

Foul Play? How Officiating and Penalties Influence Game Outcomes in the National Football League

Jun Hanvey

ECON4950, Spring 2024

Introduction

It's January 20th, 2019. Drew Brees attempts a pass to Tommylee Lewis. There's 1:49 to go in the fourth quarter of the NFC Championship. Lewis is hit, hard, by Nickell Robey-Coleman. Robey-Coleman never turns his head to look at the ball. Three officials discuss. There is no foul on the play. The New Orleans Saints settle for a field goal. They eventually lose the game, sending the Los Angeles Rams to the Super Bowl. Drew Brees, one of the most decorated and admired quarterbacks to ever play the game goes on to retire the next season, unable to lead his team to another Super Bowl. Robey-Coleman's actions in the 2018 NFC Championship, to many, appear to be an unquestionable case of defensive pass interference. Rule 8.5.2(a) of the NFL Rulebook states that "contact by a player who is not playing the ball that restricts the opponent's opportunity to make the catch" is prohibited (Goodell 32). But what was the true effect of this highly controversial no-call?

A 2019 article published by The Ringer calls it "the worst missed call in NFL playoff history." In fact, the title of the article is "Ball Does Lie: The Saints Just Lost on the Worst Missed Call in NFL Playoff History." (Heifetz) But did the New Orleans Saints truly lose due to the missed call? What inflammatory headlines and angry tweets don't mention is that the Saints kicked a field goal after the missed call, the Rams responded with one of their own, and the game went into overtime before their eventual victory. The missed call may have been egregious, but how can it definitively be said that the Saints would've gone to the Super Bowl had the defensive pass interference been called?

I intend to explore the true influence of fouls on whether a team loses or wins any given football game. A foul in the National Football League (NFL) is "any infraction of a playing rule for which a penalty is prescribed (Goodell 5)." "Penalty" is the term attributed to the punishment for a foul. In this paper, I will use the term "penalty" to refer to fouls, as I am analyzing the influence of fouls which were called and led to a team being penalized.

A simple Google search for "nfl officiating outrage" returns 79,000 results in 0.4 seconds detailing how a team or player is outraged over a call from a certain game. The outrage is understandable, given that officiating is largely subjective. While there are clearly defined rules for what is and is not allowed in a game, officials (or referees) are only human. In real time, it can be very difficult to correctly adjudicate every play or action taken by a player. This has led to a belief that referees "take sides" or that games are "rigged." After the infamous 2019 NFC championship no-call, New Orleans Saints defensive end Cameron Jordan stated "As far as I've known, we haven't had referees that go gung-ho for the Saints. We've got to put ourselves in the best position so we're not able to be touched (Cacciola)." My goal is to analyze the available data and support the theory that individual officiating decisions do not determine the outcomes of NFL games.

Literature Review

Data

These data come from Pro Football Reference and nflpenalties.com(include citations). The dataset obtained from Pro Football Reference contains information about a team's performance in a given game. The dataset from nflpenalties.com contains information about the penalties received in a given game. After merging the two together, I then generated some variables I believed may be relevant to my analysis. The data include information on the 2022-2023 NFL regular seasons. I chose those seasons in order to hopefully mitigate variations caused by institutional changes due to the COVID-19 pandemic.

