

Cvičení č. 12.: Shluková analýza

Článek Ladislava Rabušice Koho Češi nechtějí? (uveřejněn ve Sborníku prací FSS MU Sociální studia 5, 2000) se zabývá touto problematikou:

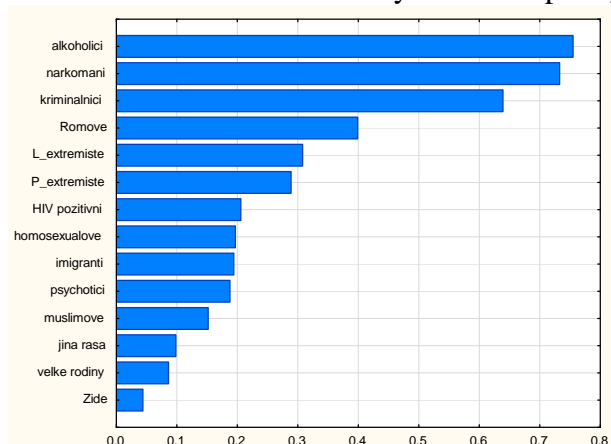
V roce 1999 proběhlo ve 24 evropských zemích sociologické šetření, v němž měli respondenti za úkol odpovědět na otázku „Můžete prosím z následujícího seznamu vybrat všechny ty, koho byste nechtěl(a) mít za sousedy?“ V seznamu byly tyto skupiny osob:

Kriminálníci, osoby jiné rasy, levicoví extrémisté, alkoholici, pravíkoví extrémisté, početné rodiny, psychotici, muslimové, imigranti, HIV pozitivní, narkomani, homosexuálové, židé, Romové.

V datovém souboru netolerance.sta jsou zaznamenány relativní četnosti vybraných skupin osob.

V České republice se výzkumu, který proběhl v květnu 1999, zúčastnilo 1908 osob.

Úkol 1.: Zaměřte se na ČR. Vytvořte sloupcový diagram tohoto tvaru:



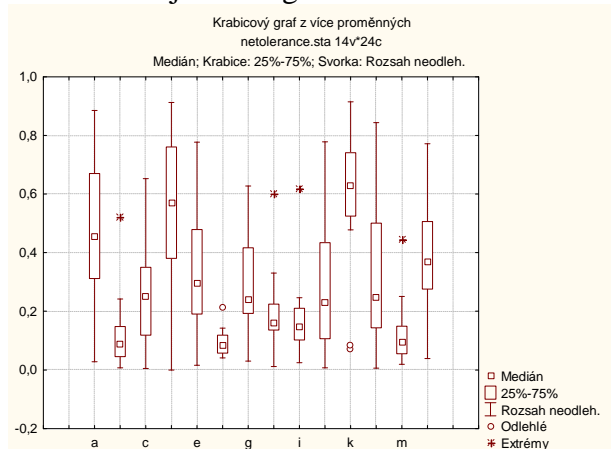
Návod: Řádek pro Českou republiku okopírujeme (se záhlavími) do nového datového souboru o 14 proměnných a jednom případě.

Soubor transponujeme: Data – Transponovat – Soubor.

Hodnoty proměnné Ceska rep. uspořádáme: Data – setřít – Přidat prom. Ceska rep. – OK.

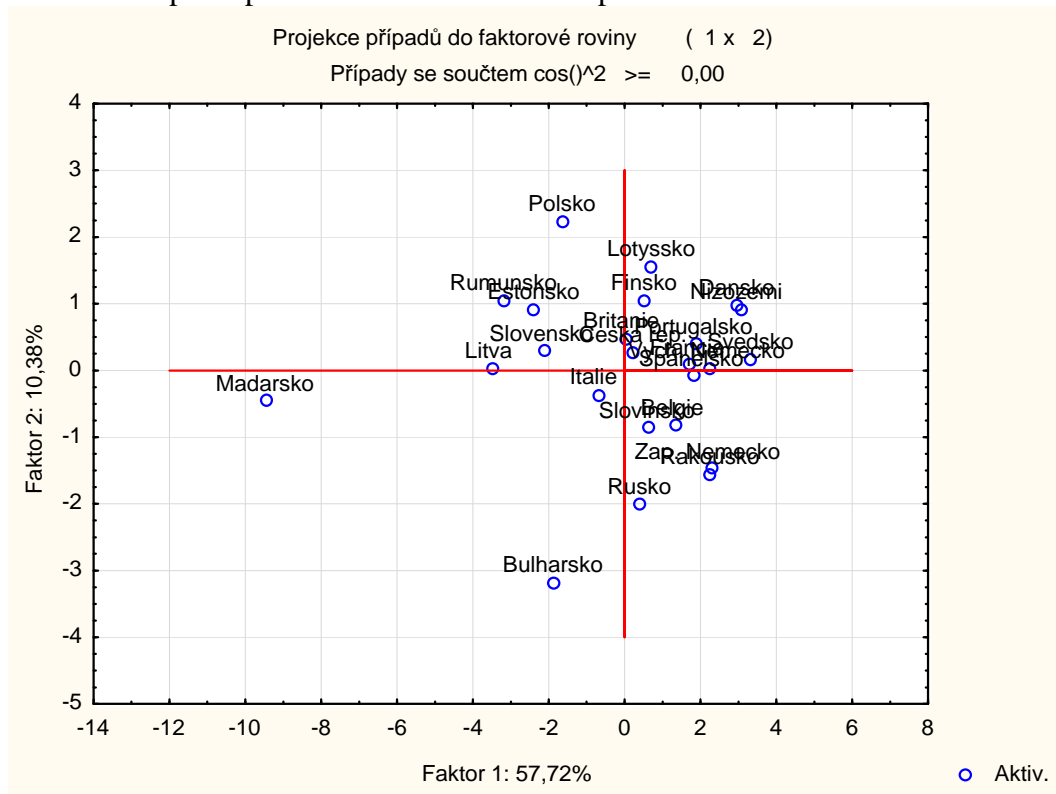
Nakreslíme sloupcový graf: Grafy – 2D grafy – Sloupcové/pruhové grafy – Proměnné Ceska rep. – O, Typ grafu Běžný, Orientace Horizontální – OK.

Úkol 2.: Do jednoho grafu nakreslete krabicové diagramy všech 14 proměnných.

