Modeling NFL Quarterback Success with College Data

by

Lewis Jones

(Under the Direction of Cheolwoo Park)

Abstract

Sports analysts often say that the quarterback is the most important and impactful position in American professional sports. Because of this popular opinion, each season a disproportionately large number of college quarterbacks are drafted into the National Football League (NFL), making at a minimum $435,000, but most making substantially more. However, of the 21 quarterbacks drafted in the first 2 rounds in the past 5 NFL drafts, only 12 threw 100 or more passes in the 2015 NFL season, suggesting that a large number of these players were drafted earlier than their talent warranted. Although finding a great quarterback can arguably help a team more than any other position, other positions can be safer bets in the draft and drafting a quarterback too early and overpaying said quarterback can set an NFL franchise back for years. In this thesis we will present a series of models, each modeling a different element of information relevant to quarterback success in the NFL in an attempt to explain what NFL teams are currently looking for in quarterback prospects and what may be a better way of judging a quarterback prospect.

Index words: Quarterback, National Football League, NFL Draft, Tobit Regression

Modeling NFL Quarterback Success with College Data

by

Lewis Jones

B.S., University of Georgia, 2014

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Science

Athens, Georgia

2016

Modeling NFL Quarterback Success with College Data

by

Lewis Jones

Approved:

Major Professors: Cheolwoo Park

Committee: Paul Schliekelman
Lynne Seymour

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

Modeling NFL Quarterback Success with College Data

Lewis Jones

April 26, 2016

Acknowledgments

First and foremost, I want and need to thank my parents for all the love, support and that one time you tried to help me with my math homework. Thank you for being the cool parents, but also for getting me a drum set for Christmas instead of that Nintendo 64 I wanted. I get it now. Thank you, Spencer for the patience, love, patience, support and patience. Thank you, Cheolwoo for all of the support, advice and for being the calming voice through this entire process. Thank you, Jonathan. I've learned so much from you in the last two years and I hope to continue to learn from you in years to come. Thank you, Dillon and Wes. Go Dawgs! Thank you, Ted. Thank you, Gammy. Thank you, Yvonne and Eric, Colin and Julia. Thank you, Sal. Thank you, friends and family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Each spring the National Football League holds a draft where the top college prospects find out for which team they will start their professional careers. The NFL draft has become a national spectacle, shown on prime time television featuring soon-to-be millionaires sporting Armani suits. The center of this event is the quarterbacks. Sports writers and analysts overwhelmingly agree that the NFL quarterback is the most important position in sports - not just in football, not just in professional sports, but in all team sports. Because of this idea, teams are willing to use top draft picks (and willing to wager millions of dollars in doing so) on selecting a quarterback early in the draft, and in the recent years this has paid off on occasion. In 2007 the Atlanta Falcons finished 4-12 and selected Matt Ryan, a quarterback from Boston College third overall in the 2008 draft. In Ryan's rookie campaign the Falcons made the playoffs with an 11-5 record. A few years later, in 2012, the Indianapolis Colts took Andrew Luck first overall after earning the first selection by going 2-14 in the 2011 season. In 2012 the Colts also made the playoffs at 11-5. The idea that a quarterback is the most important position on the field has been put on display many times, like in the

Falcons' and Colts's case. This has led to quarterbacks being drafted at disproportionately high rate, as shown in Table 1.1. (NFL, 2015)

Table 1.1: Quarterbacks Drafted By Round from 1998-2012

Round	1st	2nd	3rd	4th	5th	6th	7th	Undrafted
Selections	43	17	19	14	26	25	29	25

There are 22 starters on every football team and only one starting quarterback, however over the past ten years 26 of the 320 first round picks have been quarterbacks, meaning quarterbacks are being drafted in the first round at a rate over 75% higher than expected, including 6 quarterbacks selected first overall in the past 10 NFL drafts. However, only 18 of those 26 first rounders are still on an active roster in the NFL. Tables 1.2 and 1.3 show that from the 1986 to the 2010 season, quarterbacks selected in the first round have worst success in making the NFL All-Pro team and the second highest "bust rate" out of any position in the NFL (a player is considered a "bust" when he is drafted in the first round and fails to play at least 50 games in the NFL) (Gagnon, 2015).

Table 1.2: All-Pro Rate for First Round Draft Picks 1986-2010

Position	Drafted	All-Pro	All-Pro Rate
Linebacker	87	23	26.4%
Safety	35	9	25.7%
Center	12	3	25.0%
Guard	32	7	21.9%
Offensive Tackle	84	16	19.0%
Defensive Line	170	28	16.5%
Running Back	96	15	15.6%
Cornerback	91	14	15.4%
Wide Receiver	99	15	15.2%
Tight End	28	4	14.3%
Quarterback	57	3	5.3%

Table 1.3: Bust Rate for First Round Draft Picks 1986-2010

Position	Drafted	Busts	Bust Rate
Running Back	96	45	46.9%
Quarterback	57	23	40.4%
Defensive Line	170	60	35.3%
Wide Receiver	99	30	30.3%
Cornerback	91	24	26.4%
Guard	32	7	21.9%
Linebacker	87	19	21.8%
Tight End	28	6	21.4%
Offensive Tackle	84	16	19.0%
Safety	35	5	14.3%
Center	12	1	8.3%

Because of the drastic improvement caused by adding a quarterback to a roster, NFL organizations continuously reach for quarterbacks in the draft, selecting quarterbacks well before their skills suggest they should be drafted. However, this decision often backfires, leaving franchises with a sub-par, but well-paid quarterback who handcuffs the team from success. While drafting a young quarterback has instantly taken teams from being bottom-feeders to playoff teams in the cases of Andrew Luck and Matt Ryan, selecting a quarterback in the first round is not the only way to find production at the quarterback position. Production can be found outside of the first round. For example, in the 2015 Super Bowl XLIX, Tom Brady beat Russell Wilson, a sixth and third round draft pick, respectively. Figure 1.1 below shows that first round draft picks generally play more games in the NFL than any other round. While this may come as no surprise, drafting 22% of quarterbacks in the first round suggests the talent pool may steeply decline after the first round (Pro-Football-Reference, 2015).

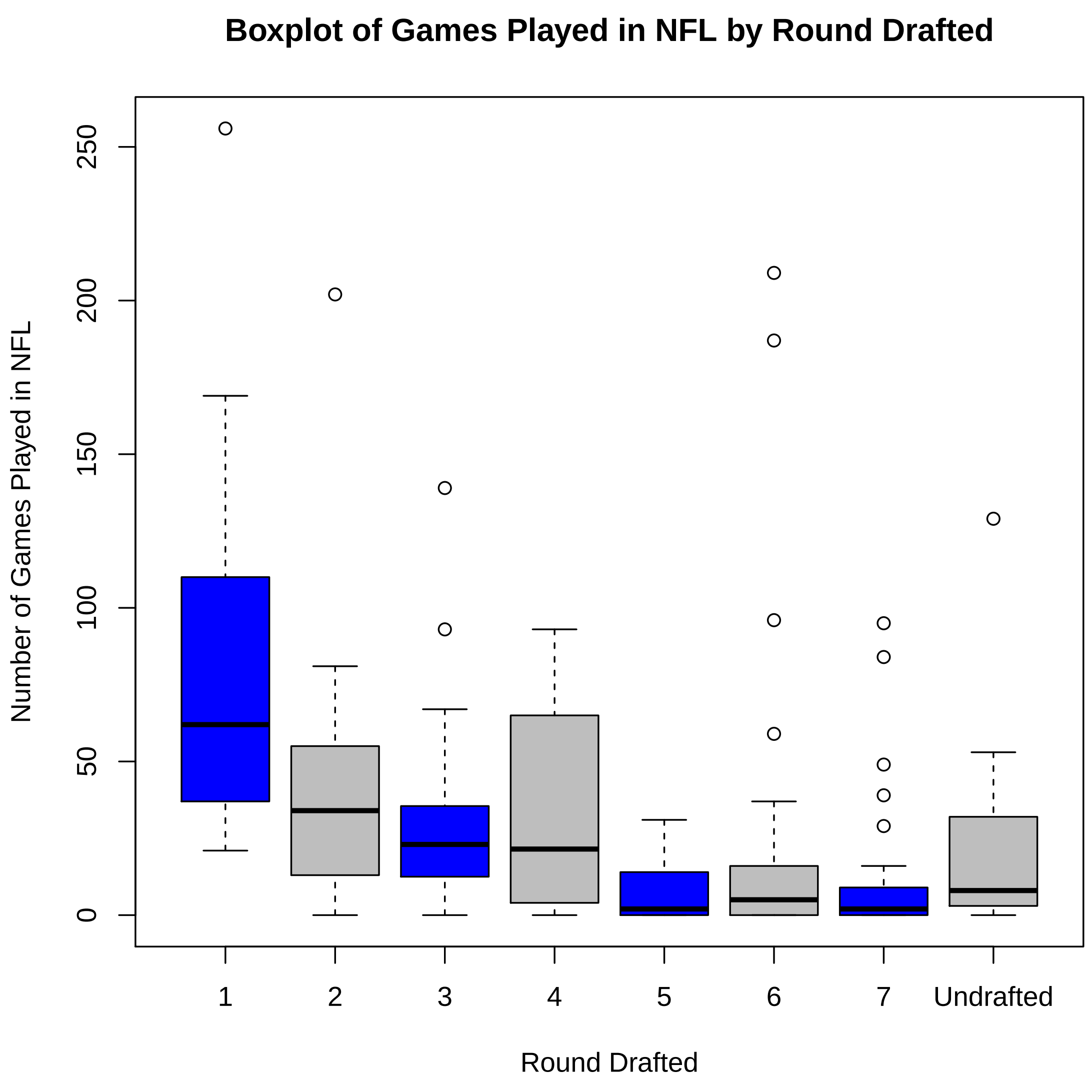

Figure 1.1: Boxplot of Games Played in the NFL by Round Drafted

Figure 1.2 shows the median Quarterback Rating (a calculated metric for quarterback play) is highest for first round draft picks, but only 7.6% higher than quarterbacks drafted in the second round. The median Quarterback Rating for first round draft picks is only 14.3% higher than quarterbacks who failed to even get drafted.

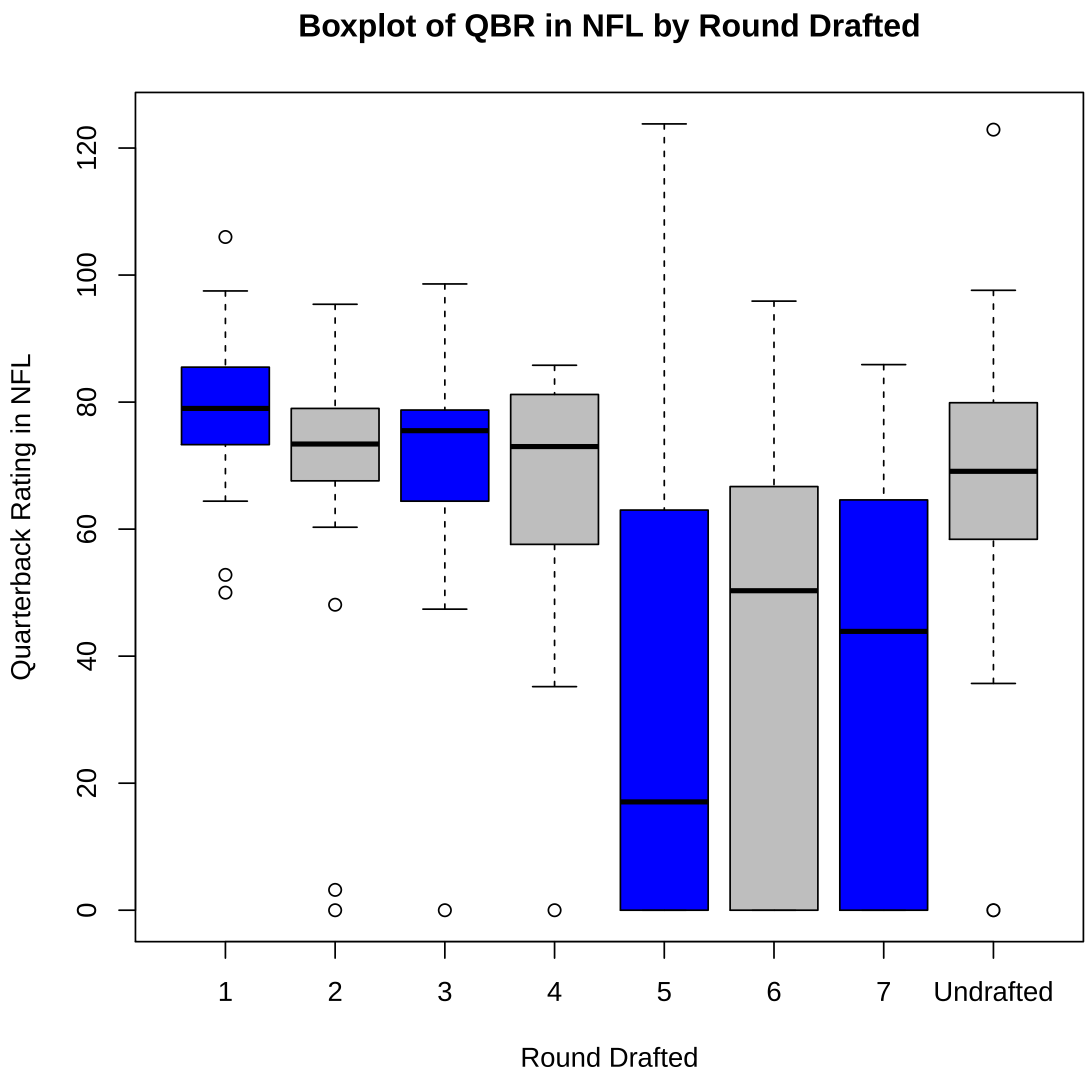

Figure 1.2: Boxplot of Quarterback Rating in the NFL by Round Drafted

As mentioned previously with Andrew Luck and Matt Ryan, sometimes organizations make brilliant first round decisions. The year after Tom Brady and Russell Wilson faced off against each other in Super Bowl XLIX, Peyton Manning and Cam Newton, both selected first overall, started at quarterback for their respective teams in Super Bowl 50. However, the lack of success from first round quarterbacks is difficult to ignore, especially when teams have found Hall of Fame caliber quarterbacks in later rounds and there are other positions, like Safety and the Offensive Line (Center, Guard and Offensive Tackle), which all have above average All-Pro Rates and below average Bust Rates.

