

STATYSTYKA OPISOWA

Dr Alina Gleska

Instytut Matematyki WE PP

28 września 2018

1

Pojęcie koncentracji może być stosowane w dwóch różnych znaczeniach:

- 1) koncentracja jako skupienie poszczególnych wartości cechy wokół średniej arytmetycznej, tzw. kurtoza; im większe zróżnicowanie zbiorowości, tym mniejsza koncentracja i odwrotnie;
- 2) koncentracja jako zjawisko nierównomiernego podziału ogólnej sumy wartości cechy pomiędzy poszczególne jednostki zbiorowości (np. gdy w nieznacznej części gospodarstw domowych skupiona jest duża część dóbr luksusowych).

Miara kurtozy dotyczy tylko rozkładów symetrycznych albo zbliżonych do symetrycznych. Ocena kurtozy odbywa się przez porównanie spłaszczenia badanego rozkładu ze spłaszczeniem rozkładu normalnego, uznawanym za wzorcowy. Wyróżniamy dwa typy rozkładów:

- rozkłady leptokurtyczne - bardziej wysmukłe niż rozkład normalny,
- rozkłady platokurtyczne - bardziej spłaszczone niż rozkład normalny.

Do pomiaru kurtozy wykorzystujemy dwie miary klasyczne, zwane miarami spłaszczenia: współczynnik kurtozy i współczynnik ekscesu.

WSPÓŁCZYNNIK KURTOZY (inaczej: koncentracji lub spłaszczenia) wyraża się wzorem:

$$K = \frac{m_4}{s^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4}.$$

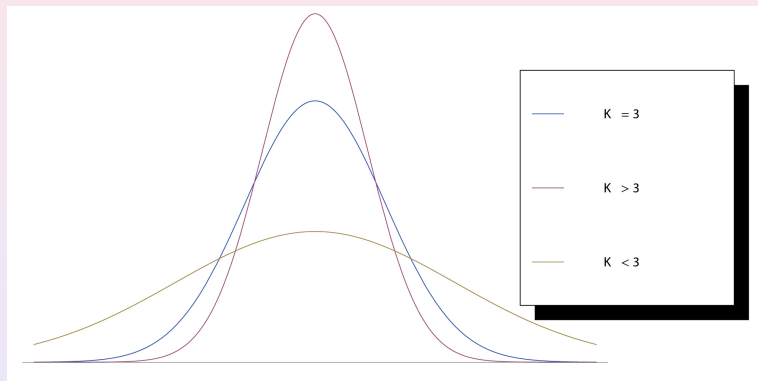
Współczynnik ten jest miarą względną (wielkością niemianowaną), może więc służyć do porównania koncentracji różnych rozkładów. Im wyższa jest wartość współczynnika kurtozy K , tym bardziej wysmukły jest rozkład, a więc silniejsza koncentracja wartości cechy wokół średniej.

Wartości współczynnika kurtozy K :

- $K = 3$ - rozkład ma taką koncentrację, jak rozkład normalny,
- $K > 3$ - rozkład leptokurtyczny,
- $K < 3$ - rozkład platokurtyczny.

WSPÓŁCZYNNIK EKSCESU $K' = K - 3$, gdzie K to współczynnik kurtozy. Zatem:

- $K' = 0$ - dla rozkładów o takiej koncentracji, jak rozkład normalny,
- $K' > 0$ - dla rozkładów leptokurtycznych,
- $K' < 0$ - dla rozkładów platokurtycznych.



Ocena koncentracji jako nierównomierności podziału
Badanie koncentracji w sensie nierównomierności podziału
znajduje zastosowanie m.in. przy badaniu dochodów, płac,
gęstości zaludnienia. Powiemy np. o koncentracji zaludnienia,
gdy w nieznacznej części miast skupiona jest duża część
ogólnej liczby ludności.

