Чернівецький національний університет імені Юрія Федьковича

Інститут фізико-технічних та комп’ютерних наук

Кафедра Математичних проблем управління і кібернетики

Лабораторна робота № 10

“Тема: Однофакторний дисперсійний аналіз”

з дисципліни

“Інтелектуальний аналіз”

Варіант - 2

Виконав:

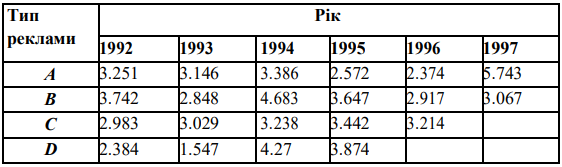
ст. гр. 341СК Гаваньо Дмитро

Прийняв:

професор Дрінь Я. М.

Чернівці – 2025

**Мета:** Ознайомитися з методикою проведення однофакторного дисперсійного аналізу



**Завдання 10.1:** Встановити вплив обсягу продаж товару в залежності від типу реклами, яка використовується, для заданого рівня значущості 𝛼 = 0.05, встановити міру впливу кожного типу реклами. Визначити параметри і побудувати щільність ймовірностей розподілу числа продаж для кожного типу реклами.

1. Визначте і введіть компоненти матриці вибіркових значень випадкової величини 𝑥𝑖𝑗, 𝑖 – рівень фактора, 𝑗 – номер спостереження.
2. Введіть значення кількості факторів і обсяг вибірки для кожного рівня фактора.
3. Обчисліть обсяг вибірки.
4. Обчисліть вибіркове середнє і середні по групах.
5. Обчисліть значення 𝑠12 і 𝑠22 і значення 𝐹𝐻.
6. Для заданого значення 𝛼 перевірте гіпотезу про незалежність кінцевого значення від рівня фактору.
7. Якщо гіпотеза незалежності від фактору відхиляється, обчисліть коефіцієнт детермінації.
8. Запишіть оцінки параметрів розподілу величини, яка досліджується, для кожного фактору.
9. Побудуйте графіки щільності ймовірностей для відповідних розподілів.

**Додаткове завдання:** Перевірте гіпотезу про рівність групових дисперсій для даних завдання 10.1.

**Хід роботи:**

**Завдання 1:**

Програмний код:

setwd("D:/Study/R/lab10")

alpha <- 0.05

data <- read.csv("selection.csv", header = TRUE, sep = ",")

selection\_matrix <- t(as.matrix(data))

c\_length <- nrow(selection\_matrix)

r\_length <- ncol(selection\_matrix)

lengths <- rowSums(!is.na(selection\_matrix))

global\_length <- sum(lengths)

print(selection\_matrix)

cat("Довжини факторів:", lengths, "\n")

row\_dot\_means <- rowMeans(selection\_matrix, na.rm = TRUE)

global\_mean <- mean(selection\_matrix, na.rm = TRUE)

row\_variances <- apply(selection\_matrix, MARGIN = 1, FUN = var, na.rm = TRUE)

df\_b <- c\_length - 1

df\_w <- global\_length - c\_length

s\_sq\_1\_W <- sum(row\_variances \* (lengths - 1))

s\_sq\_2\_B <- sum((row\_dot\_means - global\_mean)^2 \* lengths)

sd\_sq <- sum((selection\_matrix - global\_mean) ^ 2, na.rm = TRUE) # перевірка

cat("Сума квадратів відхилень всередині груп:", round(s\_sq\_1\_W,4), "\n")

cat("Сума квадратів відхилень між групами:", round(s\_sq\_2\_B,4), "\n")

cat("Загальна квадратна дисперсія:", round(sd\_sq,4), "\n")

