Offense Pass Completion as a Predictor of NFL Score Differential*

Using offense pass completion to predict how successful an NFL team is at creating score advantages

Joshua Kwon

October 8, 2025

This paper examines whether the quality of an NFL team's quarterback translates into more wins for a team. We find strong statistical evidence of a linear association between offensive pass completion percentage and score differential. However, only 24.68% of the variation in score differential is explained by pass completion. While pass completion is significantly associated with score differential, it alone does not fully capture differences in score outcomes.

Introduction

In recent years, there has been an increase in the quarterback's pay as percent of salary cap. In a 5 year span, the number of quarterbacks taking "at least 10% of the cap"(Howe 2025) increased from 7 to 23. This leaves less room on a team's salary cap for the rest of the team. Crucial positions on the defense, and perhaps even on the offense may be left neglected. Is quarterback really that important? This paper tries to quantify whether a better quarterback can deliver a greater amount of wins using score-differential and pass completion rate.

Since an NFL team wins by scoring more than their opponent at the end of a game, this paper uses score-differential as a proxy for how well a team did during a game. Thus a better high score-differential should represent a more winning team. Offense pass completion percent is also used as a measure of a quarterback's quality.

Our analysis find evidence of a significant association between pass completion and score differential. However, only 24.68% of the variation in score differential can be explained by offense pass completion percentage.

*Project repository available at: https://github.com/peteragao/MATH261A-project-template.

In this paper, we begin by introducing the reader to our data. We visualize important aspects of this data. We continue to discuss the methodology used in the analysis. We then present our findings and test the assumptions made in our regression model. Finally, we conclude with a discussion of our findings and the implications on whether the quality of an NFL team's quarterback translates into more wins for a team.

Data

The `nfl-team-statistics.csv` dataset provided by SCORE Sports Data Repository (Yurko 2023) contains statistics about the regular season performance for each NFL team from 1999 to 2022. The data was collected using the `nflreadr` package (Ho and Carl 2025)in R.

The purpose of this paper is to observe the relationship between offense completion percentage and score-differential and as such these are the two relevant variables taken from the raw dataset provided by (Yurko 2023). A total of 765 observations were recorded and analyzed in this paper.

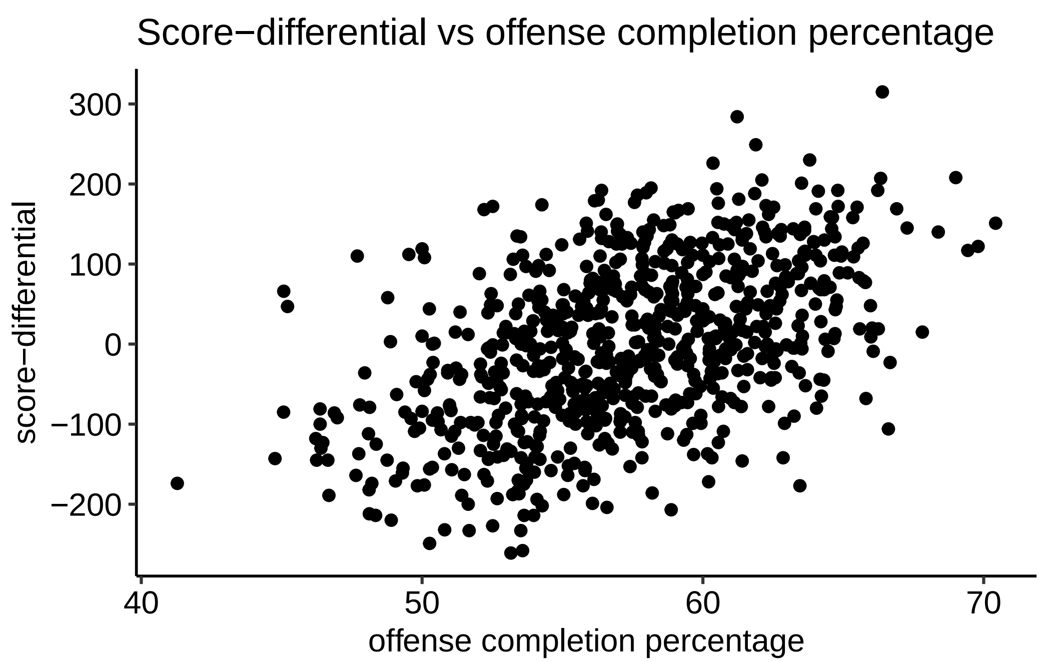

Figure 1: Scatter plot with offense completion percentage as the predictor values and score-differential as the outcome values