Owning an exceptional quarterback in the NFL is a top priority for every team, which is why the NFL draft is so important. However, teams feel such a pressure to draft the perfect quarterback that they will use a top pick on a quarterback whose talent suggests he should be selected much later in the draft. Some teams in the NFL have realized that quarterback performance is very difficult to predict and a more reasonable strategy may be to use their first few picks filling other needs and draft a quarterback in a later round. After all, quarterbacks drafted in later rounds often find more success than those drafted in the first round as shown in Figures 1.1 and 1.2. In 2012, Washington employed both of these strategies. In the same draft the Indianapolis Colts selected Andrew Luck first overall, Washington traded 5 draft picks to the St. Louis Rams for the rights to the second pick in the 2012 draft. Washington used the pick to select Robert Griffin III, a Heisman Trophy winning quarterback from Baylor. In the fourth round of the same draft Washington also selected Kirk Cousins, a quarterback from Michigan State. In the first 4 years of the duo's career, Griffin has started 35 games and thrown 40 touchdowns while Cousins has started 25 games and thrown 47 touchdowns.

So why do teams continue to draft quarterbacks higher than their talent warrants? Are teams being fooled by some statistic or metric? Is the "eye test" erroneously leading franchises to search for a specific body type? This thesis will attempt to answer these questions while also attempting to understand the traits that actually lead to success in the NFL.

1.2 Objective

The data include 198 quarterbacks, whose careers began between the 1998 season and the 2012 season. Possible predictor variables are Height, Weight and about 40 other possible predictors from each player's college career. These variables include metrics such as Touchdowns per Game, Rushing Yards per Attempt, Maximum Number of Passing Yards in a Single Season, etc. Since this thesis is asking two different questions: 1) which characteristics do NFL teams value when drafting a quarterback and 2) which characteristics lead to success in the NFL, five different models will be fit. First, a model modeling when a quarterback will be selected in the draft is built. Then, four other models modeling different benchmarks of NFL success are fit and the variables selected in each of the success models will be compared to the variables selected in the model predicting when quarterbacks will be selected in the draft. The different success models are introduced in more detail in Chapters 3 and 4.

This thesis is inspired by the work of Bill James, specifically some of Mr. James' ideas introduced in Michael Lewis' *Moneyball* (Lewis, 2003). Similar to the ideas explored in *Moneyball*, the goal of this thesis is to improve on the way prospected players are evaluated. In *Moneyball*, the Oakland Athletics, a Major League Baseball team, are attempting to build a playoff-caliber team with very little money in a league with no salary cap, hence the subtitle of *Moneyball*: *"The Art of Winning an Unfair Game"*. In the NFL there is a salary cap, but unlike baseball where no one player can lead a team for an entire season, the NFL has the quarterback position. A few quarterbacks outside of what sports analysts may consider the "top tier" have won Super Bowls, but it is difficult to imagine a team winning a Super Bowl with sub-par quarterback play.

Besides *Moneyball*, 3 other papers inspired this thesis: Addona and Wolfson (2011), Berry and White (2002) and Stimel (2009). While these papers analyze quarterback play, this thesis is also an attempt to compare the traits NFL teams think are desirable with the traits correlated with success in the NFL. This thesis proceeds as follows: Chapter 2 describes more specifically from where the data were collected and a few descriptive statistics and some of the tools used in the analysis. Chapter 3 explores the methods used in the analysis and some of the model types and the metrics used to evaluate models. Chapter 4 presents the analysis of the models and Chapter 5 examines the different models and concludes the thesis.

Chapter 2

Preliminary Data Exploration

2.1 Data Collection

Regular season college statistics and all NFL passing and rushing statistics were collected from www.totalfootballstats.com. Draft results were collected from www.NFL.com. Height and Weight were collected from www.pro-football-reference.com, although some players were not listed. If players were not listed on Pro-Football-Reference the data were gathered from Wikipedia. Playoff football stats were also gathered from Pro-Football-Reference. A data dictionary for all possible predictor variables can be found in Table 2.1.

2.2 Description of the Predictor Variables

All explanatory variables are observed from the quarterbacks' college careers. Variables described as averages are aggregated over the player's entire college career. "Maximum" variables are decided by calculating the metric for each year the quarterback played in college and selecting the maximum value. For instance, if a quarterback played four seasons of college football and threw 10 touchdowns his freshman season, 12 his sophomore, 27 his junior and 24 his senior season, his value for the variable "Maximum Passing Touchdowns in a Single Season" would be 27. This is to take into account that some players took a little while to mature to their full potential, thus bringing their overall averages down. Finally, variables described as "Total" are measured over the quarterback's entire college career. These are similar to the "Average" variables, except they are measured as a count and not as an average.

A few variables used in the model selection performed in this thesis involve Quarterback Rating. Quarterback Rating is a metric commonly used to evaluate quarterback play. Because success as a quarterback is largely affected by the rest of the team and by the offensive system the team runs, it is unfair to judge a quarterback strictly by one statistic. The Quarterback Rating (QBR) was developed to combine four commonly metrics into one. Quarterback Rating "is a linear combination of four categories: completion percentage, (average) yards per pass, (average) touchdowns per pass, and (average) interceptions per pass" (Berry and White, 2002):

$$QBR = (\tfrac{100}{6}) * (\tfrac{\text{Completion Percentage}-0.3}{0.2} + \tfrac{\text{Yards per Pass Attempt}-3}{4} + \tfrac{\text{Touchdown Percentage}}{0.05} + \tfrac{0.095-\text{Interception Percentage}}{0.04})$$

If any of the four parts are outside of the range [0,2.375], the value is truncated to the respective bound.

Table 2.1: Description of Possible College Predictor Variables

	Variable	Note
1	Average Pass Attempts per Game	
2	Average Pass Completions per Game	
3	Average Quarterback Rating	
4	Average Rushing Attempts per Game	
5	Average Rushing Yards per Game	
6	Average Interceptions per Game	
7	Average Passing Touchdowns per Game	
8	Average Passing Yards per Attempt	
9	Average Passing Yards per Completion	
10	Average Passing Yards per Game	
11	Completion Percentage	
12	Interception Percentage	Interceptions divided by Pass Attempts
13	Passing Touchdown Percentage	Touchdowns divided by Pass Attempts
14	Rushing Touchdown Percentage	Touchdowns divided by Rush Attempts
15	Maximum Completion Percentage	
16	Maximum Interception Percentage	
17	Maximum Interceptions	
18	Maximum Pass Attempts	
19	Maximum Pass Completions	
20	Maximum Passing Touchdowns	
21	Maximum Passing Yards per Attempt	
22	Maximum Passing Yards per Game	
23	Maximum Quarterback Rating	
24	Maximum Rushing Attempts	
25	Maximum Rushing Touchdowns	
26	Maximum Rushing Yards per Carry	
27	Maximum Rushing Yards per Game	
28	Maximum Rushing Yards	
29	Maximum Passing Touchdown Percentage	
30	Total Number of Passing Attempts	
31	Total Number of Passing Completions	
32	Total Number of Passing Touchdowns	
33	Total Number of Passing Yards	
34	Total Number of Rushing Attempts	
35	Total Number of Rushing Touchdowns	
36	Total Number of Rushing Yards	
37	Total Number of Games Played in College Career	
38	Total Number of Interceptions	
39	Height (In Inches)	
40	Weight (In pounds)	
41	Proportion of Pass Plays	

2.3 Distributions of Target Variables

Draft Selection

The first model fit modeled Draft Selection. The distribution of the target variable is shown below in Figure 2.1. All undrafted quarterbacks had a censored target, so their draft selection value was imputed with 254 because the maximum Draft Selection for any drafted quarterback was 253. Figure 2.1 shows a bimodal distribution, censored at 254, so a Tobit model was fit. Tobit Regression will be introduced in greater detail in Section 3.1.2.

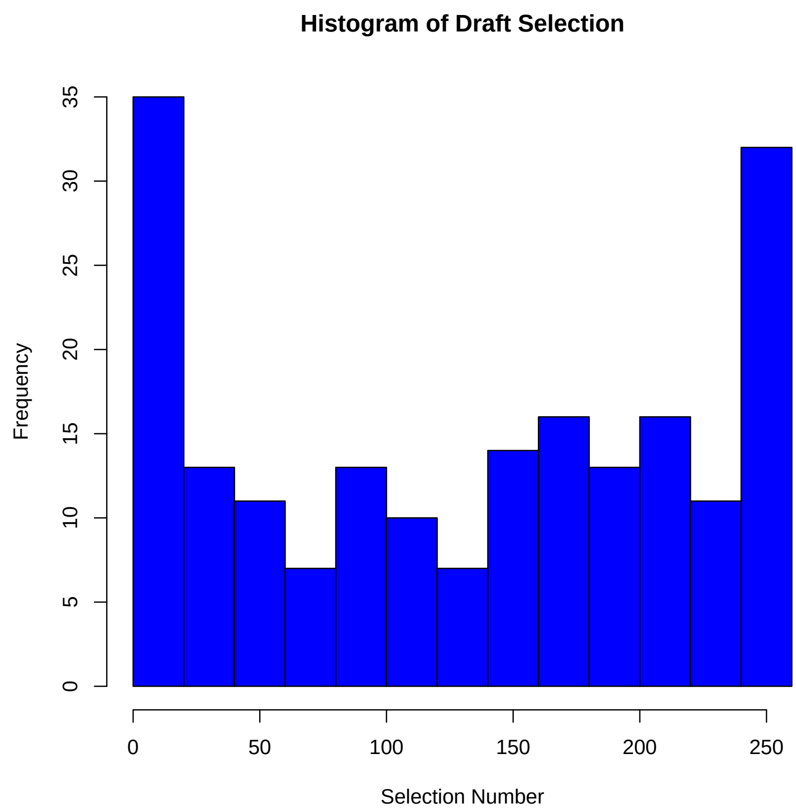

Figure 2.1: Distribution of Selection

Much of the inspiration of this thesis came from the volume at which quarterbacks are selected early in the first round of the NFL Draft. Below, Figure 2.2 shows the distribution of the draft selection spot of quarterbacks drafted in the first round. This distribution will not be modeled by itself, the distribution plot is just for a better understanding of the volume of quarterbacks drafted early in the first round.

