Sri Lanka Institute of Information Technology

Lab Submission
Lab sheet No.4

IT24100861

Fernando M.S.F.

Discrete Mathematics | IT1160

B.Sc. (Hons) in Information Technology

Exercise

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

```
setwd("C:\\Users\\it24100861\\Desktop\\IT24100861")

--(Q1)--
branch_data <- read.table("Exercise.txt",header=TRUE,sep = ",")
```

```
> setwd("C:\\Users\\it24100861\\Desktop\\IT24100861")
> branch_data <- read.table("Exercise.txt",header=TRUE,sep = ",")
```

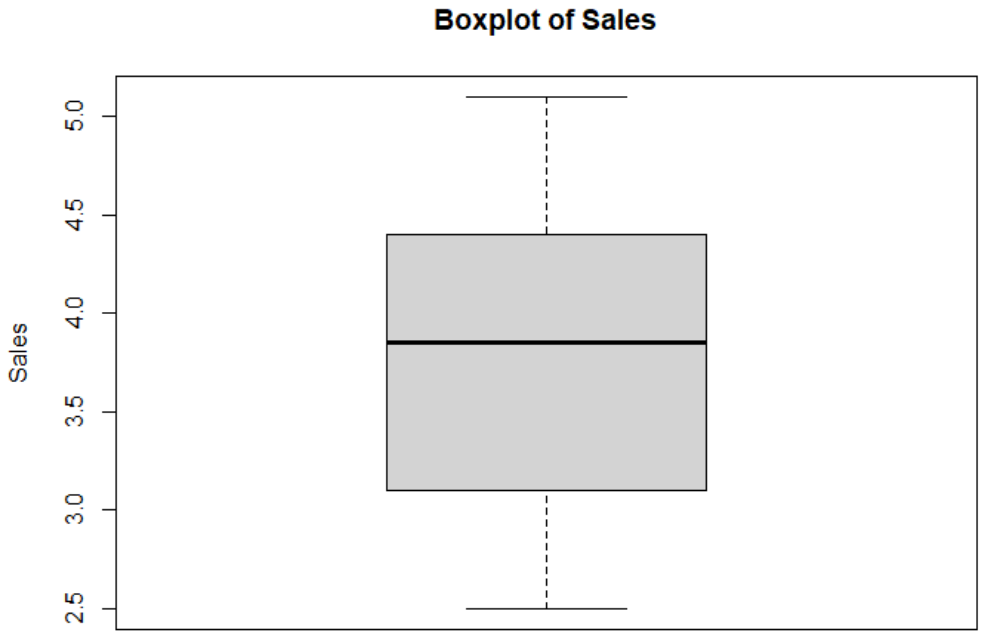
2. Identify the variable type and scale of measurement for each variable.

```
--(Q2)--
str(branch_data)
```

```
> str(branch_data)
'data.frame':   30 obs. of  4 variables:
 $ Branch        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Sales_X1      : num  3.4 4.1 2.8 5 3.7 4.5 3 4.9 3.2 2.5 ...
 $ Advertising_X2: int  120 150 90 200 110 175 95 185 105 80 ...
 $ Years_X3      : int  4 7 3 10 5 6 2 9 4 1 ...
```

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

```
--(Q3)--
boxplot(branch_data$Sales,
          main = "Boxplot of Sales",
          ylab = "Sales")
boxplot(branch_data$Sales,
          main = "Boxplot of Sales",
          ylab = "Sales")
```



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

```
--(Q4)--
summary(branch_data$Advertising_X2)

IQR(branch_data$Advertising_X2)
```

```
summary(branch_data$Advertising_X2)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 80.0   101.2   132.5   134.8   158.8   210.0
> IQR(branch_data$Advertising_X2)
[1] 57.5
```

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

```
> get.outliers <- function(x){
+     q1 <- quantile(x, 0.25)
+     q3 <- quantile(x, 0.75)
+     iqr <- q3 - q1
+
+     ub <- q3 + 1.5 * iqr
+     lb <- q1 - 1.5 * iqr
+
+     outliers <- x[x < lb | x > ub]
+
+     print(paste("Lower bound =", lb))
+     print(paste("Upper bound =", ub))
+     print(paste("Outliers:", if(length(outliers) == 0) "None" else paste(sort(outliers), collapse = ", ")))
+   }
>
>   get.outliers(branch_data$Years)
[1] "Lower bound = -3.5"
[1] "Upper bound = 14.5"
[1] "Outliers: None"
```

```
--(Q5)--
get.outliers <- function(x){
    q1 <- quantile(x, 0.25)
    q3 <- quantile(x, 0.75)
    iqr <- q3 - q1

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

    outliers <- x[x < lb | x > ub]

    print(paste("Lower bound =", lb))
    print(paste("Upper bound =", ub))
    print(paste("Outliers:", if(length(outliers) == 0) "None" else paste(sort(outliers), collapse = ", ")))
  }

  get.outliers(branch_data$Years)
```