

MATH 423 - Assignment 1, Dan Yunheum Seol - 260677676

This document was written using R Markdown.

Part (a). The following part will present summary of 3 SLR models with respective data set for each.

Simple Linear Regression for Data set 1

```
#Reading in the data

file1<-"http://www.math.mcgill.ca/yyang/regression/data/a1-1.txt"
data1<-read.table(file1,header=TRUE)

#Plotting the data with its center indicated

plot(data1$x,data1$y,pch=18, xlab = "X1", ylab = "Y1")
x1<-data1$x
y1<-data1$y
abline(v=mean(x1),h=mean(y1),lty=2)

#Fitting our model for the line of best fit

model1 <- lm(y1 ~ x1)
summary(model1)

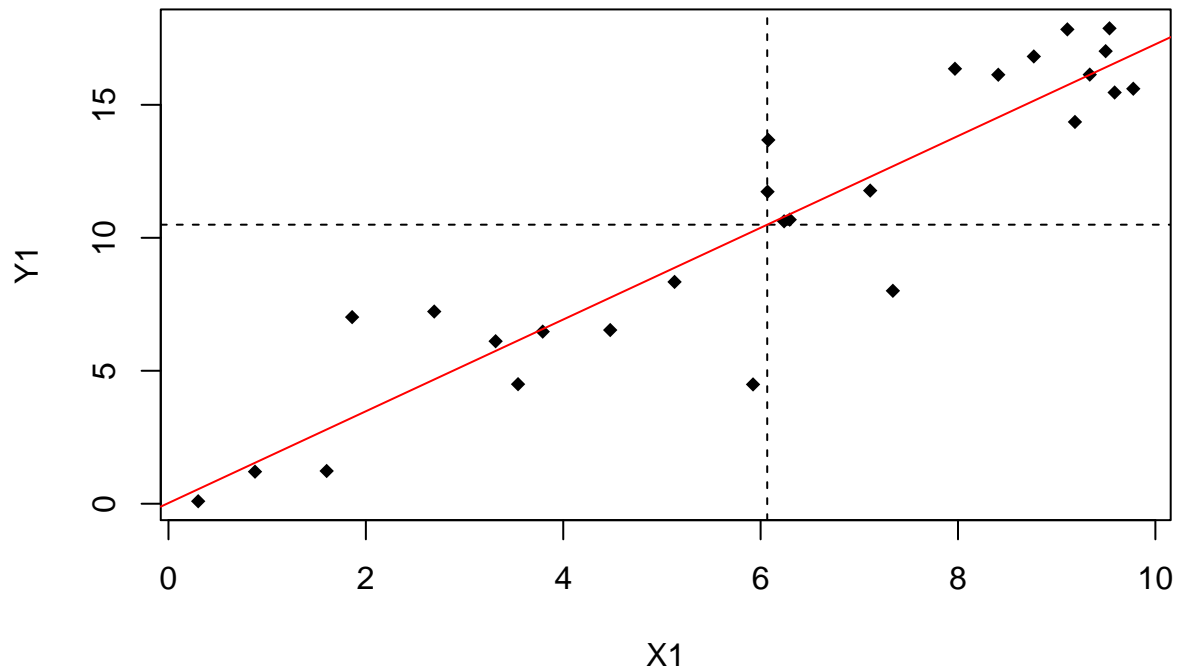
##
## Call:
## lm(formula = y1 ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.7577 -1.1595 -0.1691  1.5003  3.7808
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02677    0.96783   0.028   0.978
## x1          1.72512    0.14372  12.003 7.14e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.182 on 25 degrees of freedom
## Multiple R-squared:  0.8521, Adjusted R-squared:  0.8462
## F-statistic: 144.1 on 1 and 25 DF,  p-value: 7.145e-12

coef(model1)

## (Intercept)          x1
##  0.02676552  1.72512091

abline(coef(model1),col='red')
title('Line of best fit for Data Set 1')
```

