UNIVERSITY OF ARIZONA

Submission of Thesis to the Honors College

PREDICTING OFFENSIVE SUCCESS IN THE NATIONAL FOOTBALL LEAGUE

BY

JACOB ALLEN KRAUSE

_____

23057514

Honors Thesis

Fall 2012

Approved by:

_____

Professor Thomas Dalton

Department of Economics

Introduction

Throughout the world there are many things that can serve as a driving force for an individual. This thing motivates people to achieve things that they never before thought possible. For some it may be money, for some fame, and yet for others it could be an emotion such as love. In professional sports, athletes, coaches, executives, and teams are all motivated by one thing however, success. This success that is so highly sought after ultimately takes the form of winning a championship, but starts at a much lower level. Much like in Economic theory, this idea of prosperity or success starts at the individual level. Individuals must make decisions that are in their best interest in order for the economy to fully function. In the NFL, acting in one's best interest means being physically and mentally in the best possible shape to succeed. Only if all of the players on a given team are in top form will the team find success as a whole.

The discussion above leads to one question. How can one predict success on an individual level in the National Football League in order to make the team as a whole better? This paper will attempt to predict success on an individual basis using eleven variables that can be measured at the annual scouting combine. The scouting combine takes place before each year's NFL draft and is a chance for coaches, scouts, general managers, and owners to evaluate the talent at the combine, in order to make their team more successful in the future. These individuals use intuition, technique evaluation, and other knowledge in order to evaluate a player but I will be evaluating players on a basis that is purely statistical. I have played football for eleven years in the little league, high school, and collegiate ranks and will look to combine my knowledge of the game, economic theory and statistical analysis to produce a regression that predicts success. This paper will use data analysis in order to evaluate players on variables that are common to every football player. In my analysis I will try to predict the success of skill

position offensive players only, which are Quarterbacks, Wide Receivers, and Running Backs, as they have common goals to be achieved in games. It is my belief that these eleven variables can be used in order to predict some degree of success in the NFL.

Literature Review

When one conducts a report of this kind, it is useful to include a discussion about what others have found in performing similar studies. As far as I can tell, there haven't been any studies done with the same specific research question as in my study. I have found however, a study conducted by ESPN that predicts draft position of players using their combine statistics. This is not a prediction of NFL success but solely a prediction of what round and selection an offensive player will be drafted in. This study finds that Quarterback's draft position correlates to height at the -.27 level which means that a taller player will be drafted later on average. It will be interesting to see if my findings conclude that height correlates to success in the NFL as well. The only other study that I found that predicts success is based on fantasy football statistics, not combine statistics. This study attempts to find indicators of second year running back success based off of their statistics from the first year they played in the National Football League. As far as my research has shown, my study is the first of its kind to be performed.

Methodology

The regression model used here is not radically different from other models created that predict success in a field. The data used for the regression have come from two different databases on the internet. The first database that I used to gather data was NFLcombineresults.com, which has compiled combine data from 1999 up until the most recent Scouting Combine. They keep track of all of the variables that I have used in my regression,

plus a few others. Their dataset contains all positions on a football team, and I have chosen three positions on offense to use for my regression model. The next source that I used to compile my dataset is Pro-FootballReference.com. This source is a highly respected source in the football community as they have an extremely large dataset and keep track of every statistic in football. I used this website in order to track the Pro Bowl appearances of each player in my database.

My regression uses eleven variables in order to predict success of players in the National Football League. The independent variables that I am using for my regression are: College Conference, Position, Height, Weight, Wonderlic Score, 40 Yard Dash, Bench Press, Vertical, Broad Jump, Shuttle, and 3 Cone Drill. I believe that these variables will be able to best predict success, since all can be tested subjectively and under the same set of conditions for different athletes. I used the Combine statistics rather than statistics from individual player's pro days because of the consistency of the testing conditions. In players pro days they can be set up with advantages other players might not have at theirs. For example, players from larger and wealthier schools are able to run on different track conditions that aid their times.

The way that I will measure success in this regression is by the number of Pro Bowls attended. This dependent variable is a quality measure of success because it shows how players perform relative to their peers. Players are voted into the Pro Bowl based on a season's success and multiple Pro Bowls suggest that a player has been successful over a career and that success did not come over a short span.

I have compiled all of the data into an Excel spreadsheet and from there I produced descriptive statistics on each variable and conditional statistics with the independent variables and the dependent variable. The College Conference variable and the Position variable have

been coded in order to allow them to be used in a regression. I have broken down the College Conference variable by marking a "1" in the dataset for "Major" conferences and a "0" for "minor" conferences. The Position variable was also categorical, and I have created dummy variables in order to account for that variable. Missing values in the dataset will be handled by substituting the average value for a given variable. In this way the dataset will not be altered in any way. The regression will be run using Excel as well, with charts being outputted for each variable.

Description of Variables

The dataset that I use for this thesis contains 2,013 total observations and 12 variables. The dataset has been gathered using historical databases on the National Football League. This dataset is cross sectional in nature, as I have looked at data from a specific time range. In this section the variables will be listed and explained further, with a description of what the variable is, how it is measured, and how it is represented in my dataset. The first variable is the college conference variable.

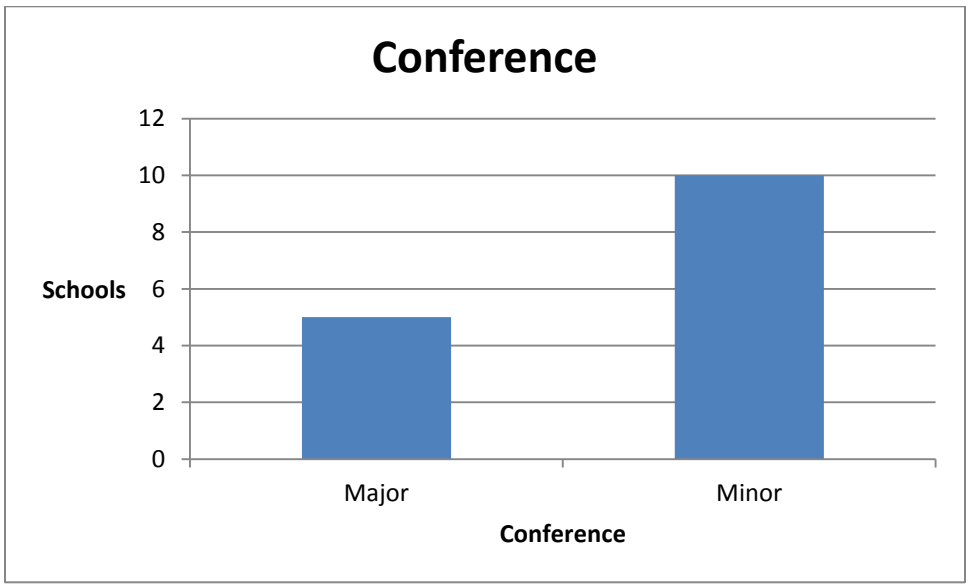


Figure 1

As can be seen in Figure 1, the Minor conferences dominate the dataset as we have far more conferences represented than we have Major conferences. There are ten total Minor conferences and only five total Major conferences. This variable will be represented by a "1" for "Major" conferences and a "0" for "Minor" conferences when taken into account in my regression.

The second variable is position, which is defined as either a Quarterback, Running Back, Wide Receiver, or a Tight End.

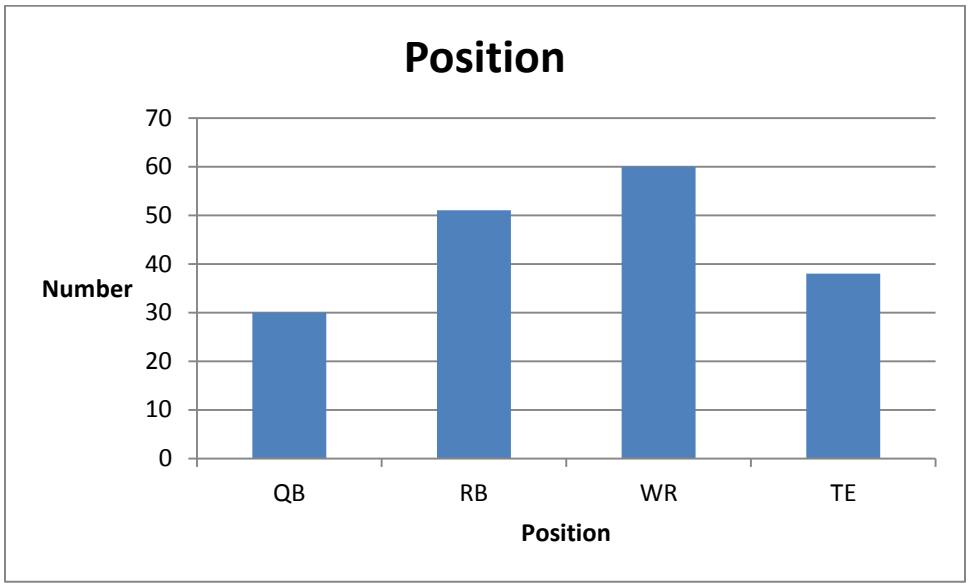


Figure 2

We can see from Figure 2 that there are more Wide Receivers than any other position in the dataset. This is because more Wide Receivers play on the field at one time than any other position, so teams need to have more on the roster. The Quarterbacks are the least represented in the dataset, as only one plays at any given time.

The next variable to be defined is the height variable.

