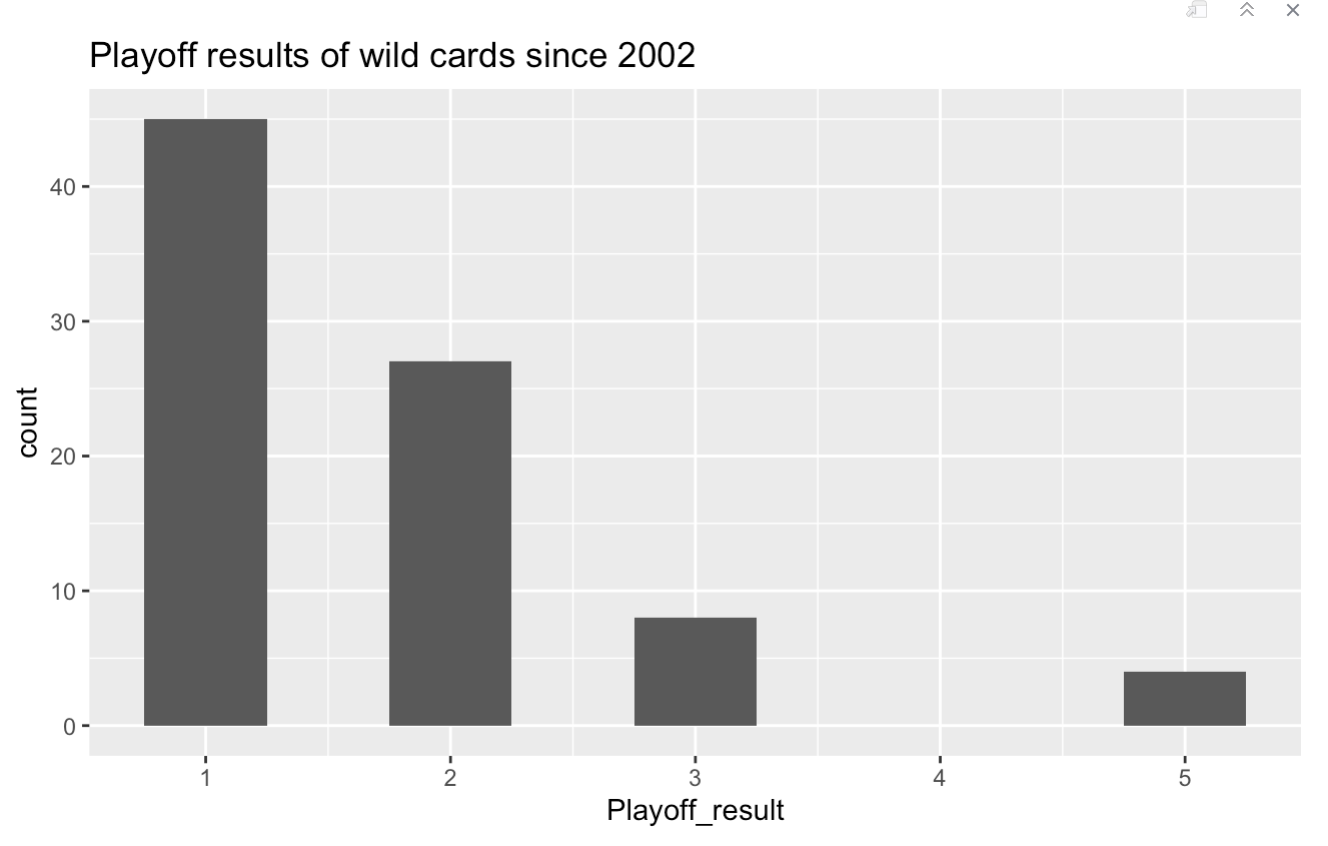
**Using Regression Modeling to Predict Wild Card Success in the NFL Playoff**

