4.2. Mediana, skala centylowa i dominanta

Poza średnią arytmetyczną rozpatrujemy też inne wielkości służące do analizy danych. Jedną z nich jest wartość środkowa – **mediana**. Aby ją wyznaczyć, porządkujemy dane od wartości najmniejszej do największej, na przykład:

$1, 2, 2, \boxed{2}, 5, 5, 6$ – medianą jest liczba 2,

$-3, -3, 0, 0, 2, \boxed{7}, 9, 11, 13, 17, 18$ – medianą jest liczba 7.

Medianą nieparzystej liczby danych jest wartość środkowa. W przypadku parzystej liczby danych medianą jest średnia arytmetyczna dwóch sąsiednich wartości środkowych, na przykład:

$1, 2, \boxed{3}, \boxed{8}, 9, 14$ – mediana jest równa $\frac{3+8}{2} = 5{,}5$.

Zauważ, że mediana dzieli dane na dwie równoliczne grupy. Dane w jednej grupie są od niej mniejsze lub jej równe, dane w drugiej – większe lub równe.

Ćwiczenie 1

Wyznacz medianę podanych liczb.

a) 1, 2, 3, 10, 20 b) 6, 10, 11, 11, 20, 7 c) 18, 6, 4, 10, 5, 4

Ćwiczenie 2

W stadninie zważono wszystkie konie i otrzymano następujące wyniki (w kilogramach):

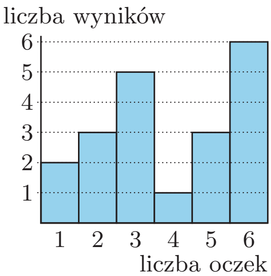
– ogiery: 530, 550, 550, 590, 565, 570, 560, 540;

– klacze: 490, 500, 510, 540, 505, 500.

Wyznacz mediany wagi: a) ogierów, b) klaczy, c) wszystkich koni w stadninie.

Ćwiczenie 3

Rzucono 20 razy kostką. Na diagramie obok przedstawiono, ile razy otrzymano poszczególne liczby oczek. Wyznacz medianę uzyskanych wyników.



Ćwiczenie 4

Podaj przykład pięciu liczb, których mediana jest:

a) większa od ich średniej arytmetycznej,

b) mniejsza od ich średniej arytmetycznej.

Ćwiczenie 5

Średnia arytmetyczna zestawu liczb: $8, 6, 3, x, 8, x+4, 4, 6, 7, 8$ jest równa 7. Ile wynosi mediana tego zestawu liczb?

Aby wskazać położenie wybranej danej względem innych danych, używa się **skali centylowej**. Podając centyl, któremu odpowiada liczba x (mówimy też: w którym mieści się liczba x), określamy, jaki jest procent liczb mniejszych lub równych tej liczbie.

Przykład 1

W poniższej tabeli podano, jakie wartości centyli odpowiadają poszczególnym wynikom procentowym uzyskanym na egzaminie maturalnym z matematyki na poziomie podstawowym przeprowadzonym w maju 2019 roku.

Wynik procentowy	Wartość centyla	Wynik procentowy	Wartość centyla	Wynik procentowy	Wartość centyla
0	1	34	21	68	64
2	1	36	23	70	66
4	1	38	26	72	69
6	1	40	28	74	71
8	1	42	30	76	74
10	2	44	33	78	76
12	3	46	35	80	78
14	4	48	38	82	80
16	5	50	40	84	82
18	7	52	43	86	85
20	8	54	45	88	87
22	10	56	48	90	89
24	12	58	51	92	91
26	13	60	53	94	93
28	15	62	56	96	96
30	17	64	58	98	98
32	19	66	61	100	100

Źródło: https://cke.gov.pl

Z tabeli można odczytać, że:

- Jeśli maturzysta uzyskał na egzaminie 30% punktów możliwych do zdobycia, to wynik ten odpowiada 17. centylowi, co oznacza, że 17% maturzystów podchodzących do tego egzaminu uzyskało wynik niższy lub taki sam, natomiast 83% maturzystów uzyskało wynik wyższy.
- Jeśli maturzysta uzyskał na egzaminie 70% możliwych do zdobycia punktów, to wynik ten odpowiada 66. centylowi, co oznacza, że 66% maturzystów podchodzących do tego egzaminu uzyskało wynik niższy lub taki sam, natomiast 34% maturzystów uzyskało wynik wyższy.

