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 30, 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. Limitations of this paper include the exclusion of other variables such as the rest of the team's performance.

Introduction

The National Football League, NFL for short, is the most popular sport in America, with 41% of U.S. adults claiming football as their "favorite sport" according to a Jones (2024) poll. This popularity has driven an increase in lucrative media deals for the league, which in turn has driven an increase in player salary demands. To prevent wealthier teams from spending their way to dominance and to maintain competition in the league, policies have been enacted to ensure a balanced competitive field, such as the draft and salary cap. The salary cap, "calculated annually based on a collectively bargained formula tied to league revenues"(Graziano 2025), limits the amount of money an NFL team can spend in a given year. This is particularly interesting due to its implications for team structure.

Despite the limit on how much each team can spend in a given year, in recent years, there has been a dramatic increase in the quarterback's pay as a percentage of the salary cap. Over 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. Additionally, "there are 11 quarterbacks that make more in annual salary than the highest-paid non-quarterback"(Geitheim 2025). Since the "highest-paid quarterbacks make over double what many of the highest-paid running backs, linebackers, safeties, offensive linemen, cornerbacks and tight ends [earn] each year"(Geitheim 2025), crucial positions on the defense, and perhaps even on the offense may be left neglected. Is the quarterback really that important? This paper tries to quantify whether a better quarterback can deliver a greater number 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 performed during a game. Thus a higher score-differential should represent a more winning team. Since every pass starts with the quarterback, pass completion percent is used as a measure of a quarterback's quality.

Our use of score-differential and offence pass completion percent is not perfect however. Score differential reflects the performance of the entire team, including the defense, which may obscure the true influence of the quarterback on game outcomes. Our use of pass completion as a proxy for a quarterback's performance also neglects the role of the receiver and their ability to influence pass completion. Additionally, lineman performance may affect how easily a quarterback is able to perform.

Our analysis finds 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 offensive pass completion percentage.

In this paper, we begin by introducing the reader to our data. We visualize important aspects of this data, then 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 the 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 for the years 1999 to 2022. Play-by-play data dates back to 1999 at the earliest and the latest year where data was collected was 2022. A total of 765 observations were recorded and analyzed in this paper. 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 offensive pass completion percentage and score-differential and as such these are the two relevant variables taken from the raw dataset provided by Yurko (2023). Score-differential is calculated by taking the total number of points scored by the team and subtracting the total number of points allowed by the team. Wins result in a positive number and losses result in a negative number. Pass

completion percent is calculated as the number of successful passes made divided by the total number of passes attempted. A pass is considered successful when a quarterback throws the ball to a receive who catches it.

We make some key assumptions for our use of pass completion and score-differential. For the purpose of this paper we assume that pass completion is dependent entirely on the quarterback and that no one team has an advantage due to their receivers alone. We also assume that all points are scored from passes. Points can be scored without utilizing the quarterback, but since the majority of points are scored through the pass completions, we ignore these sources of points. If any of our assumptions fail, we expect our analysis to change.

There are also limitations with our data. The data we have on pass completion is for each team. As a result, we cannot distinguish between quarterbacks on the same team. If a quarterback is changed in the middle of the season, we expect the team's score-differential to change, but we do not have the means to measure that.
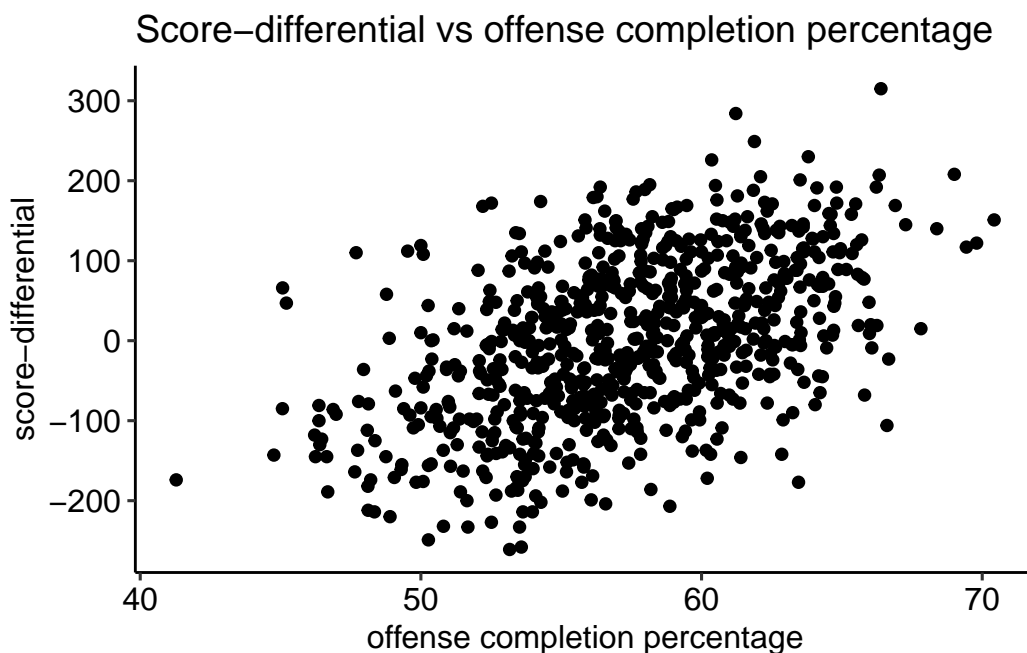

