# Analiza struktury c.d.

**Wykład 6**

elzbieta.golata@ue.poznan.pl
dr hab. Elżbieta Gołata, prof. nadzw. UEP,
Katedra Statystyki
Wydział Informatyki i Gospodarki Elektronicznej
Uniwersytet Ekonomiczny w Poznaniu

# MIARY ANALIZY STRUKTURY

KLASYCZNE | POZYCYJNE

## 1. CHARAKTERYSTYKI TENDENCJI CENTRALNEJ

- średnia arytmetyczna
- średnia geometryczna
- średnia harmoniczna
- średnia kwadratowa

- kwantyle (kwartyle, decyle, percentyle)
- dominanta (wartość najczęściej występująca, moda)

## 2. CHARAKTERYSTYKI ZRÓŻNICOWANIA - DYSPERSJI - ZMIENNOŚCI

- odchylenie przeciętne
- wariancja
- odchylenie standardowe
- klasyczny współ. zmienności

- rozstęp, obszar zmienności
- odchylenie ćwiartkowe
- odchylenie decylowe ...
- pozycyjny współ. Zmienności

## 3. CHARAKTERYSTYKI ASYMETRII - SKOŚNOŚCI

- moment trzeci centralny
- moment trzeci centralny stand.

- pozycyjny miernik asymetrii
- pozycyjny współ. asymetrii

klasyczno-pozycyjny miernik asymetrii

klasyczno-pozycyjny współczynnik asymetrii

## 4 A. CHARAKTERYSTYKI KONCENTRACJI WOKÓŁ ŚREDNIEJ

(kurtozy-ekscesu)

moment czwarty centralny

moment czwarty centralny standaryzowany

## 4 B. CHARAKTERYSTYKI KONCENTRACJI-RÓWNOMIERNOŚCI PODZIAŁU

współczynnik koncentracji K

## MIARY ASYMETRII WSPÓŁCZYNNIKI ASYMETRII

$A_{s(x)} = \frac{\bar{x} - D}{s(x)}$ klasyczno-pozycyjny współczynnik asymetrii $A_{s(x)} \in \langle -1;1 \rangle$

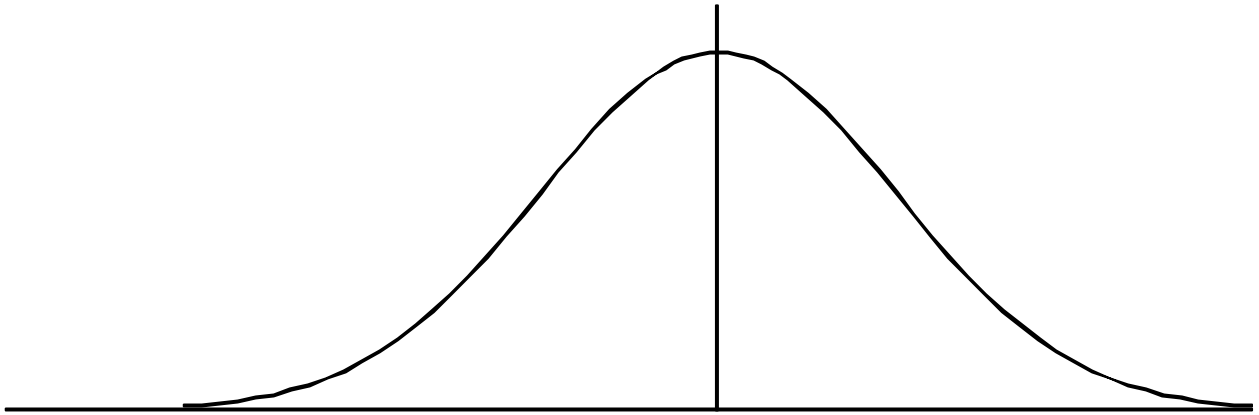
$A_{Q(x)} = \frac{Q_3 + Q_1 - 2Q_2}{2Q(x)}$ pozycyjny współczynnik asymetrii $A_{Q(x)} \in \langle -1;1 \rangle$

$A_{Q(x)} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$

$(Q_3 - Q_2) - (Q_2 - Q_1)$ $a+$

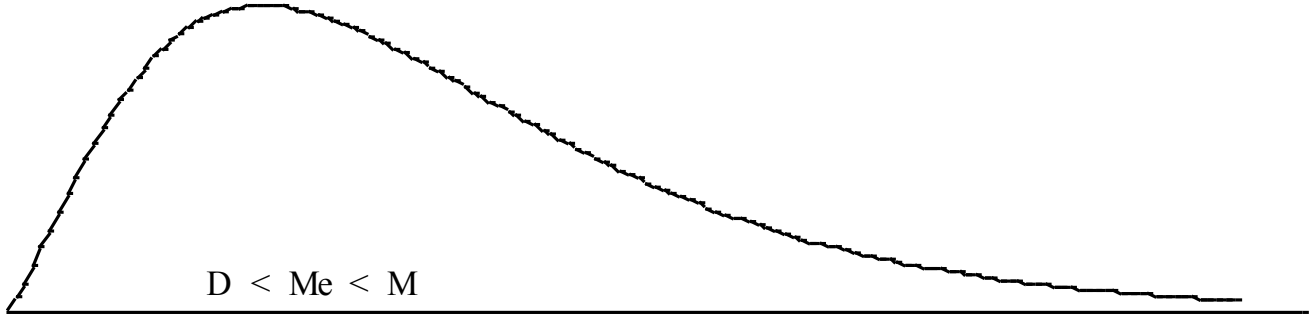
$(Q_3 - Q_2) + (Q_2 - Q_1)$ $a-$

**Rozkład symetryczny**

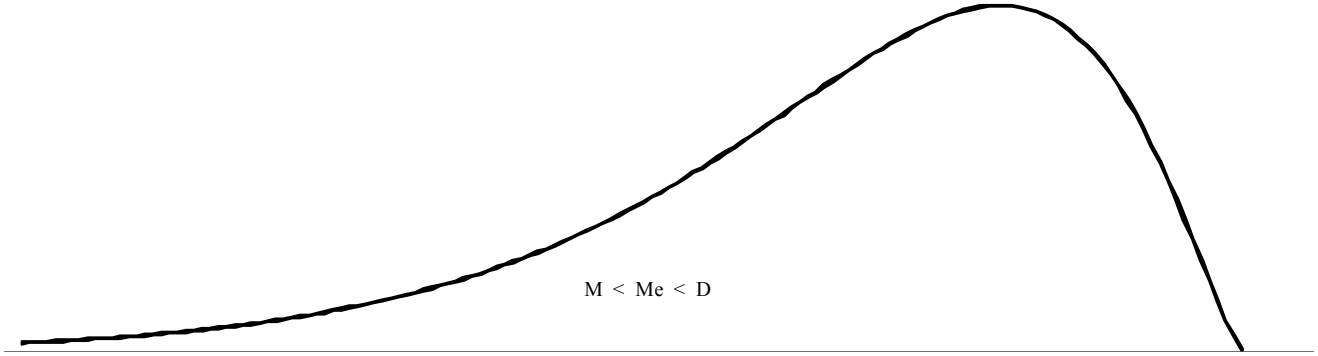


M=Me=D

**Asymetria prawostronna**




**Asymetria lewostronna**



## RACHUNEK MOMENTÓW

Dowolnym k-tym momentem rozkładu nazywamy średnią arytmetyczną z odchyleń poszczególnych wartości zmiennej $X$ od dowolnej liczby $x_0$ podniesionych do k-tej potęgi:

$$m_k = \frac{\sum (x_i - x_0)^k n_i}{n}$$

### MOMENTY ZWYKŁE

$$m_1 = \frac{\sum (x_i - 0)}{n} = \frac{\sum x_i}{n} = \bar{x}$$

$$m_2 = \frac{\sum (x_i - 0)^2}{n} = \frac{\sum x_i^2}{n} = \overline{x^2}$$

### MOMENTY CENTRALNE

$$\mu_1 = \frac{\sum (x_i - \bar{x})^1 n_i}{n} = 0$$ zawsze równy zero

$$\mu_2 = \frac{\sum (x_i - \bar{x})^2 n_i}{n} = s^2(x)$$ wariancja

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3 n_i}{n}$$ miara asymetrii

$$\mu_4 = \frac{\sum (x_i - \bar{x})^4 n_i}{n}$$ miara ekscesu

## moment trzeci centralny wyrażony w jednostkach odchylenia standardowego:

$$\alpha_3 = \frac{\frac{1}{n}\sum (x_i - \bar{x})^3}{s^3(x)} \qquad \alpha_3^{'} = \frac{\mu_3}{s^3(x) + |\mu_3|} \qquad -1 < \alpha_3^{'} < 1$$

