Offense Pass Completion as a Predictor of NFL Score Differential*

A comparision betweeen

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

---

*Project repository available at: https://github.com/peteragao/MATH261A-project-template.
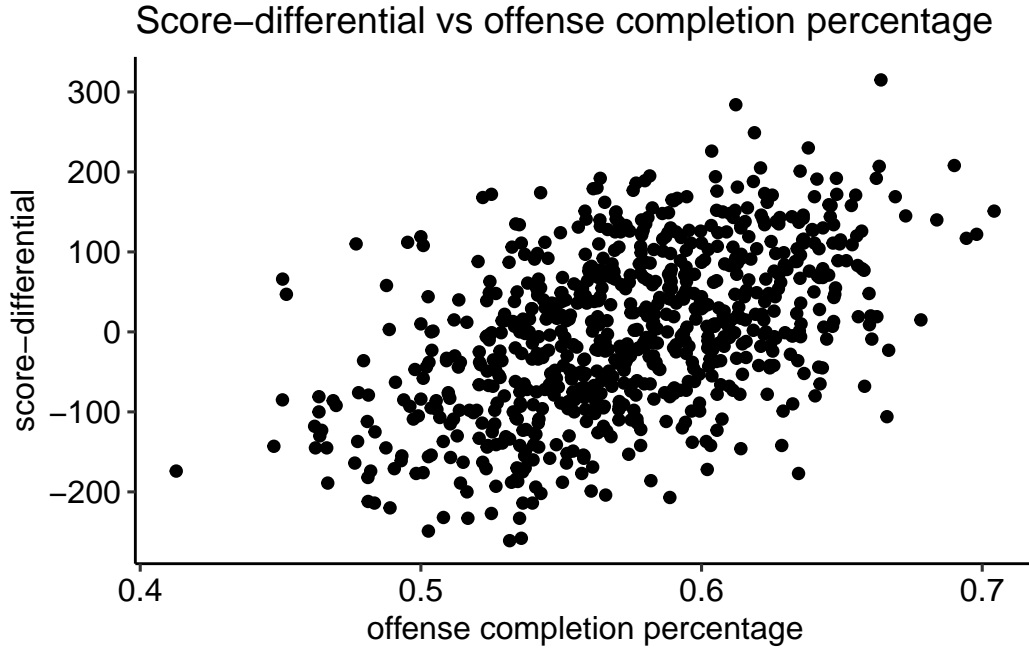
Figure 1: Scatter plot with offense completion percentage as the predictor values and score-differential as the outcome values

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.   :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.

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.

This paper uses a two sided t-test to compare the null hypothesis $\beta_1 = 0$ with the two-sided alternative hypothesis $\beta_1 \neq 0$.

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

Results

```
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)                    -618.03      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
```

weakness: Quality of receiver. This regression analysis also ignores the quality of the team's defense. Sc

Organize the quarterbacks by pay to see if paying more for quarterbacks improves score differential

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.