Faculty of Computing

Year 2 Semester 1 (2025)

IT2120 - Probability and Statistics

# Lab Sheet 05

Exercise

1.

```
setwd("C:\\Users\\IT24103987\\Desktop\\IT24103987")
getwd()

#. Import the dataset ('Exercise - Lab 05.txt') into R and store it in a data frame called "Delivery Times".
Delivery_Time <- read.table("Exercise - Lab 05.txt", header = TRUE)
head(Delivery_Time)
```

```
> setwd("C:\\Users\\IT24103987\\Desktop\\IT24103987")
> getwd()
[1] "C:/Users/IT24103987/Desktop/IT24103987"
> #. Import the dataset ('Exercise - Lab 05.txt') into R and store it in a data frame called "Delivery Times".
> Delivery_Time <- read.table("Exercise - Lab 05.txt", header = TRUE)
> head(Delivery_Time)
  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
histogram <- hist(Delivery_Time$Delivery_Time,
                  main = "Histogram of Delivery Times",
                  breaks = seq(20, 70, length = 10),
                  xlab = "Delivery Time",
                  ylab = "Frequency")
```

Histogram of Delivery Times
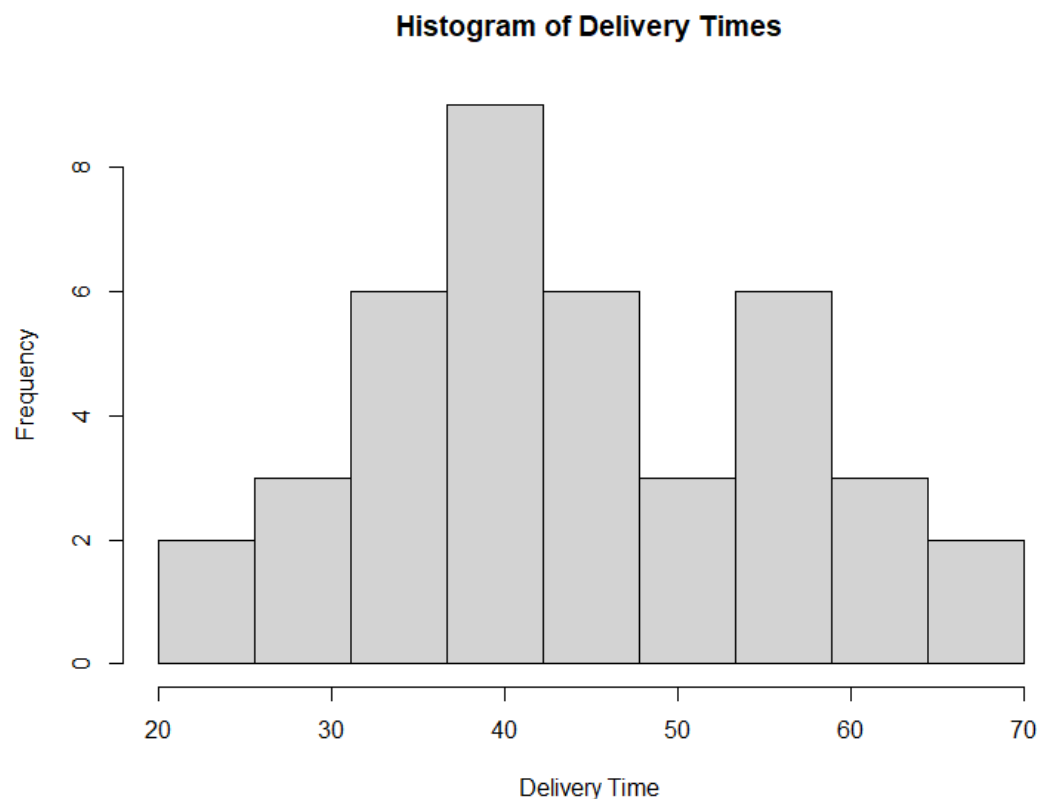
3. Comment on the shape of the distribution

Shape: The distribution is roughly bell-shaped (unimodal) with most delivery times clustering around the middle around 35–45 minutes.

Symmetry: It appears fairly symmetric, with frequencies tapering off on both sides of the central peak.

Spread: Delivery times range from about 20 minutes to 70 minutes.

In this histogram, the bars are highest in the middle (around 40), and the left side (20–30) and right side (60–70) drop down almost evenly.

So the shape is approximately normal not strongly skewed.

Peaks: The highest frequency occurs near 40 minutes.

4.

```r
#Draw a cumulative frequency polygon (ogive) for the data in a separate plot
hist_data <- hist(Delivery_Time$Delivery_Time,
                  breaks = seq(20, 70, length = 10),
                  plot = FALSE)

cum_freq <- cumsum(hist_data$counts)

plot(hist_data$mids,
     cum_freq,
     type = "o",   # "o" ensures both points and lines are drawn
     xlab = "Delivery Time",
     ylab = "Cumulative Frequency",
     main = "Cumulative Frequency Polygon (Ogive)",
     pch = 16)
```

Cumulative Frequency Polygon (Ogive)