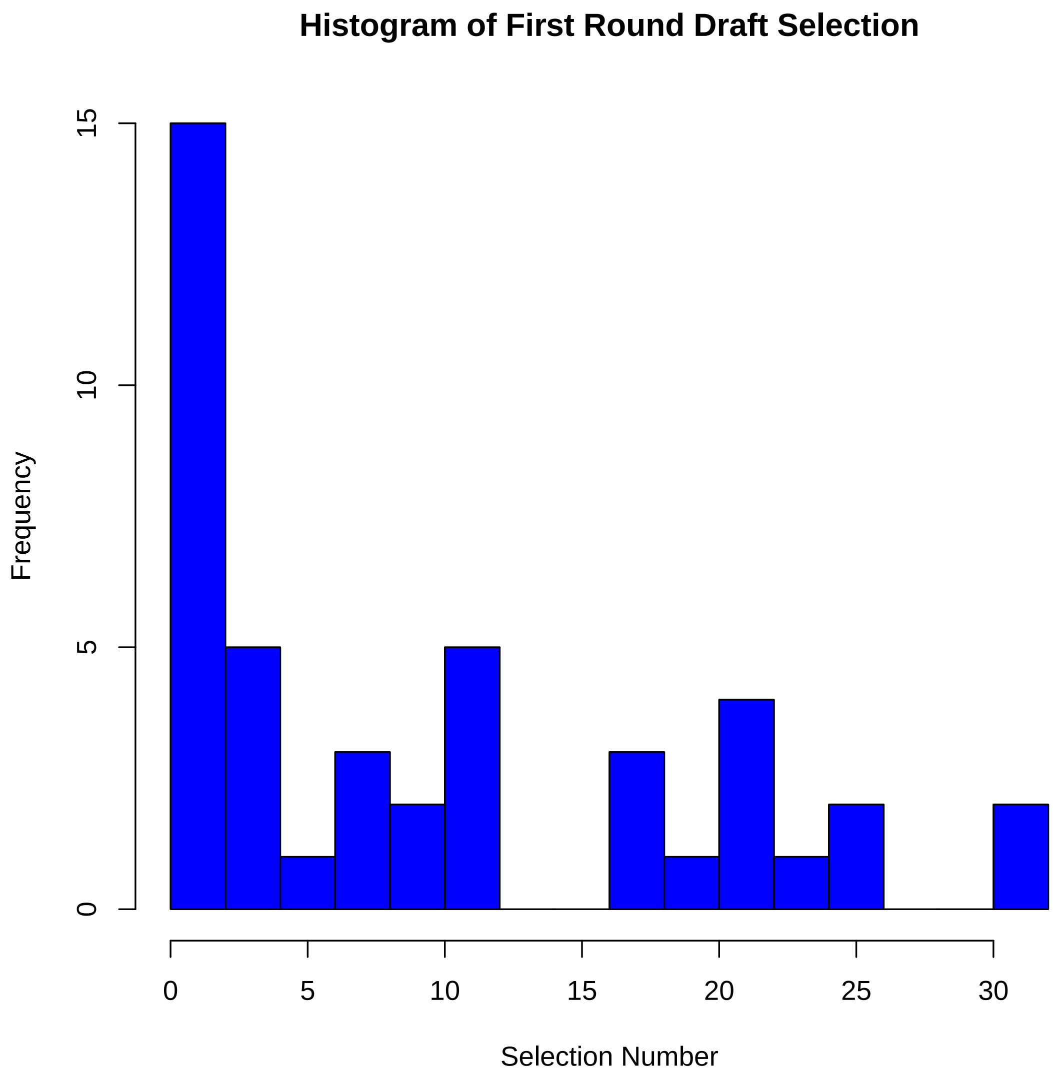

Figure 2.2: First Round Selections

16 Games Played

The second model built is a logistic model predicting whether a quarterback will play in 16 games or more (16 games corresponds to a full NFL season) in his career. Figure 2.3 shows the distribution of the binary target variable. For more information on logistic models, see Section 3.1.1.

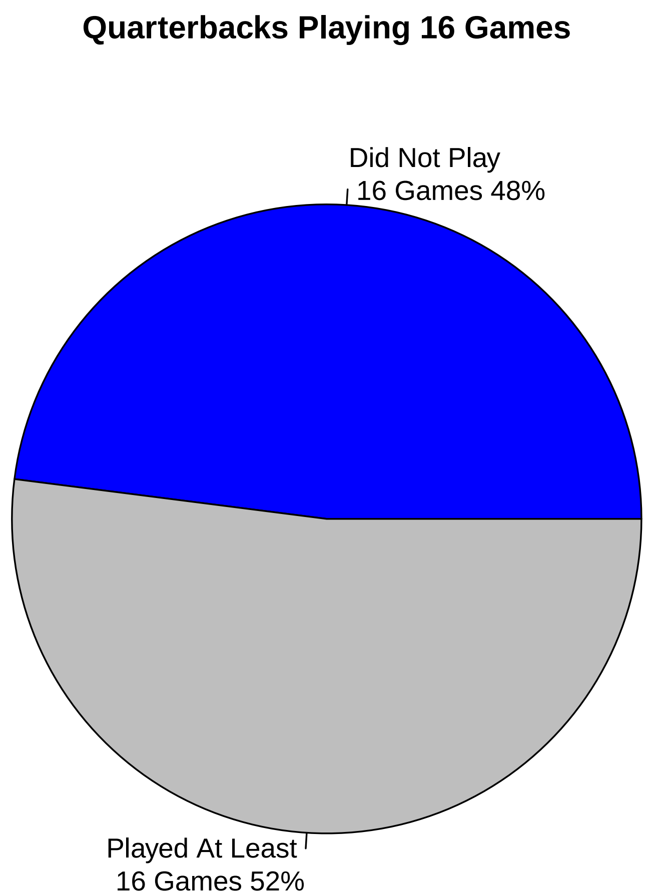

Figure 2.3: Distribution of 16 Games Played

Started Playoff Game

The third model is also a logistic model, but predicting whether a quarterback will start a playoff game in his career. Figure 2.4 shows the distribution.

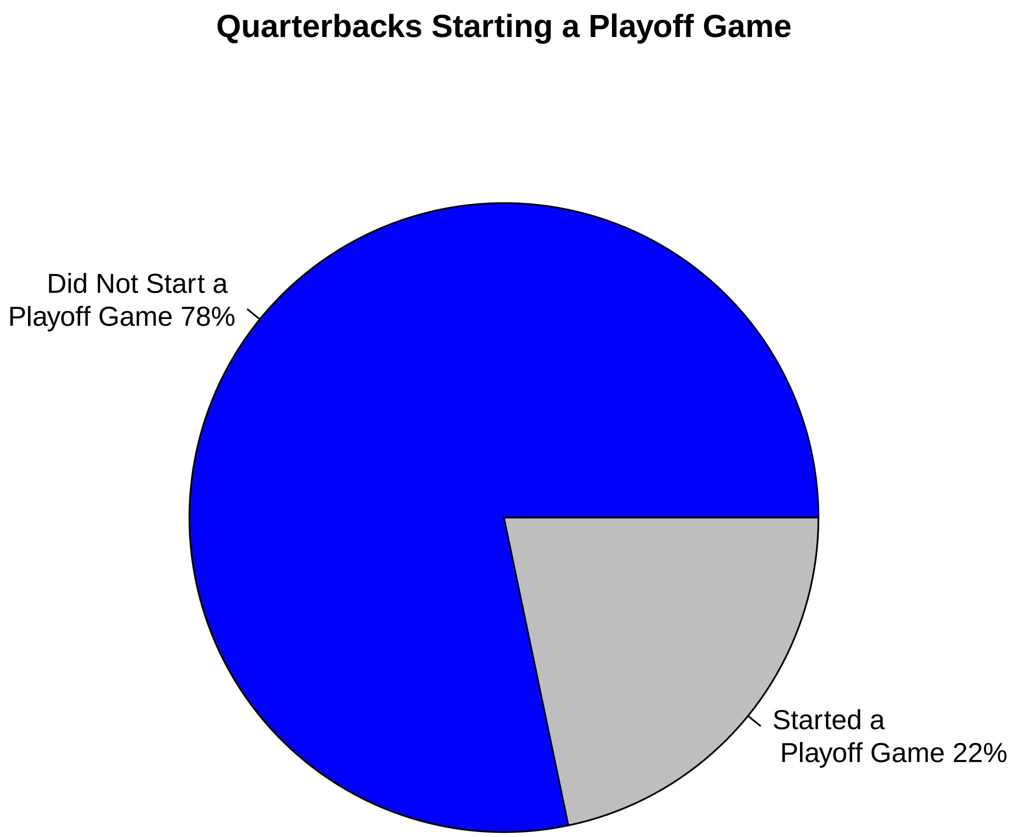

Figure 2.4: Distribution of Playoff Game Starts

Quarterback Rating

Figure 2.5 shows the distribution of Quarterback Rating, the target variable in the fourth model. Because of the large mass at 0 (caused by quarterbacks who have yet to play in an NFL game), a Tobit model is fit for this target variable. More information on the Tobit model can be found in Section 3.1.2.

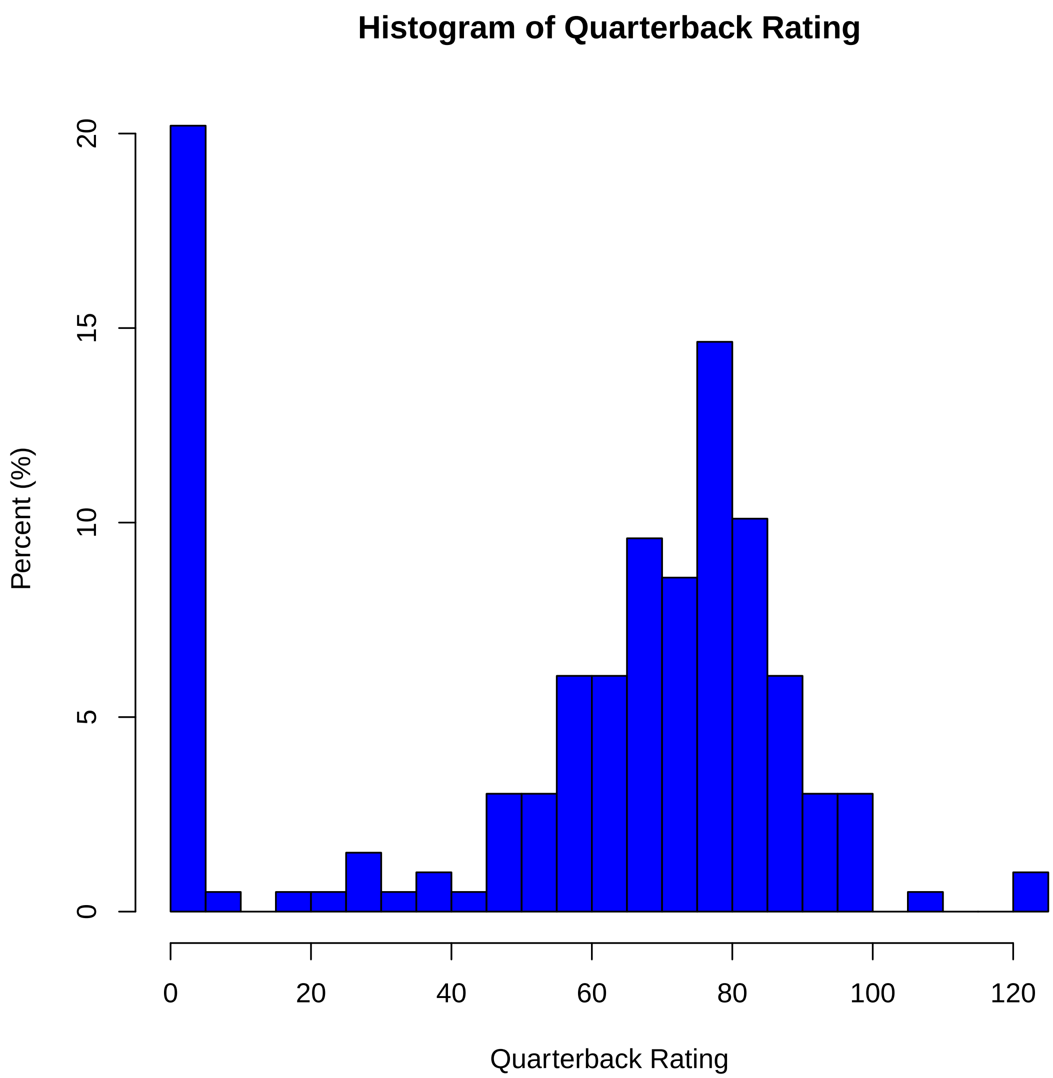

Figure 2.5: Distribution of Quarterback Rating

PC Rating

The final model built in this thesis is a Tobit model predicting the Principal Components Analysis (PCA) score of each quarterback's NFL career data. PCA provides a way to explain each quarterback's NFL play in one metric, similar to Quarterback Rating, which is a combination of only four statistics, so it is limited in power (Principal Components Analysis will be introduced in greater detail in Section 3.3). PCA provides loadings (similar to coefficients in a linear model) that explain which variables contribute towards explaining the variance in the data. The loadings in the first Principal Component were used to create a metric called PC Rating. In this case, the first Principal Component explained 56.22% of the variance, but there was not much dispersion between the loadings. Every loading was between 0.15 and 0.26 except 3 (Maximum Interception Percentage: 0.046, Rushing Touchdowns: 0.096 and Maximum Rushing Touchdowns: 0.126). The PC Rating model is an attempt to improve the Quarterback Rating metric to better explain and understand the performance of NFL quarterbacks. Figure 2.6 shows the distribution of PC Rating.

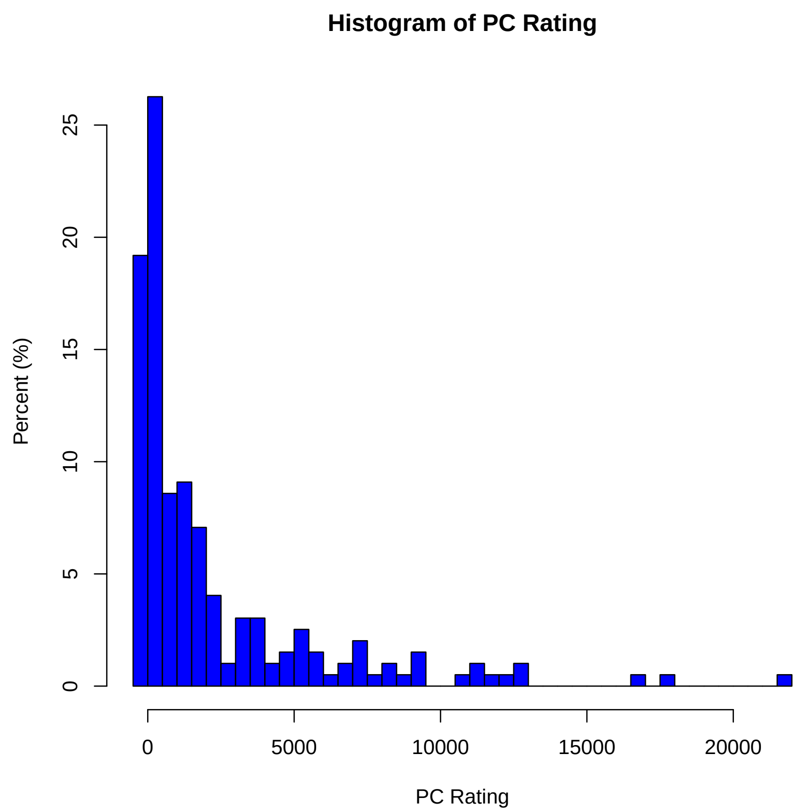

Figure 2.6: Distribution of PC Rating

After plotting PC Rating against a few different predictor variables it became apparent that the natural log of the PC Rating was more linearly related with the predictor variables. The new target variable is shown in Figure 2.7.

