26.03.2014r.

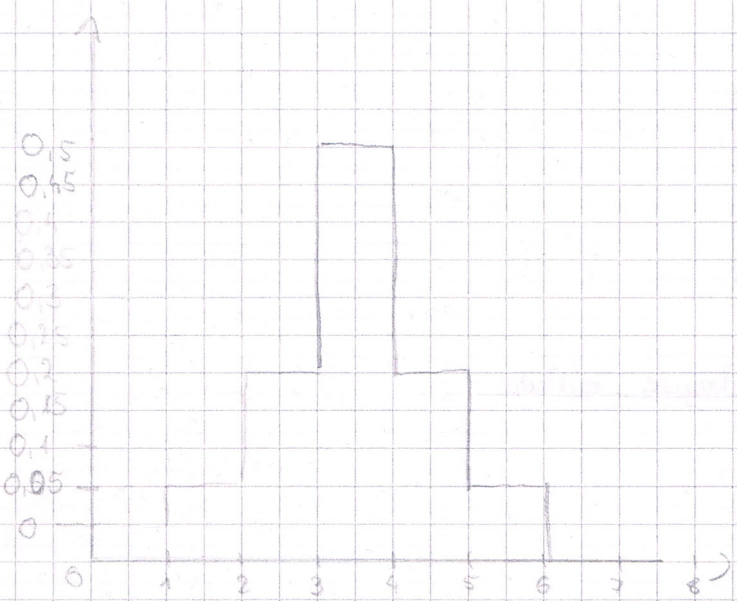
MIARY ASYMETRII

★ Rozkład symetryczny - histogram

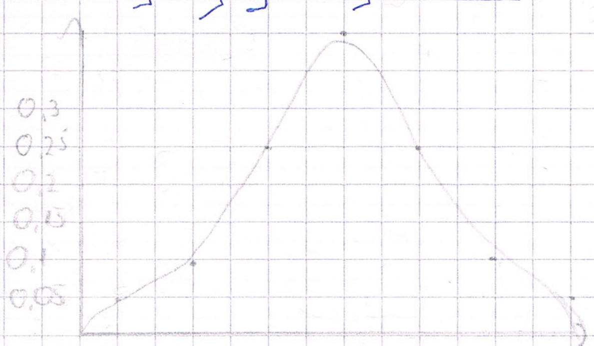
Cecha ma rozkład symetryczny, jeżeli:

$\bar{x} = Me = Mo$

Histogram cechy ciągłej symetrycznej

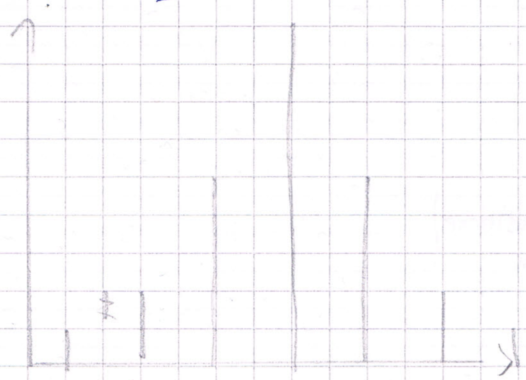


★ Rozkład symetryczny - krzywa rozkładu



★ Rozkład symetryczny - częstości

Cecha dyskretna

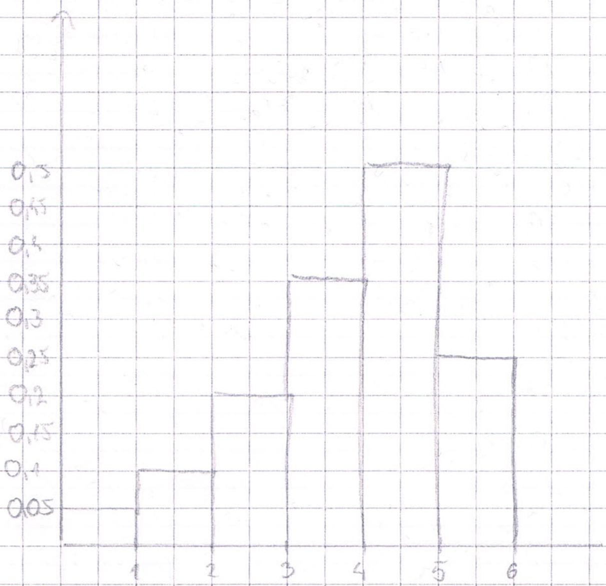


* Rozkład asymetryczny lewostronnie - histogram

Cecha ma ~~charakter~~ rozkład asymetryczny lewostronnie, jeśli

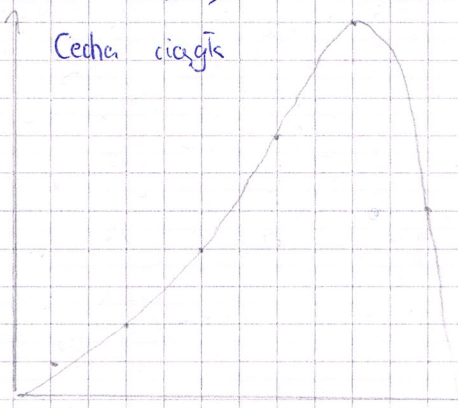
$\bar{x} < Me < Mo$

Histogram cechy ciągłej lewostronnie asymetrycznej



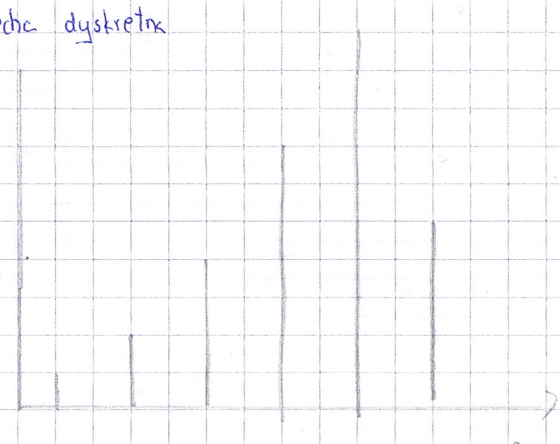
* Rozkład asymetryczny lewostronnie - krzywa rozkładu

Cecha ciągła



* Rozkład asymetryczny lewostronnie - częstość

Cecha dyskretna



* Rozkład asymetryczny prawostronnie (odbicie lustrzane)
- histogram
- krzywa
- częstość

$\bar{x} > Me > Mo$

* Współczynnik asymetrii

Współczynniki asymetrii określają kierunek skośności, a także siłę skłonności. Są miarami niemianowanymi (nie są wyrażone w jednostkach) i względnymi.

- Klasyczny współczynnik asymetrii

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

- Pozycyjny współczynnik asymetrii

$$A_Q = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1}$$

Ich znak mówi o kierunku asymetrii, a im wyższa jest ich wartość bezwzględna, tym asymetria jest silniejsza.
Pozwalają porównać asymetrię różnych rozkładów

* Moment centralny rzędu trzeciego

- Dla szeregu szczegółowego

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

- Dla szeregu rozdzielczego punktowego

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

- Dla szeregu rozdzielczego przedziałowego

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

Kierunki asymetrii za pomocą m_3:

1. $m_3 = 0$ dla rozkładu symetrycznego
2. $m_3 < 0$ dla rozkładu asymetrycznego lewostronnie
3. $m_3 > 0$ dla rozkładu asymetrycznego prawostronnie

* Indeks skośności

Kierunek i siłę asymetrii rozkładu można mierzyć indeksem skośności określonym wzorem:

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

Własności indeksu skośności

1. $-1 \leq A_m \leq 1$
2. $A_m = 0$, gdy rozkład jest symetryczny
3. $A_m < 0$, gdy rozkład jest asymetryczny lewostronnie

* <u>Przykład - dane symetryczne</u>

$$\underset{}{1},\ 2,\ \overset{Q_1}{\overset{\downarrow}{3}},\ 4,\ \overset{Me}{\overset{\downarrow}{5}},\ 6,\ \overset{Q_3}{\overset{\downarrow}{7}},\ 8,\ 9$$

$$Q = \frac{Q_3 - Q_1}{2} = \frac{7-3}{2} = 2$$

$$V_Q = \frac{Q}{Me} = \frac{2}{5} = 0{,}4$$

$$V_{Q_1Q_2} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{7-3}{7+3} = \frac{4}{10} = 0{,}4$$

$$A_Q = \frac{Q_3 + Q_1 - 2Me}{2Q} = \frac{7+3-2\cdot 5}{2\cdot 5} = 0$$

$$\bar{x} = \frac{1+2+\ldots+9}{9} = \frac{45}{9} = 5$$

$$\bar{x}_G = 4{,}15$$

$$s^2 = \frac{2(1+4+9+16)}{9} = \frac{60}{9} = 6{,}92$$

$$s = \sqrt{s^2} = 2{,}58$$

$$V_s = 0{,}52$$

$$1,\ 1,\ \overset{Q_1}{\overset{\downarrow}{1}},\ 1,\ \overset{Me}{\overset{\downarrow}{2}},\ 2,\ \overset{Q_3}{\overset{\downarrow}{5}},\ 5,\ 9$$

$$Mo = 1$$

$$Q = \frac{Q_3 - Q_1}{2} = \frac{5-1}{2} = 2$$

$$V_Q = \frac{Q}{Me} = \frac{2}{2} = 1$$

$$V_{Q_1Q_2} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{5-1}{5+1} = \frac{4}{6} = 0{,}67$$

