

WIELKA POWTÓRKA MATURALNA

Witaj, nazywam się **Julia Truss** jestem **businesswoman**, a co dla Ciebie najważniejsze **biologiem - praca w EDU TRUST to moja pasja** od ponad 9 lat. W tym czasie zarządzałam ponad 32 osobowym zespołem nauczycieli w swojej firmie. Tworzę profesjonalne produkty, które wprowadzam do szkół. **Swoją pierwszą firmę założyłam mając 18 lat. Ponad 3000 tysięcy osób korzysta z moich flipbooków**, które na rynek edukacji w Polsce dopiero wprowadziłam 13 miesięcy temu. Kocham to. Postaram się dać Ci to narzędzie w postaci mojego kursu abyś też kochał/a swoją przyszłą pracę. Proszę Cię wyznaczyć sobie konkretny cel i dążyć do tego.



Mój cel to zmiana edukacji biologii w Polsce.

Opracowuje, analizuje i interpretuje wyniki badań w oparciu o proste analizy statystyczne

- definiuje pojęcia: *minimum, maksimum, średnia arytmetyczna*
- oblicza minimum, maksimum, średnią arytmetyczną na podstawie danych uzyskanych w doświadczeniu poprawnie sporządza wykres liniowy i słupkowy
- definiuje pojęcia: *zakres wartości, średnia arytmetyczna, mediana, dominanta, odchylenie standardowe*
- oblicza dominantę, medianę, odchylenie standardowe
- wykorzystuje analizę statystyczną do opisu i interpretacji wyników badań
- udowadnia lub odrzuca na podstawie wykonanych obliczeń z użyciem mediany i odchylenia statystycznego hipotezę do przedstawionego doświadczenia lub obserwacji

Miary tendencji centralnej – pojedyncze wartości stanowiące centrum zbioru wyników

- średnia arytmetyczna
- mediana
- dominanta

Miary zmienności – zwane też miarami rozproszenia, określają, jak bardzo zróżnicowane są wyniki

- rozstęp
- odchylenie standardowe

1) Średnia arytmetyczna- suma wszystkich pomiarów podzielona przez liczbę tych pomiarów: Pomiary długości działek zewnętrznego okółka okwiatu kosańca szczecinkowego (*Iris setosa*) dały następujące wyniki w milimetrach: 51, 49, 47, 46, 50, 54, 46, 50. Średnia arytmetyczna pomiarów wynosi:

$$\bar{x} = \frac{51 + 49 + 47 + 46 + 50 + 54 + 46 + 50}{8} \text{ mm} = 49,125 \text{ mm}$$

Średnią arytmetyczną obliczysz ze wzoru:

$$\text{Średnia arytmetyczna} = \frac{\text{suma wyników}}{\text{liczba wyników}}$$

Udzielając odpowiedzi, możesz podać wynik przybliżony. Jeśli polecenie nie precyzuje, z jaką dokładnością masz podać wynik, możesz np. podać wartość z dokładnością do drugiego miejsca po przecinku. Pamiętaj też o podaniu właściwej jednostki.

2) Dominanta (moda, modalna) – najczęstsza wartość pomiaru w danym zbiorze danych.

Przykład 1.

Wielkości lęgu u gęsi: 8, 6, 5, 7, 5, 9, 3, 5.

Sprawdzamy, która liczba występuje najczęściej: 8, 6, 5, 7, 5, 9, 3, 5.

Dominantą jest tu liczba 5 (ona jako jedyna powtórzyła się trzykrotnie, pozostałe liczby wystąpiły tylko raz).

Na tej podstawie możemy stwierdzić, że podczas prowadzenia dalszych badań najczęściej będziemy obserwować w gniazdach 5 lub zbliżoną liczbę jaj.

Przykład 2

Dla małych zbiorów danych może się okazać, że więcej niż jedna wartość pomiaru występuje z największą częstością – taki zbiór danych ma więcej niż jedną dominantę.

W przypadku pomiarów długości działek okwiatu u kosańca dominantami są 46 mm i 50 mm – obydwa wyniki pomiarów występują dwukrotnie w tym zbiorze danych: 51, 49, 47, **46, 50**, 54, **46, 50**

- 3) Mediana (wartość środkowa)** – taka wartość badanej cechy, która dzieli zbiór danych na dwie równoliczne części: połowa obserwacji ma wartości mniejsze lub równe medianie, a druga połowa – równe lub większe od mediany.

