Calvin Makelky

**Do Elite Defenses or Offenses Significantly Outperform Expectations in the NFL Playoffs?**

My research question originated out of trying to find whether the age-old adage of “Defense wins championships” is true. What does this myth even mean, and what is a way that one can empirically test this myth? A common interpretation of this myth is that a NFL team needs an elite defense to win championships. It also can be interpreted to mean that NFL teams with elite defensives do better in the playoffs than elite offenses. So, does having an elite defense in the NFL mean the team will do better in the playoffs than expected?

Some subjective questions had to be answered first to carry out the research to this question. What qualifies a defense as elite? What would these teams’ expectations be based on?

When it came to formulating what an elite defense would be defined as, I considered various statistics. The first one that comes to mind is points allowed per game. The problem with this statistic is that it is affected by how fast a team plays. Teams that run the ball and take their time between running plays will naturally make it easier for their defense to give up less points because their slow offense will result in fewer possessions/chances for both teams to score. Thus, team that play fast will naturally make it easier for their opposing teams to score more, while slow teams make it harder for their opposing teams to score.

The statistics used should not be biased by the different paces teams play at. A superior statistic to points allowed per game is points allowed per drive (PAPD). Due to being based on drives (possessions) instead of games, this statistic is not affected by differing paces. There are only two possible variables out of a defense’s control that could bias this statistic: a team’s strength of schedule (SOS) and its opponent’s averaging starting position (ASP). Defenses that have a tougher schedule and/or their opponent’s offense starting closer to the end zone to score may have a significantly lower (better) PAPP statistic. To see if these two variables should be included in the analysis, I sampled a random NFL season where points allowed per drive was available from FootballOutsiders.com. I copied the data into a text file and measured the correlation between the three statistics in Statistical Analysis System (SAS). If strength of schedule and average starting position were significant in affecting a defense’s points allowed per drive, then the correlation between them and points allowed per drive would be significant (p-value below 0.05). Doing a Pearson correlation test in SAS, both SOS and ASP had very low correlation with PAPP, with SOS having a correlation of 0.171 and ASP having a correlation of 0.277. Obviously having an easier schedule or your opponent having worse field position is not going to override how good a defense will perform, so this result is expected. The question is whether it has a significant effect on PAPD. Doing a test for association between paired samples in R, the two-tailed probability of observing this relationship was not significant for SOS or ASP. ASP had a p-value of 0.124, while SOS had a p-value of 0.35. Starting position is almost significant, but is not significant because it is still above 0.1, and therefore is not worth including as an independent variable in this research. Thus, an elite defense will be based solely upon its PAPD.

What would make a defense elite in terms of its points allowed per drive? The top ten percent? The top five percent? The top one percent? Or simply the one team that is the very best? There are 32 teams in the NFL, and there were 30 teams back in 1997, which is the oldest year PPPA is available. Taken the top five or one percent would be ideal, but with only 20 years’ worth of data available, it would really limit the sample size and make it harder to make statistical conclusions. Thus, decided to take the top three teams each season in points allowed per drive. Further studies on this could play around with this number or base it on being a certain amount of standard deviations above the mean.

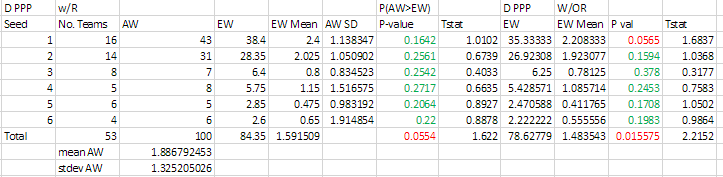
The last, somewhat subjective step, before I was able to start the research, was on what an elite defensive team’s expectations in the playoffs should be based on. The first that comes to mind is a team’s regular season record, but that would not take in the record and the toughness of their path in the playoffs. One year having a 12-4 record could mean having a bye as the one seed, while other years a 12-4 record could mean being a five seed and playing on the road instead of having a bye. Something that generally accounts for how well a team is expected to perform in the playoffs by taking into account every other teams’ records as well is playoff seeding. One seeds have the best record in their conference, and therefore are expected to do better than six seeds who barely made the playoffs.