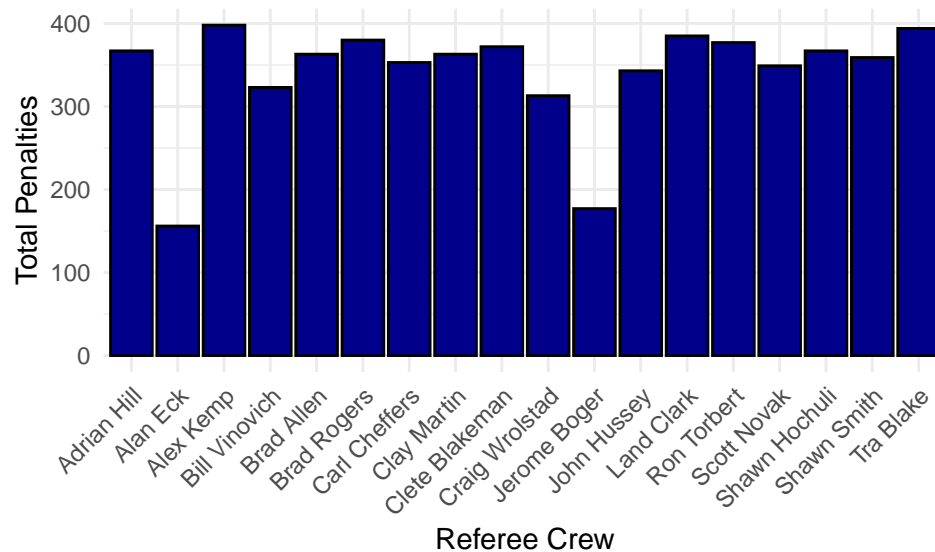
It should be noted that each game technically has two observations. I did this because my model predicts a teams' odds of winning a given game. (a binary variable) and so there had to be wins and losses in the data. So, a hypothetical game between the Dallas Cowboys and the New York Jets would have two entries: one from the Cowboys' "perspective" which lists how many penalties they received, how many points they scored, etc. and one from the Jets' "perspective" containing the same information. Here is a sample of the data:

```
## # A tibble: 6 x 19
##   ...1 date      week team opp   crew  points oYards pYards pCount   to   won
##   <dbl> <chr>    <dbl> <chr> <chr> <chr>  <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>
## 1     1 1/1/2023     17 ATL  ARI Alex~    20    298    70     5     1     1
## 2     2 1/1/2023     17 DET  CHI Adri~   41    504    42     6     0     1
## 3     3 1/1/2023     17 GB   MIN Carl~   41    315    30     3     0     1
## 4     4 1/1/2023     17 KC   DEN Tra ~    27    374    31     5     2     1
## 5     5 1/1/2023     17 LAC  LAR Clet~    31    431    19     2     0     1
## 6     6 1/1/2023     17 NE   MIA Brad~    23    249    39     6     0     1
## # i 7 more variables: tDiff <dbl>, penDiff <dbl>, ptsDiff <dbl>, ydsDiff <dbl>,
## #   pYdDiff <dbl>, home <dbl>, tie <dbl>
```

