

Analyzing Yards Gained Based on QB Position

Hanna Festa
Ted Henson
Victoria Pierce
Conner Mann
Shane Stephens

Introduction to the Data

The goal of football is to gain yardage on your opponent in pursuit of the end zone; it's a simple goal, but coaches have spent decades trying to find the most efficient way to do so. While many have introduced radical ideas, such as the Wildcat, triple option, and RPO, we theorized that perhaps the formation that the team lined up in was influential. Specifically, whether the quarterback lines up under center, or in the backfield in "shotgun" formation, could influence how successful a play would be. We hoped to find data that suggested that one formation was better than another at yardage gained, accounting for our other variables. We gathered data over the course of the NFL's Weeks 4-6, picking a random assortment of games and assigning one team member to each to watch and record every play. Our five variables were: "Play Type", "Offensive Team", "Home Team", "QB Position", and "Yards Gained". Below are the first 6 rows of the raw data that we logged:

Table 1: Preview of the data

QB Position	Play Type	Yards Gained	Offensive Team	Home Team
1	0	3	Bears	1
1	1	3	Bears	1
1	0	6	Bears	1
1	1	6	Bears	1
1	0	0	Bears	1
0	1	1	Bears	1

QB Position:

- Shotgun = 1
- Under Center = 0

Play Type:

- Run = 1
- Pass = 0

The "QB Position" variable was a binary variable indicating whether the quarterback took the snap in the shotgun formation. This "Yards Gained" variable allowed us to examine the differences in yardage the offensive team gained in a given formation. The "Play Type" variable was a binary variable indicating whether a play was a run or a pass. It was logged to account for the fact that some teams tend to run more from under center, and pass more from the shotgun. This would also allow us to examine whether running from the shotgun was truly a disadvantage for the running back. Our "Offensive Team" variable allowed us to account for different coaches and personnel, as some teams are notoriously shotgun heavy (such as the Chiefs) while some play from under center a lot (such as the Titans). Our binary "Home Team" variable measured which team was the home team; we felt this was important as QBs often say they have harder times giving calls from the shotgun in hostile environments, due to the extra distance and noise of an opposing crowd. Additionally, research across a variety of sports has demonstrated that the home team gets advantageous calls from officials.

For the rest of the analysis, we performed some variable transformations on the raw data for easier interpretation and analysis. The “QB Position” and “Play Type” variables were transformed into categorical variables with the levels “Shotgun”, “Under Center”, and “Run”, “Pass” respectively. The “Offensive Team” variable was renamed “Team” since we only consider the offensive team in this analysis. The variables “QB Position”, “Play Type”, and “Yards Gained” were renamed as follows: “QB.Position” “Play.Type”, and “Yards.Gained” for easier programming when running the anova model.

Summary of the Data

The first summary table (Table 2 shown below) examines two categorical variables, Home and Team. The number of observations (or plays) for each category is included as well as the percentage of total plays in each category. There are also summary statistics for each category which include the average, minimum, maximum and standard deviation of yards gained in each category. Unsurprisingly, the average yards gained per play for a home team is higher than that of the away team however both have a high standard deviation.

Table 2: Yards gained for Home and Team variables

Home	Number of Plays	Pct of Plays	Average	Minimum	Maximum	SD
0	57	41.01	3.895	-5	22	5.059
1	82	58.99	5.378	-27	34	8.431

Team	Number of Plays	Pct of Plays	Average	Minimum	Maximum	SD
Bears	18	12.95	5.556	-4	25	6.446
Cardinals	19	13.67	4.474	-27	31	10.57
Cowboys	13	9.353	3.615	0	12	4.073
Giants	10	7.194	3.6	0	11	4.248
Patriots	15	10.79	5.4	-7	21	8.322
Rams	15	10.79	4	-5	22	6.579
Saints	14	10.07	5.143	-10	18	7.594
Seahawks	26	18.71	5.923	-3	34	7.944
Vikings	9	6.475	3.111	0	13	4.485

