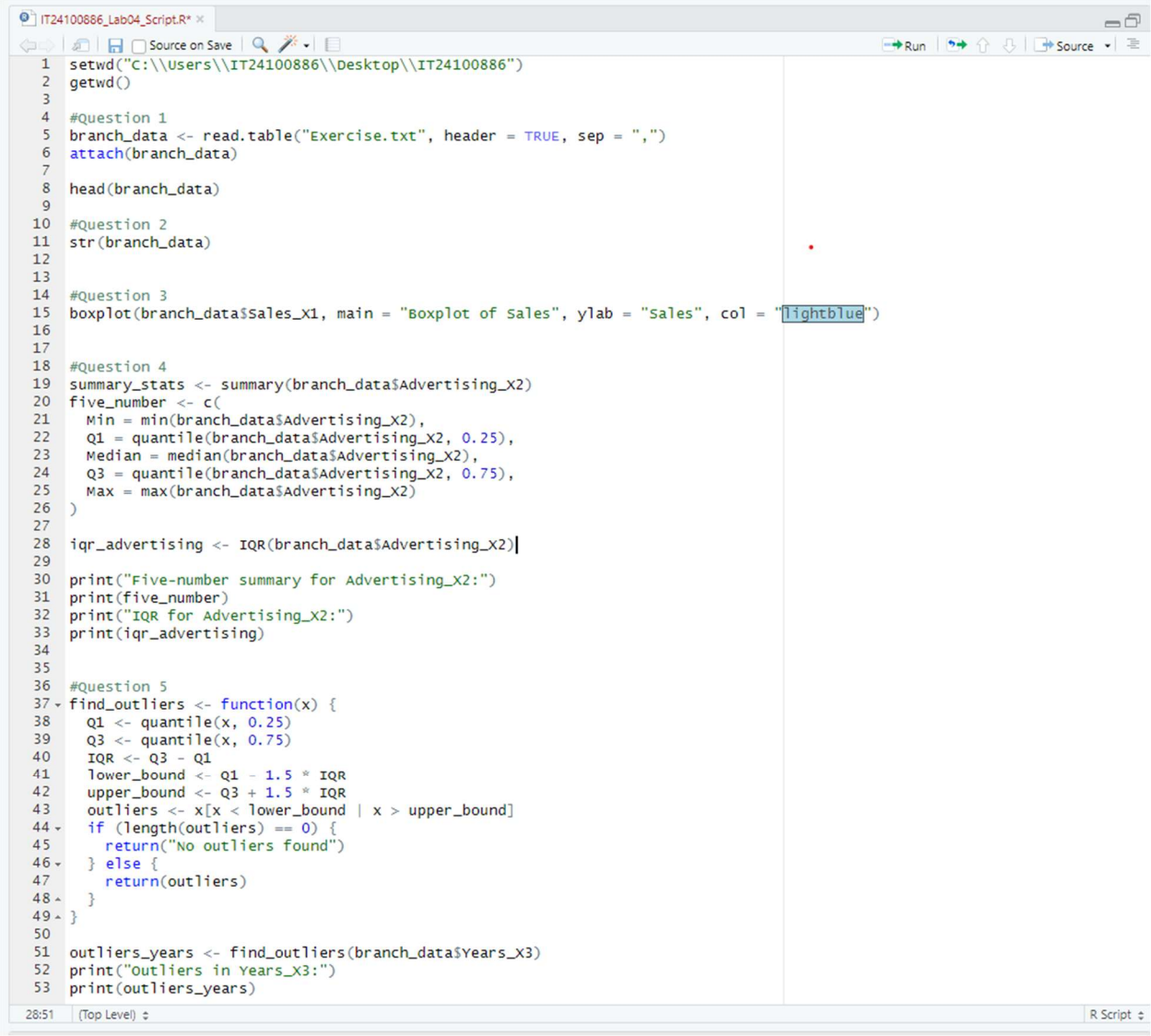
Probability and Statistics – IT2120

IT24100886 – H.A.H.E. Wickramasinghe

Lab04



```r
setwd("C:\\Users\\IT24100886\\Desktop\\IT24100886")
getwd()

#Question 1
branch_data <- read.table("Exercise.txt", header = TRUE, sep = ",")
attach(branch_data)

head(branch_data)

#Question 2
str(branch_data)


#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:")
print(five_number)
print("IQR for Advertising_X2:")
print(iqr_advertising)


#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:")
print(outliers_years)
```

```
>
> 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"
>
```