Odhalení skryté struktury a vnitřních vazeb dat metodami PCA, FA vícerozměrné statistické analýzy

Prof. RNDr. Milan Meloun, DrSc.,

Katedra analytické chemie, Univerzita Pardubice, 532 10 Pardubice, milan.meloun@upce.cz

a

Prof. Ing. Jiří Militký, CSc.,

Katedra textilních materiálů, Technická univerzita Liberec, 461 17 Liberec, jiri.militky@tul.cz

Postup analýzy vícerozměrných dat

1. Standardizace: analýze vždy předchází standardizace čili škálování proměnných.

2. Odhady parametrů polohy, rozptýlení, tvaru a intenzita vztahu mezi proměnnými:

Vyčíslení výběrové střední hodnoty každé proměnné.
Odhad kovarianční matice ***S*** a její normované podoby - korelační matice ***R*.**
Odhadu vícerozměrné šikmosti a vícerozměrné špičatosti.
Matice ***R*** obsahuje Pearsonovy párové korelační koeficienty, které se diskutují.

3. Exploratorní analýza dat EDA:

(a) Hledání podobnosti objektů vizuálními rozptylovými diagramy typu casement plot, draftsman plot, dále symbolových a profilových grafů (hvězdičky, sluníčka, obličeje, křivky, stromy),

(b) Nalezení vybočujících objektů nebo vybočujících proměnných, mnohdy nevhodných k analýze,

(c) Testy předpokladů lineárních vazeb,

(d) Testy předpokladů o datech (normalitu, nekorelovanost, homogenitu).
Ověřování normality založené na vícerozměrné šikmosti a vícerozměrné špičatosti.

4. Určení vhodného počtu latentních proměnných:

a) Matice ***S*** nebo ***R*** se rozloží na vlastní čísla a vlastní vektory.

b) Indexový graf úpatí vlastních čísel (Scree plot): určí vhodný počet latentních proměnných, které ještě dostatečně popisují proměnlivost v datech.

c) Když se latentní proměnné podaří pojmenovat a dát jim i fyzikální, biologický či jiný věcný význam, jedná se o faktory. Jinak jde o hlavní komponenty.

5. Určení struktury v proměnných (PCA a FA):

Graf komponentních vah (Plot of components weights, loadings): hledání struktury a vzájemných vazeb (korelace) proměnných se provede v grafu

a) Rozptylový diagram komponentního skóre (Scatterplot): hledání struktury v objektech a třídění objektů do shluků.

b) Dvojný graf (Biplot) je přehledným spojením obou předešlých grafů a ukáže interakci objektů a proměnných.

Určení vzájemných vazeb

(a) *struktura a vazby v proměnných*
(b) *struktura a vazby v objektech*

(1) Hledání struktury v proměnných (metrická škála): faktorová analýza FA, analýza hlavních komponent PCA a shluková analýza.

(2) Hledání struktury v objektech (metrická škála): shluková analýza.

(3) Hledání struktury v objektech (metrická i nemetrická škála): vícerozměrné škálování.

(4) Hledání struktury v objektech (nemetrická škála): korespondenční analýza.

(5) Hledání lineárních vícerozměrných modelů (metrická i nemetrická škála): většina metod vícerozměrné statistické analýzy, kde závisle proměnné se uvažují jako lineárn kombinace nezávisle proměnných.

5) ***Pravidlo:*** má-li nějaký původní znak malý či dokonce žádný rozptyl, není schopen přispívat k rozlišení mezi objekty.

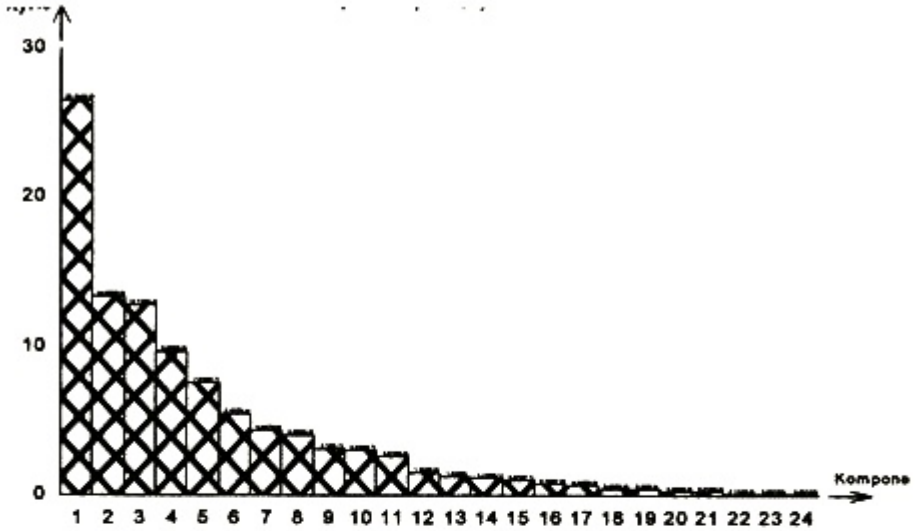
6) Využitím PCA je *snížení dimenze úlohy* čili redukce počtu znaků bez velké ztráty informace, užitím pouze prvních několika hlavních komponent.

7) *Nevyužité hlavní komponenty* obsahují malé množství informace, protože jejich rozptyl je příliš malý.

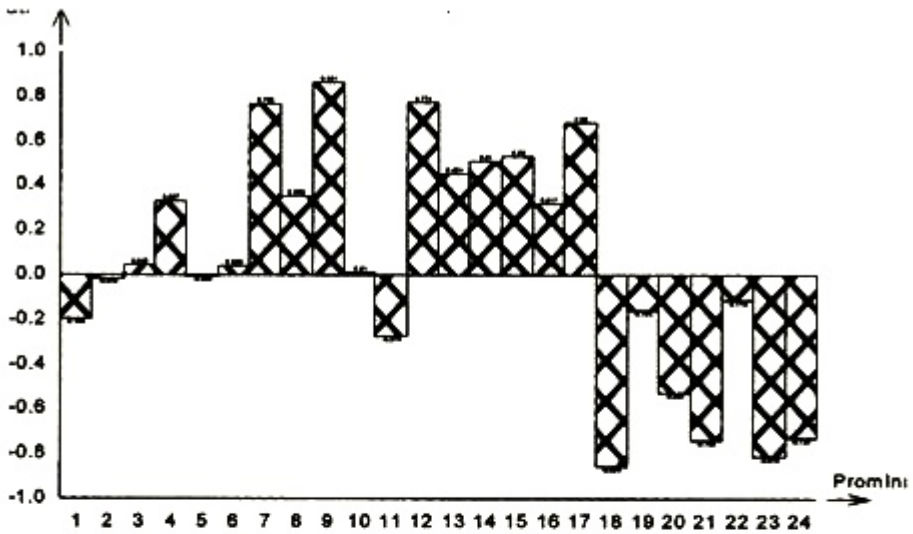
8) Hlavní komponenty jsou nekorelované.

9) První hlavní komponenta je například vhodným ukazatelem jakosti.

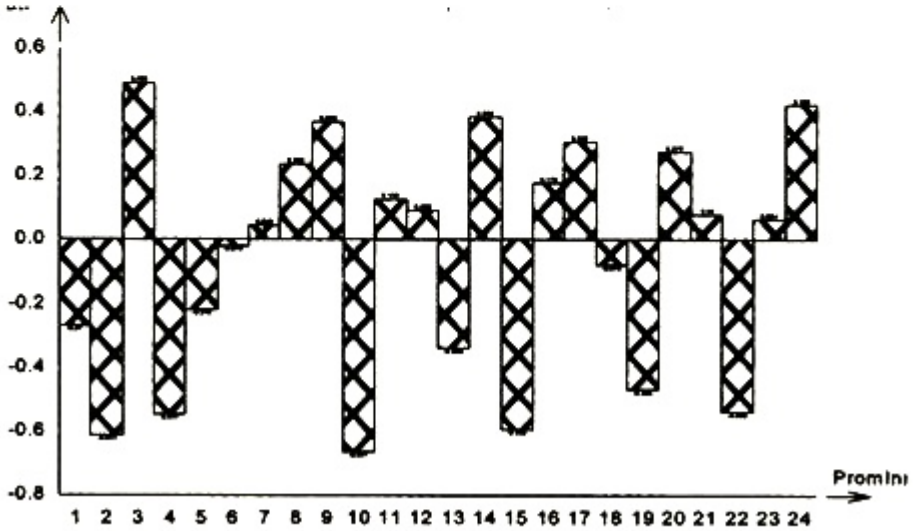
10) První dvě resp. první tři hlavní komponenty se využívají především jako techniky zobrazení vícerozměrných dat v projekci do roviny (nebo do prostoru).



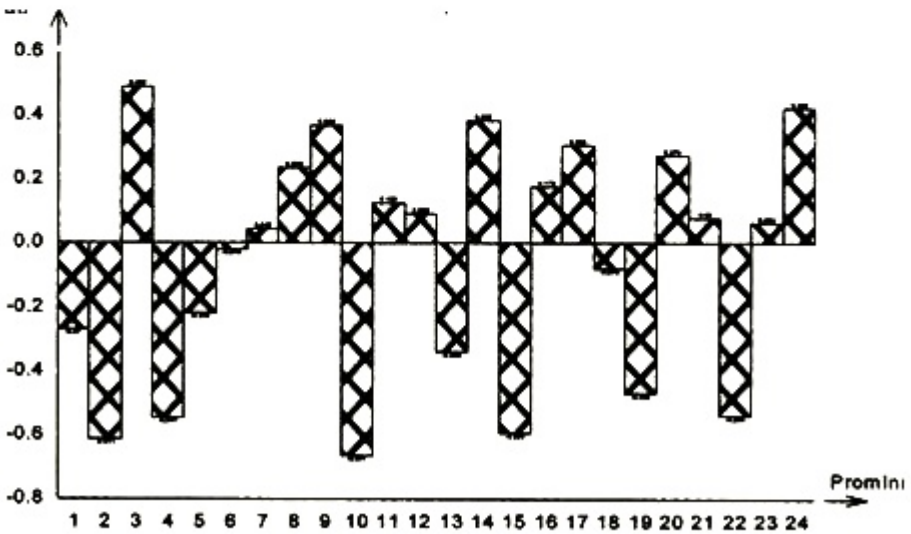
Obr. 1. Sloupcový diagram indexového grafu úpatí pro 38 objektů a 24 původních proměnných zdrojové matice Wine.



Obr. 2. Složení 1. hlavní komponenty z původních proměnných pro 38 objektů a 24 původních proměnných zdrojové matice Wine.



Obr. 3. Složení 2. hlavní komponenty z původních proměnných pro 38 objektů a 24 původních proměnných zdrojové matice Wine.



Obr. 4. Složení 3. hlavní komponenty z původních proměnných pro 38 objektů a 24 původních proměnných zdrojové matice Wine.

Výklad:

Graf komponentních vah, zátěží

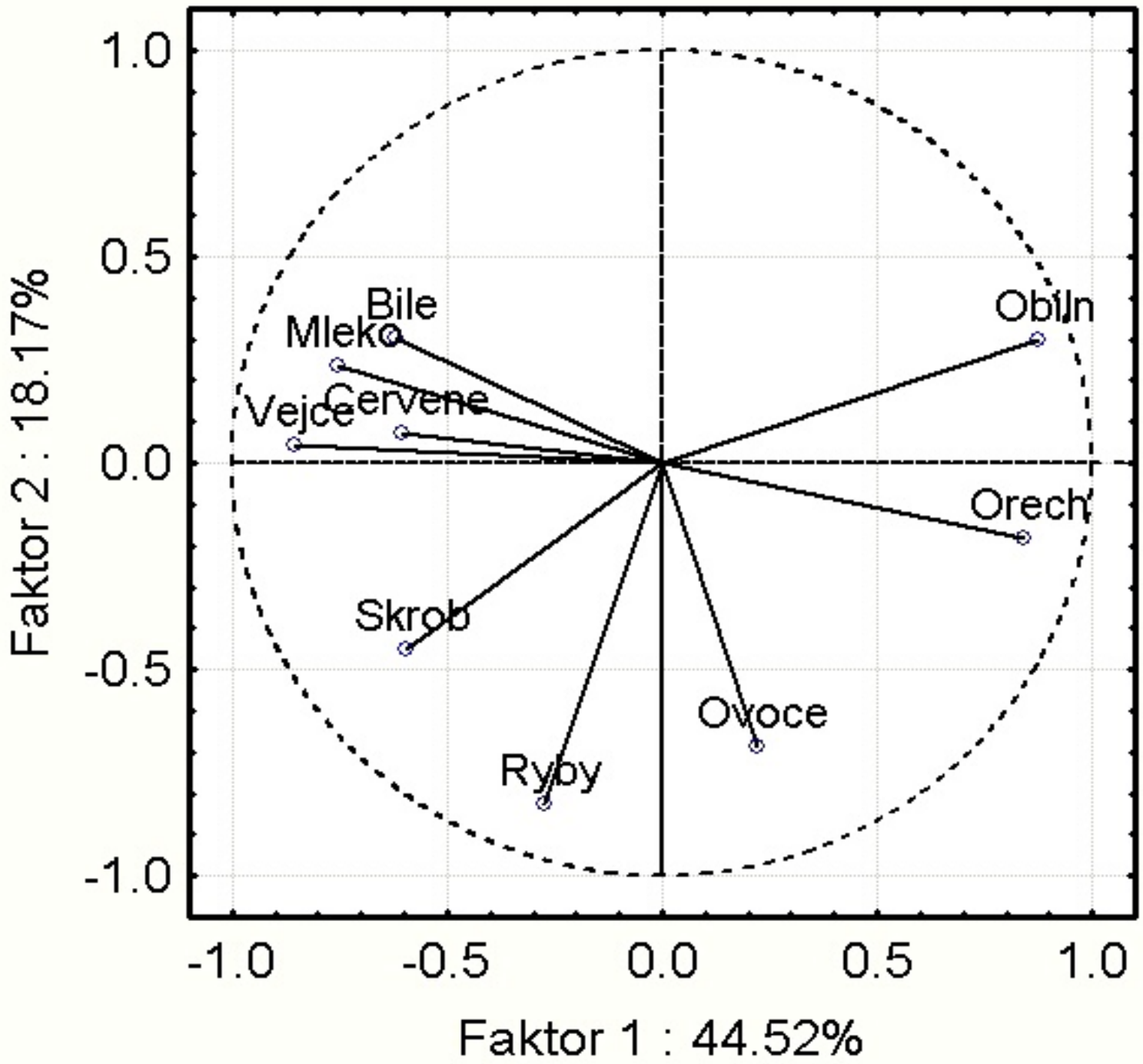
(Plot Components Weights)

Zobrazí: komponentní váhy

Porovnávají se: vzdálenosti mezi proměnnými.
Krátká znamená silnou korelaci.

Nalezneme: shluk podobných proměnných, jež spolu korelují.

Představuje: most mezi původními proměnnými a hlavními komponentami.


1.0
0.5
0.0
-0.5
-1.0
Faktor 2 : 18.17%
Bile
Mleko
Vejce
Cervene
Obilin
Orech
Skrob
Ovoce
Ryby
-1.0
-0.5
0.0
0.5
1.0
Faktor 1 : 44.52%

Rozptylový diagram komponentního skóre
(Scatterplot)

1. **Umístění objektů:** daleko od počátku jsou extrémy. Objekty nejblíže počátku jsou nejtypičtější.

2. **Podobnost objektů:** objekty blízko sebe si jsou podobné, daleko od sebe jsou si nepodobné.

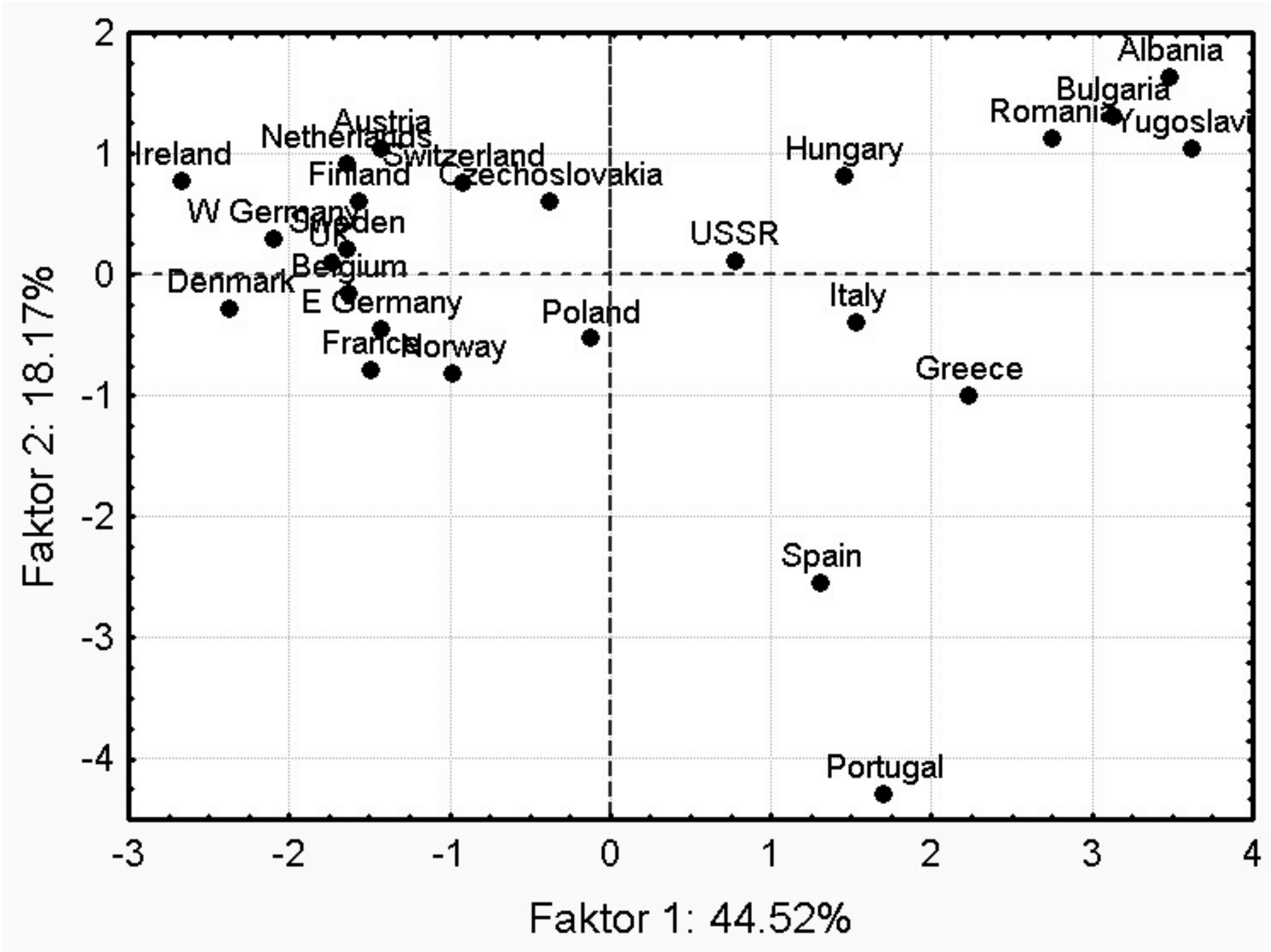
3. **Objekty v shluku:** umístěné zřetelně v jednom shluku jsou si podobné a nepodobné objektům v ostatních shlucích. Jsou-li shluky blízko sebe, znamená to značnou podobnost objektů.

