

UWAGA!

PODCZAS WYKONYWANIA ZADAŃ 1.1, 2.1, 3.1, 4.1, 4.2 NALEŻY ZAIMPLEMENTOWAĆ WŁASNE FUNKCJE, KORZYSTANIE Z WBUDOWANYCH JEST NIEDOZWOLONE (NIE DOTYCZY MECHANIZMÓW POBOCZNYCH, WYKORZYSTYWANYCH JAKO ELEMENT CAŁOŚCI SKRYPTU).

1. Średnia

1.1. Napisz funkcję obliczającą średnią arytmetyczną listy. Dane mogą być dobierane losowo.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Dla $n = 5$:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

2. Mediana

2.1. Napisz funkcję obliczającą medianę danych w liście. Dane mogą być dobierane losowo.

Dla n nieparzystego:

$$M_e = x_{\frac{n+1}{2}}$$

Dla n parzystego:

$$M_e = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$$

Dla $n = 5$:

$$M_e = x_{\frac{5+1}{2}} = x_3$$

Dla $n = 4$:

$$M_e = \frac{1}{2} \left(x_{\frac{4}{2}} + x_{\frac{4}{2}+1} \right) = \frac{1}{2} (x_2 + x_3)$$

3. Dominanta (moda / wartość modalna)

3.1. Napisz funkcję obliczającą dominantę ze zbioru danych. Dane mogą być dobierane losowo.

Możesz wykorzystać jedno z dwóch rozwiązań w zapisie danych:

- a) Dane jako lista
- b) Dane jako słownik z elementami listy oraz liczbą ich wystąpień (wyliczane przez *dictionary comprehensions*). Wartość modalna jako lista, stworzona przez *list comprehensions*.

$$D_0 = x_0 + \frac{(n_0 - n_{-1})}{(n_0 - n_{-1}) + (n_0 - n_{+1})} * h_0$$

gdzie:

x_0 – dolna granica przedziału wartości dominanty

n_0 – liczebność przedziału dominanty

n_{-1} – liczebność przedziału poprzedzającego przedział dominanty

n_{+1} – liczebność przedziału następującego po przedziale dominanty

h_0 – rozpiętość przedziału dominanty

Lub intuicyjnie:

Liczymy ile razy wystąpiły poszczególne wartości i ta, która okaże się najliczniejsza jest dominantą

4. Wariancja

- 4.1. Sprawdź, która z 4 grup wygrała zawody (wyniki uczestników grup możesz dobrać losowo – nazwy drużyn wymyślamy). Która z grup ma mniejszą wariancję?
- 4.2. Narysuj wykres przedstawiający wyniki grup jako punkty (scatters) oraz średnią wyników jako linię ciągłą (plot)

Liczenie wariancji krok po kroku:

1. Wyliczamy średnią wyników
2. Sprawdzamy rozrzut wyniku od średniej poprzez różnicę wyniku i średniej, podniesione do kwadratu
3. Sumujemy rozrzuty wyników
4. Sumę dzielimy przez liczbę wyników
5. *Uzyskaną wartość pierwiastkujemy (odchylenie standardowe)

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

ODCHYLENIE STANDARDOWE:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

5. Korelacja

Mamy dwa zbiory **x** oraz **y** oraz ich średnie: $\bar{x}$ oraz $\bar{y}$

Wówczas:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- $r_{xy} = 1$: silna korelacja
- $r_{xy} = -1$: słaba korelacja
- $r_{xy} = 0$: brak korelacji

Zakresy pomiędzy -1 a 0 oraz 0 a 1 opisuje się różnie i dzieli na podzakresy w zależności od potrzeb i dziedziny życia, jakiej dotyczą analizowane dane.

Np. w naukach medycznych¹:

- $< 0,2$ korelacja słaba
- $0,3-0,5$ korelacja dostateczna
- $0,6-0,7$ korelacja umiarkowana
- $0,8-0,9$ korelacja bardzo silna
- 1 korelacja idealna

¹ <https://predictivesolutions.pl/wspolczynnik-korelacji-r-pearsona>