

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

Дисципліна

«Ймовірнісні основи програмної інженерії»

Звіт з лабораторної роботи № 1

на тему:

«Центральні тенденції на міра дисперсії»

Виконала:	Дрозд Єлизавета Андріївна	Перевірила:	Марцафей А. С.
Група	ІПЗ-12(2)	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

Мета роботи:

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

Постановка задачі:

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.

Побудова математичної моделі:

1. При побудові таблиці частот будемо використовувати такі визначення:

Нехай

x_1, x_2, \dots, x_k – елементи, що вивчаються, які розташовані в порядку зростання, і f_1, f_2, \dots, f_k – відповідно частоти появи цих елементів.

$N = \sum_{i=1}^k f_i$ називається розміром даних.

$F_j = \sum_{i=1}^k f_j$ називається сукупною частотою x_j

$\frac{f_i}{N}$ – відносна частота x_i

2. Для знаходження медіани нам необхідні такі формули:

Suppose $x_1, x_2, x_3, \dots, x_n$ is a set of numeric data arranged in increasing or decreasing order. The *median* \tilde{x} of this set is defined as:

$$\tilde{x} = \begin{cases} \frac{x_{\frac{n+1}{2}}}{2} & \text{if } n \text{ is odd} \\ \frac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ is even} \end{cases}$$

Мода – це елемент, що зустрічається найчастіше у вибірці. Якщо кожен елемент вибірки зустрічається по одному разу, то моди не існує. Якщо декілька елементів мають однакову частоту, яка більше ніж усі інші частоти, то усі ці елементи є модами.

3. Для розрахунку дисперсії та середнього квадратичного відхилення розподілу необхідні наступні формули:

➤ The *variance* of a set of data X is defined as:

$$\text{Var}(X) = \frac{\sum_{x \in X} f_x (x - \bar{x})^2}{\sum_{x \in X} f_x} \quad \text{where } f_x \text{ is the frequency of occurrence of } x.$$

➤ The *standard deviation* of a set of data X is given by:

$$\sigma = \sqrt{\text{Var}(X)}$$

Псевдокод алгоритму:

```
1  from copy import copy
2  import math
3  import matplotlib.pyplot as plt
4
5  s = input('Enter a file name: ')
6  f1 = open(s, 'r')
7  f2 = open('output.txt', 'w')
8  items = f1.read().split()
9  items = [int(i) for i in items]
10 items.sort()
11 input_val = copy(items)
12
13
14 def table():
15     cf = 0
16     j = 0
17     while j < 2:
18         for i in items:
19             if items.count(i) > 1:
20                 items.remove(i)
21             j = j + 1
22     f2.write('Item\tFrequency\tRelative frequency\tCumulative frequency\n' + '-' * 68 + '\n')
23     for i in items:
24         cf += input_val.count(i)
25         f2.write('%s\t%s\t\t%s\t\t%s\n' % (i, input_val.count(i), input_val.count(i) / input_val.__len__(), cf))
26     f2.write('\nThe most viewed film - %s' % max(items))
```

```

29 def moda():
30     n = 0
31     mo = 0
32     for i in items:
33         if input_val.count(i) > n and input_val.count(i) > mo and input_val.count(i) != 1:
34             mo = i
35             n = input_val.count(i)
36     if mo == 0:
37         f2.write('\nNo mode')
38     else:
39         f2.write('\nModa = %s' % mo)
40
41
42 def med():
43     i = int((len(input_val) + 1) / 2)
44     m = input_val[i - 1]
45     f2.write('\nMediana = %s' % m)
46
47
48 def variance():
49     v = 0
50     sr = sum(input_val) / len(input_val)
51     i = 0
52     while i < len(input_val):
53         v += pow(input_val[i] - sr, 2)
54         i = i + 1
55     v /= len(input_val)
56     sd = math.sqrt(v)
57     f2.write('\nVariance = %s\nStandart deviation = %s' % (v, sd))

```

```

60 table()
61 moda()
62 med()
63 variance()
64 plt.hist(input_val)
65 plt.savefig('hist.png')
66 f1.close()
67 f2.close()