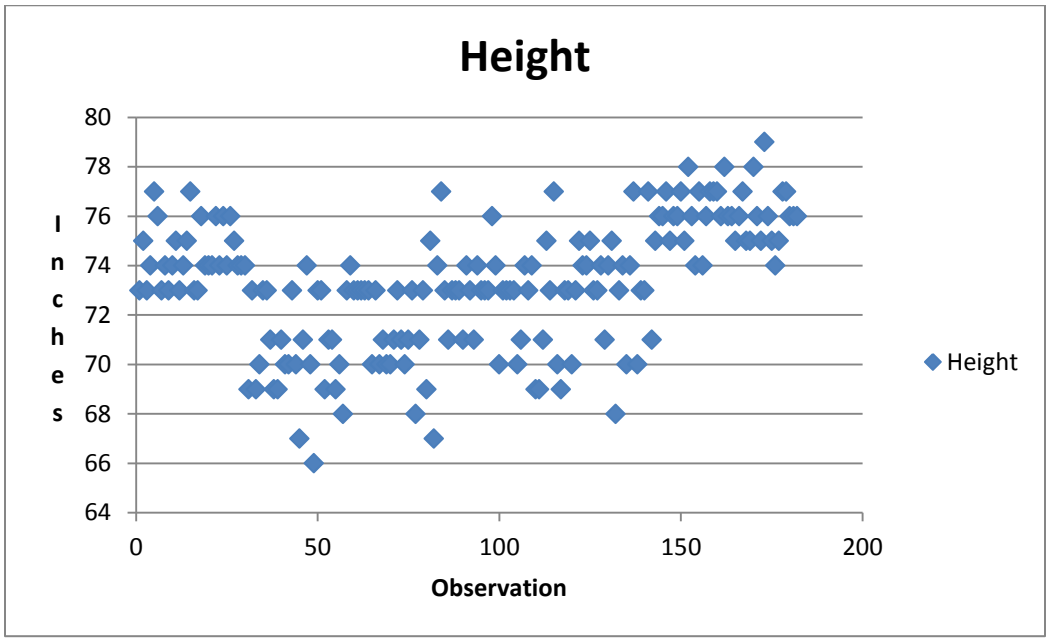


Figure 3

As we can see from Figure 3, the shortest player in the dataset is 66 inches tall while the tallest player is 79 inches tall. The majority of players fall under the 70 to 76 inch range. At the combine height is measured in front of all of the coaches and scouts, so that no foul play is involved. I predict that the height variable will be the most predictive of success for Quarterbacks and Tight Ends in the NFL, because these two players need to have a higher vantage point in order to see open players and to create leverage while blocking. Height is not as important for running backs and wide receivers because they can cover up a lack of height by being quick or fast.

The next variable is weight. The weight variable is measured in pounds and not kilograms, as all combine statistics do not use the metric system.

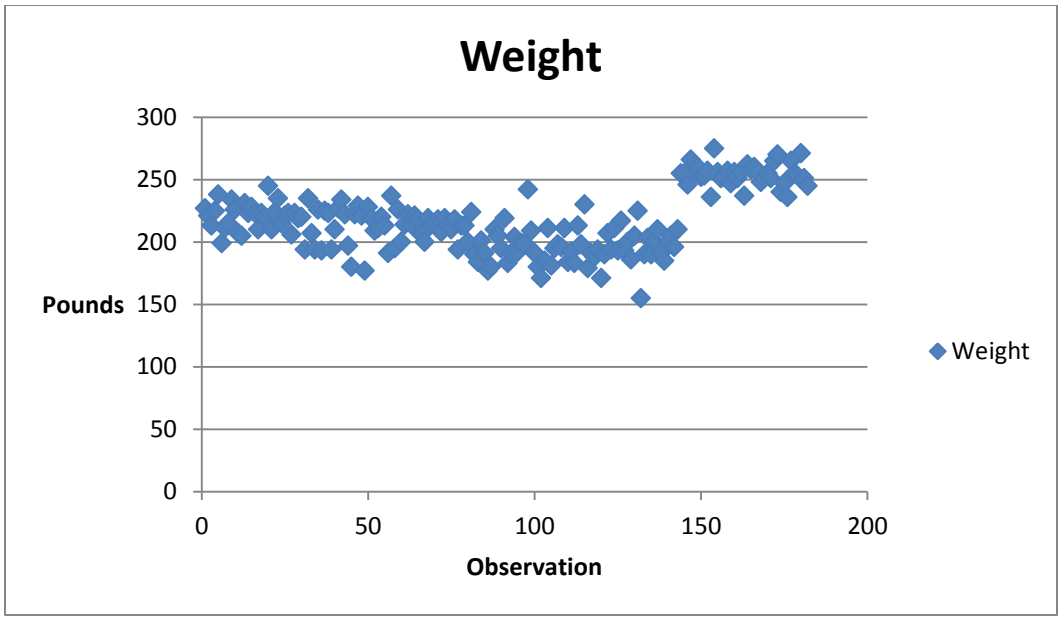


Figure 4

In Figure 4 we see that the lightest player at the combine weighs 155 pounds and the heaviest player in the dataset weighs 275 pounds. The lightest group was the wide receiver group and the heaviest group was Tight End. I believe that maximum success will be achieved by players that are of average weight for their positions. This allows players to be well rounded in what they can do in terms of performance on the field. Weight is important in the National Football League in order to prolong a career. Players that are extremely light for their position are put at a greater risk of injury. This can affect their overall success by shortening their career and thereby allowing them to be voted to fewer Pro Bowls.

The next variable to be discussed is the Wonderlic Score. This test measures intellectual ability of players attending the combine.

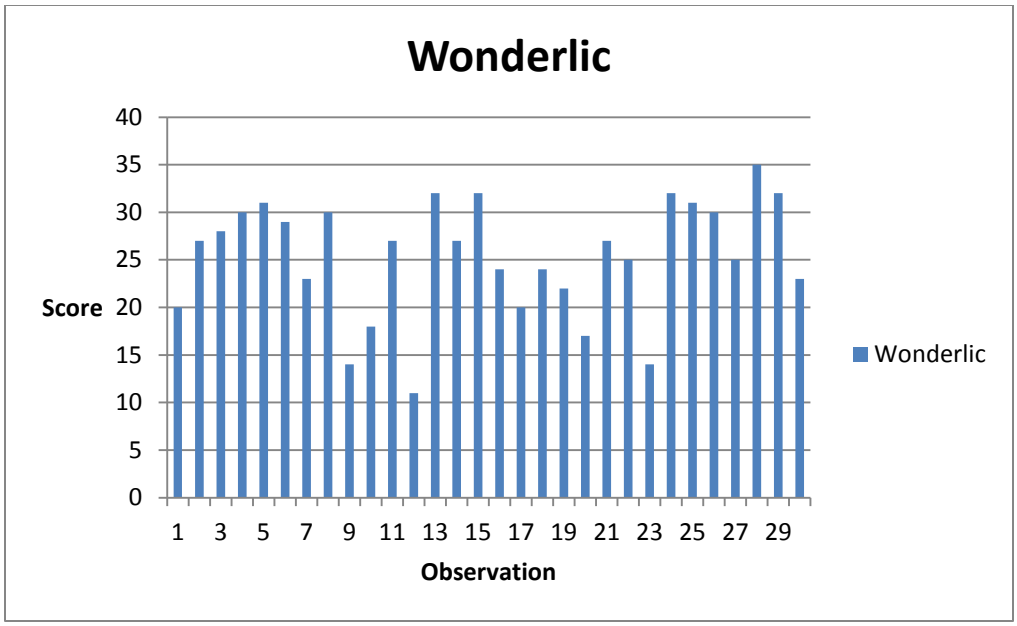


Figure 5

This test is only administered to the Quarterback position and will only be included in the regression that I run on that group of players. The Wonderlic is similar to an IQ test, but tests player's knowledge as it pertains to the game of football. The graph shows us that the highest Wonderlic score measured was 35 while the lowest score was 11. A Quarterback is said to be the most cerebral player on the field, and I feel as though this variable will correlate highly with success in the NFL. There have been many instances in the past of superstar collegiate athletes scoring low on the Wonderlic. These players success in college does not usually translate into success in the NFL as the game moves too quickly for them. In college they relied on athletic ability but in the professional realm, they need all of the skills in order to excel. A prime example of this is Vince Young. Young, a superstar at Texas, won the National Championship his last year there but saw limited success in the NFL after scoring very poorly on the Wonderlic.

One of the most important variables to scouts and coaches at the combine is the 40 yard dash. This test can either make or break a players draft status and correlates highly with draft position. The high correlation with draft position does not necessarily mean a high correlation with NFL success.

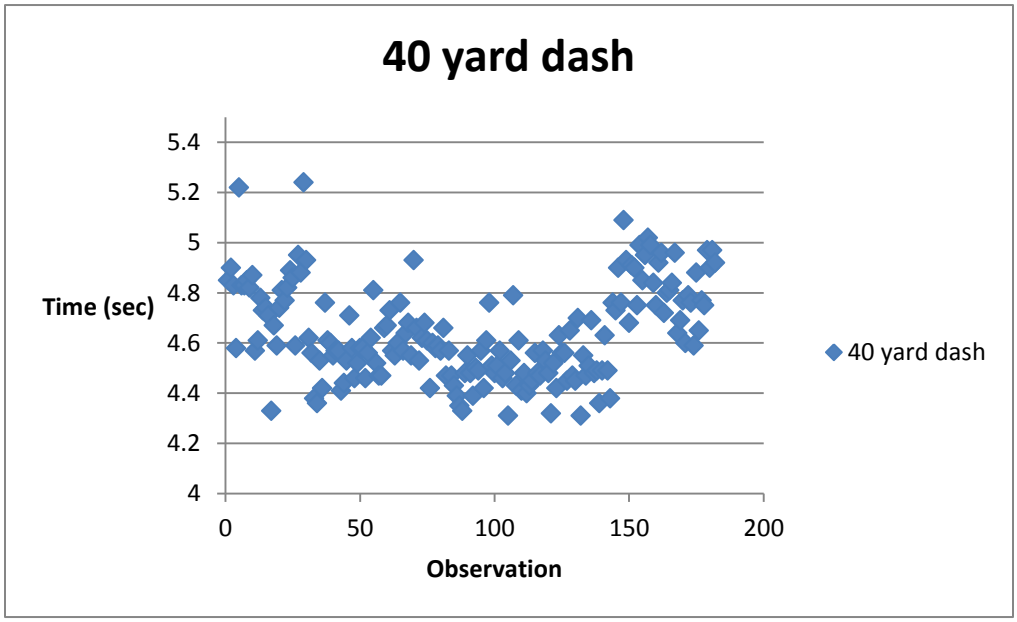


Figure 6

As we can see from Figure 6 the 40 yard dash times of the position groups in my dataset vary from a minimum of 4.31 to a maximum of 5.24. This variable is a measurement of how fast a player can run 40 yards. It is measured electronically so that no user error is added or subtracted from a player's real time. In the case of this variable time is measured in seconds. I believe that the highest correlation to success considering the 40 yard dash time of a player will be amongst the running back and wide receiver group. These players need to be fast enough to outrun other players and create space on the football field. Slow running backs will not be able to get to open holes or follow their blocks well and slow wide receivers will not be able to create separation between themselves and the defense, making it almost impossible for the quarterback

to be able to throw them the ball. While it helps for quarterbacks and tight ends to be fast, it is not a necessary skill for them to be able to succeed.

