

CHARAKTERYSTYKI LICZBOWE STRUKTURY ZBIOROWOŚCI (dok.)

1. miary położenia - *wykład 2*
2. miary zmienności (dyspersji, rozproszenia) - *wykład 3*
3. miary **asymetrii** (skośności)
4. miary **koncentracji**

MIARY ASYMETRII

Miary asymetrii charakteryzują rodzaj i stopień odstępstwa od symetrii rozkładu badanej cechy.

Miary asymetrii dzielą się podobnie jak poprzednie na miary klasyczne i pozycyjne.

1. miary klasyczne (współczynnik skośności (A_s lub A_d), współczynnik asymetrii (A)) oraz
2. miary pozycyjne (współczynnik skośności (A_Q)).

Najprostszą miarą asymetrii jest wskaźnik skośności (W_s lub W_Q). Dla miar klasycznych jest to różnica pomiędzy średnią arytmetyczną i modalną.

$$W_s = \bar{x} - M_o$$

Dla miar pozycyjnych badamy odległości obu kwartyli od mediany.

$$W_Q = (Q_{III} - M_e) - (M_e - Q_I) = Q_I + Q_{III} - 2 \times M_e$$

Jeżeli **rozkład** badanej cechy jest **symetryczny**,
to średnia jest równa modalnej,
a wskaźnik skośności jest równy zero.

$$W_s = \bar{x} - M_o = 0$$

Rozkłady badanych cech różnią się między sobą
kierunkiem i siłą asymetrii.

Jeżeli rozkład badanej cechy nie jest symetryczny, to mamy do czynienia z asymetrią rozkładu. Mówimy o dwóch rodzajach (kierunkach) asymetrii: lewo- i prawostronnej.

Dla miar klasycznych będzie to:

- asymetria lewostronna gdy

$$W_s = \bar{x} - M_o < 0 \quad \text{oraz}$$

- asymetria prawostronna gdy

$$W_s = \bar{x} - M_o > 0$$

Dla miar pozycyjnych będzie to:

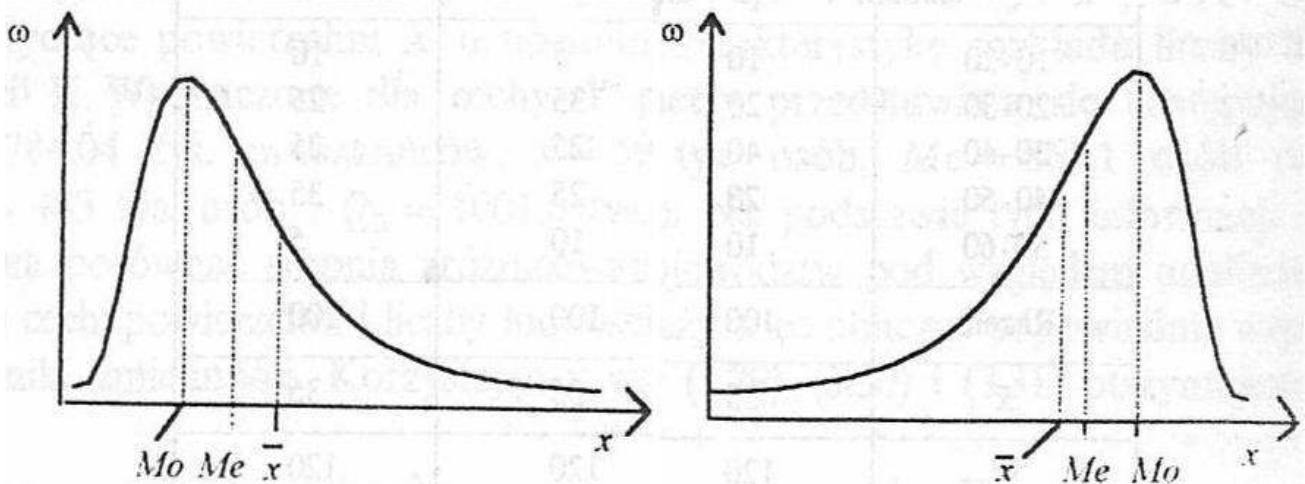
- asymetria lewostronna gdy

$$W_Q = (Q_{III} - M_e) - (M_e - Q_I) < 0 \quad \text{oraz}$$

- asymetria prawostronna gdy

$$W_Q = (Q_{III} - M_e) - (M_e - Q_I) > 0 .$$

Poniższe rysunki ilustrują rodzaje asymetrii i wzajemne relacje pomiędzy podstawowymi miarami położenia.



