NFL Prediction: Analysis of Passing EPA based on first nine weeks*

Hyungsoo Park

April 3, 2024

This paper focuses on the later Passing EPA in 2023 NFL based on the first weeks league data.

1 Introduction

This paper explores the total expected points added on passing for quarterback position players. The quarterback position is a role that passes to a receiver from an offensive lineman and passes the ball to a running back, which means the quarterback position plays an essential role in winning a football game. In addition, a team without good quarterback players has less chance to win the football games in a league.

Therefore, the total expected points added on passing for quarterback players is one of the most important factors in evaluating each quarterback player. This study focuses on the correlation between "passing_epa" and other related factors in the football game.

The data from the first nine weeks of the 2023 NFL season were used to predict the later 2023 season's total expected points added on passing.

In the following section, Section 2 explains the data used and the cleaning process, and Section 3 analyzes the built model for this study.

2 Data

The raw data used was retrieved from NFLverse(Sebastian Carl 2023). The data was cleaned by R(R Core Team 2023) and Tidyverse(**citeTidyverse?**), and analyzed by r. The raw data from Sebastian Carl (2023) contains all data from 1999. The cleaned data contains only player data

*https://github.com/kakaomonk/nfl-prediction

Table 1: Cleaned NFL data

```
# A tibble: 646 x 11
   player_id  season  week completions attempts passing_yards passing_tds
   <chr>       <dbl> <dbl>       <dbl>    <dbl>         <dbl>       <dbl>
 1 00-0023459   2023     1           0        1             0           0
 2 00-0026158   2023    13          23       44           254           2
 3 00-0026158   2023    14          26       45           311           3
 4 00-0026158   2023    15          28       44           374           2
 5 00-0026158   2023    16          27       42           368           3
 6 00-0026158   2023    17          19       29           309           3
 7 00-0026498   2023     1          24       38           334           0
 8 00-0026498   2023     2          34       55           307           1
 9 00-0026498   2023     3          18       33           269           1
10 00-0026498   2023     4          27       40           319           1
# i 636 more rows
# i 4 more variables: interceptions <dbl>, sacks <dbl>, sack_yards <dbl>,
#   passing_epa <dbl>
```

in the first 9 weeks of 2023. The variables in interest are 'passing_epa', 'player_id', 'season', 'week', 'completions', 'attempts', 'passing_yards', 'passing_tds', 'interceptions', 'sacks', and 'sack_yards'.

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.

$$y = B_0 + B_1 X_{completions} + B_2 X_{attempts} + B_3 X_{passing_yards} \tag{1}$$

$$+ B_4 X_{passing_tds} + B_5 X_{interceptions} + B_6 X_{sacks} + B_7 X_{sack_yard} \tag{2}$$

The model was ran in R(R Core Team 2023) by using rstanarm(Goodrich et al. 2022).

Table 2: Explanatory models of flight time based on wing width and wing length

```
parsnip model object


Call:
stats::lm(formula = passing_epa ~ completions + attempts + passing_yards +
    passing_tds + interceptions + sacks + sack_yards, data = data)

Coefficients:
  (Intercept)    completions       attempts  passing_yards    passing_tds
     -0.93981        0.41778       -0.75037        0.09604        1.68490
interceptions          sacks     sack_yards
     -3.81529       -1.33799       -0.09272
```

3.1.1 Model justification

The model considers 6 predictor variables (completions, attempts, passing_yards, passing_tds, interceptions, sacks, sack_yard) to predict passing_epa.

4 Results

The results from the model is in Table 2.

As a result, the paper could predict that the passing distance (passing_yards (+0.09604)), the number of touchdowns (passing_tds (1.68490)), number of sacks (sacks (-1.33799)), and yards of lost sack (sack_yards (-0.09272)) influence to the total expected points added on passing. In addition, the number of touchdowns (passing_tds (1.68490)) and number of sacks (sacks (-1.33799)) influence to the total expected points added on passing significantly.

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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

Sebastian Carl, Lee Sharpe, Ben Baldwin. 2023. *Nflverse: Easily Install and Load the 'Nflverse'*. https://CRAN.R-project.org/package=nflverse.