

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 № 2

на тему:

«Лінійне перетворення та графічне зображення даних»

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 лінійні перетворення та графічне зображення даних.

Постановка задачі:

1. Знайдіть Q_1 , Q_3 та 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.

Розроблена програма повинна зчитувати вхідні дані з файлу заданого формату та записувати дані у файл.

Побудова математичної моделі:

Правила обчислення кватилей та персентилей:

DEFINITIONS

- The k^{th} percentile, P_k , is a value that splits the data into two parts Part 1 consisting of N_1 numbers that are less than P_k and part 2 consisting of N_2 numbers that are greater than P_k . The ratio $N_1 : N_2$ is $\frac{k}{100 - k}$.
- The 25th percentile is called the *first* or *lower quartile* and denoted by Q_1 .
- The 50th percentile is called the *second* or *middle quartile* Q_2 . It is also the median of the data.
- The *third* or *upper quartile* Q_3 is the 75th percentile.

The k^{th} percentiles, the lower quartile, and the upper quartile of a data set of size N are sometimes referred to, respectively, as $\frac{k}{100}(N+1)^{\text{th}}$, $\frac{1}{4}(N+1)^{\text{th}}$, and $\frac{3}{4}(N+1)^{\text{th}}$ terms of the data.

Формула знаходження середнього значення:

СЕРЕДНЄ

Середнє \bar{x} незгрупованих даних $x_1, x_2, x_3, \dots, x_n$

отримують додаванням елементів і діленням суми на розмір даних:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} .$$

Формула знаходження стандартного відхилення:

VARIANCE AND STANDARD DEVI

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

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

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

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

Принцип лінійних перетворень:

LINEAR TRANSFORMATION

The mean, variance, and standard deviation are the most commonly used measures to extract useful information from data. Some of their properties are discussed in this section.

A set of data X is said to be linearly transformed into a set Y if the elements of X are mapped onto the elements of Y by the relation $y = ax + b \in Y$, where a and b are real numbers.

The mean and standard deviation of Y are calculated as follows:

$$\begin{aligned}\bar{y} &= \frac{\sum_{y \in Y} f_y \cdot y}{\sum_{y \in Y} f_y} = \frac{\sum_{x \in X} f_x \cdot (ax + b)}{\sum_{y \in X} f_x} = \frac{\sum_{x \in X} f_x \cdot (ax) + \sum_{x \in X} f_x \cdot (b)}{\sum_{y \in X} f_x} = \frac{a \sum_{x \in X} f_x \cdot x + b \sum_{x \in X} f_x}{\sum_{y \in X} f_x} \\ &= a \frac{\sum_{x \in X} f_x \cdot x}{\sum_{y \in X} f_x} + b \frac{\sum_{x \in X} f_x}{\sum_{y \in X} f_x}\end{aligned}$$

Hence, $\bar{y} = a \cdot \bar{x} + b$.

In a similar way, one can show that $\text{Var}(Y) = a^2 \text{Var}(X)$ and $\sigma_y = |a| \sigma_x$.

