

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра програмних систем і технологій

Дисципліна
«Ймовірнісні основи програмної інженерії»

Лабораторна робота № 2
Центральні тенденції та міра дисперсії

Виконав:	Зарожевський Євген Сергійович	Перевірила:	Марцафей А.С.
Група	ІПЗ-22	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

Мета роботи:

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.

Побудова математичної моделі:

Для визначення фільму, який був переглянутий частіше за інші, потрібно побудувати таблицю частот і знайти там фільм з найбільшою частотою.

Для визначення моди використовувалось означення:

Мода – елемент з найбільшою частотою.

Для визначення медіани використовувалась формула:

$$Median = \begin{cases} x_{\frac{n+1}{2}}, & x \% 2 == 1 \\ \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}, & x \% 2 == 0 \end{cases}$$

Для визначення дисперсії використовувалась формула:

$$Var(x) = \frac{\sum_{i=1}^n (f_i * x_i^2)}{n} - Mean^2,$$

$$Де Mean = \frac{\sum_{i=1}^n x_i}{n};$$

Для визначення середнього квадратичного відхилення використовувалась формула:

$$s_x = \sqrt{Var(x)}$$

Для побудови гістограми частот використовувалась бібліотека matplotlib мови програмування python.

Код алгоритму:

```
import math
import matplotlib.pyplot as plt

fileName = input("Enter name of file without extension: ")

f = open(f'/Users/zarozhenya/Desktop/knu/йопі/lab1_python/{fileName}.txt', "r")

quantityOfViews = int(f.readline())
data = []
dictionary = {}
for i in range(quantityOfViews):
    data.append(int(f.readline()))

f.close()
data.sort()

for i in range(quantityOfViews):
    if data[i] in dictionary:
        dictionary[data[i]] += 1
    else:
        dictionary[data[i]] = 1

f = open("/Users/zarozhenya/Desktop/knu/йопі/lab1_python/output.txt", "w")

f.write("xi\tfi\tFi\n\n")
cumulativeFrequency = 0
maxValue = 0
total = 0
variance = 0
maxMovies = []

for i in dictionary:
    total += i * dictionary[i]
    variance += dictionary[i] * (i ** 2)
    cumulativeFrequency += dictionary[i]
    f.write(f'{i}\t{dictionary[i]}\t{cumulativeFrequency}\n')

for i in dictionary:
    if dictionary[i] > maxValue:
        maxValue = dictionary[i]
        maxMovies = [i]
        continue

    if dictionary[i] == maxValue:
        maxMovies.append(i)

mean = total / cumulativeFrequency
variance /= cumulativeFrequency
variance -= mean ** 2

f.write(f'\nMovie with max views: {int(max(data))}\n')

if len(data) % 2 == 0:
    median = int((data[int(len(data) / 2) - 1] + data[int(len(data) / 2)]) / 2)
```

```

else:
    median = data[int((len(data) + 1) / 2) - 1]

f.write(f'Median: {median}\n')
f.write(f'Mode: {maxMovies} - {maxValue}\n')

f.write(f'Variance: {int(variance)}\n')
f.write(f'Standard deviation: {int(math.sqrt(variance))}\n')

plt.hist(data, rwidth=0.5)
plt.savefig('hist.png')

```

Випробування алгоритму:

Результат роботи програми при вхідних даних розміром 10:

xi	fi	Fi
----	----	----

1	3	3
---	---	---

10	1	4
----	---	---

12	2	6
----	---	---

66	2	8
----	---	---

75	1	9
----	---	---

97	1	10
----	---	----

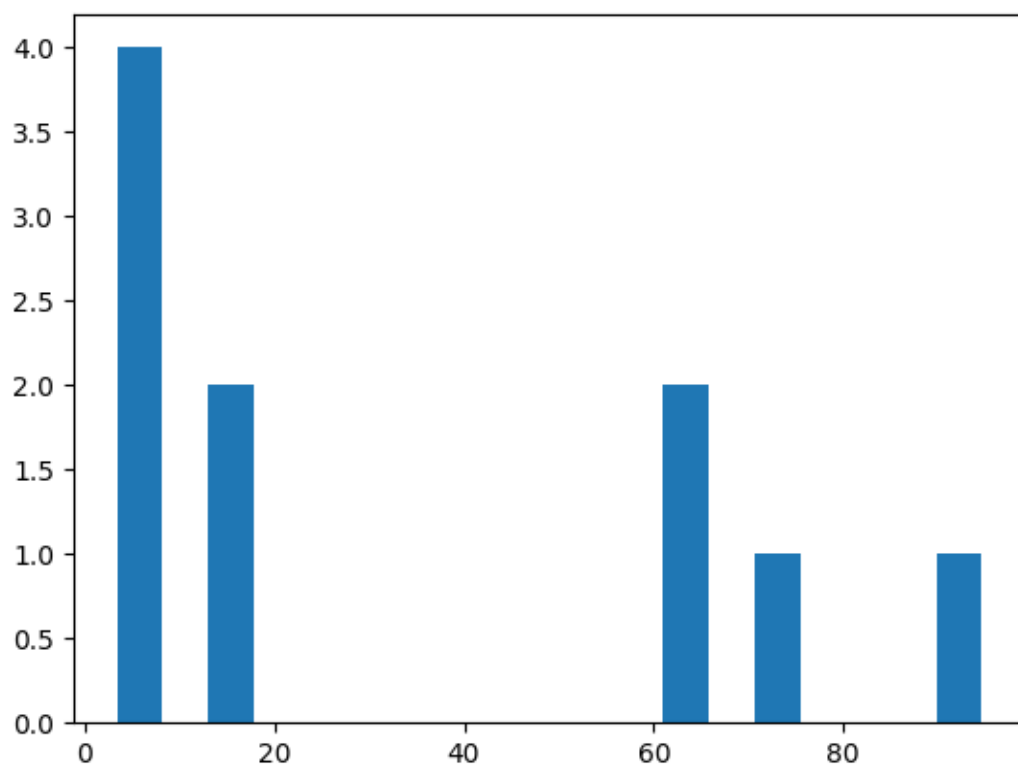
Movie with max views: 97

Median: 12

Mode: [1] - 3

Variance: 1250

Standard deviation: 35



Результат роботи програми при вхідних даних розміром 100:

```
736 1 76
738 1 77
763 1 78
768 1 79
775 1 80
777 1 81
782 1 82
784 1 83
813 1 84
817 1 85
820 1 86
821 1 87
824 1 88
832 1 89
834 1 90
858 1 91
878 1 92
879 1 93
880 1 94
923 1 95
928 1 96
945 1 97
976 1 98
984 1 99
999 1 100
```

