

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  
«Лінійне перетворення та Графічне зображення даних»

Виконав:	Сирота Ангеліна Олександрівна	Перевірила:	Вечерковська Анастасія Сергіївна
Група	ІПЗ-21	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

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

## Хід роботи

### Постановка задачі:

Написати програму, що зчитує дані з файла і виконує наступні функції:

- Знайдіть  $Q_1$ ,  $Q_3$  та  $P_{90}$
- Знайдіть середнє та стандартне відхилення цих оцінок.
- Через незадоволення низькими оцінками викладач вирішив використати шкалу форми  $y = ax + b$ , щоб відредагувати оцінки. Він хотів, щоб середнє значення масштабних оцінок становило 95, а оцінка 100, щоб залишалася рівною 100.
- Показати дані за допомогою діаграми "стовбур – листя".
- Відобразити дані за допомогою коробкового графіка.

Усі результати програма записує в окремий текстовий файл.

### Побудова математичної моделі:

**Перцентиль:** статистична міра положення, яка розподіляє впорядкований розподіл даних на сто рівних частин.

Номер елемента:  $P_k = \frac{k}{100}(N + 1)$ . Щоб знайти значення k-го персентіля необхідно знайти елемент з цим номером у вибірці

**Перший / нижній кuartиль:**  $Q_1 = P_{25}$

**Третій / вищий кuartиль:**  $Q_3 = P_{75}$

**Дисперсія** рахується за формулою:  $Var(x) = \frac{\sum_{x \in x} f_x (x - \bar{x})^2}{\sum_{x \in x} f_x}$

**Середнє квадратичне відхилення** обчислюється за формулою:  $\sigma = \sqrt{Var(x)}$

**Діаграма «стовбур - листя»:** засіб для зображення кількісних даних у графічному форматі, схожого на гістограму, що допоможе у візуалізації форми розподілу.

```
4 | 4 6 7 9
5 |
6 | 3 4 6 8 8
7 | 2 2 5 6
8 | 1 4 8
9 |
10 | 6
ключ: 6 | 3=63
```

**Коробковий графік:** засіб візуалізації в описовій статистиці груп числових даних через їх квантілі. Коробковий графік може також мати лінії, які виходять вертикально з коробки (вони називаються вусами), вони вказують величину мінливості поза верхньою та нижньою межами квантіля. Викиди може бути нанесено у вигляді точок.

### Псевдокод алгоритму:

Заповнення масиву значеннями елементів:

```
for item in f:
    arr.append(int(item))
arr = sorted(arr)           // відсортувати масив
```

Знаходження номера персентиля:

```
p = k / 100 * (len(arr) + 1) // формула для обрахунку
return p # номер
```

Знаходження значення персентиля:

```
pn = percTH(arr, k)           // номер елемента
x = math.modf(pn)             // 0 - дробова частина; 1 - ціла частина
if x[1] != 0:
    p = arr[int(x[1]) - 1] + x[0] * (arr[int(x[1])] - arr[int(x[1]) - 1])
    return p # значення
```

Масив впорядкованих елементів і їх частот:

```
global arrf
arrR = [] # масив повторюваних значень
for el in arr:
    flag = False           # визначити оцінку, як ту, яка ще не повторювалась
    f = 0                  # частота
    if len(arrR) != 0:
        for i in arrR:
            if el == i:    # якщо оцінка співпадає з тією, що вже була
                flag = True # визначити її як вже враховану
    if flag == False:      # якщо такої оцінки ще не було
        for el1 in arr:   # шукати частоту
            if el == el1:
                f += 1
        arrf.extend([[el, f]]) # заповнити масив оцінками і їх частотами
        arrR.append(el)       # внести оброблену оцінку в масив повторюваних
                                # значень
return arrf
```

Визначення середнього значення:

```
for i in range(len(arr)):
    numerator += arr[i][0] * arr[i][1] // чисельник
    denominator += arr[i][1]         // знаменник
Xave = numerator / denominator
return Xave
```

