

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

Дисципліна
«Ймовірнісні основи програмної інженерії»

Тема
«Лінійне перетворення та Графічне зображення даних»

Виконав:	Булава Геннадій Юрійович	Перевірила:	Марцафей Анна Сергіївна
Група	ІПЗ-21(2)	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

Назва теми: Лінійне перетворення та Графічне зображення даних

Мета: навчитись використовувати на практиці набуті знання про лінійні перетворення та графічне зображення даних.

Постановка задачі

1. Знайдіть Q_1 , Q_3 , P_{90}
2. Знайдіть середнє та стандартне відхилення цих оцінок.
3. 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.
4. Показати дані за допомогою діаграми "стовбур – листя".
5. Відобразити дані за допомогою коробкового графіка.
6. Зробити висновок.

Математична модель

1. Першим квантилем є $0.25 \cdot (n+1)$ -й елемент вибірки. Третім персентилем є $0.75 \cdot (n+1)$ -й елемент вибірки. Дев'яностим персентилем є $0.9 \cdot (n+1)$ -й елемент вибірки.
2. Середнє відхилення обчислюється за формулою:

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

Стандартне відхилення обчислюється за формулою:

$$\sqrt{\frac{\sum_{i=1}^n (x_i - x_{\text{ср}})^2}{n - 1}}$$

3. Розв'язуємо систему рівнянь:

$$\begin{cases} 95 = a\bar{x} + b \\ 100 = 100a + b \end{cases}$$

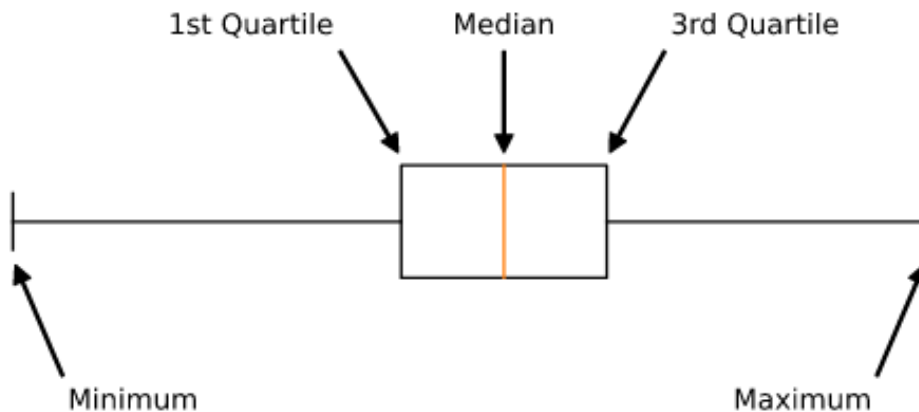
З розв'язку отримуємо значення a та b , які підставляємо у функцію $y = ax + b$. Всі оцінки обчислюємо за цією функцією.

4. Діаграма стовбур-листя будується за схемою:

5, 16, 21, 23, 23, 26, 26, 30, 32, 4:

Stem	Leaf
1	5 6
2	1 3 3 6 6
3	0 2
4	1

5. Коробкова діаграма будується за наступним зразком:



Випробування алгоритму

Вхідні файли: input_10.txt, input_100.txt

Результат для input_10.txt

Q1: 64.25

Q3: 91.25

P90: 99.5

Середнє відхилення: 14.440000000000001

Стандартне відхилення: 18.103406677565783

Кориговані оцінки: 88.3720930232558, 92.63565891472868,
93.21705426356588, 93.41085271317829, 94.18604651162791,

94.18604651162791, 96.89922480620154, 98.06201550387597,
99.03100775193798, 100.0

