

Proszę do 22 maja rozwiązać zaległe zadania. Przypominam że w każdą środę o godzinie 12.00 będą zamieszczane zadania na sprawdzian (dla chętnych) na ocenę dobrą lub bardzo dobrą. Rozwiązania zadań ze sprawdzianu należy wysłać najpóźniej do godziny 13.30 na adres e-mail kupkaandrzej@radymno.edu.pl

Statystyka opisowa

Sposoby przedstawiania danych

Dane liczbowe to zestaw n liczb opisujących jakieś zjawisko.

Dane liczbowe mogą dotyczyć poszczególnych obiektów w jakiejś zbiorowości, wyników uzyskanych w różnych latach, wyników pomiarów dokonanych w różnym czasie albo różnymi metodami. Co istotne, zakładamy, że wszystkie dane (liczby) mają tę samą naturę, opisują to samo zjawisko, mają takie same jednostki. Nie zakładamy, że poszczególne wartości liczbowe są bardzo podobne do siebie. Może tak być, ale nie musi.

Zbiór danych liczbowych można zdefiniować bardziej formalnie i można podczas nauki statystyki używać bardzo sformalizowanego języka. Uważamy, że na Twoje potrzeby wystarczy podejście bardziej nieformalne. Ewentualnie zapamiętaj, że dane, które rozważamy w statystyce, charakteryzują jakąś **cechę statystyczną** (wyniki egzaminu, wzrost osób w klasie itp.).

Rozstęp danych to różnica między wartością największą i najmniejszą.

Wartość najmniejsza, największa oraz rozstęp to najprostsze, ale całkiem przydatne charakterystyki zbioru danych. Przydadzą Ci się do weryfikacji wyników obliczeń, o czym będzie dalej.

Załóżmy, że mamy następujące liczby (dla ułatwienia uporządkowaliśmy je od najmniejszej do największej):

-0,2 0,3 0,3 0,4 0,6 0,8 0,9 1,2 2,5

Wartość największa w tym przykładzie to 2,5.

Wartość najmniejsza to -0,2.

Rozstęp danych, czyli różnica między wartością największą i najmniejszą, w naszym wypadku wynosi 2,7.

Dane w zestawie mogą się powtarzać.

Załóżmy, że mamy następujące liczby:

-0,2 **0,3** **0,3** 0,4 0,6 0,8 0,9 1,2 2,5

Widzimy, że dwie liczby (równe 0,3) się powtarzają. Nie możemy jednak utożsamić ich ze sobą ani „zbędnych” powtórzeń skreślić. To są odrębne dane, a że są równe — to przypadek. **Liczba danych** w podanym przykładzie jest równa 9.

Teoretycznie mogłoby się zdarzyć, że wszystkie dane liczbowe są równe. Weźmy pod uwagę zestaw:

3,2 3,2 3,2 3,2 3,2 3,2 3,2 3,2

Liczba danych i tutaj jest równa 9. Do wzorów będziemy podstawiać $N = 9$.

Jeśli dużo danych jest takich samych, można w tabeli lub na diagramie podać, ile razy która się powtarza.

Jeśli danych jest niewiele, możemy po prostu wymienić je wszystkie:

-0,2 0,3 0,3 0,4 0,6 0,8 0,9 1,2 2,5

Jeśli jednak niektóre dane się wielokrotnie powtarzają, na przykład

0 1 1 1 1 2 3 3 3 3 3 4 4 4 5 7

to wygodniej jest je przedstawić zbiorczo. Można to uczynić na kilka sposobów.

