

# Metody klasyfikacji i klasteryzacji obiektów wielocechowych

# Zakres szkolenia

- Podstawowe pojęcia związane z klasyfikacją wielocechową
- Proste metody porządkowania liniowego (ratingu)
- Metody grupowania (klasteryzacji)

# Przykład

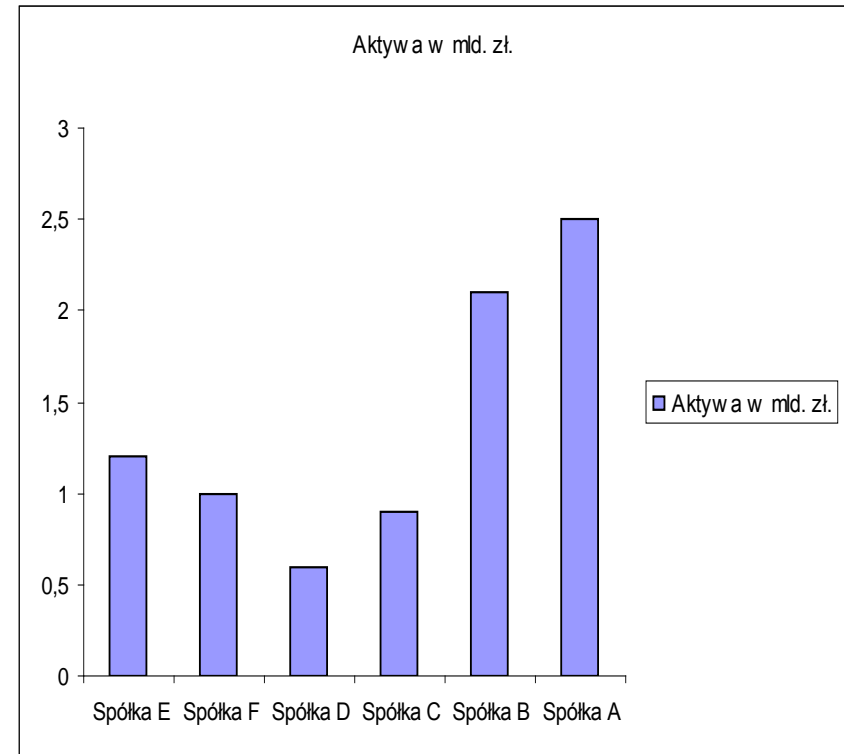
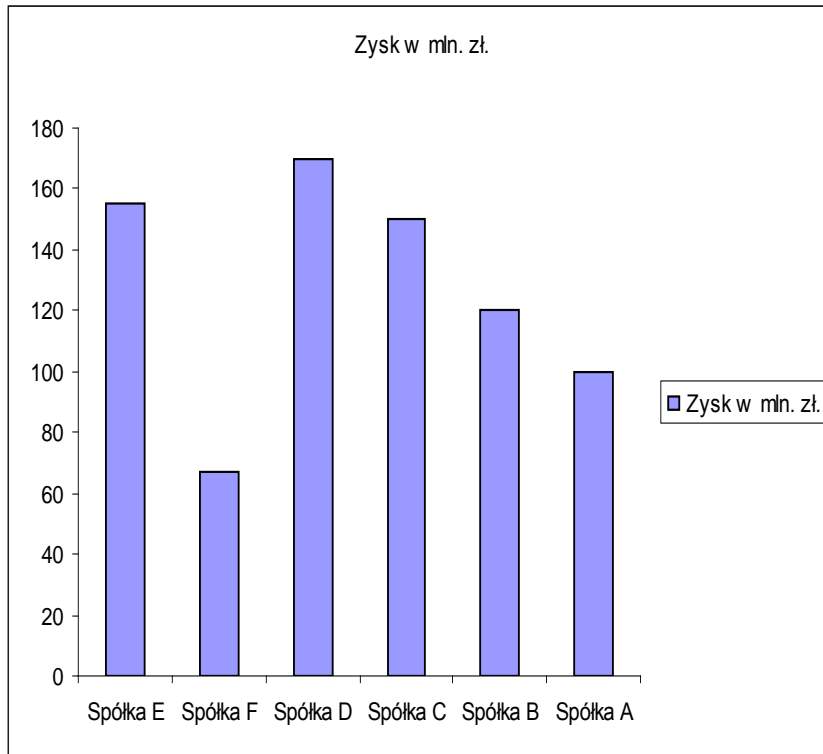
Celem badania jest ocena pewnej grupy przedsiębiorstw tej samej branży. Aby dokonać takiej oceny konieczna jest znajomość danych dotyczących pewnej kluczowej zmiennej, na przykład zysku. Znając zysk każdego przedsiębiorstwa możliwe jest pogrupowanie przedsiębiorstw w grupy lub uporządkowanie przedsiębiorstw względem wielkości zysku.