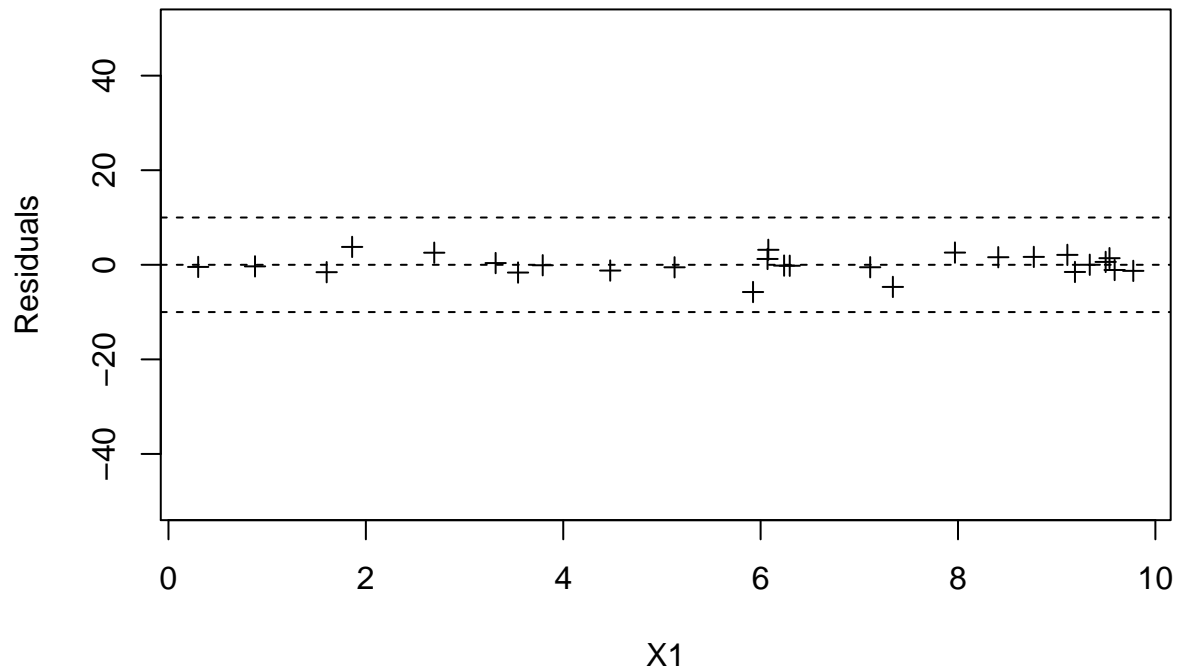
Line of best fit for Data Set 1



```
#Residuals
m1.resids<-residuals(model1)
m1.fitted<-fitted(model1)

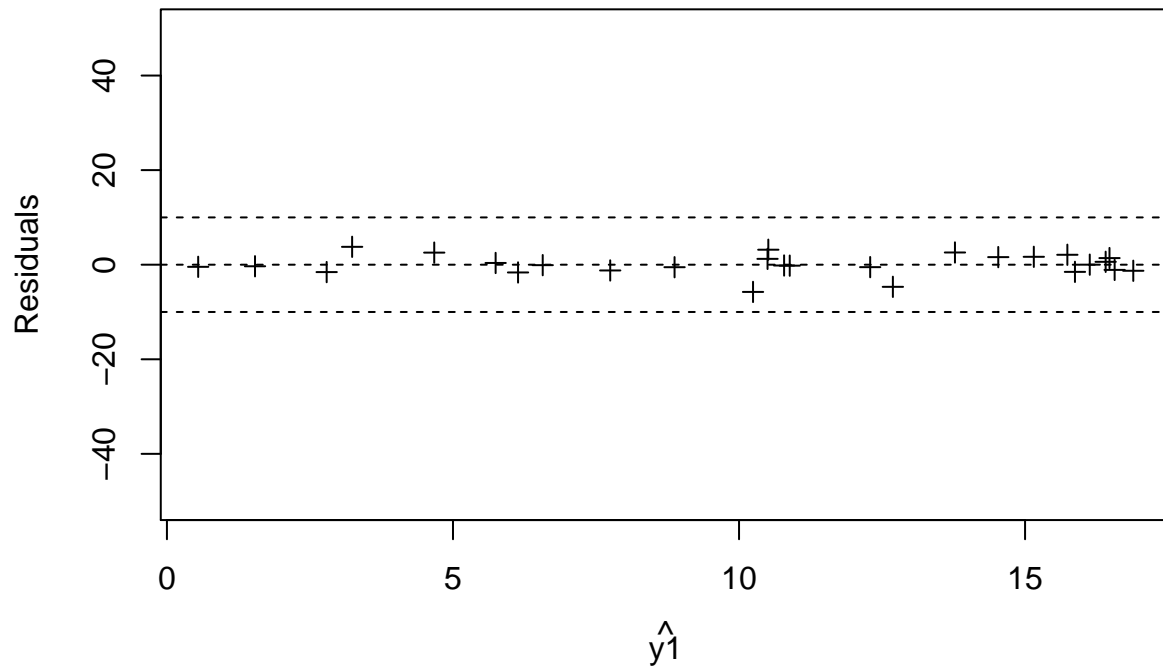
#Plots for Residual v. X1, Residual v. Fitted values to test whether our model holds
plot(x1, m1.resids,xlab='X1',ylab='Residuals',pch=3,ylim=range(-50,50));
abline(h=0,lty=2); abline(h=10, lty=2); abline(h=-10, lty=2)
title('Residuals vs X1')
```

Residuals vs X1



```
plot(m1.fitted, m1.resids,xlab=expression(hat(y1)),ylab='Residuals',
pch=3,ylim=range(-50,50));abline(h=0,lty=2); abline(h=10, lty=2); abline(h=-10, lty=2)
title('Residuals vs Fitted values')
```

Residuals vs Fitted values



mark: We can assume there is a positive linear relationship between the covariate and the outcome.

Our estimate for the intercept and the slope would be

$$\hat{\beta}_0 = 0.02676552$$

and

$$\hat{\beta}_1 = 1.72512091$$

, respectively.

Therefore, our line of best fit is given by:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \Rightarrow \hat{y} = 0.02676552 + 1.72512091x$$

R also states the result for the tests of the intercept and the slope estimates using t distribution with null and alternative hypotheses.

$$H_0 : \beta_0 = 0$$

and

$$H_a : \beta_0 \neq 0$$

$$H_0 : \beta_1 = 0$$

and

$$H_a : \beta_1 \neq 0$$

We can see that we have failed to reject the null

$$H_0$$

for the intercept estimate, but we have sufficient evidence to reject the null

$$H_0$$

for the slope estimate.

From the residual plots, we can notice from both residual plots that they are centered around zero and have constant variance. Our assumptions for the expected value of the error and homoscedasticity of variance have been kept. We could say that our residual,

$$e_i$$

, represents the behaviour our true random error,

$$\epsilon_i$$

.

Data set 2

```
file2<- "http://www.math.mcgill.ca/yyang/regression/data/a1-2.txt"
data2<-read.table(file2,header=TRUE)
plot(data2$x,data2$y,pch=18, xlab = "X2", ylab = "Y2" )
x2<-data2$x
y2<-data2$y
abline(v=mean(x2),h=mean(y2),lty=2)
model2 <- lm(y2 ~ x2)
summary(model2)
```

