

Sri Lanka Institute of Information Technology



Lab Submission Lab sheet No 5

IT24101551

Kuyilini. T

IT2120 - Probability and Statistics

B.Sc. (Hons) in Information Technology

Exercise

1. Import the dataset ('Exercise – Lab 05.txt') into R and store it in a data frame called "Delivery Times".

The screenshot displays the R Studio interface. The script editor at the top contains the following R code:

```
1  
2 getwd()  
3 setwd("C:\\Users\\IT24101551\\Desktop\\IT24101551")  
4  
5  
6 # Question 01  
7 Delivery_Times<-read.table("Exercise - Lab 05.txt",header=TRUE)  
8 print(Delivery_Times)  
9 fix(Delivery_Times)  
10
```

The console window below the script editor shows the output of the code execution:

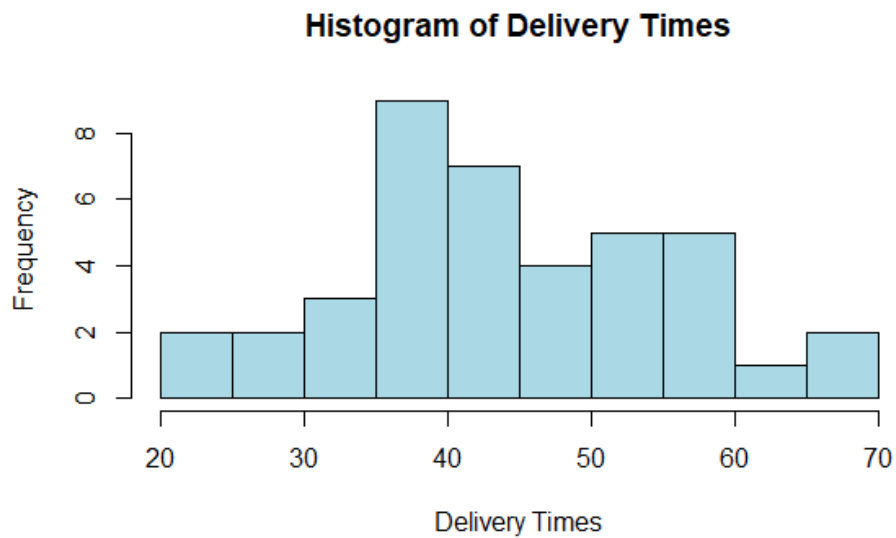
```
> # Question 01  
> getwd()  
[1] "C:/Users/IT24101551/Desktop/IT24101551"  
> setwd("C:\\Users\\IT24101551\\Desktop\\IT24101551")  
> # Question 02  
> Delivery_Times<-read.table("Exercise - Lab 05.txt",header=TRUE)  
> print(Delivery_Times)  
  Delivery_Time_.minutes.  
1                34  
2                54  
3                47  
4                29  
5                39  
6                61  
7                20  
8                40  
9                57  
10               36  
11               38  
12               44  
13               59  
14               38  
15               40  
16               40  
17               67  
18               66  
19               55  
20               48  
21               52
```

The Data Editor window at the bottom shows the data frame structure with columns: Delivery_Time_.minutes., var2, var3, var4, and var5. The data is displayed in a table format with 21 rows and 5 columns.

	Delivery_Time_.minutes.	var2	var3	var4	var5
1	34				
2	54				
3	47				
4	29				
5	39				
6	61				
7	20				
8	40				
9	57				
10	36				
11	38				
12	44				
13	59				
14	38				
15	40				
16	40				
17	67				
18	66				
19	55				

2. Draw a histogram for deliver times using nine class intervals where the lower limit is 20 and upper limit is 70. Use right open intervals.

```
IT24101551.R* x
Source on Save Run
10
11 # Question 02
12 hist(Delivery_Times$Delivery,
13       breaks = seq(20, 70, by = 5),
14       right = FALSE,
15       main = "Histogram of Delivery Times",
16       xlab = "Delivery Times",
17       ylab = "Frequency",
18       col = "lightblue",
19       border = "black")
20
21
22 hist_data <- hist(Delivery_Times$Delivery,
23                   breaks = seq(20, 70, by = 5),
24                   right = FALSE,
25                   plot = FALSE)
26
```



3. Comment on the shape of the distribution.

```
27 # Question 03
28 # The distribution appears to be slightly right-skewed with a peak around 35-40 minutes.
29
30:1 (Top Level) ⚙
```

```
Console Terminal × Background Jobs ×
R 4.2.2 · C:/Users/IT24101551/Desktop/IT24101551/ ↗
> # Question 03
> # The distribution appears to be slightly right-skewed with a peak around 35-40 minutes.
> |
```

4. Draw a cumulative frequency polygon (ogive) for the data in a separate plot.

```
31 # Question 04
32 cumulative_freq <- cumsum(hist_data$counts)
33
34
35 plot(hist_data$mids, cumulative_freq,
36       type = "o",
37       main = "Cumulative Frequency Polygon (Ogive)",
38       xlab = "Delivery Times",
39       ylab = "Cumulative Frequency",
40       pch = 16,
41       col = "blue")
42
43
```

```
42:1 (Top Level) ⚙
```

```
Console Terminal × Background Jobs ×
R 4.2.2 · C:/Users/IT24101551/Desktop/IT24101551/ ↗
> # Question 04
> cumulative_freq <- cumsum(hist_data$counts)
> plot(hist_data$mids, cumulative_freq,
+       type = "o",
+       main = "Cumulative Frequency Polygon (Ogive)",
+       xlab = "Delivery Times",
+       ylab = "Cumulative Frequency",
+       pch = 16,
+       col = "blue")
>
```

