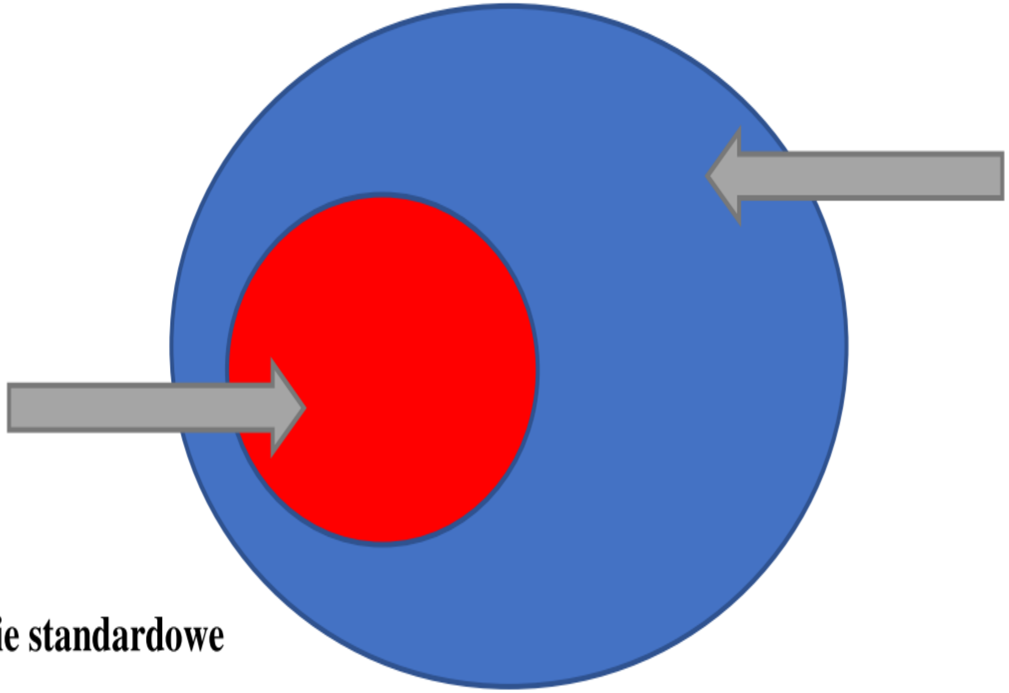
PODSTAWY PROJEKTOWANIA KONSTRUKCJI

ĆWICZENIA PROJEKTOWE

STATYSTYKA OPISOWA

POPULACJE I ZBIORY



PRÓBA

$\overline{x}$ - średnia

s - odchylenie standardowe

s^2 - wariancja

POPULACJA

μ - średnia (wartość przeciętna)

σ - odchylenie standardowe

σ^2 - wariancja

ZMIENNA LOSOWA

Zmienna losowa

dyskretna

Gdy wartości zmiennej losowej X są izolowanymi punktami na osi liczbowej (obejmują skończony przedział wartości)

- *Rzut monetą*
- *Błędy przy transmisji*
- *Wadliwe układy z linii produkcyjnej.*
- *Ilość połączeń przychodzących w ciągu 5 minut*

ciągła

Gdy wartości zmiennej losowej stanowią wszystkie punkty odcinka (obejmują przedział liczb rzeczywistych)

- *Natężenie prądu w przewodniku*
- *Temperatura*
- *Ciśnienie*

STATYSTYKA OPISOWA

Statystyki opisowe

Miary tendencji centralnej

1. Modalna – Mo – inaczej Dominanta
2. Mediana – Me
3. Średnie:
   a. Kwantyle

   Tercyle na 3 części

   Kwartyle na 4 części

   Kwintyle na 5 części

   Decyle na 10 częći

   Percentyle na 100 części

Miary rozproszenia

1. Rozstęp (najwyżej – najniżej)
2. Wariancja (wokół średniej)
3. Odchylenie standardowe (pierwiastek kwadratowy z wariancji)

Miary symetrii rozkładu

1. Skośność (lewo i prawo stronna)
2. Kurtoza (leptokurtyczna, platykurtyczna)
3. Błąd standardowy
4. Test Kołmogorowa-Smirnowa
5. Test Shapiro-Wilka

WARTOŚĆ MODALNA

Modalna- jest to najczęściej występująca wartość w badanym przypadku.

Przykład 3. Znajdź modalną ocen na świadectwie.

$$2, 3, 3, 3, 4, 4, 4, \underbrace{5, 5, 5, 5}_{\textbf{4 piątki}}, 6$$

Modalna ocen na świadectwie wynosi 5.

$$D_0 = x_0 + \frac{(n_0 - n_{-1})}{(n_0 - n_{-1}) + (n_0 - n_{+1})} \times h_o$$

Gdzie:

x_0 $-$ *dolna granica przedziału dominanty*

n_0 $-$ *liczebność przedziału domianinanty*

n_{-1} $-$ *liczebność przedziału poprzedzającego przedział dominanty*

n_{+1} $-$ *liczebność przedziału następującego po przedziale domiannty*

h_o $-$ *rozpiętość przedziału dominanty (przedziały sąsiadujące muszą mieć taką samą rozpiętość)*

MEDIANA

Mediana (wartość środkowa) – dzieli zbiorowość obserwacji na dwie części. Połowa jednostek ma wartości mniejsze lub równe medianie, a połowa ma wartości równe lub od niej mniejsze, jeżeli ilość uporządkowanych wartości jest parzysta, to mediana jest średnią arytmetyczną dwóch wartości środkowych.

$$m_e = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{gdy n jest liczbą nieparzystą} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2} & \text{gdy n jest liczbą parzystą} \end{cases}$$

ŚREDNIA

Średnia arytmetyczna

Średnia arytmetyczna liczb $a_1, a_2, \ldots, a_n$, to **iloraz sumy** *n* zmiennych **przez liczbę** tych zmiennych.

$$\boxed{S_A = \bar{a} = \frac{\sum_{k=1}^{n} a_k}{n}}$$

Średnia ważona

Przy założeniu, że prawdopodobieństwo uzyskania każdego z wyników a_i wynosi odpowiednio: p_i , to średnią ważoną definiuje się jako **iloraz sumy**: *n* iloczynów prawdopodobieństwa uzyskania wyniku przez wartość uzyskaną w doświadczeniu **przez sumę** prawdopodobieństw uzyskania poszczególnych wyników.

$$\boxed{S_W = \frac{\sum_{k=1}^{n} p_k a_k}{\sum_{k=1}^{n} p_k}}$$

Średnia geometryczna

Średnia geometryczna, nazywana również ważoną średnią geometryczną liczb $a_1, a_2, \ldots, a_n$, jest to pierwiastek *n*-tego stopnia z iloczynu *n* liczb a_k :

$$\boxed{S_g = \sqrt[n]{\prod_{k=1}^{n} a_k}}$$

ŚREDNIA

Średnia harmoniczna

Średnia harmoniczna, zwana również ważoną średnią harmoniczną, jest to iloraz ilości pomiarów, których jest "n", dla których liczymy tą naszą średnią, przez sumę odwrotności tychże liczb:

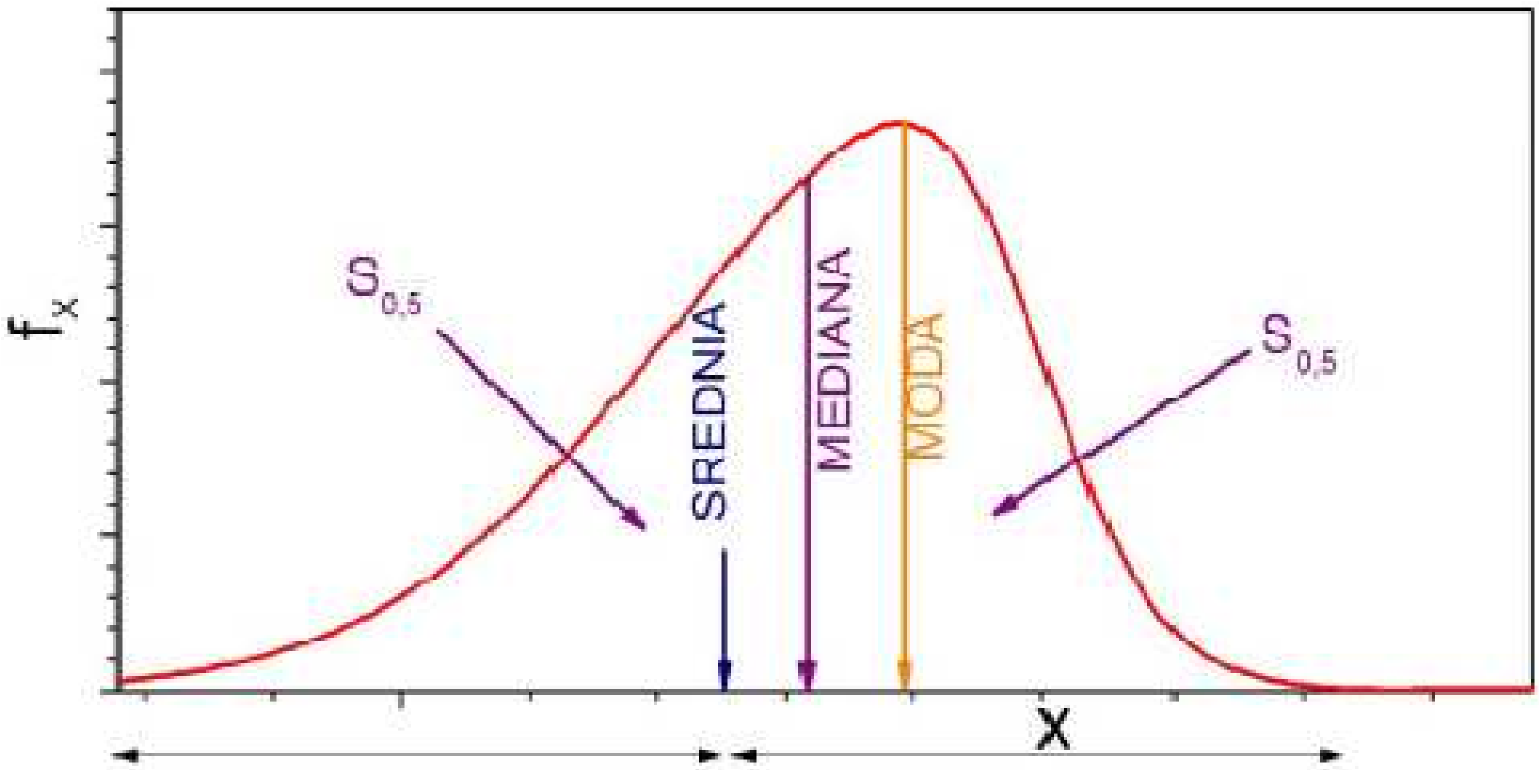
$$\boxed{S_H = \frac{n}{\sum_{k=1}^{n} \frac{1}{a_k}}}$$

Średnia kwadratowa

Średnia kwadratowa to przykład miary statystycznej liczb $a_1, a_2, \ldots, a_n$. Jest to pierwiastek **ilorazu sumy** kwadratów *n* tychże liczb **przez ich liczbę**.

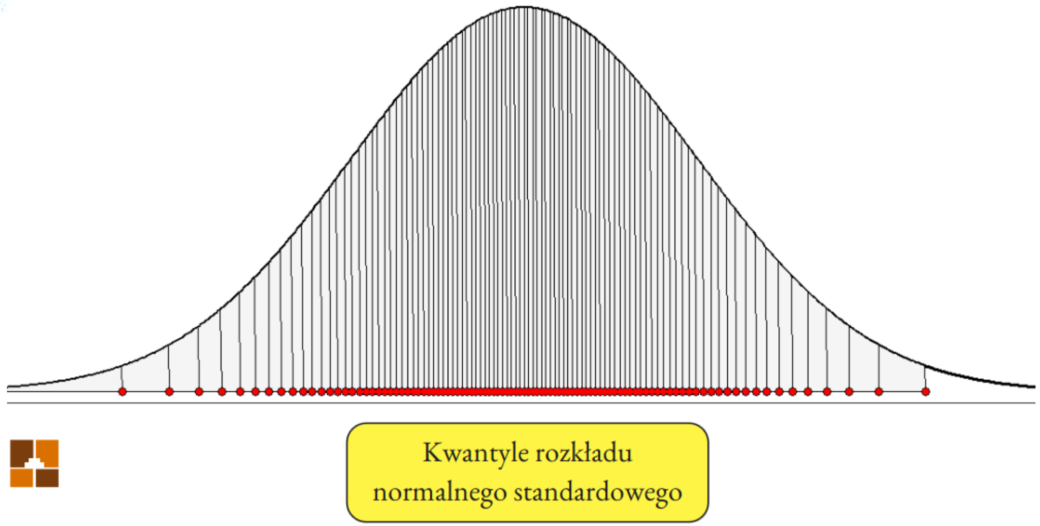
$$\boxed{S_K = \sqrt{\frac{\sum_{k=1}^{n} a_k^2}{n}}}$$