Ćwiczenie 6

a) Ile procent maturzystów podchodzących do egzaminu maturalnego z matematyki podstawowej w maju 2019 roku uzyskało wynik lepszy od maturzysty, który zdobył na tym egzaminie 50% punktów?

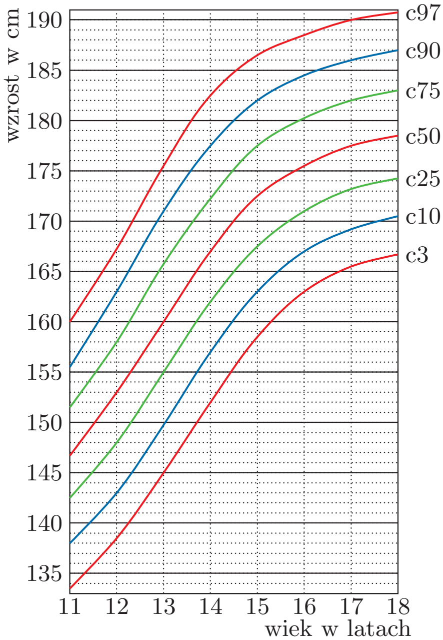
b) Jaki procentowy wynik z egzaminu maturalnego z matematyki podstawowej w maju 2019 roku odpowiada 40. centylowi, a jaki – 80. centylowi wyników egzaminu?

Przykład 2

Przedstawiona obok siatka jest siatką centylową wzrostu chłopców w wieku 11–18 lat.

Zaznaczone linie opisują centyle: 3, 10, 25, 50, 75, 90 i 97.

- W wieku 13 lat chłopiec o wzroście 160 cm mieści się w 50. centylu.
- W wieku 16 lat chłopiec o wzroście większym (lub równym) od 75% swoich rówieśników mierzy ponad 180 cm.



Siatka centylowa wzrostu chłopców w wieku 11–18 lat
Źródło: http://www.czd.pl

Ćwiczenie 7

a) Czternastoletni chłopiec ma 167 cm wzrostu. W którym centylu mieści się jego wzrost?

b) Siedemnastolatek ma 186 cm wzrostu. Ile procent jego równieśników jest od niego wyższych?

Kolejną wielkością przydatną podczas analizy danych jest **dominanta** – wartość, która występuje wśród danych najczęściej (dominanta bywa również nazywana **wartością modalną** lub **modą**).
Na przykład dla liczb: 1, 1, 2, 2, 2, 3, 5, 5, 6 dominantą jest liczba 2.
Jeśli w zestawie danych kilka liczb występuje z tą samą, najwyższą częstością, to przyjmujemy, że każda z tych liczb jest dominantą. Jeżeli natomiast wszystkie liczby występują tak samo często, to przyjmujemy, że nie ma dominanty.

Ćwiczenie 8

Sprzedawca zanotował rozmiary butów męskich, które sprzedał pewnego dnia: 42, 44, 41, 42, 43, 42, 44, 42, 45, 43, 45, 46. Wyznacz medianę i dominantę tych danych. Który rozmiar butów był kupowany najczęściej?

Ćwiczenie 9

Nauczyciel biologii zrobił zestawienie wyników trzech sprawdzianów przeprowadzonych w dwudziestoosobowej klasie (tabela obok). Po kolejnym sprawdzianie dopisał do niego nowe oceny. Wyznacz dominantę i medianę ocen w nowym zestawieniu, jeśli wiadomo, że z tego sprawdzianu:

Ocena	1	2	3	4	5	6
Liczba ocen	4	10	23	2	19	2

a) wszyscy uczniowie otrzymali ocenę dobrą,

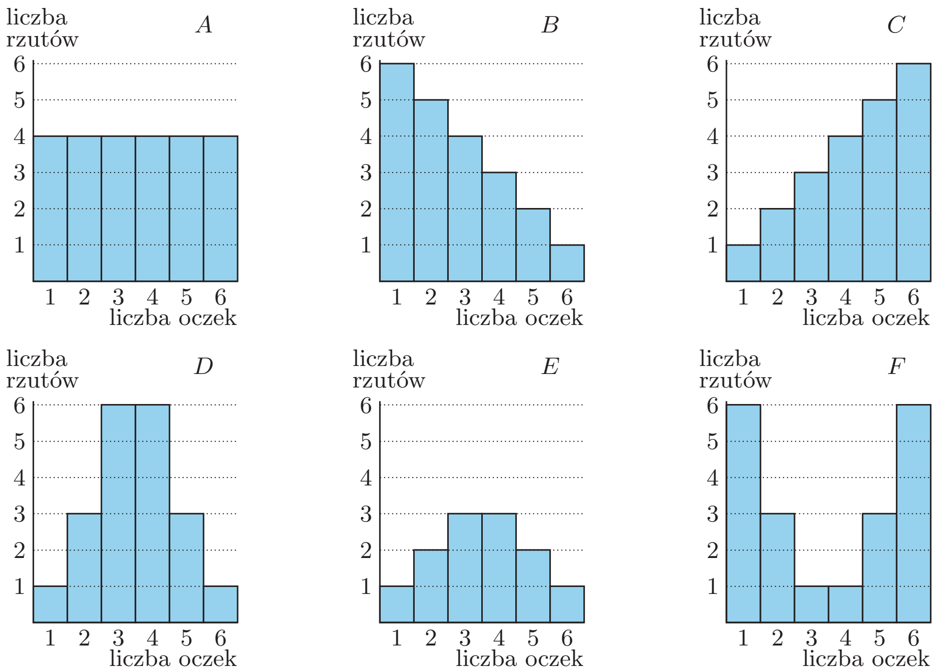
b) połowa uczniów otrzymała ocenę bardzo dobrą, a pozostali – niedostateczną.

Zadania

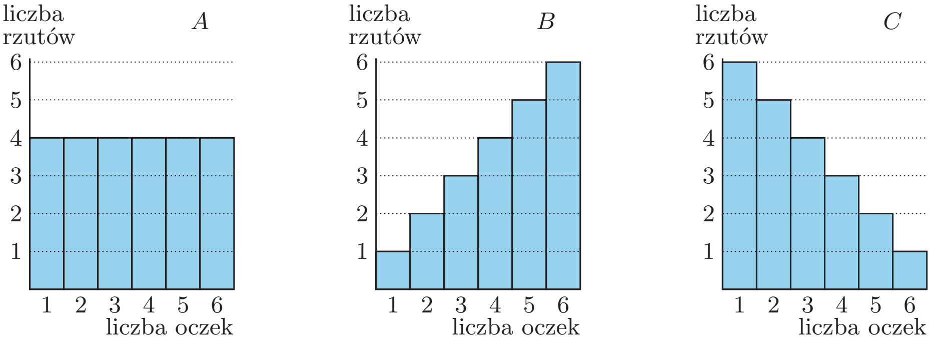
1. Oznaczmy przez $\overline{x}$, M i D odpowiednio średnią arytmetyczną, medianę i dominantę zestawu liczb: 1, 2, 3, 4, c, 2, 3, 4, 4, 4. Oblicz c, jeśli:

a) $\overline{x} = D$, b) $\overline{x} = M$.

2. Każda z osób A, B, C, D, E, F wykonała serię rzutów kostką. Na diagramach przedstawiono, ile razy wypadły poszczególne liczby oczek w seriach. Wyznacz medianę, dominantę i średnią arytmetyczną liczb oczek wyrzuconych w każdej serii.



3. Każda z osób A, B, C wykonała serię rzutów kostką. Na diagramach przedstawiono, ile razy wypadły poszczególne liczby oczek w seriach. Wyznacz medianę, dominantę i średnią arytmetyczną liczb oczek wyrzuconych łącznie przez osoby: a) A i B, b) A i C, c) B i C.



Czy wiesz, że…

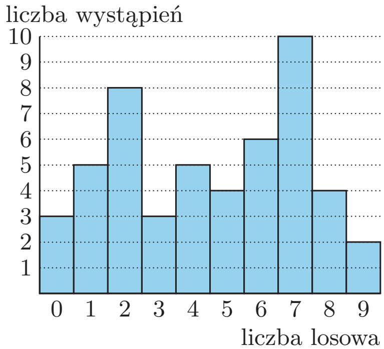
W niektórych zagadnieniach (np. w kryptografii) wykorzystuje się liczby losowe. Liczby te można wygenerować za pomocą programu komputerowego lub można skorzystać z odpowiednich tablic. Poniżej podano osiem wierszy tablicy liczb losowych, w której występują liczby od 0 do 9 odpowiednio pogrupowane.

