Digital Commons @ George Fox University

Student Scholarship - College of Business

College of Business

Spring 2022

How to Win a NFL Football Game

Ian Johnson

Grayson Ponce

Follow this and additional works at: https://digitalcommons.georgefox.edu/gfsb_student

 Part of the Business Commons, and the Sports Studies Commons

How to Win a NFL Football Game

Ian Johnson and Grayson Ponce

George Fox University

Abstract

The purpose of this research is to explain the value of turnovers lost, turnovers gained, total yards, points, and first downs when it comes to winning games. Using cross-sectional data analysis, this study analyzed all of the stats from the 2021 season. Then picked three random teams in the AFC and three teams from the NFC. Results of the study showed that both turnovers gained and lost had the largest impact.

Keywords: R squared, Winning percentage, Regression table Y1

Introduction

For this project, we wanted to find what variables have a significant impact on the winning percentage in the NFL for the 2021 season. To do this we are going to use turnovers lost, turnovers gained, total yards, points scored, and first downs. We expect that both turnovers gained and turnovers lost will have a large impact on a team's chance of winning. As football players, we have heard our coaches tell us the importance of not turning the ball over. We have seen countless games change for good and for worse because of a turnover in the game. After the regression is completed we will be able to identify the impact these variables have. The magnitude of the coefficient on these variables will show us the importance they have on the chance of winning.

The National Football League (NFL) is a multi-billion dollar business. It is the most successful sports league in North America, outcompeting all other sports in the country. Even with Covid-19 struggles the NFL made $12.2 billion in 2020 and was projected to be a $25 billion dollar business by 2027. Each of the 32 teams gets a split of the resulting profits and then makes individual profits on ticket sales, concessions, and sponsors. These parts of the business bring in millions of dollars for the individual teams. The better a team does the more money they make at the local level. Being able to inform coaches on what things they should focus on could have a dramatic impact on profits for a team. You could advise a team to invest more in defense to increase the number of turnovers they get. This would allow the team to be more competitive and raise profits for both the team and the business itself. The information could also be really important for the sports betting industry. It is estimated that there are over 100 billion dollars wagered on NFL games alone. If sports bettors know what variables affect winning percentage they will be better able to come up with more accurate bets. This could allow people to make

more money on the wagers and give betting companies the ability to create better bets for themselves (Eckstein & Brown, 2022) (Legal Sports Betting).

Data Overview

In the NFL, the organization is split into two halves, one half is AFC and the other half NFC. We used a random number generator to pick 3 teams from both the AFC and the NFC. Then downloaded all of the stats from the 2021 season for all six of the teams. The data was collected from Pro football references. The five estimators consist of turnovers lost, turnovers gained, total yards, first downs, and points scored. Each variable has 102 data points.

The Pro football reference and ESPN are two major sources that we looked at for finding data. Pro football reference data was clean and very easy to read. ESPN had more distractions and was not as clean. We have decided to go with Pro football reference as our source for data.

	Mean	Median	S.D.	Min	Max
Turnoverlost	1.461	1.000	1.264	0.0000	5.000
Tgained	1.314	1.000	1.227	0.0000	5.000
TotYd	338.2	344.0	83.53	155.0	533.0
Tm	20.26	21.00	7.091	0.0000	34.00
stD	22.80	21.00	8.813	6.000	47.00
Wins	0.4902	0.0000	0.5024	0.0000	1.000

$$(\text{winperc.}) = \beta 0 + \beta 1(\text{turnovers lost}) + \beta 2(\text{turnovers gained}) + \beta 3(\text{total yard}) + \beta 4(\text{first downs}) + \beta 5(\text{point scored}) + \upsilon$$

Variable β1 is *turnovers lost*. This is when a team has possession of the ball and then gives it or loses the ball to the other team. This stat is an offensive stat showing when a team gives their opponent another chance to score. This does not include when the team switches possession at the end of a half. The 2nd variable β2 is *turnover received*. This is when a team's

defense takes the ball from its opponent. This is a defensive stat showing when a team now gets a chance to score. The next variable is *total yards*. This is the number of yards that a team advances the ball down the field in a game. This combines both offensive and defensive yards. Variable β4 is *first downs*. This is an offensive stat that shows when an offense advanced the ball 10 yards beyond the starting line of scrimmage. The last variable β5 is *points scored*. This can be done in a number of ways in the game, but the variable measures every point scored by a team by the end of the game.

Looking further into our variables it became clear OLS assumption tests would be strenuous to run because we have a binary variable as our outcome variable. This makes scatter plots and trend lines basically pointless. That being said, looking at some of the assumptions it appears that there could be perfect multicollinearity between our variable's total yards and first downs. It is hard to interpret because gaining yards does not necessarily mean you are getting first downs. You have to gain yards in order to get first downs. It is technically possible to have a lot of yards and not many first downs but it is improbable. For this reason, we decide that because the two variables can stand alone that there wasn't multicollinearity between them. In order to make sure that we follow the heteroskedasticity assumption we will be using heteroskedastic-robust standard errors. This ensures that we come to a valid conclusion by decreasing the range of errors in our regressions.

