# 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)
sdy <- sd(y)
covxy <- mean(xy)-(meanx*meany)
coxy
r_pearson <- covxy/ (sdx * sdy)
```

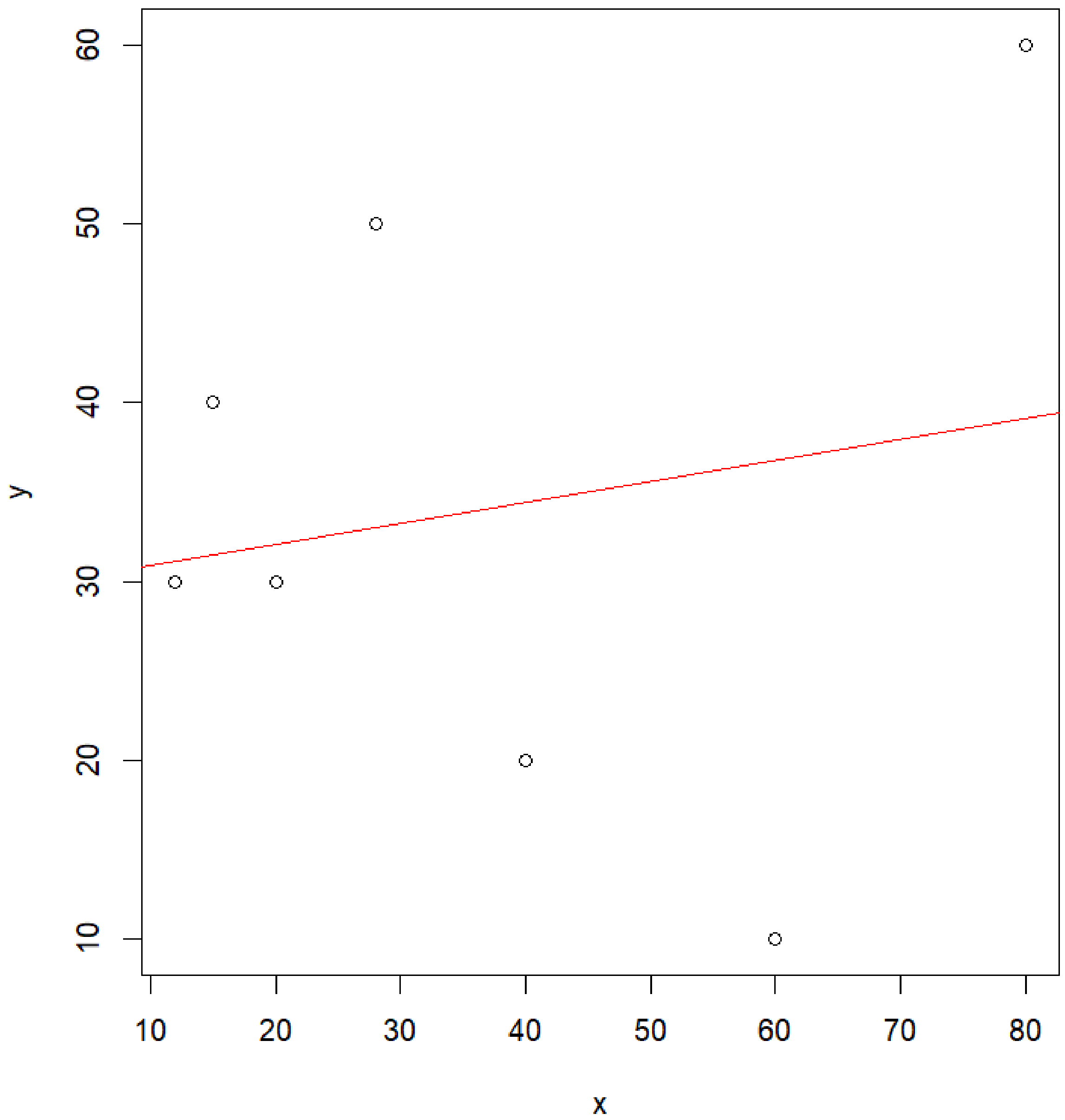
```
r_pearson
#Spearman correlation coefficient
rankx <- rank(x)
ranky <- rank(y)
d <- rankx -ranky
n <- length(x)
s=(6 * sum(d^2)) / (n * (n^2 - 1))
r_spearman <- 1-s
r_spearman
df=data.frame(c(x,meanx),c(y,meany),c(xy,meanxy))
df
df2=data.frame(x,y,rankx,ranky,d)
df2
plot(x,y)
abline(lm(y ~ x, data = df), col = "red")
```

```r
#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)
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")
```

