

STATYSTYKA OPISOWA

Plan wykładów

Próba – własności, porządkowanie, szeregi rozdzielcze, wykresy

Parametry próby (poziomu, zmienności, skośności, kształtu, koncentracji) klasyczne i pozycyjne

Porównywanie prób, wykresy parametrów

Korelacja i regresja - własności

Wanda Olech

Katedra Genetyki i Ochrony Zwierząt

Próba – własności,
porządkowanie, szeregi
rozdzielcze, wykresy

STATYSTYKA OPISOWA

Podstawowe pojęcia

- Zbiorowość statystyczna – zbiór jednostek (obserwacji) nie identycznych, ale stanowiących logiczną całość
- Zbiorowość (populacja) generalna – skończony lub nieskończony zbiór jednostek, które zamierzamy poddać obserwacji
- Próba – część (podzbiór) zbiorowości generalnej, która jest badana pod względem ustalonej cechy
- Próba losowa – próba do której losowo wybrano jednostki
- Cecha badana – właściwość jednostek zbiorowości, która może służyć do odróżnienia poszczególnych obserwacji

Własności próby

- Ma określoną liczebność (N)
- Jednorodność – należy do zbiorowości generalnej
- Reprezentatywność względem zbiorowości generalnej

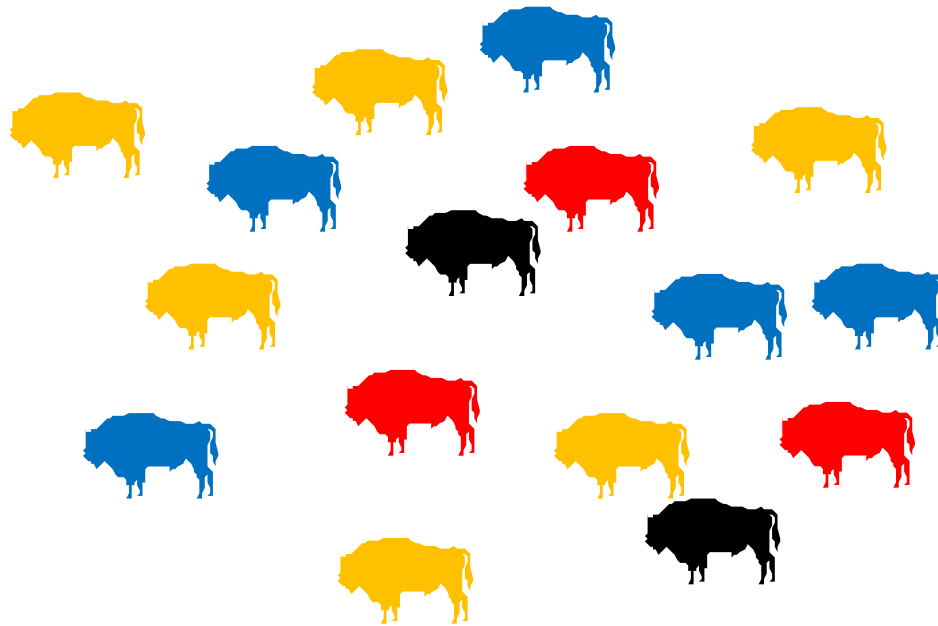
Cechy statystyczne dzielimy na

- Jakościowe – opisują jakości elementów zbiorowości, np. kolor oczu, wykształcenie, marka samochodu
- Ilościowe – wartości liczbowe
 - skokowe – wartości są oddalone od siebie, np. liczba dzieci w rodzinie,
 - ciągłe – wartości są w przedziale liczbowym, np. wzrost człowieka

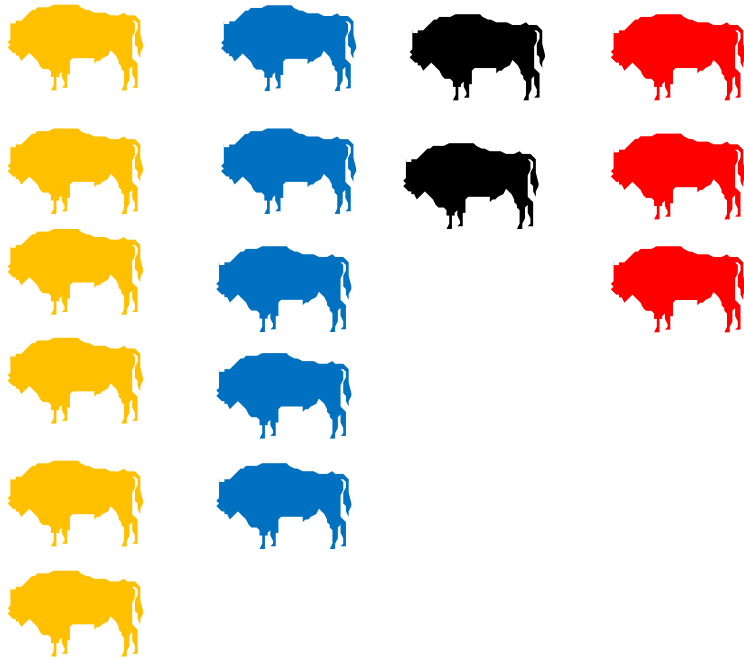
Opis próby - statystyka opisowa

- Uporządkowanie próby i prezentacja graficzna
- Parametry próby
- Porównania prób pod względem poziomu, zmienności i skośności

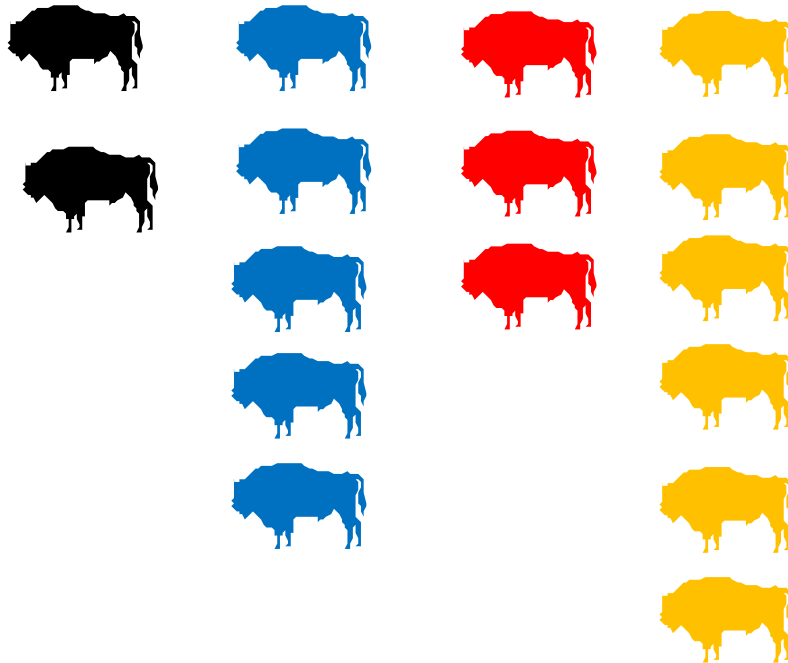
Cecha jakościowa







Cecha jakościowa



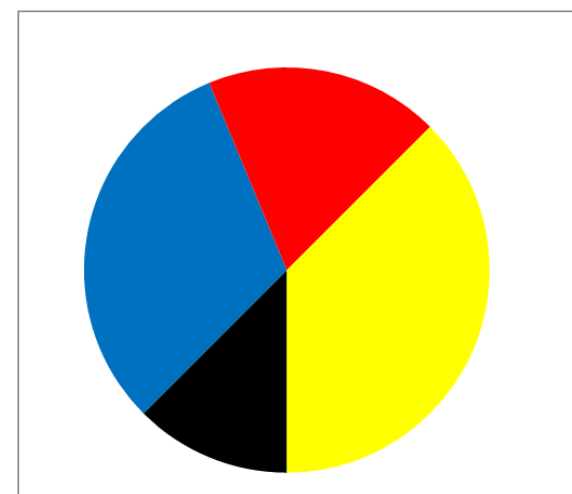
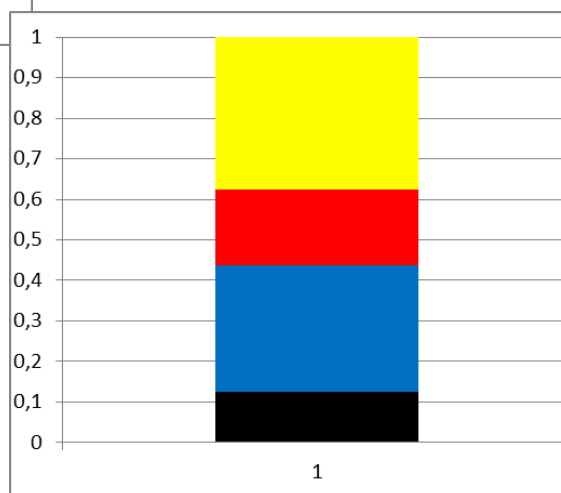
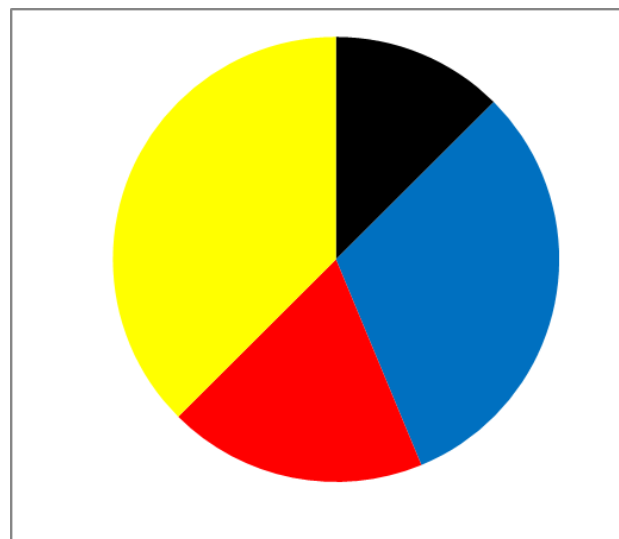
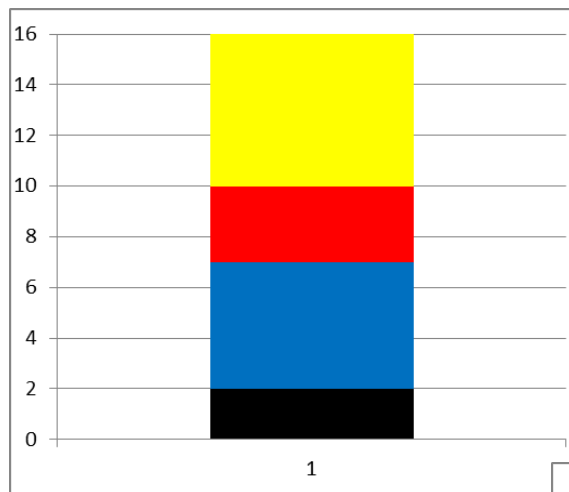
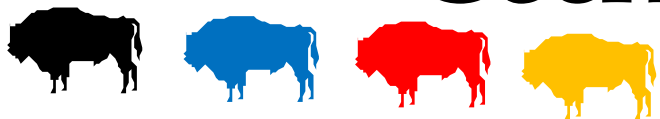
Cecha jakościowa



