

# Setting Up R Assignment

Krishna Kumar Tavva - 811283461

2023-02-19

1. Install the ISLR library using the `install.packages()` command. Call the library using the `library(ISLR)` command to ensure that the library is correctly installed (10% of total points)

```
#install.packages("ISLR") #already installed the package  
library(ISLR) #calling the library command
```

2. Create a new R-Notebook (.Rmd) file. In the first code chunk, call the ISLR library and then print the summary of the Carseats dataset. How many observations (rows) this dataset contains?

```
library(ISLR) #calling the ISLR library  
summary(Carseats) #printing the carseats summary
```

```
##      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  
##  
##  
##  
##
```

```
nrow(Carseats) #number of observations (rows) in the dataset
```

```
## [1] 400
```

3. Using the summary statistics shown above, what is maximum value of the advertising attribute?

```
max(Carseats$Advertising) #maximum value of advertising attribute
```

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

4. Calculate the IQR of the Price attribute.

```
IQR(Carseats$Price) #finding Inter Quartile Range (IQR) of Price attribute
```

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

5. Plot the Sales against Price. What do you see in there? Calculate the correlation of the two attributes. What does the sign of the correlation coefficient suggest?

```
#install.packages("ggplot2") - If the package isn't installed.
```

```
library("ggplot2") #calling the library
```

```
#Using Quick Plot function to plot sales vs price relationship
```

```
qplot(Carseats$Sales, xlab="Sales", Carseats$Price, ylab="Price", main="Sales vs Price")
```

```
## Warning: 'qplot()' was deprecated in ggplot2 3.4.0.
```



```
cor.test(x = Carseats$Sales, y = Carseats$Price) #correlation coefficient default method=Pearson
```

```
##  
## 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
```

*#Key Findings from the Correlation value*

Sales and Price attributes are having a linear association.

There is a mild negative correlation between Sales and Price attributes and, therefore, the correlation value(r) is -0.4449507.

*#Brief about Pearson and Spearman methods*

Pearson's correlation coefficient is a statistics that measures the statistical relationship, or association, between two continuous variables. Pearson method most common method of measuring the association between variables of interest. If doesn't mention method type while calculating correlation, Pearson method is default.

Whereas, the Spearman's correlation assesses monotonic relationship (whether linear or not). Unlike Pearson, Spearman Rank can be applied to both ordinal (they are already ranked) and numerical (i.e. interval) variables. With numerical variables, we need to first sort them and use their ranking instead.

Like Pearson correlation, the Spearman correlation is symmetric.

Note: Though there is a association or relationship between two variables "Correlation Does not Imply Causation".