## moment czwarty centralny wyrażony w jednostkach odchylenia standardowego:

$$\alpha_4 = \frac{\frac{1}{n}\sum (x_i - \bar{x})^4}{s^4(x)}$$

$\alpha_4 < 3$

$\alpha_4 = 3$

$\alpha_4 > 3$

$$\alpha_4^{'} = \frac{\mu_4 - s(x)^4}{\mu_4} \qquad \alpha_4^{'} \in \langle 0;1 \rangle$$

dla rozkładu normalnego $\alpha_4^{'} = 0{,}666$

## 5 - LICZBOWA SYNTEZA FIVE NUMBER SUMMARY

### SYNTETYCZNY OPIS ZBIOROWOŚCI PRZY POMOCY PIĘCIU LICZB

**Przykład**

W 2000 r. Pentor opublikował informacje dotyczące czasu oglądania telewizji przez dzieci w wieku szkolnym według różnych charakterystyk społecznych. W przykładowej próbie 20 dzieci w jednej z poznańskich szkół podstawowych otrzymano następujące informacje o przeciętnej liczbie godzin spędzanych przed telewizorem w ciągu tygodnia:

25 41 27 32 43 66 35 31 15 5
34 26 32 38 16 30 38 30 20 21

Proszę przedstawić i zinterpretować 5 – liczbową syntezę.

**Rozwiązanie:**

5 15 16 20 21 25 26 27 30 30 31 32 32 34 35 38 38 41 43 66

**WARTOŚCI EKSTREMALNE: MIN=5 GODZIN MAX=66 GODZIN**

**Kwartyle:**

- Ponieważ *n=20* *n/4=5* ***$Q_1$= 21+0,25*(25-21)=22***
- Ponieważ *n=20* *n/2=10* ***$Q_2$= 30+0,5*(31-30)=30,5***
- Ponieważ *n=20* *3/4n=15* ***$Q_3$= 35+0,75*(38-35)=37,25***

Łącznie z uzupełniającą informacją o zróżnicowaniu badanej zbiorowości, 5-liczbową syntezę można zapisać następująco:

***Min = 5*** ***$Q_1$ = 22*** ***$Q_2$ = 30,5*** ***$Q_3$ = 37,25*** ***Max = 66***

***$Q_1$ - Min = 17*** ***$Q_2$ -$Q_1$ = 8,5*** ***$Q_3$ - $Q_2$ = 6,75*** ***Max- $Q_3$=28,75***

## JEDNOSTKI O WARTOŚCIACH SKRAJNYCH - OUTLIERS

<DG, GG> jednostki, które przyjmują wartość cechy z tego przedziału **nie** są traktowane jako odstające.

$$DG = Q_1 - 1{,}5 \cdot IQR$$

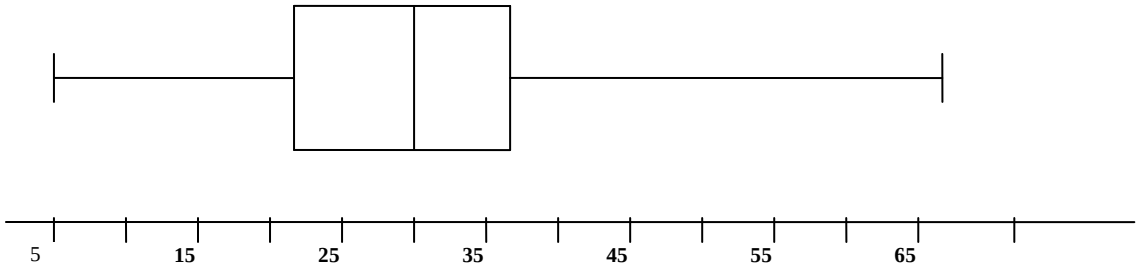
$$GG = Q_3 + 1{,}5 \cdot IQR$$

$$IQR = Q_3 - Q_1.$$

***Rozstęp między-kwartylowy*** ***IQR = 37,25 – 22 = 15,25***
***Dolna Granica*** ***$Q_1$ -1,5 · IQR = 22 - 1,5 · 15,25 = - 0,875***
***Górna Granica*** ***$Q_3$ +1,5 · IQR = 37,25 + 1,5 · 15,25 = 60,125***

## WYKRES PUDEŁKOWY – BOXPLOT

- ✧ Przedstaw 5-liczbowa syntezę
- ✧ Narysuj oś liczbową i nanieś na nią wielkości obliczone w poprzednim kroku. Powyżej osi zaznacz krótkie odcinki pionowe w miejscach odpowiadających kwartylom i połącz je tworząc prostokąt podzielony na dwie części w miejscu odpowiadającym kwartylowi drugiemu.
- ✧ Zaznacz przy pomocy krótkich odcinków pionowych wartości ekstremalne. Połącz odcinkami (tzw. wąsami) boki prostokąta odpowiadające kwartylom z wartościami minimalna i maksymalną.




***Min = 5*** ***$Q_1$ = 22*** ***$Q_2$ = 30,5*** ***$Q_3$ = 37,25*** ***Max = 66***

***$Q_1$ - Min = 17*** ***$Q_2$ -$Q_1$ = 8,5*** ***$Q_3$ -$Q_2$ = 6,75*** ***Max-$Q_3$ =28,75***

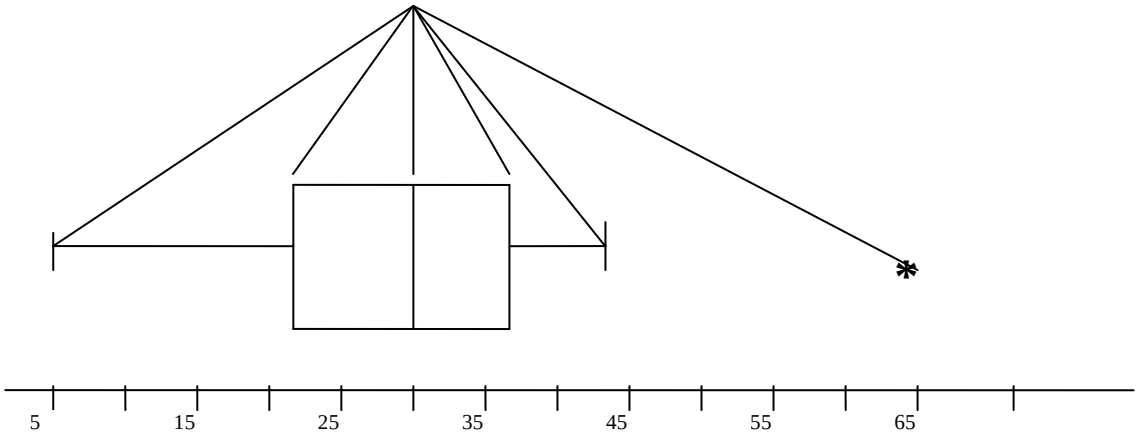
| 5 - liczbowa synteza | |
|---|---|

## ZMODYFIKOWANY WYKRES PUDEŁKOWY

- ✧ Wyznacz kwartyle
- ✧ Zidentyfikuj potencjalne wartości odstające oraz wartości *ekstremalne** nie będące wartościami odstającymi, tzn. zawarte w przedziale <*DG ; GG*>. Jeżeli w zbiorze nie występują wartości odstające, wówczas są to po prostu wartości ekstremalne tzn. *Min* i *Max*
- ✧ Narysuj oś liczbową i nanieś na nią wielkości obliczone w pierwszym kroku. Powyżej osi zaznacz krótkie odcinki pionowe w miejscach odpowiadających kwartylom i połącz je tworząc prostokąt podzielony na dwie części w miejscu odpowiadającym medianie. Zaznacz przy pomocy krótkich odcinków pionowych wartości *ekstremalne**. Połącz odcinkami (tzw. wąsami) boki prostokąta odpowiadające kwartylom z wartościami *ekstremalnymi**.
- ✧ Narysuj gwiazdkę odpowiadającą każdej potencjalnej wielkości odstającej

