

# STATYSTYKA MATEMATYCZNA

## Pojęcia wstępne

### Słownik pojęć

*Statystyka matematyczna* zajmuje się opisywaniem i analizą zjawisk masowych przy użyciu metod rachunku prawdopodobieństwa.

W statystyce matematycznej rozważa się pewien zbiór elementów podlegających badaniu statystycznemu zwany *populacją generalną* oraz pewną *cechą* (lub kilka cech) tej populacji.

Cecha populacji może być *ilościowa*, inaczej mierzalna (np.: czas świecenia żarówki, wzrost człowieka) lub *jakościowa*, inaczej niemierzalna (np.: płeć, kolor oczu).

Gdy obserwowana cecha mierzalna może przyjmować każdą liczbę z pewnego przedziału, to mamy do czynienia z *cechą ciągłą*; jeśli zaś tylko niektóre wartości, to mamy do czynienia z *cechą skokową*.

**Próbka** – skończony podzbiór populacji generalnej podlegający bezpośredniemu badaniu ze względu na interesującą nas cechę.

Jednym z ważniejszych zadań statystyki jest wnioskowanie o populacji na podstawie próbki (**wnioskowanie statystyczne**).

Aby przy wnioskowaniu statystycznym nie popełnić zbyt dużych błędów, próbka powinna być tak dobrana, aby dobrze reprezentowała populację, tzn., aby rozkład cech w próbce różnił się możliwie mało od rozkładu tych cech w populacji.

## Statystyczny rozkład z próbki

Z populacji generalnej wybieramy  $n$ -elementową próbkę. Badamy mierzalną cechę  $X$  i otrzymujemy  $n$  wyników:

$$x_1, x_2, \dots, x_n.$$

Załóżmy, że pewne wyniki powtarzają się, tzn.:

$$\begin{array}{ll} x_1 - n_1 \text{ razy,} & n_i - \text{częstości występowania} \\ x_2 - n_2 \text{ razy,} & \text{wartości cechy w próbce} \\ \vdots & \\ x_k - n_k \text{ razy.} & \sum_i n_i = n \end{array}$$

Oznaczmy

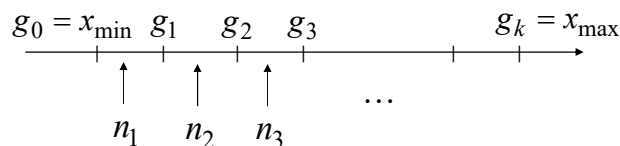
$$w_i = \frac{n_i}{n} \quad - \text{częstości względne} \quad \sum_i w_i = \sum_i \frac{n_i}{n} = \frac{1}{n} \sum_i n_i = \frac{1}{n} \cdot n = 1.$$

Wyniki badań statystycznych prezentujemy w postaci *szeregów statystycznych*. Szereg statystyczny jest to zbiór wyników obserwacji jednostek według pewnej cechy. Szereg statystyczny jest na ogół uporządkowany według rosnących lub malejących wartości.

Przy tworzeniu szeregów statystycznych (w celu ułatwienia analizy) wartości próbki często grupuje się w klasach. W tym celu przedział otrzymanych wartości

$$\langle x_{\min}, x_{\max} \rangle$$

dzielimy na podprzedziały klasowe  $(g_{i-1}, g_i)$ ,  $i = 1, 2, \dots, k$  (najczęściej jednakowej długości) zwane krótko **klasami** i dla każdej klasy znajdujemy odpowiednie częstości.



W ten sposób otrzymujemy tzw. *szereg rozdzielczy (przedziałowy)*.

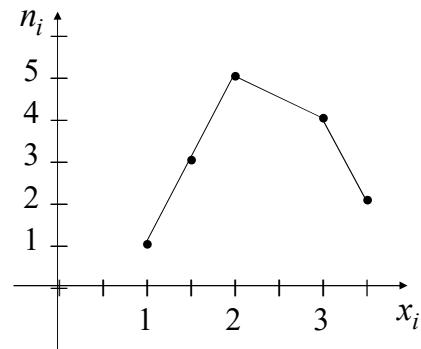
wiek pracownika $x_i$	liczba osób $n_i$	liczba dzieci $x_i$	liczba małżeństw $n_i$
20-30	7	0	6
30-40	12	1	18
40-50	16	2	34
50-60	6	3	14
60-70	2	4	9
Szereg rozdzielczy przedziałowy		Szereg rozdzielczy punktowy	

*Statystycznym rozkładem z próbki* (rozkładem empirycznym) nazywamy zestawienie wyników obserwacji cechy  $X$  i odpowiadających im częstości (lub częstości względnych)

