

Laboratorium nr 1

Podstawowa analiza zbiorów danych

1. Na stronie <https://dane.gov.pl/pl> wybierz interesujące dane, możesz skorzystać z dowolnego źródła z otwartymi danymi i pobrać dane
2. Znajdź Wartość oczekiwaną, medianę, modę, wariancję, odchylenie standardowe (wyjaśnij ich znaczenie(sens))
3. Wizualizuj pobrane dane za pomocą histogramu
4. W MS Teams (zakładka Class Notebook, zakładka odpowiadająca Twojemu Imieniu, zakładka Homework) utwórz folder 02.03.2021_LAB1. Wklej plik danych, plik kodu, raport (.docx lub .doc). Podczas obrony pracy pokaż działanie kodu programu i wyjaśnij swoje działania

Wskazówka:

Wartość oczekiwana

```
In [1]: grades = [85, 93, 45, 89, 85]
```

```
In [2]: sum(grades) / len(grades)
```

```
Out[2]: 79.4
```

Lub

```
In [3]: import statistics
```

```
In [4]: statistics.mean(grades)
```

```
Out[4]: 79.4
```

```
In [5]: statistics.median(grades)
```

```
Out[5]: 85
```

```
In [6]: statistics.mode(grades)
```

```
Out[6]: 85
```

```
In [7]: sorted(grades)
```

```
Out[7]: [45, 85, 85, 89, 93]
```

Wariancja

Mamy ogólny zestaw 10 rzutów sześciokątną kostką:

1, 3, 4, 2, 6, 5, 3, 4, 5, 2

```
In [1]: import statistics
```

```
In [2]: statistics.pvariance([1, 3, 4, 2, 6, 5, 3, 4, 5, 2])  
Out[2]: 2.25
```

Odchylenie standardowe

```
In [3]: statistics.pstdev([1, 3, 4, 2, 6, 5, 3, 4, 5, 2])  
Out[3]: 1.5
```

Lub

```
In [4]: import math
```

```
In [5]: math.sqrt(statistics.pvariance([1, 3, 4, 2, 6, 5, 3, 4, 5, 2]))  
Out[5]: 1.5
```

Wizualizacja (na przykładzie histogramów)

```
ipython --matplotlib
```

```
In [1]: import matplotlib.pyplot as plt
```

```
In [2]: import numpy as np
```

```
In [3]: import random
```

```
In [4]: import seaborn as sns
```

Biblioteki Matplotlib i Seaborn nie wyświetlają automatycznie wykresu utworzonego w skrypcie. Z tego powodu na końcu skryptu dodaje się wywołanie funkcji show biblioteki Matplotlib, która wyświetla okienko z wykresem:

```
plt.show()
```