
SEMESTR LETNI 2023

# EKONOMETRIA PRZESTRZENNA

NIEZBĘDNY PODRĘCZNIK –
OD STUDENTA DLA STUDENTA

KANAŁ
STUDENTCKI

# EKONOMETRIA PRZESTRZENNA

## DLACZEGO ZWYKŁA NIE WYSTARCZY?

Dane przestrzenne są znacznie bardziej skomplikowane w swojej strukturze, dlatego nie można ich analizować za pomocą klasycznych metod ilościowych. Wymagają specjalistycznych metod, dzięki którym można uniknąć problemów powstających na skutek wprowadzenia efektów przestrzennych do modelu ekonometrycznego.

## EKONOMETRIA

**Ekonometria zajmuje się pomiarem w gospodarce. Nie mierzy zjawisk, a zależności między zjawiskami.**

- badanie zależności zachodzących w gospodarce i relacji między zjawiskami ekonomicznymi (musimy mieć **przynajmniej dwa zjawiska**)
- tworzenie analiz ekonomicznych
- przewidywanie i stawianie prognoz : „znam zależności, oceniam skutki decyzji”
- narzędzie: **model ekonometryczny**

**Model ekonometryczny** – podstawowe narzędzie analizy, umożliwiające skwantyfikowanie (przedstawienie liczbowo) informacji i zbadanie ich ilościowych zależności; **funkcja, która jest najlepiej dopasowana do danych**.

## EKONOMETRIA PRZESTRZENNA

**Ekonometria przestrzenna zajmuje się modelowaniem ekonomicznych zjawisk, które mają odniesienie przestrzenne.**

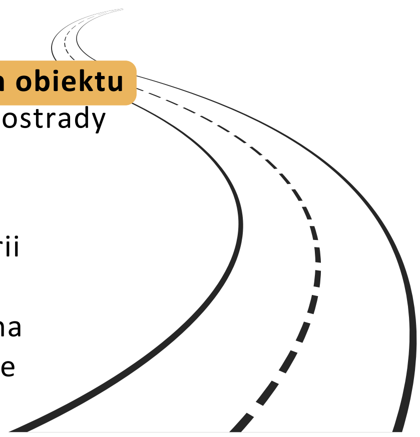
- Wykorzystanie metod ekonometrycznych w analizie ekonomiczno-przestrzennej
- Techniki umożliwiające **uniknięcie problemów metodologicznych** powstających na skutek **obecności przestrzeni w modelu**
- Uwzględnienie większej liczby prawidłowości -> **lepsze prognozy**

**Efekty przestrzeni**

**polepszające stan obiektu**
np. obecność parku, terenów zielonych

**pogarszające stan obiektu**
np. obecność autostrady w pobliżu

W klasycznej ekonometrii obchodzi nas tylko, jak jedno zjawisko wpływa na drugie, ale nie interesuje nas "sceneria".



# EKONOMETRIA PRZESTRZENNA TO...

**EKONOMETRIA**
zapewnia metody i narzędzia (modele, testy, estymacje)

**GEOGRAFIA**
zapewnia aspekt przestrzenny, lokalizację, pracujemy na danych zlokalizowanych

**EKONOMIA**
zapewnia podstawę teoretyczną, modele muszą bazować na teorii, by móc je interpretować

**KARTOGRAFIA**
zapewnia sposób prezentacji, ułatwia interpretacje i uchwycenia prawidłowości

**STATYSTYKA**
zapewnia wspólny aparat pojęciowy oraz pracowanie na zjawiskach masowych (skala -> prawidłowości)

**TAKSONOMIA**
zapewnia opisywanie obiektów (nauka o obiektach), pracujemy na obiektach wielocechowych

# POJĘCIA

**ogół**

**zbiorowość statystyczna** – ogół jednostek biorących udział w badaniu, spełniających dwa kryteria: **jednej cechy różnicującej** i **jednej cechy wspólnej.**

**jednostka obserwacji** – jednostka badawcza, **pojedynczy** element zbiorowości.

**cecha statystyczna** – to, co różnicuje jednostki badawcze, wyróżniamy **ilościowe** (wyrażone liczbowo) i **jakościowe** (wyrażone tekstowo).

**wariant cechy** – **konkretna** odmiana cechy, np. 166 cm.

**obserwacja statystyczna** – nadanie **konkretnych wariantów** (tekstowych lub liczbowych) jednostkom statystycznym, przypisanie wariantu do jednostki.

**szczegół**

**szereg statystyczny** – zbiór wartości obserwacji statystycznych, każda cecha ma nadany odpowiadający jej wariant; **ciąg wariantów cechy statystycznej, uporządkowany według określonego kryterium**

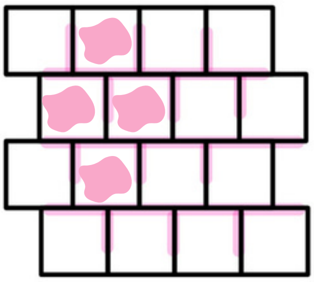
**strukturalne**
sortowanie wg. wartości konkretnej cechy, np. wzrostu dzięki temu obrazują strukturę

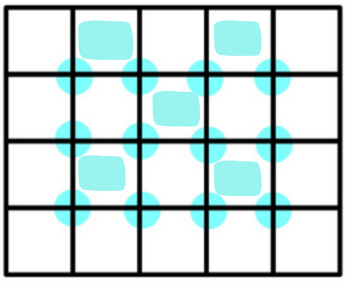
**czasowe**
sortowanie wg. pojawiania się w czasie, dzięki temu obrazują rozwój zjawiska

**przestrzenne**
sortowanie wg. lokalizacji geograficznej, dzięki temu obrazują wymiar

## SĄSIEDZTWO

**FAKTYCZNE**
wspólna granica o niezerowej długości (boki), na pierwszy rzut oka widać, że są obok siebie



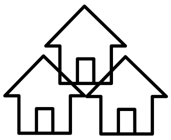


**FORMALNE**
wspólna granica o dowolnej długości, także zerowej (skosy)

# RODZAJE JEDNOSTEK BADAWCZYCH

### NATURALNE

obiekty stanowiące **wyodrębnioną morfologicznie** od otoczenia całość



**Przykład:** zabudowa jednorodzinna naturalnie różni się od blokowisk.

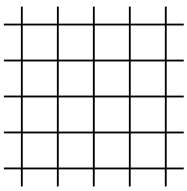
### PSEUDONATURALNE

agregaty jednostek naturalnych wyodrębnione **dla celów administracyjnych**, planistycznych, statystycznych

**Przykład:** województwa, powiaty, gminy itd. "Pseudo" - w końcu Gorzów Wielkopolski jest w lubuskim...

### SZTUCZNE

**regularne figury geometryczne** w sensie kształtu, wielkości i położenia w przestrzeni, wypełniające w całości obszar, którego dotyczy badanie.



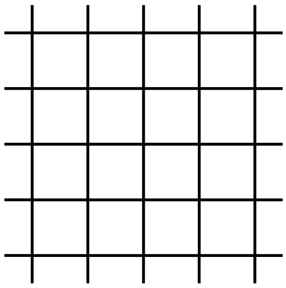
**Przykład:** siatki tworzone na obszarze przy różnych badaniach, szukaniu prawidłowości.

### DOBÓR JEDNOSTKI SZTUCZNEJ - KRYTERIA

- **Łatwość opisania w prostokątnym układzie współrzędnych**
- **Odległość między sąsiadami -** odległość od każdego punktu do każdego punktu powinna być taka sama.
- **Sąsiedztwo faktyczne i formalne -** chcemy jak najmniej formalnego, gdyż trzeba to rozstrzygać samodzielnie, a my chcemy badanie jak najbardziej obiektywne.
- **Stosunek obwodu do powierzchni (minimalna długość granic) -** minimalizujemy ryzyko spornych decyzji o sąsiedztwie.

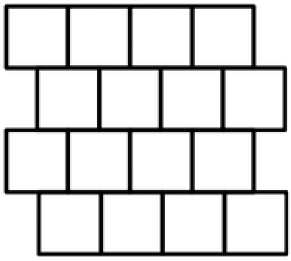
# DOBÓR JEDNOSTKI SZTUCZNEJ - PRZYKŁADY

## SIATKA KWADRATÓW



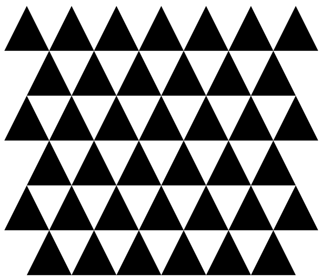
- ✔ łatwość opisania w prostokątnym układzie współrzędnych
- ✘ odległość między sąsiadami
- ✘ sąsiedztwo faktyczne i formalne
- ✔ mała długość granic

## CEGIEŁKA



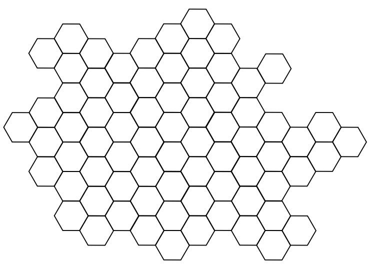
- ✘ nie do opisania w prostokątnym układzie współrzędnych
- ✘ odległość między sąsiadami
- ✔ sąsiedztwo jedynie faktyczne
- ✔ mała długość granic

## SIATKA TRÓJKĄTÓW



- ✘ nie do opisania w prostokątnym układzie współrzędnych
- ✘ odległość między sąsiadami
- ✘ sąsiedztwo faktyczne i formalne
- ✘ duża długość granic

## SIATKA HEKSAGONÓW



- ✘ nie do opisania w prostokątnym układzie współrzędnych
- ✔ odległość między sąsiadami
- ✔ sąsiedztwo jedynie faktyczne
- ✔ minimalna długość granic

**WNIOSEK:**

Najlepsza jest siatka heksagonów, najlepiej realizuje kryteria. Trójkąty wyglądają fancy, ale są do kitu. Niestety - nie da się jej opisać w układzie kartezjańskim, dlatego i tak najczęściej stosuje się kwadraty.

