

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»

| | | | |
|----------------|----------------------------|----------------|--|
| Виконав: | Дармороз Аліна Олегівна | Перевірила: | |
| Група | ІПЗ-21(1) | Дата перевірки | |
| Форма навчання | денна | Оцінка | |
| Спеціальність | 121 | | |
| 2022 | | | |

Мета: навчитись використовувати на практиці набуті знання про центральні тенденції та міри.

Постановка задачі.

1. Перш за все необхідно створити функцію зчитування даних з файлу. Дані будуть перезаписані у масив, який далі буде використовуватись для обчислення статистичних даних.

2. Для побудови таблиці частот у вхідних даних ми отримуємо масив значень, який циклом перебирається для перевірки наявності елемента в масиві. Якщо елемент наявний додається 1 до його частоти, якщо ні частота не змінюється. Отримуємо словник значень: елемент та його частота.

3. Для визначення сукупних частот маємо масив частот де переписуємо його так, що кожна частота дорівнює сумі поточної частоти та попередньої.

4. Для знаходження медіани, для вхідних даних потребуємо масив значень, якщо кількість елементів парна знаходимо індекси двох центральних елементів та виводимо середнє значення елементів з цим індексом, якщо кількість елементів непарна виводимо елемент з цим індексом.

5. Для знаходження найбільш переглянутого фільму на вході маємо масив кількості переглядів, обираємо максимальне значення та виводимо номер цього елемента.

6. Для знаходження дисперсії та середнього квадратичного відхилення на вході маємо масив частот. Знаходимо середнє значення усіх елементів та масив значень різниці кожного елемента та середнього значення піднесеного до квадрату. На виході отримаємо середнє значення створеного масиву та корінь з нього, що і буде дисперсією та середнім квадратичним відхиленням.

7. На виході отримуємо результати наших обчислень виведені на консоль та записані у файл.

Побудова математичної моделі

1. Частота – число повторів визначеного значення у вибірці

2. Сукупна частота:

$$F_j = \sum_{i=1}^k f_i$$

де F_j – сукупна частота елемента x_j , k – кількість елементів у масиві частот, f_i – частота появи цих елементів, i – індекс частоти j -го елемента.

3. Мода – елемент, який з'являється у вибірці найчастіше

4. Медіана:

$$M = \begin{cases} x_{\frac{n+1}{2}}, & 2 \text{ isn't multiple of } n \\ \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}, & \frac{n}{2} \end{cases}$$

5. Дисперсія:

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

де n – кількість елементів у масиві значень, x_i – значення i -го елементу,
 \bar{x} – середнє значення масиву елементів

6. Середнє квадратичне:

$$\sigma = \sqrt{Var}$$

Var – дисперсія

Псевдокод алгоритму.

```
input: fileName
SortOfArray( ):
    for index=0 to data do
        min = index
        for comp=index+1 to len(data) do
            if data[comp] < data[min] then
                min = comp
        if min != index then
            data[index], data[min] = data[min],
data[index]
        end if
    return data
GetData():
    f = open(fileName, 'r')
    s = f.readline()
    n = int(s)
    arr:=[]
    for line in f do
        data = line.split(' ')
        for s in data do
            if s != "":
                arr = arr+[int(s)]
            end if
        end if
    SortOfArray( ):
        for index=0 to arr do
            min = index
            for comp=index+1 to len( arr) do
                if arr[comp] < arr[min] then
                    min = comp
            if min != index then
                arr[index], arr[min] = arr[min], arr[index]
            end if
        print arr
    n=GetData()
    freqDict := {}
    FreqTable():
        a=[]
        for num in arr do
            if num in freqDict then
                freqDict[num] += 1
            else
                freqDict[num] = 1
        l=Cumulative(freqDict)
        i:=0
        k:=0
    print freqDict
        i=i+1
        k=k+1
    Cumulative():
```

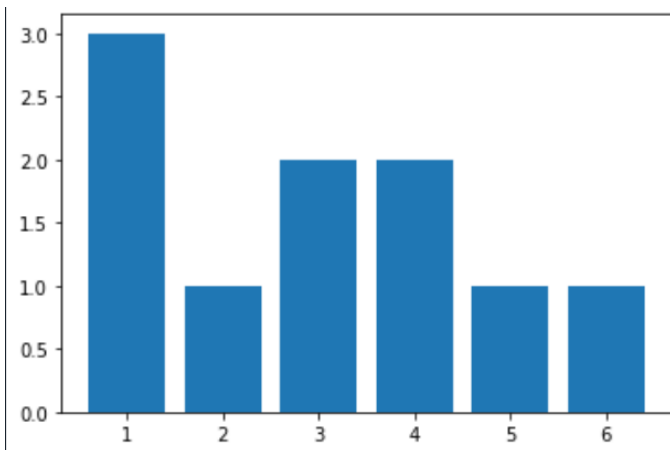
```
L:=[]
l=list(freqDict.values())
for i=1 to len(l) do
    l[i]=l[i]+l[i-1]
return l
findMedian():
    n=len(arr)
    index =n//2
    if n % 2 then
        return arr[index]
    end if
    print sum(arr[index-1:index+1])/2
    h=sum(arr[index-1:index+1])/2
findMode():
    a=list(freqDict.values())
    k=max(a)
    if k!=1 then
        for j in freqDict.keys() do
            if k=freqDict[j] then
                end if
        end if
    MostViewed(a):
        max_value := []
        maximum = max(a)
        for i = 1 to len(a)+1 do
            if a[i - 1] = maximum then
                max_value.append(i)
        if len(max_value) = len(a) then
        else if (len(max_value) == 1):
            print str(max_value[0])
        else:
            print max_value
            for i in range (0, len(max_value)):
                if i != len(max_value) - 1 then
                    print str(max_value[i])
                else:
                    print max_value[i]
    func1(a):
        x := []
        for i in range (1, len(a) + 1):
            x.append(i)
        fig, axes = plt.subplots(1, 1)
        axes.bar(x, a)
    variance():
        data=SortedArray
        n = len(data)
        mean = sum(data)/n
        deviations = [(x - mean) ** 2 for x in data]
        variance = sum(deviations) / n
        print variance
        print sqrt(variance)
```

Виконання алгоритму.

```
Введіть назву файлу input_10.txt
[1, 66, 75, 1, 1, 12, 10, 97, 12, 66]
[1, 66, 75, 1, 1, 12, 10, 97, 12, 66]
Медіана 12.0
[1, 1, 1, 10, 12, 12, 66, 66, 75, 97]
None
Дисперсія: 1250.8899999999999
Відхилення: 35.367923320432595
Таблиця частот

1:3:3
10:1:4
12:2:6
66:2:8
75:1:9
97:1:10
Найчастіше було переглянуто Фільм №10

Мода вибірки: 1
```



output_10.txt: Блокнот

Файл Редагування Формат Вигляд Довідка

```
Медіана: 12.0
Дисперсія: 1250.8899999999999
Відхилення: 35.367923320432595
Таблиця частот
1:3:3
10:1:4
12:2:6
66:2:8
75:1:9
97:1:10
Найчастіше було переглянуто Фільм №10
Мода вибірки:1
```

Висновок. При виконанні цієї лабораторної роботи було розроблено алгоритм, за допомогою якого можна обчислити статистичні значення вибірки розміщеної у файлі. Було використано на практиці знання про центральні тенденції та міри, та розроблено алгоритми для їх обчислення та роботи з ними. Реалізовано розрахунок та побудову таблиці частот та сукупних частот, знаходження моди та медіани,

розрахунок дисперсії та середньо квадратичного відхилення розподілу, а також побудову гістограми частот.