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 № 1

«Центральні тенденції та міра дисперсії»

Виконала:	Манойлова Катерина Борисівна	**Перевірила**:	Вечерковська Анастасія Сергіївна
Група	ІПЗ-21	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		

2022

Тема: центральні тенденції та міра дисперсії.
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.
Розроблена програма повинна зчитувати вхідні дані з файлу заданого формату та записувати дані у файл.

Математична модель:

Середнє значення вибірки рахується за формулою $\frac{\sum_{i=1}^{n} x_i}{n}$

Частота значень визначається кількістю повторів даного значення у вибірці.
Сукупна частота розраховується як сума сукупної частоти попереднього елемента з частотою теперішнього.
Медіана вибірки є її середнім елементом, якщо кількість елементів непарна, або обчислюється за формулою $\frac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}$, якщо кількість елементів є парною.
Проте, оскільки індексація масивів розпочинається з нуля, при реалізації у коді ця формула набуде вигляду $\frac{x_{\frac{n}{2}}+x_{\frac{n}{2}-1}}{2}$
Мода вибірки є числом, що має найбільшу частоту. Якщо таку частоту мають декілька елементів, вони обидва є модою. Якщо кожен елемент вибірки зустрічається лише один раз, моди немає. Ці властивості треба передбачити у коді.

Дисперсія вибірки рахується за формулою $\frac{\sum_{x\in X} f_x(x-\bar{x})^2}{\sum_{x\in X} f_x}$, де $\bar{x}$ – середнє значення вибірки, f_x – частота поточного елементу.
Середнє квадратичне відхилення вибірки рахується як корінь з дисперсії.

Гістограма вимагає розбиття значень на інтервали.
Було вирішено зробити інтервали розміром 4% від максимального значення, забезпечивши водночас детальність та видимість.
У вибірці з 10 елементів дуже великий розбіг значень елементів, що разом з невеликою кількістю робить побудову гістограми за інтервалами незручним, отже для неї було вирішено додатково побудувати стовпчасту діаграму, яка вказує частоту для кожного значення.

Псевдокод алгоритмів:

Знаходження медіани:

```
median():
    if к-ть елементів парна:
        med = (arr[розмір/2]+arr[розмір/2-1])/2
    else:
        med = (arr[розмір/2])
```

Знаходження частоти кожного елементу:

```
countFr():
    for el in arr:
        if el in freqarr:
            частота елемента += 1
        else:
            додаємо елемент з частотою 1
```

Знаходження сукупної частоти:

```
countCuFr():
    fr = 0
    for el in freqarr:
        fr += частота
        сукупна частота = fr
```

Знаходження моди:

```
findMod():
    if max==1:
        Немає моди
    else:
        for el in freqarr:
            if частота==max:
                Мода = el
```

Знаходження середнього абсолютного відхилення:

```
MAD():
    sum = 0
    for el in freqarr:
        sum += частота*|el-mid|
    mad = sum/к-ть елементів
```

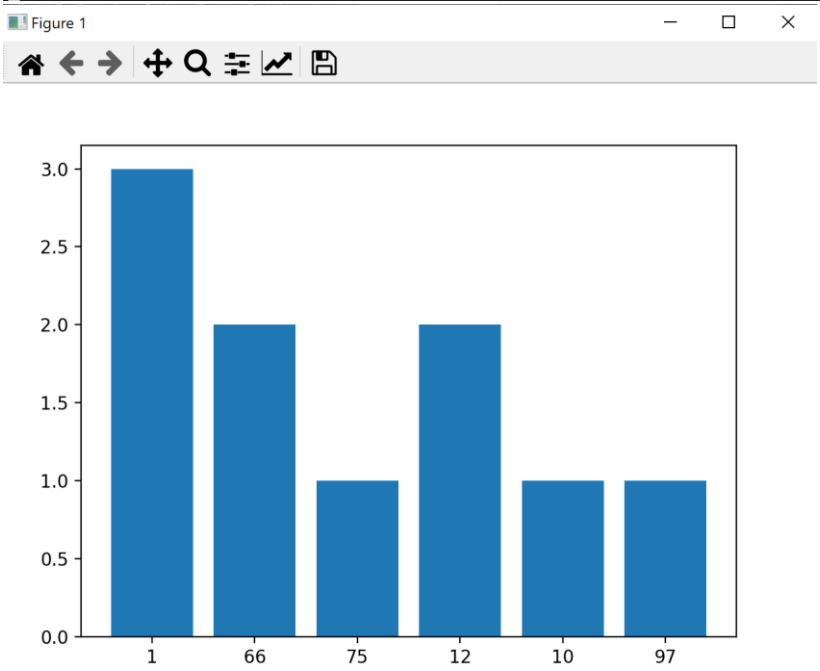
Знаходження дисперсії:

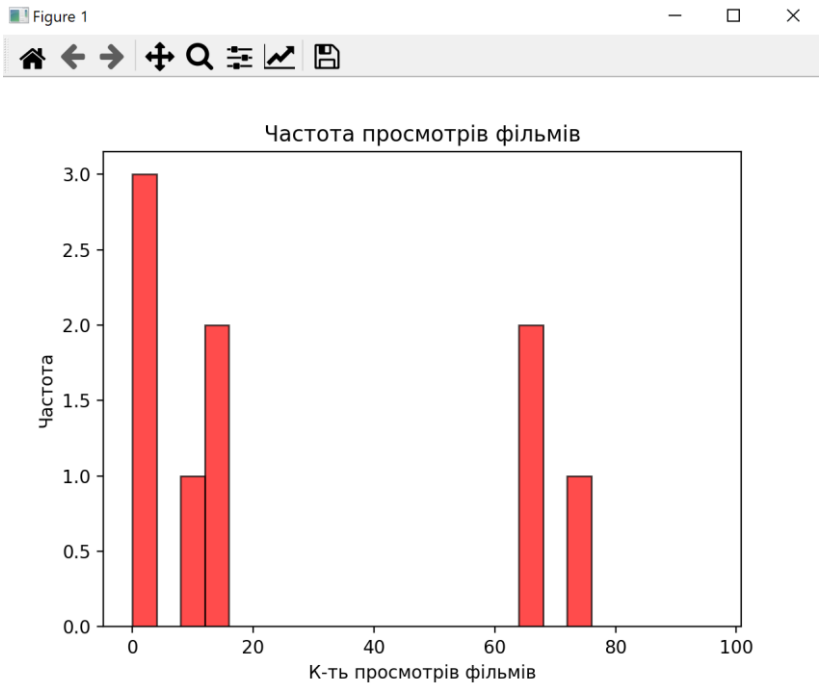
```
dispersion():
    sum = 0
    for el in freqarr:
        sum += частота*(el-mid)**2
    disp = sum/ к-ть елементів
```

Випробування алгоритму

10 елементів:

```
Введіть кількість елементів файлу (10, 100, 1000):  10
Список елементів:  1 66 75 1 1 12 10 97 12 66
Кількість елементів:  10
Найбільше переглядів  97  у фільма за індексом  7
Медіана:  6.5
Таблиця частот:
Елемент | Частота | Сукупна частота
1       |  3      |      3
66      |  2      |      5
75      |  1      |      6
12      |  2      |      8
10      |  1      |      9
97      |  1      |      10
Мода:  1
Середнє абсолютне відхилення =  33.52
Дисперсія =  1250.8899999999999
Середнє квадратичне (стандартне) відхилення =  35.367923320432595
--- час виконання: 0.1568608283996582 секунд ---
```





100 елементів:

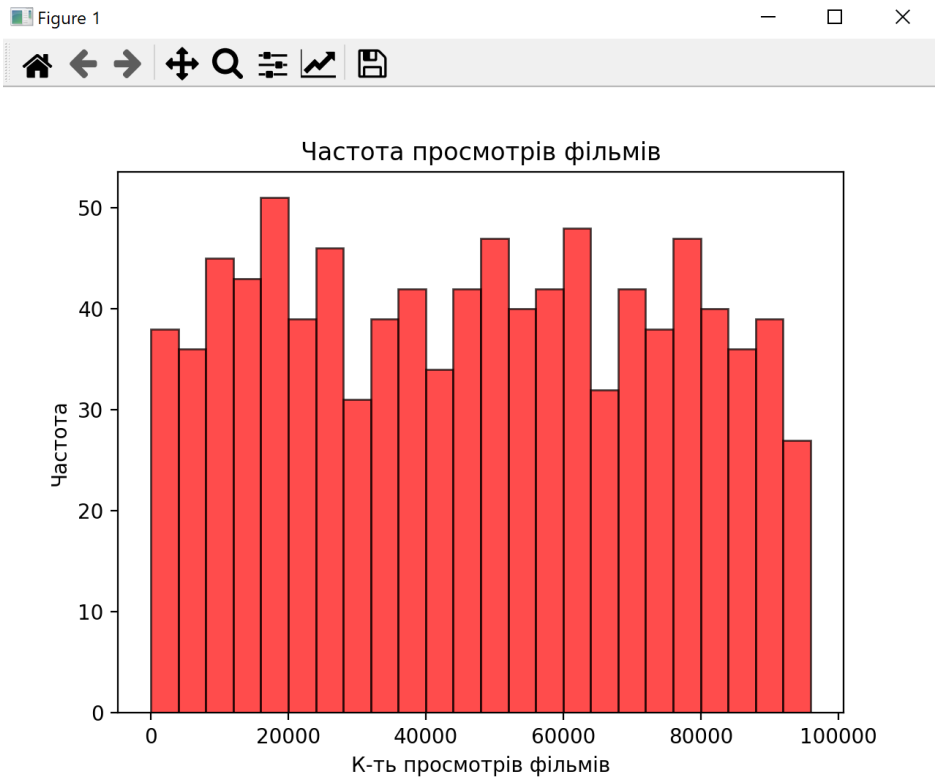
```
Введіть кількість елементів файлу (10, 100, 1000):  100
Список елементів:  642 51 97 529 46 999 317 99 880 46 79 548 361 821 71 288 51 255 429 80 657 22 817 168 688 858 162 587 775 51 566 738 763 83
2 447 414 784 355 154 251 660 250 813 382 694 613 923 362 687 571 103 79 535 162 193 198 607 91 928 676 569 503 945 777 269 47 615 685 225 22
824 553 589 22 976 384 702 612 878 820 77 834 147 879 119 736 768 146 707 450 498 119 636 612 359 984 782 22 354 607
Кількість елементів:  100
Найбільше переглядів  999  у фільма за індексом  5
Медіана:  337.0
Таблиця частот:
Елемент | Частота | Сукупна частота
642     | 1       |     1
51      | 3       |     4
97      | 1       |     5
529     | 1       |     6
46      | 2       |     8
999     | 1       |     9
317     | 1       |     10
99      | 1       |     11
880     | 1       |     12
79      | 2       |     14
548     | 1       |     15
361     | 1       |     16
821     | 1       |     17
71      | 1       |     18
288     | 1       |     19
255     | 1       |     20
429     | 1       |     21
80      | 1       |     22
657     | 1       |     23
22      | 4       |     27
817     | 1       |     28
168     | 1       |     29
688     | 1       |     30
858     | 1       |     31
162     | 2       |     33
587     | 1       |     34
775     | 1       |     35
566     | 1       |     36
738     | 1       |     37
763     | 1       |     38
763     | 1       |     38
832     | 1       |     39
447     | 1       |     40
414     | 1       |     41
784     | 1       |     42
355     | 1       |     43
154     | 1       |     44
251     | 1       |     45
660     | 1       |     46
250     | 1       |     47
813     | 1       |     48
382     | 1       |     49
694     | 1       |     50
613     | 1       |     51
923     | 1       |     52
362     | 1       |     53
687     | 1       |     54
571     | 1       |     55
103     | 1       |     56
535     | 1       |     57
193     | 1       |     58
198     | 1       |     59
607     | 2       |     61
91      | 1       |     62
928     | 1       |     63
676     | 1       |     64
569     | 1       |     65
503     | 1       |     66
945     | 1       |     67
777     | 1       |     68
269     | 1       |     69
47      | 1       |     70
615     | 1       |     71
685     | 1       |     72
225     | 1       |     73
824     | 1       |     74
553     | 1       |     75
589     | 1       |     76
976     | 1       |     77
976     | 1       |     77
384     | 1       |     78
702     | 1       |     79
612     | 2       |     81
878     | 1       |     82
820     | 1       |     83
77      | 1       |     84
834     | 1       |     85
147     | 1       |     86
879     | 1       |     87
119     | 2       |     89
736     | 1       |     90
768     | 1       |     91
146     | 1       |     92
707     | 1       |     93
450     | 1       |     94
498     | 1       |     95
636     | 1       |     96
359     | 1       |     97
984     | 1       |     98
782     | 1       |     99
354     | 1       |     100
Мода:  22
Середнє абсолютне відхилення =  265.76519999999977
Дисперсія =  89012.62360000004
Середнє квадратичне (стандартне) відхилення =  298.34983425502355
Інтервали значень:  [0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, 390, 420, 450, 480, 510, 540, 570, 600, 630, 660, 690, 720, 7
50, 780, 810, 840, 870, 900, 930, 960, 990]
--- час виконання: 0.31969690322875977 секунд ---
```

