**NFL COMBINE DATA ANALYSIS**

Ibrahim Taher and Michael Wayne

**ABSTRACT:**

The NFL Combine one of the first times prospects get to show their athletic ability to coaches and scouts. These coaches and scouts use the combine as one tool to guide their team to a successful draft. Coaches and scouts use the combine results of a player to determine whether he fits the mold of their team and if the team should use a draft pick on the player. At the end of the day, teams use the combine as one of the tools to measure success of a player in the NFL. But how good of a measure is the combine in determining success in the NFL? And the combine a good predictor of when a player is going to get picked? In this analysis, a linear model is used to answer these questions.

**INTRODUCTION:**

The National Football League is the premiere league for athletes to showcase athletic ability. College athletes work relentlessly so that one day they can hear their name being called during the draft. Once they declare draft eligibility, prospects are put under a microscope. Their film is reviewed and every weakness and strength is examined to make sure they are a candidate to play in the NFL. The prospects have to showcase their talents to coaches as well, at Pro Days, and if they are lucky to attend: the NFL Combine. The combine has several events that players participate in like the 40 yard dash, bench press, broad jump, etc. The combine is one of the many tools that is used to determine if a player should be drafted into the NFL. If the combine is an effective predictor of success, we can highlight which aspects of the combine are important for determining whether a player will have an impact in the NFL, which can be invaluable to a team’s front office, because then they can have more successful draft’s and the chances of drafting a player that will fail in the NFL will decrease. If the combine is just an effective predictor of draft position, then the average person can learn what coaches and scouts deem important when drafting for a certain position.

Here, we look into how effective the combine is in determining success via the average number of games started per season. This was used to measure success because there is no real measure of success between ALL positions. Offensive Tackles rarely catch passes and Defensive Ends don’t rush. Also, players who start games generally are considered good at the position and if they start a lot of games they are considered very good, if not excellent. In order to determine draft pick, we took the draft numbers of all the players who participated in the combine.

The dataset used consisted of all players who participated in the NFL combine from 1999 to 2015. However, we decided to use players from 2004 up until 2013, because players drafted after that probably have no started enough games to be considered successful and players who played before were vastly different than players who have been drafted recently. Also, we only used players who participated in three or more events because a prospect who only participated in two or one events did not participate enough in the combine and their combine results will not be a proper indicator of how successful they will be in the NFL. The dataset will be classified using linear regression. The recorded measurable aspects of the players and their combine results will be a part of the design matrix and the amount of games they started will be the observation vector for determining their success in the NFL. For draft position the design matrix will be the same but their draft pick will the observation vector instead.

The results will show how well of an indicator the combine results are in terms of success in the NFL. We feel that generally, players with better combine results (i.e. offensive tackles with high benches) and more ideal measurables for their position (wide receivers will be 6 foot 3 inches and around 210-230lbs) will have successful careers. Obviously there will be outlier, but the model will be a good predictor to how successful the players will be in the NFL.

**TECHNIQUE USED IN THIS PROJECT:**

The classification method used in this experiment is linear regression. Linear regression is being used because it is the premiere way a correlation between a player’s combine results and our measure of success in the NFL or their draft position can be formed. The linear regression is completed by using statmodels’ ordinary least squares (otherwise known as OLS) module. When looking at the OLS output, there are two aspects to take into consideration regarding how successful the model is for this analysis. The first are confidence intervals of the predictors and whether the confidence intervals say that the predictors are significant for how successful a player will be in the NFL or when he will be drafted. The second is the R-squared the OLS outputs because that shows how well fitted the data is.

In order to create our dataset, BeautifulSoup4 was used to scraped information from several websites. The data that was scraped includes, combine results, number of games started and draft position. The data was then written into a CSV. In order to run the OLS, the data in the CSV was put into a PANDAS dataframe. Whenever, a player was missing information for a combine event, the mean of the position for that event during that year was used as that player’s result for that event.

**DATASETS AND EXPERIMENTS:**

The dataset used for combine results and physical attributes was data scraped from nflcombineresults.com. Originally there were over 5665 players and 16 attributes scraped from the website. The original attributes that were scraped for: draft year, name, college, position, height, weight, hand size, arm length, wonderlic, 40 yard dash, bench press, vertical leap, broad jump, shuttle, 3 cone and 60 yard shuttle. However, the dataset did not provide the wonderlic, hand size and arm length for several players and it seems like the 60 yard shuttle was an event that not many players participated in. Therefore they were cut from the dataset.

