

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра програмних систем і технологій

Лабораторна робота 1
з дисципліни
«Ймовірнісні основи програмної інженерії»

Тема. «Центральні тенденції та міра дисперсії»

| | | | |
|----------------|---|----------------|-----------------------------------|
| Виконав: | Вітковський Володимир Олександрович | Оцінка | |
| Група | ІПЗ-24 | в балах | за національною шкалою |
| Форма навчання | денна | | |
| Спеціальність | Інженерія програмного забезпечення | | |
| Перевірив: | Вечерковська Анастасія Сергіївна | | |
| Дата перевірки | | | |

Мета

Навчитись використовувати на практиці набуті знання про центральні тенденції та міри.

Завдання

1. Побудувати таблицю частот та сукупних частот для переглянутих фільмів. Визначити фільм, який був переглянутий частіше за інші.
2. Знайти Моду та Медіану заданої вибірки.
3. Порахувати Дисперсію та Середнє квадратичне відхилення розподілу.
4. Побудувати гістограму частот для даного розподілу.
5. Зробити висновок з вигляду гістограми, про закон розподілу.

Технічне завдання

1.

Підключення бібліотек. Зчитування файлів із вхідними даними. Виведення таблиці частот та запис її.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4
5 filename = input("Введіть назву вхідного файла - ")
6 data = np.genfromtxt(filename, dtype='str')
7 empty_list = []
8 num0 = 0
9 datasort = sorted(data)
10 for ele in datasort:
11     if (ele not in empty_list):
12         num0 += 1
13         empty_list.append(ele)
14 print(num0)
15
16 months_num = [0 for i in range(num0)]
17 freq = [1 for i in range(num0)]
18 numm0 = 0
19 num1 = -1
20 for i in range(len(datasort)):
21     for j in range(num0):
22         if (datasort[i] == months_num[j]):
23             freq[num1] += 1
24             break
25         if (j == num0 - 1):
26             months_num[numm0] = datasort[i]
27             numm0 += 1
28             num1 += 1
29 print(months_num)
30 print(freq)
31 count = [(months_num[i], freq[i]) for i in range(num0)]
```

```
32
33 # ----- Ex.1 -----
34 totalcount = 0
35 f = open("result.txt", "w+")
36 f.write("\n    Фільм | Частота | Сукупна частота ")
37 print("    Фільм | Частота | Сукупна частота ")
38 for i in range(len(count)):
39     totalcount += count[i][1]
40     f.write("\n\t " + str(count[i][0]) + " \t|  " + str(count[i][1]) + "   |\t" + str(totalcount))
41     print("\t " + str(count[i][0]) + " \t|  " + str(count[i][1])+"   |\t" + str(totalcount))
42     f.write("\n")
43
44 totalcount = [0 for i in range(len(count))]
```

Результат виконання.

| Фільм | Частота | Сукупна частота |
|-------|---------|-----------------|
| 1 | 3 | 3 |
| 10 | 2 | 5 |
| 12 | 2 | 7 |
| 66 | 2 | 9 |
| 75 | 1 | 10 |
| 97 | 1 | 11 |

2.

Розрахування та виведення моди та медіани.

```
46 # ---- Ex.2 ----
47 data = sorted(data)
48 moda = max(set(data), key=data.count)
49 pmoda = 'Мода - ' + moda
50 print(pmoda)
51
52 medianaen = 'Медіана - ' + str(data[int(len(data)/2)])
53 print(medianaen)
```

Використані формули для моди і медіани:

$$Mo = x_0 + h \frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})}; \quad \tilde{x} = \frac{(x_k + x_{k+1})}{2}$$

Результат:

```
Мода - 1
Медіана - 12
```

3.

Розрахування та виведення дисперсії й середнє квадратичне відхилення розподілу.