A simple measure to assess the magnitude of expectations based upon seeding statistically is the number of wins in the playoffs. The one problem with this is that one and two seeds have a bye, and therefore get one less opportunity for a win than other teams. The solution to this problem was to count byes as a win. This gave the ability to compare playoff performance based on all seeds. Using NFL.com, I recorded the number of wins of every seed for the years of 1997 to 2016. Before I calculate the average wins per seed per playoffs, I assessed whether the performance of seeds in the playoffs remained relatively constant or if there was a sudden shift during a certain year that I would have to account for. Looking at the stacked graph, one can see that wins by seed remain relatively constant for every seed except possibly six seeds. There seems to be a positive shift in the number of wins by six seeds after 2004. The p-value comparing the years after 2004 and before 2005 using a t-test were insignificant though.

The next the step was to find the top three defensive teams for each season and record their seed and number of wins they had in the playoffs. I used footballoutsiders.com to find the top three teams in PAPD and NFL.com to find the playoff results. I recorded this along with the year and team name in an Excel file.

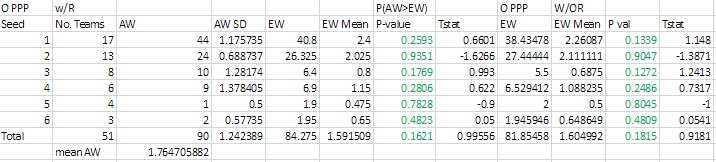
I then made a table of the number of teams sampled, the number of wins, the standard deviation of the wins, all stratified by seed. I then ran a student’s t-test using the number of wins over the number of teams as the actual mean, the standard deviation of the actual wins, the expected win mean, all for each seed and as a whole. The hypothesis test done was whether the actual win mean was higher than expected win mean. The p-value was non-significant at every seed, however, it was somewhat significant when comparing all teams regardless of seed. The p-value was 0.0554, which is moderately statistically significant. The expected win mean based on historical data was 1.592 based on the distribution of the number of teams in each seed I sampled. The actual mean was 1.887. **This means that as a whole, there is moderately statistically significant evidence that the elite (top three) NFL teams in terms of PAPD exceed win expectations in the playoffs based upon their seed when these elite teams are factored into the expectations.** Interestingly, there was not this significance when looking at specific seeds, but the small differences in each seed summed up to a significance difference as a whole.This analysis was done with replacement. As in the seed expectation mean and standard deviation was calculated using the elite defensive teams.

The result of analyzing elite defensive teams without replacement was similar, but more significant, to the outcome with replacement. To compare, I replicated the procedure above, but without replacement. This does not affect the actual win means and standard deviations, but it did change the expected win means and standard deviations. As a result, the p-values from the t-test for five of the six seeds were lower in analyzing the data without replacement than with replacement. The one seed whose p-value was higher, and thus less significant when analyzing without replacement was the three seed. On the other hand, the p-value decrease for one seeds was great enough that it became moderately statistically significant at 0.0565. The p-value analyzing all seeds without replacement decreased compared to the result with replacement. It dropped from 0.0554 with replacement, to 0.015575 without replacement. **Thus, there is highly statistical significant evidence that elite (top three) NFL teams in terms of points allowed per drive exceed win expectations in the playoffs based upon their seed when these elite teams are not factored into the expectations.** Another way to phrase this is to say that if elite defensive teams did not exceed expectations, we would obtain this observed difference or more in 2% of studies due to random sampling error.



As a bonus, I researched whether elite NFL offensive teams significantly exceeded expectations based on seed in the playoffs. This time I used points scored per drive (PSPD) instead of average points allowed per drive. Elite was once again defined as the top three teams that season in this statistic. The expected win means and standard deviations for seeds is the same as the defensive expectations used when done with replacement. With replacement, as a whole and for each individual seed, the p-values from t-tests were all above 0.16, and thus not statistically significant. **Thus, there is not statistically significant evidence that elite (top three) NFL offensive teams in terms of points scored per drive exceed expectations in the playoffs based upon their seed when these elite teams are factored into the expectations**. When looking at all teams as a whole regardless of seed, the p-value was only 0.1621.

Similar results came from performing the same analysis on elite offensive teams but without replacement. After calculating the expected win means and standard deviations for each individual seeds when removing elite offensive teams, the overall p-value regardless of seed actually increased from 0.1621 to 0.1815. Furthermore, none of the individual seeds when analyzed without replacement became statistically significant. **Thus, there is not statistically significant evidence that elite (top three) NFL offensive teams in terms of points scored per drive exceed expectations in the playoffs based upon their seed when these teams are not factored into the expectations.**



Potential problems with this analysis is sample size, assumptions that regular season performance in offense or defense roughly stays the same in the playoffs, and taking the top three team as one group rather than the exact magnitude of the “eliteness” relative to other teams that season.

The sample size for the defensive analysis was 53 elite defensive teams and 51 for elite offensive teams. Both of these are large enough when looking at all the teams as one group rather than as each seed because it is greater a sample size of 20. However, when looking at individual seeds for both the defensive and offensive analysis, the sample size was less than 20 for all seeds. This wasn’t much of a problem for one and two seeds, which nearly had sample sizes of 20. But this is potentially significant problem when trying to draw conclusions about the three, four, five, and six seeds. All had sample sizes of less than ten. I could not use years older than 1997 because of the absence of drive related data from those years. Potential remedies to this sample size problem could be to take the top five or ten teams. Unfortunately, this would dilute the meaning of elite, effectively making 5/32 or 10/32 elite, and giving less power to predict playoff performance when so many teams in the playoffs would be considered elite.

By classifying elite teams based on their regular season performance and the looking at their playoff performance, I am making the assumption that these are in fact the best offensive and defensive teams in the playoffs. The major problem with this idea is that teams can change from the beginning of the season to the playoffs. New players can play in playoffs that did not in the regular season due to injury or performance. For example, maybe the best defensive team loses its top defensive players due to injury at the end of the season. Then they are likely no longer an elite defensive team in the playoffs. There is not really anything I can do about this except to look at every individual team’s situation heading into the playoffs and subjectively removing them from the analysis if they are not deemed elite anymore. This would extremely time consuming and subjective though. More importantly, this potential problem seems unlikely to be a cause for concern if the sample size is large enough. It seems fair to assume most elite teams in the regular season are elite in the playoffs, and the few that aren’t will not make a significant impact on the analysis if the sample is big enough.

It may be better to look at how much each team is better defensively or offensively by looking at how many standard deviations they are above the mean rather than grouping them all together by top three. This way it measures exactly how elite each teams are, rather than categorically labeling them elite. This allows us to see how relatively elite the Patriots in 2007 were defensively, and not grouping a team that 1st and extremely elite with a team that was 3rd but not much better than the average NFL team. It would be hard to make conclusions for each individual teams, since the sample size would be one though. However, maybe I could establish a cutoff point in terms of standard deviation, where all teams above this point are significantly more likely to succeed playoff expectations. If I have more time, I would like to try this.

In conclusion, **there does seem to be strong evidence that elite defensive teams in the NFL exceed expectations in the playoffs.** In contrast, **there does not seem to be strong evidence that elite offensive teams in the NFL exceed expectations in the playoffs.** The evidence was moderate when done with replacement, and high when done without replacement. It makes more sense to use the without replacement results because we want to know if these elite defensive teams are separate group from all other teams in the playoffs. Thus, their results in the playoffs should not be factored into their expectations because the null hypothesis is that they perform just like every other team based upon their seed. Elite was strictly defined to mean the top three NFL teams in terms of points allowed per drive or points scored per drive. Expectations was strictly defined to what the average team in their seed since the current playoff format has been in effect. To sum it up, there is really strong evidence that elite defensive teams have a better chance of winning the championship than one would think based upon their seed, unlike for elite offensive teams.