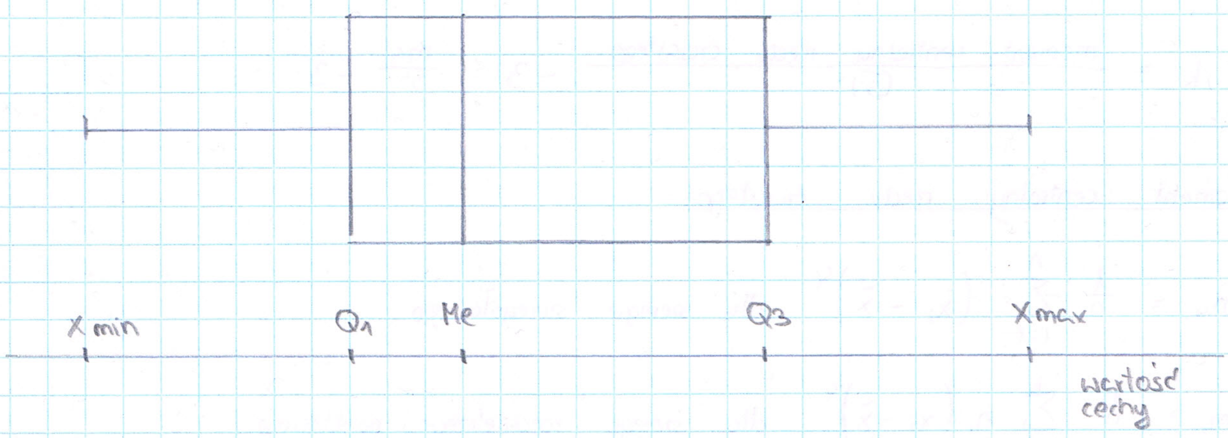
$$A_Q = \frac{Q_3 + Q_1 - 2Me}{2Q} = \frac{5+1-2\cdot 2}{2\cdot 2} = \frac{2}{4} = 0{,}5$$

2. 04. 2014 r.

WYKRES PUDEŁKOWY. KRZYWA LORENZA.

1. Graficzna prezentacja rozkładu za pomocą kwartyli..

- Wykres pudełkowy (ang. box-plot) - jest schematycznym sposobem graficznej prezentacji rozkładu za pomocą pozycyjnych miar położenia Q_1, Me, Q_3

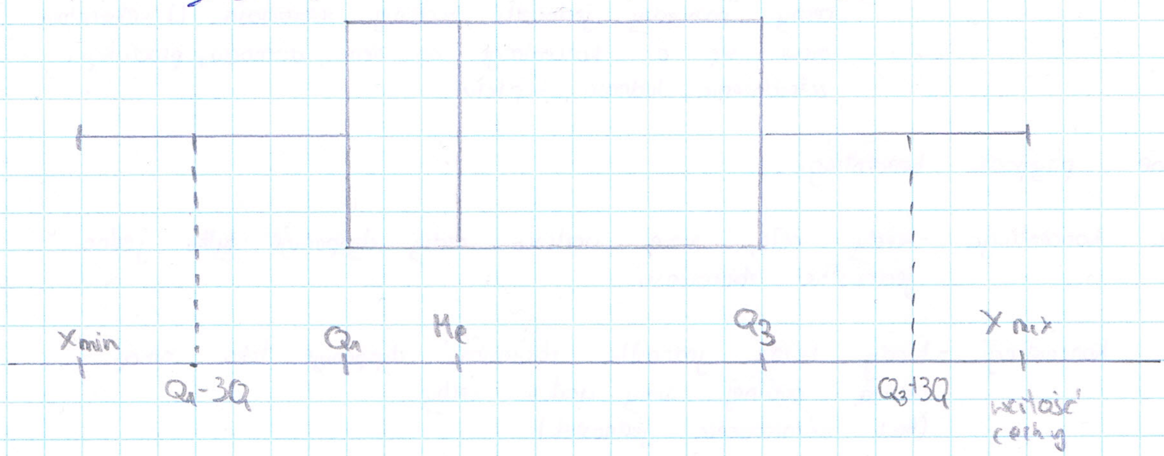


- wysokość pudełka nie ma znaczenia

Jeżeli obserwacje cechy badane nie mieszczą się w przedziale,

$$[Q_1 - 3Q \; ; \; Q_3 + 3Q]$$

to znaczymy je kropkami na wykresie pudełkowym.



