40 Time Analysis

Nicholas Sowers, Avery Ryan, Drew Polito, Tommaso Villa

Nicholas Sowers: Data Summary and Results Step Stones

Avery Ryan: Intro/Conclusion

Drew Polito: Empirical Methodology

Tommaso Villa: Data Visualization

The question at hand is to what degree, if at all, does a college football running backs 40 yard dash time dictate their success in the NFL draft. There are numerous determinants of running backs getting drafted, including production, team, size, and lastly speed. Understanding that many variables impact success makes the question important because we want to know which variables are most important and which aren't at all. If we figure this out then scouts and players can know what to look for or train for. The question can be continued after our tests because we are only looking for one variable, speed. To measure speed the 40 yard dash is the best and easiest to examine for college athletes and using those times we can see if there is a correlation and how strong the correlation is between success. In our project we will use whether or not a player gets drafted after college as our measure of success. This will be a dummy dependent variable in the experiment. The independent variable, 40 time, will be the subject of our testing and in order to create a hypothesis for this topic, we will look at only significantly fast and slow running backs. Our definition of a fast 40 time will be one that is less than 4.4 seconds and a slow 40 time for running backs will be one that is greater than 4.6 seconds. The data was gathered from pro football reference and fantasypros from between 2000 and 2019. We then used the data to run t-tests for the difference of the means using excel. Our findings implied that there was a significant difference between the means which suggests speed is an important factor in determining draftability.

To start with our data summary, it is important to reiterate that our data came from pro football reference and fantasypros. These websites contained all of the running backs that participated in the

combines, and they had other information, like where the team got drafted and when they were picked. Pro football reference covers many aspects to football like team stats, defensive stats, and game scores. We initially obtained over 500 points of data, for every running back at the combine from 2000-2019. Since what we are looking for is to see the correlations in fast and slow running backs in getting drafted, we cleaned the data to contain rbs who ran faster than a 4.4 and slower than a 4.6. This filtered the amount of data points down from 503 to 206. We formatted the data in both google sheets and excel and we formatted it by year ascending.

From Table 1 we were provided with our descriptive statistics. We decided to format multiple sets of descriptive statistics, as we deemed multiple to be appropriate. We highlighted the mean 40 Time to be around 4.64 with our cleaned data, which would be lower if we included the data points from 4.4-4.6. The median was around 4.66 which was slightly greater than the mean. It made us curious to see if the comparison could be an indicator for our data to be left skewed. We also noted how our standard errors were low for the summary stats. In the middle columns of the graph we referenced the descriptive statistics for the dummy variable for being drafted. The median is 0 here because in all of the data more people went undrafted. With the mean being .4369, less than half of the total participants in our cleaned data got drafted. The stats for the drafted dummy variable in the faster group tell a different story. The mean was .7 showing the chance of getting drafted while in this group is higher.

Moreover, our group found the visuals to tell an interesting story. Visual 1 is an introduction to what our data looks like involving frequency. More participants in our data ran over a 4.6 and the slowest time was around 5 seconds. For comparison, Peyton and Eli Manning ran around 4.88 and 4.92 seconds in the 40 Yard Dash. Visual 2 is less correlated to the relationship of 40 Time and success, but it shows the most popular schools with people in the combine in our data. Our group was not surprised to see SEC schools like LSU, Auburn, and Alabama at the top. Visual 3 showed our group the probabilities of getting drafted based on running a specific time. If you ran between a 4.3 and a 4.4, it looks like there would be a 75% chance of being drafted, while if someone ran between 4.8 and 4.9 seconds, the chances of being

