Sri Lanka Institute of Information Technology


# Lab Submission
# <Lab Sheet 5 >

**<IT24104099>**

**<Liyanage R.J.P>**

**Probability and Statistics - IT2120**

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”

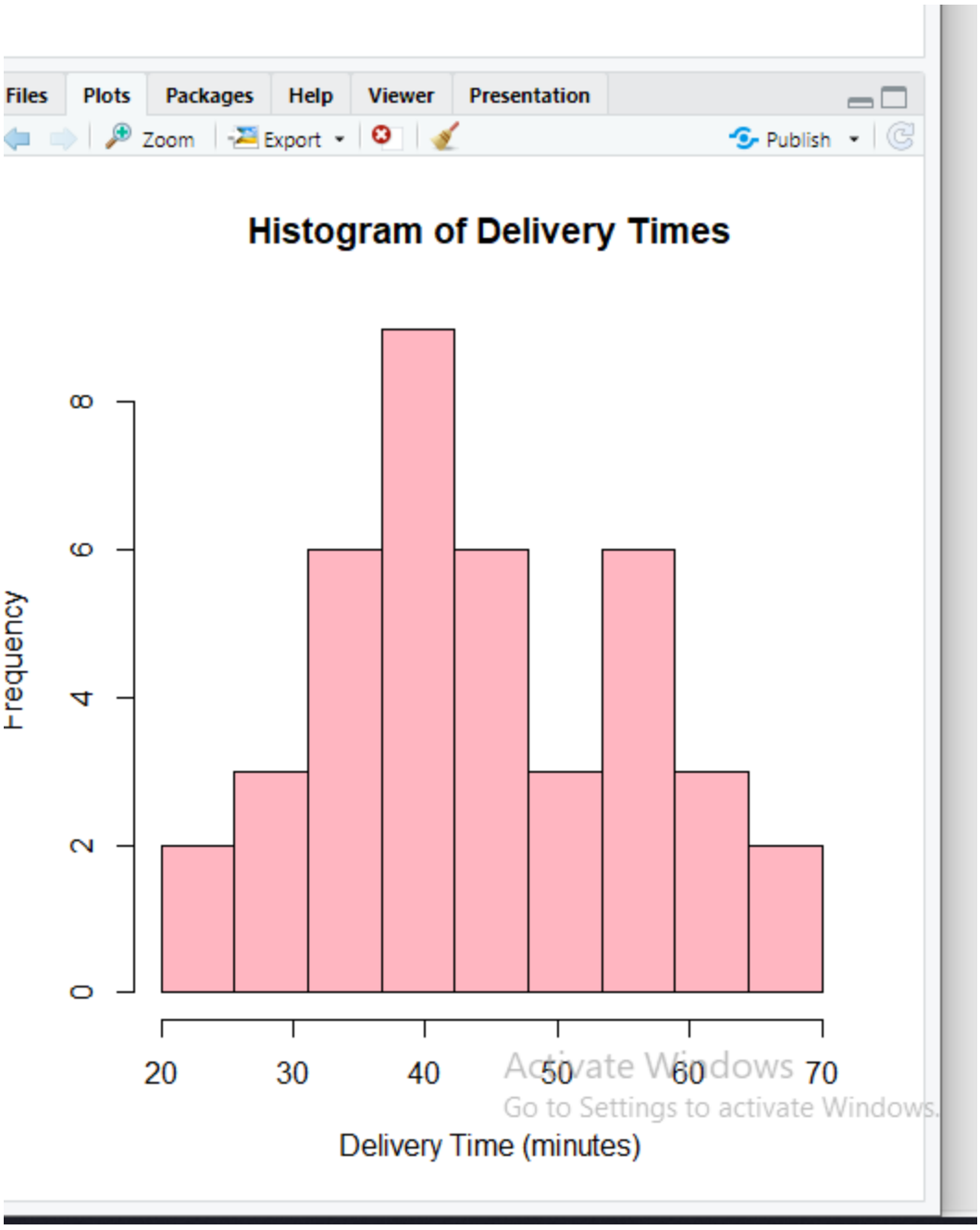
```
# 01.

Delivery_Times <- read.table('Exercise - Lab 05.txt', header = TRUE)

cat("Dataset structure:\n")
str(Delivery_Times)
cat("\nFirst few rows:\n")
head(Delivery_Times)
```

```
> # 01.
>
> Delivery_Times <- read.table('Exercise - Lab 05.txt', header = TRUE)
>
> cat("Dataset structure:\n")
Dataset structure:
> str(Delivery_Times)
'data.frame':   40 obs. of  1 variable:
 $ Delivery_Time_.minutes.: int  34 54 47 29 39 61 20 40 57 36 ...
> cat("\nFirst few rows:\n")

First few rows:
> head(Delivery_Times)
  Delivery_Time_.minutes.
1                      34
2                      54
3                      47
4                      29
5                      39
6                      61
>
```