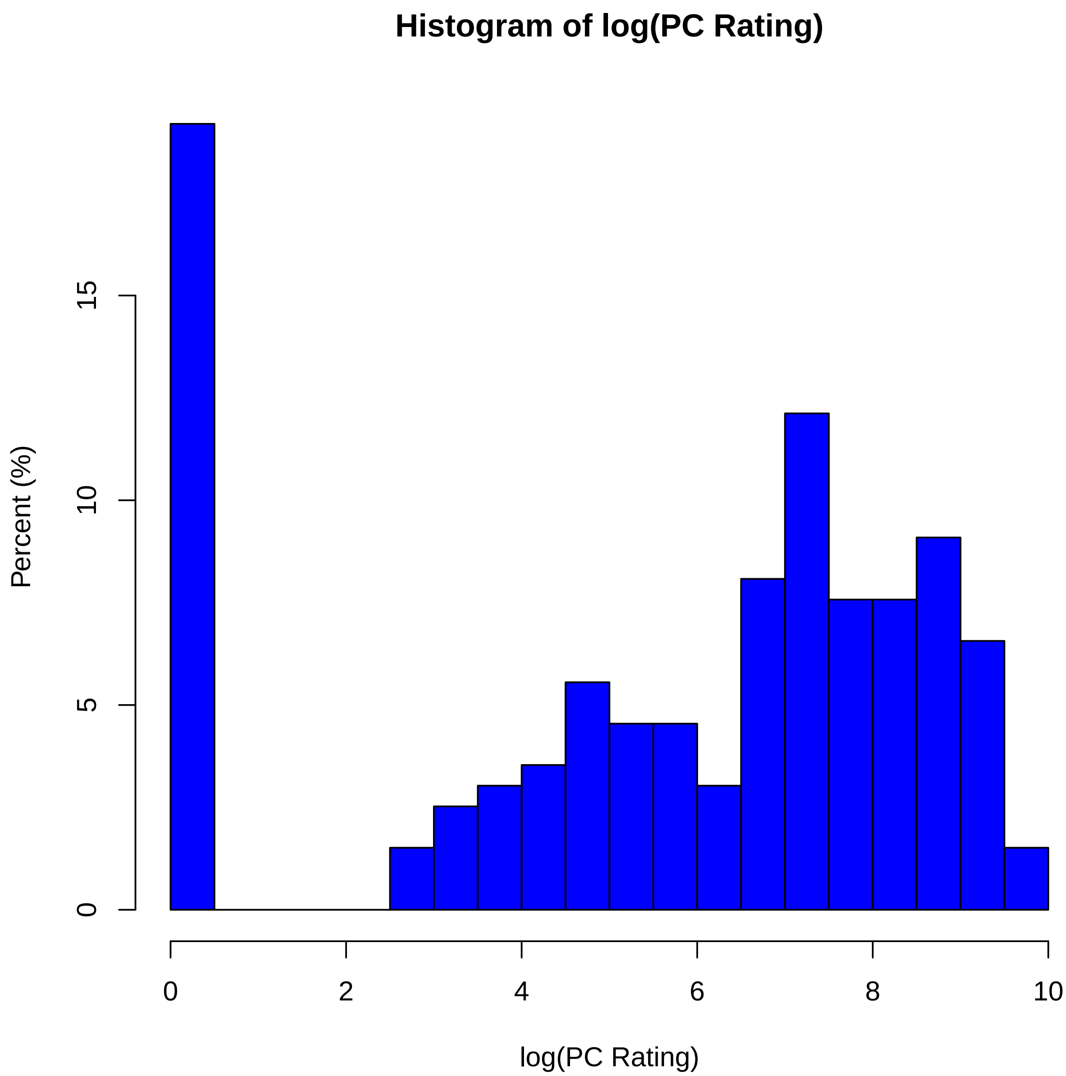

Figure 2.7: Distribution of log(PC Rating)

Chapter 3

Statistical Methods

3.1 Model Types

Logistic Regression

Logistic regression is a generalized linear model type where the target variable is binary (0,1). The response is often referred to as a success or failure, coded 1 and 0, respectively. Suppose an automotive insurer may be building a model predicting which of their customers are most likely to be in a car accident in the next year based on data collected over the past year. A customer's response is 1 if he or she was in a car accident and 0 if not. For each observation the logistic model will produce a predicted probability that the customer will be in a car accident in the next year. By using the form $\hat{p} = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_d x_d}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_d x_d}}$ each predicted response will always be bounded by (0,1). The logistic model is considered a linear model because the predicted probability of success, $\hat{p}$, can be transformed into the logit, $\ln(\frac{\hat{p}}{1-\hat{p}}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_d x_d$. In this thesis, two logistic models are built, one for modeling if quarterbacks would play at least 16 games (one full NFL season) in their careers and one model modeling if quarterbacks would start a postseason game.

Tobit Regression

Tobit regression is a way to model a distribution with left- and/or right- censoring. The model, which was introduced by Tobin (1958), describes the linear relationship(s) between a non-negative latent variable, y and a vector of independent variables $x_1, ..., x_d$. Because the Tobit model takes into consideration left- and right- censoring, y needs a lower bound, an upper bound or both and definition, y must be greater than or equal to 0. McDonald and Moffitt (1980) write, "The stochastic model underlying Tobit may be expressed by the following relationship:

$$y_i = \sum_{j=1}^{d} x_{ij}\beta_j + u_i \;\; if \;\; \sum_{j=1}^{d} x_{ij}\beta_j + u_i > 0$$
$$y_i = 0 \;\; if \;\; \sum_{j=1}^{d} x_{ij}\beta_j + u_i \leq 0.$$

...Thus, the model assumes that there is an underlying stochastic index, equal to $\sum_{j=1}^{d} x_{ij}\beta_j + u_i$ which is observed only when it is positive, and hence qualifies as an unobserved, latent variable". Two of the Tobit models built in this thesis, Quarterback Rating and log(PC Rating), were censored at 0. However, in the case of the Draft Selection Tobit model, the data were censored at an upper bound of 254. The same logic as above applies to a right-censored variable, like Draft Selection. In the Draft Selection Tobit model:

$$y_i = \sum_{j=1}^{d} x_{ij}\beta_j + u_i \;\; if \;\; \sum_{j=1}^{d} x_{ij}\beta_j + u_i < 0$$
$$y_i = 254 \;\; if \;\; \sum_{j=1}^{d} x_{ij}\beta_j + u_i \geq 254.$$

While the Tobit model still uses the regression form of:

$$y_i^* = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_d x_{id} + u_i, \; u_i \sim N(0, \sigma^2),$$

β_j coefficient should not be interpreted as the effect of x_j on y, as one would with a linear regression model. Instead, it should be interpreted as the combination of (1) the change in y of those above the limit weighted by the probability of being above the limit; and (2) the change in the probability of being above the limit, weighted by the expected value of y if above. So while each β can be interpreted as the increase per unit in $\hat{y}$, each β vector will be different from the β vector of the corresponding Least Squares model because each β takes into account the probability of an observation being censored.

3.2 Model Evaluation Methods

C-Statistic

Although logistic regression is a generalized linear model, a logistic regression model in the form $\frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_d x_d}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_d x_d}}$ is not a linear function. The R^2 metric holds an assumption that the model is linear, so R^2 is not appropriate for a logistic model. Instead, the C-Statistic is a commonly used metric which evaluates a logistic regression model (Lund and Raimi, 2011). However, the C-Statistic does not evaluate the goodness-of-fit, but instead measures how well the model rank orders the observations. Each observed success is paired with each observed failure, say $n_0 * n_1$ pairs. The C-Statistic is the proportion of pairs in which the observed success has a higher predicted probability than the observed failure, with a tie counting as half concordant. Technically the C-Statistic is bounded by [0,1]. However, a C-Statistic of 0.5 means when given any random pair of one success and one failure, the model gives a 50% chance of ordering the two observations correctly in order. The C-Statistic is especially useful in practices like marketing. A company may build a logistic model predicting which customers are most likely to use a coupon that was sent in the mail. The company wants to maximize profits and minimize costs by only sending the coupons to the top 20% of customers they think are most likely to use the coupon to buy their product. A study like this may have a very small target rate, so once the model is built, every observation may have a predicted probability less than 0.5, so every observation is predicted to be a failure. The C-Statistic can measure how well the model is separating to give a clearer understanding of which observations should be in a certain percentile of the data. For example, if the data set looked like this:

Table 3.1: Example of Scored Data for C-Statistic Calculation

ID	Target	$\hat{p}$
A	0	0.1
B	0	0.2
C	1	0.2
D	0	0.4
E	0	0.5
F	1	0.6

Each observed success (C,F) is paired with each observed failure (A,B,D,E) and the predicted probabilities are compared:

Table 3.2: Pairs for C-Statistic Calculation

Pair(0,1)	$\hat{p_0}$ --- $\hat{p_1}$	Concordance
A, C	$0.1 < 0.2$	Concordant
A, F	$0.1 < 0.6$	Concordant
B, C	$0.2 = 0.2$	Tied
B, F	$0.2 < 0.6$	Concordant
D, C	$0.4 > 0.2$	Discordant
D, F	$0.4 < 0.6$	Concordant
E, C	$0.5 > 0.2$	Discordant
E, F	$0.5 < 0.6$	Concordant

Of the 8 pairs, 5 were concordant, 2 were discordant and 1 was tied. So, this model's C-Statistic is $\frac{1*wins+0*losses+0.5*ties}{8} = \frac{5.5}{8} = 0.6875$, which means about 69% of the time, when given one failure and one success, the model can correctly choose which is the failure and which is the success.

Lift Charts

It is difficult to compare the fits of multiple models of different types. There is no common metric like R^2 or C-Statistic that can evaluate all model types. However, there are plots that can be helpful. A lift chart is a line graph that compares the predicted quantiles against the true average target for each of those quantiles. After a model is built, the data are scored by the model and ranked by the predicted target. Then, the data are split up into quantiles (traditionally 10 deciles), and the average true response is calculated for each quantile and plotted. Lift charts are used to evaluate the models in Chapter 4.

There are a few desirable characteristics to look for in a lift chart. An ideal model will be able to consistently rank order the observations. The lift chart can show how well the model is consistently separating the data by looking at the slope of the line and by the monotonicity of the line. A well-fitting model will have a steep slope and there will be few changes in trajectory. A simulated lift chart from a logistic model where the target rate is 25% is shown below in Figure 3.1. The lift curve looks fairly monotonic, with only a few slight dips at deciles 3, 5 and 7. In many cases a logistic model is fit to identify the most likely or least likely candidates for some outcome, so while a consistently fitting model is always important, the main focus may be on the tails in the lift curves. In the case of our analysis, the two logistic models built, 16+ Games Played and Playoff Start are both probably more useful in identifying quarterbacks who have the highest probability (so a team can try to acquire the quarterback for their team) or the lowest probability (so the team can avoid this quarterback). While the predicted probabilities are important, in these cases the rank-ordering of the observations is much more important.

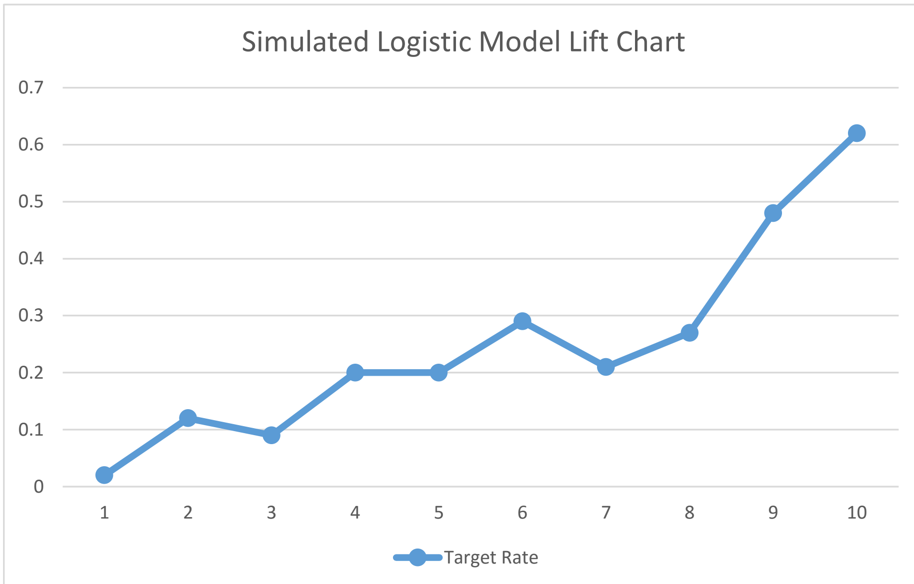

Figure 3.1: Simulated Lift Chart of a Logistic Model

While lift charts have many advantages, there are a few drawbacks. Lift charts have the potential to be misleading when not used properly. The number of quantiles is an important decision, and when comparing competing models the number of quantiles needs to be kept constant. Too few quantiles reduces variance in the quantiles and will generally produce a more monotonic line, although the slope of the line will be shallow. Too many quantiles will often produce a steeper slope, but the average target in the deciles can fluctuate greatly because outlier observations greatly affect the mean. There is little information to be learned by a lift chart with too few or too many deciles and a well-fitting model could receive a negative diagnosis.

