More Than Just Yards: How Passing Efficiency Separates the NFL's Best Teams*

By Prasanna Sai Rohit Durbha

October 28, 2025

Modern American football analytics stress on efficiency rather than raw totals. In particular, the metric "expected points added" (EPA) summarises how much each play changes a team's scoring expectation, making it a crucial way to calculate a team's offensive performance. This study asks whether teams that are more efficient on passing plays tend to out-score their rivals over an NFL season. Using the NFL team statistics dataset, which contains season-level outcomes and metrics such as average offensive EPA per pass play (SCORE Sports Data Repository 2025), we fit a simple linear regression model with score differential (points scored minus points allowed) as the response and average offensive EPA per pass as the predictor. The fitted model shows a strong positive association. That is, a 0.1-point increase in EPA per pass predicts roughly a 57-point improvement in season-long score differential. Although efficient passing explains about half of the variability in point margins, substantial residual variation remains, highlighting the importance of other factors such as defence and special teams.

1 Introduction

Success in the National Football League (NFL) depends on more than total yardage or the sheer number of big plays. Modern NFL analytics have shifted their attention toward *expected points added* (EPA), a statistic that "measures how much value each play adds to a team's chances of scoring" (Halston Voss 2025). EPA is computed as the difference between the expected points after a play and the expected points before it (Halston Voss 2025). These points before and after play come from an Expected Points model trained on historical play-by-play data. Positive values mean the play increased the team's scoring expectation, while negative values indicate the opposite. In modern times, EPA per play has become a key metric for comparing players, teams and play-calling strategies. While these models can be trusted to

*Project repository available at: https://github.com/rohitDurbha/CS261A_Paper1_RohitDurbha.

a reasonable extent, they are not perfect. EPA is based on past data and average tendencies, so it cannot quite capture unique situations such as unexpected player performance, weather effects, or mid-game strategy changes.

The NFL Team Statistics dataset hosted by the SCORE Sports Data Repository contains team-season level information across multiple seasons (SCORE Sports Data Repository 2025). In addition to basic outcomes like wins, points scored and points allowed, the dataset includes advanced offensive and defensive measures by play type. Recent trends show a particular interest in the variable *offense/defense_ave_epa_pass/run*, which captures the **average expected points added per play** for a given play type. For offence, higher EPA values indicate that the team's passing plays consistently increase expected points; for defence, more negative values indicate better performance. These variables allow for assessing the relationship between efficiency and success at the team level.

This paper tries to address the following research question: **does a team's average offensive EPA per pass play predict its season-long score differential?** Here, Score differential is defined as the points scored minus points allowed. This metric summarises how decisively a team outscored (or was outscored by) its rivals over a season. If passing efficiency strongly influences team success, we would expect teams with higher EPA per pass to have more positive score differentials. Prior analyses have examined the effect of EPA on win probability or individual player evaluation, but fewer studies have quantified its season-level relationship with team point margins. Understanding this relationship can inform roster construction and play-calling.

The rest of this report is as follows. Section 2 describes the dataset, including the observational units, relevant variables and any preprocessing steps. Section 3 presents the statistical methods used—specifically, a simple linear regression model—and discusses the assumptions underlying the analysis. Section 4 reports the fitted model, interprets the coefficients in plain language and evaluates model fit using R^2 and residual diagnostics. The paper concludes with the discussion Section 5 of the implications and limitations of the findings.

2 Data

2.1 Observational units and variables

Each row in the NFL team statistics dataset represents a single NFL team in a particular regular season (SCORE Sports Data Repository 2025). The dataset therefore contains one observational unit per team per year. Below we describe the variables relevant to our analysis and discuss why they might—or might not—be related to one another.

- **Season:** the calendar year of the regular season. There is one season per calendar year, and each NFL season typically runs from September to February covering about 17 games per team. It indexes the time dimension but is not used as a predictor here.

