

Optimal NFL Play Call in Short Yardage Situations

Ronan Connell and Dr. N.David Pifer

Introduction:

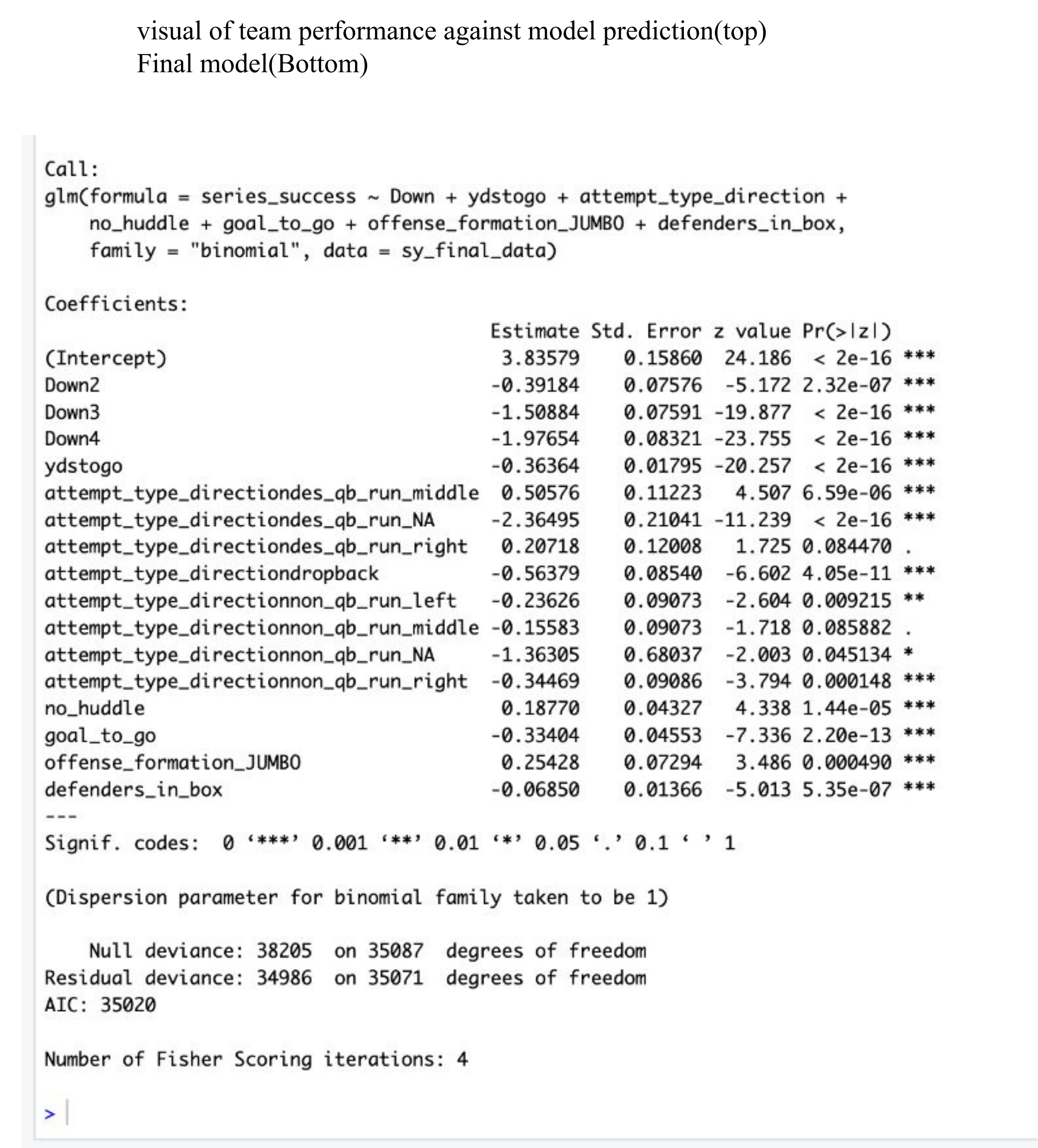
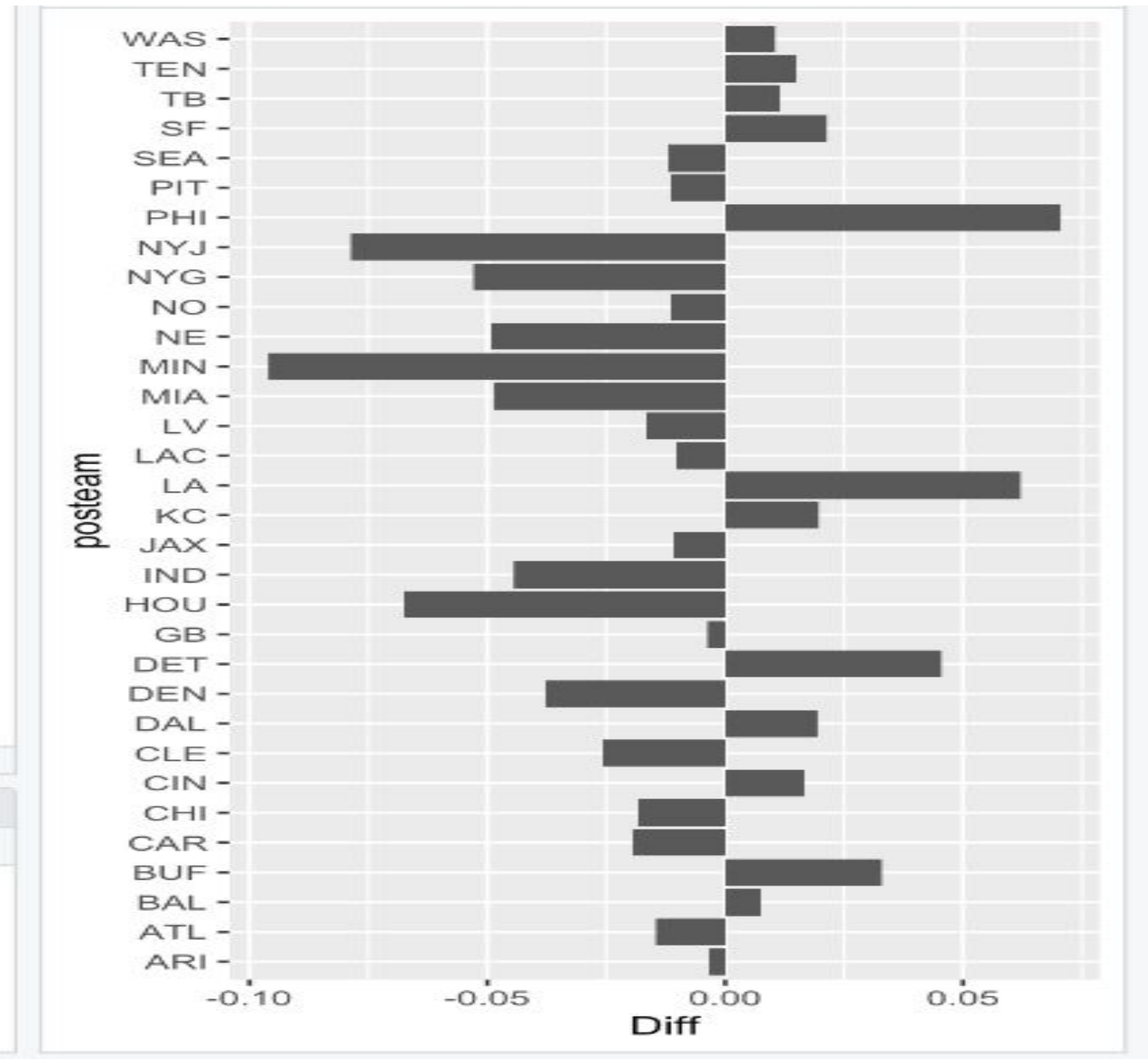
The field of football analytics has exploded in popularity over the last decade. Learning from the well-established analysis techniques in sports such as baseball, football analysts have revolutionized the way players are taught, scouted, and played against. Databases such as *nflastR* have allowed for long term tracking of key metrics and allowed for more sophisticated analysis to take place. This database is constantly being improved, with new data such as defensive schemes and sets being added within the last two years. My strong passion for football combined with my interest and analysis led me to be introduced to the field with this first research project. A trend I appeared to have noticed when watching NFL games was that teams seemed to pass more on these short yardage situations. This led to a curiosity on learning not only what the most frequently called plays were, but also the most effective. This project addresses this curiosity by focusing on the optimal play call for short yardage situations, with hopes to make improvements to play calls based on this research. While this project is still ongoing, there are already meaningful results gleaned from the limited model I have helped create.

