

Final Paper - Rough Draft

Dillon Beake

Course Code and Name

Dr, Ransom

Date

Introduction

Strength is the cornerstone of athletic development, and power is the cornerstone of athletic performance (Haff & Triplett, 2016). The athlete who is stronger can apply more force (Kraemer & Fleck, 2005, p. 49). This enhanced force application can be seen when a running back drags half the team with him before finally being tackled, when a sprinter overtakes her competitors with large strides at a faster stride frequency as she applies greater force to the ground, or when a tennis player launches a 100+ mile per hour serve. Power is defined as $\text{Work} / \text{Time}$, and Work is defined as $\text{Force} \times \text{Distance}$ (Baeckle & Earle, 2008, p. 74). In other words, power is a measure of how fast you can move as much weight as possible. So, by definition, it is understandable that these would be extremely important traits to possess as an athlete. However, much more goes into becoming a great athlete. Some other well-known attributes are hard work, discipline, perseverance, staying injury free, and a little bit of luck. Becoming a great, even a professional athlete is rare, and that rarity is very apparent when considering professional football. During the year 2022-2023 there were over one million high school football players, and of that one million, only 3% have the chance to go on and play Division I football; furthermore, of that 3%, the likelihood of being drafted into the NFL is roughly 1.5% (NCAA, 2024). Furthermore, roughly 250 football players are drafted each year. So, being drafted is quite rare and the draft decisions typically made as a result of the NFL combine are important. Furthermore, football is the quintessential game of strength and power. It's a physical game where more often than not, with all else being equal, the

stronger and more powerful player wins. However, finding the right players to successfully fill these roles is not only difficult, many times it's surprising. Tom Brady, for example, is arguably one of the best football players of all time. He did not start in college until his junior year, and he was the very last pick in the draft. This type of story is not unusual and can unfortunately go the other way as well. So, what does strength, power and the draft have to do with this project? It comes back to the concept that if all things being equal the stronger, more powerful player usually wins. They win the block, the tackle, the race to the ball, even the fight to catch the ball. I will leave the majority of the determinants of what equalizes one player to another to the coach's knowledge and experience. However, the concept of this project is to use data to evaluate which athlete might have the most potential to develop strength and power. Thus, giving a coach more insight into who the best option might be. For this task, I will be analyzing the NFL combine data from the years 2004-2024. I chose the NFL combine data because it demonstrates a variety of anthropometric measurements, e.g. height and weight: as well as, know performance indicators, e.g. vertical jump test. I am interested in assessing how much of the variance in the vertical jump is described by BMI ($\text{Weight (kg)} / \text{Height (m}^2\text{)}$), hand size, and the Arm Length / Height ratio. I chose the vertical jump because according to Baechle and Earle (Baechle & Earle, 2008), the vertical jump is directly correlated to lower body muscular strength and power. I want to test if, and to what degree, these independent variables correlate to enhanced performance in the vertical jump.

Literature Review

I will finish the Literature this week. I plan to review the following, but not limited to, articles:

- Teramoto, M., Cross, C. L., Rieger, R. H., Maak, T. G., & Willick, S. E. (2018). Predictive validity of National Basketball Association draft combine on future performance.

Journal of Strength and Conditioning Research, 32(2), 396-408.

- Robbins, D. W. (2012). Relationships between National Football League combine performance measures. *Journal of Strength and Conditioning Research*, 26(1), 226-231.
- Ackland, T. R., Schreiner, A. B., & Kerr, D. A. (1997). Absolute size and proportionality characteristics of World Championship female basketball players. *Journal of Sports Sciences*, 15(5), 485-490.
- Torres-Unda, J., Zarrazquin, I., Gil, J., Ruiz, F., Irazusta, A., Kortajarena, M., ... & Irazusta, J. (2013). Anthropometric, physiological and maturational characteristics in selected elite and non-elite male adolescent basketball players. *Journal of Sports Sciences*, 31(2), 196-203.
- Bergman, S. A., & Logan, T. D. (2016). The effect of recruit quality on college football team performance. *Journal of Sports Economics*, 17(6), 578-600.
- Dumond, J. M., Lynch, A. K., & Platania, J. (2008). An economic model of the college football recruiting process. *Journal of Sports Economics*, 9(1), 67-87.
- Berri, D. J., & Simmons, R. (2011). Catching a draft: On the process of selecting quarterbacks in the National Football League amateur draft. *Journal of Productivity Analysis*, 35(1), 37-49.

Data

My main source of data came from nflcombineresults.com. I utilized the 2004 – 2024 NFL combine data from the website. This is a large data set with sample size of over 10,000, after the missing items were accounted for (see table 1). The dependent variable is one of the NFL combine day's Key Performance Indicators (KPI), the vertical jump test. TI was

also interested in how different anthropometric measurements might explain the variance in the vertical jump test. The independent variables include BMI (weight (kg) / height² (m²)), hand size (in), and an Arm Length/Height ratio. The development of the arm-length/height variable is based on my previous recruiting experience concerning wing-span measurements, and from the research conducted by Shreiner and Kerr (**ackland1997absolute**). I decided to convert height and weight to BMI (Basal Metabolic Index) to reduce the chances of multicollinearity between height and weight, and height and the Arm Length/Height Ratio. Also, it should be noted that this ratio is a substitute for wing-span, which was not included in the data set. I am particularly interested in how much these variable correlates to vertical jump performance.

