Name: Omika DHARAMDASANI

Student Number: 1000984483

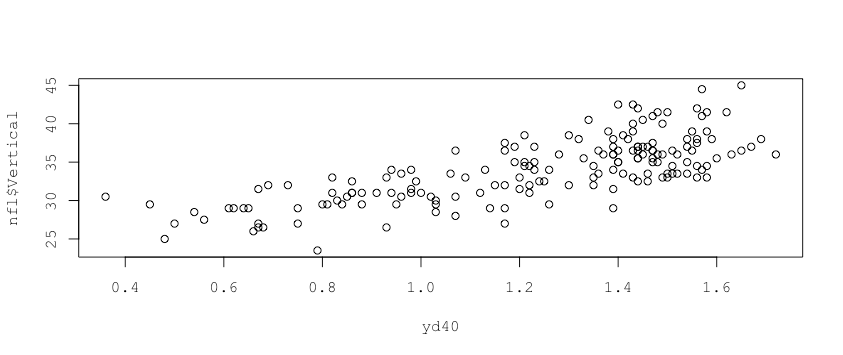
Lecture: 5101

Date: June 8, 2016

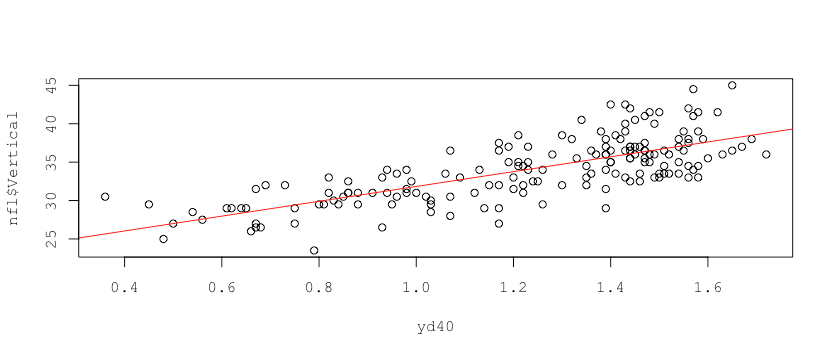
STA302 ASSIGNMENT 1

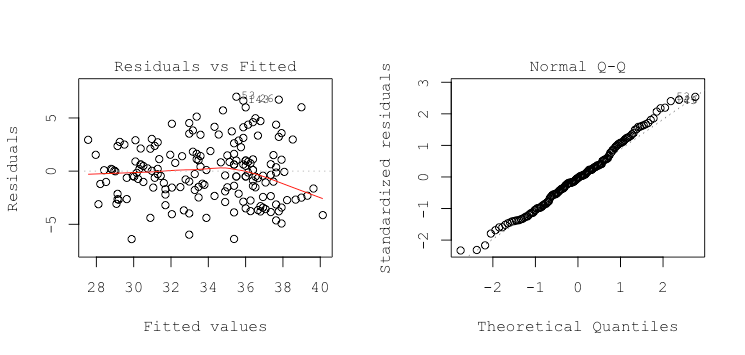
PART A:

1.

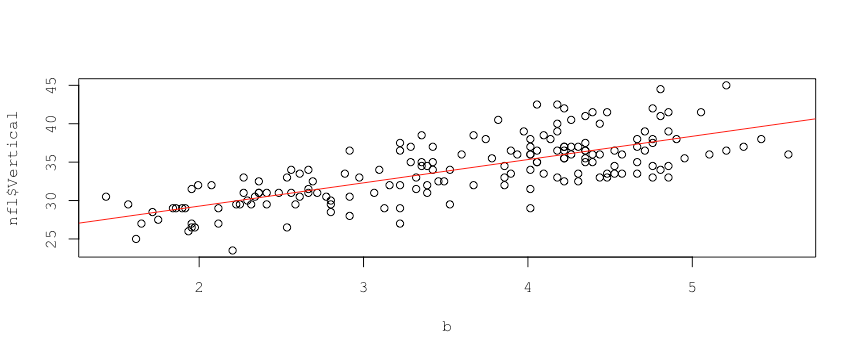


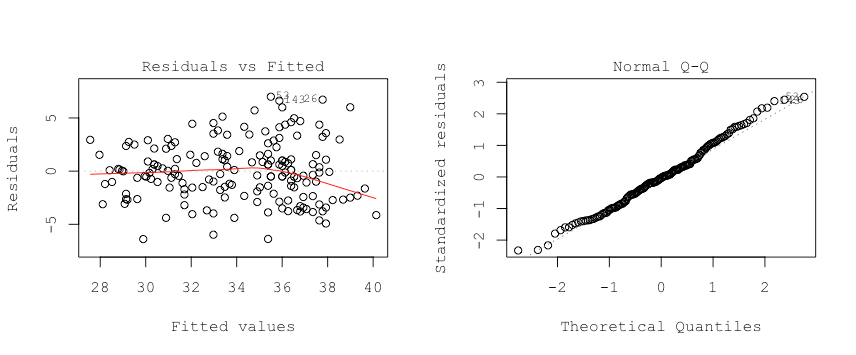
3.

