

ODCHYLENIE STANDARDOWE

Ćwiczenie 1

Oblicz średnią arytmetyczną, medianę i dominantę dla podanych zestawów danych liczbowych.

Zestaw I: 5, 5, 5, 5, 5, 5

Zestaw II: 1, 5, 5, 5, 5, 9

Zestaw III: 1, 4, 5, 5, 6, 9

Zestaw IV: 2, 3, 5, 5, 7, 8

Rozwiązując ćwiczenie, mogłeś się przekonać, że średnie arytmetyczne, mediany oraz dominanty we wszystkich czterech zestawach liczb są równe.

W pierwszym zestawie wszystkie dane są jednakowe, w drugim większość danych (poza dwoma skrajnymi) to te same liczby, a w trzecim i w czwartym zestawie dane są zróżnicowane.

O danych z zestawów III i IV możemy powiedzieć, że są bardzo rozproszone.

Trudno na pierwszy rzut oka ocenić, czy bardziej rozproszony jest zestaw III, czy zestaw IV.

Aby zmierzyć poziom rozproszenia, możemy porównać dane z ich średnią arytmetyczną.

Wielkością, która dobrze opisuje, jak bardzo dane różnią się od średniej arytmetycznej, jest odchylenie standardowe.

DEFINICJA

Odchyleniem standardowym zestawu danych liczbowych $x_1, x_2, x_3, \dots, x_n$, nazywamy liczbę σ , którą obliczamy ze wzoru:

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

gdzie $\bar{x}$ oznacza średnią arytmetyczną liczb $x_1, x_2, x_3, \dots, x_n$.

Wyrażenie $\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$ występujące w powyższym wzorze pod pierwiastkiem, nazywane jest **wariancją**.

Wariancja zatem równa się σ^2 .

Zauważ, że odchylenie standardowe jest tym większe, im bardziej dane różnią się od ich średniej arytmetycznej.

Obliczmy odchylenia standardowe zestawu III i zestawu IV.

$$\sigma_{III} = \sqrt{\frac{(1-5)^2 + (4-5)^2 + (5-5)^2 + (5-5)^2 + (6-5)^2 + (9-5)^2}{6}} = \sqrt{\frac{35}{6}} = \sqrt{\frac{17}{3}} \approx 2,38$$

$$\sigma_{IV} = \sqrt{\frac{(2-5)^2 + (3-5)^2 + (5-5)^2 + (5-5)^2 + (7-5)^2 + (8-5)^2}{6}} = \sqrt{\frac{26}{6}} = \sqrt{\frac{13}{3}} \approx 2,08$$

Odchylenie standardowe zestawu IV jest mniejsze niż zestawu III. Możemy powiedzieć, że w zestawie IV dane są mniej rozproszone niż w zestawie III.

Łatwo zauważyć, że odchylenie standardowe zestawu I jest równe 0 (żadna z danych w tym zestawie nie różni się od średniej).

Ćwiczenie 2

~~~~~

Oblicz odchylenie standardowe zestawu II.

Aby ustalić odchylenie standardowe, na ogół wygodnie jest uporządkować rachunki, zapisując wyniki kolejnych obliczeń w tabelce

## Ćwiczenie 3

~~~~~

Zmierz długości palców u jednej dłoni. Oblicz odchylenie standardowe otrzymanego zestawu danych.

Przykład 1

Poniżej przedstawiono wagi ośmiu cieląt urodzonych w pewnym gospodarstwie.

Oblicz odchylenie standardowe tego zestawu danych.

38 kg 40 kg 42 kg 42kg 45 kg 48 kg 50 kg 51 kg

Średnia arytmetyczna: $\bar{x} = \frac{38 + 40 + 42 + 42 + 45 + 48 + 50 + 51}{8} \approx 44,5$

Obliczenia:

x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
38	-6,5	42,25
40	-4,5	20,25
42	-2,5	6,25
42	-2,5	6,25
45	0,5	0,25
48	3,5	12,25
50	5,5	30,25
51	6,5	42,25

Suma: 160,0

$$\sigma = \sqrt{\frac{160}{8}} = \sqrt{20} \approx 4,47 \quad \text{Odchylenie standardowe jest równe ok.4,47 kg.}$$

Przykład 2

W tabeli zapisano wagi cieląt urodzonych na farmie. Oblicz odchylenie standardowe.

