STAT 511: Assignment #1

Rumil Legaspi

25 January 2021

Assignment Questions

---

1. KNN 4th Edition End of Chapter 1 Questions

In a regression model, β_0 = 100, β_1 = 20, and σ^2 = 25. An observation on variable will be made for = 5.

$Y_i = \beta_0$ *(intercept)* + $\beta_1 X_i$ *(slope)(independent variable)* + ϵ_i *(error)*

y = 100 + $20X_i$ + ϵ_i

(a). Can we compute the exact probability that will fall between 195 and 205? Explain The probability cannot be calculated because the for a simple linear regression model the mean of ϵ_i should equal 0. Because ϵ_i is unspecified we are missing information and cannot compute the exact probability.

(b). If the normal error regression model is applicable, can we now compute the exact probability that will fall between 195 and 205? If so, compute it. *note: ϵ_i (error term) = 0 and follows a normal distribution*

For this problem we recall:

- 1. The Z score formula since we are dealing with a normal distribution. $\frac{X-\mu}{\sigma}$
- 2. How to find the probability between 2 points given a normal distribution.

(aka find the z score which finds everything from the left and subtract it by the larger number to get the probability between a and b)

- 3. And that we are also given $\sigma^2 = 25$ *(variance)* and $\sigma = 5$ *(Standard deviation)*

SO: $P(195 \leq Y \leq 205) = P(\frac{195-200}{5} \leq \frac{X-\mu}{\sigma} \leq \frac{205-200}{5})$

$= P(-1 \leq z \leq 1)$

$= P(z < 1) - P(z < -1)$ *bigger number or b is P(z < 1)*

$= 0.841 - 0.158$ *converting numbers using pos/neg z table*

$= 0.683$

The probability that Y will fall between the 195 and 205 is roughly 0.683.

---

2. Grade Point Average Problem (Use R)

The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year ( ) can be predicted from the ACT test score ( ). See the dataset "GPA.txt". The first column is GPA. The second column is ACT.

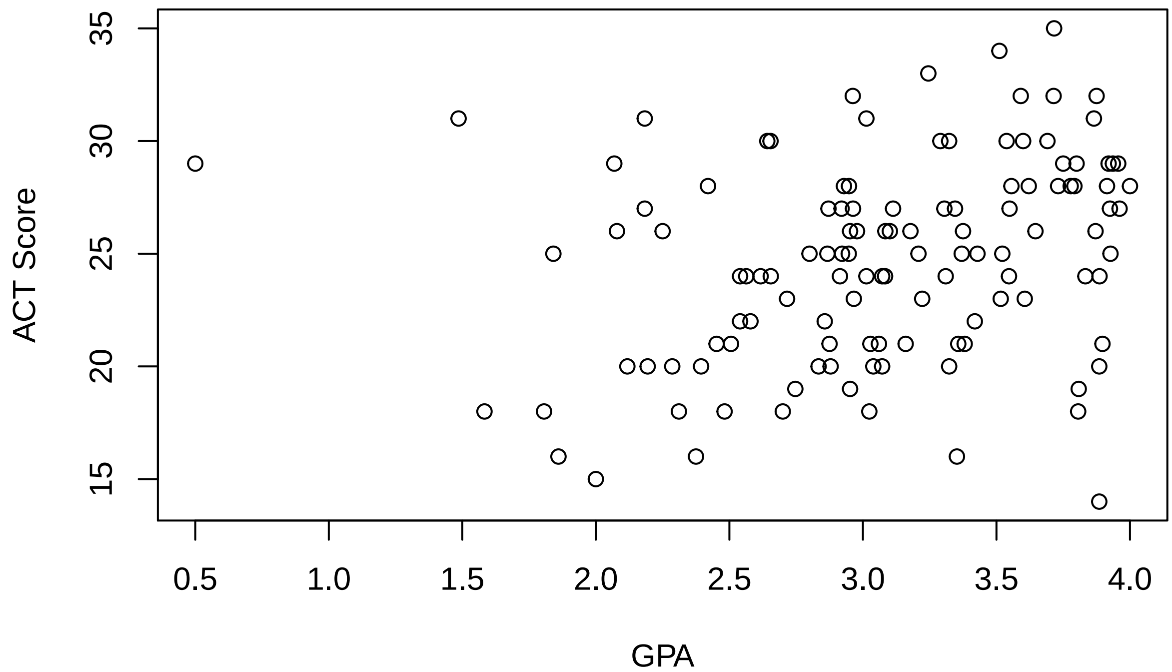
```r
setwd("C:/Users/RUMIL/Desktop/APU/STAT 511 - Millie Mao (Applied Regression Analysis)/R Files/STAT 511")


gpa_data = read.table(file = "GPA.txt", header = FALSE, sep = "")
head(gpa_data)
```

```
##      V1 V2
## 1 3.897 21
## 2 3.885 14
## 3 3.778 28
## 4 2.540 22
## 5 3.028 21
## 6 3.865 31
```

```r
#No headers, so we add
names(gpa_data) <- c("GPA", "ACT Score")
head(gpa_data)
```

```
##     GPA ACT Score
## 1 3.897        21
## 2 3.885        14
## 3 3.778        28
## 4 2.540        22
## 5 3.028        21
## 6 3.865        31
```

```r
#scatterplot
plot(gpa_data)
```




```r
lm(`ACT Score`~ GPA, data = gpa_data)
```

