Faculty of Computing

Year 2 Semester 1 (2025)

IT2120 - Probability and Statistics Lab Sheet 04

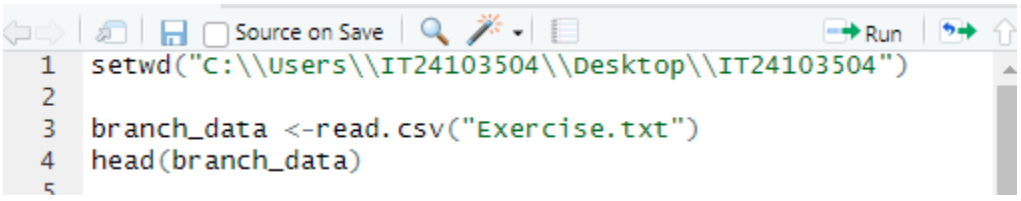
IT24103504 (Yunidu EDP)

Exercise

Instructions: Create a folder in your desktop with your registration number (Eg: "IT....."). You need to save the R script file and take screenshots of the command prompt with answers and save it in a word document inside the folder. Save both R script file and word document with your registration number (Eg: "IT....."). After you finish the exercise, zip the folder and upload the zip file to the submission link.

1. Import the dataset ('Exercise.txt') into R and store it in a data frame called "branch_data".
2. Identify the variable type and scale of measurement for each variable.
3. Obtain boxplot for sales and interpret the shape of the sales distribution.
4. Calculate the five number summary and IQR for advertising variable.
5. Write an R function to find the outliers in a numeric vector and check for outliers in years variables.

1)



2)

```
head(branch_data)

str(branch_data)
# Check column names to confirm that Sales_x1 exists
colnames(df)

# Use the correct column for summary and boxplot
summary(df[["Sales_x1"]])  # Using [[ to avoid errors wi
```

3)

```
# Boxplot with custom colors
boxplot(branch_data$Sales_x1,
        main = "Boxplot of Sales",
        ylab = "Sales",
        col = "lightblue",
        border = "darkblue")
```

4)

```
summary(branch_data$Advertising)

IQR_advertising <- IQR(branch_data$Advertising)
IQR_advertising
```

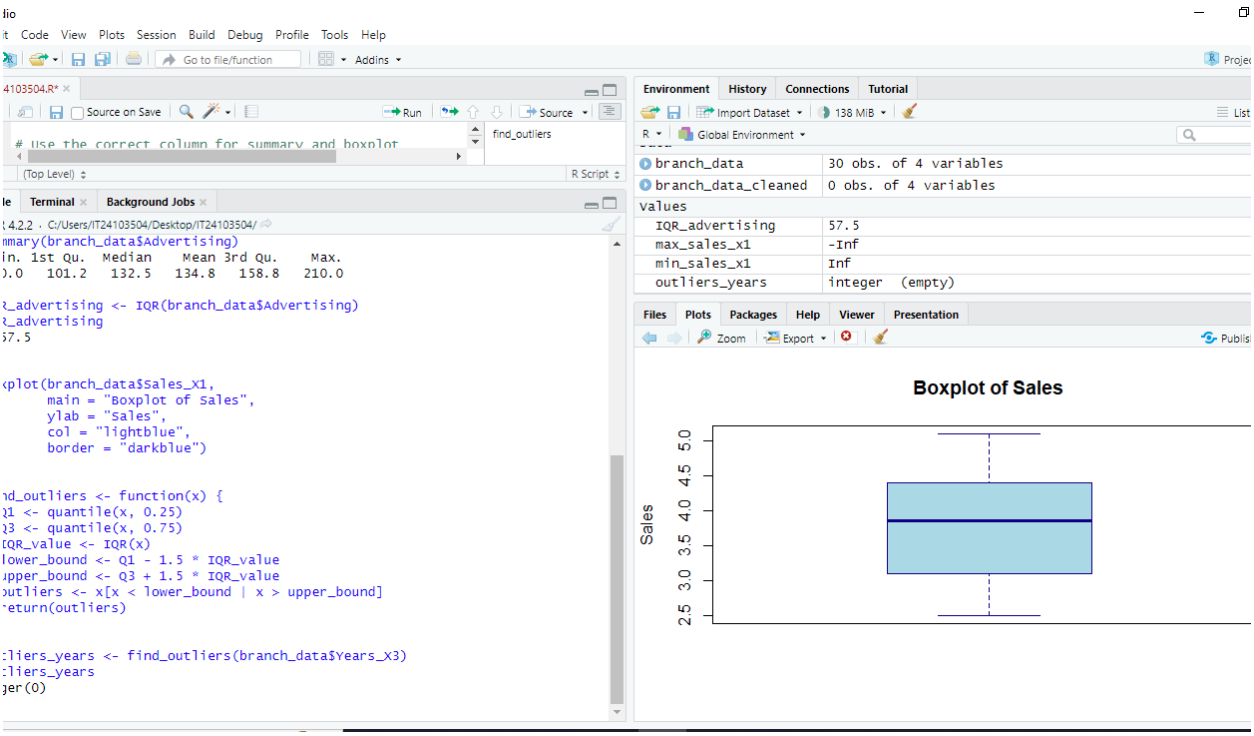
5)

```
boxplot(branch_data$Sales_X1,
        main = "Boxplot of Sales",
        ylab = "Sales",
        col = "lightblue",
        border = "darkblue")


find_outliers <- function(x) {
  Q1 <- quantile(x, 0.25)
  Q3 <- quantile(x, 0.75)
  IQR_value <- IQR(x)
  lower_bound <- Q1 - 1.5 * IQR_value
  upper_bound <- Q3 + 1.5 * IQR_value
  outliers <- x[x < lower_bound | x > upper_bound]
  return(outliers)
}

outliers_years <- find_outliers(branch_data$Years_X3)
outliers_years
```