4. **Osamělé objekty:** izolované objekty mohou být odlehlé.

5. **Odlehlé objekty:** ideálně bývají objekty rozptýlené po celé ploše diagramu. V opačném případě je špatný model.

6. **Pojmenování objektů:** výstižná jména objektů slouží k hledání hlubších souvislostí mezi objekty a vystihneme tak jejich fyzikální či biologický vztah.

7. **Vysvětlení místa objektu:** umístění objektu na ploše v diagramu může být porovnáváno s komponentními vahami původních proměnných ve dvojném grafu.


Albania
Bulgaria
Romania
Yugoslavi
Hungary
Austria
Netherlands
Switzerland
Ireland
Finland
Czechoslovakia
W Germany
Sweden
UK
USSR
Belgium
Denmark
E Germany
Italy
Poland
France
Norway
Greece
Spain
Portugal
Faktor 2: 18.17%
Faktor 1: 44.52%
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

Výklad:

Indexový graf úpatí vlastních čísel

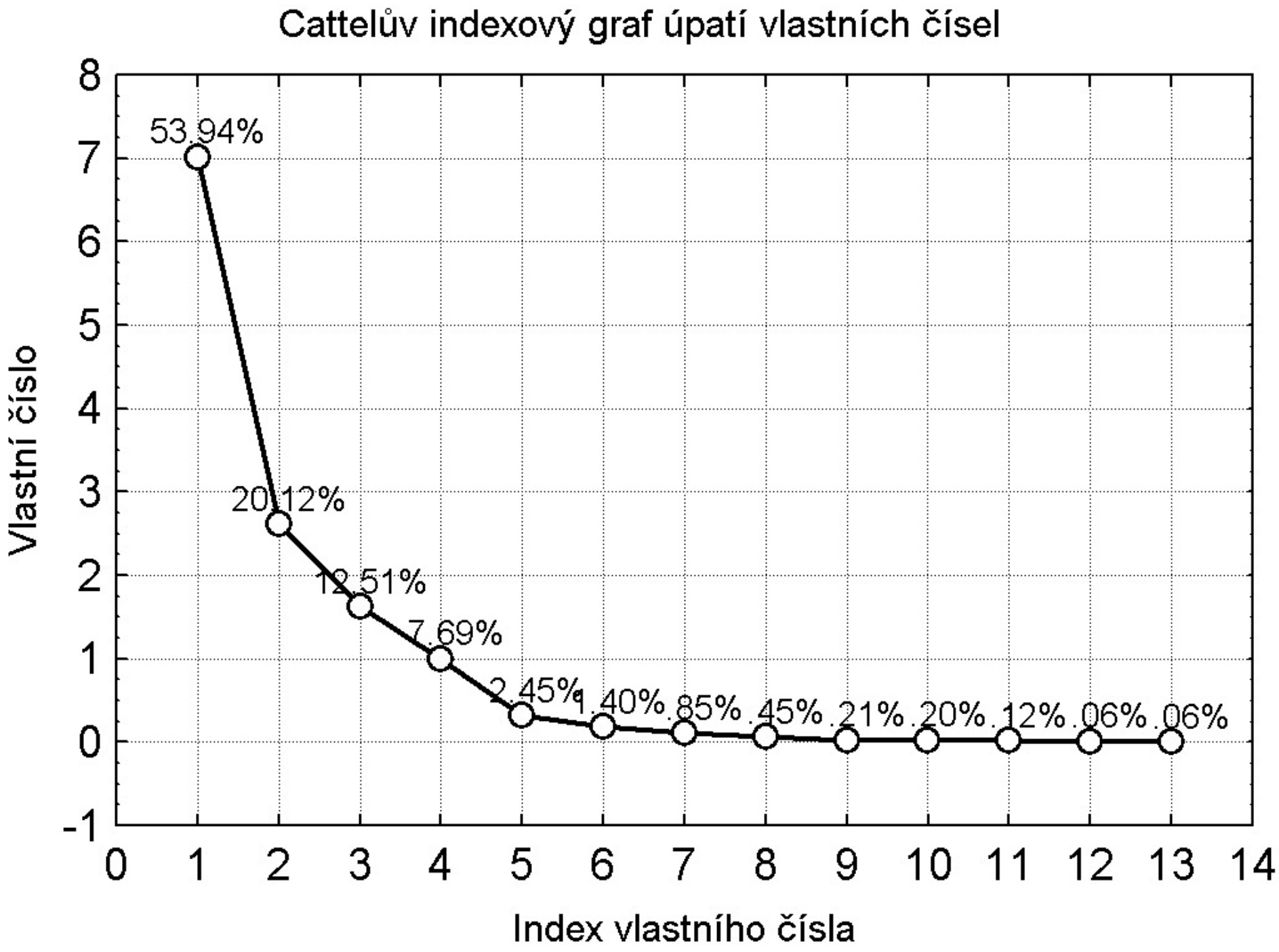
(Scree Plot)

Je to sloupcový diagram vlastních čísel proti indexu *A*.

Zobrazuje: relativní velikost jednotlivých vlastních čísel.

Využití: k určení počtu *A* "užitečných" hlavních komponent.
Graf úpatí se jeví nejobjektivnějším kritériem.

Kritérium “1": hrubším kritériem PC, jejichž vlastní číslo je větší než jedna. Graf úpatí se však jeví objektivnějším.


Cattelův indexový graf úpatí vlastních čísel
Vlastní číslo
Index vlastního čísla
53.94%
20.12%
12.51%
7.69%
2.45%
1.40%
.85%
.45%
.21%
.20%
.12%
.06%
.06%
8
7
6
5
4
3
2
1
0
-1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Diagnostika metody PCA

Maticový graf rozptylových diagramů znaků slouží k získání počáteční informace o datech, zda data potřebují škálování. V PCA postupně provádíme:

1. ***Vyšetření indexového grafu úpatí vlastních čísel*** – z hrany úpatí v tomto diagramu se určí vhodný počet hlavních komponent.

2. ***Výpočet vlastních vektorů*** – vedle číselných hodnot se užívá i názorný čárový diagram hodnot vlastních vektorů, který přehledně informuje o relativním zastoupení původních znaků x_j , $j = 1, ..., m$, v hlavních komponentách.

3. ***Výpočet komponentních vah*** – matice párových korelačních koeficientů obsahující korelace původních znaků s hlavními komponentami. Uživatel nyní vybere pouze prvních k hlavních komponent a vytvoří tak model PCA.

4. ***Vyšetření grafu komponentních vah.***

5. ***Vyšetření rozptylového diagramu komponentního skóre.***

6. ***Vyšetření dvojného grafu.***

7. ***Vyšetření reziduí*** – rezidua objektů a rezidua proměnných by měla prokazovat dostatečnou těsnost proložení.

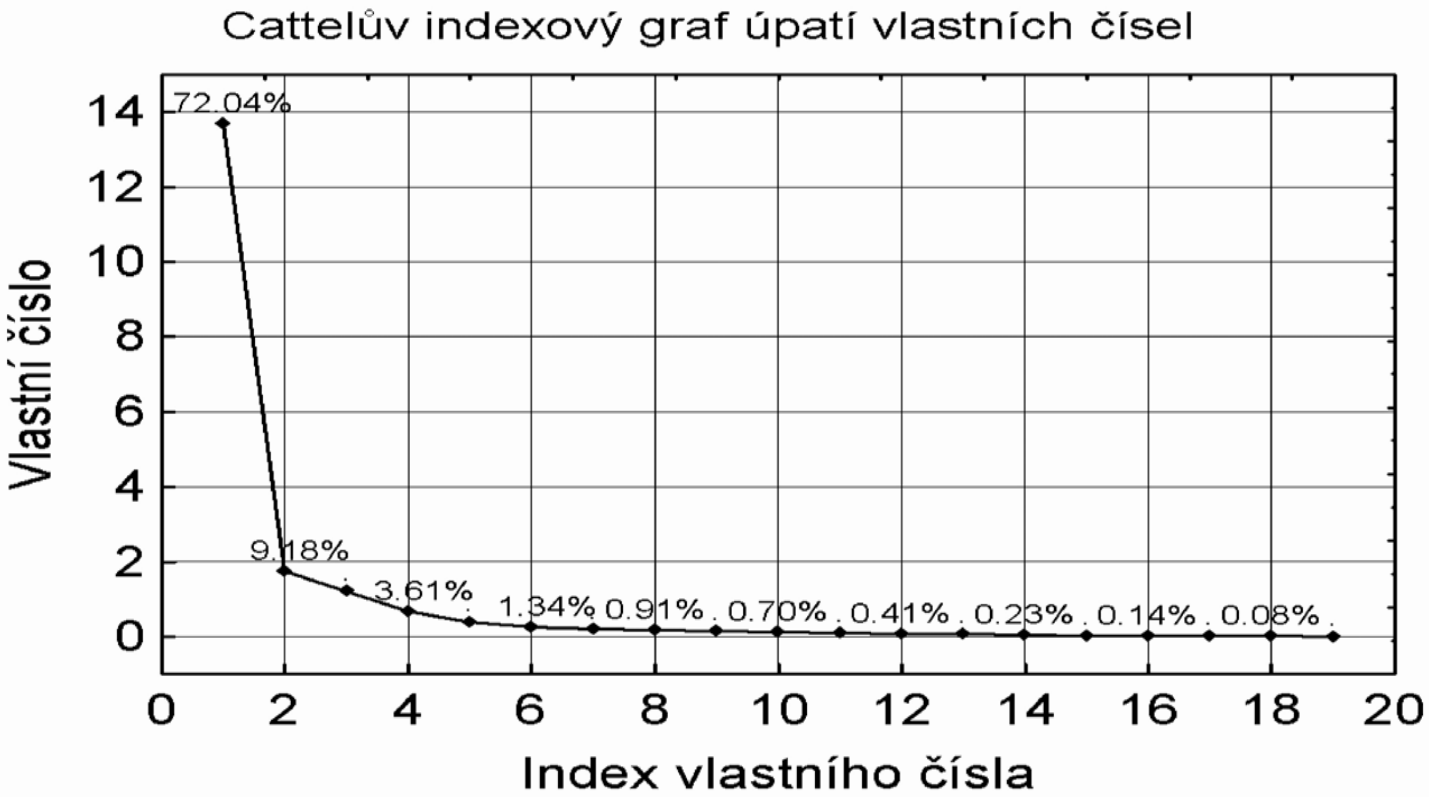
8. ***Určení významných původních znaků*** – je výhodné vyhledávat významné znaky, protože klasická metoda PCA umožňuje sice redukci počtu hlavních komponent, ale každá komponenta zůstává stále kombinací všech původních znaků.

Úloha 1. Klasifikace polétavých mšic

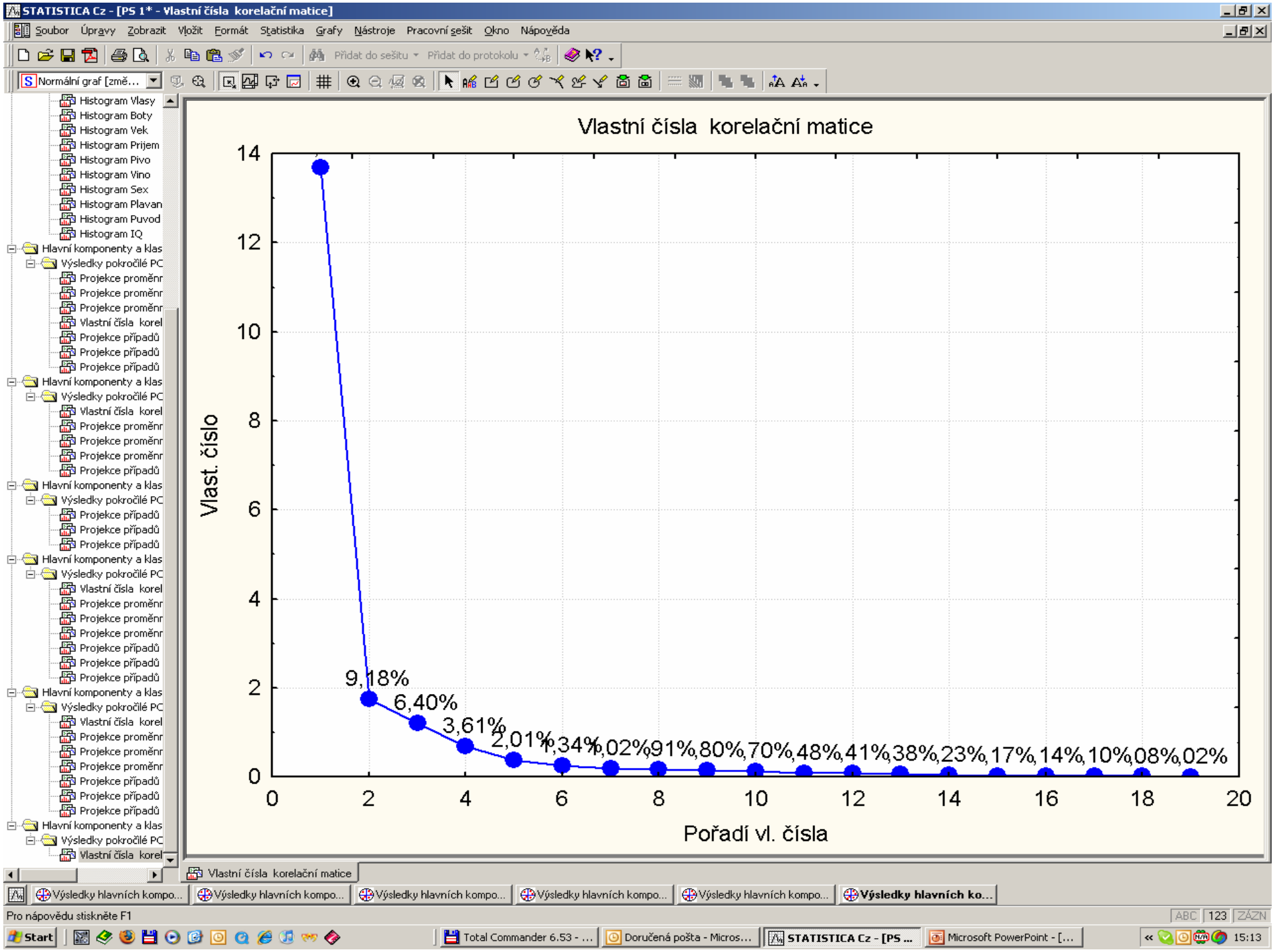
Jeffers (1967)25 studoval 40 jedinců polétavých mšic *(Alate adelges)* : 19 ukazatelů k rozlišení druhů, 14 znaků délky a šířky, 4 znaky se týkají počtu a 1 binární vyjadřuje přítomnost či absenci: x1 délka těla, x2 šířka těla, x3 délka předního křídla, x4 délka zadního křídla, x5 počet průduchů, x6 délka tykadla I, x7 délka tykadla II, x8 délka tykadla III, x9 délka tykadla IV, x10 délka tykadla V, x11 počet tykadlových ostnů, x12 délka posledního článku nohy, x13 délka holeně, tibia, x14 délka stehna, x15 délka sosáku, x16 délka kladélka, x17 počet kladélkových trnů, x18 řitní otvor, x19 počet háčků zadních křídel

