

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

**Факультет інформаційних технологій**

**Кафедра прикладних інформаційних систем**

напрямок 6.040302 «Інформатика»

(шифр і назва напрямку підготовки або спеціальності)

**Звіт**

**з лабораторної роботи №14**

**На тему: «Проведення дисперсійного аналізу ANOVA»**

**Виконав: студент 4 курсу навчання**

**групи інформатика (І-42)**

**Довбня Дмитро Володимирович**

**Київ – 2017**

**Мета:** Ознайомлення з методикою проведення дисперсійного аналізу методом ANOVA.

***1. Проведення однофакторного дисперсійного аналізу методом у GPSS та MATLAB.***

А) Проведіть аналітично однофакторний дисперсійний аналіз.

Б) Використовуючи загальноцільову систему моделювання GPSS проведіть однофакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу фактору.

Для виконання завдання Вам знадобляться наступні оператори: ANOVA, BEGIN, END, GENERATE, MATRIX, PLUS, PROCEDURE, SAVEVALUE, START, TERMINATE.

В) Використовуючи матричну систему Matlab проведіть однофакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу фактору.

Для виконання завдання Вам знадобиться функція ANOVA.

Г) Порівняйте результати отримані усіма трьома способами.

Вхідні дані:

| Рівні фактора |   |   |
|---------------|---|---|
| 1             | 2 | 3 |
| 6             | 4 | 5 |
| 8             | 5 | 6 |
| 9             | 7 | 5 |
| 6             | 8 | 6 |
| 6             | 8 | 7 |

***2. Проведення двохфакторного дисперсійного аналізу методом у GPSS та MATLAB.***

А) Проведіть аналітично однофакторний дисперсійний аналіз.

Б) Використовуючи загальноцільову систему моделювання GPSS проведіть двохфакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу фактору.

Для виконання завдання Вам знадобляться наступні оператори: ANOVA, BEGIN, END, GENERATE, MATRIX, PLUS, PROCEDURE, SAVEVALUE, START, TERMINATE.

В) Використовуючи матричну систему Matlab проведіть двохфакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу кожного із двох факторів та їх взаємодії.

|          | Фактор А |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|
|          | 1 рівень |          | 2 рівень |          | 3 рівень |          |
| Фактор В | 1 рівень | 2 рівень | 1 рівень | 2 рівень | 1 рівень | 2 рівень |
| 1        | 5        | 2        | 3        | 1        | 4        | 6        |
| 2        | 4        | 3        | 7        | 4        | 5        | 2        |
| 3        | 2        | 5        | 9        | 7        | 4        | 5        |
| 4        | 3        | 7        | 6        | 6        | 7        | 4        |
| 5        | 6        | 8        | 5        | 2        | 3        | 7        |

Г) Порівняйте результати отримані усіма трьома способами.

**Хід виконання:**

### 1. Проведення однофакторного дисперсійного аналізу методом у GPSS та MATLAB.

А) Проводимо аналітичний однофакторний дисперсійний аналіз.

$$\begin{aligned}
 & \text{ORIGIN} := 1 \quad n := 5 \quad k := 3 \quad N := k \cdot n \quad i := 1..k \\
 & x := (6 \ 8 \ 9 \ 6 \ 6 \ 4 \ 5 \ 7 \ 8 \ 8 \ 5 \ 6 \ 5 \ 6 \ 7) \\
 & S_i := \left| \sum_{n=(i-1) \cdot n+1}^{i \cdot n} x^{(n)} \right| \quad S2_i := \frac{\sum_{j=(i-1) \cdot n+1}^{n \cdot i} \left( |x^{(j)}| - X_i \right)^2}{n-1} \quad X_i := \frac{S_i}{n} \\
 & S2Total := \frac{\sum_{i=1}^N \left( |x^{(i)}| - Xtotal \right)^2}{N-1} = 1.971 \quad Xtotal := \frac{\sum X_i}{k} = 6.4 \\
 & DfTotal := N - 1 = 14 \quad SumD2Total := DfTotal \cdot S2Total = 27.6 \\
 & SumD2BG := \sum_{j=1}^k \left[ (X_j - Xtotal) \cdot n \right] - \frac{\left[ \sum_{j=1}^N \left( |x^{(j)}| \right) \right]^2}{N} = 3.6 \quad dfBG := k - 1 = 2 \\
 & SumD2WG := SumD2Total - SumD2BG = 24 \quad dfWG := N - k = 12 \\
 & S2BG := \frac{SumD2BG}{dfBG} = 1.8 \quad S2WG := \frac{SumD2WG}{dfWG} = 2 \quad F := \frac{S2BG}{S2WG} = 0.9 \quad Fk := 3.88
 \end{aligned}$$

