

Laboratorium 1 - wprowadzenie

Agnieszka Mensfelt

4 marca 2021

Statystyka

- ▶ Rachunek prawdopodobieństwa
- ▶ Statystyka
 - **opisowa**
 - wnioskowanie statystyczne

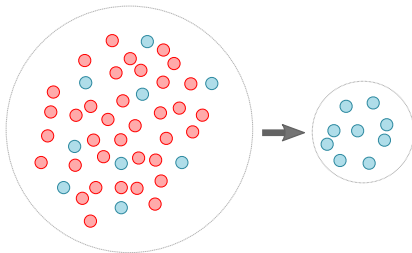
Populacja i próba

Populacja

Zbiór zawierający wszystkie elementy podlegające badaniu.

Próba

Podzbiór populacji wybrany w określony sposób.



Badanie statystyczne

- ▶ Etapy badania statystycznego
 1. Sformułowanie problemu
 2. Zbieranie danych statystycznych
 3. **Opracowanie** i analiza danych
 4. Sformułowanie wniosków
- ▶ Obserwacja i eksperyment
- ▶ Dobór próby

Skale pomiarowe

Przykładowe dane

ID	Kod województwa	Płeć	Wiek	Wzrost	Temperatura	Poziom glukozy [mg/dL]
P2	5	K	50	średni	37.7	90
P4	10	K	25	wysoki	36.6	101
P10	3	M	49	niski	35.7	72
P21	8	M	32	średni	38.1	120
...
P33	5	K	51	średni	36.7	110

Skale pomiarowe

- ▶ jakościowe
 - nominalne
 - porządkowe
- ▶ ilościowe
 - dyskretne
 - ciągłe



Szeregi rozdzielcze

Przykład

Numer rzutu	Liczba oczek
1	3
2	2
3	2
4	1
5	6
6	4
7	1
8	3
9	1
10	1
11	6
12	5
13	3
14	1
15	5
16	4
17	4
18	2
19	6
20	3
21	6
22	2
23	5
24	3
25	5
26	1
27	3
28	5
29	1
30	1



Szereg punktowy

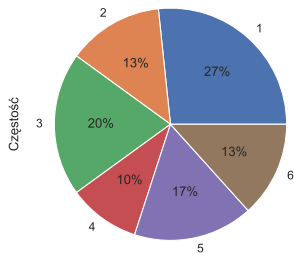
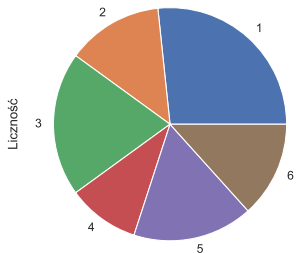
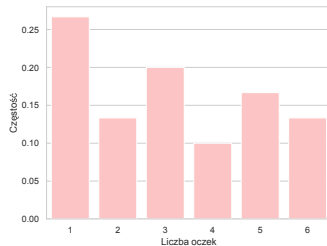
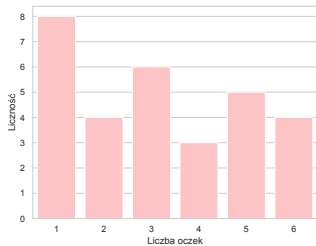
Wyniki rzutu kostką:

3 2 2 1 6 4 1 3 1 1 6 5 3 1 5 4 4 2 6 3 6 2 5 3 5 1 3 5 1 1

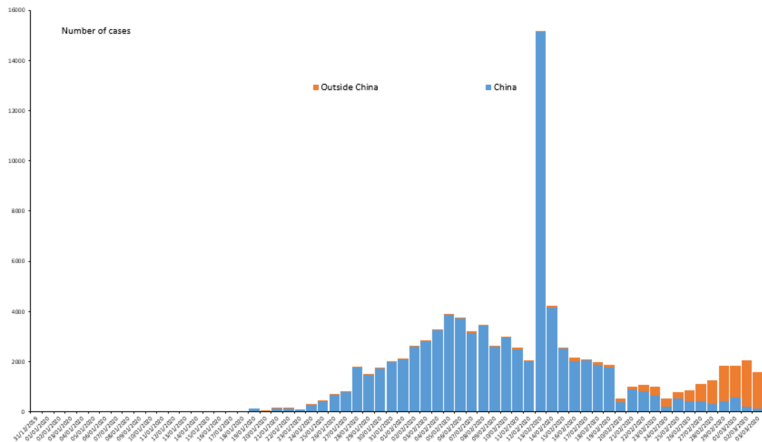
Rozkład wyników:

Wynik	1	2	3	4	5	6
Liczność	8	4	6	3	5	4
Częstość	0.27	0.13	0.20	0.10	0.17	0.13

Wykres słupkowy i kołowy



Wykres słupkowy skumulowany



Źródło: www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

Przykład 2

ID pacjenta	Wiek pacjenta
U1	32
U3	31
U10	57
U11	70
U12	72
U14	66
U15	34
U16	47
U17	42
U20	53
U21	52
U22	32
U23	50
U24	47
U25	31
U26	63
U27	55
U30	57
U31	46
U33	47
U34	39
U35	63
U36	70
U38	30
U40	30
U41	25
U50	58
U60	32
U61	48
U62	59

Szereg przedziałowy

Wiek pacjentów:

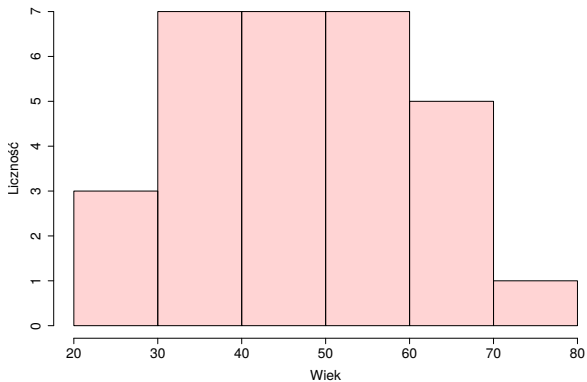
32 31 57 70 72 66 34 47 42 53 52 32 50 47 31 63 55 57 46 47 39
63 70 30 30 25 58 32 48 59

Rozkład wieku:

Przedział	Obserwacje	Liczność	Częstość
(20, 30]	25, 30, 30	3	1/10
(30, 40]	31, 31, 32, 32, 32, 34, 39	7	7/30
(40, 50]	42, 46, 47, 47, 47, 48, 50	7	7/30
(50, 60]	52, 53, 55, 57, 57, 58, 59	7	7/30
(60, 70]	63, 63, 66, 70, 70	5	1/6
(70, 80]	71	1	1/30

Histogram

Przedział	(20, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 70]	(70, 80]
Liczność	3	7	7	7	4	1



Wybór liczby przedziałów

k – liczba przedziałów (klas)

n – liczba obserwacji

- ▶ $k \approx \sqrt{n}$
- ▶ $k \approx 1 + 3.22 \log n$
- ▶ $k < 5 \log n$
- ▶ $h_0 = 2.64 \times IQR \times n^{-1/3}$
- ▶ ...

Szerokość przedziału

$$h \approx \frac{x_{max} - x_{min}}{k}$$

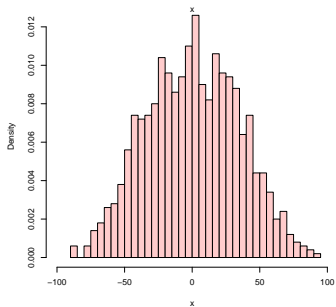
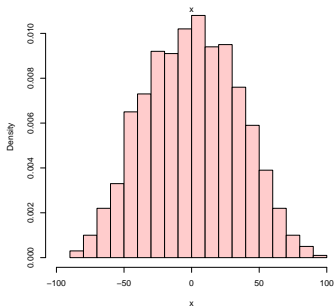
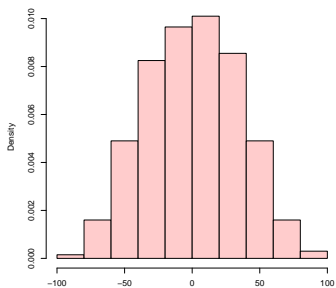
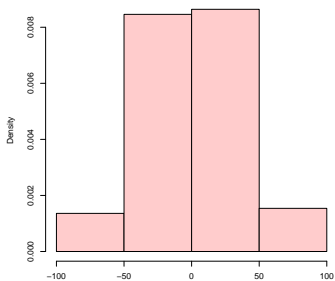
h – szerokość przedziału

x_{min} – najmniejsza obserwacja

x_{max} – największa obserwacja

k – liczba przedziałów

Wybór liczby przedziałów



Konstrukcja szeregu rozdzielczego

1. Wybór liczby przedziałów (klas) k
2. Obliczenie szerokości przedziału h
3. Ustalenie początku pierwszego przedziału
4. Utworzenie przedziałów
5. Zliczenie obserwacji należących do każdego z przedziałów