3.3 Principal Components Analysis

An Explanation of Principal Components Analysis

PCA is a dimension reduction technique that finds orthogonal vectors, called eigenvectors, that explain the underlying structure of the data by measuring the data in terms of its principal components instead of its axes (Dallas, 2013). The first eigenvector has no restrictions on its direction and is simply pointed in the direction of the most variation in the data. So if the data consisted of two variables, the first principal component would be the eigenvector shown below (all pictures taken from Dallas (2013)).

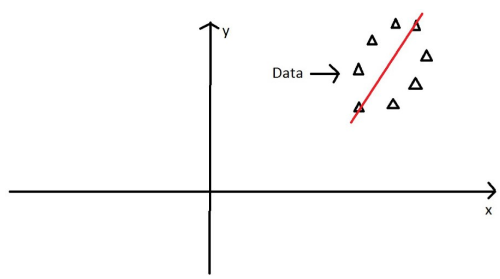

Figure 3.2: First Principal Component of Example Data

The second principal component explains the most variance, perpendicular to the first principal component.

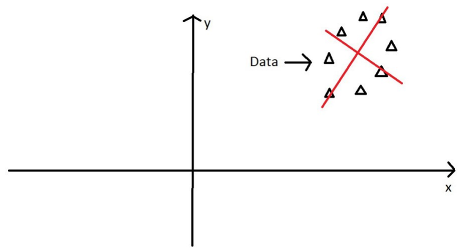

Figure 3.3: Second Principal Component of Example Data

The eigenvectors now create new axes describing the important components of the data.

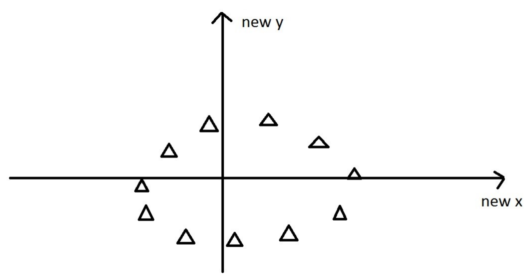

Figure 3.4: New Axes of Example Data

When using a larger data set, it is usually unnecessary to use every principal component to describe the data. At some point the variance in the data is thoroughly described and adding more principal components only describes minimal additional variance. The variance each component captures can be described in an eigenvalue. The eigenvalues all sum to d, the number of variables in the data, so the proportion each principal component describes is the eigenvalue divided by d.

How PCA was Used in This Thesis

Predicting NFL success by college success is susceptible to lurking variables. As of right now, the Quarterback Rating is the most popular single metric by which to compare quarterbacks. However, the Quarterback Rating has its limitations. It is a combination of only four statistics: completion percentage, yards per attempt, touchdown percentage and interception percentage (refer to Section 2.2 for a full explanation of Quarterback Rating). By only using four statistics it is limited in capturing the true performance of quarterbacks, but there are also lurking variables that could have an effect on a player's quarterback rating, mostly with regards to the team's running back. For instance, a higher touchdown percentage increases a player's quarterback rating, so a quarterback is penalized for handing the ball off to a running back and letting him score. Principal Components Analysis provides an alternative to Quarterback Rating to combine NFL data into one composite metric that could then be used as the target variable in a model. Principal Components Analysis was performed on 23 NFL statistics. The first principal component, which explains 56.22% of the variance, is used to score all of the observations to make a PC Rating. A Tobit model is then fit using college predictors to model the PC Rating. However, a natural log transformation was used on the PC Ratings to fit the model. This will be gone over in more detail in Section 4.5.

Chapter 4

Data Analysis

4.1 Draft Selection Model (Tobit)

To understand the potential mistakes NFL teams are making in the draft, an understanding of what NFL teams value in the draft first needs to be established. A Tobit model was fit predicting in which draft slot the quarterback would be selected. Each year there are 7 rounds in the NFL draft, but the total number of selections in the draft differs from year to year because of penalties and compensation picks. For instance, the New England Patriots were found guilty of under-inflating footballs in order to give their quarterback, Tom Brady a better grip on the football. This scandal is now commonly referred to as “Deflategate” and cost New England their first round selection in the 2016 draft. Compensation picks are awarded by the NFL in situations when a team is unable to re-sign a valuable player to a new contract. These picks are awarded every year, usually in the later rounds.

25 of the 198 quarterbacks in this study were not drafted out of college. The lowest draft selection for any quarterback who was drafted was 253, so all of the quarterbacks who were not selected in the NFL draft were analyzed as if their selection was 254, so 254 was the upper limit set for the Tobit model. Unfortunately, PROC QLIM, the SAS procedure

utilized for the Tobit models does not have any built-in selection methods, so all three of the Tobit models had to be reduced by hand. The model reductions started by fitting the model with every possible predictor and removing the least significant variable at each step. After each step the AIC and BIC were recorded to see if there was significant improvement in the model after removing an insignificant variable. Sometimes in the model reduction stage powerful variables will appear to be insignificant because of multicollinearity, so if the AIC and/or BIC did not improve significantly, that variable was kept in a list of variables to try to fit into the model after further reduction.

Once the model was reduced to a point where all of the variables were significant (usually around 10-15 variables), multicollinearity was taken into consideration. Multicollinearity occurs when two or more variables which are highly correlated are included in a model. While multicollinearity does not usually dramatically affect the predictions produced by the model, it can alter the parameters and even change the signs of the parameters. Multicollinearity can also inflate the standard error of the betas, minimizing the apparent effect of the correlated variables (Penn State, 2016). If variable importance is of interest, multicollinearity can alter the order of variable importance. A way to detect multicollinearity in a model is to calculate each variable's Variance Inflation Factor (VIF). A Variance Inflation Factor is the estimate of the proportion in which the beta's standard error is increased when there is a correlated variable in the model. The formula for each variable's Variance Inflation Factor is $\frac{1}{1-R^2}$, where R^2 is the correlation coefficient of the predictor variable at hand, modeled by the rest of the predictor variables. Variance Inflation Factors are independent of the target variable. So, if a target variable is being modeled by $X_1 - X_{10}$, the VIF of X_1 is $\frac{1}{R^2}$ of the variable X_1 modeled by $X_2 - X_{10}$. While the optimal cutoff for VIFs is a bit arbitrary, 3, which is what was used for all five models, is a commonly used cutoff, which would be an R^2 of 0.67. Finally, two-way interactions and variable transformations were tested for significance after the final model was fit, however none were selected.

The final model is shown below in Table 4.1, with the variables listed in alphabetical order.

Table 4.1: Selection Tobit Model Parameter Estimates

Parameter	Estimate	Std. Error	T-Value	p-value
Intercept	919.945	175.457	5.24	< 0.0001
Completion Percentage	-404.354	151.729	-2.66	0.0077
Passing Yards per Completion	-9.321	4.054	-2.30	0.0215
Rushing Attempts	-0.077	0.045	-1.69	0.0905
Weight	-1.810	0.519	-3.49	0.0005

Four variables were selected in the model: Completion Percentage, Passing Yards per Completion, Rushing Attempts and Weight. All four variables are negatively correlated with Draft Selection. Because players want to be drafted as early as possible, this means that all four variables are positive traits for a college quarterback to possess. Completion Percentage is actually in the form of a proportion, so the coefficient of -404.354 should be interpreted as a 1% increase in a quarterback's completion percentage leads to a prediction 4.04 draft slots earlier. The coefficient for Passing Yards per Completion is -9.321. This means according to this model a quarterback will be projected to get drafted 9.321 spots earlier by increasing his passing yards per completion by 1 yard. However, as mentioned before, all of these coefficients take into account the probability an observation will be censored, so these coefficients are not to be interpreted as if they were from a Least Squares model. Rushing Attempts is the total number of rushing attempts a quarterback had in his college career. While this variable was significant, there are a lot of potential lurking variables that could contribute to this variable being significant, so if this model was being used for predictions, Rushing Attempts should be included cautiously. Some offensive schemes in college football require the quarterback to run on a high percentage of plays, while some offensive schemes may never call for a quarterback run. The number of rushing attempts also likely depends on the running backs and wide receivers also on the team. Similar to the interpretation of the first two variables, the coefficient of -0.077 means the quarterback's draft selection projection will decrease by 0.077 spots for every rushing attempt he has in his career. So, for a quarterback to improve his draft selection by 1 spot he has to have an additional 13 times. Finally, the quarterback's weight is also a significant variable. The coefficient of -1.81 suggests heavier quarterbacks get drafted earlier. This variable could also be partially a proxy variable for the quarterback's style of play. Most quarterbacks in the NFL do not run as much as quarterbacks in college, and mobile quarterbacks are usually lighter than quarterbacks who pass more often.

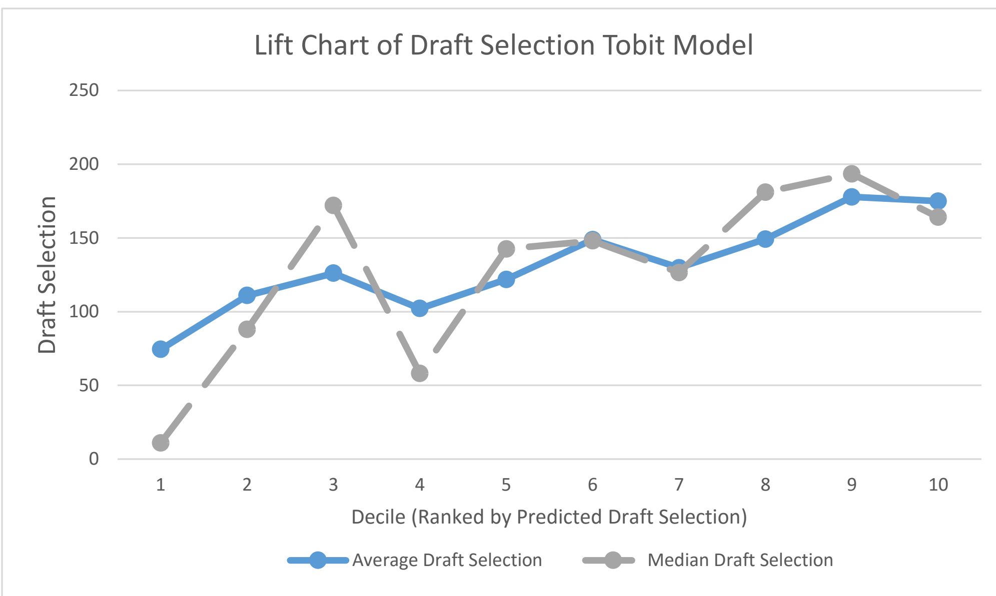

Figure 4.1: Lift Chart of Draft Selection Tobit Model

The lift chart in Figure 4.1 shows the average actual draft selection for each decile ranked by predicted draft selection. The lift curve shows a steadily increasing trend besides slight dips at the fourth and seventh deciles. The average draft selection from the tenth decile is almost 2.5 times larger than the average draft selection from the first decile. The combination of a monotonicity and a lift around 2.5 shows this model is separating players selected high in the draft from players selected low in the draft fairly well. The model is also doing a good job at finding the very high draft picks. The first decile includes 4 of the 12 first overall selections, 2 of the 3 second overall selections and 8 of the 22 players selected in the first 7 picks. Because only 20 observations were used in each decile, the variation in each decile was large, so the lift curve of the median draft selection from each decile was also plotted. However, lift curves are more commonly used to examine the mean of the target in each predicted decile.

4.2 16 Games Played Model (Logistic)

The second model is a logistic model predicting whether a quarterback will play at least 16 games in his career. Sixteen games was chosen because an NFL regular season is 16 games. While playing a full NFL season can be interpreted as a quarterback performing well enough to lead a team for an entire season, longevity in the NFL can lead to a player's success. Also, some teams will "tank" if they do not get off to a promising start. Tanking is when a team plays sub-par players in order to lose games, securing a higher draft pick in next year's NFL draft. So just because a player can play 16 games in his NFL career does not necessarily mean he was one of the best 32 quarterbacks in the NFL.

The model selection strategy was a bit different in the logistic models because PROC LOGISTIC offers backward, forward and stepwise selection in reducing the variables in the model. The first step in model reduction for the two logistic models was to run PROC

LOGISTIC with backward, forward and stepwise selection (with entry and exit alpha=0.10) and take the union of the variables selected in each model. This gave a conservative approach, careful to not exclude any possible powerfully predictive variables. Then, the full model was fit and reduced with a similar technique to the Tobit model reduction strategy. Variables were removed if they were not significant, but Variance Inflation Factors greater than 3 were also taken into account when removing variables, as well as variables whose coefficient may have changed from positive to negative or vice versa. The final logistic model predicting a quarterback plays at least 16 games in the NFL is shown in Table 4.2.