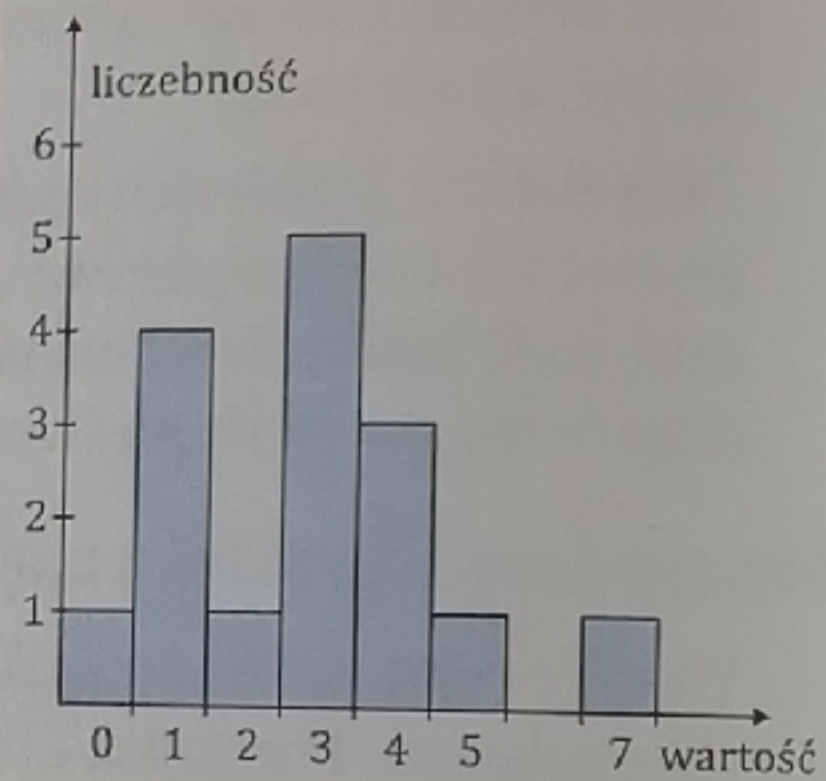
Tabela liczebności może być zredagowana w poziomie:

wartość	0	1	2	3	4	5	7
liczebność	1	4	1	5	3	1	1

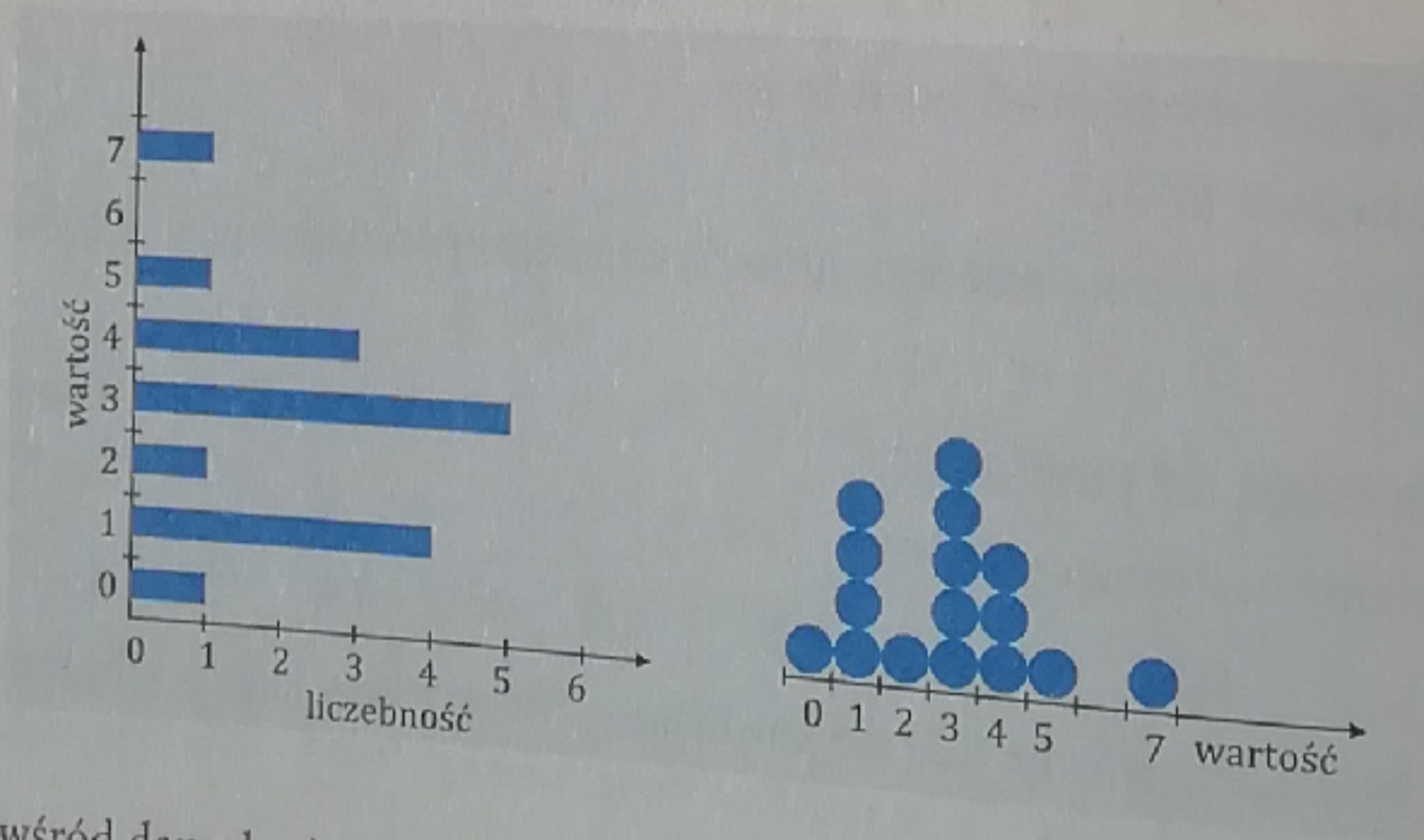
Może też być podana w pionie

wartość	liczebność
0	1
1	4
2	1
3	5
4	3
5	1
7	1

lub w postaci **diagramu**:



Słupki mogą być ustawione pionowo, poziomo, mogą mieć różny kształt, rozmaite ozdobniki — forma graficzna to kwestia drugorzędna:



Liczby, które wśród danych nie występują (np. liczba 6 w naszym przykładzie), można pomijać albo (jeśli tak chcemy) wymieniać jako mające liczebność zerową.

W tabeli lub na diagramie mogą być podane liczebności lub częstości.

Liczebność to po prostu ilość danych mających konkretną wartość liczbową. Liczebność jakiejś danej jest dowolną liczbą naturalną albo zerem.

Częstość to liczebność odniesiona do liczebności całego zbioru danych (czyli do N). Częstość wystąpienia jakiejś liczby wyraża się **ułamkiem** zwykłym, ułamkiem dziesiętnym albo **procentem**.

Tabele (albo diagramy) liczebności łatwo jest przekształcić na tabele (albo diagramy) częstości. Wystarczy podzielić liczbę wystąpień każdej kolejnej danej przez N , czyli przez sumaryczną liczbę danych (i ewentualnie pomnożyć przez 100%, jeśli chcemy mieć wyniki wyrażone w procentach).

Tabele częstości można zamienić na tabele liczebności tylko wtedy, gdy wiemy, jaka była ogólna liczba danych. Mnożąc częstości przez N uzyskujemy liczebności. Jeśli nie znamy N , taka zamiana nie jest możliwa.

Przykład 354. Tabelę liczebności zamień na tabelę częstości:

wartość	0	1	2	3	4	5	7
liczebność	1	4	1	5	3	1	1

Rozwiązanie. Podstawą jest obliczenie ogólnej liczby danych. Sumujemy poszczególne liczebności:
 $N = 1 + 4 + 1 + 5 + 3 + 1 + 1 = 16$.

Teraz wystarczy podzielić wszystkie liczebności przez N i wynik gotowy:

wartość	0	1	2	3	4	5	7
częstość	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Gdybyśmy chcieli, moglibyśmy wyrazić częstości w procentach:

wartość	0	1	2	3	4	5	7
częstość	6,25%	25%	6,25%	31,25%	18,75%	6,25%	6,25%

Mając gotowe dane, moglibyśmy też wykonać diagram częstości. ■

Suma częstości poszczególnych danych musi być równa jedności (albo 100%).

Ten fakt sprawdź zawsze, zanim przystąpisz do jakichkolwiek obliczeń. Jeśli jako sumę częstości otrzymasz 1 (albo 100%), upewnij się, że rozumiesz o co chodzi w zadaniu. Jeśliby się okazało, że nie wychodzi 1, dobrze się zastanów nad sensownością danych.

Ćwiczenie 535. Sporządzono tabelę częstości występowania usterek w samochodach opuszczających linię produkcyjną:

liczba usterek	0	1	2	3	4	5
częstość	86,3%	-----	3,1%	0,6%	0,1%	0

Niestety, liczba samochodów z jedną usterką została wpisana niewyraźnie. Jaka to była liczba?

Zamiast danych z tabeli liczebności można wstawiać dane z tabeli częstości.

Liczebność i częstość to nie to samo. Jednak w większości wypadków, takich jak obliczanie mediany czy średniej arytmetycznej, dane dotyczące liczebności możesz zastąpić danymi dotyczącymi częstości, a wynik wyjdzie taki sam. Pamiętaj tylko, aby w jednym wzorze nie występowały liczebności obok częstości. Albo stosujesz jedne, albo drugie!