

Міністерство освіти і науки України
Національний університет «Запорізька політехніка»

Кафедра програмних засобів

ЗВІТ

Дисципліна «Емпіричні методи в інформаційних технологіях»

Робота №3

Тема «Дисперсійний аналіз»

Виконав варіант 19

Студент КНТ-122

Онищенко О. А.

Прийняли

Викладач

Леоценко С. Д.

2024

МЕТА РОБОТИ

Вивчити методи дисперсійного аналізу, провести дослідження ефекту дії одного та двох якісних факторів на одну змінну відгуку відповідно, використовуючи для цього пакети статистичних програм.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

Використовуючи рекомендавну літературу та дані методичні вказівки, вивчити основні положення дисперсійного аналізу (ДА).

Для двохфакторного дисперсійного аналізу, з використанням мови програмування R використати бібліотеку datarium та вибірку jobsatisfaction.

КОД ПРОГРАМИ

```
library(datarium)
data("jobsatisfaction", package="datarium")
str(jobsatisfaction)

library(ggplot2)
ggplot(data=jobsatisfaction, aes(x=gender, y=score)) + geom_boxplot(aes(fill=education_level))

require(doby)
summaryBy(score~gender+education_level, data=jobsatisfaction, FUN=c(mean, sd, length))

plot.design(jobsatisfaction)

with(jobsatisfaction, interaction.plot(x.factor=education_level, trace.factor=gender, response=score))

# weightgain = score
# type = gender
# source = education_level

Model=lm(score~gender*education_level, data=jobsatisfaction)
summary(Model)
anova(Model)
```

РЕЗУЛЬТАТИ ВИКОНАННЯ

Нижче наведено поступовий процес виконання роботи:

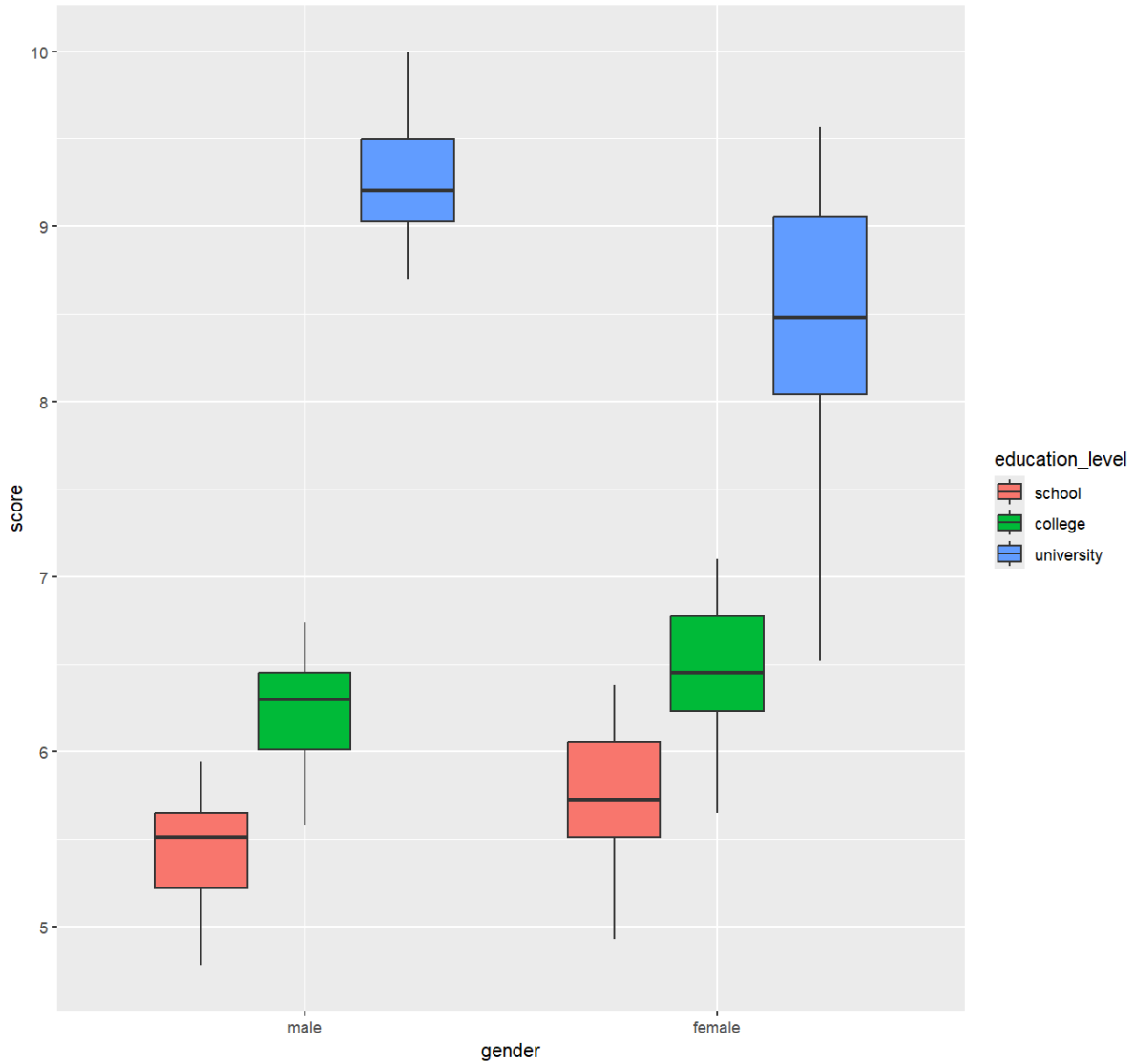


Рисунок 1.1 – Згенерований графік для двох статей на основі рівня освіти

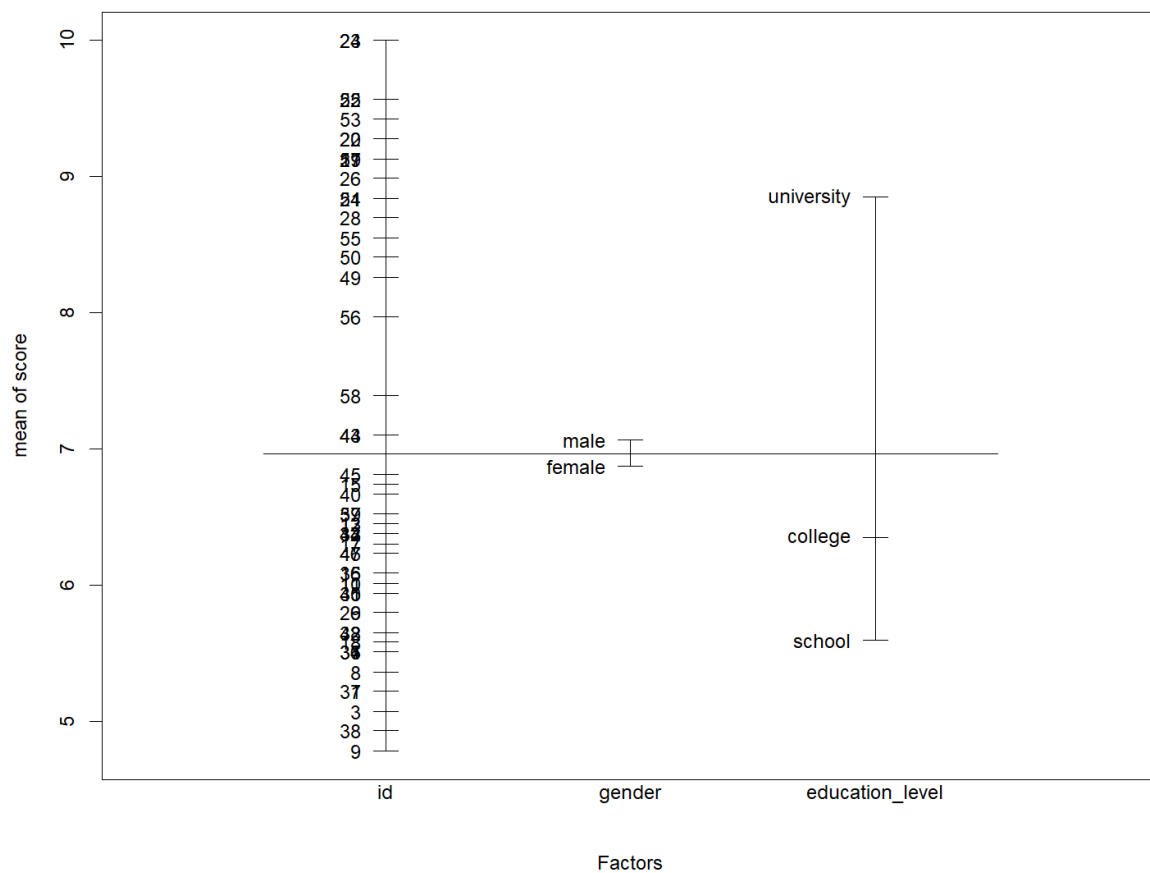


Рисунок 1.2 – Згенерований графік для jobsatisfaction

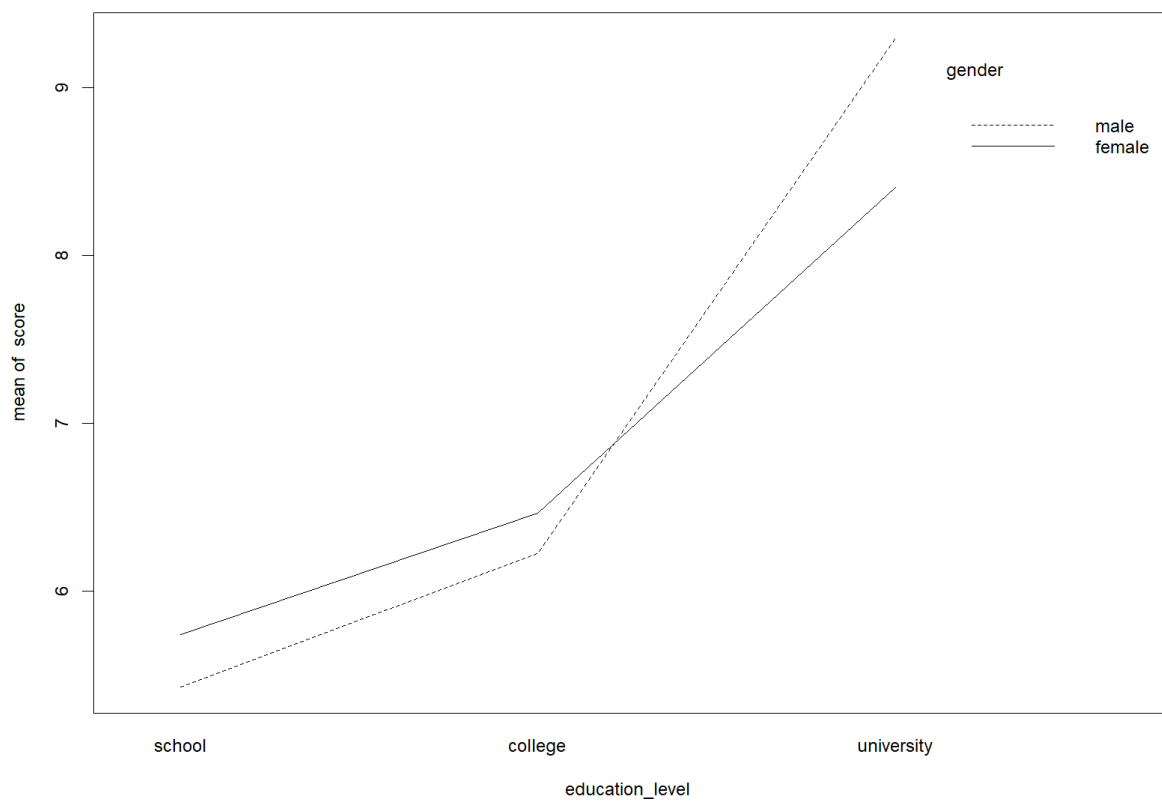


Рисунок 1.3 – Згенерований графік взаємодій

```
> anova(Model)
Analysis of Variance Table

Response: score
              Df Sum Sq Mean Sq  F value    Pr(>F)
gender          1   0.541    0.541    1.7872  0.187090
education_level  2 113.684   56.842  187.8921 < 2.2e-16 ***
gender:education_level  2   4.440    2.220    7.3379  0.001559 **
Residuals      52  15.731    0.303
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Рисунок 1.4 – Результат застосування функції *anova* до моделі

```

> summary(Model)

Call:
lm(formula = score ~ gender * education_level, data = jobsatisfaction)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8860 -0.2325 -0.0040  0.3297  1.1640

Coefficients:
              Estimate Std. Error t value
(Intercept)      5.42667    0.18334  29.599
genderfemale      0.31433    0.25272   1.244
education_levelcollege  0.79667    0.25928   3.073
education_leveluniversity 3.86533    0.25272  15.295
genderfemale:education_levelcollege -0.07467    0.35740  -0.209
genderfemale:education_leveluniversity -1.20033    0.35266  -3.404
Pr(>|t|)
(Intercept)      < 2e-16 ***
genderfemale      0.21915
education_levelcollege  0.00337 **
education_leveluniversity < 2e-16 ***
genderfemale:education_levelcollege  0.83533
genderfemale:education_leveluniversity 0.00129 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.55 on 52 degrees of freedom
Multiple R-squared:  0.8829,    Adjusted R-squared:  0.8717
F-statistic: 78.45 on 5 and 52 DF,  p-value: < 2.2e-16

```

Рисунок 1.5 – Результат застосування функції *summary* до моделі