STAT 206 Logistic Regression Final Project

Modeling Field Goal Outcomes in the NFL Using a Logistic Regression Method

Zachary Kobban

Due: December 13th, 2023

Introduction to Topic:

Kickers can have a profound impact on the outcome of NFL games. Specifically, Field goals play a crucial role in determining the final outcome of close games. If coaches were better equipped to predict the probability of a successful field goal, they would be able to make stronger strategic decisions for example, risking going for a first down versus kicking a field goal. Having this information may also equip fans to determine if their coach is making statistically sound decisions. It is too easy to say that your coach should have made a certain decision when a play goes wrong.

The dataset used in this model is from a data analytics competition put on by the NFL in 2022. The *Big Data Bowl*[1] was a challenge to college students to create additional statistics for special teams plays and players. After searching, I found that this was one of the only free, publicly available datasets on special teams. The original dataset has ~20,000 observations from all 2018-2020 special teams plays. After removing all non-field goal plays and cleaning the dataset, I was left with ~1,300 observations. From that, I was able to do some data wrangling to create a binary success/ failure variable for the outcome. I was also able to create some interesting predictor variables to be tested later in the logistic model.

It is highly likely that in a given game, the outcome of a kicker's first kick has an effect on the outcome of their second kick. This is a problem because in order to perform a logistic regression, we assume that observations are independent of each other. Under this logic, the only kicks that will be independent of each other are the first kicks for a given field goal kicker in a given game. In order to improve independence between the observations I removed every observation that was not the first kick for a given field goal kicker in a given game. One could also imagine that a kicker's performance in one game could psychologically impact their performance in the next game. Because of this, independence is not perfect, but, in order to create a model with sufficient observations, I will proceed with the logistic model.

Variables to be Studied

Result: This is the binary response variable for this model. The original dataset uses a categorical variable specialTeamsResult to report the outcome of a field goal attempt. Kicks are reported as one of the following: Blocked Kick Attempt, Downed , Kick Attempt Good, Kick Attempt No Good, and Out of Bounds. There was only one "Out of Bounds" observation that was described in the play description variable as a touchback; I removed this from my dataset. There was only one observation of a "Downed" in specialTeamsResult. This was a fake field goal that was taken as a punt; I removed this observation from our dataset. The "Non-special teams Result" were all fake field goal attempts; I removed these 14 observations from the dataset. Finally, I reported

[1] Addison Howard, Dhriti Yandapally, Michael Lopez, Sohier Dane, Thompson Bliss. (2021). NFL Big Data Bowl 2022. Kaggle. https://kaggle.com/competitions/nfl-big-data-bowl-2022

“Blocked Kick Attempt” and “Kick Attempt No Good” as missed field goal results (0). Finally, I reported “Kick Attempt Good” as a successful field goal result (1).

kicklength: This variable is defined as the length of a given field goal attempt in yards. There were 21 observations in the dataset that were reported as having a kicklength of 0. The play description variable described kicks that were not of a length zero. These I consider to be data reporting errors and because of that, I removed them from the dataset.

yardstogo: This variable is defined as the number of yards left before a first down.

quarter: The quarter of the game in which the kick is performed. Note the 5th quarter is overtime.

down: The “down” the kick is performed in.

total_seconds: This variable is taken from the gameClock variable in the original dataset. Because this variable was in the form (MM:SS), I transformed it to a single quantitative measurement, seconds.

goalDiff_abs: This variable is the absolute value of the difference between the home and away score at the time of the kick. This variable represents how close a game is at the time of the kick.

season: This will account for whether a game is in regular (0) or postseason (1). Because there are no observations in August, there are no preseason games in this dataset.

Use Questions:

For my first use question, I want to know what the model estimates as the probability of making a specific field goal. For this, I look at a field goal attempt by Ka'imi Fairbairn on December 21, 2019. In this play, the Houston Texans kick against the Tampa Bay Buccaneers. The ball lands short and is reported as out of bounds at the Buccaneers’ 10 yard line.

For my second use question, I will report a table of OR for a change from 40 yards to 50 yards. This will tell us how the model will estimate the odds of making a field goal will change from 40 to 50 yards.