| 5 - liczbowa synteza | |
|---|---|

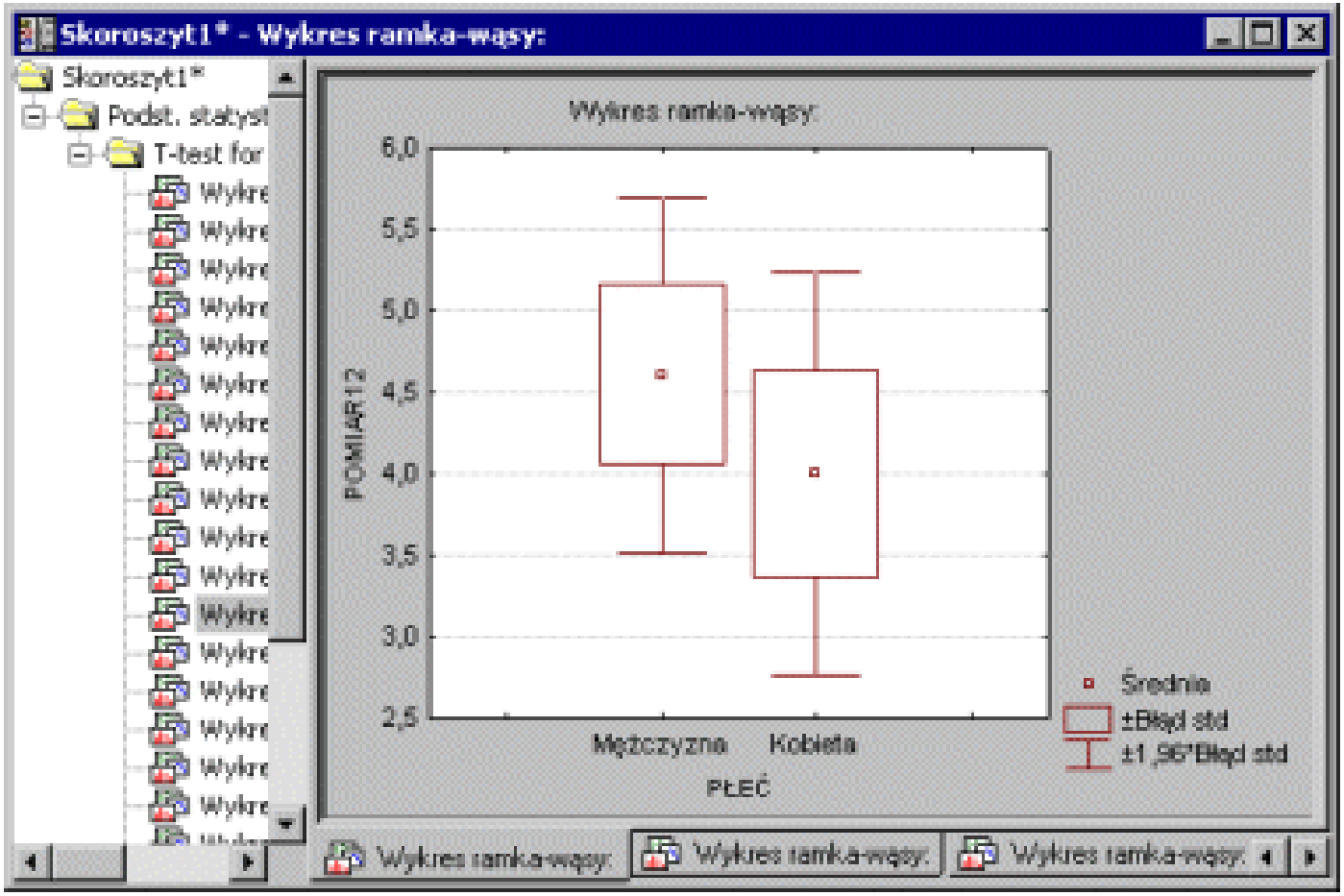
***Ekstremalna D*= Min*** ***Ekstremalna G*= 43***

