Lab2 Solution

Leanne Dong

9 August 2018

Exercise 1

(a) Plot the points (-1,8), (0, 5), (1, 6), (3, 2), (3, 4), (5,1) on a set of axes.

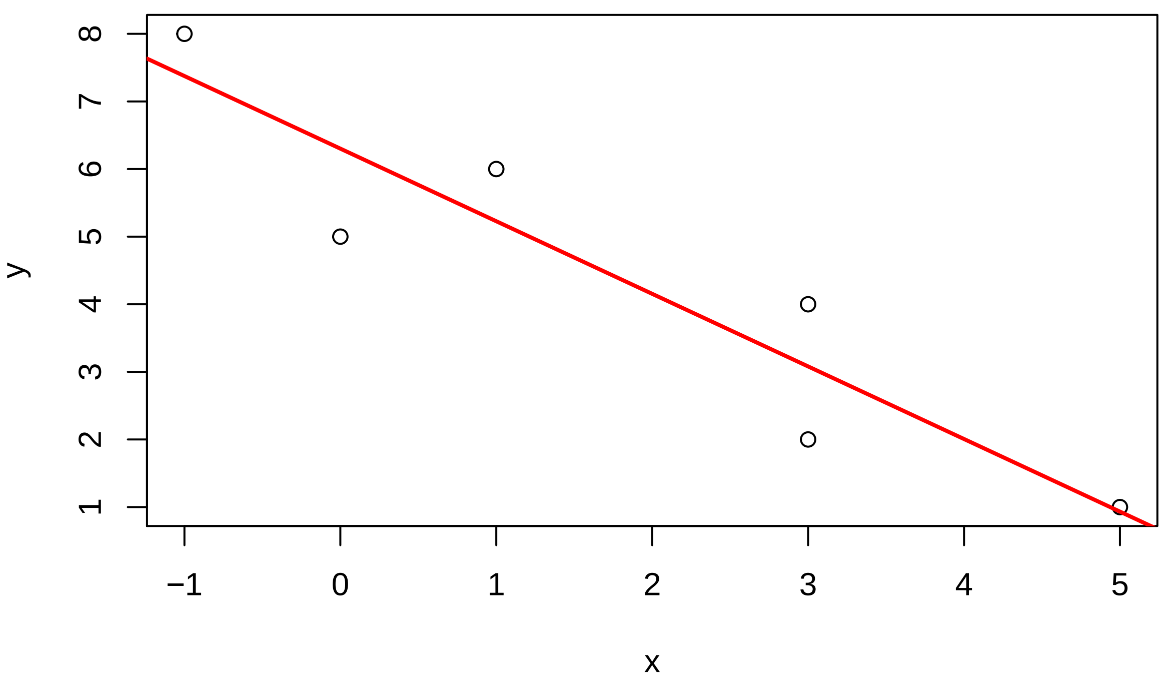
(b) Find the least squares line fit for this data and plot this on the scatterplot.

(c) Find Pearson correlation coefficient and the coefficient of determination for this scatterplot.

```
#scatterplot
x<- c(-1,0,1,3,3,5)
y<-c(8,5,6,2,4,1)
plot(x,y)
#fit least square line
fit<- lm(y~x)
fit
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##       6.302       -1.074
```

```
abline(fit,col="red",lwd=2)
```




```
# Find Pearson correlation coefficient and R^2
cor(y,x)
```

```
```

```
## [1] -0.9268551
```

```
cor(y,x)^2
```

```
## [1] 0.8590604
```

Q3

(a) What can be observed from the scatterplots of GNP vs LE and GNP vs CBR?

Both scatterplots show non-linear relationship between two corresponding variables.

(b) No, it will not be appropriate because these variables have a non-linear relationship.

(c) It does not require a linear relationship but only a monotonic relationship.

(d)

```
GNP<-c(2360,340,780,940,1120,1940,7750,2520,16090,240)
LE<-c(65,59,62,64,66,74,74,71,76,47)
CBR<-c(34,31,36,36,32,32,18,21,14,48)
cor(GNP,CBR,method="spearman")
```

```
## [1] -0.7744046
```

```
cor(GNP,LE,method="spearman")
```

```
## [1] 0.9240164
```

(e)

(f) GNP vs LE: The spearman rank coefficient comes very close this maximum possible association, indicating a very strong positive correlation between GNP and LE. People in countries with higher income levels tend to live longer than do people in countries with lower income levels.