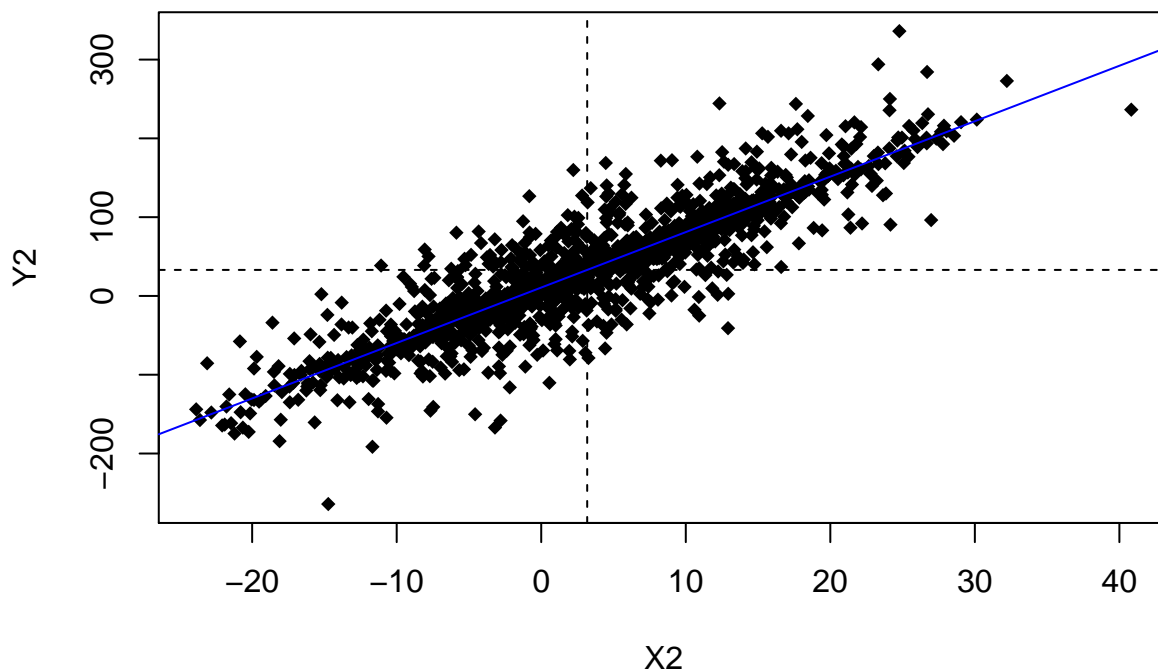
```
##
## Call:
## lm(formula = y2 ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -171.108  -13.288   -0.479   12.210  151.068
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.66085    0.98157   10.86  <2e-16 ***
## x2          7.03795    0.09253   76.06  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.6 on 1443 degrees of freedom
## Multiple R-squared:  0.8004, Adjusted R-squared:  0.8002
## F-statistic: 5785 on 1 and 1443 DF,  p-value: < 2.2e-16
```

```
coef(model2)
```

```
## (Intercept)          x2
## 10.660853      7.037952
```

```
abline(coef(model2),col='blue')
title('Line of best fit for Data Set 2')
```

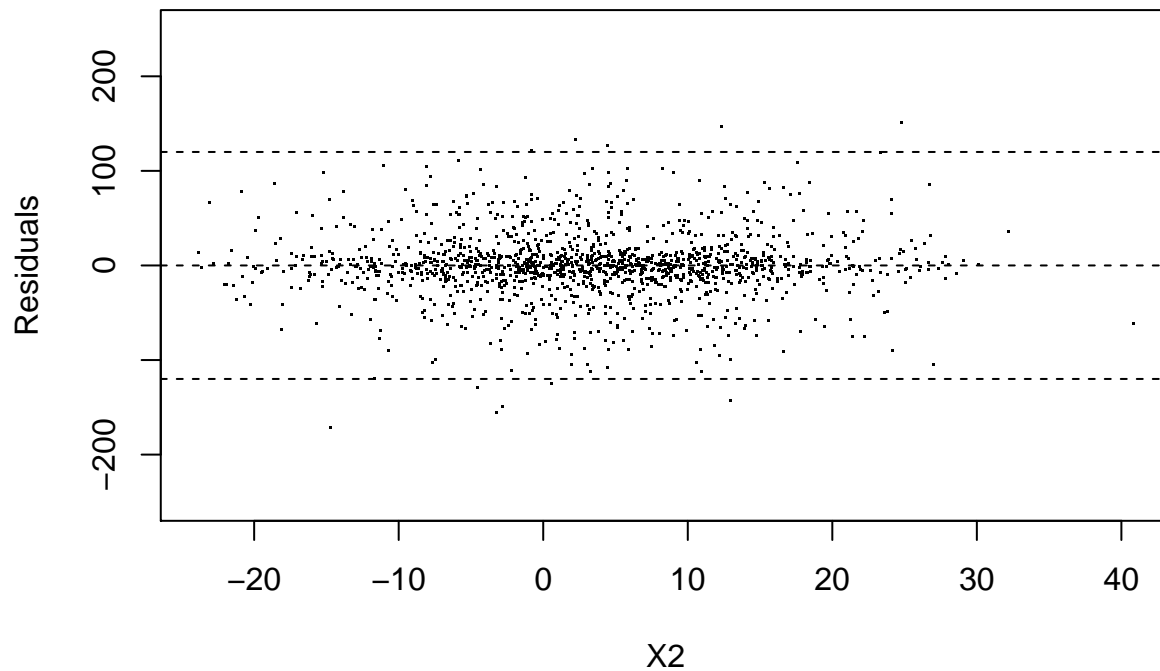
Line of best fit for Data Set 2



```
#Residuals
m2.resids<-residuals(model2)
m2.fitted<-fitted(model2)
```

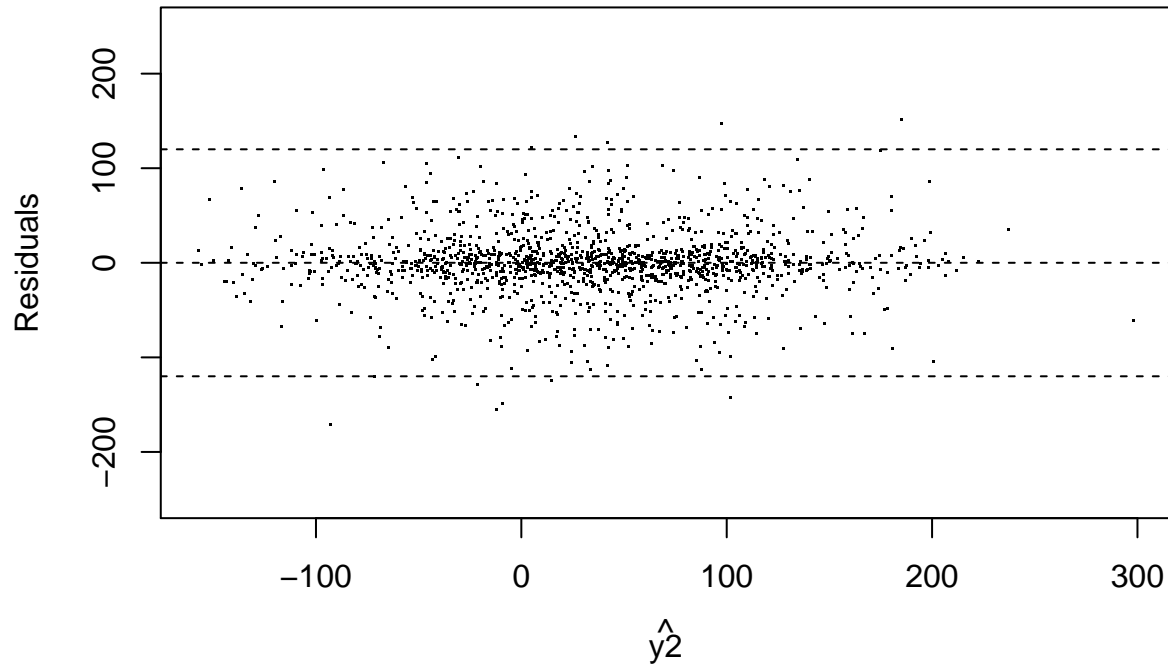
```
#Plotting residuals - Resid v X2
plot(x2, m2.resids,xlab='X2',ylab='Residuals',pch=".",ylim=range(-250,250));
abline(h=0,lty=2); abline(h=120, lty=2);
abline(h=-120, lty=2)
title('Residuals vs X2')
```

