



Wariacja

- [Wprowadzenie](#)
- [Przeczytaj](#)
- [Galeria zdjęć interaktywnych](#)
- [Sprawdź się](#)
- [Dla nauczyciela](#)



Pustynia kamienista – Sahara w Algierii

Źródło: Florence Devouard, dostępny w internecie: commons.wikimedia.org, licencja: CC BY-SA 3.0.

Na pewno wiesz, że Sahara to największa najgorętsza pustynia na Ziemi. Zapewne kojarzy ci się z bezkresnym piaszczystym obszarem. A jak jest w rzeczywistości? Jak myślisz – ile procent powierzchni Sahary pokrywa piasek?

Nie wiem jaką dasz odpowiedź, ale większość ludzi twierdzi, że około 80%, choć jest wręcz przeciwnie – tylko 15% powierzchni Sahary pokrywa piasek. Dzieje się tak dlatego, że powierzchnia pustyni jest bardzo zróżnicowana, w dużej mierze skalista.

Widać więc, że uśrednianie danych, w przypadku dużego ich zróżnicowania, może całkowicie wypaczyć obraz pewnego zjawiska czy sytuacji. Aby unikać takich pułapek, w statystyce, oprócz miar tendencji centralnej, stosuje się jeszcze inne miary. Z niektórymi z nich zapoznasz się w tym materiale.

Twoje cele

- Poznasz niektóre miary rozproszenia i porównasz je z miarami tendencji centralnej.
- Obliczysz rozstęp, odchylenie przeciętne i wariancję danych przedstawionych w różny sposób.
- Przeanalizujesz i zinterpretujesz wariancję danego zestawu danych statystycznych.

Przeczytaj

Miary rozproszenia

Rozpatrzmy wyniki dwóch serii rzutów kostką do gry.

Seria 1	Seria 2
4, 4, 4, 4, 4	1, 3, 4, 6, 6

W obu przypadkach średnia arytmetyczna liczby wyrzuconych oczek jest równa 4. Mediana obu zestawu danych też jest równa i wynosi 4. Jednak oba te zestawy danych wyraźnie się różnią. Pierwszy zestaw nie jest zróżnicowany, a drugi – zróżnicowany. Widać więc, że zastosowanie miar tendencji centralnej nie opisuje dobrze różnic między tymi zestawami.

Aby więc analiza danych była pełniejsza, warto zastosować jeszcze charakterystyki zróżnicowania (rozproszenia) danych, zwane miarami rozproszenia (dyspersji). Miary te pozwalają na określenie, jak duże są różnice (odchylenia) między poszczególnymi wartościami jednostek zbiorowości, a ich wartością przeciętną (średnią).

Definicja: Miara rozproszenia

Miary rozproszenia (rozrzutu, zmienności, dyspersji) to miary charakteryzujące stopień zróżnicowania między sobą jednostek statystycznych pod względem badanej cechy.

Przykładowe miary rozproszenia to rozstęp, wariancja, odchylenie standardowe.

Rozstęp

Rozstęp jest miarą służącą do wstępnej analizy rozproszenia.

Definicja: Rozstęp

Rozstępem (obszarem zmienności) nazywamy różnicę między największą a najmniejszą wartością cechy w szeregu statystycznym.

Rozstęp oznaczamy literą R .

$$R = x_{max} - x_{min},$$

gdzie:

x_{max} – największa wartość cechy,

x_{min} – najmniejsza wartość cechy.

Przykład 1

W tabeli przedstawiono oceny z historii uzyskane przez Anię, Ewę i Julka.

Osoba	Oceny z historii
Ania	4, 5, 6, 4, 5, 4, 6
Ewa	3, 4, 4, 4, 4, 3, 3, 3
Julka	2, 1, 3, 4, 4, 4, 3, 5

$$R = 6 - 4 = 2$$

$$R = 4 - 3 = 1$$

$$R = 5 - 1 = 4$$

Odchylenie przeciętne

Definicja: Odchylenie od średniej

Odchyleniem wartości x_i cechy statystycznej od średniej arytmetycznej \bar{x} nazywamy liczbę $|x_i - \bar{x}|$.

Przykład 2

Policzono ile bombek zawieszono na 4 choinkach stojących na Placu Ratuszowym. Otrzymano następujące wyniki: 126, 47, 24, 183.