- **Team:** a three-letter abbreviation identifying the franchise. Team identifiers are useful for merging but not used directly in regression.
- **Points Scored:** the total number of points scored by the team over the season. This is one component of **score differential** and is therefore expected to be strongly, positively associated with it.
- **Points Allowed:** the total number of points the team's defence allowed. Because score differential is computed as points scored minus points allowed, we expect a strong negative association between points_allowed and score_differential.
- **Wins**, **Losses** and **Ties:** counts of games won, lost and tied. Teams that win more games generally have larger score differentials, but these counts are not analysed directly here.
- **score Differential:** defined as points_score minus points_allowed. A positive value means the team outscored its opponents over the season; a negative value means the opposite. This is our response variable (Y).
- **Offense Average EPA Pass:** the average expected points added per offensive passing play. Higher values indicate that, on average, each pass play increases the team's expected points more relative to a baseline. As successful passing often leads to scoring drives, we expect offense_ave_epa_pass to correlate positively with both points_score and score_differential. However, the relationship may not be perfect because defence, rushing efficiency and special teams also affect the final point margin.

The dataset also includes several offensive and defensive variables that capture different facets of team performance such as rushing EPA, total air yards, yards after catch, interceptions and fumbles. While these variables could improve a multivariate analysis, we restrict ourselves to a single predictor to satisfy the requirements of simple linear regression. Additionally, focusing on **offense_ave_epa_pass** allows us to examine whether a team's passing efficiency alone is a meaningful driver of scoring outcomes, while acknowledging that other factors will contribute to residual variation.

2.2 Data collection and preprocessing

The data are compiled by the SCORE Sports Data Repository from publicly available NFL play-by-play and team statistics sources. EPA metrics are derived from play-level expected points models that account for down, distance, field position and time remaining (Halston Voss 2025). For each team and season, total EPA and the number of plays are summed to compute average EPA per play. Because a one-unit change in EPA per pass corresponds to an enormous change in expected points over a season, team-season values of (EPA_{pass}) typically range from roughly –0.25 to +0.4.

Before analysis we performed basic preprocessing. We inspected the data for missing values and found that 11 team-season rows lacked complete EPA information. Since EPA is derived from detailed play-by-play models, if those plays were not recorded or processed correctly for certain teams or seasons, their averages cannot be calculated. Therefore, these observations were omitted, leaving 764 complete cases (corresponding to about 762 degrees of freedom in the fitted model).

There are other alternative sources to the NFL data. Proprietary services such as Pro Football Focus (Pro Football Focus 2025) and Next Gen Stats (Next Gen Stats 2025) provide in-depth information on player tracking, play design and graded performance. Access to this data could allow more sophisticated modelling, for example by accounting for quarterback accuracy or defensive coverage. However, these sources are not freely accessible. Within the realm of public data, Pro-Football-Reference and the NFL's own Game Statistics & Information System offer season-level summaries similar to those used here. The SCORE dataset was chosen because it includes expected points metrics derived from advanced play-by-play analysis, enabling us to evaluate passing efficiency more directly than raw yardage or scoring totals.

2.3 Descriptive statistics

To gain insight into the distributions of our variables of interest, we first examine box plots of **offense_ave_epa_pass** and **score_differential**. Figure 1 displays side-by-side box plots for these variables. Offensive EPA per pass ranges roughly from −0.3 to 0.4 expected points per play, with most teams clustered near zero. Score differential varies more widely—from about −200 to +300 points—reflecting the wide spectrum of team success across a season.