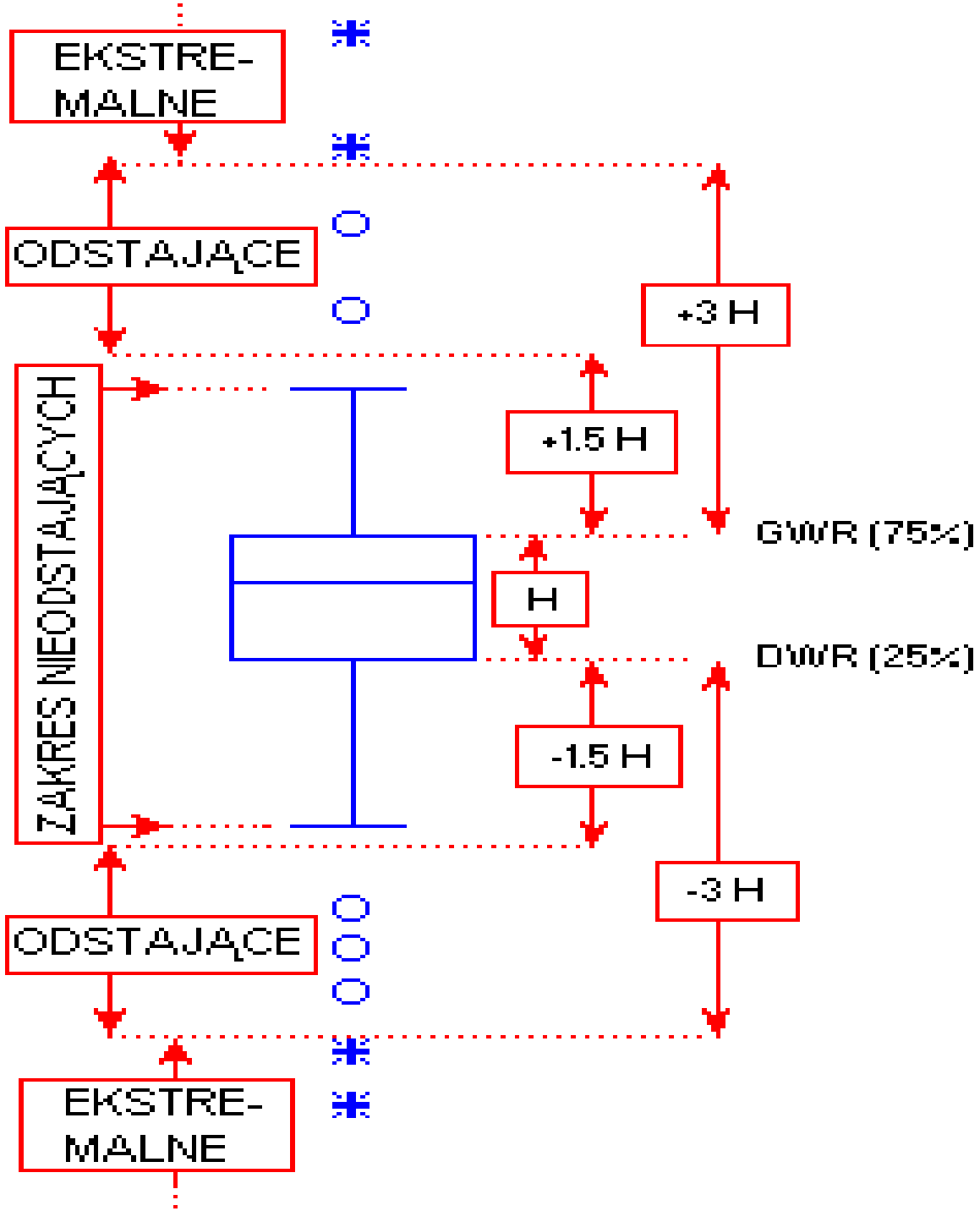



***Min = 5*** ***$Q_1$ = 22*** ***$Q_2$ = 30,5*** ***$Q_3$ = 37,25*** ***Max = 66***

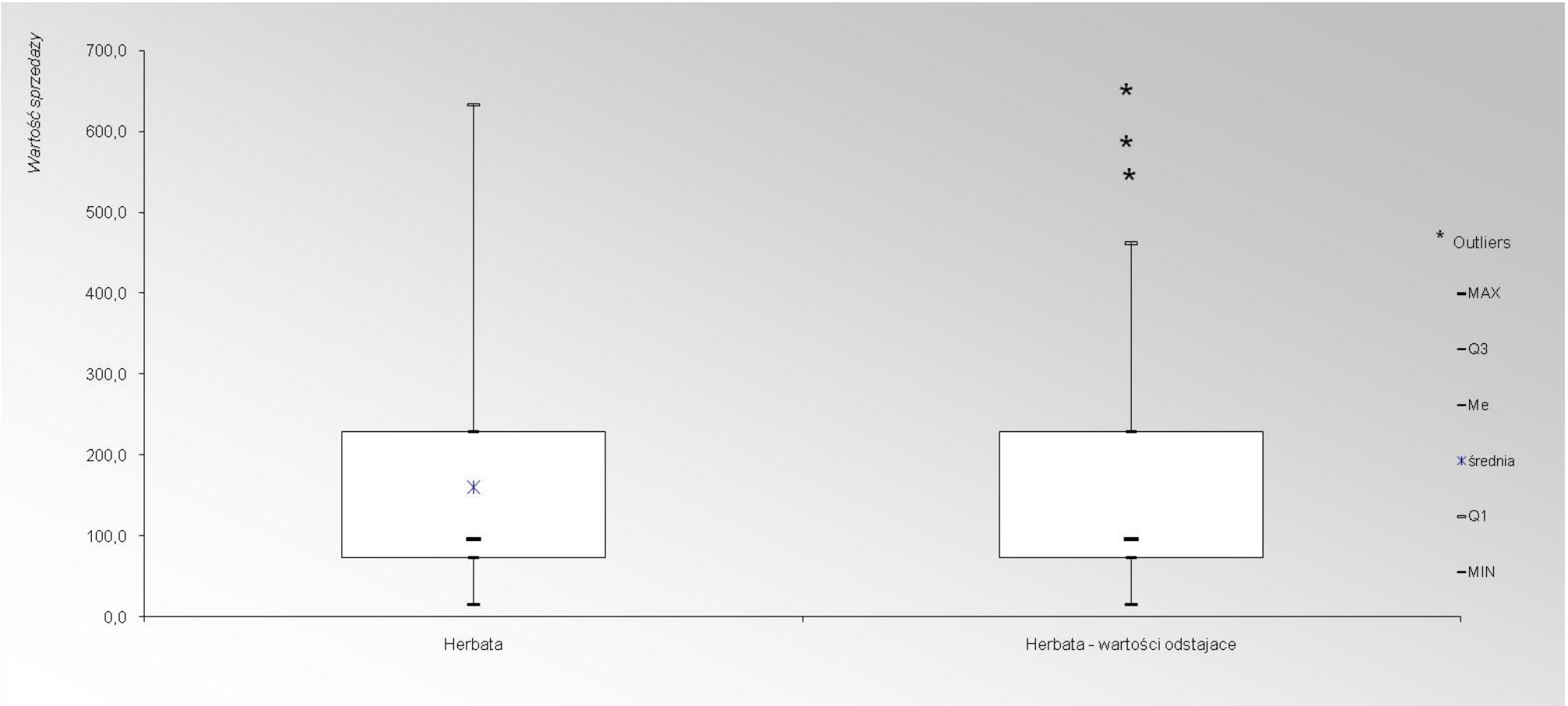
***$Q_1$ - Min = 17*** ***$Q_2$ -$Q_1$ = 8,5*** ***$Q_3$ - $Q_2$ = 6,75*** ***Max-$Q_3$=28,75***

| 5 - liczbowa synteza | |
|---|---|




EKSTRE-
MALNE
ODSTAJĄCE
ZAKRES NIEODSTAJĄCYCH
ODSTAJĄCE
EKSTRE-
MALNE
+3 H
+1.5 H
GWR (75%)
H
DWR (25%)
-1.5 H
-3 H

## 5 - liczbowa synteza




| Statystyka opisowa | Podstawowe rodzaje badań statystycznych |
|---|---|

# PRZYKŁADY → EXCEL

Wyniki ankiety socjologicznej przeprowadzonej w październiku 2006 roku przez studentów socjologii UAM w grupie 500 studentów z różnych uczelni wyższych z województwa wielkopolskiego dostarczyły następujących informacji o przeciętnym czasie spędzanym tygodniowo w Internecie

| Tygodniowy czas pracy w Internecie | Liczba studentów |
|---|---|
| $x_i$ | $n_i$ |
| Poniżej 2 godzin | 13 |
| 2-3 | 37 |
| 3-5 | 57 |
| 5-10 | 115 |
| 10-15 | 147 |
| 15-20 | 63 |
| 20-30 | 46 |
| 30-50 | 14 |
| 50 i więcej | 8 |

Powierzchnię gospodarstw rolnych w województwie wielkopolskim opisuje poniższy szereg:

| Powierzchnia gospodarstw rolnych (w ha) | Odsetek gospodarstw |
|---|---|
| 0 – 3 | 20 |
| 3-6 | 30 |
| 6-10 | 25 |
| 10-15 | 15 |
| 15 i więcej | 10 |
| Razem | 100 |

**Zdefiniuj zbiorowość statystyczną, jednostkę oraz scharakteryzować badaną cechę.**
**Przeprowadź kompleksową analizę struktury**

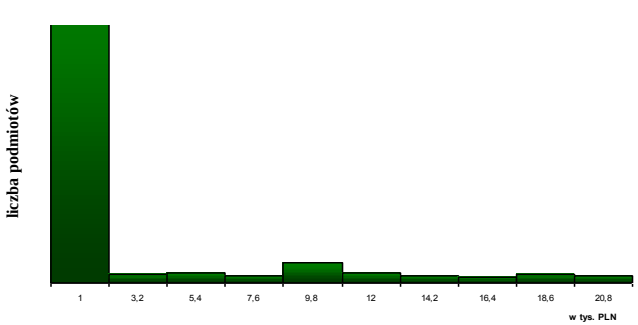
W banku X poddano obserwacji wysokość udzielanych kredytów konsumpcyjnych. Dla 200 wybranych i udzielonych kredytów konsumpcyjnych w 2006 roku otrzymano poniższy szereg:

| Wysokość kredytu (w tys. zł) | Liczba kredytów |
|---|---|
| 0-5 | 20 |
| 5-10 | 40 |
| 10-15 | 50 |
| 15-20 | 40 |
| 20-25 | 30 |
| 25-30 | 20 |
| Razem | 200 |