Dla porównania kierunku i siły asymetrii w dwóch lub więcej zbiorowościach stosujemy **współczynniki skośności**.

$$A_s = \frac{\bar{x} - M_o}{s}$$

dla miar klasycznych

$$A_Q = \frac{Q_I + Q_{III} - 2 \times M_e}{2Q}$$

dla miar pozycyjnych

Do klasycznych miar asymetrii należy również współczynnik asymetrii (A). **Uwaga!!! Jest on pracochłonny w liczeniu.**

$$A = \frac{m_3}{s^3}$$

gdzie: s – odchylenie standardowe

Licznik powyższego ułamka (m_3) wyliczamy odmiennie dla każdego sposobu pogrupowania materiału statystycznego. I tak:

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

- szereg szczegółowy

$$m_3 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^3 n_i$$

- szereg rozdzielczy punktowy

$$m_3 = \frac{1}{n} \sum_{i=1}^k (\dot{x}_i - \bar{x})^3 n_i$$

- szereg rozdzielczy przedziałowy

PRZYKŁAD 1 (Przykład 7 z wykładu 3 – praca domowa)**Płace (stawka godzinowa) w firmach A, B i C**

<i>klasa</i>	Stawka [zł/godz.]		liczba pracowników (n_i)		
i	x_{0i}	x_{1i}	firma A	firma B	firma C
1	2	4	15	15	20
2	4	6	30	105	50
3	6	8	60	75	50
4	8	10	30	75	70
5	10	12	15	30	10
×	<i>razem</i>				
<i>średnia</i>					
<i>wariancja</i>					
<i>odchylenie standardowe</i>					
<i>modalna</i>					
<i>kwartyl I</i>					
<i>kwartyl II (mediana)</i>					
<i>kwartyl III</i>					
<i>odchylenie ćwiartkowe</i>					
<i>wskaźnik skośności (klas.)</i>					
<i>wskaźnik skośności (pozyc.)</i>					
<i>współcz. skośności (klas.)</i>					
<i>współcz. skośności (pozyc.)</i>					
<i>współcz. asymetrii (A)</i>					
<i>(licznik A, tj. m_3)</i>					

PRZYKŁAD 1a (przykładowe *obliczenia dla firmy C*)

