Assingment 11

ABDUL RAUF

2023-01-19

Contents

* LAB - 11 *

```
__NAME__:-**ABDUL RAUF**
__ENRL NO__:-**GL6092**
__ROLL NO__:-**22DSMSA116**
```

1 Assingment 1

1.1 Read the data from excel

```r
advert<-read.csv("Advertising.csv")
head(advert)
```

```
##   X    TV radio newspaper sales
## 1 1 230.1  37.8      69.2  22.1
## 2 2  44.5  39.3      45.1  10.4
## 3 3  17.2  45.9      69.3   9.3
## 4 4 151.5  41.3      58.5  18.5
## 5 5 180.8  10.8      58.4  12.9
## 6 6   8.7  48.9      75.0   7.2
```

```r
names(advert)
```

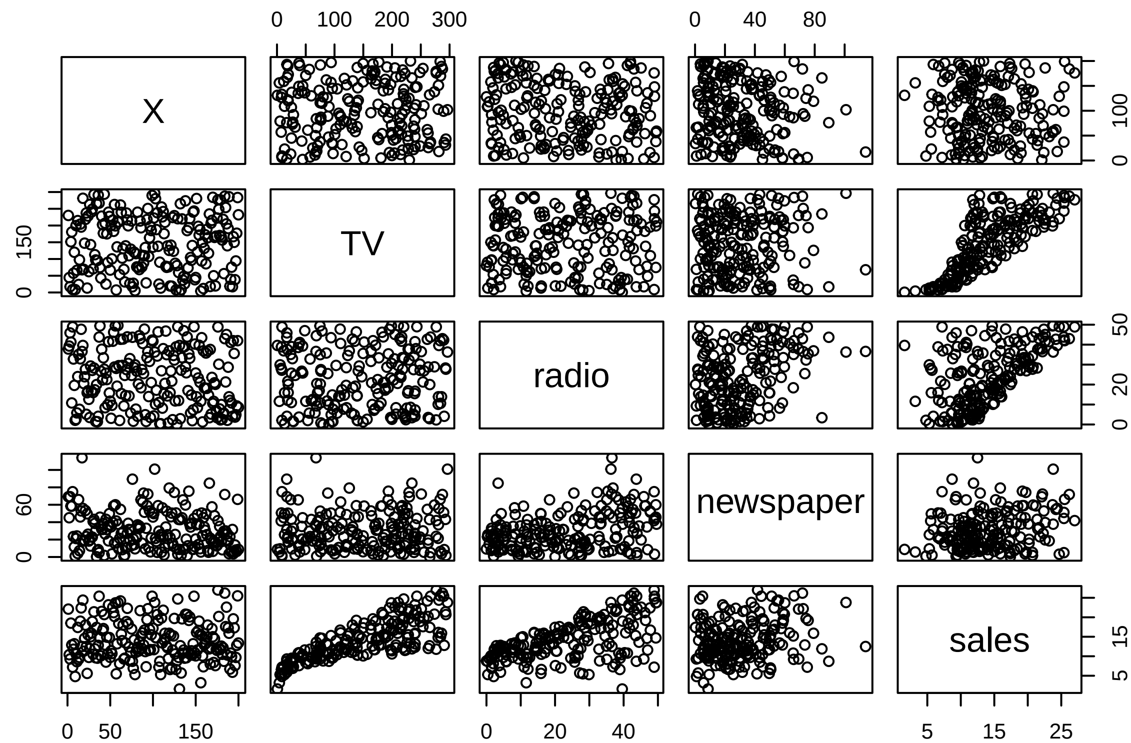
```
## [1] "X"         "TV"        "radio"     "newspaper"
## [5] "sales"
```

```r
attach(advert)
```

```
## The following objects are masked from advert (pos = 4):
##
##     newspaper, radio, sales, TV, X

## The following objects are masked from advert (pos = 7):
##
##     newspaper, radio, sales, TV, X
```

```r
pairs(advert)
```




```
dim(advert)
```

```
## [1] 200   5
```

1.2 Fitting the model

```
m1<-lm(sales~TV+radio+newspaper,data = advert)
m1
```

```
##
## Call:
## lm(formula = sales ~ TV + radio + newspaper, data = advert)
##
## Coefficients:
## (Intercept)           TV        radio    newspaper
##    2.938889     0.045765     0.188530    -0.001037
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = sales ~ TV + radio + newspaper, data = advert)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8277 -0.8908  0.2418  1.1893  2.8292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
## TV           0.045765   0.001395  32.809   <2e-16 ***
## radio        0.188530   0.008611  21.893   <2e-16 ***
## newspaper   -0.001037   0.005871  -0.177     0.86
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.686 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

```r
AIC(m1)
```

```
## [1] 782.3622
```

- Except Newspaper all the variables are significant as their p- values are less than **0.05** .
- 89.56% of the variability of sales is explained by the model.

