IT24101099

Madushika B.G.C.

PS Lab05

```
setwd("C:\\Users\\ASUS\\Desktop\\PS_LAB05")
getwd()
```

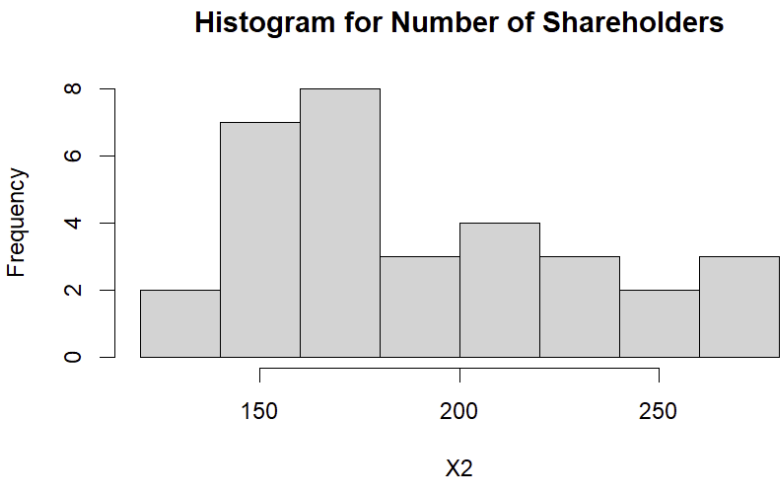
```
> setwd("C:\\Users\\ASUS\\Desktop\\PS_LAB05")
> getwd()
[1] "C:/Users/ASUS/Desktop/PS_LAB05"
```

```
#Importing the data set
data <- read.table("Data.txt",header =TRUE,sep = ",")
##View the filr in a separate  window
fix(data)
#Close the data window before you run the rest of the commands.
#Unless rest of the commands won't run.

#Attach the filr into R. so,you can call the variables by their names
attach(data)
```

```
> data <- read.table("Data.txt",header =TRUE,sep = ",")
> fix(data)
> attach(data)
The following objects are masked from data (pos = 4):

    Company, Number_of_Shareholders.thousands.
```

```
##Part1
#Rename the variables (column headings ) of the data set as X1 and X2
names(data)<-c("X1","X2")
##Attach the file into R again as we renamed the variables.
attach (data)
##Obtain histogram for number of shareholders
hist(X2,main="Histogram for Number of Shareholders")
```

```
> names(data)<-c("X1","X2")
> attach (data)
The following objects are masked from data (pos = 4):
```




```
##Part 2
#Using "breaks command we can define the number of classes we need i
#along with the lower limit and upper limit
#Using "right" command we can define whether classes have closed int
histogarm <- hist(X2,main="Histogram for Number of Shareholders",bre
##Check how each argument inside "hist" command works using "help" c
?hist
```

```
> hist(X2,main="Histogram for Number of Shareholders")
> histogarm <- hist(X2,main="Histogram for Number of Shareholders",breaks
= seq(130,270,length=8),right = FALSE)
> ?hist
```

resulting object of class "histogram" is plotted by plot.histogram, before it is returr

Usage

```
hist(x, ...)

## Default S3 method:
hist(x, breaks = "Sturges",
     freq = NULL, probability = !freq,
     include.lowest = TRUE, right = TRUE, fuzz = 1e-7,
     density = NULL, angle = 45, col = "lightgray", border = NULL,
     main = paste("Histogram of" , xname),
     xlim = range(breaks), ylim = NULL,
     xlab = xname, ylab,
     axes = TRUE, plot = TRUE, labels = FALSE,
     nclass = NULL, warn.unused = TRUE, ...)
```

Arguments

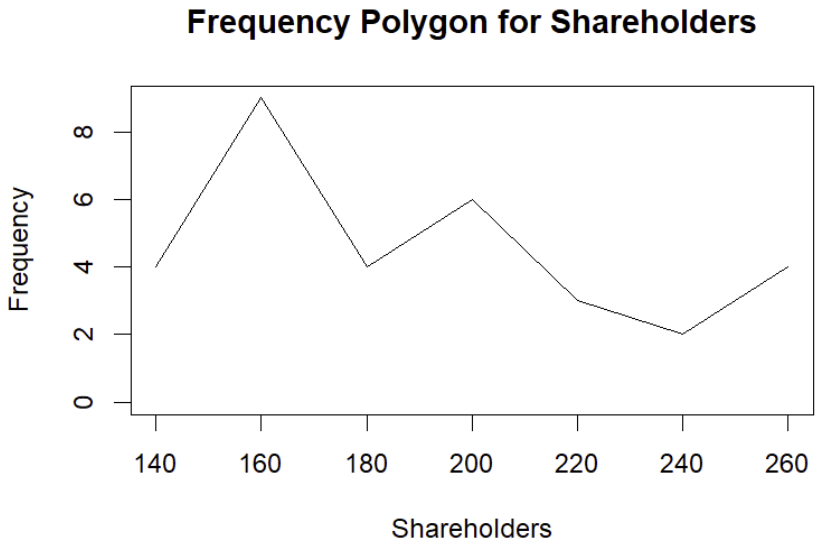
```
##Part3
##Assign class limits of the frequency distribution into a variable called "breaks"
breaks <- round(histogarm$breaks)
#Assign class frequencies of the histogram into a variable called"freq"
freq <- histogarm$counts
#Assign mid point of each class into a variable called"mids"
mids <- histogarm$mids

#Creating the variable called "Classes" for the frequency distribution
classes <- c()

##Creating a "for" loop to assign classes of the frequency distribution into "Classes" variable c
for (i in 1: length(breaks)-1){
  classes[i] <- paste0("[",breaks[i],",",breaks[i+1],")")
}
```

```
##Creating a "for" loop to assign classes of the frequency distribution into "Classes" variable cı
for (i in 1: length(breaks)-1){
  classes[i] <- paste0("[",breaks[i],",",breaks[i+1],")")
}
#Obtaining frequency distribution by combining the values of "Classes" & "freq" variables
##"cbind" command used to merge the columns with the columns with same length
cbind(Classes = classes,Frequency = freq)
```

```
> breaks <- round(histogarm$breaks)
> freq <- histogarm$counts
> mids <- histogarm$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,] "[130,150)" "4"
[2,] "[150,170)" "9"
[3,] "[170,190)" "4"
[4,] "[190,210)" "6"
[5,] "[210,230)" "3"
[6,] "[230,250)" "2"
```

```
#Part 4
##Draw  frequency polygon to the same plot.
lines(mids,freq)

#Draw frequency polygon in a new plot.
plot(mids,freq,type ='l',main = "Frequency Polygon for Shareholders",xlab = "Shareholders",ylab =
```

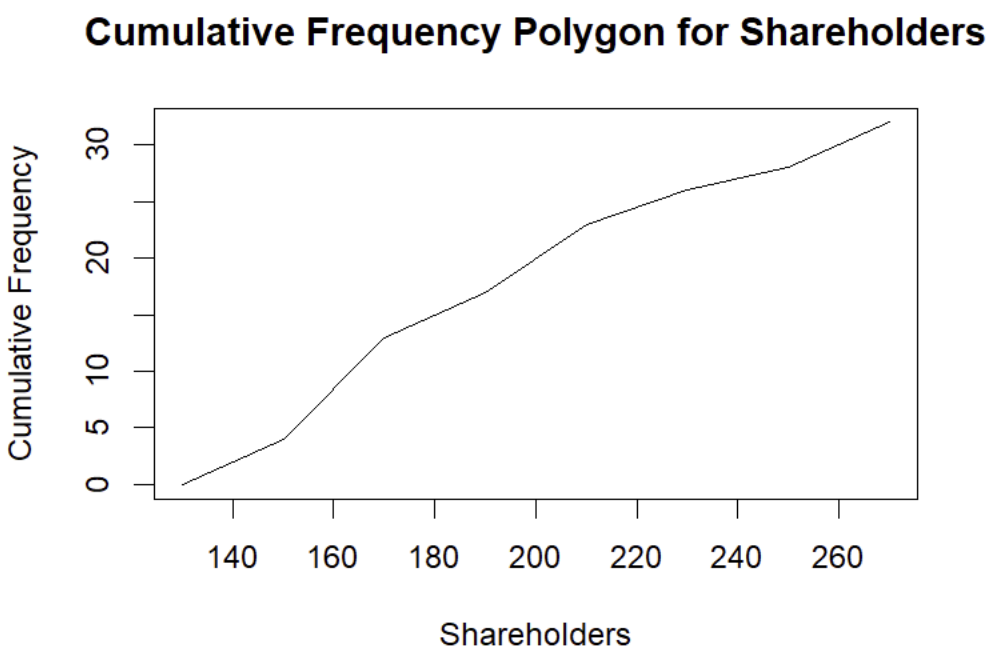
```
> lines(mids,freq)
> plot(mids,freq,type ='l',main = "Frequency Polygon for Shareholders",xlab = "Shareholders",ylab = "Fr
equency",ylim = c(0,max(freq)))
```




```
#Part 5
##Using "cumsum" command we can get cumulative frequencies
cum.freq <- cumsum(freq)
#Creating a null variable called "new"
new <- c()
## Using "for" loop to store cumulative frequencies in order to get the ogive
for(i in 1:length(breaks)){
  if(i==1){
    new[i] = 0
  }else{
    new[i]= cum.freq[i-1]
  }
}
#Draw cumulative frequency polygon in a new plot
plot(breaks,new,type="l",main = "Cumulative Frequency Polygon for Shareholders",xlab
#Obtain upper limit of each class along with its cumulative frequency in a table
cbind(Upper = breaks,CumFreq = new)
```

```
> lines(mids,freq)
> plot(mids,freq,type ='l',main = "Frequency Polygon for Shareholders",xlab = "Shareholder
s",ylab = "Frequency",ylim = c(0,max(freq)))
> cum.freq <- cumsum(freq)
> new <- c()
> for(i in 1:length(breaks)){
+   if(i==1){
+     new[i] = 0
+   }else{
+     new[i]= cum.freq[i-1]
+   }
+ }
> plot(breaks,new,type="l",main = "Cumalative Frequency Polygon for Shareholders"",xlab =
"Shareholders",ylab ="Cumulative Frequency",ylim=c(0,max(cum.freq)))
```

```
> plot(breaks,new,type="l",main = "Cumulative Frequency Polygon for Shareholders",xlab =
"Shareholders",ylab ="Cumulative Frequency",ylim=c(0,max(cum.freq)))
> cbind(Upper = breaks,CumFreq = new)
     Upper CumFreq
[1,]   130       0
[2,]   150       4
[3,]   170      13
[4,]   190      17
[5,]   210      23
[6,]   230      26
[7,]   250      28
[8,]   270      32
>
```



Exercise

(1)

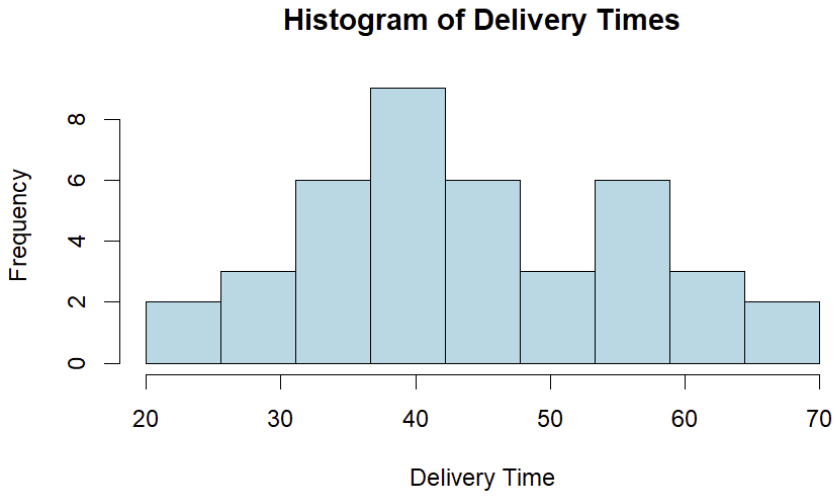
```
# Import dataset (assuming the file is in working directory)
Delivery_Times <- read.table("Exercise - Lab 05.txt", header = TRUE)

head(Delivery_Times)
```

```
> setwd("C:\\Users\\ASUS\\Desktop\\PS_LAB05")
> Delivery_Times <- read.table("Exercise - Lab 05.txt", header = TRUE)
> head(Delivery_Times)
  Delivery_Time_.minutes.
1                      34
2                      54
3                      47
4                      29
5                      39
6                      61
```

(2)

```
# Convert column to numeric
Delivery_Times$Time <- as.numeric(as.character(Delivery_Times$Time))

# Define breaks (20 to 70 with 9 equal intervals)
breaks <- seq(20, 70, length.out = 10)

# Draw histogram
hist(Delivery_Times$Time,
     breaks = breaks,
     right = FALSE,
     col = "lightblue",
     border = "black",
     main = "Histogram of Delivery Times",
     xlab = "Delivery Time",
     ylab = "Frequency")
```

```
> # Define breaks (20 to 70 with 9 equal intervals)
> breaks <- seq(20, 70, length.out = 10)
> # Draw histogram
> hist(Delivery_Times$Time,
+      breaks = breaks,
+      right = FALSE,
+      col = "lightblue",
+      border = "black",
+      main = "Histogram of Delivery Times",
+      xlab = "Delivery Time",
+      ylab = "Frequency")
```



(3)

Comment on the shape of the distribution.

Bell shaped/Normal because the bars are all tall in the middle and tapper off

(4)

```
# Build frequency table
freq_table <- hist(Delivery_Times$Time,
                   breaks = breaks,
                   right = FALSE,
                   plot = FALSE)

# Cumulative frequencies
cum_freq <- cumsum(freq_table$counts)

# Ogive plot
plot(breaks[-1], cum_freq, type = "o", pch = 16, col = "blue",
     main = "Cumulative Frequency Polygon (Ogive)",
     xlab = "Delivery Time",
     ylab = "Cumulative Frequency")
```

```
> # Build frequency table
> freq_table <- hist(Delivery_Times$Time,
+                    breaks = breaks,
+                    right = FALSE,
+                    plot = FALSE)
> # Cumulative frequencies
> cum_freq <- cumsum(freq_table$counts)
> # Ogive plot
> plot(breaks[-1], cum_freq, type = "o", pch = 16, col = "blue",
+      main = "Cumulative Frequency Polygon (Ogive)",
+      xlab = "Delivery Time",
+      ylab = "Cumulative Frequency")
>
```