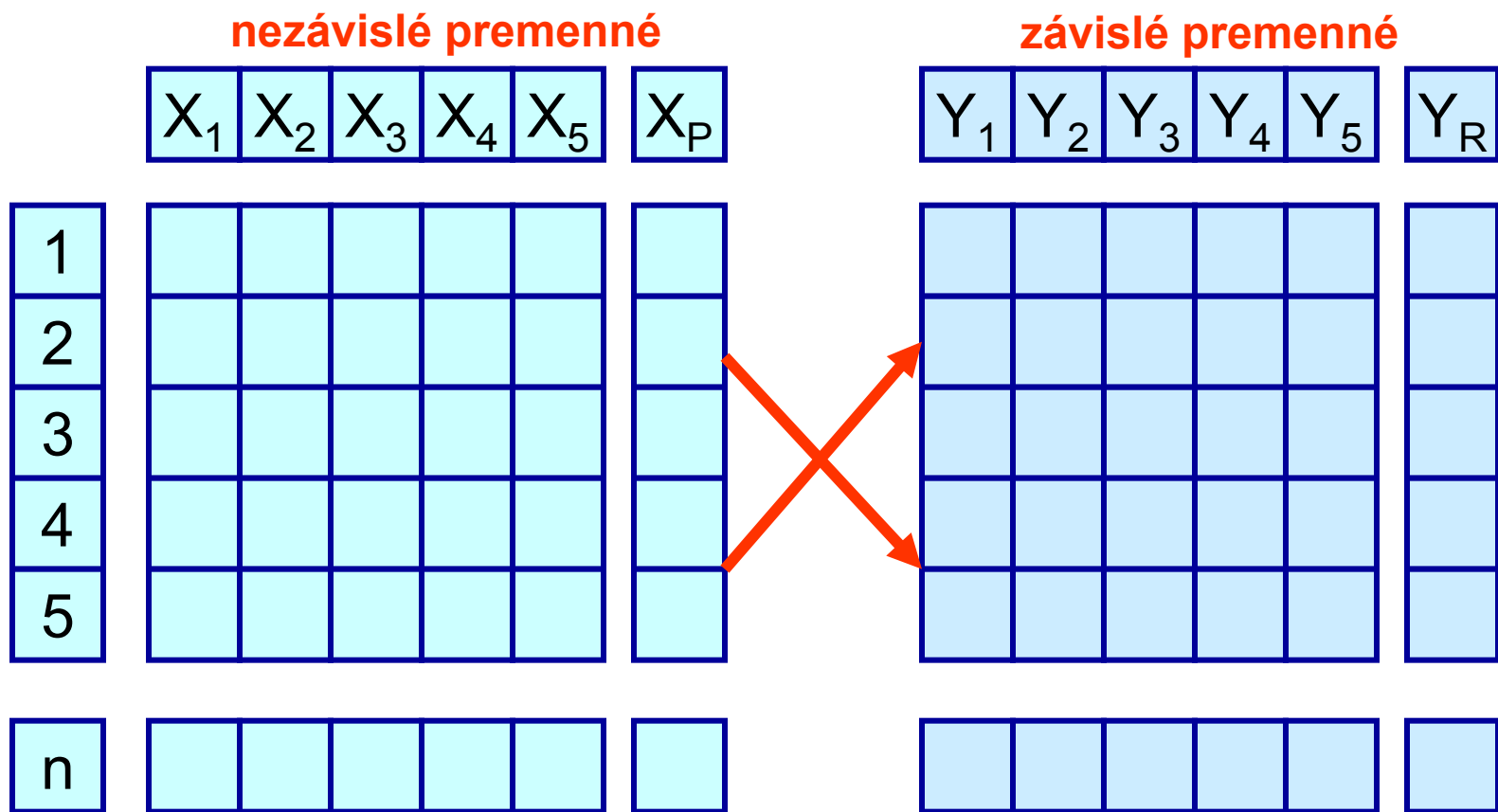


Viacrozmerné metódy všeobecne

Viacrozmerné metódy



Metódy analýzy závislostí

Viacrozmerné metódy

- **Metódy analýzy závislostí**
 - premenné možno rozdeliť do **dvoch** skupín
 - cieľom je dokázať prítomnosť/neprítomnosť **závislosti** medzi týmito dvomi skupinami
 - ak možno jednu skupinu označiť za závislé premenné a druhú za nezávislé, potom cieľom je určiť ako nezávislé premenné pôsobia na závislé premenné