Cecha jakościowa

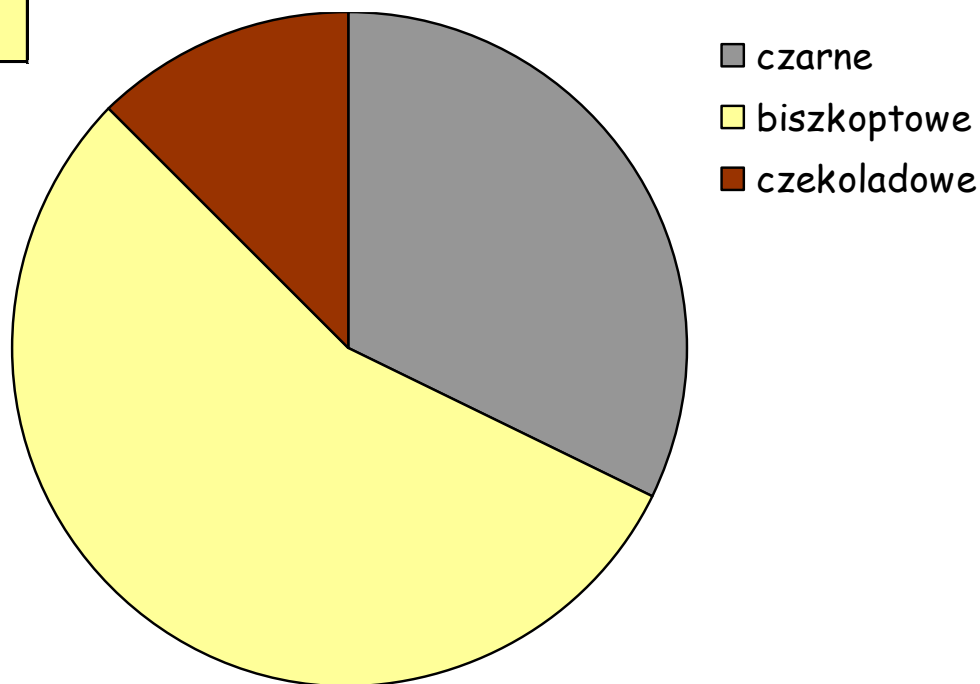
	n_i	w_i
	2	0,125
	5	0,3125
	3	0,1875
	6	0,375
	16	1

Cecha jakościowa



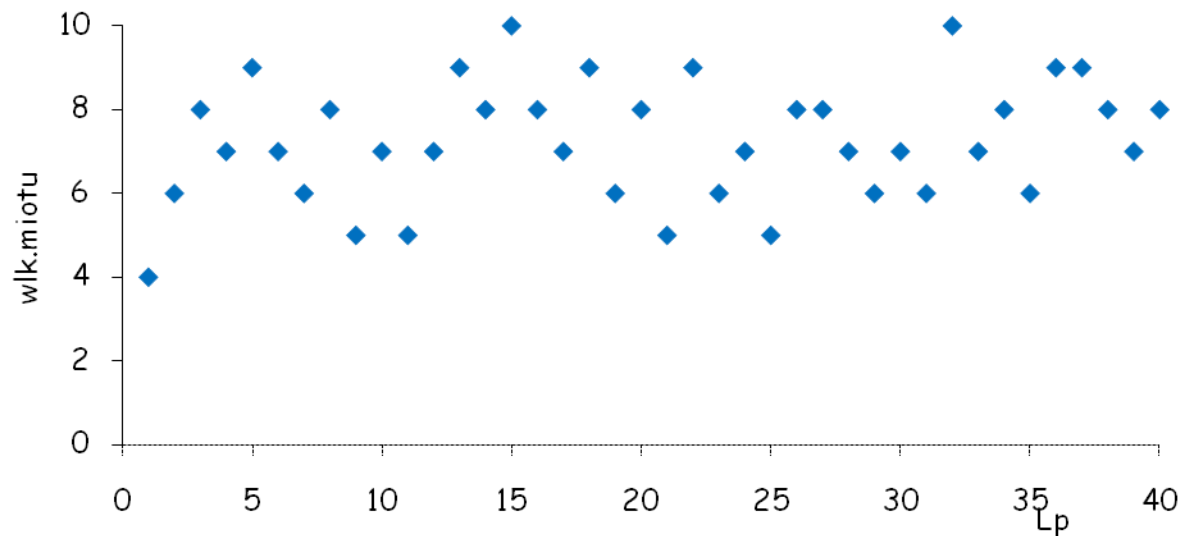
Cecha jakościowa - umaszczzenie

umaszczzenie	liczba psów	% psów
	n_i	w_i
czarne	18	32,1
biskoptowe	31	55,4
czekoladowe	7	12,5
RAZEM	56	100,0



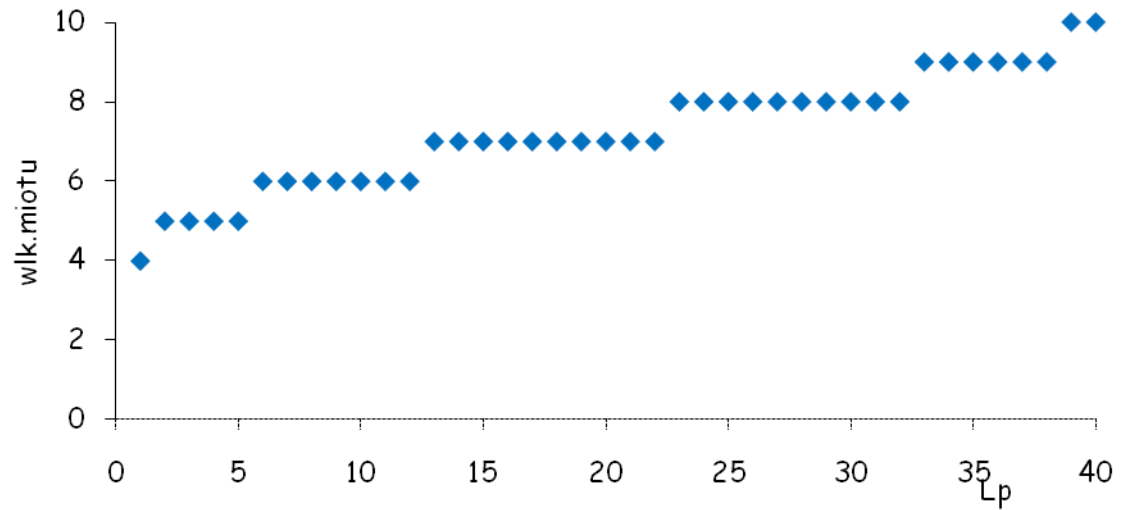
Cecha ilościowa skokowa – wielkość miotu

4	6	8	7	9
7	6	8	5	7
5	7	9	8	10
8	7	9	6	8
5	9	6	7	5
8	8	7	6	7
6	10	7	8	6
9	9	8	7	8



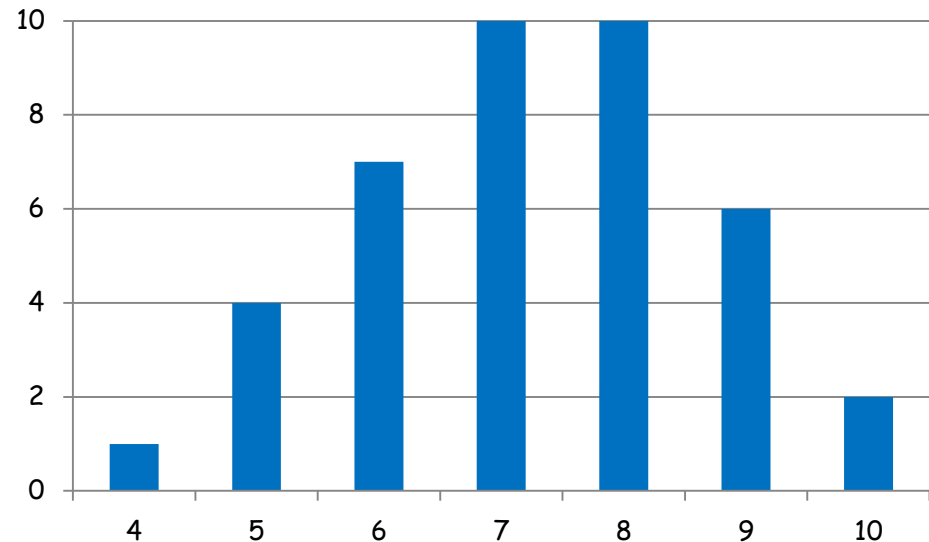
Uporządkowanie cechy ilościowej skokowej