Residuals vs X2



```
#Plotting resid - Resid V hat(Y2)
plot(m2.fitted, m2.resids,xlab=expression(hat(y2)),ylab='Residuals',
pch=".",ylim=range(-250,250));abline(h=0,lty=2);abline(h=120, lty=2); abline(h=-120, lty=2)
title('Residuals vs Fitted values')
```

Residuals vs Fitted values



Remark: We can assume there is a positive linear relationship between the covariate and the outcome.

Our estimate for the intercept and the slope would be

$$\hat{\beta}_0 = 10.66085$$

and

$$\hat{\beta}_1 = 7.03795$$

, respectively.

Therefore, our line of best fit is given by:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \Rightarrow \hat{y} = 10.66085 + 7.03795x$$

R also states the result for the tests of the intercept and the slope estimates using t distribution with null and alternative hypotheses.

$$H_0 : \beta_0 = 0$$

and

$$H_a : \beta_0 \neq 0$$

$$H_0 : \beta_1 = 0$$

and

$$H_a : \beta_1 \neq 0$$

We can see that but we have sufficient evidence to reject the null

$$H_0$$

for both the intercept estimate and the slope estimate.

From the residual plots, we can notice from both residual plots that they are centered around zero and have constant variance. Our assumptions for the expected value of the error and homoscedasticity of variance have been kept. We could say that our residual,

$$e_i$$

, represents the behaviour our true random error,

$$\epsilon_i$$

N.B I have also added plots for Residuals v. Fitted value to get my result confirmed.

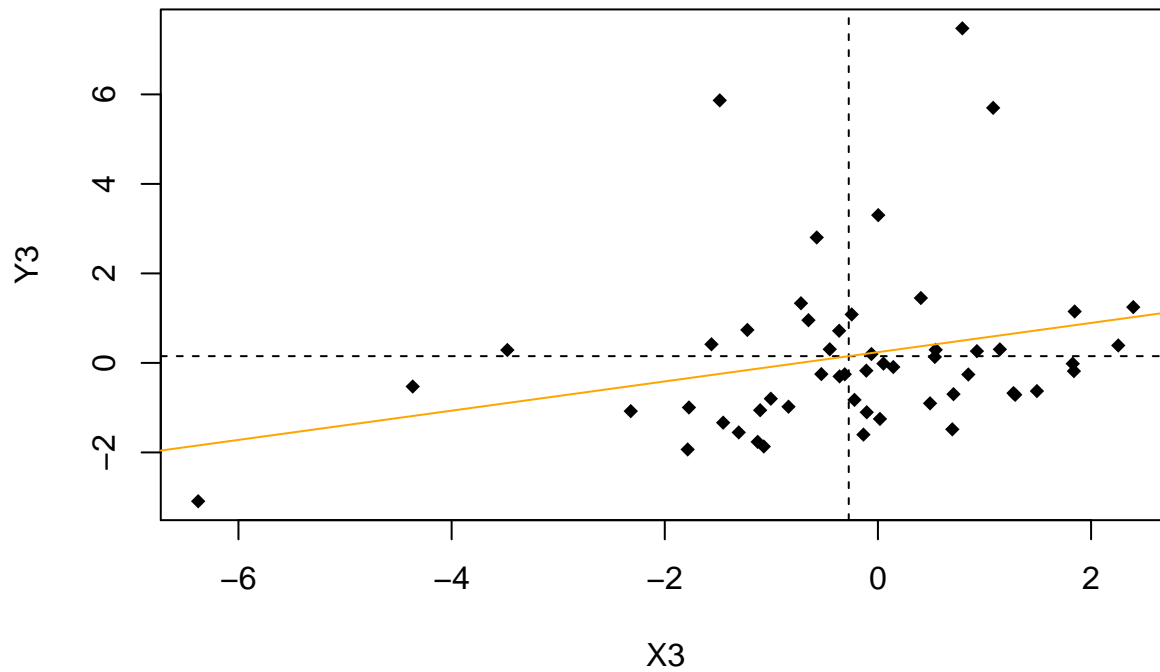
Data set 3

```
file3<- "http://www.math.mcgill.ca/yyang/regression/data/a1-3.txt"
data3<-read.table(file3,header=TRUE)
plot(data3$x,data3$y,pch=18, xlab = "X3", ylab="Y3")
x3<-data3$x
y3<-data3$y
abline(v=mean(x3),h=mean(y3),lty=2)
model3 <- lm(y3 ~ x3)
summary(model3)
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9524 -1.1700 -0.3927  0.5966  6.9769
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2403     0.2602   0.924   0.360
## x3              0.3268     0.1635   1.998   0.051 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.866 on 51 degrees of freedom
## Multiple R-squared:  0.07263,    Adjusted R-squared:  0.05444
## F-statistic: 3.994 on 1 and 51 DF,  p-value: 0.05101
coef(model3)
```

```
## (Intercept)          x3
##  0.2403328    0.3267628
abline(coef(model3),col='orange')
title('Line of best fit for Data Set 3')
```

