

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний аерокосмічний університет ім. М.Є. Жуковського
“Харківський авіаційний інститут”

Кафедра комп’ютерних систем, мереж і кібербезпеки

Лабораторна робота № 4

з дисципліни “Обробка даних засобами Python”

Дослідження статистичних методів в Python

Виконав студент гр. 555iM Литвинов.О.А

(підпис, дата)

Перевірив к.т.н зав.кафедри каф. 301

(науковий ступінь, вчене звання, посада)

Дергачов.К.Ю

(підпис, дата)

(П.І.Б.)

2023

Мета роботи: набути практичний досвід використання функцій, які надає вбудована бібліотека `statistics`. Також, відпрацювати вміння працювати з цими інструментами, використовуючи довільний набір даних згідно варіанту завдання.

Варіант 8:

Варіант 8

Початкове положення (123; 345)

№ етапу маршруту	Істинний курс, градус	Швидкість польоту, км / год	Час руху ЛА, хв
1	59	360	59
2	89	200	5
3	46	45	6
4	160	100	50
5	10	155	8
6	50	163	1
7	45	124	18
8	169	146	3
9	128	143	6
10	10	150	16

Результат

Хід роботи

1. Виконати демонстраційні приклади із даними, сформованими особисто.
2. Для одного набору даних (згідно варіанту) отримати оцінку середнього значення, моди, медіани, середньоквадратичного відхилення та дисперсії.
3. У висновках пояснити різницю між оцінками середнього значення, моди та медіани, дати смислову оцінку величині дисперсії.

Код завдання 1:

```
import statistics
import random
from fractions import Fraction as F
from decimal import Decimal as D

print("TASK 1")

print('statistics.mean([34, 12, 65, 15]) = '
      f'{statistics.mean([34, 12, 65, 15])}')
```

```

print('statistics.mean([F(10, 15), F(47, 93), F(5, 1)]) = '
      f'{statistics.mean([F(10, 15), F(47, 93), F(5, 1)])}')
print('statistics.mean([D("3.4"), D("1.2"), D("6.5"), D("1.5")]) = '
      f'{statistics.mean([D("3.4"), D("1.2"), D("6.5"), D("1.5")])}')

print(f'{"-" * 100}')
print('statistics.mean([random.randint(1, 100) for x in range(1, 1001)]) = '
      f'{statistics.mean([random.randint(1, 100) for x in range(1, 1001)])}')
print(
    'statistics.mean([random.triangular(1, 100, 70) for x in range(1, 1001)]) = '
    f'{statistics.mean([random.triangular(1, 100, 70) for x in range(1, 1001)])}'
)

print(f'{"-" * 100}')
print('statistics.mode([random.randint(1, 100) for x in range(1, 1001)]) = '
      f'{statistics.mode([random.randint(1, 100) for x in range(1, 1001)])}')
print('statistics.mode([random.randint(1, 100) for x in range(1, 1001)]) = '
      f'{statistics.mode([random.randint(1, 100) for x in range(1, 1001)])}')
print('statistics.mode([random.randint(1, 100) for x in range(1, 1001)]) = '
      f'{statistics.mode([random.randint(1, 100) for x in range(1, 1001)])}')
print('statistics.mode(["F-35", "F-16", "SU-27", "F-35", "Cessna 172"]) = '
      f'{statistics.mode(["F-35", "F-16", "SU-27", "F-35", "Cessna 172"])}')

print(f'{"-" * 100}')
print('statistics.median([random.randint(1, 100) for x in range(1, 50)]) = '
      f'{statistics.median([random.randint(1, 100) for x in range(1, 51)])}')
print(
    'statistics.median_grouped([random.randint(1, 100) for x in range(1, 50)]) = '
    f'{statistics.median_grouped([random.randint(1, 100) for x in range(1, '
51)])}'
)
print(
    'statistics.median_high([random.randint(1, 100) for x in range(1, 50)]) = '
    f'{statistics.median_high([random.randint(1, 100) for x in range(1, 51)])}'
)
print(
    'statistics.median_low([random.randint(1, 100) for x in range(1, 50)]) = '
    f'{statistics.median_low([random.randint(1, 100) for x in range(1, 51)])}'
)

print(f'{"-" * 100}')
data = [1, 2, 3, 4, 5, 6, 7, 8, 9]
print(f'data = {data}')
print('statistics.pvariance(data) = '
      f'{statistics.pvariance(data)}')
print('statistics.pstdev(data) = '
      f'{statistics.pstdev(data)}')
print('statistics.variance(data) = '
      f'{statistics.variance(data)}')

more_data = [3, 4, 5, 5, 5, 5, 5, 6, 6]

```

```

print(f'more_data = {more_data}')
print('statistics.pvariance(more_data) = '
      f'{statistics.pvariance(more_data)}')
print('statistics.pstdev(more_data) = '
      f'{statistics.pstdev(more_data)}')
print('statistics.variance(more_data) = '
      f'{statistics.variance(more_data)}')

some_fractions = [F(5, 6), F(2, 7), F(11, 76)]
print(f'some_fractions = {some_fractions}')
print('statistics.pvariance(some_fractions) = '
      f'{statistics.pvariance(some_fractions)}')
print('statistics.pstdev(some_fractions) = '
      f'{statistics.pstdev(some_fractions)}')
print('statistics.variance(some_fractions) = '
      f'{statistics.variance(some_fractions)}')