$$W_s = \bar{x} - M_o = 7 - 8,5 = -1,5$$

$$W_Q = Q_I + Q_{III} - 2 \times M_e = 5,2 + 8,86 - 2 \times 7,2 = -0,34$$

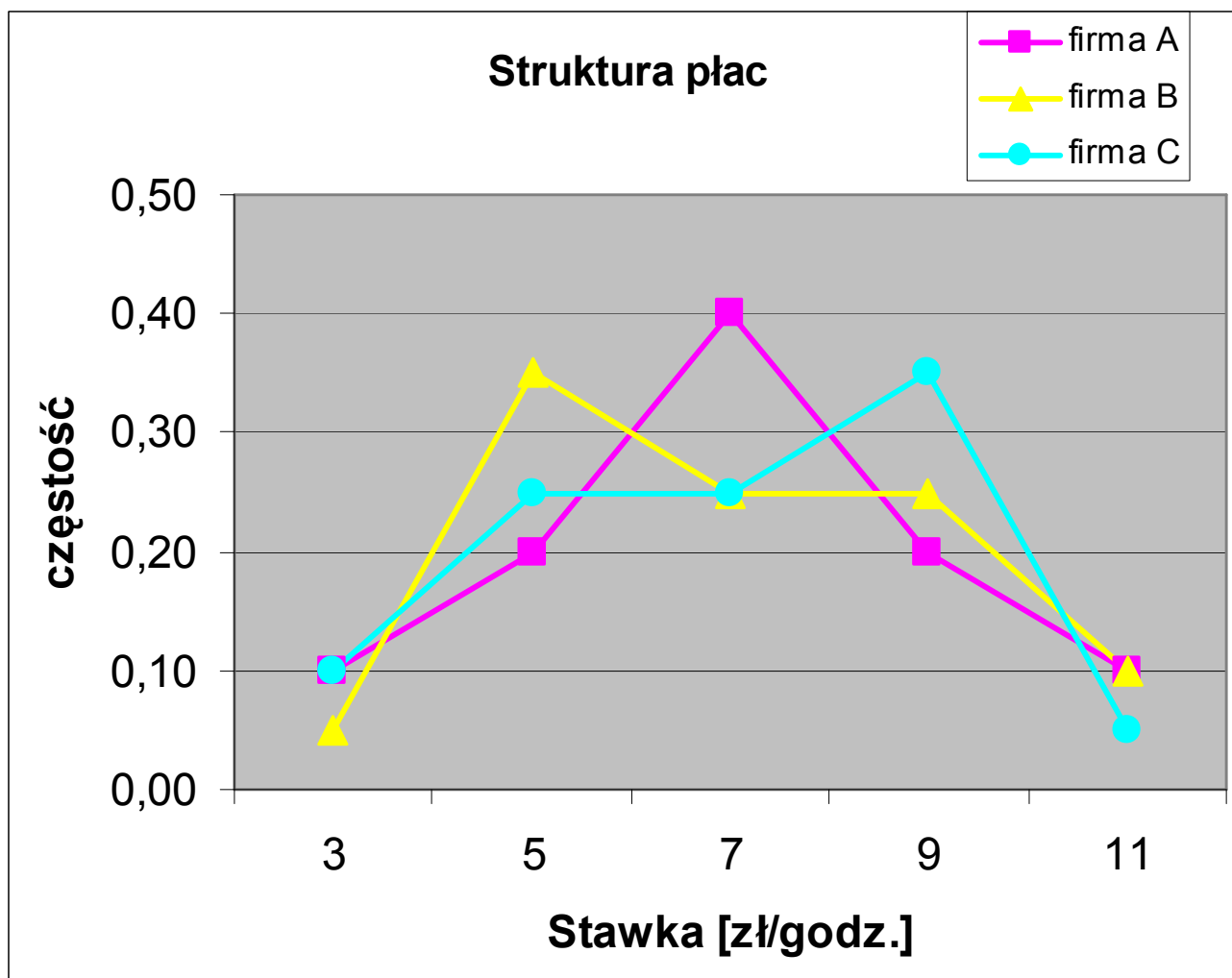
$$A_s = \frac{\bar{x} - M_o}{s} = \frac{-1,5}{2,19} = -0,68$$

$$A_Q = \frac{Q_I + Q_{III} - 2 \times M_e}{2Q} = \frac{-0,34}{2 \times 1,83} = -0,09$$

Obliczanie współczynnika asymetrii (A)

<i>klasa</i>	Stawka [zł/godz.]		<i>środek klasy</i>	<i>obliczanie m_3 we współczynniku asymetrii (firma C)</i>			
	x_{0i}	x_{1i}		n_i	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^3$	$(\dot{x}_i - \bar{x})^3 n_i$
1	2	4	3	20			
2	4	6	5	50			
3	6	8	7	50			
4	8	10	9	70	2	8	560
5	10	12	11	10	4	64	640
×	<i>razem</i>		×		×	×	

$$A = \frac{m_3}{s^3} = \frac{-480 / 200}{(2,19)^3} = -0,23$$



MIARY KONCENTRACJI

Trzy dotychczas omówione grupy miar (tj. miary położenia, rozproszenia i asymetrii) w sposób wyczerpujący opisują strukturę badanej zbiorowości.

Uzupełnieniem tego opisu są miary koncentracji. Istnieje bowiem ścisły związek pomiędzy koncentracją a rozproszeniem: im mniejsze rozproszenie tym większa koncentracja. I na odwrót.

Zjawisko koncentracji może być rozważane jako nierównomierny podział ogólnej sumy wartości cechy pomiędzy poszczególne jednostki badanej zbiorowości.

Do oceny stopnia koncentracji stosujemy dwie metody.