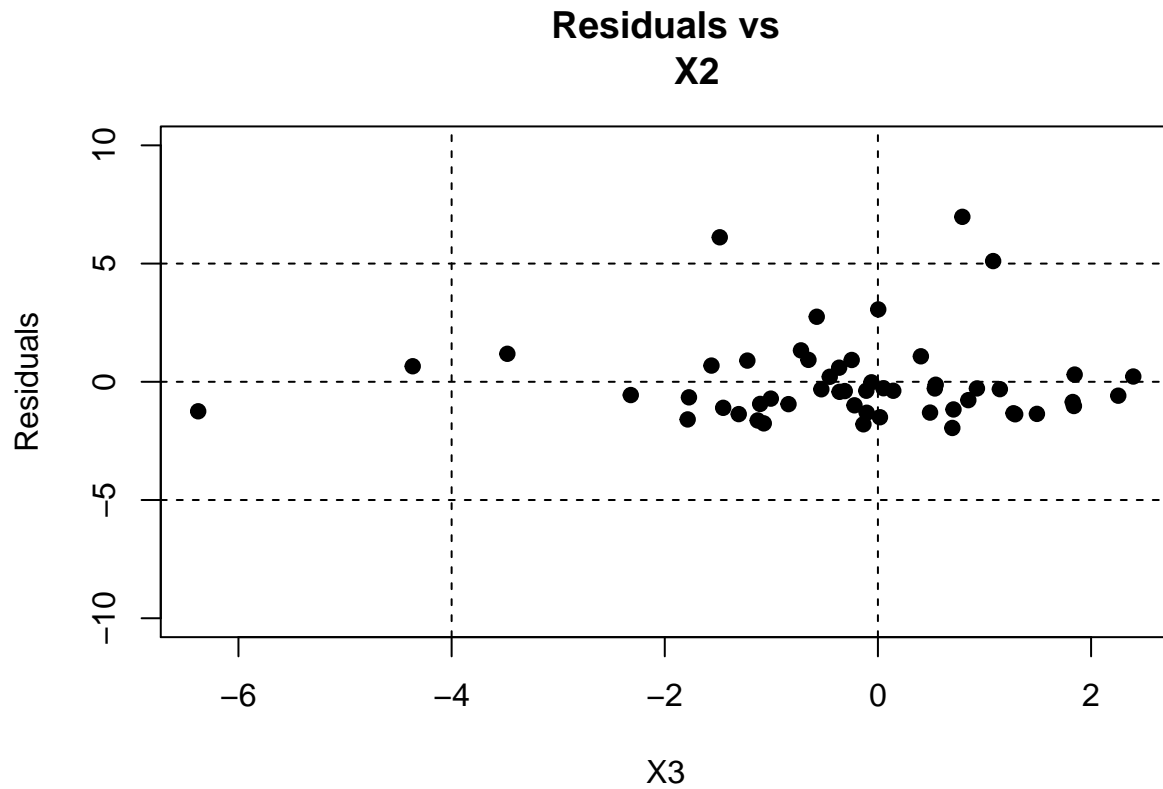

Line of best fit for Data Set 3



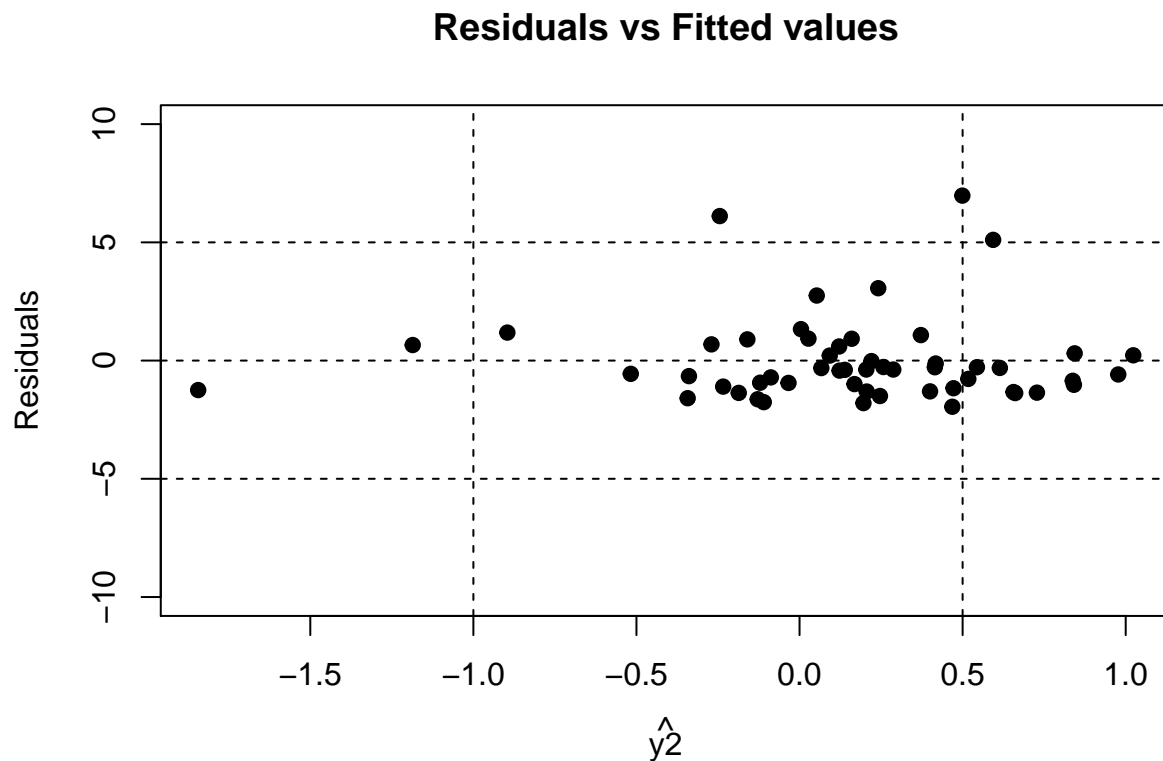
#Residuals

```
m3.resids<-residuals(model3)
m3.fitted<-fitted(model3)
```

```
#Plots for Residuals v. X3 and Residuals v. Fitted Values
plot(x3, m3.resids,xlab='X3',ylab='Residuals',
pch=19,ylim=range(-10,10));abline(h=0,lty=2);abline(h=5, lty=2);
abline(h=-5,lty=2);
abline(v=0, lty=2); abline(v=-4.0,lty=2)
title('Residuals vs
      X2')
```



```
plot(m3.fitted, m3.resids,xlab=expression(hat(y2)),ylab='Residuals',
pch=19,ylim=range(-10,10));abline(h=0,lty=2);abline(h=5, lty=2);
abline(h=-5,lty=2); abline(v=0.5, lty=2); abline(v=-1.0, lty=2)
title('Residuals vs Fitted values')
```



Remark: We might be able to assume that there is a weak positive linear relationship between the covariate and

the outcome.