Визначення середнього квадратичного відхилення:

```
arrf = frequency(arr)           // масив з частотами
Xave = average(arrf)            // середнє значення

# дисперсія
for i in range(len(arrf)):
    numerator += arrf[i][1] * math.pow(arrf[i][0] - Xave, 2)
    denominator += arrf[i][1]
dis = numerator / denominator

# середнє квадратичне відхилення
msd = math.sqrt(dis)
print(msd)
```

Побудова шкали оцінок:

```
// 100 = a*100 + b")
// 95 = a*x~ + b")

Xavg = average(arrf)           // середнє значення x
x = numpy.array ([[100, 1], [Xavg, 1]])
y = numpy.array ([[100], [95]])
res = numpy.linalg.solve(x, y) // розв'язує систему рівнянь

print("Шкала: y =", *numpy.round(res[0], 3), "* x +", *numpy.round(res[1], 3))
```

Створити словник:

```
for i in range(len(arr)):
    el = arr[i] / 10
    x = math.modf(el)

    if(curKey != int(x[1]) and First != True):

        dict_[curKey] = values
        values = []
        First = False

    values.append(int(round(x[0], 1) * 10))
    curKey = int(x[1])

dict_[curKey] = values
return dict_
```

Вивести діаграму стовбур-листя:

```
dict_ = createDict(arr)

print("----- Діаграма стовбур-листя -----")
print("Ключ: 1|1 = 11")
print()

for key, value in dict_.items():
    print(key, "\t", *value)
```

Побудова шкали оцінок:

```
pyplot.title("Коробкова діаграма")

pyplot.boxplot(x = arr,
               patch_artist = True,                      // власний колір
               widths = 0.2,
               boxprops = {'color':'black','facecolor':'#69f745'}, // колір
               medianprops = {'linestyle':'--','color':'black'}) // медіана

pyplot.show()
```

**Випробування алгоритму:**

Набір з 10 оцінок:

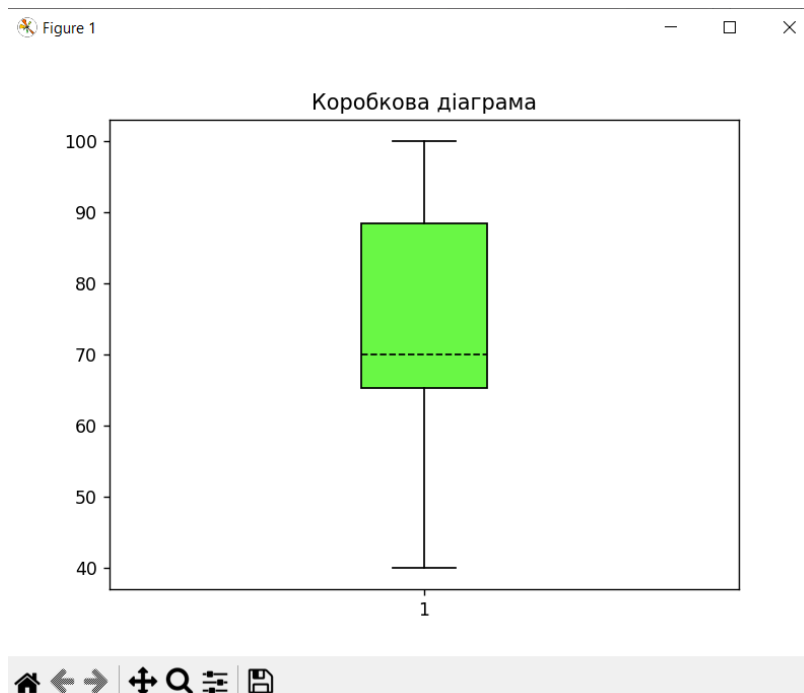
```
Введіть значення кількості елементів у вхідному файлі (10/100):
10
Q1: 64.25
Q3: 91.25
P90: 99.5

Середнє квадратичне відхилення: 17.174

 $y = a \cdot x + b$ 
 $y_{\sim} = a \cdot x_{\sim} + b$ 
Оцінка 100 лишається 100:       $100 = a \cdot 100 + b$ 
Середнє значення оцінок = 95:     $95 = a \cdot x_{\sim} + b$ 
Шкала:  $y = 0.194 \cdot x + 80.62$ 

----- Діаграма стовбур-листя -----
Ключ: 1|1 = 11

4      | 0
6      | 2 5 6
7      | 0 0
8      | 4
9      | 0 5
10     | 0
```