7:53 (Top Level)

```
Console   Terminal ×   Background Jobs ×
R 4.2.2 · C:/Users/IT24103504/Desktop/IT24103504/
> setwd("C:\\Users\\IT24103504\\Desktop\\IT24103504")
>
> branch_data <-read.csv("Exercise.txt")
> 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
>
> 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 ...
> # Check column names to confirm that Sales_x1 exists
> colnames(df)
NULL
>
> # Use the correct column for summary and boxplot
> summary(df[["Sales_x1"]])  # Using [[ to avoid errors with non-standard names
```

```
> summary(branch_data$Advertising)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   101.2   132.5   134.8   158.8   210.0
>
> IQR_advertising <- IQR(branch_data$Advertising)
> IQR_advertising
[1] 57.5
>
>
> boxplot(branch_data$Sales_X1,
+         main = "Boxplot of Sales",
+         ylab = "Sales",
+         col = "lightblue",
+         border = "darkblue")
>
>
> find_outliers <- function(x) {
+   Q1 <- quantile(x, 0.25)
+   Q3 <- quantile(x, 0.75)
+   IQR_value <- IQR(x)
+   lower_bound <- Q1 - 1.5 * IQR_value
+   upper_bound <- Q3 + 1.5 * IQR_value
+   outliers <- x[x < lower_bound | x > upper_bound]
+   return(outliers)
+ }
>
> outliers_years <- find_outliers(branch_data$Years_X3)
> outliers_years
integer(0)
>
```


Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Source on Save
Run
Source
Use the correct column for summary and boxplot
find_outliers
(Top Level)
R Script
Terminal
Background Jobs
C:/Users/IT24103504/Desktop/IT24103504/
mmary(branch_data$Advertising)
in. 1st Qu. Median Mean 3rd Qu. Max.
0.0 101.2 132.5 134.8 158.8 210.0
_advertising <- IQR(branch_data$Advertising)
_advertising
57.5
plot(branch_data$Sales_X1,
main = "Boxplot of Sales",
ylab = "Sales",
col = "lightblue",
border = "darkblue")
d_outliers <- function(x) {
Q1 <- quantile(x, 0.25)
Q3 <- quantile(x, 0.75)
IQR_value <- IQR(x)
lower_bound <- Q1 - 1.5 * IQR_value
upper_bound <- Q3 + 1.5 * IQR_value
outliers <- x[x < lower_bound | x > upper_bound]
return(outliers)
liers_years <- find_outliers(branch_data$Years_X3)
liers_years
ger(0)
Environment
History
Connections
Tutorial
Import Dataset
138 MiB
Global Environment
branch_data 30 obs. of 4 variables
branch_data_cleaned 0 obs. of 4 variables
values
IQR_advertising 57.5
max_sales_x1 -Inf
min_sales_x1 Inf
outliers_years integer (empty)
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Boxplot of Sales
Sales
2.5
3.0
3.5
4.0
4.5
5.0