# F = (SSB / DFB) / (SSW / DFW) => (SSB \* DFW) / (SSW \* DFB)

F\_H <- s\_sq\_2\_B \* df\_w / s\_sq\_1\_W \* df\_b

cat("Значення Фішера:", round(F\_H,4), "\n")

x\_a <- qf(1 - alpha, df\_b, df\_w)

cat("Критичне значення:", round(x\_a,4), "\n")

independence\_lower = F\_H < x\_a

cat("--- Гіпотеза H0: F\_H < x\_a (незаліжність від фактору):", F\_H, "|", x\_a, "\n")

cat("--- Гіпотеза", ifelse(independence\_lower, "доведена | F\_H < x\_a", "відкинута | F\_H >(!!!) x\_a"), "\n")

det\_coef <- s\_sq\_2\_B / sd\_sq

cat("Коефіцієнт детермінації:", round(det\_coef,4), "\n")

cat("Математичне сподівання для кожного фактору:", round(row\_dot\_means,4), "\n")

cat("Дисперсія для кожного фактору:", round(row\_variances,4), "\n")

# Графіки

factors <- rownames(selection\_matrix)

names(row\_dot\_means) <- factors

names(row\_variances) <- factors

sds <- sqrt(row\_variances)

x\_min <- min(row\_dot\_means - 3 \* sds)

x\_max <- max(row\_dot\_means + 3 \* sds)

x\_range <- seq(x\_min, x\_max, length.out = 500)

y\_max <- dnorm(row\_dot\_means["C"], mean = row\_dot\_means["C"], sd = sds["C"]) \* 1.1

png("Graphic/lab10\_1\_Distribution\_Graph.png", width = 800, height = 600)

plot(x = range(x\_range),

     y = c(0, y\_max),

     type = "n",

     main = "Щільність ймовірності (Нормальний розподіл) за факторами",

     xlab = "Значення",

     ylab = "Щільність ймовірності")

colors <- c("blue", "green", "red", "purple")

for (i in 1:length(factors)) {

  factor\_name <- factors[i]

  mean\_val <- row\_dot\_means[factor\_name]

  sd\_val <- sds[factor\_name]

  curve(dnorm(x, mean = mean\_val, sd = sd\_val),

        from = x\_min, to = x\_max,

        col = colors[i], lwd = 2, add = TRUE)

  abline(v = mean\_val, col = colors[i], lty = 2)

}

legend("topright",

       legend = paste0(factors,

                       " (M=", round(row\_dot\_means, 2),

                       ", Var=", round(row\_variances, 2), ")"),

       col = colors,

       lty = 1,

       lwd = 2,

       title = "Фактор")

dev.off()

F\_crit <- qf(1 - alpha, df\_b, df\_w)

png("Graphic/lab10\_1\_F\_distribution.png", width = 800, height = 600)

curve(df(x, df1 = df\_b, df2 = df\_w),

      from = 0, to = F\_crit \* 2,

      col = "darkred", lwd = 2,

      main = paste("F-розподіл (df1 =", df\_b, ", df2 =", df\_w, ")"),

      xlab = "F-статистика",

      ylab = "Щільність ймовірності")

abline(v = F\_crit, col = "blue", lty = 3)

text(x = F\_crit \* 1.3, y = max(df(seq(0, F\_crit \* 1.5, length.out=100), df1=df\_b, df2=df\_w))/2,

     labels = paste("F\_крит:", round(F\_crit, 3)), col = "blue")

dev.off()

cat(" ========== Додатково ==========\n")

selection\_matrix\_no\_na <- selection\_matrix[!is.na(selection\_matrix)]

data\_long <- data.frame(

  Value = c(selection\_matrix\_no\_na),

  Factor = rep(rownames(selection\_matrix), times = lengths)

)

bartlett\_result <- bartlett.test(Value ~ Factor, data = data\_long)

p\_value <- bartlett\_result$p.value

cat("P-значення тесту Бартлетта:", round(p\_value, 4), "\n")

cat("--- Гіпотеза H0: p < alpha (рівність групових дисперсій):", p\_value, "|", alpha, "\n")

if (p\_value < alpha) {

  cat("--- Гіпотеза відхилена | Дисперсії груп статистично нерівні.\n")

} else {

  cat("--- Гіпотеза доведена | Дисперсії груп статистично рівні\n")

}

Результат виконання:

[,1] [,2] [,3] [,4] [,5] [,6]