Viacrozmerné metódy

Nezávislá premenná (premenné)		Závislá premenná (premenné)			
		Jedna		Viac ako jedna	
		Kvantitatívna	Kvalitatívna	Kvantitatívna	Kvalitatívna
Jedna	Kvantitatívna	Regresná analýza	Logistická regresia Diskriminačná analýza	Kanonická korelačná analýza	Viacskupinová diskriminačná analýza
	Kvalitatívna	t-test	Diskrétna diskriminačná analýza	MANOVA	Diskrétna viacskupinová diskriminačná analýza
Viac ako jedna	Kvantitatívna	Viacnásobná regresná analýza	Logistická regresia Diskriminačná analýza	Kanonická korelačná analýza	Viacskupinová diskriminačná analýza
	Kvalitatívna	ANOVA	Diskrétna diskriminačná analýza Conjoint analýza	MANOVA	Diskrétna viacskupinová diskriminačná analýza

Metódy analýzy závislostí

Viacrozmerne metody

premenne

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_R
1												
2												
3												
4												
5												
n												

Metody analyzy skrytych vzťahov

Viacrozmerné metódy

- **Metódy analýzy skrytých vzťahov**
 - premenné **nemožno** logicky rozdeliť do **dvoch** skupín na závislé a nezávislé
 - cieľom je pochopiť alebo identifikovať **prečo a ako** sú premenné **navzájom korelované** t.j. ako sa navzájom ovplyvňujú
 - ak sú premenné navzájom prepojené – korelované, možno rovnaký objem informácií vystihnúť menším počtom premenných – **zníženie dimenzie**

Viacrozmerné metódy

Počet premenných	Typ údajov	
	Kvantitatívne	Kvalitatívne
Dve	Jednoduchá korelácia	Analýza dvojrozmerných kontingenčných tabuliek
		Loglineárne modely
Viac ako dve	Analýza hlavných komponentov	Analýza viacrozmerných kontingenčných tabuliek
	Faktorová analýza	Loglineárne modely
		Korešpondenčná analýza

Metódy analýzy skrytých vzťahov

Viacrozmerné metódy

premenné

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_R
1												
2												
3												
4												
5												
n												

Klasifikačné metódy

Viacrozmerné metódy

- **Klasifikačné metódy**

- v porovnaní s predchádzajúcimi metódami sú predmetom analýzy **pozorovania** nie premenné
- cieľom je
 - identifikovať podobné pozorovania
 - zoskupit' ich **do skupín**, ktoré sú vnútorne homogénne ale navzájom odlišné
 - určiť **pravidlo**, podľa ktorého možno do určených skupín zaradiť ďalšie pozorovania na základe podobnosti

Viacrozmerné metódy

- **Viacrozmerné metódy**
 - **metódy analýzy skrytých vzťahov (zníženia dimenzie)**
 - analýza hlavných komponentov
 - faktorová analýza
 - **klasifikačné metódy**
 - zhuková analýza
 - **metódy analýzy závislostí**
 - kanonická korelačná analýza
 - diskriminačná analýza

Zhluková analýza

Cluster Analysis (CA)

Zhluková analýza

- **Charakteristika CA**

- rozklad súboru na niekoľko homogénnych podsúborov
 - štatistické jednotky v jednom zhluku sú čo najpodobnejšie
 - štatistické jednotky rôznych zhlukov sú navzájom čo najodlišnejšie
- zoskupit' X_i $i=(1, \dots, n)$ do zhlukov c_1, \dots, c_q ($2 < q < n$)

Zhluková analýza

- **Postup CA**
 - **výber miery vzdialenosti**
 - **výber druhu zhukovacieho postupu**
 - **výber zhukovacej metódy**
 - **určenie počtu významných zhukov**
 - **interpretácia zhukov**

Zhluková analýza

- **výber miery vzdialenosti**
 - miery podobnosti
 - koeficienty asociácie
 - koeficienty korelácie
 - miery nepodobnosti - vzdialenosti
 - pozitivita
 - symetria
 - trojuholníková rovnosť
 - Euklidovská vzdialenosť

$$D_{ij} = \sqrt{\sum (x_{ik} - x_{jk})^2}$$

Zhluková analýza

- **výber druhu zhlukovacieho postupu**

- **hierarchický postup**

- úplná analýza štatistického súboru
- výsledok možno zobrazit' pomocou hierarchického grafu - dendrogramu

- **aglomeratívne**

- v prvom kroku je každá jednotka samostatným zhlukom
- v poslednom kroku sú všetky jednotky v jednom zhluku

- **diverzívne**

- v prvom kroku sú všetky jednotky v jednom zhluku
- v poslednom kroku každá jednotka predstavuje zhluk

