PS_IT24100543

Lab sheet 04

1. Import the dataset (’Exercise.txt’) into R and store it in a data frame called ”branch data”.

```
setwd("C:\\Users\\it24100543\\Desktop\\IT24100543_lab04")
branch_data<-read.table("Exercise.txt", header=TRUE, sep=",")
fix(data)
```

2. Identify the variable type and scale of measurement for each variable.

Variable Type :

```
> class(branch_data$Branch)
[1] "integer"
> class(branch_data$Sales_X1)
[1] "numeric"
> class(branch_data$Advertising_X2)
[1] "integer"
> class(branch_data$Years_X3)
[1] "integer"
```

Scale of measurement :

Branch- Ordinal

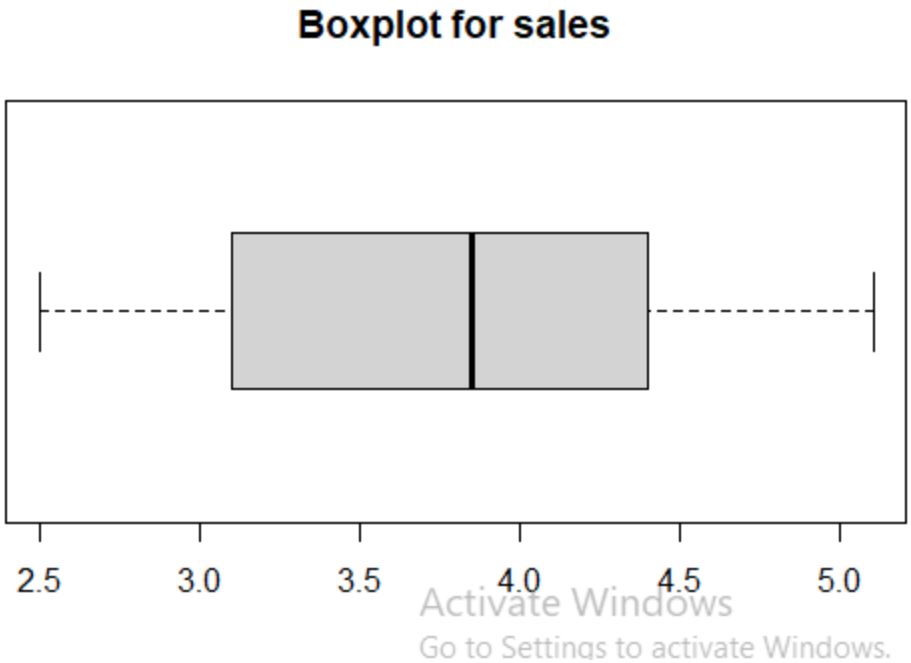
Sales- Nominal,

Advertising-Ordinal

Years-Ordinal

3. Obtain boxplot for sales and interpret the shape of the sales distribution.

```
boxplot(branch_data$Sales_X1,main="Boxplot for sales",outline= TRUE, oupatch=8,horizontal=TRUE)
```



Shape : Symmetrical

4. Calculate the five number summary and IQR for advertising variable.

```
> #q4-five number summary
> summary(branch_data$Branch)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    8.25   15.50   15.50   22.75   30.00
> summary(branch_data$Sales_X1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.500   3.125   3.850   3.790   4.375   5.100
> summary(branch_data$Advertising_X2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   101.2   132.5   134.8   158.8   210.0
> summary(branch_data$Years_X3)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    3.25    5.50    5.70    7.75   12.00
```

```
> #IQR
>
> IQR(branch_data$Branch)
[1] 14.5
> IQR(branch_data$Sales_X1)
[1] 1.25
> IQR(branch_data$Advertising_X2)
[1] 57.5
> IQR(branch_data$Years_X3)
[1] 4.5
```

5. Write an R function to find the outliers in a numeric vector and check for outliers in years variables.

```
#q5-outliers in a neumeric vector

get.outliers<- function(z){
  q1<-quantile(z)[2]
  q3<-quantile(z)[4]
  iqr<- q3-q1

  ub<-q3+1.5*iqr
  lb<-q1+1.5*iqr

  print(paste("Upper Bound = ", ub))
  print(paste("Lower Bound = ", lb))
  print(paste("Outliers: ", paste(sort(z[z<lb | z>ub]),collapse = ",")))
}
```

```
> get.outliers(branch_data$Sales_X1)
[1] "Upper Bound =  6.25"
[1] "Lower Bound =  5"
[1] "Outliers:  2.5,2.6,2.7,2.8,2.9,3,3,3.1,3.2,3.3,3.4,3.5,3.6,3.7,3.8,3.9,3.9,4,4.1,4.2,4.2,4.3,4.4,4.5,4.6,4.7,4.8,4.9"
> get.outliers(branch_data$Years_X3)
[1] "Upper Bound =  14.5"
[1] "Lower Bound =  10"
[1] "Outliers:  1,1,2,2,2,3,3,3,4,4,4,5,5,5,5,6,6,6,6,7,7,7,8,8,9,9"
```