1. Metoda numeryczna – wyznaczanie odpowiednich wskaźników liczbowych (współczynnik skupienia inaczej kurtoza, współczynnik koncentracji Lorenza).
2. Metoda graficzna – wykreślanie i analiza tzw. krzywej koncentracji Lorenza.

Współczynnik skupienia (kurtoza)

Kurtoza (K) należy do klasycznych miar koncentracji.

Uwaga!!! Jest ona pracochłonna w liczeniu.

$$K = \frac{m_4}{s^4} \quad \text{gdzie: } s - \text{odchylenie standardowe}$$

Licznik powyższego ułamka (m_4) wyliczamy odmiennie dla każdego sposobu pogrupowania materiału statystycznego. I tak:

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \quad - \text{szereg szczegółowy}$$

$$m_4 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^4 n_i \quad - \text{szereg rozdzielczy punktowy}$$

$$m_4 = \frac{1}{n} \sum_{i=1}^k (\dot{x}_i - \bar{x})^4 n_i \quad - \text{szereg rozdzielczy przedziałowy}$$

Im większa wartość kurtozy (K), tym większa koncentracja (diagram wyższy i smuklejszy).

Zjawiska społeczne, gospodarcze, przyrodnicze ... są najczęściej opisywane tzw. rozkładem normalnym (przykłady diagramów takiego rozkładu pokazano w wykładzie 3 na stronach 3 i 4).

Kurtoza w rozkładzie normalnym jest zawsze równa trzy ($K=3$).

W praktyce policzoną kurtozę porównujemy z kurtozą rozkładu normalnego. I tak jeżeli:

- $K > 3$ - rozkład badanej cechy jest wyższy i smuklejszy od rozkładu normalnego
- $K < 3$ - odwrotnie; niższy i bardziej rozłożysty

PRZYKŁAD 2 (dane z przykładu 1 – firma A; *w domu policz dla pozostałych firm*)

Płace (stawka godzinowa) w firmie A

<i>klasa</i>	Stawka [zł/godz.]		<i>środek klasy</i>	<i>obliczanie m_4 w kurtozie (firma A)</i>			
<i>i</i>	x_{0i}	x_{1i}	\dot{x}_i	n_i	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^4$	$(\dot{x}_i - \bar{x})^4 n_i$
1	2	4	3	15			
2	4	6	5	30			
3	6	8	7	60			
4	8	10	9	30			
5	10	12	11	15			
×	<i>razem</i>		×		×	×	

$$K = \frac{m_4}{s^4} = \frac{8640/150}{(2,19)^4} = 2,5$$

WNIOSEK

$K < 3$ - koncentracja wokół średniej stawki godzinowej w firmie A jest mniejsza niż w przypadku rozkładu normalnego (diagram jest niższy i bardziej rozłożysty niż w rozkładzie normalnym); rozproszenie jest większe niż w rozkładzie normalnym.