The next variable in my regression will be used for two groups of players only in the position specific regression. This variable is bench press which measures the number of times that a player can bench press 225 pounds.

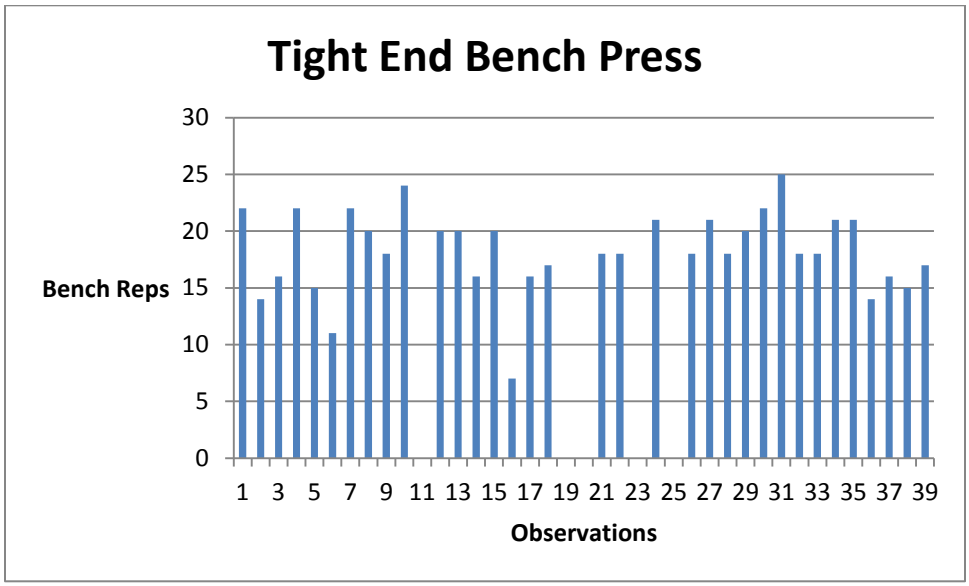


Figure 7

The tight end group has a maximum bench press of 25 repetitions and a minimum of 7 repetitions. This variable will be an important factor for success amongst tight ends because this group needs to be able to block defenders for extended periods of time as well as fight through press coverage at the line of scrimmage and go out for a pass. Tight ends that cannot do these things will not be successful in the NFL.

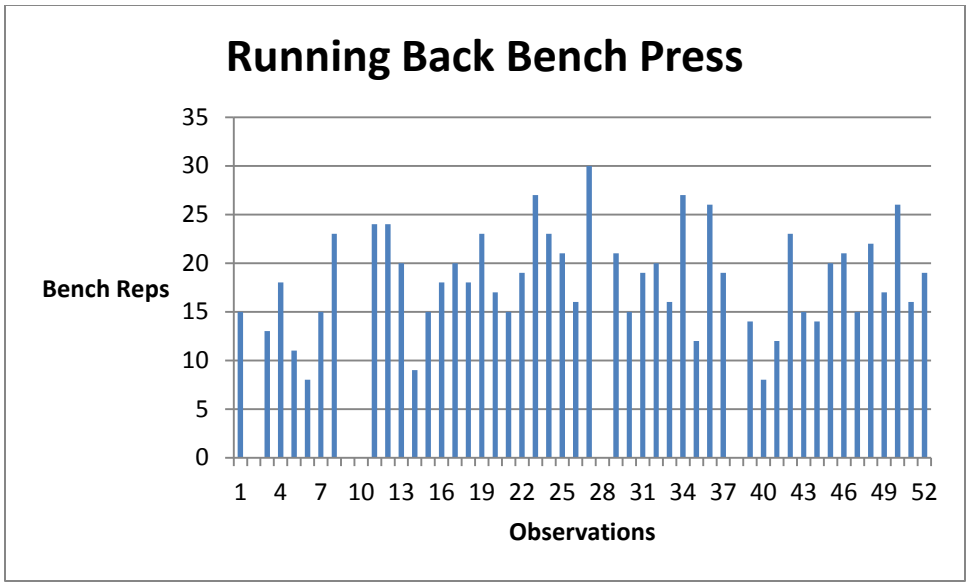


Figure 8

The bench press is less important for running backs but will still be a decent predictor of success. Running backs do have to be able to block defenders, but not on the same consistency as tight ends. Running backs are not pressed at the line like tight ends are so will not need to fight through press coverage. Running backs do need to have good upper body strength in order to break tackles, which is a vital skill for a running back's success in the NFL. The bench press is conducted by one coach for every player, so that no player gets special treatment throughout the process. It is performed in a room full of scouts and coaches, by one player at a time using the same bench press for each individual.

The next variable that I chose for my regression is the Vertical Jump. This exercise measures explosiveness in players which is valuable for almost every position on the football field. A player's vertical jump is measured using an upright bar with pegs sticking out at the top. A player stands with both feet planted on the ground underneath the pole and then explodes upward trying to tap the highest peg that he can. The number of inches is called out by the coach

that is overseeing the drill and the players reach is then measured by another coach. The height of a players reach is then subtracted from the height of the players jump and that difference is a players vertical leap. This measurement is taken in inches, which is common amongst the combine statistics.

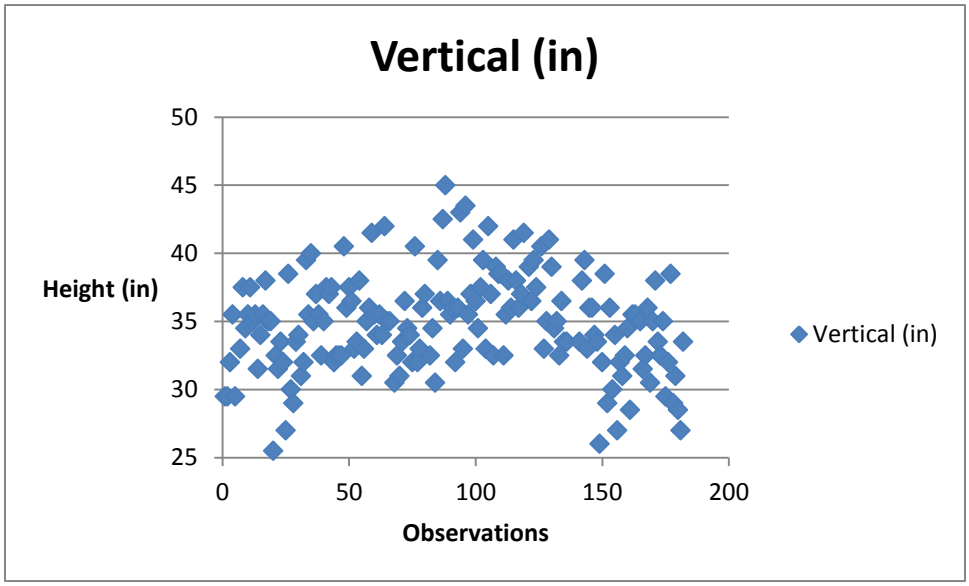


Figure 9

As we can see from Figure 9, the maximum vertical jump is 45 inches while the minimum vertical jump is 26 inches. The majority of players jumped in the 33 through 36 inch range. This measurement will show the most correlation to success amongst the wide receiver and running back group, as these players need to stop and start running very often. By doing so, running backs and wide receivers can elude and run away from defenders before they are able to close in around the player. Being able to achieve top speed in just a few steps is crucial to these position groupings because of the above reasoning. Wide receivers need a high vertical jump so that they are able to jump up and get the ball at the highest point possible. This means that they can make more catches throughout a game because the defender won't have a chance to knock

the pass away. Receivers like the Detroit Lion's Calvin Johnson uses his extremely high vertical to his advantage as he wins many jump balls for touchdowns and has already been to 2 Pro Bowls in his 5 year career.

The next variable is similar to the last one, but measures explosiveness horizontally instead of vertically. This variable is the Broad Jump, which is also measured in inches. In the broad jump, a player stands at a painted line on the track with both feet planted on the ground. He then explodes forward jumping as far out in front of him as he can, and the distance is measured between where he started and where his foot hits the ground after the jump.

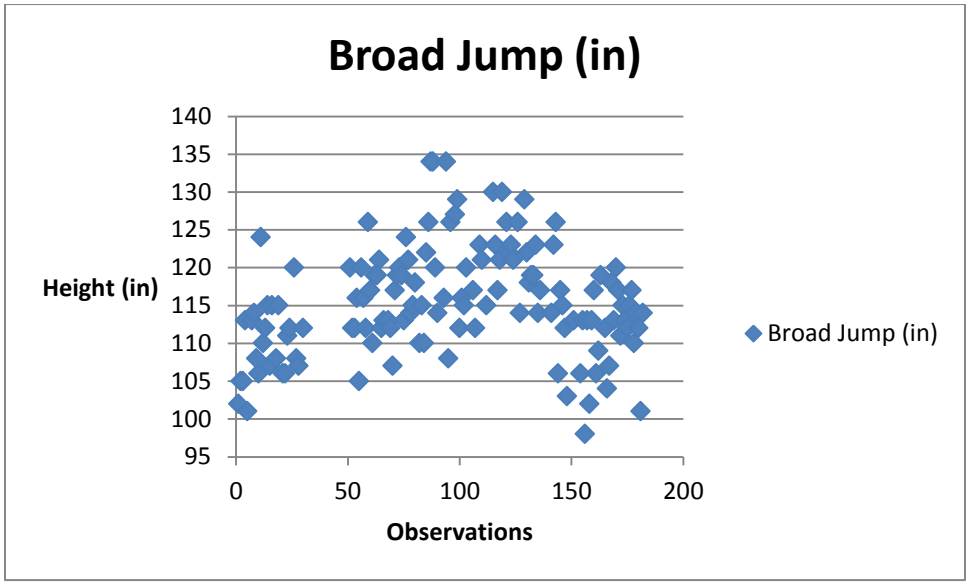


Figure 10

As we can see from Figure 10 the longest jump performed in the 2001 or 2002 combine was 134 inches while the shortest jump performed was 98 inches. The average of all the broad jumps performed was 115 inches. Since the Vertical and Broad Jump are very similar in what they measure it will be very interesting to see which one has a higher correlation with success after the regression is performed. This variable will probably have the highest correlation of

success with the wide receiver group and the running back group, for the same reasons as described above. I also believe this to have more of a correlation to quarterback success than the vertical variable, because quarterbacks need to be able to slide in the pocket to avoid pass rushers.