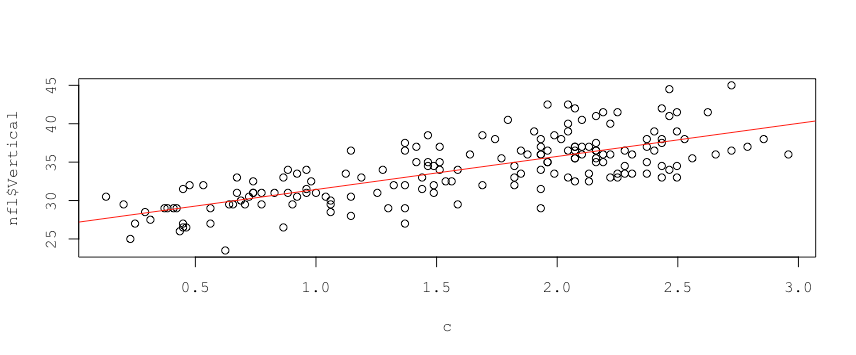


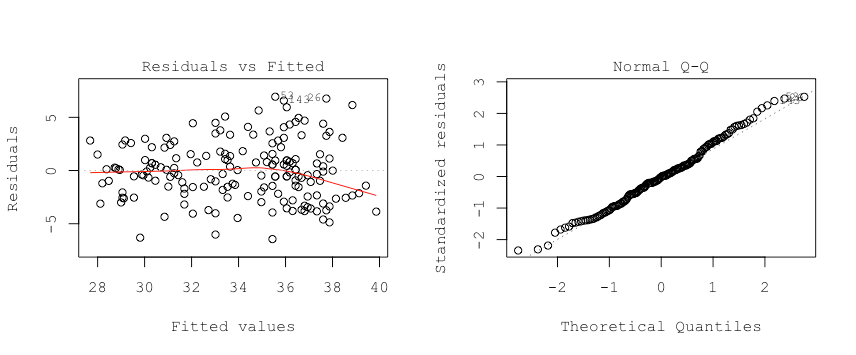
4

5. Transforming X:





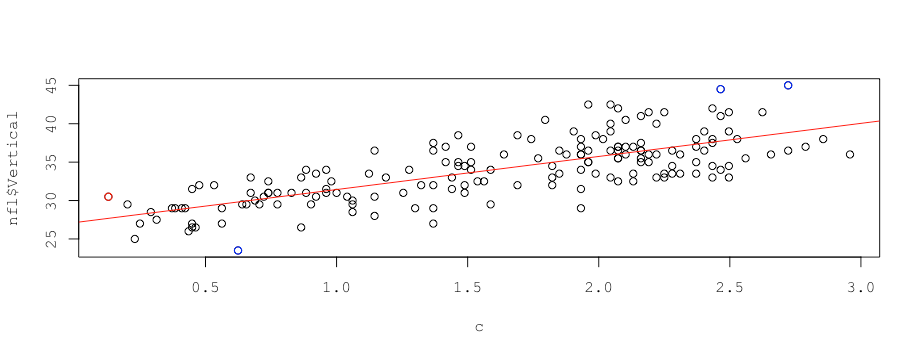
6.



7.

|  | R-squared | SSE | t-statistic | Slope estimate (b1) |
| --- | --- | --- | --- | --- |
| Linear Model | 0.5501 | 1298.1 | 14.42 | 9.6632 |
| Exponential | 0.5489 | 1301.6 | 14.38 | 3.0302 |
| Squared | 0.552 | 1292.8 | 14.47 | 4.3054 |
|  |  |  |  |  |

8. According to the summary statistics shown above, the squared model seems to be the best one because it has the highest R-squared and the lowest SSE. Hence, we will work on the squared model for this question. The player with leverage greater than (average leverage)\*2.5 is Danny Shelton and the players with the absolute DFFITs greater than 0.3 are Byron Jones, Chris Conley and Trenton Brown.



The player with the high leverage seems to have a shorter time in 40yd dash and also a lower vertical jump than the other players. The high DFFITs point have a higher 40-yard dash time than most others and also a higher vertical time. They also deviate farthest away from the trendline with regards to Y. The blue point below the trendline is also farthest away from the trendline, but in the opposite direction.