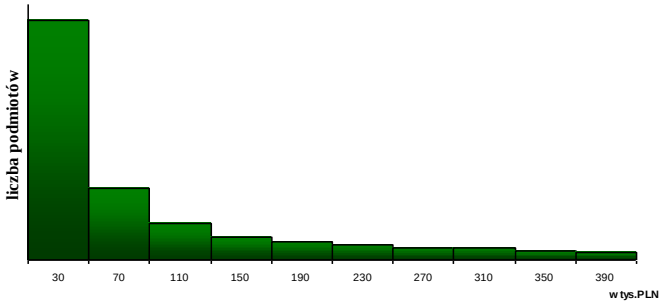
| Współczynnik Giniego | |
|---|---|

# 1. Koncentracja jednostek zbiorowości wokół wartości średniej (kurtoza)
# 2. Nierównomierność rozkładu wartości globalnej cechy na poszczególne jednostki zbiorowości:

**Dochód** **Podatki** **Sprzedaż**




Rys.1 Rozkład podmiotów gospodarczych według sumy wypłaconych wynagrodzeń w roku, SP3, 2001




Rys.2 Rozkład podmiotów gospodarczych według przychodów, SP-3, 2001

bezpośrednio związana z dyspersją i asymetrią – im silniejsza asymetria i większe zróżnicowanie jednostek, tym koncentracja jest większa

**2 skrajne przypadki** **- Skrajne przypadki nie występują**

- **brak koncentracji** – na każdą jednostkę przypada taka sama część ogólnej sumy wartości
(każdy pracownik w przedsiębiorstwie otrzymuje taką samą cześć łącznego funduszu płac

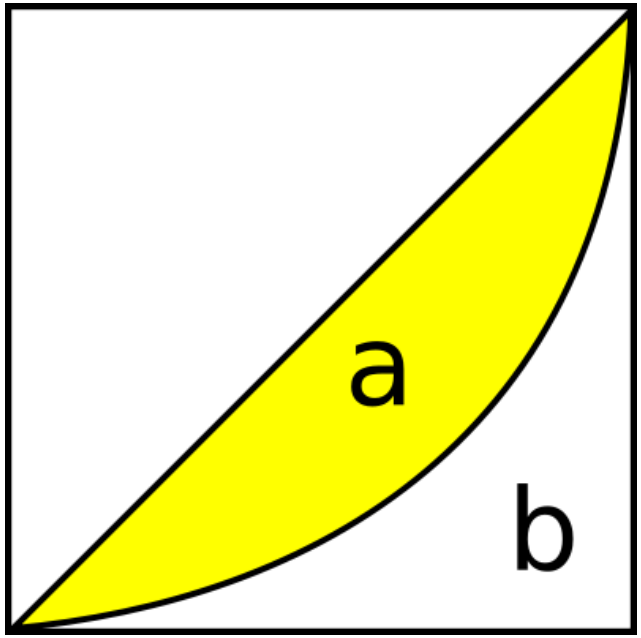
- **koncentracja zupełna** (całkowita)
łączny fundusz cechy przypada na daną jednostkę zbiorowości (łączny fundusz płac przypada na jednego pracownika

## 2 metody badania siły koncentracji: graficzna i analityczna

### KRZYWA LORENZA - metoda graficzna

$$G = \frac{a}{(a+b)}$$

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



# Współczynnik Giniego - metoda analityczna

Współczynnik zaproponowany w 1912 r. przez włoskiego statystyka Corrado Giniego (1884-1965) jest miarą stopnia nierówności rozkładu dochodów w danym społeczeństwie.

- współczynnik Giniego przyjmuje wartości z przedziału [0; 1],
- wartość zerowa współczynnika wskazuje na pełną równomierność rozkładu dochodów,
- wzrost wartości współczynnika oznacza wzrost nierówności dochodowych,
- wartość równa jedności sytuacja absolutnej koncentracji majątku (dochodu) w posiadaniu jednego tylko gospodarstwa

W przypadku szeregu szczegółowego zastosujemy wzór:

$$G(x) = 1 - \frac{1}{n^2 \cdot \bar{x}} \sum_{i=1}^{n} \left( x_i + 2 \sum_{s=1}^{i-1} x_s \right)$$

•współczynnik Giniego przyjmuje wartości z przedziału [0; 1],

| **Współczynnik Giniego** | |
|---|---|

W przypadku szeregu strukturalnego zastosujemy wzór:

$$G(x) = 1 - \sum_{i=1}^{k} \left[ \sum_{s=1}^{i-1} z_s(x) + \sum_{s=1}^{i} z_s(x) \right] w_i$$

gdzie:

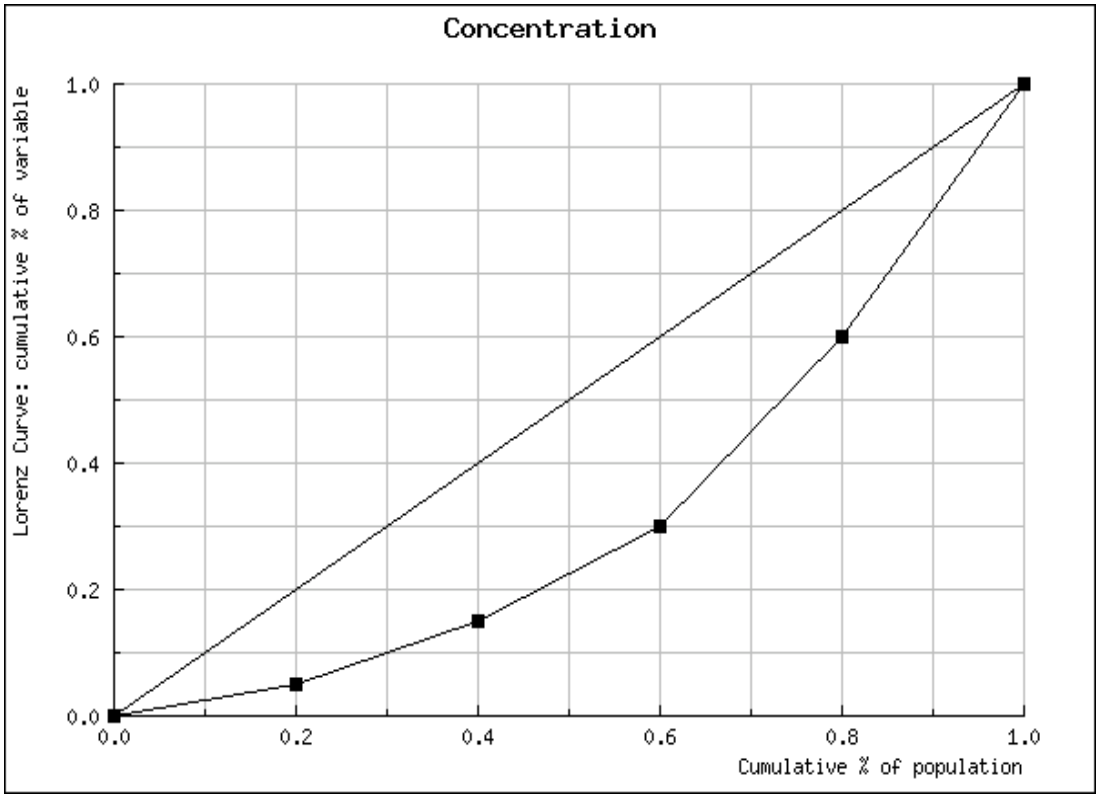
$$w_i = \frac{n_i}{n}$$

$$z_i(x) = \frac{x_i n_i}{\sum_{i=1}^{k} x_i n_i}$$

Metoda przybliżona sprowadza się do obliczania pól figur geometrycznych:

trójkąta i trapezów wyznaczonych przedziałami klasowymi

Zastosujemy wówczas wzór:




$$G(x) = \frac{0{,}5 - \sum_{i=1}^{k} \left( \frac{{}_{cum}z_i + {}_{cum}z_{i-1}}{2} \right) w_i}{0{,}5}$$

| **Współczynnik Giniego** | |
|---|---|

Rozkład płatników podatku dochodowego według wysokości dochodu na podstawie budżetu 2003, Irlandia