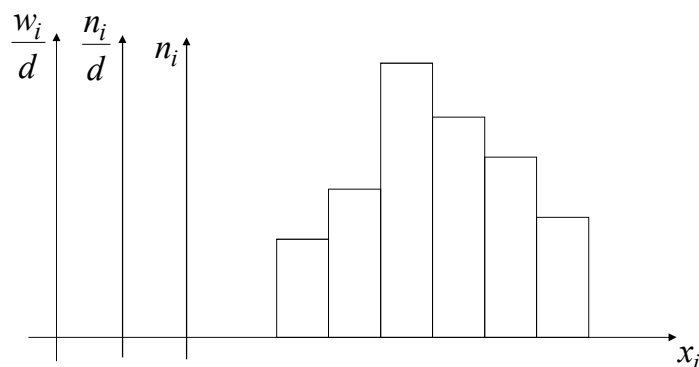
### Wykresy rozkładu statystycznego

*Wielobok częstości* – łamana, której odcinki łączą punkty  $(x_i, n_i)$  dla szeregu punktowego lub punkty  $(\hat{x}_i, n_i)$ , gdzie są  $\hat{x}_i$  środkami przedziałów w szeregu przedziałowym.

$x_i$	1	1,5	2	3	3,5
$n_i$	1	3	5	4	2



*Histogram częstości* (lub częstości względnych) – figura schodkowa składająca się z prostokątów, których podstawami są podprzedziały klasowe, a wysokości mogą być równe  $n_i$ , lub  $\frac{n_i}{d}$  ( $d$  – szerokość podprzedziałów), lub  $\frac{w_i}{d} = \frac{n_i}{nd}$ .



## Ważniejsze charakterystyki rozkładów statystycznych

Niech dana będzie próbka:  $x_1, x_2, \dots, x_n$ .

### 1. Średnia arytmetyczna

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Jeżeli w próbie wynik pomiaru  $x_i$  wystąpił  $n_i$  razy,  $i=1, \dots, k$ , gdzie  $\sum_{i=1}^k n_i = n$ , to średnią arytmetyczną można wyliczyć ze wzoru:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_i$$

Średnia ta bywa nazywana *średnią arytmetyczną ważoną*.

Dla danych pogrupowanych w szereg rozdzielczy przedziałowy o  $k$  przedziałach klasowych, średnią arytmetyczną wyznaczamy ze wzoru:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \dot{x}_i n_i$$

$\dot{x}_i$  – środek  $i$ -tego przedziału wyznaczamy ze wzoru:

$$\dot{x}_i = \frac{1}{2}(g_{i-1} + g_i)$$

$n_i$  – licznosc  $i$ -tej klasy, czyli liczba elementów próbki, które znalazły się w danym przedziale klasowym

## 2. Mediana empiryczna

**Medianą** (ozn.  $\tilde{x}_{me}$ ) empiryczną nazywamy liczbę, która rozdziela uporządkowaną niemalejąco próbkę na dwie równe grupy, tzn.

$$\tilde{x}_{me} = \begin{cases} x_{(n+1)/2}, & \text{gdy } n \text{ nieparzyste} \\ \frac{1}{2}(x_{(n/2)} + x_{(n/2)+1}), & \text{gdy } n \text{ parzyste} \end{cases}$$

### Przykłady

$$3 \quad 3,4 \quad 3,7 \quad 3,7 \quad 4,1 \quad \boxed{4,2} \quad 4,4 \quad 4,4 \quad 4,5 \quad 4,8 \quad 5,1$$

$$\tilde{x}_{me} = 4,2$$

$$3 \quad 3,4 \quad 3,7 \quad 3,7 \quad 4,1 \quad \boxed{4,2} \quad \boxed{4,4} \quad 4,4 \quad 4,5 \quad 4,8 \quad 5,1 \quad 5,5$$

$$\tilde{x}_{me} = \frac{4,2 + 4,4}{2} = 4,3$$

## 3. Wariancja empiryczna i odchylenie standardowe

Wariancja empiryczna stanowi miarę rozproszenia (dyspersji) aktualnie zaobserwowanej próbki względem średniej arytmetycznej.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Przy obliczaniu wariancji empirycznej można posłużyć się wzorem:

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

Jeżeli w próbkę wynik pomiaru  $x_i$  wystąpił  $n_i$  razy,  $i=1, \dots, k$ , to wariancję można wyliczyć ze wzoru:

$$s^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i \quad \text{lub} \quad s^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i - \bar{x}^2$$

Dla szeregu rozdzielczego przedziałowego o  $k$  klasach otrzymujemy:

$$s^2 = \frac{1}{n} \sum_{i=1}^k (\dot{x}_i - \bar{x})^2 n_i \quad \text{lub} \quad s^2 = \frac{1}{n} \sum_{i=1}^k \dot{x}_i^2 n_i - \bar{x}^2$$

Pierwiastek kwadratowy z wariancji nazywamy *odchyleniem standardowym* (empirycznym odchyleniem standardowym):

$$s = \sqrt{s^2}$$

#### 4. Współczynnik zmienności

$$V = \frac{s}{\bar{x}} \cdot 100\%$$

Współczynnik zmienności jest względną miarą rozproszenia i bywa użyteczny przy porównywaniu zmienności cech o tych samych lub różnych mianach (jednostkach). Duża wartość współczynnika zmienności świadczy o niejednorodności (dużej dyspersji) danej zbiorowości.