Table 4.2: 16 Games Played Model Parameter Estimates

Parameter	Estimate	Std. Error	T-Value	p-value	Std. Est.
Intercept	-15.400	6.887	-2.236	0.0253	—
Pass Yards per Game	0.009	0.003	3.313	0.0009	0.314
Height	0.220	0.095	2.308	0.0209	0.200
Proportion of Pass Attempts	-3.437	1.772	-1.940	0.0524	-0.1839

The model produces a C-Statistic of 0.649, meaning when given a player who has played at least 16 games and player who has not played at least 16 games, the model can correctly select the player who has played at least 16 games about 65% of the time. The model selected three variables: Passing Yards per Game, Height and Percentage of Pass Attempts. All three variables are positively correlated with the probability a quarterback will play at least 16 games in his NFL career, however because this is a logistic model, the coefficients are only interpretable by themselves with everything else at a fixed position. That is to say, the increase in predicted probability from one variable will change as the value of the other predictors in the model change. Passing Yards per Game is positively affects the probability a quarterback will start at least 16 games in his career, so a high number of passing yards per game is a positive trait for a college quarterback to possess. Height is also a positive indicator of the likelihood a quarterback will start at least 16 games in his NFL career. The third variable in this model is Proportion of Pass Attempts. This is the proportion of passing attempts a quarterback has out of the summed number of passing attempts and rushing attempts. This is a negative indicator of the number of games played in the NFL, suggesting having a balanced proportion of rushing and passing attempts increases a quarterback's chances of starting more games.

While a few variables included in this model and the Draft Selection model are correlated, no variables were used in both models. Only three pairs of variables in the two models had correlations more extreme than 0.5 (Pass Yards per Game & Completion Percentage: 0.501, Height & Weight: 0.528, Proportion of Pass Attempts & Rushing Attempts: -0.603). While playing at least 16 games in an NFL career does not necessarily translate to proficiency at the quarterback position, it is important to observe the difference in variables chosen in these two models. The contrast in these two models, although different types of regression models, shows that since the 1998 draft, teams are placing value in characteristics that do not necessarily translate to success, at least in regards to playing time.

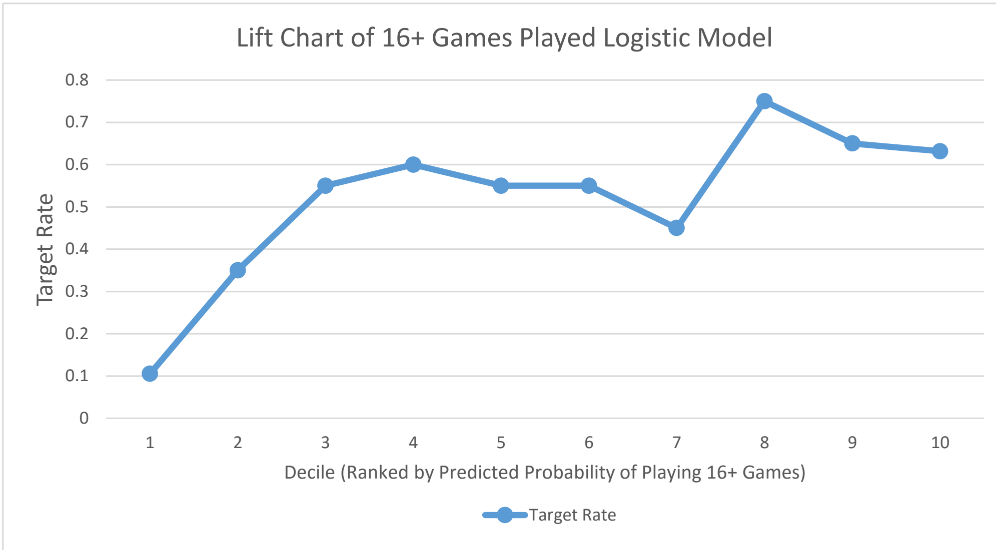

Figure 4.2: Lift Chart of 16 Games Played Model

The lift chart in Figure 4.2 shows a general upward trend with a bit of uncertainty in the middle. The model is doing a very good job at finding the players who did not play 16 or more games in their NFL careers, however has a bit of trouble finding players who definitely have played 16 or more games in the NFL. Of the 95 players who have not played 16 or more games in the NFL, the lowest ranked 10% includes 17 of the 95 players (17.9%) and the lowest ranked 20% include 30 of the 95 players (31.6%). So while a player ranked in the top decile according to this model, meaning being selected as the top 10% most likely to play at least 16 games in their NFL career, only 63% actually have played at least 16 games. However, 89% of the group being isolated as the 10% least likely to play at least 16

NFL games, so using this model NFL teams can be fairly confident about the potential lack of success of quarterback, but maybe not the potential success of a quarterback.

4.3 Playoff Start (Logistic)

The second logistic model fit calculated the probability that a quarterback would start at least one playoff game in his career. While gaudy statistics are fun for Fantasy Football team owners and easy material for sports radio hosts to discuss, they often do not translate into postseason success, which is what fans and the team ultimately want. While it is very natural to assume that leading the NCAA in a common statistic like passing yards or touchdowns would translate to winning, this is often an error in the eyes of the fans and the franchises, alike. Often, the quarterbacks with the best statistics fail to win in the playoffs and even frequently miss the playoffs altogether. In the 2015-2016 NFL season, only 3 of the top 10 quarterbacks in passing yards made the playoffs. Drew Brees, the quarterback of the New Orleans Saints, led the entire NFL in passing yards, but the Saints finished the season at 7-9. This model was fit to attempt to understand what actually translates to postseason success and then compare the variables chosen in the model to those chosen in the Draft Selection model to ultimately understand if NFL teams value the same traits that lead to success in the NFL playoffs. The model parameters are shown below in Table 4.3.

Table 4.3: Playoff Game Start Model Parameter Estimates

Parameter	Estimate	Std. Error	T-Value	p-value	Std. Est.
Intercept	-14.105	4.056	-3.477	0.0005	—
Max Rush Yards per Carry	-0.229	0.107	-2.142	0.0322	-0.521
Rush Yards per Game	0.031	0.012	2.684	0.0076	0.3473
Completion Percentage	10.673	4.034	2.646	0.0082	0.2852
Weight	0.030	0.013	2.224	0.0260	0.2183

The model introduced in Table 4.3 produces a C-Statistic of 0.698. The high C-Statistic along with the relatively monotonic lift chart shown in Figure 4.3 shows that this model is rank-ordering well. Again, because this is a logistic regression model, the coefficients cannot be directly interpreted as a linear relationship with the dependent variable. However, the signs of the parameters do specify a positive or negative correlation with starting at least one playoff game. Four variables were selected in this model: Maximum Rushing Yards per Carry, Rushing Yards per Game, Completion Percentage and Weight. The first interesting observation about this model is that the two most important variables are both rushing statistics. It is also interesting that while averaging more rushing yards per game increases a quarterback's probability to start a playoff game, earning a higher maximum average rushing yards per carry actually decreases the quarterback's probability to start a playoff game. Completion Percentage and Weight are also included in the model as positive indicators as success. Both of these parameters were also included in the Draft Selection model as indicators of success (with negative coefficients because a lower draft selection is favorable). While the model displayed in Table 4.3 shows an importance of rushing statistics, two of the four variables included in the model were also included in the Draft Selection model, suggesting similar traits between success in reaching the playoffs and the characteristics of a player NFL teams are looking for.

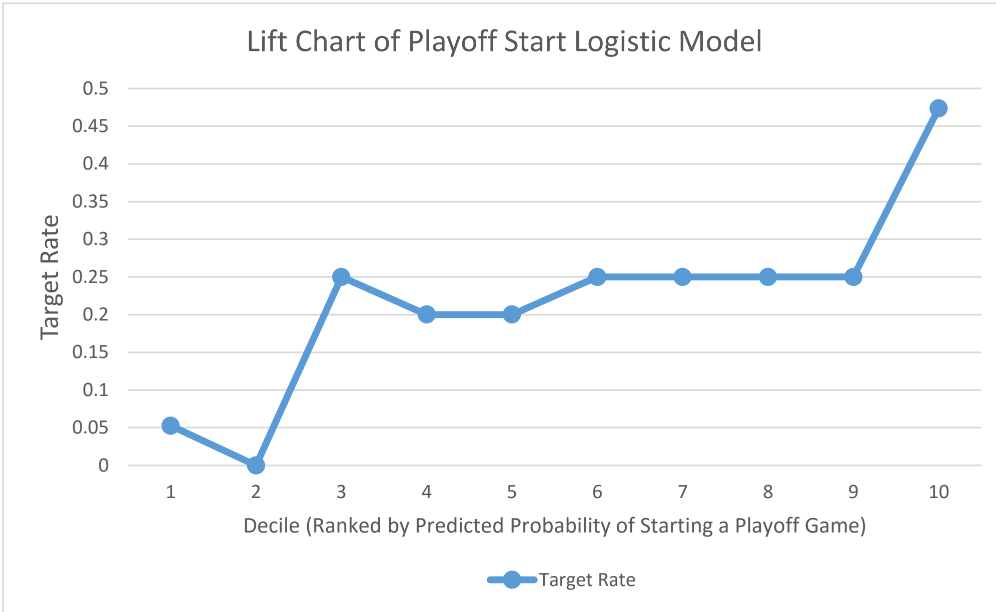

Figure 4.3: Lift Chart of Playoff Start Model

The lift chart shown in Figure 4.3 shows great separation in the bottom 20% and the top 10%, but the middle 70% plateaus between a target rate of 20% and 25%. Of the 198 players in this study, 43 have started a playoff game and 155 have not, giving an overall target rate of 22%. Only 1 player ranked in the least likely 20% actually started a playoff game, meaning this model was able to capture 25% of the players who have not started a playoff game. While the top decile only has a target rate of 47% this translates to capturing 9 of the 43 players (21%) in just the top 10% predicted.

4.4 Quarterback Rating (Tobit)

As mentioned in Section 2.2, Quarterback Rating is a common metric that is used to evaluate quarterback play. It takes into account four statistics: completion percentage, yards per pass, touchdown percentage, and interception percentage. While the goal of any team is to win in the playoffs, it is unfair to completely judge a quarterback's success by his playoff appearances. There are great quarterbacks on bad teams who would likely make it to the playoffs had they been surrounded with a different team. While the Quarterback Rating metric is limited by only four statistics, it is generally thought of as the gold standard to measure overall quarterback play.

Thirty-nine of the 198 quarterbacks in the data set had a Quarterback Rating of 0 due to not recording any statistics in an NFL game, so the distribution of Quarterback Rating had a large mass at 0. Because of the mass, a Tobit regression model was fit instead of a Least Squares model, using the same model reduction strategy as the Draft Selection model. The final model parameters are shown below in Table 4.4.

Table 4.4: Quarterback Rating Tobit Model Parameter Estimates

Parameter	Estimate	Std. Error	T-Value	p-value
Intercept	-228.161	128.749	-1.77	0.0764
Rush Yards per Attempt	-2.505	1.213	-2.06	0.0389
Max Completion Percentage	0.933	0.420	2.22	0.0265
Height	3.106	1.696	1.83	0.0671
Max Interception Percentage	-1.428	0.638	-2.24	0.0251

The Tobit model for Quarterback Rating includes four variables: Rush Yards per Attempt, Max Completion Percentage, Height and Max Interception Percentage. Completion Percentage and Interception Percentage are both included in calculating Quarterback Rating, but it is interesting that the model chose a quarterback's Maximum Completion Percentage in college and Maximum Interception Percentage in college to help model Quarterback

Rating in the NFL. This suggests that a quarterback should not be ignored if he has a low overall Completion Percentage or a high overall Interception Percentage, but has one great year in college. It suggests that a player should be evaluated based on his maximum value in college. Again, both of these variables are proportions, so a 1% increase in a player's Maximum Completion Percentage in college will increase his predicted NFL Quarterback Rating by 0.00933 points, while a 1% increase in a player's Maximum Completion Percentage in college will decrease his predicted NFL Quarterback Rating by 0.01428 points. Rush Yards per Attempt was a negative contributer, suggesting a mobile quarterback is not a trait found in successful NFL quarterbacks. Again, while having a fast quarterback is always a desirable trait, this is most likely a proxy variable for a quarterback's style of play. A quarterback has to be a proficient passer to be successful in the NFL, and while being a great rusher would be a great complimentary trait to possess with being a great passer, it is very uncommon. Usually in college football, teams choose to either have a "pocket passer" quarterback or a mobile quarterback. The style of play in the NFL fits the pocket passer much better, so while being able to run is not a negative trait for a quarterback to possess, being a mobile quarterback is usually indicative of a quarterback being a sub-par passer. Lastly, Height is also a positive contributer in predicting NFL Quarterback Rating, agreeing with the previous three models that bigger quarterbacks are more desirable.

While this model did not include any of the same variables as the Draft Selection model, two pairs of variables in the two models were highly correlated. As was the case in the 16 Games Played Logistic model, Height and Weight have a correlation coefficient of 0.528. Completion Percentage, used in the Draft Selection model and Max Completion Percentage used in this model are also highly correlated, with a correlation coefficient of 0.714.