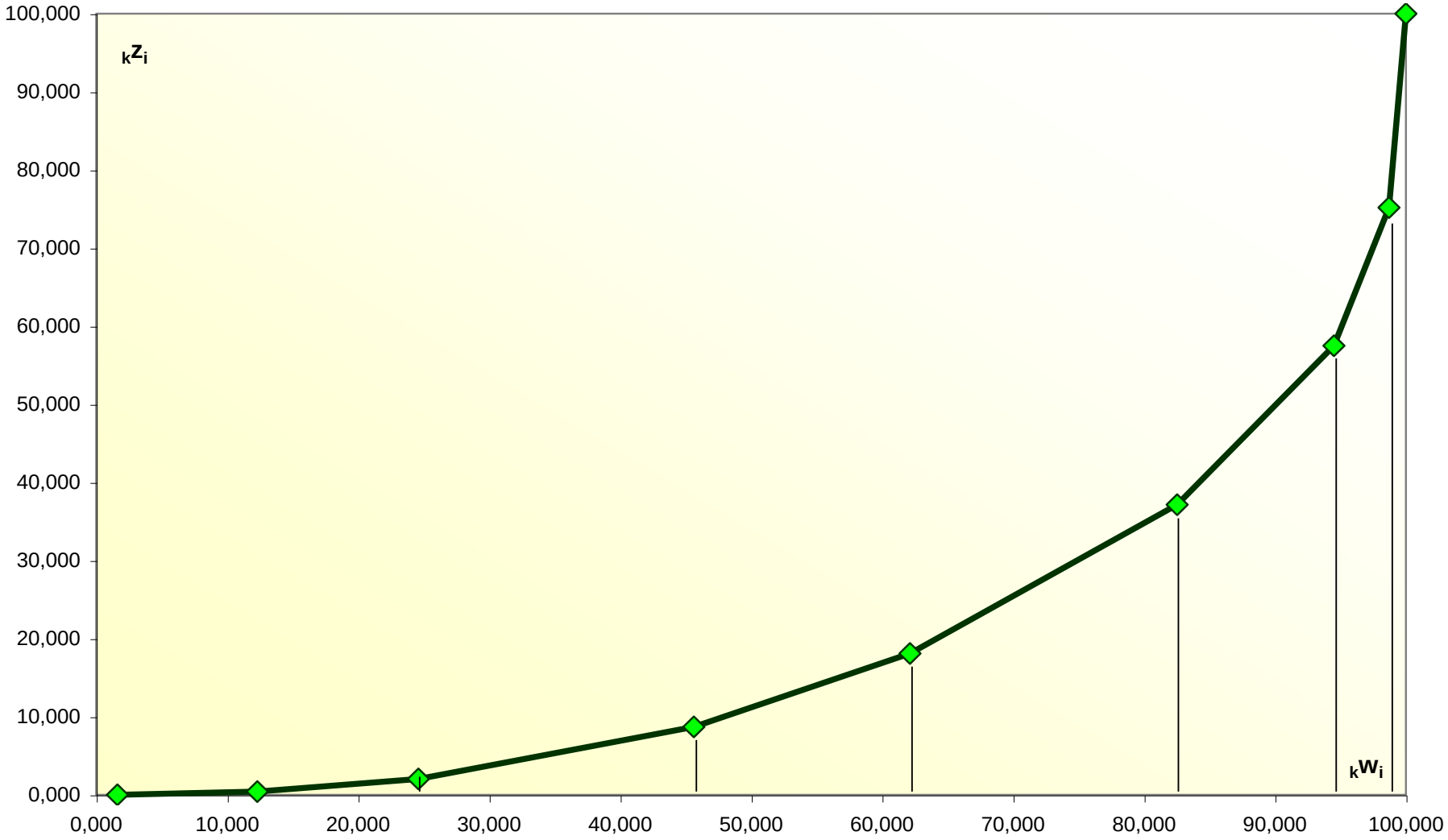
| **Dochód** | | **Liczba płatników (w tys.)** | **Wartość podatku (w tys. £)** | $w_i$ | $z_i$ | ${}_kw_i$ | ${}_kz_i$ | $[({}_kz_i+{}_kz_{i-1})/2]w_i$ | $({}_kz_i+{}_kz_{i-1})w_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 4615 | 5000 | 0,5 | 10 | 1,618 | 0,008 | 1,618 | 0,008 | 0,007 | 0,0130399 |
| 5000 | 7500 | 3,3 | 510 | 10,680 | 0,411 | 12,298 | 0,419 | 2,281 | 4,5613621 |
| 7500 | 10000 | 3,8 | 1990 | 12,298 | 1,604 | 24,595 | 2,023 | 15,014 | 30,028315 |
| 10000 | 15000 | 6,5 | 8270 | 21,036 | 6,665 | 45,631 | 8,687 | 112,645 | 225,2906 |
| 15000 | 20000 | 5,1 | 11660 | 16,505 | 9,396 | 62,136 | 18,084 | 220,925 | 441,84968 |
| 20000 | 30000 | 6,3 | 23670 | 20,388 | 19,075 | 82,524 | 37,159 | 563,148 | 1126,2965 |
| 30000 | 50000 | 3,7 | 25220 | 11,974 | 20,324 | 94,498 | 57,482 | 566,621 | 1133,2416 |
| 50000 | 100000 | 1,3 | 21950 | 4,207 | 17,689 | 98,706 | 75,171 | 279,045 | 558,09008 |
| 100000 | i więcej | 0,4 | 30810 | 1,294 | 24,829 | 100,000 | 100,000 | 113,379 | 226,7589 |
| **Razem** | | **30,9** | **124090** | **100** | **100** | | | **1873,065** | **3746,130** |

Źródło: Obliczenia wlasne na podstawie danych Inland Revenue, T. Bradley, Essential Statistics for Economics, Business and Management, John Wiley & Sons, 2007