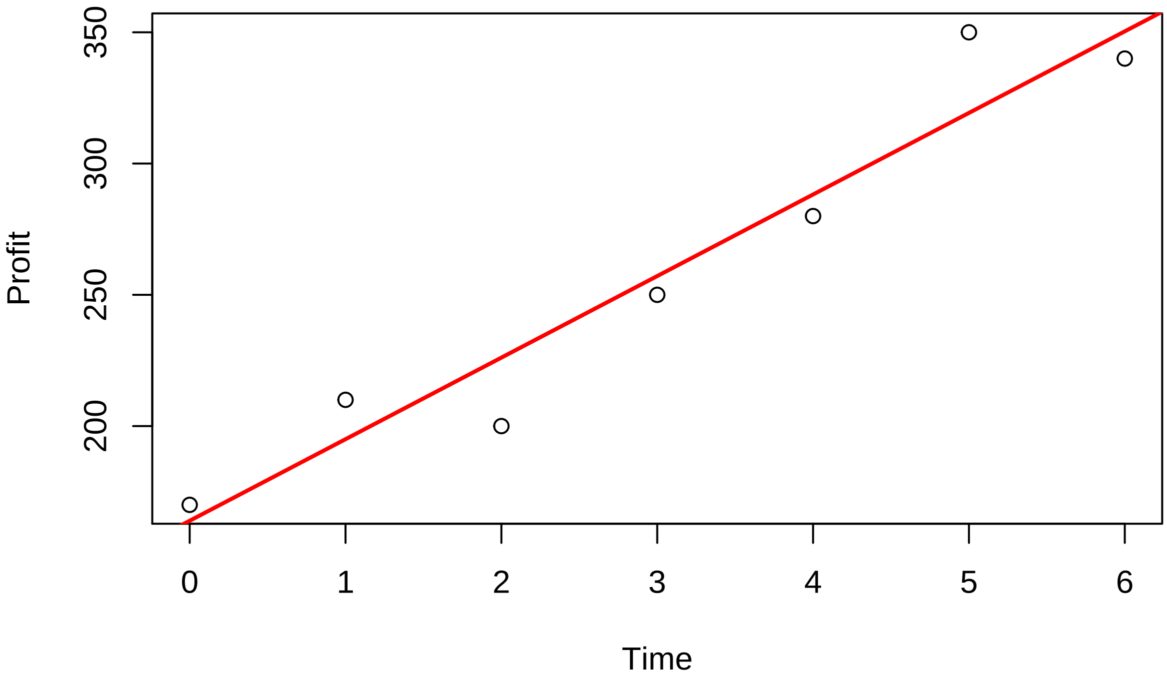
(ii) The Spearman rank coefficient is fairly large negative number, indicating a strong negative correlation between GNP and CBR. Among these 10 countries, higher values of GNP are generally associated with lower values of the birth rate.

Q4 The table gives the profit figures for a small company over a 7 year period.

```
Time<-c(0,1,2,3,4,5,6)
Profit<-c(170,210,200,250,280,350,340)
#scatter plot
plot(Time,Profit)
#find least square line
fit<- lm(Profit~Time)
fit
```

```
##
## Call:
## lm(formula = Profit ~ Time)
##
## Coefficients:
## (Intercept)         Time
##      163.93        31.07
```

```
abline(fit,col="red",lwd=2)
```




```
# Find Pearson correlation coefficient and R^2
cor(Profit,Time)
```

```
## [1] 0.9631062
```

```
cor(Profit,Time)^2
```

```
## [1] 0.9275735
```

The least square line is

$$\text{Profit} = 163.93 + 31.07 * \text{Time}$$

The slope suggests profit is increasing by \$31000 per annum.

(e) The predicted profit in 2019 is $163.93 + 31.07 * 7 = 381.42(000)$

```
163.93+31.07*7
```

```
## [1] 381.42
```

Q6 The table gives measurements of maple trees dimensions were taken with the trunk diameter measurements being taken at 1.5 m height above ground.