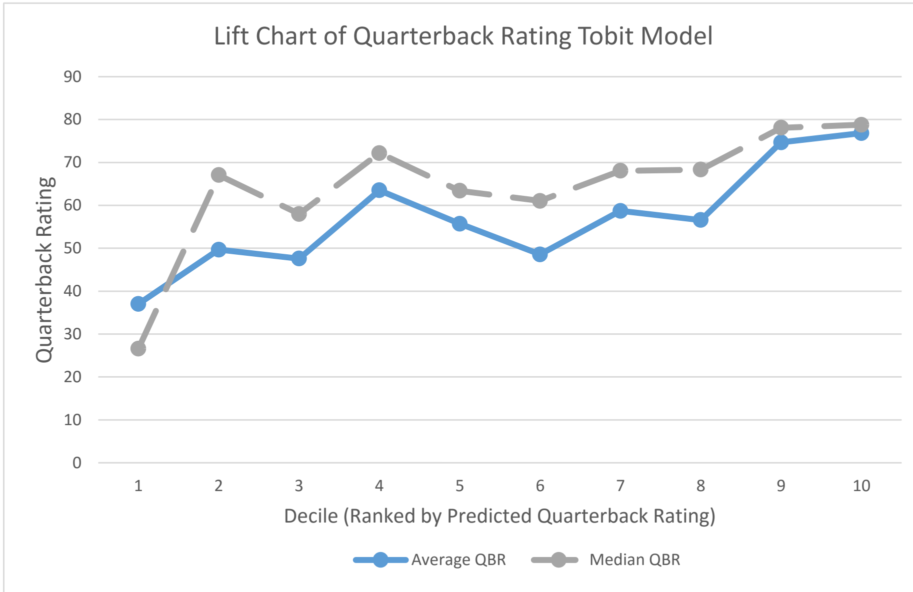

Figure 4.4: Lift Chart of QBR Tobit Model

The lift chart for the Quarterback Rating model, shown in Figure 4.4 shows a monotonic upward trend, but also has a few slight dips in deciles 3 and 8 and a rather large decrease in average Quarterback Rating in deciles 5 and 6. Because averages can be heavily influenced by just a few observations and each decile has only 20 observations, the median Quarterback Rating in each decile was also plotted. The median lift curve actually captures a smaller observation in the first decile and a larger observation in the tenth decile, but actually follows a very similar pattern to the mean lift curve. The lift of this model is just a bit over 2, which is fairly low, suggesting the model has fairly weak predictive power, and although the lift curve shows a monotonic upward trend, the slight dips in the curve show the model is fairly inconsistently separating.

4.5 Principal Components Rating (Tobit)

Because of the limitations in Quarterback Rating, Principal Components Analysis was used to create a metric that best explained quarterback play in the NFL. Each player was scored based on the first Principal Component. This score metric had a very large range and was highly right skewed (see distribution in Section 2.3.5). The PC Rating metric was plotted against a few different variables and looked as if a log transformation on PC Rating would produce a more linear relationship between the PC Rating and the independent variables. Again, because a large percentage of quarterbacks did not have any NFL experience there was a large mass in PC Rating at 0. Because the log of 0 is undefined, and because of the very large range in PC Ratings, these quarterbacks were assigned a log(PC Rating) of 0 and this log transformed variable was then used as the target variable in the final Tobit model in this analysis. The model is shown below in Table 4.5.

Table 4.5: log(PC Rating) Tobit Model Parameter Estimates

Parameter	Estimate	Std. Error	T-Value	p-value
Intercept	-26.757	12.212	-2.19	0.0284
Height	0.404	0.162	2.49	0.0129
Pass Yards per Game	0.0084	0.0043	1.96	0.0501

The PCA model only used two variables, making it the most reduced model in this analysis. Height was again a positive predictor as well as Pass Yards per Game, which are both of these variables were also used in the 16 Games Played Logistic model.

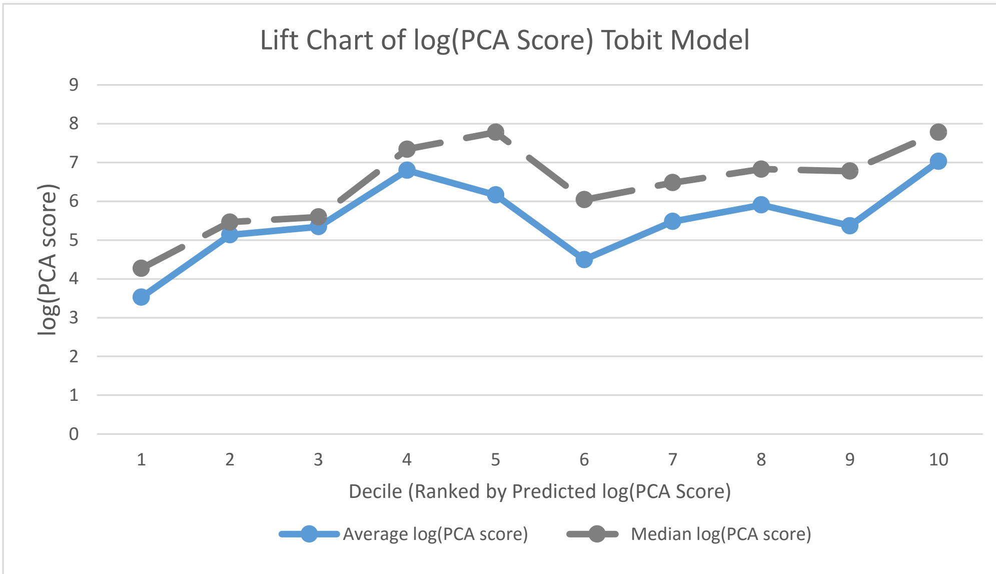

Figure 4.5: Lift Chart of log(PC Rating) Tobit Model

Figure 4.5 shows a very inconsistent fitting model. While the first decile has the lowest average log(PC Rating) and the tenth decile does offer the highest log(PC Rating), the fourth decile has the second highest average log(PC Rating) and the fifth decile has the second highest median. The lift of this model is positive, but it is weak and while the tails may be separating well, the middle 80% looks more like noise than a well fitting model.

Chapter 5

Conclusion

The motivation behind this thesis was inspired by the staggering number of quarterbacks drafted very early and not living up to expectations in recent years. NFL teams have set themselves back years by drafting a quarterback too early and having to pay an outrageous amount of money for him, to only be let down by his lack of production on the field. The goal of this thesis was to first identify some of the characteristics NFL teams found important when selecting quarterbacks in the NFL draft. Then, identify which characteristics actually translate to success. If those two goals could be obtained then this thesis could perhaps provide some insight into some potential warning signs that could prevent teams from over-drafting a quarterback and self-inflicting themselves with major setbacks. Forty-one variables were available for selection in the 5 models. Twelve variables were used in the five models, including 4 variables that were used more than once. Table 5.1 shows a chart of each variable used in each model.

Table 5.1: Variables Used in Each Model

	Selection	16+ GP	Playoff Start	QBR	PCA
Completion Percentage	-		+		
Passing Yards per Completion	-				
Passing Yards per Game		+			+
Max Completion Percentage				+	
Max Interception Percentage				-	
Proportion of Pass Attempts		-			
Rushing Attempts	-				
Rushing Yards per Carry				-	
Rushing Yards per Game			+		
Max Rush Yards per Carry			-		
Height		+		+	+
Weight	-		+		

The first model built was a Tobit regression modeling Draft Selection. This model was used as a baseline to determine if the information found important was also found to be powerful information in the "success" models. Four variables were selected in the model: Completion Percentage, Yards per Completion, Rushing Attempts and Weight, all of which were positive indicators of success. This is the only model where a smaller target is better, because players want to get drafted as early as possible, so a variable having a negative coefficient is desirable and should be thought of as a positive indicator of success. This model had a fairly monotonic lift curve which means the model is fitting consistently and is doing a good job at separating players selected early in the draft from players selected late in the draft.

The second model fit was a logistic model predicting whether or not a quarterback had played at least 16 games in his career. Sixteen games was chosen because the NFL regular season is 16 games long and provided a split where 52% of players had played at least 16 games and 48% had not. The model did not use any of the same variables as the Draft Selection model, however each variable selected in the 16+ Games Played model was correlated with a variable in the Draft Selection model by a magnitude of at least 0.5. So while the exact variables chosen are not the same, the traits that NFL teams find desirable, and worthy of spending a top draft pick and millions of dollars on, are similar to the traits linked to longevity in the NFL. The lift chart for this model in Figure 4.2 shows that this model does a very good job at finding players who are very unlikely to play at least 16 games, but struggles at finding players who are very likely to play 16 games.

The probability a quarterback would start a playoff game in his NFL career was then modeled using logistic regression. While the previous model shows positive traits towards longevity, what every player, coach and fan want is success in the playoffs. The goal of this model is to find variables that are predictive of postseason success and not just impressive statistics, and then to be able to use this model to separate the players who are likely to earn a start in a playoff game from the rest of the group. Two of the 4 variables used in this model were also used in the Draft Selection model. Weight and Completion Percentage are both positive indicators of a quarterback being drafted early and a quarterback starting a playoff game. The other two variables used in the model, Rushing Yards per Game and Maximum Rushing Yards per Carry were a bit surprising considering the emphasis put on quarterback's passing statistics. Also surprising is that Rushing Yards per Game is a positive indicator for Playoff Start while Maximum Rushing Yards per Carry is a negative indicator (these two variables are also positively and negatively correlated, respectively, in the univariate case). While NFL teams are not drafting based on the exact same variables that lead to starting a playoff game, the similarity in the variables used in the models shows that teams do value the same quarterback characteristics that lead to playoff success. The lift chart created by this model shows a very interesting trend: the model fits very well in the tails, but the middle 70% is almost just fitting the average. This model is able to identify 25% of players who have not started a playoff game in just the first 20% and 21% of players who have started a playoff game in just the top 10%. The lift curve sits at 25% from the third decile to the ninth decile, expect for deciles 4 and 5 where it drops to 20%. Usually lift charts show more of a consistent upward trend, however since this model is aimed to find quarterbacks who will definitely start a playoff game and who will definitely not start a playoff game, the main focus of this model is the have good separation in the tails.

The fourth model built was another Tobit model, this time modeling Quarterback Rating (QBR). Quarterback Rating is a very common metric used by NFL analysts (a description of Quarterback Rating can be found in Section 2.2). This model includes 4 variables, none of which were used in the Draft Selection model. The QBR Tobit model includes two positive indicators (Maximum Completion Percentage and Height) and two negative indicators (Rushing Yards per Attempt and Maximum Interception Percentage).

While the information deemed important in this model does not necessarily resemble the information deemed important in the Draft Selection model, there is important information to be learned from this model. Completion Percentage and Interception Percentage are both used to calculate Quarterback Rating, so including players's Maximum Completion Percentage and Maximum Interception Percentage from college to model their Quarterback Rating from their NFL career shows that a quarterback who has shown he can complete a high percentage of passes in college is likely to keep that trait in the National Football League. However, if a quarterback shows he throws interceptions at a high rate in college, this model says he is unlikely to correct this mistake in the NFL. The lift curve produced by this model shows a general upward trend, however there are a few dips, showing a fairly inconsistent model fit.

The final Tobit model built in this thesis modeled the log of the Principal Component Rating (PC Rating). Principal Components Analysis (PCA) was utilized to attempt to provide a more robust alternative metric to Quarterback Rating, which is calculated with only 4 statistics. The target variable was log transformed after plotting the target variable against a few predictors and noticing the relationship was much more linear after taking the log of the PC Rating. The model included only two variables: Passing Yards per Game and Height, both positive indicators. Unfortunately, the lift chart of the log(PC Rating) model was very inconsistent, showing very volatile trends in predicting a quarterback's log(PC Rating).

While this thesis showed progress and answered a few broad questions, more observations and more variables would be desirable and would hopefully lead to stronger conclusions. Perhaps the largest piece of information overlooked in this study was the college program the quarterback came from. This includes everything from the style of offense the quarterback ran in college to the level of defensive play the quarterback faced. When examining the quarterbacks in the NFL, it becomes very apparent that the NFL is possibly much more of a team-driven league than originally thought and quarterbacks may be a product of the system. The focus of the NFL and especially the NFL draft is all about the quarterback, but if nothing else, perhaps this thesis concludes that there is not one recipe for success in the National Football League which may be reason enough to use early draft picks to select players at more reliable positions and take a chance on a quarterback in a later round. While it is difficult to compare lift charts of different types of models, the 16+ Games Played logistic model is probably the most informative model, being able to find 25% of the quarterbacks who have not played in a playoff game in the first 20% of the model and identifying 21% of the quarterbacks who have played in a playoff game in the top 10% of the model.

As difficult as it is to admit, perhaps the key to success in the NFL is more intangible than statisticians would like to believe. As statisticians, mathematicians and self-proclaimed number-gurus, we like to believe that everything can be quantified. And maybe everything can be quantified, but maybe just not we can be accessible via the internet. Quarterbacks are leaders of the team and maybe their personal and leadership qualities, along with other intangibles are more important to team success than tangible metrics from college statistics.