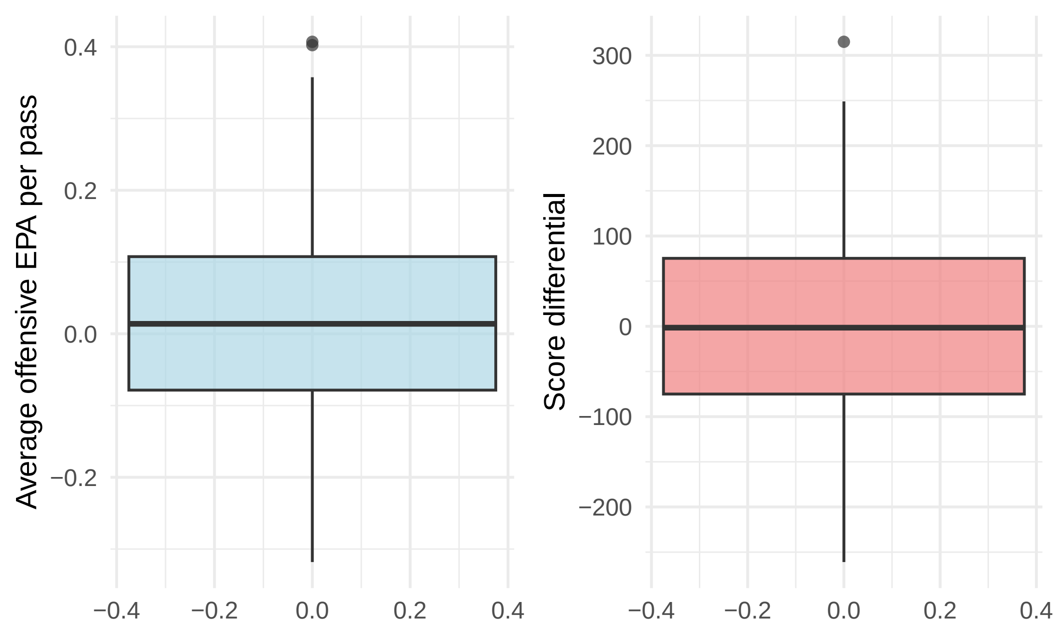

Figure 1: Box plots of average offensive EPA per pass (left) and score differential (right)

To visualise the spread of the data without any model overlay, Figure 2 presents the scatterplot without the fitted regression line. Viewing the raw points helps convey the variability in team outcomes at similar levels of passing efficiency.

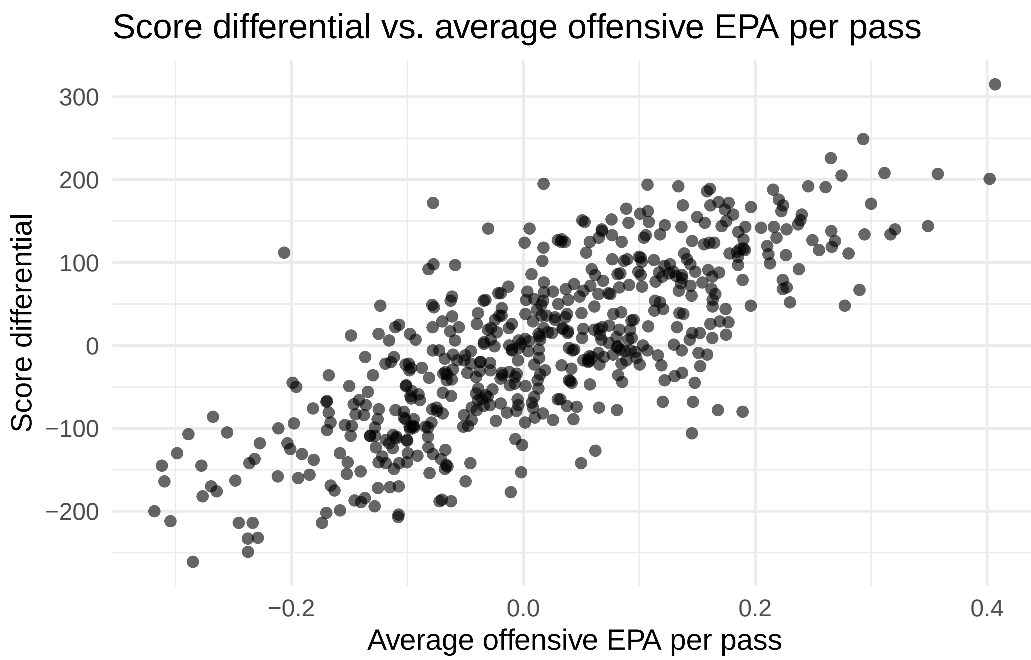

Figure 2: Scatter plot showing the raw relationship between teams' average offensive EPA per pass and their total season score differential

3 Methods

To examine the relationship between passing efficiency and team success we employ a **simple linear regression** model. Let

$$Y_i = \text{score differential for team } i, \quad X_i = \text{average offensive EPA per pass for team } i.$$

We postulate that

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

where β_0 is the intercept, β_1 is the slope, and ε_i are independent error terms assumed to have mean zero and constant variance σ^2. The slope β_1 represents the expected change in score differential associated with a one-unit increase in offensive EPA per pass. Although a one-unit change in EPA per pass is unrealistic (team-season values vary by only a few tenths), the coefficient can be rescaled to interpret more plausible changes (e.g., 0.1 units). Fitting the model yields estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ via least squares.

