STAT 145 – Mini Project 1

Predicting NFL Team Wins from QB Salary

Timothy Jimothy

I decided to analyze the relationship between NFL team wins in a season and the salary of the teams Quarterback. Thus for my project I have:

- Individual – NFL Teams
- Response Variable – Team wins (2015-16 Season)
- Explanatory Variable – Starting QB Salary (In millions of dollars)

I began by randomly selecting 12 NFL Teams, and finding the both variables (from the 15-16 season) for each team. A Scatterplot of the data is given below.
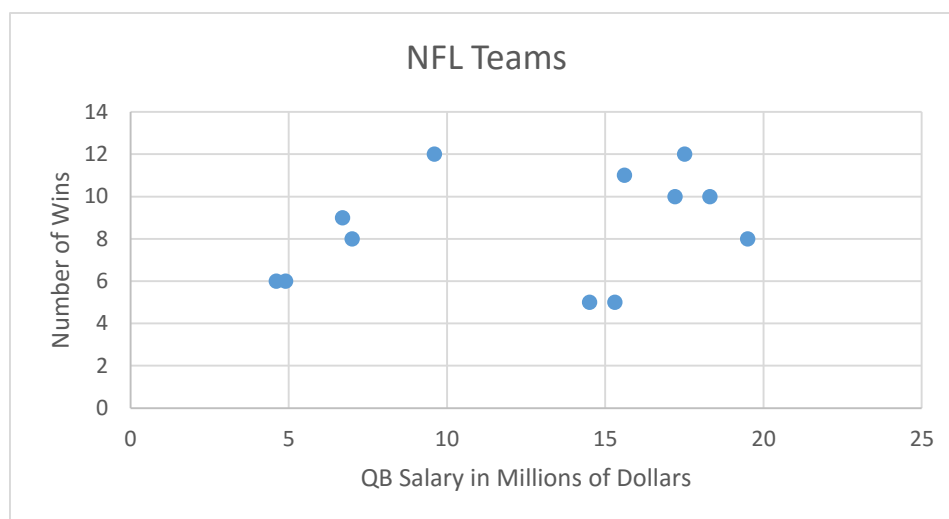
I was expecting to see a positive linear correlation. While there may be some positive correlation, it appears to be very weak. It also seems roughly linear.

Using Microsoft Excel, I calculated the following:

$$\bar{x} = 12.56 \; Million \; \$ \qquad\qquad \bar{y} = 8.5 \; Wins$$

$$S_x = 5.59 \; Million \; \$ \qquad\qquad S_y = 2.58 \; Wins$$

$$r = \; 0.31 \qquad\qquad\qquad R^2 = \; .10$$

As was stated above, there appears to be a positive correlation, but it is fairly weak. The R^2 value of .10, tells me that 10% of the variance in number of wins can be explained by change in the Quarterback Salary.

To find the equation of the regression line, I remember the formulas for a and b:
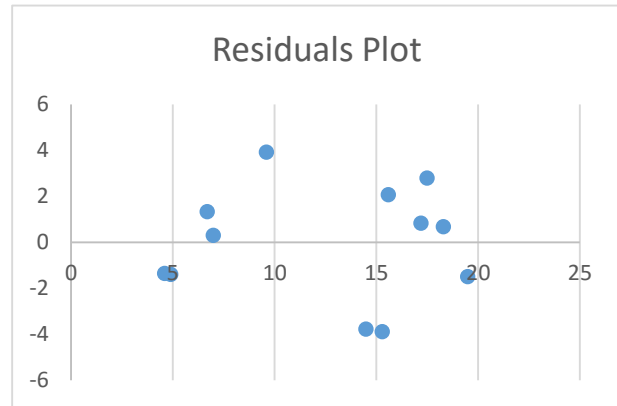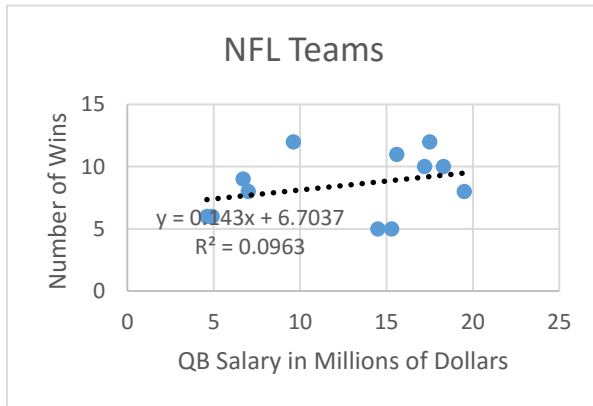
$$b = r\frac{S_y}{S_x} = 0.143$$

$$a = \bar{y} - b\bar{x} = 6.704$$

Therefore the Regression Equation is

$$\hat{y} = 6.704 + 0.143x$$

Which looks like the following (Left). A plot of the residuals is also given (right).

The residual plots looks a little bit like a frowny face with two dots in the bottom middle. It doesn't look like a random scattering of points, which is what I hoped it would look like for a good Regression model.

I can use my model for prediction, but I don't think that it is particularly reliable. For example, the Arizona Cardinals Quarterback Carson Palmer made 7.375 Million dollars during the 2015-16 season. According to the regression line, we predict that the Cardinals should have won 7.75 games. But the Cardinals actually won 13 games, way above their predicted value.

Sometimes it works a little better. For example, the Colts Quarterback Andrew Luck made 7 million dollars last season, so our Regression Model predicts that the Colts should have won 7.7 games. Since they actually won 8 games, the Colts had a residual of only .3. In this case, the regression line gave a pretty nice prediction.