Probability and Statistics Lab 8

Name: Heli Vijay Naliapara
Roll no.: K068

1.
```
mean_x=65
mean_y=67
sd_x = 2.5
sd_y = 3.5
r = 0.75
x = 70
byx = r*(sd_y/sd_x)
y = byx*(x-mean_x)+mean_y
print(y)
```

OUTPUT:

```
> mean_x=65
> mean_y=67
> sd_x = 2.5
> sd_y = 3.5
> r = 0.75
> x = 70
> byx = r*(sd_y/sd_x)
> y = byx*(x-mean_x)+mean_y
> print(y)
[1] 72.25
>
```

Question 1

$r = 0.75$

	Calcutta (X)	Mumbai (Y)
Mean	$\bar{x} = 65$	$\bar{y} = 67$
S.D.	$\sigma_x = 2.5$	$\sigma_y = 3.5$

$$b_{yx} = r\left(\frac{\sigma_y}{\sigma_x}\right) = 0.75 \times \frac{3.5}{2.5} = 1.05$$

$$b_{xy} = r\left(\frac{\sigma_x}{\sigma_y}\right) = 0.75 \times \frac{2.5}{3.5} = 0.53$$

We know,

y on $x \longrightarrow \quad Y - \bar{Y} = b_{yx}(X - \bar{X})$

$X = 70$ (given)

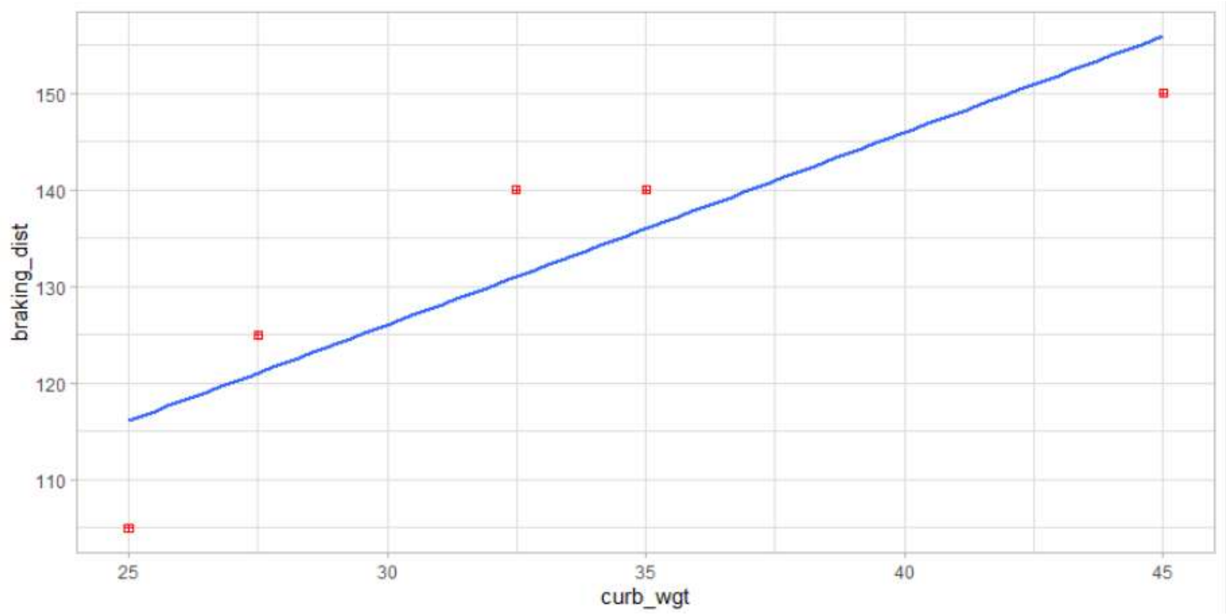
$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$
$$Y - 67 = 1.05(70 - 65)$$
$$Y - 67 = 1.05(5)$$
$$Y = 67 + 5.25 = 72.25$$

2.
```
curb_wgt = c(25,27.5,32.5,35,45)
braking_dist = c(105,125,140,140,150)
cor(curb_wgt, braking_dist)
ggplot()+aes(x=curb_wgt,y=braking_dist)+geom_point(shape=12,co
l="red")+geom_smooth(method=lm,se=FALSE,linetype="solid")+them
e_light()
```

OUTPUT:

```
> curb_wgt = c(25,27.5,32.5,35,45)
> braking_dist = c(105,125,140,140,150)
> cor(curb_wgt, braking_dist)
[1] 0.8834641
> ggplot()+aes(x=curb_wgt,y=braking_dist)+geom_point(shape=12,col="red")+geom_smooth(method=lm,se=FALSE,linety
pe="solid")+theme_light()
`geom_smooth()` using formula = 'y ~ x'
> 
```

PLOT:



3. head(Auto)
   data=lm(Auto$mpg~Auto$cylinder+Auto$displacement+Auto$horsepow
   er+Auto$weight+Auto$acceleration+Auto$year+Auto$origin)
   summary(data)

OUTPUT:

```
> head(Auto)
  mpg cylinders displacement horsepower weight acceleration year origin                      name
1  18         8          307        130   3504         12.0   70      1 chevrolet chevelle malibu
2  15         8          350        165   3693         11.5   70      1         buick skylark 320
3  18         8          318        150   3436         11.0   70      1        plymouth satellite
4  16         8          304        150   3433         12.0   70      1             amc rebel sst
5  17         8          302        140   3449         10.5   70      1               ford torino
6  15         8          429        198   4341         10.0   70      1          ford galaxie 500
> data=lm(Auto$mpg~Auto$cylinder+Auto$displacement+Auto$horsepower+Auto$weight+Auto$acceleration+Aut
o$origin)
> summary(data)

Call:
lm(formula = Auto$mpg ~ Auto$cylinder + Auto$displacement + Auto$horsepower + 
    Auto$weight + Auto$acceleration + Auto$year + Auto$origin)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.5903 -2.1565 -0.1169  1.8690 13.0604 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -17.218435   4.644294  -3.707  0.00024 ***
Auto$cylinder      -0.493376   0.323282  -1.526  0.12780    
Auto$displacement   0.019896   0.007515   2.647  0.00844 ** 
Auto$horsepower    -0.016951   0.013787  -1.230  0.21963    
Auto$weight        -0.006474   0.000652  -9.929  < 2e-16 ***
Auto$acceleration   0.080576   0.098845   0.815  0.41548    
Auto$year           0.750773   0.050973  14.729  < 2e-16 ***
Auto$origin         1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,	Adjusted R-squared:  0.8182 
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

4. head(Carseats)
   data =
   lm(Carseats$Sales~Carseats$Price+Carseats$Income+Carseats$Adve
   rtising)
   summary(data)

OUTPUT:

```
> head(Carseats)
  Sales CompPrice Income Advertising Population Price ShelveLoc Age Education Urban  US
1  9.50       138     73          11        276   120       Bad  42        17   Yes Yes
2 11.22       111     48          16        260    83      Good  65        10   Yes Yes
3 10.06       113     35          10        269    80    Medium  59        12   Yes Yes
4  7.40       117    100           4        466    97    Medium  55        14   Yes Yes
5  4.15       141     64           3        340   128       Bad  38        13   Yes  No
6 10.81       124    113          13        501    72       Bad  78        16    No Yes
> data = lm(Carseats$Sales~Carseats$Price+Carseats$Income+Carseats$Advertising)
> summary(data)

Call:
lm(formula = Carseats$Sales ~ Carseats$Price + Carseats$Income + 
    Carseats$Advertising)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.4568 -1.5420 -0.0753  1.4975  6.7877 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          12.172701   0.682736  17.829  < 2e-16 ***
Carseats$Price       -0.053836   0.005051 -10.658  < 2e-16 ***
Carseats$Income       0.011066   0.004276   2.588     0.01 *  
Carseats$Advertising  0.120237   0.017985   6.685 7.85e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.382 on 396 degrees of freedom
Multiple R-squared:  0.2938,	Adjusted R-squared:  0.2884 
F-statistic: 54.91 on 3 and 396 DF,  p-value: < 2.2e-16
```

```
5. A = matrix(data=c(1,6,3,2),ncol = 2,byrow = TRUE)
   B = matrix(data=c(6,10),ncol = 1)
   solve(A,B)
   byx = -1/6
   bxy = -2/3
   r = sqrt(byx*bxy)
   print(r)
```

OUTPUT:

```
> A = matrix(data=c(1,6,3,2),ncol = 2,byrow = TRUE)
> B = matrix(data=c(6,10),ncol = 1)
> solve(A,B)
     [,1]
[1,]  3.0
[2,]  0.5
> byx = -1/6
> bxy = -2/3
> r = sqrt(byx*bxy)
> print(r)
[1] 0.3333333
> 
```

Q5:

$x + 6y = 6 \leftarrow y$ on x

$3x + 2y = 10 \leftarrow x$ on y

$6y = 6 - x$

$y = 1 - \frac{1}{6}x$

$b_{yx} = -\frac{1}{6}$

$3x + 2y = 10$

$3x = 10 - 2y$

$x = \frac{10}{3} - \frac{2}{3}y$

$b_{xy} = -\frac{2}{3}$

$r^2 = b_{yx} \cdot b_{xy}$

$= -\frac{1}{6} \times -\frac{2}{3} = \frac{1}{9}$ $\Rightarrow r = \frac{1}{3}$

coefficient of correlation, $r = \frac{1}{3}$