

ASSIGNMENT-4

NAME: M Gyanada Chowdary

REG.NO: 21BCE7727

1. Compute and interpret separately with Karl Pearson and Spearman correlation coefficient for the ing advertising cost and product sales (in lakhs) of 8 companies are selected at random:

Advertising cost	15	20	28	12	80	20	60	40
Product sales	40	30	50	30	60	30	10	20

CODE:

```
#21BCE7727
```

```
#M Gyanada Chowdary
```

```
#x=advertising cost
```

```
#y=Product sales
```

```
x<- c(15, 20, 28, 12, 80, 20, 60, 40)
```

```
y <- c(40, 30, 50, 30, 60, 30, 10, 20)
```

```
xy<-x*y
```

```
#Pearson correlation coefficient
```

```
meanx <- mean(x)
```

```
meany <- mean(y)
```

```
meanxy<-mean(xy)
```

```
sdx <- sd(x)
```

```
sd y <- sd(y)
```

```
covxy <- mean(xy)-(meanx*meany)
```

```
coxy
```

```
r_pearson <- covxy/ (sdx * sdy)
```

```
r_pearson
```

```
#Spearman correlation coefficient
```

```
rankx <- rank(x)
```

```
ranky <- rank(y)
```

```
d <- rankx -ranky
```

```
n <- length(x)
```

```
s=(6 * sum(d^2)) / (n * (n^2 - 1))
```

```
r_spearman <- 1-s
```

```
r_spearman
```

```
df=data.frame(c(x,meanx),c(y,meany),c(xy,meanxy))
```

```
df
```

```
df2=data.frame(x,y,rankx,ranky,d)
```

```
df2
```

```
plot(x,y)
```

```
abline(lm(y ~ x, data = df), col = "red")
```

```

1 #21BCE7727
2 #M Gyanada Chowdary
3 #x=advertising cost
4 #y=Product sales
5 x<- c(15, 20, 28, 12, 80, 20, 60, 40)
6 y <- c(40, 30, 50, 30, 60, 30, 10, 20)
7 xy<-x*y
8 #Pearson correlation coefficient
9 meanx <- mean(x)
10 meany <- mean(y)
11 meanxy<-mean(xy)
12 sdx <- sd(x)
13 sdy <- sd(y)
14 covxy <- mean(xy)-(meanx*meany)
15 coxy
16 r_pearson <- covxy/ (sdx * sdy)
17 r_pearson
18 #Spearman correlation coefficient
19 rankx <- rank(x)
20 ranky <- rank(y)
21 d <- rankx -ranky
22 n <- length(x)
23 s=(6 * sum(d^2)) / (n * (n^2 - 1))
24 r_spearman <- 1-s
25 r_spearman
26 df=data.frame(c(x,meanx),c(y,meany),c(xy,meanxy))
27 df
28 df2=data.frame(x,y,rankx,ranky,d)
29 df2
30 plot(x,y)
31 abline(lm(y ~ x, data = df), col = "red")

```

Output:

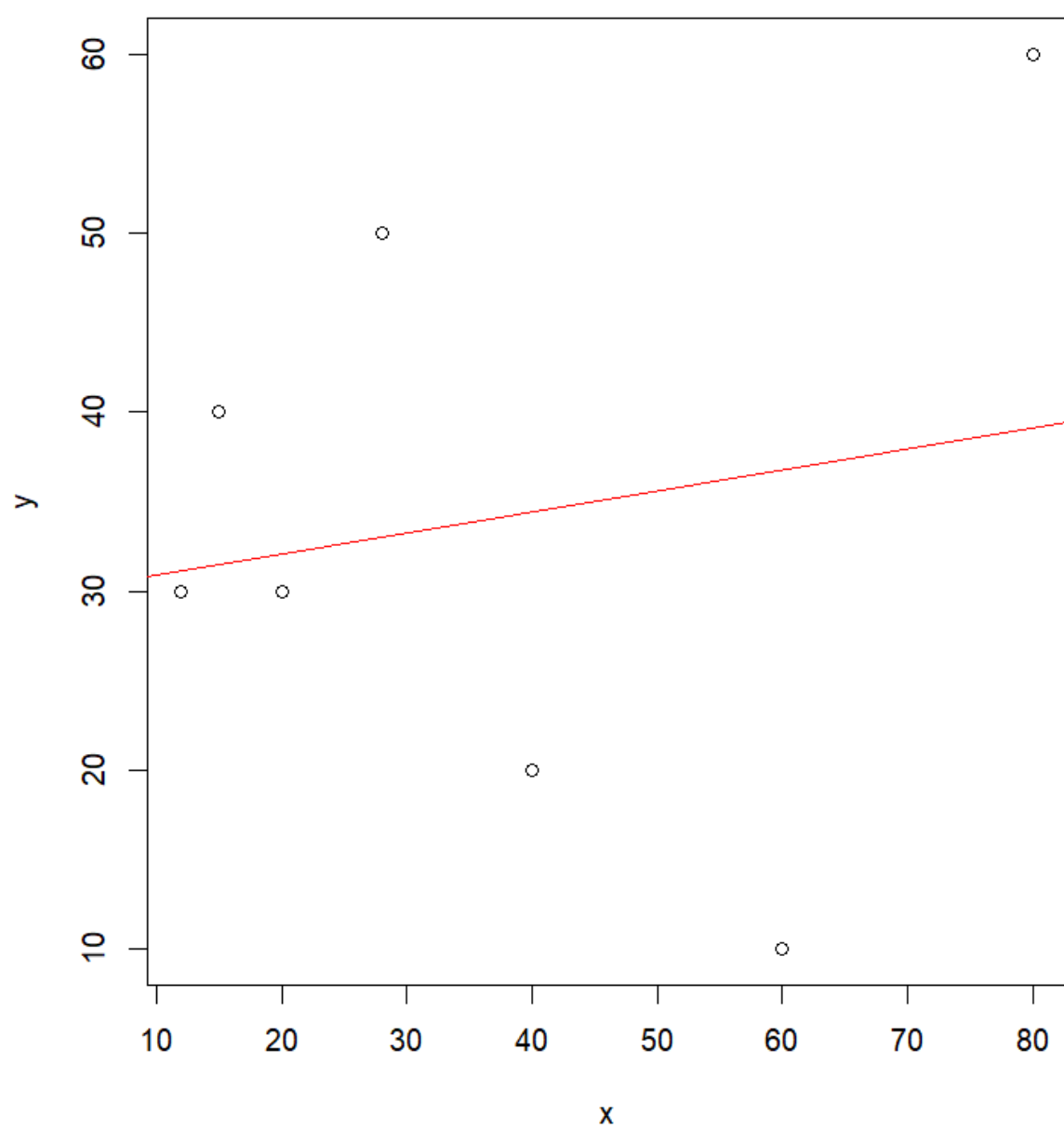
```

> #21BCE7727
> #M Gyanada Chowdary
> #x=advertising cost
> #y=Product sales
> x<- c(15, 20, 28, 12, 80, 20, 60, 40)
> y <- c(40, 30, 50, 30, 60, 30, 10, 20)
> xy<-x*y
> #Pearson correlation coefficient
> meanx <- mean(x)
> meany <- mean(y)
> meanxy<-mean(xy)
> sdx <- sd(x)
> sdy <- sd(y)
> covxy <- mean(xy)-(meanx*meany)
> covxy
[1] 59.84375
> r_pearson <- covxy/ (sdx * sdy)
> r_pearson
[1] 0.1547421
> #Spearman correlation coefficient
> rankx <- rank(x)
> ranky <- rank(y)
> d <- rankx -ranky
> n <- length(x)
> s=(6 * sum(d^2)) / (n * (n^2 - 1))
> r_spearman <- 1-s
> r_spearman
[1] 0.0297619
> df=data.frame(c(x,meanx),c(y,meany),c(xy,meanxy))
> df
   c.x..meanx.  c.y..meany.  c.xy..meanxy.
1      15.000      40.00      600
2      20.000      30.00      600
3      28.000      50.00     1400
4      12.000      30.00      360
5      80.000      60.00     4800
6      20.000      30.00      600
7      60.000      10.00      600
8      40.000      20.00      800
9      34.375      33.75     1220

> df2=data.frame(x,y,rankx,ranky,d)
> df2
   x  y rankx ranky  d
1 15 40   2.0    6 -4.0
2 20 30   3.5    4 -0.5
3 28 50   5.0    7 -2.0
4 12 30   1.0    4 -3.0
5 80 60   8.0    8  0.0
6 20 30   3.5    4 -0.5
7 60 10   7.0    1  6.0
8 40 20   6.0    2  4.0
> plot(x,y)
> abline(lm(y ~ x, data = df), col = "red")
> |