2. Informacje odczytane z wykresu

Z wykresu pudełkowego odczytujemy:

I. Kwartyle Q_1 i Q_3 - pionowe brzegi pudełka
II. Medianę Me - linia dzieląca pudełko
III. Zmienność - szerokość pudełka
IV. Skośność rozkładu - odległość lini dzielącej pudełko od pionowego brzegu pudełka
V. Wartości cechy znacznie odbiegające od wartości typowych, zwanych wartościami odstającymi, a na wykresie zaznaczonymi kropkami

3. Współczynnik koncentracji

Rodzaje koncentracji:

* Koncentracja wartości cechy wokół średniej - miarą tego rodzaju koncentracji jest kurtoza
* Koncentracja jako nierównomierny podział łącznej sumy wartości cechy pomiędzy jednostki zbiorowości. Miarą tego rodzaju koncentracji jest współczynnik Giniego

Koncentrację wartości cechy wokół średniej mierzy się współczynnikiem koncentracji, zwanym też KURTOZĄ, określonym wzorem:

$$k = \frac{\text{moment centralny rzędu czwartego}}{(S)^4} - 3 = \frac{m_4}{S^4} - 3$$

4. Moment centralny rzędu czwartego:

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

$$m_4 = \frac{1}{n} \sum_{i=1}^{k} n_i (x_i - \bar{x})^4 \quad \text{dla szeregu rozdzielczego punktowego}$$

$$m_4 = \frac{1}{n} \sum_{i=1}^{k} n_i (\mathring{x}_i - \bar{x})^4 \quad \text{dla szeregu rozdzielczego przedziałowego}$$

Kurtoza dla rozkładu normalnego, tzn gdy krzywą gęstości jest Gaussa, wynosi zero

5. Koncentracja wartości cechy

- informacje o nierównomiernym rozdziale łącznego funduszu cechy pomiędzy jednostki badanej zbiorowości. W ekonomii mówi się o koncentracji np. ziemi, dochodu, produkcji, zatrudnienia, ludności, kapitału

Skrajne przypadki koncentracji:

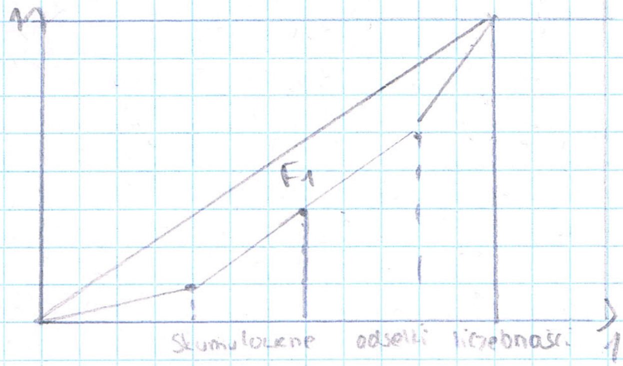
* Pełna koncentracja - kiedy całą sumą wartości cechy dysponuje tylko jedna jednostka zbiorowości
* Brak koncentracji - kiedy każda jednostka zbiorowości dysponuje taką samą częścią ogólnej sumy wartości cechy.
  (brak zróżnicowania jednostek)

Miarą koncentracji wartości cechy jest krzywa Lorenza, oraz oszacowany na podstawie krzywej Lorenza współczynnik Giniego.

6. Stopień koncentracji

Można ocenić przez porównanie częstości występowania jednostek w różnych przedziałach wartości cechy z udziałem wartości cechy w poszczególnych przedziałach w łącznym funduszu cechy

7. Wielobok koncentracji Lorenza



8. Opis rysunku

- Na osi odciętych ⟶ zaznaczamy skumulowany częstości względne liczebności (skumulowane odsetki liczebności)
- Na osi odciętych zaznaczamy skumulowane częstości względne sumy wartości cechy (skumulowane odsetki funduszu cechy)
- Linia łamana na rysunku nazywa się wielobokiem (krzywa) koncentracji Lorentza
- Jeśli nie ma koncentracji, to krzywa Lorenza pokrywa się z przekątną kwadratu, tzn. skumulowane częstości względne jednostek są równe względnym sumom wartości cechy w każdym z przedziałów wartości cechy.
- Jeżeli występuje pełna koncentracja, to krzywa Lorentza pokrywa się z bokiem kwadratu leżącym na osi, tzn pole figury F_2 wynosi zero

9 Współczynnik Giniego

Im większe pole figury F_1 (a tym samym mniejsze pole figury F_2), tym koncentracja wartości cechy badanej jest silniejsza.

Siłę koncentracji mierzymy współczynnikie Giniego określonym wzorem:

$$k = \frac{\text{pole } F_1}{\text{pole } F_1 + \text{pole } F_2} = \frac{\frac{1}{2} - \text{pole } F_2}{\frac{1}{2}} = 1 - 2|F_2|$$

$$0 \le k \le 1$$

$k = 0$, kiedy nie ma koncentracji

$k = 1$, gdy jest pełna koncentracja