Krzywa koncentracji Lorenza

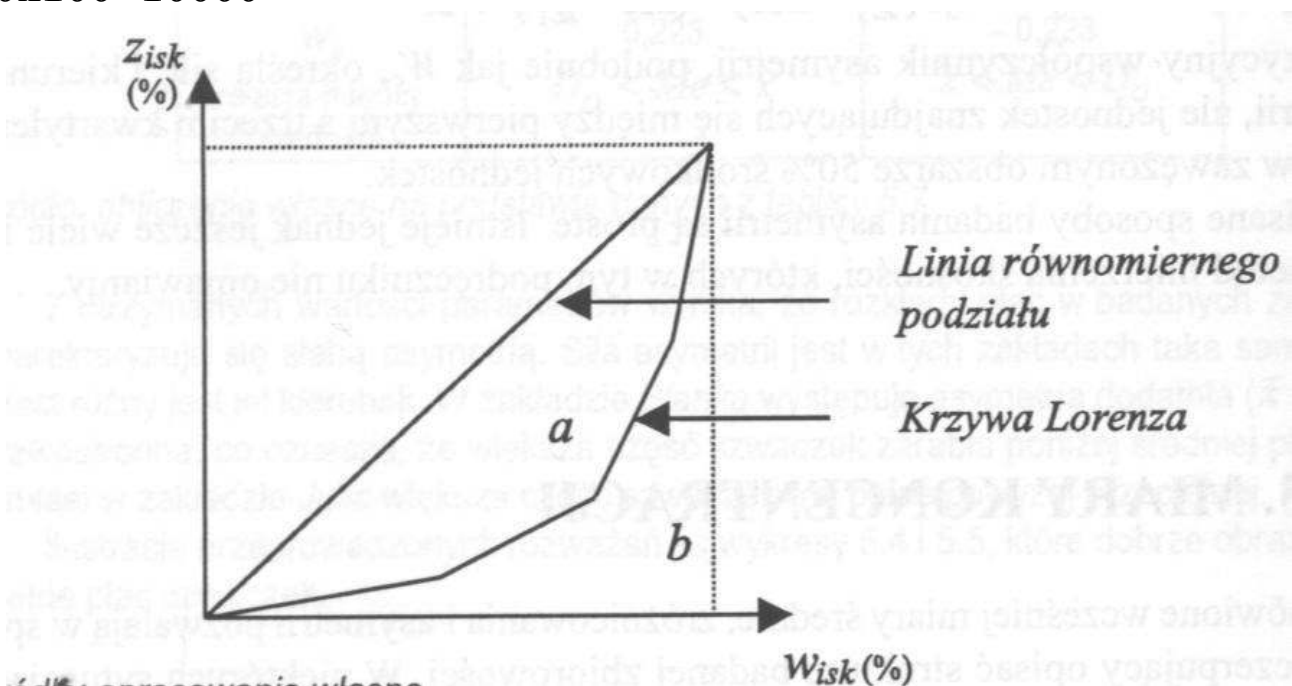
Dane pogrupowane są w szereg rozdzielczy przedziałowy.

Krzywą koncentracji Lorenza rysujemy wykorzystując:

- skumulowaną częstość dla liczebności ($W_{i sk}$) oraz
- skumulowaną częstość dla wartości cechy ($Z_{i sk}$);
wartość cechy obliczamy w każdej klasie jako iloczyn $n_i Z_i$ (tak jak przy liczeniu średniej)

Obie częstości wyrażamy w % .

Kwadrat w którym rysujemy krzywą Lorenza ma powierzchnię $100 \times 100 = 10000$



Krzywą Lorenza otrzymujemy nanosząc na powyższym wykresie dla każdej klasy punkt o współrzędnych ($W_{i sk}, Z_{i sk}$).

Następnie łączymy te punkty odcinkami. Punkt ($W_{1 sk}, Z_{1 sk}$) łączymy dodatkowo z punktem (0, 0).

Im większa jest powierzchnia pola (a), tym większa jest koncentracja w badanym zjawisku.