The next variable that will be used in the regression is the 20 yard shuttle. This drill measures quickness in players. The drill puts three cones five yards apart in a straight line, with the player standing in front of the middle cone. When the whistle is blown, the player runs to the cone on his right, touches the line, and then runs all the way to the cone that is ten yards away. He then touches the line there and sprints back through the middle cone and the time is stopped.

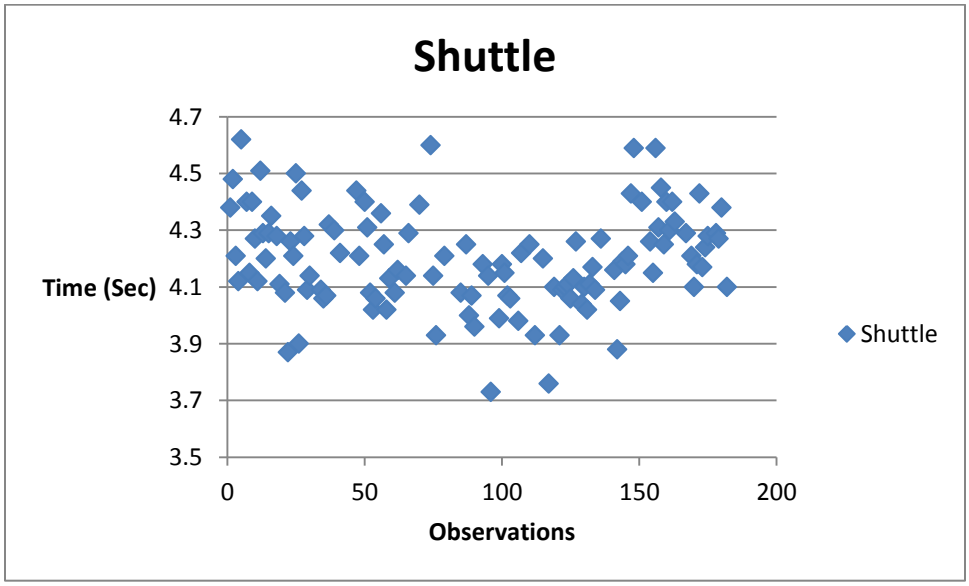


Figure 11

As we can see from Figure 11, the maximum time from the 2001 and 2002 combines was 4.62 seconds while the minimum time was 3.73 seconds. The average of all the players at the combine was 4.19 seconds. I believe that this drill will have the highest correlation of success amongst the running back, tight end, and wide receiver groups as all of these players need to be

quick in order to block, run routes, or run with the ball. The shuttle is not only a measurement of how quickly players can stop to avoid players, but also a measurement of how quickly they can hit full speed which is useful to run away from players in small spaces on the field.

The next and last independent variable used in the regression is the three cone drill. This drill measures quickness and agility of players as they have to bend and maneuver around cones. In this drill there is a total of 30 yards run and five changes of direction involved, not all of the same type like in the 20 yard shuttle.

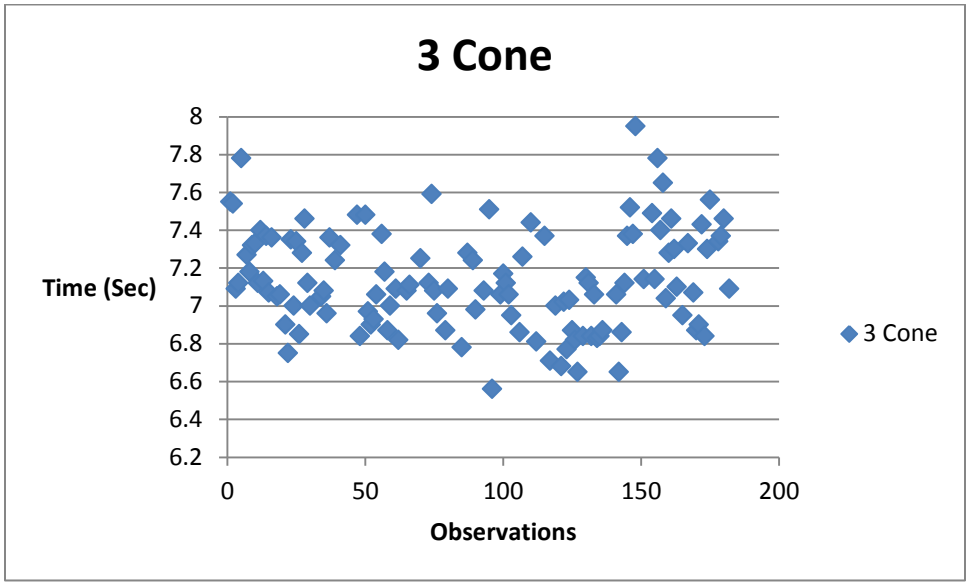


Figure 12

As we can see from Figure 12, the maximum time for a player at either combine was 7.95 seconds while the minimum was 6.56 seconds. The average for all players at the combines was 7.14 seconds. I believe that this will also show the highest correlation of success in the tight end, running back, and wide receiver groups, with running back and receivers showing a higher correlation than the tight ends. These groups need the most athleticism in order to thrive on the

field so drills like the vertical, broad jump, 20 yard shuttle, and 3 cone drill are of vital importance to them.

The last variable to describe is my dependent variable, which is Number of Pro Bowls. This variable is how success will be measured in this regression and is a good indicator of a player's success on the field, the more pro bowls the more success a player enjoys.

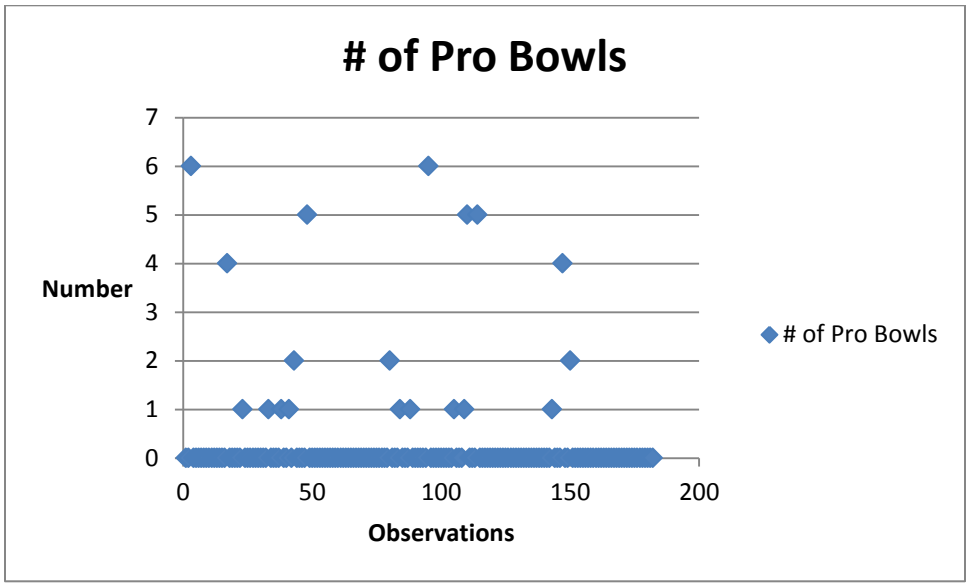


Figure 13

Figure 13 shows that success is hard to come by in the NFL. Only a select few players are able to achieve it which is why being able to predict success is such an important factor of teams winning games and building successful franchises. Teams that waste draft picks on players that are not successful set the franchise back multiple years and keep them from being competitive. The teams that enjoy success in the NFL are the teams that can consistently draft successful players and build on a previous year's draft class. The graph shows us that the most success enjoyed in the database came in the form of 6 pro bowls by a specific player, and this happened twice out of the two combines measured. The minimum number of pro bowls by a

player was 0, which happened164 times. There are a total of 50 pro bowls between the 183 players that were involved in the regression.

Conditional Statistics

This section will show conditional statistics for each of the independent variables separately with the dependent variable. The first conditional statistic that will be evaluated is success of players by conference.

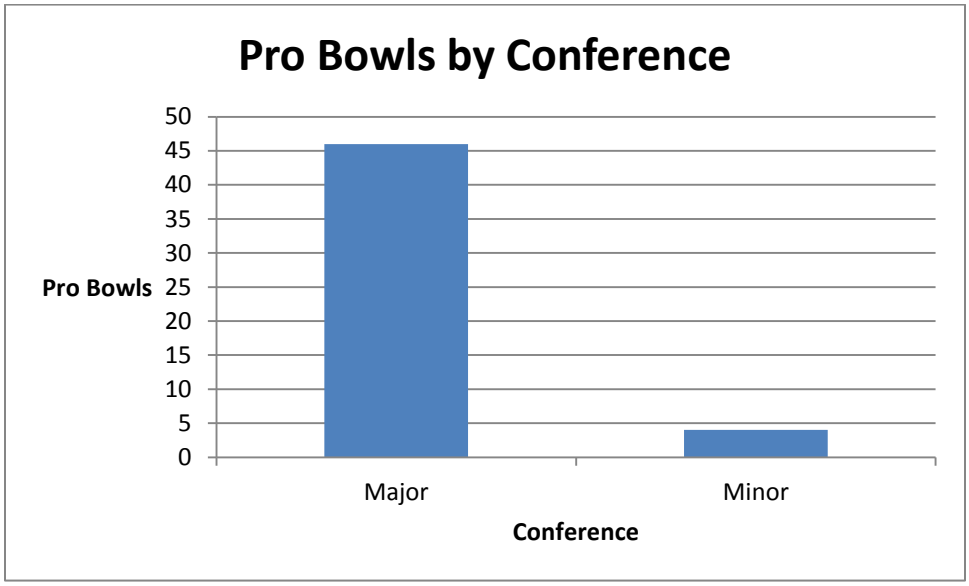


Figure 14

In Figure 14 I have broken down the number of Pro Bowl appearances from each conference in the dataset. We can infer from the graph that there are many more Pro Bowl players and Pro Bowls that are from the Major conferences than from the minor conferences, even though there are more minor conferences in the dataset. This should lead to a positive correlation between the conference that a player plays in and his chance of reaching a Pro Bowl and becoming a successful NFL player.