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.

```r
# 2.

breaks <- seq(20, 70, length.out = 10)

cat("\nClass intervals (right open):\n")
intervals <- paste0("(", head(breaks, -1), ", ", tail(breaks, -1), "]")
print(intervals)

hist(Delivery_Times$Delivery_Time_.minutes.,
     breaks = breaks,
     right = TRUE,  # Right open intervals
     main = "Histogram of Delivery Times",
     xlab = "Delivery Time (minutes)",
     ylab = "Frequency",
     col = "lightpink",
     border = "black",
     xlim = c(20, 70))
```

```
> breaks <- seq(20, 70, length.out = 10)
>
> cat("\nClass intervals (right open):\n")

Class intervals (right open):
> intervals <- paste0("(", head(breaks, -1), ", ", tail(breaks, -1), "]")
> print(intervals)
[1] "(20, 25.5555555555556]"              "(25.5555555555556, 31.1111111111111]" "(31.1111111111111, 36.6666666666667]"
[4] "(36.6666666666667, 42.2222222222222]" "(42.2222222222222, 47.7777777777778]" "(47.7777777777778, 53.3333333333333]"
[7] "(53.3333333333333, 58.8888888888889]" "(58.8888888888889, 64.4444444444444]" "(64.4444444444444, 70]"
>
>
> hist(Delivery_Times$Delivery_Time_.minutes.,
+      breaks = breaks,
+      right = TRUE,  # Right open intervals
+      main = "Histogram of Delivery Times",
+      xlab = "Delivery Time (minutes)",
+      ylab = "Frequency",
+      col = "lightpink",
+      border = "black",
+      xlim = c(20, 70))
>
```




3. Comment on the shape of the distribution.

```r
# 3.

cat("\n3. Shape of the distribution:\n")
# Analyze the distribution shape from the histogram
dist_shape <- "The distribution appears to be approximately symmetric with a slight right skew. "
dist_shape <- paste0(dist_shape, "Most delivery times are concentrated between 35-55 minutes. ")
dist_shape <- paste0(dist_shape, "There are fewer deliveries at the extremes (very fast or very slow delivery times).")
cat(dist_shape, "\n")
```

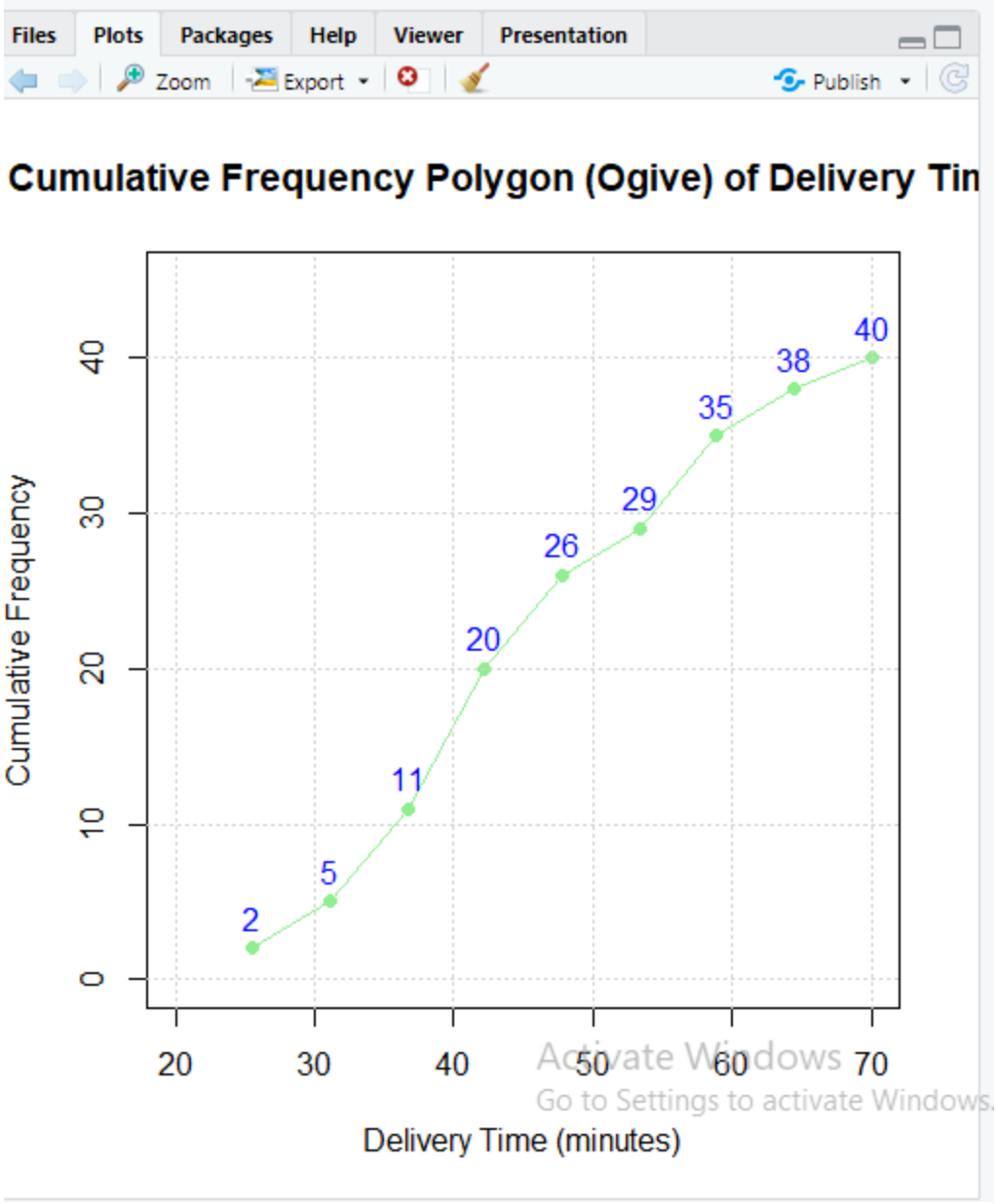
```
+      xlim = c(20, 70))
> # 3.
>
> cat("\n3. Shape of the distribution:\n")

3. Shape of the distribution:
> # Analyze the distribution shape from the histogram
> dist_shape <- "The distribution appears to be approximately symmetric with a slight right skew. "
> dist_shape <- paste0(dist_shape, "Most delivery times are concentrated between 35-55 minutes. ")
> dist_shape <- paste0(dist_shape, "There are fewer deliveries at the extremes (very fast or very slow delivery times).")
> cat(dist_shape, "\n")
The distribution appears to be approximately symmetric with a slight right skew. Most delivery times are concentrated between 35-55 minutes. There are fewer d
eliveries at the extremes (very fast or very slow delivery times).
>
```