Afterwards, when looking at the dataset it seemed like several players did not participate in several combine events. So, any player that did not participate in at least three events was also removed from the data. Also, all players drafted 2014 onwards were also cut from the dataset because they simply did not have enough game starts to be considered successful, even though they may be successful. It can be said that older players will be considered more successful because of more time in the league, but many players drafted rarely start and therefore players that have been in the league for 10 years may have as many starts as a player who got drafted two seasons ago. Players who participated in the combine from 1999 to 2003 were cut from the dataset as well. The motivation for this was that the NFL is a constantly changing game, with different needs at every position. The dataset needed to be large enough that players who already made a significant impact on the NFL were included, but players that were successful in the late 20th and early 21st century would not be as successful in NFL today due to the increase in athletic ability demanded at each position.

The games started by every player who participated in the NFL combine was scraped from pro-football-reference.com. Any player who did not start a game in the NFL was given a default result of 0. Draft position was also scraped from pro-football-reference.com. For all players, their draft position was subtracted from 257 (i.e. a player drafted first overall would have 256 as their draft position) and that was passed into the linear regression as their draft positions. Therefore any player that was not drafted was given a draft position of 0. The motivation behind this was it was deemed that 0 would be a more appropriate value given to the undrafted player than 257 or any other value after 256.

For seeing whether the combine could measure the successful of a player in the NFL, the combine results and the physical measurable of players who participated in the combine were the considered the design matrix and the numbers of game started were put in as the observation vector. The dataset was split into training data for the model and testing data using cross validation. 60% of the data was used as the training data. The same was done for determining the draft position of a player except the draft positions of players was put in as the observation vector.

There was also the question that whether the college that a player was drafted had any effect on his success in the National Football League. This data was from statssen.com. The data that was scraped was the AP Poll. This was used as opposed to the pure ranking because it was determined that the AP Poll was a truer estimate for how good a college was during the year a player was drafted.

OLS was used on the training data with the corresponding design matrix and observation vectors for each position and the outputs were examined for any insignificant predictors. Those predictors were removed and OLS was reran. The outputs were reexamined and for any output that had an R-squared that was below .7, it was determined that the combine was not a significant contributor for predicting draft position and/or success in the NFL. The new OLS models were used on the testing data next and the error for the predicted values and actual values was calculated and put on a histogram.

**RESULTS:**

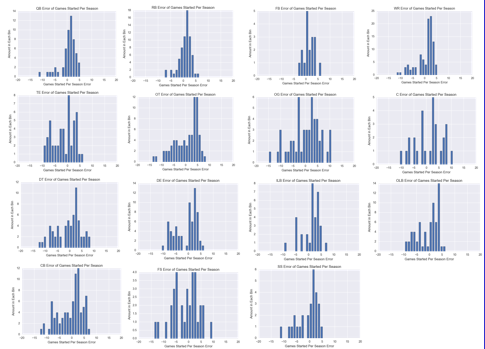
*Games Started:*



**Figure 1.** The OLS outputs for games started per season. Above are all the predictors and their coefficients for each position.

The OLS linear regression model was deemed an ineffective predictor of success in the NFL, using games started as the predicted value. Looking at figure 1, we notice that the R-squared values, which is how well fitted the data is to the linear model is across the board, very low. This is most likely because the combine is mainly a test of athletic ability, but it does not test what most football players, coaches and scouts call Football IQ. A player can be extremely athletic, but if they do not have any sort of football intelligence then they will be more prone to make incorrect decisions after the play beings. It is interesting to note that the quarterback position is not fitted well at all. This is most likely because the combine again does not test football intelligence, but also it does not test their throwing ability and it does not test their ability to react in the pocket. Also the coefficient for the bench press of the quarterbacks was not given a value because quarterbacks generally do not compete in the bench press competition.