PART B:

4. SUMMARY TABLE FOR RECEIVING YARDS AGAINST 6 MINUS DASH TIME

Call:

lm(formula = new2$rec\_YDS ~ yardDash)

Residuals:

Min 1Q Median 3Q Max

-254.46 -145.22 -55.60 95.97 810.54

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -214.1 220.5 -0.971 0.3359

yardDash 299.7 153.8 1.949 0.0566 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 219 on 52 degrees of freedom

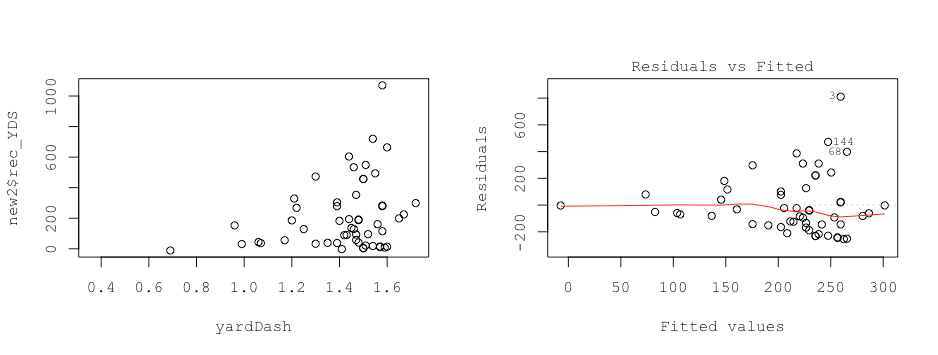
(161 observations deleted due to missingness)

Multiple R-squared: 0.0681, Adjusted R-squared: 0.05018

F-statistic: 3.8 on 1 and 52 DF, p-value: 0.05665

5. The slope coefficient is 299.7. This means that the receiving yards of a player will change by 299.7 units on average if the 6 - 40 yard dash times change by 1 second. No, the relationship is not statistically significant because the p-value is greater than 0.05.

6.



The scatterplot seems to predict a non-linear relationship, as the points seem to be following an upward curve. Outliers are definitely visible in the scatterplot and the observation on the top right particularly stands out. The residual vs fitted line has observations close together for small fitted values and spread farther apart as fitted values increase, which seems to indicate increasing variance. The line on the residual vs. fitted graph is somewhat flat, but not towards the end. The residuals do not seem to add up to 0 because there seem to be more points below the 0 than above. Therefore, the assumptions for linear regression are not met.

7.

Call:

lm(formula = new2$rush\_ATTG ~ new2$Overall)

Residuals:

Min 1Q Median 3Q Max

-6.8501 -4.2898 0.8997 3.0112 10.6691

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.12955 1.55571 4.583 7.07e-05 \*\*\*

new2$Overall -0.01986 0.01248 -1.592 0.122

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.792 on 31 degrees of freedom

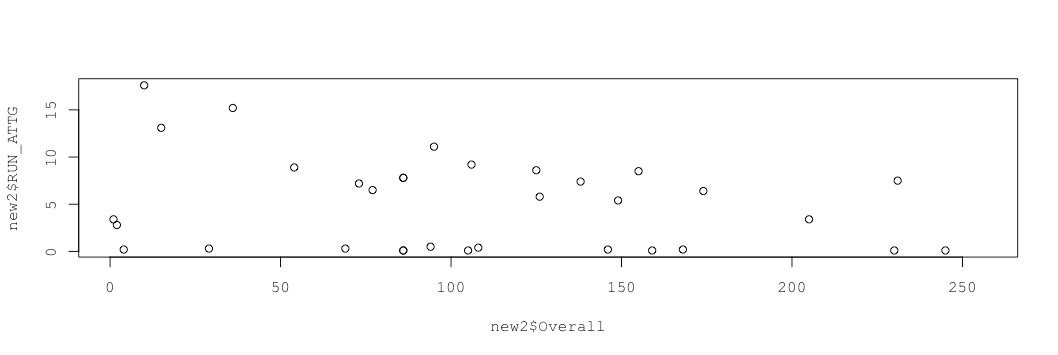
(182 observations deleted due to missingness)

Multiple R-squared: 0.07555, Adjusted R-squared: 0.04573

F-statistic: 2.533 on 1 and 31 DF, p-value: 0.1216

The slope coefficient is -0.01986, which implies an inverse relationship between overall draft rank and attempts per game.

The attempts per game of a player will decrease by 0.01986 units on average if the overall draft rank increases by 1. No, the relationship is not statistically significant because the p-value is greater than 0.05.

8.

We see a few outliers present, particularly in the top left corner. The relationship looks non-linear because there seems to be a downward trend. The observations are spread out for lower values of X and closer together as X increases, indicating non-constant variance of the residuals. All these factors challenge the linearity assumption.