Exploratory Data Analysis (EDA)

kicklength:

This is the variable I predicted to have the largest impact on the logit model.

Result = 0	Result = 1

46.86979	37.03262

Figure 1: Mean of kicklength by Result

Just by looking at the mean kicklength by result, we can see that this term will likely be a part of the model.

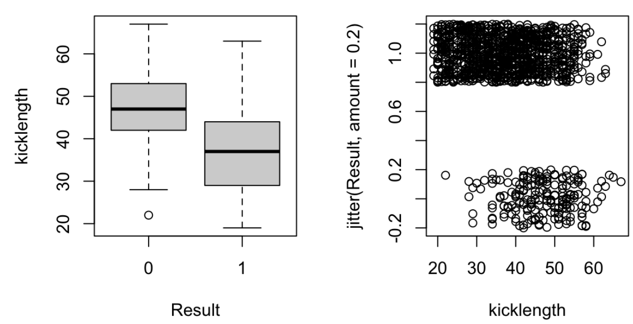

Figure 2: Plots of kicklength against Result

The Box Plot and Scatter plot representation of this relationship tells us the same story. We can expect shorter kicks to be easier to make.

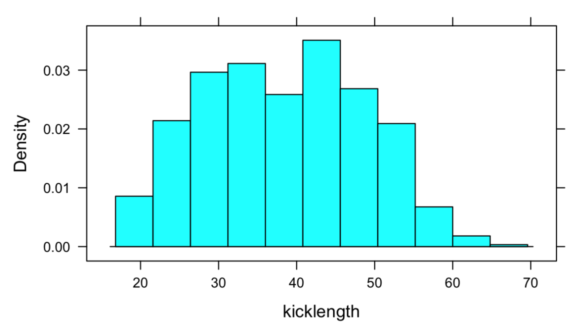

Figure 3: kicklength histogram

Interestingly, kicklength appears to be approximately normal.

yardstogo:

Result = 0	Result = 1
8.442708	7.670084

Figure #: Mean of yardstogo by Result

By looking at the favstats, there seemed to be a less stark difference between the means of yardstogo.

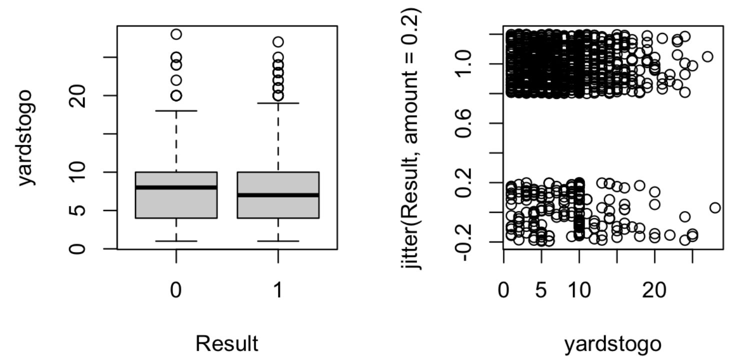

Figure 4: Plots of yardstogo against Result

quarter:

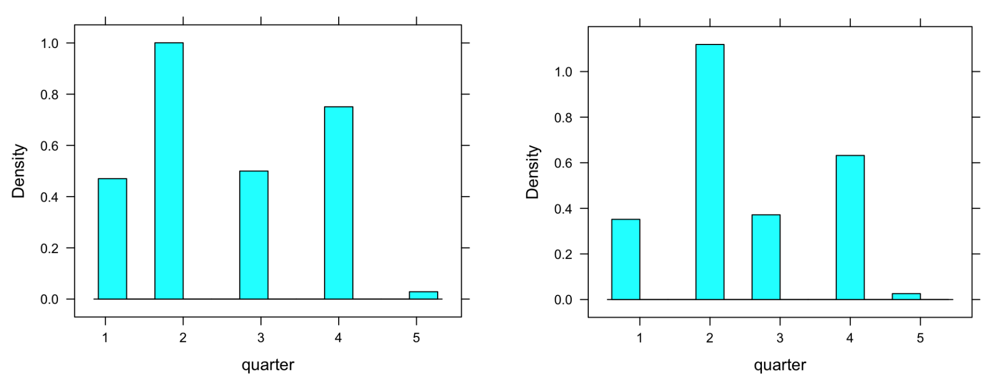

Figure 5:(left) Histogram of the frequency of each quarter before removing all observations besides a kickers fist attempt in a game
Figure 6: (right) Histogram of the frequency of each quarter after removing all observations besides a kickers fist attempt in a game