(a). Obtain the least squares estimates of β_0 and β_1. Write down the estimated regression equation.

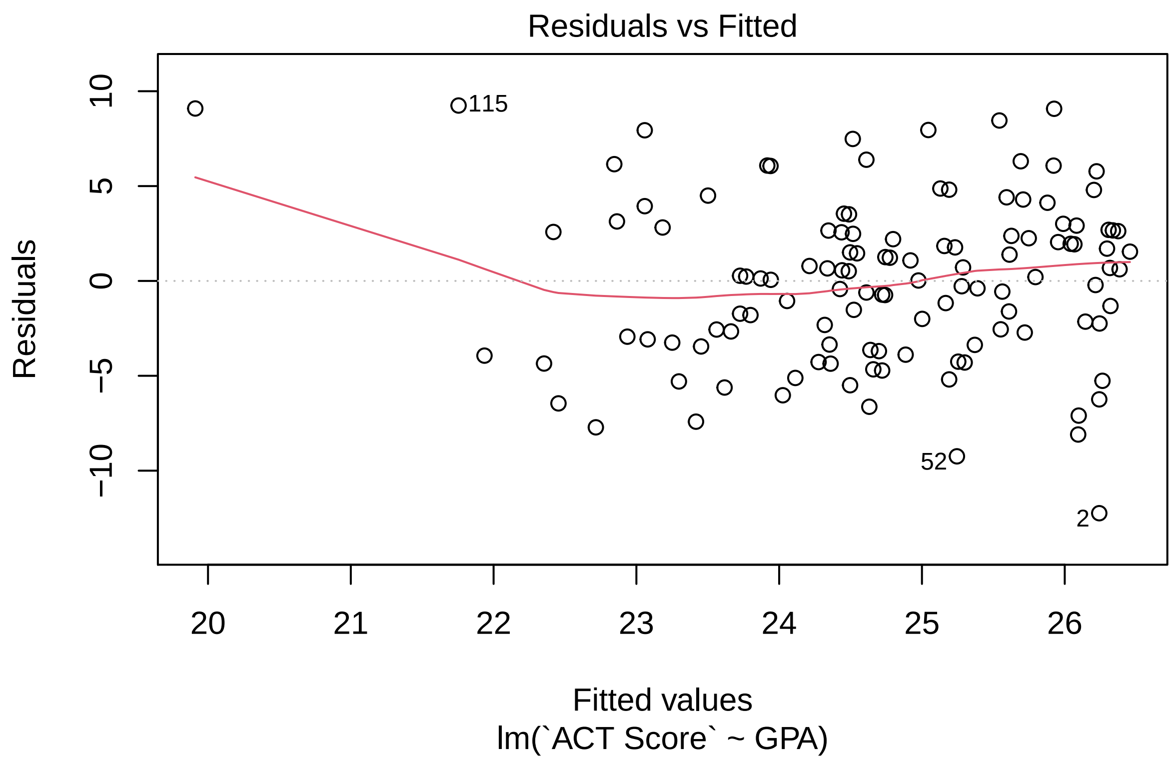
```
##
## Call:
## lm(formula = `ACT Score` ~ GPA, data = gpa_data)
##
## Coefficients:
## (Intercept)          GPA
##       18.98         1.87
```