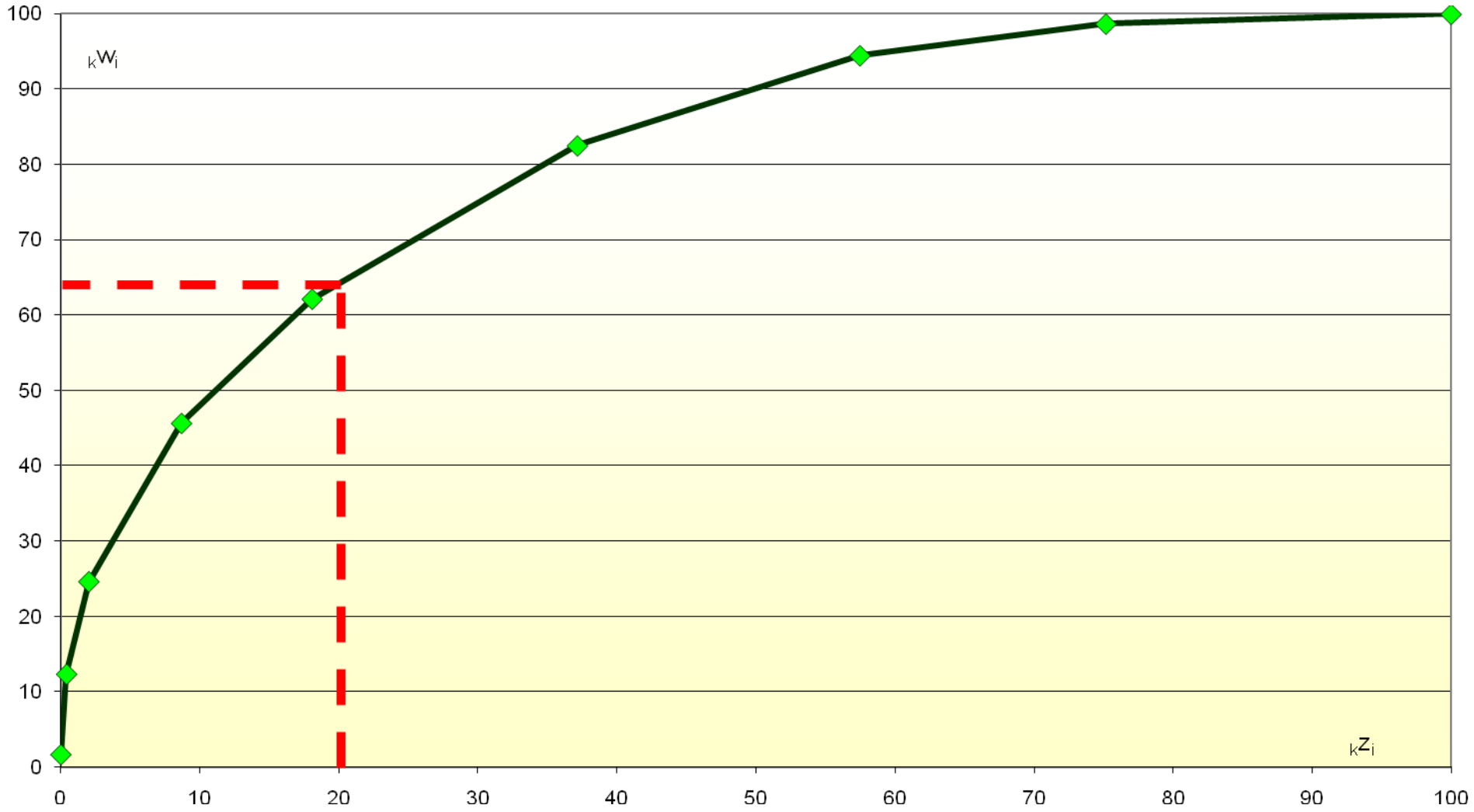
$$G(x)=\frac{5000-\sum_{i=1}^{k}\left(\frac{{}_{cum}z_i+{}_{cum}z_{i-1}}{2}\right)w_i}{5000} \qquad G(x)=\frac{5000-1873{,}065}{5000}=0{,}625$$

$$G(x)=1-\sum_{i=1}^{k}\left[\sum_{s=1}^{i-1}z_s(x)+\sum_{s=1}^{i}z_s(x)\right]w_i \qquad G(x)=10000-3746{,}130=6253{,}8699$$

## Współczynnik Giniego



## Współczynnik Giniego



# PRAKTYCZNA UŻYTECZNOŚĆ WSPÓŁCZYNNIKA GINIEGO

## 1. Human Development Index (HDI) jako syntetyczny miernik jakości życia:

- oczekiwane trwanie życia
- analfabetyzm wśród dorosłej ludności
- współczynnik skolaryzacji
- dochód narodowy na osobę

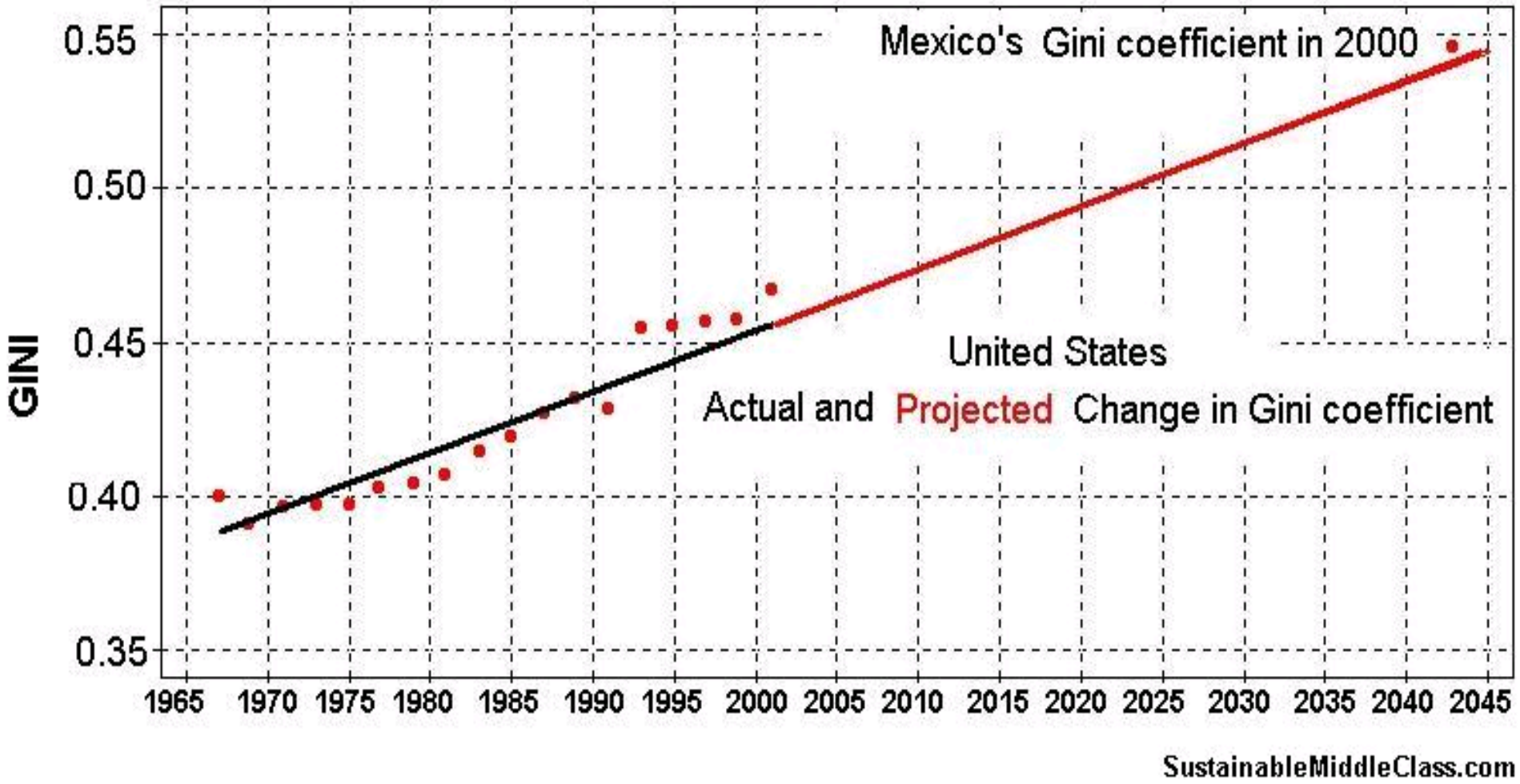
## 2. W praktyce współczynnik Giniego przyjmuje wartości:

- od około 0,2 dla byłych państw socjalistycznych – społeczeństw egalitarnych np. Bułgaria, Węgry, Słowacja, Czechy, Polska
- do ponad 0,6 dla krajów Ameryki Środkowej i Południowej o ukształtowanych elitach gospodarczych

