Чернівецький національний університет імені Юрія Федьковича

Інститут фізико-технічних та комп’ютерних наук

Кафедра Математичних проблем управління і кібернетики

Лабораторна робота № 1

“Тема: Основи програмування в R”

з дисципліни

“Інтелектуальний аналіз”

Варіант - 2

Виконав:

ст. гр. 341СК Гаваньо Дмитро

Прийняв:

професор Дрінь Я. М.

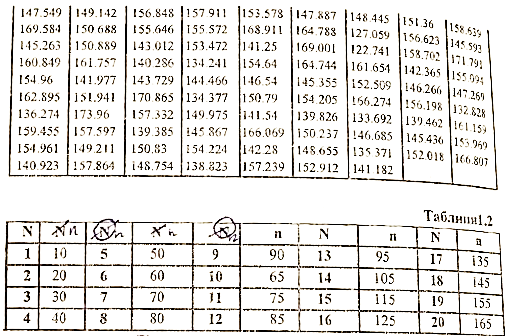
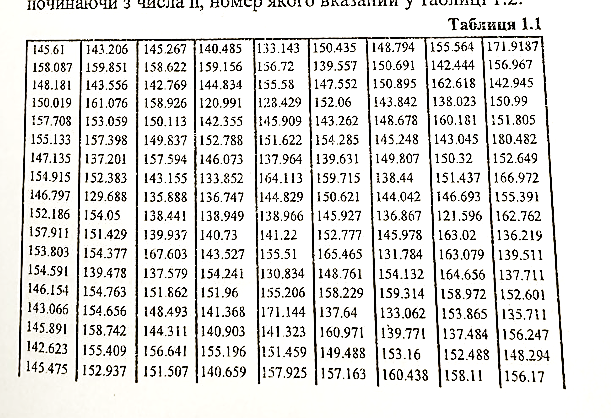
Чернівці – 2025

**Мета:** Ознайомитися з основними функціями R, а саме із введенням та зчитуванням даних, обробкою даних, основними методами візуалізації даних.

**Завдання:** Обчисліть максимальне, мінімальне значення і розмах для заданої вибірки. Виконайте групування для значень m=10,20, побудуйте відповідні гістограми, полігони частот, полігони накопичених частот. Виконайте обчислення для 100 чисел із таблиці 1.1, починаючи з числа п, номер якого вказаний у таблиці 1.2.

**Додаткова інформація:**

Дані для роботи програми необхідно внести у файл data.csv, що повинен знаходитися разом із файлом прогрмами в одній папці.



**Хід роботи:**

Програмний код:

setwd("D:\\Study\\R") # set working directory

n <- 20 # starting index

y\_length <- 100 # data length

m <- 5 # num of intervals

# -----------------------------------------------------

y <- read.csv("numbers\_list.csv")

#1

y\_vec <- as.numeric(unlist(y[[1]]))

y\_vec <- y\_vec[!is.na(y\_vec)]

end <- min(n + y\_length - 1, 250)

y\_vec <- y\_vec[n:end]

#2

y\_sorted <- sort(y\_vec)

print\_first <- 3

print\_last <- 3

cat("Selection of", y\_length, "numbers from general selection starting from index", n, # nolint

":\n")

cat(y\_sorted[1:print\_first], sep = ", ")

cat("\n...\n")

cat(y\_sorted[(y\_length - print\_last + 1):y\_length], sep = ", ")

#3

min\_val <- min(y\_sorted)

max\_val <- max(y\_sorted)

delta <- (max\_val - min\_val) / m

cat("\nDelta (interval width):", delta, "\n")

#4

if (y\_length %% m != 0)

stop("\n >>> Data length must be divisible by number of intervals")

x <- c()

for (j in 1:m) {

interval\_center\_offset <- delta \* (j - 0.5)

xj <- min\_val + interval\_center\_offset

x <- c(x, xj)

}

cat("\nCenters of intervals:\n")

cat(x, sep = ", ")

#5

breaks <- seq(min\_val, max\_val + 1e-8, length.out = m + 1)

freq\_vector <- table(cut(y\_sorted, breaks = breaks, right = FALSE))

cat("\nFrequencies of intervals:\n")

print(as.matrix(freq\_vector))

#6

png("Graphic\\!!!!Histogram.png")

barplot(freq\_vector, names.arg = round(x, 2),

xlab = "Interval Centers", ylab = "Frequencies",

main = "Histogram of Frequencies", col = "lightblue")

dev.off()

png("Graphic\\!!!FreqPolygon.png")

plot(x = x, y = as.matrix(freq\_vector), type = "b", pch = 19, col = "red",

xlab = "Interval", ylab = "Frequency",

main = "Frequency polygon")

dev.off()