**Uwaga.** Podane charakterystyki rozkładów statystycznych są bardzo nielicznymi przykładami charakterystyk liczbowych stosowanych w statystyce do opisu próbki. Oprócz podanych można przykładowo wymienić: średnia harmoniczna, średnia geometryczna, moda, odchylenie przeciętne, współczynnik asymetrii, współczynnik skupienia.

**Przykład.**

Waga (w kg) 20 poborowych, którzy stanęli w pewnym dniu przed komisją poborową, była następująca:

76, 67, 80, 73, 69, 73, 64, 85, 77, 74, 76, 76, 58, 63, 74, 69, 81, 74, 71, 83.

**A.** Na podstawie tego zbioru danych oblicz: średnią arytmetyczną, medianę, odchylenie standardowe, współczynnik zmienności.

**B.** Utwórz szereg rozdzielczy z klasami prawostronnie domkniętymi o jednakowej rozpiętości równej 5 kg i dolnej granicy pierwszego przedziału klasowego równej 55 kg. Przedstaw ten rozkład graficznie za pomocą histogramu i wieloboku częstości. Dla otrzymanego szeregu rozdzielczego oblicz średnią arytmetyczną oraz odchylenie standardowe.

**A**

Dane po uporządkowaniu:

58, 63, 64, 67, 69, 69, 71, 73, 73, 74, 74, 74, 76, 76, 76, 77, 80, 81, 83, 85.

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = \frac{1}{20} \cdot 1463 = 73,15$$

$$\tilde{x}_{me} = \frac{74 + 74}{2} = 74$$

$$s^2 = \frac{1}{20} \sum_{i=1}^{20} x_i^2 - \bar{x}^2 = \frac{1}{20} \cdot 107899 - (73,15)^2 = 5394,95 - 5350,923 = 44,0275$$

$$s = \sqrt{s^2} = \sqrt{44,0275} \approx 6,6353$$

$$V = \frac{s}{\bar{x}} \cdot 100\% = \frac{6,635322}{73,15} \cdot 100\% \approx 9\%$$



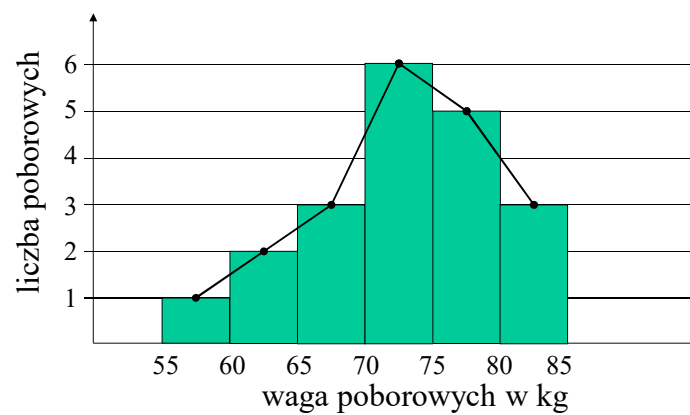
**B**

58, 63, 64, 67, 69, 69, 71, 73, 73, 74, 74, 74, 76, 76, 76, 77, 80, 81, 83, 85.

Szereg rozdzielczy:

Waga w kg	Liczba poborowych $n_i$	Częstość poborowych $w_i$
55–60	1	0,05
60–65	2	0,1
65–70	3	0,15
70–75	6	0,3
75–80	5	0,25
80–85	3	0,15
Razem	20	1

Histogram oraz wielobok częstości rozkładu wagi poborowych:



Waga w kg	Środki klas $\dot{x}_i$	Liczba poborowych $n_i$
55-60	57,5	1
60-65	62,5	2
65-70	67,5	3
70-75	72,5	6
75-80	77,5	5
80-85	82,5	3
	Razem	20

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \dot{x}_i n_i = \frac{1}{20} \sum_{i=1}^6 \dot{x}_i n_i = \frac{1}{20} \cdot 1455 = 72,75$$

$$s^2 = \frac{1}{n} \sum_{i=1}^k \dot{x}_i^2 n_i - \bar{x}^2 = \frac{1}{20} \sum_{i=1}^6 \dot{x}_i^2 n_i - \bar{x}^2 =$$

$$= \frac{1}{20} \cdot 106755 - 5292,5625 = 46,1875$$

$$s = \sqrt{s^2} = \sqrt{46,1875} \approx 6,7961$$