

IT2120 - Probability and Statistics

Lab Sheet 05

IT24103526 - Senaratne P.A.R.T.

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 on the left contains the following code:

```
1 setwd("C:\\Users\\HP\\Desktop\\IT24103526")
2 getwd()
3
4 Delivery_Times <- read.table("Exercise - Lab 05.txt", header = TRUE, sep = ",")
5 fix(Delivery_Times)
6 head(Delivery_Times)
7 str(Delivery_Times)
8
```

The console on the bottom left shows the output of the commands:

```
R - R 4.5.1 - C:/Users/HP/Desktop/IT24103526/
> Delivery_Times <- read.table("Exercise - Lab 05.txt", header = TRUE, sep = ",")
> View(Delivery_Times)
> head(Delivery_Times)
Delivery_Time_.minutes.
1      34
2      54
3      47
4      29
5      39
6      61
> str(Delivery_Times)
'data.frame': 40 obs. of 1 variable:
 $ Delivery_Time_.minutes.: int 34 54 47 29 39 61 20 40 57 36 ...
> fix(Delivery_Times)
```

The Environment pane on the right shows the loaded data frame:

Object	Class	Attributes
branch_data	data.frame	30 obs. of 4 variables
Delivery_Times	data.frame	40 obs. of 1 variable
find_outliers	function	function (x)

The Data Editor window is open, showing the first 19 rows of the 'Delivery_Times' data frame. The columns are 'Delivery_Time_.minutes.', 'var2', 'var3', 'var4', and 'var5'. The data values for 'Delivery_Time_.minutes.' are: 34, 54, 47, 29, 39, 61, 20, 40, 57, 36, 11, 38, 12, 44, 13, 59, 14, 38, 15, 40, 16, 40, 17, 67, 18, 66, 19, 55.

- The screenshot displays the RStudio interface. The top-left pane shows the R script editor with the following code:

```
9  
10 breaks_seq <- seq(20, 70, length.out = 10) # Breaks: 20, 25.5556, ..., 70  
11 hist(Delivery_Times$Delivery_Time..minutes., breaks = breaks_seq, right = FALSE,  
12      main = "Histogram of Delivery Times", xlab = "Delivery Time (minutes)",  
13      ylab = "Frequency", col = "lightblue", border = "black")  
14 print(breaks_seq)  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30
```

The bottom-left pane shows the console output:

```
> breaks_seq <- seq(20, 70, length.out = 10) # Breaks: 20, 25.5556, ..., 70  
> hist(Delivery_Times$Delivery_Time..minutes., breaks = breaks_seq, right = FALSE,  
+      main = "Histogram of Delivery Times", xlab = "Delivery Time (minutes)",  
+      ylab = "Frequency", col = "lightblue", border = "black")  
> print(breaks_seq)  
[1] 20.00000 25.55556 31.11111 36.66667 42.22222 47.77778 53.33333 58.88889 64.44444 70.00000
```

The top-right pane shows the Environment window with the following variables:

Variable	Value
branch_data	30 obs. of 4 variables
Delivery_Times	40 obs. of 1 variable
freq_table	9 obs. of 2 variables
hist_info	List of 6
breaks_seq	num [1:10] 20 25.6 31.1 36.7 42.2 ...
find_outliers	function (x)

The bottom-right pane shows a histogram titled "Histogram of Delivery Times". The x-axis is labeled "Delivery Time (minutes)" and ranges from 20 to 70. The y-axis is labeled "Frequency" and ranges from 0 to 8. The histogram consists of 10 bars with a width of 5 units each. The bars are light blue with black borders. The frequencies for each bin are approximately: 20-25: 2, 25-30: 3, 30-35: 6, 35-40: 8, 40-45: 6, 45-50: 3, 50-55: 6, 55-60: 3, 60-65: 3, 65-70: 2.

```

32
33 # Optional: Get frequency table from hist for verification
34 hist_info <- hist(Delivery_Times$Delivery_Time_.minutes., breaks = breaks_seq, right = FALSE, plot = FALSE)
35 freq_table <- data.frame(Interval = paste("[" , round(breaks_seq[-length(breaks_seq)], 2), " , " , round(breaks_seq[-1], 2), " )", sep = "" ),
36                               Frequency = hist_info$counts)
37 print(freq_table)
38
39
40
41
42
43
44
45
46
47
48
49

```

32:1 (Top Level) ↕

Console Terminal Background Jobs

R • R 4.5.1 • C:/Users/HP/Desktop/IT24103526/ ↗

```

>
> # Optional: Get frequency table from hist for verification
> hist_info <- hist(Delivery_Times$Delivery_Time_.minutes., breaks = breaks_seq, right = FALSE, plot = FALSE)
> freq_table <- data.frame(Interval = paste("[" , round(breaks_seq[-length(breaks_seq)], 2), " , " , round(breaks_seq[-1], 2), " )", sep = "" ),
+                               Frequency = hist_info$counts)
> print(freq_table)

```

	Interval	Frequency
1	[20, 25.56)	2
2	[25.56, 31.11)	3
3	[31.11, 36.67)	6
4	[36.67, 42.22)	9
5	[42.22, 47.78)	6
6	[47.78, 53.33)	3
7	[53.33, 58.89)	6
8	[58.89, 64.44)	3
9	[64.44, 70)	2

3. Comment on the shape of the distribution.

The shape of the distribution is unimodal with a primary peak in the interval [36.67, 42.22) minutes, where 9 deliveries occur. It appears approximately symmetric around the center but with a slight right skew, as the tail extends to higher values (up to 67 minutes) and there is a secondary cluster or minor peak in the [53.33, 58.89) interval with 6 deliveries. Overall, most delivery times are concentrated between 30 and 50 minutes, indicating moderate variability without extreme outliers.

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



