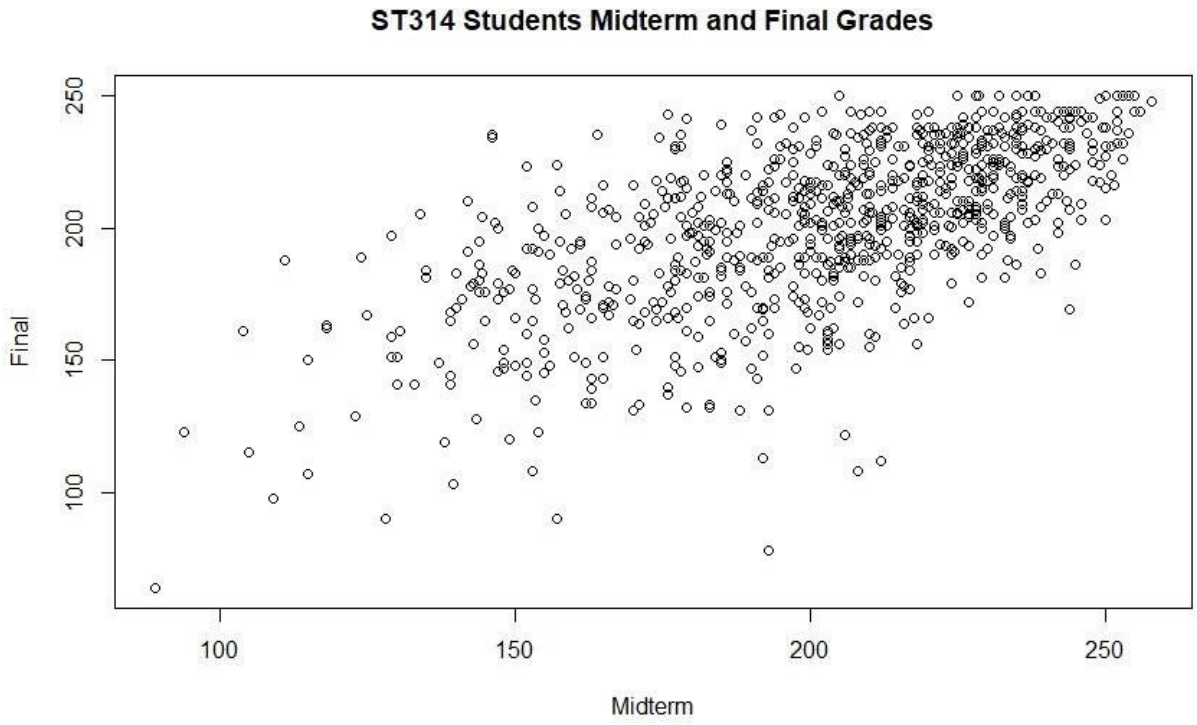
ST 314 - Data Analysis 8

Ivan Timothy Halim

March 6, 2019

Part 1

a.



There is a moderately strong, positive, linear association between the midterm grades and the final grades. There are plenty of outliers such as the student with a midterm grade of 190 and a final grade of less than 100.

b. $r = 0.636$
The correlation between midterm grades and final grades is moderately strong.

