Sri Lanka Institute of Information Technology

Lab Submission
05

IT24102798

Sooriyabandara U.R.G.W.K.

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”.

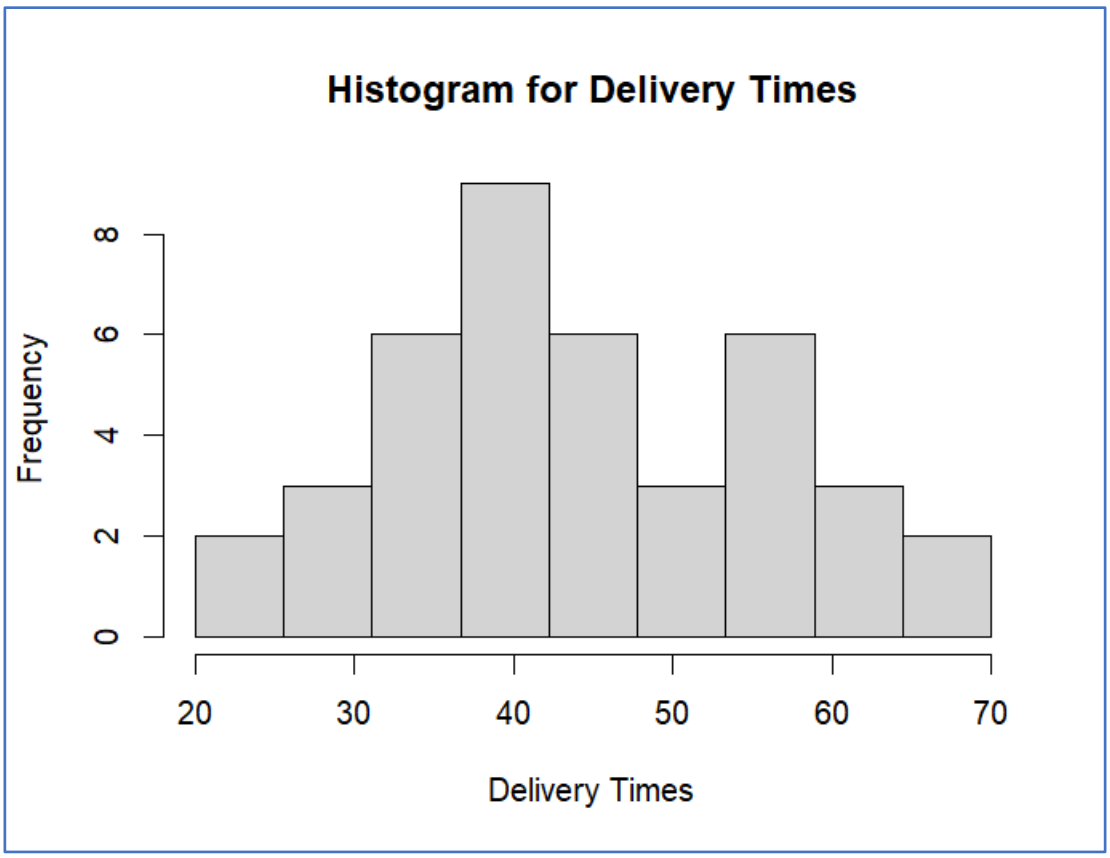
```
5
6  # Question 01
7
8  data <- read.table("Exercise - Lab 05.txt", header = TRUE, sep = ",")
9  Delivery_Times <- data.frame(data)
10 names(Delivery_Times) <- c("Delivery_Time")
11 attach(Delivery_Times)
12
```

```
> # Question 01
>
> data <- read.table("Exercise - Lab 05.txt", header = TRUE, sep = ",")
> Delivery_Times <- data.frame(data)
> names(Delivery_Times) <- c("Delivery_Time")
> attach(Delivery_Times)
```

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.

```
13
14 # Question 02
15
16 str(Delivery_Times)
17 histogram_for_delivery_times <- hist(Delivery_Time,
18                                       main = "Histogram for Delivery Times",
19                                       xlab = "Delivery Times",
20                                       ylab = "Frequency",
21                                       breaks = seq(20, 70, length = 10),
22                                       right = FALSE)
23
```

```
>
> # Question 02
>
> str(Delivery_Times)
'data.frame':   40 obs. of  1 variable:
 $ Delivery_Time: int  34 54 47 29 39 61 20 40 57 36 ...
> histogram_for_delivery_times <- hist(Delivery_Time,
+                                      main = "Histogram for Delivery Times",
+                                      xlab = "Delivery Times",
+                                      ylab = "Frequency",
+                                      breaks = seq(20, 70, length = 10),
+                                      right = FALSE)
>
```



