NPTEL DATA SCIENCE FOR ENGINEERS

ASSIGNMENT-6

1)

Solution:

$$r_{xy} = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\Sigma x_i^2 - n\bar{x}^2)}\sqrt{(\Sigma y_i^2 - n\bar{y}^2)}} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

```
> X=c(4.5,4.2,3.21,2.1,9.8)
> Y=c(2.3,2.1,1.5,1.0,6.8)
> correlation_using_function=cor(X,Y)
> correlation_using_formula=cov(X,Y)/(sd(X)*sd(Y))
> print(correlation_using_function)
[1] 0.9947401
> print(correlation_using_formula)
[1] 0.9947401
```

2) Solution:
From the given linear regression model, the value 1.98 represents the intercept

3) Solution:
Homoscedasticity in regression analysis is the condition in which the error variance remains the same

4) Solution:
For the best linear regression model, R^2 value should be equal to 1

5) Solution:

a. There is a strong evidence of a relationship between salary and years of experience

c. **The null hypothesis can be rejected**

6)

Solution:

```
> model<-lm(formula=auto$mpg~.,data=auto)
> summary(model)

Call:
lm(formula = auto$mpg ~ ., data = auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.2011 -1.9157 -0.0812  1.7341 15.0246

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 40.5619792  0.6461532   62.77   <2e-16 ***
weight      -0.0062905  0.0001984  -31.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.032 on 298 degrees of freedom
Multiple R-squared:  0.7714,    Adjusted R-squared:  0.7706
F-statistic:  1005 on 1 and 298 DF,  p-value: < 2.2e-16
```

7) Solution:

```
> model<-lm(formula=auto$mpg~.,data=auto)
> summary(model)

Call:
lm(formula = auto$mpg ~ ., data = auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.2011 -1.9157 -0.0812  1.7341 15.0246

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 40.5619792  0.6461532   62.77   <2e-16 ***
weight      -0.0062905  0.0001984  -31.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.032 on 298 degrees of freedom
Multiple R-squared:  0.7714,    Adjusted R-squared:  0.7706
F-statistic:  1005 on 1 and 298 DF,  p-value: < 2.2e-16
```

8) Solution:

```
Call:
lm(formula = auto$mpg ~ ., data = auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.2011 -1.9157 -0.0812  1.7341 15.0246

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 40.5619792  0.6461532   62.77   <2e-16 ***
weight      -0.0062905  0.0001984  -31.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.032 on 298 degrees of freedom
Multiple R-squared:  0.7714,    Adjusted R-squared:  0.7706
F-statistic:  1005 on 1 and 298 DF,  p-value: < 2.2e-16
```