Б) Використовуючи GPSS проводимо двохфакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу фактору.

```
Rez MATRIX ,1,3,5
Rez_Residuals TABLE ,-5,.5,20
GENERATE ,,,1
PLUS (FillMatrix())
SAVEVALUE StErr,(ANOVA(Rez,3,2))
TERMINATE 1
START 1
PROCEDURE FillMatrix() BEGIN
Rez[1,1,1]=6;Rez[1,1,2]=8;Rez[1,1,3]=9;Rez[1,1,4]=6;Rez[1,1,5]=6;
Rez[1,2,1]=4;Rez[1,2,2]=5;Rez[1,2,3]=7;Rez[1,2,4]=8;Rez[1,2,5]=8;
Rez[1,3,1]=5;Rez[1,3,2]=6;Rez[1,3,3]=5;Rez[1,3,4]=6;Rez[1,3,5]=7;
END;
```

#### ANOVA

| Source of Variance | Sum of Squares | Degrees of Freedom | Mean Square | F     | Critical Value of F (p=.05) |
|--------------------|----------------|--------------------|-------------|-------|-----------------------------|
| A                  | 0.000          | 0                  |             |       |                             |
| B                  | 3.600          | 2                  | 1.800       | 0.900 | 3.89                        |
| AB                 | 0.000          | 0                  |             |       |                             |
| Error              | 24.000         | 12                 | 2.000       |       |                             |
| Total              | 27.600         | 14                 |             |       |                             |

  

| Treatment Level | Count | Mean  | Minimum | Maximum | 95% C.I. (SE)    |
|-----------------|-------|-------|---------|---------|------------------|
| A B             |       |       |         |         |                  |
| 1 1             | 5     | 7.000 | 6.000   | 9.000   | [ 5.622, 8.378 ] |
| 1 2             | 5     | 6.400 | 4.000   | 8.000   | [ 5.022, 7.778 ] |
| 1 3             | 5     | 5.800 | 5.000   | 7.000   | [ 4.422, 7.178 ] |

В) Використовуючи Matlab проводимо двохфакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу кожного із двох факторів та їх взаємодії.

```
a =
```

```

6     4     5
8     5     6
9     7     5
6     8     6
6     8     7
```

```
>> anova(a)
```

```
ans =
```

```
0.4323
```

| Source  | SS   | df | MS  | F   | Prob>F |
|---------|------|----|-----|-----|--------|
| Columns | 3.6  | 2  | 1.8 | 0.9 | 0.4323 |
| Error   | 24   | 12 | 2   |     |        |
| Total   | 27.6 | 14 |     |     |        |

Г) Порівняння результатів отриманих трьома способами.

|       | Sum of Squares |      |      | Degrees of Freed |      |    | Mean Square |      |     | F   |      |     | Fk<br>(p=0.05) |      | ML<br>(Prob>F) |
|-------|----------------|------|------|------------------|------|----|-------------|------|-----|-----|------|-----|----------------|------|----------------|
|       | A              | gpss | ML   | A                | gpss | ML | A           | gpss | ML  | A   | gpss | ML  | A              | gpss |                |
| A     |                | 0    |      |                  | 0    |    |             |      |     |     |      |     |                |      |                |
| B     | 3.6            | 3.6  | 3.6  | 2                | 2    | 2  | 1.8         | 1.8  | 1.8 | 0.9 | 0.9  | 0.9 | 3.88           | 3.89 | 0.4323         |
| AB    |                | 0    |      |                  | 0    |    |             |      |     |     |      |     |                |      |                |
| Error | 24             | 24   | 24   | 12               | 12   | 12 |             |      |     |     |      |     |                |      |                |
| Total | 27.6           | 27.6 | 27.6 | 14               | 14   | 14 |             |      |     |     |      |     |                |      |                |

Висновок: Провівши двофакторний дисперсійний аналіз даних методом

ANOVA та аналітичним методом, ми отримали однакові результати.

Оскільки  $p > 0.05$  ми можемо говорити про статистичну незначущість впливу фактора B

## 2. Проводимо двофакторний дисперсійний аналізу методом у GPSS та MATLAB.

A) Проведіть аналітично однофакторний дисперсійний аналіз.