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19
21.2	11	7.5	4.8	5	2	2	2.8	2.8	3.3	3	4.4	4.5	3.6	7	4	8	0	3
20.2	10	7.5	5	5	2.3	2.1	3	3	3.2	5	4.2	4.5	3.5	7.6	4.2	8	0	3
20.2	10	7	4.6	5	1.9	2.1	3	2.5	3.3	1	4.2	4.4	3.3	7	4	6	0	3
22.5	8.8	7.4	4.7	5	2.4	2.1	3	2.7	3.5	5	4.2	4.4	3.6	6.8	4.1	6	0	3
20.6	11	8	4.8	5	2.4	2	2.9	2.7	3	4	4.2	4.7	3.5	6.7	4	6	0	3
19.1	9.2	7	4.5	5	1.8	1.9	2.8	3	3.2	5	4.1	4.3	3.3	5.7	3.8	8	0	3.5
20.8	11.4	7.7	4.9	5	2.5	2.1	3.1	3.1	3.2	4	4.2	4.7	3.6	6.6	4	8	0	3
15.5	8.2	6.3	4.9	5	2	2	2.9	2.4	3	3	3.7	3.8	2.9	6.7	3.5	6	0	3.5
16.7	8.8	6.4	4.5	5	2.1	1.9	2.8	2.7	3.1	3	3.7	3.8	2.8	6.1	3.7	8	0	3
19.7	9.9	8.2	4.7	5	2.2	2	3	3	3.1	0	4.1	4.3	3.3	6	3.8	8	0	3
10.6	5.2	3.9	2.3	4	1.2	1	2	2	2.2	6	2.5	2.5	2	4.5	2.7	4	1	2
9.2	4.5	3.7	2.2	4	1.3	1.2	2	1.6	2.1	5	2.4	2.3	1.8	4.1	2.4	4	1	2
9.6	4.5	3.6	2.3	4	1.3	1	1.9	1.7	2.2	4	2.4	2.3	1.7	4	2.3	4	1	2
8.5	4	3.8	2.2	4	1.3	1.1	1.9	2	2.1	5	2.4	2.4	1.9	4.4	2.3	4	1	2
11	4.7	4.2	2.3	4	1.2	1	1.9	2	2.2	4	2.5	2.5	2	4.5	2.6	4	1	2
18.1	8.2	5.9	3.5	5	1.9	1.9	1.9	2.7	2.8	4	3.5	3.8	2.9	6	4.5	9	1	2
17.6	8.3	6	3.8	5	2	1.9	2	2.2	2.9	3	3.5	3.6	2.8	5.7	4.3	10	1	2
19.2	6.6	6.2	3.4	5	2	1.8	2.2	2.3	2.8	4	3.5	3.4	2.5	5.3	3.8	10	1	2
15.4	7.6	7.1	3.4	5	2	1.9	2.5	2.5	2.9	4	3.3	3.6	2.7	6	4.2	8	1	3
15.1	7.3	6.2	3.8	5	2	1.8	2.1	2.4	2.5	4	3.7	3.7	2.8	6.4	4.3	10	1	2.5
16.1	7.9	5.8	3.7	5	2.1	1.9	2.3	2.6	2.9	5	3.6	3.6	2.7	6	4.5	0	1	2
19.1	8.8	6.4	3.9	5	2.2	2	2.3	2.4	2.9	4	3.8	4	3	6.5	4.5	0	1	2.5
15.3	6.4	5.3	3.3	5	1.7	1.6	2	2.2	2.5	5	3.4	3.4	2.6	5.4	4	0	1	2
14.8	8.1	6.2	3.7	5	2.2	2	2.2	2.4	3.2	5	3.5	3.7	2.7	6	4.1	0	1	2
16.2	7.7	6.9	3.7	5	2	1.8	2.3	2.4	2.8	4	3.8	3.7	2.7	5.7	4.2	0	1	2.5
13.4	6.9	5.7	3.4	5	2	1.8	2.8	2	2.6	4	3.6	3.6	2.6	5.5	3.9	0	1	2
12.9	5.8	4.8	2.6	5	1.6	1.5	1.9	2.1	2.6	5	2.8	3	2.2	5.1	3.6	9	1	3
12	6.5	5.3	3.2	5	1.9	1.9	2.3	2.5	3	5	3.3	3.5	2.6	5.4	4.3	8	1	2
14.1	7	5.5	3.6	5	2.2	2	2.3	2.5	3.1	5	3.6	3.7	2.8	5.8	4.1	0	1	2
16.7	7.2	5.7	3.5	5	1.9	1.9	2.5	2.3	2.8	5	3.4	3.6	2.7	6	4	0	1	2.5
14.1	5.4	5	3	5	1.7	1.6	1.8	2.5	2.4	5	2.7	2.9	2.2	5.3	3.6	8	1	2
10	6	4.2	2.5	5	1.6	1.4	1.4	2	2.7	6	2.8	2.5	1.8	4.8	3.4	8	1	2
11.4	4.5	4.4	2.7	5	1.8	1.5	1.9	1.7	2.5	5	2.7	2.5	1.9	4.7	3.7	8	1	2
12.5	5.5	4.7	2.3	5	1.8	1.4	1.8	2.2	2.4	4	2.8	2.6	2	5.1	3.7	8	0	2
13	5.3	4.7	2.3	5	1.6	1.4	1.8	1.8	2.5	4	2.7	2.7	2.1	5	3.6	8	1	2
12.4	5.2	4.4	2.6	5	1.6	1.4	1.8	2.2	2.2	5	2.7	2.5	2	5	3.2	6	1	2
12	5.4	4.9	3	5	1.7	1.5	1.7	1.9	2.4	5	2.7	2.7	2	4.2	3.7	6	1	2
10.7	5.6	4.5	2.8	5	1.8	1.4	1.8	2.2	2.4	4	2.7	2.6	2	5	3.5	8	1	2
11.1	5.5	4.3	2.6	5	1.7	1.5	1.8	1.9	2.4	5	2.6	2.5	1.9	4.6	3.4	8	1	2
12.8	5.7	4.8	2.8	5	1.6	1.4	1.7	1.9	2.3	5	2.3	2.5	1.9	5	3.1	8	1	2

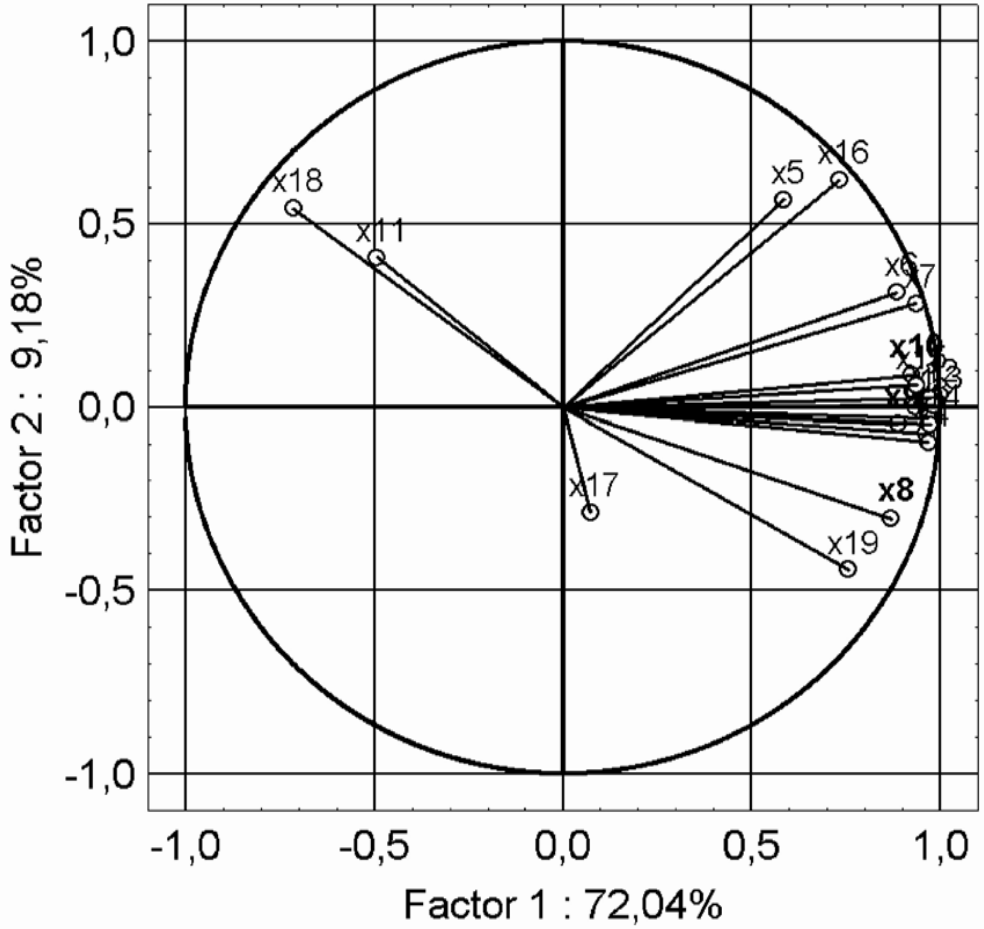
1. Cattelův indexový graf úpatí vlastních čísel: z 19 znaků lze snížit rozměrnost na první dvě hlavní komponenty, které popisují přes 81% původní proměnlivosti v datech.



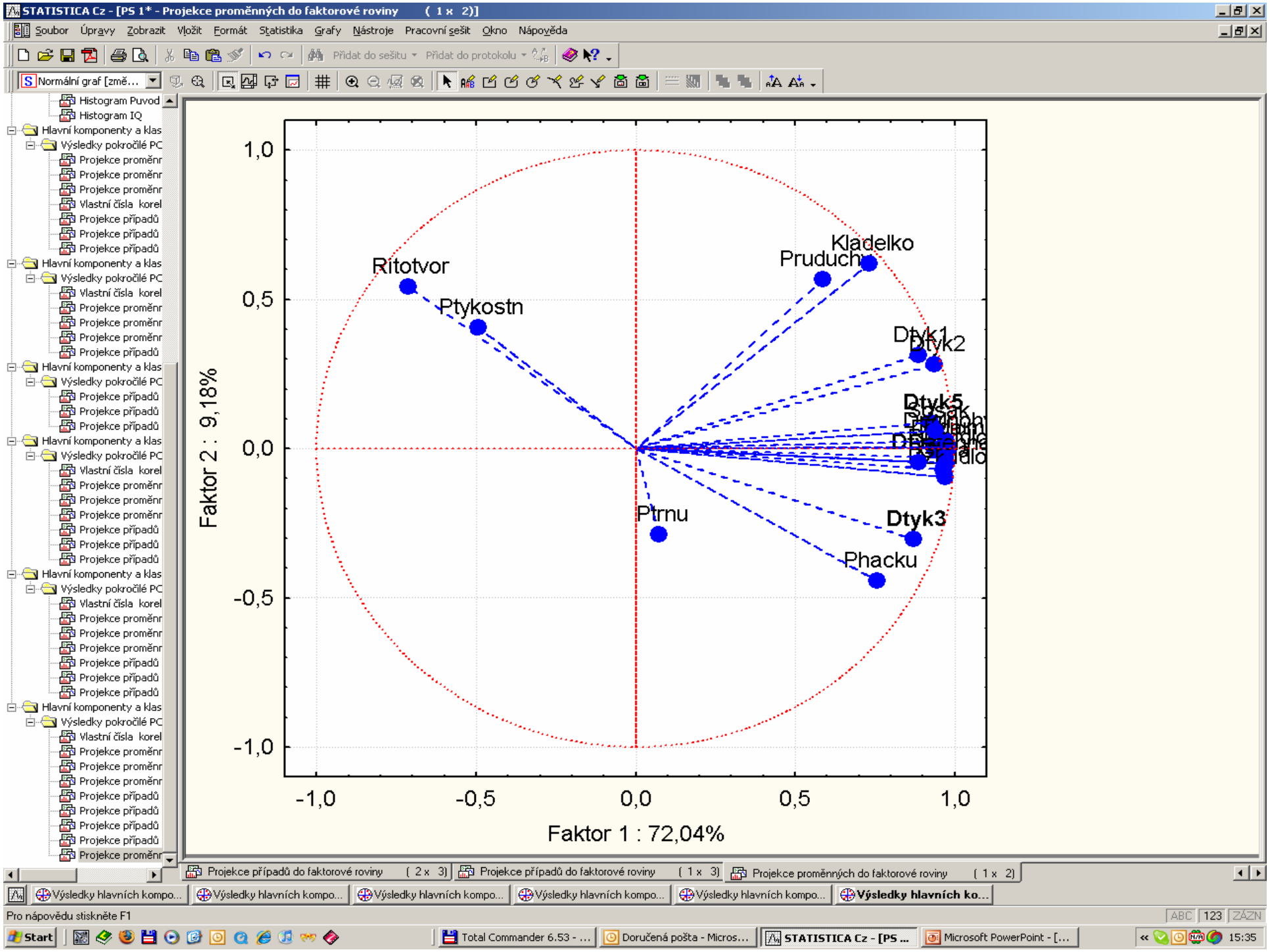
Obr. 4.23 Cattelův indexový graf úpatí vlastních čísel Scree plot dat *Mšice* (STATISTICA).

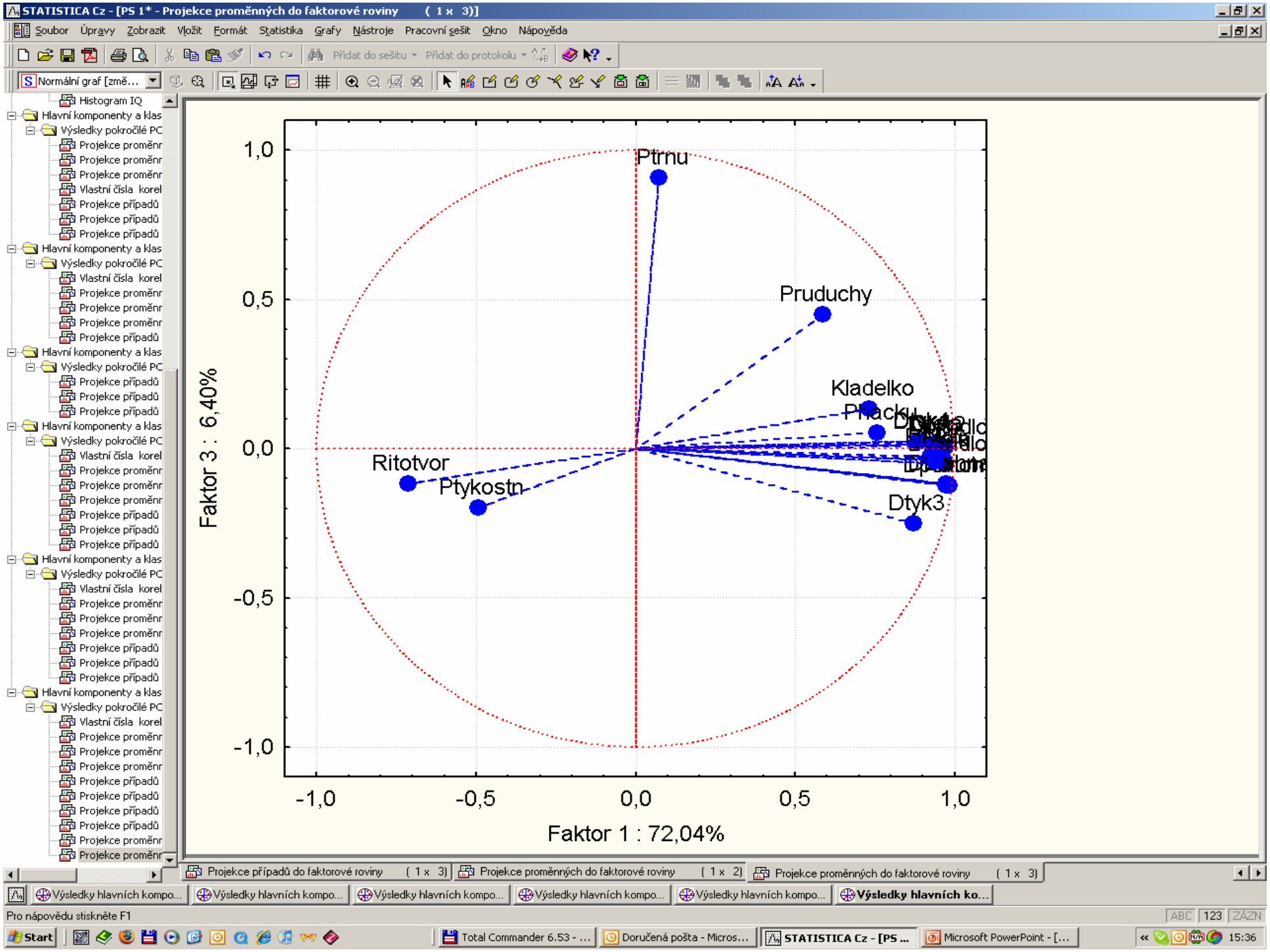

STATISTICA Cz - [PS 1* - Vlastní čísla korelační matice]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Vlastní čísla korelační matice
Vlast. číslo
Pořadí vl. čísla
9,18%
6,40%
3,61%
2,01%
1,34%
,02%
91%
80%
70%
48%
41%
38%
23%
17%
14%
10%
08%
02%
Pro nápovědu stiskněte F1