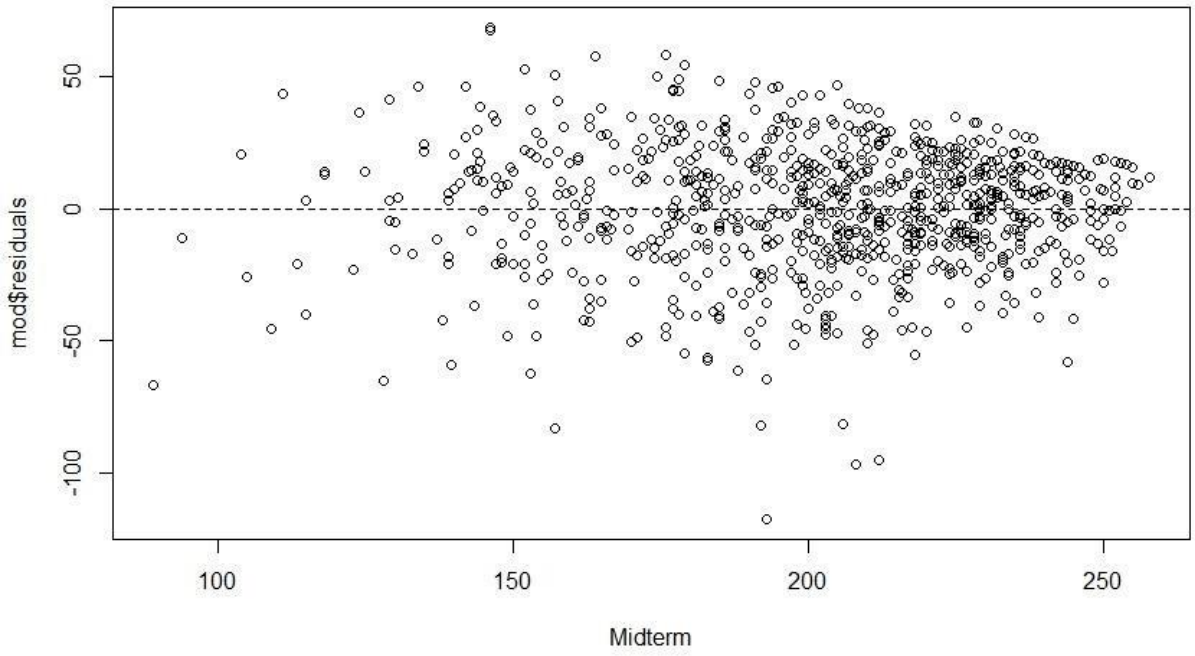
Part 2

a.

```
Call:
lm(formula = Final ~ Midterm)

Residuals:
     Min       1Q   Median       3Q      Max
-117.571  -13.617    1.962   16.059   68.671

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 75.49269    5.28281   14.29   <2e-16 ***
Midterm      0.62217    0.02587   24.05   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.69 on 850 degrees of freedom
Multiple R-squared:  0.4049,    Adjusted R-squared:  0.4042
F-statistic: 578.4 on 1 and 850 DF,  p-value: < 2.2e-16
```

Model: $\hat{y} = 0.622x + 75.49$

b.



Linearity condition: Linear relationship must exist between x and y.
True, there doesn't seem to be any "U" shaped or curved patterns in the residuals. The residual plot shows a fairly random pattern.

Constant Variation Condition: For each value of x, the distribution of the response variable has the same spread.
False, the residual plot shows a heteroscedastic (cone-shaped) pattern.

Normality Condition: For each value of x, the distribution of the response variable is normally distributed.
True, there seems to be an equal of residuals above and below the reference line.

Part 3

a. $H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

b. $t\text{-}stat = 24.05$
$df = 850$
$p\text{-}value < 2 \times 10^{-16}$

c. There is a strong evidence to conclude that the midterm grades and the final grades are statistically significant. Reject the null hypothesis based on a significance level of 0.05. ($t\text{-}stat = 24.05$, $df = 850$, $p\text{-}value < 2 \times 10^{-16}$)

d. $95\%\ CI = b_1 \pm (t_{n-(k+1),\frac{\alpha}{2}})\,(SE_{b1})$
$= 0.62217 \pm 1.984(0.02587)$
$= (0.571, 0.673)$

The LSRL estimates that for every one point increase in midterm grade, the final grade will increase by 0.622, with a 95% confidence interval between 0.571 and 0.673

Part 4

a. $\hat{y} = 0.622(200) + 75.49 = 199.89$

b. (198.3, 201.5) is the confidence interval.
(153.4, 246.5) is the prediction interval.
The confidence interval is narrower than the prediction interval.

c. The 95% Confidence Interval estimates the average final grade of all students with a midterm grade of 200 to be between 198.3 and 201.5, with a point estimate of 200.

The 95% Prediction Interval predicts the average final grade of a students with a midterm grade of 200 to be between 153.4 and 246.5, with a point estimate of 200.

d. Midterm score: 218
Final score: 0.622(218) + 75.49 = 211

e. The predicted value is not a reasonable score to assume for my final grade. Although, the association between midterm grades and final grades is moderately strong, the prediction interval is too wide for me to make any reasonable judgement about my final grade.