Waga	33	35	38	40	43	46	48	50	51
Ilość wskazań	5	6	6	7	6	10	6	3	1

$$\bar{x} = \frac{5 \cdot 33 + 6 \cdot 35 + 6 \cdot 38 + 7 \cdot 40 + 6 \cdot 43 + 10 \cdot 46 + 6 \cdot 48 + 3 \cdot 50 + 1 \cdot 51}{50} = 41.8$$

c_k to ilość wskazań

c_k	x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$	$c_k(x_k - \bar{x})^2$
5	33	-8,8	77,44	387,20
6	35	-6,8	46,24	277,44
6	38	-3,8	14,44	86,64
7	40	-1,8	3,24	22,68
6	43	1,2	1,44	8,64
10	46	4,2	17,64	176,40
6	48	6,2	38,44	230,64
3	50	8,2	67,24	201,72
1	51	9,2	84,64	84,64

Suma: 1476

$$\sigma = \sqrt{\frac{1476}{50}} \approx 5,43 \quad \text{Odchylenie standardowe jest równe ok. 5,43 kg.}$$

Na podstawie powyższych danych i wyników obliczeń możemy stwierdzić, że cielęta z gospodarstwa ważyły przeciętnie więcej niż urodzone na farmie.

Dane w pierwszym przykładzie były mniej rozproszone niż dane w przykładzie drugim, tzn. wagi cieląt z gospodarstwa mniej różniły się od średniej niż w przypadku cieląt urodzonych na farmie.

W obu powyższych przykładach dane liczbowe dotyczyły zwierząt tego samego gatunku i tych samych ich cech. Mogliśmy zatem bezpośrednio porównać ich parametry statystyczne. Gdyby jednak jeden zestaw danych dotyczył cieląt, a drugi na przykład prosiąt, to porównywanie parametrów statystycznych byłoby bardziej skomplikowane. Łatwo bowiem zauważyć, że odchylenie standardowe równe 1 kg w wypadku cieląt oznacza niewielkie

rozproszenie, a odchylenie 1 kg w wypadku prosiąt oznacza duże zróżnicowanie danych (prosięta po urodzeniu ważą nie więcej niż 1,5 kg).

ZADANIA

1. Oblicz odchylenie standardowe zestawu danych:

- a) 2, 4, 5, 9, 10
- b) 10, 20, 20, 30, 40, 40, 50
- c) 4, -3, -1, 0, 2, 3
- d) 1, 1, 1, 4, 7, 8, 9, 9

2. Liczby w zestawach danych oznaczają wyniki pięciu rzutów kostką. Spróbuj odgadnąć, w którym z trzech zestawów średnia jest największa, a w którym najmniejsze jest odchylenie standardowe. Następnie sprawdź swoje przypuszczenia, wykonując odpowiednie obliczenia,

a)	Zestaw I	1	2	3	4	5
	Zestaw II	1	2	4	5	6
	Zestaw III	1	2	3	5	6

b)	Zestaw I	1	1	1	6	6
	Zestaw II	1	1	6	6	6
	Zestaw III	1	1	3	6	6

c)	Zestaw I	1	1	1	1	6
	Zestaw II	1	6	6	6	6
	Zestaw III	3	3	3	3	3

d)	Zestaw I	1	2	2	2	6
	Zestaw II	1	3	3	3	6
	Zestaw III	1	1	3	6	6

3. Oblicz średnie arytmetyczne i odchylenia standardowe dwóch podanych zestawów danych. Porównaj obliczone wielkości.

- a) Czasy finalistów w biegu na 100 m podczas Olimpiady w Sydney w roku 2000.
 kobiety: 10,75 s, 11,12 s, 11,18 s, 11,19 s, 11,20 s, 11,21 s, 11,21 s, 11,29 s
 mężczyźni: 9,87 s, 9,99 s, 10,04 s, 10,08 s, 10,09 s, 10,13 s, 10,17 s.

b) Średnia miesięczna temperatura powierzchni wód.

Miesiąc	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Morze Bałtyckie	2,7°	1,7°	,9°	3,7°	7,5°	12,8°	15,8°	15,9°	14,2°	11,5°	8,9°	4,9°
Morze Czerwone	25,8°	25,4°	26,3°	27,7°	29,2°	30,3°	31,0°	31,4°	31,6°	30,9°	28,6°	26,7°

c) Miejsca Adama Małysza i Svena Hannawalda w ostatnich 10 konkursach Pucharu Świata 2003 roku w skokach narciarskich.

Miejsce konkursu	Sapporo	Tauplitz	Willingen	Oslo	Lahti	Planica
Adam Małysz	6	3 4	– –	1	1 1	2 4
Sven Hannawald	–	2 1	1 36	14	3 10	4 2