```

GRAPH:



2. IC chips production (in thousands) and electricity consumption (in KW/day) for a particular week of the company are given as follows:

IC chips production	451	358	431	506	499	529	564
Electricity consumption	248	226	247	277	305	583	299

- (a) Estimate the linear regression line.
- (b) Estimate the quadratic regression curve.
- (c) Estimate the electricity consumption for the day that company produces 4,95,000 IC chips.
- (d) Draw the scattered plot with line and curve.

CODE:

```
x <- c(451, 358, 431, 506, 499, 529, 564)
y <- c(248, 226, 247, 277, 305, 583, 299)

# (a) Linear regression using formula
n <- length(x)
slope <- (n * sum(x*y) - sum(x) * sum(y)) / (n * sum(x^2) - sum(x)^2)
intercept <- mean(y) - slope * mean(x)
cat("Linear Regression Equation: y =", round(slope, 3), "* x +", round(intercept, 3),
"\n")

# (b) Quadratic regression using formula
x2 <- x^2
x3 <- x^3
x4 <- x^4
sx <- sum(x)
sx2 <- sum(x2)
sx3 <- sum(x3)
sx4 <- sum(x4)
sy <- sum(y)
```

```

sxy <- sum(x*y)
sx2y <- sum(x2*y)
a <- matrix(c(n, sx, sx2, sx, sx2, sx3, sx2, sx3, sx4), ncol = 3)
b <- c(sy, sxy, sx2y)
coefficients <- solve(a, b)

cat("Quadratic Regression Equation: y =", round(coefficients[3], 3), "* x^2 +",
round(coefficients[2], 3), "* x +", round(coefficients[1], 3), "\n")

# (c) Estimate electricity consumption for 495000 IC chips production
x_new <- 495
y_lin <- slope * x_new + intercept
y_quad <- coefficients[3] * x_new^2 + coefficients[2] * x_new + coefficients[1]
cat("Electricity consumption for 495000 IC chips production (linear regression):",
round(y_lin, 3), "\n")
cat("Electricity consumption for 495000 IC chips production (quadratic regression):",
round(y_quad, 3), "\n")

# (d) Scatter plot with linear regression line and quadratic regression curve
plot(x, y)
abline(slope, intercept, col = "red")
curve(coefficients[3] * x^2 + coefficients[2] * x + coefficients[1], add = TRUE, col =
"blue")

quad_coef <- lm(y ~ poly(x, 2, raw=TRUE))$coefficients
a <- quad_coef[1]
b <- quad_coef[2]
c <- quad_coef[3]

# Prediction for 495000 IC chips production
x_pred <- 495
y_pred_linear <- slope * x_pred + intercept

```

```

y_pred_quad <- a * x_pred^2 + b * x_pred + c
abline(intercept, slope, col="red")
curve(a*x^2 + b*x + c, col="green", add=TRUE)