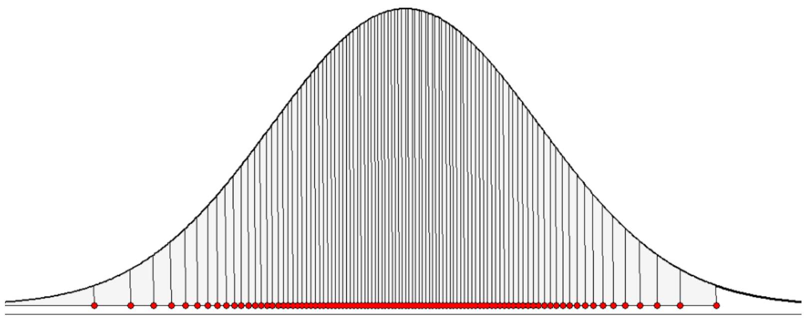
ŚREDNIA



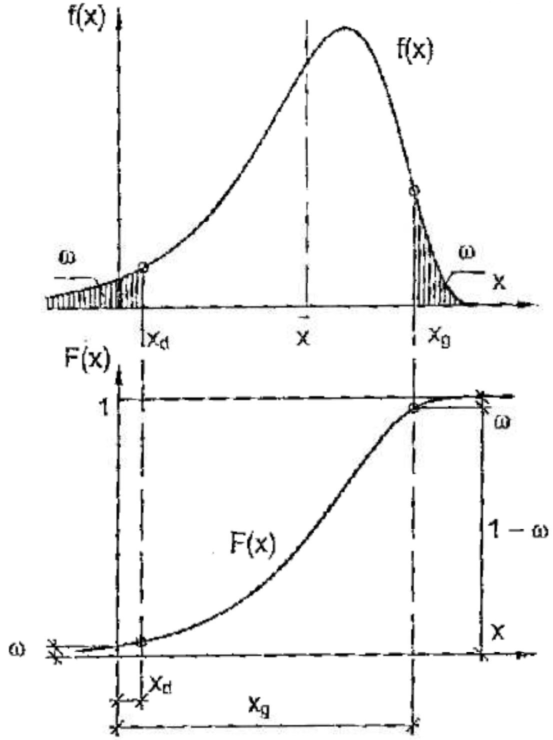
KWANTYLE


Poziom Zaufania
95%
Średnia
dolna granica
górna granica


Kwantyle rozkładu normalnego standardowego



Kwantyle rozkładu normalnego standardowego

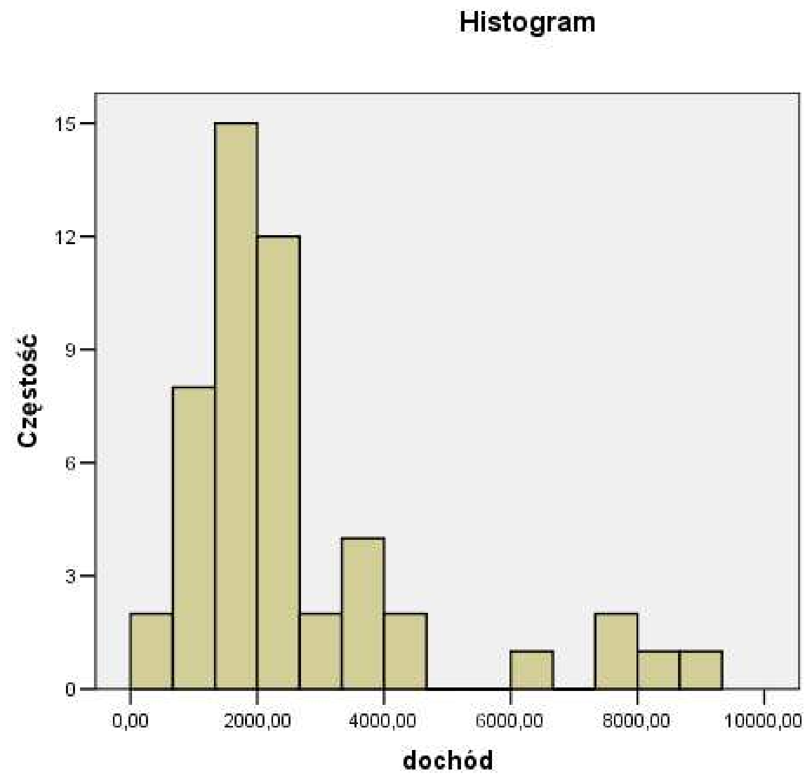


Rys. 1.4. Kwantyle dolny x_d i górny x_g zmiennej losowej $\underline{x}$

ROZSTĘP

Rozstęp - różnica pomiędzy wartością maksymalną, a minimalną cechy - jest miarą charakteryzującą empiryczny obszar zmienności badanej cechy, nie daje on jednak informacji o zróżnicowaniu poszczególnych wartości cechy w zbiorowości.

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



WARIANCJA

Wariancja - jest to średnia arytmetyczna kwadratów odchyleń poszczególnych wartości cechy od średniej arytmetycznej zbiorowości.

Współczynnik zmienności - jest ilorazem bezwzględnej mary zmienności cechy i średniej wartości tej cechy, jest wielkością niemianowaną, najczęściej podawaną w procentach.

$$S_X^2 = \frac{\sum(X - \bar{X})^2}{N}$$

$S_X^2 -$ wariancja

X – wartości obserwacji

$\bar{X}$ – średnia wartości obserwacji

N – liczba obserwacji

$$\upsilon = \frac{s}{\underline{x}} \cdot 100\%$$

ODCHYLENIE PEŁNOMOCNICTWA

***Odchylenie standardowe (zróżnicowanie poszczególnych wartości od średniej)* -** jest to pierwiastek kwadratowy z wariancji. Stanowi miarę zróżnicowania o mianie zgodnym z mianem badanej cechy, określa przeciętne zróżnicowanie poszczególnych wartości cechy od średniej arytmetycznej.

Jeżeli zbiorowość danych którą posiadamy jest tylko próbką z dużej zbiorowości odchylenie standardowe jest oszacowaniem (estymatorem), oznaczane jako „s” i liczone ze wzoru:

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

Jeżeli zbiorowość danych, którą posiadamy jest całą populacją odchylenie standardowe jest dokładne, oznaczane jako „σ” i liczone ze wzoru:

$$\sigma = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

SKOŚNOŚĆ

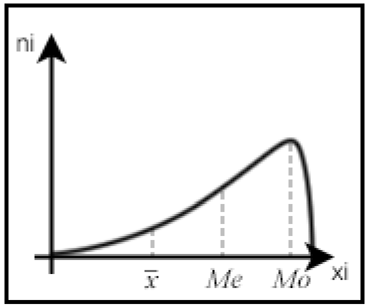
$$A = \frac{\bar{X} - Mo}{S_X}$$

A – współczynnik skośności

$\bar{X}$ – średnia wartości obserwacji

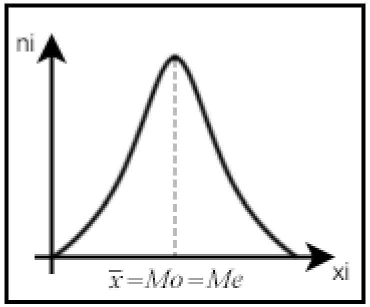
Mo – dominanta (moda)

