SLIIT UNI
THE KNOWLEDGE UNIVERSITY

Faculty of Computing

Year 2 Semester 1 (2025)

IT2120 - Probability and Statistics Lab Sheet 04

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

```
# Set working directory (replace with your actual path)
setwd("C:\\Users\\IT24100320\\Desktop\\IT24100320")

# Import the dataset
branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
print(branch_data)
```

Filter

	Branch	Sales_X1	Advertising_X2	Years_X3
1	1	3.4	120	4
2	2	4.1	150	7
3	3	2.8	90	3
4	4	5.0	200	10
5	5	3.7	110	5
6	6	4.5	175	6
7	7	3.0	95	2
8	8	4.9	185	9
9	9	3.2	105	4
10	10	2.5	80	1
11	11	3.9	130	5

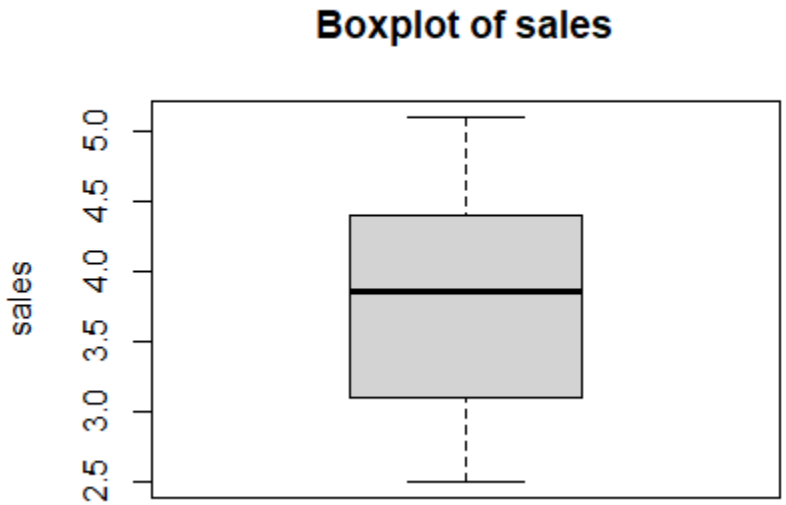
2. Identify the variable type and scale of measurement for each variable.

```
8  str(branch_data)
9  summary(branch_data)
38:1  (Top Level)
```

```
Console  Terminal  Background Jobs
R 4.2.2 · C:/Users/IT24100320/Desktop/IT24100320/
> 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 ...
> summary(branch_data)
     Branch         Sales_X1     Advertising_X2     Years_X3
 Min.   : 1.00   Min.   :2.500   Min.   : 80.0   Min.   : 1.00
 1st Qu.: 8.25   1st Qu.:3.125   1st Qu.:101.2   1st Qu.: 3.25
 Median :15.50   Median :3.850   Median :132.5   Median : 5.50
 Mean   :15.50   Mean   :3.790   Mean   :134.8   Mean   : 5.70
 3rd Qu.:22.75   3rd Qu.:4.375   3rd Qu.:158.8   3rd Qu.: 7.75
 Max.   :30.00   Max.   :5.100   Max.   :210.0   Max.   :12.00
```

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

```
11  boxplot(branch_data$Sales_X1, main = "Boxplot of sales", ylab = "sales")
12
```



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

```
13 five_num_summary <- fivenum(branch_data$Advertising_X2)
14 iqr_advertising <- IQR(branch_data$Advertising_X2)
15 print(five_num_summary)
16 print(iqr_advertising)
```

38:1 (Top Level)

Console Terminal Background Jobs

R 4.2.2 · C:/Users/IT24100320/Desktop/IT24100320/

```
> five_num_summary <- fivenum(branch_data$Advertising_X2)
> iqr_advertising <- IQR(branch_data$Advertising_X2)
> print(five_num_summary)
[1]  80.0 100.0 132.5 160.0 210.0
> print(iqr_advertising)
[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(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 = ",")))
}

# Check for outliers in the 'years' variable
get.outliers(Years_X3)
```

```
> # Check for outliers in the 'years' variable
> get.outliers(Years_X3)
[1] "Upper Bound =  14.5"
[1] "Lower Bound =  -3.5"
[1] "Outliers: "
```