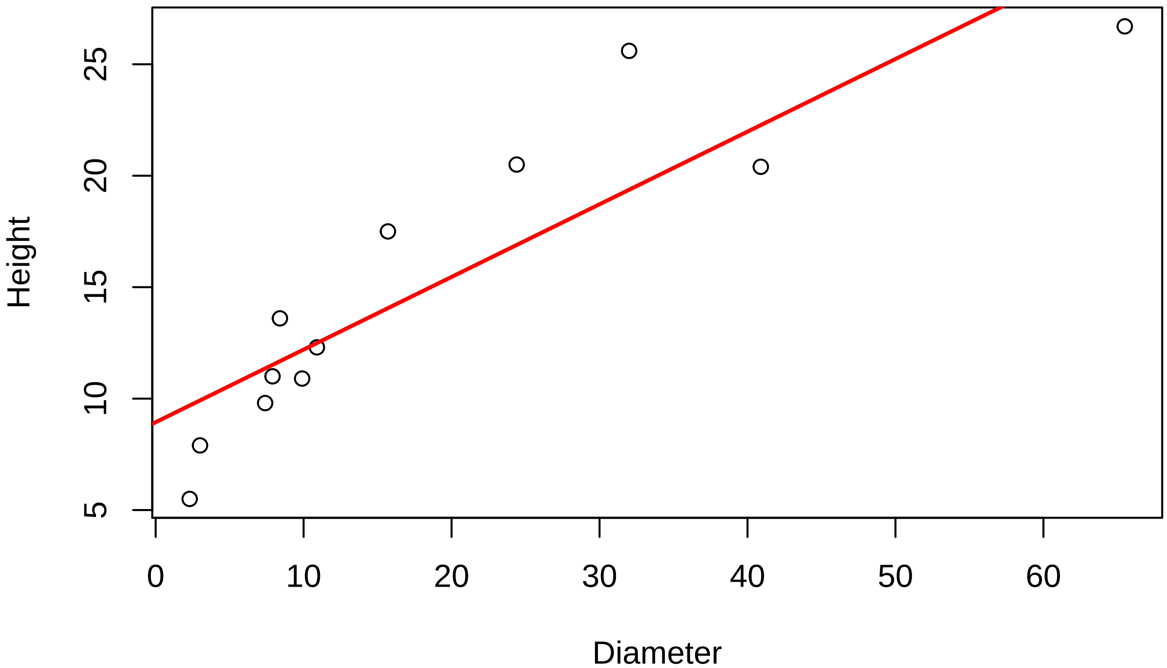
Solution

```
Diameter<-c(2.3,3.0,7.4,7.9,8.4,9.9,10.9,15.7,24.4,32.0,40.9,65.5)
Height<-c(5.5,7.9,9.8,11.0,13.6,10.9,12.3,17.5,20.5,25.6,20.4,26.7)
plot(Diameter,Height)
#fit least square line
fit<- lm(Height~Diameter)
fit
```

```
##
## Call:
```

```
## lm(formula = Height ~ Diameter)
##
## Coefficients:
## (Intercept)     Diameter
##      8.9410       0.3259
```

```r
abline(fit,col="red",lwd=2)
```




```r
# Find Pearson correlation coefficient and R^2
cor(Height,Diameter)
```

```
## [1] 0.8892891
```

```r
cor(Height,Diameter)^2
```

```
## [1] 0.7908351
```

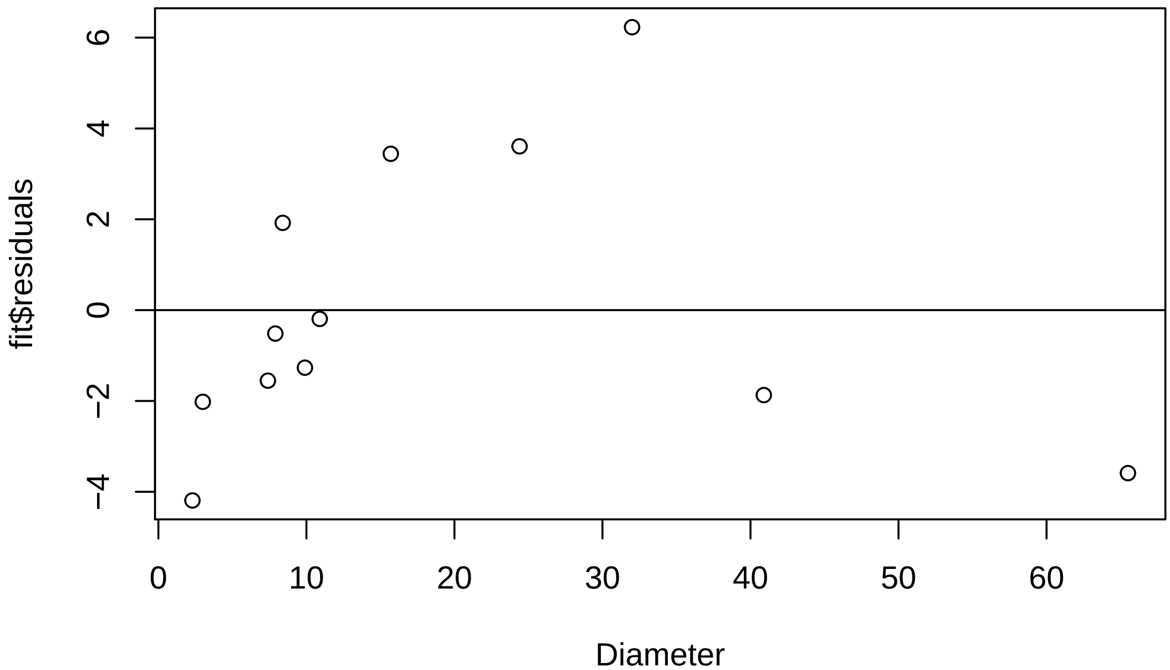
(b) The data seems curved in scatterplot, though correlation and coefficient of determination are quite high.