```

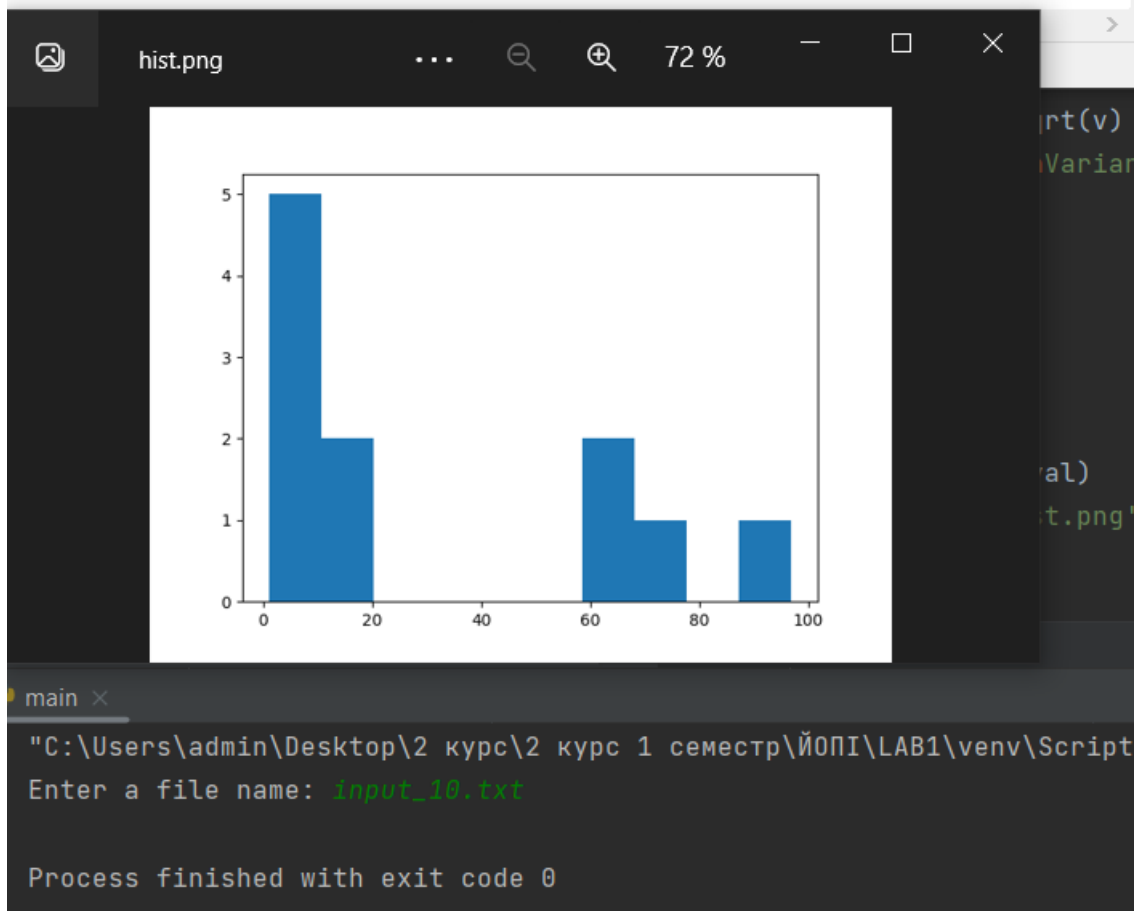
Випробування алгоритму:

Результат роботи програми при введенні даних із файлу input_10.txt:

output.txt: Блокнот

Item	Frequency	Relative frequency	Cumulative frequency
1	3	0.2727272727272727	3
10	2	0.18181818181818182	5
12	2	0.18181818181818182	7
66	2	0.18181818181818182	9
75	1	0.09090909090909091	10
97	1	0.09090909090909091	11

The most viewed film - 97
Moda = 1
Mediana = 12
Variance = 1185.1735537190084
Standart deviation = 34.42634970075986