As we see on the graph on the left, the majority of field goal attempts in the NFL in 2018-2020 are in the 2nd and 4th quarters. This is still true after we remove all observations besides a kicker's fist attempt in a game. This change changes the proportions of attempted kicks in each quarter, but does not drastically impact the overall picture.

down:

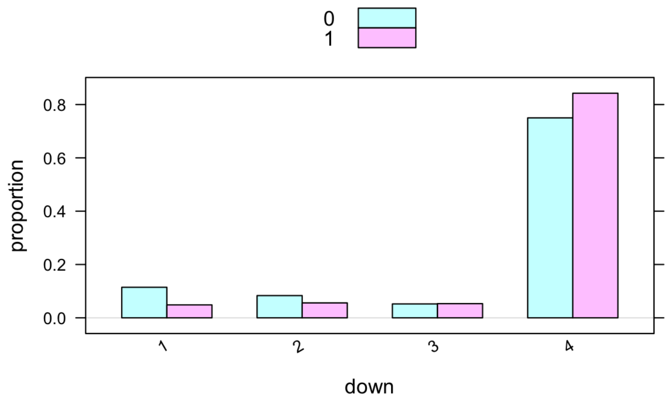

Figure 7: Bar Graph of quarter against the proportion of made field goal attempts

It is important to note the dominance of fourth down kick attempts here. Also important is the fact that the proportion of successful kicks on the fourth down is higher than the proportion of successful kicks on any other down. This may be due to the fact that kicks attempted on earlier downs are often forced. For example, if a team is running out of time before half time or full time.

total_seconds:

Result = 0	Result = 1
182.9418	220.4238

Figure 8: Mean of total_seconds by Result

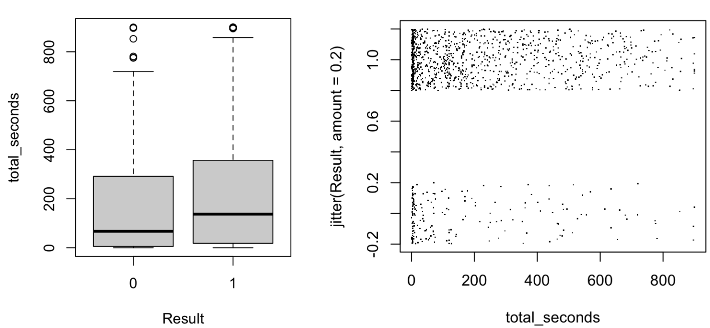

Figure 9: Plots of total_seconds against Result

The means, histograms and scatter plots here don't seem to tell us much about the predictive power of total seconds. The mean of total_seconds for made field goal attempts is slightly higher.

season:

	Result = 0	Result = 1
season = 0	191	1054
season = 1	1	19

Figure 10: Table of season and field goal results.

I encountered problems later in my analysis with the variable season. Because there is only one observation in the postseason where the field goal attempt was missed, attempting to create interaction terms with the variable season can be difficult.