The next conditional statistic is the number of Pro Bowls by position group in the dataset. We see here that the Wide Receiver class from the 2001 and 2002 National Football League combines have the most Pro Bowls as a group. The high number indicates that these players are easier to judge based on their skill set and might be the safest offensive position to draft in order to make an NFL franchise better.

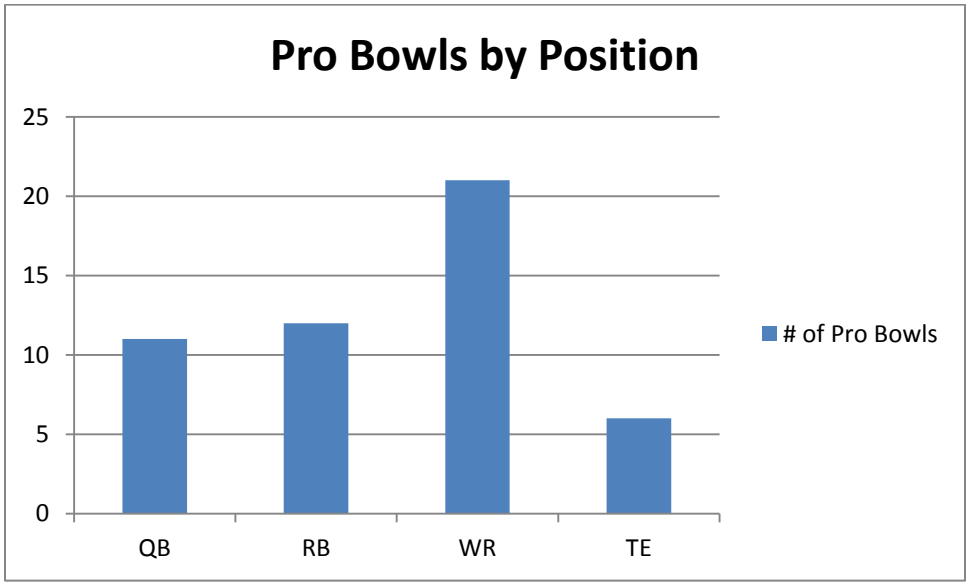


Figure 15

Figure 15 shows that Wide Receivers had the most Pro Bowls amongst them, at 21, while the next easiest group to scout were the Running Backs, having 12 Pro Bowls between them. Quarterbacks and Tight Ends have the least amount of Pro Bowls in the two combines measured, indicating that those two positions might not have as strong of correlation as the Running Back and Wide Receiver positions.

The next variable is the height variable. It seems as though the majority of success seen in the National Football League amongst Quarterbacks, Running Backs, Tight Ends, and Wide Receivers happens when a player is of average height for the position groups. We can see from

the graph below that the highest total of Pro Bowls lands right in the middle of the graph at the 73 inch mark. We can also see an overall bell shaped curve in the middle indicating that the highest probability for success lies in the 68 inch to 77 inch range. The outliers in the dataset don't seem to have enjoyed much NFL success in their career.

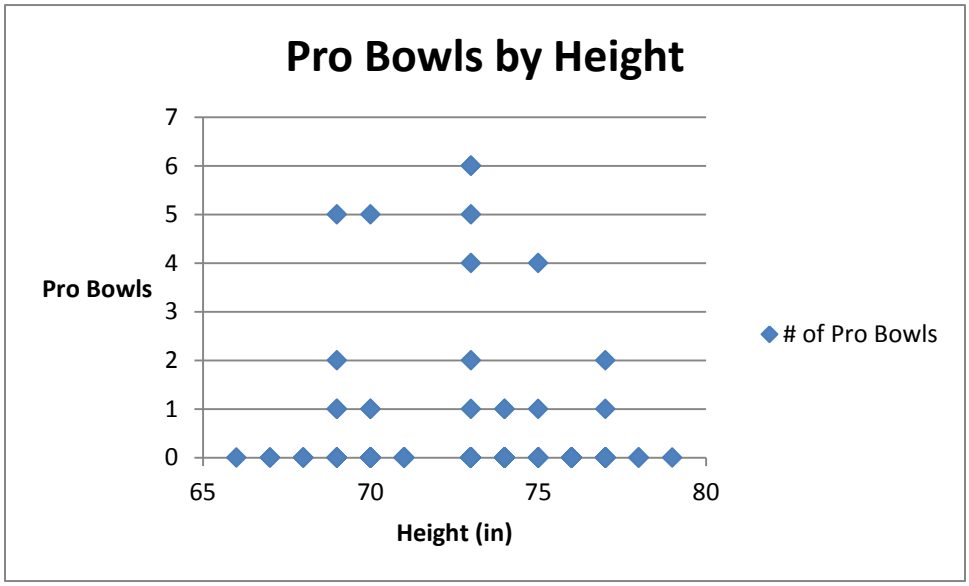


Figure 16

The next variable in the conditional statistics section is the other physical descriptor in the dataset, weight.

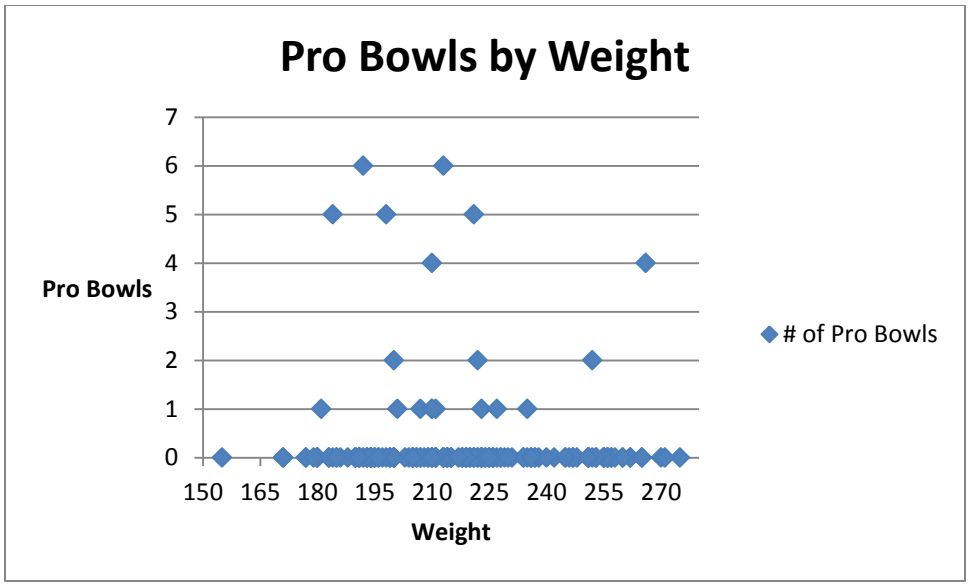


Figure 17

Figure 17 displays a similar pattern to the height variable in that the majority of success comes from the middle of the weight axis. We can see that from the range between 180 and 225 pounds there were a total of 42 Pro Bowls. As I predicted in my descriptive statistics section, a player in these position groups must be well rounded in his physical tools so that he can do all things well, instead of only one thing. This means that a player cannot be too heavy or too light, if he wants to achieve success in his career.

Looking at the height and weight variables together however, we can see an interesting phenomena taking place. In a league where players are commended for their size and stature, we see quite the opposite happening. The players that seem to be the most successful are the players of average height and weight, not the players that are outliers. The monstrous players that are thought to be the most successful in the NFL are the ones with the least amount of Pro Bowls in my dataset. Amongst players in the position groupings that I have assembled, basing success

solely on height and weight, a player that is 6'1" and weighs 200 pounds will have more success than a player that is 6'5" with a weight of 240 or a player that is 5'6" and weighs 175.

The next variable in my dataset solely pertains to the Quarterback position as they were the only ones tested. This variable is the Wonderlic Score, a measure of intelligence in the NFL.

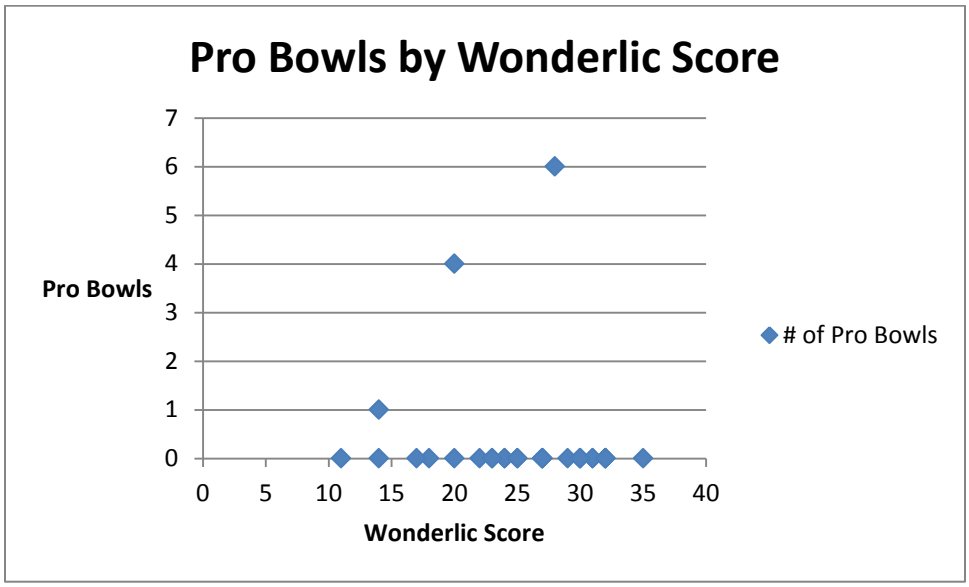


Figure 18

The correlation between success and Wonderlic score in my dataset is not clear, as the majority of players did not attend the Pro Bowl. However, amongst players that did attend the Pro Bowl we can see a clear pattern. The higher the Wonderlic score is, the more Pro Bowls that player attended. This variable will only be used in the specific position regression that will be run and not in the overall regression. It would be interesting to see if there was a correlation between Wonderlic scores and all of the position groups as we would have more observations to factor in, however time and resource constraints have made this a more difficult task than is currently possible.