4	5	5	5	5
6	6	6	6	6
6	6	7	7	7
7	7	7	7	7
7	7	8	8	8
8	8	8	8	8
8	8	9	9	9
9	9	9	10	10



Cecha ilościowa skokowa – szereg rozdzielczy punktowy

x_i	n_i	w_i
4	1	0,025
5	4	0,100
6	7	0,175
7	10	0,250
8	10	0,250
9	6	0,150
10	2	0,050
razem	40	1



Średnia arytmetyczna (ważona)

x_i	n_i	w_i
4	1	0,025
5	4	0,100
6	7	0,175
7	10	0,250
8	10	0,250
9	6	0,150
10	2	0,050
razem	40	1

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

$$\begin{aligned} \bar{x} &= \frac{1}{40} (4 \cdot 1 + 5 \cdot 4 + 6 \cdot 7 + 7 \cdot 10 + 8 \cdot 10 + 9 \cdot 6 + 10 \cdot 2) = \\ &= \frac{1}{40} (4 + 20 + 42 + 70 + 80 + 54 + 20) = \frac{1}{40} \cdot 290 = 7,25 \end{aligned}$$

Cecha ilościowa – parametry

Dzielimy na parametry:

- 👤 Poziomu (tendencji centralnej)
- 👤 Zmienności (zróźnicowania, dyspersji)
- 👤 Asymetrii (skośności)
- 👤 Kurtozy (kształtu)
- 👤 Koncentracji

Oraz dzielimy na

- 👤 Klasyczne
- 👤 Pozycyjne

Cecha ilościowa - parametry

Średnia arytmetyczna

4	6	8	7	9
7	6	8	5	7
5	7	9	8	10
8	7	9	6	8
5	9	6	7	5
8	8	7	6	7
6	10	7	8	6
9	9	8	7	8

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j = \frac{1}{N} (x_1 + x_2 + \dots + x_N)$$

$$\bar{x} = \frac{1}{40} (4 + 6 + 8 + \dots + 7 + 8) = \frac{1}{40} \cdot 290 = 7,25$$

Średnia - własności

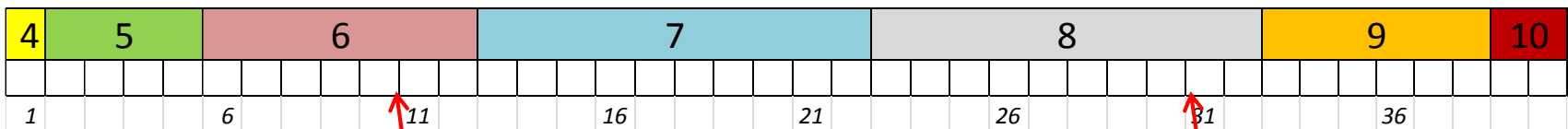
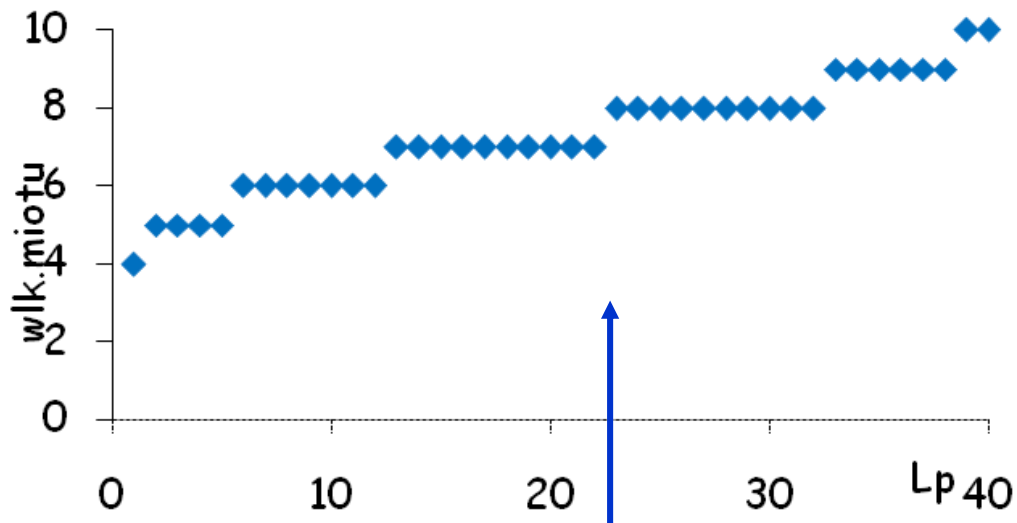
Średnia arytmetyczna

4	7	5
	2	2

$$\bar{x} = \frac{1}{5} \times (4+7+5+2+2) = 4$$

	odchylenie
2	-2
2	-2
4	0
5	1
7	3

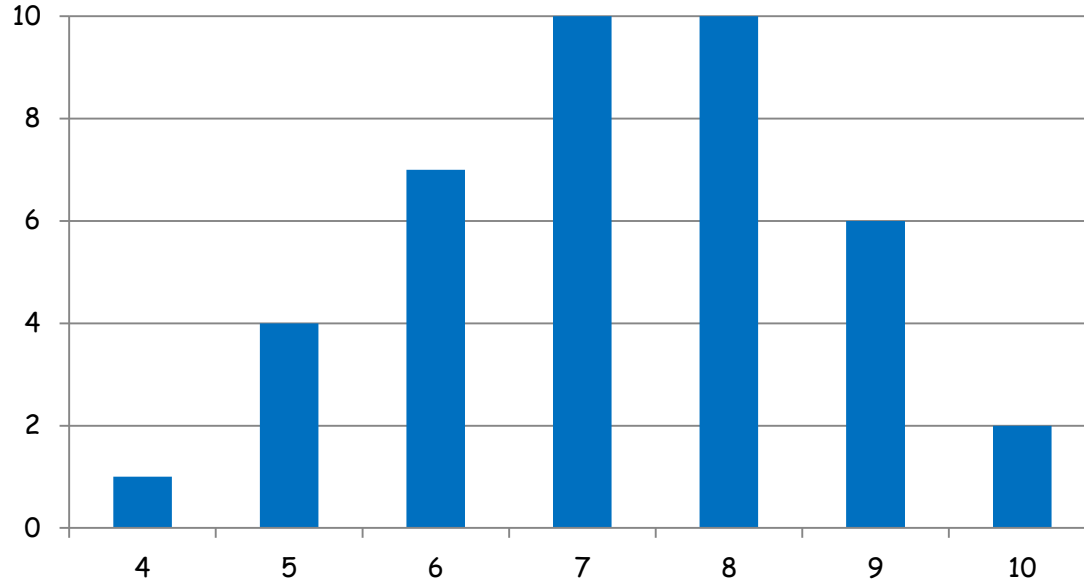
Kwantyle w tym mediana – wartość środkowa



$$Q_1=6$$

$$Q_3=8$$

Dominanta – wartość modalna, występującą najczęściej



D=7 ?

D=8 ?

Rozstęp i odchylenie ćwiartkowe

$$RS = x_{\max} - x_{\min}$$

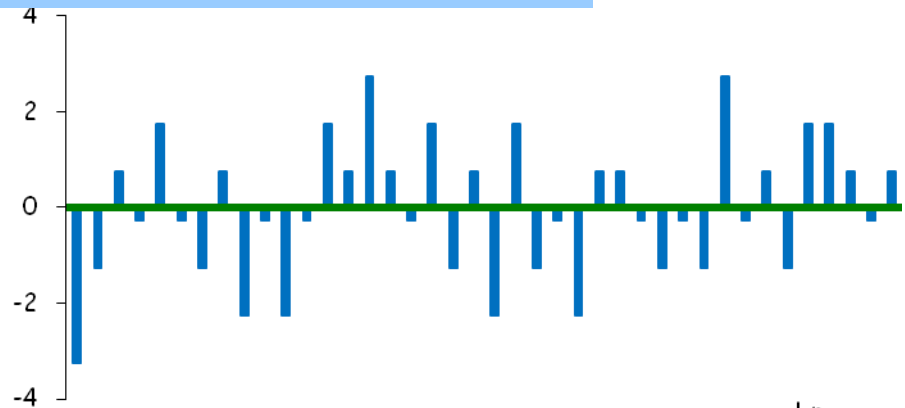
$$Q_x = 0,5 \times (Q_3 - Q_1)$$

Wariancja i odchylenie standardowe

$$S^2 = \frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2 = \frac{1}{N-1} \left(\sum_{j=1}^N x_j^2 - \frac{\left(\sum_{j=1}^N x_j \right)^2}{N} \right) \quad S = \sqrt{S^2}$$

$$S^2 = \frac{1}{39} [(4 - 7,25)^2 + \dots + (8 - 7,25)^2] = \frac{1}{39} 81,5 = 2,09$$

$$S = 1,445$$



Lp

Wskaźnik (współczynnik) zmienności

służy do porównań zróżnicowania w zbiorowościach

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

$$V = \frac{1,445}{7,25} = 0,1993$$

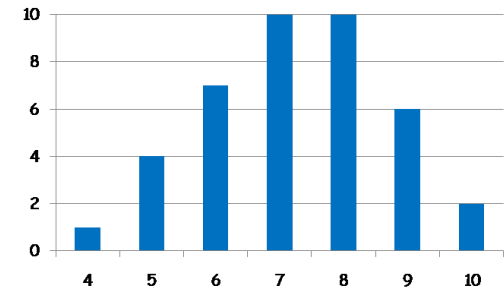
Współczynnik asymetrii (skośności)

ocenia asymetrię rozkładu próby

$$A = \frac{\frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^3}{S^3}$$

$$A'' = \frac{(\bar{x} - D)}{S}$$

$$A = -0,1303$$



Porównania prób

średnia	17,8	867,8
odch.st.	4	4
v	0,225	0,005

Porównania prób

średnia	17,8	867,8
odch.st.	4	195,2
v	0,225	0,225

Porównania prób

średnia	7,250	25,0
odch.st.	1,446	4,60
V	0,199	0,184
A	-0,130	0,650

Rozkład wartości cechy

x_i	p_i
2	0,22
5	0,46
7	0,32
suma	1

$$\begin{aligned}\bar{x} &= 2 \times 0,22 + 5 \times 0,46 + 7 \times 0,32 = \\ &= 0,44 + 2,3 + 2,24 = 4,98\end{aligned}$$