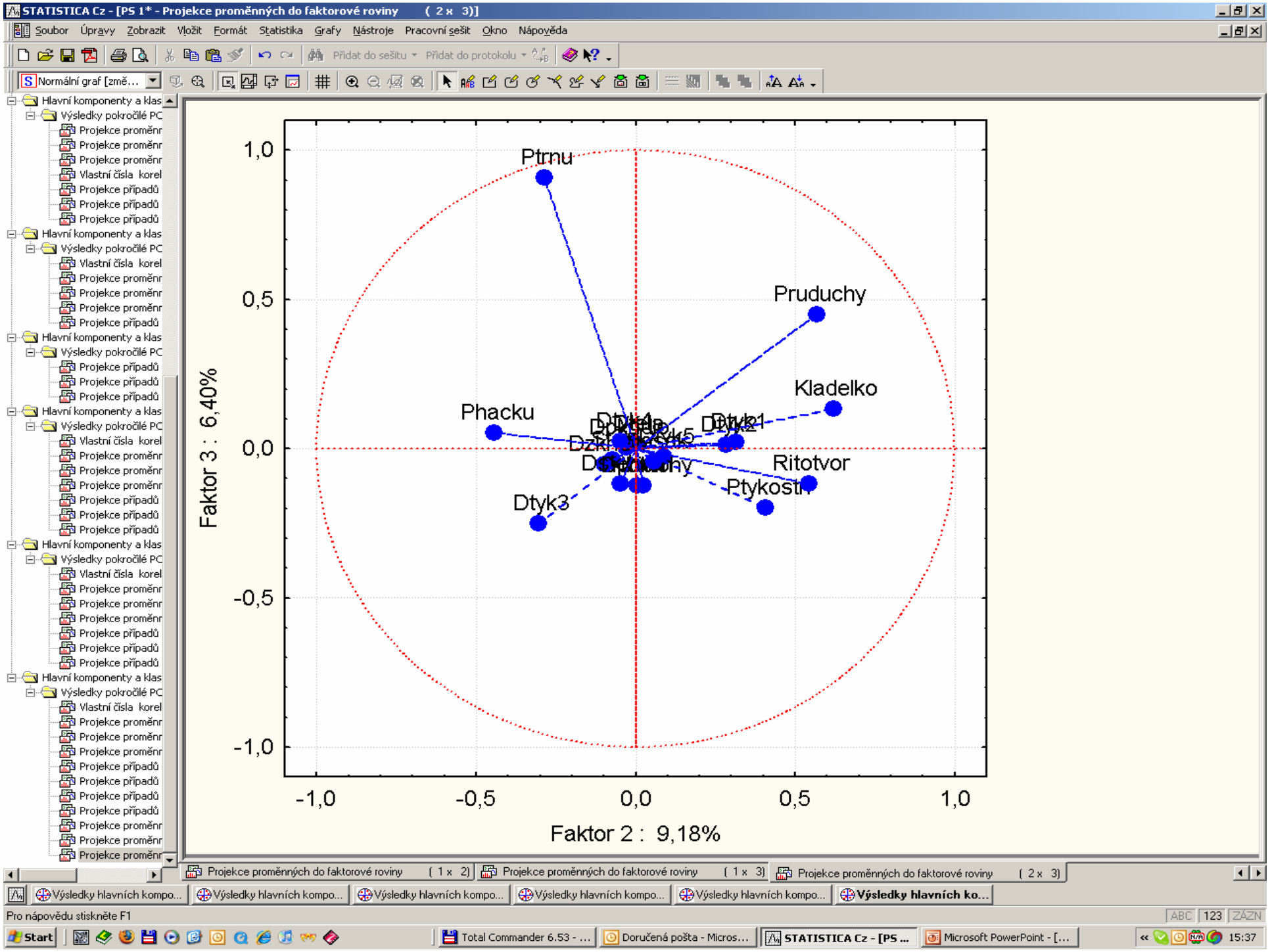
2. Graf komponentních vah: roztřídí 19 znaků: vedle shluku společných znaků jsou x_1 a x_{17} odlehlé od ostatních. Od shluku jsou odděleny znaky x_2 a x_3, a dále x_{11} a x_{13}. Znaky x_2 a x_3 spolu pozitivně korelují, dále x_{11} a x_{18} spolu pozitivně korelují ale negativně korelují se x_1, x_2 a x_3. Znak x_1 pozitivně koreluje s x_2, a x_1 koreluje s x_3.



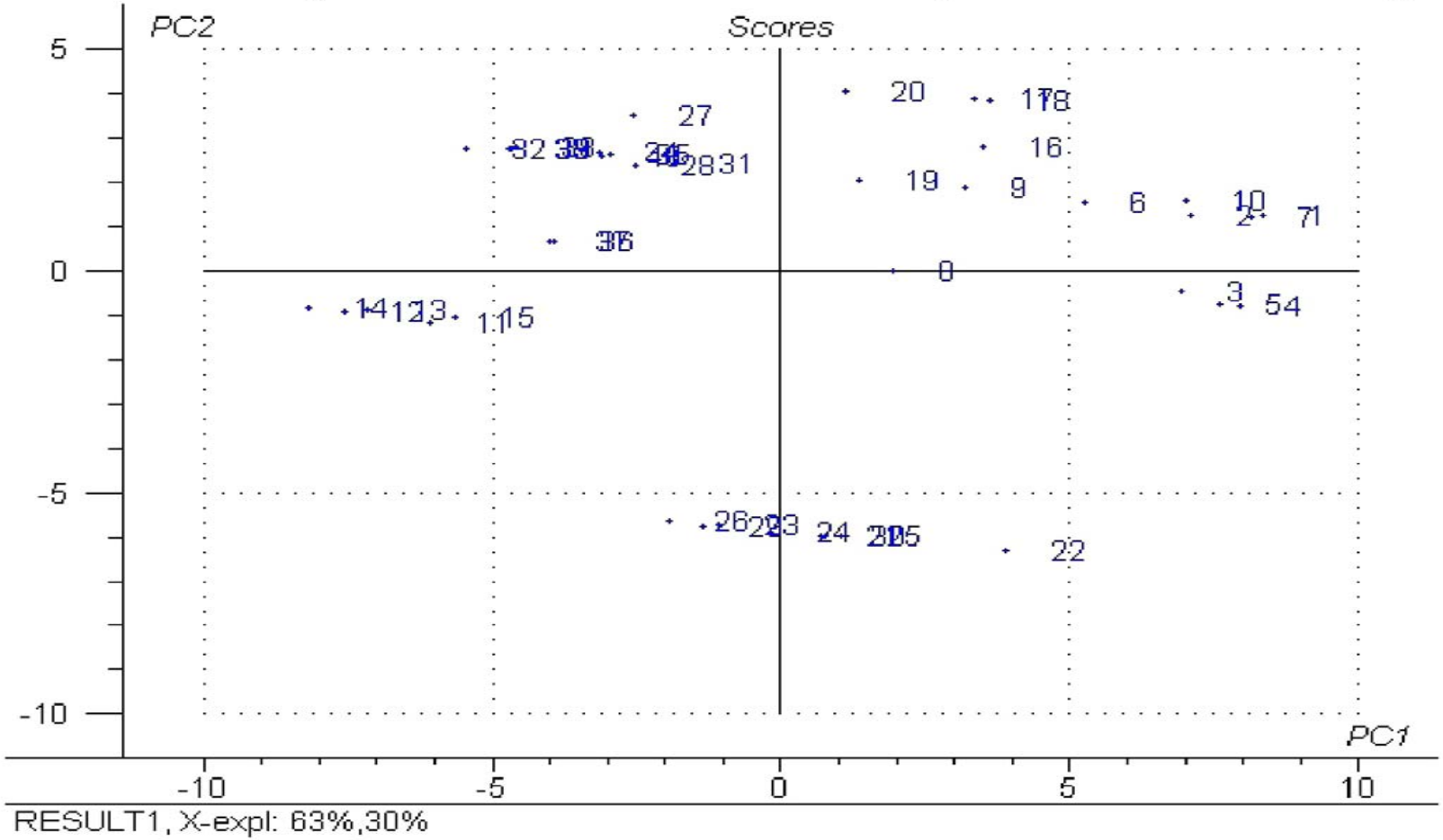
Obr. 4.24 Graf komponentních vah 1 a 2 zdrojové matice dat *Mšice* (STATISTICA).


STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 2)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Ritotvor
Ptykostn
Kladelko
Pruduch
Dtyk1
Dtyk2
Dtyk5
Ptrnu
Dtyk3
Phacku
1,0
0,5
0,0
-0,5
-1,0
Faktor 2 : 9,18%
Faktor 1 : 72,04%
Projekce případů do faktorové roviny (2 x 3)
Projekce případů do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (1 x 2)
Pro nápovědu stiskněte F1

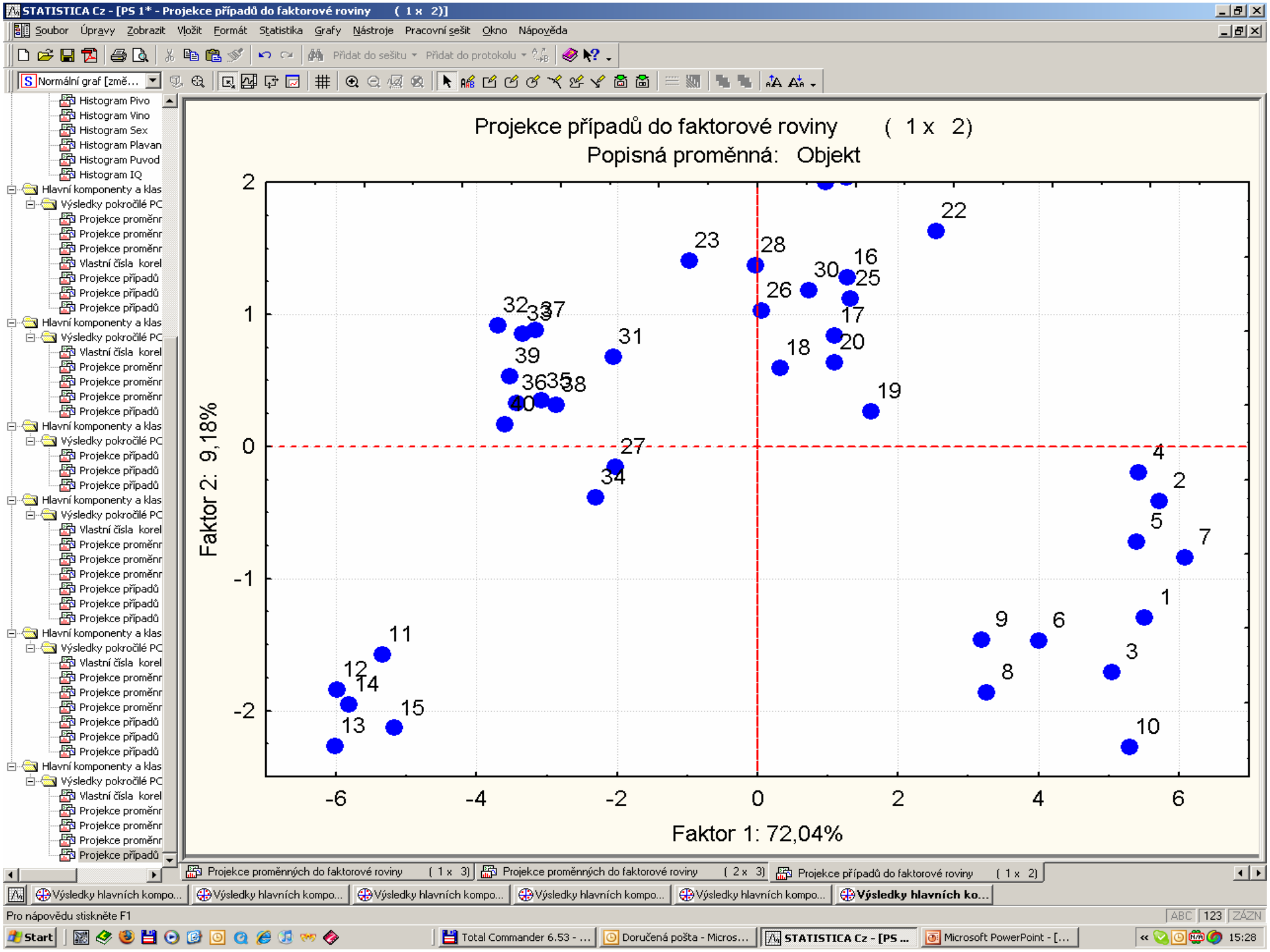

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Ptrnu
Pruduchy
Kladelko
Ritotvor
Ptykostn
Dtyk3
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 6,40%
Faktor 1 : 72,04%
Projekce případů do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Pro nápovědu stiskněte F1

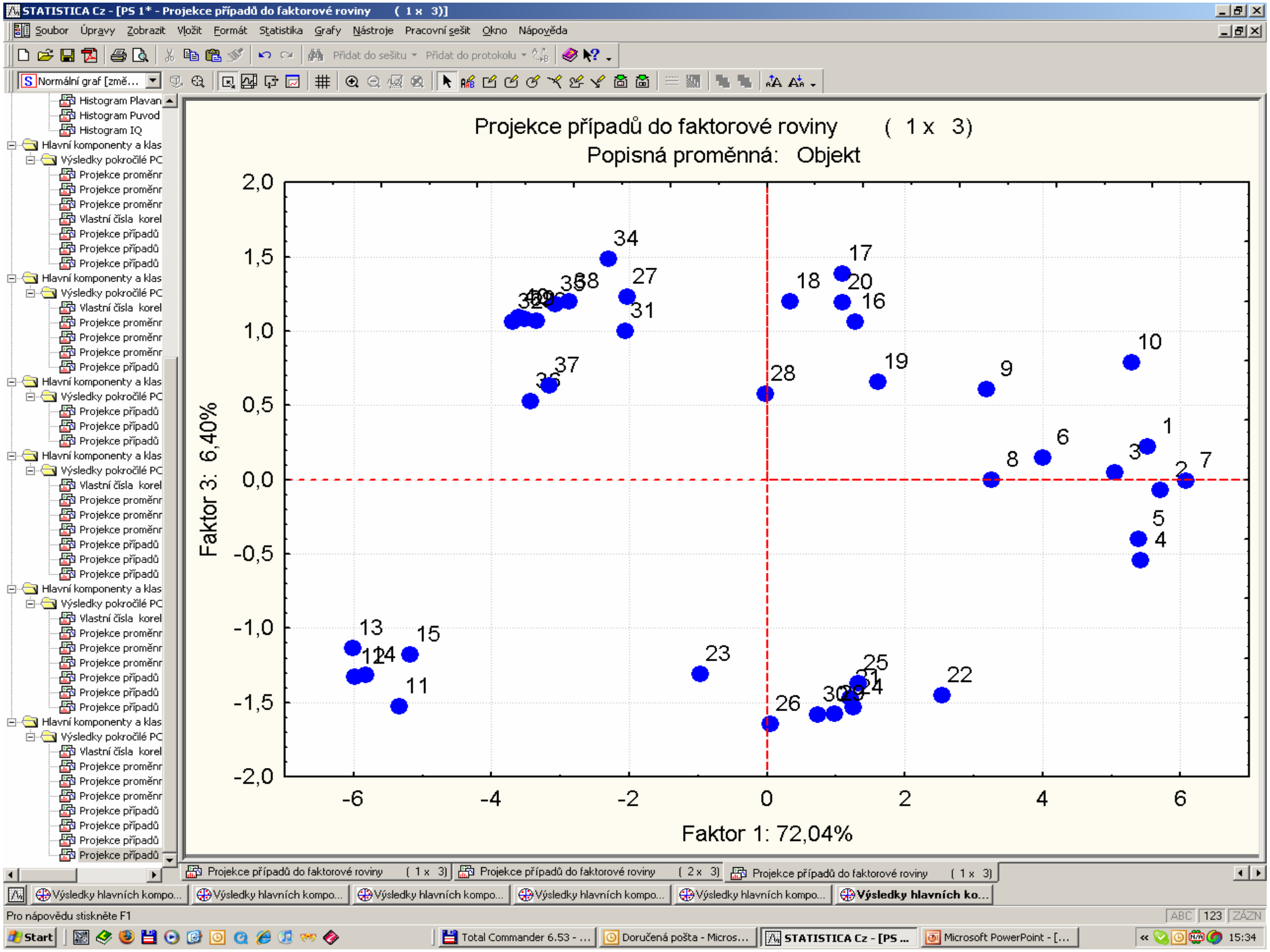

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (2 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Ptrnu
Pruduchy
Kladelko
Phacku
Ritotvor
Ptykostr
Dtyk3
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 6,40%
Faktor 2 : 9,18%
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Pro nápovědu stiskněte F1

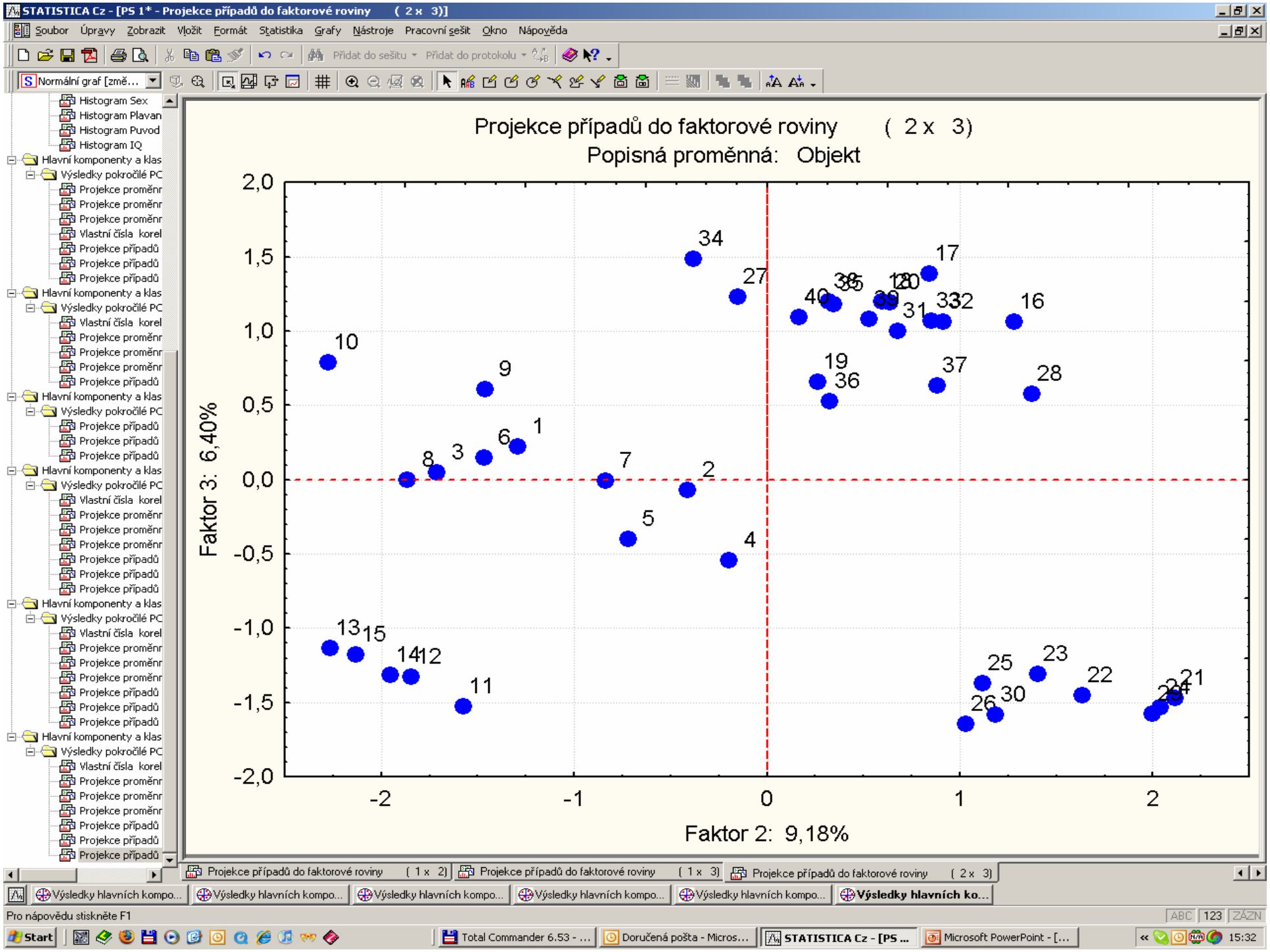
3. Rozptylový diagram komponentního skóre: mšice jsou roztříděny do 4 shluků. Závěr je v souhlase se taxonomických tříděním z biologie.