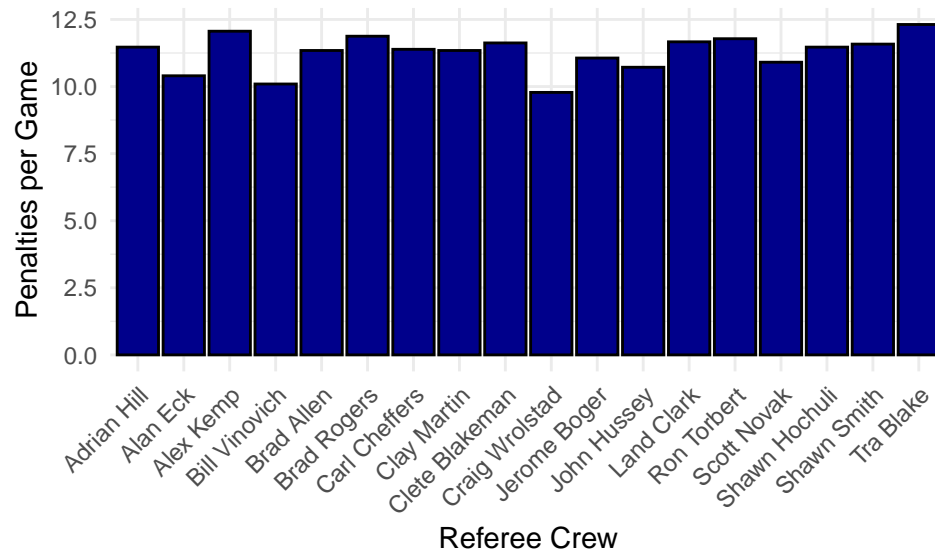
Exploratory Data Analysis

Summary Statistics for Primary Variables of Interest

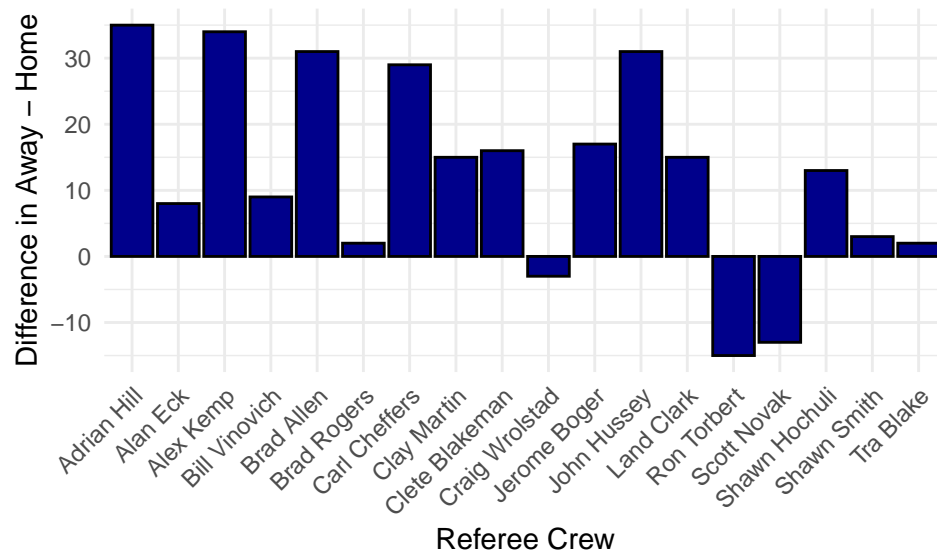
Variable	Mean	SD	Minimum	25%	Median	75%	Maximum
Points	22	9.7	0	16	21	28	70
Offensive Yards	336	82	58	281	334	391	726
Penalty Yards	47	25	0	29	44	61	151
Penalties	5.7	2.5	0	4	5	7	14
Turnovers	1.3	1.2	0	0	1	2	6
Turnover Diff.	0	1.8	-6	-1	0	1	6
Penalty Diff.	0	3.3	-10	-2	0	2	10
Point Diff.	0	14	-50	-7	0	7	50
Offensive Yards Diff.	0	119	-468	-81	0	81	468
Penalty Yard Diff.	0	31	-116	-21	0	21	116



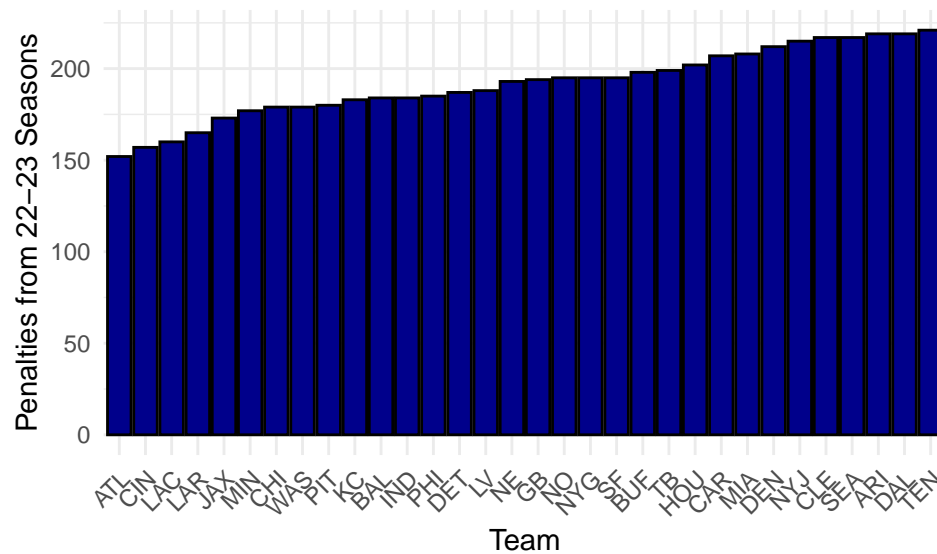
It should be noted that both Alan Eck and Jerome Boger only officiated games for one of the two seasons in my data. To account for the disparity in games between different crews, I also have plotted average penalties per game by crew.

