Final Project:

What Predicts the NFL Point Spread

By

Norman Tam

STAT7520

My name is Norman Tam, and I'm the only person in the group to do this project. It is a project researching what indicators or predictors can contribute to the point spread wins. Every game counts towards the point spread total, which at the end of the season should total up to 16 games, that that's a total of wins, losses, and push that add up to 16 for each team. The goal is to find a model that has the best predictors I can find that contribute the best information towards teams winning the point spread or beating the point spread as it's called in gambling circles.

On a personal note, I would like to use this model as my own and find out which predictors are the most significant both numerical and p-value wise contribute to winning the point spread from the perspective of a gambler, so I would like to test whether the odds of winning are true in most cases. The data comes from sports sites like covers.com, which gives the basic performance statistics like Yards per Game, Rushing per Game, Third Down Conversions, Time of Possession, but most importantly they have the important variable we need, which will be the dependent variable, Against the Spread Wins and Losses.

The other site that I have drawn my data from is TeamRankings.com, which has a plethora of all kinds of stats, I chose Power Rankings, which is how a team ranks in the top 32 against other ranked teams, as well as other important stats, like Turnover Margin, Average Margin of Victory in terms of Points, Red Zone Touchdowns, etc. So it is a rich dataset which are drawn from all 32 NFL teams, but also drawn over the years, from 2006 to 2022, so it is a panel data set which follows individual teams over time as well as being able to see the variation between teams over time to see how they've performed. So a Random Effects model will be used to see if there are significant random effects in the model compared to a fixed effects.

The model I have chosen is the binomial response model. I have chosen this model because specifically, it may make the most intuitive sense to model the dependent variable, the Against the Spread Wins and Losses as the trial number of successes and the number of failures as the spread wins and losses are a count total that is limited to the number of games per season. So each year we have the number of spread wins and the number of spread losses for each team that total up to 16 each and every year from 2006 to 2022. So the binomial might be the right choice for this model, at least when it comes to fixed effects models.

The dependent response variable is as mentioned, the number of point spread wins and losses per team in a given season. The predictor variables are numerous, and I will initially use them all until I use a backwards stepwise method to remove seemingly insignificant variables in the model. They are in general number of Yards per Game, which it is total passing and rushing yards, Points per Game, number of Third Down conversions, number of Defensive Sacks, QBR Rating, Power Rankings from 1-32, Turnover Margin, Red Zone efficiency or Touchdowns, and Average Number of Point Margin victories, or how far the average spread of their points is between the team and their opponents. These are seem like valuable performance stats that predicts the likely success of a team at least winning their games. So it may make sense for these stats to increase the probability of winning the point spread, though obviously not always.

The link function of choice will be eta or the log-link function that comes with the binomial distribution. So when I get the coefficient results from the estimation, I will use an exponential function to get the number from a log-odds scale to an odds ratio scale that is I'm guessing more readily interpretable.

Labeling any of the time variant terms of non-factor level terms as random effects might be another way to look at the model rather than seeing them as all fixed effects, because we would like to see not only how teams performance within their own record but compare how they do against other teams of that particular time period. We can probably do this by looking at their standard deviance differences or variation in the whole population of 32 teams and seeing how significant their statistical variance is. That way we can see if there is a contribution to the point spread within the same team group or between groups. Although, admittingly, we will have to assume that there is no correlation between the variables, which may be hard to do as at some level they are all related, at least when it comes to the team performing well.

I have ran a random effects model, and used teams as the factor level of choice for the fixed effects to focus on. Calculating the ICC, we see that we get a small number, possibly indicating that the correlation both between and within the same group has little to no correlation as the ICC is more close to zero than it is to one with ICC calculated as 0.001776149.

But before we do the random effects model, we'll start at the beginning with a very simple basic model, the linear regression model. As we run the model, we see some interesting results. To start in particular, the minimum value shows a negative number, even though ATSWin is a positive number, I'm not sure if it's counting ATSLoss to get that negative number. But it is negative regardless. The max isn't all that high either, logging in at least 2 wins maybe? Maybe it's factoring other variables as it comes up with the minimum and maximum estimate.

```
Call:
lm(formula = newats$ATSWin ~ ., data = newats)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2156 -0.4443  0.1571  0.3547  1.9949 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11.1563048  0.8780262  12.706  < 2e-16 ***
YardsG       0.0010671  0.0022546   0.473  0.63618    
RushG       -0.0023163  0.0043043  -0.538  0.59071    
RushP       -0.0042310  0.1305905  -0.032  0.97417    
PassG       -0.0056014  0.0028701  -1.952  0.05151 .  
QBR          0.0047167  0.0047508   0.993  0.32125    
Sacks        0.0039488  0.0033243   1.188  0.23542    
Third        0.0275116  0.0090678   3.034  0.00253 ** 
PossG        0.0103208  0.0249874   0.413  0.67975    
PtsG         0.0256491  0.0157441   1.629  0.10389    
PR.1.5      -0.0022752  0.0032689  -0.696  0.48674    
PR.6.10      0.0020893  0.0032537   0.642  0.52106    
PR11.16      0.0036997  0.0033206   1.114  0.26573    
PR17.22      0.0035287  0.0033248   1.061  0.28902    
PR23.32      0.0001656  0.0008711   0.190  0.84927    
ATS         -5.1240976 11.1347097  -0.460  0.64557    
ATSW         0.0804716  0.1113232   0.723  0.47009    
ATSLoss     -0.7693474  0.0294730 -26.103  < 2e-16 ***
AvgM        -0.0172497  0.0091414  -1.887  0.05972 .  
TOM          0.0706644  0.0760887   0.929  0.35347    
RZ          -0.0183174  0.0939096  -0.195  0.84543    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.6605 on 523 degrees of freedom
Multiple R-squared:  0.8787,   Adjusted R-squared:  0.874
F-statistic: 189.4 on 20 and 523 DF,  p-value: < 2.2e-16

                    2.5 %        97.5 %
(Intercept)   9.431413329  1.288120e+01
YardsG       -0.003361992  5.496240e-03
RushG        -0.010772222  6.139581e-03
RushP        -0.260777384  2.523154e-01
PassG        -0.011239658  3.690542e-05
QBR          -0.004616319  1.404981e-02
Sacks        -0.002581793  1.047944e-02
Third         0.009697868  4.532540e-02
PossG        -0.038767172  5.940887e-02
PtsG         -0.005280354  5.657854e-02
PR.1.5       -0.008696909  4.146605e-03
PR.6.10      -0.004302547  8.481221e-03
PR11.16      -0.002823762  1.022307e-02
PR17.22      -0.003002828  1.006025e-02
PR23.32      -0.001545702  1.876971e-03
ATS         -26.998348532  1.675015e+01
ATSW         -0.138224032  2.991672e-01
ATSLoss      -0.827247407 -7.114473e-01
AvgM         -0.035207950  7.086313e-04
TOM          -0.078812628  2.201414e-01
RZ           -0.202803715  1.661690e-01

Call:
lm(formula = newats$ATSWin ~ PassG + QBR + Third + PR11.16 +
    ATSW + ATSLoss + AvgM, data = newats)

Coefficients:
(Intercept)        PassG          QBR        Third      PR11.16         ATSW
ATSLoss
  11.425523    -0.003981     0.009647     0.026992     0.005349     0.030022
-0.770187
       AvgM
  -0.011071

Step:  AIC=-449.51
newats$ATSWin ~ PassG + QBR + Third + PR11.16 + ATSW + ATSLoss +
    AvgM

          Df Sum of Sq    RSS     AIC
<none>                 231.19 -449.51
- PR11.16  1      1.40 232.59 -448.23
- AvgM     1      1.54 232.73 -447.89
- QBR      1      2.67 233.86 -445.26
- Third    1      5.24 236.43 -439.32
- PassG    1      6.10 237.29 -437.35
- ATSW     1     17.28 248.47 -412.29
- ATSLoss  1    323.80 554.99   24.88

Call:
lm(formula = newats$ATSWin ~ PassG + QBR + Third + PR11.16 +
    ATSW + ATSLoss + AvgM, data = newats)

Coefficients:
(Intercept)        PassG          QBR        Third
  11.425523    -0.003981     0.009647     0.026992
    PR11.16         ATSW      ATSLoss         AvgM
   0.005349     0.030022    -0.770187    -0.011071
```