```

```

1 x <- c(451, 358, 431, 506, 499, 529, 564)
2 y <- c(248, 226, 247, 277, 305, 583, 299)
3
4 # (a) Linear regression using formula
5 n <- length(x)
6 slope <- (n * sum(x*y) - sum(x) * sum(y)) / (n * sum(x^2) - sum(x)^2)
7 intercept <- mean(y) - slope * mean(x)
8 cat("Linear Regression Equation: y =", round(slope, 3), "* x +", round(intercept, 3), "\n")
9
10 # (b) Quadratic regression using formula
11 x2 <- x^2
12 x3 <- x^3
13 x4 <- x^4
14 sx <- sum(x)
15 sx2 <- sum(x2)
16 sx3 <- sum(x3)
17 sx4 <- sum(x4)
18 sy <- sum(y)
19 sxy <- sum(x*y)
20 sx2y <- sum(x2*y)
21 a <- matrix(c(n, sx, sx2, sx, sx2, sx3, sx2, sx3, sx4), ncol = 3)
22 b <- c(sy, sxy, sx2y)
23 coefficients <- solve(a, b)
24 cat("Quadratic Regression Equation: y =", round(coefficients[3], 3), "* x^2 +", round(coefficients[2], 3), "* x +", round(coefficients[1], 3), "\n")
25
26 # (c) Estimate electricity consumption for 495000 IC chips production
27 x_new <- 495
28 y_lin <- slope * x_new + intercept
29 y_quad <- coefficients[3] * x_new^2 + coefficients[2] * x_new + coefficients[1]
30 cat("Electricity consumption for 495000 IC chips production (linear regression):", round(y_lin, 3), "\n")
31 cat("Electricity consumption for 495000 IC chips production (quadratic regression):", round(y_quad, 3), "\n")
32
33 # (d) Scatter plot with linear regression line and quadratic regression curve
34 plot(x, y)
35 abline(slope, intercept, col = "red")
36 curve(coefficients[3] * x^2 + coefficients[2] * x + coefficients[1], add = TRUE, col = "blue")
37
38 quad_coef <- lm(y ~ poly(x, 2, raw=TRUE))$coefficients
39 a <- quad_coef[1]
40 b <- quad_coef[2]
41 c <- quad_coef[3]
42
43 # Prediction for 495000 IC chips production
44 x_pred <- 495
45 y_pred_linear <- slope * x_pred + intercept
46 y_pred_quad <- a * x_pred^2 + b * x_pred + c
47 abline(intercept, slope, col="red")
48 curve(a*x^2 + b*x + c, col="blue", add=TRUE)

```

OUTPUT:

R 4.2.2 . ~/

```
> x <- c(451, 358, 431, 506, 499, 529, 564)
> y <- c(248, 226, 247, 277, 305, 583, 299)
>
> # (a) Linear regression using formula
> n <- length(x)
> slope <- (n * sum(x*y) - sum(x) * sum(y)) / (n * sum(x^2) - sum(x)^2)
> intercept <- mean(y) - slope * mean(x)
> cat("Linear Regression Equation: y =", round(slope, 3), "* x +", round(intercept, 3), "\n")
Linear Regression Equation: y = 0.934 * x + -133.47
>
> # (b) Quadratic regression using formula
> x2 <- x^2
> x3 <- x^3
> x4 <- x^4
> sx <- sum(x)
> sx2 <- sum(x2)
> sx3 <- sum(x3)
> sx4 <- sum(x4)
> sy <- sum(y)
> sxy <- sum(x*y)
> sx2y <- sum(x2*y)
> a <- matrix(c(n, sx, sx2, sx, sx2, sx3, sx2, sx3, sx4), ncol = 3)
> b <- c(sy, sxy, sx2y)
> coefficients <- solve(a, b)
> cat("Quadratic Regression Equation: y =", round(coefficients[3], 3), "* x^2 +", round(coefficients[2], 3), "* x +", round(coefficients[1], 3), "\n")
Quadratic Regression Equation: y = 0.001 * x^2 + 0.199 * x + 32.324
>
> # (c) Estimate electricity consumption for 495000 IC chips production
> x_new <- 495
> y_lin <- slope * x_new + intercept
> y_quad <- coefficients[3] * x_new^2 + coefficients[2] * x_new + coefficients[1]
> cat("Electricity consumption for 495000 IC chips production (linear regression):", round(y_lin, 3), "\n")
Electricity consumption for 495000 IC chips production (linear regression): 329.097
> cat("Electricity consumption for 495000 IC chips production (quadratic regression):", round(y_quad, 3), "\n")
Electricity consumption for 495000 IC chips production (quadratic regression): 326.584
>
> # (d) Scatter plot with linear regression line and quadratic regression curve
> plot(x, y)
> abline(slope, intercept, col = "red")
> curve(coefficients[3] * x^2 + coefficients[2] * x + coefficients[1], add = TRUE, col = "blue")
>
> quad_coef <- lm(y ~ poly(x, 2, raw=TRUE))$coefficients
> a <- quad_coef[1]
> b <- quad_coef[2]
> c <- quad_coef[3]
>
> # Prediction for 495000 IC chips production
> x_pred <- 495
> y_pred_linear <- slope * x_pred + intercept
> y_pred_quad <- a * x_pred^2 + b * x_pred + c
> abline(intercept, slope, col="red")
> curve(a*x^2 + b*x + c, col="blue", add=TRUE)
> |
```

GRAPH:

