|  |  |  |  |
| --- | --- | --- | --- |
| 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:** | Пономаренко Андрій Сергійович | **Перевірила**: | Вечерковська Анастасія Сергіївна |
| Група | ІПЗ-25мс | Дата перевірки |  |
| Форма навчання | денна | Оцінка |  |
| Спеціальність | 121 |
| 2022 | | | |

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

**Завдання:**

1. Знайдіть Q1 ,Q2, P90

2. Знайдіть середнє та стандартне відхилення цих оцінок та Z-оцінки.

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.

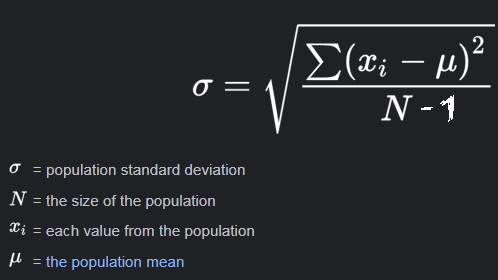
4. Показати дані за допомогою діаграми "стовбур – листя".

5. Відобразити дані за допомогою коробкового графіка.

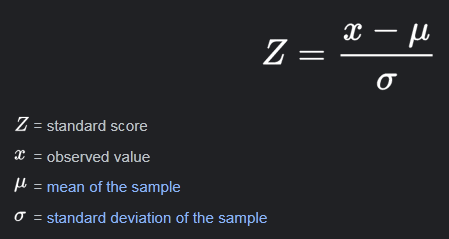
**Рішення:**

1.Спочатку я створив функцію cusquantile що знаходить довільний квантиль з введеними даними за цією формулою

Далі підставляємо туди 25 75 та 90 по завданню. Результат було перевірено самостійно та за допомогою функції вбудованої в numpy.

2.Середнє знаходиться додаванням всіх значень і ділення на кількість. Стандартне відхилення ми знаходим спочатку знаходячи верхню частину цієї формули у for лупі 

Всі інші операції робляться при записуванні devsum у файл

Також за допомогою for ми знаходимо Z-оцінки за цією формулою і виводимо все це у файл.

3.а та b шкали y=ax+b ми знаходимо за допомогою системи

100=100\*a+b

95=Sx\*а+b

Що ми переводимо у матрицю

Та вирішуємо за допомогою numpy.linalg.solve.

Зміненні дані виводимо та дивимось що вони справді задовольняють завдання.

4.Використані бібліотеки pandas та stemgraphic,переводимо наші данні у тип series та використовуємо функцію stem\_graphic.

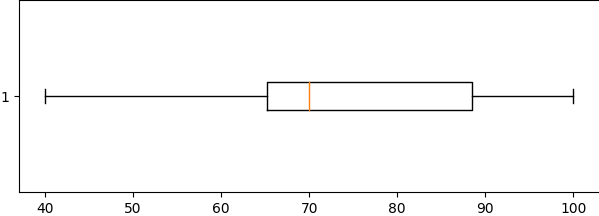
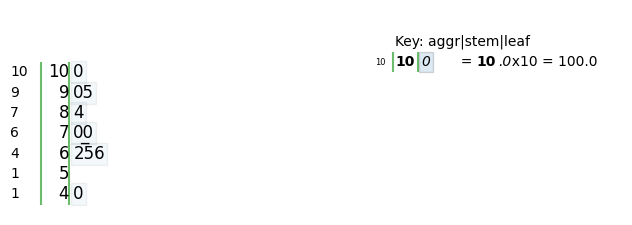
5.Використовуємо matplotlib.pyplot.boxplot для коробкового графіка. У параметрах робим його вертикальним

**Код програми:**

#used libraries  
from numpy import \*  
from numpy.linalg import \*  
from matplotlib.pyplot import \*  
from stemgraphic import \*  
import pandas as pd  
  
#functions  
def cusquantile(cqnda,cloc):#custom function for finding quantiles  
 lower = int(floor((len)\*cloc))-1  
 upper = int(ceil((len)\*cloc))-1  
 return cqnda[lower]+(cqnda[upper]-cqnda[lower])\*cloc  
  
#start  
filename = input()  
f = open(filename,"r")#opens file  
len = int(f.readline())#takes file length from first line  
datarray = []#initialises data array  
for x in range(len):  
 datarray.append(int(f.readline()))  
f.close()#closes input file  
datarray = sort(datarray)  
  