# RODZAJE ZMIENNYCH

ze wzg. na różnicowanie jednostek statystycznych

**STAŁE**
tylko jeden wariant jest odpowiedni

**ZMIENNE**
wiele wariantów możliwych, odpowiednich

ze wzg. na stosunek do badanego zjawiska (względne)

**STYMULANTY**
pożądane wysokie wartości z punktu widzenia ogólnej charakterystyki badanego zjawiska (np. PKB)

**NOMINANTY**
pożądana określony przedział wartości, każde odchylenie jest zjawiskiem niekorzystnym z punktu widzenia danego zjawiska (np. bezrobocie)

**DESTYMULANTY**
pożądane niskie wartości z punktu widzenia ogólnej charakterystyki badanego zjawiska (np. poziom zanieczyszczenia środowiska)

ze wzg na przynależność do danych

**SUROWE**
wyrażone w jednostkach pomiaru (mają miano)

**ZNORMALIZOWANE**
nie posiadają jednostki; wyrażone w formie niemianowanej, **porównywalne**, np. odnoszące się do średniej

ze wzg. na charakter wariantów (mierzalność)

**ILOŚCIOWE**
warianty są liczbami (kwantyfikacja)

**JAKOŚCIOWE**
warianty są tekstem (opisywanie)

rodzaje ilościowych ze wzg. na sposób występowania

**skokowa (dyskretna)**
ściśle określone wartości ze skończonych przedziałów, **kryterium:** pomiędzy dwiema dowolnymi wartościami **nie możemy** znaleźć wariantów pośrednich, np. piętra

**ciągła**
nieskończenie wiele wartości z pewnego naturalnie uwarunkowanego przedziału; **kryterium:** pomiędzy dwiema dowolnymi wartościami **możemy** znaleźć wariant pośredni, np. czas, wzrost

rodzaje ilościowych ze wzg. na sposób wyrażenia wartości

**absolutne**
wyrażane są w naturalnych jednostkach miary (związanych z wielkością a nie natężeniem); możemy je bezpośrednio zmierzyć lub policzyć

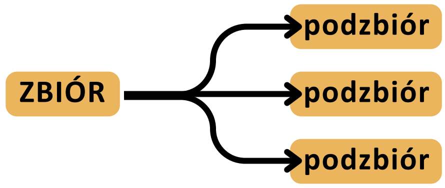
**relatywne**
wyrażane są w postaci relacji dwóch wielkości, z których jedna jest traktowana jako zmienna i odnoszona (relacjonowana) jest do drugiej traktowanej jako wielkość stała np.: na km2, na ha, wartości procentowe, promilowe

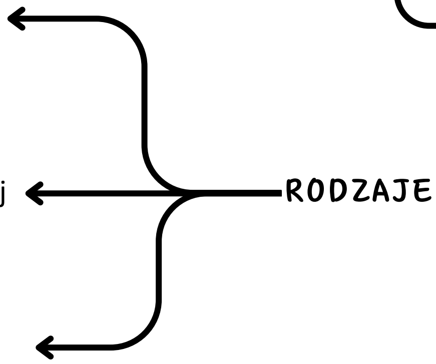
**UWAGA! Musimy zwracać uwagę na wymagania danego współczynnika czy macierzy. Niektóre wymagają jedynie wartości absolutnych, niektóre relatywnych!**

# METODY KLASYFIKACJI PRZESTRZENNEJ

## KLASYFIKACJA PRZESTRZENNA

Podział zbiorowości na podgrupy.




**klasyfikacja jednocechowa**
bierzemy pod uwagę daną cechę określoną dla pewnej liczby obiektów (generalizacja ułatwia prowadzenie analiz).

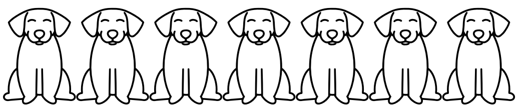
**klasyfikacja pseudowielocechowa**
polega na sprowadzeniu wielu cech do jednej „metacechy” o charakterze syntetycznym.

**klasyfikacja wielocechowa**
polega na analizie większej liczby cech, które opisują złożoność danego problemu.

**Klasyfikację można uznać za poprawną,**
**jeżeli zostaną spełnione wszystkie poniższe kryteria:**

### 1. KRYTERIUM PODOBIEŃSTWA:

- **warunek homogeniczności wewnętrznej** – obiekty należące do tej samej klasy powinny być do siebie jak najbardziej podobne.



- **warunek heterogeniczności zewnętrznej** – obiekty należące do różnych klas powinny być do siebie jak najmniej podobne.

### 2. KRYTERIUM FORMALNE:

- **warunek zupełności (adekwatności) –** każda jednostka badawcza musi być sklasyfikowana i włączona do jednego z podzbiorów **(wszyscy mają domek)**.
- **warunek rozłączności –** wyróżnione podzbiory muszą się wzajemnie wykluczać, jednostka nie może należeć do dwóch grup **(nie można mieć dwóch domków).**
- **warunek niepustości –** żaden z wyróżnionych podzbiorów nie może być pusty **(nie ma domków bez mieszkańców).**

## METODA POSTĘPOWANIA PRZY KLASYFIKACJI

1. **Uporządkować obiekty liniowo** według wartości danej cechy od najmniejszej do największej (posortować rosnąco).
2. **Określić liczbę klas** - na ile klas powinniśmy podzielić naszą zbiorowość, żeby wykryć istotne prawidłowości? Patrz: sposoby określenia liczby klas.
3. **Wydzielić klasy** - przydzielić poszczególne obserwacje do odpowiednich klas, stosując jedną z metod.

**SPOSOBY NA WYDZIELENIE LICZBY KLAS**

$k \approx \sqrt{N}$ $\quad$ $k \approx 1 + 3{,}322 \times \log N$

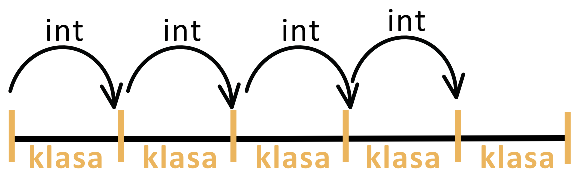
$k \approx \sqrt{\frac{N}{2}}$ $\quad$ $k \le 5 \times \log N$

Wyniki wychodzą zbliżone różnymi metodami, możemy zdecydować indywidualnie, który najbardziej pasuje nam do badań.

## SPOSOBY NA WYDZIELENIE KLAS

### Równe interwały

1. Obliczenie rozstępu **R = Xmax - Xmin**.
2. Obliczenie interwału **I = R/k**.
3. Wyznaczenie granic klas:
   **Xmin + I, Xmin + 2*I ....**




### Równoliczne klasy

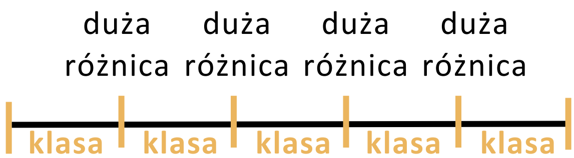
1. **Bazuje na podziale kwantylowym,** dzielącym uporządkowaną zbiorowość statystyczną na określone, **równe części pod względem liczby jednostek** statystycznych.
2. W zależności od przyjętej liczby klas k obliczamy właściwe kwantyle.
3. **Wartości właściwych kwantyli = granice.**


q - dany kwantyl; rodzaje ze względu na ilość klas:

mediana: 2 klasy
kwartyle: 4 klasy
kwintyle: 5 klas
decyle: 10 klas

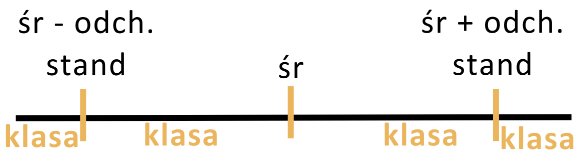
### Największe różnice

1. Obliczamy różnice pomiędzy kolejno po sobie występującymi wartościami.
2. **Miejsca największych różnic = granica.**

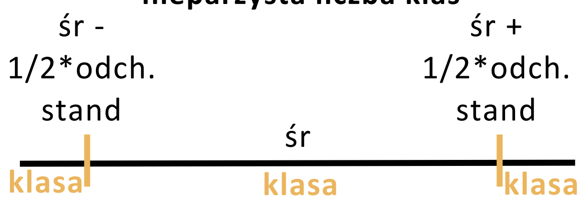



### Średnia i odchylenie standardowe

**parzysta liczba klas**




**nieparzysta liczba klas**




**Arbitralnie** wybieramy kryterium podziału samodzielnie i przydzielamy jednostki.

# NIERÓWNOŚCI PRZESTRZENNE

### POZYTYWNY WPŁYW
motor rozwoju, chęć wyrównania efektywności regionów i powoływanie w tym celu różnych inicjatyw

### NEGATYWNY WPŁYW
początkowe przewagi i słabości z czasem ulegają kumulacji i nierówności zaczynają się pogłębiać, np. drenaż mózgów

## DLACZEGO MIERZYMY NIERÓWNOŚCI??
Wynika to z klasycznych teorii i założeń ekonomicznych, zgodnie z którymi przedmiotem ekonomii jest racjonalna i nastawiona na zysk działalność człowieka. Relacja efektów do nakładów powinna być więc jak najwyższa. Przy pojawieniu się efektywności, pojawiają się nierówności.