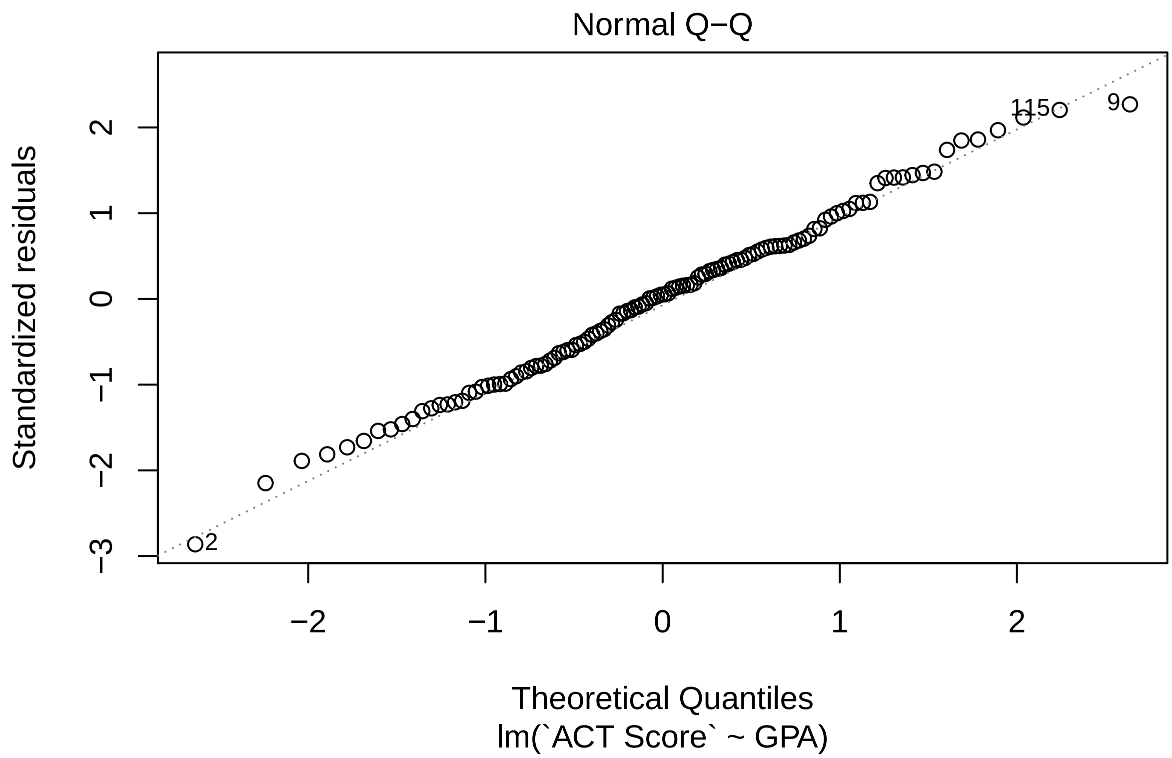
```r
gpa_lm = lm(`ACT Score`~ GPA, data = gpa_data)
summary(gpa_lm)
```

