

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

Дисципліна
«Комп'ютерна графіка та візуалізація»

Лабораторна робота № 1
«Центральні тенденції та міра дисперсії»

Виконав:	Бойко Костянтин Богданович	Перевірила:	
Група	ІПЗ-24(1)	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		

2022

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.

Формат вхідних/вихідних даних

Розроблена програма повинна зчитувати вхідні дані з файлу заданого формату та записувати дані у файл.

Вхідний файл представляє собою текстовий файл із $M+1$ рядків. Перший рядок містить число: M , де M —кількість переглядів i -того фільму.

До документа завдання також додаються приклад вхідних файлів різної розмірності.

Хід роботи

1. Побудувати таблицю частот та сукупних частот для переглянутих фільмів. Визначити фільм, який був переглянутий частіше за інші.

Для виконання лабораторної роботи будемо використовувати мову програмування Python(версія 3.10) та бібліотеки, numpy – для зчитування вхідних даних з txt файлу, вбудовану, collections – для групування вхідних даних та matplotlib – для графічної побудови.

Вхідні дані беремо з наданих файлів до лабораторної роботи за допомогою коду, де виводиться назва .txt файлу та вхідні дані зберігаються у змінну

```
4   FileName = input("Enter name of .txt file: ")
5   Data = np.genfromtxt(FileName, dtype='str')
```

Таблиця частот будуюмо за допомогою відповідного коду. Вона складається з стовпчиків: Назва і кількість переглядів, Частота повторень відповідних даних та Сукупна частота (зображує к-сть переглядів помножену на к-сть повторень(частоту)).

```

34 f = open("result.txt", "w+")
35 f.write("\n Name & Views | Count | Total count")
36 f.write("\n -----+-----+-----")
37 print("Name & Views| Count | Total Count ")
38 print("-----+-----+-----")
39 for i in range(len(count)):
40     totalcount += count[i][1]
41     f.write("\n\t " + str(count[i][0]) + " \t| " + str(count[i][1]) + " \t| " + str(totalcount))
42     print("\t " + str(count[i][0]) + " \t| " + str(count[i][1]) + " \t| " + str(totalcount))
43 f.write("\n")

```

```

Enter name of .txt file: input_10.txt
Name & Views| Count | Total Count
-----+-----+-----
      1      |    3    |    3
     10      |    2    |    5
     12      |    2    |    7
     66      |    2    |    9
     75      |    1    |   10
     97      |    1    |   11

```

2. Знайти моду та медіану

Фільм з найбільшою кількістю переглядів визначаємо за допомогою коду:

```

46 for i in range(int(len(freq))):
47     if (freq[i] == max(freq)):
48         maxnumber = ('Number = ' + str(NameOfFilms[i]) + ' Freq = ' + str(freq[i]))
49     Moda = 'Mode of entry data: ' + maxnumber
50 print(Moda)

```

```
Mode of entry data: Number = 1 Freq = 3
```

Моду розраховували за формулою:

$$Mo = x_0 + h \frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})};$$

Медіану розраховували за допомогою коду:

```
53 Mediana = 'Mediana of entry data = ' + str(datasort[int(len(Data) / 2)])
54 print(Mediana)
```

Результат:

```
Mediana of entry data = 12
```

Для розрахунку використовували таку формулу:

$$\tilde{x} = \frac{(x_k + x_{k+1})}{2}$$

3. Розраховуємо Дисперсію та Середнє квадратичне відхилення розподілу

За допомогою наведеного коду ми розраховуємо дисперсію та середнє квадратичне відхилення розподілу

```
56 sumX = 0
57 for i in range(int(len(freq))):
58     sumX += int(NameOfFilms[i]) * freq[i]
59 sumX = sumX/int(len(Data))
60 AverageSum = "Average value: " + 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 disp = "Dispersion: " + str(round(varX, 3))
68 print(disp)
69 varX = np.sqrt(varX)
70 averagequad = "Mean square deviation: " + str(round(varX, 3))
71 print(averagequad)
```

Ось який результат буде виводитись:

```
Average value: 31.909
Dispersion: 1303.691
Mean square deviation: 36.107
```

Обчислювали за допомогою формул:

Середнє арифметичне вибірки

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

Середньоквадратичне відхилення

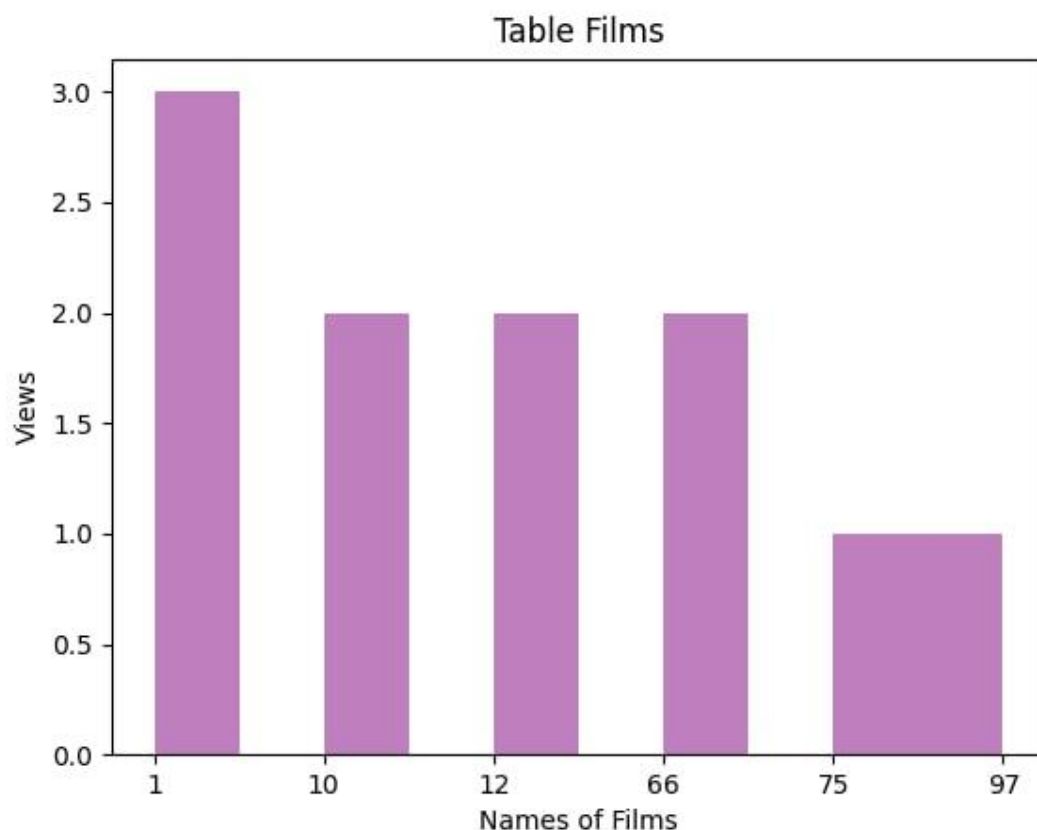
$$S_0 = \sqrt{\frac{n}{n-1} S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

Дисперсія буде дорівнювати прикориневому виразу, зображеному вище

4. Побудувати гістограму частот для даного розподілу

За віссю x у нас знаходяться назви фільмів, у вигляді номерів, а за віссю y - кількість переглядів. Ось як виглядає код та самий вигляд гістограми:

```
78 f, ax = plt.subplots()
79 plt.hist(sorted(Data), bins=int(10), facecolor='purple', align='mid', alpha=0.5)
80 plt.ylabel("Views")
81 plt.xlabel("Names of Films")
82 plt.title("Table Films")
83 plt.show()
```



Усі дані, після виконання коду, додаються до текстового документу та зберігаються.



result — Блокнот

Файл Правка Формат Вид Справка

Name & Views	Count	Total count
1	3	3
10	2	5
12	2	7
66	2	9
75	1	10
97	1	11

Name & Views	Count	Total count
1	3	3
10	2	5
12	2	7
66	2	9
75	1	10
97	1	11

Mode of entry data: Number = 1 Freq = 3

Mediana of entry data = 1

Average value: 31.909

Dispersion: 1303.691

Mean square deviation: 36.107

5. Зробити висновок з вигляду гістограми, про закон розподілу.

Проаналізувавши отримані гістограми під час виконання лабораторної роботи, можна прийти до висновку, що закон розподілу характеризує випадкову величину з точки зору теорії ймовірностей. Розподіл ймовірностей тісно зв'язаний з рядами розподілу частот. Якщо розглядати ряди розподілу (користуючись термінологією теорії ймовірностей) як перелік можливих результатів або груп вимірів і відповідних їм частот кожного результату, то аналогічне визначення можна дати і розподілу ймовірностей. Це перелік можливих результатів або груп вимірів, але, замість спостережуваної частоти, тут вказані ймовірності появи кожного результату.

Висновок:

Під час виконання першої лабораторної роботи було опановано методи обчислення деяких статистичних даних, використано на практиці набуті знання про центральні тенденції та міри, отримано нові навички з побудови діаграм, проаналізовано виконану роботу та повторено використання бібліотек у мові програмування Python.