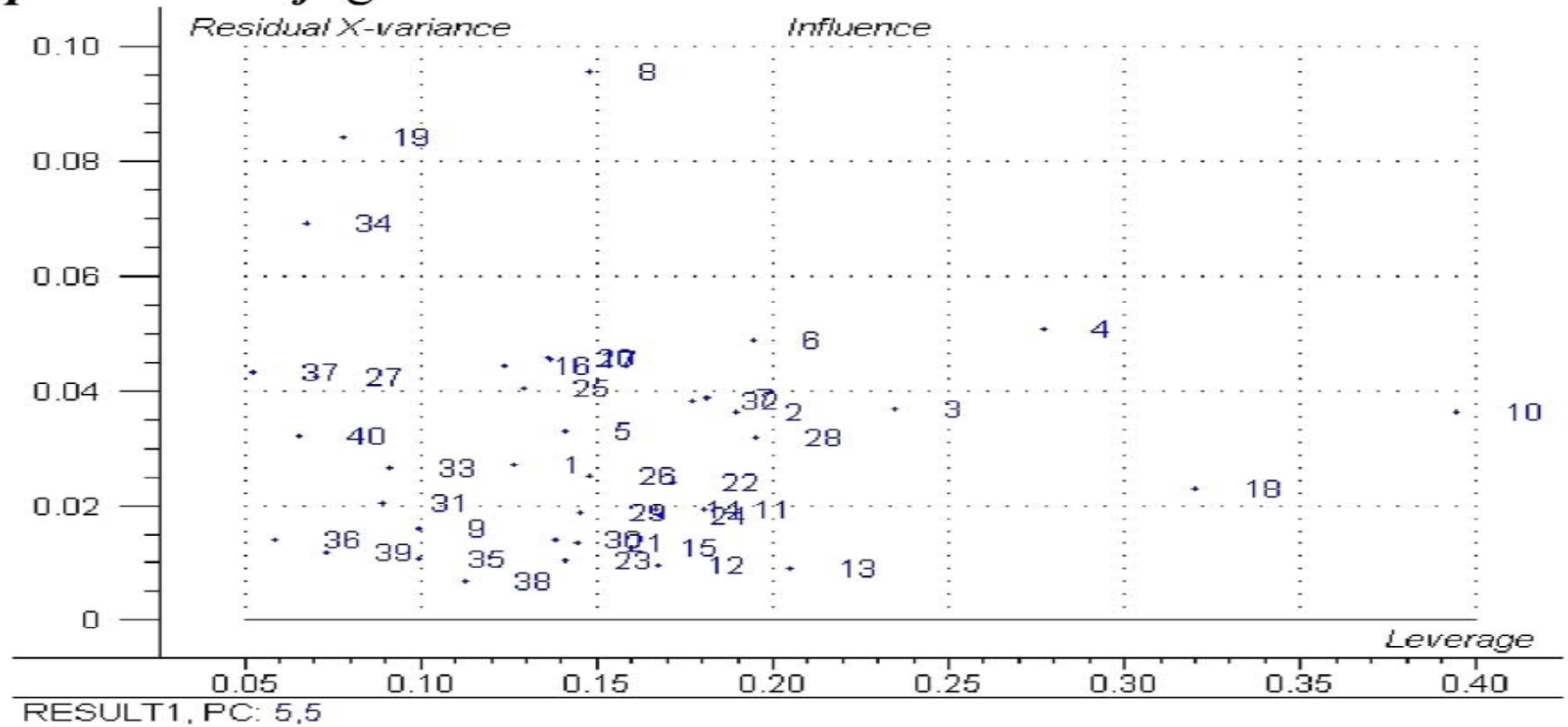
Obr. 4.25 Rozptylový diagram komponentního skóre dat *Mšice* (UNSCRAMBLER).


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 2)]
Projekce případů do faktorové roviny (1 x 2)
Popisná proměnná: Objekt
Faktor 2: 9,18%
Faktor 1: 72,04%
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Projekce případů do faktorové roviny (1 x 2)


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 3)]
Projekce případů do faktorové roviny (1 x 3)
Popisná proměnná: Objekt
Faktor 3: 6,40%
Faktor 1: 72,04%
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-6
-4
-2
0
2
4
6


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (2 x 3)]
Projekce případů do faktorové roviny (2 x 3)
Popisná proměnná: Objekt
Faktor 3: 6,40%
Faktor 2: 9,18%
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2
-1
0
1
2
Projekce případů do faktorové roviny (1 x 2)
Projekce případů do faktorové roviny (1 x 3)
Projekce případů do faktorové roviny (2 x 3)
Pro nápovědu stiskněte F1

4. Analýza vlivných bodů: analýzou reziduí indikovány vlivné body, tj. ***odlehlé objekty*** nesouhlasící s navrženým modelem PCA při ***horním okraji*** grafu, a ***extrémní objekty,*** které souhlasí s navrženým modelem PCA a jsou při ***pravém okraji*** grafu.



Obr. 4.26 Graf vlivných bodů statistické analýzy reziduí dat *Mšice* (UNSCRAMBLER).

○ *Závěr:* PCA je užitečná při taxonomickém třídění mšic: nalezeny 4 shluky mšic.

PŘÍKLAD 9.4 *Vytvoření dendrogramu neuroleptik*

Neuroleptika redukují nežádoucí účinky přebytečného dopaminu a liší se ve svých účincích: potlačují nervozitu, záchvaty, třes, ospalost, parkinsonismus, vynechávání menstruace, vyrážky, zvýšené slinění atd. Cílem je provést klasifikaci neuroleptik do shluků podobných účinků.

○ ***Data*:** Data ***Neuroleptika*** (převrácená hodnota mediánové účinné dávky *1/ED50* [kg/mg]):
Lek název neuroleptika,
Nervoz potlačení nervozity,
Stereo potlačení stereotypního chování,
Tres potlačení záchvatu a třesu a
Usmr dávka smrtícího účinku.

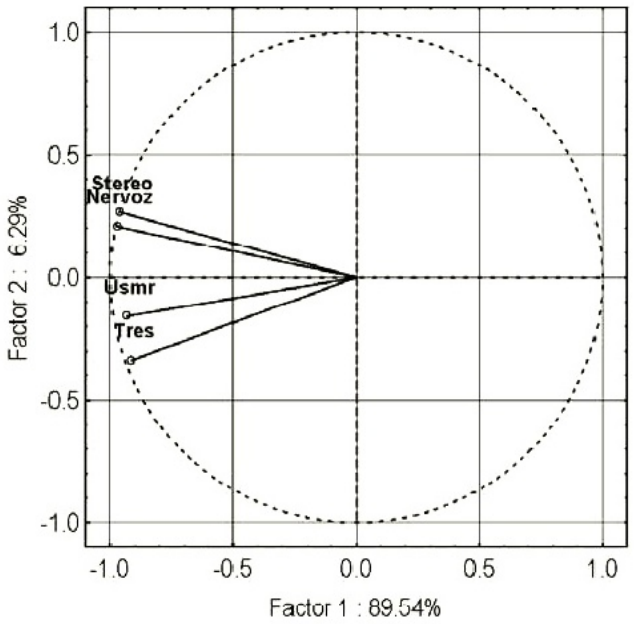
Lek	*Nervoz*	*Stereo*	*Tres*	*Usmr*
1 Chlorphromazine	3.846	3.333	1.111	1.923
2 Promazine	0.323	0.213	0.108	1.429
3 Trifluperazine	27.027	17.857	0.562	0.14
4 Fluphenazine	17.857	15.385	1.695	1.075
5 Perphenazine	27.027	27.027	1.961	2.083
6 Thioridazine	0.244	0.185	0.093	1.333
7 Pifluthixol	142.857	142.857	20.408	163.934
8 Thiothixene	4.348	4.348	0.047	0.345
9 Chorprothixene	5.882	2.941	4.545	4.167
10 Spiperone	62.5	47.619	11.765	0.847
11 Haloperidol	52.632	62.5	1.282	0.568
12 Azaperone	2.941	1.282	2.222	3.03
13 Pipamperone	0.327	0.187	1.724	0.397
14 Pimozide	20.408	20.408	0.107	0.025
15 Metitepine	15.385	10.204	10.204	27.027
16 Clozapine	0.161	0.093	0.327	0.323
17 Perlapine	0.323	0.323	0.37	0.067
18 Sulpiride	0.047	0.047	0.003	0.001
19 Butaclamol	10.204	9.091	1.471	0.025
20 Molindone	7.692	7.692	0.14	38138

○ ***Řešení:*** Po vyhledání optimální tvorby dendrogramu sestrojíme dendrogram podobnosti znaků a dendrogram podobnosti objektů.

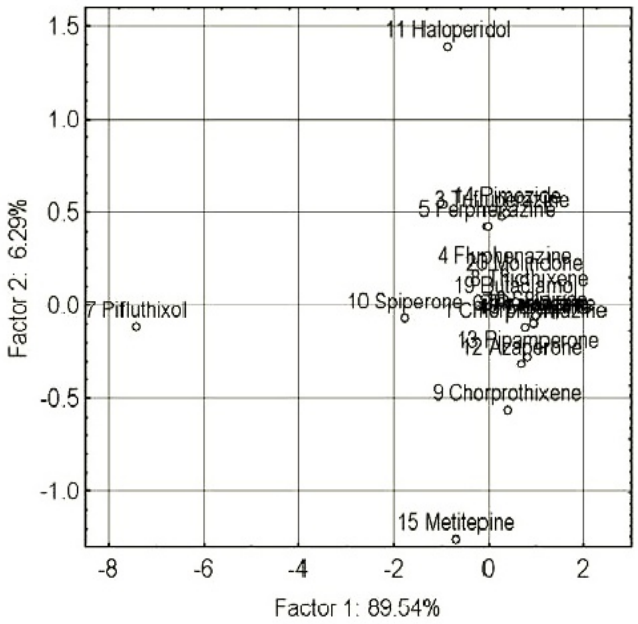
Nejvyšší hodnota kofenetického korelačního koeficientu ***CC*** a nejnižší hodnota obou kritérií delta, ***Delta(0.5)*** **a** ***Delta(1.0)***, vybrala **metodu skupinového průměru** (software NCSS2004).

1. Nejbližšího souseda, *Kofenetická korelace CC:* 0.988598, *Delta(0.5):* 0.474238, *Delta(1.0):* 0.391993.
2. Nejvzdálenějšího souseda: *Kofenetická korelace CC:* 0.982795, *Delta(0.5):* 0.178589, *Delta(1.0):* 0.183477;
3. Párový průměr, *Kofenetická korelace CC:* 0.988876, *Delta(0.5):* 0.177810, *Delta(1.0):* 0.188781;
4. **Skupinový průměr**, *Kofenetická korelace CC:* 0.987356, *Delta(0.5):* 0.137455, *Delta(1.0):* 0.125290;
5. Těžiště, *Kofenetická korelace CC:* 0.984750, *Delta(0.5):* 0.175238, *Delta(1.0):* 0.166599;
6. Median, *Kofenetická korelace CC:* 0.984215, *Delta(0.5):* 0.452308, *Delta(1.0):* 0.428346;
7. Wardova metoda, *Kofenetická korelace CC:* 0.979285, *Delta(0.5):* 0.549394, *Delta(1.0):* 0.492716.

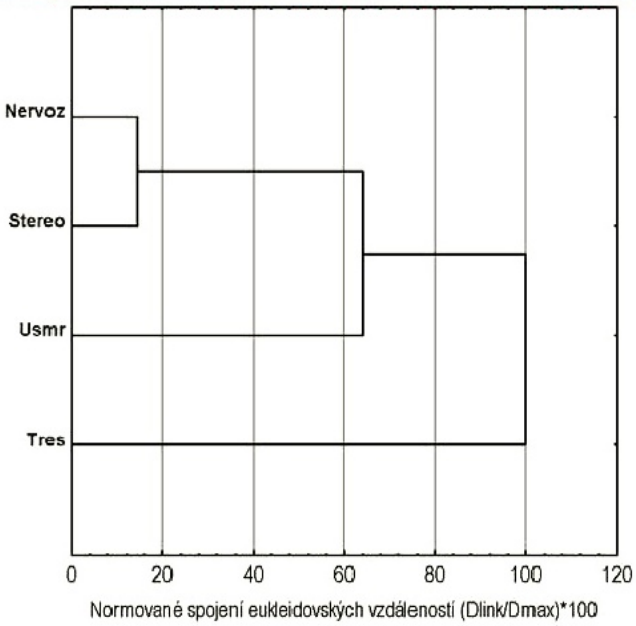
Metoda skupinového průměru v dendrogramu podobnosti objektů:
první shluk obsahuje 12 objektů 1, 8, 12, 9, 2, 6, 16, 17, 18, 13, 19, 20,
druhý shluk 5 objektů 3, 4, 14, 5, 15,
třetí shluk 2 objekty 10 a 11,
čtvrtý shluk obsahuje jeden objekt, a to 7.



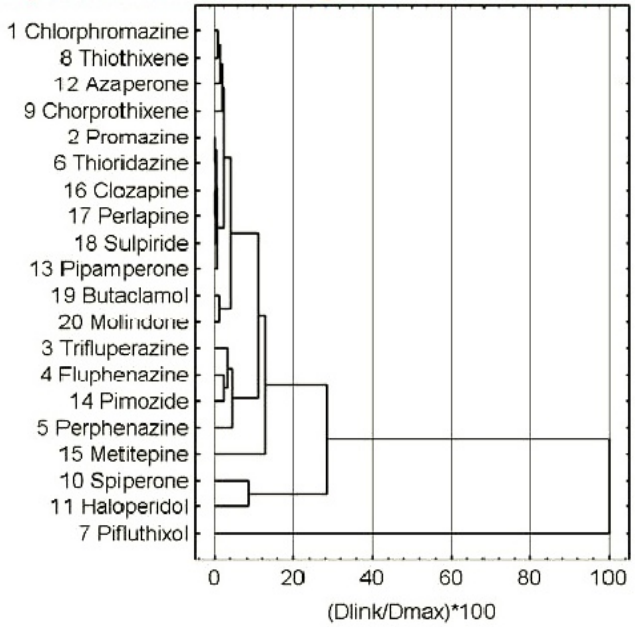
Graf komponentních vah znaků matice dat *Neuroleptika*.