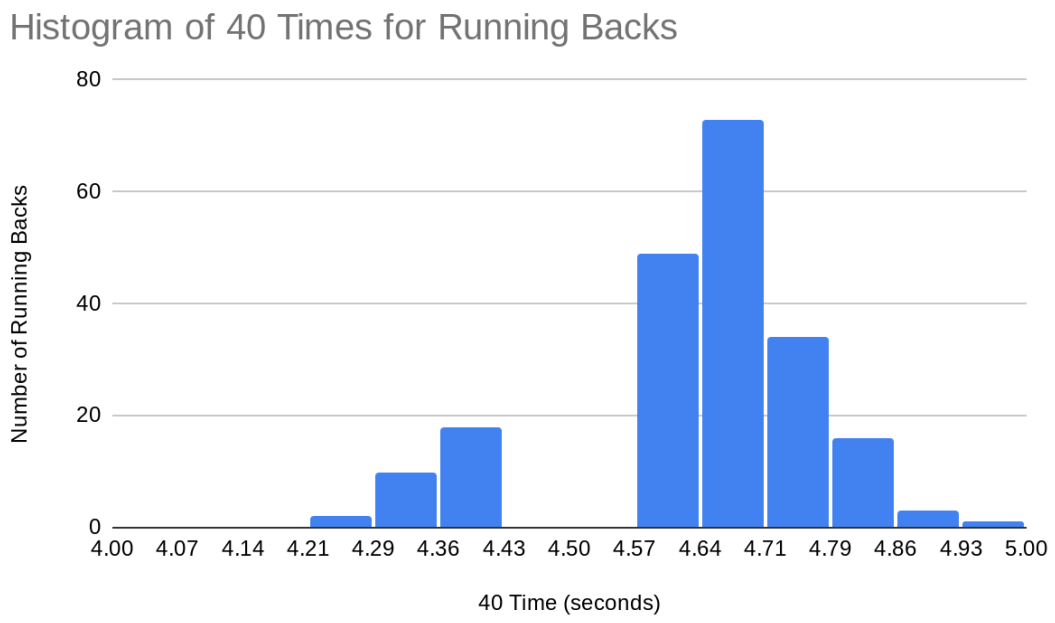
drafted go down to 25%. The red points on the graph represent what actually happened, as a participant in the combine either gets drafted or does not.

Table 1: Summary Stats

Summary Stats 40 Time	Values	All Data: Drafted Dummy Variable	Values 2	Under 4.4: Drafted Dummy Variable	Values 3
Mean 40 Time	4.6407	Mean	**0.4369**	Mean	**0.7**
Standard Error	0.00937	Standard Error	0.03464	Standard Error	0.085096
Median	4.66	Median	0	Median	1
Standard Deviation	0.13453	Standard Deviation	0.49721	Standard Deviation	0.46609
Count	206	Count	206	Count	30

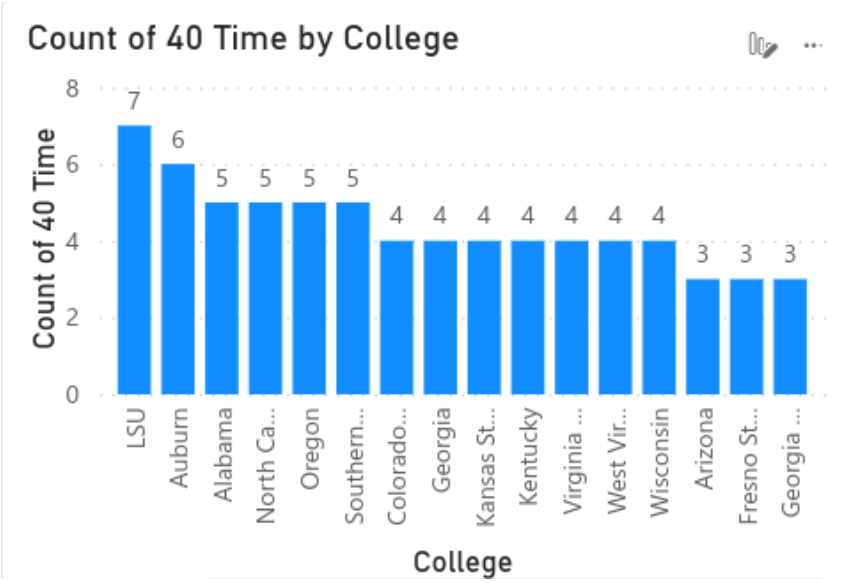
Note: The bolded means represent the average of being drafted to the NFL. One means they got drafted, and zero means they were undrafted. With the running backs who ran under a 4.4, the mean was higher than the mean of all cleaned data, making a case for correlation.