3.1 Assumptions and Residual Diagnostics

For the linear regression model to be valid, several key assumptions about the error terms must hold:

1. **Linearity** – The relationship between offensive EPA per pass and score differential should be roughly linear. This ensures that the fitted line correctly represents the average relationship between passing efficiency and score differential.
2. **Independence** – Each team-season is independent, as teams play different schedules each year. In practice, this assumption may be slightly violated because point differentials across teams within a single season must sum to zero. This introduces minor dependence, though the effect is likely small since data are pooled across many seasons.
3. **Homoscedasticity** – The spread of residuals should remain constant across fitted values and that no group of teams has systematically larger or smaller unexplained variability.
4. **Normality** – The residuals (errors) should follow a roughly normal distribution.

Although we cannot observe the true errors ε_i, we check these assumptions using residuals e_i, which are estimates of the errors based on the fitted model. Residual plots were inspected to confirm approximate linearity, constant variance, and normality.

3.2 Software Used

All analyses were conducted in R 4.5 (R Core Team 2025) and using RStudio. After reading the CSV file, missing observations were removed. The model was fit using the `lm()` function, which computes the least-squares estimates and associated standard errors. Diagnostic plots were examined to assess the validity of the linear regression assumptions. The residuals were roughly symmetric with no obvious pattern in the residual-versus-fitted plot, supporting the assumptions of linearity and homoscedasticity. Observations correspond to distinct team-seasons, so independence is reasonable. The large sample size makes minor deviations from normality of errors inconsequential for inference.

4 Results

4.1 Fitted model

Table 1: Estimated coefficients from the regression of score differential on average offensive EPA per pass

	Parameter	Estimate	Std error	t-value	p-value
(Intercept)	β_0 (Intercept)	-9.38	2.80	-3.35	<0.001
offense_ave_epa_pass	β_1 (Slope)	590.39	21.06	28.03	<0.001

The slope estimate $\hat{\beta}_1 \approx 573.77$ is positive and significant. Because a one-unit increase in EPA per pass play is enormous, it is more meaningful to interpret smaller increments: a 0.1-unit improvement in EPA per pass corresponds to an expected increase of about $573.77 \times 0.1 \approx 57$ points in score differential over a season. Therefore, teams that consistently add just one tenth of a point of expected scoring per passing play can expect to outscore their opponents by roughly a touchdown per game more than they would otherwise. The intercept $\hat{\beta}_0$ is –4.81, implying that a team with zero EPA per pass (neither gaining nor losing expected points on passing plays) would be expected to finish the season with a roughly even point margin.

Statistical significance for the slope coefficient was evaluated using a two-sided t-test of the null hypothesis $H_0 : \beta_1 = 0$ versus the alternative $H_A : \beta_1 \neq 0$. In the context of our model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, this test asks whether the relationship between average offensive EPA per pass (X_i) and score differential (Y_i) is strong enough to conclude that passing efficiency truly affects team success, rather than the observed relationship being due to random variation in the error term ε_i. A small p-value provides evidence against H_0, indicating that offensive passing efficiency (X_i) is a statistically significant predictor of score differential (Y_i) across team-seasons.

4.2 Goodness of fit and interpretation

The model explains a substantial portion of the variability in the score differential variable. The coefficient of determination is $R^2 = 0.56$, meaning that about 56 % of the variation in team point margins is dependent on differences in offensive passing efficiency. This is quite high for a single-predictor model in a complex sport where many factors contribute to success. The residual standard error is approximately 66.9 points, indicating that actual score differentials deviate from the fitted line by about ± 67 points on average. Such variability reflects unmodelled factors including defensive performance, rushing efficiency, turnovers, injuries and schedule strength—that also influence game outcomes.