Methods:

This research was conducted using nflreadr, a comprehensive data tool that provides detailed play-by-play data for all NFL games since 1999. The primary objective was to analyze the effectiveness of play calls in short-yardage situations.

- Data Collection & Filtering
 - The dataset was filtered to include only plays where the yards-to-go for a first down was less than four.
 - Special teams plays (field goals, punts, etc.) were excluded, as the focus was solely on conversion attempts.
 - Data Includes all plays that meet this criteria from 2016-2024
- Feature Selection & Data Processing
 - The dataset was categorized by play type (run vs. pass).
 - Further breakdowns included run direction, offensive formation, defensive alignment (men-in-box), quarterback vs. running back rushes, and scrambles.
 - Some missing data, such as player position for specific plays, was manually imputed in Excel and re-uploaded into nflreadr for further analysis.
- Statistical Analysis
 - Logistic regression models were used to assess the probability of a successful conversion based on various play characteristics.
 - All statistical analyses were conducted in R, where I further converted results to make them easier to understand

This methodology enabled the creation of a data-driven model that provides insight into the optimal play-calling strategies in short-yardage scenarios. I then assessed the reliability of the model with a Brier Score, which measures the reliability of probabilistic predictions from 0-1, with a lower score meaning a stronger model fit. My final model had a brier score of 0.1631129, which is a good result and means my model was a good fit I then compared NFL teams actual results to the results predicted by my model to see if teams on average over or underperformed my model, the graph of which can be seen to the right.



Preliminary Results:

While my research is ongoing we have successfully determined there are certain factors that have statistical significance when evaluating the effectiveness of playcall in short-yardage situations. Run plays are 8.28% more likely to yield a successful result than pass plays. QB runs to the middle are 5.99% more likely to yield a successful result than QB runs to the left. No-huddle is the most effective tempo, being 3.06% more likely to yield a successful conversion than standard tempo. Jumbo packages are the most effective packages, being 4.14% more likely to yield a successful conversion compared to an empty set. Compared to first down, second down is 3.91% less likely to yield a successful result, third down is 21.92% less likely to yield a successful result, and fourth down is 32.22% less likely to yield a successful result. For each yard farther from the first-down line, probability of success decreases by 5.92%. For each defender in the box, probability of success decreases by 1.12%. Goal-to-go situations are 5.44% less likely to yield a successful result compared to non-goal-to-go situations. To the left is a chart of how well NFL teams performed against the model, predicting success using the above factors. Notably, the Eagles, Rams, and Lions substantially outperformed the model while the Texans, Jets, and Vikings substantially underperformed the model.

Conclusion:

Throughout this year, I have extensively examined short-yardage situations in the NFL. Through this research I helped create a model which is relatively effective at predicting successful outcomes in short-yardage situations. This model predicts that the most effective playcall in short yardage situations would be a no-huddle, jumbo package, QB run up the middle, on first down, to convert a first down. What is important to note is that football is played with humans not robots and as a result has variation in team success. Across the eight years of data used to create this model, the average results have been used to make predictions. Certain circumstances such as momentum and morale were not able to be factored into this model. Football play call is also something that goes two ways. If a team can correctly guess what play is being run, they have a substantially higher likelihood of stopping it. As a result, defaulting to the optimal play call described by this model may allow the opposing defense to counter it, making the play call less effective. This model is a basic one that allows for the noticing of broad trends, and still needs to be tested at predicting specific future results. This is my first introduction to the realm of football analytics and I plan on continuing to improve this model during future years at FSU.

References

- Microsoft Corporation. (n.d.). *Microsoft Excel* [Software]. Microsoft. Available at: <https://www.microsoft.com/en-us/microsoft-365/excel>
- R. J. S. (2021). *nflreadr: NFL data for R*. GitHub repository. Available at: <https://github.com/RJStats/nflreadr>
- L. A. P. (2020). *nflfastR: Fast Analysis of NFL Play-by-Play Data*. GitHub repository. Available at: <https://github.com/nflverse/nflfastR>