What surprises me are the statistically insignificant results for some of these performance variables as well as it's unintuitive signs. Yards per Game is positive, so the sign is in the right direction as negative yards per game doesn't necessarily make you win the point spread, whether favorite or underdog. However this positive coefficient isn't all that big, even close to zero, not to mention it is not statistically significant. The Rushing statistics are all negative which is strange and nonsensical, as well as the Passing per Game statistic, which shows a negative relationship with the point spread as well, but it it barely statistically significant. However if we add this to the 11 point spread wins intercept, it may all be positive numbers since they have such a negligible impact.

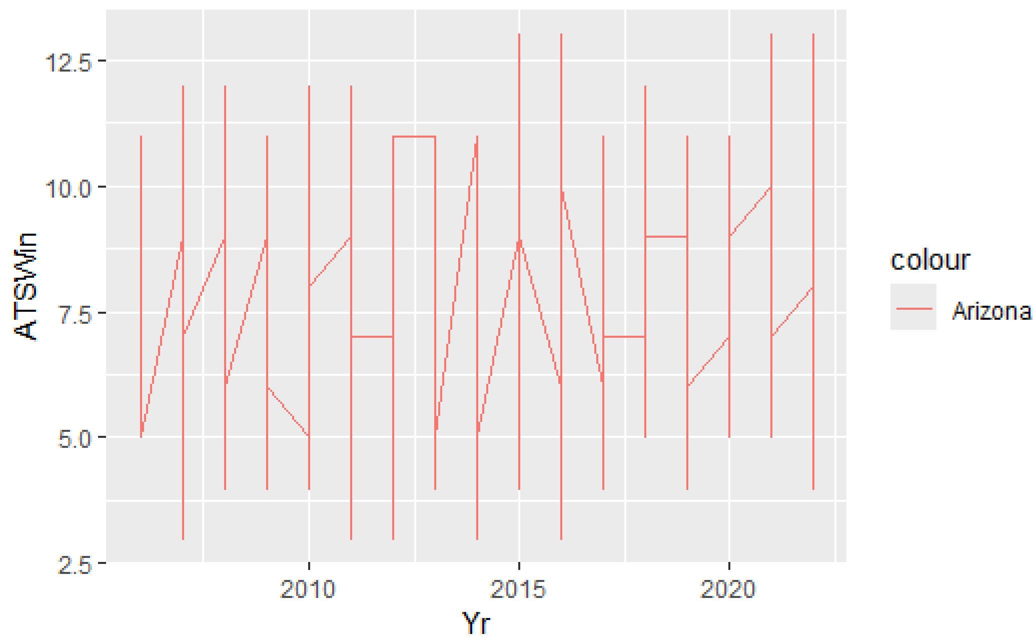
QBR and number of Sacks also don't contribute statistically to the point spread. However when we get to Third Down conversions, Possessions per Game which is in minutes per game, and Points per Game, we see better increases with the point spread even though it's only 1 to 2 %. However, only Third Down and possibly Points per Game seem statistically significant. All the Power Rankings don't seem to show any statistically significant contribution towards the point spread.

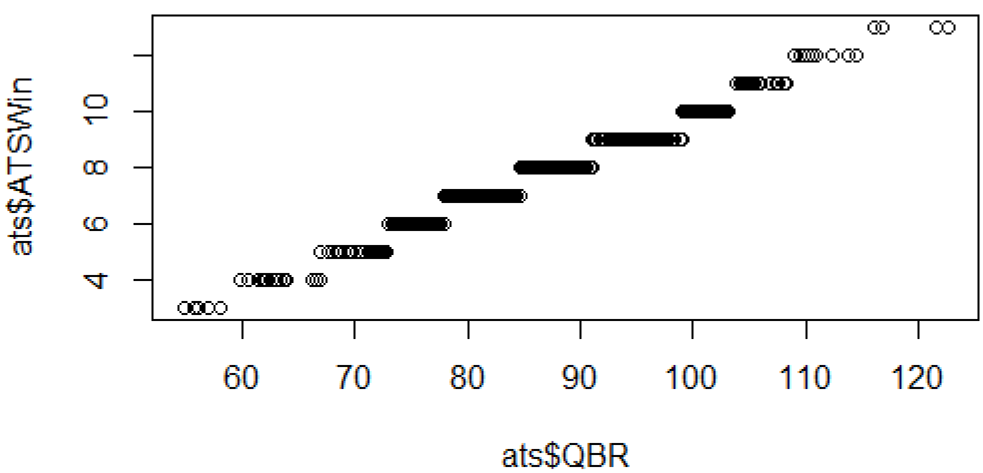
I've included ATS stats as well, and only Against the Spread Losses seem to be statistically and economically significant. So as Losses pile up against the spread one at a time, the point spread loses 7.69% of the time. I'm guessing as you lose against the spread, the losses might increase as time goes by. But we'll need a better model to confirm that. Average Margin of Victory is the only significant variable compared to Turnover Margin and Red Zone Efficiency, but it is negative which is a strange result because it is more intuitive to think that as you increase the point margin between one team and the others on average you would think that team would likely have a better chance at beating the point spread.