Appendix A

Boxplots of Variables Used in Models

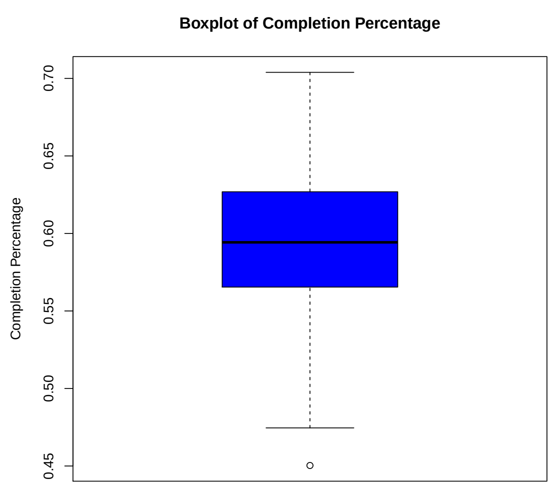

Figure A.1: Boxplot of Completion Percentage

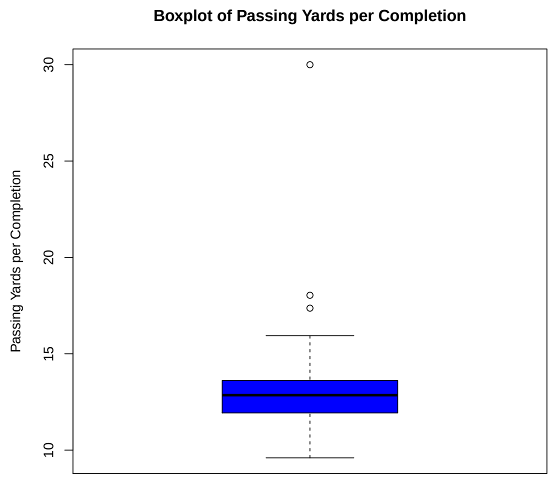

Figure A.2: Boxplot of Passing Yards per Completion

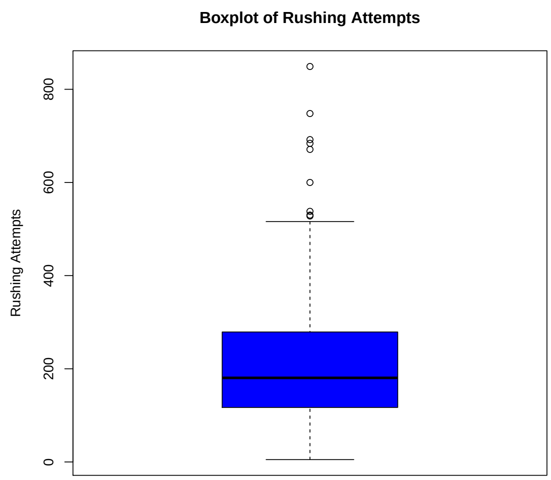

Figure A.3: Boxplot of Rushing Attempts

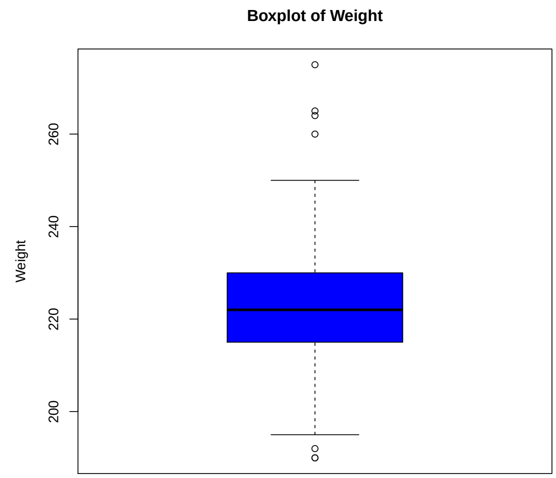

Figure A.4: Boxplot of Weight

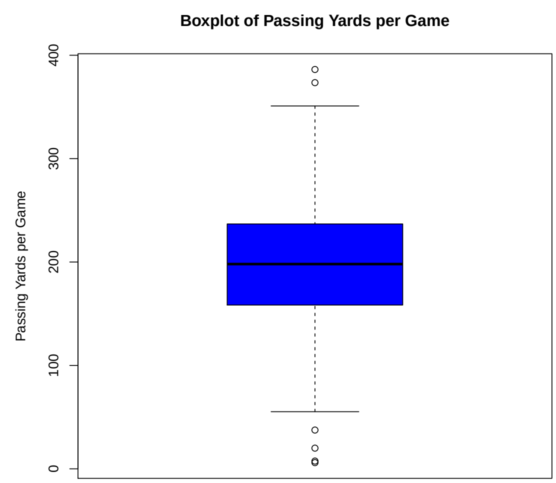

Figure A.5: Boxplot of Passing Yards per Game

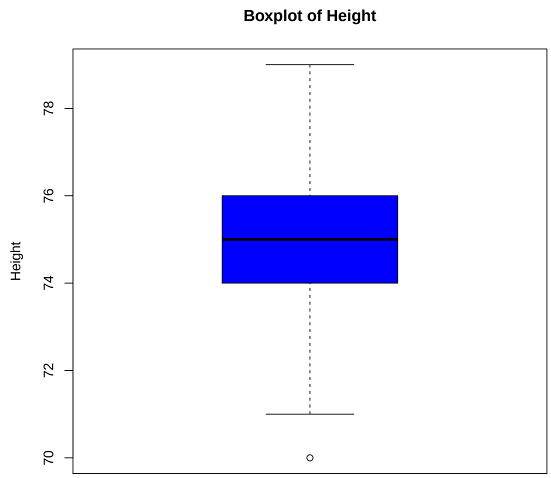

Figure A.6: Boxplot of Height

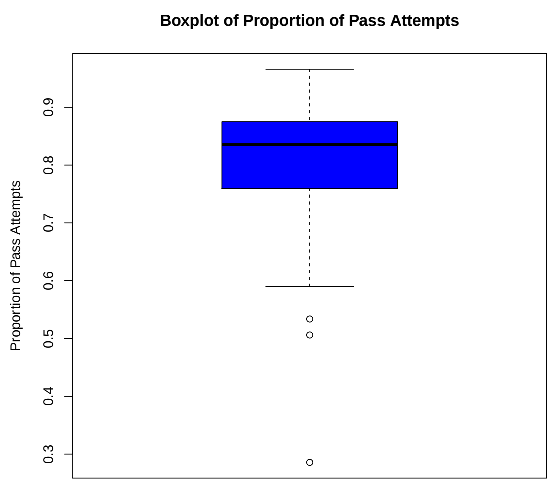

Figure A.7: Boxplot of Proportion of Passing Attempts

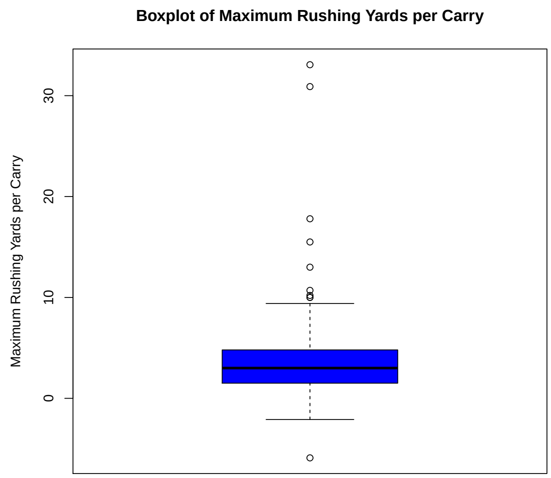

Figure A.8: Boxplot of Maximum Rushing Yards per Carry

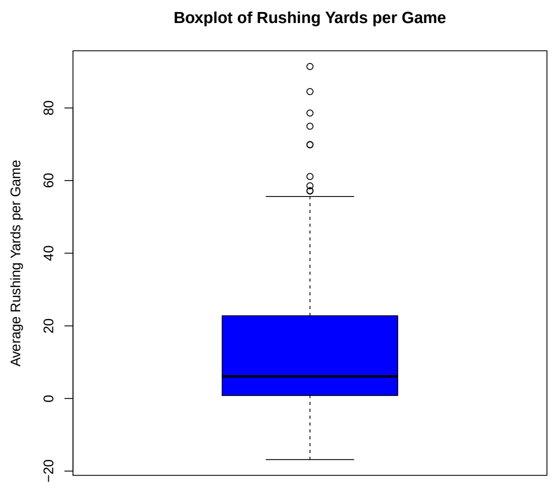

Figure A.9: Boxplot of Rushing Yards per Game

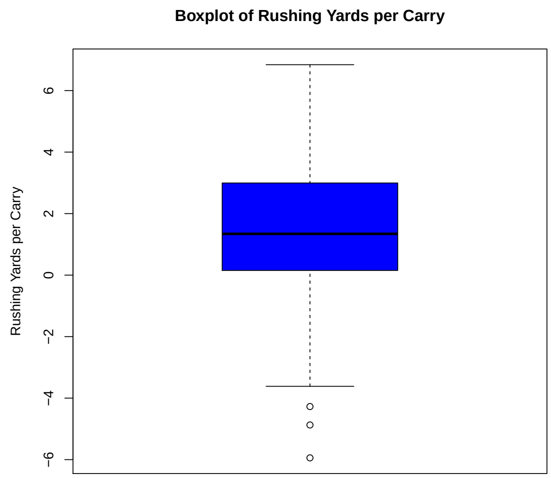

Figure A.10: Boxplot of Rushing Yards per Attempt

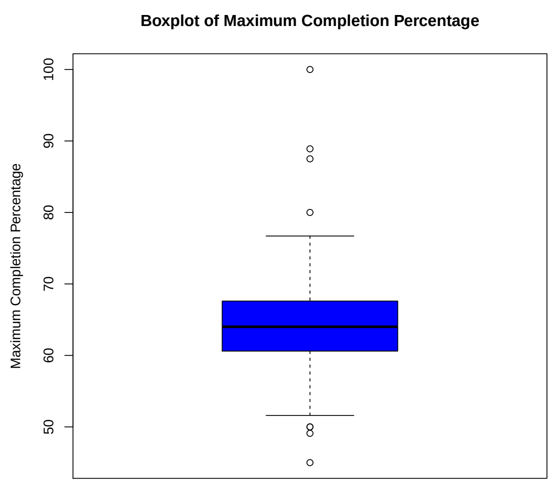

Figure A.11: Boxplot of Maximum Completion Percentage

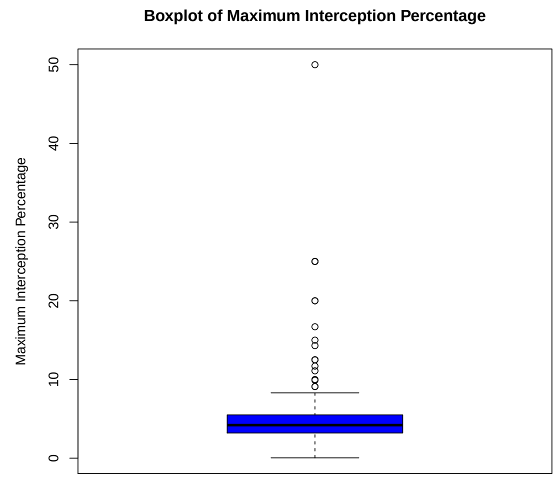

Figure A.12: Boxplot of Maximum Interception Percentage

Bibliography

Addona, V., S. R. H. and Wolfson, J. (2011). The Quarterback Prediction Problem: Forecasting the Performance of College Quarterbacks Selected in the NFL Draft. *Journal of Quantitative Analysis*, 7:1–19.

Berry, S. and White, C. (2002). Tiered Polychotomous Regression: Ranking NFL Quarterbacks. *The American Statistician*, 56(1):10–21.

Dallas, G. (2013). Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction. `https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/` [Accessed: January 2016].

Gagnon, B. (2015). Which Positions Are the Safest, Riskiest at the Top of the NFL Draft? `http://bleacherreport.com/articles/2441018-which-positions-are-the-safest-riskiest-at-the-top-of-the-nfl-draft` [Accessed: January 2016].

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company, New York, NY.

Lund, B. and Raimi, S. (2011). Before Logistic Modeling - A Toolkit for Identifying and

Transforming Relevant Predictors. http://www.mwsug.org/proceedings/2011/stats/MWSUG-2011-SA03.pdf [Accessed: March 2016].

McDonald, J. and Moffitt, R. (1980). The Uses of Tobit Analysis. *The Review of Economics and Statistics*, 62(2):318–321.

NFL (2015). http://www.nfl.com/draft/history/fulldraft?type=position [Accessed: January 2015].

Penn State, E. C. o. S. (2016). 12.4 - Detecting Multicollinearity Using Variance Inflation Factors. https://onlinecourses.science.psu.edu/stat501/node/347 [Accessed: March 2016].

Pro-Football-Reference (2015). http://www.pro-football-reference.com/ [Accessed: January 2015].

Stimel, D. (2009). A Statistical Analysis of NFL Quarterback Rating Variables. *Journal of Quantitative Analysis*, 5:1–22.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econimetrica*, 26(1):24–36. http://www.jstor.org/stable/1907382?seq=1#page_scan_tab_contents [Accessed: March 2016].