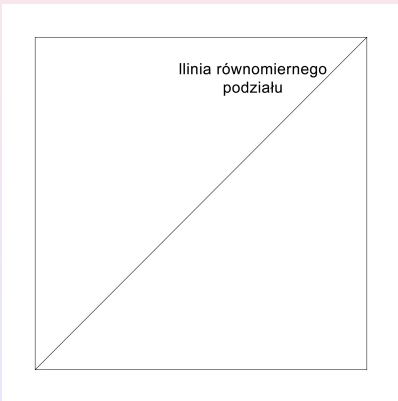
W skrajnych sytuacjach możemy mieć do czynienia z:

- koncentracją całkowitą - gdy jedna jednostka skupia całą globalną sumę wartości cechy (np. jedno miasto skupia całą ludność) - jest to tzw. **rozkład elitarny** ("jeden bierze wszystko"),
- brakiem koncentracji (całkowitym rozdrobnieniem) - gdy na każdą jednostkę przypada taka sama część globalnej sumy wartości cechy (np. w każdym mieście mieszka taki sam odsetek ludności) - jest to tzw. **rozkład egalitarny** ("wszyscy dostają tyle samo").

Koncentrację w znaczeniu nierównomierności podziału można ocenić:

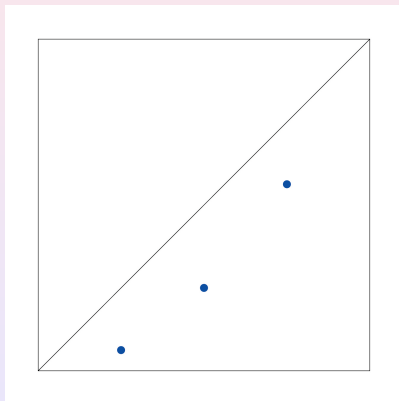
- metodą graficzną - za pomocą tzw. **krzywej koncentracji Lorenza**,
- metodą analityczną - obliczając tzw. **współczynnik koncentracji Giniego**.

Krzywa koncentracji Lorenza



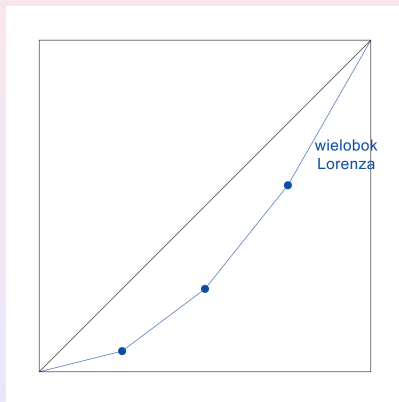
1. Rysujemy kwadrat o boku 1 (lub 100%) i jego przekątną
2. Zaznaczamy punkty (w_i^{skum}, z_i^{skum}) , gdzie w_i^{skum} , z_i^{skum} - skumulowane wskaźniki struktury dla x i y
3. Łączymy narysowane punkty otrzymując wielobok Lorentza
4. Wygładzamy wielobok, otrzymując krzywą Lorentza

Krzywa koncentracji Lorenza



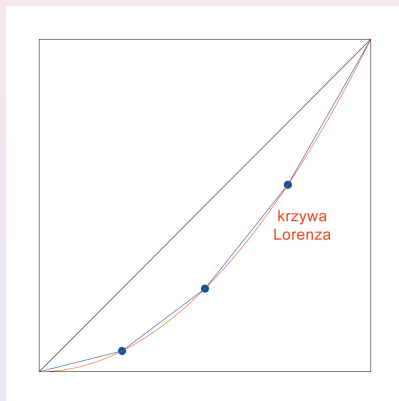
1. Rysujemy kwadrat o boku 1 (lub 100%) i jego przekątną
2. Zaznaczamy punkty (w_i^{skum}, z_i^{skum}) , gdzie w_i^{skum} , z_i^{skum} - skumulowane wskaźniki struktury dla x i y
3. Łączymy narysowane punkty otrzymując wielobok Lorenza
4. Wygładzamy wielobok, otrzymując krzywą Lorenza