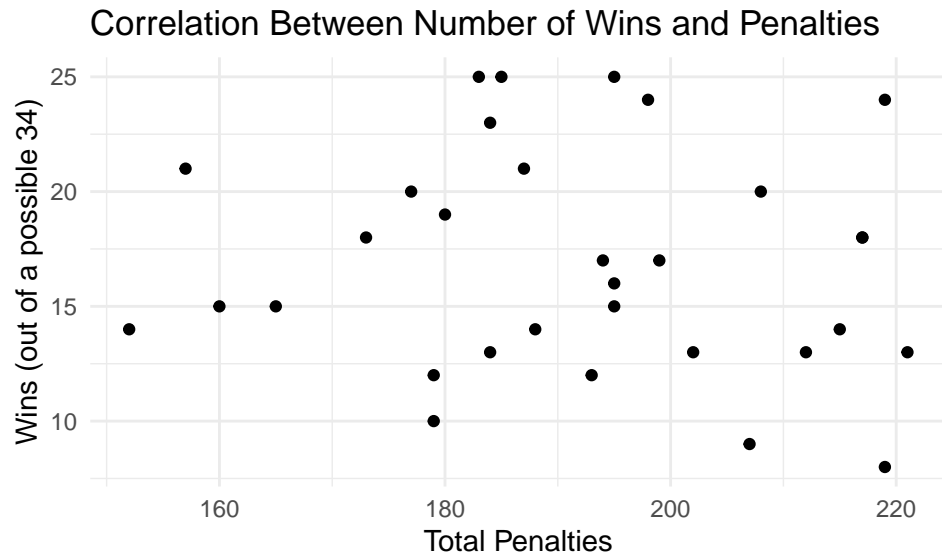
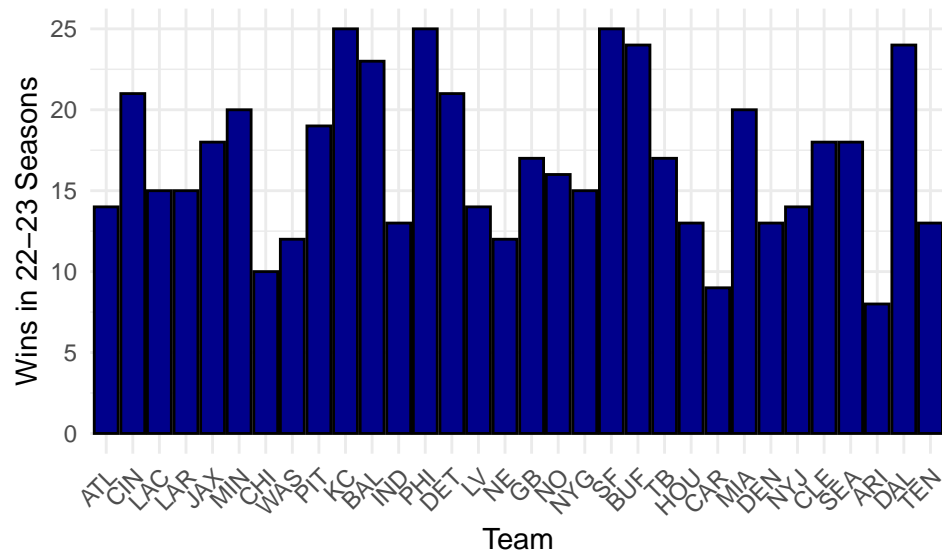


Each bar represents the difference between penalties called on the away team and penalties called on the home team over the two seasons by crew. It should be noted that many of the referees officiated roughly 30 games over the course of the two seasons.



This next graph represents total penalties over the two seasons for all 32 teams. It is arranged in ascending order. The following graph charts how many wins each team had. It is arranged in the same order. As you can see, there is a lot of fluctuation, giving some insight into the effect of penalties on winning.





Models

To run these models, I split my data into a training set and a testing set with 70% of the original data and 30% of the original data, respectively. Until I determine my final model, all models and analysis will be based on the training set only.

```
penMod <- glm(won ~ pCount + pYards + penDiff + pYdDiff, data = train,
              family = "binomial")
summary(penMod)
```

```
##
## Call:
## glm(formula = won ~ pCount + pYards + penDiff + pYdDiff, family = "binomial",
##      data = train)
```

```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4410  -1.1729  -0.9617   1.1723   1.3941
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.060951   0.220712   0.276   0.782
## pCount      -0.026716   0.077142  -0.346   0.729
## pYards       0.001803   0.007829   0.230   0.818
## penDiff      0.024930   0.056099   0.444   0.657
## pYdDiff     -0.007802   0.005684  -1.373   0.170
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1055.0  on 760  degrees of freedom
## Residual deviance: 1048.2  on 756  degrees of freedom
## AIC: 1058.2
##
## Number of Fisher Scoring iterations: 4
```

```
y = predict(penMod, train, type = "response")

y_hat0 <- ifelse(y > .50, 1, 0) |>
  factor(levels = levels(train$won))

lgCm0 <- confusionMatrix(data = y_hat0, reference = train$won, positive = "1")
lgCm0
```