Перевірка: середнє значення коригованих оцінок = 95

4 | 0

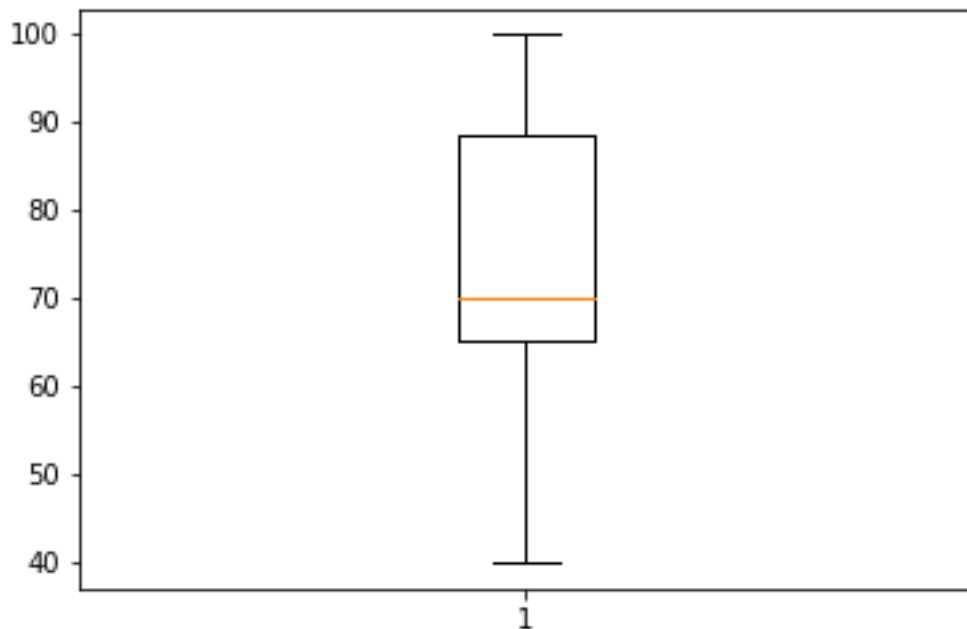
6 | 2 5 6

7 | 0 0

8 | 4

9 | 0 5

10 | 0



Результат для input_100.txt

Q1: 36.5

Q3: 78.75

P90: 91.9

Середнє відхилення: 21.243199999999999

Стандартне відхилення: 25.280615003387382

Кориговані оцінки: 89.29590865842056, 89.5337773549001, 89.77164605137965, 89.77164605137965, 89.89058039961941, 90.00951474785919, 90.00951474785919, 90.3663177925785, 90.3663177925785, 90.72312083729781, 90.72312083729781, 90.72312083729781, 90.72312083729781, 90.72312083729781, 91.07992388201713, 91.07992388201713, 91.07992388201713, 91.1988582302569, 91.43672692673644, 91.79352997145575, 92.26926736441484, 92.26926736441484, 92.26926736441484, 92.38820171265462, 92.38820171265462, 92.62607040913416, 92.62607040913416, 92.98287345385347, 93.10180780209325, 93.33967649857279, 93.45861084681256, 93.57754519505234, 93.57754519505234, 93.6964795432921, 93.93434823977165, 94.05328258801143, 94.17221693625119, 94.17221693625119, 94.17221693625119, 94.29115128449097, 94.41008563273074, 94.5290199809705, 94.64795432921028, 94.76688867745005, 94.76688867745005, 94.88582302568982, 95.00475737392959, 95.00475737392959, 95.24262607040913, 95.24262607040913, 95.3615604186489, 95.3615604186489, 95.3615604186489, 95.3615604186489, 95.59942911512846, 95.71836346336822, 95.837297811608, 95.95623215984777, 96.07516650808753, 96.19410085632731, 96.19410085632731, 96.19410085632731, 96.31303520456709, 96.43196955280685, 96.43196955280685, 96.55090390104662, 96.78877259752616, 96.78877259752616, 97.14557564224549, 97.14557564224549, 97.26450999048525, 97.26450999048525, 97.26450999048525, 97.38344433872503, 97.38344433872503, 97.50237868696479, 97.50237868696479, 97.62131303520457, 97.74024738344434, 97.74024738344434, 97.85918173168412, 97.85918173168412, 97.85918173168412, 97.97811607992388, 97.97811607992388, 98.21598477640343, 98.45385347288297, 98.45385347288297, 98.57278782112274, 98.92959086584206, 99.04852521408182, 99.04852521408182, 99.28639391056137, 99.64319695528069, 99.64319695528069, 99.76213130352046, 99.88106565176022, 99.88106565176022, 100.0, 100.0