$$\begin{aligned}
 & \text{ORIGIN} := 1 \quad x := (5 \ 4 \ 2 \ 3 \ 6 \ 2 \ 3 \ 5 \ 7 \ 8 \ 3 \ 7 \ 9 \ 6 \ 5 \ 1 \ 4 \ 7 \ 6 \ 2 \ 4 \ 5 \ 4 \ 7 \ 3 \ 6 \ 2 \ 5 \ 4 \ 7) \\
 & n := 5 \quad k := 6 \quad \text{CountA} := 3 \quad i := 1..6 \quad \text{CountB} := 2 \quad N := 30 \quad N_a := 10 \quad a := 1..3 \quad b := 1..2 \\
 & S_a := \sum_{n=(i-1) \cdot 5 + 1}^{i \cdot 5} x^{(n)} \quad N_b := 15 \quad N_t := 30 \\
 & S_{a_a} := S_{a \cdot 2 - 1} + S_{a \cdot 2} \quad S_{b_b} := S_b + S_{b+2} + S_{b+4} \quad X_t := \frac{\sum S_{a_a}}{N_t} = 4.733 \\
 & S2_{\text{Total}} := \frac{\sum_{i=1}^{N_t} (|x^{(i)}| - X_t)^2}{N_t - 1} = 4.133 \\
 & Df_{\text{Total}} := N_t - 1 = 29 \quad \text{SumD2}_{\text{Total}} := Df_{\text{Total}} \cdot S2_{\text{Total}} = 119.867 \\
 & \text{SumD2}_{\text{BG}} := \sum_{j=1}^k \left[ \frac{(S_j \cdot S_j)}{n} \right] - \frac{\left[ \sum_{j=1}^{N_t} (|x^{(j)}|) \right]^2}{N_t} = 13.867 \\
 & df_{\text{BG}} := \text{CountA} \cdot \text{CountB} - 1 = 5 \quad S2_{\text{BG}} := \frac{\text{SumD2}_{\text{BG}}}{df_{\text{BG}}} = 2.773 \\
 & df_{\text{WG}} := N - \text{CountA} \cdot \text{CountB} = 24 \\
 & \text{SumD2}_{\text{WG}} := \text{SumD2}_{\text{Total}} - \text{SumD2}_{\text{BG}} = 106 \\
 & S2_{\text{WG}} := \frac{\text{SumD2}_{\text{WG}}}{df_{\text{WG}}} = 4.417 \\
 & \text{SumD2}_A := \frac{\left[ \sum_a (S_{a_a})^2 \right]}{n \cdot \text{CountB}} - \frac{\left[ \sum_{j=1}^{N_t} (|x^{(j)}|) \right]^2}{N_t} = 1.267 \\
 & df_A := \text{CountA} - 1 = 2 \quad S2_A := \frac{\text{SumD2}_A}{df_A} = 0.633 \\
 & \text{SumD2}_B := \frac{\left[ \sum_b (S_{b_b})^2 \right]}{n \cdot \text{CountA}} - \frac{\left[ \sum_{j=1}^{N_t} (|x^{(j)}|) \right]^2}{N_t} = 0.533 \\
 & df_B := \text{CountB} - 1 = 1 \quad S2_B := \frac{\text{SumD2}_B}{df_B} = 0.533 \\
 & df_{\text{AB}} := df_A \cdot df_B = 2 \quad \text{SumD2}_{\text{AB}} := \text{SumD2}_{\text{BG}} - \text{SumD2}_A - \text{SumD2}_B = 12.067 \\
 & S2_{\text{AB}} := \frac{\text{SumD2}_{\text{AB}}}{df_{\text{AB}}} = 6.033
 \end{aligned}$$

### Критерії Фішера

$$F_a := \frac{S^2_A}{S^2_{WG}} = 0.143 \quad df_A = 2 \quad FK_a := 3.4$$

$$F_b := \frac{S^2_B}{S^2_{WG}} = 0.121 \quad df_B = 1 \quad FK_b := 4.26$$

$$F_{ab} := \frac{S^2_{AB}}{S^2_{WG}} = 1.366 \quad df_{AB} = 2 \quad FK_{ab} := 3.4$$

$$F_{bg} := \frac{S^2_{BG}}{S^2_{WG}} = 0.628 \quad df_{BG} = 5 \quad FK_{bg} := 2.62$$

$$df_{WG} = 24$$

Б) Використовуючи GPSS проводимо двохфакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу фактору.

```
Rez MATRIX ,3,2,5
Rez_Residuals TABLE ,-5,.5,20
GENERATE ,,,1
PLUS (FillMatrix())
SAVEVALUE StErr,(ANOVA(Rez,3,2))
TERMINATE 1
START 1
```

```
PROCEDURE FillMatrix() BEGIN
Rez[1,1,1]=5;Rez[1,1,2]=4;Rez[1,1,3]=2;Rez[1,1,4]=3;Rez[1,1,5]=6;
Rez[1,2,1]=2;Rez[1,2,2]=3;Rez[1,2,3]=5;Rez[1,2,4]=7;Rez[1,2,5]=8;

Rez[2,1,1]=3;Rez[2,1,2]=7;Rez[2,1,3]=9;Rez[2,1,4]=6;Rez[2,1,5]=5;
Rez[2,2,1]=1;Rez[2,2,2]=4;Rez[2,2,3]=7;Rez[2,2,4]=6;Rez[2,2,5]=2;

Rez[3,1,1]=4;Rez[3,1,2]=5;Rez[3,1,3]=4;Rez[3,1,4]=7;Rez[3,1,5]=3;
Rez[3,2,1]=6;Rez[3,2,2]=2;Rez[3,2,3]=5;Rez[3,2,4]=4;Rez[3,2,5]=7;
END;
```