Methodology

Our methodology for this project started with using the linear version of each variable in a single regression. In order to achieve this, we created a formula with the variables *turnovers lost, turnovers gained, total yards, total points, first downs,* and our outcome variable *winning*

percentage. Because our outcome variable is a binary variable everything is measured on a scale between 0 and 1. This means that every game that has percentages going over a 1 or under 0 will be considered as such.

(winperc.) = β0 + β1(***turnovers lost)*** + β2(***turnovers gained)*** + β3(***total yard)*** + β4(***first downs)*** + β5(***point scored)*** + υ

We believed that turnovers lost would be the only variable with a negative correlation. Likewise, the other four variables would have a positive correlation. In order to make sure these variables were the best fit we ran a number of regressions on the variables and created graphs individually. We tested each variable in its linear, squared, and log forms. Because our outcome variable is a binary variable the graphs were unable to tell us which of these variables was the most accurate. This led us to use whichever variable had the highest squared value as our most accurate variable. We found that turnovers lost, total points, and first downs had the highest R squared in their original form. Turnovers gained and total yards had higher R squared values in the log form. With further inspection, we found that the regression on turnovers gained was omitting 29 of our 102 observations in the log form. We then decided to go back and look at our data finding that the squared version of turnovers gained had the highest R squared value. With this data, we ran many regressions. Unfortunately, with the new squared variable in our data both the original and the squared version became insignificant at even the 10% level. This caused us to believe that the linear form of turnovers gained would be the best for our formula. Using the linear form of turnovers gained, we were able to create a new formula using the log version of the variable total yards.

$$\textbf{\textit{winperc.}} = \beta 0 + \beta 1(\textbf{\textit{turnovers lost}}) + \beta 2(\textbf{\textit{turnovers gained}}) + \beta 3(ln\ \textbf{\textit{total yard}}) + \beta 4(\textbf{\textit{first downs}}) + \beta 5(\textbf{\textit{point scored}}) + \upsilon$$

After running a few regressions we wanted to see the results if we grouped or singled out certain variables. The first one we wanted to test was the effect turnovers have on the game. We had already predicted that these would have a large impact on the outcome variable. This caused us to want to run a regression with only turnovers lost and turnovers gained. Doing this would allow us to see how they react by themselves separate from the whole regression. The next regression we looked to add was a single regression on total yards. We had predicted that this would have a positive correlation. To our surprise, a practice regression showed that it was negatively correlated with winning percentage. This seemed illogical because you need to get yards in order to be successful in a football game. Because of this, we thought that separating the variable might give us greater insight into what the variable might show.

Results

Starting our regressions we first wanted to see what effect turnovers had on winning percentage. From past experience, we predicted that turnovers would have a large impact on whether a team wins or loses. After running the regression we found that our expectations mirrored our results. We found that the variable turnovers lost had a negative correlation with winning. For every turnover lost in a football game ceteris paribus, you decrease your chances of winning by 17.4 percentage points. This could have a dramatic effect on a football game given that the maximum for turnovers lost is five. Even at the median of one, decreasing your chances of winning by 17.4% is huge. It has a mean of 1.4. This doesn't explain much to us because turnovers can only be whole numbers. That being said, it lets us know that there are a number of games that have two or more turnovers. For the variable turnovers gained we found that it had a

positive correlation with winning. For every turnover gained a team increases its chance of winning by 9.9 percentage points. Turnovers gained also had a maximum of five and a median of 1. We found it captivating that losing possession of the ball had nearly double the effect on winning that gaining possession of the ball did. Both variables and the constant were statistically significant at the 1% level and had an R squared value of .29. This tells us that the variation in both turnovers lost and turnovers gained explains 29% of the variation in winning percentage. This shows that turnovers are a huge piece in whether a team wins or loses an NFL football game.

The second regression we did was to help understand our data further. In our first test regression with all our variables together we found that total yards had a negative correlation. This seemed odd considering that getting yards is usually attributed as a good thing in football. To try and get a better understanding of what was going on we ran a regression with total yards as the only variable. This new regression showed that total yards was positively correlated with winning when it was the only variable. The correlation changed from a negative -.19% to a positive .2% per yard. It also had an R squared of .12 explaining 12% of the variation in winning percentage. This led us to conclude that there might be more to this variable since its correlation changed between the regressions. Further, inspection could be necessary in order to gain a concrete view of the effect on this variable and cause us to change this to a log variable in the last regression. Looking at this practically we might suggest that total yards is positively correlated with winning percentage but when influenced by a number of other variables its effect is less important.

The third regression in the table is all of our variables in their linear form. We found that all the variables together had an R squared value of .51. This means that the variation in the five

variables explains about 51% of the variation in winning percentage. Turnovers lost, total yards, total points, and first downs were all significant at the 1% level. The only variable that was different was turnovers gained. This was significant at the 5% level, well within our parameters for the regression. Both total points and first downs both had a positive correlation on winning percentage. They also had similar effects with an increase in 1 unit of either variable, you increase your chances of winning by about 2 percentage points. It is also important to note that the constant is not statistically significant when all the variables are in their original form.