Псевдокод алгоритму:

```
1  from math import *
2  import numpy as np
3  import stemgraphic
4  import matplotlib.pyplot as plt
5
6  s = input('Enter a file name: ')
7  f1 = open(s, 'r')
8  f2 = open('output.txt', 'w')
9  items = f1.read().split()
10 items = [int(i) for i in items]
11 items.sort()
12
13
14 def int_round(num):
15     num = int(num + (0.5 if num > 0 else -0.5))
16     return num
17
18
19 def task1():
20     q1_index = 1 / 4 * (len(items) + 1)
21     q3_index = 3 / 4 * (len(items) + 1)
22     p90_index = 90 / 100 * (len(items) + 1)
23     q1 = items[trunc(q1_index) - 1] + q1_index % 1 * (items[trunc(q1_index)] - items[trunc(q1_index) - 1])
24     q3 = items[trunc(q3_index) - 1] + q3_index % 1 * (items[trunc(q3_index)] - items[trunc(q3_index) - 1])
25     p90 = items[trunc(p90_index) - 1] + p90_index % 1 * (items[trunc(p90_index)] - items[trunc(p90_index) - 1])
26     f2.write("First quartile = %s\nThird quartile = %s\n90th percentile = %s" % (q1, q3, p90))
27     f2.write("\n-----")
```

```

30 def task2():
31     variance = 0
32     mean = sum(items) / len(items)
33     i = 0
34     while i < len(items):
35         variance += pow(items[i] - mean, 2)
36         i = i + 1
37     variance /= len(items)
38     standart_deviation = sqrt(variance)
39     f2.write('\nMean = %s\nStandart deviation = %s' % (mean, standart_deviation))
40     f2.write("\n-----\n")
41
42
43 def task3():
44     mean = sum(items) / len(items)
45     left = np.array([[100, 1], [mean, 1]])
46     right = np.array([100, 95])
47     a = np.linalg.solve(left, right)[0]
48     b = np.linalg.solve(left, right)[1]
49     i = 0
50     f2.write("Before\tAfter\n")
51     while i < len(items):
52         f2.write("%s\t" % items[i])
53         items[i] = int_round(items[i] * a + b)
54         f2.write("%s\n" % items[i])
55         i = i + 1

```

```

58 task1()
59 task2()
60 task3()
61 stemgraphic.stem_graphic(items, scale=10)
62 plt.savefig('stem.png')
63 fig = plt.figure()
64 plt.boxplot(items)
65 fig.savefig('boxplot.png')
66 f1.close()
67 f2.close()
68

```

Випробування алгоритму:

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

```

"C:\Users\admin\Desktop\2 курс\2 курс 1 семестр\ІІОПІ\LAB2\venv\Scripts\python.exe" "C:\Users\admin\Desktop\2 курс\2 курс 1 семестр\ІІОПІ\LAB2/main.py"
Enter a file name: input_10.txt
Process finished with exit code 0

```

output.txt: Блокнот

Файл Редагування Формат Вигляд Довідка

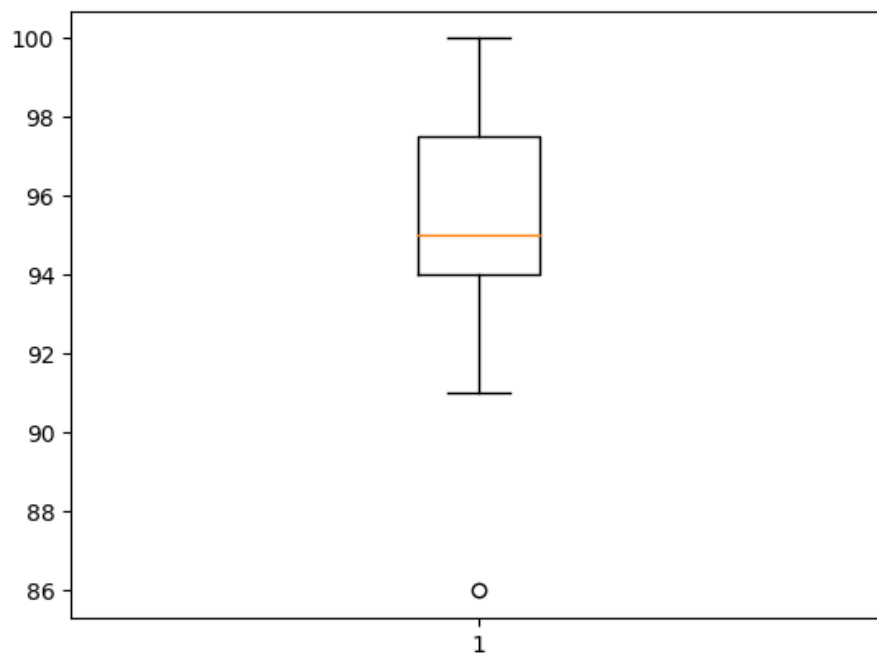
First quartile = 62.0
Third quartile = 90.0
90th percentile = 99.0