#measurements to txt  
f = open("output from("+filename+").txt","w")  
f.write("P25="+str(cusquantile(datarray,0.25))+'\n')  
f.write("P75="+str(cusquantile(datarray,0.75))+'\n')  
f.write("P90="+str(cusquantile(datarray,0.90))+'\n')  
f.write("average="+str(average(datarray))+'\n')  
devsum = 0.0  
for i in range(len):  
 devsum = devsum + (float(datarray[i])-(float(datarray.sum())/len))\*\*2#half of the standard deviation formula  
f.write("standard deviation = "+str(sqrt(devsum/(len-1)))+'\n')#second half  
zscores = [(value - (datarray.sum()/len)) / sqrt(devsum/(len-1)) for value in datarray]  
f.write("Z-scores="+str(zscores)+'\n')  
  
#figure out the scale formula for 95  
A = array([[100,1],[average(datarray),1]])  
B = array([100,95])  
a,b = solve(A,B)  
f.write("a and b needed to scale grades to 95:\na="+str(a)+'\nb='+str(b))  
f.write("\nfunction:\ny="+str(a)+'\*x+'+str(b)+'\nresulting data:')  
for k in range(len):  
 f.write('\n'+str(round(datarray[k]\*a+b)))  
f.close()  
  
#diagrams  
y = pd.Series(datarray)  
fig, ax = stem\_graphic(y,delimiter\_color='C2',median\_alpha=0,alpha=0.05,break\_on=30)  
savefig("leafstem plot from("+filename+").png",bbox\_inches='tight',pad\_inches=0)#saves file  
clf()#erases the previous plot  
fig = boxplot(datarray,vert=False)  
savefig("boxplot plot from("+filename+").png",bbox\_inches='tight',pad\_inches=0)#saves file

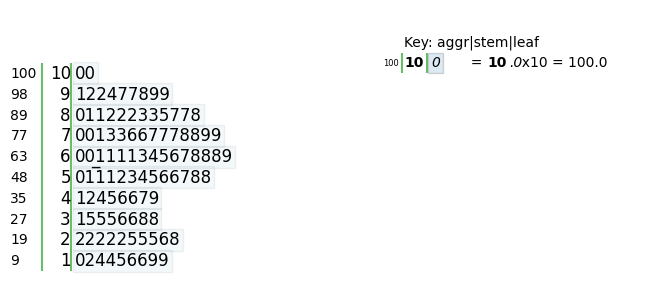
**Рішення задач 4,5,1,2,3**

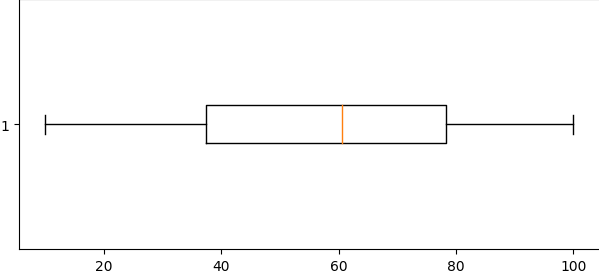
**input\_10.txt**



*P25=62.75  
P75=88.5  
P90=95.0  
average=74.2  
standard deviation = 18.103406677565783  
Z-scores=[-1.8891471980454122, -0.6739063104138606, -0.508191643918649, -0.4529534217535785, -0.23200053309329638, -0.23200053309329638, 0.541334577217691, 0.8727639102081142, 1.1489550210334667, 1.4251461318588194]  
a and b needed to scale grades to 95:  
a=0.19379844961240322  
b=80.62015503875968  
function:  
y=0.19379844961240322\*x+80.62015503875968  
resulting data:  
88  
93  
93  
93  
94  
94  
97  
98  
99  
100*

**input\_100.txt**

****

****

*P25=36.0  
P75=78.0  
P90=91.0  
average=57.96  
standard deviation = 25.280615003387382  
Z-scores=[-1.8971057465798906, -1.8179937471395282,*

…

*, 1.6629342282364177, 1.6629342282364177]  
a and b needed to scale grades to 95:  
a=0.1189343482397716  
b=88.10656517602284  
function:  
y=0.1189343482397716\*x+88.10656517602284  
resulting data:  
89  
90  
90  
90  
90  
90  
90*

… *99  
99  
100  
100  
100  
100  
100  
100  
100*

(… це пропуски даних через їх занадто багату кількість для звіту,повний файл:https://github.com/Andrew-Ponomarenko/laboratorni-iz-yopi/blob/main/lab2/output%20from(input\_100.txt).txt)

**Висновок:**

Я застосував свої знання про методи знаходження властивостей даних та параметрів для лінійних перетворень з допомогою python 3.8 та бібліотек numpy,matplotlib,pandas та stemgraphic.Після повторної перевірки результатів все є правильним.