### ANOVA

| Source of Variance | Sum of Squares | Degrees of Freedom | Mean Square | F     | Critical Value of F (p=.05) |
|--------------------|----------------|--------------------|-------------|-------|-----------------------------|
| A                  | 1.267          | 2                  | 0.633       | 0.143 | 3.40                        |
| B                  | 0.533          | 1                  | 0.533       | 0.121 | 4.26                        |
| AB                 | 12.067         | 2                  | 6.033       | 1.366 | 3.40                        |
| Error              | 106.000        | 24                 | 4.417       |       |                             |
| Total              | 119.867        | 29                 |             |       |                             |

  

| Treatment Level | Count | Mean  | Minimum | Maximum | 95% C.I. (SE)    |
|-----------------|-------|-------|---------|---------|------------------|
| A B             |       |       |         |         |                  |
| 1 1             | 5     | 4.000 | 2.000   | 6.000   | ( 2.026, 5.974 ) |
| 1 2             | 5     | 5.000 | 2.000   | 8.000   | ( 3.026, 6.974 ) |
| 2 1             | 5     | 6.000 | 3.000   | 9.000   | ( 4.026, 7.974 ) |
| 2 2             | 5     | 4.000 | 1.000   | 7.000   | ( 2.026, 5.974 ) |
| 3 1             | 5     | 4.600 | 3.000   | 7.000   | ( 2.626, 6.574 ) |
| 3 2             | 5     | 4.800 | 2.000   | 7.000   | ( 2.826, 6.774 ) |

В) Використовуючи Matlab проводимо двохфакторний дисперсійний аналіз даних методом ANOVA та визначить рівень впливу кожного із двох факторів та їх взаємодії.

```
>> b'

ans =

     5     4     2     3     6     3     7     9     6     5     4     5     4     7     3
     2     3     5     7     8     1     4     7     6     2     6     2     5     4     7

>> anova2(b,5)

ans =

    0.7312    0.8671    0.2742
```

| Source      | SS      | df | MS      | F    | Prob>F |
|-------------|---------|----|---------|------|--------|
| Columns     | 0.533   | 1  | 0.53333 | 0.12 | 0.7312 |
| Rows        | 1.267   | 2  | 0.63333 | 0.14 | 0.8671 |
| Interaction | 12.067  | 2  | 6.03333 | 1.37 | 0.2742 |
| Error       | 106     | 24 | 4.41667 |      |        |
| Total       | 119.867 | 29 |         |      |        |

Г) Порівняння результатів отриманих трьома способами.

|       | Sum of Squares |         |         | Degrees of Freed |      |    | Mean Square |       |       | F     |       |       | Fk (p=0.05) |      | p ML   |
|-------|----------------|---------|---------|------------------|------|----|-------------|-------|-------|-------|-------|-------|-------------|------|--------|
|       | A              | gpss    | ML      | A                | gpss | ML | A           | gpss  | ML    | A     | gpss  | ML    | A           | gpss |        |
| A     | 1.267          | 1.267   | 1.267   | 2                | 2    | 2  | 0.633       | 0.633 | 0.633 | 0.143 | 0.143 | 0.143 | 3.4         | 3.4  | 0.8671 |
| B     | 0.533          | 0.533   | 0.533   | 1                | 1    | 1  | 0.533       | 0.533 | 0.533 | 0.121 | 0.121 | 0.121 | 4.26        | 4.26 | 0.7312 |
| AB    | 12.067         | 12.067  | 12.067  | 2                | 2    | 2  | 6.033       | 6.033 | 6.033 | 1.366 | 1.366 | 1.366 | 3.4         | 3.4  | 0.2742 |
| Error | 106            | 106     | 106     | 24               | 24   | 24 | 4.417       | 4.417 | 4.417 |       |       |       |             |      |        |
| Total | 119.867        | 119.867 | 119.867 | 29               | 29   | 29 |             |       |       |       |       |       |             |      |        |

Висновок: Провівши двофакторний дисперсійний аналіз даних методом ANOVA та аналітичним методом, ми отримали результати які не містять відхилень.

Оскільки для фактора А та В  $p > 0.05$  ми можемо говорити про статистичну незначущість впливу факторів А та В на результати експерименту не залежно один від одного. Взаємодія факторів АВ також має  $p > 0.05$  тому можна зробити висновок що вплив взаємодії факторів А та В також не має статистичної значущості.