IT2120 - Probability and Statistics

Lab Sheet 04

IT24104234 - Dhanujaya G.D.I.T

```
1  setwd("C:\\Users\\it24104234\\Desktop\\IT24104234")
2
3  # Import dataset
4  branch_data <- read.table("Exercise.txt", header = TRUE)
5
```

```
> setwd("C:\\Users\\it24104234\\Desktop\\IT24104234")
>
> # Import dataset
> branch_data <- read.table("Exercise.txt", header = TRUE)
>
```

Q1

```
##Question 01
branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")

fix(branch_data)

attach(branch_data)
```

Data Editor

File Edit Help

	Branch	Sales_X1	Advertising_X2	Years_X3	var5	var6	var7
1	1	3.4	120	4			
2	2	4.1	150	7			
3	3	2.8	90	3			
4	4	5	200	10			
5	5	3.7	110	5			
6	6	4.5	175	6			
7	7	3	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			
12	12	4.2	140	7			
13	13	2.7	100	3			
14	14	3.6	125	4			
15	15	4.8	190	8			
16	16	3.3	115	5			
17	17	4	135	6			
18	18	5.1	210	12			
19	19	3.8	145	6			

Q2

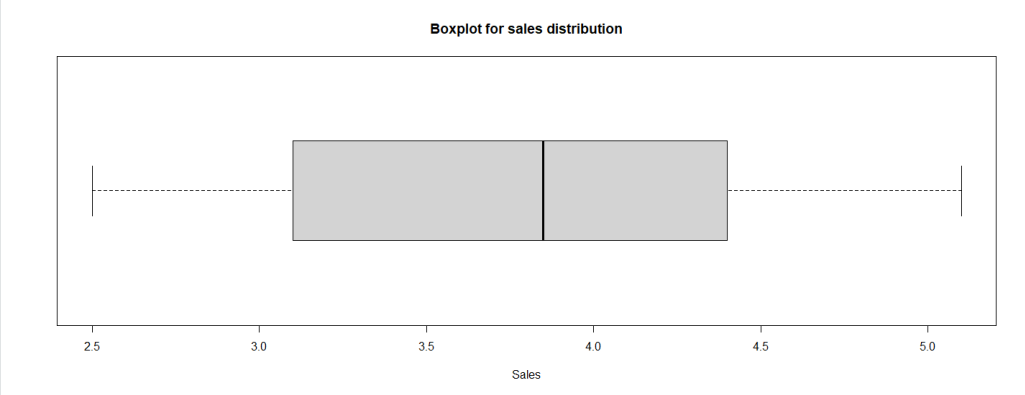
```
##Question 02
str(branch_data)

sapply(branch_data, class)
```

```
> attach(branch_data)
> str(branch_data)
'data.frame':   30 obs. of  4 variables:
 $ Branch        : num  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: num  120 150 90 200 110 175 95 185 105 80 ...
 $ Years_X3      : num  4 7 3 10 5 6 2 9 4 1 ...
>
> sapply(branch_data, class)
       Branch       Sales_X1 Advertising_X2       Years_X3
    "numeric"      "numeric"      "numeric"      "numeric"
>
```

Q3

```
##Question 03
boxplot(Sales_X1,main="Boxplot for sales distribution",xlab="Sales",outline=TRUE,outpch=8,horizontal=TRUE)
```



Q4

```
##Question 04
quantile(Advertising_X2)

IQR(Advertising_X2)
```

```
> ##Question 04
> quantile(Advertising_X2)
    0%    25%    50%    75%   100%
 80.00 101.25 132.50 158.75 210.00
>
> IQR(Advertising_X2)
[1] 57.5
>
```

Q5

```r
##Question 05s
find_outliers <- function(x) {
  Q1 <- quantile(x, 0.25)
  Q3 <- quantile(x, 0.75)
  IQR <- Q3 - Q1

  lower_bound <- Q1 - 1.5 * IQR
  upper_bound <- Q3 + 1.5 * IQR

  outliers <- x[x < lower_bound | x > upper_bound]

  return(outliers)
}

find_outliers(Years_X3)
```

```
> ##Question 05s
> find_outliers <- function(x) {
+   Q1 <- quantile(x, 0.25)
+   Q3 <- quantile(x, 0.75)
+   IQR <- Q3 - Q1
+
+   lower_bound <- Q1 - 1.5 * IQR
+   upper_bound <- Q3 + 1.5 * IQR
+
+   outliers <- x[x < lower_bound | x > upper_bound]
+
+   return(outliers)
+ }
>
> find_outliers(Years_X3)
numeric(0)
>
```