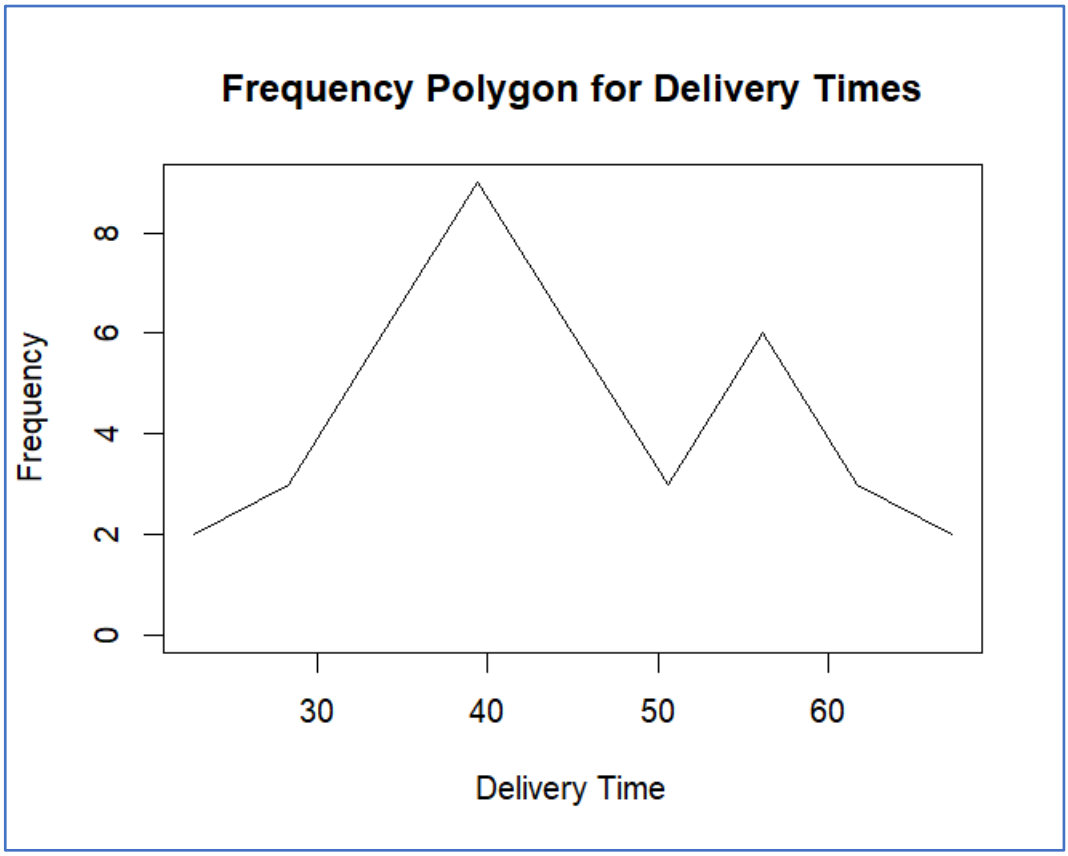
3. Comment on the shape of the distribution.
   - The 4th bar has the most frequency, which is 8 while the 1st and 9th bars have the least frequency, which is 2.
   - The graph is slightly skewed to the left side.

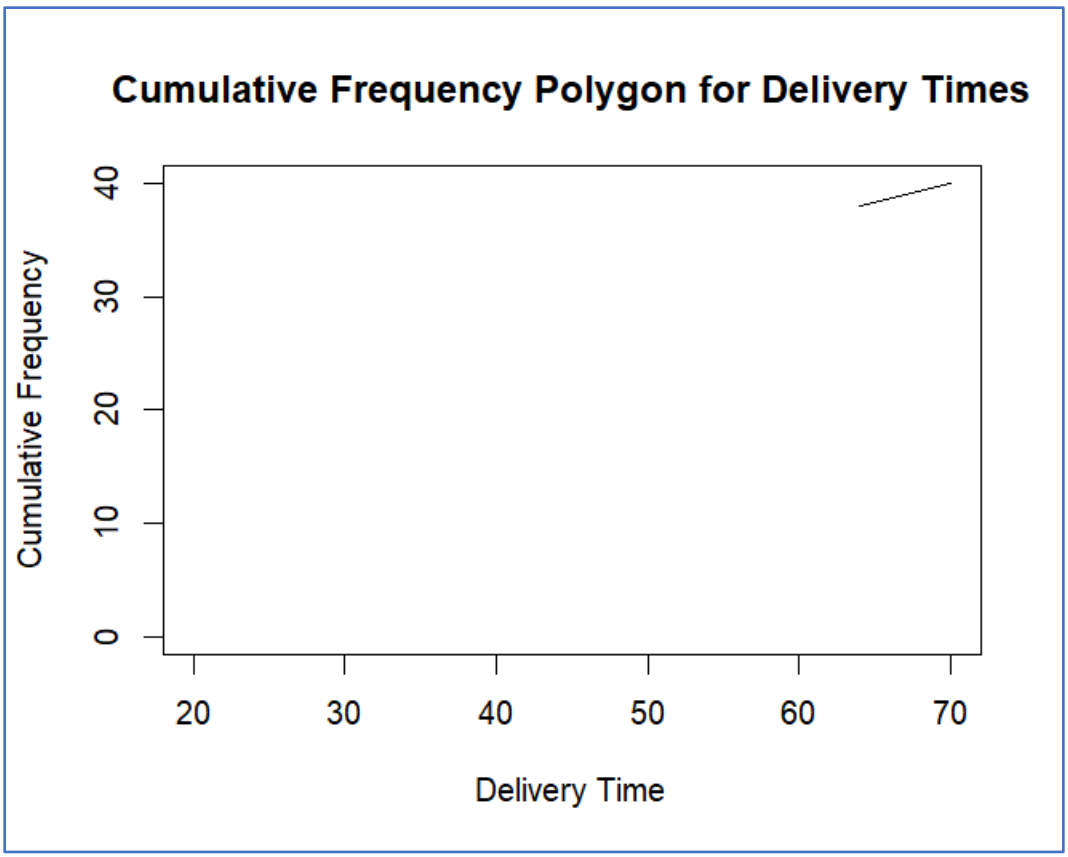
4. Draw a cumulative frequency polygon (ogive) for the data in a separate plot.

```r

# Question 04

breaks <- round(histogram_for_delivery_times$breaks)
freq <- histogram_for_delivery_times$counts
mids <- histogram_for_delivery_times$mids

classes <- c()

for (i in 1:(length(breaks) - 1)) {
  classes[i] <- paste0("[", breaks[i], ",", breaks[i + 1], ")")
}

cbind(Classes = classes, Frequency = freq)

lines(mids, freq)

plot(mids, freq, type = 'l', main = "Frequency Polygon for Delivery Times", xlab = "Delivery Time", ylab = "Frequency", ylim = c(0, max(freq)))

cum.freq <- cumsum(freq)

new <- c()

for(i in i:length(breaks)) {
  if (i == 1) {
    new[i] = 0
  } else {
    new[i] = cum.freq[i-1]
  }
}

plot(breaks, new, type = 'l', main = "Cumulative Frequency Polygon for Delivery Times", xlab = "Delivery Time", ylab = "Cumulative Frequency", ylim = c(0, max(cum.freq)))

cbind(Upper = breaks, CumFreq = new)

```

```
>
> # Question 04
Warning messages:
1: In doTryCatch(return(expr), name, parentenv, handler) :
  display list redraw incomplete
2: In doTryCatch(return(expr), name, parentenv, handler) :
  invalid graphics state
3: In doTryCatch(return(expr), name, parentenv, handler) :
  invalid graphics state
>
> breaks <- round(histogram_for_delivery_times$breaks)
> freq <- histogram_for_delivery_times$counts
> mids <- histogram_for_delivery_times$mids
>
> classes <- c()
>
> for (i in 1:(length(breaks) - 1)) {
+   classes[i] <- paste0("[", breaks[i], ",", breaks[i + 1], ")")
+ }
>
> cbind(Classes = classes, Frequency = freq)
      Classes   Frequency
 [1,] "[20,26)" "2"
 [2,] "[26,31)" "3"
 [3,] "[31,37)" "6"
 [4,] "[37,42)" "9"
 [5,] "[42,48)" "6"
 [6,] "[48,53)" "3"
 [7,] "[53,59)" "6"
 [8,] "[59,64)" "3"
 [9,] "[64,70)" "2"
>
> lines(mids, freq)
>
> plot(mids, freq, type = 'l', main = "Frequency Polygon for Delivery Times", xlab = "Delivery Time", ylab = "Frequency", ylim = c(0, max(freq)))
>
```

```
>
> cum.freq <- cumsum(freq)
>
> new <- c()
>
> for(i in i:length(breaks)) {
+   if (i == 1) {
+     new[i] = 0
+   } else {
+     new[i] = cum.freq[i-1]
+   }
+ }
>
> plot(breaks, new, type = 'l', main = "Cumulative Frequency Polygon for Delivery Times", xlab = "Delivery Time", ylab = "Cumulative Frequency", ylim = c(0, max(cum.freq)))
>
> cbind(Upper = breaks, CumFreq = new)
      Upper CumFreq
 [1,]    20      NA
 [2,]    26      NA
 [3,]    31      NA
 [4,]    37      NA
 [5,]    42      NA
 [6,]    48      NA
 [7,]    53      NA
 [8,]    59      NA
 [9,]    64      38
[10,]    70      40
>
```




Cumulative Frequency Polygon for Delivery Times
Cumulative Frequency
0
10
20
30
40
20
30
40
50
60
70
Delivery Time