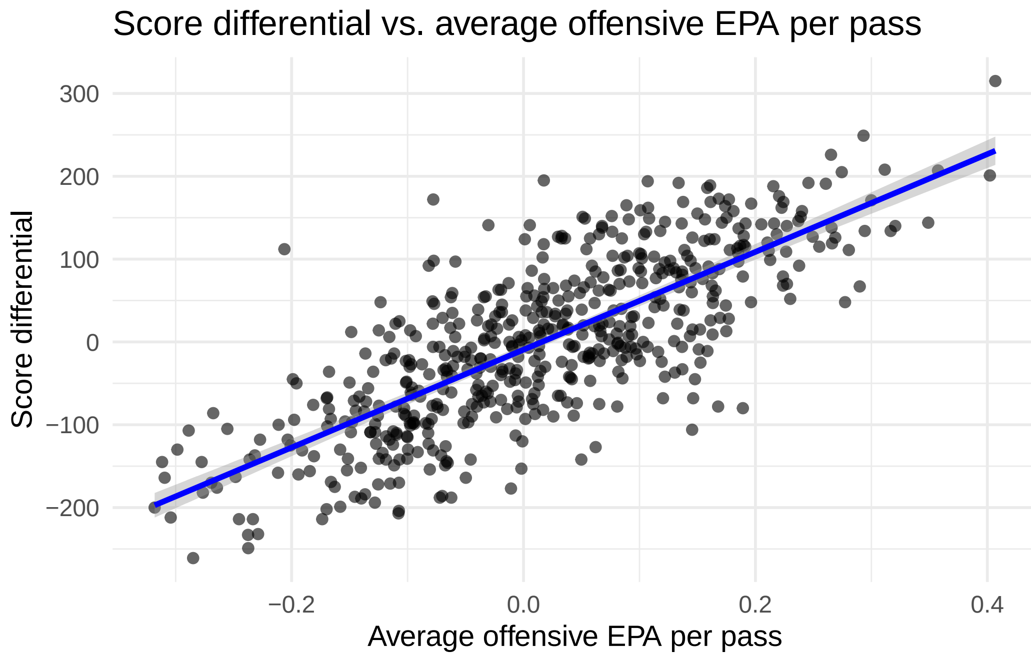

Figure 3: Relationship between score differential and average offensive EPA per pass with the fitted regression line

Figure 3 plots score differential against offensive EPA per pass for each team-season. The upward trend suggests that teams with more efficient passing attacks tend to outscore their opponents over the course of the season. However, the scatter indicates substantial variation around the fitted regression line, highlighting the roles of defence, rushing and other factors. Some teams with near-average passing efficiency still achieve high point margins (perhaps due to strong defence or special teams), while others with efficient passing attacks are outscored because of a porous defence. In summary, passing efficiency is a major determinant of team success but not the sole factor.