Przykład 1: parzysta liczba wyników

Aby wyznaczyć medianę, najwygodniej zacząć od uporządkowania zbioru danych od wartości najmniejszych do największych. To krok pierwszy. Przywołajmy nasz prosty przykład wielkości lęgu u gęsi: 8, 6, 5, 7, 5, 9, 3, 5. Po uporządkowaniu otrzymujemy taki zestaw danych: 3, 5, 5, 5, 6, 7, 8, 9. W następnym kroku liczymy, ile mamy danych – jest 8 liczb. Dla zbiorów danych o parzystej liczbie wyników medianę oblicza się jako średnią z dwóch środkowych wartości. 3, 5, 5, **5, 6**, 7, 8, 9 W ostatnim kroku musimy zatem dodać do siebie dwie sąsiadujące liczby – czwartą i piątą z kolei 5 – i podzielić na dwa.

$$\frac{5+6}{2} = 5,5.$$

Średnia arytmetyczna z dwóch środkowych wartości stanowi jednocześnie środek przedziału wyznaczonego przez te dwie wartości. Dzięki obliczeniu mediany możemy powiedzieć, że połowa gęsi złożyła co najwyżej 5 jaj, a druga połowa – 6 lub więcej jaj. Możemy też stwierdzić, że gęś, która złożyła 8 jaj, znajduje się w tej połowie próby, która miała wyższe wyniki lęgu. Badana próba jest stosunkowo mała, ale w dużym przybliżeniu możemy stwierdzić, że przy prowadzeniu dalszych badań z takim samym prawdopodobieństwem będziemy odnajdywać lęgi małe – liczące do 5 jaj, oraz lęgi duże – co najmniej 6 jaj.

Przykład 2: nieparzysta liczba wyników

W przypadku zbiorów danych o nieparzystej liczbie wyników medianę stanowi pojedyncza wartość środkowa uporządkowanego zbioru, a więc jest równa wartości jednego z uzyskanych wyników.

Poszukajmy mediany w następującym zestawie 11 danych z obserwacji lęgowych wykonanych w innej lokalizacji: 2, 3, 4, 5, 5, **8**, 9, 9, 11, 11, 12 Środkowa – szósta z kolei 6 – liczba to 8. Mediana wynosi zatem **8**.

4) Rozstęp

różnica między maksymalnym (największym) i minimalnym (najmniejszym) wynikiem pomiaru.

Znalezienie maksimum i minimum jest łatwiejsze po uporządkowaniu zbioru danych, ale nie okazuje się to konieczne. Przykład 1. W naszym prostym przykładzie wielkości lęgu u gęsi: 8, 6, 5, 7, 5, **9, 3**, 5, rozstęp wynosi $9 - 3 = 6$

Zakres zmienności to zakres liczb pomiędzy wartością najmniejszą a wartością największą w próbie. Na przykład dla zbioru liczb: 1,2; 1,3; 0,2; 3,7; 0,8; 6,1 zakres zmienności mieści się w przedziale od 0,2 do 6,1.

5) Odchylenie standardowe

miara skupienia/rozrzutu wyników wokół wartości średniej

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_i - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

s – odchylenie standardowe

$x_1, x_2, \dots, x_i, \dots, x_n$ – kolejne wyniki pomiarów

\bar{x} – średnia arytmetyczna

n – liczba pomiarów

Wartość odchylenia standardowego wyznacza typowe wartości badanej cechy zawierające się w przedziale $\bar{x} \pm s$. Jednocześnie przyjmuje się, że wystąpienie wartości cechy spoza przedziału $\bar{x} \pm 3s$ jest mało prawdopodobne. Jeżeli zmienność badanej cechy jest dobrze przybliżona rozkładem normalnym⁷ i pobrano próbę liczącą co najmniej kilkanaście obserwacji, to można przewidzieć, że około 2/3 przyszłych wartości pomiarów będzie mieściło się w przedziale $\bar{x} \pm s$, a około 99% – w przedziale $\bar{x} \pm 3s$.

Na egzaminie maturalnym nie musisz obliczać odchylenia standardowego, tylko wiedzieć, jak interpretować wartości odchylenia w zadaniach. Odchylenie standardowe informuje o tym, jak bardzo wartości w próbie są rozrzucone wokół wartości średniej. Niska wartość odchylenia standardowego świadczy o tym, że wartości pomiarów w próbie są zbliżone do wartości średniej, a wysoka wartość odchylenia standardowego świadczy o dużej zmienności w wynikach pomiarów odbiegających (znacznie) od wartości średniej. Na podstawie odchylenia standardowego możesz oszacować, czy wartości uzyskane w różnych próbach różnią się od siebie na tyle, aby potwierdzić hipotezę badawczą, czy raczej, aby ją odrzucić.