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    0    1
##              0 200 183
##              1 182 196
##
##              Accuracy : 0.5204
##              95% CI : (0.4842, 0.5564)
##      No Information Rate : 0.502
##      P-Value [Acc > NIR] : 0.1639
##
##              Kappa : 0.0407
##
## Mcnemar's Test P-Value : 1.0000
##
##              Sensitivity : 0.5172
##              Specificity : 0.5236
##              Pos Pred Value : 0.5185
##              Neg Pred Value : 0.5222
##              Prevalence : 0.4980
##              Detection Rate : 0.2576
##      Detection Prevalence : 0.4967
##              Balanced Accuracy : 0.5204
##
```

```
##          'Positive' Class : 1
##
```

This first regression model estimates $won = \beta_0 + \beta_1(\text{penaltycount}) + \beta_2(\text{penaltyyards}) + \beta_3(\text{penaltydifferential}) + \beta_4(\text{penaltyyarddifferential})$. This model only predicts wins and losses with 52.04% accuracy, and none of the coefficients are significant at any level. So, not off to a great start. (insert model summaries)

```
##
## Call:
## glm(formula = won ~ oYards + points + to + tDiff + penDiff +
##      ydsDiff + pYdDiff + home, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.72638  -0.38625  -0.00798   0.36353   2.59076
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.807867   0.849851   3.304 0.000953 ***
## oYards       -0.023738   0.003389  -7.005 2.47e-12 ***
## points        0.247099   0.026938   9.173 < 2e-16 ***
## to           0.021005   0.162385   0.129 0.897077
## tDiff       -1.037467   0.137819  -7.528 5.16e-14 ***
## penDiff      -0.057014   0.062704  -0.909 0.363213
## ydsDiff       0.025607   0.002428  10.548 < 2e-16 ***
## pYdDiff      -0.013638   0.006460  -2.111 0.034771 *
## home1       -0.130621   0.246011  -0.531 0.595449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1054.96  on 760  degrees of freedom
## Residual deviance:  444.21  on 752  degrees of freedom
## AIC: 462.21
##
## Number of Fisher Scoring iterations: 7

##      oYards  points      to      tDiff  penDiff  ydsDiff  pYdDiff      home
## 4.246605 2.439732 1.975705 2.393211 2.756736 3.606326 2.673496 1.057289
```

All VIFs for this model are < 5 , which is great. We can assume that multicollinearity is not going to be an issue here. Furthermore, it seems safe to assume that the coefficients are relatively trustworthy. (insert model summaries)

```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    0    1
##              0 333  53
##              1  49 326
##
##              Accuracy : 0.866
```

```

##                95% CI : (0.8397, 0.8894)
##      No Information Rate : 0.502
##      P-Value [Acc > NIR] : <2e-16
##
##                Kappa : 0.7319
##
##      McNemar's Test P-Value : 0.7664
##
##      Sensitivity : 0.8602
##      Specificity : 0.8717
##      Pos Pred Value : 0.8693
##      Neg Pred Value : 0.8627
##      Prevalence : 0.4980
##      Detection Rate : 0.4284
##      Detection Prevalence : 0.4928
##      Balanced Accuracy : 0.8659
##
##      'Positive' Class : 1
##

```

(insert confusion matrices) The model predicts wins and losses with 87.92% accuracy, which is pretty good. I used a cutoff of 0.5 probability of winning to determine if a team should be classified as winning or losing. The only penalty-related coefficient significant at any level ($p = 0.0348$) is penalty yard differential. A 1-yard increase in penalty yard differential leads to a predicted change in win probability by a factor of $e^{-0.0137}$, or 0.9865. This equates to a roughly 1.5% reduction in winning odds. It is important to remember that the standard deviation of that variable is 31, meaning a one standard deviation increase in penalty yard differential would lead to a (follow up on this because it feels wrong.)

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```

##
## Call:
## glm(formula = hWin ~ oYards + points + to + tDiff + penDiff +
##      ydsDiff + pYdDiff + home + crew, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.47602  -0.00007   0.00000   0.00854   2.48417
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.573e+01  1.003e+03  -0.026   0.9795
## oYards        -2.600e-02  5.565e-03  -4.672  2.98e-06 ***
## points         2.628e-01  4.369e-02   6.015  1.79e-09 ***
## to             5.741e-02  2.520e-01   0.228   0.8198
## tDiff         -1.277e+00  2.179e-01  -5.859  4.65e-09 ***
## penDiff       -9.678e-02  9.944e-02  -0.973   0.3304
## ydsDiff        2.837e-02  3.915e-03   7.247  4.26e-13 ***
## pYdDiff       -1.791e-02  1.032e-02  -1.736   0.0826 .
## home1         2.727e+01  1.003e+03   0.027   0.9783
## crewAlan Eck   2.775e+00  1.677e+00   1.655   0.0979 .
## crewAlex Kemp  1.587e+00  1.244e+00   1.276   0.2019
## crewBill Vinovich 1.224e+00  1.301e+00   0.940   0.3470