Our estimate for the intercept and the slope would be

$$\hat{\beta}_0 = 0.2403$$

and

$$\hat{\beta}_1 = 0.3268$$

, respectively.

Therefore, our line of best fit is given by:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \Rightarrow \hat{y} = 0.2403 + 0.3268x$$

R also states the result for the tests of the intercept and the slope estimates using t distribution with null and alternative hypotheses.

$$H_0 : \beta_0 = 0$$

and

$$H_a : \beta_0 \neq 0$$

$$H_0 : \beta_1 = 0$$

and

$$H_a : \beta_1 \neq 0$$

We can see that but we have failed to reject the null

$$H_0$$

for both the intercept estimate and the slope estimate.

Both residual plots suggest that our homogeneity of variance assumption might have been broken. We can notice the variance is actually increasing as X3 increases. However, it is rather difficult to make correct judgment due to a significant number of outliers.

N.B I have also added plots for Residuals v. Fitted value to get my result confirmed.

Part b

From class, we know

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \end{bmatrix}$$

so

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \end{bmatrix}$$

and

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix}^{-1} = \frac{1}{n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2} \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & -\sum_{i=1}^n x_{i1} \\ -\sum_{i=1}^n x_{i1} & n \end{bmatrix}$$

then we have

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \frac{1}{n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2} \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & -\sum_{i=1}^n x_{i1} \\ -\sum_{i=1}^n x_{i1} & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \end{bmatrix}$$

$$\hat{\beta}_0 = \frac{1}{n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2} \left[\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_{i1} \sum_{i=1}^n x_{i1} y_i \right] = 1/S_{xx} [\bar{y} \sum_{i=1}^n x_{i1}^2 - \bar{x} \sum_{i=1}^n x_{i1} y_i]$$

$$\hat{\beta}_1 = \frac{1}{n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2} \left[n \sum_{i=1}^n x_{i1} y_i - \sum_{i=1}^n x_{i1} \sum_{i=1}^n y_i \right] = \frac{1}{S_{xx}} \left[\sum_{i=1}^n x_{i1} y_i - \frac{1}{n} \sum_{i=1}^n x_{i1} \sum_{i=1}^n y_i \right]$$

(i) Location shift :

$$x_{i1new} \leq x_{i1} - m$$

Replace all

$$x_{i1}$$

as

$$x_{i1} - m$$

, then

$$\begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{bmatrix} = \frac{1}{n \sum_{i=1}^n (x_{i1} - m)^2 - (\sum_{i=1}^n (x_{i1} - m))^2} \begin{bmatrix} \sum_{i=1}^n (x_{i1} - m)^2 & -\sum_{i=1}^n (x_{i1} - m) \\ -\sum_{i=1}^n (x_{i1} - m) & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n (x_{i1} - m) y_i \end{bmatrix}$$

$$\begin{aligned} \tilde{\beta}_0 &= \frac{1}{n \sum_{i=1}^n (x_{i1} - m)^2 - (\sum_{i=1}^n (x_{i1} - m))^2} \left[\sum_{i=1}^n (x_{i1} - m)^2 \sum_{i=1}^n y_i - \sum_{i=1}^n (x_{i1} - m) \sum_{i=1}^n (x_{i1} - m) y_i \right] \\ &= \frac{1}{\sum_{i=1}^n (x_{i1} - m)^2 - \frac{1}{n} (\sum_{i=1}^n (x_{i1} - m))^2} \left[\frac{1}{n} \sum_{i=1}^n (x_{i1} - m)^2 \sum_{i=1}^n y_i - \frac{1}{n} \sum_{i=1}^n (x_{i1} - m) \sum_{i=1}^n (x_{i1} - m) y_i \right] \end{aligned}$$

$$\tilde{\beta}_1$$

$$\begin{aligned} &= \frac{1}{n \sum_{i=1}^n (x_{i1} - m)^2 - (\sum_{i=1}^n (x_{i1} - m))^2} \left[n \sum_{i=1}^n (x_{i1} - m) y_i - \sum_{i=1}^n (x_{i1} - m) \sum_{i=1}^n y_i \right] \\ &= \frac{1}{\sum_{i=1}^n (x_{i1} - m)^2 - \frac{1}{n} (\sum_{i=1}^n (x_{i1} - m))^2} \left[\sum_{i=1}^n (x_{i1} - m) y_i - \frac{1}{n} \sum_{i=1}^n (x_{i1} - m) \sum_{i=1}^n y_i \right] \end{aligned}$$

Remark

$$\begin{aligned} &\sum_{i=1}^n (x_{i1} - m)^2 - \frac{1}{n} (\sum_{i=1}^n (x_{i1} - m))^2 \\ &= \sum_{i=1}^n (x_{i1}^2 - 2mx_{i1} + m^2) - (\bar{x} - m) (\sum_{i=1}^n x_{i1} - nm) \\ &= \sum_{i=1}^n x_{i1}^2 - 2nm\bar{x} + nm^2 - n\bar{x}^2 + nm\bar{x} + nm\bar{x} - nm^2 \\ &= \sum_{i=1}^n x_{i1}^2 - n\bar{x}^2 = S_{xx} \end{aligned}$$

then

$$\tilde{\beta}_1 = \frac{1}{S_{xx}} \left[\sum_{i=1}^n (x_{i1} - m) y_i - \frac{1}{n} \sum_{i=1}^n (x_{i1} - m) \sum_{i=1}^n y_i \right]$$