S_X – odchylenie standardowe



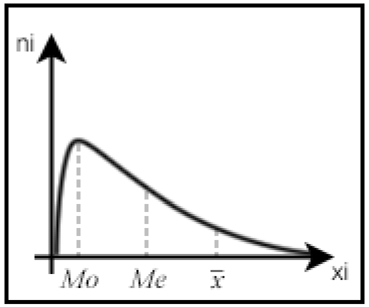
Asymetria lewostronna

$\bar{x}$ < Me < Mo



Rozkład symetryczny

$\bar{x}$ = Me = Mo

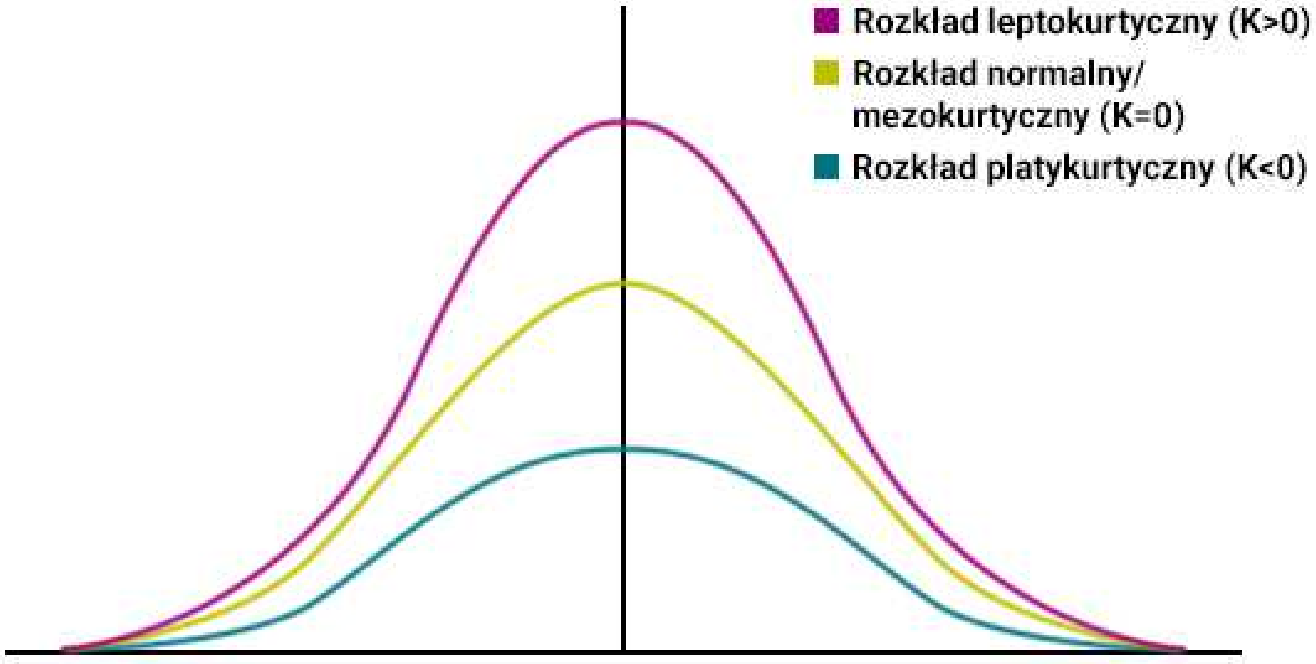


Asymetria prawostronna

