

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

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

ЛАБОРАТОРНА РОБОТА № 1

Центральні тенденції та міра дисперсії

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.

За поданими даними запишемо в файл таблицю частот та сукупних частот, а також знайдемо фільм, який був переглянутий частіше за інші:

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3
4
5 def task1():
6     file = input("Enter a name of the file: ")
7
8     reader = open(file).read().splitlines()
9     del reader[0]
10    [int(i) for i in reader]
11    writer = open("output.txt", "w")
12
13    cumulFreq, count = 0, 0
14
15    writer.write("  f(i)\t\tF(i)\n")
16    for line in reader:
17        cumulFreq += int(line)
18        writer.write(str(count) + "  " + line + "\t\t" + str(cumulFreq) + "\n")
19        count += 1
20
21    writer.write("Max value: " + str(max(reader)))
```

Маємо наступний вивід для файлів **input_10.txt**, **input_100.txt** та **input_1000.txt** відповідно:

	f(i)	F(i)
0)	1	1
1)	66	67
2)	75	142
3)	1	143
4)	1	144
5)	12	156
6)	10	166
7)	97	263
8)	12	275
9)	66	341
Max value: 97		

	f(i)	F(i)
0)	642	642
1)	51	693
2)	97	790
3)	529	1319
4)	46	1365
5)	999	2364
6)	317	2681
7)	99	2780
8)	880	3660
9)	46	3706
10)	79	3785
11)	548	4333
12)	361	4694
13)	821	5515
14)	71	5586
15)	288	5874
16)	51	5925
17)	255	6180
18)	429	6609
19)	80	6689
20)	657	7346
21)	22	7368
22)	817	8185
23)	168	8353
24)	688	9041
25)	858	9899
26)	162	10061
27)	587	10648
28)	775	11423
29)	51	11474
30)	566	12040
31)	738	12778
32)	763	13541
33)	832	14373

34)	447	14820
35)	414	15234
36)	784	16018
37)	355	16373
38)	154	16527
39)	251	16778
40)	660	17438
41)	250	17688
42)	813	18501
43)	382	18883
44)	694	19577
45)	613	20190
46)	923	21113
47)	362	21475
48)	687	22162
49)	571	22733
50)	103	22836
51)	79	22915
52)	535	23450
53)	162	23612
54)	193	23805
55)	198	24003
56)	607	24610
57)	91	24701
58)	928	25629
59)	676	26305
60)	569	26874
61)	503	27377
62)	945	28322
63)	777	29099
64)	269	29368
65)	47	29415
66)	615	30030
67)	685	30715
68)	225	30940

69)	22	30962
70)	824	31786
71)	553	32339
72)	589	32928
73)	22	32950
74)	976	33926
75)	384	34310
76)	702	35012
77)	612	35624
78)	878	36502
79)	820	37322
80)	77	37399
81)	834	38233
82)	147	38380
83)	879	39259
84)	119	39378
85)	736	40114
86)	768	40882
87)	146	41028
88)	707	41735
89)	450	42185
90)	498	42683
91)	119	42802
92)	636	43438
93)	612	44050
94)	359	44409
95)	984	45393
96)	782	46175
97)	22	46197
98)	354	46551
99)	607	47158
Max value: 999		

(input_1000.txt)

```
999) 69601 48984296
Max value: 99970
```

2. Знайти Моду та Медіану заданої вибірки.

Моду та медіану знайдемо за наступними формулами:

Для моди

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

для послідовності x_1, x_2, x_3, \dots ,

Медіана - елемент, який найбільше зустрічається в послідовності.

Отже, розрахуємо моду та медіану:

```

24 def task2():
25     file = input("Enter a name of the file: ")
26
27     reader = open(file).read().splitlines()
28     del reader[0]
29     writer = open("output.txt", "w")
30
31     median, mode = 0, 0
32     middle = int(len(reader)/2) - 1
33
34     if len(reader) % 2 == 0:
35         median = ((int(reader[middle]) + int(reader[middle + 1]))) / 2
36     else:
37         median = int(reader[middle + 1]) / 2
38
39     mode = (max(set(reader), key=lambda reader.count))
40     writer.write("Mode is " + str(mode) + "\nMedian is " + str(median))

```

Маємо такі результати для файлів **input_10.txt**, **input_100.txt** та **input_1000.txt** відповідно:

```

Mode is 1
Median is 6.5

```

```

Mode is 22
Median is 337.0

```

```

Mode is 40617
Median is 35071.0

```

3. Порахувати Дисперсію та Середнє квадратичне відхилення розподілу.

Дисперсію розрахуємо за наступною формулою:

$$D = \frac{\sum (x - \mu)^2}{n}$$

Де μ - середнє значення послідовності.

Середнє квадратичне відхилення - значення, що являє собою значення дисперсії в квадратному корені.

Отже, за наступним кодом розрахуємо Дисперсію та середнє квадратичне відхилення:

```

43 def task3():
44     file = input("Enter a name of the file: ")
45
46     reader = open(file).read().splitlines()
47     del reader[0]
48     writer = open("output.txt", "w")
49
50     average, dispersion = 0, 0
51     for i in range(len(reader)):
52         average += int(reader[i])
53
54     average /= len(reader)
55     for i in range(len(reader)):
56         dispersion += pow((int(reader[i]) - average), 2)
57
58     dispersion /= len(reader)
59     writer.write("Dispersion is " + str(dispersion) + "\nQuadratic deviation is " + str(np.sqrt(dispersion)))

```

Маємо такі результати для файлів **input_10.txt**, **input_100.txt** та **input_1000.txt** відповідно:

```

Dispersion is 1250.89
Quadratic deviation is 35.3679233204326

```

```

Dispersion is 89012.623600000004
Quadratic deviation is 298.34983425502355

```

```

Dispersion is 801811586.6903838
Quadratic deviation is 28316.27776898623

```

4. Побудувати гістограму частот для даного розподілу.

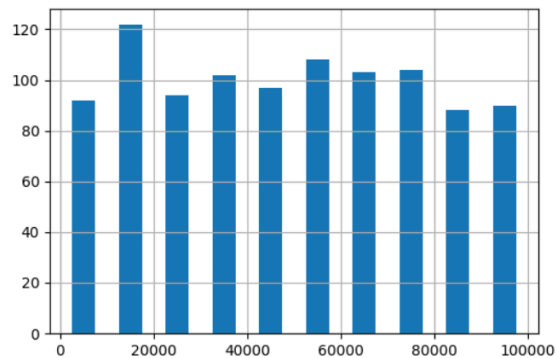
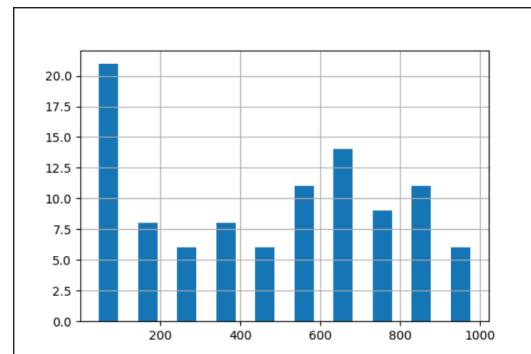
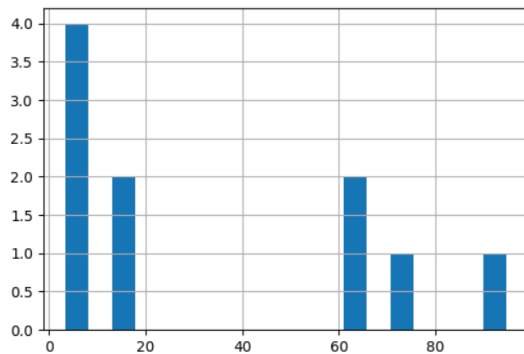
Напишемо програму для побудови гістограми частот для даного розподілу:

```

62 def task4():
63     fig = plt.figure(figsize=(6, 4))
64     ax = fig.add_subplot()
65
66     file = input("Enter a name of the file: ")
67
68     reader = open(file).read().splitlines()
69     del reader[0]
70     writer = open("output.txt", "w")
71     y = [int(i) for i in reader]
72
73     ax.hist(y, rwidth=0.5)
74     ax.grid()
75     plt.show()

```

Маємо такі гістограми для файлів **input_10.txt**, **input_100.txt** та **input_1000.txt** відповідно:



Висновок:

Протягом даної лабораторної роботи було засвоєно та використано набуті знання про центральні тенденції та міри. Було побудовано таблицю частот, розраховано моду та медіану, дисперсію та середнє квадратичне відхилення за поданою інформацією. Також було ознайомлено з гістограмою частот та побудовано дану діаграму за поданою інформацією.