



POLSKIE TOWARZYSTWO  
STATYSTYCZNE

Projekt realizowany  
z Narodowym Bankiem Polskim  
w ramach programu edukacji ekonomicznej

NBP

Narodowy Bank Polski

# SIGMA KWADRAT

# LUBELSKI KONKURS STATYSTYCZNO- DEMOGRAFICZNY



**UMCS**  
UNIWERSYTET MARII CURIE-SKŁODOWSKIEJ  
W LUBLINIE



**POLSKIE TOWARZYSTWO  
STATYSTYCZNE**

Projekt realizowany  
z Narodowym Bankiem Polskim  
w ramach programu edukacji ekonomicznej

**NBP**

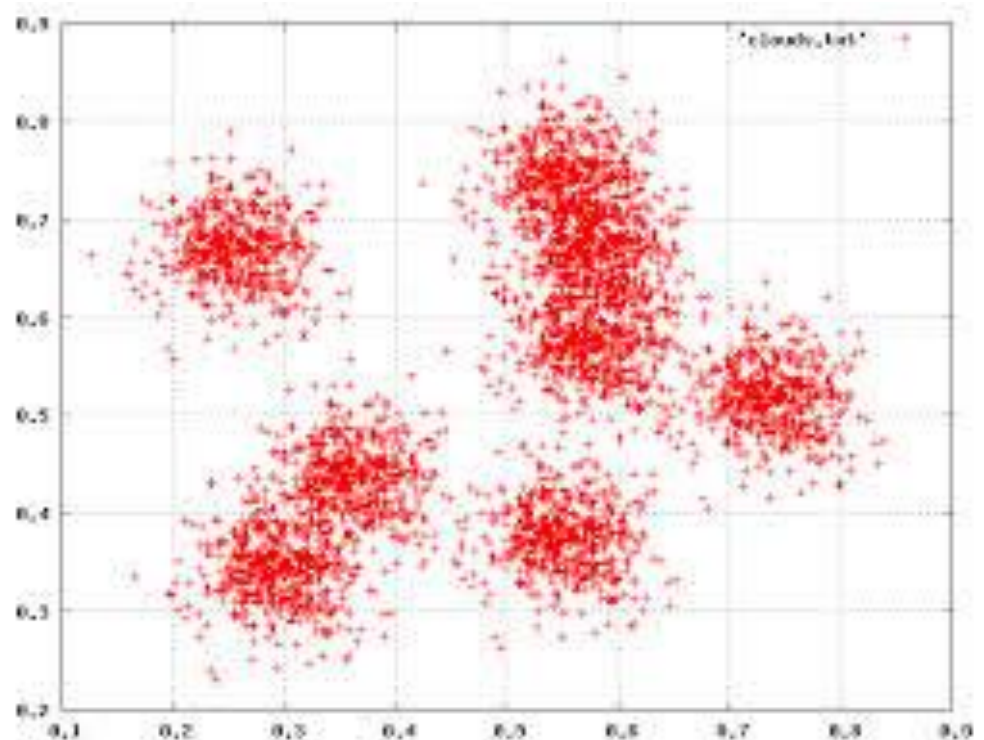
Narodowy Bank Polski

# ***ANALIZA SKUPIEŃ***

# Analiza skupień

Analiza skupień (grupowanie obiektowe) polega na dzieleniu (zazwyczaj wielowymiarowego) zbioru danych na grupy w taki sposób, by elementy w tej samej grupie były do

siebie podobne,  
a jednocześnie jak najbardziej odmienne od elementów z pozostałych grup.



# Analiza skupień

Ważnym krokiem jest wybór metody pomiaru odległości która określi stopień podobieństwa i niepodobieństwa obiektów.

Miarę odległości w analizie skupień definiowana jest na wiele sposobów. Najczęściej stosuje się odległość euklidesową.