Zhluková analýza

- **výber druhu zhlukovacieho postupu**
 - **nehierarchický postup**
 - vopred určený počet zhlukov
 - iteratívny postup určenia ich stredov
 - postupné zaradovanie jednotiek k vybraným stredom

Zhluková analýza

- **výber zhlukovacej metódy**
 - **hierarchické – aglomeratívne**
 - **metóda najbližšieho suseda (single linkage)**
 - poznáme maticu euklidovských vzdialeností
 - kritérium spojenia = $\min d_{ij}$
 - vyberieme spojenie, prepočítame, atď.
 - **tzv. reťaziaci efekt**

	S1	S2	S3	S4	S5	S6
S1	0	2	181	221	625	821
S2	2	0	145	181	557	745
S3	181	145	0	2	136	250
S4	221	181	2	0	106	212
S5	625	557	136	106	0	26
S6	821	745	250	212	26	0

Zhluková analýza

- **výber zhlukovacej metódy**
 - **hierarchické – aglomeratívne**
 - **metóda najvzdialenejšieho suseda (complete linkage)**
 - poznáme maticu euklidovských vzdialeností
 - kritérium spojenia = $\max d_{ij}$
 - vyberieme spojenie, prepočítame, atď.
 - konzistentné, izolované zhluky

	S1	S2	S3	S4	S5	S6
S1	0	2	181	221	625	821
S2	2	0	145	181	557	745
S3	181	145	0	2	136	250
S4	221	181	2	0	106	212
S5	625	557	136	106	0	26
S6	821	745	250	212	26	0

Zhluková analýza

- **výber zhlukovacej metódy**

- **hierarchické – aglomeratívne**

- **metóda priemernej väzby suseda (average linkage)**

- poznáme maticu euklidovských vzdialeností
- po inicializačnom spojení je vzdialenosť medzi dvomi clustrami vypočítaná ako priemer vzdialeností ich prvkov

	S1	S2	S3	S4	S5	S6		S1&S2	S3	S4	S5	S6
S1	0	2	181	221	625	821						
S2	2	0	145	181	557	745	S1	0	163	201	591	783
S3	181	145	0	2	136	250	S1&S2	163	0	2	136	250
S4	221	181	2	0	106	212	S4	201	2	0	106	212
S5	625	557	136	106	0	26	S5	591	136	106	0	26
S6	821	745	250	212	26	0	S6	783	250	212	26	0

Zhluková analýza

- **výber zhlukovacej metódy**
 - **hierarchické – aglomeratívne**
 - **centroidná metóda (centroid method)**
 - poznáme maticu euklidovských vzdialeností
 - inicializačné spojenie
 - každý zhuk je nahradený “priemerným prvkom”, ktorý sa používa pri výpočte vzdialeností
 - nevýhodou sú inverzie
 - **mediánová metóda**
 - reprezentantom zhuku je jeho medián
 - vylepšenie centroidnej metódy

Zhluková analýza

- **určenie počtu významných zhlukov**
 - **RMSSTD** zhluku – vnútroskupinová variabilita
 - nižšia hodnota -> nižšia variabilita -> vyššia homogenita
 - **R-Squared**
 - pomer
 - medziskupinovej sumy štvorcov
 - celkovej sumy štvorcov
 - čím väčšia je medziskupinová variabilita, tým väčšie sú rozdiely medzi skupinami a tým homogénnejšie sú skupiny
 - $R^2=0$ – skupiny sú rovnaké
 - $R^2=1$ – skupiny sú rozdielne

Zhluková analýza

- **určenie počtu významných zhlukov**
 - **semipartial R-Squared (spojených zhlukov)**
 - pomer
 - vnútroskupinová variabilita(1) – vnútroskupinová variabilita (2)
 - celková variabilita
 - ak je rozdiel rovný 0 -> spojené boli úplne zhodné skupiny
 - rozdiel vyjadruje tzv. stratu homogenity
 - **vzdialenosť' zhlukov (CD)**
 - ak je CD malé číslo -> spojené boli podobné skupiny
 - **CCC – cubic clustering criterium > 2**

Zhluková analýza

- **určenie počtu významných zhlukov**
 - **RMSSTD - homogenita nového zhluku**
 - **malé**
 - **SPR – homogenita spojených zhlukov**
 - **malé**
 - **RS – heterogenita zhlukov**
 - **vysoké**
 - **CD – homogenita spojených zhlukov**
 - **malé**

Zhluková analýza

- **interpretácia zhlukov**
 - dendrogram
 - grafická analýza
 - popisná štatistika

CA - Príklady

