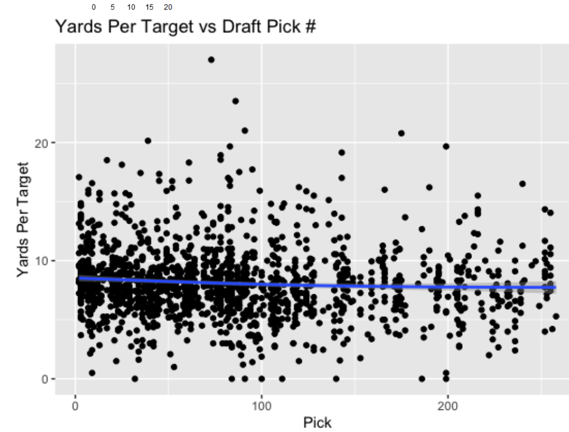
To negate the large variance due to some players getting passed the ball a small number of times, we decided to only count players above a certain threshold of targets. Without the threshold, we would inaccurately be measuring the receivers' success. As can be seen in the figure to the right, The distribution of yards per target over a career stabilizes after removing the players with only 1 target. Past that, the distribution of yards per target is practically identical. The data that we work with from here on has all players with only 1 target removed.

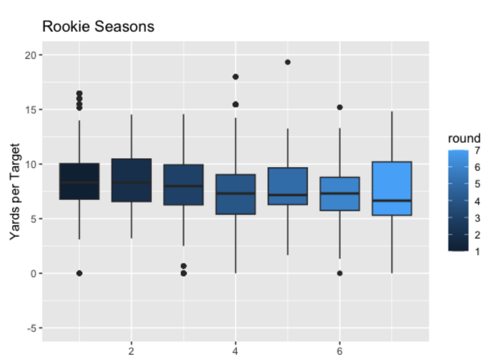


Now that we have adjusted the dataset to exclude any of these outlier events, we can further discuss the distribution of yards per target. After some investigation, we found the yards per target for each round to be roughly normal, which can be seen in the figure to the right. We split these distributions based on the round the receiver was selected in an attempt to compare different groups. The idea that yards per target is normally distributed across each round will aid us in the later parts of our analysis.



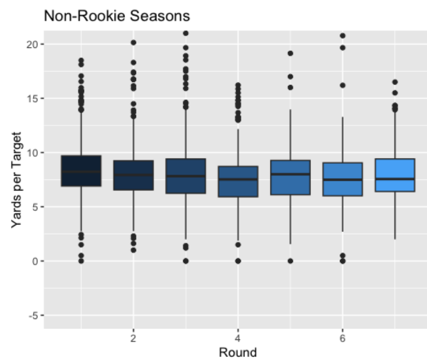
With an understanding of how yards per target is distributed, we then can turn to analyze how our rookie data and draft position affect the yards per target of a receiver in a given season. There were a total of 1,707 individual non-rookie seasons spanning across 418 unique wide receivers. The figure to the right shows the relationship between the yards per target and draft pick number for each of our receiver's seasons. Most noticeable is the heterogeneity of our variance across pick numbers, which will justify our need for a hierarchical model in the later parts. Additionally, the correlation between yards per target and pick number is significantly different from 0, and allows us to conclude that those drafted with higher pick numbers (rounds 5,6,7) will have lower yards per target. To further expand on the different effects that a draft round has on yards per target, the figures below show the distributions of yards per target, for non-rookie and rookie seasons, based on draft round.






```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   6.812   8.212   8.432   9.855  16.480
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
3.200   6.281   7.740   8.080   9.634  14.538
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   5.625   7.433   7.542   9.567  16.000
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   5.125   6.974   6.979   8.709  18.000
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   4.576   7.145   7.328   9.657  19.333
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   4.708   6.903   7.033   8.260  15.333
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   4.333   6.059   6.871   8.667  31.500
```

It is clear from the two plots that variances differ across the seven unique rounds, which was validated with an ANOVA test. More importantly, the yards per target in non-rookie seasons appears to be significantly larger than that of rookie seasons, as one can see that across all rounds, the quantiles are larger in the non-rookie seasons.




```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   6.911   8.232   8.547   9.698  18.500
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   6.558   7.936   8.123   9.238  20.143
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   6.246   7.821   8.182   9.408  27.000
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   5.923   7.524   7.695   8.719  16.214
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   6.115   8.000   7.859   9.273  19.147
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000   6.011   7.492   7.621   9.052  20.778
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
2.000   6.407   7.556   8.015   9.412  16.500
```

While the correlation value of .108 appears low, we can conclude that it is significantly different from zero, which means that there is a positive association between the rookie yards per target and a season's yard per target for a given player. This finding shows that those receivers who perform better in their rookie season will, on average, perform better in future years than those rookies who do not.

Our initial work with the data allowed us to adjust for random outcomes, as well as establish a relationship between the yards per target of a rookie season and the yards per target after a rookie season. There also is a clear trend based on the round a player is selected that must be taken into account. With this, we can now turn towards creating a more representative distribution in order to project the future yards per target of a wide receiver in the NFL.