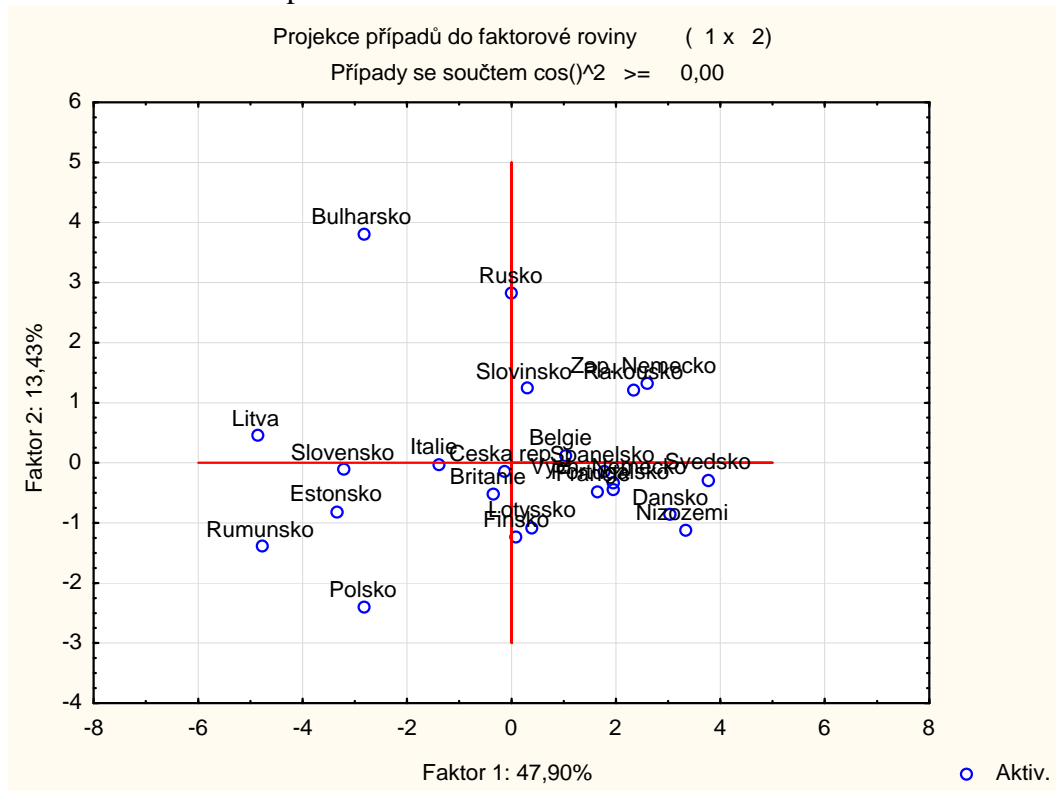


Vzhledem k velmi rozdílné variabilitě proměnných se jeví vhodnější pracovat se standardizovanými daty.

Úkol 3.: Na ploše prvních dvou hlavních komponent znázorníte rozmístění zemí.

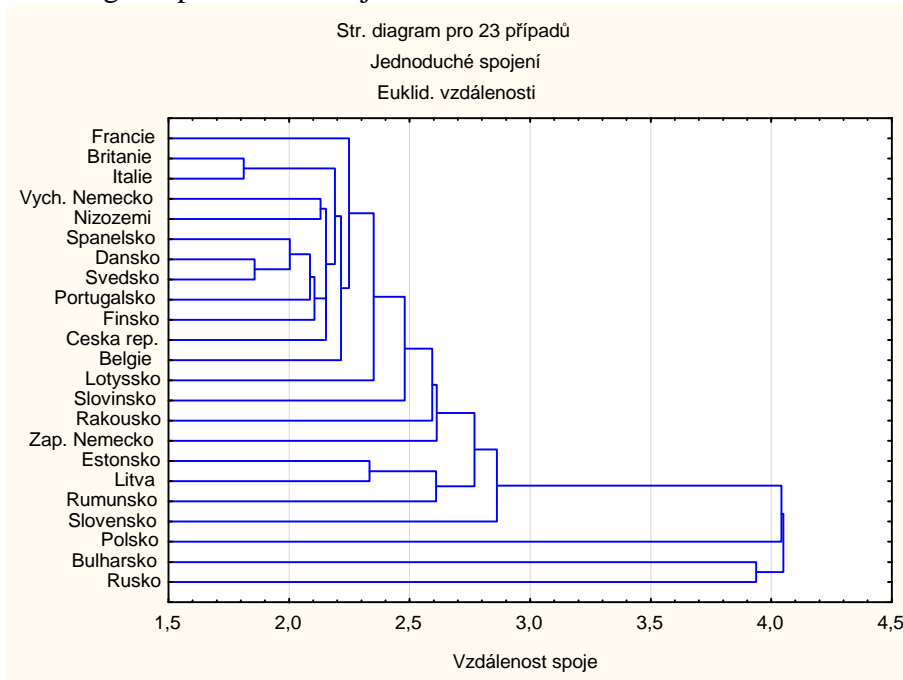


Maďarsko se jeví jako odlehle pozorování. Z dalších analýz ho vyloučíme. Znovu provedeme metodu hlavních komponent a dostaneme toto rozmístění zemí:

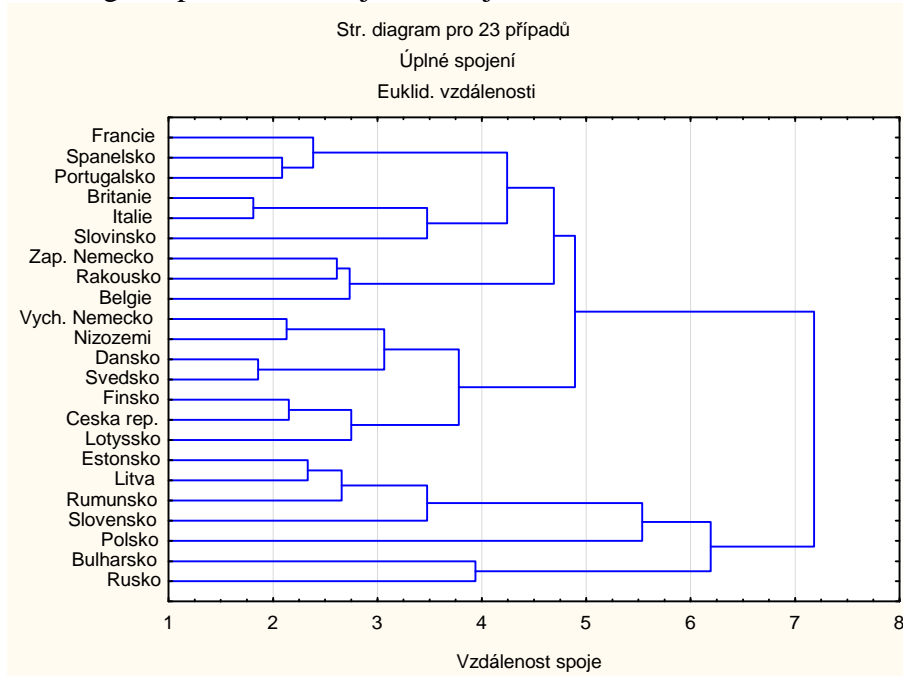


Úkol 4.: Použijte metodu nejbližšího souseda, nejbzdálenějšího souseda, metodu průměrné vazby a Wardovu metodu pro nalezení shluků zemí podobných z hlediska tolerance. Výsledky znázorněte pomocí dendrogramů.

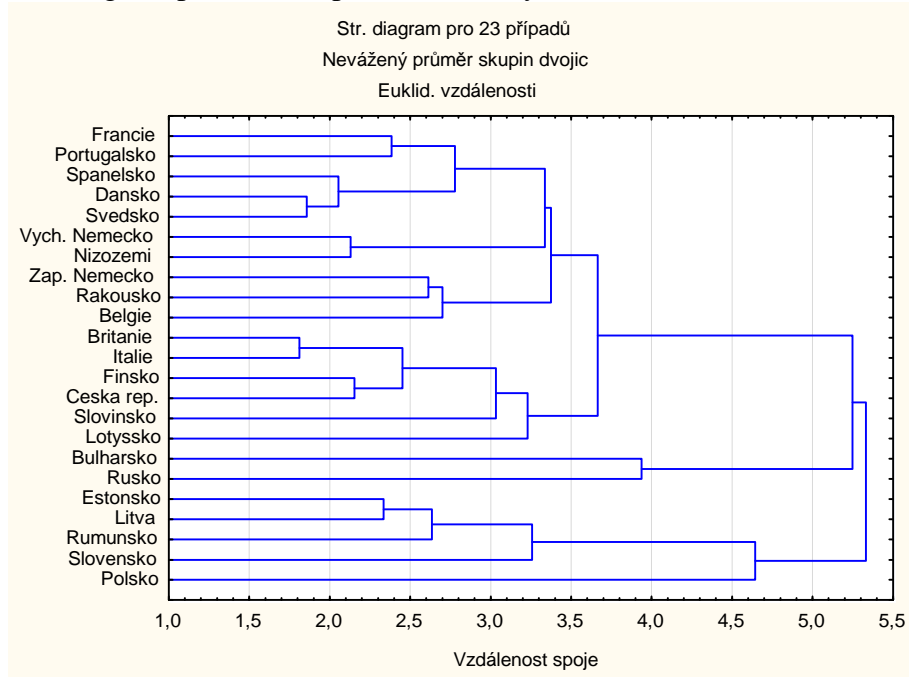
Dendrogram pro metodu nejbližšího souseda:



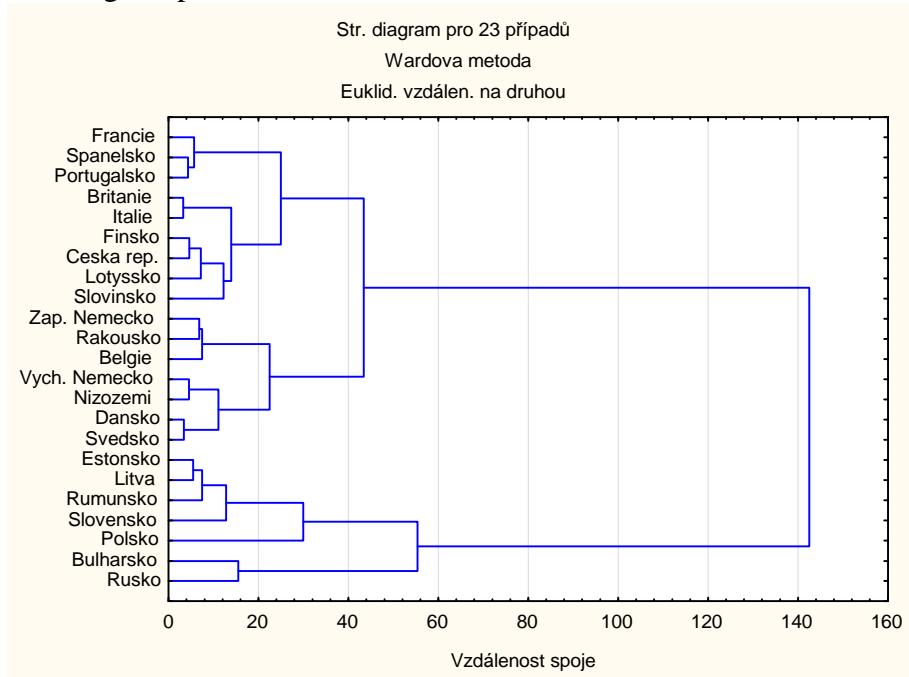
Dendrogram pro metodu nejbzdálenějšího souseda:



Dendrogram pro metodu průměrné vazby:



Dendrogram pro Wardovu metodu:



Úkol 5.: Pro Wardovu metodu určete 4 shluky navzájem si podobných zemí.

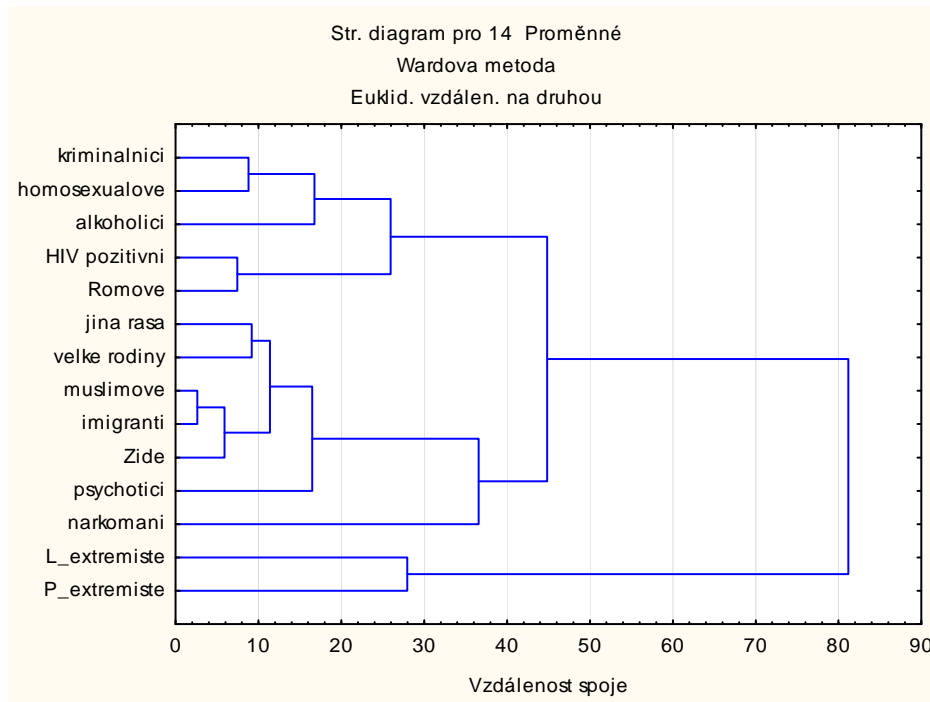
Shluk č. 1: Francie, Španělsko, Portugalsko, Velká Británie, Itálie, Finsko, ČR, Lotyšsko, Slovinsko