Mean = 68.36363636363636
Standart deviation = 24.67340388969735

Before	After
10	86
40	91
62	94
65	94
66	95
70	95
70	95
84	97
90	98
95	99
100	100

11		10	0
10		9	144555789
1		8	6

Key: aggr|stem|leaf
11 | 10 | 0 = 10.0 x 10 = 100.0



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

```
"C:\Users\admin\Desktop\2 курс\2 курс 1 семестр\ЙОПІ\LAB2\venv\Scripts\python.exe" "C:/Users/admin/Desktop/2 курс/2 курс 1 семестр/ЙОПІ/LAB2/main.py"
Enter a file name: input_100.txt

Process finished with exit code 0
```

First quartile = 37.0

Third quartile = 79.0

90th percentile = 92.0

Mean = 58.37623762376238

Standart deviation = 25.372805074127573

Before	After
--------	-------

10	89
----	----

12	89
----	----

14	90
----	----

14	90
----	----

15	90
----	----

16	90
----	----

16	90
----	----

19	90
----	----

19	90
----	----

22	91
----	----

22	91
----	----

22	91
----	----

22	91
----	----

22	91
----	----

25	91
----	----

25	91
----	----

25	91
----	----

26	91
----	----

28	91
----	----

31	92
----	----

35	92
----	----

35	92
----	----

35	92
----	----

36	92
----	----

36	92
----	----

38	93
----	----

38	93
----	----

41	93
----	----

42	93
----	----

44	93
----	----

45	93
----	----

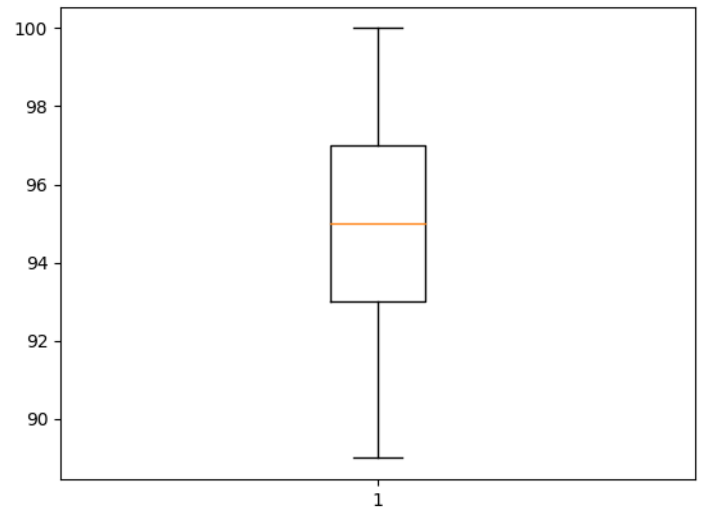
Файл Редагування Формат Вигляд Довідка

46	94
46	94
47	94
49	94
50	94
51	94
51	94
51	94
52	94
53	94
54	94
55	95
56	95
56	95
57	95
58	95
58	95
60	95
60	95
61	95
61	95
61	95
61	95
63	96
64	96
65	96
66	96
67	96
68	96
68	96
68	96
69	96
70	96
70	96
71	97
73	97
73	97
76	97
76	97

77	97
77	97
77	97
78	97
78	97
79	97
79	97
80	98
81	98
81	98
82	98
82	98
82	98
83	98
83	98
85	98
87	98
87	98
88	99
91	99
92	99
92	99
94	99
97	100
97	100
98	100
99	100
99	100
100	100
100	100
100	100



Key: aggr|stem|leaf
 $101 | 10 | 0 = 10.0 \times 10 = 100.0$
101 | 10 | 00000000
93 | 9 | 0000000111111111222222333333444444444444555555555555666666
2 | 8 | 99



Висновки:

Під час виконання цієї лабораторної роботи я навчилася використовувати здобуті знання про лінійні перетворення та графічне зображення даних на практиці за допомогою мови програмування Python.