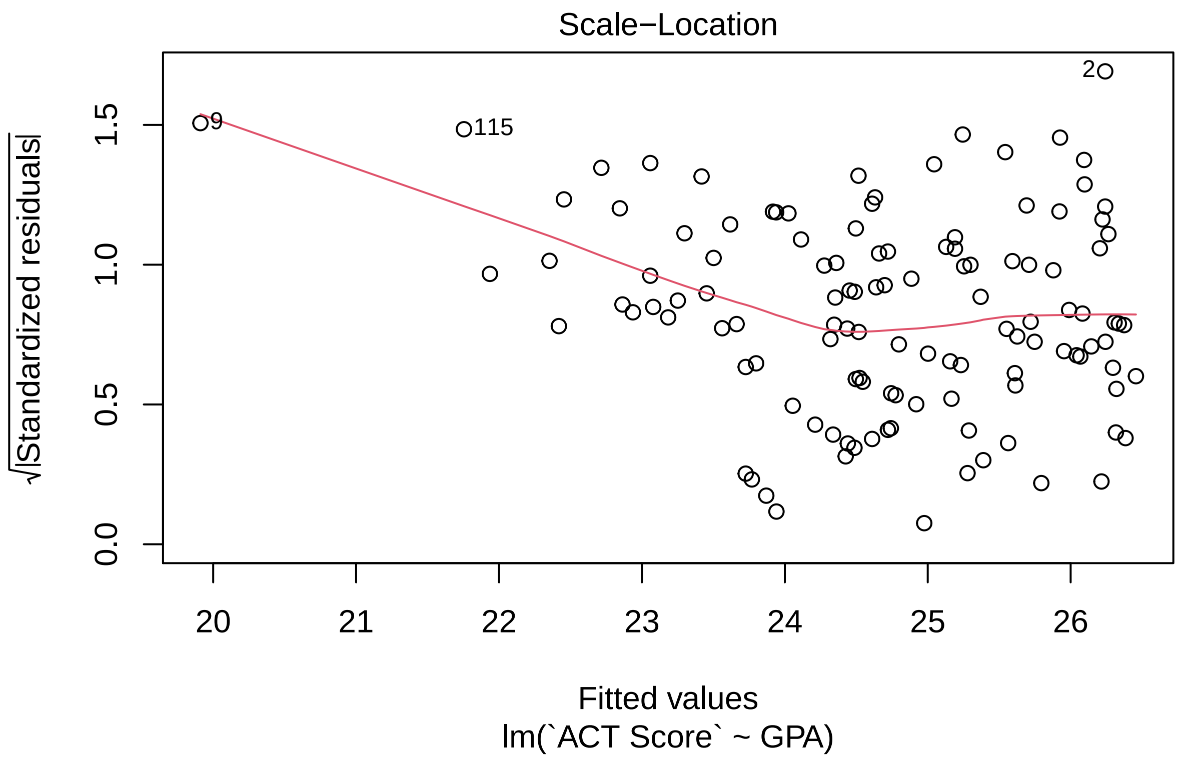
```
##
## Call:
## lm(formula = `ACT Score` ~ GPA, data = gpa_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

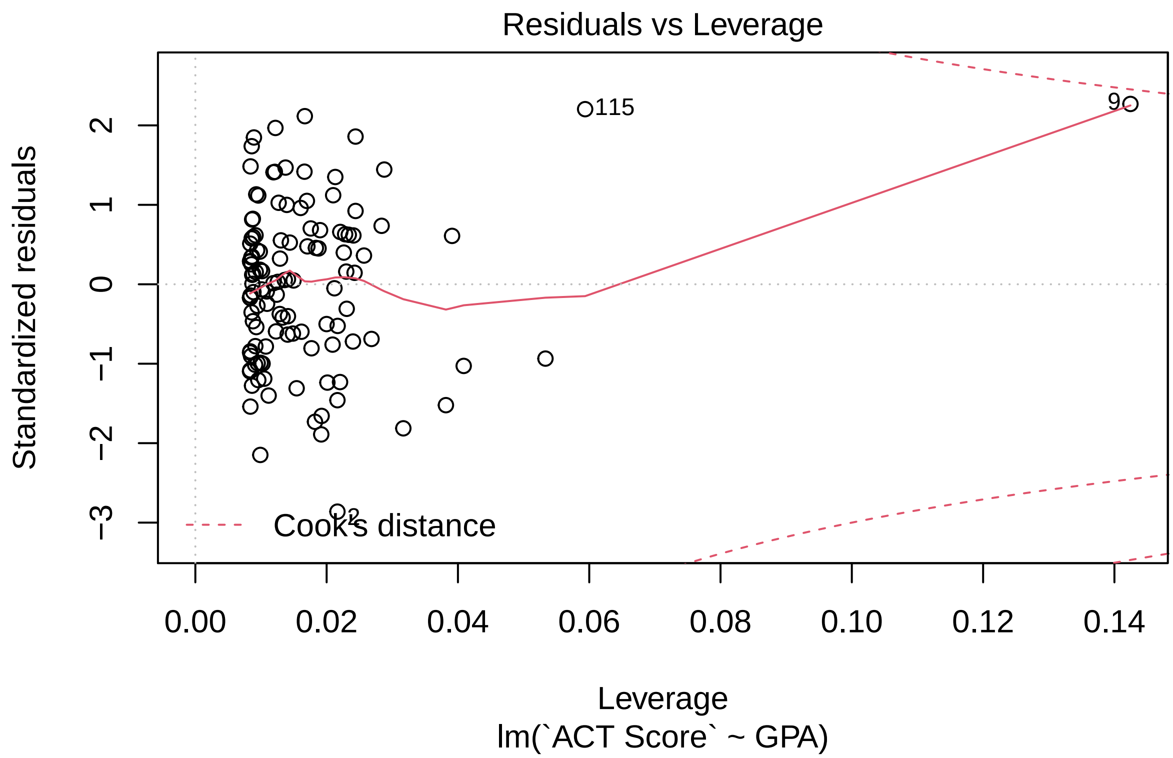
```
## -12.242  -3.276   0.218   2.657   9.245
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.9754     1.9322   9.821  < 2e-16 ***
## GPA           1.8704     0.6153   3.040  0.00292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.325 on 118 degrees of freedom
## Multiple R-squared:  0.07262,    Adjusted R-squared:  0.06476
## F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917
```