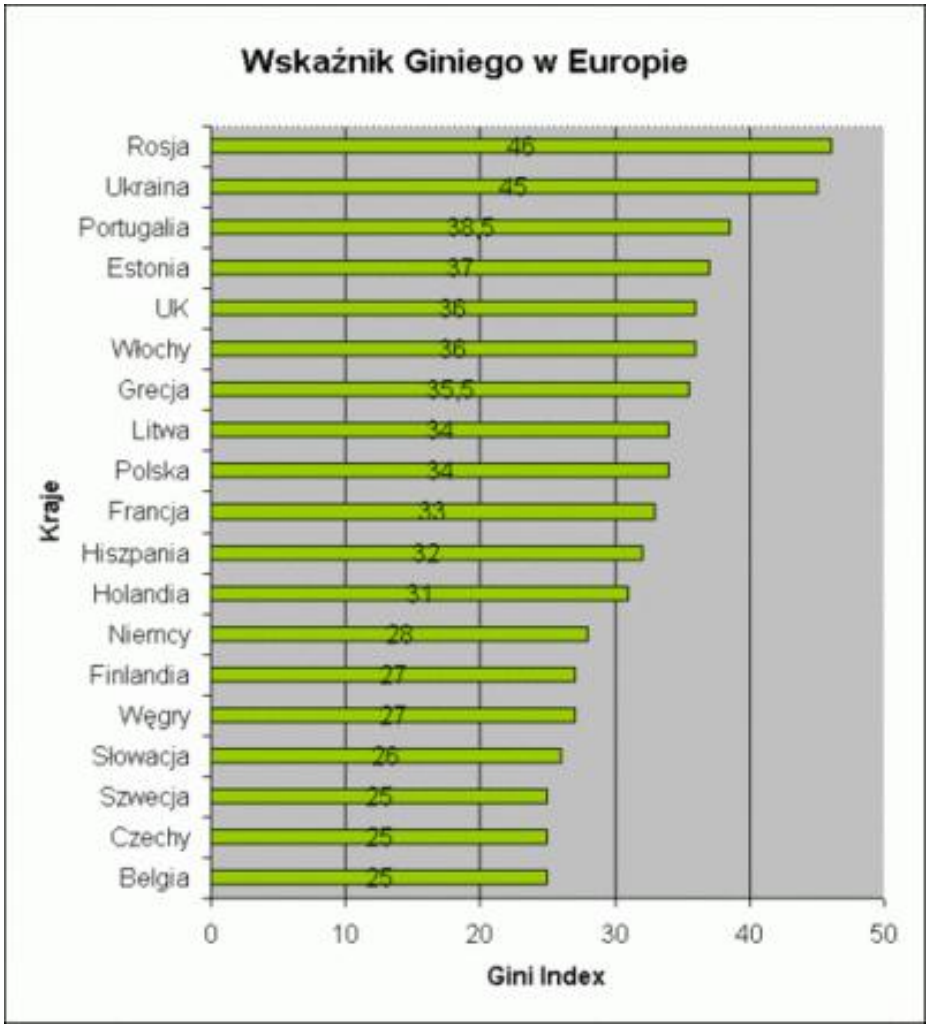
## 3. Analiza wartości współczynnika Giniego jest użyteczna w badaniu trendów zmian.

- Wskazuje np. wzrost równości społecznych na Kubie w latach 1953 do 1986 – G(x) zmniejszył się z 0,55 do 0,22.
- Pogłębianie się nierówności społecznych w USA w ostatnich czterech dekadach, w których zaobserwowano wzrost współczynnika Giniego z 0,35 w latach 70. do 0,45 obecnie, przy dalszym wzroście
- W większości państw europejskich i Kanadzie, współczynnik ten przyjmuje wartość około 0,30.
- W Japonii i niektórych krajach azjatyckich współczynnik Giniego wynosi około 0,35 - 0,40
- W większości krajów afrykańskich G(x) przekracza 0,45.

| Współczynnik Giniego | |
|---|---|

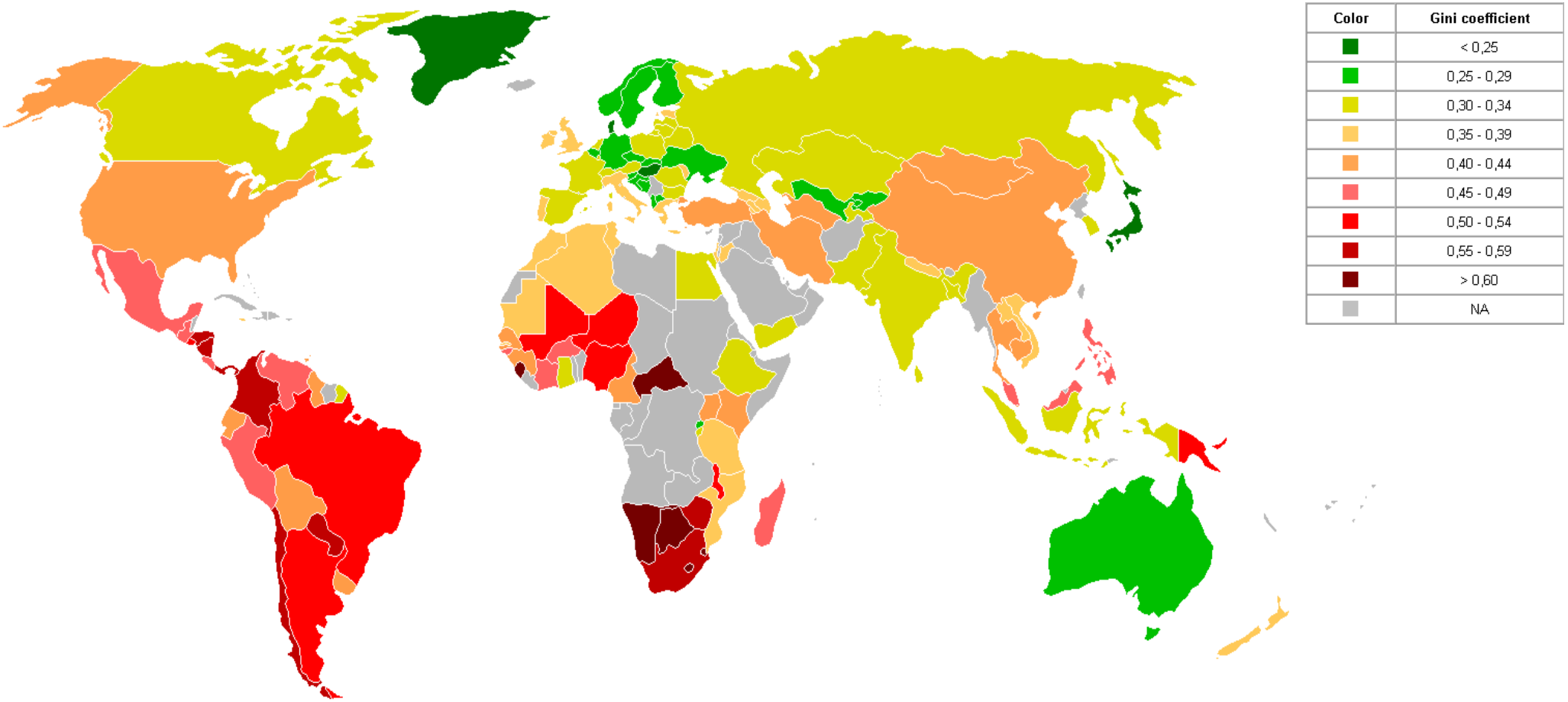


| Współczynnik Giniego | |
|---|---|




**Na podstawie informacji z 2005 roku wiadomo, iż w Polsce wskaźnik Giniego kształtuje się na poziomie 0, 31-0, 33**

## Współczynnik Giniego



# ZASADA PARETO

## Zasada 80/20 głosi, że 80% wyników wypływa tylko z 20% przyczyn

**Schemat leżący u podstaw tej zasady został odkryty w roku 1897, włoskiego ekonomistę**

**Vilfreda Pareto**

O znacznej koncentracji mówi się, gdy 20% jednostek skupia 80% wartości (G(x)>0,64)

- 20% produktów firmy daje jej 80% zysków
- 20% klientów przynosi nam 80% wartości sprzedaży
- 20% kryminalistów popełnia 80% przestępstw
- 20% kierowców powoduje 80% wypadków
- 20% słownictwa wystarczy by móc czytać 80% tekstów w języku
- 20% powierzchni dywanu przypada na 80% zużycia
- 20% ubrań nosimy przez 80% czasu
- 20% naszej pracy daje 80% efektów
- 20% naszego życia daje nam 80% szczęścia

| Współczynnik Giniego | |
|---|---|

# Dziękuję za uwagę