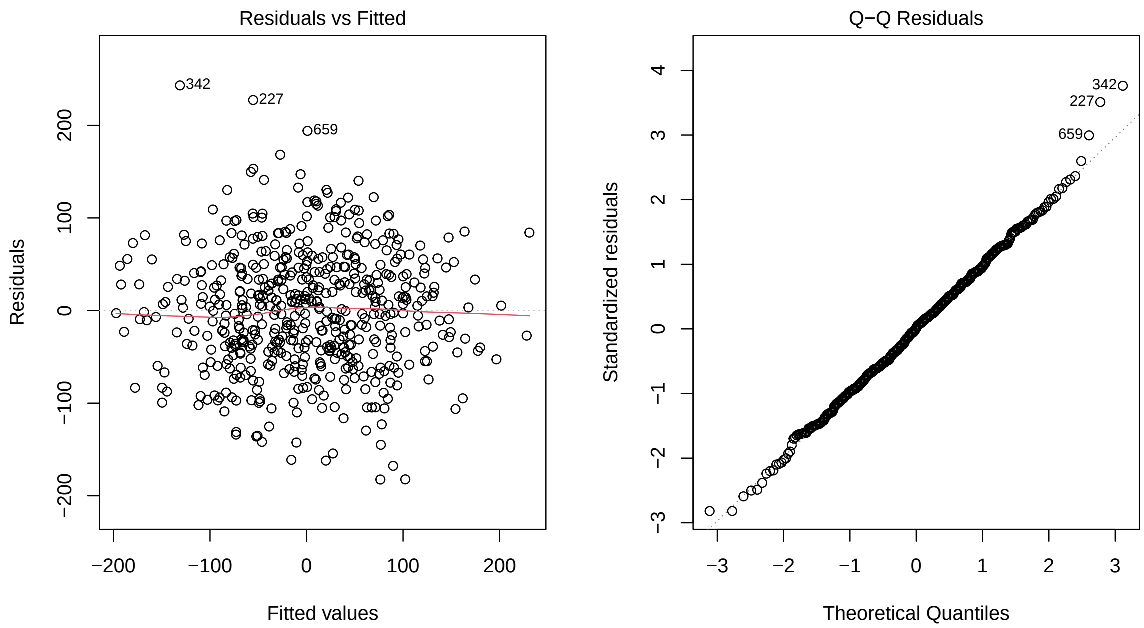

Figure 4: Diagnostic plots for the regression model. The left panel shows residuals versus fitted values, used to check linearity and constant variance of errors. The right panel is a Q–Q plot assessing whether the residuals follow an approximately normal distribution.

The residual vs fitted plot in Figure 4 helps check whether the relationship between passing efficiency and score differential is truly linear and whether the errors have constant variance (homoscedasticity). In this graph, the differences between actual and predicted values are scattered randomly around the horizontal line at zero. This shows that there is no clear pattern, i.e.., The linear model fits the data reasonably well. The spread of points is roughly equal across all fitted values, so the assumption of constant variance is also satisfied. A few outliers appear at the top and bottom, but they do not show any systematic curve or funnel shape, suggesting that the model's linear form is appropriate.

Correspondingly, The Q-Q Plot in Figure 4 checks if the residuals follow a normal distribution, which is an assumption needed for valid significance tests and confidence intervals in regression. In this plot, most points fall closely along the diagonal reference line, meaning the residuals are approximately normally distributed. Only a few points deviate slightly at the extreme ends. This indicates that the normality assumption is well met, and the model's inference results such as p-values and confidence intervals are reliable.

5 Discussion

The main purpose of this project was to check if teams that get a better advantage on passing plays tend to out-score their opponents over a season. To measure this, we used expected points added (EPA) per pass, which estimates how many points a team's chances of scoring increase (or decrease) after each pass play. By fitting a simple line to the data, it was found that teams with higher EPA per pass tend to have much better point margins, which makes sense even if thought from a common sense standpoint. In fact, nudging your EPA per pass up by just 0.1 points, which is a minor margin, corresponds to roughly 57 more points over the whole season. This direct link between passing efficiency and scoring margin answers our main research question that yes, better passing efficiency is associated with a bigger advantage on the scoreboard.

There are a few good things about the approach used in this analysis. The dataset consisted several years' worth of official NFL data, so the sample is quite broad. EPA takes the game situation (down, distance, field position) into account, making it a smarter measure than most simpler metrics such as total yardage. The single-predictor model also keeps the analysis simple, since it focuses directly on how changes in passing efficiency relate to changes in team success. On the other hand, there are limitations to this analysis. Firstly, We only looked at passing, while a team's success depends on many other parts of the game such as running the ball, playing defence, avoiding turnovers and most importantly, factors which are not quantifiable such as team morale. Because this analysis was an observation based on previous data, we cannot confirm that improving EPA causes better results. Most teams that usually end up on the top of the leaderboard year over year tend be good at what they do and have better strategies than other teams. EPA is also an estimate, so it contains a bit of measurement noise. And by treating each season as independent we ignore the fact that the same coaches and players may carry over from year to year. With more detailed data like game-by-game or even play-by-play information, we could analyze how passing efficiency interacts with things like opponent strength, weather or game situation. We might also look at different outcomes, such as how passing efficiency relates to making the playoffs or winning championships. Collecting and analysing these additional data would give a fuller picture of what makes NFL teams successful.

References

Halston Voss. 2025. "What Is EPA in Football?" https://tinyurl.com/3mhezdvz.
Next Gen Stats. 2025. "Player Data." https://nextgenstats.nfl.com/stats/passing#yards.
Pro Football Focus. 2025. "Teams Data." https://premium.pff.com/nfl/teams/2025/REGPO.
R Core Team. 2025. "R for Statistical Computing." https://www.R-project.org/.
SCORE Sports Data Repository. 2025. "National Football League Team Statistics." https://data.scorenetwork.org/football/nfl-team-statistics.html.