Assignment1

2019150445 BaekRok

2020 9 21

1.

(a)

```
X<-c(2,3,3,3,4,5)
Y<-c(0,1,1,2,2,3)
lm(Y~X)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Coefficients:
## (Intercept)            X
##     -1.6250       0.9375
```

(b)

```
lm(Y~X-1)
```

```
##
## Call:
## lm(formula = Y ~ X - 1)
##
## Coefficients:
##      X
## 0.4861
```

2.

(a)

```
EDU<-c(12,20,20,14,16,16,18,14,12,16,15,10)
Income<-c(35,80,78,45,57,65,59,63,57,66,73,23)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

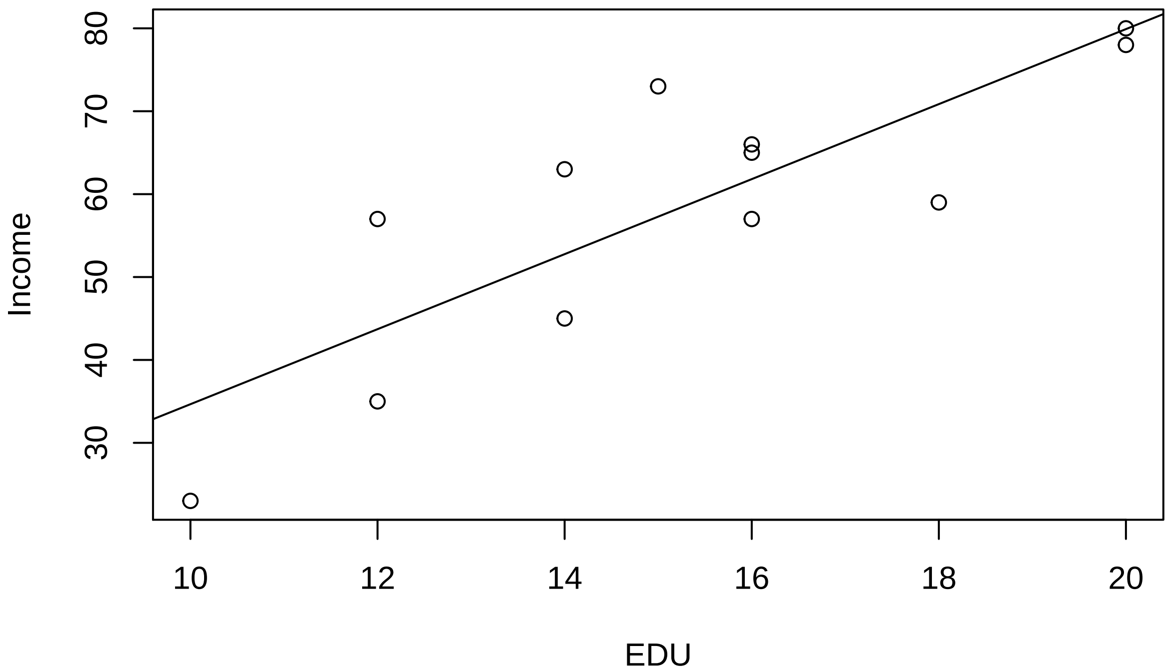
```
lm(Income~EDU) %>% summary()
```

```
##
## Call:
## lm(formula = Income ~ EDU)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.860  -7.998  -0.909   5.702  15.714
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -10.585     15.041  -0.704 0.497647
## EDU            4.525      0.968   4.674 0.000875 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.978 on 10 degrees of freedom
## Multiple R-squared:  0.686,  Adjusted R-squared:  0.6546
## F-statistic: 21.85 on 1 and 10 DF,  p-value: 0.0008753
```

There does appear to be a positive linear relationship between EDU and Income. By a simple linear regression model, income increases by 4.525(1000USD) when one unit(1 year) of EDU increase.

(b)

```
plot(Income~EDU)
lm(Income~EDU) %>% abline()
```



The equation of the regression line is

$$(\text{Income in 1000 USD}) = -10.585 + 4.525 \times (\text{EDU in years}).$$

(c)

```r
cov(EDU,Income)
```

```
## [1] 43.70455
```

```r
sd(EDU)
```

```
## [1] 3.107908
```

```r
sd(Income)
```

```
## [1] 16.97837
```

3

(a)