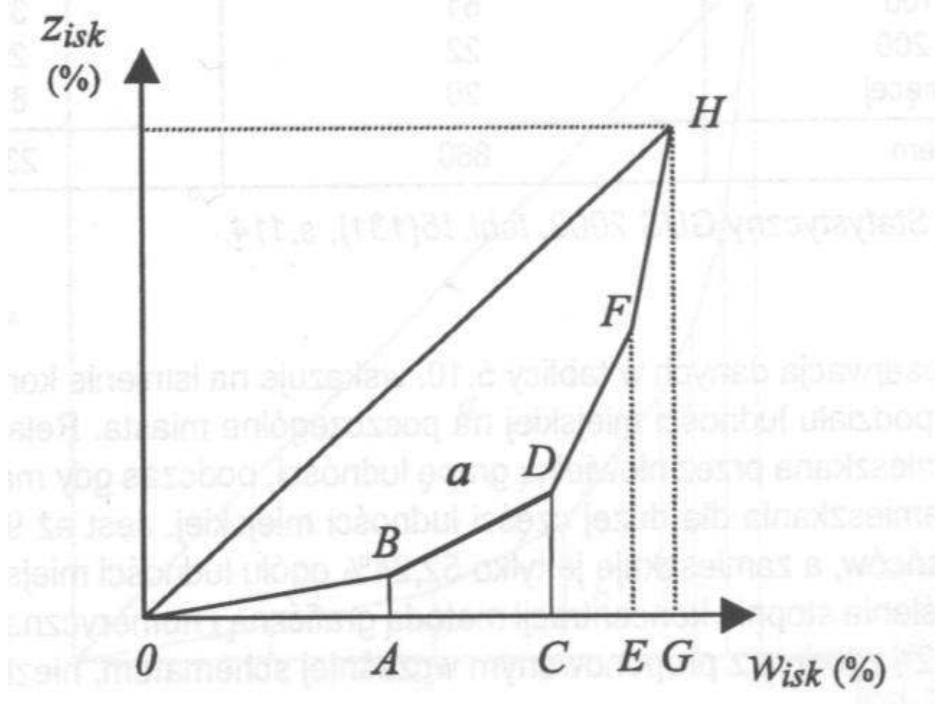
Współczynnik koncentracji Lorenza

Aby liczbowo wyrazić wielkość koncentracji wyliczamy tzw. współczynnik koncentracji Lorenza (KL). Jest on równy stosunkowi pola (a) do pola powierzchni połowy kwadratu (5000):

$$KL = \frac{a}{5000}$$

Ponieważ łatwiej jest policzyć pole (b), to pole (a) wyznaczamy z różnicy $a=5000-b$.

Pole (b) jest sumą pól trapezów prostokątnych (dla pierwszej klasy jest to trójkąt prostokątny).



Ostateczny wzór na współczynnik koncentracji Lorenza (KL) ma postać:

$$KL = \frac{5000 - b}{5000} = 1 - \frac{b}{5000}$$

$KL \rightarrow 1$ oznacza silną koncentrację

$KL \rightarrow 0$ oznacza słabą koncentrację

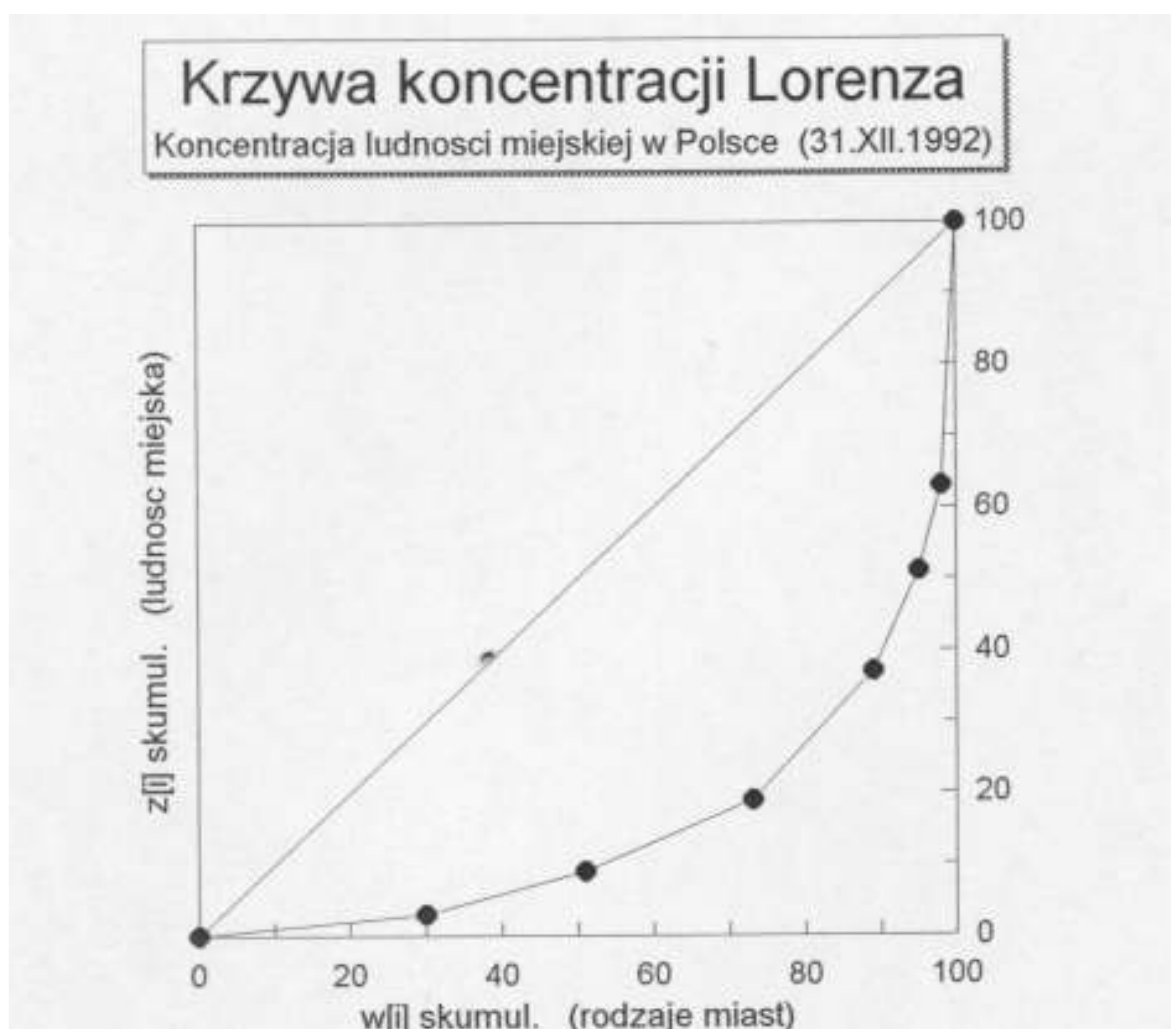
PRZYKŁAD 3 (Miasta i ludność w miastach – stan na 31.12.1992)