Methodology and Results

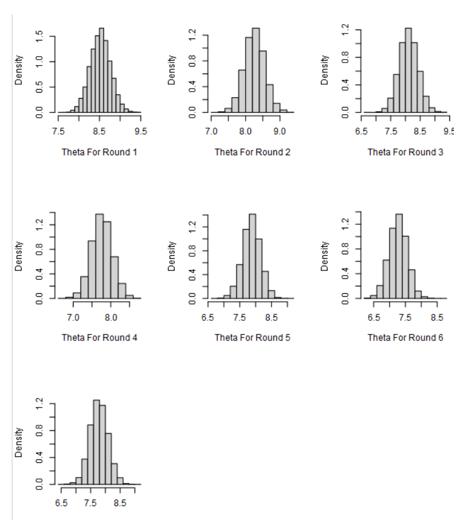
We decided to use a normal model for each round to model yards per target because the data is roughly normal, as was shown above. Our priors for round i and sampling model are as follows:

$$1/\sigma_i^2 \sim Gamma(\nu_0, \nu_0\sigma_{0i}^2/2)$$

$$\theta_i|\sigma_i^2 \sim Normal(\mu_{0i}, \sigma_i^2/\kappa_0)$$

$$Y_{1i}, \ldots, Y_{ni} \sim Normal(\theta_i, \sigma_i^2)$$

The priors were chosen to be conjugate and the prior parameters were decided based on the yards per target in the rookie data. With this model, we calculated the joint posteriors for each round and then with that found the marginal posterior of θ given the data using Monte Carlo, which is seen in the figure below and to the right. Figure 1, which is attached at the end of the report, plots the joint posterior.



By looking closer at our distribution for θ across rounds, we can discover differences in yards per target. The $P(\theta_R > \theta_C)$ table (on the following page) shows the probability that the mean for the row is higher than the mean for the column. For example, the first value of .7394 means that there is a probability of .7394 that the true career average yards per target for players picked in round one is higher than that of players picked in round two.

Our posterior predictive distribution is vital to understanding the projection of future players. The table $P(\tilde{Y}_R > \tilde{Y}_C)$ below shows the probability that a randomly selected player from the row has a higher YPT than a random player from the column. For example, the first value of .5187 means that there is a probability of .5187 that the career average yards per target for a randomly picked player in round one is higher than that of a randomly picked player in round two. This table is very important for teams deciding what round to draft a wide receiver.

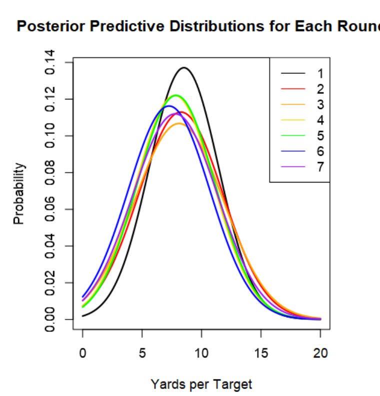
$P(\theta_R > \theta_C)$	2	3	4	5	6	7
1	.7394	.8399	.9793	.96	.999	.9693
2		.6242	.8879	.8264	.9892	.8666
3			.7943	.7202	.9746	.7737
4				.3906	.8811	.4717
5					.9298	.5815
6						.1129

$P(\tilde{Y}_R > \tilde{Y}_C)$	2	3	4	5	6	7
1	.5187	.5288	.5675	.5561	.6093	.5551
2		.5126	.5393	.5343	.5859	.5414
3			.5283	.5204	.5695	.5203
4				.4885	.5369	.4922
5					.5478	.5037
6						.4569

We can also look at the confidence intervals for the new mean yards per target for each of the seven rounds in order to determine how to evaluate players and predict results. The table below shown to the right represents the 95% posterior confidence intervals for our sampled theta for each round. It is clear that higher yards per target are contained within rounds 1-3, as well as that the later rounds have wider intervals, which is likely due to the smaller sample size in later rounds. We would estimate that a receiver's mean yards per target in the first round would be approximately 8.54 in seasons after his rookie season, which can give us expectations after a rookie season. These intervals allow us to identify a player's performance and see if they are performing as they should given their draft round and rookie success.

Round	95% CI
1	(8.05, 9.03)
2	(7.72, 8.71)
3	(7.55, 8.74)
4	(7.26, 8.28)
5	(7.49, 8.56)
6	(6.86, 7.98)
7	(7.31, 8.38)

The following figure plots the posterior predictive distributions for each round on the same graph. It is congested, but the important thing to note is that round 1 is the most different of all the rounds, with a slightly higher mean and a slightly smaller variance. Many of the other rounds are similar, but there is a general trend of a decreasing mean as the round increases.