	Zysk w mln. zł.
Spółka E	155
Spółka F	67
Spółka D	170
Spółka C	150
Spółka B	120
Spółka A	100

Rozwiązanie zagadnienia komplikuje się, jeżeli zmienną, na podstawie której należy dokonać oceny nie jest tylko zysk lecz pojawia się zmienna dodatkowa, np. wartość aktywów.

	Zysk w mln. zł.	Aktywa w mld. zł.
Spółka E	155	1,2
Spółka F	67	1
Spółka D	170	0,6
Spółka C	150	0,9
Spółka B	120	2,1
Spółka A	100	2,5

Problem klasyfikacji tych przedsiębiorstw potwierdza poniższy wykres.



Kolejny problem pojawia się jeżeli byśmy chcieli dodać kolejną zmienną np. obciążenia kredytowe. O ile zmienne zysk i aktywa są zmiennymi „pozytywnymi” (atrakcyjność przedsiębiorstwa rośnie wraz ze ich wzrostem), to zmienna obciążenia kredytowe jest zmienna „negatywną” (atrakcyjność przedsiębiorstwa maleje wraz ze jej wzrostem). Pojawia się zatem problem odpowiedniego jej przekształcenia.

Zmienne „pozytywne” w analizie danych określane są mianem stymulanty

Zmienne „negatywne” w analizie danych określane są mianem destymulanty

Podstawowe problemy doprowadzenia do porównywalności, które należy rozwiązać, aby stosować mierniki syntetyczne to:

- sprowadzić wszystkie zmienne do jednej uniwersalnej „jednostki” (np. dwie zmienne, pierwsza wyrażona w zł. druga np. w liczbie pracowników) Normalizacja (standaryzacja zmiennej).
- przekształcić zmienne na stymulanty.

## NORMALIZACJA

Umożliwia sprowadzenie wyników różnych zmiennych do porównywalności. Przekształca wyniki pomiarów na skalę rozkładu normalnego ze średnią 0.

$$z_{ij} = \left( \frac{x_{ij} - a}{b} \right)^p$$

$$p = 1; a = \bar{x}_j; b = S(x_j)$$

	WRynkowa
Chinamob	90167
telmexic	25223
skTeleco	15276
KoTeleco	13892
TP.S.A.	7282
Tmalaysi	6457
KtFretel	5704
TelNotre	5556
M-Cell	4290
MATAV	3257
T_Argent	3091
BrasilTI	2875
CeskyTel	2855
AdvInfSr	2558
Philipin	2168
<b><u>średnia</u></b>	<b><u>12710</u></b>
<b><u>SD</u></b>	<b><u>22351</u></b>

	WRynkowa znormalizowana
Chinamob	3,47
telmexic	0,56
skTeleco	0,11
KoTeleco	0,05
TP.S.A.	-0,24
Tmalaysi	-0,28
KtFretel	-0,31
TelNotre	-0,32
M-Cell	-0,38
MATAV	-0,42
T_Argent	-0,43
BrasilTI	-0,44
CeskyTel	-0,44
AdvInfSr	-0,45
Philipin	-0,47



# Destymulanty i stymulanty

Przekształcenie destymulanty w stymulanty może przebiegać zgodnie z poniższą formułą:

$$x_{ij} = \max_i x_{ij} - x_{ij}$$

Lub:

$$x_{ij} = \frac{1}{x_{ij}}$$

# Proste metody syntetyczne

Dla celów stworzenia prostego rankingu można posłużyć się prostą metodą porządkowania opisaną w poniższych krokach:

1. Z badać zróżnicowanie badanych cech (za pomocą współczynnika zmienności; przyjmuje się w badaniach „dobra” zmienna ma wartość  $V_x > 10\%$ )
2. Przekształcić destymulanty na stymulanty
3. Dokonać normalizacji
4. Obliczyć miernik  $s_i$  ze wzoru:

$$s_i = \frac{1}{p} \sum x_{ij}$$

wyższy wynik miernika  $s_i$  oznacza lepszą pozycję obiektu (spółki) w rankingu

## Zadanie.

Korzystając z danych:

Przeprowadź prosta klasyfikację i przygotuj ranking przedsiębiorstw uwzględniający wszystkie posiadane zmienne korzystając z prostej formuły miernika  $s_i$

Zmienne w analizie:

WRynkowa	CAkcji	Zmiana	Wksięgow	C/Z	Rentowno	sprzedaż	zysk	aktywa	dochódAk
----------	--------	--------	----------	-----	----------	----------	------	--------	----------

Założenie: dla zmiennych branych pod uwagę powinien zachodzić warunek:

$$|Vx| > 10\%$$

Modyfikacja uwzględniająca różne wagi poszczególnych zmiennych.