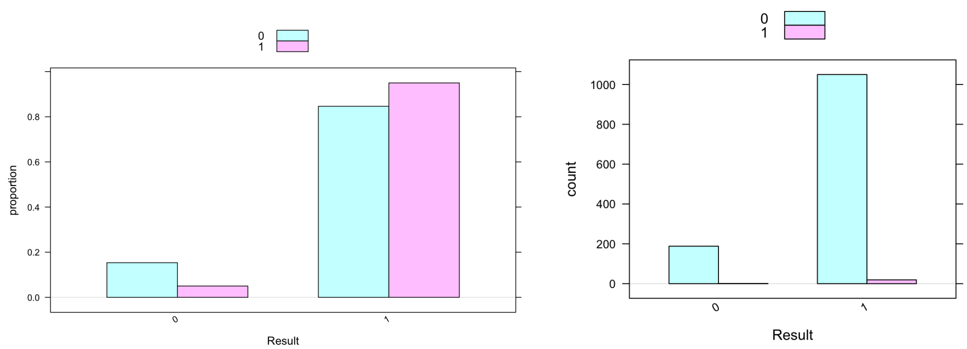

Figure 11: (left) Proportion of seasons based on the field goal attempt Result
Figure 12: (right) Count of seasons based on the field goal attempt Result

This graph shows us that there is a slightly higher probability of making a goal in the postseason. This makes sense because the teams who made it to the postseason are likely to have better kickers. Also, the number of observations in the postseason are much lower.

goalDiff_abs:

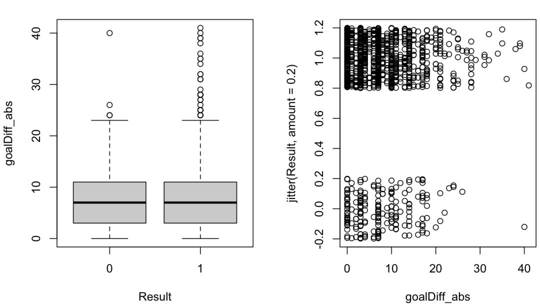

Figure 13-14:(left) Box and (right) Jitter plots for goalDiff_abs based on field goal attempt Result

Goal diff absolute value does not lead to any obvious conclusions from these preliminary graphs.

It is interesting that the made field goal attempts have more outliers. This is likely because most of the data points are made field goals. Note that the means and positions of the quantiles are almost identical.

Transformations:

kicklength:

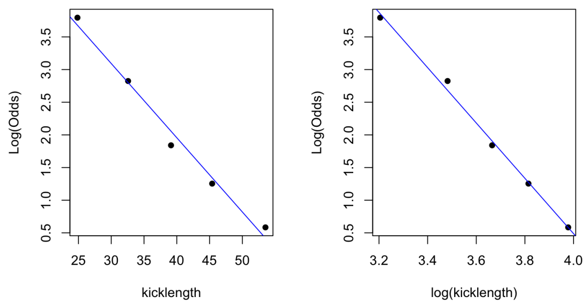

Figure 15: Transformations for kicklength

Because the log of kicklength leads to a slightly more linear relationship with the logit, I decided to apply the log transformation.

yardstogo:

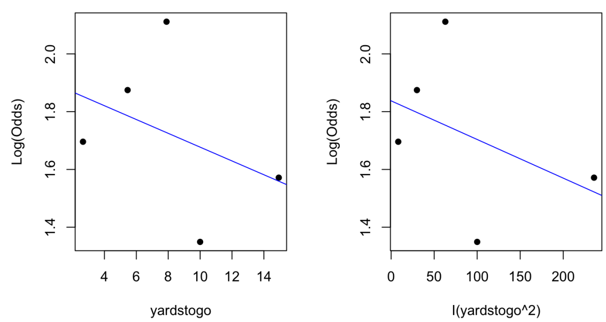

Figure 16-17: Transformations for yardstogo

The variable yardstogo resisted attempts to be made linear through transformations. Because none of the transformations were particularly strong and in order to improve the interpretation of the model, I decided that this variable did not satisfy the linearity condition and therefore needed to be removed from the model.

total_seconds:

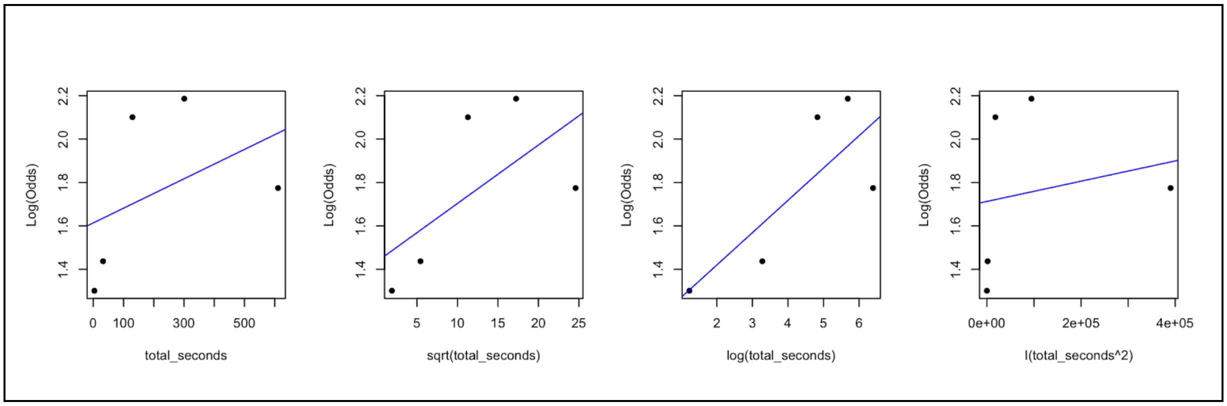

Figure 18-21: Transformations for total_seconds

For total_seconds, the log transformation was deemed the most appropriate. There were some observations for which the field goal was reportedly taken at the zeroth second of the game. These 17 observations are most likely reporting errors so I removed them from my dataset. This also avoids errors related to taking the log of zero.

goalDiff_abs:

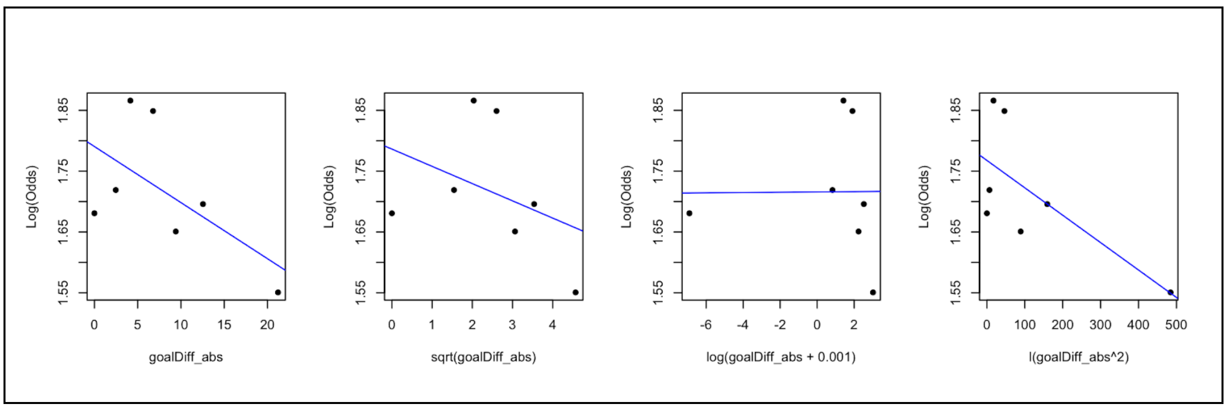

Figure 21-25:

For goalDiff_abs, I decided that not applying a transformation was best to preserve linearity and to keep the model simple to interpret.

Interactions:

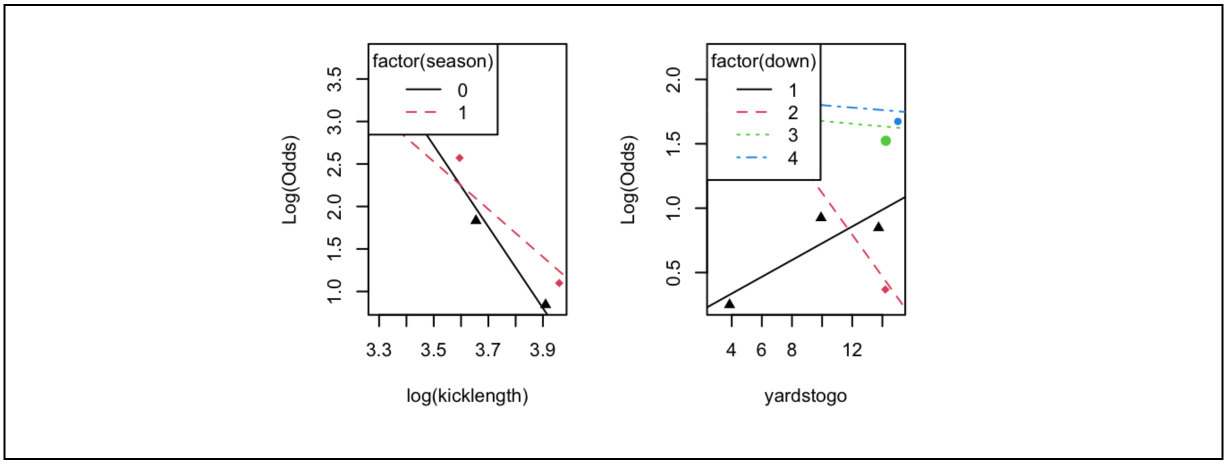

Figure 26: (left) Interaction between log(kicklength) factor(season)
Figure 27: (right) Interaction between yardstogo factor(down)

The interactions for this model were a bit of a headache. With now three linear quantitative variables and three categorical variables, there were many interactions to consider. Running all of the interaction plots showed me that every combination of categorical and quantitative variables resulted in either an interaction or an error. Because of this, I decided to run my original model with every interaction between categorical and quantitative variables.

Conditions

Linearity:
We showed in the transformations section that this condition was met by either applying the appropriate transformation or removing a non-linear variable from the model.

Independence:
Although we attempted to satisfy this independence by removing points after a given kicker's first field goal attempt in a game, I am not fully convinced that this condition is satisfied. As I mentioned in the introduction, there are other concerns for independence. For example, a kicker's performance in past games may impact their future performance. Hot or cold streaks may psychologically impact the probability of success. Overall, I don't think this condition is fully satisfied.

Randomness:
No kicker has a 100% or 0% chance of making a field goal for the range of the data.

The Logistic Regression Model

When I attempted to input all of my transformed variables and their interactions into forwards and backwards selection, I received multiple errors that were caused by having too many terms in a model without enough data points. To do a preliminary pruning of my model, I looked at the summary of my full model and decided to include every term that was significant at the 0.01 level. For example because factor(down)3:goalDiff_abs had a p-value of 0.0684, I included factor(down)*goalDiff_abs in my pruned model. This then allowed me to run the backwards, forward and "both" stepwise model selection technique on my model. From there, I found that all selection techniques resulted in a one predictor model:

$$log(odds_{Result}) = Intercept + log(kicklength)$$

In order to check that the one predictor model is sufficient, I then ran a drop in deviance. This compares the first model that we started with (the model that did not fit in our backwards/forwards selection methods) with our one predictor model. That drop in deviance test resulted in a G value of 39.42 with 33 degrees of freedom, giving us a p-value of 0.2047. Because this is above the significance level of 0.05, I fail to reject the null hypothesis. In context, this means that from a statistical perspective, there is nothing in the larger model that is not captured by the smaller model.

Final Model

The final model after running the backwards selection procedure is:

$$log(odds_{Result}) = Intercept + log(kicklength)$$

$$log(odds_{Result}) = 18.702 - 4.565(log(\widehat{kicklength}))$$

It is important to note that this model is limited in its predictive power to only a kicker's first kick in a given NFL game in the 2018-2020 seasons.

Addressing the Use questions

1. *For my first use question, I want to know what the model estimates as the probability of making a specific field goal. For this, I look at a field goal attempt by Ka'imi Fairbairn on December 21, 2019. In this play, the Houston Texans kick against the Tampa Bay Buccaneers. The ball lands short and is reported as out of bounds at the Buccaneers' 10 yard line.*

With the one predictor model, we are able to quite easily answer both use questions.