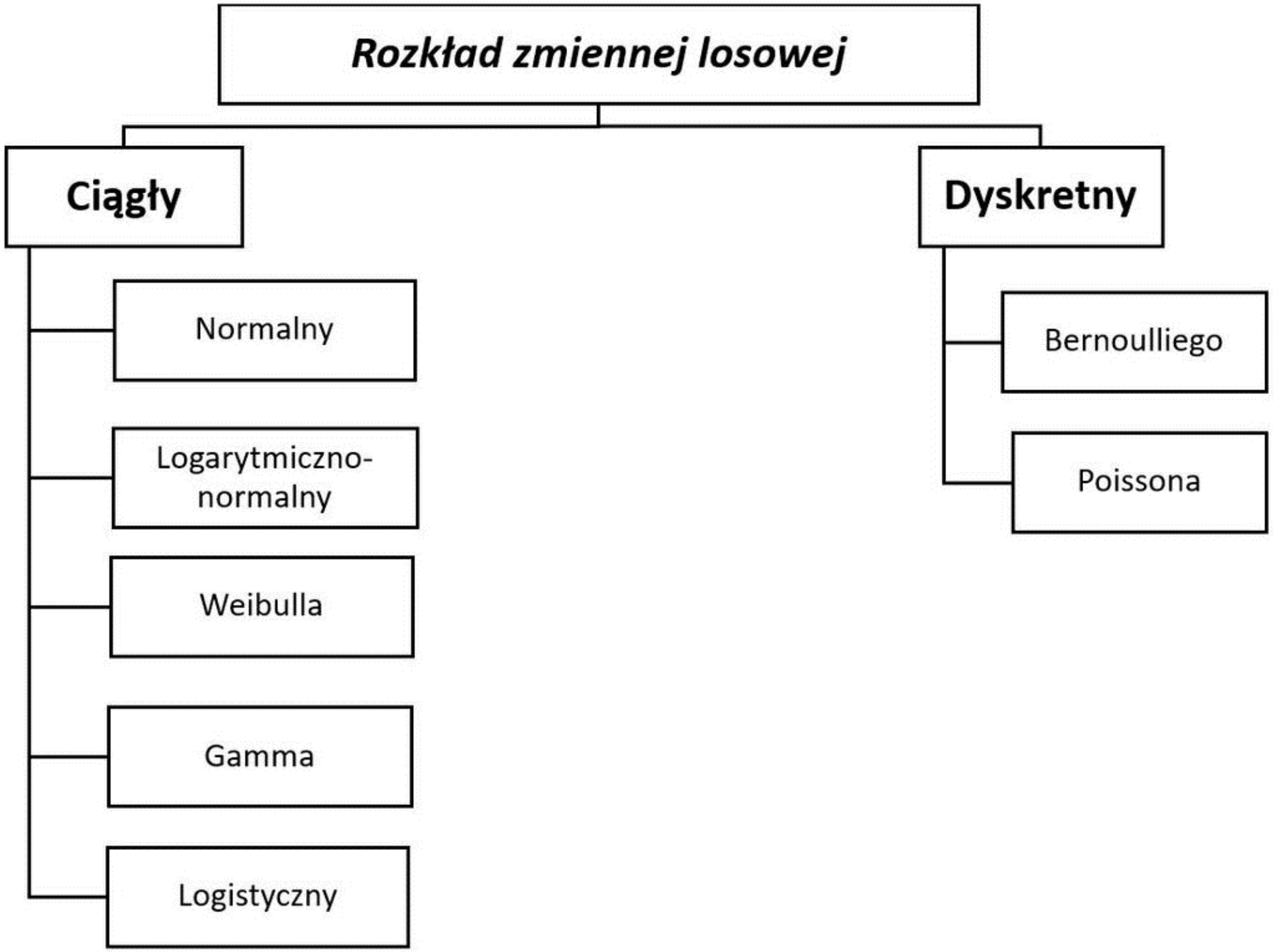
$\bar{x}$ > Me > Mo

KURTOZA

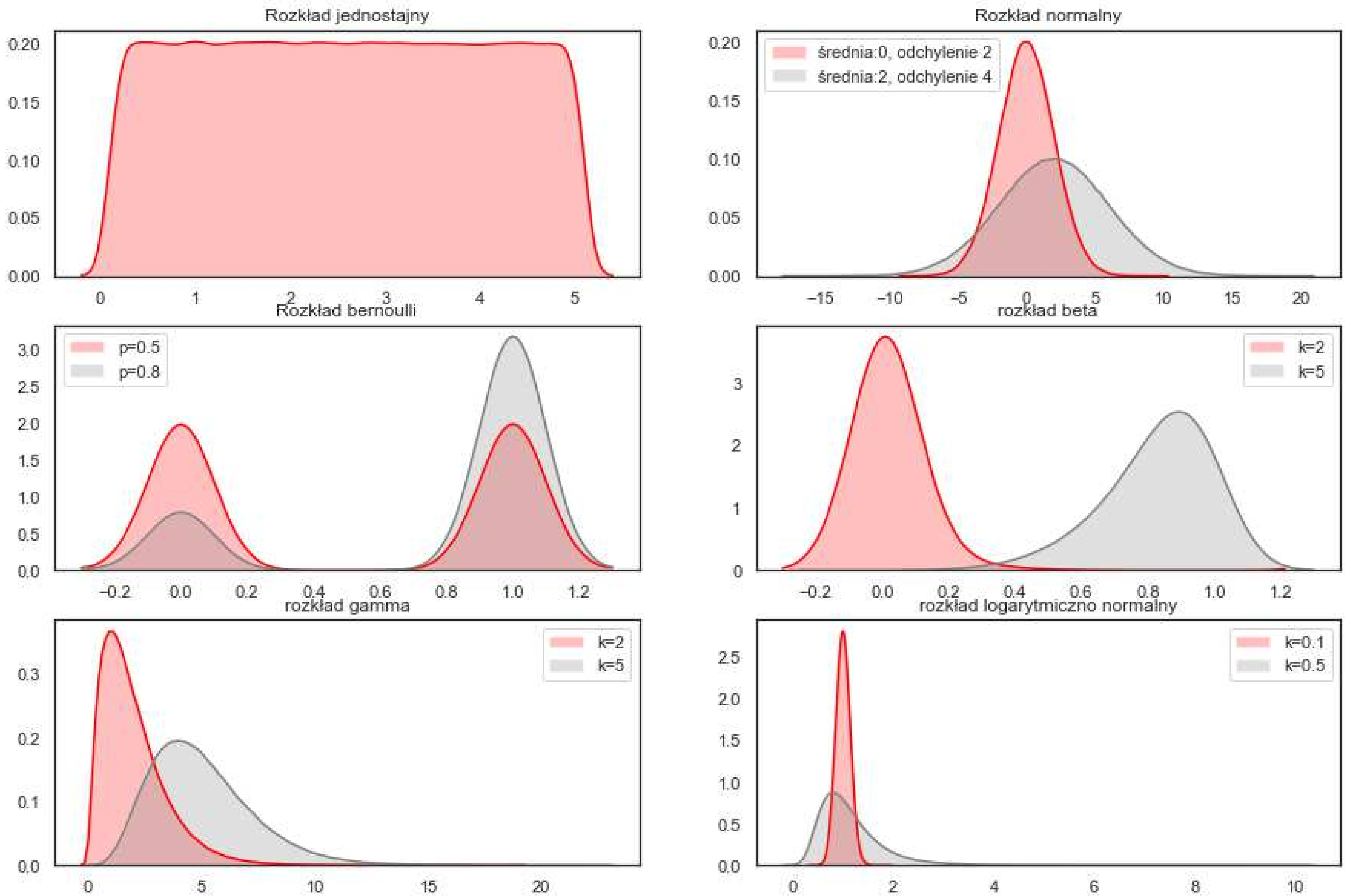
Kurtoza charakteryzuje **względną szczytowość lub płaskość rozkładu w porównaniu z rozkładem normalnym**.