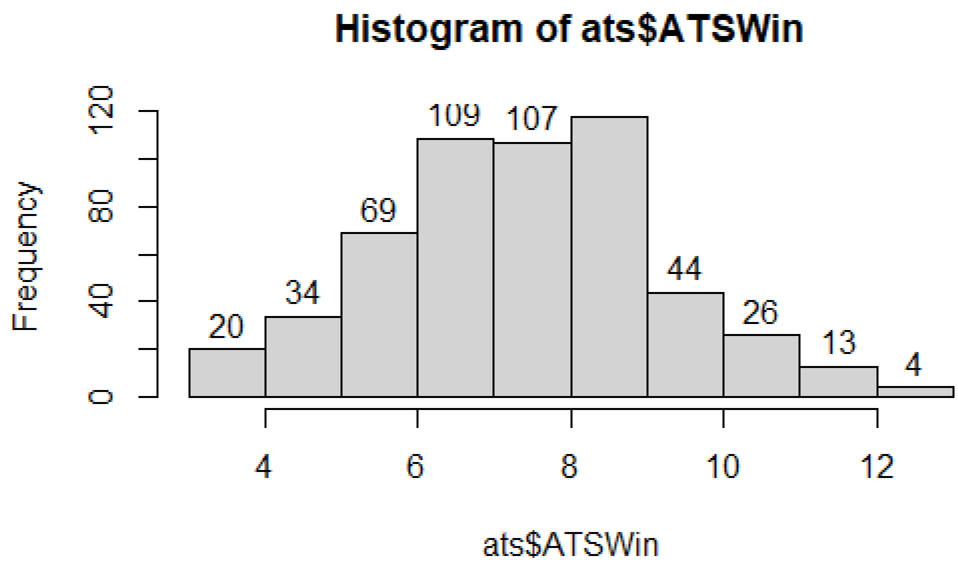
So almost most of the predictor variables are not great at predicting the point spread because of their statistical insignificance, not to mention having questionable negative relationships with the point spread. I suspect this is due to omitted variable bias, where there could be some correlated variables or proxies that could be included, but researching them will take much valuable time, so we can save that for another time.

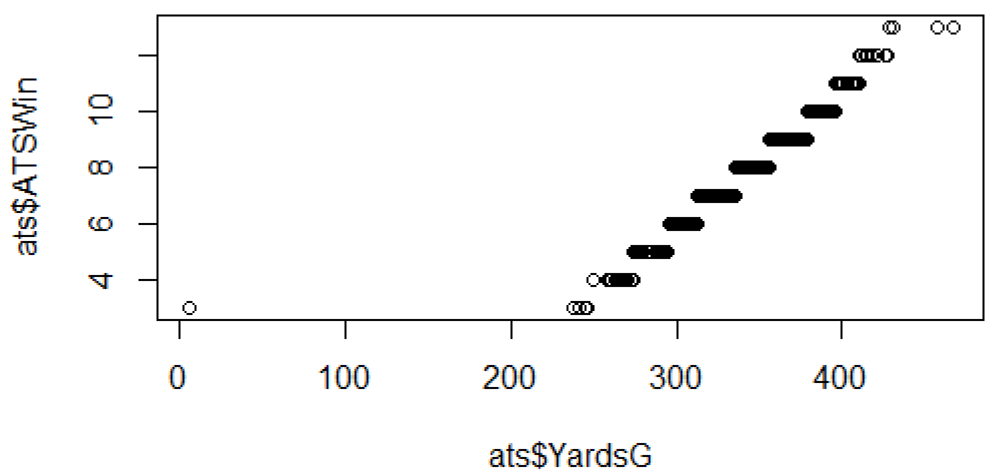
I've ran a stepwise elimination method that reduces the model down to surprisingly 7 predictor variables out of 25 or 26, some of which are factor variables. So Pass per Game, QBR, Third Down Conversions, Power Ranking 11-16, ATSW, ATSLoss, and AvgM were the ones that survived the test. The R-value is lower at 66% from around 80% but that's because of the reduction of parameters into the model. I've also ran some probably unimportant diagnostics plotting predictor variables against the dependent variable ATSWins, as well as a histogram that shows a normal distribution with 8 wins against the spread for every team as the average.

The qqplots show a step wise progression probably due to the discrete data inherent in the dependent data. Even if it is a binomial data, it deals with the number of successes and failures which we will label against the spread wins and losses respectively. Lastly, I've shown a graph from 2006 to 2022 for one team, the Arizona Cardinals and despite it being a discrete data set it shows a continuous graph which shows a somewhat up and down pattern, but really doesn't show much of a pattern.