Результат роботи програми при введенні даних із файлу input_100.txt:

output.txt: Блокнот

— □ ×

Файл	Редагування	Формат	Вигляд	Довідка
Item	Frequency	Relative frequency	Cumulative frequency	
22	4	0.039603960396039604	4	
46	2	0.019801980198019802	6	
47	1	0.009900990099009901	7	
51	3	0.0297029702970297	10	
71	1	0.009900990099009901	11	
77	1	0.009900990099009901	12	
79	2	0.019801980198019802	14	
80	1	0.009900990099009901	15	
91	1	0.009900990099009901	16	
97	1	0.009900990099009901	17	
99	1	0.009900990099009901	18	
100	1	0.009900990099009901	19	
103	1	0.009900990099009901	20	
119	2	0.019801980198019802	22	
146	1	0.009900990099009901	23	
147	1	0.009900990099009901	24	
154	1	0.009900990099009901	25	
162	2	0.019801980198019802	27	
168	1	0.009900990099009901	28	
193	1	0.009900990099009901	29	
198	1	0.009900990099009901	30	
225	1	0.009900990099009901	31	
250	1	0.009900990099009901	32	
251	1	0.009900990099009901	33	
255	1	0.009900990099009901	34	
269	1	0.009900990099009901	35	
288	1	0.009900990099009901	36	
317	1	0.009900990099009901	37	
354	1	0.009900990099009901	38	
355	1	0.009900990099009901	39	
359	1	0.009900990099009901	40	
361	1	0.009900990099009901	41	
362	1	0.009900990099009901	42	
382	1	0.009900990099009901	43	
384	1	0.009900990099009901	44	
414	1	0.009900990099009901	45	
429	1	0.009900990099009901	46	
447	1	0.009900990099009901	47	
450	1	0.009900990099009901	48	

output.txt: Блокнот

Файл	Редагування	Формат	Вигляд	Довідка
498	1	0.009900990099009901	49	
503	1	0.009900990099009901	50	
529	1	0.009900990099009901	51	
535	1	0.009900990099009901	52	
548	1	0.009900990099009901	53	
553	1	0.009900990099009901	54	
566	1	0.009900990099009901	55	
569	1	0.009900990099009901	56	
571	1	0.009900990099009901	57	
587	1	0.009900990099009901	58	
589	1	0.009900990099009901	59	
607	2	0.019801980198019802	61	
612	2	0.019801980198019802	63	
613	1	0.009900990099009901	64	
615	1	0.009900990099009901	65	
636	1	0.009900990099009901	66	
642	1	0.009900990099009901	67	
657	1	0.009900990099009901	68	
660	1	0.009900990099009901	69	
676	1	0.009900990099009901	70	
685	1	0.009900990099009901	71	
687	1	0.009900990099009901	72	
688	1	0.009900990099009901	73	
694	1	0.009900990099009901	74	
702	1	0.009900990099009901	75	
707	1	0.009900990099009901	76	
736	1	0.009900990099009901	77	
738	1	0.009900990099009901	78	
763	1	0.009900990099009901	79	
768	1	0.009900990099009901	80	
775	1	0.009900990099009901	81	
777	1	0.009900990099009901	82	
782	1	0.009900990099009901	83	
784	1	0.009900990099009901	84	
813	1	0.009900990099009901	85	
817	1	0.009900990099009901	86	
820	1	0.009900990099009901	87	
821	1	0.009900990099009901	88	
824	1	0.009900990099009901	89	
832	1	0.009900990099009901	90	
834	1	0.009900990099009901	91	

858	1	0.009900990099009901	92
878	1	0.009900990099009901	93
879	1	0.009900990099009901	94
880	1	0.009900990099009901	95
923	1	0.009900990099009901	96
928	1	0.009900990099009901	97
945	1	0.009900990099009901	98
976	1	0.009900990099009901	99
984	1	0.009900990099009901	100
999	1	0.009900990099009901	101