ROZKŁADY ZMIENNEJ LOSOWEJ



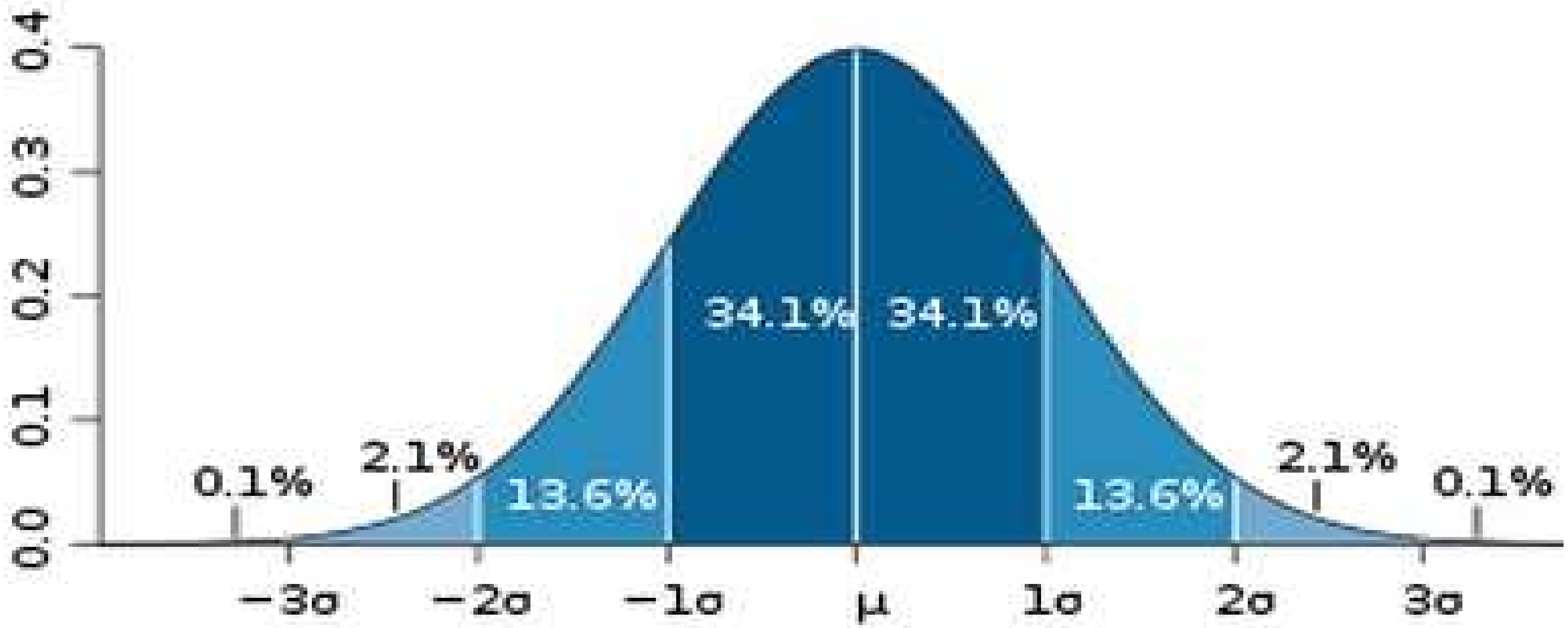
ROZKŁADY ZMIENNEJ LOSOWEJ



ROZKŁADY ZMIENNEJ LOSOWEJ

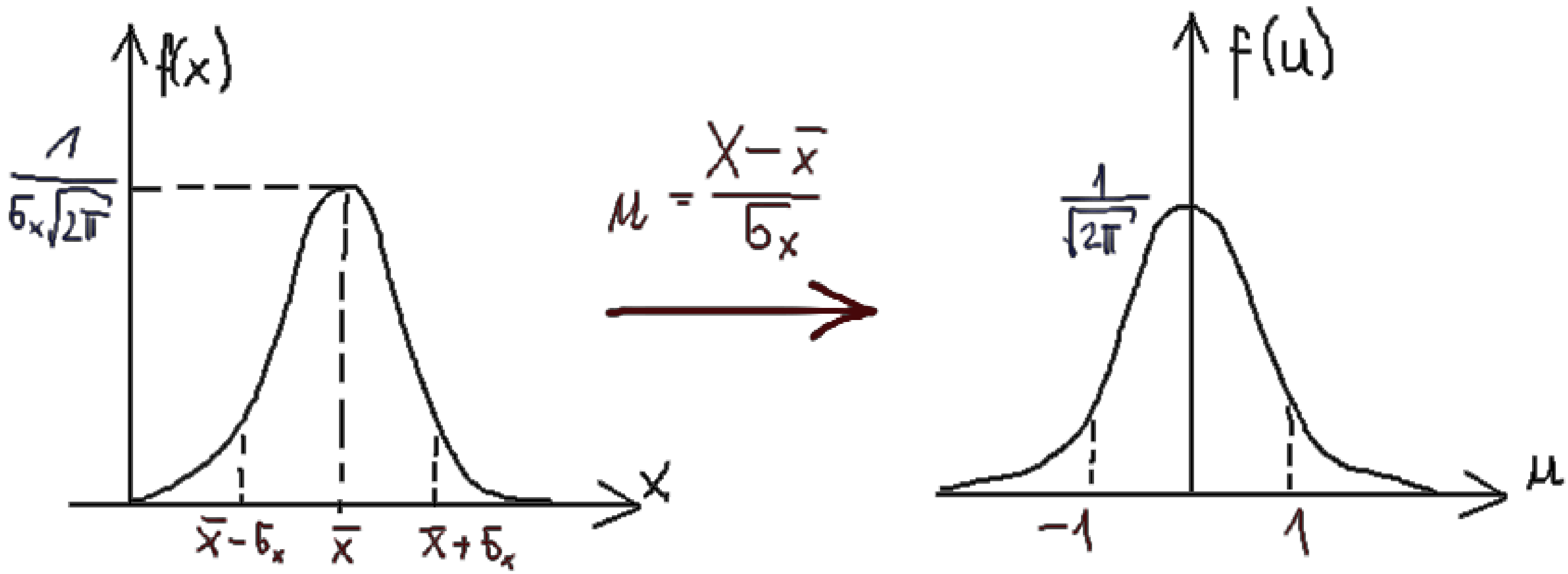
Rozkład:	Ocena wartości centralnej:	Nazwa:
normalny „N"	$\bar{x} \simeq \frac{1}{n} \sum_i x_i$	– w. oczekiwana
log-normalny „LN"	$\breve{x} \simeq \sqrt[n]{\prod_i x_i}$	– mediana
Gumbela (dla max.) „G"	$\check{x} \simeq \ln\left(\frac{1}{n} \sum_i e^{-\frac{x_i}{u}}\right)^{-u}$	– moda (w. modalna)
Weibulla (dla min.) „W"	$\hat{x} \simeq \left(\frac{1}{n} \sum_i \sqrt[\omega]{x_i}\right)^{\omega}$	– w. charakterystyczna