A 3.251 3.146 3.386 2.572 2.374 5.743

B 3.742 2.848 4.683 3.647 2.917 3.067

C 2.983 3.029 3.238 3.442 3.214 NA

D 2.384 1.547 4.270 3.874 NA NA

Довжини факторів: 6 6 5 4

Сума квадратів відхилень всередині груп: 14.7454

Сума квадратів відхилень між групами: 0.6652

Загальна квадратна дисперсія: 15.4106

Значення Фішера: 2.3008

Критичне значення: 3.1968

--- Гіпотеза H0: F\_H < x\_a (незаліжність від фактору): 2.300801 | 3.196777

--- Гіпотеза доведена | F\_H < x\_a

Коефіцієнт детермінації: 0.0432

Математичне сподівання для кожного фактору: 3.412 3.484 3.1812 3.0187

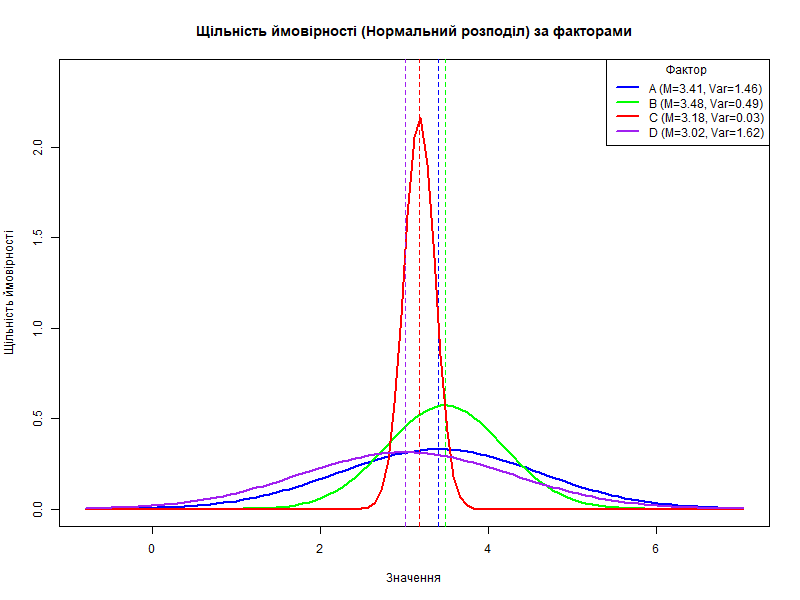
Дисперсія для кожного фактору: 1.4628 0.4861 0.0337 1.622

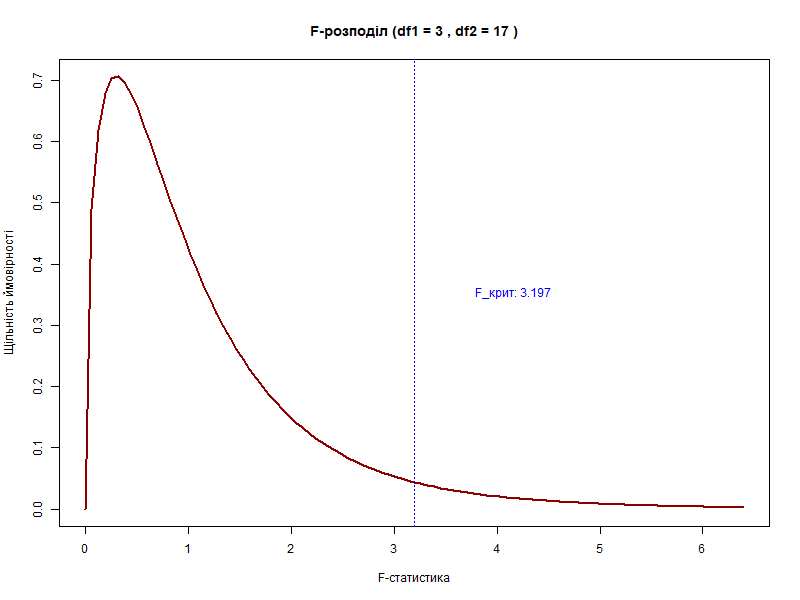
========== Додатково ==========

P-значення тесту Бартлетта: 0.1751

--- Гіпотеза H0: p < alpha (рівність групових дисперсій): 0.1750678 | 0.05

--- Гіпотеза доведена | Дисперсії груп статистично рівні





**Висновок:** яознайомився з методикою проведення однофакторного дисперсійного аналізу.