МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М. Є. Жуковського

«Харківський авіаційний інститут»

Факультет радіоелектроніки, комп'ютерних систем та інфокомунікацій

**Лабораторна робота № 4 `**

з «Обробка даних засобами Python»

(назва дисципліни)

тема: «Дослідження статистичних методів в Python»

Виконав: студент 5 курсу групи № 555вМн

напряму підготовки (спеціальності)

123 Системне програмування

(шифр і назва напряму підготовки (спеціальності))

Логачов М.Г.

(прізвище й ініціали студента)

Прийняв: д.т.н., доцент

Дергачов К.Ю.

(посада, науковий ступінь, прізвище й ініціали)

Національна шкала: \_\_\_\_\_\_\_\_\_\_

Кількість балів: \_\_\_\_\_

Оцінка: ECTS \_\_\_\_\_

Харків – 2023

Лабораторна робота №4

**Тема**: Дослідження статистичних методів в Python.

**Мета роботи :** Сформувати у здобувачів компетенції використання засобів бібліотеки Statistics для вирішення прикладних завдань зі статистичного аналізу.

**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

Демонстраційні приклади використання модуля статистичного аналізу Python

Text

Description automatically generated

Отриманий результат

Text

Description automatically generated

Дані для обчислювання згідно варіанту (№3).

Table

Description automatically generated

Статистичний аналіз даних допомагає нам зрозуміти всю інформацію загалом. У цього є програми у багатьох областях, таких як біостатистика та бізнес-аналітика. Замість того, щоб проходити через окремі точки даних, лише один погляд на їхнє колективне середнє значення або дисперсію може виявляти тенденції та функції, які ми могли б пропустити, спостерігаючи всі дані в необробленому форматі. Це також робить порівняння між двома великими наборами даних більш простим та зрозумілим. Зважаючи на ці потреби, Python надав нам модуль статистики, який містить в собі функції для знаходження середнього значення(*statistics.mean( )*), значення моди(*statistics.mode( )*), дисперсії (*statistics.pvariance( )*) та інш.

Отримані результати обчислення згідно варіанту представлено у таблиці 1.

Таблиця 1. Результат обчислення

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № п/п |  | Середнє значення | Мода | Медіана | Сер.квадр. відхилення | Дисперсія |
| 1 | Істинний курс, градус | 68.4 | 100 | 46.0 | 81.32 | 5951.64 |
| 2 | Швидкість польоту, км/год | 113.6 | 58 | 79.0 | 85.391 | 6562.44 |
| 3 | Час руху ЛА, хв | 16.3 | 16 | 16.0 | 9.452 | 80.41 |

Програмний код можна знайти за посиланням на [GitHub](https://github.com/lohachovKhai/geodata_processing/tree/main/Lab_4)

Висновки

Середнє значення, медіана та мода є мірою центральної тенденції і часто використовуються для вимірювання великих наборів даних, де потрібно провести аналіз та інтерпретувати результати. Середнє, медіана та мода - це три міри середніх показників, які показують розподіл даних від середнього або середнього. Ці методи широко використовуються в статистиці, тоді як середнє значення даних є найбільш широко застосовуваним методом серед трьох.

Середнє значення - це проста сума чисел спостережень у масиві, яка ділиться на кількість спостережень.

Медіана - це середнє число в наборі масиву даних, яке відокремлює вищий набір даних від нижчого.

Окрім середнього та медіанного, існує ще один метод, який часто використовується для вимірювання центральної тенденції - це режим або мода. Режим / мода - це значення, яке найчастіше зустрічається у наборі даних, режим має перевагу перед середнім та середнім значенням, оскільки його можна знайти як для числового, так і для категоризованого набору даних.

Дисперсія — це міра розсіяння значень випадкової величини щодо середнього значення розподілу. У простому розумінні, дисперсія дозволяє виміряти наскільки далеко випадкові значення розподілені від їх середнього значення. Великі значення дисперсії для швидкості та істинного курсу про більші відхилення значень випадкової величини від центру розподілу.