(c) The correlation coefficient is 0.89. And the R^2 is 0.79.

(d)

$$\text{height} = 8.94 + 0.33 * \text{diameter}$$

```r
#Residual plot
plot(Diameter,fit$residuals)
abline(h=0)
```



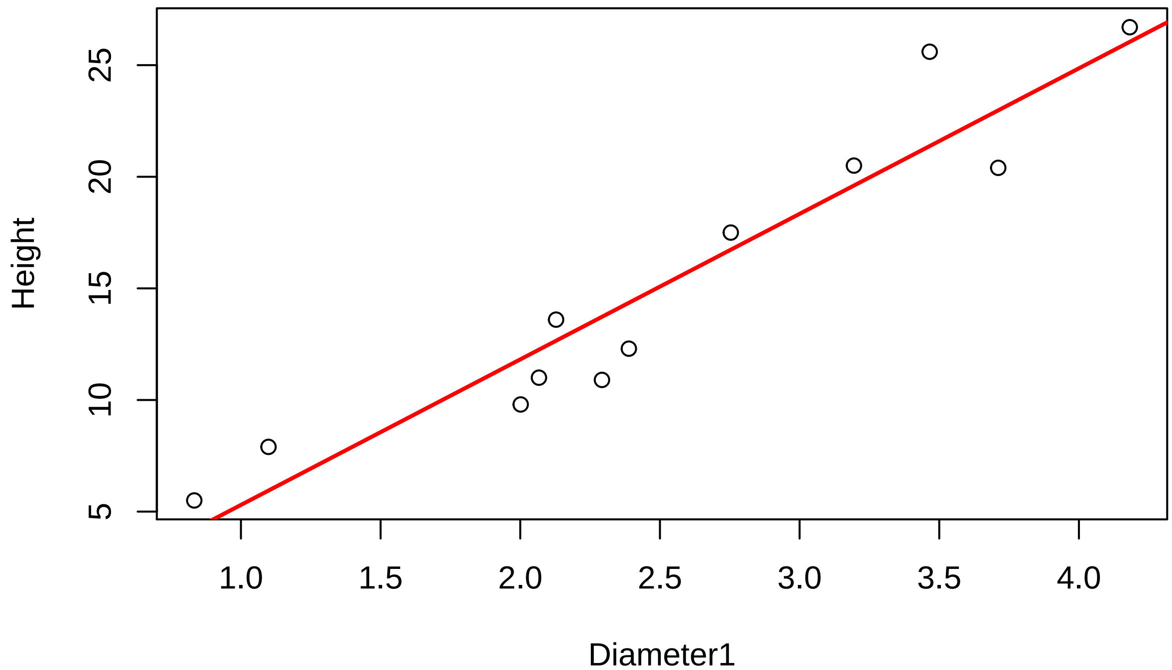

```
Diameter1<-log(Diameter)
Diameter1
```

```
##  [1] 0.8329091 1.0986123 2.0014800 2.0668628 2.1282317 2.2925348 2.3887628
##  [8] 2.7536607 3.1945831 3.4657359 3.7111301 4.1820501
```

```
plot(Diameter1,Height)
#fit least square line
fit<- lm(Height~Diameter1)
fit
```

```
##
## Call:
## lm(formula = Height ~ Diameter1)
##
## Coefficients:
## (Intercept)    Diameter1
##      -1.221        6.520
```

```
abline(fit,col="red",lwd=2)
```




```
# Find Pearson correlation coefficient and R^2
cor(Height,Diameter1)
```

```
## [1] 0.9505538
```

```
cor(Height,Diameter1)^2
```

```
## [1] 0.9035525
```

This shows an improvement over the initial model which had a coefficient of determination of only 0.7908, and is clear from the scatter of points about the trendline that the residual plot would be better.

Our least square line is

$$\text{height} = -1.221 + 6.520 * \log(\text{Diameter})$$

(e) A calculation of the Spearman correlation will not yield a model- only an idea of an extent relationship.