1.	46016	24742	21311	88342	67778	20741	26755	55382	96777	06729
2.	89311	86439	32128	17700	90725	01936	23678	54622	27342	96439
3.	09006	57476	64080	47646	68020	32924	68500	81779	17120	57784
4.	84642	87958	96983	80086	19451	53725	82606	54516	62617	67000
5.	88645	63989	45783	33657	54733	68580	97232	98073	27454	36174
6.	75815	34604	83791	91421	09176	49317	17045	79972	65081	32075
7.	37019	94048	44462	48948	29999	19107	51184	89352	00122	07222
8.	01324	69795	95403	20891	89075	82476	02147	19961	84203	02724

4. Wykonano diagram dla liczb z pierwszego wiersza powyższej tablicy liczb losowych.

a) Oblicz ich średnią arytmetyczną oraz medianę. Któremu centylowi odpowiada liczba 2, a któremu liczba 7?

b) Wykonaj analogiczny diagram dla liczb z drugiego wiersza tablicy. Wyznacz ich średnią arytmetyczną, medianę i dominantę. Któremu centylowi odpowiada liczba 3?



5. Poniżej przedstawiono tabelę zawierającą skalę centylową wyników egzaminu maturalnego z historii (poziom rozszerzony, maj 2019 roku).

Wynik procentowy	Wartość centyla	Wynik procentowy	Wartość centyla	Wynik procentowy	Wartość centyla
0	1	34	53	68	90
2	2	36	56	70	91
4	3	38	59	72	92
6	6	40	62	74	93
8	10	42	65	76	95
10	13	44	67	78	95
12	17	46	69	80	96
14	20	48	71	82	97
16	24	50	74	84	98
18	27	52	76	86	98
20	31	54	78	88	99
22	34	56	80	90	99
24	38	58	82	92	99
26	41	60	84	94	99
28	44	62	85	96	100
30	47	64	87	98	100
32	50	66	88	100	100

Źródło: https://cke.gov.pl

a) Jaki wynik procentowy gwarantował znalezienie się w gronie 16% maturzystów, którzy uzyskali najlepsze wyniki na tym egzaminie?

b) Wynik procentowy pewnego maturzysty równy jest medianie wyników egzaminu, a wynik jego kolegi odpowiada 80. centylowi. O ile punktów procentowych różnią się ich wyniki z tego egzaminu?

Powtórzenie

6. W klasie Marty 10% uczniów otrzymało ze sprawdzianu ocenę bardzo dobrą, a pozostali uzyskali ocenę dobrą albo dostateczną. Średnia wszystkich ocen wyniosła 3,6. Wyznacz medianę i dominantę ocen z tego sprawdzianu.

7. Znajdź taką liczbę całkowitą x, aby poniższy zestaw danych miał tylko jedną dominantę i aby była ona równa jego średniej arytmetycznej.

a) $4, x, 2, 5, 4, 4, 5, 12, 6, 5, 3$　　b) $1, x, 2x, 1, 6, 9, 1, 6, 9, 6, 9$

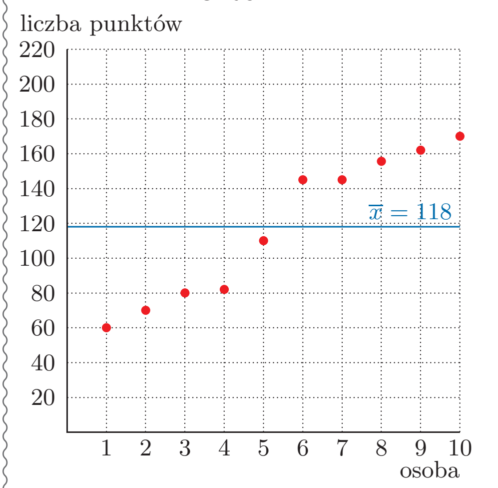
8. Skorzystaj z tabeli zamieszczonej przy zadaniu 5. i podaj, ile procent maturzystów podchodzących do egzaminu z historii w maju 2019 roku uzyskało wynik lepszy od maturzysty, który zdobył na tym egzaminie 68% punktów.