Visual 1: Graph of Data Distribution



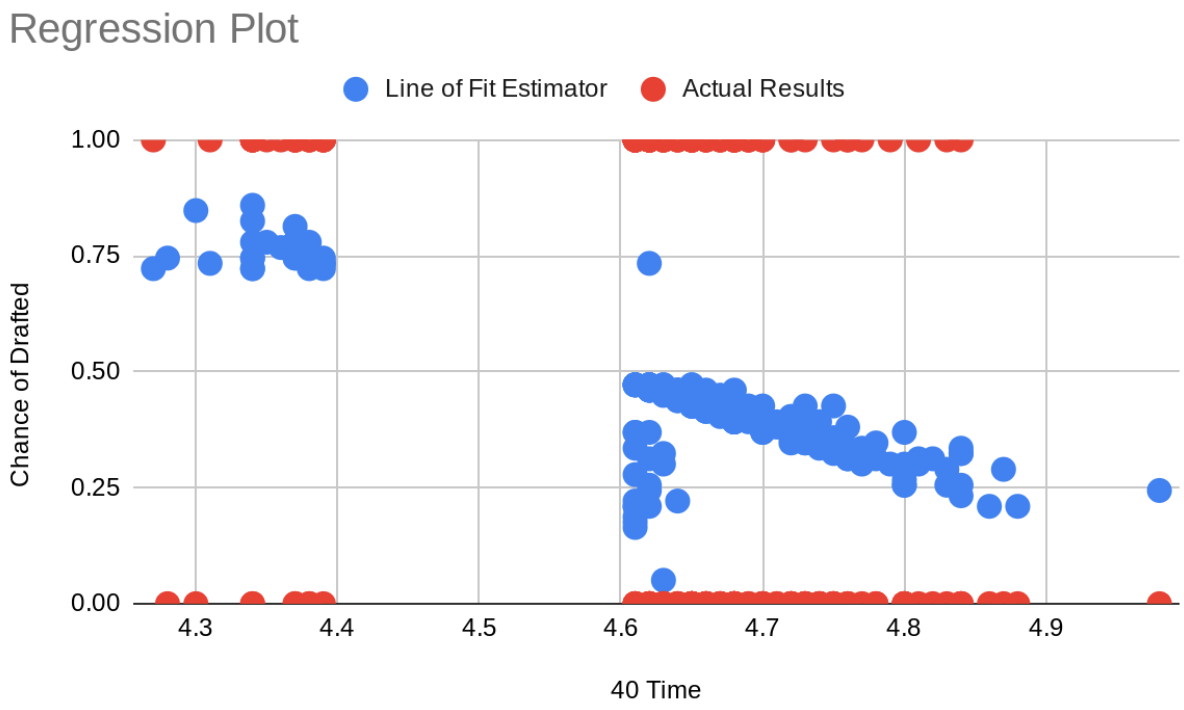
Note: There is no data from 4.43 to 4.57 in this graph because it represents our cleaned data, which is faster than 4.4 and slower than 4.6.

Visual 2: Frequency of 40 Yard Dash Times By School



These colleges had the most people at the combine from 2000-2019 with our cleaned data.

Visual 3: Regression Line



For empirical methodology we used our goal of determining the impact of speed on draftability for prospective NFL running backs, and our methods were designed to explore that relationship. As such, the data range analyzed includes RB 40 times from the NFL combine from 2000-2019, excluding those running backs who ran between a 4.4 second and 4.6 second 40-yard dash. Our reasoning behind this exclusion was the scope of our analysis: in particular, we intended to focus on if being a "notably fast" (for our purposes, a less than 4.4 second 40) or a "notably slow" (more than 4.6 second 40) running back had a statistically significant impact on draftability for NFL teams. To that end, our methodology was hypothesis testing on the draftability of notably fast and notably slow running backs, specifically meaning whether or not they were drafted whatsoever. For this, we conducted hypothesis t-testing. Furthermore, regression analysis was conducted to determine the relationship between the time that a running back recorded in the 40 and the average percent chance that they would be drafted.

