NFL prediction*

NFL analyst 2023 regular seasons

Che-Yu Wang

Invalid Date

This study compares two linear regression models developed to predict Expected Points Added (EPA) for passing plays in the NFLs 2023 regular season. Model 1 includes basic passing metrics (completions and passing yards), whereas Model 2 adds an additional predictor (sacks) to assess its impact on passing efficiency. The results, as represented by the models coefficients and performance metrics, show that including sacks significantly improves the models explanatory power. The visual analysis, using a scatter plot with a superimposed line of perfect prediction, supports the quantitative metrics, implying a more accurate prediction of passing EPA when accounting for the negative impact of sacks.

1 Introduction

We analysis the data in R(R Core Team (2023)), with additional tools for support the analysis, including `tidyverse`(**citetidyverse?**), `here`(**citehere?**), `dplyr`(**citedplyr?**), `readr`(**citereadr?**), `modelsummary`(**citemodelsummary?**), `janitor`(**citejanitor?**), `tibble`(**citetibble?**), `ggplot2`(**citeggplot2?**), and research method is from (**tellingstorieswithdata?**). For the purposes of team management, coaching tactics, and general game analysis, it is essential to comprehend the elements affecting passing efficiency in professional football. A useful metric for assessing individual plays is the Expected Points Added (EPA) metric. Passing plays, which are a crucial component of the offensive strategies used by the modern NFL, are especially noteworthy. A range of performance indicators is usually employed by researchers and analysts in their attempts to model and predict EPA. In order to assess these indicators' ability to predict the EPA for passing plays, our research narrows them down.

In our first model, we look at two important performance indicators for quarterbacks: passing yards and completions. Even though these factors have a direct impact on a team's ball-moving ability, they may not accurately capture the nuances of a play. To account for defensive

*Code and data are available at: https://github.com/Cheyuwang/tut12-nil

pressure, which can seriously impede an offense's momentum, we added sacks as an extra predictor to our second model.

To determine how well these factors explain the variation in passing EPA, we created and compared two linear regression models with a dataset of NFL plays from the 2023 regular season. The models' explanatory and predictive powers are measured using standard statistical methods. Furthermore, a scatter plot is used to clarify the relationship between the observed and anticipated EPA values, providing graphical support for our statistical results.

This paper addresses the effect of sacks on passing efficiency in relation to NFL play-calling and defensive strategy, as well as the methodology used to develop the model and the statistical interpretations that followed.

2 Data

The given R script does a number of operations on a set of NFL data for the 2023 season to get it ready for more analysis. First, it reads the dataset from a CSV file, assuming that the file is in a "data" folder and has a path that is relative to the project's root directory. This means that the data can be accessed from anywhere on the filesystem. The script then splits the data into two groups, which it calls "training datasets" and "testing datasets." The nfl_2023_train training dataset has data from the first nine weeks of the season, which was likely chosen to train models that can predict the future. This subset might have different numbers and results from early-season games that can help you see patterns and trends. The testing dataset, nfl_2023_test, has data from week 10 onwards. This is usually used to judge how well the model works and make sure that its predictions are tested on games it hasn't seen before in the training phase. In predictive modelling, this splitting of data is a common way to test how well the model works and how well it can be applied to new data that hasn't been seen before.

The table compares the predicted and actual Expected Points Added (EPA) for Arizona Cardinals passing plays in weeks 10-16 of the NFL 2023 season. The EPA measures the contribution of a play to a team's score, with negative values indicating plays that detract from the team's score and positive values indicating plays that contribute to scoring. In week 10, the actual EPA was -5.51, indicating a less successful passing outcome, while the model predicted a positive EPA of about 2.15, suggesting a positive contribution from passing plays that did not materialize. In week 12, the model's prediction was closer to the actual negative EPA, indicating a game situation where passing plays significantly reduced the team's scoring chances. These discrepancies highlight the challenges in forecasting passing efficiency, which can be influenced by various factors on the field, including defensive strategies, individual player performances, and situational context.