Na przykład w ciągu wartości:

0, 0, 1, 1, 2, 2, 3, 3, 4, 4, 4, 4, 4, 5, 5, 6, 6, 7, 7, 8, 8

- średnia arytmetyczna wynosi: **4**

- odchylenie standardowe wynosi: ok. 2,45*

A dla ciągu wartości:

2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6

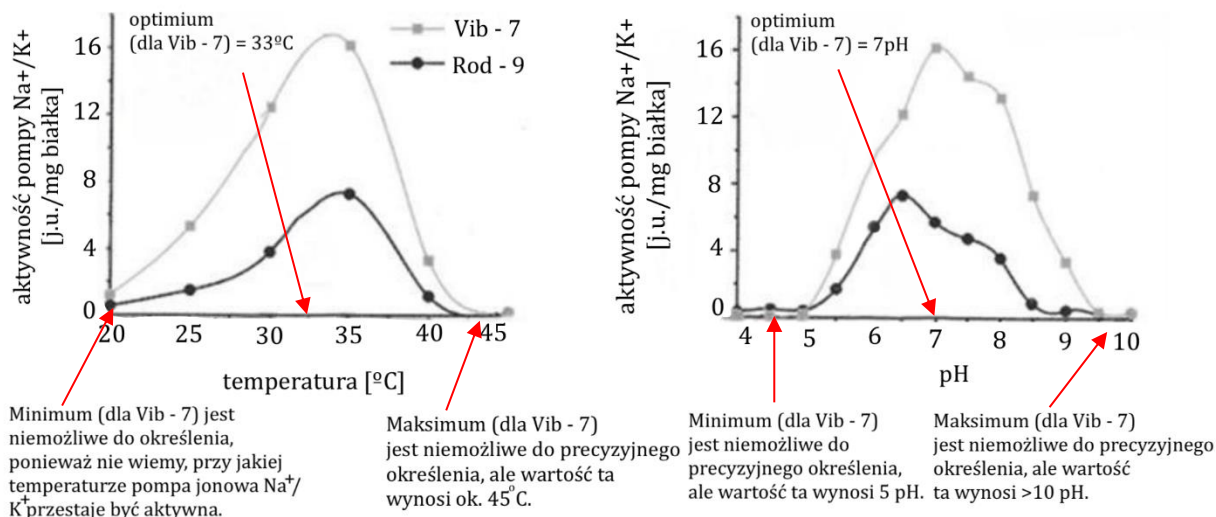
- średnia arytmetyczna również wynosi: **4**

- natomiast odchylenie standardowe wynosi ok. 1,26*

Oba ciągi wartości mają takie same średnie, jednak różnią się wartościami odchylenia standardowego. Wartość odchylenia standardowego w drugim ciągu jest niższa niż wartość odchylenia standardowego w pierwszym ciągu, dlatego można stwierdzić, że w ciągu drugim wartości są bardziej skupione wokół średniej niż w ciągu pierwszym.

Wnioskowanie polega na udzieleniu odpowiedzi na postawione pytanie badawcze lub zweryfikowaniu sformułowanych hipotez na podstawie analizy uzyskanych wyników doświadczenia lub obserwacji.

1. Wyznaczanie minimum i maksimum



Minimum to najniższa wartość uzyskana w badanej próbie, a maksimum to najwyższa wartość. Zakres wartości między minimum a maksimum określa się mianem **zakresu zmienności**. Z kolei dominanta to wartość, która jest najczęstsza w badanej próbie.

Wyznaczenie minimum, maksimum, zakresu zmienności i dominanty w tabeli

Długość orzeszków [mm]	Liczba orzeszków
1,4	15
1,5	73
1,6	379
1,7	637
1,8	612
1,9	355
2	111
2,1	17
2,2	1