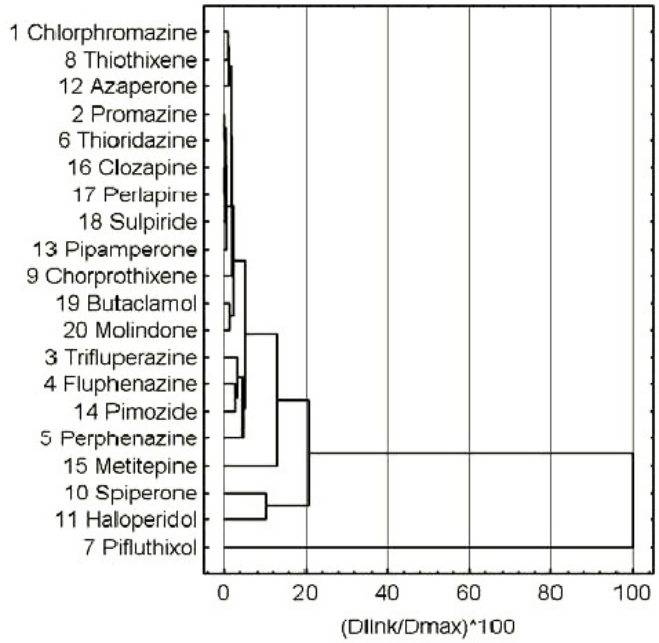
Graf komponentního skóre objektů matice dat *Neuroleptika*



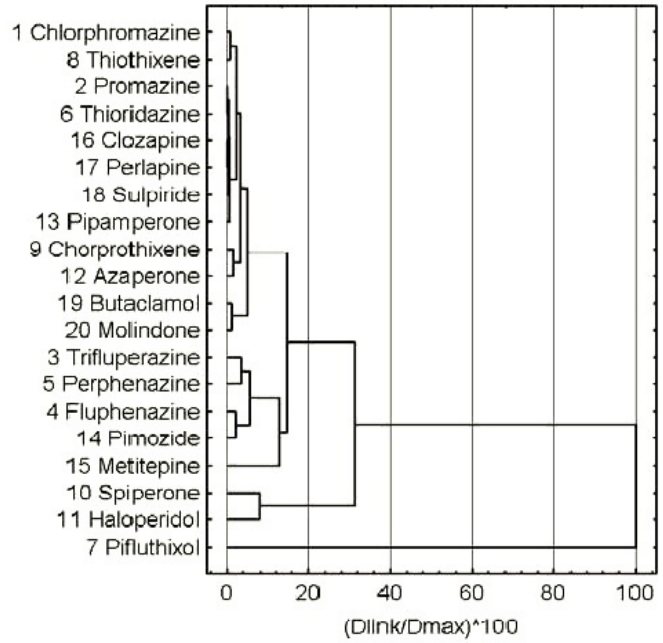
Dendrogram znaků metodou skupinového průměru



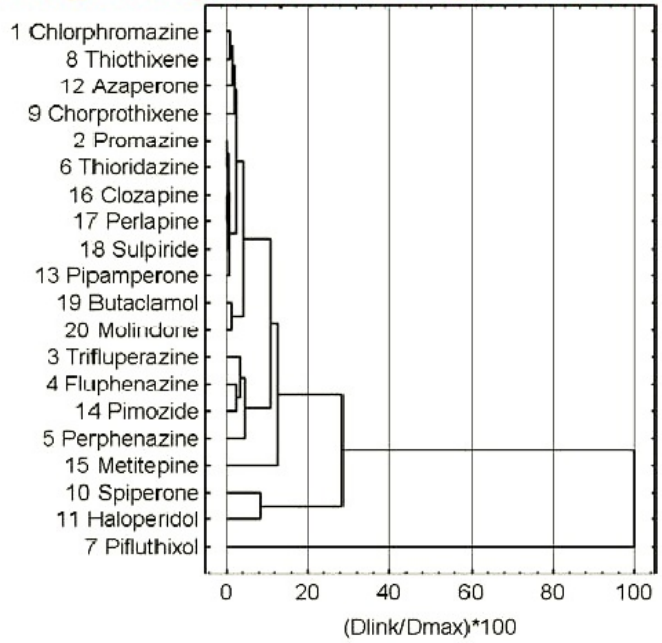
Dendrogram objektů metodou skupinového průměru



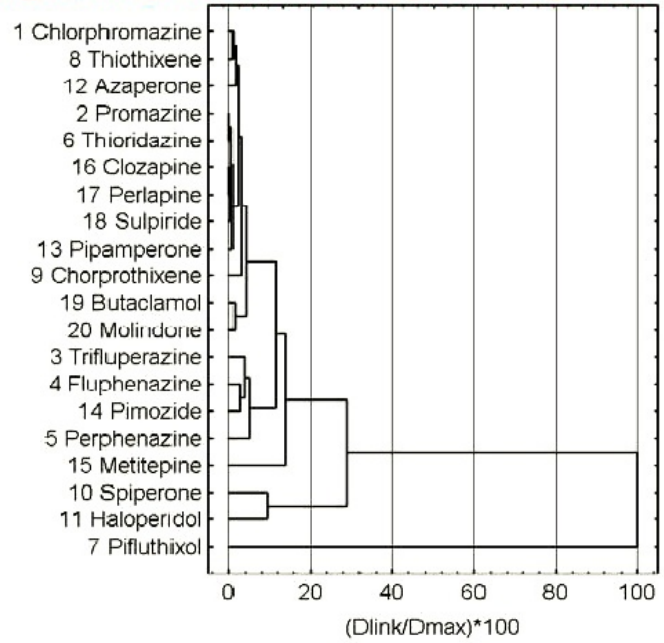
Dendrogram objektů metodou nejbližšího souseda



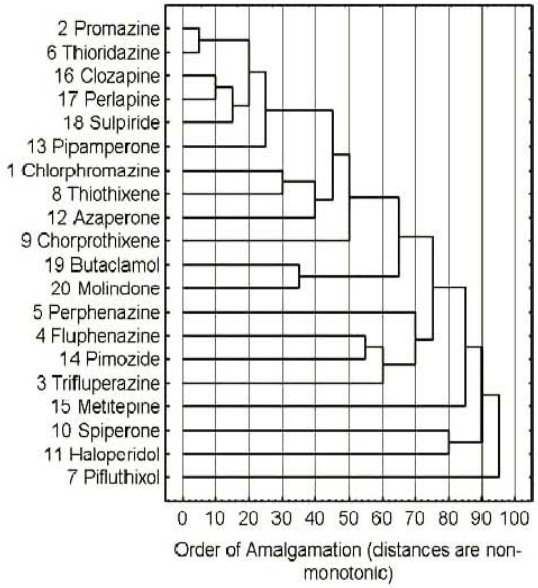
Dendrogram objektů metodou nejvzdálenějšího souseda



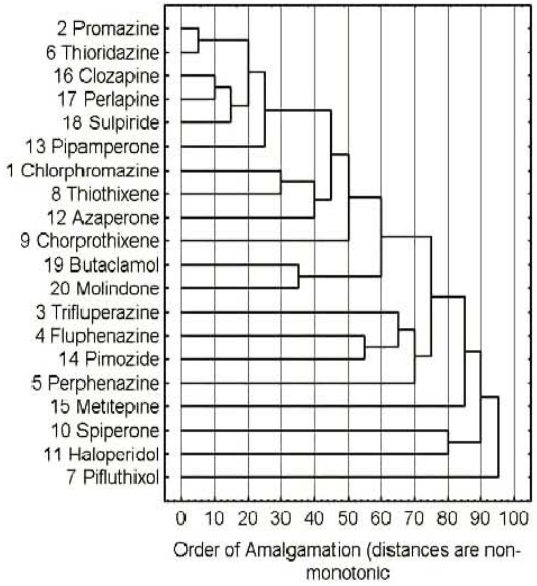
Dendrogram objektů metodou párového průměru



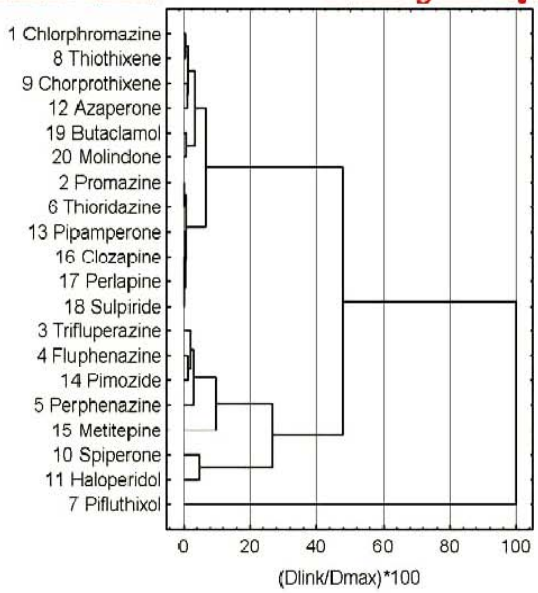
Dendrogram objektů metodou skupinového průměru



Dendrogram objektů metodou neváženého těžiště

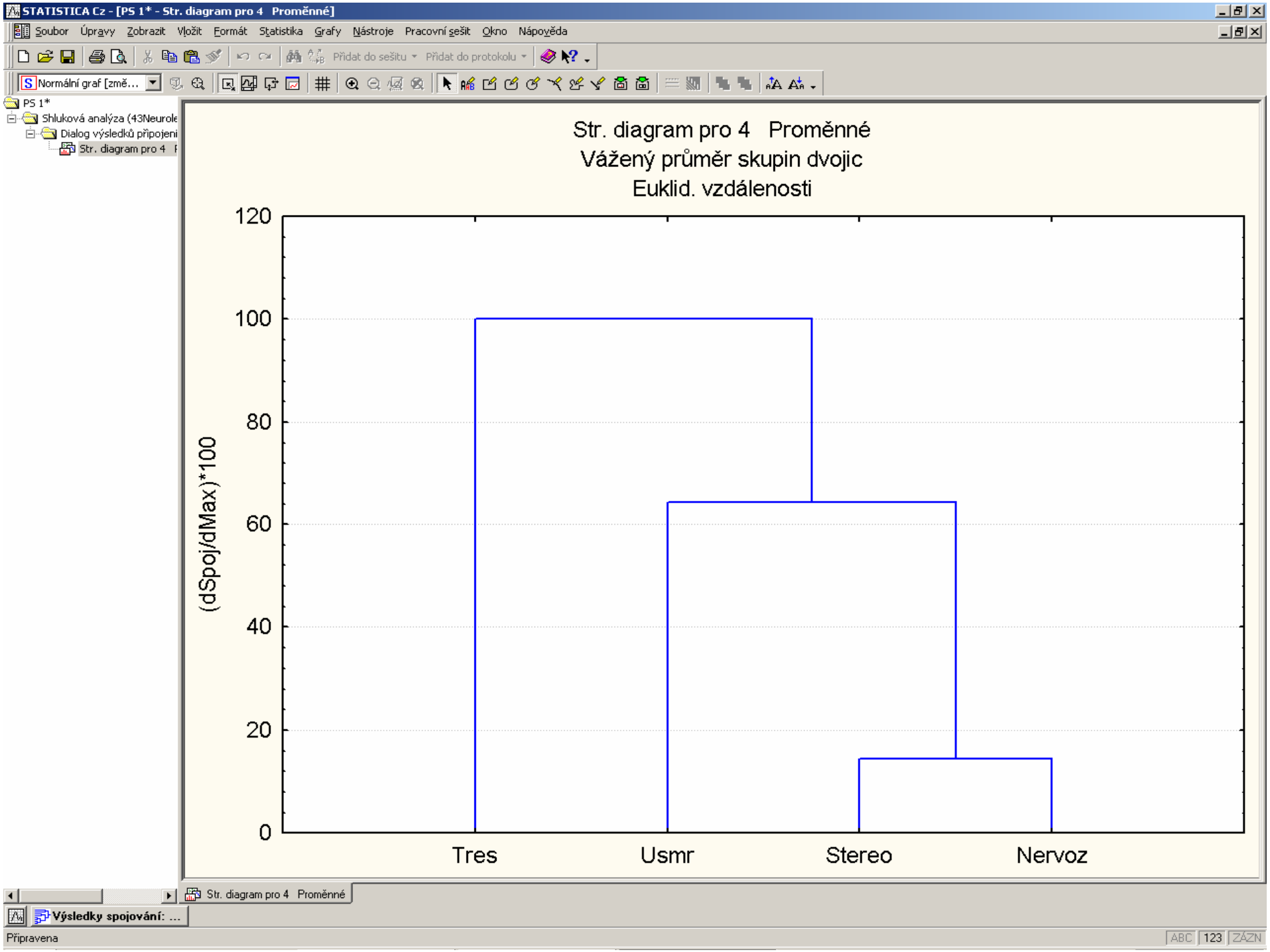


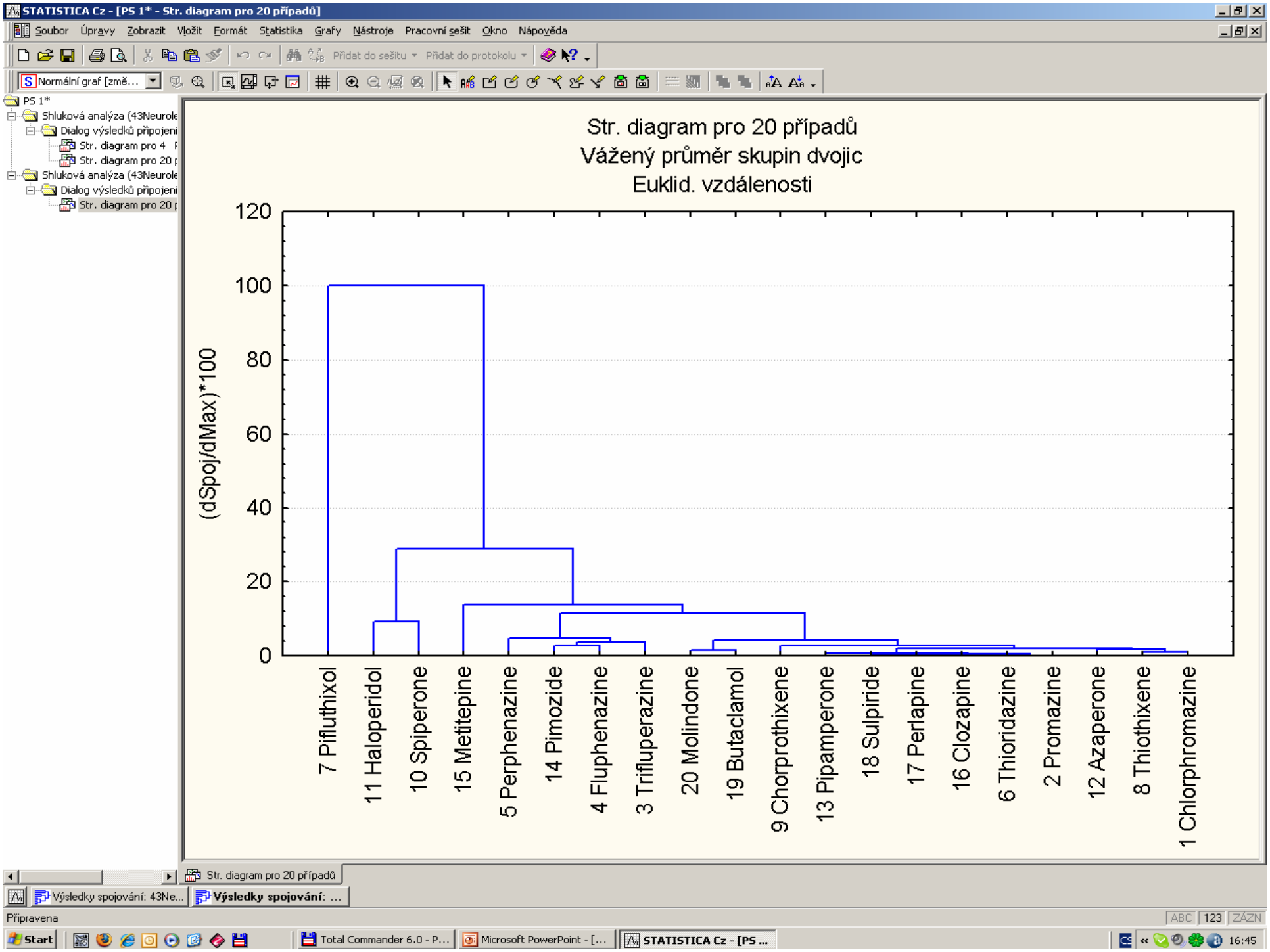
Dendrogram objektů metodou váženého těžiště (mediánu).