$$\begin{aligned}
&= \frac{1}{S_{xx}} \left[\sum_{i=1}^n x_{i1} y_i - m \sum_{i=1}^n y_i - \frac{1}{n} \sum_{i=1}^n x_{i1} \sum_{i=1}^n y_i + m \sum_{i=1}^n y_i \right] \\
&= \frac{1}{S_{xx}} \left[\sum_{i=1}^n x_{i1} y_i - \frac{1}{n} \sum_{i=1}^n x_{i1} \sum_{i=1}^n y_i \right] = \hat{\beta}_1
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\beta}_0 &= \frac{1}{S_{xx}} \left[\frac{1}{n} \sum_{i=1}^n y_i \sum_{i=1}^n (x_{i1} - m)^2 - \frac{1}{n} \sum_{i=1}^n (x_{i1} - m) \sum_{i=1}^n (x_{i1} - m) y_i \right] \\
&= \frac{1}{S_{xx}} \left[\bar{y} \sum_{i=1}^n (x_{i1} - m)^2 - (\bar{x} - m) \sum_{i=1}^n (x_{i1} - m) y_i \right] \\
&= \frac{1}{S_{xx}} \left[\bar{y} \sum_{i=1}^n (x_{i1} - m)^2 - (\bar{x} - m) \sum_{i=1}^n (x_{i1} - m) y_i \right] \\
&= \frac{1}{S_{xx}} \left[\bar{y} \sum_{i=1}^n (x_{i1} - m)^2 - \bar{y} n (\bar{x} - m)^2 + \bar{y} n (\bar{x} - m)^2 - (\bar{x} - m) \sum_{i=1}^n (x_{i1} - m) y_i \right] \\
&= \bar{y} - \left(\frac{\sum_{i=1}^n (x_{i1} - m) y_i - \bar{y} n (\bar{x} - m)}{S_{xx}} \right) (\bar{x} - m) \\
&= \bar{y} - \frac{\sum_{i=1}^n x_{i1} y_i - m \sum_{i=1}^n y_i - \frac{\sum_{i=1}^n x_{i1} y_i}{n} + m \sum_{i=1}^n y_i}{S_{xx}} \\
&= \bar{y} - \frac{\sum_{i=1}^n x_{i1} y_i - \frac{\sum_{i=1}^n x_{i1} y_i}{n}}{S_{xx}} \\
&= \bar{y} - \hat{\beta}_1 (\bar{x} - m) = \hat{\beta}_0 + \hat{\beta}_1 m
\end{aligned}$$

so

$$\hat{\beta}_1$$

remains the same whereas

$$\hat{\beta}_0$$

increases by

$$\hat{\beta}_1 m$$

Let's look at a numerical example.

Suppose $m = 3$

```
length(x1)
```

```
## [1] 27
```

```
#forming a vector m = (m, m, m, ...)
```

```
m <- rep(3, 27)
```

```
m
```

```
## [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
```

```

#shifting location by m
x1_dash <- x1-m

#fitting new model
model1_dash <- lm(y1 ~ x1_dash)
b1_dash <- coef(model1_dash)[2]
b0_dash <- coef(model1_dash)[1]

#our $beta_0$
b0_dash

## (Intercept)
##      5.202128

#Our theory suggests our $beta_0$ should be equal to
mean(y1) - b1_dash*(mean(x1_dash))

## x1_dash
## 5.202128

paste("The shifted model has coefficients :", coef(model1_dash))

## [1] "The shifted model has coefficients : 5.20212824866401"
## [2] "The shifted model has coefficients : 1.72512090818717"

paste("The original model has coefficients :", coef(model1))

## [1] "The original model has coefficients : 0.0267655241025053"
## [2] "The original model has coefficients : 1.72512090818717"

summary(model1_dash)

##
## Call:
## lm(formula = y1 ~ x1_dash)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.7577 -1.1595 -0.1691  1.5003  3.7808
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.2021     0.6088   8.545 6.93e-09 ***
## x1_dash         1.7251     0.1437  12.003 7.14e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.182 on 25 degrees of freedom
## Multiple R-squared:  0.8521, Adjusted R-squared:  0.8462
## F-statistic: 144.1 on 1 and 25 DF,  p-value: 7.145e-12

```

It seems that our numerical example agrees with what the theory has predicted.

(ii) rescaling

Replace all

$$x_{i1}$$

as

$$lx_{i1}$$

, then since we know

$$\begin{aligned} \sum_{i=1}^n lx_{i1} &= l \sum_{i=1}^n x_{i1} \\ \begin{bmatrix} \star\beta_0 \\ \star\beta_1 \end{bmatrix} &= \frac{1}{nl^2 \sum_{i=1}^n x_{i1}^2 - (l \sum_{i=1}^n x_{i1})^2} \begin{bmatrix} l^2 \sum_{i=1}^n x_{i1}^2 & -l \sum_{i=1}^n x_{i1} \\ -l \sum_{i=1}^n x_{i1} & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ l \sum_{i=1}^n x_{i1} y_i \end{bmatrix} \\ \star\beta_0 &= \frac{1}{nl^2 \sum_{i=1}^n x_{i1}^2 - l^2 (\sum_{i=1}^n x_{i1})^2} [l^2 \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n y_i - l^2 \sum_{i=1}^n x_{i1} \sum_{i=1}^n x_{i1} y_i] \\ &\quad l^2 \end{aligned}$$

would cancel out, so

$$\star\beta_0 = \hat{\beta}_0$$

The intercept remains intact, and for the slope

$$\star\beta_1 = \frac{1}{nl^2 \sum_{i=1}^n x_{i1}^2 - l^2 (\sum_{i=1}^n x_{i1})^2} [nl \sum_{i=1}^n x_{i1} y_i - l \sum_{i=1}^n x_{i1} \sum_{i=1}^n y_i]$$

If you reduce

$$l$$

from the numerator, you will have

$$\star\beta_1 = \frac{1}{nl \sum_{i=1}^n x_{i1}^2 - l (\sum_{i=1}^n x_{i1})^2} [n \sum_{i=1}^n x_{i1} y_i - \sum_{i=1}^n x_{i1} \sum_{i=1}^n y_i] = \frac{1}{l} \hat{\beta}_1$$

so our slope will shrink by

$$\frac{1}{l}$$

Let's take a numerical example once again, with

$$l = 3$$

```
x2_dash <- x2*3
model2_dash <- lm(y2 ~x2_dash)

#our original intercept
coef(model2)[1]

## (Intercept)
##      10.66085

#our original slope
coef(model2)[2]

##          x2
##      7.037952
```

```
paste("The rescaled model has coefficients :", coef(model2_dash))

## [1] "The rescaled model has coefficients : 10.6608534071323"
## [2] "The rescaled model has coefficients : 2.34598408427181"

paste("The original model has coefficients :", coef(model2))

## [1] "The original model has coefficients : 10.6608534071323"
## [2] "The original model has coefficients : 7.03795225281544"

paste("If you multiply 1/3 to our original slope", 1/3*coef(model2)[2])

## [1] "If you multiply 1/3 to our original slope 2.34598408427181"
```

It shows that what our theory suggested holds.