Minimum – najniższy wynik



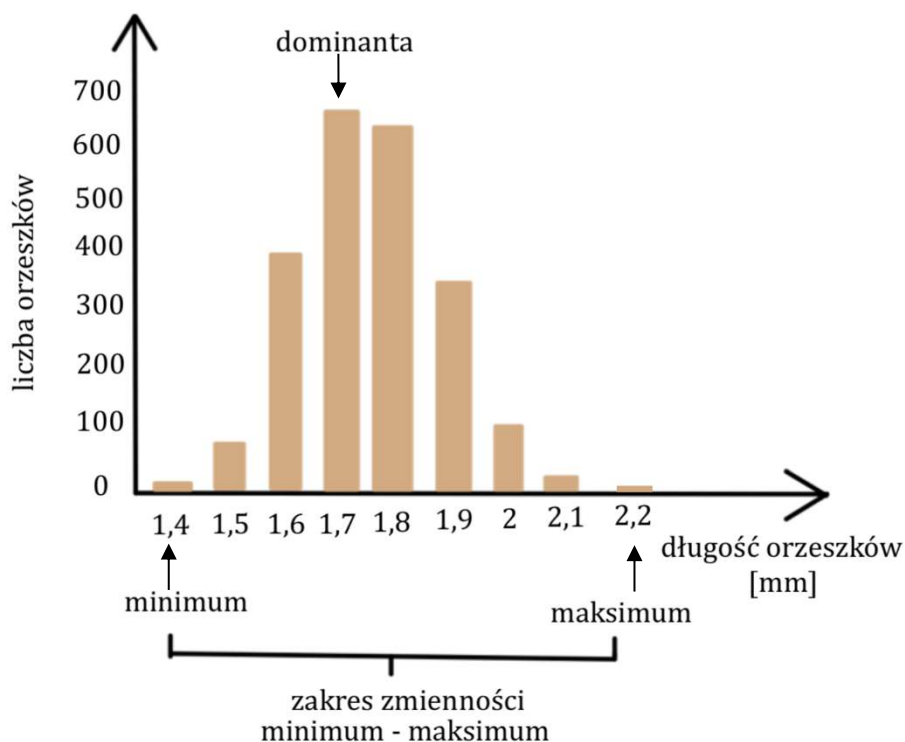
Dominanta – najczęstszy wynik



Maksimum – najwyższy wynik



Wyznaczenie minimum, maksimum, zakresu zmienności i dominanty na wykresie



2. Obliczanie średniej arytmetycznej

Aby obliczyć średnią arytmetyczną, zsumuj wartości wszystkich pomiarów, a następnie podziel otrzymaną sumę przez liczbę pomiarów.

$$\bar{X} = \frac{\text{suma wartości wszystkich pomiarów}}{\text{liczba pomiarów}}$$

$$\bar{X} = \frac{160+162+202+163+167+160+161}{7} = 167,86 \text{ cm}$$

3. Obliczanie mediany

160cm, 160cm, 161cm, 162cm, 163cm, 167cm, 202cm.

Liczba pomiarów jest nieparzysta, czyli medianę stanowi wartość środkowa: **162cm**

Jeśli liczba pomiarów jest parzysta, to medianę oblicza się przez wyznaczenie średniej arytmetycznej z dwóch środkowych pomiarów: **161 cm oraz 162 cm**.

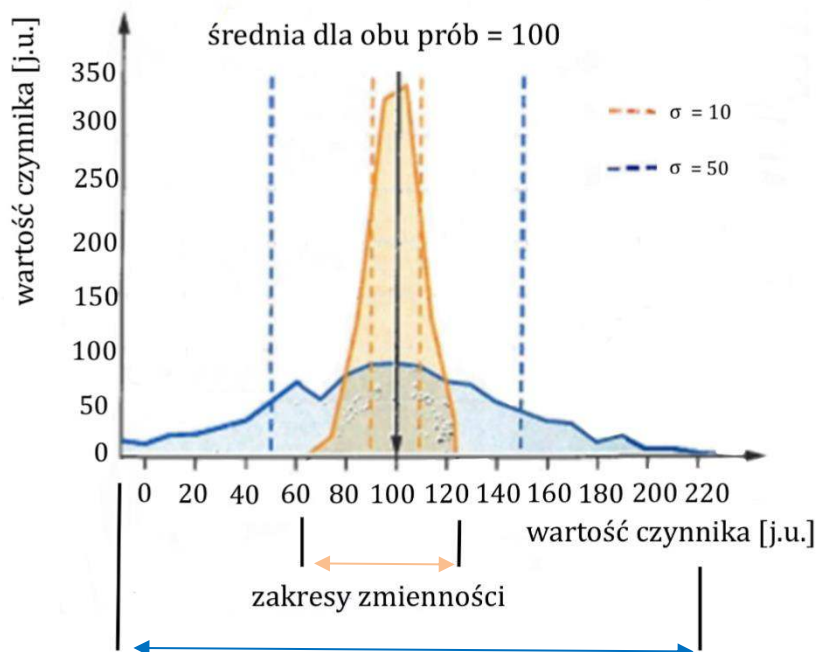
Średnia arytmetyczna a mediana

W powyższym przykładzie wartość średniej arytmetycznej i wartość mediany różnią się od siebie. Wartość średniej arytmetycznej zależy od wartości każdego pomiaru w próbie, a wyniki skrajnie odbiegające od średniej (np. 202cm) znacząco wpływają na wartość średniej (zwłaszcza gdy liczba pomiarów jest niewielka). Wartości skrajne nie mają natomiast wpływu na wartość mediany.