**Wymiary nierówności:**
- wymiar ekonomiczny (dochód, zatrudnienie, warunki bytowe),
- wymiar społeczny (wykształcenie, prestiż, styl życia),
- wymiar polityczny (udział we władzy, zaangażowanie obywatelskie),
- wymiar czasowy (cykle koniunkturalne),
- wymiar przestrzenny (zróżnicowanie przestrzenne gospodarki).

## MIARY NIERÓWNOŚCI

### KLASYCZNY WSPÓŁCZYNNIK ZMIENNOŚCI
- miara relatywna - odnosi się do średniej
- może przyjmować wartości większe od 1 - **nie jest unormowany!!**
- **interpretacja:** "odchylenie standardowe stanowi V% średniej arytmetycznej"
- **V=0** – brak zróżnicowania wartości zmiennej
- **V>0** – rośnie zróżnicowanie cechy (rośnie dyspersja)

$$V = \frac{S_x}{\bar{x}}$$

**BUMP: ON NAPRAWDĘ NIE JEST UNORMOWANY!**
**MOŻE PRZEKRACZAĆ 1 (TO JEDNO Z PYTAŃ NA EGZAMIN)**

### WAŻONY WSPÓŁCZYNNIK ZMIENNOŚCI
- ważony ludnością, używany, kiedy porównywane jednostki przestrzenne znacznie różnią się liczbą ludności (np. mazowieckie i opolskie)
- **pi** - liczba ludności w i-tej jednostce przestrzennej (opolskie)
- **P** - liczba ludności całego badanego obszaru (Polska)

ważenie

$$V_W = \frac{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{p_i}{P}}}{\bar{x}}$$

## WSPÓŁCZYNNIK GINIEGO DLA DANYCH PRZESTRZENNYCH

- najpopularniejsza w naukach ekonomicznych miara nierówności
- miara graficzno-liczbowa
- reprezentacją graficzną jest tzw. **Krzywa Lorenza**
- **unormowany na przedziale <0,1>**
- **G* = 0** – rozkład idealnie wyrównany (wszyscy mamy tyle samo pieniążków)
- **G* = 1** – pełna nierówność rozkładu (jedna osoba ma wszystkie pieniążki, a my nie mamy nic)

$$G^* = 1 - \sum_{i=1;k=0}^{N} w_i \left[u_{i(k+1)} + u_{i(k)}\right]$$

**wada:** wrażliwość na liczbę jednostek uwzględnionych w badaniu.

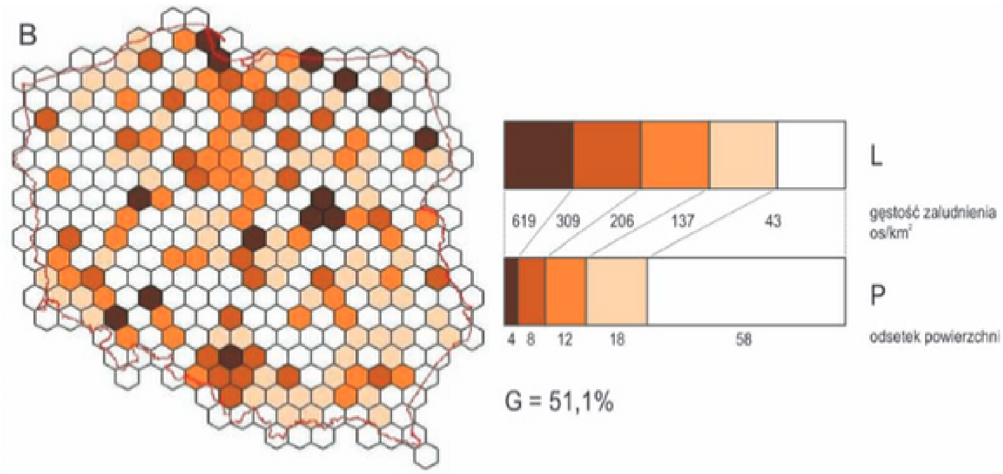
**zaleta:** łatwość interpretacyjna dzięki unormowaniu na przedziale

**Współczynnik Giniego jest unormowany na przedziale <0, 1> !!!**

**UWAGA! Zagrożeniem są zmienne posiadające ujemne wartości.**
**Przy odpowiednim doborze jednostek mogłaby nam wyjść zerowa średnia.**

## MAPA KONCENTRACJI

Mapa koncentracji jest uzupełnieniem Krzywej Lorenza. **Przedstawia przestrzenne progi udziałów badanej cechy w formie kwantyli stopnia koncentracji.** Wartości skumulowanych udziałów analizowanej zmiennej dzieli się na kwintyle (co 20%) a następnie oszukuje się odpowiadające im wartości skumulowanych udziałów zmiennej wagowej.




## KRZYWA LORENZA

- **graficzna interpretacja współczynnika Ginniego**.
- poziom nierówności obrazowany jest przez odchylenie krzywej Lorenza od linii rozkładu równomiernego.

# ETAPY KONSTRUKCJI KRZYWEJ LORENZA:

1.**Nadajemy wagi pi obiektom** - najczęściej jest to liczba ludności.
2. Konstruujemy tabelę pomocniczą z badanymi jednostkami przestrzennymi, np. województwa Polski.
3. Przypisujemy wartości badanej zmiennej **xi** oraz wagi **pi** naszym jednostkom - **muszą być to wartości absolutne!**

| jednostki | xi - cecha, której poziom nierówności badamy | pi - nasza waga, w tym przypadku liczba ludności |
|---|---|---|
| lubuskie | 2086,67 zł | 1,017 miliona |
| wielkopolskie | 1880,91 zł | 3,489 miliona |

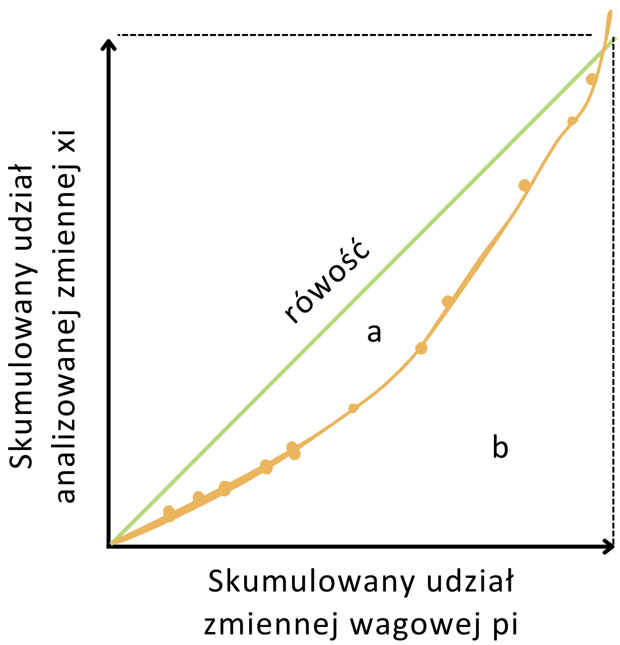
4. **Obliczamy iloraz wskaźnika natężenia** (**xi/pi**) i **sortujemy** badane jednostki rosnąco według jego wartości.
6. **Obliczamy udziały** (**pi/P**) wartości zmiennej wagowej **pi** dla i-tej jednostki przestrzennej w ogólnej sumie wartości zmiennej wagowej **P -** *czyli jak liczba ludności naszego województwa ma się do ludności całej Polski.*
7. **Obliczamy udziały ui** wartości analizowanej zmiennej **xi** dla i-tej jednostki $ui = \frac{xi}{\sum_{i=1}^{N} xi}$ przestrzennej w całkowitej sumie wartości tej zmiennej dla wszystkich jednostek - *czyli jak dochód na osobę w naszym województwie ma się do sumy dochodu na osobę z wszystkich województw.*
8. **Kumulujemy udziały zmiennej wagowej pi oraz zmiennej analizowanej xi** (dodajemy kolejne wartości narastająco). **Kumulacje muszą wyjść równe 1.**
11. Zestawiamy poszczególne pary liczb z szeregów skumulowanych i **wykreślamy jako punkty** na płaszczyźnie prostokątnego układu współrzędnych.
12. Łączymy punkty krzywą. **Otrzymana krzywa jest Krzywą Koncentracji Lorenza.**

## Wykorzystanie krzywej Lorenza do policzenia współczynnika Giniego:

**a** – pole koncentracji
**b** – dopełnienie pola koncentracji do połowy kwadratu
**a + b** – pole połowy kwadratu

**G* = a/(a+b)**

**Im większe pole koncentracji tym większe nierówności.** Linia prosta wskazuje równość, więc każde odchylenie od niej jest oznaką nierówności.



# MODELE WZAJEMNEGO ODDZIAŁYWANIA

## SKĄD TAKA KONCEPCJA??

Na podstawie bezpośrednich obserwacji empirycznych ustalono, że poszczególne zjawiska społeczno-gospodarcze oddziałują na siebie analogicznie do zjawisk fizycznych. Zależność między regionem a regionem przypomina zależności między masami obiektów.