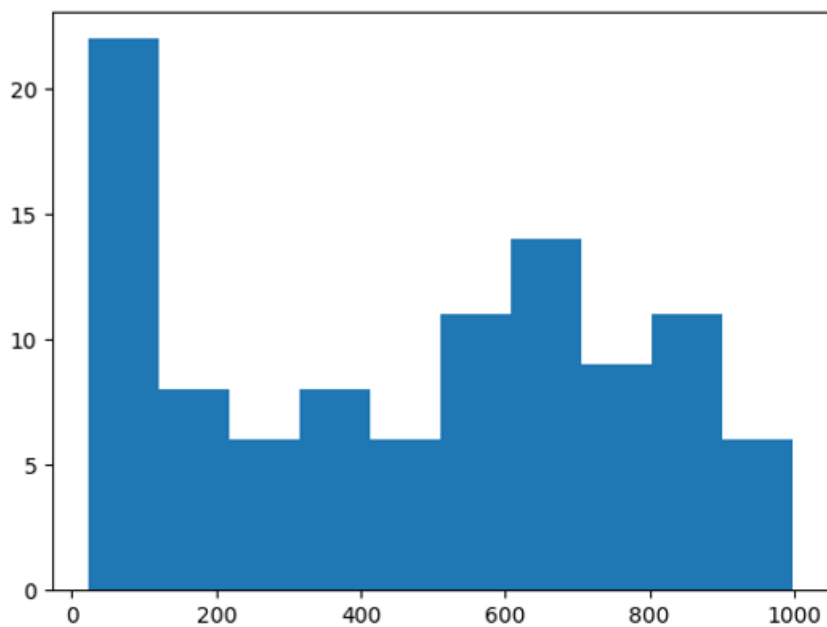
The most viewed film - 999

Moda = 22

Mediana = 529

Variance = 89484.82188020783

Standart deviation = 299.1401375278948



```
"C:\Users\admin\Desktop\2 курс\2 курс 1 семестр\ИОПИ\LAB1\ven
```

```
Enter a file name: input_100.txt
```

```
Process finished with exit code 0
```

Результат роботи програми при введенні даних із файлу input_1000.txt:

output.txt: Блокнот

Файл Редагування Формат Вигляд Довідка

99272	1	0.000999000999000999	994
99403	1	0.000999000999000999	995
99575	1	0.000999000999000999	996
99696	1	0.000999000999000999	997
99808	1	0.000999000999000999	998
99820	1	0.000999000999000999	999
99968	1	0.000999000999000999	1000
99970	1	0.000999000999000999	1001

The most viewed film - 99970

Moda = 14023

Mediana = 49311

Variance = 803308470.689839

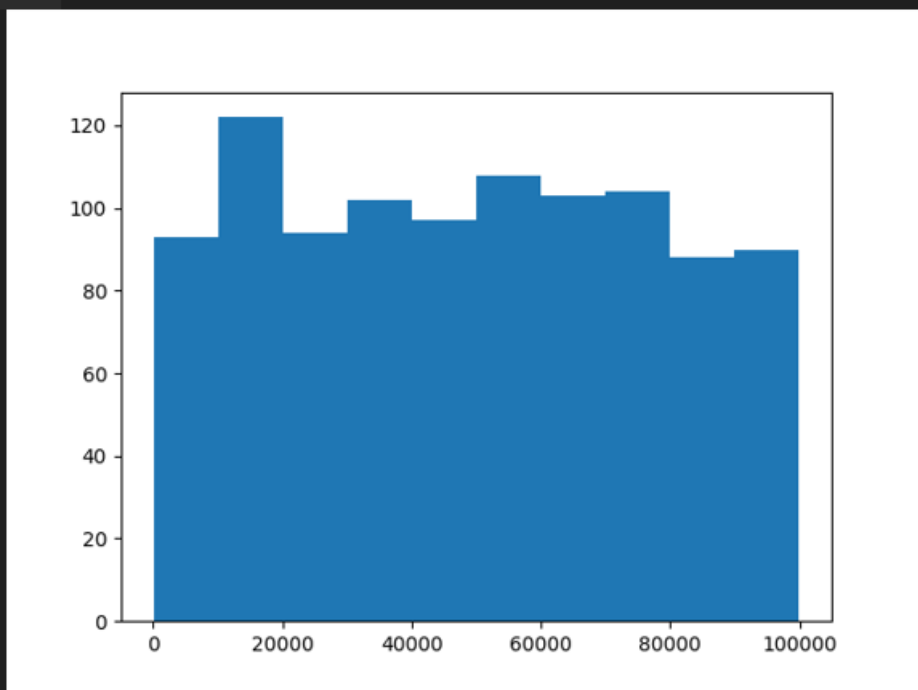
Standart deviation = 28342.696955121242



hist.png



89 %



Enter a file name: input_1000.txt

Process finished with exit code 0

Висновки:

Під час виконання цієї лабораторної роботи я навчилася використовувати здобуті знання про центральні тенденції та міри на практиці за допомогою мови програмування Python.