The next variable in the regression is the 40 yard dash time, the chart below shows the amount of Pro Bowls broken down by 40 yard dash time.

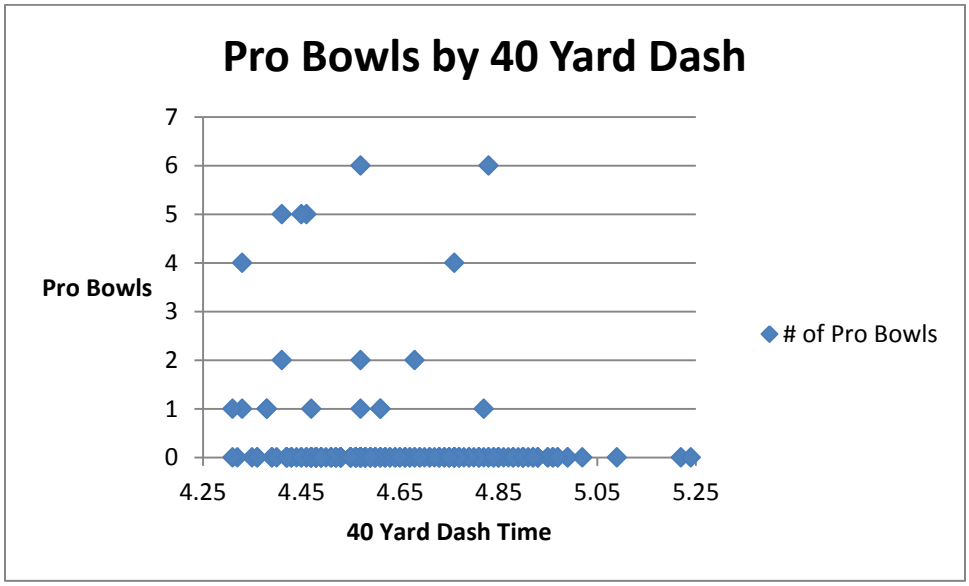


Figure 19

Figure 19 shows a good correlation between a lower 40 yard dash time and a higher amount of Pro Bowls. This was an easy variable to predict since speed is such a valued commodity in the NFL. The faster a player is, the harder it is for the defense to catch up with them and bring them to the ground. We can see from the graph that there are no Pro Bowl appearances from players that ran above a 4.85 40 yard dash time. The three individuals that ran the lowest 40 yard dash times all reached the pro bowl, and one of them reached the Pro Bowl 4 times. I expect this variable to have the highest correlation to success in my regression.

The next variable in the regression has nothing to do with how fast or quick a player is, but how strong they are. The chart below shows the relationship between Pro Bowl appearances and Bench Press in the dataset.

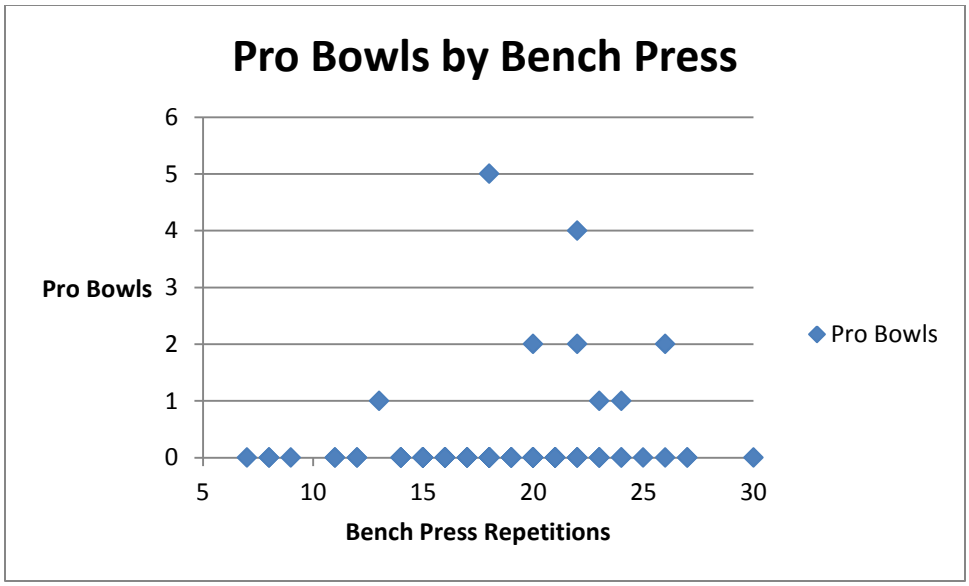


Figure 20

This variable as well as the last variable shows a relatively strong correlation between the strength of a player and the number of Pro Bowls selected to. As the Bench Press repetitions rise, so do Pro Bowl appearances. The least amount of Bench Press repetitions performed by a Pro Bowl player was thirteen, and that player went to only one Pro Bowl. The majority of Pro Bowls in the dataset come from players who performed in the 18 to 26 range. The ability to break tackles is crucial for every position and it is no surprise that we see a correlation between strength and success in the NFL.

The next conditional chart is the Pro Bowls by Vertical chart that can be seen below. This interaction of these two variables is a bit more unclear than the previous interactions, but we can still see a slight pattern emerge.

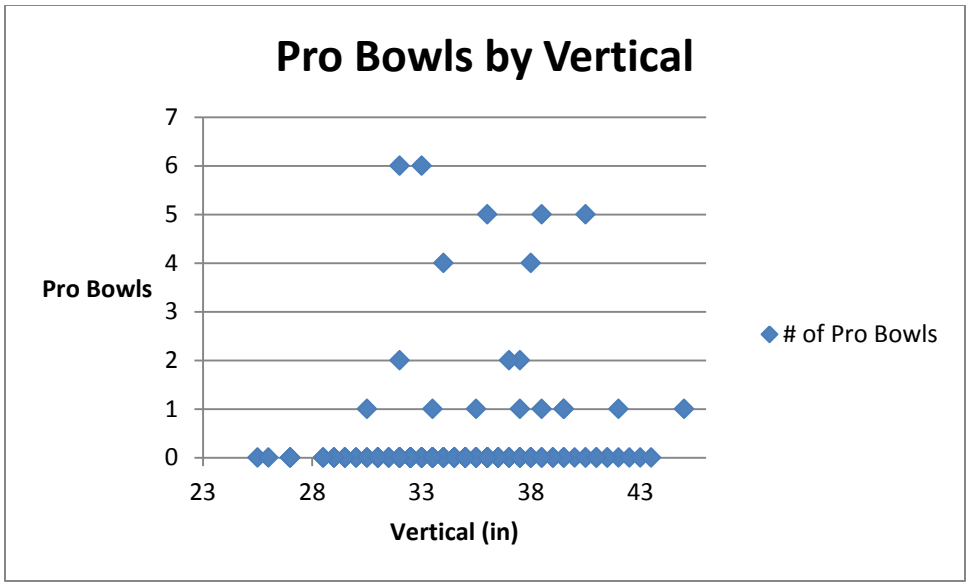


Figure 21

The majority of success achieved on this chart happens between 32 and 40 inches jumped. There are players that have reached Pro Bowls outside of this range but only three Pro Bowls total lie outside of it. The Vertical leap probably has a higher correlation of success in other sports, basketball for example, but is still a skill that can be used on the football field.

The next chart shows the amount of Pro Bowls given the players Broad Jump statistics. This chart seems to show a higher correlation between Pro Bowls and a shorter distance jumped. The chart shows that there are eighteen Pro Bowls attended by players who jumped from 100 inches to 110 inches. The past two variables seem to indicate that explosiveness does not correlate well to success in the NFL. I was expecting to find quite the opposite, as these drills are heavily scrutinized by coaches and scouts at the combine.

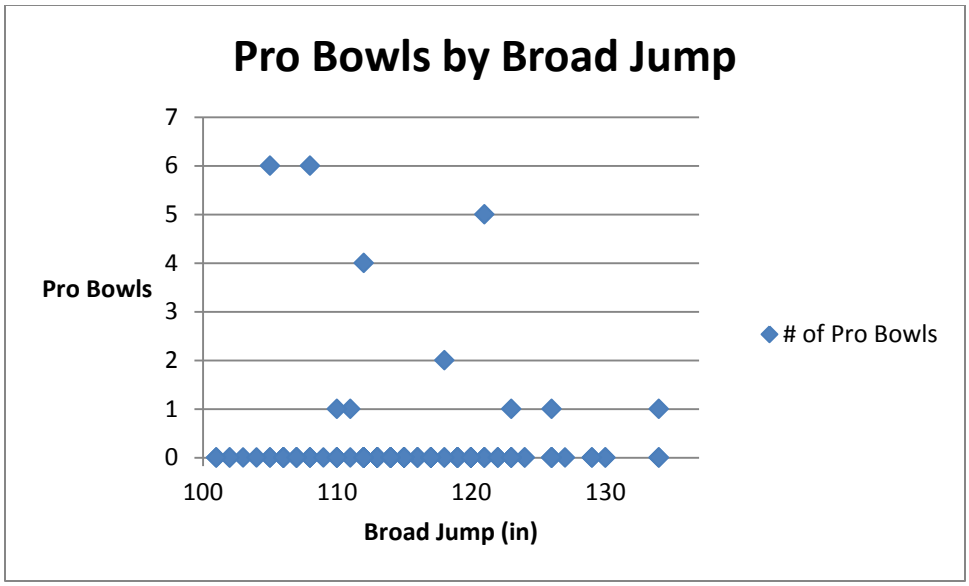


Figure 22

The next variable is the 20 Yard Shuttle, which as I mentioned earlier measures quickness in skill position players.