4. Odchylenie standardowe

Wartość odchylenia standardowego (sigma) przedstawia, w jakim stopniu wyniki przeprowadzonych pomiarów odbiegają od wartości średniej arytmetycznej obliczonej na ich podstawie. Na przykład jeśli wartość odchylenia standardowego wynosi 10, to przyjmuje się, że ok. 2/3 wyników powinno znajdować się w przedziale między wynikami większymi o 10 od średniej arytmetycznej a wynikami mniejszymi o 10 od średniej arytmetycznej.

Jeśli próba liczy wiele pomiarów, to na podstawie tabeli trudno ocenić, czy wartości poszczególnych wyników są zbliżone do wartości średniej arytmetycznej. W takiej sytuacji wyniki badań warto przedstawić za pomocą wykresu. Poniższy wykres przedstawia dane z dwóch badań.



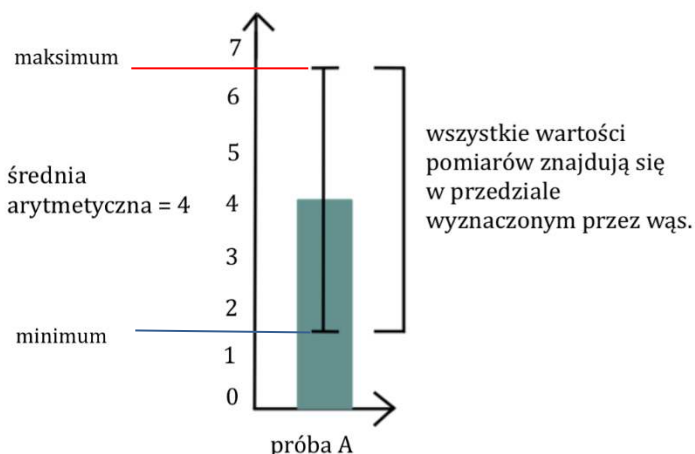
Wyniki obu badań (oznaczone na wykresie kolorami pomarańczowym i niebieskim) mają taką samą średnią arytmetyczną. W próbie pomarańczowej (o odchyleniu standardowym sigma= 10) wartości pomiarów są zbliżone do wartości średniej – zakres zmienności (minimum - maksimum) jest wąski. W próbie niebieskiej (o odchyleniu standardowym sigma = 50) wyniki pomiarów znacznie odbiegają od wartości średniej – zakres zmienności (minimum - maksimum) jest szerszy niż w próbie pomarańczowej.

5. słupki błędu (tzw. Wąsy)

Na wykresach często zaznacza się słupki błędu (tzw. wąsy). Są to pionowe linie zakończone poprzecznymi kreskami, odzwierciedlające różnicowanie wartości pomiarów w próbie. Wąsy mogą przedstawiać m.in. zakres zmienności lub odchylenie standardowe.

Wąsy przedstawiające zakres zmienności

Dla analizowanej próby średnia arytmetyczna wynosi 4, minimum -1,5, a maksimum - 6,5. Wąs obejmuje zakres zmienności, dlatego jest on zaznaczony między minimum a maksimum.



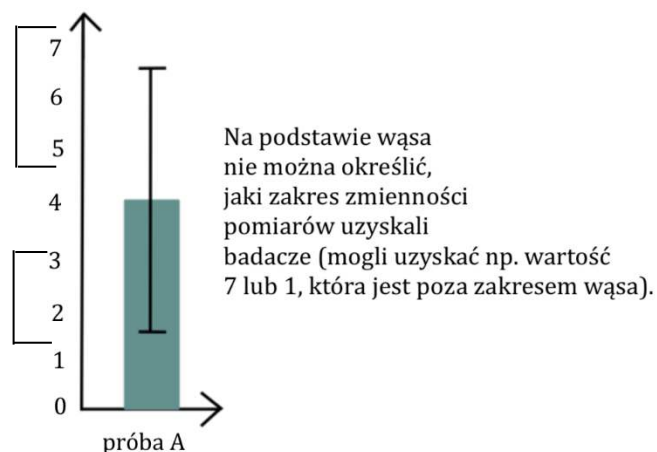
Wąsy przedstawiające odchylenie standardowe

Dla analizowanej próby średnia arytmetyczna wynosi 4, a odchylenie standardowe - 2,5. Wąs obejmuje przedział powyżej wartości średniej (+ 2,5) i poniżej wartości średniej (- 2,5).

Wąs odchylenia standardowego zaznacza się nad wartością średnią...