To supplement our posterior predictive distribution for each round, we also built a Bayesian Regression model in an attempt to predict a receiver's yards per target of a given season based on the yards per target of his rookie season. In order to adjust for non-constant variance, we used hierarchical modeling to ensure that each draft round was grouped together. By comparing the coefficients for each of the seven rounds, we will be able to compare the relationship between rookie and future success across rounds and determine what receivers are easier to predict after year one.

Since yards per target for both rookie and non-rookie seasons are normally distributed, we can impose weak prior distributions for β (normal) and σ^2 (inverse gamma) in order to approximate the joint posterior distribution and ensure that it is more reflective of our data. The following below illustrate the two distributions that we will use to approximate the joint posterior distribution of β and σ^2.

$$\beta^{s+1} : MVN(m, v) \;\; where$$
$$v = (\Sigma_0^{-1} + X^T X/\sigma^2)^{-1} \;\; and \;\; m = V * (\Sigma_0^{-1} * \beta_0 + X^T y/\sigma^2)$$
$$\sigma^{2(s+1)} : Inv - Gamma([\nu_0 + n]/2, [v_0 * \sigma^2 + SSR(\beta^{s+1})]/2)$$

To understand the idea of projecting players, it is important to compare the coefficient estimates relative to each round, as well as the variances in each group. For starters, our lowest β estimates come in rounds 2 and 3, meaning that these receivers have less improvement in yards per target from their rookie year to future seasons. Therefore, one's performance in their rookie season can be more reflective of future performance than those drafted in other rounds. To build on this idea, our σ^2 estimates are largest in rounds 3,6, and 7, meaning that predicting their future success from player to player is difficult after one season. While the risk is lower in the later rounds, there are chances for high rewards.

Our models can give teams insight as to how to draft wide receivers. For example, using our posterior distribution, we can conclude that if a team is looking at picking a wide receiver during the draft, they should pick one either in the first round or in the later rounds. The difference between the first round and the later rounds is relatively large compared to the differences between the other rounds. Therefore, picking a wide receiver in the second round gives roughly the same probability of success as picking a wide receiver in the later rounds. Taking into account the variance as well, one can conclude that picking a wide receiver in the later rounds is more difficult. If there is a player that the front office is confident in, it would be in their best interest to take them in the late rounds; Otherwise, it is best suggested to stay away in order to avoid some of the low outcome events.

Conclusion

The analysis outlined above serves as an approach to combine prior information, such as draft information and rookie year statistics, with observed data in order to update our beliefs on the quality of player a team possesses. This modeling allows us in the future to be able to predict the success of a wide receiver, in terms of yards per target, after just one season, as well as compare how different receivers of different rounds develop.

While our modeling is effective, it is important to understand that it is a simplistic approach to model a complex issue. There are a multitude of factors that scouts and general managers take into account to measure a wide receiver's impact. Yards per target gives us a measure that tells us how impactful a receiver is every time he is targeted, so any value a receiver adds on plays where he is not targeted is unaccounted for. For example, a wide receiver can provide additional value to a team, such as by rushing, blocking, etc. that is not encapsulated through yards per target. So, one must not use our modeling as a “end all be all” to evaluate wide receivers completely, but rather combine it with other statistics that tell more of a story.

Our analysis did not consider any players that were not drafted. It is a possibility for players to go undrafted and then have successful careers. A future model could improve on our work by considering such players.

Overall, our new and updated distribution for the yards per target and regression model aimed to predict yards per target enabled us to see the impacts of both draft position and rookie season success on the future performance of wide receivers. Understanding these distributions in relation to how a receiver performs in his rookie season can give a team a competitive advantage in terms of player evaluation, as well as allow them to alter future draft strategy.

References

Carl S, Baldwin B (2022). _nflfastR: Functions to Efficiently Access NFL Play by Play Data_. R package version 4.4.0, <https://CRAN.R-project.org/package=nflfastR>

Ho T, Carl S (2022). _nflreadr: Download 'nflverse' Data_. R package version 1.3.0, <https://CRAN.R-project.org/package=nflreadr>..

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods (Springer Texts in Statistics)*. Springer.

Sharpe, L. (2020). Nfldata. Github. https://github.com/nflverse/nfldata/blob/master/DATASETS.md

Contributions

Each of us worked as a group and all met frequently to discuss our ideas and formulate a project plan. We all contributed to planning the project as well as the final paper and presentation, with us focusing on the specific areas:

- Anthony Paolillo: Data Cleaning, Rookie and Non-Rookie EDA, Bayesian Regression, Interpretation
- Seth Corbridge: Non-Rookie EDA, Joint Posterior Distribution, Posterior Predictive
- Krittin Tangboriboonrat: EDA, Bayesian Regression, Interpretation

Overall, we felt that we all worked together well to produce a final project that is applicable and representative of multiple Bayesian techniques that we learned throughout the semester.

Codes

The codes are provided through the GitHub repository in this link.
https://github.com/boatwonder/STATS551-NFL
The readme file explains how to run the codes required for the analysis. The codes are stored in the Project Codes folder.

Figure 1: Posterior Distribution for θ