Obliczymy średnią arytmetyczną liczby bombek i dla każdego wyniku podamy odchylenie od średniej liczb bombek.

Rozwiązanie:

Obliczamy średnią arytmetyczną:

$$\bar{x} = \frac{126+47+24+183}{4} = 95$$

Obliczamy odchylenie od średniej.

Argumenty i Wartości				
x_i	126	47	24	183
$ x_i - \bar{x} $	$ 126 - 95 = 31$	$ 47 - 95 = 48$	$ 24 - 95 = 71$	$ 183 - 95 = 88$

Wniosek:

Największe odchylenie od średniej jest w przypadku choinki, na której zawieszono 183 bombki.

Miarą rozproszenia, która uwzględnia wszystkie dane rozkładu (a nie poszczególne elementy – tak jak odchylenie od średniej), jest odchylenie przeciętne (średnie).

Definicja: Odchylenie przeciętne

Odchyleniem przeciętnym zestawu danych statystycznych x_1, x_2, \dots, x_n od ich średniej arytmetycznej \bar{x} nazywamy liczbę:

$$d = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

Przykład 3

Obliczymy odchylenie przeciętne dla zestawu danych z Przykładu 2.

$$d = \frac{31+48+71+88}{4} = 59,5 \approx 60$$

Możemy powiedzieć, że liczby bombek na poszczególnych choinkach różnią się o ok. 60 od średniej dla wszystkich choinek.

Wariancja

Podstawową miarą zmienności obserwowanych wyników jest wariancja. Wariancja informuje o tym, jak duże jest zróżnicowanie wyników w danym zbiorze danych – czy wyniki są bardziej czy mniej skoncentrowane wokół średniej.

Definicja: Wariancja

Wariancją zestawu danych statystycznych x_1, x_2, \dots, x_n nazywamy średnią arytmetyczną kwadratów odchyleń od ich średniej arytmetycznej \bar{x} .

Wariancję oznaczamy symbolem σ^2 (σ – sigma) i określamy wzorem:

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

Przykład 4

W loterii fantowej wzięły udział 3 osoby. Każda wyciągnęła 10 losów. Pierwsza z osób wyciągnęła 4 losy pełne, druga 6, a trzecia 2. Obliczymy wariancję wyciągnięcia losów pełnych.

Obliczamy średnią arytmetyczną liczb: 4, 6, 2.

$$\bar{x} = \frac{4+6+2}{3} = 4$$

Obliczamy wariancję.

$$\sigma^2 = \frac{(4-4)^2 + (6-4)^2 + (2-4)^2}{3} = \frac{0+4+4}{3} = \frac{8}{3} \approx 2,7$$

Wariancja jest równa w przybliżeniu 2, 7.

Przykład 5

Obliczmy wariancję dla zestawu danych zapisanych w tabeli liczebności.

Argumenty i Wartości				
Wartość x_i cechy	2	4	6	10
Liczebność n_i	2	5	1	2

Obliczamy średnią arytmetyczną.

$$\bar{x} = \frac{2 \cdot 2 + 4 \cdot 5 + 6 \cdot 1 + 10 \cdot 2}{2 + 5 + 1 + 2} = \frac{4 + 20 + 6 + 20}{10} = 5$$

Obliczmy wariancję.

$$\sigma^2 = \frac{2 \cdot (2-5)^2 + 5 \cdot (4-5)^2 + 1 \cdot (6-5)^2 + 2 \cdot (10-5)^2}{10}$$

$$\sigma^2 = \frac{2 \cdot 9 + 5 \cdot 1 + 1 \cdot 1 + 2 \cdot 25}{10} = \frac{74}{10} = 7,4$$

Wariancja dla podanego zestawu danych jest równa 7, 4.

Przykład 6

Obliczmy rozstęp, średnią arytmetyczną, odchylenie przeciętne i wariancję dla zestawu danych: $-4, 10, -1, 0, 5$.