```r
Length<-c(5.1,4.9,4.7,4.6,5.0,5.4,4.6,5.0,4.4,4.9,5.4,4.8)
Width<-c(3.5,3.0,3.2,3.1,3.6,3.9,3.4,3.4,2.9,3.1,3.7,3.4)
mean(Length)
```

```
## [1] 4.9
```

```r
mean(Width)
```

```
## [1] 3.35
```

```r
cor(Length,Width)
```

```
## [1] 0.8081269
```

Correlation between Length and Width is 0.8081269, which is positive and close to 1. So we can say Length and Width have positive and strong linear relationship.

(b)

```r
lm(Width~Length)
```

```
##
## Call:
## lm(formula = Width ~ Length)
##
## Coefficients:
## (Intercept)       Length
##     -0.5135       0.7885
```

The fitted simple linear regression line is

$$\text{Width} = -0.5135 + 0.7885 \times \text{Length}$$

Estimated Intercept of the regression line is -0.5135, and slope of the regression line is 0.7885. That is, when one unit of Length increases, on average Width increases by 0.7885

(c)

```r
lm(Width~Length) %>% summary() %>% .$sigma
```

```
## [1] 0.1853272
```

(d)

We have to test

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0$$

In order to do this, we have to compute p-value.

```r
lm(Width~Length) %>% summary() #To find p-value.
```

```
##
## Call:
## lm(formula = Width ~ Length)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35000 -0.04712 -0.01058  0.13558  0.28654
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5135     0.8921  -0.576  0.57762
## Length        0.7885     0.1817   4.339  0.00147 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1853 on 10 degrees of freedom
## Multiple R-squared:  0.6531, Adjusted R-squared:  0.6184
## F-statistic: 18.82 on 1 and 10 DF,  p-value: 0.001469
```

p-value is 0.0147 which is less than 0.01. So we can say there is linear relationship between Length and Width and thus We can reject H_0.

Or we can compute p-value through $\hat{\beta}_1$ & $s.e.(\hat{\beta}_1)$.

$$t = 0.7885/0.1817 = 4.339$$

And thus,

```r
2*(1-pt(4.339,length(Length)-2))
```

```
## [1] 0.001468675
```

(e)

We can compute 95% confidence interval when $x_0 = 4.8$ by this code.

```r
lm(Width~Length) %>% predict(interval="confidence",newdata=data.frame(Length=4.8))
```

```
##        fit     lwr      upr
## 1 3.271154 3.14526 3.397047
```

Or we can compute CI by hand.

The SE of $\hat{\mu}_0$:

$$s.e.(\hat{\mu}_0) = \hat{\sigma}\sqrt{(1/n + (x_0 - \bar{x})^2 / \sum(x_i - \bar{x}^2)}$$

The confidence interval for mean response is

$$\hat{\mu}_0 \pm t_{(n-2,\alpha/2)} \times s.e.(\hat{\mu}_0)$$

Then we can estimated mean Width when Length(x_0)=4.8

$$\hat{\mu}_0 = -0.5135 + 0.7885 \times 4.8 = 3.271$$

And from above equation,

$$s.e.(\hat{\mu}_0) = 0.0565$$

We have to construct 95% confidence interval, which means $\alpha = 0.05$.

Thus, 95% confidence interval is

$$(3.271 \pm t_{12-2,0.025} \times 0.0565) = (3.271 \pm 2.228 \times 0.056) = (3.145, 3.397)$$

(f)

We can compute 95% prediction when $x_0 = 4.8$interval by this code.

```
lm(Width~Length) %>% predict(interval='predict',newdata=data.frame(Length=4.8))
```

```
##        fit      lwr      upr
## 1 3.271154 2.839455 3.702853
```

Note that

$$\hat{y}_0 = \hat{\mu}_0 = 3.271$$

In prediction interval,

$$s.e.(\hat{y}_0) = \hat{\sigma}\sqrt{1 + 1/n + (x_0 - \bar{x})^2 / \sum (x_i - \bar{x})^2}$$

Since $x_0 = 4.8$, $s.e.(\hat{y}_0) = 0.1937$.

The (1-α) prediction value for the predicted value y_0 at $x = x_0$ is

$$\hat{y}_0 \pm t_{n-2,\alpha/2} \times s.e.(\hat{y}_0)$$

Then 95% prediction interval is

$$(3.271 \pm 2.228 \times 0.1937) = (2.839, 3.703)$$