```
55 # ---- Ex.3 ----
56 sumX = 0
57 for i in range(int(len(freq))):
58     sumX += int(months_num[i]) * freq[i]
59 sumX = sumX/int(len(data))
60 averagesum = "Середнє значення - "+str(round(sumX, 3))
61 print(averagesum)
62
63 varX = 0
64 for i in range(int(len(data))):
65     varX += (int(data[i]) - sumX)**2
66 varX = varX/(int(len(data))-1)
67 disperesion = "Дисперсія - " + str(round(varX, 3))
68 print(disperesion)
69 varX = np.sqrt(varX)
70 averagequad = "Середнє квадратичне відхилення - " + str(round(varX, 3))
71 print(averagequad)
72
73
74 f.write(pmoda + "\n" + medianaen + "\n" + disperesion + "\n" + averagequad)
75 f.close()
```

Використані формули для дисперсії й середнє квадратичне відхилення розподілу.

Середнє квадратичне відхилення розподілу

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n).$$

$$S_0 = \sqrt{\frac{n}{n-1} S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

Дисперсія

$$S_0 = \left(\sqrt{\frac{n}{n-1} S^2} \right)^2$$

Результат:

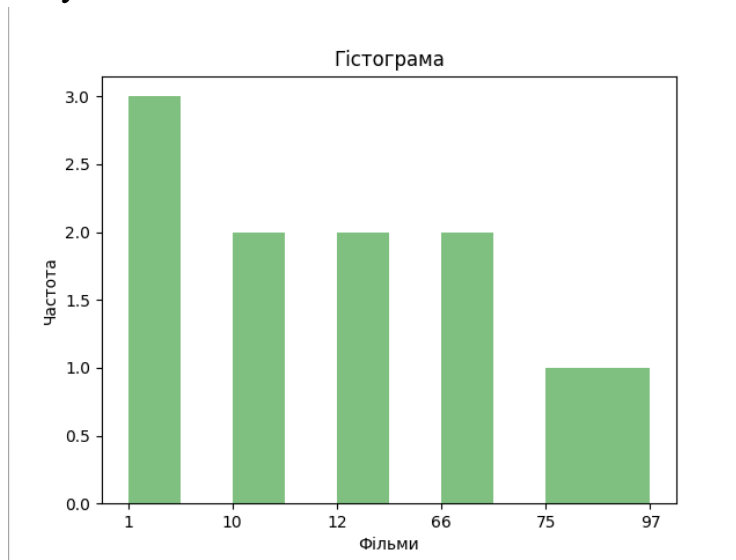
```
Дисперсія - 1303.691
Середнє квадратичне відхилення - 36.107
```

4.

Гістограма частот.

```
77 # ---- Ex.4 ----
78 f, ax = plt.subplots()
79 plt.hist(data, facecolor='green', align='mid', alpha=0.5)
80 plt.xlabel("Фільми")
81 plt.ylabel("Частота")
82 plt.title("Гістограма")
83 plt.show()
```

Результат:



Виведення та збереження результатів у текстовому файлі.

| Фільм | Частота | Сукупна частота |
|-------|---------|-----------------|
| 1 | 3 | 3 |
| 10 | 2 | 5 |
| 12 | 2 | 7 |
| 66 | 2 | 9 |
| 75 | 1 | 10 |
| 97 | 1 | 11 |

Мода - 1
 Медіана - 12
 Дисперсія - 1303.691
 Середнє квадратичне відхилення - 36.107

5.

Створивши та відобразивши гістограму частот, я дійшов висновку, що гістограма є графічним зображенням залежності частоти попадання елементів вибірки від відповідного інтервалу групування, а також, способом графічного представлення табличних даних та приблизним представлення розподілу числових даних.

Висновок

Навчився використовувати на практиці набуті знання про центральні тенденції та міри, а також, будувати таблицю частот та сукупних частот. Здобув навички знаходження моди та медіани заданої вибірки. Навчився рахувати дисперсію та середнє квадратичне відхилення розподілу. Крім того, здобув знання побудови гістограми частот.