Często w rzeczywistości poszczególne zmienne (cechy) nie wpływają w jednakowym stopniu na ostateczny wynik (mają różną wagę).

Na podstawie tej sugestii pojawiła się metoda syntetyczna z możliwością nadania wag poszczególnym zmiennym. Wadą tej metody jest konieczność znajomości wag przed przystąpieniem do obliczeń.

# Syntetyczny miernik wzorcowy (metoda porządkowania liniowego Hellwiga)

## Metoda porządkowania liniowego

Metoda ta należy do grupy metod porządkowania liniowego. Wynikiem jej jest uporządkowanie obiektów od najgorszego do najlepszego (utworzony zostaje ranking). Cechą charakterystyczną tej metody jest to, iż wszystkie zmienne są jednakowo ważne (mają te same wagi).

Zastosowanie metody Hellwiga należy rozpocząć od określenia które z cech są stymulantami, a które destymulantami.

Po przeprowadzeniu procedury normalizacji należy wyznaczyć wzorce dla poszczególnych cech:

Dla destymulant :  $x_{0k} = \min\{x_{ij}\}$

Dla stymulant :  $x_{0k} = \max\{x_{ij}\}$

Obliczenia należy wykonać zgodnie z formułami:

$$S_0 = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_{i0} - \bar{d}_0)^2}$$

$$d_i = 1 - \frac{d_{i0}}{d_0}$$

$$\bar{d}_0 = \frac{1}{n} \sum_{i=1}^n d_{i0}$$

$$d_{i0} = \sqrt{\sum_i (x_{ij} - x_{0j})^2}$$

$$d_0 = \bar{d}_0 + 2 \cdot S_0$$

Wynikiem procedury będzie ranking obiektów (firm). Każdy obiekt w wyniku obliczeń uzyska wartość  $d$  z przedziału  $\langle 0; 1 \rangle$ .

Im wyższy wynik  $d$  tym obiekt (firma) jest wyżej w klasyfikacji.

## Zadanie.

Zastosuj miernik syntetyczny do klasyfikacji spółek.

Wyniki:

spółki	Miernik syntetyczny Hellwiga $d_i$
telmexic	0,47
Chinamob	0,36
skTeleco	0,27
KoTeleco	0,23
AdvInfSr	0,23
TP.S.A.	0,19
T_Argent	0,19
TelNotre	0,18
KtFretel	0,18
MATAV	0,15
Philipin	0,15
BrasilTI	0,14
Tmalaysi	0,14
M-Cell	0,11
CeskyTel	0,07

W literaturze wyróżnia się dwie podstawowe klasy metod klasyfikacji:

1. Klasyfikację, na podstawie której dokonuje się tylko przydzielenia obiektu do grupy (tworzone są np. trzy grupy: firmy słabe, przeciętne, dobre)
2. Porządkowanie liniowe, na podstawie którego tworzy się ranking (każdy obiekt (firma) ma indywidualną pozycję w rankingu. Istnieje możliwość utworzenia na podstawie tego rankingu grup.

Tworzenie grup na podstawie rankingu.

Podział na grupy metodą naukową możliwy jest poprzez zastosowanie kwartyli na danych dotyczących wskaźników  $d_i$

**Zadanie.**

Bazując na rankingu utworzonym dla spółek utwórz 4 grupy firm stosując dobór do grupy na podstawie kwartyli.

spółki	$d_i$
telmexic	0,47
Chinamob	0,36
skTeleco	0,27
KoTeleco	0,23
AdvInfSr	0,23
TP.S.A.	0,19
T_Argent	0,19
TelNotre	0,18
KtFretel	0,18
MATAV	0,15
Philipin	0,15
BrasilTI	0,14
Tmalaysi	0,14
M-Cell	0,11
CeskyTel	0,07

	wartość $d_i$
kwartyl 1	0,141
kwartyl 2	0,184
kwartyl 3	0,227



**Zadanie.**

Bazując na rankingu utworzonym dla spółek utwórz 4 grupy firm stosując dobór do grupy na podstawie kwartyli.

spółki	$d_i$
telmexic	0,47
Chinamob	0,36
skTeleco	0,27
KoTeleco	0,23
AdvInfSr	0,23
TP.S.A.	0,19
T_Argent	0,19
TelNotre	0,18
KtFretel	0,18
MATAV	0,15
Philipin	0,15
BrasilTI	0,14
Tmalaysi	0,14
M-Cell	0,11
CeskyTel	0,07

	wartość $d_i$
kwartyl 1	0,141
kwartyl 2	0,184
kwartyl 3	0,227