Figure 1: Scatter plot with offense completion percentage as the predictor values and score-differential as the outcome values

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

Table 1: Summary Statistics of Offenseive Completion Percentage and Score-Differential

Variable	Min	1st Qu.	Median	Mean	3rd Qu.	Max
offense	41.28	54.09	57.3	57.32	60.61	70.43
score	-261.00	-74.00	1.0	0.00	72.00	315.00

The table and Figure 1 above shows that on average, an NFL team 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 the offense completion percentage of the ith observation and Y_i represents the score-differential of the ith outcome. β_0 represents the intercept coefficient, which indicates the expected score-differential when the offense has a completion percentage of 0. β_1 represents the slope coefficient, which indicates the expected increase in score-differential for every one 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 there is a statistically significant association between our predictor, pass completion percentage, and outcome, score differential. We define the null hypothesis as $H_0 : \beta_1 = 0$, which states that there is no linear relationship between pass completion and score differential. We define the alternative hypothesis as $H_A : \beta_1 \neq 0$, which states that there is a linear relationship between pass completion and score differential. For our hypothesis test, we assume that the error terms, ε_i, are independent, normally distributed with mean 0, and have constant variance σ^2.

4

Additionally, we assume that the relationship between pass completion percentage and score differential is linear. We use a 5% significance level to evaluate the test.

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.

Table 2: Regression coefficients for predicted score differential.

Parameter	Estimate
B0	-618.0257
B1	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.
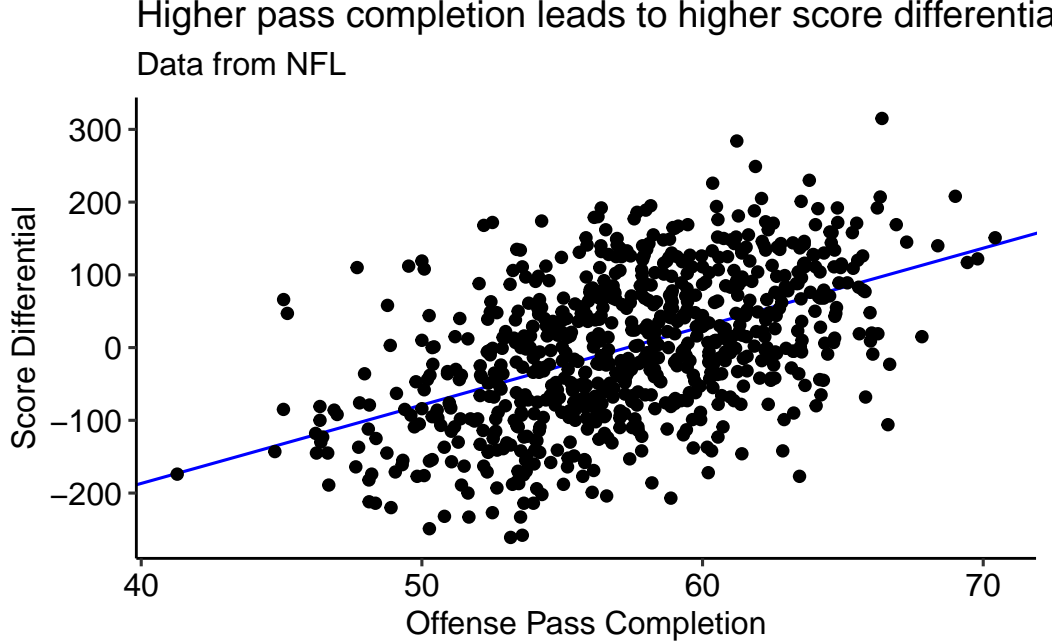
Figure 2: Scatter plot with offense completion percentage as the predictor values and score-differential as the outcome values with fitted regression line included

We can use a two sided t-test to test the null hypothesis, calculated by $t^* = b_1/\hat{se}(b_1)$, where b_1 is the estimated slope and $\hat{se}(b_1)$ is its standard error. However, the $lm()$ gives a P-value we can use. We use this to test at a 5% significant level. For this test to be valid, we assume that the relationship between pass completion percentage and score-differential is linear, the residuals are independent, the residuals are approximately normally distributed, and the variance of residuals is constant.

Table 3: Model summary statistics for predicted score differential

Metric	Value
P-value	6.34e-49
R squared	2.47e-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 and the tests we performed.
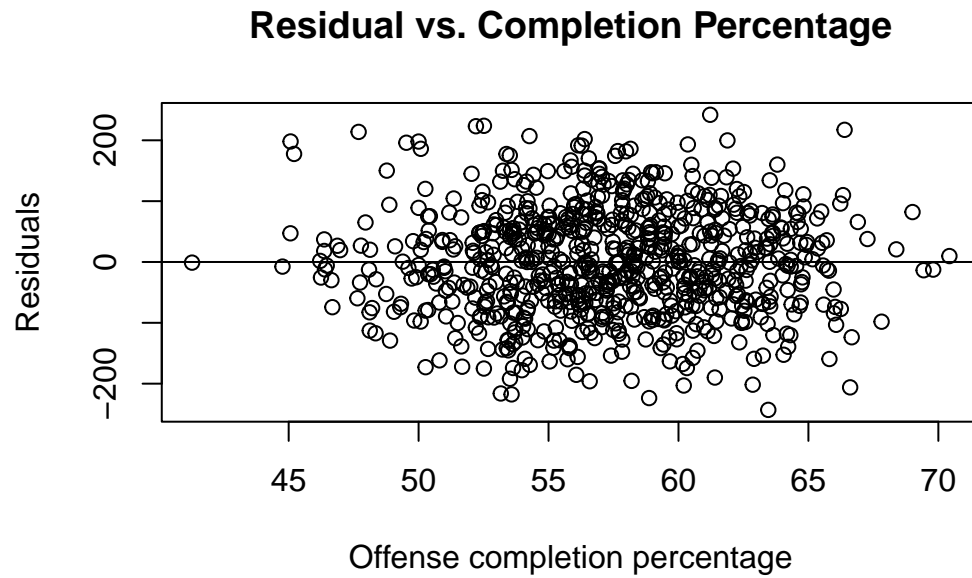
Residual vs. Completion Percentage

Figure 3: A residual plot for score differential predicted by offense pass completion.

The residual plot in Figure 3 displays the residuals of the model against the fitted values. The residuals are scattered around 0, with no pattern, which indicates that the assumptions of linearity and homoscedasticity are satisfied.
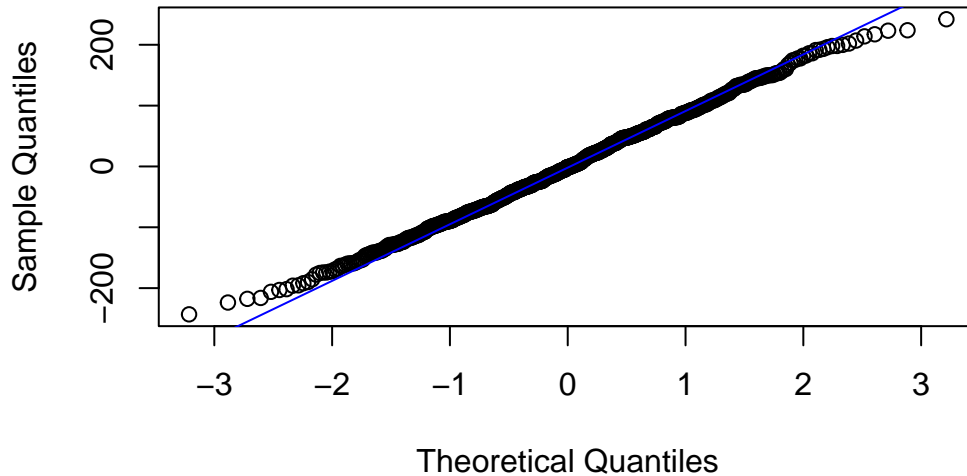
Q–Q Plot of Residuals

Figure 4: Residuals vs. Offense Completion Percentage.

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.

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 performance and score differential as well as include measures of the receiver's influence on pass completion and games outcome. If a team's quarterback changes within the season, future estimates should separate the score-differentials for each quarterback.

References

Geitheim, Eva. 2025. "Highest-Paid Quarterbacks in the NFL: Full Breakdown." https://www.si.com/nfl/highest-paid-quarterbacks-football-full-breakdown?utm_source=gol.

Graziano, Dan. 2025. "NFL Salary Cap Increases Substantially for 2nd Consecutive Year." https://www.espn.com/nfl/story/_/id/43921969/nfl-salary-cap-increases-substantially-2nd-consecutive-year.

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/.

Jones, Jeffrey M. 2024. "Football Retains Dominant Position as Favorite u.s. Sport." https://news.gallup.com/poll/610046/football-retains-dominant-position-favorite-sport.aspx.

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.