```r
m2<-lm(sales~TV+radio,data = advert)
m2
```

```
##
## Call:
## lm(formula = sales ~ TV + radio, data = advert)
##
## Coefficients:
## (Intercept)           TV        radio
##     2.92110      0.04575      0.18799
```

```r
summary(m2)
```

```
##
## Call:
## lm(formula = sales ~ TV + radio, data = advert)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7977 -0.8752  0.2422  1.1708  2.8328
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.92110    0.29449   9.919   <2e-16 ***
## TV           0.04575    0.00139  32.909   <2e-16 ***
```

```
## radio        0.18799    0.00804  23.382   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.681 on 197 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8962
## F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

```
AIC(m2)
```

```
## [1] 780.3941
```

- 89.62% of the variability of sales is explained by the model.
- m2 has the lowest AIC value and in this model all the variables are significant.
- adjusted R sq and MSE of m2 have approximately same value as compared to the m1 so m2 is better as compared to model m1.

1.3 Bootstraping

```
library(boot)
set.seed(1)
advert.fn<-function(advert,index)
  coef(lm(sales~TV+radio,data=advert,subset=index))
advrt.obj<-boot(data = advert,statistic = advert.fn ,R=200)
advrt.obj
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = advert, statistic = advert.fn, R = 200)
##
##
## Bootstrap Statistics :
##       original        bias    std. error
## t1* 2.92109991 -4.812417e-03 0.310339672
## t2* 0.04575482  7.985856e-05 0.002003148
## t3* 0.18799423 -9.421676e-05 0.010484828
```

Estimates of parameters are as follows:-

- $SE(\hat{\beta}_0)$ = **0.3133033**
- $SE(\hat{\beta}_1)$ = **0.002003**
- $SE(\hat{\beta}_2)$ = **0.010484**

Estimates of the bais of the parameter are as follows:-

- $Bais(\hat{\beta}_0)$ = **-4.812417e-03**
- $Bais(\hat{\beta}_1)$ = **7.985856e-05**
- $Bais(\hat{\beta}_2)$ = **-9.421676e-05**

2 Assingment 2

2.1 Read the data from excel

```
sol<-read.csv("solubility.csv")
head(sol)
```

```
##   Observation.Number     y  X1  X2  X3
## 1                  1 0.222 7.3 0.0 0.0
## 2                  2 0.395 8.7 0.0 0.3
## 3                  3 0.422 8.8 0.7 1.0
## 4                  4 0.437 8.1 4.0 0.2
## 5                  5 0.428 9.0 0.5 1.0
## 6                  6 0.467 8.7 1.5 2.8
```

```
names(sol)
```

```
## [1] "Observation.Number" "y"
## [3] "X1"                 "X2"
## [5] "X3"
```