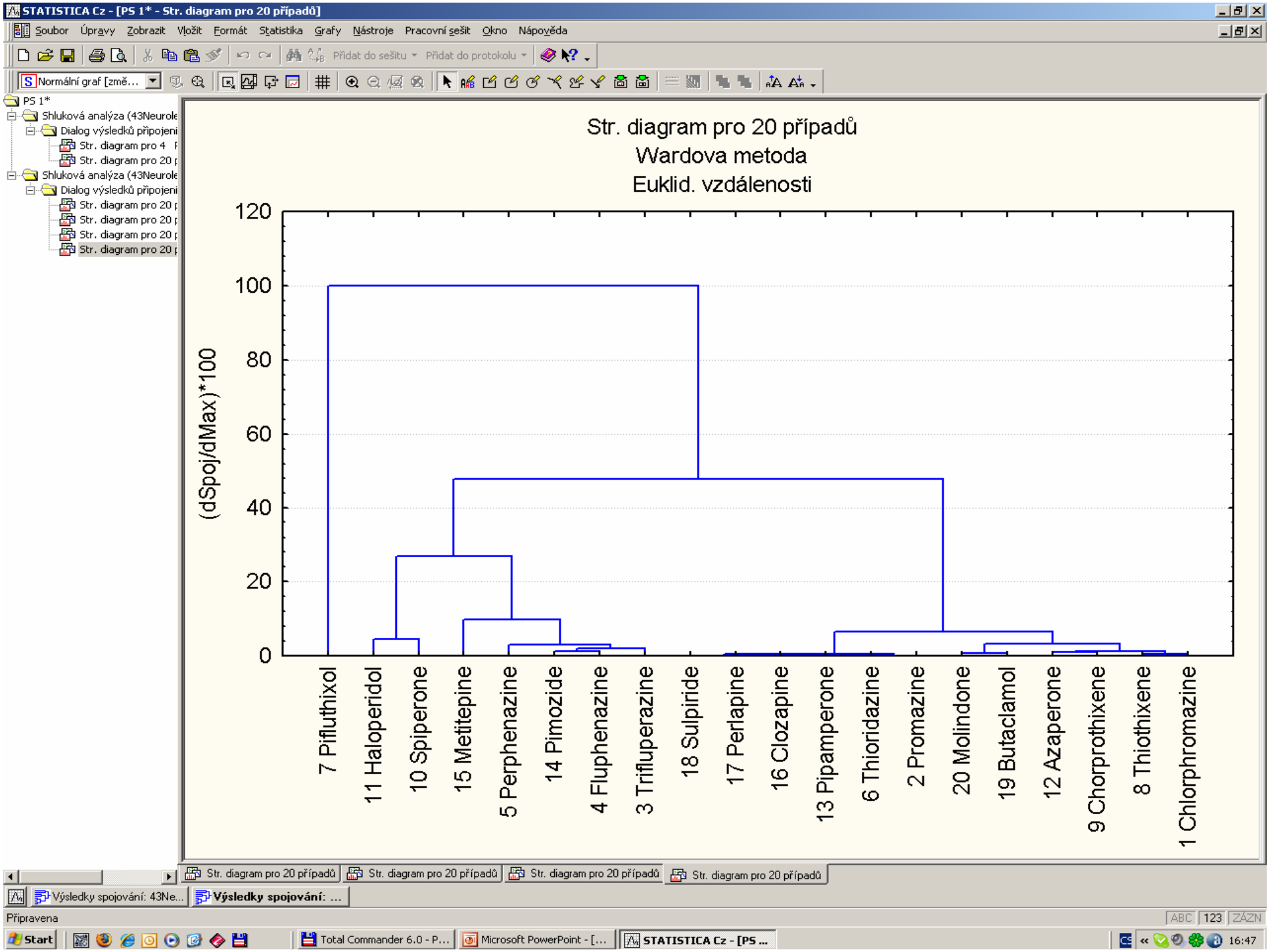


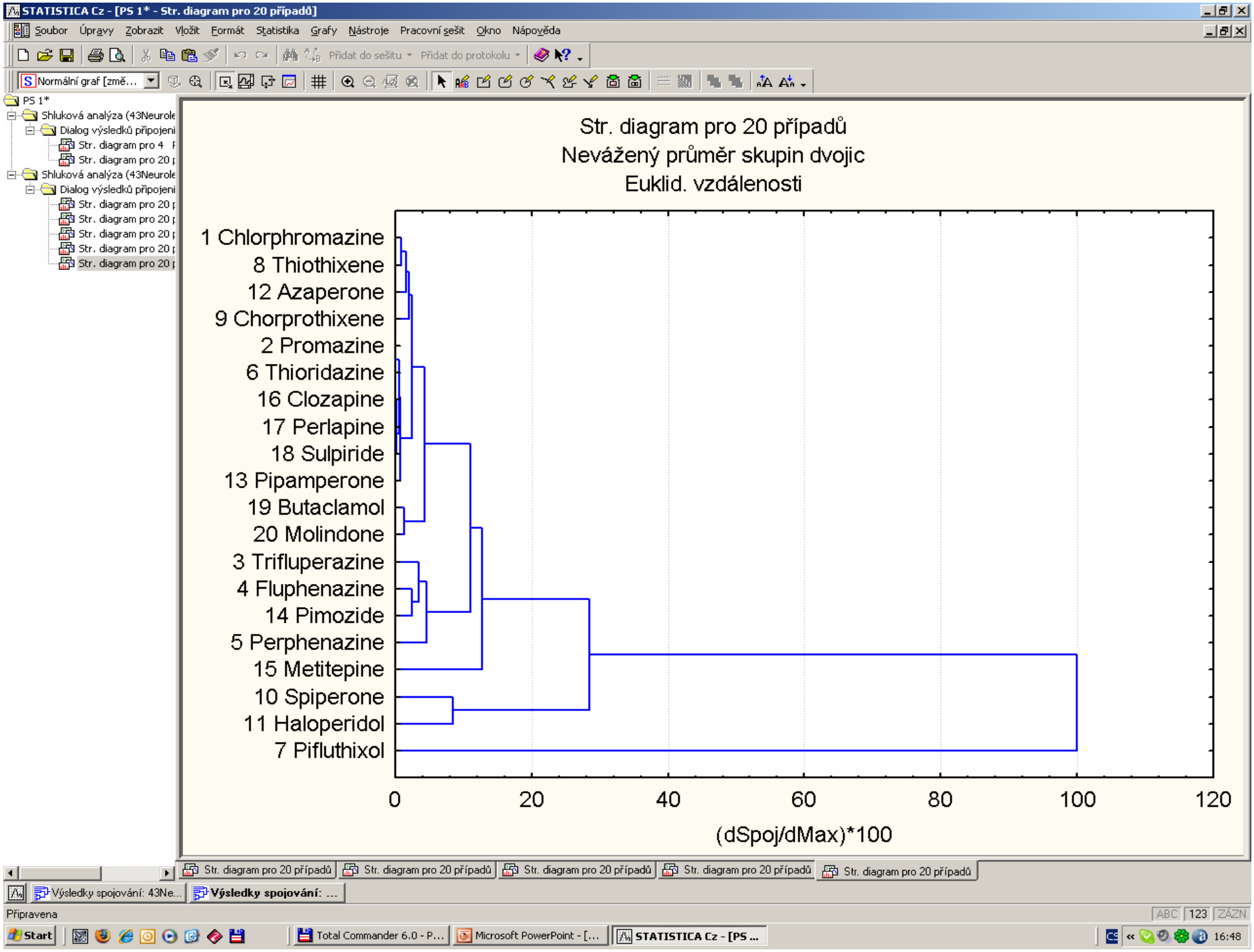
Dendrogram objektů metodou Wardovou

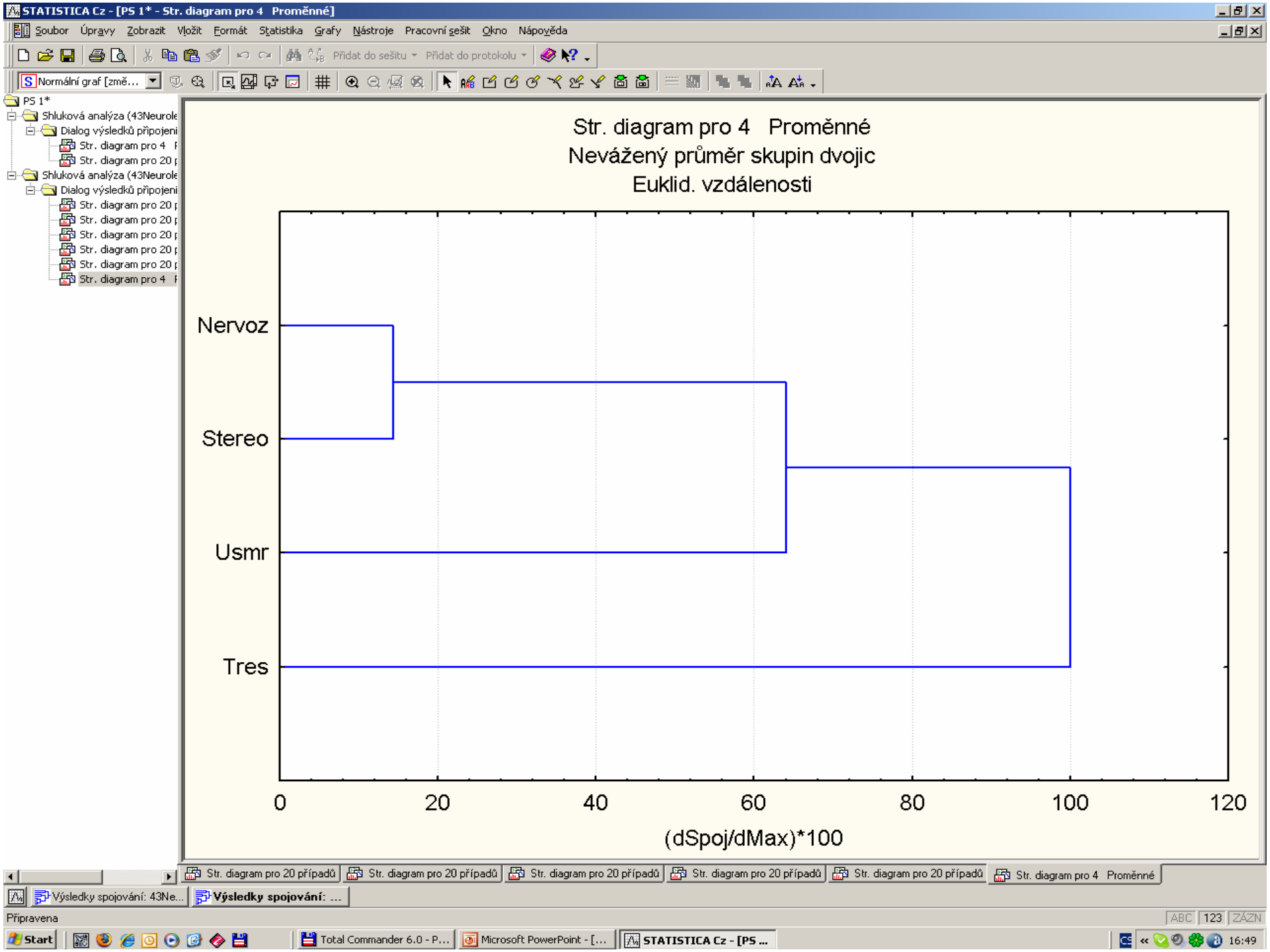
Závěr: Nejvhodnější tvorba dendrogramu je metodami párového průměru a skupinového průměru.

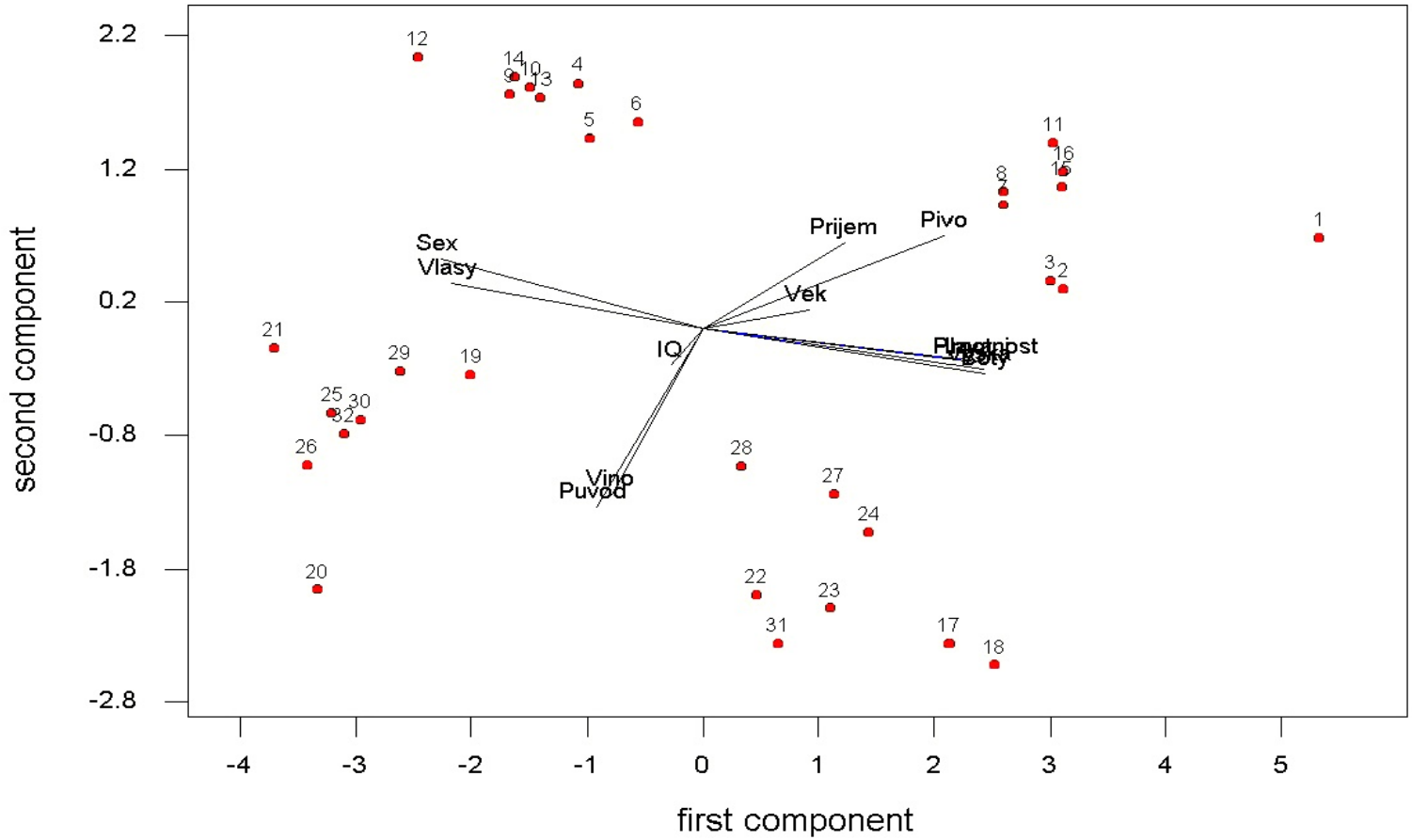

STATISTICA Cz - [PS 1* - Str. diagram pro 4 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 4 P
Str. diagram pro 4 Proměnné
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
Tres
Usmr
Stereo
Nervoz
Str. diagram pro 4 Proměnné
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN


STATISTICA Cz - [PS 1* - Str. diagram pro 20 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 4
Str. diagram pro 20
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 20
Str. diagram pro 20 případů
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
7 Pifluthixol
11 Haloperidol
10 Spiperone
15 Metitepine
5 Perphenazine
14 Pimozide
4 Fluphenazine
3 Trifluperazine
20 Molindone
19 Butaclamol
9 Chorprothixene
13 Pipamperone
18 Sulpiride
17 Perlapine
16 Clozapine
6 Thioridazine
2 Promazine
12 Azaperone
8 Thiothixene
1 Chlorphromazine
Str. diagram pro 20 případů
Výsledky spojování: 43Ne...
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
16:45


STATISTICA Cz - [PS 1* - Str. diagram pro 20 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 4
Str. diagram pro 20
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20 případů
Wardova metoda
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
7 Pifluthixol
11 Haloperidol
10 Spiperone
15 Metitepine
5 Perphenazine
14 Pimozide
4 Fluphenazine
3 Trifluperazine
18 Sulpiride
17 Perlapine
16 Clozapine
13 Pipamperone
6 Thioridazine
2 Promazine
20 Molindone
19 Butaclamol
12 Azaperone
9 Chorprothixene
8 Thiothixene
1 Chlorphromazine
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Výsledky spojování: 43Ne...
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
16:47


STATISTICA Cz - [PS 1* - Str. diagram pro 20 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 4
Str. diagram pro 20
Shluková analýza (43Neurole
Dialog výsledků připojeni
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20
Str. diagram pro 20 případů
Nevážený průměr skupin dvojic
Euklid. vzdálenosti
1 Chlorphromazine
8 Thiothixene
12 Azaperone
9 Chorprothixene
2 Promazine
6 Thioridazine
16 Clozapine
17 Perlapine
18 Sulpiride
13 Pipamperone
19 Butaclamol
20 Molindone
3 Trifluperazine
4 Fluphenazine
14 Pimozide
5 Perphenazine
15 Metitepine
10 Spiperone
11 Haloperidol
7 Pifluthixol
0
20
40
60
80
100
120
(dSpoj/dMax)*100
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Str. diagram pro 20 případů
Výsledky spojování: 43Ne...
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
16:48


STATISTICA Cz - [PS 1* - Str. diagram pro 4 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Str. diagram pro 4 Proměnné
Nevážený průměr skupin dvojic
Euklid. vzdálenosti
Nervoz
Stereo
Usmr
Tres
0
20
40
60
80
100
120
(dSpoj/dMax)*100
Připravena



Obr. 4.6 Dvojný graf Biplot zdrojové matice dat *Lidé* (SCAN).

● *Závěr:* Analytická korelace neznamená nutně fyzikální korelaci v příčinném slova smyslu. Hmotnost, velikost bot a výška *musí* obecně korelovat u mladistvých ale věk nebude korelovat s výškou u osob zralého věku. IQ nezávisí na zeměpisném původu a fyzických charakteristikách. Je třeba zdůraznit, že metoda hlavních komponent poskytuje *úplný a nezaujatý pohled* na data, do kterého nevkládáme žádné předpojaté myšlenky.

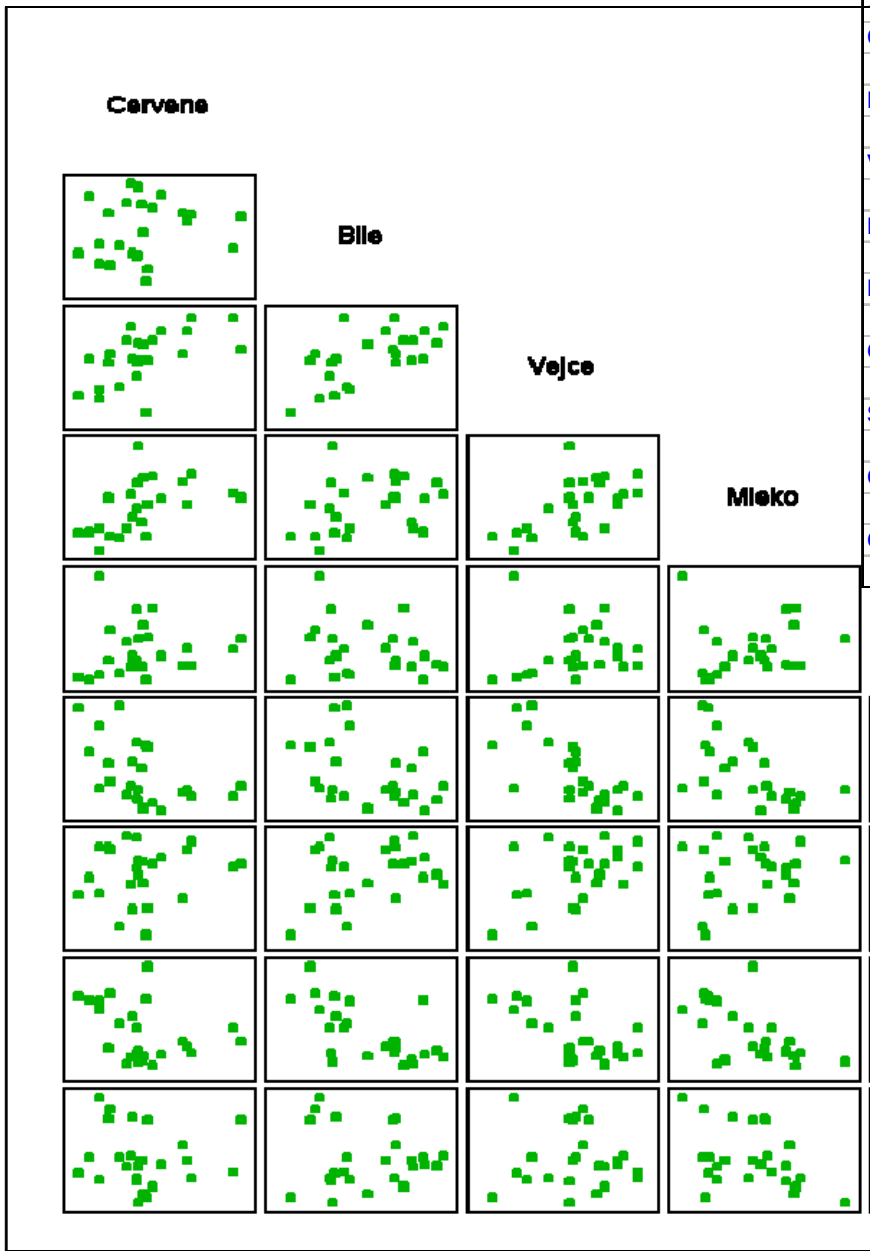
Úloha 3. Sledování spotřeby proteinů v Evropě

Sledovaná spotřeba proteinů v 25 zemích formou spotřeby 9 druhů potravin je předmětem vyšetření.