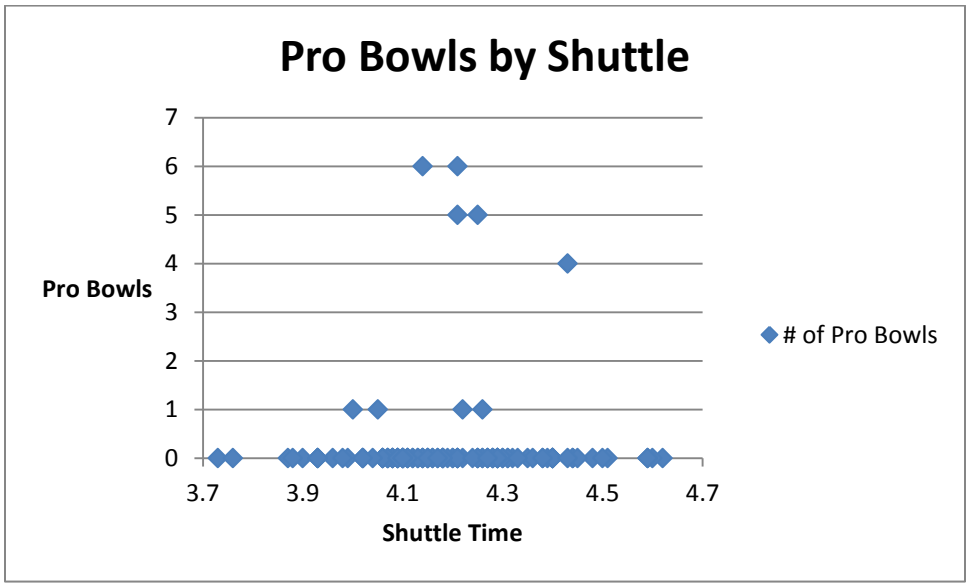


Figure 23

In this graph we can see that a shuttle time between 4.1 and 4.3 is the optimal time for success in the NFL. In between this range there are 24 total Pro Bowls, and the players in this range seem to attend the Pro Bowl more than one time. We can also see that there were no Pro Bowls attended by players that either ran higher than a 4.4 second shuttle or less than a 4 second shuttle.

The final variable in the dataset is similar to the last variable, and that is the 3 Cone Drill which measures quickness and agility. The chart below shows similar statistics to the 20 Yard Shuttle chart in the sense that the majority of Pro Bowls come from the middle of the chart. The area between 6.7 seconds and 7.5 seconds contains all of the Pro Bowls in the dataset, with outliers never attending a Pro Bowl.

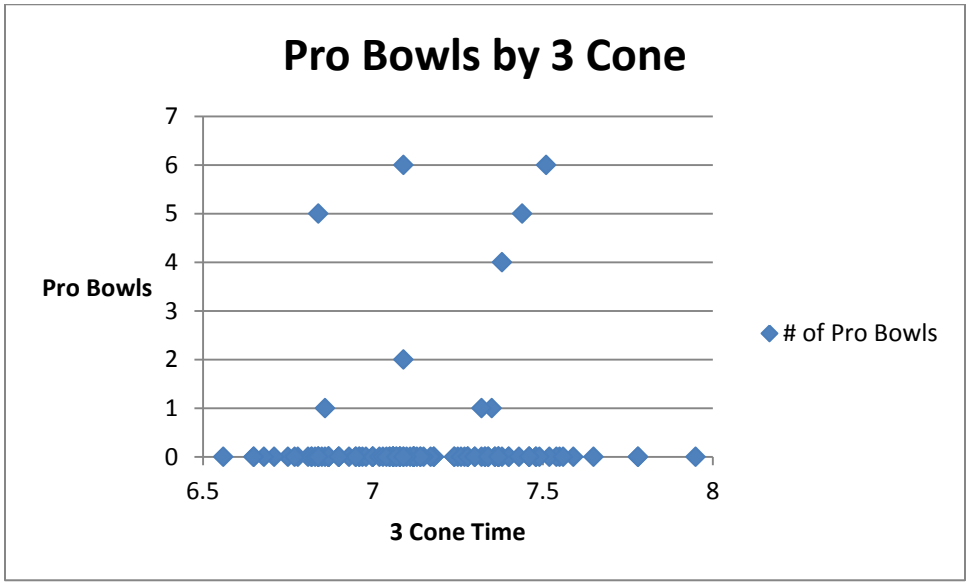


Figure 24

Conditional statistics are useful to estimate the relationship of each independent variable on the dependent variable. While useful, the charts seen above are not specific enough to determine a solid relationship. In the next section, using regression analysis we will see exactly

how important each of the variables are in predicting success in the NFL. Based on the conditional charts, 40 yard dash time looks to have the highest correlation to success at this point. I also believe that College Conference will have a high correlation amongst players in the 2001 and 2002 scouting combines.

Regression Analysis

After modeling my dataset, I have run five different regressions; one for the entire dataset, one for the Quarterback position, one for the Running Back position, one for the Wide Receiver position, and one for the Tight End position. The results are as follows.

Quarterback Analysis

A regression was run on the Quarterback position using the variables Conference, Position, Wonderlic, Height, Weight, 40 Yard Dash, Vertical, Broad Jump, Shuttle, and 3 Cone. The variables are also defined in that order on the screen shot of the regression results. X Variable 1 is Conference, X Variable 2 is Position, X Variable 3 is Wonderlic, X Variable 4 is Height, X Variable 5 is Weight, X Variable 6 is 40 Yard Dash, X variable 7 is Vertical, X Variable 8 is Broad Jump, X Variable 9 is Shuttle, X Variable 10 is 3 Cone.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	62.51643486	29.83130198	2.095666	0.049739	0.0788022	124.95407	0.078802238	124.9540675
X Variable 1	-0.086029309	0.676369049	-0.1271	0.900198	-1.502733	1.3306739	-1.50273251	1.330673893
X Variable 2	0	0	65535	#NUM!	0	0	0	0
X Variable 3	0.043827401	0.054155159	0.809293	#NUM!	-0.069527	0.157175	-0.069520165	0.157175452
X Variable 4	-0.437077492	0.28527898	-1.53211	0.141978	-1.034173	0.1600183	-1.03417326	0.160018276
X Variable 5	-0.016548396	0.029793968	-0.55543	0.585083	-0.078908	0.0458111	-0.07890789	0.045811095
X Variable 6	2.298462444	1.980378477	1.15032	0.260239	6.444489	1.8475639	6.44448875	1.847563858
X Variable 7	-0.000409968	0.101447517	-0.00404	0.996818	-0.212742	0.2119221	-0.21274206	0.211922125
X Variable 8	-0.092675341	0.060596578	-1.52938	0.142618	-0.219505	0.0341518	-0.21950541	0.034154756
X Variable 9	-0.318526779	3.296245581	-0.09663	0.92403	-7.217648	6.5805945	-7.21764807	6.580594512
X Variable 10	-0.603158068	2.564524803	-0.23518	0.816532	-5.970979	4.7646633	-5.97097947	4.764563335

Table 1

As we can see from the Table 1, the two variables that have the most effect on whether or not a Quarterback reaches a Pro Bowl and has success throughout his career are 40 yard dash with a coefficient of 2.29 and 3 cone which has a coefficient of negative .6. We also see that none of the variables have a P-value of less than .05 which means that none of the variables are significant contributors to Quarterback success in the NFL. This means that the quarterback position cannot be evaluated efficiently based on combine numbers. Coaches, scouts, general managers, and other coaches must add other factors into their evaluation of whether or not a quarterback will be successful in the NFL.

Running Back Regression

A regression was run on the Running Backs from the two combines as well. The variables used in the regression were Position, Height, Weight, 40 Yard Dash, Vertical, Broad Jump, Shuttle, 3 Cone, and Bench Press. The results are as follows with the variables defined in the order they appear above. X Variable 1 is Position, X Variable 2 is Height, X Variable 3 is Weight, X Variable 4 is 40 Yard Dash, X Variable 5 is Vertical, X Variable 6 is Broad Jump, X variable 7 is Shuttle, X Variable 8 is 3 Cone, and X Variable 9 is Bench Press.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	17.87732937	7.659709021	2.333944	0.024457	2.41941074	33.335248	2.419410742	33.33524799
X Variable	0	0	65535	#NUM!	0	0	0	0
X Variable	-0.118856499	0.059110985	-2.01073	#NUM!	-0.2381473	0.0004343	-0.2381473	0.000434299
X Variable	0.010820392	0.009038126	1.197194	0.237944	-0.0074193	0.0290601	-0.00741928	0.029060068
X Variable	-1.723528191	1.053960686	-1.63529	0.109463	-3.850507	0.4034506	-3.85050697	0.403450584
X Variable	0.141922326	0.045871882	3.093885	0.003507	0.04934912	0.2344955	0.04934912	0.234495532
X Variable	-0.052241126	0.032354872	-1.61463	0.11388	-0.1175359	0.0130537	-0.1175359	0.01305365
X Variable	4.034014406	1.434800589	2.811551	0.007462	1.13846959	6.9295592	1.13846959	6.929559223
X Variable	-2.716689524	1.020048384	-2.66329	0.010925	-4.7752305	-0.658149	-4.7752305	-0.65814854
X Variable	-0.007635693	0.021767703	-0.35078	0.727506	-0.0515647	0.0362933	-0.0515647	0.036293311

Table 2

As we can see from the Table 2, the two variables that influence whether or not a Running Back will have success in the NFL are the 20 Yard Shuttle with a coefficient of 4.03 and the 3 Cone Drill with a coefficient of negative 2.71. We also see that three of the variables, Vertical, Shuttle, and 3 Cone are statistically significant and have a P-value of less than .05. This means that these three variables are good predictors of success for Running Backs in the NFL. These numbers make sense in the sense that they measure explosiveness and quickness in an athlete, which are things that Running Backs need in order to produce at the professional level.