```

```

TASK 1
statistics.mean([34, 12, 65, 15]) = 31.5
statistics.mean([F(10, 15), F(47, 93), F(5, 1)]) = 574/279
statistics.mean([D("3.4"), D("1.2"), D("6.5"), D("1.5")]) = 3.15
-----
statistics.mean([random.randint(1, 100) for x in range(1, 1001)]) = 50.579
statistics.mean([random.triangular(1, 100, 70) for x in range(1, 1001)]) = 55.64089369079807
-----
statistics.mode([random.randint(1, 100) for x in range(1, 1001)]) = 20
statistics.mode([random.randint(1, 100) for x in range(1, 1001)]) = 70
statistics.mode([random.randint(1, 100) for x in range(1, 1001)]) = 76
statistics.mode(["F-35", "F-16", "SU-27", "F-35", "Cessna 172"]) = F-35
-----
statistics.median([random.randint(1, 100) for x in range(1, 50)]) = 47.5
statistics.median_grouped([random.randint(1, 100) for x in range(1, 50)]) = 44.5
statistics.median_high([random.randint(1, 100) for x in range(1, 50)]) = 73
statistics.median_low([random.randint(1, 100) for x in range(1, 50)]) = 41
-----
data = [1, 2, 3, 4, 5, 6, 7, 8, 9]
statistics.pvariance(data) = 6.666666666666667
statistics.pstdev(data) = 2.581988897471611
statistics.variance(data) = 7.5
more_data = [3, 4, 5, 5, 5, 5, 6, 6]
statistics.pvariance(more_data) = 0.7654320987654321
statistics.pstdev(more_data) = 0.8748897637790901
statistics.variance(more_data) = 0.8611111111111112
some_fractions = [Fraction(5, 6), Fraction(2, 7), Fraction(11, 76)]
statistics.pvariance(some_fractions) = 1011151/11462472
statistics.pstdev(some_fractions) = 0.2970084900869067
statistics.variance(some_fractions) = 1011151/7641648

```

Рисунок 1 – Завдання 1

Код завдання 2:

```

import statistics

class Colors:
    DEFAULT = '\033[0m'
    CHANGED = '\033[95m'

```

```
def statistics_calc(data):
    print(f'{Colors.DEFAULT}Mean = {statistics.mean(data)}')
    print(f'Mode = {statistics.mode(data)}')
    print(f'Median = {statistics.median(data)}')
    print(f'St_deviation = {statistics.stdev(data)}')
    print(f'Variance = {statistics.variance(data)}')

print(f'{Colors.CHANGED}Degree:')
statistics_calc([59, 89, 46, 160, 10, 50, 45, 169, 128, 10])
print(f'{Colors.CHANGED}Velocity:')
statistics_calc([360, 25, 158, 200, 160, 63, 240, 58, 65, 50])
print(f'{Colors.CHANGED}Time:')
statistics_calc([5, 4, 18, 3, 6, 16, 18, 12, 6, 16])
```

```
Degree:
Mean = 76.6
Mode = 10
Median = 54.5
St_deviation = 57.84308275171909
Variance = 3345.822222222222
Velocity:
Mean = 137.9
Mode = 360
Median = 111.5
St_deviation = 106.60565963712558
Variance = 11364.766666666666
Time:
Mean = 10.4
Mode = 18
Median = 9.0
St_deviation = 6.186005711819758
Variance = 38.266666666666666
```

Рисунок 2 – Завдання 2

Висновок:

При виконанні цієї роботи використовувалися стандартні методи бібліотеки statistics, а точніше:

1. **Середнє значення(mean):** Це статистична міра, що представляє собою середнє арифметичне всіх значень в наборі. Воно чутливе до викидів, тобто значень, які суттєво відрізняються від інших. Якщо у нас, наприклад, є набір даних про доходи, великі викиди можуть значно вплинути на середнє значення, роблячи його не представницьким для всього набору.
2. **Мода(mode):** Це значення чи значення, які зустрічаються найчастіше в наборі. Мода особливо корисна, коли вам цікаво саме те значення, яке найбільше повторюється. Мода може бути корисною для категоріальних даних або в тих випадках, коли ви хочете знати конкретне значення, яке найчастіше зустрічається в наборі.

3. **Медіана(median):** Це значення, яке знаходиться в середині впорядкованого набору даних. Медіана не чутлива до великих викидів, оскільки вона не залежить від конкретних значень, а лише від їх порядку. Вона особливо корисна в тих випадках, коли у вас є великі викиди, і ви хочете отримати репрезентативне значення "середнього" для вашого набору даних.
4. **Дисперсія(variance):** Дисперсія вимірює розсіювання даних відносно їхнього середнього значення. Велика дисперсія свідчить про те, що дані мають тенденцію розташовуватися далеко від середнього значення, що може вказувати на великі варіації в наборі. Маленька дисперсія свідчить про те, що дані мають тенденцію групуватися близько до середнього значення. Оцінка дисперсії допомагає нам розуміти ступінь розподілу даних і визначити, наскільки набір даних є однорідним або різноманітним.

Інакше кажучи ці параметри вказують на різні аспекти вибірки: типове, найчастіше, середнє значення та розкид даних