1000 елементів:

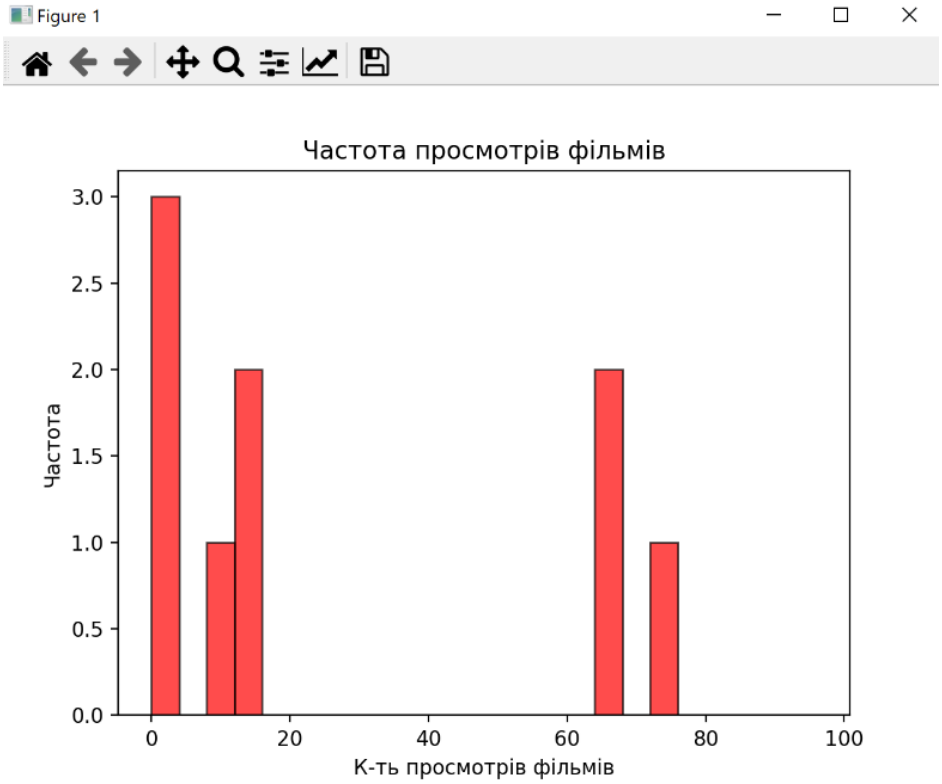
```
Введіть кількість елементів файлу (10, 100, 1000):  1000

Кількість елементів:  1000
Найбільше переглядів  99970  у фільма за індексом  924
Медіана:  35071.0
Таблиця частот:
Елемент | Частота | Сукупна частота
54831   | 1       |      1
76418   | 1       |      2
57391   | 1       |      3
58530   | 1       |      4
48009   | 1       |      5
53988   | 1       |      6
60817   | 1       |      7
36476   | 1       |      8
18316   | 1       |      9
84352   | 1       |      10
20598   | 1       |      11
51505   | 1       |      12
32702   | 1       |      13
79484   | 1       |      14
46431   | 1       |      15
55168   | 1       |      16
5646    | 1       |      17
32180   | 1       |      18
94377   | 1       |      19
51172   | 1       |      20
98419   | 1       |      21
51768   | 1       |      22
80007   | 1       |      23
98249   | 1       |      24
```