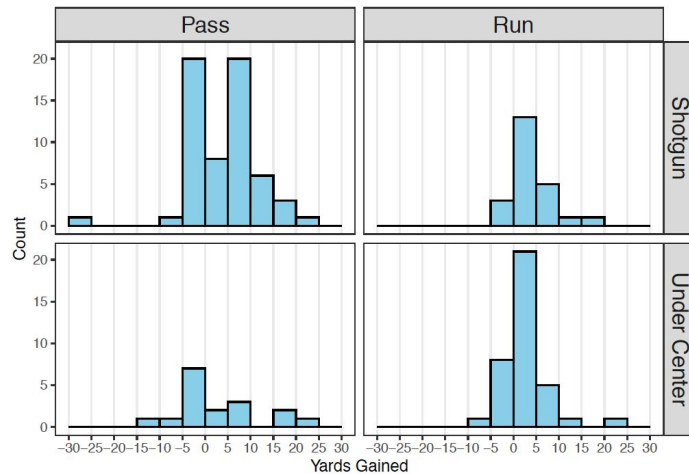
The second summary table, labeled as Table 3 below, depicts the average, minimum, maximum and standard deviation of yards gained within each combination of QB position and play type. Average yards gained with a shotgun QB position are higher for both play types than under center. The number of plays recorded within each QB position and play type combination is also included.

Table 3: Yards gained for all combinations of QB position and play type

QB Position	Play Type	Number of Plays	Average	Minimum	Maximum	SD
Shotgun	Pass	62	5.903	-27	34	8.79
Shotgun	Run	23	5.087	-3	18	4.358
Under Center	Pass	17	4.118	-10	22	8.609
Under Center	Run	37	2.973	-5	21	4.531

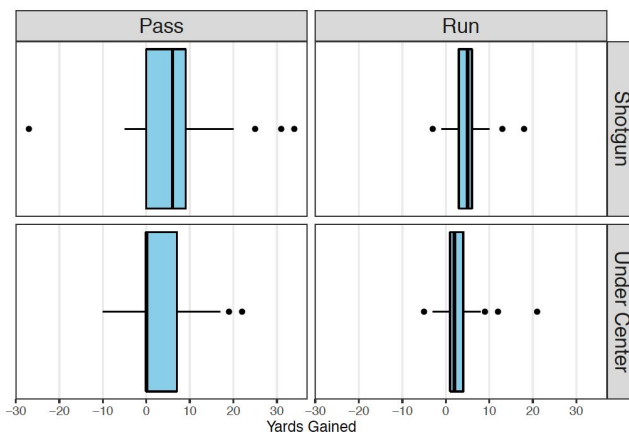
The figures below illustrate the distribution of yards gained within QB position and play type combinations. The first figure illustrates the distributions using histograms with 5 yard intervals.

Figure 1: Yards gained in QB position and play type combinations shown as histograms



The second figure shown below illustrates the distribution of yards gained within each QB position and play type combination using box plots. It appears more clear in this plot that the distributions for shotgun plays are higher than under center plays.

Figure 2: Yards gained in QB position and play type combinations shown as box plots



Insights from the Data

Looking at the data in Table 3 and Figure 2, we found that shotgun plays had a higher average number of yards gained in general than when the team lined up under center. This is the most evident when running the ball. Running the ball with the QB in the shotgun position has an average of 5.087 yards gained, while running the ball with the QB lined up under center only has an average of 2.973 yards

gained. Additionally, our data show that the highest average yard gain occurs when the QB is in the shotgun position and passes the ball, with an average of 5.903 yards gained. Looking at Table 2, we found that the home team had an average yards gained of 5.378 while the away team average was 3.895. Based on the collected data we came up with two hypotheses that we wanted to test for statistical significance seeing as standard deviation values in our summary statistics were high. The first hypothesis was that lining up in the shotgun will generate more yards than lining up under center. The second hypothesis was that when the home team lines up in the shotgun, they will gain more yards than when the away team lines up in the shotgun.

We used ANOVA and Tukey tests to check the significance of the discrepancies we observed in our data and the results of both tests are shown below in tables 4 and 5. Though the summary statistics show slightly higher numbers for shotgun plays, the Anova and Tukey HSD test in tables 4 and 5 both show that we do not have significant evidence to conclude that the QB position is statistically significant in determining the number of yards gained at an alpha level of 0.05. With that being said, the Anova model calculated the p-value to be 0.076 (Table 4) and the Tukey HSD model calculated the adjusted p-value to be 0.066 (Table 5) for QB position. Though we can not conclude a significant difference does exist, we do have a substantial trend towards significance that the position of the QB has an effect on the number of yards gained. Furthermore, in our analysis of table 3 and both figures, we thought that the interaction between play type and QB position would be significant seeing as play type seemed to affect how yards gained changed from shotgun to under center. However, the p-values for the interaction of play type and QB position resulting from the anova test and Tukey test (Tables 4 and 5) are not significant so we did not find statistical evidence of this claim either. Likewise, the ANOVA test only had a p-value of 0.199 (Table 4) for the QB position and Home vs. Away. None of the p-values associated with the Home variable or the interaction between QB position and Home were significant at an alpha level of .05 or even .1. Thus, our second assertion is not significantly supported either. In fact, (as seen in Tables 4 and 5) none of the variable interactions we measured were significantly related under a 95% confidence interval.

