Mark Nemat

Project B

STA 3064

1. **Motivation: Provide only one paragraph as to why your study is of interest and the potential benefits of fitting a statistical model.**

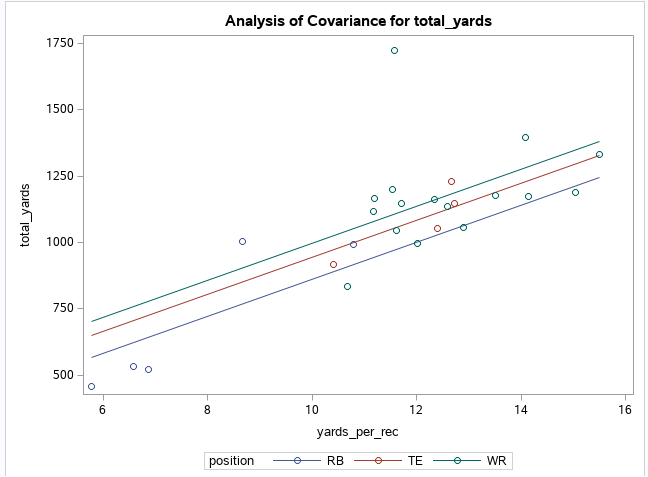
My objective for Project B is to determine if yards-per-catch influences their total catching yardage. I will also look to see which skill position WR, TE, or RB has the most impact when receiving the football. I’m interested in this study because I was the starting QB of my high school football team. This study can be beneficial to fit in a statistical model because we can learn if yards-per-catch affects to total yardage and we can learn to which skill position is the best to throw to.

1. **Data Description: In just one paragraph describe the nature of the data and the way it was produced or collected. Cite the source(s) of your data. Clearly provide explanations of all variables that will be used for the study and state the number of observations.**

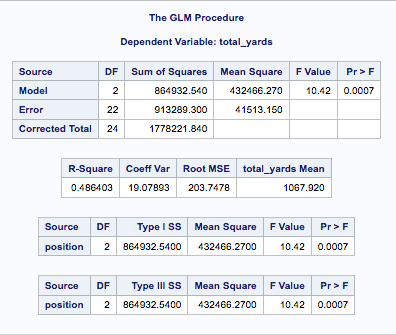
I got my data from NFL.com. This cite gave me access to every receiving statistic during the 2019 football season. The data I gather from this cite was the players name (first and last), players position, player total catches, player’s total yardage, and players yards-per-catch. I only chose the top 25 players with most reception, which means I have 25 observations in my study. My response variable is total receiving yards. I have one categorical predictor variable which is player position (WR, TE, or RB). I have a qualitative predictor variable which is total receiving yards and total catches.

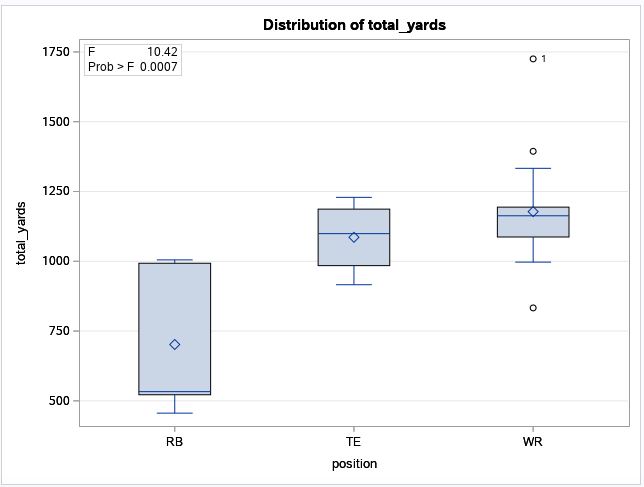
<https://www.nfl.com/stats/player-stats/category/receiving/2019/REG/all/receivingreceptions/desc>

1. **Data Exploration: Include in your report the data step or import code used to get your data into a SAS data set. Comment on any additional data manipulation that was necessary. Using your SAS data set, produce a graphical display involving all variables. For example, for each category of your categorical variable, you could construct a scatterplot relating your quantitative predictor to your response. Provide only key plots in the report and note any interesting characteristics in the relationships revealed by the above graphics.**

****

1. **Model Fitting and Analysis: Your analysis here should be performed in two phases: ANOVA modeling and ANCOVA modeling. Use the following items to guide this portion of your analysis:  ANOVA Modeling. Only include your categorical variable(s) as the predictor(s) (exclude all quantitative variables except for your response variable). In your analysis, complete the following items:**
   1. **Fit the appropriate one-way, two-way, etc. ANOVA model and clearly identify any significant factors that result from the analysis.**

****

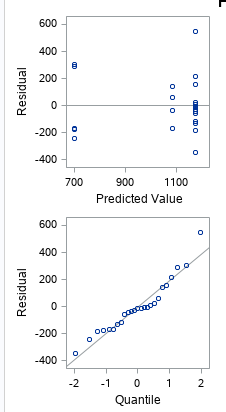
****

The P-value is less than .05, which indicate that the position is significant. When looking at the data, the RB position is the lowest however the total yardage is very spread out. The TE and WR position look similar. Those position tend to get an average of 1000-1250 yards per season. The position with typically the highest total yardage is the WR. One WR had almost of total of 1750, which is significantly higher than any TE and RB.