Rozwiązanie:

Zapisujemy dane w postaci uporządkowanego szeregu statystycznego.

x_i	x_1	x_2	x_3	x_4	x_5
	-4	-1	0	5	10

Obliczamy rozstęp.

$$R = x_5 - x_1 = 10 - (-4) = 14$$

Obliczamy średnią arytmetyczną.

$$\bar{x} = \frac{-4 - 1 + 0 + 5 + 10}{5} = 2$$

Obliczamy odchylenie przeciętne.

$$d = \frac{|-4-2| + |10-2| + |-1-2| + |0-2| + |5-2|}{5} = \frac{6+8+3+2+3}{5} = 4,4$$

Obliczamy wariancję.

$$\sigma^2 = \frac{(-4-2)^2 + (10-2)^2 + (-1-2)^2 + (0-2)^2 + (5-2)^2}{5}$$

$$\sigma^2 = \frac{6^2 + 8^2 + 3^2 + 2^2 + 3^2}{5} = \frac{122}{5} = 24,4$$

Na podstawie wartości uzyskanych parametrów zauważamy, że zróżnicowanie danych jest duże (różnica między wartością największą a najmniejszą to aż 14, wariancja to 24,4).

Słownik

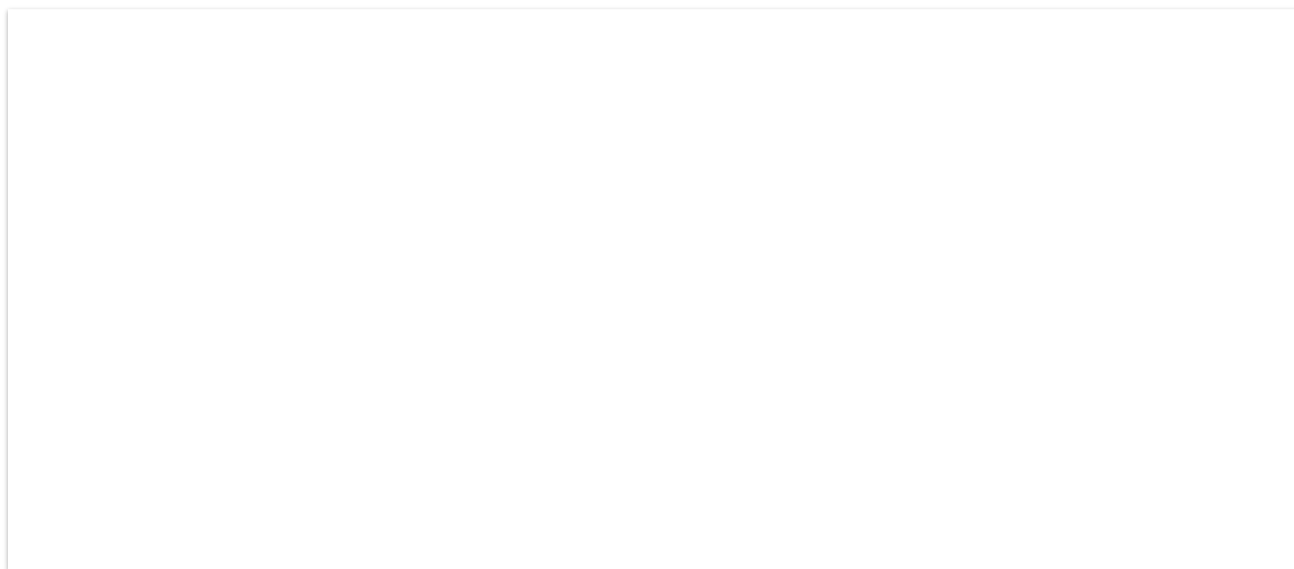
wariancja zestawu danych statystycznych

średnia arytmetyczna kwadratów odchyleń od ich średniej arytmetycznej \bar{x}

Galeria zdjęć interaktywnych

Polecenie 1

Zapoznaj się z przykładami wyznaczania miar rozproszenia pokazanymi w galerii zdjęć interaktywnych. Zinterpretuj w każdym przypadku uzyskane wyniki.



Polecenie 2

Dla zestawu danych: $-6, -2, 0, 2, 6, 12$ oblicz rozstęp, średnią arytmetyczną, odchylenie przeciętne, wariancję.

Sprawdź się

Pokaż ćwiczenia:   

Ćwiczenie 1



Ćwiczenie 2



Ćwiczenie 3



Ćwiczenie 4



Ćwiczenie 5



Ćwiczenie 6



Ćwiczenie 7



Wariancja danych zapisanych w tabelce poniżej jest równa:

Argumenty i Wartości				
Wartość x_i cechy	1	2	4	10
Liczebność n_i	2	5	2	1

Ćwiczenie 8



Maciek w tym semestrze otrzymał z geografii trzy czwórki, dwie piątki, cztery trójki i szóstkę. Oblicz wariancję dla tych danych.

