The NFL draft is coming to Pittsburgh. So, to not fall behind on the latest statistics, I wanted to answer one of the questions I hear year after year. Does a greater number of years in college equate to better success in the NFL, measured by contract size? To answer this, the main research question for my project is. Does the number of years a Quarterback plays in college entering the NFL affect their largest contract in the NFL?

To answer this question, the data I collected comprised quarterbacks drafted in the first round between 2011 and 2022 who have played 1-5 years in college and their largest contract. The data I collected comes from two sources: Over The Cap, which provided me with the contract history of each NFL player, and ESPN, which provided me with the college stats and years played in college. The biggest problem with collecting the data was determining the difference between the years started and the years of eligibility. While sitting behind a veteran quarterback is helpful to players, I chose to limit the data to years started by having a minimum of seven games started.

Because I created the data and kept the variables to one total variable, I didn't clean my data. However, just to be safe and make sure my data was clean and accurate, I used Gemini to help create the cleaning code. The code used was to remove any unwanted characters in my data. The code I used did four important things: replace spaces with underlines, make all columns lowercase, remove characters that aren't letters or numbers, and remove leading/trailing spaces.

My method to answer this question was by using a linear regression model. This model specifies "largest_contract" as the dependent variable and "years_in_college" as the sole predictor. By running this model, I was able to determine how each additional year played in college would affect the expected contract size. The intercept represents the estimated contract value for an athlete who spent zero years in college, while the slope represents value changes with additional years played. There are many limitations to the model I ran; the model assumes homoscedastic errors and independence. These assumptions are present due to the small sample size and the heavy-tailed distribution present due to larger contracts. Because there is only one predictor, years in college, the model could be improved by adding more relevant stats, such as a player's performance in college.

The regression results indicate that years_in_college is not a significant predictor of a quarterback's largest contract. The estimated slope coefficient is 1.52×10^7, meaning that for each additional year played in college, there is an increase of around $15 million in contract value. However, this estimate is highly uncertain as the p-value for the coefficient is 0.523. This fails the 95% confidence interval as the ranges span from negative to positive values. This

shows that the effect of additional years played has on the contract value could be negative, positive, or effectively zero. The model also has a very low R-squared value of 0.014, proving that only about 1.4% of the variation in contract values is explained by years played in college.

A residual plot would likely show scatter with no pattern, showing that the model does not explain a correlation between the two variables. Overall, the model's validity is limited. The model fails to utilize any other predictors that would play a bigger role in determining how to measure a quarterback's success (performance in college). The small sample of 31 quarterbacks also restricts the power of this model. While the model does show a small increase in contract value, the results should not be interpreted as evidence that years in college have a meaningful effect on contract size.

To conclude, the purpose of my model was to determine whether the number of years a quarterback played in college affects their overall success in the NFL, measured by their contract value. To test this, I gathered the largest contracts of first-round quarterbacks from 2011-2022 to eliminate any quarterbacks who are still on their rookie contract. I also eliminated factors of quarterbacks who sat on the bench or red-shirted to only take into account the years they played as a starter in college. The results of my model show that years in college are not a significant predictor of contract size. The effect was small relative to the variation in the data and was statistically insignificant. The model explained almost zero correlation between the difference in contract values across quarterbacks. I do still believe that years_in_college could affect a quarterback's success, however, not by itself. If I were to run another model, I would aim to include other predictors such as average passing yards per game and turnover ratio.

https://overthecap.com/position/quarterback
https://www.espn.com/nfl/player/stats/_/id/3046779/type/college-football
https://www.sports-reference.com/cfb/players/kenny-pickett-1.html