From Figure 1 we can see a correlation between the increase in offense completion percentage and score-differential. This paper will quantify this correlation.

```
offense_completion_percentage score_differential
Min.   :41.28                 Min.   :-261
1st Qu.:54.09                 1st Qu.: -74
Median :57.30                 Median :   1
Mean   :57.32                 Mean   :   0
3rd Qu.:60.61                 3rd Qu.:  72
Max.   :70.43                 Max.   : 315
```

The table and Figure 1 above shows that on average, an NFL quarterback will complete 57% of their pass attempts. We also note the concentration around this percentage as only about 3% of pass completion separates the 1st and 3rd quartile from the mean. Additionally note the even distribution of score differential around the mean with the 1st and 3rd quartile 74 and 72 points from the mean respectively. Outliers are present in the data as seen in Figure 1.

Methods

This paper will fit the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

to understand the relationship between the predictor variable, offense completion percentage, and the outcome variable, score-differential. In this model, X_i represents offense completion percentage of the ith observation and Y_i represents the score-differential of the ith outcome.β_0 represents the intercept coefficient, what we expect score-differential to be when the offense has a completion percentage of 0. β_1 represents the slope coefficient, what we expect the increase in score-differential will be for every percent increase in completion percentage. In this model, we assume the error term, ε_i to be random with mean 0 and finite variance σ^2.

The parameters, β_0 and β_1, can be estimated using the method of least squares which minimizes the sum of the squared deviations $Q = \sum_{i=1}^{n} = (Y_i - \beta_0 - \beta_1 X_i)^2$. This requires the assumption that the errors are uncorrelated and have equal variances with mean 0. From the OLS parameter estimates, the regression function can be estimated with $\hat{Y_i} = b_0 + b_1 X_i$. This gives the fitted values for case i. The fitted model created fits a linear line that best predicts score differential, from offense completion percentage.

With our fitted model, we can perform a hypothesis test to determine whether or not there is a statistically significant association between our predictor, pass completion, and outcome, score differential. We also test how well our regression model explains the variability in score differential.

Finally, we evaluate a residual vs fitted plot in Figure 3 and a Q-Q plot in Figure 4 to determine whether the assumptions of our regression model hold.

I implemented this analysis using the R programming language (R Core Team 2025).

Results

Using the $lm()$ function in fits our model using the ordinary least squares method. The results are shown below.

```
       B0          B1
-618.0257   10.7814
```

β_0 represents the intercept coefficient and has a value of -618.03. Thus, score-differential can be predicted to be -618.03 when the offense has a pass completion percentage of 0%. β_1 represents the slope coefficient and has a value of 10.7814. When pass completion percentage increases by a percent, we should expect an average increase in score differential of 10.7814 .

We now return to the scatter plot from the beginning of the paper to better illustrate the linear association.