Rozkład wartości cechy

x_i	$w_i (p_i)$
1	0,3
4	0,5
7	0,1
10	0,1

$$\bar{x} = \sum_{i=1}^k x_i \cdot w_i$$

$$\begin{aligned}\bar{x} &= 1 \cdot 0,3 + 4 \cdot 0,5 + 7 \cdot 0,1 + 10 \cdot 0,1 \\ &= 0,3 + 2 + 0,7 + 1 = 4\end{aligned}$$

Wariancja

x_i	n_i	w_i
1	6	0,3
4	10	0,5
7	2	0,1
10	2	0,1
suma	20	1

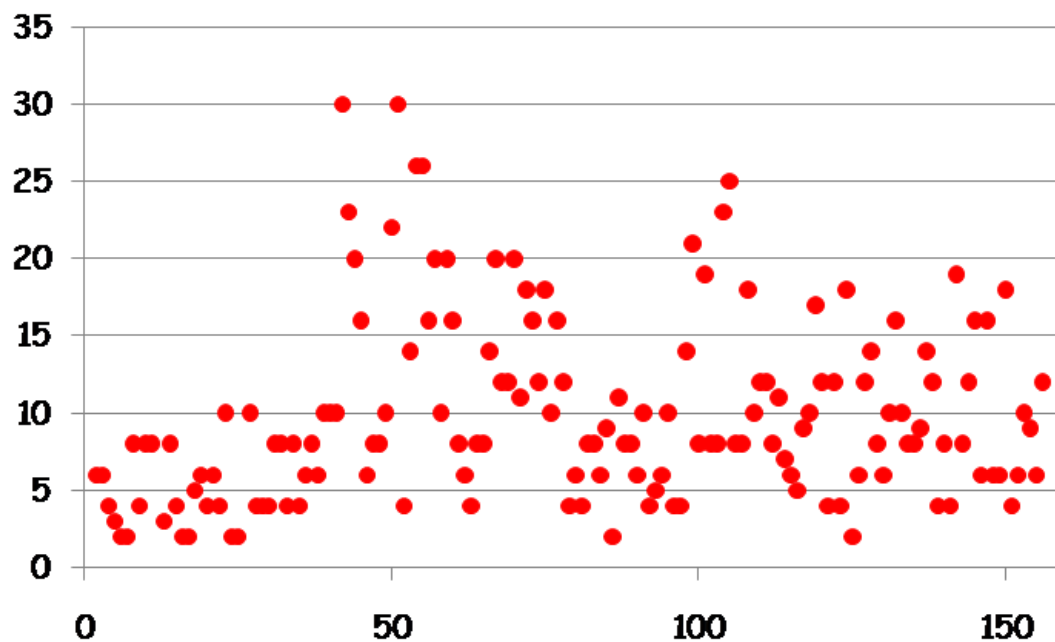
$$S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^k n_i \cdot (x_i - \bar{x})^2$$

$$S^2 = \sum_{i=1}^k w_i \cdot (x_i - \bar{x})^2 = \sum_{i=1}^k w_i \cdot x_i^2 - \left(\sum_{i=1}^k w_i \cdot x_i \right)^2$$

$$\begin{aligned} S^2 &= 1 \cdot 0,3 + 16 \cdot 0,5 + 49 \cdot 0,1 + 100 \cdot 0,1 - 16 \\ &= 0,3 + 8 + 4,9 + 10 - 16 = 7,2 \end{aligned}$$

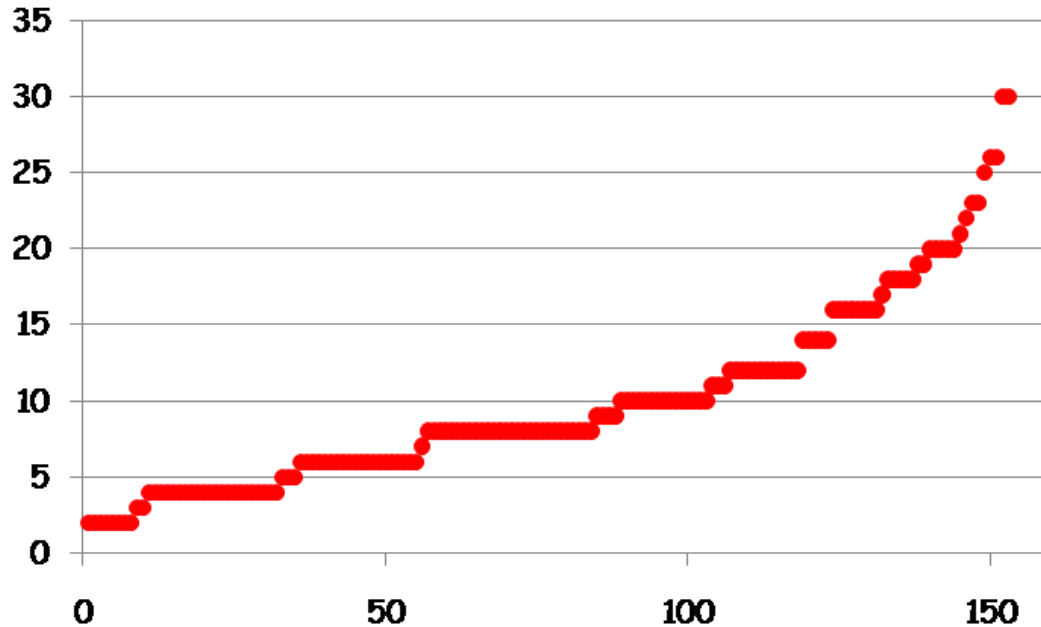
Cecha ilościowa ciągła – wydajność dzienna

4 17 4 14 6 8 6 8 8 10 22
30 26 4 14 6 9 6 8 8 10 20
2 6 4 11 6 9 6 4 8 9 20
26 4 25 11 6 5 6 4 8 9 10
20 2 23 14 6 5 6 4 8 8 5
4 19 4 12 6 4 6 4 8 8 2
23 20 4 12 6 4 4 4 16 10 4
21 4 19 10 4 2 3 2 18 12 4
12 18 12 10 4 2 3 2 18 12 12
8 6 8 16 12 16 12 10 8 6 10
8 6 8 16 11 14 10 10 4 8 6
8 6 8 10 20 16 18 10 30 10 8
8 12 8 12 14 8 18 10 16 8
8 7 8 6 8 10 12 16 8 16



Wykres rozrzutu

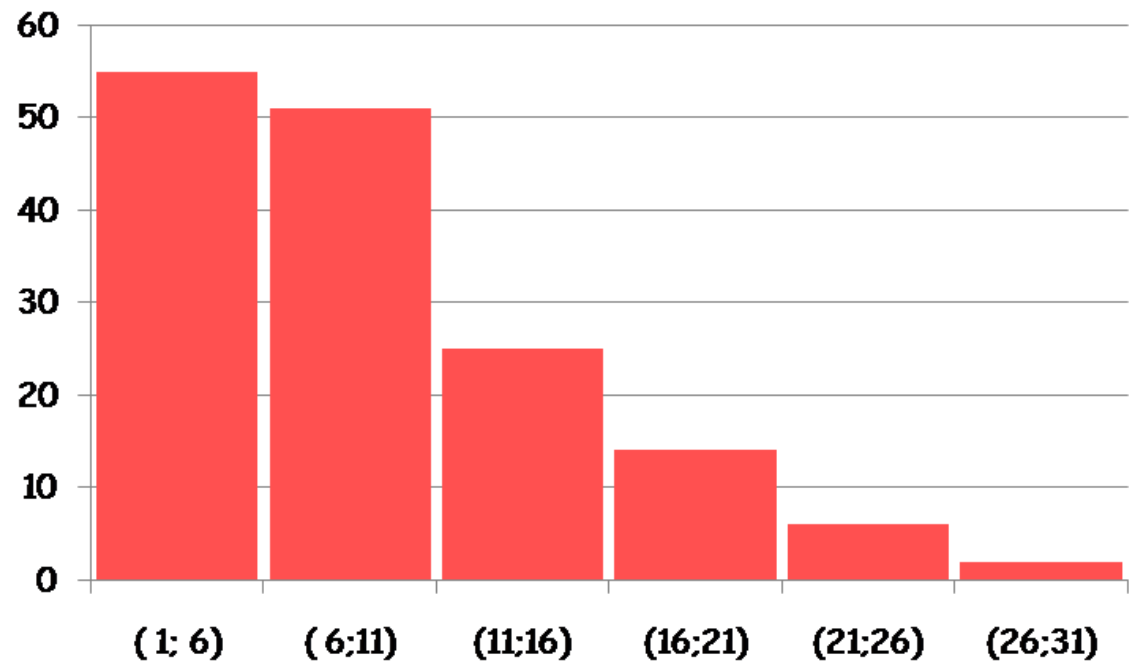
Cecha ilościowa ciągła – wydajność dzienna



