

Probability and Statistics – IT2120

IT24100886 – H.A.H.E. Wickramasinghe

Lab04

```
IT24100886_Lab04_Script.R x
1 setwd("C:\\Users\\IT24100886\\Desktop\\IT24100886")
2 getwd()
3
4 #Question 1
5 branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
6 attach(branch_data)
7
8 head(branch_data)
9
10 #Question 2
11 str(branch_data)
12
13
14 #Question 3
15 boxplot(branch_data$Sales_X1, main = "Boxplot of Sales", ylab = "Sales", col = "lightblue")
16
17
18 #Question 4
19 summary_stats <- summary(branch_data$Advertising_X2)
20 five_number <- c(
21   min = min(branch_data$Advertising_X2),
22   Q1 = quantile(branch_data$Advertising_X2, 0.25),
23   Median = median(branch_data$Advertising_X2),
24   Q3 = quantile(branch_data$Advertising_X2, 0.75),
25   Max = max(branch_data$Advertising_X2)
26 )
27
28 iqr_advertising <- IQR(branch_data$Advertising_X2)
29
30 print("Five-number summary for Advertising_X2:")
31 print(five_number)
32 print("IQR for Advertising_X2:")
33 print(iqr_advertising)
34
35
36 #Question 5
37 find_outliers <- function(x) {
38   Q1 <- quantile(x, 0.25)
39   Q3 <- quantile(x, 0.75)
40   IQR <- Q3 - Q1
41   lower_bound <- Q1 - 1.5 * IQR
42   upper_bound <- Q3 + 1.5 * IQR
43   outliers <- x[x < lower_bound | x > upper_bound]
44   if (length(outliers) == 0) {
45     return("No outliers found")
46   } else {
47     return(outliers)
48   }
49 }
50
51 outliers_years <- find_outliers(branch_data$Years_X3)
52 print("Outliers in Years_X3:")
53 print(outliers_years)
54
55 28:51 (Top Level) R Script
```

```

>
> head(branch_data)
  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
>
> #Question 2
> 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 ...
>
>
> #Question 3
> boxplot(branch_data$Sales_X1, main = "Boxplot of sales", ylab = "sales", col = "lightblue")
>
>
> #Question 4
> summary_stats <- summary(branch_data$Advertising_X2)
> five_number <- c(
+   Min = min(branch_data$Advertising_X2),
+   Q1 = quantile(branch_data$Advertising_X2, 0.25),
+   Median = median(branch_data$Advertising_X2),
+   Q3 = quantile(branch_data$Advertising_X2, 0.75),
+   Max = max(branch_data$Advertising_X2)
+ )
>
> iqr_advertising <- IQR(branch_data$Advertising_X2)
>
>
> print("Five-number summary for Advertising_X2:")
[1] "Five-number summary for Advertising_X2:"
> print(five_number)
   Min Q1.25% Median Q3.75%   Max
 80.00 101.25 132.50 158.75 210.00
> print("IQR for Advertising_X2:")
[1] "IQR for Advertising_X2:"
> print(iqr_advertising)
[1] 57.5
>

```

```

> #Question 5
> 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]
+   if (length(outliers) == 0) {
+     return("No outliers found")
+   } else {
+     return(outliers)
+   }
+ }
>
>
> outliers_years <- find_outliers(branch_data$Years_X3)
> print("Outliers in Years_X3:")
[1] "Outliers in Years_X3:"
> print(outliers_years)
[1] "No outliers found"
>

```