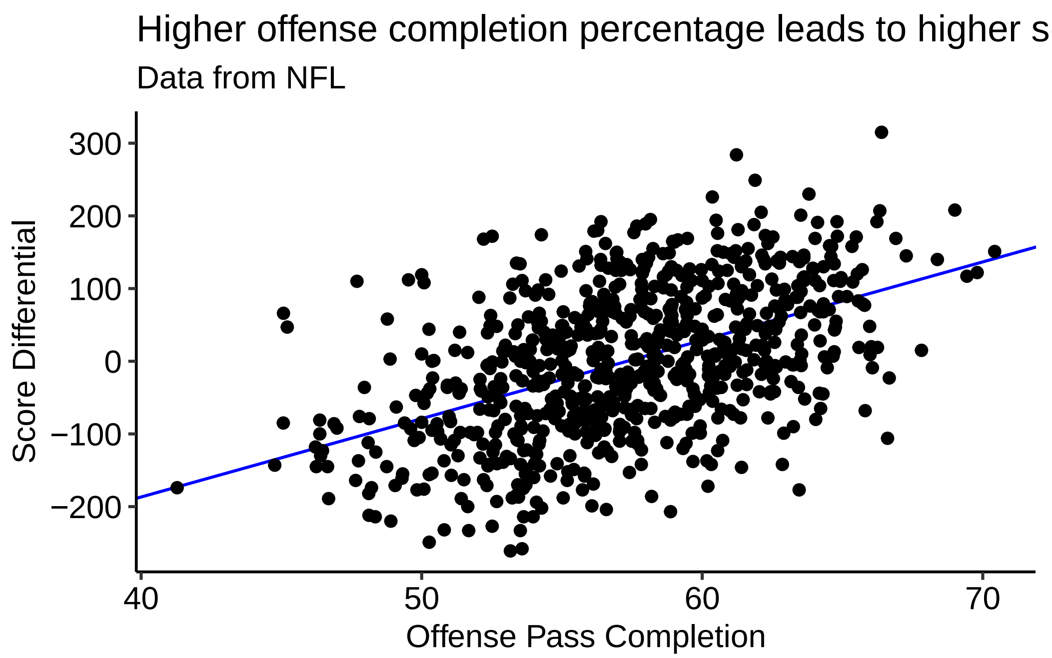

Figure 2: Scatter plot with offense completion percentage as the predictor values and score-differential as the outcome values with fitted regression line included

```
Call:
lm(formula = score_differential ~ offense_completion_percentage,
```

```
    data = nfl)

Residuals:
     Min       1Q   Median       3Q      Max 
-243.110  -64.738   -1.134   60.790  241.983 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 -618.0257    39.2139  -15.76   <2e-16 ***
offense_completion_percentage   10.7814     0.6818   15.81   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 87.62 on 763 degrees of freedom
Multiple R-squared:  0.2468,    Adjusted R-squared:  0.2458 
F-statistic:   250 on 1 and 763 DF,  p-value: < 2.2e-16
```

In order to assess whether or not there is a linear association between X and Y, this paper defines the null and alternative hypotheses as $H_0 : \beta_1 = 0$ and $H_A : \beta_1 \neq 0$ respectively.

We can use a two sided t-test to test the null hypothesis, calculated by $t^* = b_1/\hat{se}(b_1)$, but the $lm()$ gives a P-value we can use. We use this to test at a 5% significant level and is shown below.

```
     P-value    R squared 
6.343907e-49 2.468071e-01 
```

The P-value measures the probability of observing the data if our null hypothesis, that there is no relationship between pass completion and score differential, is true. We can use this to determine whether the relationship between these two variables is statistically significant. The R-squared value explains how much of the variability in score differential can be attributed to pass completion. We use this to establish how good of a fit our model is to the data. We discuss this further in the Discussions section.

Analysis below reaffirms the assumptions made in our linear regression model.

The Figure 3 figure shows random scatter of residuals around 0, which indicates that the residuals have a constant variance. From the pattern of residuals we can conclude that our assumptions of linearity and homoscedasticity hold true in our linear regression model.

The quantile-quantile (Q-Q) plot shown in Figure 4 checks whether our residuals from the regression model are normally distributed. While most of the points align with the fitted line, there is slight deviation at the tails of the regression line. These deviations are small enough that we maintain our assumption of normally distributed residuals.