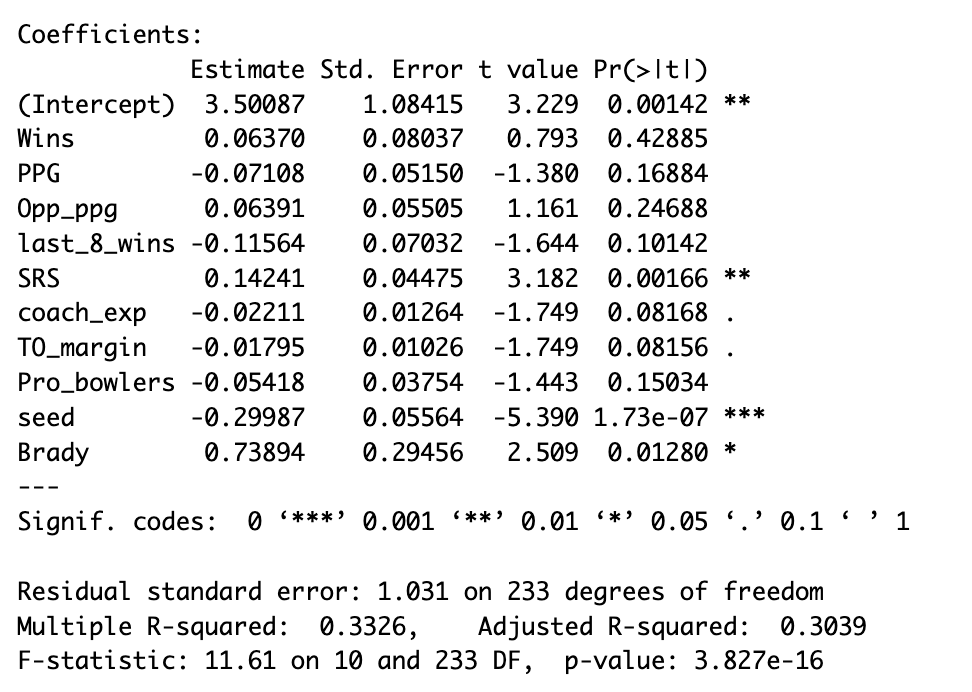
We’ve all heard the phrase, “Defense wins championships,” but how true is this? This paper explores what factors can be used to predict whether or not an NFL team will make a deep playoff run. Most of the time, wild cards lose in the first round, which is what one would likely predict. They’re playing against a division champion on the road, and it’s not an easy game. Figure 1 shows the results of Wild Card teams in the NFL Playoffs. In the last 20 years, 39 wild card teams out of 84 have won and advanced to the divisional round, where their odds of winning are even worse. However, sometimes those teams keep winning despite the odds, and four out of the 84 wild cards in the past 20 years have gone on to win the Super Bowl. Why did these teams go on a run to win the Lombardi? What did they possess that was different from most wild card teams that lost in the first round? The goal of this paper is to develop a model that explains the factors that result in Wild Card success in the NFL playoffs.

Figure 1. Playoff results of Wild Cards since 2002



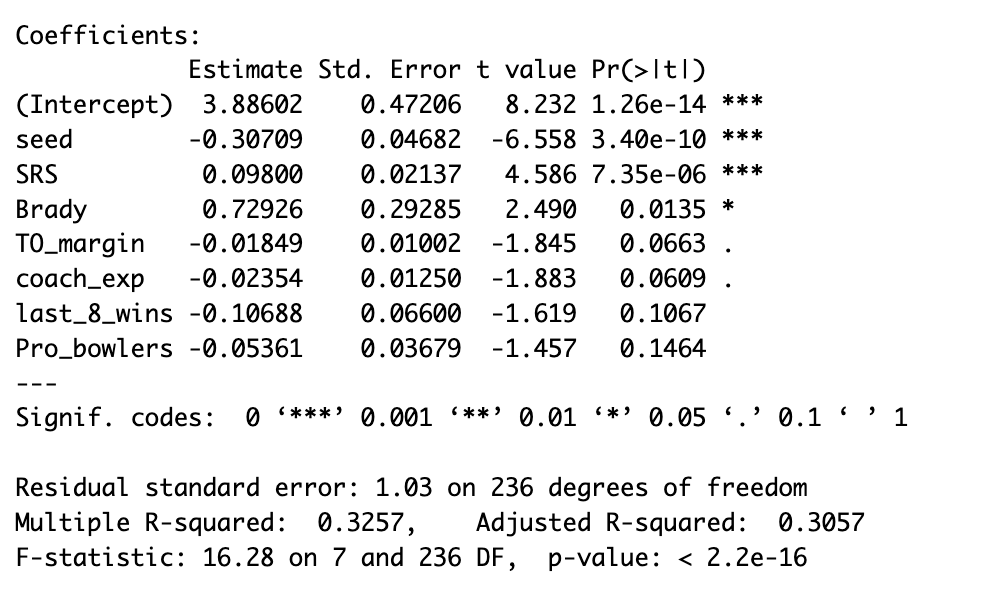
I collected data from 10 independent variables. Wins was the number of games the team won in the regular season. PPG, which was the points per game the team scored, Opp\_ppg, which was the number of points scored on them per game. Last\_8\_wins measured how many of their last eight games the team won to see if teams that got hot late did better in the postseason. SRS is a metric that combines a team’s margin of victory with their strength of schedule to determine how good they are, with 0 being average. Coach\_exp tells us how many years a team’s coach has been a head coach in the NFL, not including years as an interim coach. TO\_margin is a team’s turnover margin. Pro\_bowlers is the number of pro bowlers the team had. Seed is what seed the team is in their conference, and Brady tells us if Tom Brady played for the team. Then to measure playoff success, I created a variable, Playoff\_result. 1 meant that the team lost in the Wild Card round, 2 meant the team lost in the Divisional round, and so on until a 5, for a Super Bowl win. I collected this data for the last 20 NFL playoffs to try and see which of these variables explained success in the playoffs.