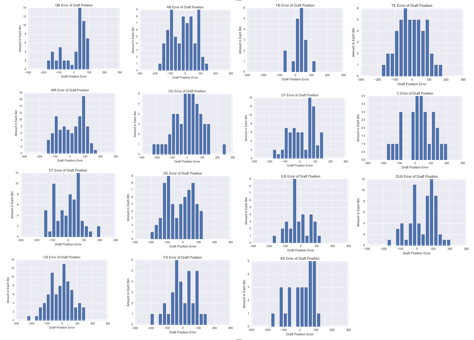
Generally, when R-squared values are low it means that there are insignificant predictors being passed through the linear model. However, when significant or pseudo-significant predictors were determined the R-squared values decreased. This led to the assumption that significant predictors could not be determined and that in the end, the combine was not a good predictor of success in the National Football League.

**Figure 2.** Errors between the predicted values from the OLS and the actual values of the amount of games started per season for each position. The order from left and right was quarterback, running back, full back, wide receiver, tight end, offensive tackle, offensive guard, center, defensive tackle, defensive end, inside linebacker, outside linebacker, cornerback, free safety and strong safety.

 We continue to find evidence that the combine is not a good predictor of games started based on figure 2, which shows the errors between the predicted values and the actual values. Across the board, the errors are not densely packed near 0, but rather sparse across the histogram. Since our parameter for success was average games started per season, anything deviation that is beyond 4 games per season can be considered an inaccurate estimation because 4 a game difference between the actual and the predicted is a 25% difference, which we can definitely consider significant difference and a difference that shows huge inaccuracies. Across the board, the errors are sparse, signifying that there is a huge dichotomy between the actual success of a player and the predicted success based on combine results, AP Poll Points and physical attributes.

**Figure 3.**  The OLS outputs for draft position. Above are all the predictors and their coefficients for each position.

When the OLS module was run on the draft position, the results were slightly more promising in terms how well fit the data was, but it was not considered good enough. The only decent fit position was the cornerback position, which when understanding the position’s role in a defense has some ground. Cornerbacks need to be very athletic. They need to change speed and direction depending on which direction the wide receiver, or player in their zone is heading and they need strength to press off an opposing player to make a play such as a pass deflection or even an interception. A similar situation to what happened with the draft position as with the OLS of the game started. When removing the insignificant values the R-squared actually went down.



**Figure 4.** Errors between the predicted values from the OLS and the actual values of the amount of draft for each position. The order from left and right was quarterback, running back, full back, tight end, wide receiver, offensive guard, offensive tackle, center, defensive tackle, defensive end, inside linebacker, outside linebacker, cornerback, free safety and strong safety.

We continue to find evidence that the combine is not a good predictor of draft position based on figure 4, which shows the errors between the predicted values and the actual values. Across the board, the errors are not densely packed near 0, but rather sparse across the histogram. Since our parameter for success was average games started per season, anything deviation that is beyond 32 picks can be considered an inaccurate estimation because a 32 pick difference between the actual and the predicted is roughly an entire round, which we can definitely consider significant difference and a difference that shows huge inaccuracies. Across the board, the errors are sparse, signifying that there is a huge dichotomy between the actual success of a player and the predicted draft position based on combine results, AP Poll Points and physical attributes.

**CONCLUSION:**

Essentially the conclusion is simple, the combine is not as great of a predictor of NFL success and draft position as we once thought. Scouts, coaches and analysts put too much emphasis on combine results, when in actuality there have to be other aspects of a player that determine his success in the NFL and his draft position. It is possible that the main indicator of where a player will go in the draft is their play at the college level. Players who succeed at the college level generally are drafted high and players who do not play as well, or come from smaller schools, generally are drafted lower. For quarterbacks, such as Marcus Mariota of the Tennessee Titans playing in an NFL-style offense at the college level helped him go number 2 overall when he was drafted.

For success in the NFL there are probably many more aspects of a player that go into his success at the professional level. One that comes to mind is football IQ, because having high intelligence in regards to the game means that a player can make smarter decisions faster and make better moves in the short duration of a play. Also, their adaptability to pro level also probably has a huge impact on their NFL success because if they cannot change their play from the college level to the pro level, they are generally not going to be able to succeed due to the quality of players being much higher at the professional level.

If the NFL wants to keep the combine and make it a better indicator of NFL success, then the officials need to make a call about how they are going to overhaul it. For one, they could possibly make the combine more position specific, giving different events to different players so that they can showcase their athletic ability as a predictor. Also, possibly trying to find a measure of football IQ (aside from the wonderlic, because that is not a good indicator) would also be beneficial. In short, do not look to the combine for how judging a players talents, rather look elsewhere.