*Частину виведених елементів таблиці пропущено

```
3008    | 1       |      998
64327   | 1       |      999
69601   | 1       |      1000
Мода:  14023
Мода:  40617
Мода:  93548
Середнє абсолютне відхилення =  24455.196224
Дисперсія =  801811586.6903838
Середнє квадратичне (стандартне) відхилення =  28316.27776898623
Інтервали значень:  [0, 2999, 5998, 8997, 11996, 14995, 17994, 20993, 23992, 26991, 29990, 32989, 35988, 38987, 41986, 44985, 47984, 50983, 53
982, 56981, 59980, 62979, 65978, 68977, 71976, 74975, 77974, 80973, 83972, 86971, 89970, 92969, 95968, 98967]
--- час виконання: 0.4591035842895508 секунд ---
```

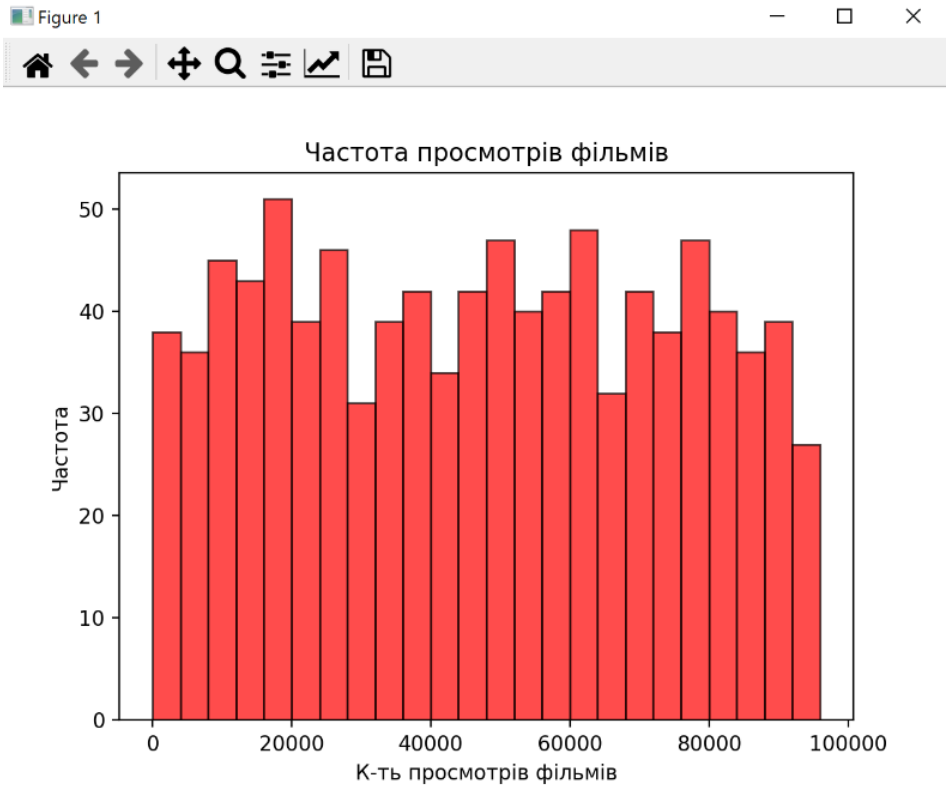


Аналіз гістограм:



Дана гістограма має невеликий кластер в інтервалі [8; 16), та три прогалини в інтервалах [4; 8), [16; 64) та [68; 72).
Вона не є симетричною.

Дана гістограма не має прогалин при розмірі інтервалів 40 просмотрів. Не є симетричною. Найменшу частоту мають інтервали [200; 240) та [720; 760), найбільшу - [40; 80)



Дана гістограма не має прогалин при розмірі інтервалів 40 просмотрів. Не є симетричною. Найбільшу частоту має інтервал [16000; 20000), найменшу - [92000; 96000)

Висновок

Під час виконання лабораторної роботи було розроблено програму зчитування та аналізу даних з документу та запису вихідних даних. Було побудовано алгоритми знаходження медіани, частоти елементів заданої вибірки, сукупної частоти, моди вибірки, середнього абсолютного відхилення, дисперсії та стандартного відхилення. Було створено алгоритм розбиття вибірки на інтервали заданого розміру та побудовано гістограми на основі цих даних.