#7

c\_freq <- cumsum(freq\_vector)

c\_freq\_vector <- as.matrix(c\_freq)

cat("Cumulative frequencies of intervals:\n")

print(c\_freq\_vector)

#8

rel\_c\_freq <- c\_freq / y\_length

rel\_c\_freq\_vector <- as.matrix(rel\_c\_freq)

cat("Relative cumulative frequencies of intervals:\n")

print(rel\_c\_freq\_vector)

png("Graphic\\!!CFreqPolygon.png")

plot(x = x, y = c\_freq, type = "b", pch = 19, col = "red",

xlab = "Interval", ylab = "Cumulative frequency",

main = "Cumulative Frequency polygon")

dev.off()

png("Graphic\\!RelCFreqPolygon.png")

plot(x = x, y = rel\_c\_freq, type = "b", pch = 19, col = "red",

xlab = "Interval", ylab = "Relative cumulative frequency",

main = "Relative cumulative Frequency polygon")

dev.off()

Результат виконання:

*M = 5*

Selection of 100 numbers from general selection starting from index 20 :

123.7678, 127.7597, 129.7796

...

167.7833, 171.2923, 176.5734

Number of intervals: 5

Delta (interval width): 10.56112

Centers of intervals:

129.0484, 139.6095, 150.1706, 160.7317, 171.2928

Frequencies of intervals:

[,1]

[124,134) 5

[134,145) 27

[145,155) 38

[155,166) 26

[166,177) 4

Cumulative frequencies of intervals:

[,1]

[124,134) 5

[134,145) 32

[145,155) 70

[155,166) 96

[166,177) 100

Relative cumulative frequencies of intervals:

[,1]

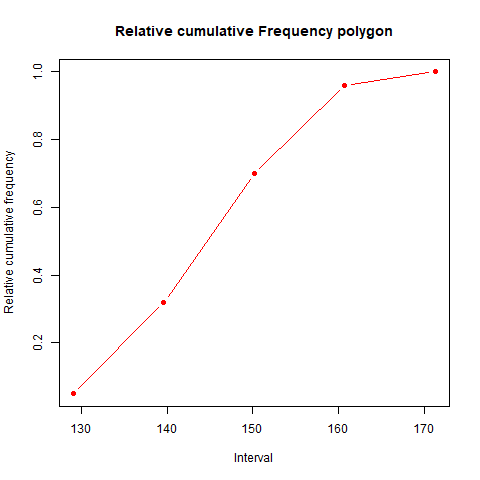
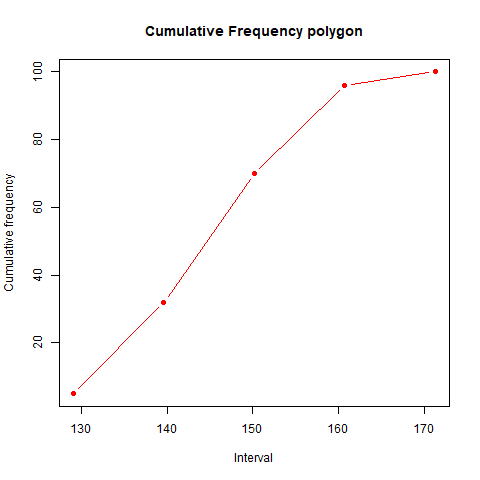
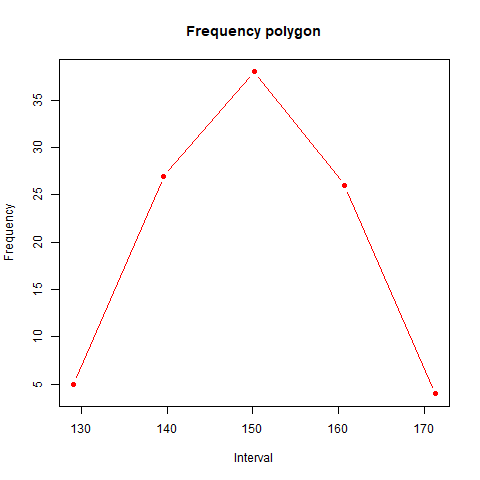
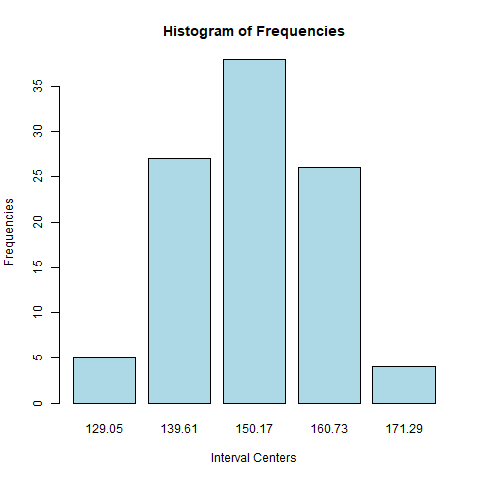
[124,134) 0.05

[134,145) 0.32

[145,155) 0.70

[155,166) 0.96

[166,177) 1.00



*M = 10*

123.7678, 127.7597, 129.7796

...