## Procedura analizy skupień:

1. Określenie zbioru danych
2. Standaryzacja danych
3. Utworzenie macierzy odległości między obiektami
4. Utworzenie grup obiektów z najmniejszymi odległościami w ramach grupy

# Analiza skupień

Przykład. Wykorzystanie metody Warda do zbadania zróżnicowania regionalnego rolnictwa w Polsce.

Zastosowano 5 zmiennych:

x1 - średnia powierzchnia gospodarstwa rolnego

x2 - zbiory warzyw gruntowych w przeliczeniu na 1 mieszkańca

x3 - pogłowie bydła w przeliczeniu na 100 ha użytków rolnych

x4 - pogłowie trzody chlewnej w przeliczeniu na 100 ha gruntów ornych

x5 - zużycie nawozów mineralnych w przeliczeniu na czysty składnik na 1 ha użytków rolnych w kg



UMCS



Projekt sfinansowany  
z Narodowego Banku Polskiego  
w ramach programu oświaty i nauki

NBP Narodowy Bank Polski

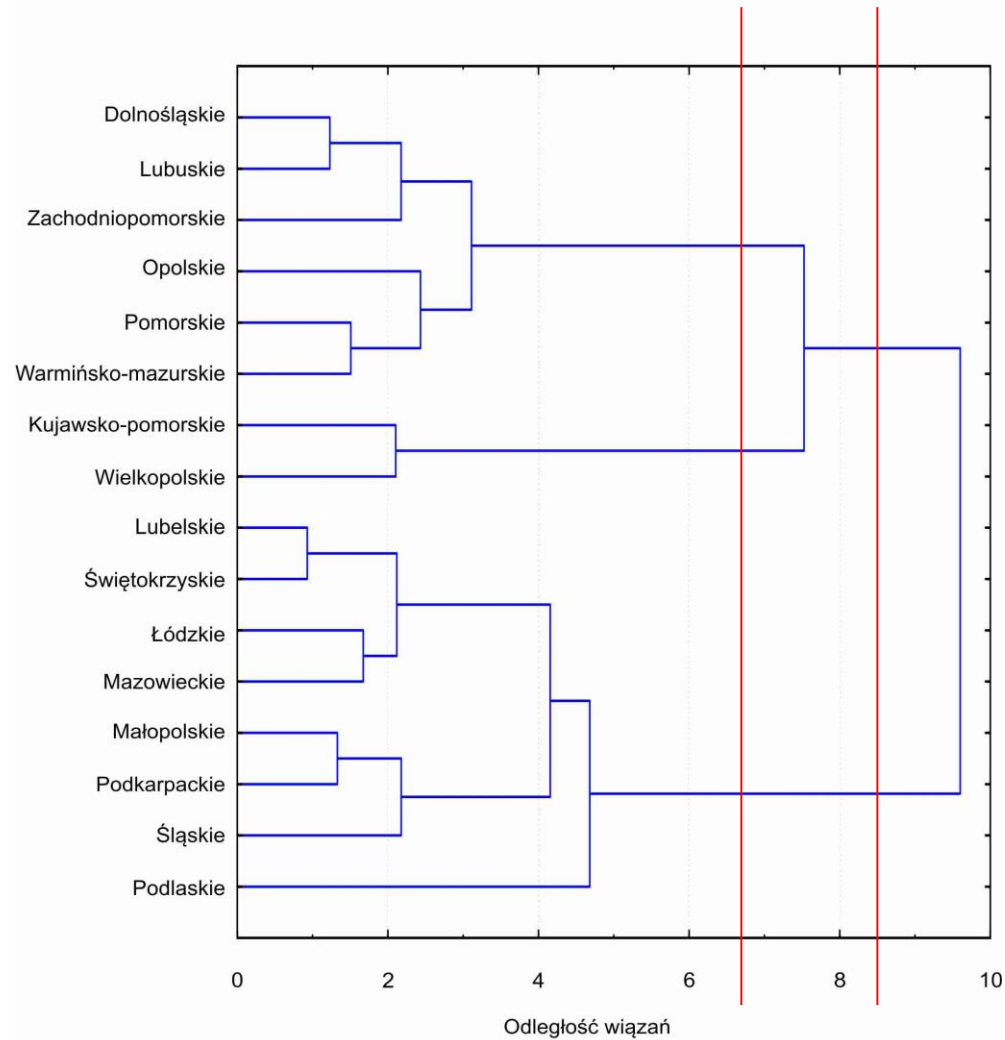
# Analiza skupień

Przykład. Zmienne zostały poddane standaryzacji. Następnie do obliczeń odległości i skupień wykorzystuje się programy statystyczne np. STATISTICA, SPSS. W efekcie otrzymujemy wykres w postaci drzewa na którym możemy zidentyfikować obiekty najbardziej podobne do siebie.

# Analiza skupień

Przykład.

RYS. 2. DENDROGRAM WYGENEROWANY METODĄ WARDA



Źródło: Obliczenia własne przy pomocy pakietu STATISTICA 9.1.



UMCS



# Analiza skupień

Przykład.

RYS. 4. ZRÓŻNICOWANIE REGIONALNE ROLNICTWA



Źródło: Obliczenia własne przy pomocy pakietu STATISTICA 9.1.



# Analiza skupień

## Przykład. Analiza utworów na Liście Przebojów Trójki<sup>a</sup> w latach 1982-2003

### Cechy statystyczne:

- liczba tygodni w pierwszej trzydziestce
- liczba tygodni w pierwszej dwudziestce
- liczba tygodni w pierwszej dziesiątce
- liczba tygodni w pierwszej piątce
- liczba tygodni na pierwszej pozycji
- najwyższa pozycja na liście
- liczba punktów uzyskana w czasie pobytu na liście
- pozycja na której utwór debiutował
- pozycja, którą utwór zajmował w ostatnim tygodniu pobytu na liście