średnia arytmetyczna = 4

...i pod wartością średnią.



$$4 + 2,5 = 6,5$$

2/3 wyników pomiędzy

$$4 - 2,5 = 1,5$$

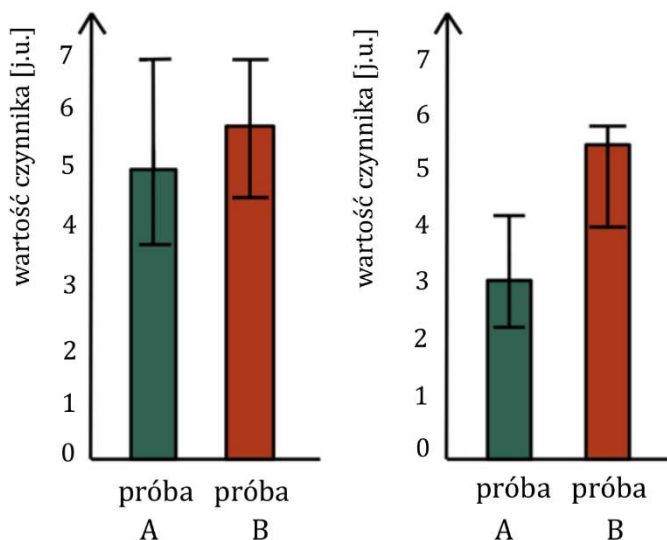
6. Interpretacja odchylenia standardowego

Analiza wartości liczbowej odchylenia standardowego

- wysoka wartość odchylenia standardowego oznacza, że wartości pomiarów są oddalone od wartości średniej.
- niska wartość odchylenia standardowego świadczy o tym, że wartości pomiarów w próbie są zbliżone do wartości średniej.

Analiza wykresu z wąsami wyznaczającymi odchylenie standardowe

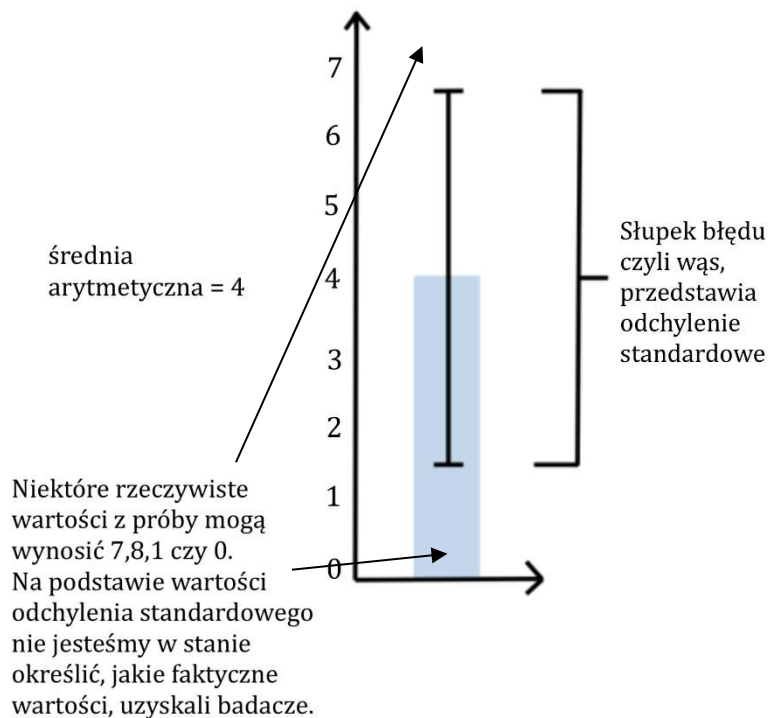
Jeśli przedział wartości wskazany przez wąs w jednej próbie zazębia się (lub w pełni pokrywa) z przedziałem wskazanym przez wąs w drugiej próbie, to można stwierdzić, że nie występują znaczące różnice między wartościami uzyskanymi w obu próbach.