**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:**



60
50
40
30
20
10
y
10
20
30
40
50
60
70
80
x

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)
```

```r
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)
```

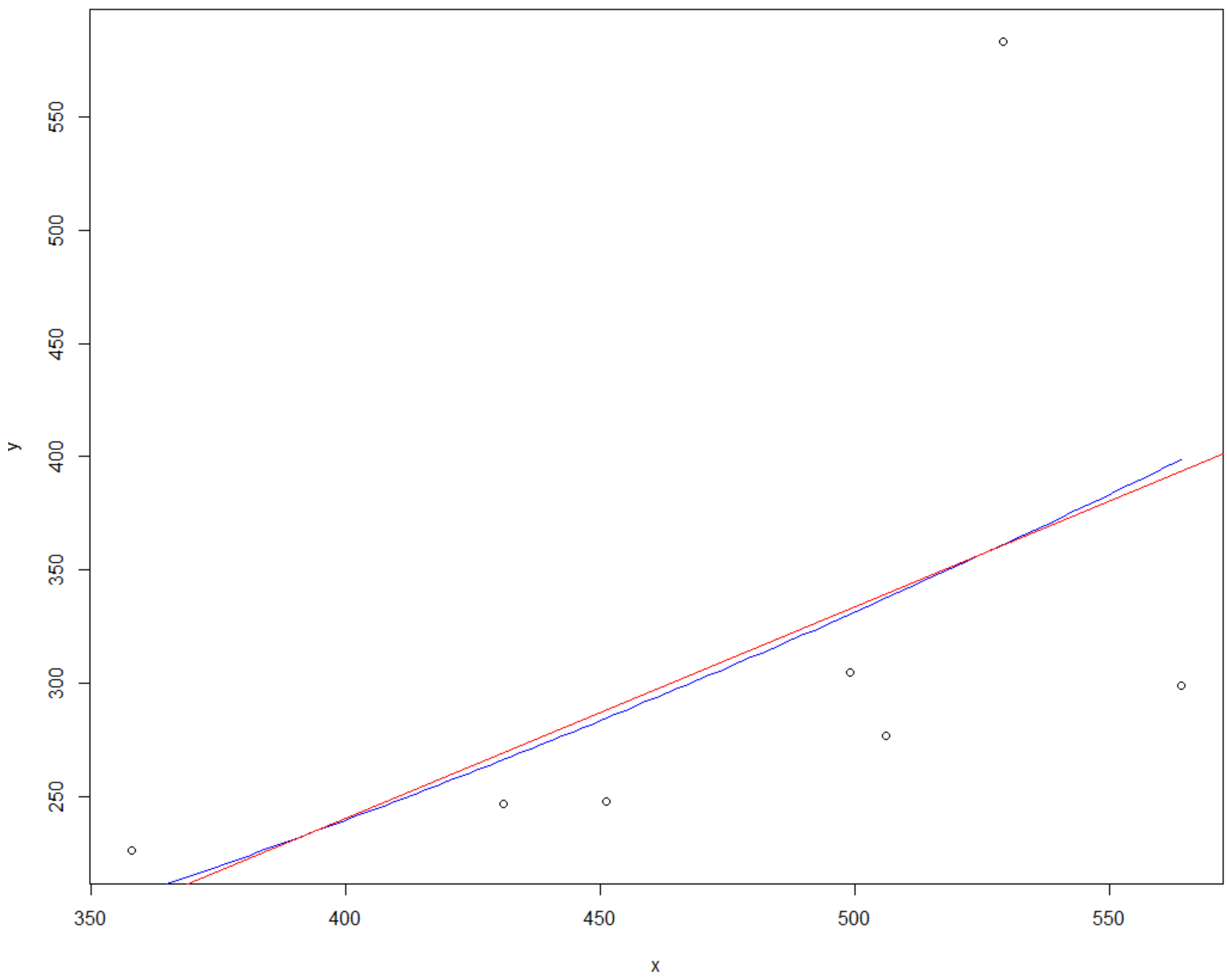
```r
x <- c(451, 358, 431, 506, 499, 529, 564)
y <- c(248, 226, 247, 277, 305, 583, 299)

# (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],

# (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="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)
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:



y
550
500
450
400
350
300
250
350
400
450
500
550
x