Title Slide

Good evening. My name is Chris Izenour and I will present to you my analysis and findings related to analysis of team scoring trends and quarterback performance between 1970 and 2016.

Next slide

The data for my analysis originated from NFL.com. Kendall Gillies scraped the data and placed it on Kaggle.

My analysis focused on researching the following questions. Conventional wisdom holds that teams score more points today than at any time since the AFL-NFL merger in 1970. I wanted to see if scoring had in fact increased. Conventional wisdom also holds that it is easier to play quarterback now than at any time previously due to changes team offensive tactics. I wanted to see if quarterbacks have improved in their performance as measured by a handful of statistical ratios. I hypothesized that team scores have increased as a result of improved quarterback play.

Next slide

During the Exploratory Data Analysis phase, I found evidence of relationships and data that seems to confirm much conventional wisdom about quarterback play and team scoring. The top figure is a series of boxplots for team scoring each year from 1970 – 2016. As you can see, there is a general upward trend in average team score over the time period covered. Something I noticed with the boxplots is that since the late nineties {click the mouse}, the whiskers of the boxplots have shown decreasing variation than in the years prior. My interpretation of this is that on average, team scores are becoming more similar in recent years than in previous years {click the mouse}.

Coming back to the increase in points over time, {click the mouse} I’ve overlayed a line to better show the increase from the 1970 median value of approximately 18.5 points to the 2016 median value of approximately 22.5 points {click the mouse}. This four-point increase represents a 22% increase over the time period covered {click the mouse}.

Below the boxplots are a series of scatterplots and residuals depicting the relationship between pairs of variables. Again, {click the mouse} team score shows a linear increase over time when plotted against year. The residual plot below the scatterplot shows a lack of heteroskedasticity {click the mouse}.

{click the mouse} Completion percentage shows a linear relationship with year as well. As time has gone by, quarterbacks have become more proficient at completing pass attempts. Again, the residuals show a lack of heteroskedasticity {click the mouse}.

Plotting Team Score by Completion Percentage {click the mouse}, a linear trend between the variables is visible. As completion percentage increases, so does team score. The residual plot shows a more concentrated grouping of residual data points, but there is no clear trend or pattern discernable in the residual data points {click the mouse}.

Plotting {click the mouse} Attempts per TD by Year again shows that over time, quarterbacks take, on average, fewer pass attempts between touchdown passes {click the mouse}.

Next Slide

I utilized four models during the supervised learning phase to find the model that would best predict team score given the variables and underlying dataset. I trained and fitted a linear model, k-nearest neighbors, decision tree, and random forest.

The linear model’s RMSE came in the lowest at 2.063, but the other three were essentially the same, ranging from 2.11 to 2.235.

The R-Squared value of the linear model came back as 0.5, meaning that quarterback performance accounted for 50% of the increase in the previously discussed average team score since 1970. This is remarkable considering that a quarterback is 1 of 11 players on the field at a time for the offense. This finding lends credence to the notion that a quarterback is the most important player of a football team. Both the decision tree and random forest found that pass attempts per touchdown was the most important variable in determining team score.

Here, the plot on the left shows how the models differ in predicted average team score given changes in pass attempts per touchdown. On the right is a set of tables depicting an arbitrarily chosen section of observations, index 400-405. The top table depicts the mean statistics for the chosen range. The bottom table depicts the Actual Mean Team Score for that observation followed by the predicted team score for the different models utilized.

Next Slide

Pending any questions, this concludes my presentation.