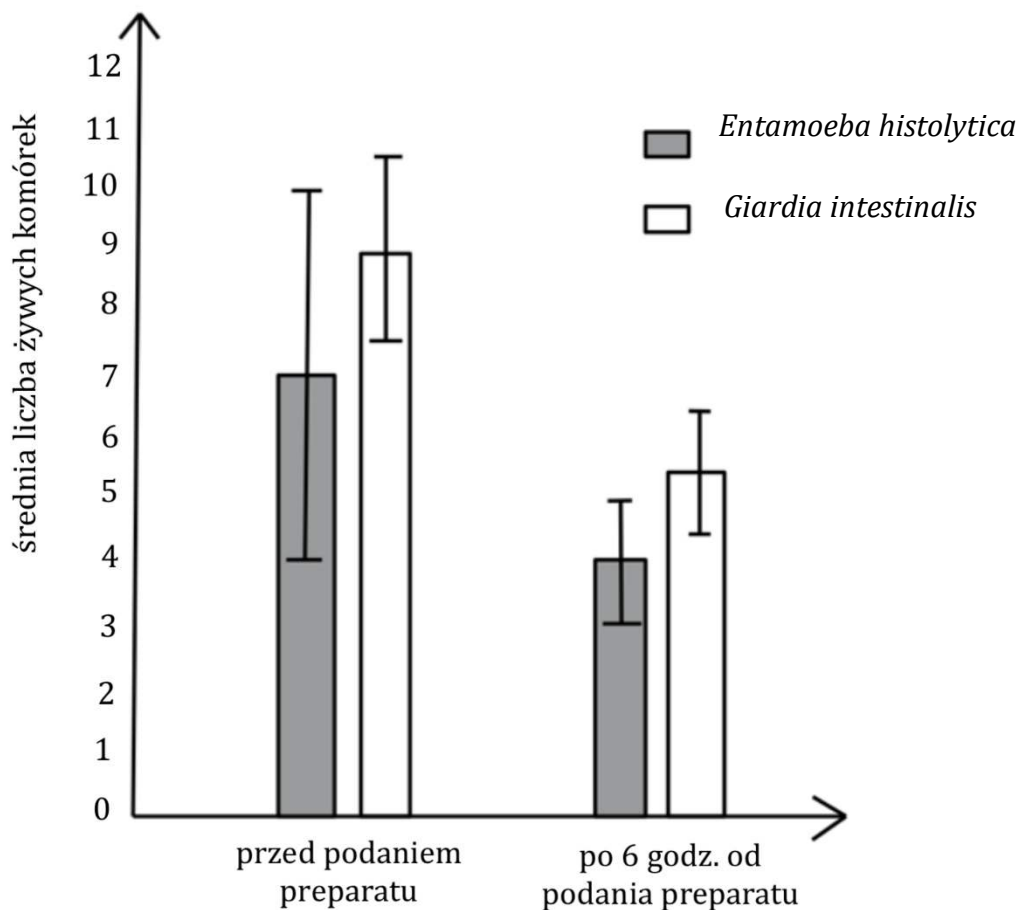
Jeśli wąsy w obu próbach nie pokrywają się, to można przypuszczać, że występują znaczące różnice między wartościami uzyskanymi w obu próbach.

Zwykle w zadaniach odchylenie standardowe przedstawia się w postaci słupków błędów, czyli wąsów na wykresie. Porównanie wartości odchylenia standardowego dwóch prób umożliwia oszacowanie, czy wyniki uzyskane w tych próbach znacząco się różnią.

W przypadku odchylenia standardowego słupki błędu, czyli wąsy, określają przedział, w którym znajdują się typowe wyniki pomiarów. Część wyników zawsze leży dalej od średniej niż jedno odchylenie standardowe. Na przykład jeśli średnia arytmetyczna dla określonej próby wynosi 4, a odchylenie standardowe wynosi 2,5 to na wykresie słupek będzie miał wysokość 4, a wąsy odchylenia standardowego będą znaczyć obszar pomiędzy wartością 1,5 a 6,5. Należy jednak pamiętać, że wśród rzeczywistych wyników uzyskanych przez badaczy mogą znaleźć się także wartości leżące poza przedziałem jednego odchylenia standardowego, takie jak 0 czy 8.



Przeprowadzono doświadczenie, którego celem było sprawdzenie skuteczności nowego leku przeciwpasożytniczego mogącego wykazywać aktywność przeciwko pełzakowi czerwonki (*Entamoeba histolytica*) oraz ogoniastkowi jelitowemu (*Giardia intestinalis*), popularnie zwanemu lamblia. W tym celu przygotowano hodowle laboratoryjne obu gatunków pasożytów, do których dodano taką samą ilość testowanego preparatu przeciwpasożytniczego. Skuteczność leku oceniono na podstawie żywotności badanych protistów. Obliczono średnie arytmetyczne oraz odchylenie standardowe dla pomiarów przeprowadzonych przed zastosowaniem preparaty oraz po 6 godz. od jego podania. Wyniki eksperymentu przedstawiono na wykresie. Słupki błędów wskazują odchylenie standardowe.



Zadanie

Określ, w stosunku do którego pasożyta – pełzaka czerwonki czy ogoniastka jelitowego – badany lek przeciwpasożytniczy wykazuje wyższą skuteczność. Odpowiedź uzasadnij, odnosząc się do przedstawionych na wykresie wartości odchylenia standardowego.

Lek przeciwpasożytniczy wykazuje wyższą skuteczności w przypadku pełzaka czerwonki, ponieważ po jego zastosowaniu zakres wartości odchylenia standardowego zmniejszył się bardziej niż w przypadku ogoniastka jelitowego.

