SEM 5

(2) Dzień dobry Kontynuujemy statystykę – przeczytaj część teoretyczną i rozwiąż zadania.

PRZYKŁAD 1.

W dwóch równoległych klasach liceum odbył się sprawdzian z matematyki. Każdy uczeń mógł otrzymać maksymalnie 35 punktów. Wyniki sprawdzianu zebrano w tabeli.

Liczba punktów		13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Liczba uczniów	Ia			5	5	4	1	4	0	3	4	1	3			
	Ib	1	1	5	3	2	5	4	1	1	2	2	1	1	1	1

Średnia arytmetyczna wyników w klasie Ia jest równa 18,77, a w klasie Ib – 18,80. Średnie niewiele różnią się od siebie, w przeciwieństwie do wyników w poszczególnych klasach. Wyniki sprawdzianu w klasie Ib są bardziej rozproszone niż w klasie Ia.
Rozstęp wyników, czyli różnica między największym a najmniejszym wynikiem, w klasie Ib wynosi 27 – 13 = 14, a w klasie Ia: 24 – 15 = 9.

Rozproszenie wyników opisujemy za pomocą dwóch miar: wariancji i odchylenia standardowego.

Definicja

Wariancją n danych liczbowych $x_1, x_2, x_3, \ldots, x_n$ o średniej arytmetycznej $\overline{x}$ nazywamy liczbę $\sigma^2 = \frac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + (x_3-\overline{x})^2 + \ldots + (x_n-\overline{x})^2}{n}$.
Symbol σ czytamy jako *sigma*.

PRZYKŁAD 2.

Obliczmy wariancję danych: 28, 29, 33, 35, 36, 37.

Średnia arytmetyczna: $\overline{x} = \frac{28+29+33+35+36+37}{6} = 33$.

Wariancja: $\sigma^2 = \frac{(28-33)^2 + (29-33)^2 + (33-33)^2 + (35-33)^2 + (36-33)^2 + (37-33)^2}{6} \approx 11{,}7$.

Zauważmy, że gdyby dane były podane w pewnych jednostkach, np. w metrach, to wariancja byłaby wyrażona kwadratem tej jednostki, czyli w m^2.

Definicja

Odchyleniem standardowym n danych liczbowych $x_1, x_2, x_3, \ldots, x_n$ nazywamy liczbę σ równą pierwiastkowi kwadratowemu z wariancji.

$$\sigma = \sqrt{\frac{(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + (x_3-\overline{x})^2 + \ldots + (x_n-\overline{x})^2}{n}}$$

Odchylenie standardowe określa, o ile wszystkie dane różnią się średnio od ich średniej arytmetycznej. Odchylenie standardowe dla wielkości podanych w jednostkach fizycznych jest wyrażone tą samą jednostką.

PRZYKŁAD 3.

Obliczmy odchylenie standardowe dla zbioru danych: 5, 6, 8, 9.

Średnia arytmetyczna: $\overline{x} = 7$.

Wariancja: $\sigma^2 = \frac{(5-7)^2 + (6-7)^2 + (8-7)^2 + (9-7)^2}{4} = 2{,}5$.

Odchylenie standardowe: $\sigma = \sqrt{2{,}5} \approx 1{,}58$.
Gdy obliczymy odchylenie standardowe σ, to możemy określić przedział $(\overline{x} - \sigma; \overline{x} + \sigma)$, do którego należy około 66% wszystkich danych. W przykładzie 3. jest to przedział $\left(7 - \sqrt{2{,}5}; 7 + \sqrt{2{,}5}\right)$.

Im mniejsza jest wartość odchylenia standardowego, tym więcej liczb jest bliskich średniej arytmetycznej, czyli liczby są bardziej skupione wokół średniej.

PRZYKŁAD 4.

W pierwszych dwóch kolumnach tabeli zamieszczono informacje pogrupowane w przedziały prawostronnie domknięte na temat wieku mieszkańców pewnego osiedla. Wyznaczmy odchylenie standardowe tego zestawu danych.

Wielkości dane		Wielkości obliczone			
Wiek mieszkańców	Liczba mieszkańców n_i	Środek przedziału x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$n_i \cdot (x_i - \overline{x})^2$
(0; 10⟩	4468	5	–24	576	2573568
(10; 20⟩	4754	15	–14	196	931784
(20; 30⟩	4578	25	–4	16	73248
(30; 40⟩	3485	35	6	36	125460
(40; 50⟩	4325	45	16	256	1107200
(50; 60⟩	3926	55	26	676	2653976

Obliczamy średnią arytmetyczną na podstawie danych pogrupowanych w rozłączne klasy.

$$\overline{x} = \frac{4468 \cdot 5 + 4754 \cdot 15 + 4578 \cdot 25 + 3485 \cdot 35 + 4325 \cdot 45 + 3926 \cdot 55}{4468 + 4754 + 4578 + 3485 + 4325 + 3926} \approx 29$$

Przeciętny wiek mieszkańców z grupy danych wynosi 29 lat.

Uzupełniamy kolejne kolumny i obliczamy odchylenie standardowe.

$$\sigma = \sqrt{\frac{2\,573\,568 + 931\,784 + 73\,248 + 125\,460 + 1\,107\,200 + 2\,653\,976}{4468 + 4754 + 4578 + 3485 + 4325 + 3926}} = \sqrt{\frac{7\,465\,236}{25\,536}} = 17{,}1$$

Wiek mieszkańców tej grupy przeciętnie różni się od średniej arytmetycznej o około 17 lat.

A GDYBY SPRAWDZIAN BYŁ TERAZ?

1. Na zajęciach wychowania fizycznego uczniowie rzucali piłką lekarską i uzyskali następujące wyniki (w metrach): 6,4; 7,1; 7,2; 5,5; 5,2; 6,9; 7,6; 4,4; 6,9; 6,2; 8,3; 6,9; 6,1; 5,8; 6,3; 6,4. Wskaż zdanie prawdziwe.

A. Mediana uzyskanych wyników jest równa 4,7.

B. Wyniki rzutów różnią się średnio o około 0,92 m od średniej arytmetycznej wyników.

C. Średnia arytmetyczna wyników rzutu piłką jest równa dominancie tych danych.

D. Wariancja wyników jest dwukrotnie większa od odchylenia standardowego uzyskanego dla danych wielkości.

2. W szpitalu w ciągu tygodnia urodziło się 22 dzieci, których waga wynosiła: 3820 g, 3220 g, 3630 g, 3600 g, 3490 g, 3290 g, 2980 g, 3140 g, 2700 g, 3440 g, 3280 g, 3540 g, 3120 g, 3190 g, 2650 g, 3610 g, 3530 g, 4280 g, 3640 g, 4050 g, 3960 g, 2940 g.

a) Oblicz średnią wagę narodzonych dzieci.

b) Wskaż medianę i modę tych danych.

c) Oblicz wariancję i odchylenie standardowe tych danych.