167.7833, 171.2923, 176.5734

Number of intervals: 10

Delta (interval width): 5.28056

Centers of intervals:

126.4081, 131.6886, 136.9692, 142.2498, 147.5303, 152.8109, 158.0914, 163.372, 168.6526, 173.9331

Frequencies of intervals:

[,1]

[124,129) 2

[129,134) 3

[134,140) 15

[140,145) 12

[145,150) 19

[150,155) 19

[155,161) 15

[161,166) 11

[166,171) 3

[171,177) 1

Cumulative frequencies of intervals:

[,1]

[124,129) 2

[129,134) 5

[134,140) 20

[140,145) 32

[145,150) 51

[150,155) 70

[155,161) 85

[161,166) 96

[166,171) 99

[171,177) 100

Relative cumulative frequencies of intervals:

[,1]

[124,129) 0.02

[129,134) 0.05

[134,140) 0.20

[140,145) 0.32

[145,150) 0.51

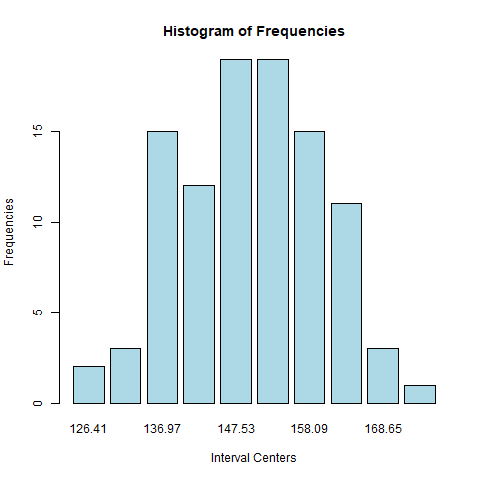
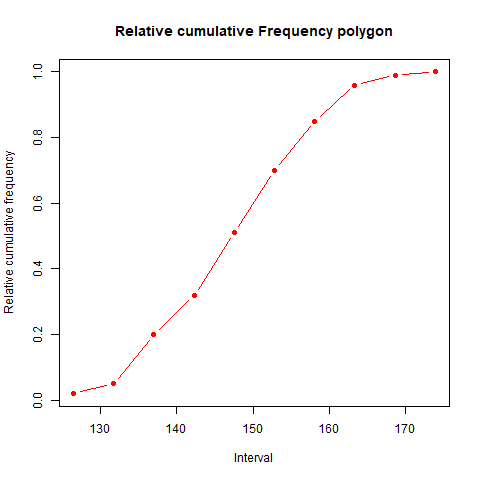
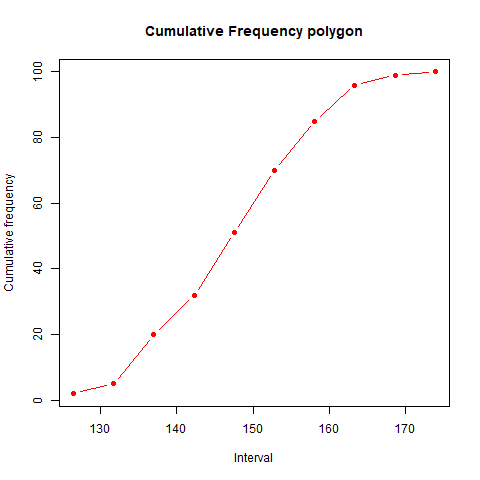
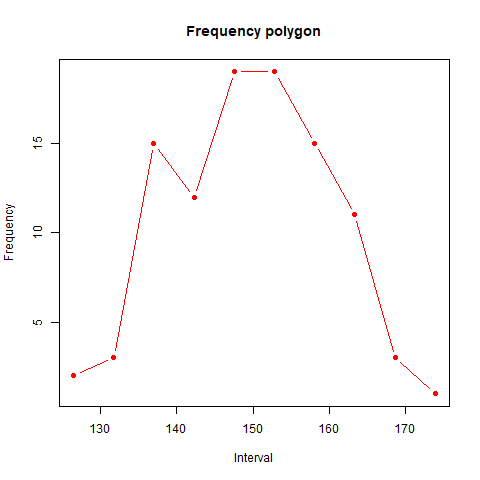
[150,155) 0.70

[155,161) 0.85

[161,166) 0.96

[166,171) 0.99

[171,177) 1.00

****

**Висновок:** я ознайомився з основними функціями R, а саме із введенням та зчитуванням даних, обробкою даних, основними методами візуалізації даних.