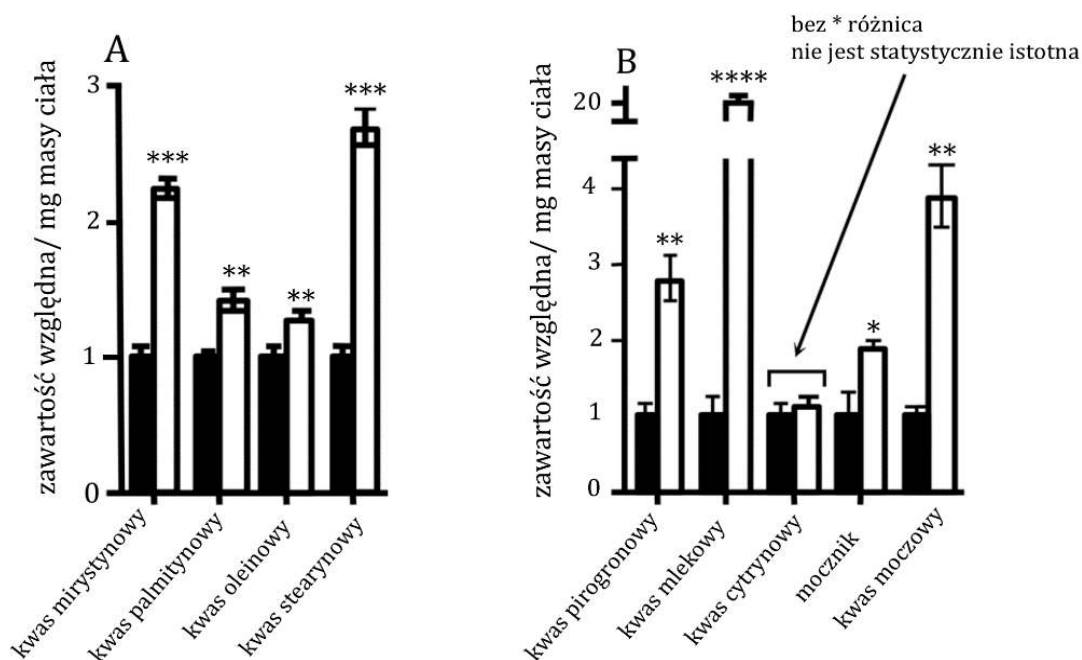
Istotność statystyczna

Co to jest istotność statystyczna?

Istotność statystyczna to **ocena, czy wynik występuje przypadkiem**. Gdy wynik jest statystycznie istotny, oznacza to, że jest mało prawdopodobne, aby wystąpił przypadkiem, z powodu losowych fluktuacji.

Co oznacza poziom istotności 0,05?

Na przykład poziom istotności 0,05 **wskazuje na 5% ryzyko stwierdzenia, że istnieje różnica między wynikami badań a hipotezą zerową, gdy w rzeczywistości takiej różnicy nie ma**. Poziom istotności musi zostać określony w protokole badania w części dotyczącej statystyki.



Informacja do wykresu

■ żywka standardowa □ żywka o wysokiej zawartości tłuszczów nasyconych

- słupki błędu oznaczają błąd standardowy

- symbolem * określono wyniki testów istotności statystycznej dla porównań par średnich:

* $p < 0,05$

** $p < 0,01$

*** $p < 0,001$

**** $p < 0,0001$

różnica statystycznie największa

Na wykresie zaznaczono za pomocą gwiazdek także kody istotności statystycznej.

Im niższa p-wartość, tym większe przekonanie badacza równe $1-p$ (nigdy nie można być pewnym na 100%), że znak różnicy między średnimi z prób jest taki sam jak znak różnicy „prawdziwych” średnich. Zwykle przyjmuje się, że p-wartość poniżej 0,05 oznacza wynik istotny statystycznie, związany z co najwyżej 5% ryzykiem błędu.

Niska p-wartość jest argumentem za tym, że jeżeli w próbie badawczej średnia z próby jest większa niż w próbie kontrolnej, to także „prawdziwa” średnia w próbie badawczej jest większa niż w próbie kontrolnej. Z kolei wysoka p-wartość wcale nie oznacza, że różnic między „prawdziwymi” średnimi nie ma (przy odpowiednio licznych próbach prawie zawsze stają się wyraźne choćby bardzo małe różnice).

Jest to tylko informacja, że określenie znaku różnicy „prawdziwych” średnich na podstawie małej próby jest niepewne – badania nie prowadzą do żadnych konkretnych wniosków pod tym względem