Grupy miast wg liczby ludności (w tys.)	Liczba miast	Ludność w miastach (w tys.)
x_i	n_i	$x_i n_i$
<i>poniżej 5</i>	253	788
5 – 10	176	1239
10 – 20	178	2544
20 – 50	136	4140
50 – 100	50	3390
100 – 200	22	2849
200 i więcej	20	8751
razem	835	23701

Średnie miasto $\bar{x} = \frac{23701}{835} = 28,4$ tys. mieszkańców.

Grupy miast wg liczby ludności (w tys.)	odsetek miast (%)	odsetek ludności w miastach (%)
x_i	w_i	z_i
<i>poniżej 5</i>	30,3	3,3
5 – 10	21,1	5,2
10 – 20	21,3	10,7
20 – 50	16,3	17,5
50 – 100	6,0	14,3
100 – 200	2,6	12,0
200 i więcej	2,4	37,0
razem	100,0	100,0

Grupy miast wg liczby ludności (w tys.)	skumulowany odsetek miast (%)	skumulowany odsetek ludności w miastach (%)
x_i	$w_{i\ sk}$	$z_{i\ sk}$
<i>poniżej 5</i>	30,3	3,3
5 – 10	51,4	8,5
10 – 20	72,7	19,2
20 – 50	89,0	36,7
50 – 100	95,0	51,0
100 – 200	97,6	63,0
200 i więcej	100,0	100,0
razem	×	×



Na zakończenie policzymy współczynnik koncentracji Lorenza.

Grupy miast wg liczby ludności (w tys.)	odsetek miast (%)	skumulowany odsetek ludności w miastach (%)	<u>obliczanie pola (b)</u> suma pól trójkąta i trapezów	rodzaj figury
x_i	w_i	$z_{i\ sk}$	$\frac{w_i(z_{i\ sk} + z_{i-1\ sk})}{2}$	
<i>poniżej 5</i>	30,3	3,3		<i>trójkąt</i>
5 – 10	21,1	8,5		<i>trapez</i>
10 – 20	21,3	19,2		<i>trapez</i>
20 – 50	16,3	36,7		<i>trapez</i>
50 – 100	6,0	51,0		<i>trapez</i>
100 – 200	2,6	63,0		<i>trapez</i>
200 i więcej	2,4	100,0		<i>trapez</i>
razem	100,0	×		×

Pole (b) wynosi 1532,0.

Współczynnik koncentracji Lorenza wynosi:

$$KL = 1 - \frac{b}{5000} = 1 - \frac{1532}{5000} = 0,694$$

WNIOSEK:

W grudniu 1992 ludność Polski zamieszkująca miasta miała tendencję do koncentrowania się w miastach o średniej wielkości 28,4 tys. mieszkańców.

Potwierdzają to:

- duża wartość współczynnika koncentracji KL oraz
- wyraźny „brzuch” krzywej koncentracji Lorenza.