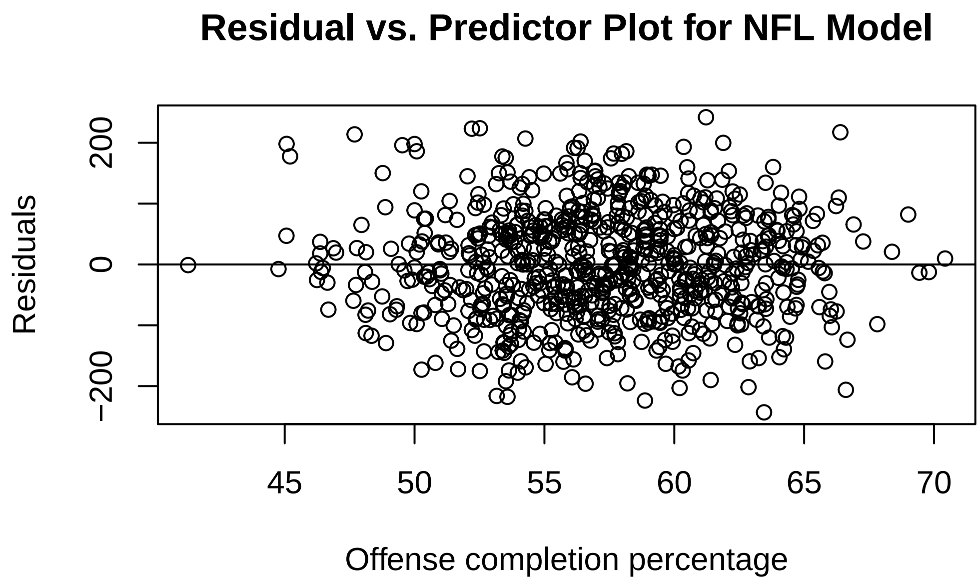

Figure 3: A residual vs predictor plot for NFL model using offense pass completion as a predictor.

Discussion

The p-value found was $6.343907e-49 < 0.05$, which is far lower than our established significance level of 0.05. This allows us to reject the null hypothesis and conclude that there is strong statistical evidence of a linear association between offense pass completion percentage and score differential. Despite this, our R^2 value of 0.2468 indicates that only 24.68% of the variation in score differential can be explained by offense pass completion percentage. So while pass completion is strongly associated with score differential, pass completion alone cannot fully explain the difference in score differential. Other factors, such as the team's defense and the opponent's strength, may play a greater role in determining the outcome of an NFL game.

While the regression itself shows a strong association between the two variables, several assumptions made in this analysis warrant caution. For example, we use pass completion as a proxy for a quarterback's performance. This, however, neglects the role of the receiver and their ability to influence pass completion. Additionally, score differential reflects the performance of the entire team, including the defense, which may obscure the true influence of the quarterback on game outcomes.

Future analyses of this topic should include discussions of the interaction between defensive

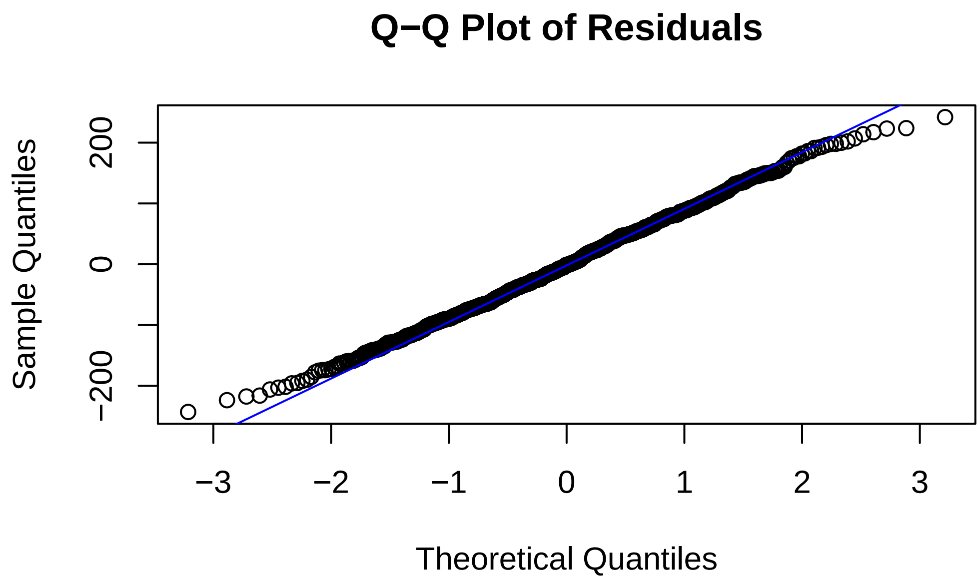

Figure 4: A residual vs predictor plot for NFL model using offense pass completion as a predictor.

performance and score differential as well as include measures of the receiver's influence on pass completion and games outcome.

References

Ho, Tan, and Sebastian Carl. 2025. *Nflreadr: Download 'Nflverse' Data*. https://CRAN.R-project.org/package=nflreadr.

Howe, Jeff. 2025. "Why Explosion of NFL QB Contracts Has Drastically Altered How Teams Build Rosters." https://www.nytimes.com/athletic/6576049/2025/08/28/nfl-qb-contracts-mahomes-allen-prescott-burrow/.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Yurko, Ron. 2023. "National Football League Team Statistics." https://data.scorenetwork.org/football/nfl-team-statistics.html.