Our analysis included t-tests to compare draft statistics for "fast" and "slow" running backs and determine if there was a statistically significant difference between the two, using the following formula. Where:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s^2(\frac{1}{n_1} + \frac{1}{n_2}))}}$$

X1, X2: The means of the respective groups ("fast" and "slow")

S^2: The pooled standard error of the "fast" and "slow" running back draft statistics

N1, N2: The number of observations in the respective groups ("fast" and "slow")

Our analysis is looking for confidence at a 95% confidence interval, where t = 1.96.

Furthermore, we also conducted regression analysis of the "fast" and "slow" times to determine the relationship, if any, between speed and draftability of running backs, using the following formula. Where:

$$Y = b_0 + b_1 \cdot X$$

Y: The dependent variable, in this case a dummy variable representing whether a player was (1) or was not (0) drafted. We took the aggregate of this as the percent chance of being drafted given a certain time.

b_0: The constant, in this case representing the percentage chance of a player being drafted given a 40 time of 0. Seeing as this time is physically impossible, we saw fit to disregard the constant in our analysis.

b_1: The coefficient for speed in the 40-yard dash, indicating how much the change in one's time impacted the percentage chance of being drafted.

X: The independent variable, in this case the time that a running back has in the 40-yard dash.

The results of our testing produced coefficient and standard error estimates, as well as an R squared value for the data. The intercept of our data was about 5.63 with a standard error of 1.15. The interpretation of the intercept is that when a player runs a zero second 40 time he has an estimated 563% chance of getting drafted. A lot here is quite illogical as both running a zero second 40 yard dash and obtaining over a 100% chance of getting drafted is impossible. It is important to note that the intercept in this case should not be taken as a literal, but rather an estimation that makes our regression line fit the data. In terms of significance in the model, the coefficient is significant at the 5% level because 5.63/1.15=4.90, which results in a t-stat greater than the critical value. Adding on, the 40 time statistic has a coefficient of about -1.12 and a standard error of 0.249. The interpretation of the coefficient is that for each additional second of a running back's 40 time, there is a predicted 112% decrease in the probability that the player is drafted. Once again this value seems illogical but considering the range of 40 times for running backs is less than 1 second, we can adjust our interpretation to a 0.1 second increase in 40 time for a running back decreases the predicted probability of getting drafted by 11.2%. This interpretation appears more reasonable for the data. The coefficient is also statistically significant because 1.12/0.249=4.50 which is also greater than its t critical value. The final part of our results of importance is the R squared value which represents the goodness of fit in the model. Our estimation gave us an R squared value of 0.65 which means that our model has a moderately good fit considering R squared ranges from 0 and 1 with 1 being a perfect fit. Overall, the real world implication of the model is that we

are 95 percent confident in the negative correlation between additional time in the 40 yard dash and draftability so speed should be considered an important trait for collegiate running backs.

Table 2: Correlation Statistics

Categories/Variables	Coefficients	Standard Error
Intercept	5.631569655	1.154861093
40 Time	-1.119663492	0.248685211
R Square	0.650541872	
Observations	205	

To conclude, after performing our tests and examining the data we came to the decision that speed is a significant variable relating to success for running backs after college. The t score of 2.864 being greater than 1.96 contributes to that assumed difference in draftability at the 5 percent level. This means that focusing on speed training is a smart idea for collegiate running backs looking to turn pro as achieving a 40 time time under 4.4 seconds is beneficial. Additionally, organizations should be more aware of their emphasis on speed so that they can differentiate between real game speed and combined speed in the future. The results, however, are not perfect and can be improved. Controlling for other factors like size, school, and production would help us avoid omitted variable bias and isolate the effect of speed more effectively. We could also collect more data to ensure we are not at risk of other errors. Nonetheless, our final conclusions from our tests is that speed and success are positively correlated.

Sources:

https://www.fantasypros.com/nfl/stats/combine/rb.php?year=2000

https://www.pro-football-reference.com/draft/2019-combine.htm