Faculty of Computing

Year 2 Semester 1 (2025)

IT2120 - Probability and Statistics   Lab Sheet 05

Lab Exercise 5 (Descriptive Statistics) The numbers of shareholders for a selected of large companies (in thousands) are given below.

Company	Number of Shareholders (thousands)
Pan American World Airways	144
General Public Utilities	266
Occidental Petroleum	177
Middle South Utilities	133
DaimlerChrysler	209
Standard Oil of California	264
Bethlehem Steel	160
Long Island Lighting	143
RCA	246
Greyhound Corporation	151
Pacific Gas & Electric	239
Niagara Mohawk Power	204
E.I. du Pont de Nemours	204
Westinghouse Electric	195
Union Carbide	176
BankAmerica	175
Northwest Utilities	200
Standard Oil (India)	173
Atlantic Richfield	195
Detroit Edition	220
Eastman Kodak	251
Dow Chemical	137
Pennsylvania Power	150
American Electric Power	262
Ohio Edition	158
Transamerica Corporation	162
Colombia Gas System	165
International Telephone	223
Union Electric	158
Virginia Electric and Power	162
Public Service Electric & Gas	225
Consumer Power	161

Before starting the lab sheet, you need to create a folder in your desktop and save all your working inside the folder. Set the working directory to that folder using the following command:

```
setwd("paste the path of the folder")
```

Eg:- `setwd("D:\\2025 - Sem 2\\IT2120\\Lab Sessions\\Lab 05")`

The number of shareholders is to be organized into a frequency distribution.

1. Draw a histogram for the above data.

```
##Setting the directory
setwd("D:\\2025 - Sem 2\\IT2120 - New\\Lab Sessions\\Lab 05")

##Importing the data set
data<-read.table("Data.txt",header=TRUE,sep = ",")

##view the file 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 file into R. So, you can call the variables by their names.
attach(data)
```

```
##Part 1
##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")
```

2. Draw a histogram using seven classes where the lower limit is 130 and an upper limit of 270.

```
##Part 2
##Using "breaks" command we can define number of classes we need in the histogram
##along with lower limit and upper limit.
##Using "right" command we can define whether classes have closed intervals or open intervals.
histogram<-hist(X2,main="Histogram for Number of Shareholders",breaks = seq(130, 270,length = 8),right = FALSE)

##Check how each argument inside "hist" command works using "help" command as follows
?hist
```

3. Construct the frequency distribution for the above specification.

```
##Part 3
##Assign class limits of the frequency distribution into a variable called "breaks"
breaks <- round(histogram$breaks)
##Assign class frequencies of the histogram into a variable called "freq"
freq <- histogram$counts
##Assign mid point of each class into a variable called "mids"
mids <- histogram$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 crated above.
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 same length
cbind(Classes = classes, Frequency = freq)
```

4. Portray the distribution in the form of a frequency polygon.

```
##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 = "Frequency", ylim = c(0,max(freq)))
```

5. Portray the distribution in a cumulative frequency polygon (ogive).

```
##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 = "Cumalative Frequency Polygon for Shareholders",
     xlab = "Shareholders", ylab = "Cumulative Frequency", ylim = c(0,max(cum.freq)))
##Obtain upper limit of each class along with its cumulative frequency in  a table
cbind(Upper = breaks, CumFreq = new)
```

Exercise

Instructions: Create a folder in your desktop with your registration number (Eg: "IT......."). You need to save the R script file and take screenshots of the command prompt with answers and save it in a word document inside the folder. Save both R script file and word document with your registration number (Eg: "IT........"). After you finish the exercise, zip the folder and upload the zip file to the submission link.

1. Import the dataset ('Exercise – Lab 05.txt') into R and store it in a data frame called "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.

3. Comment on the shape of the distribution.

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