Wartości uporządkowane

Cecha ilościowa ciągła – szereg rozdzielczy przedziałowy

$(x_{0i}; x_{1i}]$	n_i	w_i
(1; 6>	54	0,3553
(6;11>	51	0,3355
(11;16>	25	0,1645
(16;21>	14	0,0921
(21;26>	6	0,0395
(26;31>	2	0,0132
Razem	152	1



Histogram

Tworzenie przedziałowego szeregu rozdzielczego

Liczba przedziałów (górną granicą):

$$k = 1 + 3 \times \log N$$

Rozpiętość przedziału:

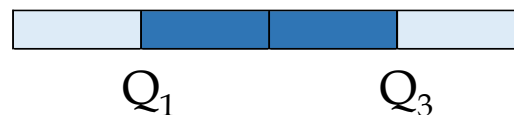
$$(x_{\max} - x_{\min}) / k$$

N	k
20	4,90
40	5,81
60	6,33
80	6,71
100	7,00
120	7,24
140	7,44
160	7,61
180	7,77
200	7,90

Rozpiętość przedziału:

$$2,64 \times Q_x \times N^{-1/3}$$

N	Q _{x=5}	Q _{x=20}
20	4,86	19,45
40	3,86	15,44
60	3,37	13,49
80	3,06	12,25
100	2,84	11,38
120	2,68	10,70
140	2,54	10,17
160	2,43	9,73
180	2,34	9,35
200	2,26	9,03



Średnia arytmetyczna

```
4 17 4 14 6 8 6 8 8 10 22
30 26 4 14 6 9 6 8 8 10 20
2 6 4 11 6 9 6 4 8 9 20
26 4 25 11 6 5 6 4 8 9 10
20 2 23 14 6 5 6 4 8 8 5
4 19 4 12 6 4 6 4 8 8 2
23 20 4 12 6 4 4 4 16 10 4
21 4 19 10 4 2 3 2 18 12 4
12 18 12 10 4 2 3 2 18 12 12
8 6 8 16 12 16 12 10 8 6 10
8 6 8 16 11 14 10 10 4 8 6
8 6 8 10 20 16 18 10 30 10 8
8 12 8 12 14 8 18 10 16 8
8 7 8 6 8 10 12 16 8 16
```

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j = \frac{1}{N} (x_1 + x_2 + \dots + x_N)$$

$$\bar{x} = \frac{1}{152} (4 + 26 + 12 + \dots + 5 + 2) = \frac{1}{152} \cdot 1506 = 9,9079$$

Średnia

$(x_0; x_1)$	n_i
(0;2>	4
(2;4>	11
(4;6>	17
(6;8>	29
(8;10>	46
(10;12>	36
(12;14>	30
(14;16>	16
(16;18>	9
(18;20>	2

$$\begin{aligned}\bar{x} &= \frac{1}{200} (1 \cdot 4 + 3 \cdot 11 + 5 \cdot 17 + 7 \cdot 29 + 9 \cdot 46 + \\ &+ 11 \cdot 36 + 13 \cdot 30 + 15 \cdot 16 + 17 \cdot 9 + 19 \cdot 2) = \\ &= \frac{1}{200} (4 + 33 + 85 + 203 + 414 + 396 + 390 + 240 + 153 + 38) = \\ &= \frac{1}{200} \cdot 1956 = 9,78\end{aligned}$$

Średnia arytmetyczna – przedziałowy szereg rozdzielczy

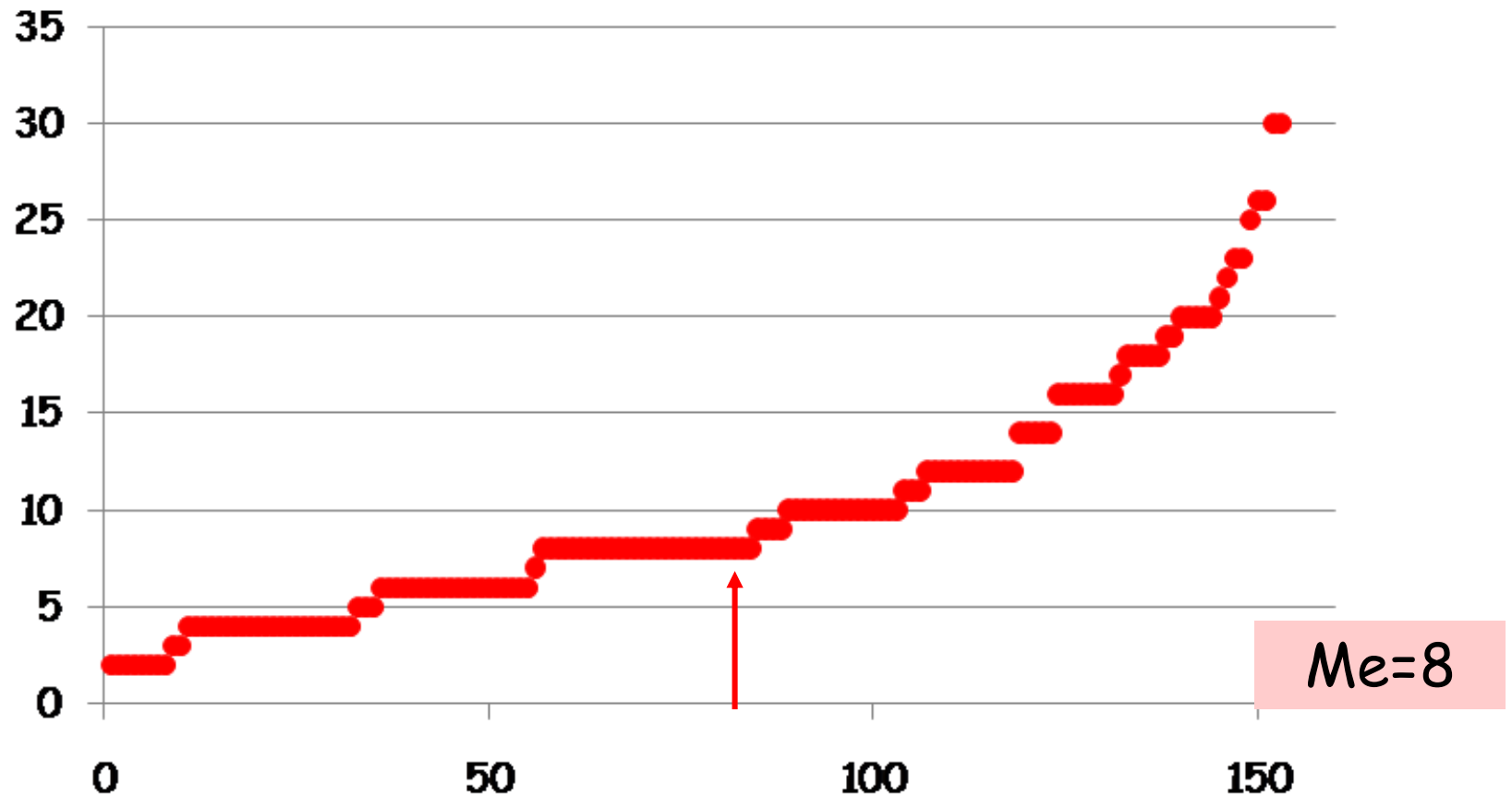
$$\bar{x} = \frac{1}{152} \cdot 1506 = 9,9079$$

$(x_{0i}; x_{1i}]$	x_i	n_i	w_i
(1; 6>	3,5	54	0,3553
(6;11>	8,5	51	0,3355
(11;16>	13,5	25	0,1645
(16;21>	18,5	14	0,0921
(21;26>	23,5	6	0,0395
(26;31>	28,5	2	0,0132
Razem		152	1

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

$$\begin{aligned} \bar{x} &= \frac{1}{152} (3,5 \cdot 54 + 8,5 \cdot 51 + 13,5 \cdot 25 + 18,5 \cdot 14 + 23,5 \cdot 6 + 28,5 \cdot 2) = \\ &= \frac{1}{152} (189 + 433,5 + 337,5 + 259 + 141 + 57) = \frac{1}{152} \cdot 1417 = 9,32237 \end{aligned}$$

