Чернівецький національний університет імені Юрія Федьковича

Інститут фізико-технічних та комп’ютерних наук

Кафедра Математичних проблем управління і кібернетики

Лабораторна робота № 9

“Тема: Лінійна регресія”

з дисципліни

“Інтелектуальний аналіз”

Варіант - 2

Виконав:

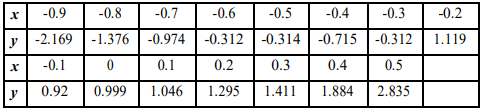
ст. гр. 341СК Гаваньо Дмитро

Прийняв:

професор Дрінь Я. М.

Чернівці – 2025

**Мета:** Ознайомитися з методикою побудови вибіркового рівняння лінійної регресії та оцінками його параметрів



**Завдання 9.1:** Для заданої в умові вибірки обчисліть регресію і знайдіть надійні інтервали коефіцієнтів регресії і дисперсії для заданої надійної ймовірності. Обчисліть смугу і коридор регресії. Зобразіть вибірку графічно на одному графіку з лінією регресії. Зобразіть графічно смугу і коридор регресії

1. Введіть задану вибірку.
2. Знайдіть точкові оцінки математичного сподівання обидвох змінних.
3. Обчисліть точкову незсунену оцінку невідомої дисперсії.
4. Знайдіть коефіцієнти регресії.
5. Побудуйте графік лінії регресії і зобразіть на ньому експериментальні точки.
6. Обчисліть значення критерію для оцінки коефіцієнта регресії 𝑎0.
7. Знайдіть надійний інтервал для 𝑎0.
8. Обчисліть значення критерію для оцінки коефіцієнта регресії 𝑎1.
9. Знайдіть надійний інтервал для 𝑎1.
10. Обчисліть значення критерію для оцінки дисперсії.
11. Знайдіть надійний інтервал для дисперсії.
12. Обчисліть коридор регресії.
13. Зобразіть на графіку лінію регресії і межі коридору для неї.
14. Обчисліть надійну область для всієї регресії.
15. Зобразіть на графіку лінію регресії і надійну область для неї.

**Хід роботи:**

**Завдання 1:**

Програмний код:

setwd("D:/Study/R/lab9")

data\_df <- read.csv("selection.csv",

                    header = TRUE,

                    sep = ",",

                    dec = ".",

                    stringsAsFactors = FALSE

                    )

alpha <- 0.05

selection\_x <- data\_df$X

selection\_y <- data\_df$Y

n <- length(selection\_x)

dot\_mean\_x <- mean(selection\_x)

dot\_mean\_y <- mean(selection\_y)

dot\_sd\_x <- sd(selection\_x)

dot\_sd\_y <- sd(selection\_y)

# Y ~ X - формула: Y залежить від X

model <- lm(Y ~ X, data = data\_df)

png("Graphic/lab9\_1\_Regression\_Plot.png" , width = 800, height = 600)

plot(selection\_x, selection\_y,

     main = "Граік регресії Y ~ X",

     type = "o",

     xlab = "X",

     ylab = "Y",

     pch = 19,

     col = "blue"

     )

abline(model, col = "red", lwd = 2)

dev.off()

t\_a <- qt(alpha / 2, df = n - 2)

cat("Квантиль t-розподілу для alpha =", alpha, "та df =", n - 2, ":", x\_a, "\n")

summary\_output <- summary(model)

s\_epsilon <- summary\_output$sigma

SSX <- sum((selection\_x - dot\_mean\_x)^2)

T\_CRIT <- c(t\_a , t\_a)

cat("Tочкове значення для a0:", coef(model)["(Intercept)"], "\n")

a0\_hat <- coef(model)["(Intercept)"]

SE\_a0 <- summary\_output$coefficients["(Intercept)", "Std. Error"]

a0\_l <- a0\_hat - t\_a \* SE\_a0

a0\_r <- a0\_hat + t\_a \* SE\_a0

cat("Довірчий інтервал a =",alpha,"для a0: [", a0\_l, ", ", a0\_r, "]\n")

cat("Tочкове значення для a1:", coef(model)["X"], "\n")

a1\_hat <- coef(model)["X"]

SE\_a1 <- summary\_output$coefficients["X", "Std. Error"]

a1\_l <- a1\_hat - t\_a \* SE\_a1

a1\_r <- a1\_hat + t\_a \* SE\_a1

cat("Довірчий інтервал a =",alpha,"для a1: [", a1\_l, ", ", a1\_r, "]\n")

chi\_la <- qchisq(alpha / 2, df = n - 2)

chi\_ra <- qchisq(1 - alpha / 2, df = n - 2)

dx\_l <- (n - 2) \* dot\_sd\_y ^ 2 / chi\_la

dx\_r <- (n - 2) \* dot\_sd\_y ^ 2 / chi\_ra

cat("Квантилі хі-квадрат розподілу для alpha =", alpha, ":", chi\_la, ", ", chi\_ra, "\n")

cat("Довірчий інтервал для дисперсії y при a =", alpha, ": [", dx\_l, ", ", dx\_r, "]\n")