**Carey, 1859**

Im większa jest liczba ludności zgromadzonej na danym obszarze, tym większa będzie tam siła przyciągania.

pozytywny wpływ ludności (licznik)

**Ravenstein, 1885**

Wielkość migracji do ośrodków miejskich zmniejsza się wraz z odległością.

$$Mij = \frac{f(Pi)}{dij}$$

negatywny wpływ odległości (mianownik)

**Young, 1924**

„SIŁA ATRAKCJI”

$$M_{ij} = k\frac{Z_i}{d_{ij}^2}$$

- **Mij** - wielkość migracji z j do i
- **Zi** - siła atrakcji i-tego miejca
- **dij** - ofległość z j do i
- **k** - stała proporcjonalności

**Janowski, 1909**

PRZESTRZENNE PRAWO POPYTU I PODAŻY

Ośrodki przyciągają się proporcjonalnie ze względu na podaż i popyt.

$$f = \frac{m \cdot m_i}{r^2}$$

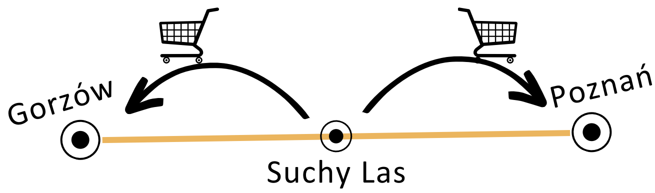
- **f** - siła wzajemnego oddziaływania dwóch miejscowości
- **m** - wielkość popytu/podaży w miejscowości
- **r** - odległość między miejscowościami

**Reilly, 1929, 1931**

## PRAWO GRAWITACJI HANDLU DETALICZNEGO

Pozwala określić proporcję, w jakiej dzielone będą zakupy ośrodka między dwa ośrodki handlu. **Ile osób z miasta c kupuje w mieście a w stosunku do kupujących w mieście b.**

$$\frac{B_a}{B_b} = \frac{P_a}{P_b} \cdot \left(\frac{D_b}{D_a}\right)^2$$




*"Na 100 osób robiących zakupy w Poznaniu, 84 robią zakupy w Gorzowie."*

- **Ba, Bb** - wielkość zakupów w miastach a i b
- **Pa, Pb** - liczba ludności w miastach a i b
- **Da, Db** - odległości z miasta a i b do miejscowości pośredniej c

zasięg oddziaływania

najdalsza odległość, z jakiej ludziom jeszcze chce się jeździć po zakupy do danego miasta

$$D_a = \frac{D_{ab}}{1+\sqrt{\frac{P_b}{P_a}}}$$

## MODEL POTENCJAŁU

- na atrakcyjność danej jednostki wpływa ona sama oraz całe otoczenie dookoła niej
- w modelu potencjału sumujemy wpływy wszystkich jednostek znajdujących się w otoczeniu naszej
- **całkowity potencjał i-tej jednostki przestrzennej to suma jej potencjału oraz wszystkich potencjałów j-tych jednostek przestrzennych**
- np. na potencjał Poznania wpływa sam Poznań, ale też Rokietnica, Suchy Las, Tarnowo Podgórne itd.

$${}_iV_j = G\frac{P_j}{d_{ij}}$$

potencjał w i-tym punkcie wywołany przez masę pj

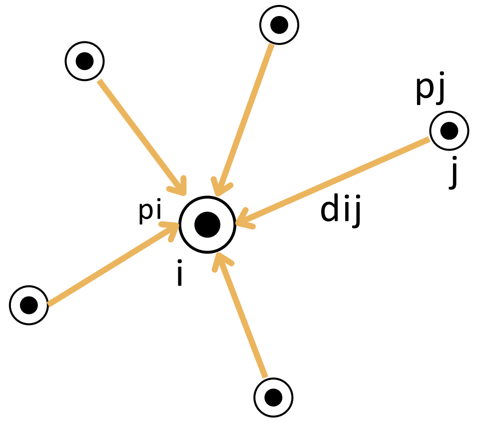
$${}_iV = {}_iV_1 + {}_iV_2 + {}_iV_3 + \ldots + {}_iV_n$$

$${}_iV = G\sum_{j=1}^{n}\frac{P_j}{d_{ij}}$$

całkowity potencjał i-tej jednostki jako suma potencjałów

$${}_iV = G\frac{P_i}{d_{ii}} + G\sum_{i \neq j}^{n}\frac{P_j}{d_{ij}}$$

potencjał własny i-tej jednostki też jest składnikiem sumy



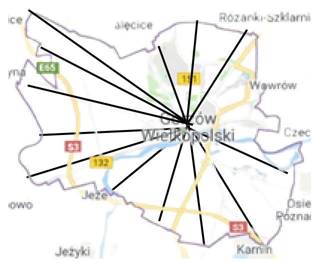

Każda z tych kropeczek ma swój potencjał, który wpływa na potencjał innych.

## PROBLEMY W MODELU POTENCJAŁU

### PROBLEM POTENCJAŁU WŁASNEGO

- dzielimy masę jednostki przez **współczynnik redukcji** (odległość dii - np. z Gorzowa do Gorzowa), który musi być różny od zera (takie są zasady matematyki, nie dzielimy przez zero)
- **I rozwiązanie:** przyjęcie 1 jako dii albo (1+dii)
- **II rozwiązanie (lepsze):**

uwzględnienie rozmiarów jednostki, poprzez wyznaczenie średniej odległości centrum miasta do jego granic




wycznaczenie wartości promienia koła o powierzchni tego miasta

## PROBLEM PRZEJŚCIA Z UKŁADU POWIERZCHNIOWEGO DO PUNKTOWEGO

- badamy jednostki o charakterze powierzchniowym, jednak interesuje nas ich punktowe położenie -> musimy więc przejść na układ punktowy
- **PROBLEM:** jak wybrać punkt reprezentujący jednostkę?
- **I rozwiązanie: centroid** - centralny (średni) punkt

**centroid** przestrzenny odpowiednik średniej arytmetycznej

obszaru

**centroid powierzchniowy** geometryczny środek ciężkości.

$$x_C = \frac{\sum_{i=1}^{n}(m_i x_i)}{\sum_{i=1}^{n} m_i} \qquad y_C = \frac{\sum_{i=1}^{n}(m_i y_i)}{\sum_{i=1}^{n} m_i}$$

średni punkt, czyli współrzędna X-owa i Y-owa

**zaleta:** szybki do policzenia.

**centroid ludnościowy** uwzględnia nieregularne rozmieszczenie ludności na danym obszarze.

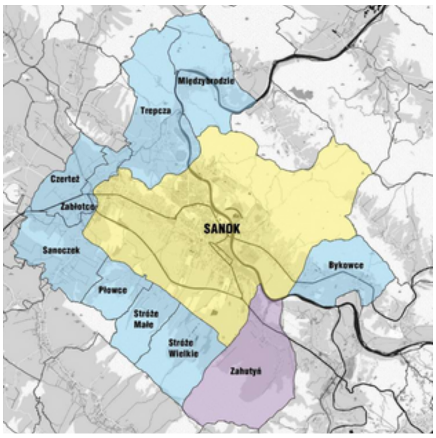
**wada:** musimy mieć szczegółowe dane, a nie zawsze mamy do takich dostęp (dlatego powierzchniowy jest popularniejszy).

**centroid miasta-wsi centralnego** stolica jednostki administracyjnej.

**centroid miasta lub wsi** centrum miasta (wsi) centralnego.

**wada:** nie zawsze stolica jednostki jest dobrym reprezentantem, może mieć mniej mieszkańców i niższy potencjał; problem gmin obwarzankowych.

## OBWARZANEK I MEDOID




**Gmina obwarzankowa**
nieciągła przestrzennie lub posiadająca stolicę na obszarze innej gminy.

**Rozwiązanie: medoid**

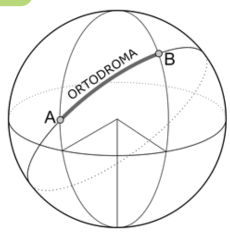
1. Mierzymy odległość od każdego punktu do wszystkich pozostałych.
2. Medoid to ten punkt, którego suma odległości do wszystkich pozostałych jest najmniejsza.

## PROBLEM POMIARU ODLEGŁOŚCI

- Jak mierzyć odległość, by wyniki dobrze odzwierciedlały rzeczywistość?

**METRYKA** **sposób pomiaru odległości**, najczęściej wykorzystujemy dystans fizyczny, może być też czasowy lub ekonomiczny.

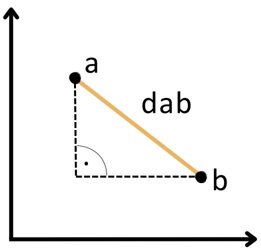
### ortodroma




odległość fizyczna, po łuku

**wada:** zakłada, że Ziemia jest idealną sferą, podczas gdy jej kształt jest bardziej zbliżony do elipsoidy

### metryka euklidesowa (kartezjańska)

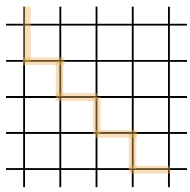



$$d(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

odległość po linii prostej, mniejsze niepewności przy gęstej siatce dróg

**wada:** zawsze mniejsza od faktycznej, nie uwzględnia zróżnicowania terenu

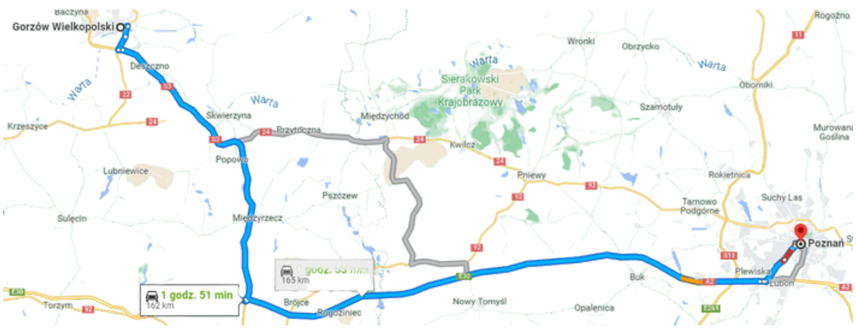
### metryka miejska (Manhattan)



odległość mierzona na siatce równoległych i prostopadłych dróg

**wada:** zawsze większa od faktycznej, ograniczona do prostopadłych ruchów

### metryka drogowa




faktyczna zmierzona droga (np. jak w Google Maps)

**wada:** czasochłonna i wymagająca obliczeniowo - czy warto?

## ZASTOSOWANIA MODELU POTENCJAŁU

- **potencjał ludnościowy** - zyskujemy prawdziwy obraz rozmieszczenia ludności
- **pomiar poziomu peryferyjności** - to, że coś jest daleko, nie oznacza, że jest peryferyjne, ponieważ liczą się też zasoby. **Potencjał = odległość + zasoby** (IKEA a Naramowice)
- **wykrywanie zróżnicowania** dostępności przestrzennej usług (szkolnictwo/ludność)
- **pomiar przestrzennej zmienności** dochodu, popytu i podaży rynkowej
- **miara możliwości zaspokojenia popytu** (potencjał dochodu/potencjał ludności)

## WNIOSKI Z MODELU POTENCJAŁU

- **Nawet gdy dane miejsce nie ma dużego potencjału, ale obiekty dookoła mają, to potencjał tego miejsca znacznie wzrasta**.
- Potencjał jednostek o niskim potencjalne **wzrasta gwałtownie**, gdy znajdują się w pobliżu jednostki z dużym potencjałem.
- **Duże miasta zyskują mniej potencjału** od pomniejszych miejscowości.
- Np. Poznań zwiększa potencjał Rokietnicy. W drugą stronę nie obserwujemy tej zależności z taką intensywnością - nikt nie jedzie do Poznania rykoszetem, bo jego celem jest Rokietnica (sorry, Rokietnica).

