

### Cvičení č. 3.: Aplikace shlukové analýzy

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?“

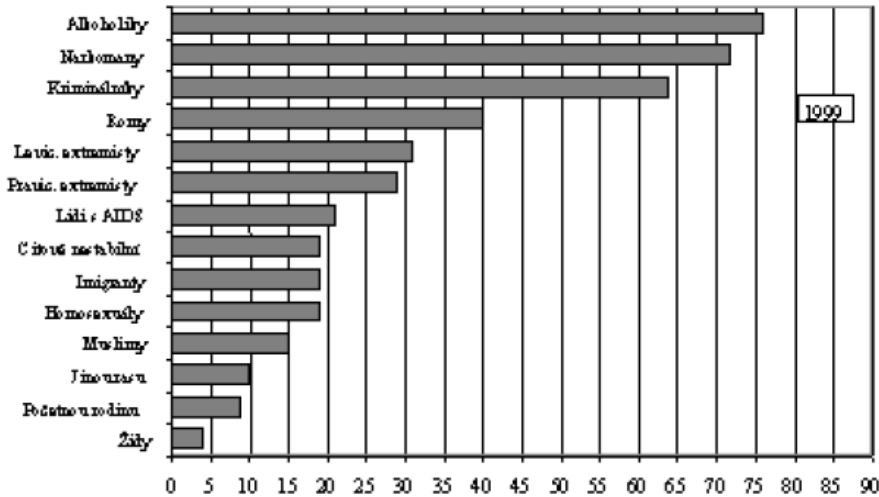
V seznamu byly tyto skupiny osob:

Kriminálníci, osoby jiné rasy, levicoví extrémisté, alkoholici, pravicoví extrémisté, početné rodiny, citově nestabilní lidé, muslimové, imigranti, lidé s AIDS, narkomani, homosexuálové, židé, Romové.

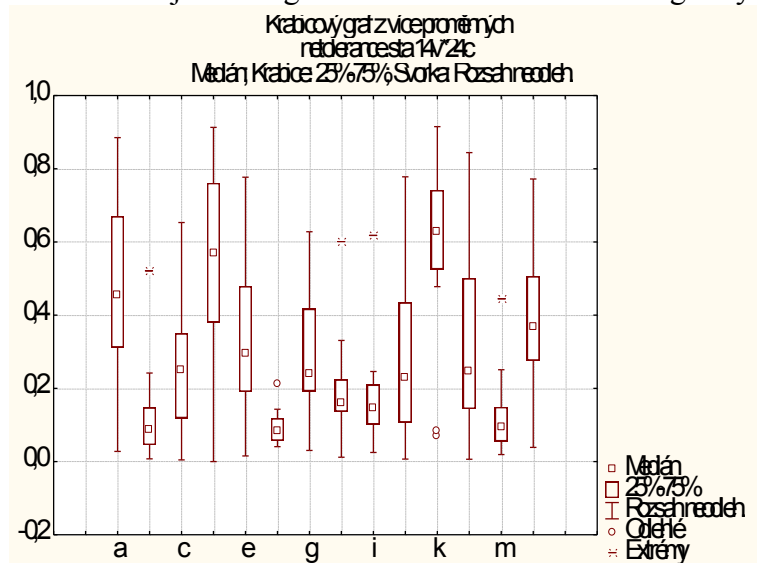
V datovém souboru netolerance.sta jsou zaznamenány relativní četnosti vybraných skupin osob.