* 1. **When factor(s) are found to be significant, employ appropriate post hoc procedures (e.g., Tukey comparisons or contrasts) to identify where differences exist.**

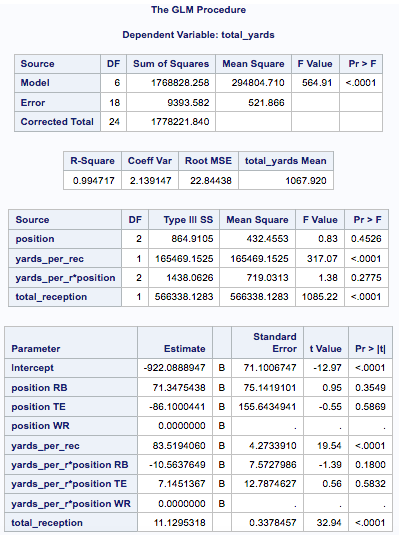
As stated above, the Position is significant when looking at total yardage. When looking at the mean of total yardage from the Tukey Method graph, it shows how the WR and TE are grouped together with estimate of 1000+ and the RB position estimate at around 700.

* 1. **Check the model assumptions through residual analysis. Explore transformations if necessary.**

****

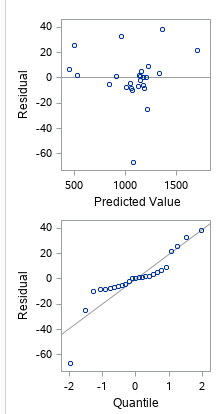
By looking at the Residual vs PV plot, It looks like constant variance in your residual vs. predicted plot as the spread seems trend vertically. I would guess the the left line in RB, middle right line represents TE, and far right line that has the most point represent WR. The Linearity is not met in the first plot because the data points are not scattered around. By looking at the Residual vs Quartile plot, I would say that the normality is met but there is a little more curvature than I want. No transformation needed. The graphs didn’t seem to improve my model.

1. **ANCOVA Modeling. Now all predictors (categorical and quantitative) are included in the model. Use the following items to guide your study:** 
   1. **Fit the appropriate ANCOVA model to your data and identify any significant predictors. You may either use a traditional regression model with indicator variables for the categorical predictors or a general linear model procedure {PROC GLM).**

****

**By looking at the GLM procedure, the model has a .0001 p-value which mean the model is significant. After looking my variables, the model shows that the yards\_per\_rec and total\_reception is significant with p-value below .05. However, yards\_per\_rec\*position, is not significant. Which means the yards per reception does not affect the three skilled position.**

* 1. **Check model assumptions for the ANVOCA fit via residual analysis.**

****

By looking at the Residual vs PV plot, It looks like constant variance in your residual vs. predicted plot as the spread is mostly in the middle and spread out around the top. Most of the points are clumped in the middle. The Linearity is met in the first plot because the data points are scattered around the plot but not by much. By looking at the Residual vs Quartile plot, I would say that the normality is not met. There is a little bit of a curve to the point that could interfere with the normality. I noticed that were are a few outliers.

**c. Compare your ANCOVA results to what your findings suggested in fitting the ANOVA model. Did using the quantitative predictor(s) clarify anything concerning your categorical variable? Did it seem to improve the fit in any way? Briefly explain.**

When comparing the models, I learn that yards per reception is significant, but also the total reception is very significant. My residual plots look a lot cleaner, with constant variance and linearity in the first plots. The normality seemed to get worse.

1. **Conclusions: Write just one paragraph about how your best statistical model from above can help enlighten your understanding of the original problem. Your discussion should be framed in non-statistical terms, that is, what action, if any, does your analysis suggest that you take in the context of this problem (e.g., what are the practical implications of using your model). In this section, your findings should be communicated in such a way that a person who has just a little statistical background can understand your results.**

In this study, we looked to determine if yards-per-catch influences their total catching yardage. As well as, looking to see which skill position WR, TE, or RB has the most impact when receiving the football. From the information that we gather, the total reception and yards per reception from a skilled player has a significant impact on the total receiving yardage. If a skilled position player (WR, TE, or RB) have a high total reception and high yards per reception, they will typically have a large total receiving yard during the season. What I learned, is how that affects each skilled position. Out of the three skilled position, the RB typically has the lowest total receiving yardage a season compared to the TE and WR position. When comparing the TE and WR position, the total reception and yards per reception are somewhat similar, however, the WR position have the advantage, because they tend to gain more reception, which accumulate more total receiving yards. From all the data I gathered, my conclusion is that yards per reception and total receptions have an impact on the total season receiving yardage. When interpreting the skilled positions, yards per reception do not have an impact on the three skilled position. I did conclude that, the WR position is the best position to thrown to, because they tend to get the most reception and higher yards per reception, which allow the WR position to gain more receiving yards in the NFL. From my study, I learned that the position that the most affective receiving impact is the WR position, with the TE position coming at a close second. If I presented my data to a NFL team, they would leave knowing that they should throw the ball more to the WR position.