12.5
10.0
7.5
5.0
2.5
ATSWin
2010
2015
2020
Yr
colour
Arizona


4
6
8
10
ats$ATSWin
60
70
80
90
100
110
120
ats$QBR





Before moving onto the binomial model, which is my model of choice for this project, let's recall that I decided to try to run a Poisson regression, even though the count data itself has no limits. I've ran a Poisson regression as well as run an AIC like reduction model test and ran a second Poisson regression to include in a comparison anova test of both models, and lastly run a prediction. The results show that the model is even more statistically insignificant both coefficient and p-value wise except for ATSLoss. After running a stepwise test we are left with QBR, Rushing per Attempt, Points per Game, and Passing Yards per Game. When we run an anova, we get a really low p-value suggesting that model 1 or the null model is probably not better than the fuller model.

```
Call:
glm(formula = newats$ATSWin ~ ., family = poisson(), data = newats)
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.507e+00  4.778e-01   5.248 1.54e-07 ***
YardsG       1.197e-04  1.141e-03   0.105    0.916    
RushG       -9.353e-05  2.281e-03  -0.041    0.967    
RushP       -1.100e-02  7.032e-02  -0.156    0.876    
PassG       -6.651e-04  1.489e-03  -0.447    0.655    
QBR          4.917e-04  2.561e-03   0.192    0.848    
Sacks        4.994e-04  1.817e-03   0.275    0.783    
Third        3.595e-03  4.863e-03   0.739    0.460    
PossG       -6.960e-04  1.358e-02  -0.051    0.959    
PtsG         3.624e-03  8.475e-03   0.428    0.669    
PR.1.5      -1.121e-04  1.781e-03  -0.063    0.950    
PR.6.10     -9.233e-05  1.736e-03  -0.053    0.958    
PR11.16      7.036e-04  1.815e-03   0.388    0.698    
PR17.22      1.496e-04  1.793e-03   0.083    0.933    
PR23.32      9.808e-05  4.580e-04   0.214    0.830    
ATS         -1.177e+00  6.009e+00  -0.196    0.845    
ATSW         1.547e-02  6.009e-02   0.258    0.797    
ATSLoss     -9.717e-02  1.579e-02  -6.153 7.59e-10 ***
AvgM        -2.517e-03  4.955e-03  -0.508    0.612    
TOM          1.551e-02  4.112e-02   0.377    0.706    
RZ           7.837e-03  5.108e-02   0.153    0.878    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 245.911  on 543  degrees of freedom
Residual deviance:  35.413  on 523  degrees of freedom
AIC: 2194.3

Number of Fisher Scoring iterations: 4

Warning: F test assumes 'quasipoisson' familySingle term deletions

Model:
newats$ATSWin ~ YardsG + RushG + RushP + PassG + QBR + Sacks + 
    Third + PossG + PtsG + PR.1.5 + PR.6.10 + PR11.16 + PR17.22 + 
    PR23.32 + ATS + ATSW + ATSLoss + AvgM + TOM + RZ
        Df Deviance    AIC  F value    Pr(>F)    
<none>       35.413 2194.3                       
YardsG   1   35.424 2192.3   0.1644  0.685324    
RushG    1   35.414 2192.3   0.0248  0.874832    
RushP    1   35.437 2192.3   0.3612  0.548093    
PassG    1   35.616 2192.5   3.0051  0.083592 .  
QBR      1   35.450 2192.3   0.5443  0.460984    
Sacks    1   35.488 2192.3   1.1152  0.291451    
Third    1   35.959 2192.8   8.0672  0.004683 ** 
PossG    1   35.415 2192.3   0.0388  0.843970    
PtsG     1   35.596 2192.4   2.6997  0.100966    
PR.1.5   1   35.417 2192.3   0.0585  0.808960    
PR.6.10  1   35.416 2192.3   0.0418  0.838150    
PR11.16  1   35.563 2192.4   2.2194  0.136892    
PR17.22  1   35.420 2192.3   0.1029  0.748524    
PR23.32  1   35.458 2192.3   0.6708  0.413162    
ATS      1   35.451 2192.3   0.5667  0.451926    
ATSW     1   35.479 2192.3   0.9794  0.322799    
ATSLoss  1   72.926 2229.8 554.0195 < 2.2e-16 ***
AvgM     1   35.671 2192.5   3.8068  0.051579 .  
TOM      1   35.555 2192.4   2.1009  0.147816    
RZ       1   35.436 2192.3   0.3475  0.555759    
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
glm(formula = newats$ATSWin ~ ATSW + ATSLoss, family = poisson(),
    data = newats)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.610849   0.230900  11.307  < 2e-16 ***
ATSW         0.004025   0.002467   1.631    0.103
ATSLoss     -0.098486   0.015091  -6.526 6.74e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 245.911  on 543  degrees of freedom
Residual deviance:  37.707  on 541  degrees of freedom
AIC: 2160.6

Number of Fisher Scoring iterations: 4

       1        2        3        4        5
7.570111 7.570111 9.550543 9.693834 6.589472
         1          2          3          4          5
 0.4298890  0.4298890  0.4494572  0.3061656 -0.5894723

Call:
glm(formula = newats$ATSWin ~ ats$QBR + ats$RushP + ats$PtsG +
    ats$PassG, family = poisson, data = newats)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.0740073  0.2080461   9.969  < 2e-16 ***
ats$QBR      0.0041708  0.0023256   1.793  0.07290 .
ats$RushP   -0.1115282  0.0400376  -2.786  0.00534 **
ats$PtsG     0.0329840  0.0063286   5.212 1.87e-07 ***
ats$PassG   -0.0028684  0.0006265  -4.579 4.68e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 245.91  on 543  degrees of freedom
Residual deviance: 177.93  on 539  degrees of freedom
AIC: 2304.8

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model 1: newats$ATSWin ~ ats$QBR + ats$RushP + ats$PtsG + ats$PassG
Model 2: newats$ATSWin ~ YardsG + RushG + RushP + PassG + QBR + Sacks +
    Third + PossG + PtsG + PR.1.5 + PR.6.10 + PR11.16 + PR17.22 +
    PR23.32 + ATS + ATSW + ATSLoss + AvgM + TOM + RZ
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       539    177.928
2       523     35.413 16   142.51 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Now the binomial model will obviously be different. I have combined ATSWin and ATSLoss as one dependent variable in the glm formula and specified it as a binomial family model. I run all the variables in the regression. So it also is not economically significant in the coefficients as well as the p-values, with Points per Game and Red Zone efficiency being decent candidates for the lowest p-values the binomial model has. If we were to interpret some of the more seemingly important variables, we would first calculate the odds ratio of the log odds of the intercept of -1.282, which comes to around 0.2775 odds of winning the point spread if the team were not to perform. If we were to add 0.2775 to Points per Game and Red Zone efficiency we would get:

$$0.2775 + e^{0.0238} = 1.3016$$

Odds of winning the point spread, which I'm guessing is around 30% odds probability. With Red Zone efficiency it would be lower at 1.150, which is half the odds of Points per Game. So it seems Points per Game is preferred for now, but both are important. We now run a stepwise test and we end up choosing a model with Year, Yards per Game, Sacks, Points per Game, Average Point Margin between teams, Turnover Margin, and Red Zone efficiency as the significant predictor variables. This time Yards per Game shows a negative relationship still even as we added a couple more variables and removed the Power Rankings, as well as the number of Sacks and Red Zone efficiency. Further investigation would be needed to figure out why that is the case. The latter variables Average Point Margin, Turnover Margin, and Red Zone efficiency does show larger coefficients and have low p-values, suggesting these could be the variables that make or break a point spread. And unlike the Poisson model where we reject the smaller model, this time the p-value is higher than 5% so we cautiously don't reject the null that the smaller model is zero or insignificant. When we run a confidence interval, we see that zeroes are included in variables like Yards per Game, in pretty much all of them except number of Third Downs and ATSLoss, which is strange, cause I would have hoped that the latter variables were ones to use for betting against the spread.

```
Call:
glm(formula = cbind(newats$ATSWin, newats$ATSLoss) ~ ., family = binomial,
    data = newats)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.282e+00  6.274e-01  -2.043   0.0410 *
YardsG      -9.112e-04  1.794e-03  -0.508   0.6115
RushG        1.297e-03  3.366e-03   0.385   0.7000
RushP       -6.898e-02  1.016e-01  -0.679   0.4973
PassG       -7.015e-04  2.267e-03  -0.309   0.7570
QBR          1.493e-03  3.713e-03   0.402   0.6877
Sacks       -1.260e-03  2.598e-03  -0.485   0.6278
Third        2.921e-03  7.089e-03   0.412   0.6802
PossG       -5.103e-03  1.937e-02  -0.263   0.7922
PtsG         2.038e-02  1.216e-02   1.676   0.0938 .
PR.1.5       1.584e-03  2.553e-03   0.621   0.5348
PR.6.10      1.398e-03  2.534e-03   0.552   0.5812
PR11.16      5.681e-06  2.594e-03   0.002   0.9983
PR17.22      5.658e-05  2.593e-03   0.022   0.9826
PR23.32      9.705e-05  6.681e-04   0.145   0.8845
ATS          1.700e+00  8.709e+00   0.195   0.8452
ATSW         1.580e-02  8.712e-02   0.181   0.8561
AvgM         6.082e-03  7.063e-03   0.861   0.3892
TOM         -2.913e-02  5.933e-02  -0.491   0.6234
```

```
RZ          -1.352e-01  7.192e-02  -1.879   0.0602 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 470.38  on 543  degrees of freedom
Residual deviance: 121.11  on 524  degrees of freedom
AIC: 1891.6

Number of Fisher Scoring iterations: 3

Warning: NaNs producedWarning: NaNs produced          predout
ats$ATSWin  no yes
        3    5   0
        4   13   2
        5   33   1
        6   61   8
        7   98  11
        8   46  61
        9    7 111
        10   4  40
        11   0  26
        12   2  11
        13   0   4