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

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

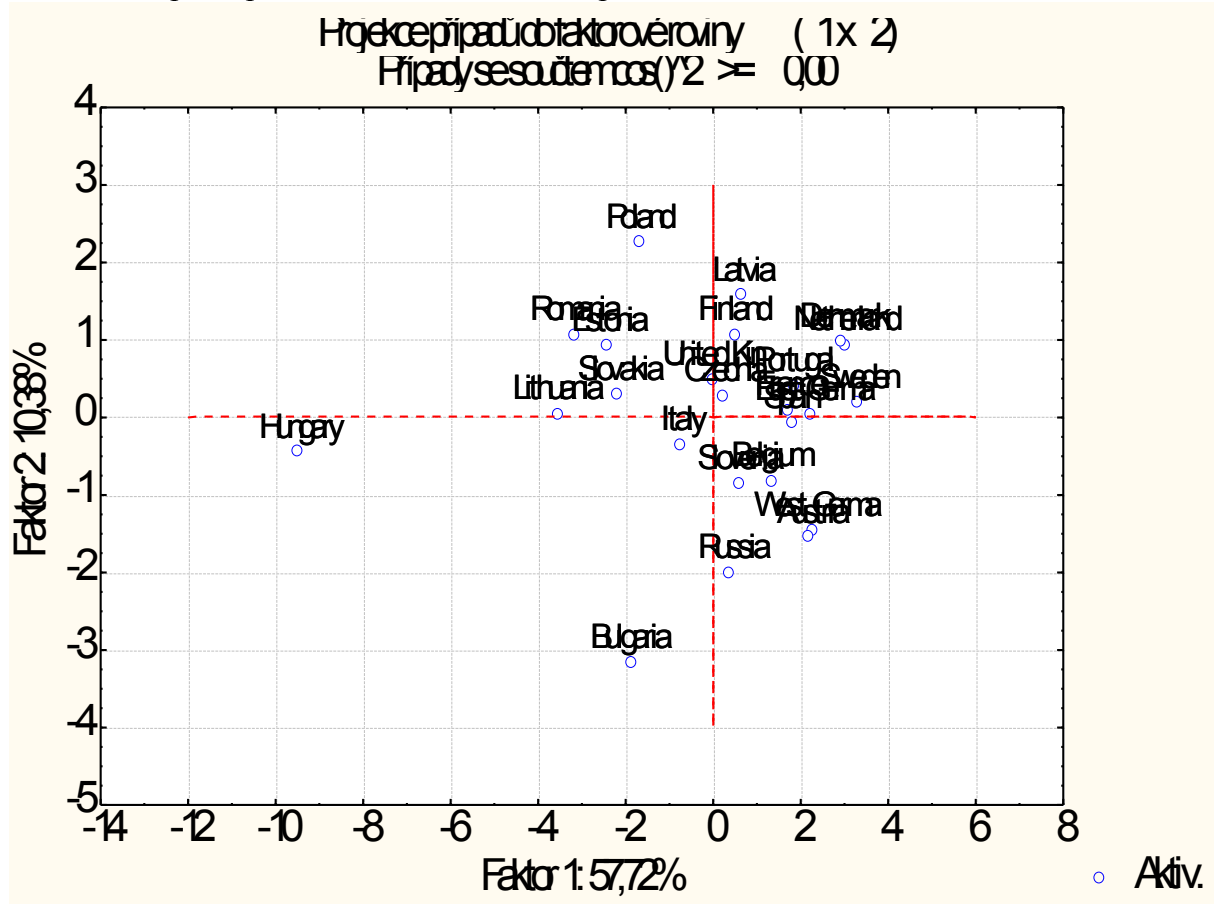


**Úkol 2.:** Do jednoho grafu nakreslete krabicové diagramy všech 14 proměnných.



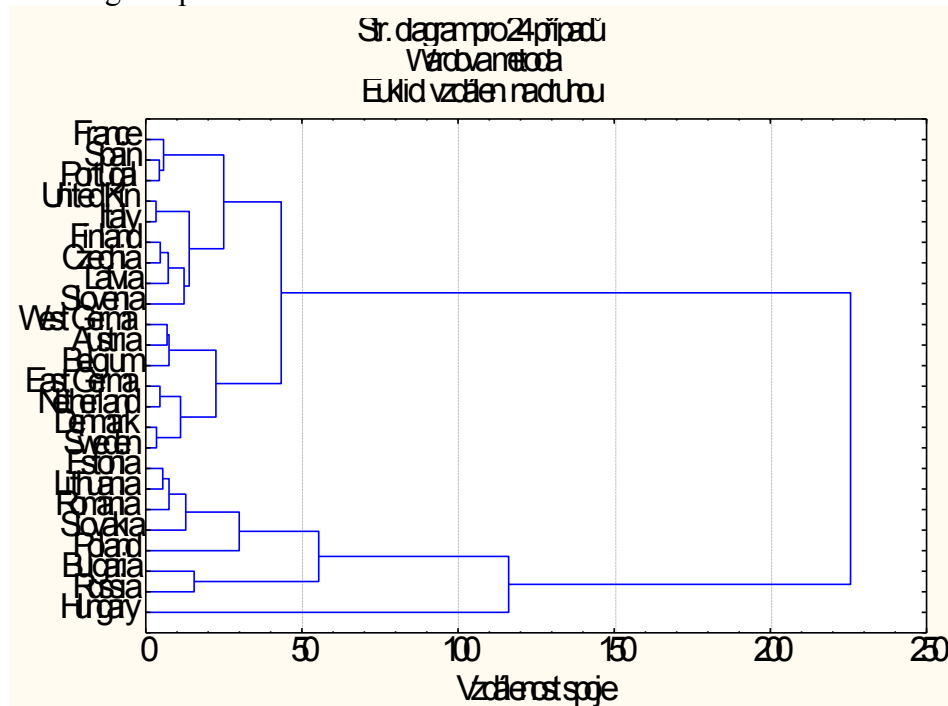
Vzhledem k velmi rozdílné variabilitě proměnných se jeví vhodnější pracovat se standardizovanými daty.