Wide Receiver Regression

The Wide Receivers were put into a regression from the combines in 2001 and 2002. The variables used in the Wide Receiver regression were Position, Height, Weight, 40 Yard Dash, Vertical, Broad Jump, Shuttle, and 3 Cone Drill. The variables are listed in that order in the regression screen shot below. X Variable 1 is Position, X Variable 2 is Height, X Variable 3 is Weight, X Variable 4 is 40 Yard Dash, X Variable 5 is Vertical, X Variable 6 is Broad Jump, X variable 7 is Shuttle, and X Variable 8 is 3 Cone.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.6896024	10.5489476	0.065372	0.948129	-20.47841	21.857614	-20.4784096	21.8576144
X Variable 1	0	0	65535	#NUM!	0	0	0	0
X Variable 2	-0.0549352	0.093121329	-0.58993	#NUM!	-0.2417968	0.1319264	-0.2417968	0.13192643
X Variable 3	0.005931	0.014270358	0.41562	0.679399	-0.0227045	0.0345666	-0.02270453	0.03456661
X Variable 4	-1.8879083	1.759200226	-1.07316	0.288153	-5.4180018	1.6421852	-5.41800178	1.64218525
X Variable 5	0.052753	0.083214211	0.633942	0.528897	-0.1142285	0.2197345	-0.11422853	0.21973453
X Variable 6	-0.0541609	0.040568888	-1.33503	0.187681	-0.1355683	0.0272466	-0.13556831	0.02724655
X Variable 7	-1.644838	1.698586485	-0.96836	0.337349	-5.0533011	1.7636252	-5.05330113	1.76362516
X Variable 8	3.1663194	0.961052526	3.294637	0.001778	1.23782644	5.0948124	1.237826438	5.0948124

Table 3

Table 3 indicates that the two variables that influence Wide Receiver success in the NFL the most are the 40 Yard Dash with a coefficient of -1.88 and the 3 Cone Drill with a coefficient of 3.16. In the Wide Receiver regression results there is only one variable that shows to be statistically significant at the 95% level. This variable is the 3 Cone Drill with a P-value of .0017. This variable tests quickness and agility which is useful for Wide Receivers as they often cut in and out while running routes. Having a fast time in this drill indicates that a Receiver is quick and agile enough to create space between himself and the Defensive Back who is trying to cover him. The Receivers who produce most consistently in the National Football League have the ability to run away from defenders and create a zone for the Quarterback to get them the ball.

Tight End Regression

A regression was also run for the Tight End position. The variables used in the Tight End regression were Position, Height, Weight, 40 Yard Dash, Vertical, Broad Jump, Shuttle, 3 Cone, and Bench Press. These variables are displayed in the same order in the regression screen shot below. X Variable 1 is Position, X Variable 2 is Height, X Variable 3 is Weight, X Variable 4 is 40 Yard Dash, X Variable 5 is Vertical, X Variable 6 is Broad Jump, X variable 7 is Shuttle, X Variable 8 is 3 Cone, and X Variable 9 is Bench Press.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-5.645186555	12.49224026	-0.451895452	0.654706	-31.19468662	19.90431351	-31.19468662	19.90431351
X Variable 1	0	0	65535	#NUM!	0	0	0	0
X Variable 2	-0.032930577	0.115312265	-0.285577404	#NUM!	-0.26877064	0.202909486	-0.26877064	0.202909486
X Variable 3	0.017036063	0.014113033	1.207115697	0.237142	-0.011828329	0.045900456	-0.011828329	0.045900456
X Variable 4	-1.287455154	1.564625087	-0.822852174	0.417307	-4.48747276	1.912562452	-4.48747276	1.912562452
X Variable 5	-0.031106913	0.047350513	-0.656949861	0.516393	-0.127949585	0.06573576	-0.127949585	0.06573576
X Variable 6	0.027717002	0.03627193	0.764144693	0.450954	-0.046467423	0.101901428	-0.046467423	0.101901428
X Variable 7	1.749344072	1.626359364	1.075619639	0.290965	-1.576934308	5.075622452	-1.576934308	5.075622452
X Variable 8	-0.02132322	0.882193967	-0.024170671	0.980882	-1.825612471	1.782966031	-1.825612471	1.782966031
X Variable 9	0.043522246	0.0420391	1.035280161	0.309094	-0.042457367	0.129501859	-0.042457367	0.129501859

Table 4

In the Tight End regression results the two variables with the most effect on Tight End success in the NFL are 40 Yard Dash with a coefficient of -1.28 and the 20 Yard Shuttle with a coefficient of 1.74. While these are the two variables with the most effect on the dependent variable, neither these nor any other variable were statistically significant at the 95% level. This means that other factors will have to be used when evaluating a Tight End at the NFL combine.

Offensive Skill Position Regression

All of the players in the dataset were placed into a regression equation and the results are shown below. The variables used in this regression were Conference, QB, WR, RB, TE, Height, Weight, 40 Yard Dash, Vertical, Broad Jump, Shuttle, and 3 Cone Drill. The QB, WR, RB, TE, variables were created as dummy variables in order to be able to put Position into the regression.

X Variable 1 is Conference, X Variable 2 is QB, X Variable 3 is WR, X Variable 4 is RB, X Variable 5 is TE, X Variable 6 is Height, X variable 7 is Weight, X Variable 8 is 40 Yard Dash, X Variable 9 is Vertical, X Variable 10 is Broad Jump, X Variable 11 is Shuttle, and X Variable 12 is 3 Cone.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	8.846052711	5.496492241	1.609399654	0.109384	-2.004114807	19.69622023	-2.004114807	19.69622023
X Variable 1	0.189297859	0.163443869	1.158182683	0.248414	-0.13334307	0.511938788	-0.13334307	0.511938788
X Variable 2	0.248977191	0.307619513	0.809367354	0.419435	-0.358268867	0.856223249	-0.358268867	0.856223249
X Variable 3	-0.298504829	0.256134901	-1.165420362	0.245481	-0.804119404	0.207109745	-0.804119404	0.207109745
X Variable 4	0	0	65535	#NUM!	0	0	0	0
X Variable 5	-0.028190438	0.417622577	-0.06750219	#NUM!	-0.852584382	0.796203506	-0.852584382	0.796203506
X Variable 6	-0.056019961	0.048418604	-1.156992495	0.248899	-0.151599095	0.039559172	-0.151599095	0.039559172
X Variable 7	0.006685398	0.00696882	0.959329925	0.338755	-0.00707117	0.020441965	-0.00707117	0.020441965
X Variable 8	-1.870945642	0.734732518	-2.546430974	0.011771	-3.321319908	-0.420571375	-3.321319908	-0.420571375
X Variable 9	0.059804311	0.031490608	1.899115777	0.059243	-0.002358676	0.121967297	-0.002358676	0.121967297
X Variable 10	-0.042735205	0.01873127	-2.281490036	0.02376	-0.079711044	-0.005759366	-0.079711044	-0.005759366
X Variable 11	-0.071726213	0.876424725	-0.081839559	0.934871	-1.801803262	1.658350836	-1.801803262	1.658350836
X Variable 12	0.81683843	0.558631935	1.462212199	0.145529	-0.285910343	1.919587203	-0.285910343	1.919587203

Table 5

As we can see from Table 5, the two variables that have the most impact on the success of an offensive skill player in the NFL are variables 3 and 8, which are the Wide Receiver variable with a coefficient of negative .29 and 40 Yard Dash with a coefficient of -1.87. There are two variables in the dataset that are statistically significant at a 95% confidence interval and these two variables are 40 Yard Dash which has a P-value of .011 and the Broad Jump with a P-value of .023. These two variables can be considered safe to judge a player on at the combine with that player's success being in consideration. With this regression we can predict a players success in the NFL based on the combine statistics from 2001 and 2002, taking into consideration Pro Bowl appearances of the players who attended the combine.

Taking the regression equation .189a+.248b-.298c+0d-.028e-.056f+.006g-1.87h+.059i-.042j-.071k+.816l+E we can then predict a player's relative success throughout his career in the NFL. A Quarterback from a Major Conference that is 6'2", 220 lbs, that runs a 4.5 second 40 yard dash, has a vertical of 32", broad jump of 120", runs a 4 second shuttle and a 7.5 second 3 Cone drill is predicted to have a success rating of -7.03. The prediction interval for my regression has a standard error of 1.07967, which means that it is accurate on average up to an amount of 1.07967.

List of Works Cited

1. Pro-Football-Reference.com n.pag. Pro Football Reference. Web. 9 Nov 2012. <http://www.pro-football-reference.com/>.

2. "NFL Combine Data." n.pag. NFL Combine Results. Web. 9 Nov 2012. <http://nflcombineresults.com/>.

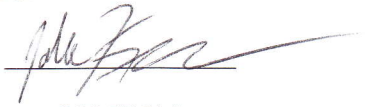
UNIVERSITY OF ARIZONA

Submission of Thesis to the Honors College

PREDICTING OFFENSIVE SUCCESS IN THE NATIONAL FOOTBALL LEAGUE

BY

JACOB ALLEN KRAUSE

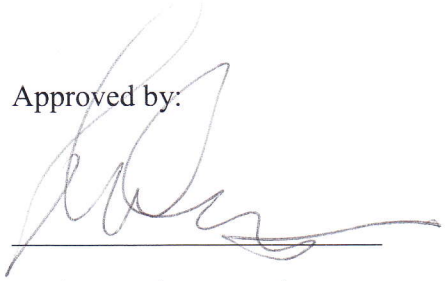


23057514

Honors Thesis

Fall 2012

Approved by:



Professor Thomas Dalton

Department of Economics

Jacob Allen Krause

Honors Thesis

December 5, 2012

Abstract

Predicting Offensive Success in the NFL

This thesis was created in order to find correlations in the National Football League between player's physical abilities and the success that those players have in the NFL. Economic incentives in this business push teams towards competitive greatness in order to stay relevant. The regression model that I have set up could be used in scouting and evaluating players in order to predict which players will have the most success in the National Football League. The Thesis combines knowledge about the game itself, data management, statistics, and economic theory in order to produce the desired results.