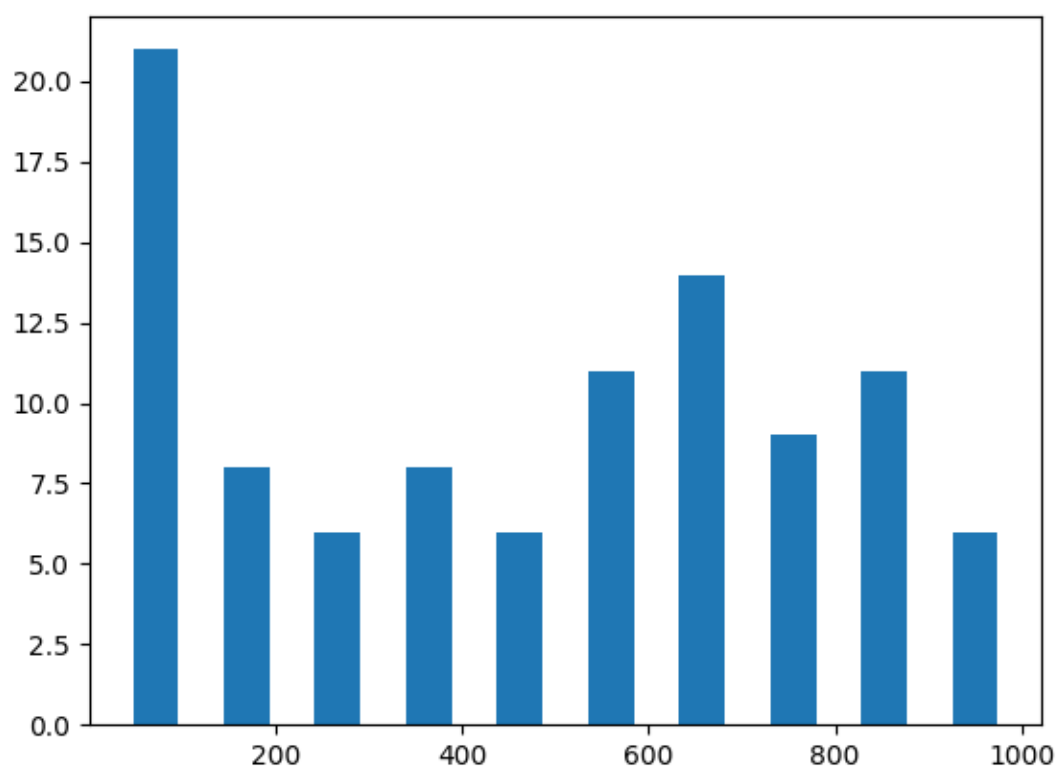
Movie with max views: 999

Median: 532

Mode: [22] - 4

Variance: 89012

Standard deviation: 298



Результат роботи програми при вхідних даних розміром 1000:

```
77871 1 775
97990 1 976
97991 1 977
98046 1 978
98228 1 979
98249 1 980
98345 1 981
98419 1 982
98521 1 983
98668 1 984
98728 1 985
98790 1 986
99024 1 987
99172 1 988
99189 1 989
99193 1 990
99246 1 991
99256 1 992
99272 1 993
99403 1 994
99575 1 995
99696 1 996
99808 1 997
99820 1 998
99968 1 999
99970 1 1000
```

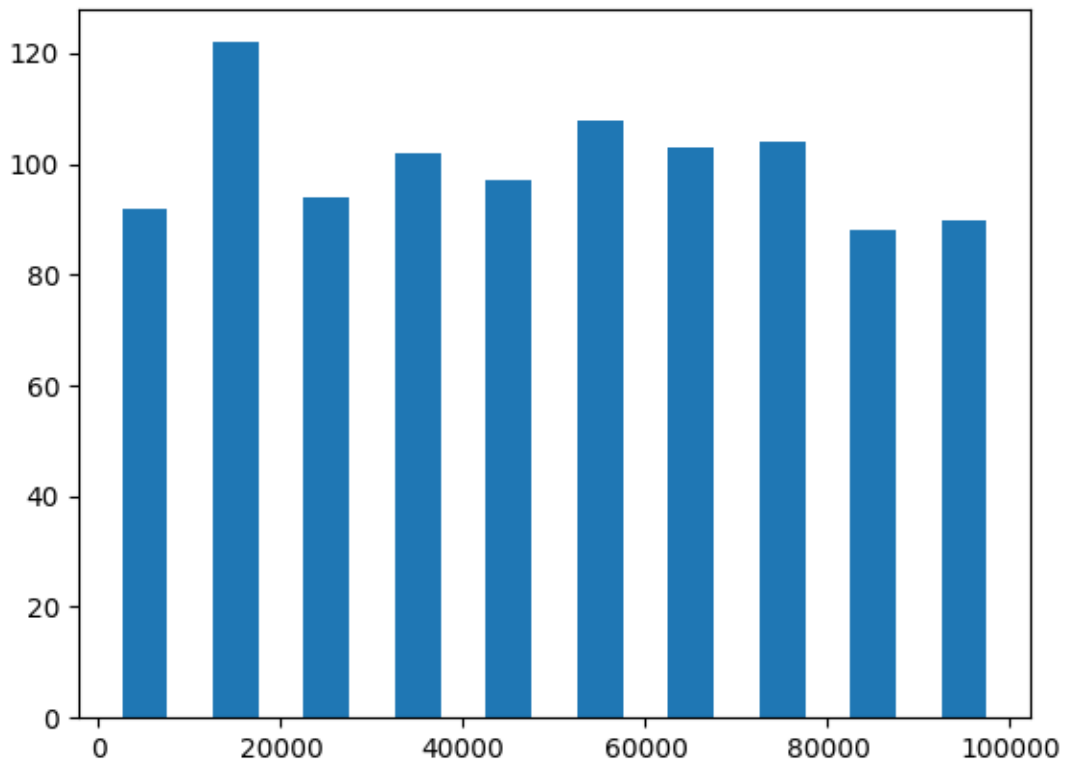
Movie with max views: 99970

Median: 49335

Mode: [14023, 40617, 93548] - 2

Variance: 801811586

Standard deviation: 28316



Висновок:

Виконано завдання першої лабораторної роботи. Опановані навички використання на практиці знань про центральні тенденції та міри. Розроблена програма будує таблицю та гістограму частот, знаходить моду, медіану, дисперсію та середнє квадратичне відхилення для розподілу незалежно від його об'єму (тестувались розподіли розміром 10, 100 та 1000).