4.3. Odchylenie standardowe

Przykład 1

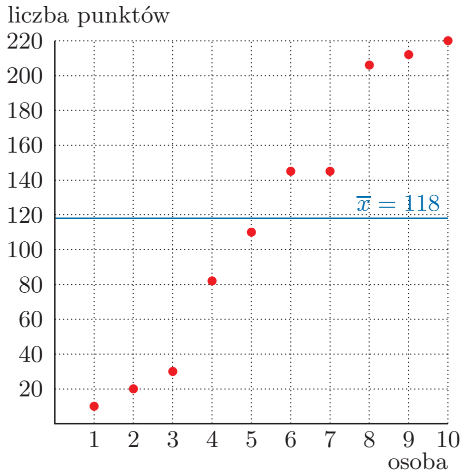
W dwóch dziesięcioosobowych grupach studentów przeprowadzono ten sam egzamin oceniany w skali 0–220 punktów. Wyniki otrzymane w grupach A i B przedstawiono na diagramach.

GRUPA A



Wyniki w grupie A: 60, 70, 80, 82, 110, 145, 145, 156, 162, 170

GRUPA B



Wyniki w grupie B: 10, 20, 30, 82, 110, 145, 145, 206, 212, 220

Obydwa zestawy danych mają średnią arytmetyczną równą 118, medianę – 127,5 oraz dominantę – 145. Jednak rozproszenie wyników w grupie B jest znacznie większe niż w grupie A. Jako miarę rozproszenia danych wokół ich średniej przyjmuje się **odchylenie standardowe**.

Definicja

Wariancją liczb: $x_1, x_2, \ldots, x_n$ nazywamy liczbę:

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

gdzie $\overline{x}$ jest średnią arytmetyczną liczb: $x_1, x_2, \ldots, x_n$.

Odchyleniem standardowym liczb: $x_1, x_2, \ldots, x_n$ nazywamy liczbę σ określoną za pomocą wzoru:

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

Uwaga. Wariancja jest równa σ^2 i jest w ten sposób oznaczana.

Przykład 2

Oblicz wariancję i odchylenie standardowe danych: 4, 9, 11, 13, 13.

Obliczamy średnią i wariancję:

$$\overline{x} = \frac{4+9+11+13+13}{5} = 10$$

$$\sigma^2 = \frac{(4-10)^2+(9-10)^2+(11-10)^2+(13-10)^2+(13-10)^2}{5} = \frac{56}{5} = 11{,}2$$

Zatem odchylenie standardowe $\sigma = \sqrt{11{,}2} \approx 3{,}35$.

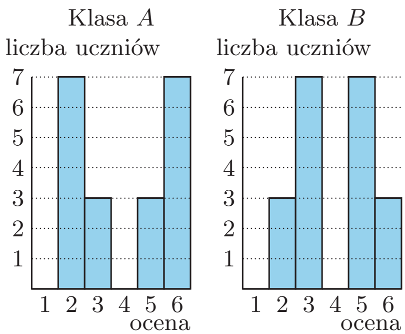
Ćwiczenie 1

Oblicz wariancję i odchylenie standardowe danych:

a) 4, 5, 6, 7, 8; b) 3, 6, 6, 6, 9; c) 8, 12, 13, 13, 14.

Przykład 3

Na diagramach przedstawiono oceny semestralne z muzyki w dwóch dwudziestoosobowych klasach. Średnia ocen w obu klasach jest równa 4. Pokaż, że odchylenie standardowe ocen jest większe w klasie A.



Obliczając wariancję w klasie A, wykorzystujemy pogrupowanie danych:

$$\sigma_A^2 = \frac{7\cdot(2-4)^2+3\cdot(3-4)^2+3\cdot(5-4)^2+7\cdot(6-4)^2}{20} = \frac{62}{20} = 3{,}1$$

Odchylenie standardowe $\sigma_A = \sqrt{3{,}1} \approx 1{,}76$.

Analogicznie obliczamy wariancję w klasie B:

$$\sigma_B^2 = \frac{3\cdot(2-4)^2+7\cdot(3-4)^2+7\cdot(5-4)^2+3\cdot(6-4)^2}{20} = \frac{38}{20} = 1{,}9$$

Odchylenie standardowe $\sigma_B = \sqrt{1{,}9} \approx 1{,}38$. Zatem $\sigma_A > \sigma_B$.

Ćwiczenie 2

Sprawdź, dla którego zestawu danych, A czy B, odchylenie standardowe jest większe.

A: 1, 1, 1, 2, 2, 3, 4, 4, 5, 5, 5 B: 1, 2, 2, 3, 3, 3, 4, 4, 5

Ćwiczenie 3

Oblicz odchylenia standardowe zestawów liczb A i B. Sformułuj wniosek.

A: 1, 2, 2, 3, 3, 3, 4, 4, 4, 4 B: 11, 12, 12, 13, 13, 13, 14, 14, 14, 14