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

```r
# 4.

freq <- hist(Delivery_Times$Delivery_Time_.minutes., breaks = breaks, plot = FALSE)$counts
cum_freq <- cumsum(freq)

cat("\nFrequency distribution:\n")
freq_table <- data.frame(Interval = intervals, Frequency = freq, Cumulative = cum_freq)
print(freq_table)


plot(breaks[-1], cum_freq,  # Use right endpoints
     type = "o",
     pch = 16,
     col = "lightgreen",
     main = "Cumulative Frequency Polygon (Ogive) of Delivery Times",
     xlab = "Delivery Time (minutes)",
     ylab = "Cumulative Frequency",
     xlim = c(20, 70),
     ylim = c(0, max(cum_freq) + 5))

grid()


text(breaks[-1], cum_freq, labels = cum_freq, pos = 3, col = "blue")

# Save the plots
png("delivery_times_histogram.png", width = 800, height = 600)
hist(Delivery_Times$Delivery_Time_.minutes.,
```

```r
# Save the plots
png("delivery_times_histogram.png", width = 800, height = 600)
hist(Delivery_Times$Delivery_Time_.minutes.,
     breaks = breaks,
     right = TRUE,
     main = "Histogram of Delivery Times",
     xlab = "Delivery Time (minutes)",
     ylab = "Frequency",
     col = "lightblue",
     border = "black",
     xlim = c(20, 70))
dev.off()

png("delivery_times_ogive.png", width = 800, height = 600)
plot(breaks[-1], cum_freq,
     type = "o",
     pch = 16,
     col = "red",
     main = "Cumulative Frequency Polygon (Ogive) of Delivery Times",
     xlab = "Delivery Time (minutes)",
     ylab = "Cumulative Frequency",
     xlim = c(20, 70),
     ylim = c(0, max(cum_freq) + 5))
grid()
text(breaks[-1], cum_freq, labels = cum_freq, pos = 3, col = "yellow")
dev.off()

cat("\nAnalysis completed successfully!\n")
cat("Plots have been saved as 'delivery_times_histogram.png' and 'delivery_times_ogive.png'\n")
```

```
> png("delivery_times_ogive.png", width = 800, height = 600)
> plot(breaks[-1], cum_freq,
+      type = "o",
+      pch = 16,
+      col = "red",
+      main = "Cumulative Frequency Polygon (Ogive) of Delivery Times",
+      xlab = "Delivery Time (minutes)",
+      ylab = "Cumulative Frequency",
+      xlim = c(20, 70),
+      ylim = c(0, max(cum_freq) + 5))
> grid()
> text(breaks[-1], cum_freq, labels = cum_freq, pos = 3, col = "yellow")
> dev.off()
RStudioGD
        2
>
> cat("\nAnalysis completed successfully!\n")

Analysis completed successfully!
> cat("Plots have been saved as 'delivery_times_histogram.png' and 'delivery_times_ogive.png'\n")
Plots have been saved as 'delivery_times_histogram.png' and 'delivery_times_ogive.png'
>
```


Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Cumulative Frequency Polygon (Ogive) of Delivery Tin
Cumulative Frequency
0
10
20
30
40
2
5
11
20
26
29
35
38
40
20
30
40
50
60
70
Delivery Time (minutes)
Activate Windows
Go to Settings to activate Windows.