Kwantyle w tym mediana – wartość środkowa

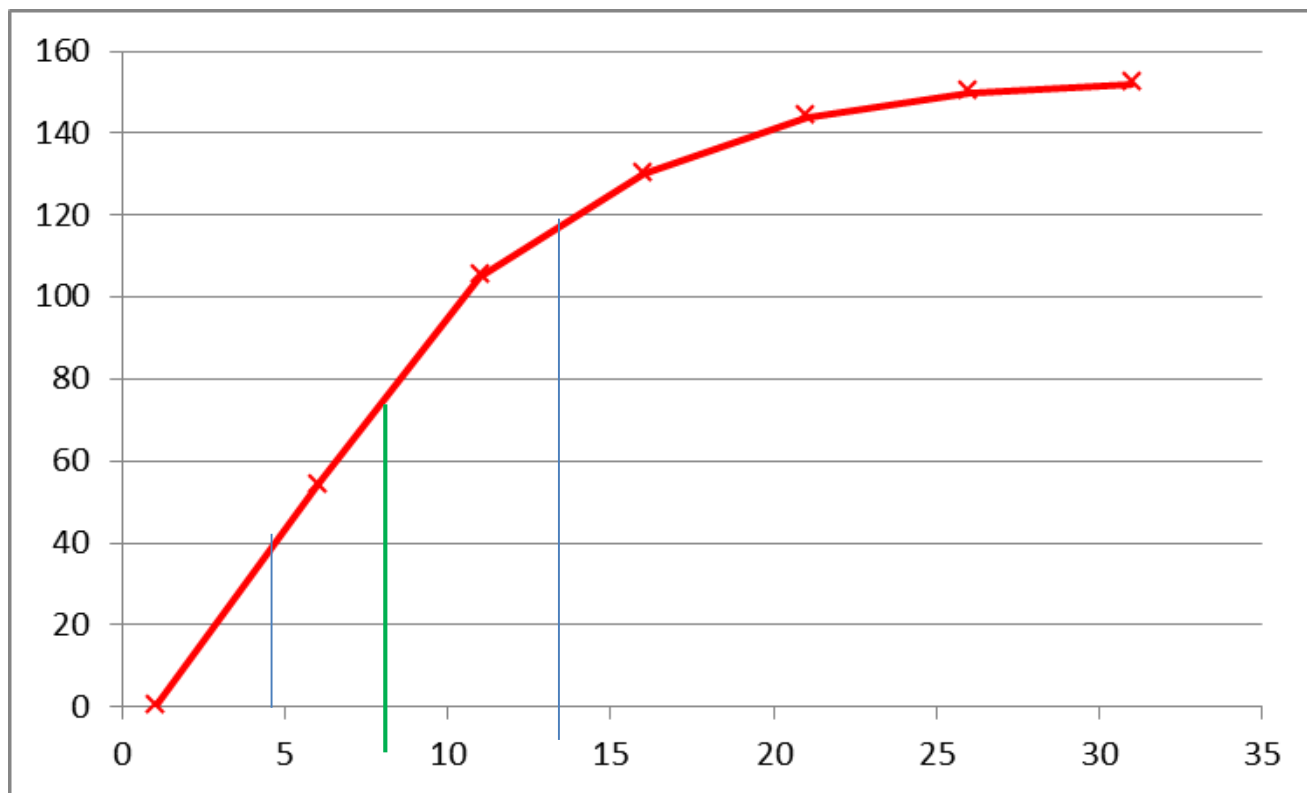


$(x_{0i}; x_{1i}]$	n_i
(1; 6>	54
(6;11>	51
(11;16>	25
(16;21>	14
(21;26>	6
(26;31>	2
Razem	152

Kwantyle

$$Me = x_{Me0} + \frac{\frac{N}{2} - \sum_{i=1}^{Me-1} n_i}{n_{Me}} \cdot h_{Me}$$

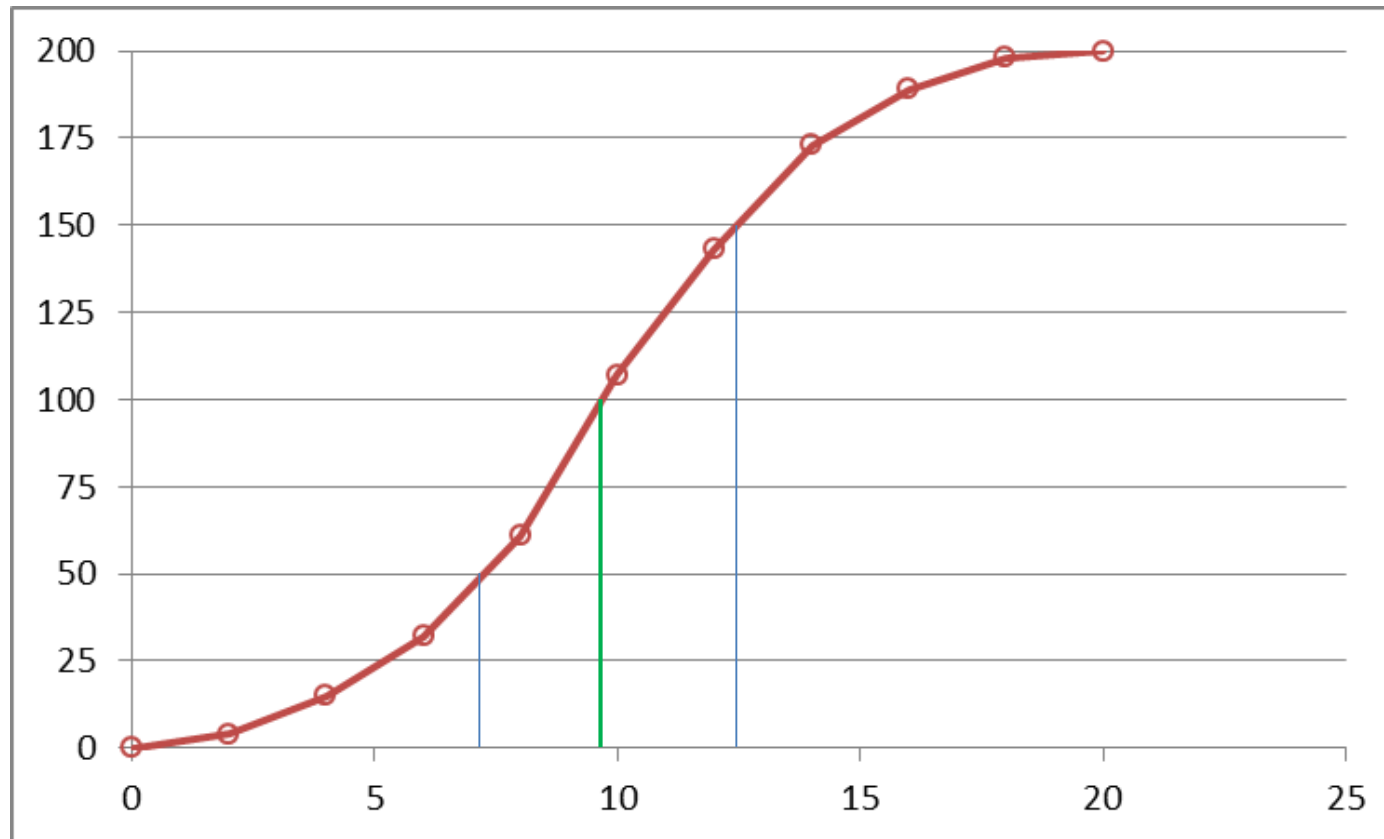
$$Me = 6 + \frac{76 - 54}{51} \cdot 5 = 6 + \frac{22}{51} \cdot 5 = 6 + 2,157 = 8,157$$



Kwantyle

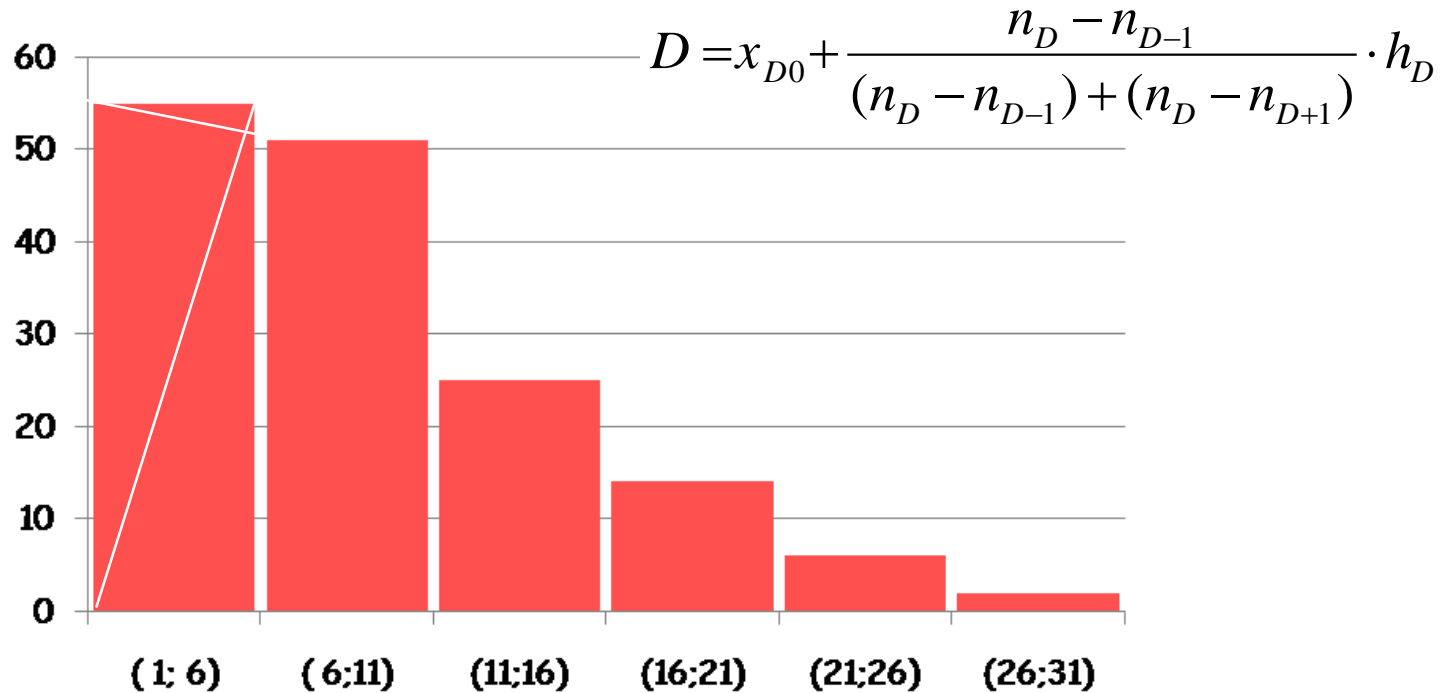
$(x_0; x_1)$	n_i
(0;2>	4
(2;4>	11
(4;6>	17
(6;8>	29
(8;10>	46
(10;12>	36
(12;14>	30
(14;16>	16
(16;18>	9
(18;20>	2

$$Me = 8 + \frac{100 - 61}{46} \cdot 2 = 8 + \frac{39}{46} \cdot 2 = 8 + 1,696 = 9,696$$



Dominanta

$(x_{0i}; x_{1i})$	n_i
(1; 6>	54
(6;11>	51
(11;16>	25
(16;21>	14
(21;26>	6
(26;31>	2
Razem	152