Набір зі 100 оцінок:

```

Введіть значення кількості елементів у вхідному файлі (10/100):
100
Q1: 36.5
Q3: 78.75
P90: 91.9

Середнє квадратичне відхилення: 25.154

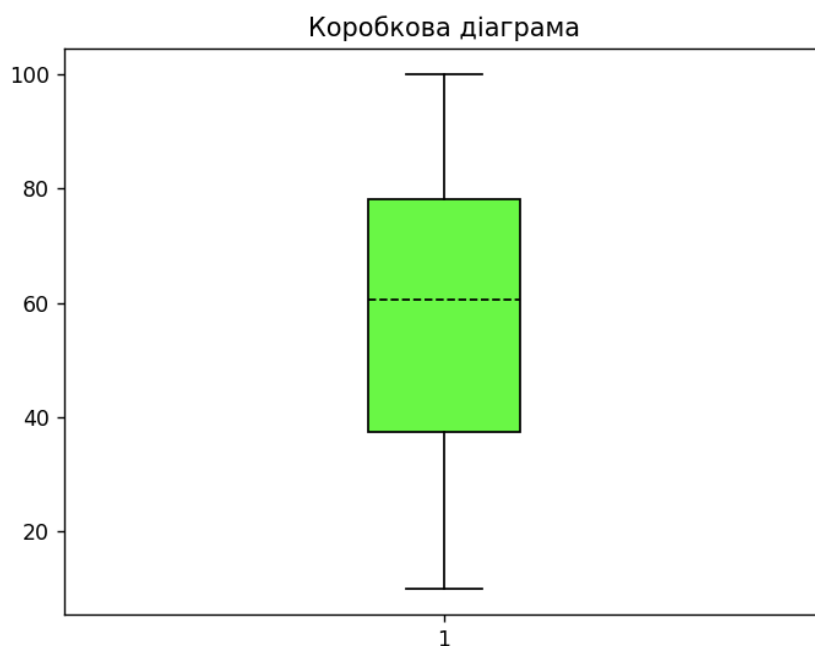
 $y = a \cdot x + b$ 
 $y_{\sim} = a \cdot x_{\sim} + b$ 
Оцінка 100 лишається 100:  $100 = a \cdot 100 + b$ 
Середнє значення оцінок = 95:  $95 = a \cdot x_{\sim} + b$ 
Шкала:  $y = 0.119 \cdot x + 88.107$ 

----- Діаграма стовбур-листя -----
Ключ: 1|1 = 11

1      | 0 2 4 4 5 6 6 9 9
2      | 2 2 2 2 2 5 5 6 8
3      | 1 5 5 5 6 6 8 8
4      | 1 2 4 5 6 6 7 9
5      | 0 1 1 1 2 3 4 5 6 6 7 8 8
6      | 0 0 1 1 1 1 3 4 5 6 7 8 8 8 9
7      | 0 0 1 3 3 6 6 7 7 7 8 8 9 9
8      | 0 1 1 2 2 2 3 3 5 7 7 8
9      | 1 2 2 4 7 7 8 9 9
10     | 0 0

```

Figure 1



**Висновок:** в ході цієї лабораторної роботи було зчитано дані з вхідних файлів і обчислено 1-ий і 3-ій квартилі і персентиль 90-ий. Для утвореної вибірки обчислено стандартне відхилення. Створено шкалу оцінок, що відповідає поставленим умовам. Дані показано у вигляді діаграми «стовбур-листя» і коробкового графіку. Вихідні дані записувались в окремий текстовий файл.