The fourth regression used an altered equation from the previous. We learned that both turnovers gained and total yards had a higher R squared when we change the variable type. Turnovers gained turned into a squared variable and total yards turned into a log variable. Used multiple regressions trying to use the squared version of the variable turnovers gained. These attempts ended up giving us disappointing results. We found that when we added the squared variable that both the original form and the squared form of turnovers gained became insignificant. To fix this we used the original form for turnovers gained and the log form for total yards. Running this new equation we found a number of interesting things. One was that the constant changed dramatically from .003 to 2.7. It also changed from not being statistically significant to being significant at the 5% level. Another interesting thing to note is that the log variable for total yards was dramatically different compared to the linear variable. Total yards changed from -.001 to -.57. This changes their effect on winning percentage dramatically. Turnovers lost, total yards, total points, and first downs were all significant at the 1% level while turnovers gained was significant at the 5% level. The new effect on total yards would suggest that for a one percent increase in yards you decrease your chances of winning by .57 percentage points. This might not look very practical but if you were to greatly increase the percentage of

yards you get it could have a big impact on the game. If you were to increase total yards by 10% you would increase your chances of winning by 5.7 percentage points. After changing the regression we expected to have a higher R squared but ended up decreasing it slightly with a new R squared of .507 compared to the previous .513.

After looking at these numbers there are a number of actions both an NFL team and a sports better could take. The first is to focus on both turnovers gained and lost. Both these had a large impact on winning percentage. In order to take advantage of this, we would advise NFL teams to invest in good defensive linemen and a defensive coordinator that can disrupt a team. This would cause an increase in turnovers gained increasing your chances of winning by six percent turnover. The second action would be to invest in a good quarterback. The quarterback handles the ball on every play and is the main contributor to whether you turn the ball over or not. A good quarterback could dramatically decrease the number of turnovers lost.

When betting you would want to bet on teams that have low turnovers lost and have a high chance of gaining multiple turnovers. The next thing we would advise is investing in a high-powered offense. Getting multiple good receivers and a running back could dramatically increase the numbers of points you score and the first downs you get. Since both of these increase your chance of winning by about two percent unit they could have a dramatic impact on a game. This is because total points has a max of 34 and first downs has a max of 47. Even with the average number of points being 20 you still increase your chances of winning by 40%. When betting on NFL Games you might want to focus on the offense capability of a team. How many of these variables do they typically get and how do they compete vs. different defenses.

$$\textbf{\textit{winperc.}} = \beta 0 + \quad -.0944(\textbf{\textit{turnovers lost}}) + \ .0601(\textbf{\textit{turnovers gained}}) \ + \ ln \ -.5714(\textbf{\textit{total yard}}) \ + \ .0272(\textbf{\textit{first downs}}) + \ .0266(\textbf{\textit{point scored}}) \ + \ \upsilon\text{^}$$

Regression Table

Regressor	1	2	3	4
Constant	.6135*** (.0909)	-.2314 (.1604)	.0032 (.1656)	2.7002** (1.0378)
Turnovers lost	-.1740*** (.0279)		-.0902*** (.0285)	-.0944*** (.0286)
Turnovers gained	.0996*** (.0308)		.0567** (.0285)	.0601** (.0282)
Total Yards		.0021*** (.0004)	-.0019*** (.0007)	
Points			.0269*** (.0065)	.0266*** (.0067)
First downs			.0285*** (.0055)	.0272*** (.0050)
Log total yards				-.5714*** (.1997)
R Squared	.29	.12	.513	.511

Conclusion

As stated previously, we expected that both turnovers gained and turnovers lost will have the largest impact on the team's chance of winning. Many of the estimators within the model are shown to be statistical significance. The estimators that we picked to put in the model are all important factors in winning an NFL football game. This could help inform Football bets, coaches, and players to make better decisions for the season to come.

Finally, we recommend that the NFL teams as a whole continue to find new and innovative ways to increase or decrease the numbers of the estimators of our model to help increase winning percentages. Teams should focus on trying to create a high-power offense. Just this alone can help a team gain a higher chance of winning. Some helpful tools to create a high-power offense could be creating new drills, better use of technology, more study on the game, and etc. These are just a few of the ways that teams in the NFL can do to better the chances of winning a Football game.

References

Legal Sports Betting. (n.d.). ***How Much Money Is Bet On Sports | How Much Is Bet On SB56*****. Legal Sports Betting. Retrieved February 22, 2022, from** **https://www.legalsportsbetting.com/how-much-money-do-americans-bet-on-sports/**

Eckstein, J., & Brown, J. R. (2022, January 27). ***How the NFL Makes Money: TV is King, Streaming and Gambling on Horizon*****. Investopedia. Retrieved February 22, 2022, from** https://www.investopedia.com/articles/personal-finance/062515/how-nfl-makes-money.asp

Pro-Football-Reference.com. 2022. ***Pro Football Statistics and History | Pro-Football-Reference.com*****. [online] Available at:** <https://www.pro-football-reference.com/> [Accessed 21 April 2022].