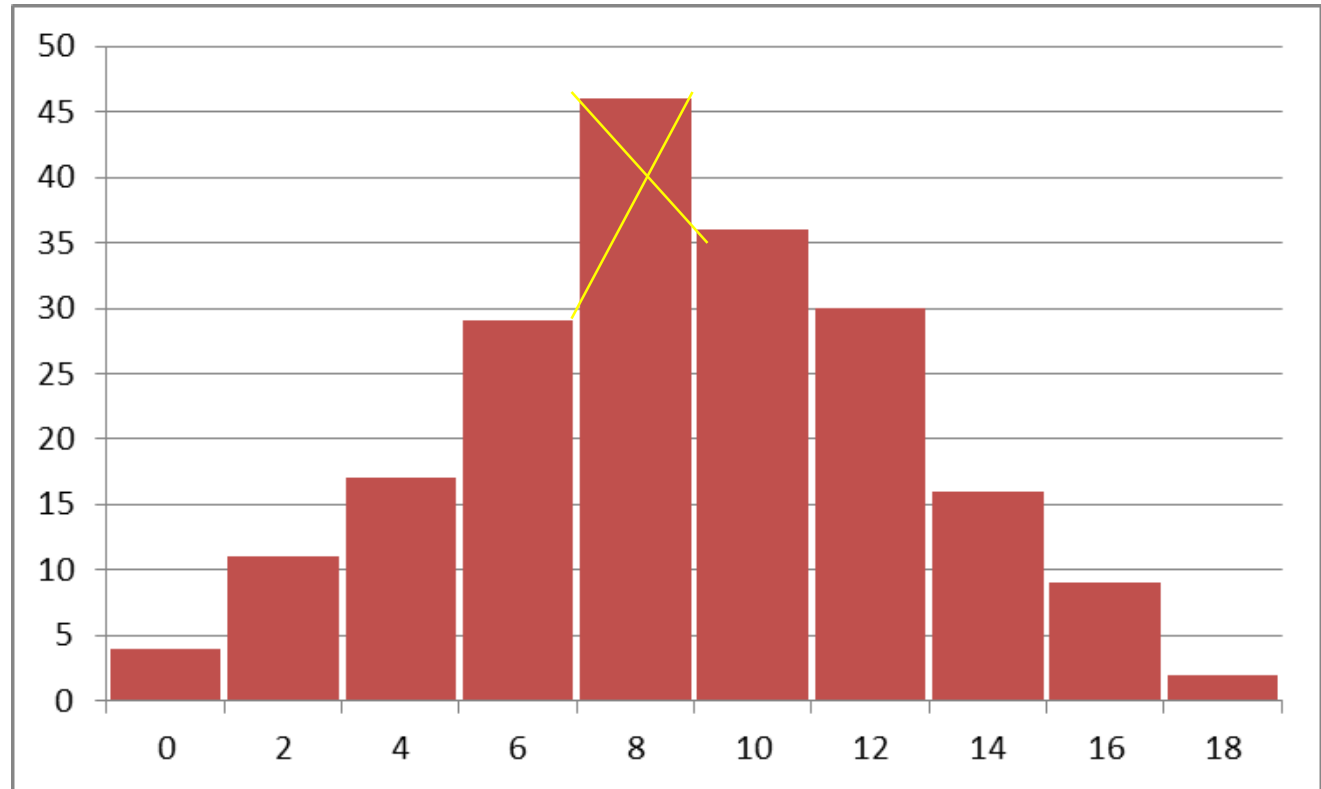
$$D = 1 + \frac{54 - 0}{(54 - 0) + (54 - 51)} \cdot 5 = 1 + \frac{54}{57} \cdot 5 = 1 + 4,737 = 5,737$$

Dominanta

$(x_0; x_1)$	n_i
(0;2>	4
(2;4>	11
(4;6>	17
(6;8>	29
(8;10>	46
(10;12>	36
(12;14>	30
(14;16>	16
(16;18>	9
(18;20>	2

$$D = x_{D0} + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D$$

$$D = 8 + \frac{46 - 29}{(46 - 29) + (46 - 36)} \cdot 2 = 8 + \frac{17}{27} \cdot 2 = 8 + 1,26 = 9,26$$



Wariancja – przedziałowy szereg rozdzielczy

$(x_{0i}; x_{1i})$	\dot{x}_i	n_i	w_i
(1; 6>	3,5	54	0,3553
(6;11>	8,5	51	0,3355
(11;16>	13,5	25	0,1645
(16;21>	18,5	14	0,0921
(21;26>	23,5	6	0,0395
(26;31>	28,5	2	0,0132
Razem		152	1

$$S^2 = \frac{1}{N - 1} \cdot \sum_{i=1}^k n_i \cdot (\dot{x}_i - \bar{x})^2$$

$$\bar{x} = 9,32237$$

$$\begin{aligned}\bar{x} &= \frac{1}{151} \{ (3,5^2 \cdot 54 + 8,5^2 \cdot 51 + 13,5^2 \cdot 25 + 18,5^2 \cdot 14 + 23,5^2 \cdot 6 + 28,5^2 \cdot 2) - \frac{1}{152} \cdot 1417^2 \} = \\ &= \frac{1}{151} \{ (661,5 + 3684,75 + 4556,25 + 4791,5 + 3313,5 + 1624,5) - 13209,796 \} = \\ &= 35,9086\end{aligned}$$

Przy użyciu danych
wyjściowych

$$S^2 = 36,0709$$

$$S = 6,0059$$

Wskaźnik (współczynnik) zmienności

służy do porównań zróżnicowania w zbiorowościach

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

$$V = \frac{6,0059}{9,9079} = 0,6062$$

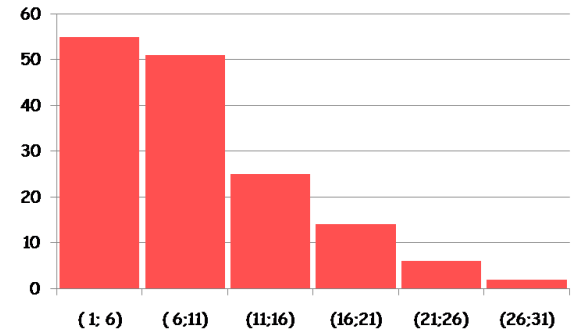
Współczynnik asymetrii (skośności)

ocena asymetrię rozkładu próby

$$A = \frac{\frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^3}{S^3}$$

$$A = 1,127$$

$$A'' = \frac{(\bar{x} - D)}{S}$$



$$A'' = 0,6945$$

Przedział typowej zmienności

$$(\bar{x} - S; \bar{x} + S)$$

Czy liczba 10 jest typowa?

średnia	7,250	9,908
odch.st.	1,446	6,006

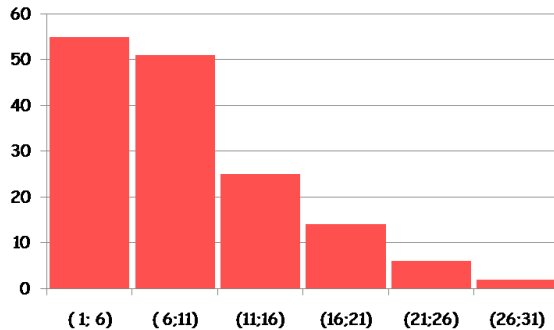
$$(7,25 - 1,446; 7,25 + 1,446) = (5,804; 8,696)$$

$$(9,908 - 6,006; 9,908 + 6,006) = (3,902; 15,914)$$

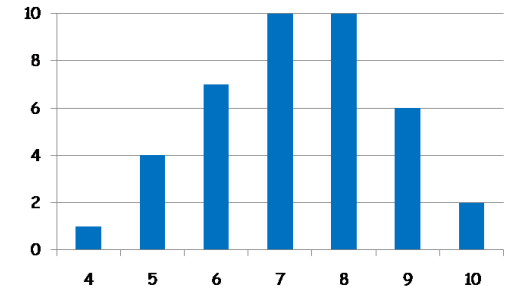
Współczynnik kurtozy (kształtu)

ocenia kształt wykresu rozkładu próby

$$W = \frac{\frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^4}{s^4} - 3$$



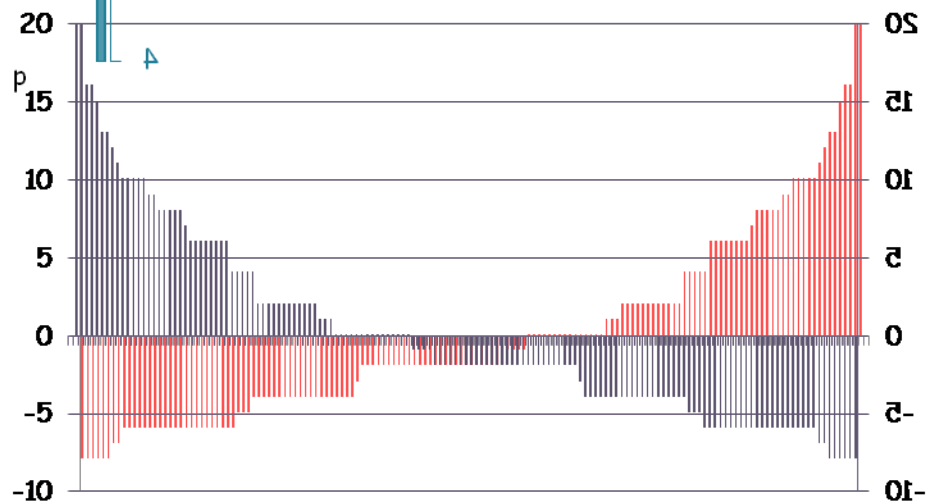
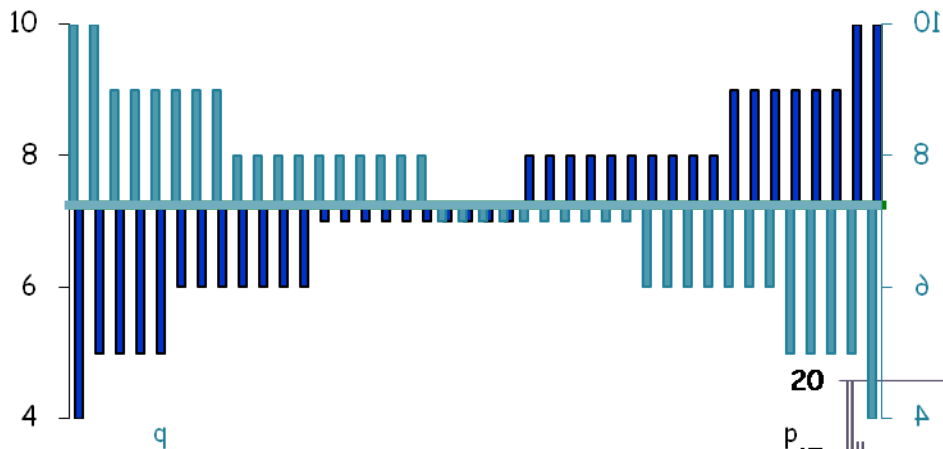
$$W = 1,044$$



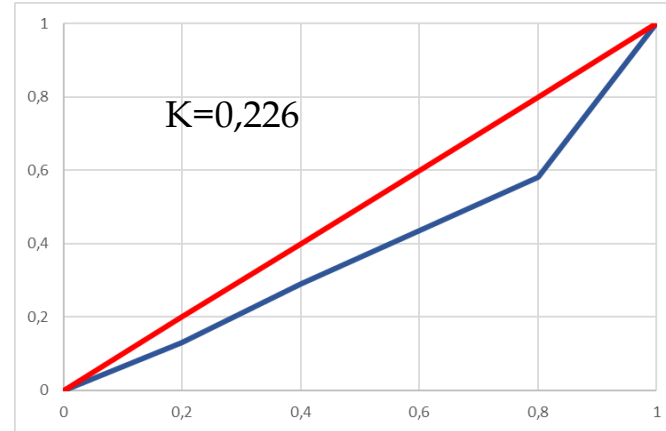
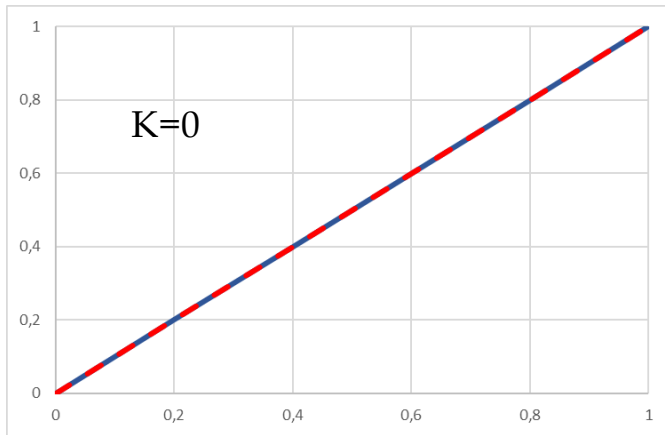
$$A = -0,471$$

Porównania prób

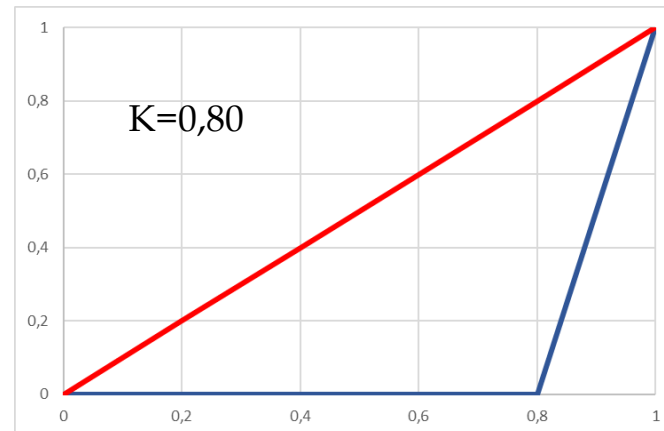
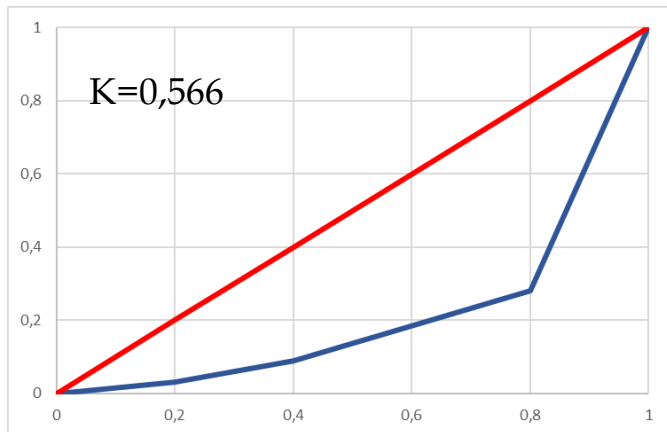
średnia	7,250	9,908
odch.st.	1,446	6,006
V	0,199	0,606
A	-0,130	1,127
W	-0,471	1,044



Współczynnik koncentracji

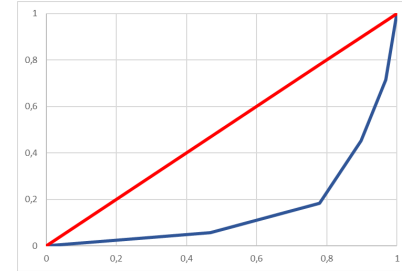


Współczynnik koncentracji (K) to proporcja pola pomiędzy linią równomiernego podziału (czerwona) i krzywą koncentracji Lorenza (niebieska) do pola trójkąta $(0,5)$, czyli $0 \leq K \leq 1$



Współczynnik koncentracji

grupa dochodu	liczba osób	wartość zarobków	częstość empiryczna		kumulowana częstość	
			osób	zarobków	osób	zarobków
do 5 tys.	234	800	0,468	0,057	0,468	0,057
od 5 do 20 tys.	156	1800	0,312	0,128	0,780	0,184
20 do 100 tys	59	3800	0,118	0,270	0,898	0,454
100 do 200 tys.	35	3700	0,070	0,262	0,968	0,716
powyżej 200 tys.	16	4000	0,032	0,284	1	1
	500	14100	1	1		



Pole trójkąta:
 $0,5 \cdot 0,468 \cdot 0,057 = 0,0133$

Pole trapezu 1.
 $0,5 \cdot (0,057 + 0,184) \cdot 0,312 = 0,0376$

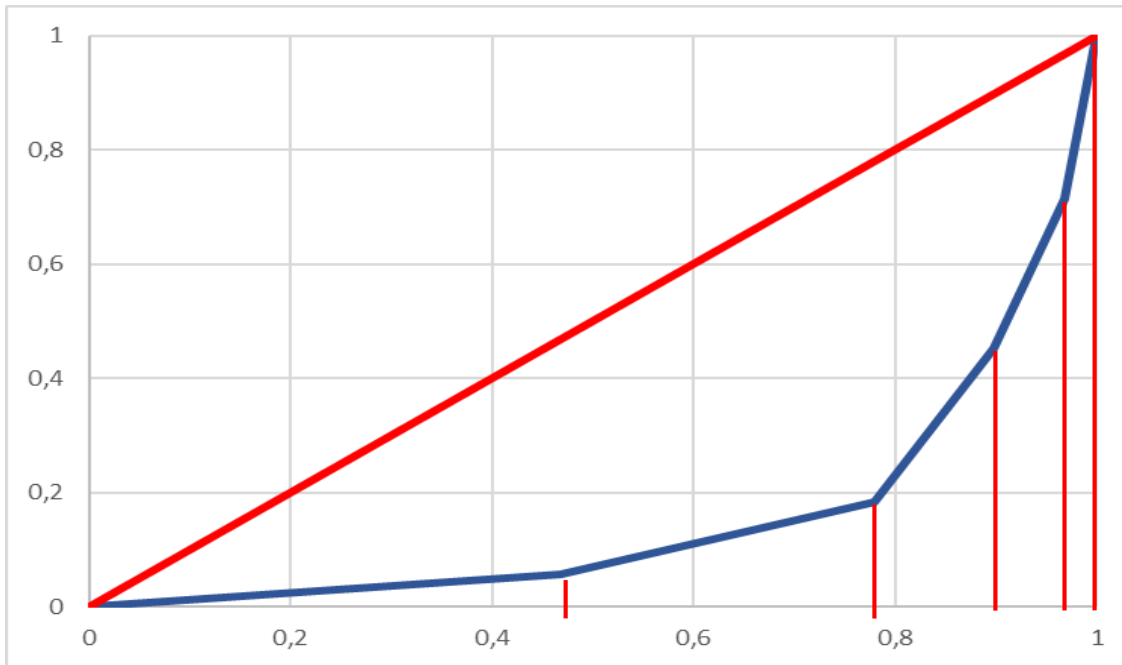
Pole trapezu 2.
 $0,5 \cdot (0,184 + 0,454) \cdot 0,118 = 0,0377$

Pole trapezu 3.
 $0,5 \cdot (0,454 + 0,716) \cdot 0,070 = 0,0410$

Pole trapezu 4.
 $0,5 \cdot (0,716 + 1) \cdot 0,032 = 0,0275$

$$K = 1 - \frac{0,0133 + 0,0376 + 0,0377 + 0,0410 + 0,0275}{0,5}$$

$$= 1 - \frac{0,157}{0,5} = 1 - 0,314 = 0,684$$



Przykład

Uzeregować według poziomu cechy i wielkości zmienności trzy próby o podanych parametrach

parametr	Próba I	Próba II	Próba III
średnia	4,1	3,9	4,5
odch.st.	0,25	0,46	0,51
V	6,1%	11,8%	11,3



Aquila pomarina