Final Project Summary

"NFL Combine" Player Analysis

Dan Zylkowski

Bellevue University

DSC 530: Data Exploration and Analysis

May 30, 2020

As an avid NFL fan, I had a general feeling that players had gotten larger and faster in recent years. This project was spawned from the idea of how to quantify the physical evolution of NFL players since 1987. I used Exploratory Data Analysis (EDA) to investigate the following research question:

Hypothesis: Have NFL players gotten bigger, faster, and stronger in the past 33 years?

The outcome of the EDA

The histograms of several of the variables appeared bell-shaped, but only the Height variable followed a normal distribution. The histograms indicated the existence of outliers in several of the variables. Since each position group has different physical attributes, the positions with similar sizes were grouped into three separate data frames, and then outliers were identified for each position grouping. For each variable, outliers were defined as any value less than the mean minus three times the standard deviation or greater than the mean plus three times the standard deviation. After researching the 224 outliers, 12 values were corrected.

The Position groupings were then separated into two time periods: (1987-1999) and (2000-2020). For each position, the percentage change between the two time periods for each measured variable (metric) was visualized with bar plots. The following changes in means were observed:

- **~ 20% increase** in the Bench_Press repetitions across almost all positions.
- **~ 5% increase** in the Vert_Leap Height across most positions.
- **~ 3% increase** in the Broad_Jump distance across most positions.
- **~ 3% increase** in Weight across almost all positions.
- **~ 1.5% decrease** in the Forty_Yard times across most positions.
- **~ 1% decrease** in the Shuttle times across most positions.

The Bench_Press, Vert_Leap, and Weight averages increased across all positions, while the Broad_Jump average increased in all but one position (C), and the Forty_Yard average decreased in all but two positions (OG and DT).

Of all positions, the Tight End (TE) position had the largest percentage change for the Broad_Jump, Vert_Leap, and Forty_Yard. Additionally, Tight Ends had the second-largest percentage change in Height (behind Wide Receivers), and Shuttle time (behind Outside Linebackers). These changes could be due to how the Tight End position has evolved from a primary role of blocking to a new role as a larger version of a Wide Receiver.

The PMF of the Bench_Press variable gave visual confirmation of the difference in Bench_Press repetitions. The 1987-1999 group of players had a higher percentage of players performing fewer reps, and the 2000-2020 group had a higher percentage of players performing higher reps. These outcomes align with the 20% observed increase in the bench press variable across most position groups.

The CDFs of player weights visually confirmed that players from 1987-1999 are lighter than players from 2000-2020 throughout the distribution. Similarly, the CDFs of the Forty_Yard variable demonstrated that players from 2000-2020 are faster than players from 1987-1999 throughout the distribution.

The normal probability plot showed that the Height variable followed the normal distribution model from -3 standard deviations up to almost +2 standard deviations before deviating at both tails. This means that the normal distribution is a good model for the height variable.

Scatterplots displayed a strong linear relationship between the Forty_Yard and Weight variables, and between the Vert_Leap and Broad_Jump variables. The strong correlations were confirmed by the Pearson r's of 0.83 and 0.82, respectively.

Hypothesis testing confirmed that the observed results of players getting bigger, faster, and stronger were all statistically significant. Furthermore, the observed results that players got taller, jumped higher and farther, and increased lateral quickness are also statistically significant. To recap, none of the observed differences in players combine results from 1987-1999, and players combine results from 2000-2020 occurred by chance.

A linear regression analysis was conducted between the Forty_Yard variable vs. the Bench_Press + Shuttle variables. The R-squared = 0.629 indicated that the Bench_Press and Shuttle variables explain 62.9% of the variation in the Forty_Yard variable. The results were statistically significant, which means these variables are better predictors than using the mean of the Forty_Yard variable as a predictor.

What do you feel was missed during the analysis?

I wish I could have analyzed the differences between multiple time periods. Finding the largest year over year differences in the variables for each position would have been interesting as well. I think there are a lot of interesting insights that were unexplored in the data.

Were there any variables you felt could have helped in the analysis?

I think the variables used allowed for a thorough analysis of differences between the two player groups.

Were there any assumptions made you felt were incorrect?

The analysis included only players that participated in the NFL combine. While most of the players in the NFL participate in the combine, there are other players that do not. Sometimes players are injured and cannot participate, and some prefer to skip the combine and perform specific tests during a personal "Pro day." The analysis also does not include how player size, strength, and speed evolve in the years after they enter the NFL.

What challenges did you face, what did you not fully understand?

The main challenge was how to handle the position groupings and how observed changes could be quantified between the positions. After much struggle, a breakthrough came when I was able to visualize the changes for each position. After I was able to do that, the project progressed in an intuitive way.

References:

1. Downey, A. (2015). Think stats: Exploratory data analysis.

2. NFL Combine Data - dataset by sportsvizsunday. (2019, January 13). Retrieved May 30, 2020, from https://data.world/sportsvizsunday/nfl-combine-data

3. 33 Years of NFL Scouting Combine Data. (2020, March 03). Retrieved May 30, 2020, from https://nflcombineresults.com/

4. Pro Football Statistics and History. (n.d.). Retrieved May 30, 2020, from https://www.pro-football-reference.com/

5. Evolution of the NFL Player. (n.d.). Retrieved May 30, 2020, from https://operations.nfl.com/the-players/evolution-of-the-nfl-player/

6. NFL Scouting Combine. (2020, April 23). Retrieved May 30, 2020, from https://en.wikipedia.org/wiki/NFL_Scouting_Combine