Перевірка: середнє значення коригованих оцінок = 95

1 | 0 2 4 4 5 6 6 9 9

2 | 2 2 2 2 2 5 5 5 6 8

3|15556688

4|12456679

5|0111234566788

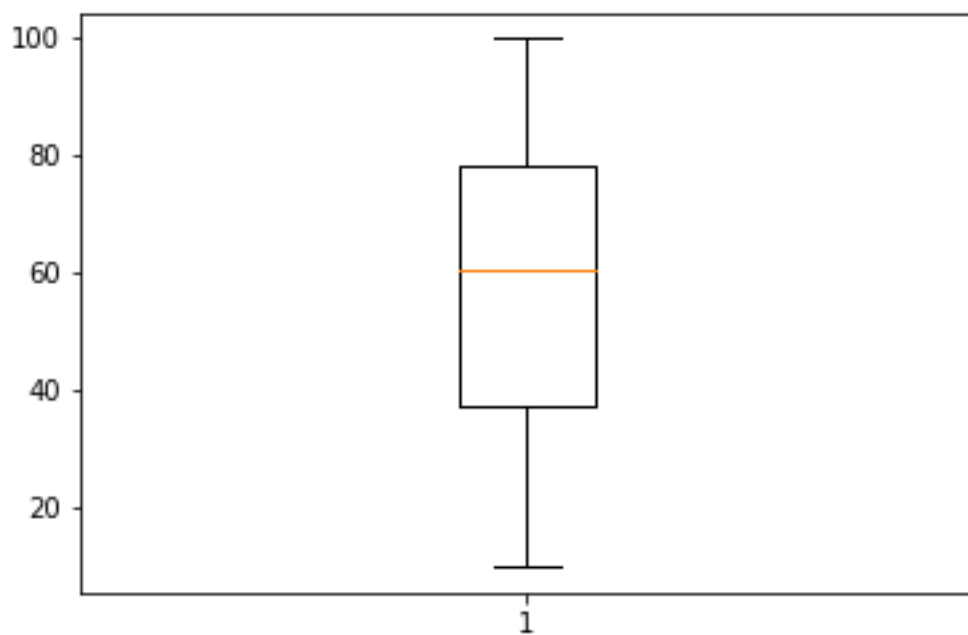
6|001111345678889

7|00133667778899

8|011222335778

9|122477899

10|00



Псевдокод

```
import math
import numpy
import matplotlib
a = []
print('Введіть назву файлу')
name = input()
file = open(name, mode="r")
i = int(0)
m = int(-1)
with open(name, "r") as file:
    for line in file.readlines():
        if (i > 0):
            a.append(int(line))
        else:
            m = int(line)
            i = i + 1
file.close()
output_name = 'output_'+str(m)+'_txt'

a.sort()

with open(output_name, "a") as f:
    q1 = (m+1)/4
    cq1 = math.floor(q1)
    dq1 = q1-cq1
    q1 = a[cq1-1]+dq1*(a[cq1]-a[cq1-1])
    f.write('Q1: ' + str(q1) + '\n')
    q3 = 0.75*(m+1)
    cq3 = math.floor(q3)
    dq3 = q3-cq3
    q3 = a[cq3-1]+dq3*(a[cq3]-a[cq3-1])
    f.write('Q3: ' + str(q3) + '\n')
    p90 = 0.9*(m+1)
    cp90 = math.floor(p90)
    dp90 = p90-cp90
    p90 = a[cp90-1]+dp90*(a[cp90]-a[cp90-1])
    f.write('P90: ' + str(p90) + '\n')

with open(output_name, "a") as f:
    f.write('\n\n')
    sum = int(0)
    for i in range(0, m):
        sum = sum + a[i]
    av = sum/m
    sum = int(0)
    for i in range(0, m):
        sum = sum + abs(a[i] - av)
    f.write('Середнє відхилення: ' + str(sum/m) + '\n')
    sum = int(0)
    for i in range(0, m):
        sum = sum + math.pow((a[i]-av), 2)
    f.write('Стандартне відхилення: ' + str(math.sqrt(sum/(m - 1))) + '\n')

with open(output_name, "a") as f:
    f.write('\n\n')
    sum = int(0)
    for i in range(0, m):
        sum = sum + a[i]
    av = sum/m
    l = numpy.array([[av,1],[100,1]])
    r = numpy.array([95,100])
    res = numpy.linalg.solve(l,r)
    f.write('Кориговані оцінки: ')
    sum = int(0)
    res1 = []
    for i in range(0, m-1):
        res1.append(res[0]*a[i]+res[1])
        f.write(str(res1[i]) + ', ')
        sum = sum + res1[i]
    res1.append(res[0]*a[m-1]+res[1])
    f.write(str(res1[m-1]) + '\n')
    sum = sum + res1[m-1]
    f.write('Перевірка: середнє значення коригованих оцінок = ' + str(round(sum/m)) + '\n')

with open(output_name, "a") as f:
    f.write('\n\n')
    if (a[0] < 100):
        c = int(str(a[0])[0])
        f.write('\n' + str(c) + ' | ')
    else:
        c = 10
        f.write('\n' + '10 | ')
    for i in range(0, m):
        if (a[i] < 100):
            c1 = int(str(a[i])[0])
        else:
            c1 = 10
        if (c1 != c):
            f.write('\n' + str(c1) + ' | ')
            c = c1
        if (len(str(a[i])) == 2):
            f.write(str(a[i])[1] + ' ')
        else:
            f.write('0 ')

matplotlib.pyplot.boxplot(a)
matplotlib.pyplot.savefig('output_figure_'+str(m)+'_png')
```

Висновок

В ході виконаної роботи було опрацьовано знаходження квартилей та персентилей, обчислення середнього та стандартного відхилень. Було опрацьовано використання лінійних перетворень. Також було опрацьовано графічне відображення даних за допомогою діаграми стовбур-листя та коробкового графіка.