In order to predict the estimated probability that this field goal was missed, we first need to find the logit predicted by our model for a 26 yard field goal attempt. Note that the notation for log represents the natural log in this statistical context.

$log(odds_{Result}) = 18.702 - 4.565(log(\widehat{kicklength})) = 18.702 - 4.565log((26)) = 3.8287893$

$odds_{Result} = e^{3.8287893} = 46.0068041$

$probability = \frac{e^{B_0+B_1X}}{1+e^{B_0+B_1X}}\frac{46.0068041}{1+46.0068041} = 0.978726484 = 97.87\%$

This makes sense because there is only one observation in our dataset below 26 yards that did not count as a successful field goal. Let's try for a longer field goal. One of the most epic NFL final field goal moments is the "Wide Right" event from Super Bowl XXV in 1991. In the game, Buffalo Bills kicker Scott Norwood attempted a 47-yard field goal that could have won the game, but the kick went wide right by less than a yard, leading the New York Giants to win by a single point, 20-19. What does our model predict his probability of success to be?

$log(odds_{Result}) = 18.702 - 4.565(* log(\widehat{kicklength})) = 18.702 - 4.565 * log((47)) = 1.1260762$

$odds_{Result} = e^{1.1260762} = 3.08353356$

$probability = \frac{e^{B_0+B_1X}}{1+e^{B_0+B_1X}} = \frac{3.08353356}{1+ 3.08353356} = 0.755114049 = 75.51\%$

What about the longest successful field goal attempt in NFL history?

$log(odds_{Result}) = 18.702 - 4.565(log(\widehat{kicklength})) = 18.702 - 4.565log((66)) = - 0.423773897$

$odds_{Result} = e^{-0.423773897} = 0.654571866$

$probability = \frac{e^{B_0+B_1X}}{1+e^{B_0+B_1X}} = \frac{0.654571866}{1+0.654571866} = 0.395614043 = \sim40\%$

What about the longest unsuccessful field goal attempt in NFL history?

$log(odds_{Result}) = 18.702 - 4.565(log(\widehat{kicklength})) = 18.702 - 4.565log((76)) = - 1.0677977$

$odds_{Result} = e^{-1.0677977} = 0.343764758$

$probability = \frac{e^{B_0+B_1X}}{1+e^{B_0+B_1X}} = \frac{0.343764758}{1+0.343764758} = 0.255822127 = \sim25.58\%$

2. *For my second use question, I will report a table of OR for a change from 40 yards to 50 yards. This will tell us how the model will estimate the odds of making a field goal will change from 40 to 50 yards.*

$OR = e^{\Delta log(odds)} = e^{\widehat{B}_1(log(40+10)-log(40))} = 0.36108195951$

This tells us that the odds of making a field goal decrease by a factor of 0.36108195951 if we go from a 40 yard field goal to a 50 yard field goal.

Further Discussion:

When looking at the logistic regression models of other statisticians, I noticed a big difference. They had better data. Things that one could expect to have a big impact on a kicker's ability were variables that I did not have access to in my dataset. For example, variables like the weather, altitude, and if the kicker was iced, all play important roles in the models created by other statisticians.

Independence is still a concern. In order to fully assure myself that outcomes are independent, l would have to do an experiment where I assign kickers to teams, kick distances, and game situations. Doing this would remove the real game pressure element of these kicks and would mean that our model would not reflect the NFL's kicking environment. In a way, ensuring independence for this model is impossible.

Conclusion:

In the end, the model seems to be anticlimactic. After all of the work to find, clean and analyze the data, it seems odd that the result is a one predictor model. As I discussed above, this is likely because the dataset I used did not include the strongest variables to predict field goal success. That one predictor model is:

$$log(odds_{Result}) = Intercept + log(kicklength)$$

It is important to note that this model is only applicable to a given kicker's first kick in a given NFL game in the 2018-2020 seasons.

Citation of the Data's Source:

Addison Howard, Dhriti Yandapally, Michael Lopez, Sohier Dane, Thompson Bliss. (2021). NFL Big Data Bowl 2022. Kaggle. https://kaggle.com/competitions/nfl-big-data-bowl-2022