ROZKŁAD NORMALNY



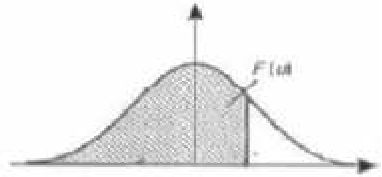
$$U = \frac{X - x}{\sigma_x}$$

ROZKŁAD NORMALNY



ROZKŁAD NORMALNY

Dystrybuanta rozkladu normalnego

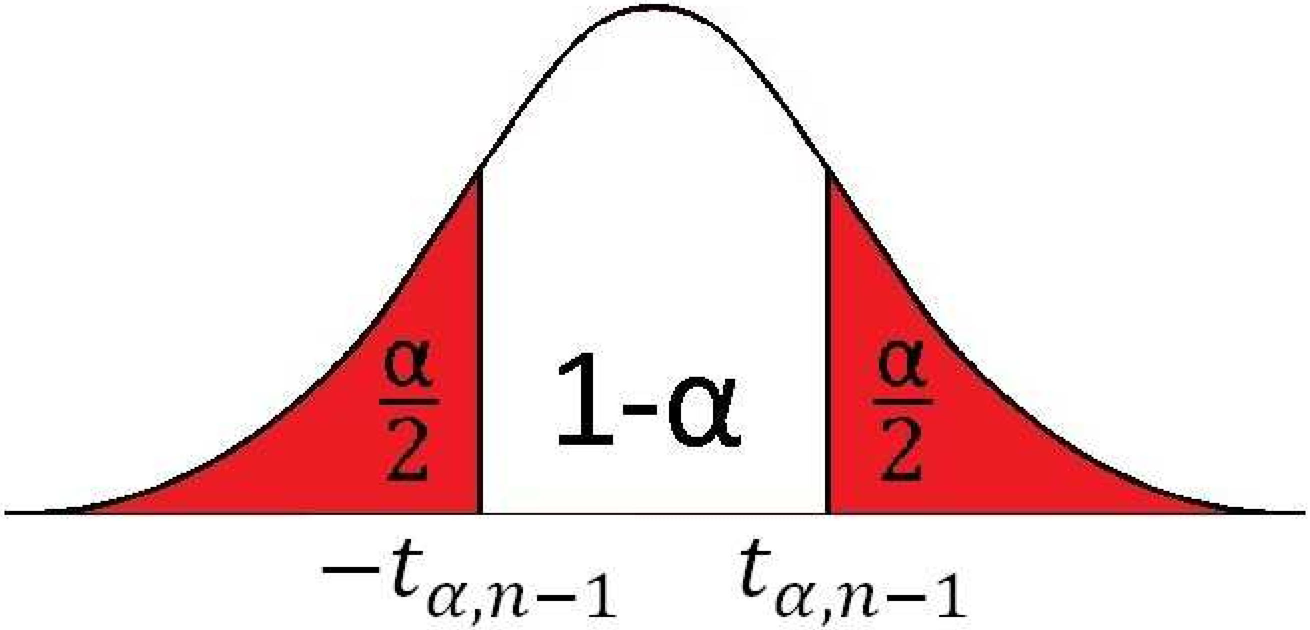


$P(U < 1) = \varphi(1) = 0{,}8413$

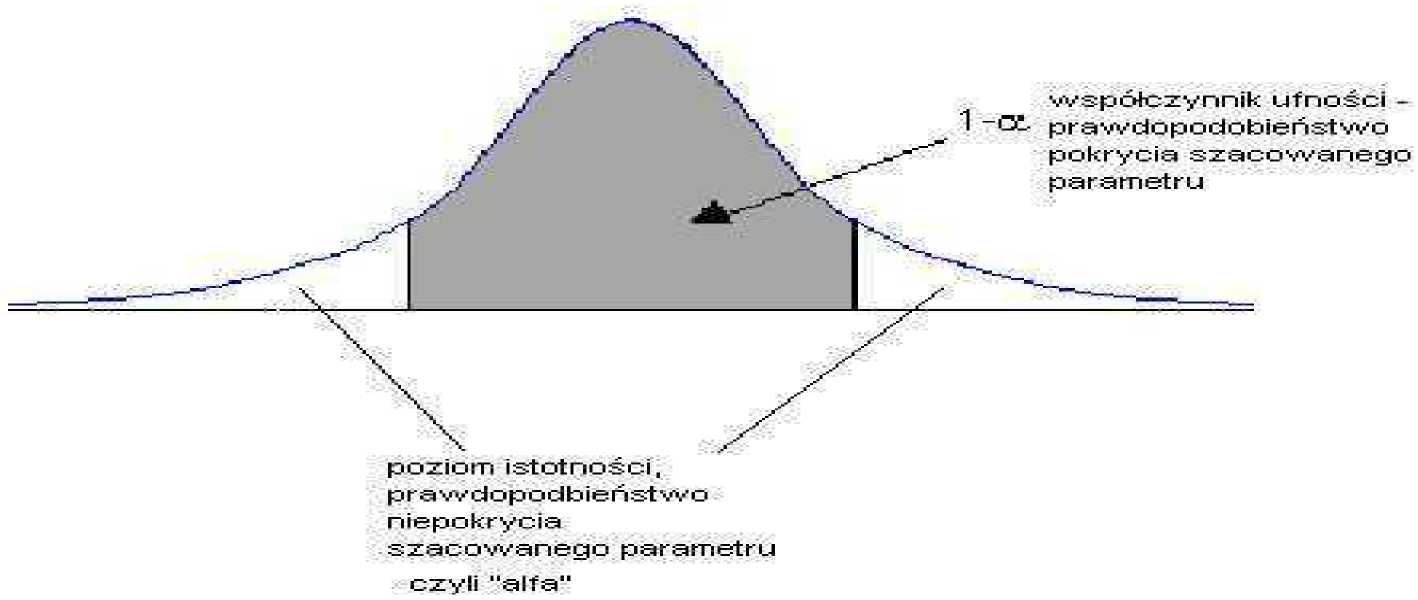
$F(u) = P(U \leq u)$, dla $u \geq 0$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441

ROZKŁAD T-STUDENTA



POZIOM UFNOŚCI I ISTOTNOŚCI



PORZĄDKOWANIE DANYCH

Pomierzone średnice 40-tu amonitów [cm] wynoszą:

3,2	3,7	3,9	3,4	3,1	3,1	3,1	3,9	3,5	3,3
3,6	3,8	3,7	3,0	3,5	3,2	3,5	3,7	3,9	3,6
3,4	2,9	3,2	3,4	2,9	3,6	3,7	3,3	3,4	4,0
3,8	3,7	3,3	2,9	3,1	3,2	3,6	3,5	3,3	3,4

podsumujmy:

wszystkich przypadków: n = **40**;

przybliżeniem ilości klas będzie pierwiastek kwadratowy z 40, k = ok. **6**;

wartość minimalna x_{min} = **2,9**;

wartość maksymalna x_{max} = **4,0**;

Różnica pomiędzy $x_{max} - x_{min} = \Delta X$ = **1,1**;

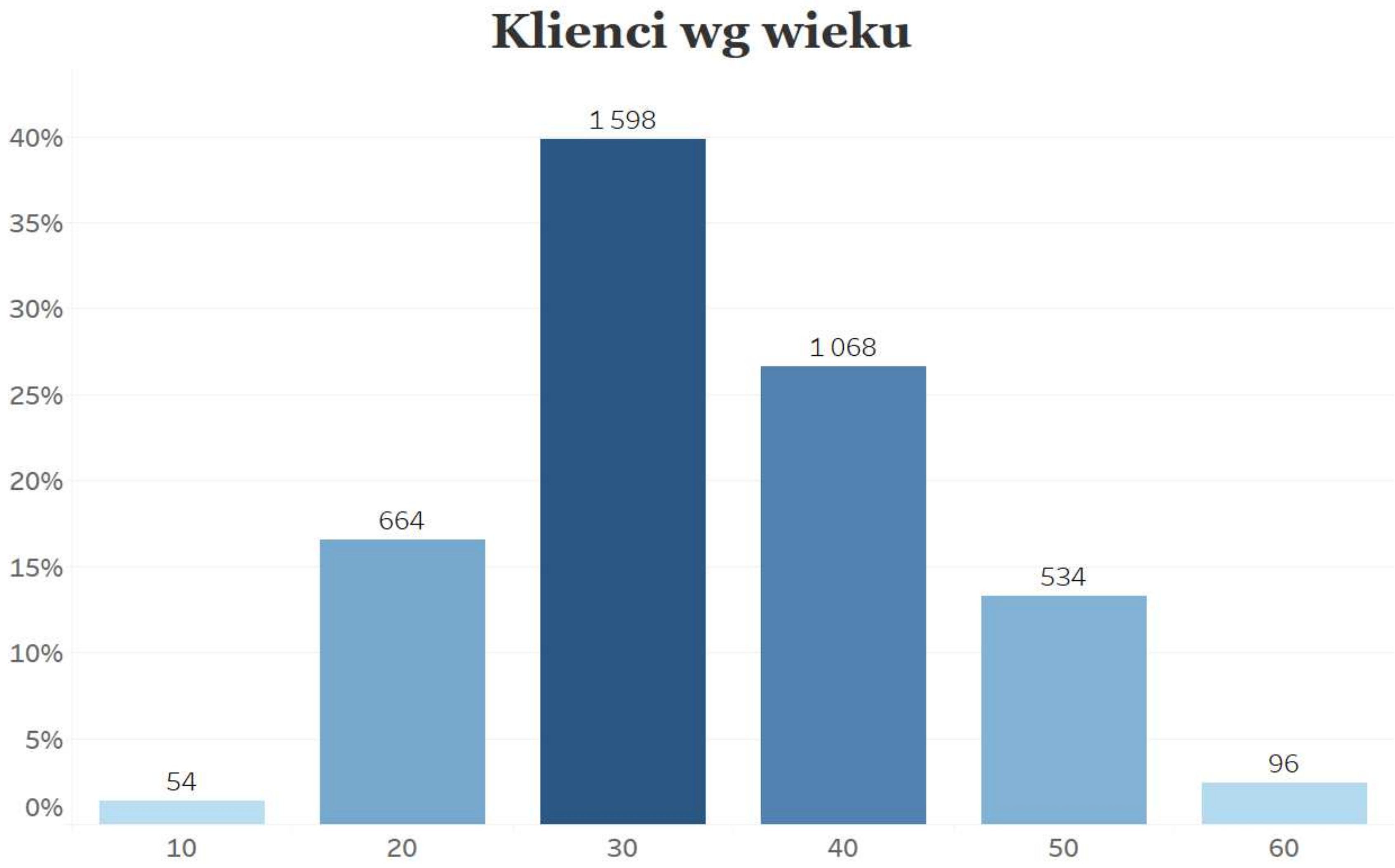
szerokość przedziału klasowego: $\Delta X / k$ = **0,18** (dla wygody wartość tą rozszerzamy do **0,2**);

Szereg rozdzielczy przedstawiono poniżej w formie tabelarycznej.