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



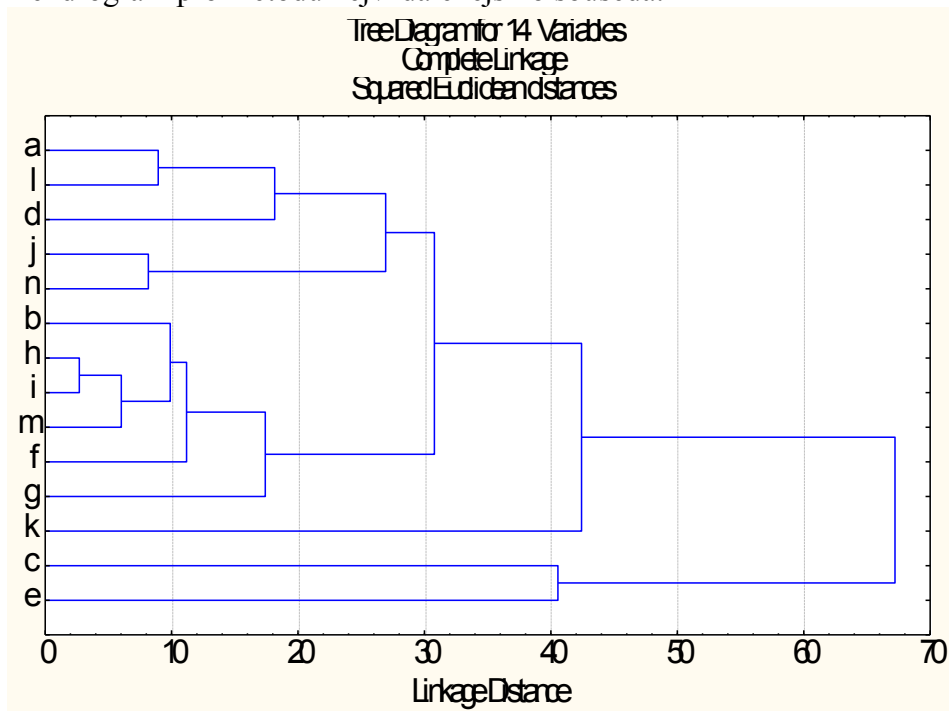
**Úkol 4.:** Použijte metodu nejbližšího souseda, nejbzdálenějšího souseda, metodu průměrné vazby a Wardovu metodu pro nalezení shluků zemí podobných z hlediska tolerance. Výsledky znázorněte pomocí dendrogramů.