  3   4   5   6   7   8   9  10  11  12  13 
  5  15  34  69 109 107 118  44  26  13   4 
        1         2         3         4         5         6 
0.5204461 0.4981550 0.4962121 0.5285171 0.4944238 0.4620253 
        7         8         9        10        11        12 
0.5413534 0.4603774 0.5074074 0.4832714 0.4528302 0.5526316 
       13        14        15        16        17        18 
0.4566038 0.5338346 0.4774436 0.5185185 0.4944238 0.4851852 
       19        20        21        22        23        24 
0.4572491 0.5000000 0.5225564 0.5168539 0.4681648 0.5730337 
       25        26        27        28        29        30 
0.5477941 0.4901961 0.5050167 0.5214286 0.5112782 0.4681648 
       31        32        33        34        35        36 
0.4760000 0.4796748 0.5204461 0.4981550 0.4962121 0.5285171 
       37        38        39        40        41        42 
0.4944238 0.4620253 0.5413534 0.4603774 0.5074074 0.4832714 
       43        44        45        46        47        48 
0.4528302 0.5526316 0.4566038 0.5338346 0.4774436 0.5185185 
       49        50        51        52        53        54 
0.4944238 0.4851852 0.4572491 0.5000000 0.5225564 0.5168539 
       55        56        57        58        59        60 
0.4681648 0.5730337 0.5477941 0.4901961 0.5050167 0.5214286 
       61        62        63        64        65        66 
0.5112782 0.4681648 0.4760000 0.4796748 0.5204461 0.4981550 
       67        68        69        70        71        72 
0.4962121 0.5285171 0.4944238 0.4620253 0.5413534 0.4603774 
       73        74        75        76        77        78 
0.5074074 0.4832714 0.4528302 0.5526316 0.4566038 0.5338346 
       79        80        81        82        83        84 
0.4774436 0.5185185 0.4944238 0.4851852 0.4572491 0.5000000 
       85        86        87        88        89        90 
0.5225564 0.5168539 0.4681648 0.5730337 0.5477941 0.4901961 
       91        92        93        94        95        96 
0.5214286 0.5214286 0.5112782 0.4681648 0.4760000 0.4796748 
       97        98        99       100       101       102 
0.5204461 0.4981550 0.4962121 0.5285171 0.4944238 0.4620253 
      103       104       105       106       107       108 
0.5413534 0.4603774 0.5074074 0.4832714 0.4528302 0.5526316 
```

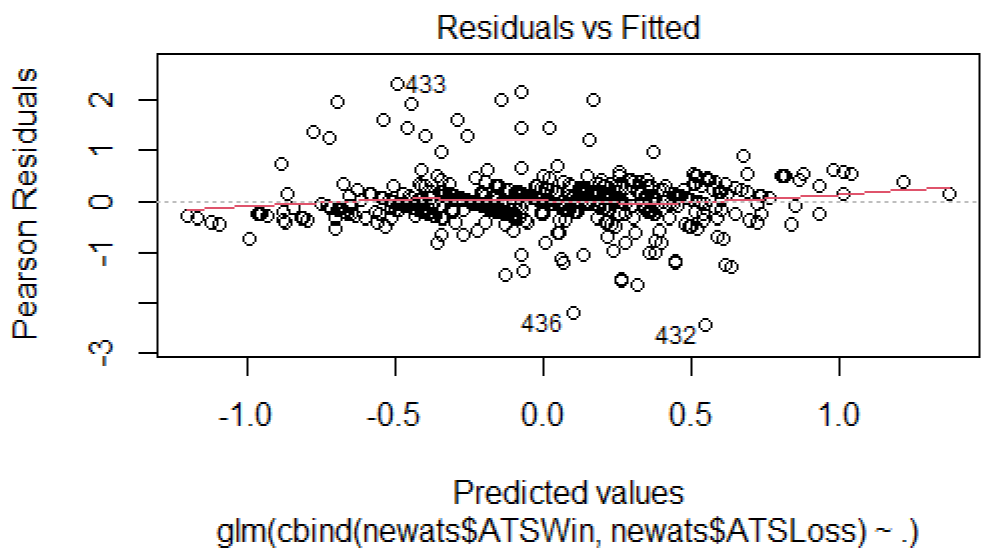
```
      109       110       111       112       113       114 
0.4566038 0.5338346 0.4774436 0.5185185 0.4944238 0.4851852 
      115       116       117       118       119       120 
0.4572491 0.5000000 0.5225564 0.5168539 0.4681648 0.5730337 
      121       122       123       124       125       126 
0.5477941 0.4901961 0.5050167 0.5214286 0.5112782 0.4681648 
      127       128       129       130       131       132 
0.4760000 0.4796748 0.5204461 0.4981550 0.4962121 0.5285171 
      133       134       135       136       137       138 
0.4944238 0.4620253 0.5413534 0.4603774 0.5074074 0.4832714 
      139       140       141       142       143       144 
0.4528302 0.5526316 0.4566038 0.5338346 0.4774436 0.5185185 
      145       146       147       148       149       150 
0.4944238 0.4851852 0.4572491 0.5000000 0.5225564 0.5168539 
      151       152       153       154       155       156 
0.4681648 0.5730337 0.5477941 0.4901961 0.5050167 0.5214286 
      157       158       159       160       161       162 
0.5112782 0.4681648 0.4760000 0.4796748 0.5204461 0.4981550 
      163       164       165       166       167       168 
0.4962121 0.5285171 0.4944238 0.4620253 0.5413534 0.4603774 
      169       170       171       172       173       174 
0.5074074 0.4832714 0.4528302 0.5526316 0.4566038 0.5338346 
      175       176       177       178       179       180 
0.4774436 0.5185185 0.4944238 0.4851852 0.4572491 0.5000000 
      181       182       183       184       185       186 
0.5225564 0.5168539 0.4681648 0.5730337 0.5477941 0.4901961 
      187       188       189       190       191       192 
0.5050167 0.5214286 0.5112782 0.4681648 0.4760000 0.4796748 
      193       194       195       196       197       198 
0.5204461 0.4981550 0.4962121 0.5285171 0.4944238 0.4620253 
      199       200       201       202       203       204 
0.5413534 0.4603774 0.5074074 0.4832714 0.4528302 0.5526316 
      205       206       207       208       209       210 
0.4566038 0.5338346 0.4774436 0.5185185 0.4944238 0.4851852 
      211       212       213       214       215       216 
0.4572491 0.5000000 0.5225564 0.5168539 0.4681648 0.5730337 
      217       218       219       220       221       222 
0.5477941 0.4901961 0.5050167 0.5214286 0.5112782 0.4681648 
      223       224       225       226       227       228 
0.4760000 0.4796748 0.5204461 0.4981550 0.4962121 0.5285171 
      229       230       231       232       233       234 
0.4944238 0.4620253 0.5413534 0.4603774 0.5074074 0.4832714 
      235       236       237       238       239       240 
0.4528302 0.5526316 0.4566038 0.5338346 0.4774436 0.5185185 
      241       242       243       244       245       246 
0.4944238 0.4851852 0.4572491 0.5000000 0.5225564 0.5168539 
      247       248       249       250       251       252 
0.4681648 0.5730337 0.5477941 0.4901961 0.5050167 0.5214286 
      253       254       255       256       257       258 
0.5112782 0.4681648 0.4760000 0.4796748 0.5204461 0.4981550 
      259       260       261       262       263       264 
0.4962121 0.5285171 0.4944238 0.4620253 0.5413534 0.4603774 
      265       266       267       268       269       270 
0.5074074 0.4832714 0.4528302 0.5526316 0.4566038 0.5338346 
      271       272       273       274       275       276 
0.4774436 0.5185185 0.4944238 0.4851852 0.4572491 0.5000000 
      277       278       279       280       281       282 
0.5225564 0.5168539 0.4681648 0.5730337 0.5477941 0.4901961 
      283       284       285       286       287       288 
0.5050167 0.5214286 0.5112782 0.4681648 0.4760000 0.4796748 
      289       290       291       292       293       294 
0.5204461 0.4981550 0.4962121 0.5285171 0.4944238 0.4620253 
      295       296       297       298       299       300 
0.5413534 0.4603774 0.5074074 0.4832714 0.4528302 0.5526316 
```

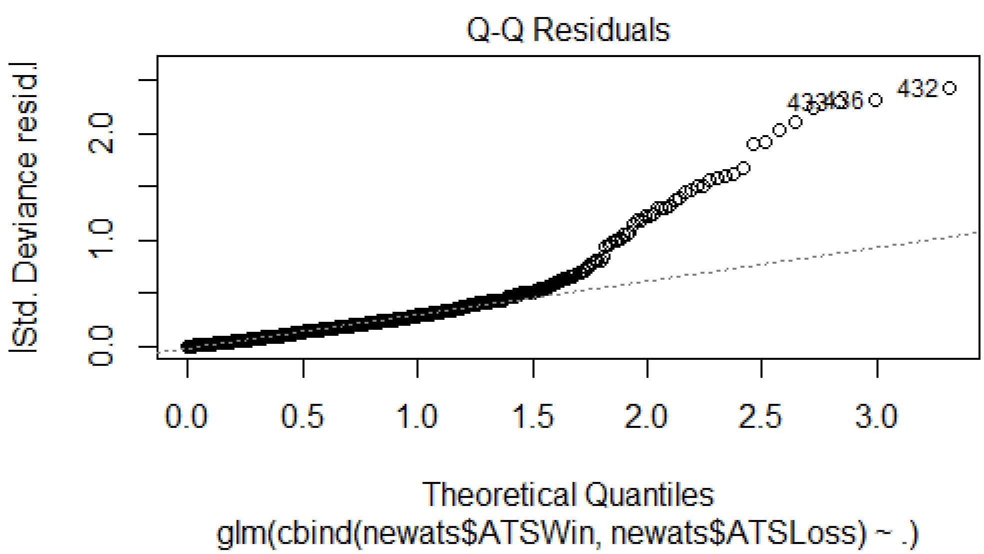
```
      301       302       303       304       305       306 
0.4566038 0.5338346 0.4774436 0.5185185 0.4944238 0.4851852 
      307       308       309       310       311       312 
0.4572491 0.5000000 0.5225564 0.5168539 0.4681648 0.5730337 
      313       314       315       316       317       318 
0.5477941 0.4901961 0.5050167 0.5214286 0.5112782 0.4681648 
      319       320       321       322       323       324 
0.4760000 0.4796748 0.5204461 0.4981550 0.4962121 0.5285171 
      325       326       327       328       329       330 
0.4944238 0.4620253 0.5413534 0.4603774 0.5074074 0.4832714 
      331       332       333       334       335       336 
0.4528302 0.5526316 0.4566038 0.5338346 0.4774436 0.5185185 
      337       338       339       340       341       342 
0.4944238 0.4851852 0.4572491 0.5000000 0.5225564 0.5168539 
      343       344       345       346       347       348 
0.4681648 0.5730337 0.5477941 0.4901961 0.5050167 0.5214286 
      349       350       351       352       353       354 
0.5112782 0.4681648 0.4760000 0.4796748 0.5204461 0.4981550 
      355       356       357       358       359       360 
0.4962121 0.5285171 0.4944238 0.4620253 0.5413534 0.4603774 
      361       362       363       364       365       366 
0.5074074 0.4832714 0.4528302 0.5526316 0.4566038 0.5338346 
      367       368       369       370       371       372 
0.4774436 0.5185185 0.4944238 0.4851852 0.4572491 0.5000000 
      373       374       375       376       377       378 
0.5225564 0.5168539 0.4681648 0.5730337 0.5477941 0.4901961 
      379       380       381       382       383       384 
0.5050167 0.5050167 0.5214286 0.5112782 0.4681648 0.4760000 
      385       386       387       388       389       390 
0.5204461 0.4981550 0.4962121 0.5285171 0.4944238 0.4620253 
      391       392       393       394       395       396 
0.4620253 0.5413534 0.4603774 0.5074074 0.4832714 0.4528302 
      397       398       399       400       401       402 
0.5526316 0.4566038 0.5338346 0.4774436 0.5185185 0.4944238 
      403       404       405       406       407       408 
0.4851852 0.4572491 0.5225564 0.5168539 0.4681648 0.5730337 
      409       410       411       412       413       414 
0.5477941 0.4901961 0.5050167 0.5050167 0.5214286 0.5112782 
      415       416       417       418       419       420 
0.4681648 0.4796748 0.5204461 0.4981550 0.4962121 0.5285171 
      421       422       423       424       425       426 
0.4944238 0.4620253 0.4620253 0.5413534 0.4603774 0.5074074 
      427       428       429       430       431       432 
0.4832714 0.4528302 0.5526316 0.4566038 0.5338346 0.4774436 
      433       434       435       436       437       438 
0.5185185 0.4944238 0.4851852 0.4572491 0.5225564 0.5168539 
      439       440       441       442       443       444 
0.4681648 0.5730337 0.5477941 0.4901961 0.5050167 0.5214286 
      445       446       447       448       449       450 
0.5112782 0.4681648 0.4760000 0.4796748 0.5204461 0.4981550 
      451       452       453       454       455       456 
0.4962121 0.5285171 0.4944238 0.4620253 0.4620253 0.5413534 
      457       458       459       460       461       462 
0.4603774 0.5074074 0.4832714 0.4528302 0.5526316 0.4566038 
      463       464       465       466       467       468 
0.5338346 0.4774436 0.5185185 0.4944238 0.4851852 0.4572491 
      469       470       471       472       473       474 
0.5225564 0.5168539 0.4681648 0.5730337 0.5477941 0.5050167 
      475       476       477       478       479       480 
0.5050167 0.5214286 0.5112782 0.4681648 0.4760000 0.4796748 
      481       482       483       484       485       486 
0.5204461 0.4981550 0.4962121 0.5285171 0.4944238 0.4620253 
      487       488       489       490       491       492 
0.5413534 0.4603774 0.5074074 0.4832714 0.4528302 0.5526316 
```

```
      493       494       495       496       497       498 
0.4566038 0.5338346 0.4774436 0.5185185 0.4944238 0.4851852 
      499       500       501       502       503       504 
0.4572491 0.5000000 0.5225564 0.5168539 0.4681648 0.5730337 
      505       506       507       508       509       510 
0.5477941 0.4901961 0.5050167 0.5214286 0.5112782 0.4681648 
      511       512       513       514       515       516 
0.4760000 0.4796748 0.5204461 0.4981550 0.4962121 0.5285171 
      517       518       519       520       521       522 
0.4944238 0.4620253 0.5413534 0.4603774 0.5074074 0.4832714 
      523       524       525       526       527       528 
0.4528302 0.5526316 0.4566038 0.5338346 0.4774436 0.5185185 
      529       530       531       532       533       534 
0.4944238 0.4851852 0.4572491 0.5000000 0.5225564 0.5168539 
      535       536       537       538       539       540 
0.4681648 0.5730337 0.5477941 0.4901961 0.5050167 0.5214286 
      541       542       543       544 
0.5112782 0.4681648 0.4760000 0.4796748 
        1         2         3         4         5 
0.4906713 0.4956002 0.6138487 0.6305149 0.4136873 
        1         2         3         4         5 
0.9633688 0.9825543 1.5896584 1.7064689 0.7055744 
ats$TOM
-1.8 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1   -1 -0.9 -0.8 -0.7 -0.6 
   4    5    8   10   22    8   56   25   93  170   78  144 
-0.5 -0.4 -0.3 -0.2 -0.1    0  0.1  0.2  0.3  0.4  0.5  0.6 
 115  191  335  275  365  200  417  234  287  281  210  233 
 0.7  0.8  0.9    1  1.1  1.2  1.3  1.4  1.5  1.6  1.7 
 125  147  114   10    8   27   36   11   11    7   12 
Warning: F test assumes 'quasibinomial' familySingle term deletions

Model:
cbind(newats$ATSWin, newats$ATSLoss) ~ YardsG + RushG + RushP + 
    PassG + QBR + Sacks + Third + PossG + PtsG + PR.1.5 + PR.6.10 + 
    PR11.16 + PR17.22 + PR23.32 + ATS + ATSW + AvgM + TOM + RZ
        Df Deviance    AIC F value    Pr(>F)    
<none>       121.11 1891.6                      
YardsG   1   121.38 1889.8  1.1318 0.2878785    
RushG    1   121.26 1889.7  0.6437 0.4227281    
RushP    1   121.58 1890.0  1.9934 0.1585791    
PassG    1   121.21 1889.7  0.4121 0.5212060    
QBR      1   121.28 1889.7  0.6991 0.4034752    
Sacks    1   121.35 1889.8  1.0170 0.3137030    
Third    1   121.28 1889.7  0.7349 0.3916851    
PossG    1   121.18 1889.6  0.3002 0.5839691    
PtsG     1   123.92 1892.4 12.1538 0.0005310 ***
PR.1.5   1   121.50 1889.9  1.6677 0.1971366    
PR.6.10  1   121.42 1889.9  1.3168 0.2516976    
PR11.16  1   121.11 1889.6  0.0000 0.9963681    
PR17.22  1   121.11 1889.6  0.0021 0.9638185    
PR23.32  1   121.14 1889.6  0.0913 0.7625916    
ATS      1   121.15 1889.6  0.1649 0.6848482    
ATSW     1   121.15 1889.6  0.1423 0.7061729    
AvgM     1   121.86 1890.3  3.2082 0.0738482 .  
TOM      1   121.36 1889.8  1.0430 0.3076005    
RZ       1   124.65 1893.1 15.2945 0.0001041 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
           1            2            3            4 
 0.009328657  0.004399802  0.011151312 -0.005514872 
           5 
-0.013687251 
Analysis of Deviance Table
```

```
Model 1: cbind(newats$ATSWin, newats$ATSLoss) ~ PtsG + ATS + RZ
Model 2: cbind(newats$ATSWin, newats$ATSLoss) ~ YardsG + RushG + RushP +
    PassG + QBR + Sacks + Third + PossG + PtsG + PR.1.5 + PR.6.10 +
    PR11.16 + PR17.22 + PR23.32 + ATS + ATSW + AvgM + TOM + RZ
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       540     127.03
2       524     121.11 16   5.9202   0.9889

Call:
glm(formula = cbind(ats$ATSWin, ats$ATSLoss) ~ Yr + YardsG +
    QBR + Sacks + Third + PtsG + AvgM + TOM + RZ, family = binomial,
    data = ats)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.107e+01  1.016e+01  -1.089  0.27596
Yr           5.722e-03  5.099e-03   1.122  0.26183
YardsG      -3.043e-03  1.018e-03  -2.990  0.00279 **
QBR          2.847e-03  3.638e-03   0.783  0.43389
Sacks       -5.456e-03  2.513e-03  -2.172  0.02989 *
Third        9.872e-04  6.718e-03   0.147  0.88318
PtsG         4.866e-02  1.134e-02   4.293 1.76e-05 ***
AvgM         2.848e-02  6.714e-03   4.242 2.21e-05 ***
TOM          1.086e-01  5.685e-02   1.911  0.05605 .
RZ          -3.374e-01  7.171e-02  -4.705 2.53e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 470.38  on 543  degrees of freedom
Residual deviance: 303.19  on 534  degrees of freedom
AIC: 2053.6

Number of Fisher Scoring iterations: 3


Call:
glm(formula = cbind(ats$ATSWin, ats$ATSLoss) ~ Yr + YardsG +
    Sacks + PtsG + AvgM + TOM + RZ, family = binomial, data = ats)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.354e+01  9.653e+00  -1.402   0.1608
Yr           6.992e-03  4.838e-03   1.445   0.1484
YardsG      -2.818e-03  9.648e-04  -2.920   0.0035 **
Sacks       -5.303e-03  2.427e-03  -2.185   0.0289 *
PtsG         5.367e-02  9.710e-03   5.527 3.25e-08 ***
AvgM         2.849e-02  6.683e-03   4.263 2.02e-05 ***
TOM          1.152e-01  5.483e-02   2.101   0.0357 *
RZ          -3.399e-01  7.126e-02  -4.770 1.84e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 470.38  on 543  degrees of freedom
Residual deviance: 303.92  on 536  degrees of freedom
AIC: 2050.4

Number of Fisher Scoring iterations: 3

Analysis of Deviance Table
```

```
Model 1: cbind(ats$ATSWin, ats$ATSLoss) ~ Yr + YardsG + QBR + Sacks + 
    Third + PtsG + AvgM + TOM + RZ
Model 2: cbind(ats$ATSWin, ats$ATSLoss) ~ Yr + YardsG + Sacks + PtsG + 
    AvgM + TOM + RZ
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       534     303.19                     
2       536     303.93 -2  -0.7387   0.6912
```





As you can see, from tabulating the number of point spread wins, we see that 9 wins is the highest average for all teams for all seasons. Trying to use the model as a prediction in modl4, we see that the average probability for all 544 observations or for all teams is between .45 to .55 so somewhere in the middle of 0 and 1 in probability that teams are likely to do against the point spread. So around 50% of the time they will do well against the spread.

Lastly, I've ran a random effects model on the dataset, using team as the random intercept for repeated measures. We fit the random effects model using the predictor variables that were significant from the binomial model. So, Passing Yards per Game, Points per Game, Average Point Margin, Turnover Margin, and Red Zone efficiency. Looking at the between Teams variance and standard deviation, we see that surprisingly there is less variation between teams than within, which is pretty close to zero for the Random Effects. But when we look at the residual variance and standard deviance within each Team, although it is more than the Random Effects, it is still small. It looks like the Random Effects model does a better job at explaining the dataset than the Fixed Effects. As was mentioned, the ICC is real low, so there is no evidence of variation within teams or correlations within the whole population especially in the response variable.

So in essence, there is no correlation and more independence between seasons for the point spread. The fixed effects of the model is kind of the same as linear model except with different standard deviations and in turn different confidence intervals. At 0 seasons, we see that the average point spread win is 7 wins, which doesn't make intuitive sense. But when we add the exponential odds ratio you see decreases with Passing Yard per Game and Red Zone efficiency and positive gains in point spread with Points per Game, Average Point Margin and Turnover Margin. Again, to change the signs of the negative coefficients we will need to find its proxy variables that's correlated pending further investigation.

We fit another random effects model, but include Year this time as a possible important variable and keep that fixed for the slope, and have Team as the Random Effect intercepts. We see that there is a strong negative correlation between Year and Team which is pretty much -1, and it shows a singular error. Looking at the variance, we see that between Teams variance is quite low, and even lower for Year. When we run an anova test, we see that we can reject the model that the fixed effect model is more important with a p-value of 0.006953. So, it may make sense that we have rejected the null that the fixed effects model is more important as there seems to be correlations between variables in this model.

So all in all, despite having a low ICC, which may indicate that the random effects are strong, I would say it makes intuitive sense that a random effects model be used rather than a purely fixed effects model, even if it doesn't make sense to say that teams that win the point spread doesn't perform well using the predictor variable metrics especially when turning the ball over multiple times per game, having low time or possession or less yards or points per game, etc. Each team of the team variable has to be correlated with other variables to produce a point spread win in my opinion.

```
Linear mixed model fit by REML ['lmerMod']
Formula:
ATSWin ~ 1 + (1 | Team) + PassG + PtsG + AvgM + TOM + RZ
   Data: ats

REML criterion at convergence: 2026.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.65787 -0.68313 -0.04304  0.63390  3.12948

Random effects:
 Groups   Name        Variance Std.Dev.
 Team     (Intercept) 0.004124 0.06422
 Residual             2.317753 1.52242
Number of obs: 544, groups:  Team, 32

Fixed effects:
             Estimate Std. Error t value
(Intercept)  7.172567   0.501168  14.312
PassG       -0.008997   0.002399  -3.751
PtsG         0.210971   0.023892   8.830
AvgM         0.092677   0.019552   4.740
TOM          0.509543   0.162844   3.129
RZ          -1.145656   0.204821  -5.593

Correlation of Fixed Effects:
      (Intr) PassG  PtsG   AvgM   TOM
PassG -0.350
PtsG  -0.308 -0.597
AvgM   0.544  0.087 -0.253
TOM   -0.162  0.174 -0.094 -0.557
RZ    -0.379 -0.141 -0.164 -0.515  0.106
[1] 0.001776149
Linear mixed model fit by REML ['lmerMod']
Formula:
ATSWin ~ 1 + (1 | Team) + Yr + PassG + PtsG + AvgM + TOM + RZ
   Data: ats

REML criterion at convergence: 2025.9

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.63494 -0.67409 -0.01014  0.62728  3.03482

Random effects:
 Groups   Name        Variance Std.Dev.
 Team     (Intercept) 0.01346  0.116
 Residual             2.28253  1.511
Number of obs: 544, groups:  Team, 32

Fixed effects:
              Estimate Std. Error t value
(Intercept) -69.148605  28.144839  -2.457
Yr            0.038103   0.014045   2.713
PassG        -0.009584   0.002401  -3.992
PtsG          0.211124   0.023926   8.824
AvgM          0.105396   0.019960   5.280
TOM           0.478375   0.162378   2.946
RZ           -1.310972   0.213141  -6.151

Correlation of Fixed Effects:
      (Intr) Yr     PassG  PtsG   AvgM   TOM
Yr    -1.000
```

```
PassG  0.070 -0.076
PtsG  -0.006  0.000 -0.590
AvgM  -0.206  0.215  0.066 -0.248
TOM    0.061 -0.064  0.178 -0.093 -0.557
RZ     0.277 -0.283 -0.113 -0.165 -0.538  0.119
large : ATSWin ~ 1 + (1 | Team) + Yr + PassG + PtsG + AvgM + TOM + RZ
small : ATSWin ~ 1 + (1 | Team) + PassG + PtsG + AvgM + TOM + RZ
          stat      ndf      ddf F.scaling  p.value
Ftest   7.3431   1.0000 521.7651         1 0.006953 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
boundary (singular) fit: see help('isSingular')
Linear mixed model fit by REML ['lmerMod']
Formula:
ATSWin ~ (1 + Yr | Team) + Yr + PassG + PtsG + AvgM + TOM + RZ
   Data: ats

REML criterion at convergence: 2025.9

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.64988 -0.67619 -0.01531  0.62884  3.03856

Random effects:
 Groups   Name        Variance  Std.Dev.  Corr
 Team     (Intercept) 2.274e+00 1.5078957
          Yr          6.700e-07 0.0008186 -1.00
 Residual             2.277e+00 1.5090162
Number of obs: 544, groups:  Team, 32

Fixed effects:
              Estimate Std. Error t value
(Intercept) -69.406857  28.122026  -2.468
Yr            0.038239   0.014034   2.725
PassG        -0.009650   0.002406  -4.010
PtsG          0.211235   0.024040   8.787
AvgM          0.106079   0.019995   5.305
TOM           0.476620   0.162447   2.934
RZ           -1.312273   0.213633  -6.143

Correlation of Fixed Effects:
      (Intr) Yr     PassG  PtsG   AvgM   TOM
Yr    -1.000
PassG  0.071 -0.077
PtsG  -0.006  0.001 -0.587
AvgM  -0.206  0.215  0.065 -0.249
TOM    0.062 -0.065  0.179 -0.093 -0.557
RZ     0.277 -0.283 -0.112 -0.171 -0.534  0.119
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
Please use `linewidth` instead.Linear mixed model fit by maximum likelihood
['lmerMod']
Formula: ats$ATSWin ~ 1 + (1 | ats$Team)
   Data: newats

     AIC      BIC   logLik deviance df.resid
  2224.2   2237.1  -1109.1   2218.2      541

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.62283 -0.51598  0.06707  0.62111  2.75901
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 ats$Team (Intercept) 0.03793  0.1948
 Residual             3.41942  1.8492
Number of obs: 544, groups:  ats$Team, 32

Fixed effects:
            Estimate Std. Error t value
(Intercept)  7.85672    0.08645   90.88
```

As a bonus, I've tried fitting a time series model to see how much correlation the dependent variable have with prior years within the same team or between teams but couldn't get the graphs to draw. More on this later when I get the time.