

Assignment 1

keerthi Tiyyagura

2023-09-24

#Setting up the ISLR Package #Giving the install.packages("ISLR") command in comments because it's averting me from knitting to pdf.

```
library(ISLR)
```

#Calling the ISLR library

```
summary(Carseats)
```

```
##      Sales      CompPrice      Income      Advertising
## Min.   : 0.000   Min.   : 77   Min.   : 21.00   Min.   : 0.000
## 1st Qu.: 5.390   1st Qu.:115   1st Qu.: 42.75   1st Qu.: 0.000
## Median : 7.490   Median :125   Median : 69.00   Median : 5.000
## Mean   : 7.496   Mean   :125   Mean   : 68.66   Mean   : 6.635
## 3rd Qu.: 9.320   3rd Qu.:135   3rd Qu.: 91.00   3rd Qu.:12.000
## Max.   :16.270   Max.   :175   Max.   :120.00   Max.   :29.000
##      Population      Price      ShelfLoc      Age      Education
## Min.   : 10.0   Min.   : 24.0   Bad   : 96   Min.   :25.00   Min.   :10.0
## 1st Qu.:139.0   1st Qu.:100.0   Good  : 85   1st Qu.:39.75   1st Qu.:12.0
## Median :272.0   Median :117.0   Medium:219   Median :54.50   Median :14.0
## Mean   :264.8   Mean   :115.8           Mean   :53.32   Mean   :13.9
## 3rd Qu.:398.5   3rd Qu.:131.0           3rd Qu.:66.00   3rd Qu.:16.0
## Max.   :509.0   Max.   :191.0           Max.   :80.00   Max.   :18.0
##      Urban      US
## No :118   No :142
## Yes:282   Yes:258
##
##
##
##
```

#Printing the summary of the Carseats dataset

```
View(Carseats)
```

#It displays the data frame in the viewer to visually inspect the dataset(just used this command for my clarification)

```
str(Carseats)
```

```
## 'data.frame': 400 obs. of 11 variables:
## $ Sales : num 9.5 11.22 10.06 7.4 4.15 ...
## $ CompPrice : num 138 111 113 117 141 124 115 136 132 132 ...
## $ Income : num 73 48 35 100 64 113 105 81 110 113 ...
## $ Advertising: num 11 16 10 4 3 13 0 15 0 0 ...
## $ Population : num 276 260 269 466 340 501 45 425 108 131 ...
## $ Price : num 120 83 80 97 128 72 108 120 124 124 ...
## $ ShelveLoc : Factor w/ 3 levels "Bad","Good","Medium": 1 2 3 3 1 1 3 2 3 3 ...
## $ Age : num 42 65 59 55 38 78 71 67 76 76 ...
## $ Education : num 17 10 12 14 13 16 15 10 10 17 ...
## $ Urban : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 2 2 1 1 ...
## $ US : Factor w/ 2 levels "No","Yes": 2 2 2 2 1 2 1 2 1 2 ...
```

```
#Printing the Structure of the Carseats dataset #This dataset contains 400 observations of 11 variables
```

```
head(Carseats,8)
```

```
## Sales CompPrice Income Advertising Population Price ShelveLoc Age Education
## 1 9.50 138 73 11 276 120 Bad 42 17
## 2 11.22 111 48 16 260 83 Good 65 10
## 3 10.06 113 35 10 269 80 Medium 59 12
## 4 7.40 117 100 4 466 97 Medium 55 14
## 5 4.15 141 64 3 340 128 Bad 38 13
## 6 10.81 124 113 13 501 72 Bad 78 16
## 7 6.63 115 105 0 45 108 Medium 71 15
## 8 11.85 136 81 15 425 120 Good 67 10
## Urban US
## 1 Yes Yes
## 2 Yes Yes
## 3 Yes Yes
## 4 Yes Yes
## 5 Yes No
## 6 No Yes
## 7 Yes No
## 8 Yes Yes
```

```
#Printing the first 8 rows of Carseats dataset
```

```
max(Carseats$Advertising)
```

```
## [1] 29
```

```
#Maximum value of Advertising attribute
```

```
IQR(Carseats$Price)
```

```
## [1] 31
```

```
#IQR(interquartile) of Price attribute
```

```
plot(Carseats$Sales,Carseats$Price,main = "Running price during the sales",xlab = "Total Sales",ylab="Total Price")
abline(lm(Carseats$Price~Carseats$Sales),col="Red",lwd=2)
```



#Scatter Plot of Sales against Price #This Plot shows the regression line of Sales vs Price and it has a negative slope ,implying a negative correlation ,and the value ranges from 0 to -1 #Also here we can observe that the regression line's quality of fit with points is good

```
correlation<-cor.test(Carseats$Sales,Carseats$Price,method="pearson")
correlation
```

```
##
## Pearson's product-moment correlation
##
## data: Carseats$Sales and Carseats$Price
## t = -9.912, df = 398, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.5203026 -0.3627240
## sample estimates:
## cor
## -0.4449507
```

#Calculating the correlation of two attributes #The correlation coefficient value is -0.444,it is indicating a negative direction.If one variable increases and other variable decreases the strength of the relationship will be moderate.