Dla nauczyciela

Autor: Justyna Cybulska

Przedmiot: Matematyka

Temat: Wariancja

Grupa docelowa:

III etap edukacyjny, liceum, technikum, zakres rozszerzony

Podstawa programowa:

XII. Rachunek prawdopodobieństwa i statystyka. Zakres podstawowy.

Uczeń:

3) oblicza średnią arytmetyczną i średnią ważoną, znajduje medianę i dominantę;

4) oblicza odchylenie standardowe zestawu danych (także w przypadku danych odpowiednio pogrupowanych), interpretuje ten parametr dla danych empirycznych.

Kształtowane kompetencje kluczowe:

- kompetencje w zakresie rozumienia i tworzenia informacji
- kompetencje matematyczne oraz kompetencje w zakresie nauk przyrodniczych, technologii i inżynierii
- kompetencje cyfrowe
- kompetencje osobiste, społeczne i w zakresie umiejętności uczenia się

Cele operacyjne:

Uczeń:

- wyznacza rozstęp, odchylenie od średniej, odchylenie przeciętne i wariancję zestawu danych
- interpretuje i operuje informacjami przedstawionymi za pomocą tekstu lub tabeli
- dobiera odpowiedni model matematyczny do sytuacji z kontekstem realistycznym

Strategie nauczania:

- konstruktywizm

Metody i techniki nauczania:

- puzzle eksperckie
- kot i mysz

Formy pracy:

- praca indywidualna
- praca w grupach
- praca całego zespołu klasowego

Środki dydaktyczne:

- komputery z dostępem do Internetu w takiej liczbie, żeby każdy uczeń miał do dyspozycji komputer

Przebieg lekcji

Faza wstępna:

1. Uczniowie przypominają poznane miary tendencji centralnej.
2. Nauczyciel przedstawia dwa zestawy danych (np. 10, 10, 10 i 1, 10, 19), których średnia arytmetyczna i mediana są takie same.
Dyskusja – czy miary tendencji centralnej dobrze opiszą rozrzut danych.
3. Nauczyciel podaje temat i cele zajęć, proponuje kryteria sukcesu.

Faza realizacyjna:

1. Praca w grupach metodą puzzli eksperckich. Poproszeni o to kilka dni wcześniej uczniowie, przygotowali materiały dotyczące istoty i sposobu obliczania rozstępu, odchylenia od średniej, odchylenia przeciętego i wariancji. Ich zadaniem jest przekazanie zdobytych informacji grupom, tak aby każdy uczeń potrafił obliczyć dane wielkości.
2. Uczniowie w parach zapoznają się z galerią zdjęć interaktywnych, a następnie metodą kot i mysz rozwiązują proponowane w materiale ćwiczenia interaktywne. Przy czym mysz stara się jak najlepiej rozwiązać zadania, a kot, sprawdza ich poprawność – po 2 nieudanych próbach – kot „łapie mysz”, która wypada z gry. Aby gra toczyła się dalej – teraz mysz staje się kotem i procedura się powtarza.

Faza podsumowująca:

1. Podsumowaniem zajęć jest dyskusja – czy poznane miary rozproszenia dobrze charakteryzują zebrane dane – co pomaga, a czego brakuje do dobrej interpretacji danych.
2. Uczniowie – eksperci opowiadają o pracy grup, dzielą się swoimi spostrzeżeniami, zwracają uwagę na mocne i słabe strony pracy.
3. Uczniowie dokonują oceny koleżeńskiej partnerów z zabawy – kot i mysz.

Praca domowa:

Nauczyciel poleca, aby uczniowie poszukali w dostępnych źródłach informacji innego, niż poznany na zajęciach, wzoru na wariancję i udowodnili go.

Materiały pomocnicze:

- [Miary rozproszenia](#)
- [Pojęcie prawdopodobieństwa](#)

Wskazówki metodyczne:

Z galerią zdjęć interaktywnych uczniów mogą zapoznać eksperci, omawiając odpowiednie pojęcia.