The lack of statistical significance in our data could be attributed to the size of our dataset. If we had to do this project again we would collect a larger sample size so that we could lower the variance and therefore make more assertive claims. By collecting data from more than 5 games on more than 10 teams we might be better able to see if being the home team has an affect on yards gained from the shotgun position. If we wanted to dig deeper into success of the shotgun we could also look at data of different quarterbacks to specifically see which are more successful with runs or passes, at home or away and in shotgun or under center. Additionally, we would have logged our data as categorical when possible rather

than using 0-1 designations for more intuitive analysis and printing of tables. By the same token, we would have been more consistent with our variable names throughout the analysis. This investigation indicates that further analysis should be done to understand the statistical significance of the effect of QB position on yards gained with the hypothesis that yards gained increases on shotgun plays.

Table 4: ANOVA Test results

```
###Summary of Anova Model

##              Df Sum Sq Mean Sq F value Pr(>F)
## QB.Position    1   182   182.21   3.210  0.076 .
## Play.Type      1    26    25.70   0.453  0.503
## Home          1    43    43.19   0.761  0.385
## Team          8    88    10.96   0.193  0.991
## QB.Position:Play.Type 1     0     0.08   0.001  0.970
## QB.Position:Home     1    95    94.89   1.672  0.199
## Play.Type:Home       1    20    20.42   0.360  0.550
## QB.Position:Team     8   374    46.78   0.824  0.583
## Play.Type:Team       7   343    49.00   0.863  0.538
## QB.Position:Play.Type:Home 1    37    37.48   0.660  0.418
## QB.Position:Play.Type:Team 2    39    19.32   0.340  0.712
## Residuals        106   6017    56.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 5: Tukey Test results

```
###Tukey HSD

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Yards.Gained ~ QB.Position * Play.Type * Home, data = data)
##
## $QB.Position
##              diff          lwr          upr          p adj
## Under Center-Shotgun -2.34902 -4.856365 0.1583259 0.0660858

##
## $Play.Type
##              diff          lwr          upr          p adj
## Run-Pass -0.7925771 -3.259931 1.674776 0.5262363
##
## $Home
##              diff          lwr          upr          p adj
## 1-0 1.101828 -1.382884 3.586541 0.3819645
##
## $`QB.Position:Play.Type`
##              diff          lwr          upr          p adj
## Under Center:Pass-Shotgun:Pass -1.9203125 -7.109359 3.268734 0.7706253
## Shotgun:Run-Shotgun:Pass -0.9234185 -5.550893 3.704056 0.9543293
## Under Center:Run-Shotgun:Pass -2.9106627 -6.848114 1.026789 0.2231877
## Shotgun:Run-Under Center:Pass 0.9968939 -5.065386 7.059173 0.9735978
## Under Center:Run-Under Center:Pass -0.9903503 -6.543839 4.563138 0.9667344
## Under Center:Run-Shotgun:Run -1.9872442 -7.019992 3.045504 0.7336212
##
## $`QB.Position:Home`
##              diff          lwr          upr          p adj
## Under Center:0-Shotgun:0 -0.7157014 -5.737436 4.306033 0.9825129
## Shotgun:1-Shotgun:0 2.2084877 -2.165599 6.582574 0.5557639
## Under Center:1-Shotgun:0 -1.0447393 -6.260087 4.170608 0.9538353
## Shotgun:1-Under Center:0 2.9241891 -1.399029 7.247407 0.2972661
## Under Center:1-Under Center:0 -0.3290379 -5.501797 4.843721 0.9983787
## Under Center:1-Shotgun:1 -3.2532269 -7.799901 1.293447 0.2495539
##
```