## ILORAZ POTENCJAŁÓW

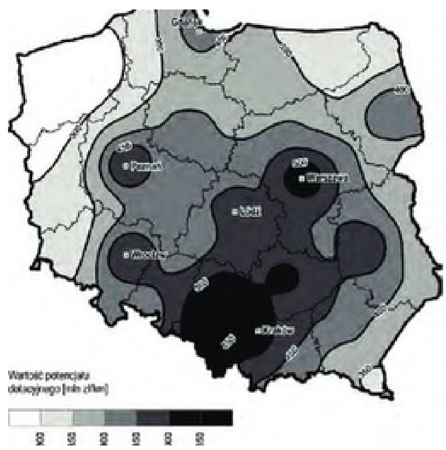
**Stosunek - jak potencjał zjawiska przekłada się na potencjał ludności.** Powstaje poprzez podzielenie potencjału danego zjawiska przez potencjał ludnościowy danej jednostki.

- **Miara systemowa**, uwzględniająca relacje danego miejsca ze wszystkimi innymi miejscami.
- Zależy od rozkładu wartości w całym analizowanym obszarze.
- Zmienna o ciągłym rozkładzie przestrzennym, miara płynna.
- Do obliczeń bierzemy wartości bezwzględne.

**Miara systemowa** wskaźnik lub metryka używana do oceny efektywności, wydajności lub jakości systemu lub procesu.

### METODA POSTĘPOWANIA

1. **Obliczenie potencjału zjawiska Vj** (np. potencjał szkół publicznych).
2. **Obliczenie potencjału demograficznego Vd**.
3. **Obliczenie ilorazu Vi/Vd** (jaki jest stosunek potencjału szkół publicznych do potencjału ludności na danym obszarze).



## MODEL GRAWITACJI

Pozwala na ustalenie wzajemnego przyciągania między miastami - **obciążenie regionu**, jak będzie przebiegać migracja.

$$I_{ij} = G\frac{P_i \times P_j}{d_{ij}^{\ 2}}$$

- **Iij -** siła przyciągania między jednostkami
- **Pi, Pj** - masy jednostek, najczęściej liczba ludności, **wartości absolutne**
- **wi, wj** - wagi jednostek, skłonność do interakcji (najczęściej pieniądze), **wartości relatywne**

$$I_{ij} = G\frac{(w_i P_i)\cdot(w_j P_j)}{d_{ij}^2}$$

## ZASTOSOWANIA MODELU GRAWITACJI

- **Pomiar ciążenia regionów do ośrodka miejskiego** - jak dane miasto przyciąga ludzi z jego okolic. Można sprawdzić czy przyciąganie miast obszarów sąsiednich pokrywa się z wyznaczonymi granicami jednostek administracyjnych.
- pomiar oddziaływania ośrodka miejskiego na zaplecze
- wykrycie powiązań między miastami a regionami
- wyznaczenie zasięgów oddziaływania (zasięgów rynkowych)

## PRZYGOTOWANIE DO MODELU GRAWITACJI I MODELU POTENCJAŁU

1. Ustalamy obszar badania.
2. Ustalamy podstawowe jednostki badawcze.
3. Wybieramy masy jednostek (co będziemy badać, jaką zmienną), np. PKB, liczba ludności, wielkość produkcji, wielkość sprzedaży itd. **UWAGA! Muszą być to wartośći absolutne.**
4. Mierzymy odległości między jednostkami przestrzennymi.
5. Do dzieła! Mamy już wszystko.

# ZALEŻNOŚCI PRZESTRZENNE

**Pierwsze prawo geografii, W. Tobler (1970)** Wszystkie obiekty są ze sobą powiązane, ale bliższe obiekty są bardziej zależne od siebie niż odległe.

↓

**WNIOSEK**
**Zjawiska przestrzenne mają tendencję do grupowania się, są powiązane ze sobą w sposób nielosowy.**

↓

Nie ma niezależnych obserwacji.

↓

Odchodzimy od podstawowych postulatów klasycznej ekonometrii.

↓

**Musimy uwzględnić strukturę powiązań (siłę oraz zależności), dlatego tworzymy macierze wag przestrzennych.**

## MACIERZ WAG PRZESTRZENNYCH

**Macierz wag przestrzennych jest to macierz kwadratowa o wymiarach nxn, której elementy odzwierciedlają strukturę powiązań między danymi obiektami.**

SĄSIEDZTWO – **MACIERZE SĄSIEDZTWA**

ODLEGŁOŚĆ – **MACIERZE ODLEGŁOŚCI**

POWIĄZANIA – **MACIERZE PRZEPŁYWÓW I PRZYNALEŻNOŚCI**

**UWAGA!** Macierze tworzymy według określonych **zasad i jest to jedno z pytań na egzamin.** Elementy głównej przekątnej (Poznań - Poznań, element A - element A), przyjmują wagi równe 0.

## MACIERZE SĄSIEDZTWA

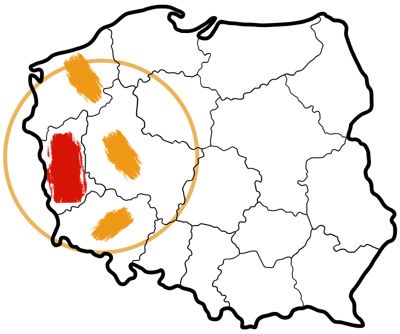
- **czy ktoś jest czy nie jest sąsiadem danej jednostki przestrzennej.**
- Jezeli obiekt **i** jest sasiadem **j-tego,** to do macierzy wstawiamy 1, a jezeli nie sąsiadują - to 0.
- **Pytanie:** jak zdefiniować sąsiedztwo?

$$Wij = \begin{cases} 0 - i,j \text{ nie są sąsiadami} \\ 1 - i,j \text{ są sąsiadami.} \end{cases}$$

# DEFINICJE SĄSIEDZTWA

## 1. Definicja bazująca na pomiarze odległości dzielącej jednostki.

określamy promień wokół jednostki i każda inna jednostka, która znajduje w jego zasięgu jest sąsiadem badanej jednostki.



**Jeżeli jakaś część nie wchodzi w obszar wyznaczony:**

- zaliczamy do sąsiadów, jeżeli przynajmniej połowa powierzchi zalicza się do obszar
- zaliczamy do sąsiadów jeżeli punkt środkowy (centroid) wchodzi w obszar to cały

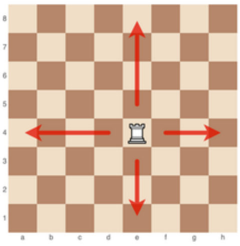
**Problemy:** metryka, arbitralny wybór promienia, decyzja dotycząca jednostek, które tylko częściowo miesza się w danym promieniu.

## 2. Definicja bazująca na przyległości obiektów.

Określamy sąsiedztwo na podstawie tego, z kim jednostka dzieli granice.

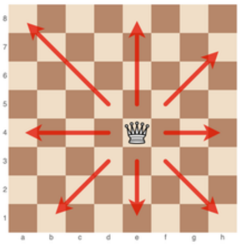
### Macierz Rook

**uznanie jedynie sąsiedztwa faktycznego** (wspólny odcinek o niezerowej długości).



### Macierz Queen

**uznanie również sąsiedztwa formalnego** (dowolny odcinek, również ten o zerowej długości – punkt wspólny).

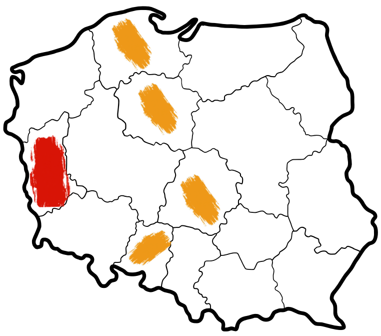


**Problemy:** sąsiedztwo formalne (sąsiedztwo w punkcie).

## 3. Definicja bazująca na sąsiedztwie k-tego stopnia.

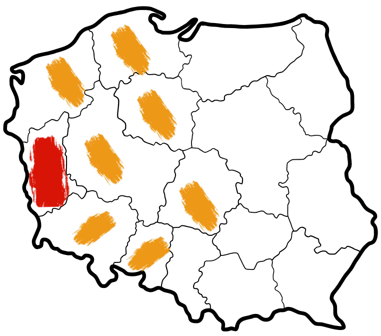
Na podstawie algorytmu k-najbliższych sąsiadów – każda jednostka ma jednakową liczbę sąsiadów.

### k-tego



np. k=2

### do k-tego włącznie



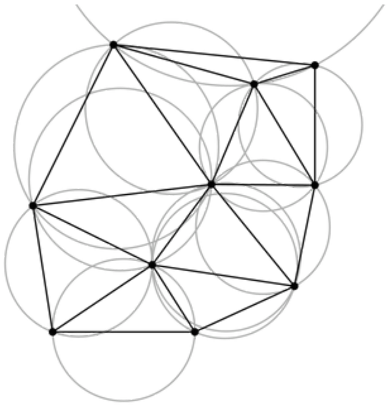
np. do k=2

**Problemy:** sąsiedztwo formalne (sąsiedztwo w punkcie).

### 4. Sąsiedztwo wyznaczone triangulacją Delauny:

Żaden z punktów zbioru nie trafia do wewnątrz okręgu opisanego na trójkącie jakiegokolwiek innego trójkąta powstałego podczas triangulacji: sąsiadami są jednostki, do których możemy dotrzeć krawędziami, obiekty o charakterze punktowym.

1. **Łączymy każdy istotny punkt** (np. sklepy w mieście) tak, aby utworzyć **trójkąty**.
2. Następnie **opisujemy koło** na trójkącie.
3. **Jeżeli jakiś inny punkt wchodzi w obszar koła to znaczy, że triangulacja została źle przeprowadzona.**



Triangulacja poprawna

**Kluczowe punkty nie wchodzą w obszary cudzych kół.** Jeśeli triangulacja została przeprowadzona pomyślnie możemy zacząć wyznaczać sąsiedztwa k-tego stopnia.

**Problemy:** Gęste skupiska punktów mogą prowadzić do generowania małych trójkątów, podczas gdy obszary rzadziej zaludnione mogą prowadzić do większych trójkątów, co wpływa na jakość triangulacji.

---

## MACIERZE ODLEGŁOŚCI

- **mierzenie odległości do wszystkich pozostałych jednostek –** kodowanie odbywa się na zasadzie pomiaru dystansu **d** pomiędzy obiektami **i** i **j**.
- wykorzystujemy funkcję malejących odległości – wagi obliczane są z wykorzystaniem funkcji malejącej wykładniczo lub potęgowo.

$$wij = e^{-\alpha dij}$$

$$wij = dij^{-\alpha}$$

macierz odwrotnej odległości (tzw. inverse distance)

### METODA POSTĘPOWANIA

1. **Określenie używanej <u>metryki (sposobu pomiaru).</u>**
2. **Określenie postaci analitycznej funkcji.**
3. Określenie wartości parametru alfa, określającego siłę spadku powiązań.
4. Wyznaczenie treshold distance - odległość mierzona jest do pewnego progu, wszystko poza progiem równe jest 0.

## MACIERZE PRZEPŁYWÓW I PRZYNALEŻNOŚCI

### Sąsiedztwo wynikające ze wspólnych interakcji (przepływów)

jakie jednostki mają między sobą do czynienia, wagi odzwierciedlają rzeczywiste relacje między jednostkami (np. w zakresie dojazdów do pracy, transferów finansowych, wymiany handlowej itp.)

### Sąsiedztwo wynikające z przynależności do tej samej grupy

o sąsiedztwie nie decyduje bliskość geograficzna a przynależność do tej samej grupy/klasy (np. typy funkcjonalne miast)

## STANDARYZACJA

### Standaryzacja wag wierszami

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$$

- każdy z elementów macierzy przyjmuje wartości z przedziału <0,1>
- elementy każdego wiersza sumują się do 1

### Standaryzacja wag globalnie

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

- każdy z elementów macierzy przyjmuje wartości z przedziału <0,1>
- wszystkie elementy sumują się do 1

## WYGŁADZANIE PRZESTRZENNE ZMIENNEJ

- uzyskanie zmiennej o lepszych (bardziej stabilnych i odszumionych) wartościach.
- zapożyczanie w tym celu informacji z regionów sąsiednich

### lokalnie ważona średnia (smooth)

Średnia arytmetyczna wartości **xj** zmiennej X dla badanego obiektu i obiektów z nim sąsiadujących, obecnych w macierzy wag.

$$smooth(x_i) = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} w_{ij}}$$

**UWAGA! Do policzenia smooth zmieniamy zerowe wartości głównej przekątnej macierzy wag na 1!**

## ZASADY MACIERZY WAG

- **zastosowanie (dowolnej) macierzy wag przestrzennych jest lepszym rozwiązaniem niż przyjęcie założenia o niezależności jednostek w przestrzeni**
- **elementy głównej przekątnej (Poznań - Poznań, element A - element A), przyjmują wagi równe 0**
- w badaniach przestrzennych należy brać pod uwagę relatywnie dużą próbę obserwacji (np. n > 60)
- lepiej jest stosować macierze wag o zbyt małej liczbie powiązań niż zbyt dużej
- modele (i macierze) o prostszej formie (np. sąsiedztwo niższego rzędu) są bardziej preferowane od tych bardziej skomplikowanych (np. sąsiedztwo wyższego rzędu)
- w przypadku regularnych siatek (jednostek sztucznych) liczba sąsiadów danej jednostki powinna się mieścić w przedziale od 4 do 6
- **należy wybrać tę macierz wag przestrzennych, która maksymalizuje wartość współczynnika I Morana lub gwarantuje uzyskanie najwyższego dopasowania modelu ekonometrycznego do danych empirycznych**
- należy unikać macierzy wag przestrzennych ze zbyt dużą liczbą powiązań z uwagi na ograniczony zasięg interakcji

# AUTOKORELACJA PRZESTRZENNA

**Pierwsze prawo geografii, W. Tobler (1970)**

Wszystkie obiekty są ze sobą powiązane, ale bliższe obiekty są bardziej zależne od siebie niż odległe.

**WNIOSEK**

**Zjawiska przestrzenne mają tendencję do grupowania się, są powiązane ze sobą w sposób nielosowy.**

Przestrzenne grupowanie się obiektów o podobnych wartościach.

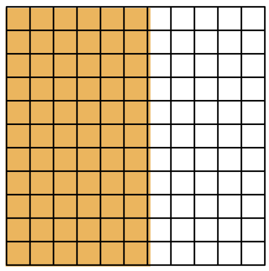
**Autokorelacja przestrzenna**

## AUTOKOREALCJA PRZESTRZENNA

**Określenie stopnia skorelowania obserwowanej wartości zmiennej w danej lokalizacji z wartością tej zmiennej w innej lokalizacji.**

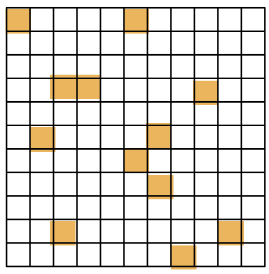
**Autokorelacja dodatnia**

**Wysokie wartości z wysokimi, niskie z niskimi**, np. wzrost PKB i wzrost zatrudnienia.



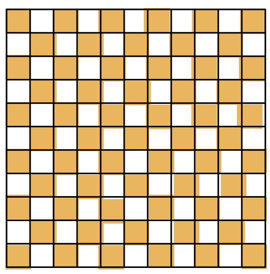
**Brak autokorelacji**

**Losowość, dana wartość nie wpływa na drugą.**



**Autokorelacja ujemna**

**Wysokie wartości z niskimi, niskie z wysokimi**, np. bezrobocie i wzrost gospodarczy.



## GLOBALNA STATYSTYKA I-MORANA

Bada, czy wartości zmiennej są skorelowane przestrzennie na całym obszarze analizy.

**n** – liczba obiektów przestrzennych
**xi, xj** – wartości zmiennej X dla obiektów i oraz j
**$\bar{x}$** – średnia arytmetyczna wartość zmiennej X
**wij** – elementy przestrzennej macierzy wag

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

**Potrzebujemy:** wartości zmiennej xji, **wiedzę o sąsiedztwie** z macierzy wag przestrzennych wij.

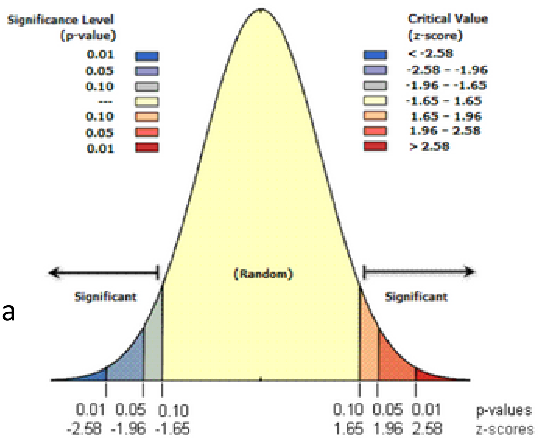
**UWAGA! Nie zaleca się przeprowadzania analizy Morana dla obiektów nie posiadających sąsiedztwa** (obiektów opisanych w macierzy wag wyłącznie wartością 0). Obiekty takie można wykluczyć z analizy dezaktywując je, lub przeprowadzić analizę wybierając inny sposób definiowania sąsiedztwa (inną macierz wag). **Bez sąsiadów nasza jednostka nie ma z kim korelować!**

### Istotność statystyczna

**H0: brak autokorelacji**

**Statystyka testowa:**

$$Z_I = \frac{I - E(I)}{\sqrt{Var(I)}} \qquad E(I) = -\frac{1}{n-1}$$




**I > 0, Z > 0** – dodatnia autokorelacja przestrzenna (podobne wartości są blisko siebie)

**I < 0, Z < 0** – ujemna przestrzenna autokorelacja (podobne wartości są oddalone od siebie)

**I =~ 0, Z =~ 0** – brak autokorelacji przestrzennej

**Wartości statystyki I-Morana są najczęściej z przedziału <-1,1>!!!**

## OPÓŹNIENIE PRZESTRZENNE (SPATIAL LAG)

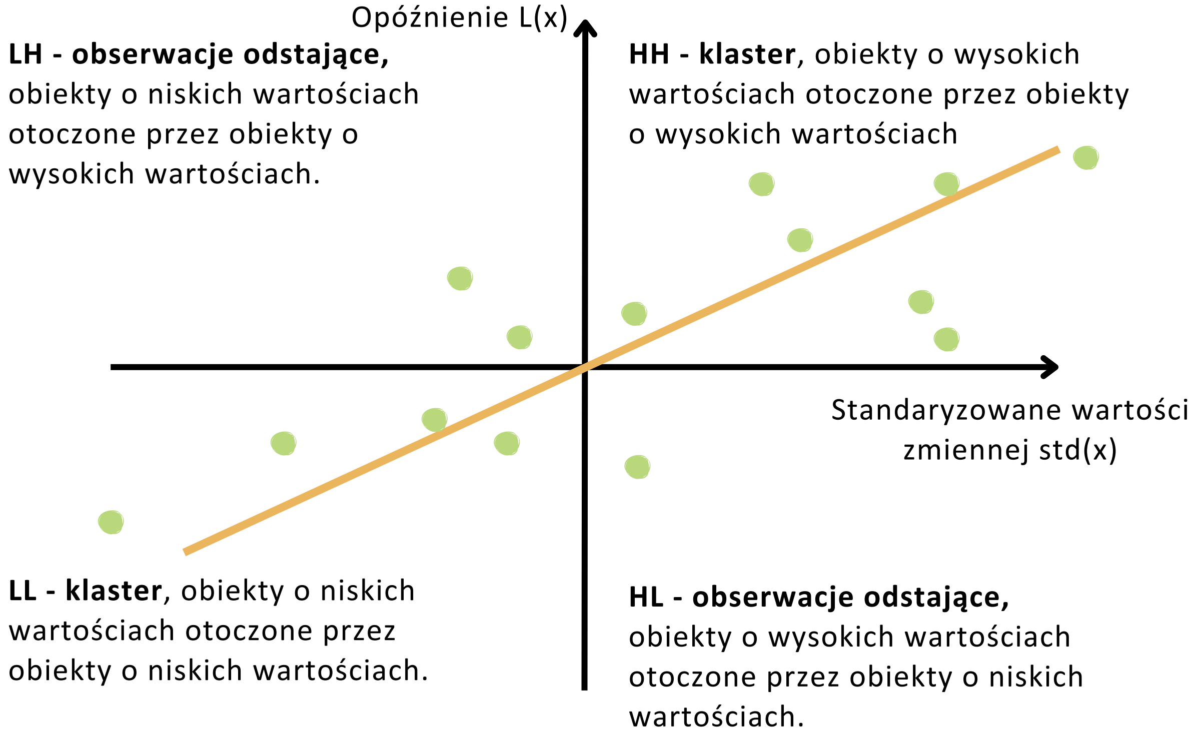
Średnia ważona ze standaryzowanych wartości dla sąsiadujących obiektów.

**n** – liczba obiektów przestrzennych
**xj** – wartość zmiennej X dla obiektu j
**x** – średnia arytmetyczna wartość zmiennej X
**Sx** – odchylenie standardowe zmiennej X
**wij*** – elementy przestrzennej macierzy wag (standaryzowanej wierszami)

$$L(x_i) = \sum_{j=1}^{n} w_{ij}^{*} std(x_j)$$

$$std(x_j) = \frac{x_j - \bar{x}}{S_x}$$

## Moranowski wykres rozporoszenia




### INTERPRETACJA

**AUTOKORELACJA UJEMNA** -
Punkty rozłożone są głównie w ćwiartce drugiej (LH) i czwartej (HL).
**AUTOKORELACJA DODATNIA** -
Punkty rozłożone głównie w ćwiartce pierwszej (HH) i trzeciej (LL)
**BRAK AUTOKORELACJI** -
Punkty rozłożone równomiernie we wszystkich czterech ćwiartkach.

**Klastry** zbiory punktów (lub jednostek) o podobnych cechach lub charakterystykach, które są bardziej podobne do siebie nawzajem niż do punktów z innych klastrów.

## LOKALNA STATYSTYKA I-MORANA

Analizuje podobieństwo wartości zmiennej w przestrzeni dla poszczególnych jednostek analizowanych. W odróżnieniu od globalnej statystyki I-Morana, dostarcza informacji na poziomie lokalnym, wskazując, które jednostki wykazują istotne lokalne skupiska wartości.

**n** – liczba obiektów przestrzennych xi
**xj** - wartości zmiennej X dla obiektów i oraz j
**$\bar{x}$** – średnia arytmetyczna wartość zmiennej X
**wij** – elementy przestrzennej macierzy wag

$$I_i = \frac{(x_i-\bar{x})\sum_{j=1}^{n} w_{ij}(x_j-\bar{x})}{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

## INTERPRETACJA

**Istotne statystycznie obiekty High-High** (obiekty o wysokich wartościach otoczone przez obiekty o wysokich wartościach) zaznaczone są na mapie kolorem czerwonym;

**Istotne statystycznie obiekty Low-Low** (obiekty o niskich wartościach otoczone przez obiekty o niskich wartościach) zaznaczone są na mapie kolorem niebieskim;

**Istotne statystycznie obiekty Low-High** (obiekty o niskich wartościach otoczone przez obiekty o wysokich wartościach) zaznaczone są na mapie kolorem jasno-niebieskim;

**Istotne statystycznie obiekty High-Low** (obiekty o wysokich wartościach otoczone przez obiekty o niskich wartościach) zaznaczone są na mapie kolorem jasno-czerwonym.

## GLOBALNA STATYSTYKA C-GEARY'EGO

jedna z możliwych alternatyw dla statystyki globalnej Morana. Podobnie jak analiza Morana bada ona stopień intensywności danej cech xij w obiektach przestrzennych opisanych za pomocą macierzy wag o elementach wij.

**n** – liczba obiektów przestrzennych xi
**xj** - wartości zmiennej X dla obiektów i oraz j
**$\bar{x}$** – średnia arytmetyczna wartość zmiennej X
**wij** – elementy przestrzennej macierzy wag

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

**UWAGA! Nie zaleca się przeprowadzania analizy Morana dla obiektów nie posiadających sąsiedztwa** (obiektów opisanych w macierzy wag wyłącznie wartością 0). Obiekty takie można wykluczyć z analizy dezaktywując je, lub przeprowadzić analizę wybierając inny sposób definiowania sąsiedztwa (inną macierz wag). **Bez sąsiadów nasza jednostka nie ma z kim korelować!**

### Istotność statystyczna

**H0: brak autokorelacji**

**Statystyka testowa:**

$$Z_C = \frac{C - E(C)}{\sqrt{Var(C)}} \qquad E(C) = 1$$

**Wartości statystyki C są z pzedziału <0, 2>!!!**

**0 < C < 1, Z > 0** dodatnia autokorelacja przestrzenna (podobne wartości są blisko siebie)

**1 < C < 2, Z < 0** ujemna przestrzenna autokorelacja (podobne wartości są oddalone od siebie)

**C ≈ 1, Z ≈ 0** brak autokorelacji przestrzennej

## LOKALNE STATYSTYKI GEARY'EGO

Wynik tej statystyki dla każdej jednostki wskazuje, czy dana jednostka jest częścią skupiska wartości podobnych do otoczenia (autokorelacja dodatnia), czy różni się od swojego otoczenia (autokorelacja ujemna). Identyfikacja obszarów HH, LL, HL, LH.

**n** – liczba obiektów przestrzennych xi
**xj** - wartości zmiennej X dla obiektów i oraz j
**x̄** – średnia arytmetyczna wartość zmiennej X
**zi, zj** - odchylenia zmiennej x od średniej
**wij** – elementy przestrzennej macierzy wag

$$C_i = \sum_{j=1}^{n} w_{ij}(z_i - z_j)^2$$

$$z_i = x_i - \bar{x} \qquad z_j = x_j - \bar{x}$$

## OGÓLNA STATYSTYKA GETISA I ORDA

Umożliwia wykrywanie lokalnej koncentracji wartości wysokich (hot spots) i niskich (cold spots) w sąsiadujących obiektach oraz bada istotność statystyczną tej zależności.

**Założenia:**
- sąsiedztwo zdefiniowane na podstawie **odległości dzielącej jednostki**
- **d –** dystans, w obrębie którego spodziewane jest pojawienie się skupień (korelacji)
- **wij(d) –** elementy przestrzennej macierzy wag macierz wag z elementami binarnymi 0-1 (bez standaryzacji)

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j} \qquad i \neq j$$

**Istotność statystyczna**

**H0: brak autokorelacji**

**Statystyka testowa:**

$$Z_G = \frac{G - E(G)}{\sqrt{Var(G)}} \qquad E(G) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}{n(n-1)}$$

**G > E(G)** grupowanie się wysokich wartości (hot spots).

**G < E(G)** grupowanie się niskich wartości (cold spots).

## LOKALNA STATYSTYKA GETISA I ORDA

Umożliwia wykrywanie lokalnej koncentracji wartości wysokich (hot spots) i niskich (cold spots) w sąsiadujących obiektach oraz bada istotność statystyczną tej zależności.

**Założenia: te same,**

$$G_i = \frac{\sum_{j \neq i}^{n} w_{ij}(d)x_j}{\sum_{j \neq i}^{n} x_j} \qquad G_i^* = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}$$

**Istotność statystyczna**

**H0: brak autokorelacji**

**Statystyka testowa:**

$$Z_{G_i} = \frac{G_i - E(G_i)}{\sqrt{Var(G_i)}} \qquad Z_{G_i^*} = \frac{G_i^* - E(G_i^*)}{\sqrt{Var(G_i^*)}}$$

$$E(G_i) = \frac{\sum_{j \neq i}^{n} w_{ij}(d)}{n-1} \qquad E(G_i) = \frac{\sum_{j=1}^{n} w_{ij}(d)}{n}$$

**Duże Z** grupowanie się wysokich wartości (hot spots).

**Niskie Z** grupowanie się niskich wartości (cold spots).

**Z koło 0** brak grupowania się jednostek.

# REGRESJA PRZESTRZENNA

## DLACZEGO AUTOKORELACJA PSUJE????????

Składnik losowy przestaje być losowy, kiedy zaczyna zależeć od rozmieszczenia w przestrzeni. Gdy występuje autokorelacja przestrzenna, nasze oszacowania stają się nic nie warte. **Przykład:** jeśli podczas egzaminu usiedliśmy obok kogoś z dobrymi ocenami, to super. Gorzej, jeśli usiądziemy obok kogoś, kto niezbyt się nauczył. **Gdy nasz wynik zależy od naszego otoczenia, a nie od naszego MODELU nauki, to nasze przygotowanie staje się nic nie warte.**

**Autokorelacja w modelu może dotyczyć:**

- **Zmiennej zależnej - y wpływa na y**
  wartości zmiennej Y z innych lokalizacji wpływają na kształtowanie się tej zmiennej w innych lokalizacjach.
- **Zmiennych niezależnych - x wpływają na y**
  na wartość zmiennej Y z danej lokalizacji wpływają wartości zmiennych X z innych lokalizacji.
- **Składnika losowego - los wpływa na y**
  gdy w modelu pominięto lub nie można uwzględnić pewnych zmiennych przestrzennie autoskorelowanych.

### DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH

**mierzalne**
możemy je precyzyjnie i obiektywnie zmierzyć, nie chcemy zakłamanych wartości

**o charakterze relatywnym**
opisują różnice między obserwacjami

**kompletne**
nie brakuje nam żadnych danych, wszystkie obserwacje powinny zawierać wartości dla wszystkich zmiennych

**stosunkowo silnie skorelowane ze zmienną objaśnianą**
skoro chcemy ich użyć do tłumaczenia zmiennej objaśnianej, muszą być z nią powiązane

**wewnętrznie zróżnicowane**
nasze zmienne mają różne wartości w obrębie zbioru danych. Gdyby wszystkie obserwacje miały te same wartości, nie wniosłyby istotnej informacji do modelu

**stosunkowo słabo skorelowane z innymi zmiennymi objaśniającymi**
nie potrzebne nam dwie zmienne tłumaczące zmienną objaśnianą w podobny sposób

**WNIOSEK:**
Jeżeli w którychś modelach wyjdzie nam autokorelacja, to musimy użyć modeli regresji przestrzennej. Modele regresji przestrzennej wprowadzają zmienną W - opóźnienie przestrzenne (średnia z wartości y dla sąsiadów).

# RODZAJE MODELI

## MODEL AUTOREGRESJI PRZESTRZENNEJ

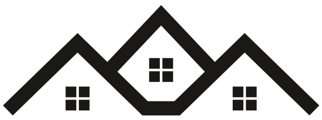
**Wartości y z sąsiedztwa wpływają na to, co się dzieje u nas, wartości x nie mają wpływu.**

Mamy dodatkową zmienną $\boldsymbol{Wy}$ z parametrem **ρ (współczynnik autoregresji przestrzennej),** obrazującymi siłę wpływu zmiennych **y** z sąsiedztwa na naszą.

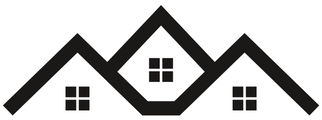
$$y = \rho W y + X\beta + \varepsilon$$

$$\rho\left(\sum_{s=1}^{n} w_{rs} y_s\right)$$

**Przykład:** Mieszkanie w drogiej okolicy jest droższe, niezależnie od wartości zmiennych typu wielkość. Po prostu ceny mieszkań dookoła są wysokie.



**Gorzów**
30m2
300k



**Warszawa**
20m2
1,5 mln

## MODEL BŁĘDU PRZESTRZENNEGO

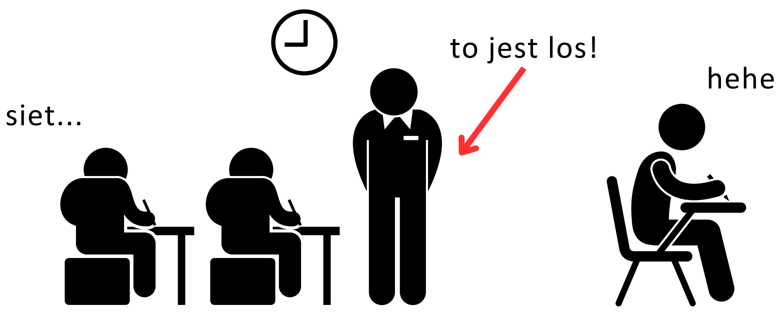
**Mamy jakiś czynnik, który jest nie do przewidzenia, ale wpływa na pewien obszar naszej zbiorowości.**

Dochodzi nam składnik losowy **u**, uwzględniający macierz wag przestrzennych W.

$$y = X\beta + u, \quad u = \lambda W u + \varepsilon$$

$$\lambda\left(\sum_{s=1}^{n} w_{rs} u_s\right)$$

**Przykład:** Wykładowca podczas egzaminu postanawia stanąć akurat przy nas...



## MODEL REGRESJI KRZYŻOWEJ

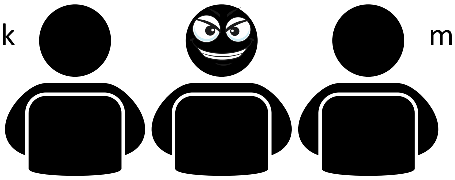
**Cechy naszych sąsiadów, wpływa na nasze wartości. Wartości x wpływają na naszą wartość y, wartości y sąsiadów nie mają wpływu.**

Opóźniamy macierzą W nasze zmienne X - (nie musimy opóźniać wszystkich).

$$y = X\beta + WX * \gamma + \varepsilon$$

**Przykład:** Znajomi obok których siedzę umieją na egzamin (wysokie x=wiedza), to mój wynik y będzie wyższy.

mądry ziomek



mądra ziomka

## PRZESTRZENNY MODEL DURBINA

**autoregresja przestrzenna + autoregresja krzyżowa**

$$y = \rho Wy + X\beta + WX\gamma + \varepsilon$$

Czyli nasza lokalizacja (położenie) ma znaczenie oraz wartości zmiennych naszych sąsiadów. Liczba zmiennych egzogenicznych działających lokalnie (X) oraz z innych lokalizacji X* może być różna.

**Przykład:** Wszyscy uczniowie w klasie dostali wysokie oceny (y) i są mądrzy (x), więc nauczyciel lepiej ocenia moją pracę (jakbym też był mądry).

## PRZESTRZENNY MODEL DURBINA DLA BŁĘDU

**autoregresja błędu przestrzennego + autoregresja krzyżowa**

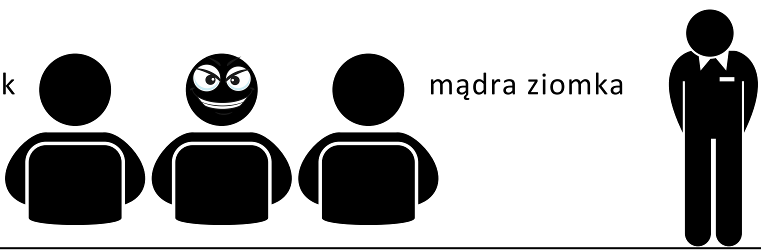
$$y = X\beta + WX\gamma + u$$

Czyli nieprzewidziane sytuacje mają znaczenie oraz wartości zmiennych naszych sąsiadów. Liczba zmiennych egzogenicznych działających lokalnie (X) oraz z innych lokalizacji X* może być różna.

**Przykład:** Stoi nad nami nauczyciel podczas egzaminu i wszyscy studenci dookoła są mądrzy.

to jest los!

mądry ziomek



mądra ziomka

## PROCEDURA DOBORU MODELU

Metodę Najmniejszych Kwadratów możemy zastosować do modelu regresji liniowej lub do modelu regresji krzyżowej.

1. Zaczynamy do modelu regresji liniowej MNK.
2. Testowanie autokorelacji składnika losowego.
3. Testy diagnostyczne mnożnika Lagreange'a.
4. Wybór modelu regresji przestrzennej.

**Testy pozwalają odpowiedzieć na pytanie, jaki model regresji przestrzennej powinniśmy zastosować: model błędu przestrzennego czy model autokorelacji przestrzennej.**

## OCENA DOBROCI DOPASOWANIA MODELU

**model klasyczny**

WSPÓŁCZYNNIK DETERMINACJI R2

DOPASOWANY WSPÓŁCZYNNIK DETERMINACJI ADJR2

**model przestrzenny**

PSEUDO-R2
nie jest unormowany!!

LOGARYTM WIARYGODNOŚCI
stymulanta - wysokie wartości świadczą o lepszym modelu

AIC
destymulanta - niskie wartości świadczą o lepszym modelu

SC
destymulanta - niskie wartości świadczą o lepszym modelu

**kryteria informacyjne**
nie są unormowane, służą do porównywania modeli między sobą


# POWODZENIA! ŻYCZYMY MIŁEJ NAUKI <3

ALEKSANDRA BENOWSKA
MAŁGORZATA TALAGA

**WSZYSTKIE ŹRÓDŁA PODLINKOWANE W MIEJSCACH UŻYCIA + PREZENTACJE PANA DR WOJCIECHA KISIAŁY :)**

KANAŁ STUDENTCKI