Empirical Methods

The first regression model I used was the linear regression model. The linear regression model is written as follows:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + E$$

The second regression model I used was log regression model. The general log regression is written as follows:

$$\log(Y) = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + E$$

Finally, I used the k fold cross validation model. The k fold cross-validation model is written as follows:

$$CV(k) = 1 \frac{1}{k \sum_{i=1}^k L_i}$$

The linear regression model consists of the dependent variable (Y), the independent variables (X), and coefficients (B). In this case the Y . or the dependent variable of interest is the Vertical Jump. X_1 , X_2 , and X_3 are the independent variables, and the variables that will be used to predict Y . The independent variables are BMI (X_1), Arm Length/ Height Ratio (X_2), and Hand Size (X_3). B_0 is the y-intercept, this is the value of y when all the in-

dependent variables are zero. B_1, B_2, B_3 are the coefficients for each predictor variable. The coefficient can be interpreted as when holding all other variable constant, one unit change in the independent variable changes the Y variable by the amount of the coefficient (Source). For example, if $X_1 = \text{Arm Length/Height ratio}$ and $B_1 = 2.25$, for every one unit change of BMI (X_1) than Vertical Jump (Y) is changed by 2.25 being multiplied by the new value of BMI (X_1).

The log regression model was used to address the possibility of a non-linear relationship. The data comes from a fixed group of high-level athletes. The data is not taken from the general population. However, the large sample size should offset most of the issues of using such a closely related sample. Also, the athletes are all different shapes and sizes which is similar to the general population. When checking for skewness and kurtosis the variables of interest in the data were only slightly outside of the bounds of normalcy (see table 2). However, the log regression model helps analyze a non-linear relationship by taking the dependent (Y) variable and transforming it by taking its logarithm. The logarithm of vertical jump (Y) will help to compress the wide range of results seen by the thousands of NFL combine participants over the past 20 years; while, creating a linear relationship between vertical jump and the predictor variables.

The k fold cross validation model was used to determine which of the regression models was the best fit for the NFL combine data. Also, which of the models was better at predicting the variance in vertical jump. I will further flesh the k fold cross-validation model details out over the week.

Discussion

The Linear Regression model was used to predict Vertical Jump based on three independent variables including BMI, Arm-Length/Height Ration, and Hand Size. All three independent variables were statistically significant predictors of Vertical Jump, with p-values <0.05 . The

coefficient for BMI was -0.57; therefore, a one unit increase in BMI will result in a 0.57 in. decrease in vertical jump. The coefficient for Arm.Length/Height Ratio was 35.91; therefore, for every increase in one unit of this ration results in a 35.91 in. increase in vertical jump. The coefficient for Hand Size was -0.18; therefore, with every one unit increase in hand size there will be a 0.18 in. decrease in vertical jump. The intercept is 36.47, which represents the vertical jump when all independent variables are zero. The 5-fold cross-validation results for the linear regression model showed an RMSE of 3.298681. The R-squared value was 0.3841652. This indicates that 38.42% of the variance in vertical jump can be explained by the three independent variables. The results indicate that the Arm-length/Height ratio has the largest impact on Vertical Jump. It also makes sense that a larger hand size, and increased BMI will decrease vertical jump, because a larges hand, and increased BMI indicate a larger body size. Furthermore, the k value cross-validation shows moderate predictive performance.

The log regression showed similar results to the linear regression model. All three independent variables were statistically significant, with p-values being less than 0.05. In order to compare the effects of the log coefficients back to the original scale of vertical jump they will need to be exponentiated, a multiplicative change. The coefficient for BMI was -0.02; therefore, for every one unit increase in BMI results in a 0.02 decrease in the logarithm of Vertical Jump. The multiplicative change for this coefficient is $\exp(-0.02) =$ approximately 0.98, therefore for every .02 increase in BMI there will be a 0.98 inch decrease in Vertical Jump. The coefficient for Arm-Length/Height Ratio is 1.10; therefore, for every one unit increase in Arm-Length/Height Ratio results in a 1.10 increase in the logarithm of Vertical Jump. The multiplicative change for this coefficient is $\exp(1.10) =$ approximately 3.00; therefore for every one unit increase in Arm-Length/Heighth ratio results in a 3.00 in. increase in vertical jump. This is still very promising and also more realistic that tan the 36+ inch increase that was reported in the linear regression model. The coefficient for Hand Size was -0.01; therefore, for every one unit increase in hand size results in a 0.01 decrease in the logarithm of

Vertical Jump. The intercept was 3.62, which represents the expected logarithm of Vertical Jump when all independent variables are zero. The 5-fold cross-validation results showed an average RMSE of 0.1036537, which is much lower than what we saw with linear regression model. The average R-squared was 0.3937993, which indicates that 39.38% of the variance in the logarithm of Vertical Jump can be explained by the three independent variables. The results of the k fold cross-validation show that the log regression model has a lower RMSE, and a slightly higher R-Squared value when compared to the linear regression model. Both models have similar explanatory powers; however, the log regression model appears to be the better fit.

References

- Baechle, T., & Earle, R. (2008). *Essentials of strength training and conditioning* (3rd ed.). Human Kinetics.
- Haff, G., & Triplett, N. (Eds.). (2016). *Essentials of strength training and conditioning* (4th ed.). Human Kinetics.
- Kraemer, W., & Fleck, S. (2005). *Strength training for young athletes*. Human Kinetics.
- NCAA. (2024, April). Football: Probability of competing beyond high school.