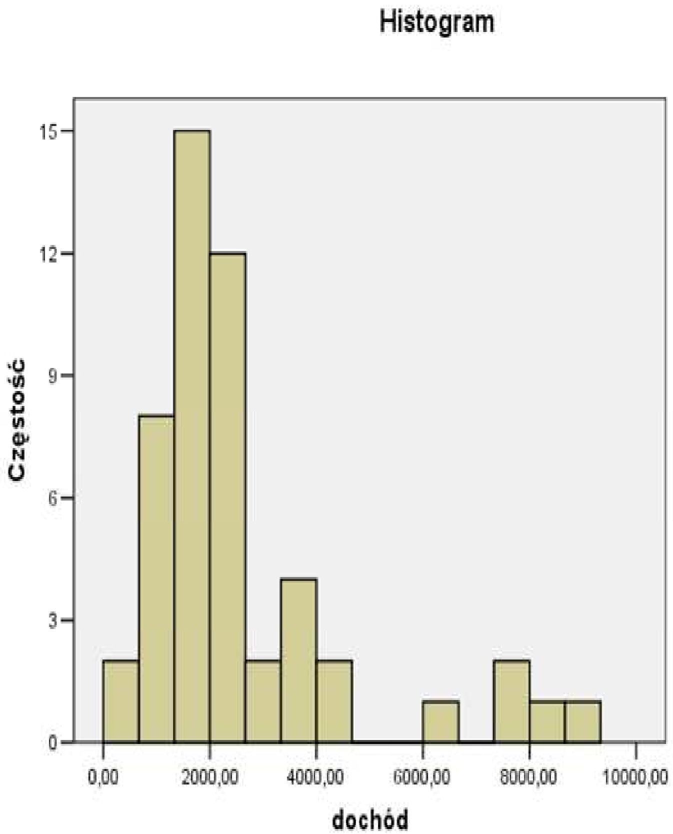
Szereg rozdzielczy:

lp.	**przedział**	**liczebności (l_i)**	**częstości (p_i) [%]**
1.	2,85-3,05	4	10
2.	3,05-3,25	8	20
3.	3,25-3,45	9	22,5
4.	3,45-3,65	8	20
5.	3,65-3,85	7	17,5
6.	3,85-4,05	4	10
		Σ = 40	Σ = 100%

PORZĄDKOWANIE DANYCH



PORZĄDKOWANIE DANYCH



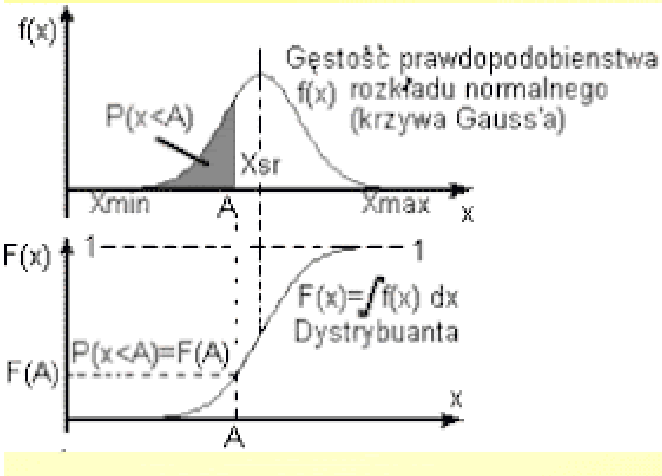
Podstawą do zbudowania szeregu rozdzielczego jest odpowiednie pogrupowanie wyników w klasy. W tym celu konieczne jest ustalenie dla danej próby: **rozstęp próby, ilości klas, długości klas, początku klasy dolnej**.

Rozstęp próby: $R = max - min$

Liczba klas: $k \approx \sqrt{n}$ lub $k \approx 1 + 3{,}322 \log n$

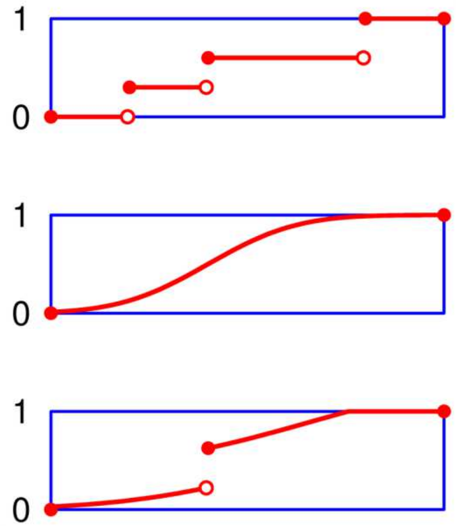
Długość klas: $b = R/k$

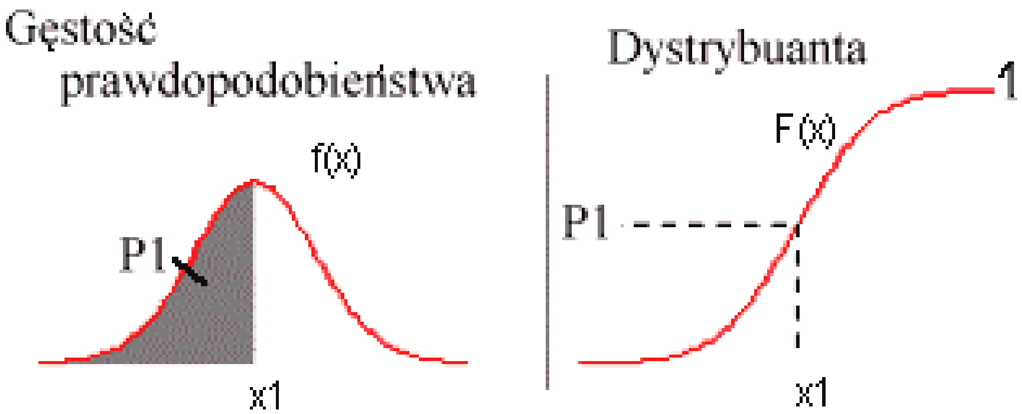
PORZĄDKOWANIE DANYCH



Funkcja f(x) gęstości (ang.: density) rozkładu prawdopodobieństwa jest pochodną dystrybuanty F(x)

Dystrybuanta F(x) czyli rozkład skumulowany jest więc całką funkcji f(x) gęstości rozkładu prawdopodobieństwa





DZIĘKUJĘ ZA UWAGĘ!