Table 1. Initial model



Since so many variables weren’t significant and because of potential collinearity. I ran forward selection, backward elimination, and stepwise regression to determine which variables to use. All three methods resulted in the same seven variables being used: seed, SRS, Brady, TO\_margin, coach\_exp, last\_8\_wins, and Pro\_bowlers. The eliminated variables were PPG, and Opp\_ppg, which I assume was because teams with good offenses still depended on their defense to make stops, and good defenses still had to rely on their offense to score points. The other eliminated variable was Wins which was probably because of the colinearity it had with seed. I believe seed was kept because it was better at predicting who won games because seed determines home-field advantage, not wins. Here is the new model below:

Table 2. Final Model

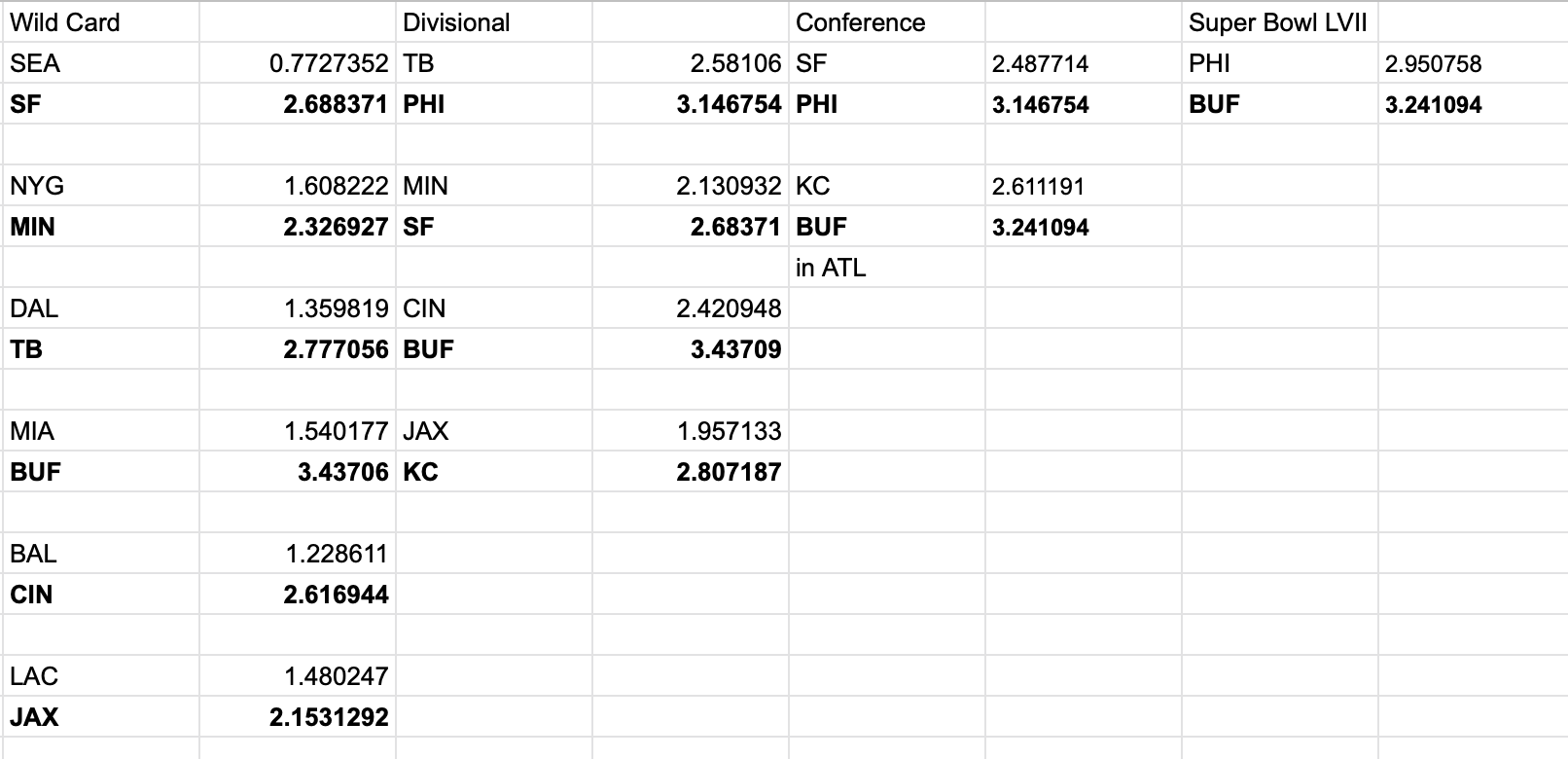


With this model, we have three variables that are statistically significant at the 𝛼 < .05 level. Seed was the most significant with P < .001. SRS and Brady are also significant at that level. The other four aren’t quite statistically significant but are close. This model has an r-squared of 0.3257, meaning it explains 32.57% of the variance in Playoff\_result. It explains enough to tell us what helps in the playoffs, but we can tell there is still a lot of randomness involved as well. There were many interesting aspects of this model. For example, coaches with more experience actually tend to perform worse, teams with high turnover margins do worse, and teams with more pro bowlers do worse.

I then looked at the four Wild Card teams that won the Super Bowl to see if this model might explain their success. The 2020 Tampa Bay Buccaneers had an SRS of 9.39 compared to the average of 5.26 and had Brady under center. The 2010 Green Bay Packers had an SRS of 10.94, and a coach with only five years of experience. The 2007 New York Giants had a turnover margin of -9 and only one pro-bowl