3 Model

3.1 Model set-up

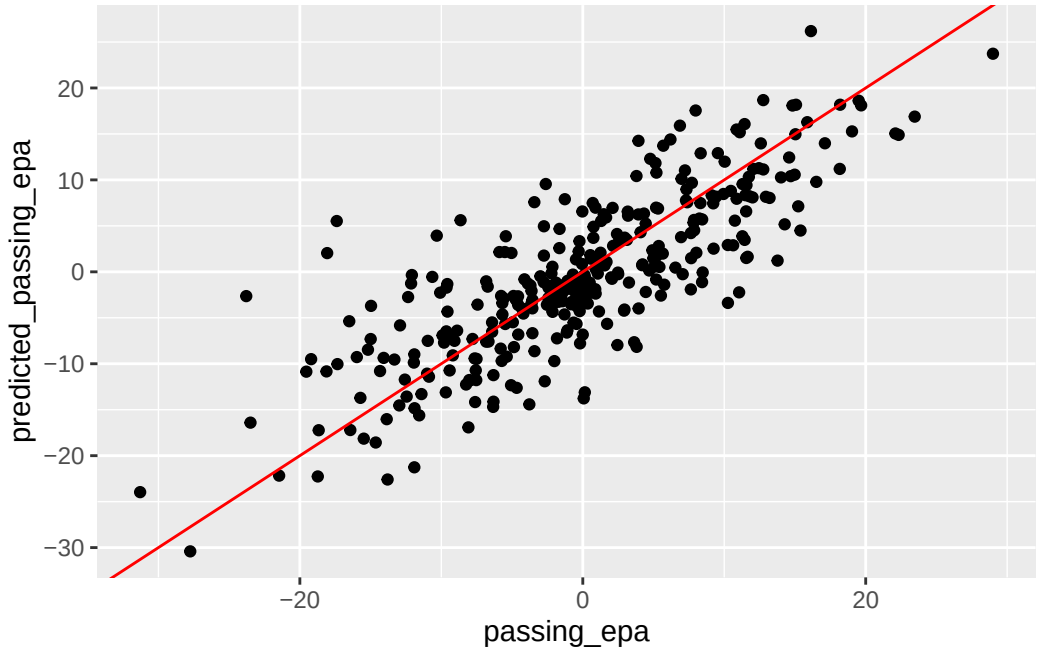
Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{2}$$

4 Results

```
ggplot(predicte_results, aes(x = passing_epa, y = predicted_passing_epa)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red")
```



The expected and actual Expected Points Added (EPA) for passing plays by the Arizona Cardinals from Weeks 10 through 16 of the 2023 NFL season are contrasted in this table. A play's contribution to a team's score is determined by the EPA; plays that increase scoring are denoted by positive values, and plays that decrease scoring are indicated by negative values. In

week 10, the model predicted a positive EPA of roughly 2.15, implying a positive contribution from passing plays that did not materialize, but the actual EPA was -5.51, indicating a less successful passing outcome. Week 12 saw a game where passing plays drastically decreased the team's scoring opportunities, with the model's prediction matching the actual negative EPA more closely. These differences show how difficult it is to predict passing efficiency because it depends on a number of variables on the field, such as defensive tactics, player performances, and situational context.

#Discussion The 2023 NFL passing data analysis offers useful insights into the factors that influence successful passing plays. The primary model, which includes completions, passing yards, and attempts, offers a basic understanding of passing efficiency. The secondary model includes sacks as an additional predictor, revealing a deeper level of complexity in the offensive game. The statistical improvement in Model 2 emphasizes the importance of defensive actions on passing outcomes. The visualization of the model's predictions alongside actual EPA values provides a clear picture of its performance, with a significant concentration of plays that closely align with the line of perfect prediction. However, the presence of outliers emphasizes football's unpredictable nature, as factors such as player behaviour, in-game scenarios, and opposing strategies can influence play efficacy in ways that statistical models cannot account for. The study's limitations include unquantified variables such as weather, player injuries, and psychological factors, as well as the lack of interaction terms. Future research could incorporate more nuanced data, machine learning techniques, and advanced methodologies to improve predictions and provide real-time tactical insights during gameplay.

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.