```r
plot(gpa_lm)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(`ACT Score` ~ GPA)
115
9
2
−3
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
20
21
22
23
24
25
26
Fitted values
lm(`ACT Score` ~ GPA)
9
115
2



(b). Plot the estimated regression line and the data points. Does the estimated regression function appear to fit the data well?

(c). Obtain a point estimate of the mean freshman GPA for students with ACT test score $=$ 30.

(d). What is the estimated change in the mean response when the ACT score increases by one point?

---

3. Refer to the GPA problem in Question 2. (Use R)

(a). Obtain the residuals $\hat{\epsilon_1}$. Do they sum to zero?

(b). Estimate the error variance σ^2 and standard deviation σ. In what units is σ expressed?

---

4. Refer to the GPA problem in Question 2.

(a). Interpret $\hat{\epsilon_0}$ in your estimated regression function. Does $\hat{\epsilon_0}$ provide any relevant information here? Explain.

(b). Interpret $\hat{\epsilon_1}$ in your estimated regression function.

(c). Verify that your fitted regression line goes through the point ($\bar{X}$, $\bar{Y}$). (Use R)

---

5. Muscle Mass Problem (Use R)

A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each l0-year age group, beginning with age 40 and ending with age 79. is age, and is a measure of muscle mass. See the dataset "Muscle.txt". The first column is muscle mass. The second column is women's age.

(a). Obtain the estimated regression equation.

(b). Interpret β_0 in your estimated regression function. Does β_0 provide any relevant information here? Explain.