player. On the outside, this might look like a disadvantage, but according to the model, this may help. This might be because the Giants manage to make the playoffs with a bad turnover margin, meaning if they can manage not to turnover the ball, they could be even more successful. Finally, the 2006 Pittsburgh Steelers had an SRS of 7.82, but all the other variables suggest they would do poorly. Their coach had 14 years of experience, they had a turnover margin of +7, and they had five pro bowlers, which suggests there is still a lot of randomnesses. Even though most of these teams had a high SRS, there were plenty of other Wild Card teams who had a high SRS but lost in the Wild Card round.

This model could be used to predict the upcoming 2022 NFL Playoffs. The output will give us the expected round that the team loses in, and we will assume the team with the higher value advances. In order to account for home-field advantage, we add a value of 2.00 to the home team’s SRS value before running the model. Table 1 provides the results when this model was run, showing the Bills defeating the Eagles in Super Bowl LVII. Unfortunately, this did not help us predict many upsets as no road team won a game, despite the 2-seed Bills winning it all. This is because the Bills defeated the Chiefs at the neutral site AFC Championship and then the Eagles at the Super Bowl in Arizona.

Figure 2. The model’s prediction for the 2022 NFL Playoffs



As we found with our model, there is still a lot of randomness involved in playoff games, so this likely won’t be how it plays out, which is one limitation of the model. Another limitation is that it’s possible for a team’s Playoff\_result to be less than 1, which shouldn’t be possible. This happened with the Seahawks due to being a 7-seed, a low SRS, and a coach with a lot of experience. It’s impossible to predict every matchup correctly. However, this model does give us a little extra insight into which teams should be on upset alert.