Shluk č. 2: Západní Německo, Rakousko, Belgie, Východní Německo, Nizozemí, Dánsko, Švédsko

Shluk č. 3: Estonsko, Litva, Rumunsko, Slovensko, Polsko

Shluk č. 4: Bulharsko, Rusko

Úkol 6.: Proveďte shlukovou analýzu pro proměnné.
Dendrogram pro Wardovu metodu:



Proměnné roztřídíme do čtyř shluků.

Shluk č. 1: kriminálníci, homosexuálové, alkoholici, HIV pozitivní, Romové

Shluk č. 2: osoby jiné rasy, velké rodiny, psychotici, muslimové, imigranti, Židé

Shluk č. 3: narkomani

Shluk č. 4: levicoví a pravicoví extrémisté

Úkol 7.: Použijte metodu k-průměrů k nalezení 4 shluků navzájem si podobných zemí a uložte skupinovou příslušnost do datového souboru. K určení významnosti jednotlivých proměnných proveďte analýzu rozptylu. Nakreslete graf průměrů všech 4 shluků a pokuste se o interpretaci.

Rozdíly oproti Wardově metodě: Francie a Portugalsko byly zařazeny do shluku se Záp. Německem,

Výsledek analýzy rozptylu:

Proměnná	Analýza rozptylu (netolerance.sta)					
	Mezisk. SČ	sv	Vnitřní SČ	sv	F	význam. p
kriminalnici	15,62198	3	3,73719	19	26,47424	0,000001
jina rasa	3,96231	3	3,40403	19	7,37203	0,001799
L_extremiste	1,64007	3	15,82236	19	0,65648	0,588798
alkoholici	13,16491	3	7,89391	19	10,56230	0,000263
P_extremiste	4,49299	3	16,89420	19	1,68434	0,204084
velke rodiny	7,20290	3	7,33896	19	6,21591	0,004016
psychotici	9,57379	3	9,74601	19	6,22142	0,004000
muslimove	5,68687	3	3,47000	19	10,37951	0,000290
imigranti	4,53931	3	2,51944	19	11,41087	0,000167
HIV pozitivni	10,45365	3	5,74605	19	11,52208	0,000158
narkomani	18,30150	3	2,57925	19	44,93922	0,000000
homosexualove	13,86812	3	3,75787	19	23,37266	0,000001
Zide	5,78665	3	5,16341	19	7,09779	0,002164
Romove	6,26289	3	13,69391	19	2,89654	0,061976

Na hladině významnosti 0,05 nejsou významné pouze proměnné L_extremisté, P_extremisté a Romové. Podle hodnot statistiky F lze soudit, že na zařazování zemí do shluků se nejvíce podílí proměnné narkomani, kriminálníci a homosexuálové.

Graf průměrů:

