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

September 24, 2025

Measuring the importance of an NFL team's offense by

Introduction

Increase in quarterback pay as % of salary cap (**CBS?**). Less room for defense or other positions. Is quarterback really that important?

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

While intuitive, that a better offense should translate to a better chance at winning, games are won not only by proficient scoring, but by competent defending as well.

This paper serves to address intuition of

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

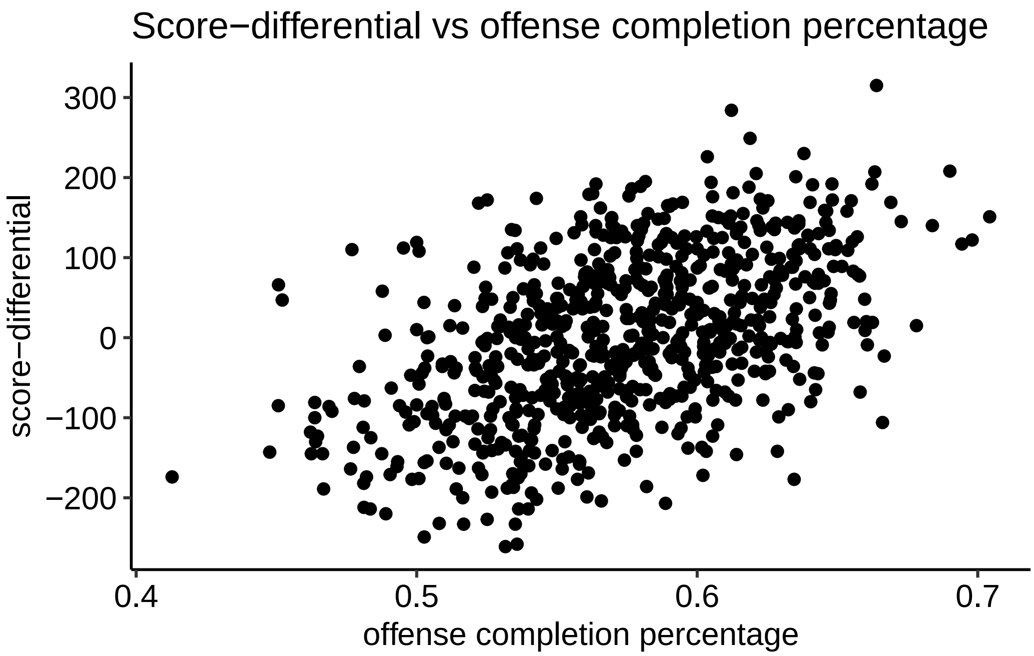

Figure 1: Scatter plot with offense completion percentage as the predictor values and score-differential as the outcome values

dataset provided by (Yurko 2023). A total of 765 observations were recorded and analyzed in this paper.

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.   :0.4128                Min.   :-261
1st Qu.:0.5409                1st Qu.: -74
Median :0.5730                Median :   1
Mean   :0.5732                Mean   :   0
3rd Qu.:0.6061                3rd Qu.:  72
Max.   :0.7043                Max.   : 315
```

The table and Figure 1 above show a concentration of observations around 57% offense completion percentage with the 1st and 3rd quartile being within 3% of 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.

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

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, $beta_0$ and $beta_1$, 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.

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 1078.1405
```

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 1078.14. When pass completion percentage goes to 100%, an increase of 1078.14 would be expected.

Revisualize the scatter plot from the beginning of the paper to better illustrate the linear association.

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

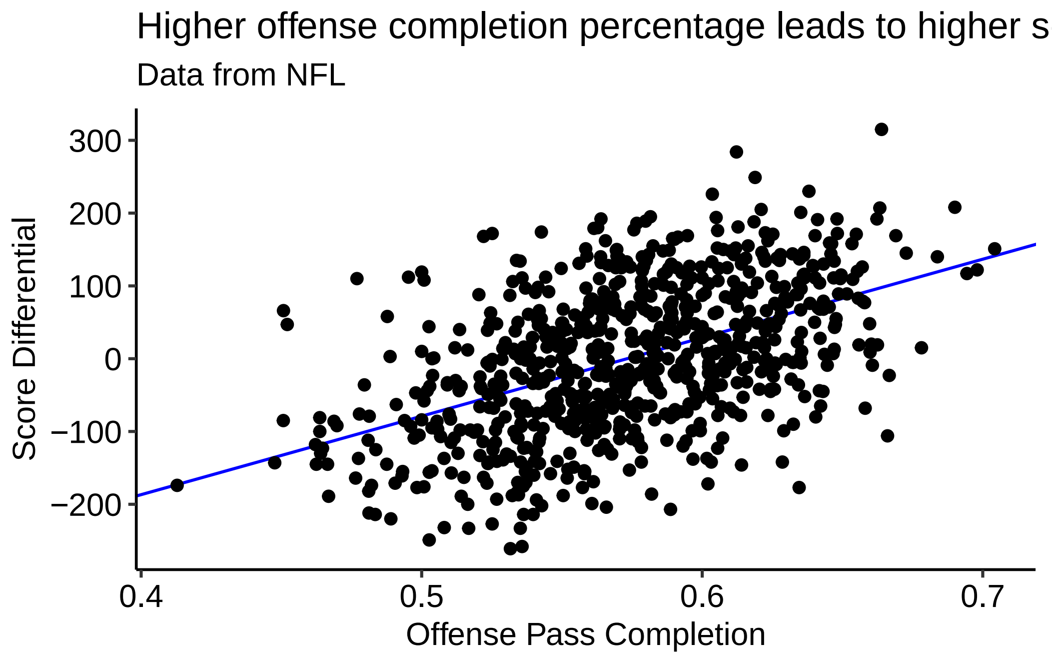

Figure 2

```
    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)                    -618.03      39.21  -15.76   <2e-16 ***
offense_completion_percentage  1078.14      68.18   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
6.343907e-49
```

We find a p-value of $6.343907e-49$.

Discussion

The p-value achieved $6.343907e-49 < 0.05$. This is much lower than our significance level, so we reject the null hypothesis. There is statistical evidence that there is a linear association between offense pass completion percentage and score differential.

Analysis below reaffirms the assumptions made in this paper.

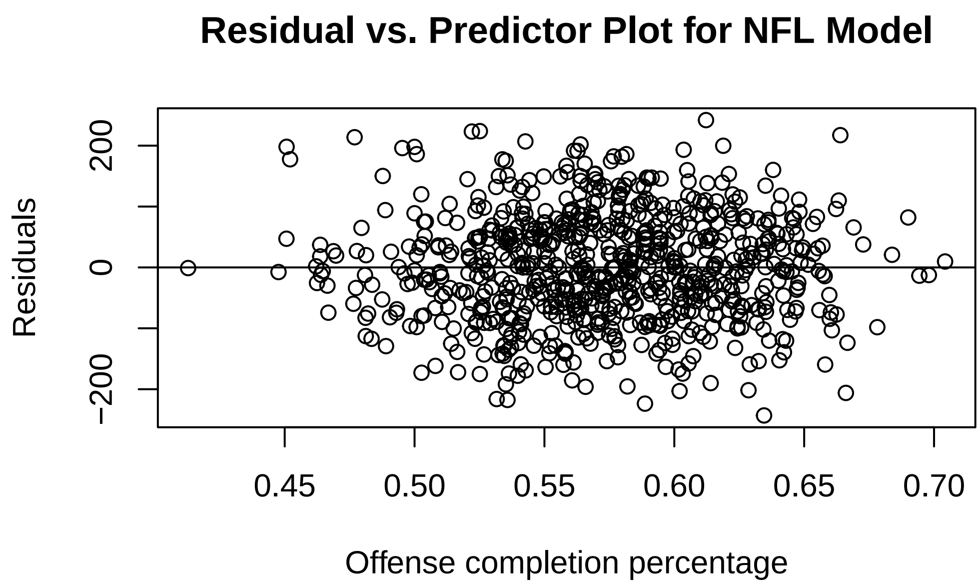

Figure 3: A residual vs predictor plot for NFL model using offense pass completion as a predictor.

The Figure 3 figure shows random scatter around 0, which indicates that the residuals have a constant variance. This is important for showing linearity and homoscedasticity.

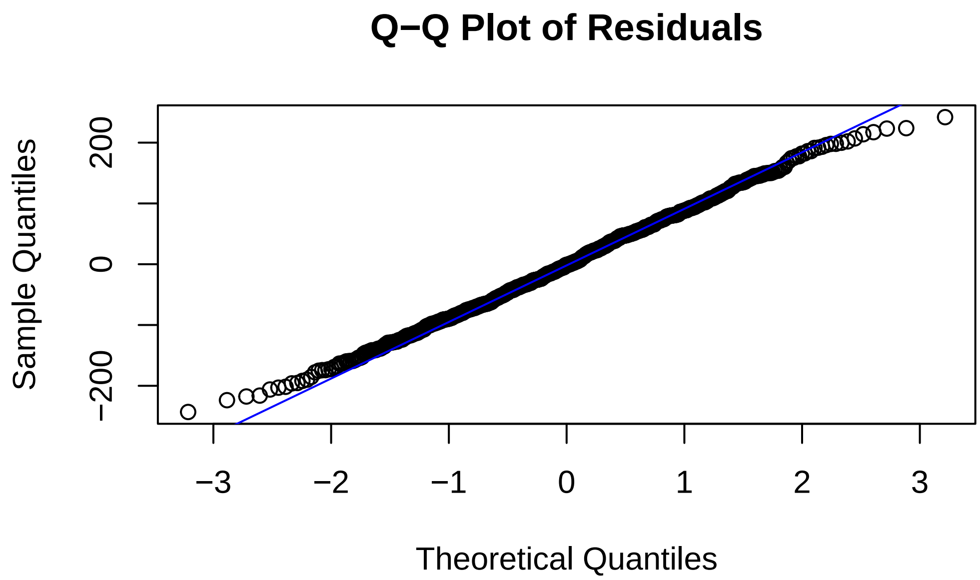

Figure 4: A residual vs predictor plot for NFL model using offense pass completion as a predictor.

The quantile-quantile plot shown in Figure 4 checks if the data is normally distributed. From Figure 4 slight deviation is seen from the tails at the ends of the regression line which would indicate outliers in the data, but it does not appear to be very influential in the dataset.

While the regression itself shows a strong association between the two variables used, other assumptions made in this paper pose questions about the conclusions made. This paper uses pass completion as a proxy for a judge for quarterbacks, but this neglects the quality of the receiver. Additionally score differential can be influenced by the quality of the defense, masking the true influence of the quarterback.

References

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

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.