Krzywa koncentracji Lorenza



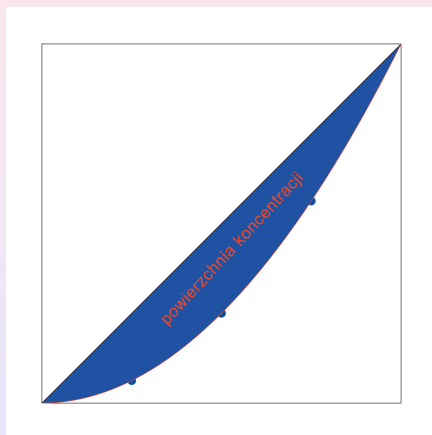
1. Rysujemy kwadrat o boku 1 (lub 100%) i jego przekątną
2. Zaznaczamy punkty (w_i^{skum}, z_i^{skum}) , gdzie w_i^{skum} , z_i^{skum} - skumulowane wskaźniki struktury dla x i y
3. Łączymy narysowane punkty otrzymując wielobok Lorenza
4. Wygładzamy wielobok, otrzymując krzywą Lorenza

Krzywa koncentracji Lorenza

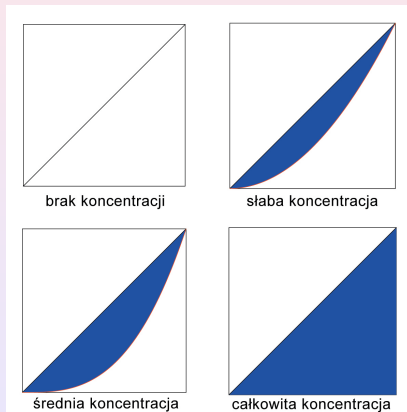


1. Rysujemy kwadrat o boku 1 (lub 100%) i jego przekątną
2. Zaznaczamy punkty (w_i^{skum}, z_i^{skum}) , gdzie w_i^{skum} , z_i^{skum} - skumulowane wskaźniki struktury dla x i y
3. Łączymy narysowane punkty otrzymując wielobok Lorenza
4. Wygładzamy wielobok, otrzymując **krzywą Lorenza**

Krzywa koncentracji Lorenza



Krzywa koncentracji Lorenza



WSPÓŁCZYNNIK KONCENTRACJI GINIEGO
(czasami zwany współczynnikiem koncentracji Lorenza) bazuje
na polu powierzchni figury zawartej między krzywą Lorenza a
linią równomiernego podziału (tzw. powierzchni Lorenza).

Współczynnik koncentracji Giniego oblicza się jako iloraz pola figury a oraz całego trójkąta AOB będącego obrazem całkowitej koncentracji. Trójkąt ten jest połową kwadratu o boku 100, zatem ma pole powierzchni równe 5000. Stąd

$$G = \frac{a}{5000} \quad \text{lub} \quad G = \frac{a}{0,5} = 2a,$$

jeśli używamy zwykłych wskaźników struktur (bez mnożenia przez 100).

Współczynnik Giniego pozwala jednoznacznie, za pomocą jednej liczby, ocenić siłę koncentracji badanego zjawiska, jak również porównać koncentrację kilku zjawisk (jest miarą względną).

$G \in [0, 1]$, przy czym jeśli G jest liczbą z przedziału:

- $0 - 0,3$ – słaba koncentracja,
- $0,3 - 0,6$ – umiarkowana koncentracja,
- $0,6 - 1$ – silna koncentracja.

Przypadki skrajne:

- $G = 0$ dla braku koncentracji,
- $G = 1$ dla całkowitej koncentracji.

W praktyce łatwiej jest policzyć pole figury b leżącej pod krzywą koncentracji Lorenza. Wówczas współczynniki Giniego przyjmują postać:

$$G = \frac{5000 - b}{5000} \quad \text{lub} \quad G = \frac{0,5 - b}{0,5} \quad \text{odpowiednio.}$$

PRZYKŁAD KONSTRUKCJI KRZYWEJ LORENZA I
OBLICZENIA WSPÓŁCZYNNIKA GINIEGO