Dendrogram pro Wardovu metodu:



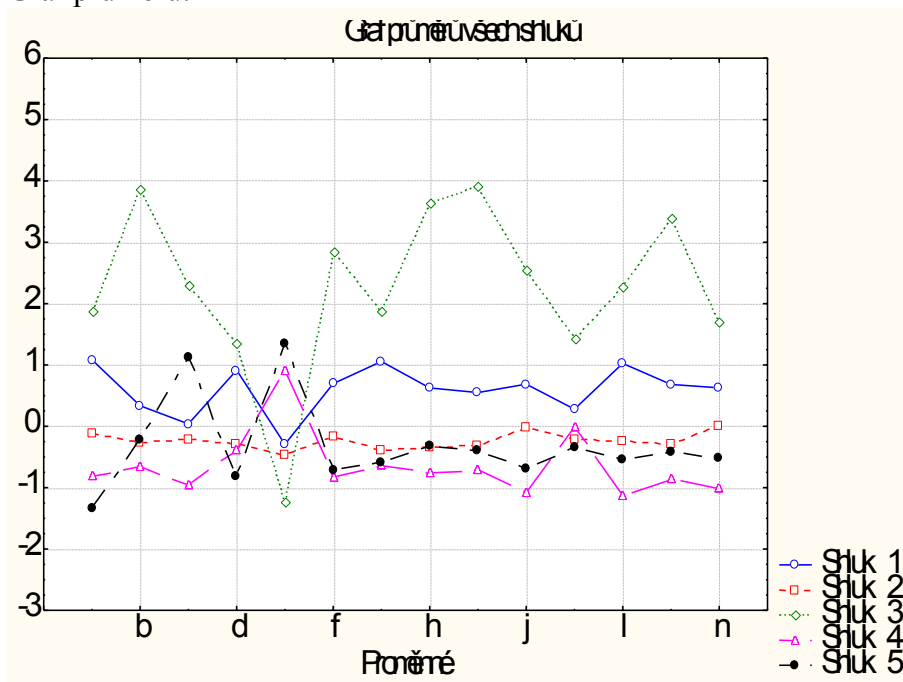
**Úkol 5.:** Určete 5 shluků navzájem si podobných zemí.

**Úkol 6.:** Proveďte shlukovou analýzu pro proměnné.  
Dendrogram pro metodu nejvzdálenějšího souseda:



**Úkol 7.:** Použijte metodu k-průměrů k nalezení 5 shluků navzájem si podobných zemí a uložte skupinovou příslušnost do datového souboru. K určení významnosti jednotlivých proměnných proveďte analýzu rozptylu. Nakreslete graf průměrů všech 5 shluků a pokuste se o interpretaci.

Graf průměrů:



## Příklad k samostatnému řešení:

(Příklad je převzat z knihy M. Meloun, J. Militký, M. Hill: Počítačová analýza vícerozměrných dat. Academia Praha 2005)

U 12 velmi slavných amerických hráčů košikové byly v sezóně 1989 zjištěny hodnoty osmi proměnných.

Výška – výška hráče v cm

Hmotnost – hmotnost hráče v kg

FgPct – první antropometrická charakteristika

FtPct – druhá antropometrická charakteristika

Body – průměrný počet dosažených bodů

Doskoky - průměrný počet doskoků

Asistence – průměrný počet asistencí

Fauly – průměrný počet faulů

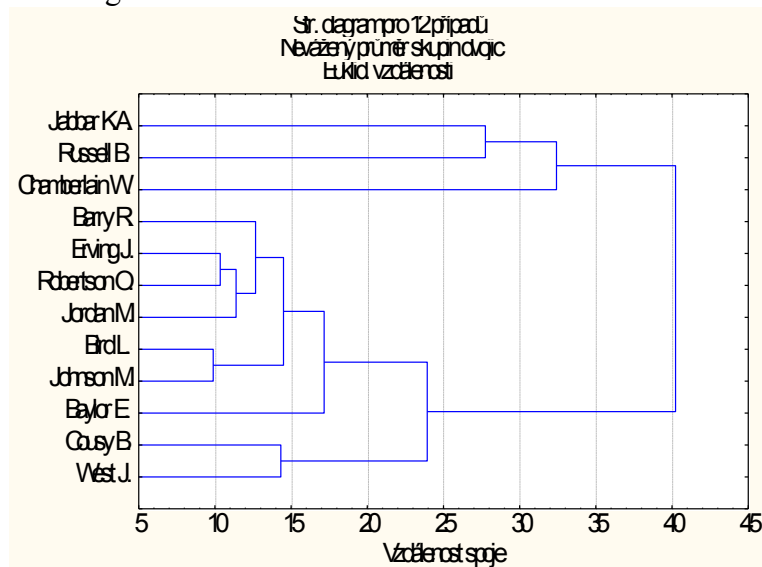
Data jsou uložena v souboru hraci.sta.

	1	2	3	4	5	6	7	8	9
	Jméno hráče	Výšk	Hmotn	Fgp	Ftp	Boc	Doskc	Asister	Fau
1	Jabbar K.	218	105	55	72	24	11	3	3
2	Barry R.	200	93	44	90	23	6	4	3
3	Baylor E.	195	102	43	78	27	13	4	3
4	Bird L.	205	100	50	88	25	10	6	2
5	Chamberlain W.	216	125	54	51	30	22	4	2
6	Cousy B.	184	79	37	80	18	5	7	2
7	Erving J.	199	91	50	77	24	8	4	2
8	Johnson M.	205	98	53	83	19	7	11	2
9	Jordan M.	198	89	51	84	32	6	5	3
10	Robertson O.	195	95	48	83	25	7	9	2
11	Russell B.	207	100	44	56	15	22	4	2
12	West J.	189	82	47	81	27	5	6	2

Metodou průměrné vazby s euklidovskými vzdálenostmi najděte 3 skupiny hráčů podobných vlastností. Výsledek ověřte metodou k-průměrů.

Výsledky

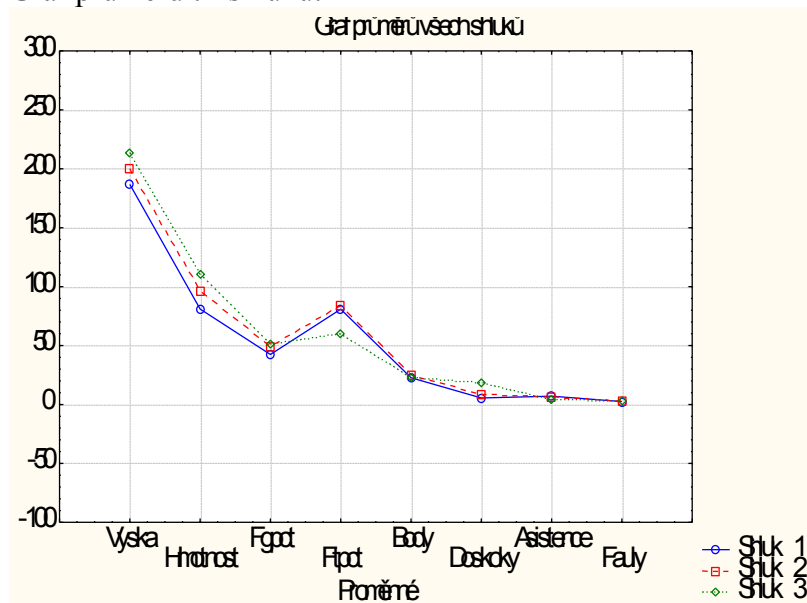
Dendrogram:



Rozdělení hráčů do 3 shluků metodou k-průměrů:

Cleny shluku cislo 1 (hrac a vzdálenosti od příslušné Shluk obsahuje 2 příp.	
	Vzdálenost
Cousy B.	2,532
West J.	2,532
Cleny shluku cislo 2 (hrac a vzdálenosti od příslušné Shluk obsahuje 7 příp.	
	Vzdálenost
Barry R.	2,995
Baylor E.	4,557
Bird L.	3,089
Erving J.	2,877
Johnson M.	3,738
Jordan M.	3,819
Robertson C.	1,951
Cleny shluku cislo 3 (hrac a vzdálenosti od příslušné Shluk obsahuje 3 příp.	
	Vzdálenost
Jabbar K.A.	5,967
Chamberlain W.	6,905
Russell B.	6,030

Graf průměrů tří shluků:



Tabulka ANOVA:

Analýza rozptylu (hraci.sta)					
Proměření	Mezi s' SC	Vnitřní s' SC	F	vyzna p	
Vyska	905,4	194,4	20,95	0,000	
Hmotnost	1051,2	505,9	9,34	0,006	
Fgpct	97,2	207,9	2,10	0,177	
Ftpct	1232,2	368,0	15,07	0,001	
Body	16,2	249,3	0,29	0,752	
Doskovy	287,4	127,7	10,12	0,004	
Asistence	15,6	44,9	1,56	0,261	
Faly	0,2	0,92	1,32	0,312	