- liczba tygodni pomiędzy debiutem a osiągnięciem najwyższej pozycji
- zmiana pozycji w pierwszym tygodniu pobytu na liście
- zmiana pozycji w ciągu dwóch pierwszych tygodni pobytu na liście

*a Wykorzystano pracę A. Sokołowskiego i M. Niedźwieckiego Lista Przebojów Trójki w obiektywie STATISTICA*



# Analiza skupień

## Przykład. Analiza utworów na Liście Przebojów Trójki<sup>a</sup>

Stosując metody analizy skupień ostatecznie uzyskano 6 grup:

1. Elita – Weterani – 91 utworów. Średnia pobytu na liście 29 tygodni. Wszystkie utwory weszły do pierwszej 5. Tylko jeden nie osiągnął pierwszego miejsca.

Najbardziej typowe utwory z tej grupy (znajdujące się najbliżej środka ciężkości w wielowymiarowej przestrzeni).

Wieża radości, wieża samotności – Sztywny Pal Azji

Animal Instinct – The Cranberries

La Isla Bonita – Madonna

Without You – Mariah Carey

Nothing Else Matters – Metallica

Californication – Red Hot Chili Peppers

Skóra – Aya R.L.

Desert Rose – Sting & Cheb Mami

Wonderful Life – Black

In The End – Linkin Park



# Analiza skupień

## Przykład. Analiza utworów na Liście Przebojów Trójki<sup>a</sup>

Grupa 2 – bardzo popularne – 339 utworów. Średnio przebywały po 18 tygodni na liście. Prawie wszystkie (99,7%) osiągnęły Top 5, a około 2/3 doszło do pozycji pierwszej

Cisza, ja i czas – Hey

The Zephyr Song – Red Hot Chili Peppers

With Or Without You – U2

Anioł wie – Urszula

Sprzedawcy marzeń – Myslovitz

Trudno nie wierzyć w nic – Raz Dwa Trzy

Hunting High & Low – a-ha

In The Shadows – The Rasmus

Heavy Metal Świat – TSA

The Sun Always Shine On TV – a-ha



# Analiza skupień

## Przykład. Analiza utworów na Liście Przebojów Trójki<sup>a</sup>

Grupa 3. Średnio popularne – niski debiut. 724 utwory. Pojedynczy utwór z tej grupy przebywał na liście średnio 11 tygodni, uzyskując 196 punktów. Nieco połowa z nich doszła do pierwszej piątki, a 7,5% nawet do pozycji pierwszej. Charakterystyczny dla tej grupy jest debiut na niskiej pozycji, średnio 26.

Yeke Yeke – Mory Kante

You Don't Understand Me – Roxette

Roxanne – George Michael

Soplicowo – Grzegorz Turnau & Stanisław Soyka

Close To Me - The Cure

Po to jesteś na świecie – Maanam

Ballada Emanuel – Wilki

Heledore Babe – Hey

Błagam, nie odmawiaj – Obywatel G.C.

About The Girl - Nirvana



# Analiza skupień

## Przykład 3. Analiza utworów na Liście Przebojów Trójki<sup>a</sup>

Grupa 4. Średnio popularne – wysoki debiut. W grupie tej są 383 utwory, czyli 9,5%. Średni pobyt na liście to 10 tygodni, a średnia zdobytych punktów to 183. Ponad połowa utworów tej grupy osiągnęła pierwszą piątkę, a 15% doszło do pozycji 1. Średnio utwory tej grupy debiutowały bardzo wysoko – na pozycji 14.

Marchewkowe pole – Lady Pank

Baby You're Mine – Basia

Osobno – Lady Pank

Tokyo – Various Manx

Clocks – Coldplay

Simple Song – Papa Dance i Petr Kotvald

Gyöngyhajú lány – Omega

Wieje piaskiem od strony wojny – Maanam

Should I Stay Or Should I Go – The Clash

Malinowy król – Urszula



# Analiza skupień

## Przykład. Analiza utworów na Liście Przebojów Trójki<sup>a</sup>

### Grupa 5. Mało popularne – Magma

W tej grupie jest aż 31,2% wszystkich utworów. Jest ich tu 1261. Przebywały na liście średnio po 7 tygodni, uzyskując średnio 74 punkty. Tylko 1% utworów tej grupy doszło do pierwszej piątki. Do pierwszej dziesiątki zestawienia tygodniowego doszło 21%. Żaden z utworów tej grupy nie osiągnął pozycji pierwszej.

Dzień się budzi – Basia

Girls – David Bowie

Summer In The City – Joe Cocker

Love To Be Loved – Peter Gabriel

Dworzec – Maryla Rodowicz

Secretly – Skunk Anansie

Hard Woman – Mick Jagger

Change – Lisa Stansfield

I Wanna Take Forever Tonight – Peter Cetera & Cristal Bernard

Kotysanka dla narzeczonej tapicera – Bielizna



# Analiza skupień

## Przykład. Analiza utworów na Liście Przebojów Trójki<sup>a</sup>

Grupa 6. Przemykające. To utwory, które przemknęły przez listę. Były na niej średnio trzy tygodnie, uzyskując średnio 15 punktów. Żaden z utworów tej grupy nie wszedł do pierwszej dziesiątki, a tylko 12,6% osiągnęło Top 20.

Hungry Like A Wolf – Duran Duran

Rozmowa bez słowa – Stare Dobre Małżeństwo

Kto zamiast mnie – Urszula

Maska – moje ja – Closterkeller

Stay In The Night – Benjamin Orr

Jan Serce – Seweryn Krajewski

Windy Town – Chris Rea

Stand Up (Kick Love Into Motion) – Def Leppard

Angel In My Heart – Mick Jagger

White Lines – Duran Duran



UMCS  
UNIWERSYTET MEDYCZNY W LUBLINIE



Projekt sfinansowany  
z Narodowego Banku Polskiego  
w ramach programu edukacji ekonomicznej  
NBP Narodowy Bank Polski