(c). Interpret $\hat{\beta_1}$ in your estimated regression function.

(d). Plot the estimated regression function and the data points. Does a linear regression function appear to give a good fit here? Does your plot support that muscle mass decreases with age?

(e). Obtain a point estimate of the difference in the mean muscle mass for women differing in age by one year.

(f). Obtain a point estimate of the mean muscle mass for women aged = 60 years.

(g). Find the estimate of error variance σ^2

---

6. Special regression models

(a). What is the implication for the regression model $Y_i = \beta_0 + \epsilon_i$? How does it plot on a graph?

(b). What is the implication for the regression model $Y_i = \beta_1 X_i + \epsilon_i$? How does it plot on a graph?

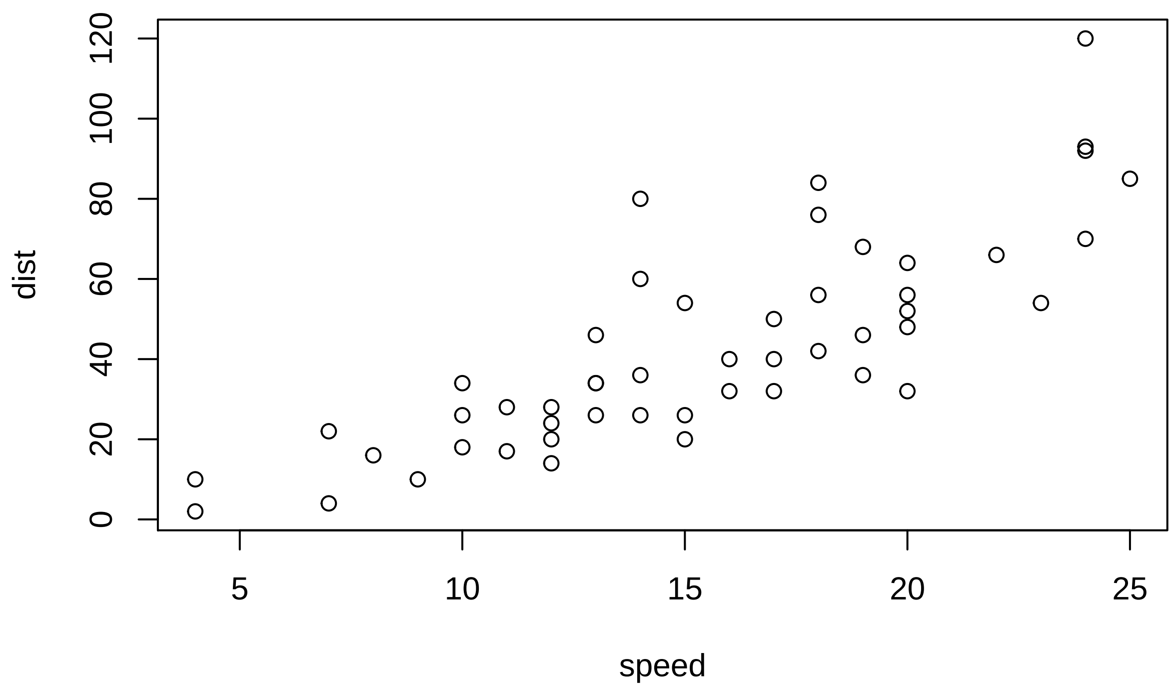
Latex Notation
σ
σ^2
β_1
β_0

$\hat{\beta} = 1.02$
$\hat{\beta}_1$
$\hat{\beta}_0$
$\hat{\epsilon}_1$
$(\bar{X}\ , \bar{Y})$
$Y_i\ \beta_0\ \epsilon_i\ \beta_i\ X_i$
$\frac{X-\mu}{\sigma}$ Z score
$P(195 \leq Y \leq 205) = P(\frac{195-200}{5} \leq \frac{X-\mu}{\sigma} \leq \frac{205-200}{5})$ more z score

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
plot(cars)
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.