```



```

## crewBrad Allen      1.564e+00  1.298e+00  1.205  0.2283
## crewBrad Rogers     2.407e+00  1.316e+00  1.829  0.0674 .
## crewCarl Cheffers   1.299e+00  1.244e+00  1.044  0.2963
## crewClay Martin     1.184e+00  1.232e+00  0.961  0.3365
## crewClete Blakeman  1.666e+00  1.240e+00  1.343  0.1791
## crewCraig Wrolstad  1.362e+00  1.304e+00  1.044  0.2963
## crewJerome Boger    -9.843e-02  1.370e+00 -0.072  0.9427
## crewJohn Hussey     3.170e+00  1.333e+00  2.378  0.0174 *
## crewLand Clark      4.643e-01  1.210e+00  0.384  0.7012
## crewRon Torbert     2.126e+00  1.302e+00  1.632  0.1026
## crewScott Novak     2.333e+00  1.319e+00  1.769  0.0769 .
## crewShawn Hochuli   2.309e+00  1.346e+00  1.715  0.0864 .
## crewShawn Smith     7.836e-01  1.253e+00  0.625  0.5318
## crewTra Blake       1.164e+00  1.220e+00  0.954  0.3401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 888.77  on 760  degrees of freedom
## Residual deviance: 201.30  on 735  degrees of freedom
## AIC: 253.3
##
## Number of Fisher Scoring iterations: 20

##              GVIF Df GVIF^(1/(2*Df))
## oYards  4.940503  1      2.222724
## points  3.039232  1      1.743339
## to      2.018388  1      1.420700
## tDiff   2.724817  1      1.650702
## penDiff 3.373398  1      1.836681
## ydsDiff 4.300991  1      2.073883
## pYdDiff 3.325411  1      1.823571
## home    1.000001  1      1.000001
## crew    1.977011 17      1.020249

## Confusion Matrix and Statistics
##
##              Reference
## Prediction  0    1
##      0 360 195
##      1  22 184
##
##              Accuracy : 0.7148
##              95% CI : (0.6813, 0.7467)
##      No Information Rate : 0.502
##      P-Value [Acc > NIR] : < 2.2e-16
##
##              Kappa : 0.4287
##
##      McNemar's Test P-Value : < 2.2e-16
##
##              Sensitivity : 0.4855
##              Specificity : 0.9424

```

```
##          Pos Pred Value : 0.8932
##          Neg Pred Value : 0.6486
##          Prevalence     : 0.4980
##          Detection Rate  : 0.2418
##          Detection Prevalence : 0.2707
##          Balanced Accuracy : 0.7139
##
##          'Positive' Class : 1
##
```

To assess potential reference bias, I created variable named “hWin” which equals 1 if the winning team was also the home team. This model assesses how a teams odds of winning when they are the home team change based on referee crew. It will include all the same variables as the previous model, with 18 dummy variables (one for each crew.)

Conclusion

Works Cited

Goodell, Roger. “OFFICIAL PLAYING RULES OF THE NATIONAL FOOTBALL LEAGUE.” 2023. 2.14.1(a)

Cacciola, Scott. “Missed Call Dooms Saints, Thrills Rams and Pains New Orleans.” The New York Times, The New York Times, 21 Jan. 2019, www.nytimes.com/2019/01/20/sports/saints-rams-call-officials.html.

Heifetz, Danny. “Ball Does Lie: The Saints Just Lost on the Worst Missed Call in NFL Playoff History.” The Ringer, The Ringer, 21 Jan. 2019, www.theringer.com/nfl/2019/1/20/18190982/saints-rams-nfc-championship-game-missed-pass-interference-nickell-robey-coleman-tommylee-lewis.

Sports Reference LLC. “2022 NFL Weekly League Schedule.” Pro-Football-Reference.com - Pro Football Statistics and History. 28 Mar. 2024, <https://www.pro-football-reference.com/>.

Sports Reference LLC. “2023 NFL Weekly League Schedule.” Pro-Football-Reference.com - Pro Football Statistics and History. 28 Mar. 2024, <https://www.pro-football-reference.com/>.