(iii)

From class, we know

$$\begin{aligned} E[\hat{\beta}_0|X] &= \beta_0 E[\hat{\beta}_1|X] = \beta_1 \\ \text{Var}[\hat{\beta}_0|X] &= \frac{\sigma^2}{nS_{xx}} \left[\sum_{i=1}^n x_{i1}^2 \right] = \frac{\sigma^2}{n} \left[1 + \frac{n\bar{x}^2}{S_{xx}} \right] \\ \text{Var}[\hat{\beta}_1|X] &= \frac{\sigma^2}{S_{xx}} \end{aligned}$$

Location shift:

$$\tilde{\beta}_0 = \hat{\beta}_0 + \hat{\beta}_1 m \tilde{\beta}_1 = \hat{\beta}_1$$

so

$$E[\tilde{\beta}_0|X] = E[\hat{\beta}_0 + \hat{\beta}_1 m|X] = \beta_0 + \beta_1 m E[\tilde{\beta}_1|X] = E[\hat{\beta}_1|X] = \beta_1$$

Its intercept estimate is no longer an unbiased estimator, but the slope estimate remains unbiased.

For variance, since

$$S_{xx}$$

remains the same under location shift

$$\begin{aligned} \text{Var}[\tilde{\beta}_0|X] &= \frac{\sigma^2}{nS_{xx}} \left[\sum_{i=1}^n (x_{i1} - m)^2 \right] = \frac{\sigma^2}{n} \left[1 + \frac{n(\bar{x} - m)^2}{S_{xx}} \right] \\ \text{Var}[\tilde{\beta}_1|X] &= \text{Var}[\hat{\beta}_1|X] = \frac{\sigma^2}{S_{xx}} \end{aligned}$$

We can notice that our estimate of the intercept shrinks.

Rescaling:

$$\star\beta_0 = \hat{\beta}_0 \star \beta_1 = \frac{1}{l} \hat{\beta}_1$$

so

$$E[\star\beta_0|X] = E[\hat{\beta}_0|X] = \beta_0 E[\star\beta_1|X] = E\left[\frac{1}{l} \hat{\beta}_1|X\right] = \frac{1}{l} \beta_1$$

Its intercept estimate remains unbiased but the slope estimate shrinks, making itself a biased estimator.
For variance,

$$\text{Var}[\star\beta_0|X] = \text{Var}[\hat{\beta}_0|X] = \frac{\sigma^2}{nS_{xx}} \left[\sum_{i=1}^n x_{i1}^2 \right] = \frac{\sigma^2}{n} \left[1 + \frac{n\bar{x}^2}{S_{xx}} \right]$$

and

$$\text{Var}[\star\beta_1|X] = \text{Var}\left[\frac{1}{l}\hat{\beta}_1|X\right] = \frac{\sigma^2}{l^2 S_{xx}}$$

The variance for the intercept estimate remains the same, but our variance for the slope intercept shrinks.

Extra Question

Part 1.

$$\|Y - X\beta\| = \sqrt{(Y - X\beta)^2} = \sqrt{\sum_{i=1}^n (y_i - \vec{x}_i\beta)^2}$$

$$\text{argmin}(\|Y - X\beta\|) = \text{argmin}\left(\sqrt{\sum_{i=1}^n (y_i - \vec{x}_i\beta)^2}\right)$$

$$\sqrt{\sum_{i=1}^n (y_i - \vec{x}_i\beta)^2} = \sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2} = S(\beta_0, \beta_1)$$

We set the partial derivatives to 0

For the intercept :

$$\begin{aligned} 0 &= \frac{\partial S}{\partial \beta_0} = \frac{1}{2\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2}} * -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1}) \\ &= \frac{1}{2\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2}} * -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1}) = -\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})}{\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2}} \\ &= -\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})}{\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2}} = -\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})}{\|Y - X\beta\|} \\ &=> \frac{1}{\|Y - X\beta\|} \sum_{i=1}^n y_i = \frac{1}{\|Y - X\beta\|} (n\beta_0 + (\sum_{i=1}^n x_{i1})\beta_1) \end{aligned}$$

, and for the slope :

$$0 = \frac{\partial S}{\partial \beta_1} = \frac{1}{2\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2}} * -2 \sum_{i=1}^n x_{i1} (y_i - \beta_0 - \beta_1 x_{i1})$$

$$\begin{aligned}
&= -\frac{\sum_{i=1}^n x_{i1}(y_i - \beta_0 - \beta_1 x_{i1})}{\sqrt{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1})^2}} = -\frac{\sum_{i=1}^n x_{i1}(y_i - \beta_0 - \beta_1 x_{i1})}{\|Y - X\beta\|} \\
&=> \frac{1}{\|Y - X\beta\|} \sum_{i=1}^n x_{i1}y_i = \frac{\sum_{i=1}^n (x_{i1})\beta_0 + \sum_{i=1}^n x_{i1}^2\beta_1}{\|Y - X\beta\|} \\
&\frac{1}{\|Y - X\beta\|} \begin{bmatrix} n & \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \frac{1}{\|Y - X\beta\|} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \end{bmatrix} \\
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \frac{\|Y - X\beta\|}{(n \sum_{i=1}^n x_{i1}^2 - (\sum_{i=1}^n x_{i1})^2)} \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & -\sum_{i=1}^n x_{i1} \\ -\sum_{i=1}^n x_{i1} & n \end{bmatrix} \frac{1}{\|Y - X\beta\|} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \end{bmatrix}
\end{aligned}$$

Which is our equation for SLR least square estimates, since the Euclidean norms would cancel out.

So our β_0 and β_1 would be essentially the same least square estimates, but there would be some different characteristics. Firstly, the exaggerated difference between the fitted value and our actual outcome by using some of squares gets reduced.

Part 2.

$$\sum_{i=1}^n |y_i - \bar{x}_i \beta|$$

The fitted value whose absolute value of the difference from the actual outcome value is larger than 1 gets more accurately represented. Thus, this robust regression is more resilient to outliers with large scale residuals.