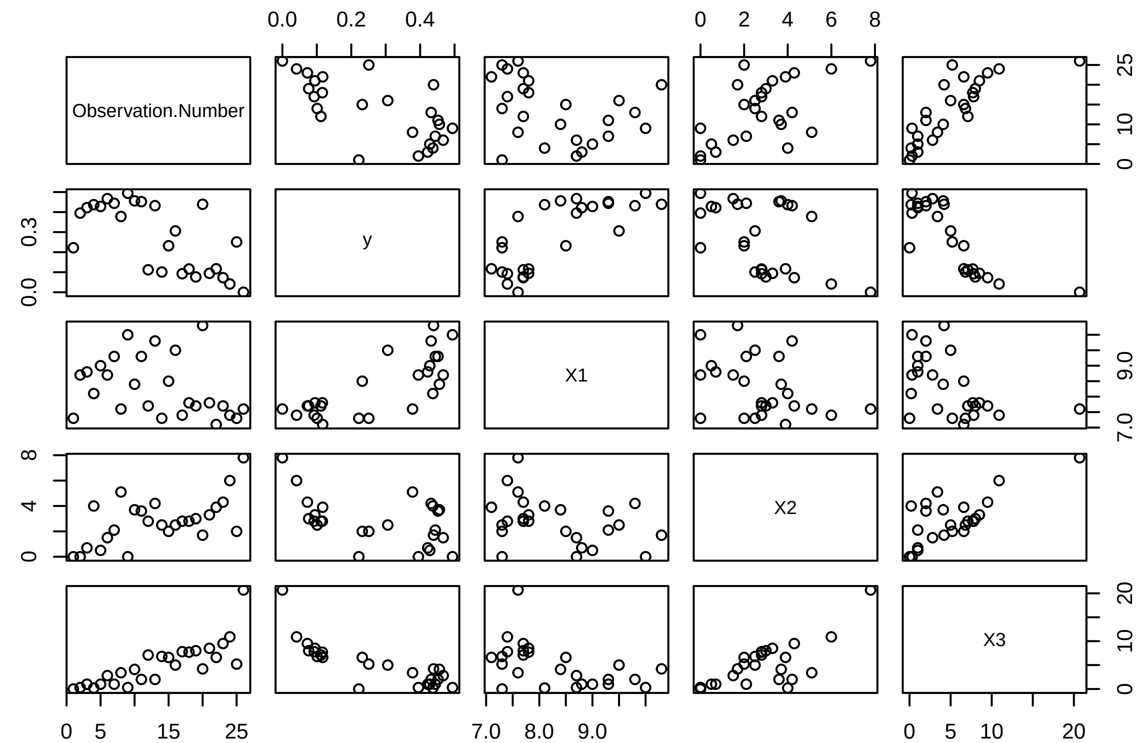
```
attach(sol)
```

```
## The following object is masked _by_ .GlobalEnv:
##
##     y
```

```
## The following objects are masked from sol (pos = 4):
##
##     Observation.Number, X1, X2, X3, y
```

```
## The following objects are masked from sol (pos = 6):
##
##     Observation.Number, X1, X2, X3, y
```

```
pairs(sol)
```




```
dim(sol)
```

```
## [1] 26  5
```

2.2 Fitting the model

```
M1<-lm(y~X1+X2+X3,data = sol)
M1
```

```
##
## Call:
## lm(formula = y ~ X1 + X2 + X3, data = sol)
##
## Coefficients:
## (Intercept)           X1           X2           X3
##    -0.36931      0.08651      0.02441     -0.02858
```

```
summary(M1)
```

```
##
## Call:
## lm(formula = y ~ X1 + X2 + X3, data = sol)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09187 -0.04660 -0.01395  0.03137  0.12707
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.369306   0.143646  -2.571   0.0174 *
## X1           0.086510   0.016003   5.406 1.98e-05 ***
## X2           0.024412   0.010226   2.387   0.0260 *
## X3          -0.028577   0.004477  -6.383 2.01e-06 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06581 on 22 degrees of freedom
## Multiple R-squared:  0.8667, Adjusted R-squared:  0.8486
## F-statistic: 47.69 on 3 and 22 DF,  p-value: 8.535e-10
```

```
AIC(M1)
```

```
## [1] -62.04607
```

- 84.86% of the variability of y is explained by the model. and all the variables are significant.

2.3 Bootstraping

```
library(boot)
set.seed(1)
sol.fn<-function(sol,index)
  coef(lm(y~X1+X2+X3,data=sol,subset=index))
sol.obj<-boot(data = sol,statistic = sol.fn ,R=200)
sol.obj
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = sol, statistic = sol.fn, R = 200)
##
##
## Bootstrap Statistics :
##        original        bias    std. error
## t1* -0.36930592  0.0401583977 0.187210926
## t2*  0.08651029 -0.0034272314 0.018998236
## t3*  0.02441173 -0.0005539604 0.010711207
## t4* -0.02857704 -0.0022745225 0.006702752
```

Estimates of parameters are as follows:-

- $SE(\hat{\beta}_0)$ = **0.187210926**
- $SE(\hat{\beta}_1)$ = **0.018998236**
- $SE(\hat{\beta}_2)$ = **0.010711207**
- $SE(\hat{\beta}_3)$ = **0.006702752**

Estimates of the bais of the parameter are as follows:-

- $Bais(\hat{\beta}_0)$ = **0.0401583977**
- $Bais(\hat{\beta}_1)$ = **-0.0034272314**
- $Bais(\hat{\beta}_2)$ = **-0.0005539604**
- $Bais(\hat{\beta}_2)$ = **-0.0022745225**