X\_range <- seq(min(selection\_x), max(selection\_x), length.out = 100)

new\_data <- data.frame(X = X\_range)

ci\_y0 <- predict(model,

                 newdata = new\_data,

                 interval = "confidence",

                 level = 1 - alpha)

png("Graphic/lab9\_1\_Regression\_CI\_Plot.png" , width = 800, height = 600)

plot(selection\_x, selection\_y, pch = 16, col = "blue",

     main = "Регресійна пряма та 95% Надійний інтервал для Y0",

     xlab = "Незалежна змінна X",

     ylab = "Залежна змінна Y",

     ylim = range(selection\_y, ci\_y0))

abline(model, col = "red", lwd = 2)

# 2. Надійний інтервал для середнього Y0

lines(X\_range, ci\_y0[, "lwr"], col = "darkgreen", lty = 2, lwd = 2)

lines(X\_range, ci\_y0[, "upr"], col = "darkgreen", lty = 2, lwd = 2)

# 3. Легенда

legend("topleft",

       legend = c("Вибіркові дані",

                  "Регресійна пряма",

                  paste0(100 \* (1 - alpha), "% Надійний інтервал для E[Y|X]")

                  ),

       col = c("blue", "red", "darkgreen"),

       pch = c(16, NA, NA),

       lty = c(NA, 1, 2),

       lwd = c(NA, 2, 2))

dev.off()

# Функція обчислює нижню межу: y = a0\_hat + a1\_hat \* x - [T\_CRIT \* s \* SE\_Factor]

calculate\_lower\_bound <- function(x\_vector, a0, a1, n\_val, x\_bar\_val, SSX\_val, s\_val, t\_crit\_val) {

  SE\_factor <- sqrt((1 / n\_val) + (x\_bar\_val^2 / SSX\_val))

  ME <- t\_crit\_val \* s\_val \* SE\_factor

  Y\_hat <- a0 + a1 \* x\_vector

  lower\_bound <- Y\_hat - ME

  return(lower\_bound)

}

Y\_lower\_bound <- calculate\_lower\_bound(

    x\_vector = X\_range,

    a0 = a0\_hat, a1 = a1\_hat, n\_val = n, x\_bar\_val = dot\_mean\_x,

    SSX\_val = SSX, s\_val = s\_epsilon, t\_crit\_val = T\_CRIT

)

Y\_upper\_bound <- calculate\_lower\_bound(

    x\_vector = X\_range,

    a0 = a0\_hat, a1 = a1\_hat, n\_val = n, x\_bar\_val = dot\_mean\_x,

    SSX\_val = SSX, s\_val = s\_epsilon, t\_crit\_val = -T\_CRIT

)

png("Graphic/lab9\_1\_Regression\_Bound\_Plot.png" , width = 800, height = 600)

plot(selection\_x, selection\_y, pch = 16, col = "blue",

     main = "Регресійна пряма та лінія, обчислена за Вашою формулою",

     xlab = "Незалежна змінна X",

     ylab = "Залежна змінна Y",

     ylim = range(selection\_y, Y\_lower\_bound, Y\_upper\_bound))

abline(model, col = "red", lwd = 2)

lines(X\_range, Y\_lower\_bound, col = "darkorange", lty = 2, lwd = 3)

lines(X\_range, Y\_upper\_bound, col = "darkorange", lty = 2, lwd = 3)

legend("topleft",

       legend = c("Вибіркові дані",

                  "Регресійна пряма",

                  "Межі"

                  ),

       col = c("blue", "red", "darkorange"),

       pch = c(16, NA, NA),

       lty = c(NA, 1, 2),

       lwd = c(NA, 2, 3))

dev.off()

Результат виконання:

Квантиль t-розподілу для alpha = 0.05 та df = 13 : -1.770933

Tочкове значення для a0: 0.9305524

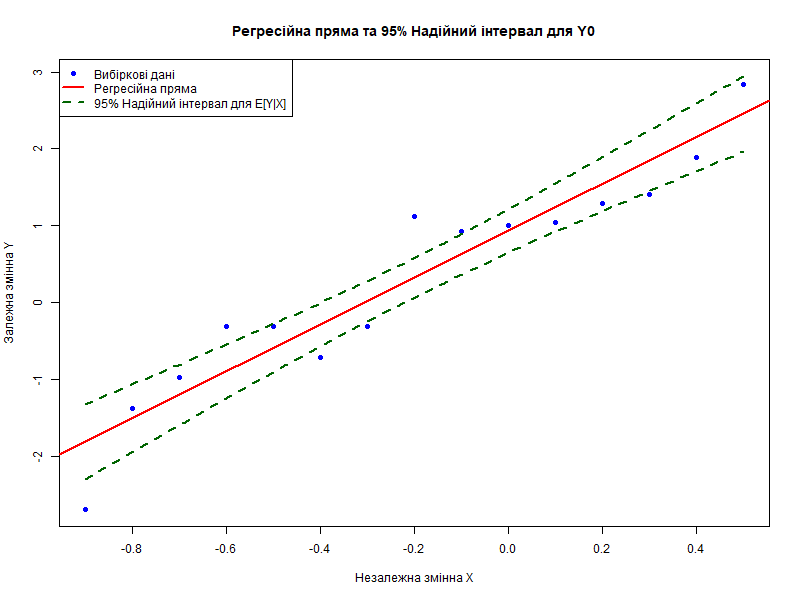
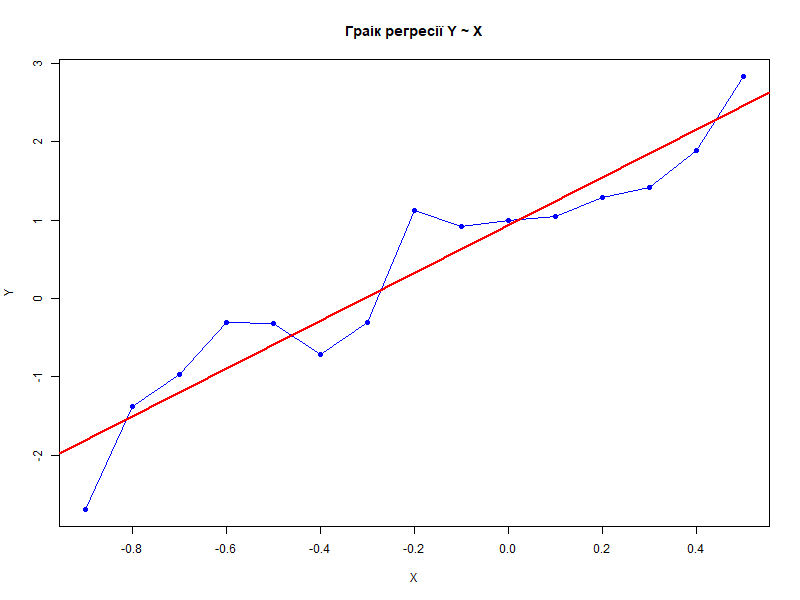
Довірчий інтервал a = 0.05 для a0: [ 1.214727 , 0.6463781 ]

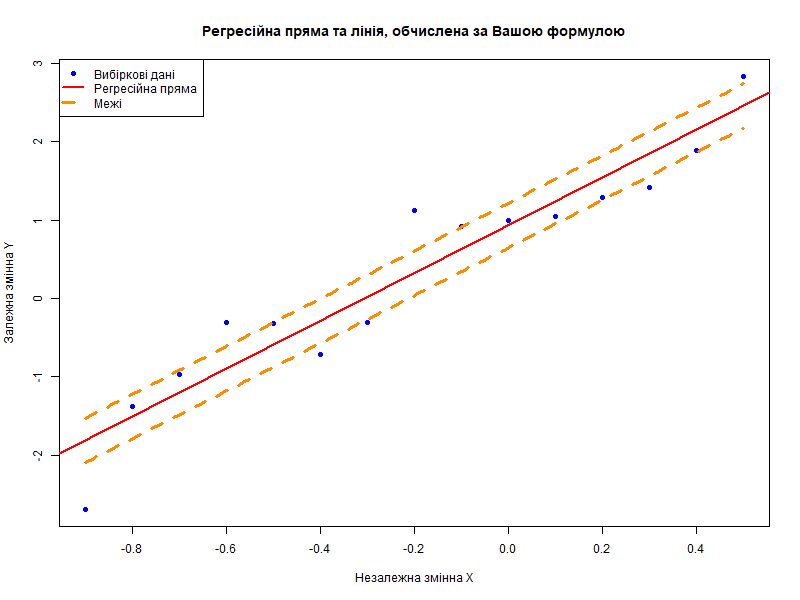
Tочкове значення для a1: 3.047429

Довірчий інтервал a = 0.05 для a1: [ 3.644314 , 2.450543 ]

Квантилі хі-квадрат розподілу для alpha = 0.05 : 5.008751 , 24.7356

Довірчий інтервал для дисперсії y при a = 0.05 : [ 5.335836 , 1.080462 ]





**Висновок:** яознайомився з методикою побудови вибіркового рівняння лінійної регресії та оцінками його параметрів