Data: *i* značí index, ***Cervene*** udává červené maso, **Bile** maso, **Vejce, Mleko**, **Ryby**, **Obilniny, Skrob**, **Orechy**, **Ovoce** a zelenina

	Objekty	Proměnné								
i	**Stát**	**Cervene**	**Bile**	**Vejce**	**Mleko**	**Ryby**	**Obilniny**	**Skrob**	**Orechy**	**Ovoce**
1	Albania	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
2	Austria	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
3	Belgium	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
4	Bulgaria	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
5	Czechoslov.	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
6	Denmark	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
7	E Germany	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
8	Finland	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
9	France	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5
10	Greece	10.2	3	2.8	17.6	5.9	41.7	2.2	7.8	6.5
11	Hungary	5.3	12.4	2.9	9.7	0.3	40.1	4	5.4	4.2
12	Ireland	13.9	10	4.7	25.8	2.2	24	6.2	1.6	2.9
13	Italy	9	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7
14	Netherlands	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7
15	Norway	9.4	4.7	2.7	23.3	9.7	23	4.6	1.6	2.7
16	Poland	6.9	10.2	2.7	19.3	3	36.1	5.9	2	6.6
17	Portugal	6.2	3.7	1.1	4.9	14.2	27	5.9	4.7	7.9
18	Romania	6.2	6.3	1.5	11.1	1	49.6	3.1	5.3	2.8
19	Spain	7.1	3.4	3.1	8.6	7	29.2	5.7	5.9	7.2
20	Sweden	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2
21	Switzerland	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9
22	UK	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3
23	USSR	9.3	4.6	2.1	16.6	3	43.6	6.4	3.4	2.9
24	W Germany	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8
25	Yugoslavia	4.4	5	1.2	9.5	0.6	55.9	3	5.7	3.2

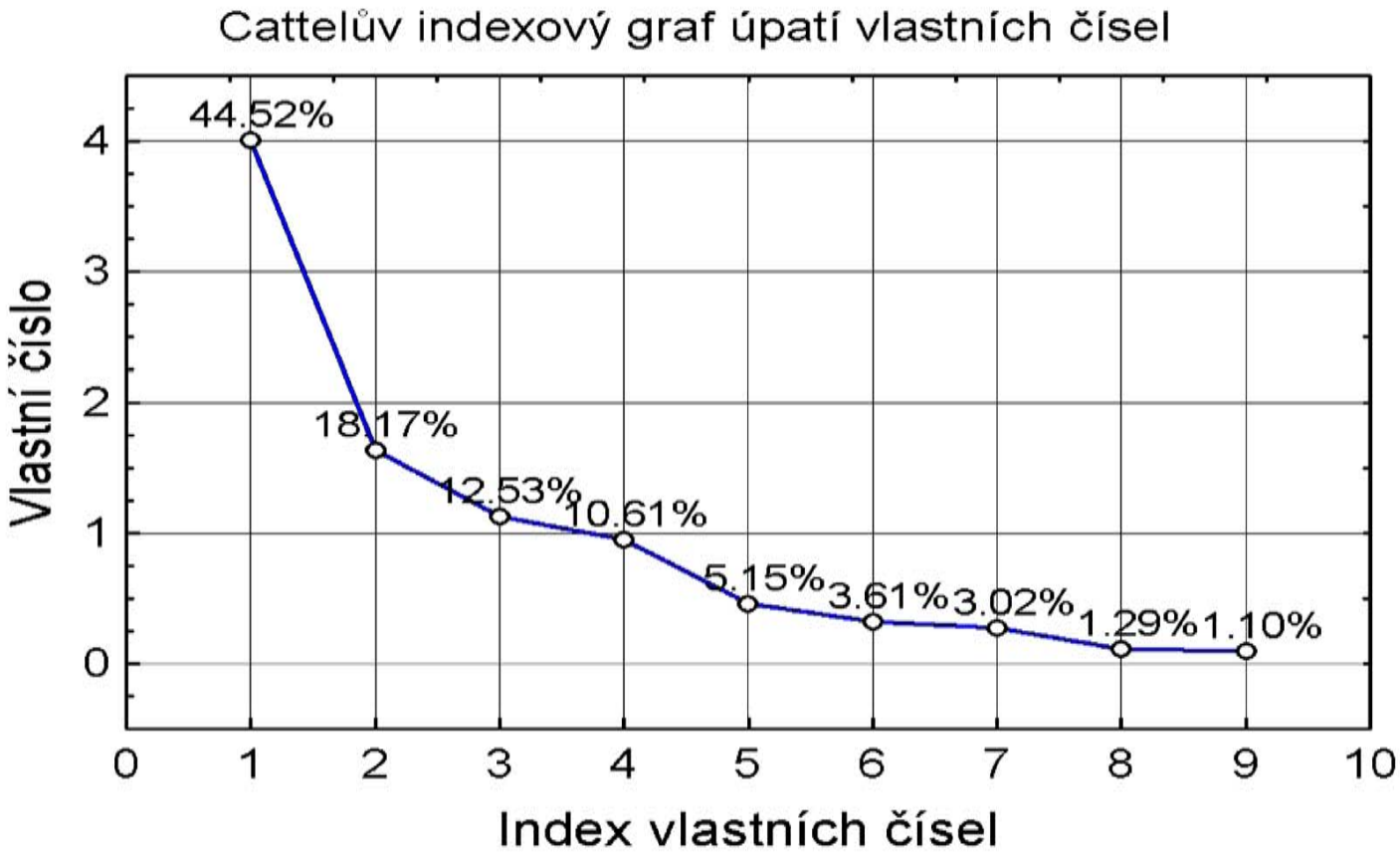
Test významnosti korelace v korelační matici



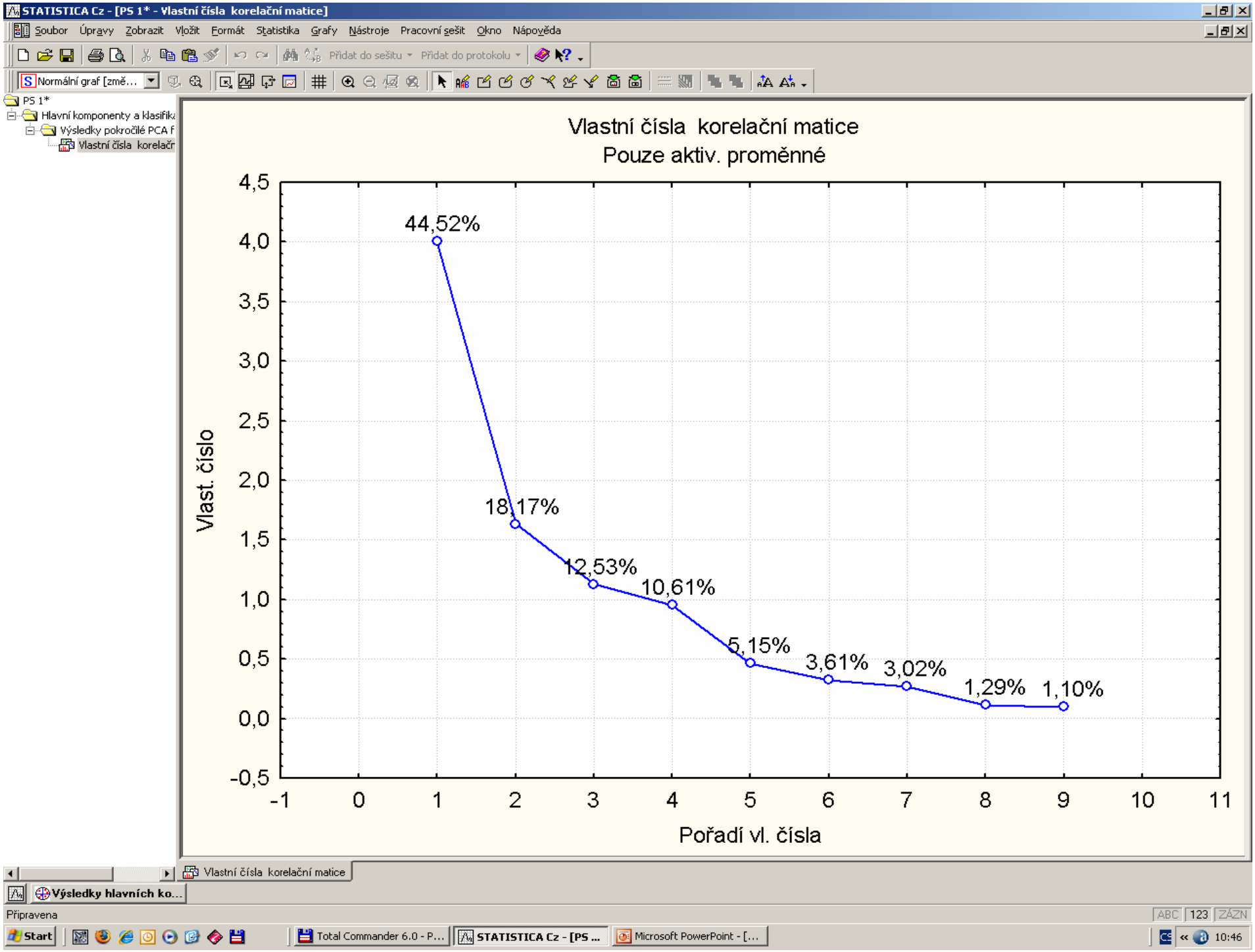
	Cervene	Bile	Vejce	Mleko	Ryby	Obilniny	Skrob	Orechy	Ovoce
Cervene									
Bile	0.153								
	0.4653								
Vejce	0.5856	0.6204							
	0.0021	0.0009							
Mleko	0.5029	0.2815	0.5755						
	0.0104	0.1728	0.0026						
Ryby	0.061	-0.234	0.0656	0.1379					
	0.7722	0.2602	0.7555	0.511					
Obilniny	-0.4999	-0.4138	-0.7124	-0.5927	-0.5242				
	0.0109	0.0398	0.0001	0.0018	0.0071				
Skrob	0.1354	0.3138	0.4522	0.2224	0.4039	-0.5333			
	0.5186	0.1267	0.0232	0.2853	0.0453	0.0061			
Orechy	-0.3494	-0.635	-0.5598	-0.6211	-0.1472	0.651	-0.4743		
	0.0869	0.0007	0.0036	0.0009	0.4827	0.0004	0.0166		
Ovoce	-0.0742	-0.0613	-0.0455	-0.4084	0.2661	0.0465	0.0844	0.375	
	0.7244	0.7709	0.829	0.0427	0.1985	0.8251	0.6883	0.0648	

Pod hodnotami korelačních koeficientů jsou vypočtené hladiny významnosti P. Je-li $P < 0.05$, je korelace významná.

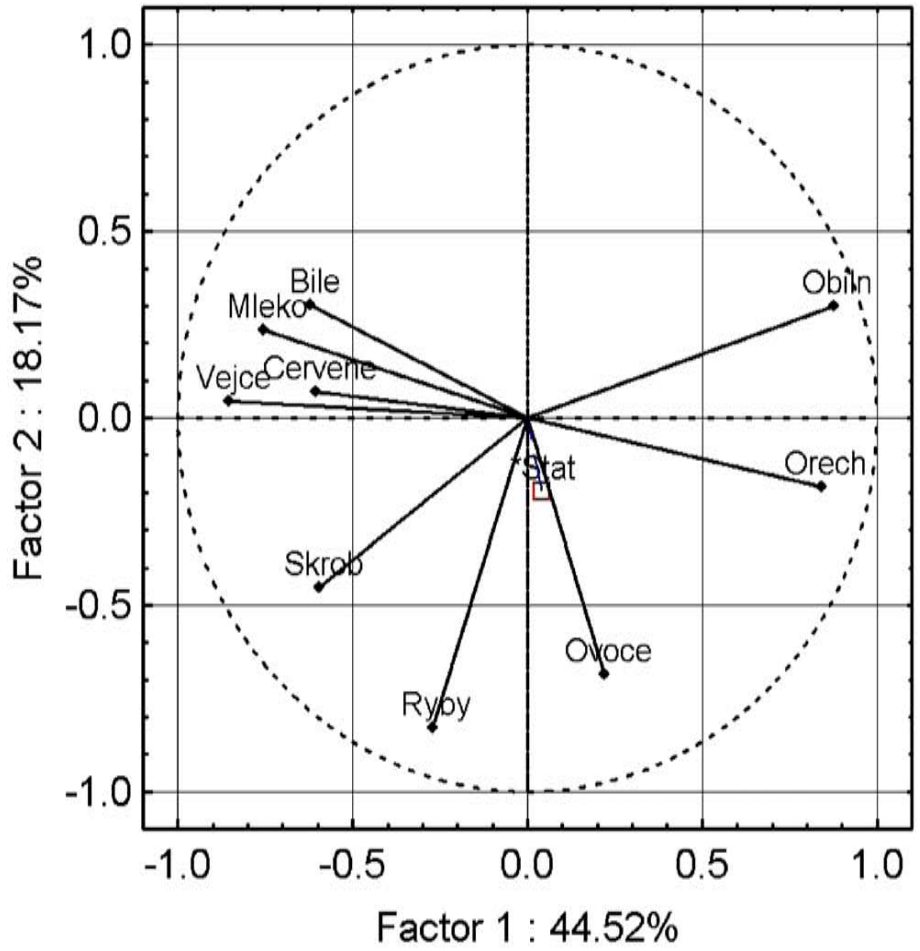
1. Cattelův indexový graf úpatí vlastních čísel: první hlavní komponenta (44.52% celkové proměnlivosti) a druhá hlavní komponenta (18.17% celkové proměnlivosti) dohromady dostatečně popíší proměnlivost v datech.



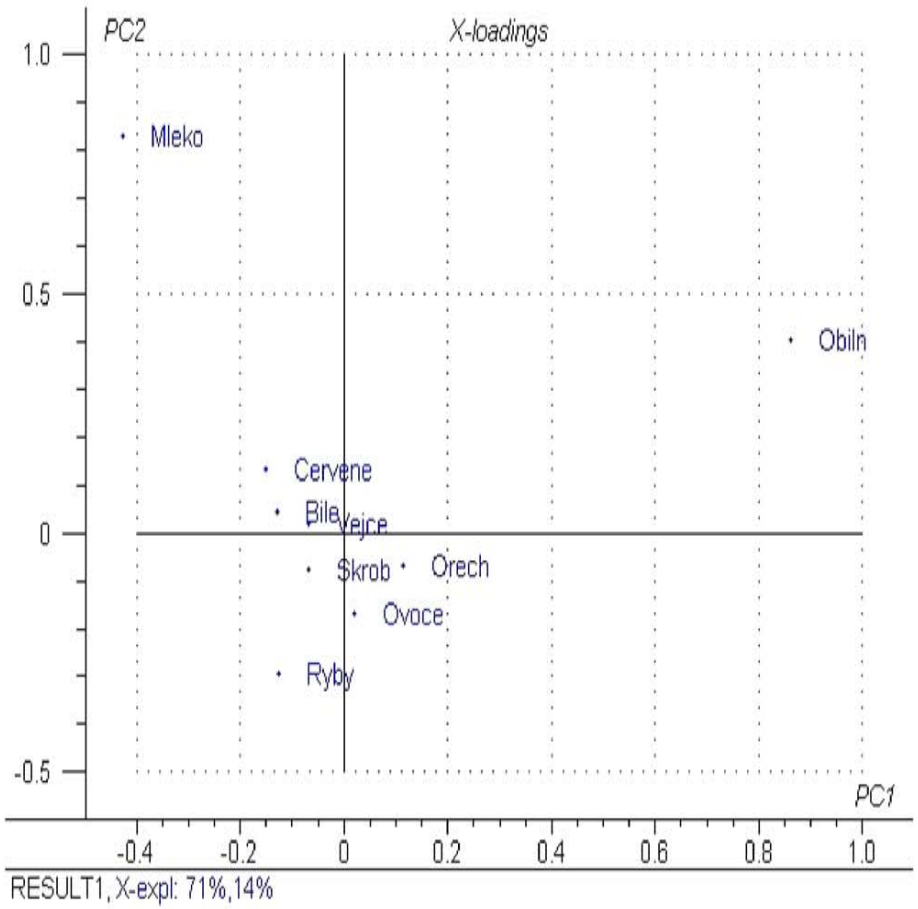
Obr. 4.15 Cattelův indexový graf úpatí celkového reziduálového rozptylu zdrojové matice dat *Proteiny* (STATISTICA).


STATISTICA Cz - [PS 1* - Vlastní čísla korelační matice]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačr
Vlastní čísla korelační matice
Pouze aktiv. proměnné
Vlast. číslo
Pořadí vl. čísla
44,52%
18,17%
12,53%
10,61%
5,15%
3,61%
3,02%
1,29%
1,10%
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1
0
1
2
3
4
5
6
7
8
9
10
11
Vlastní čísla korelační matice
Výsledky hlavních ko...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
10:46

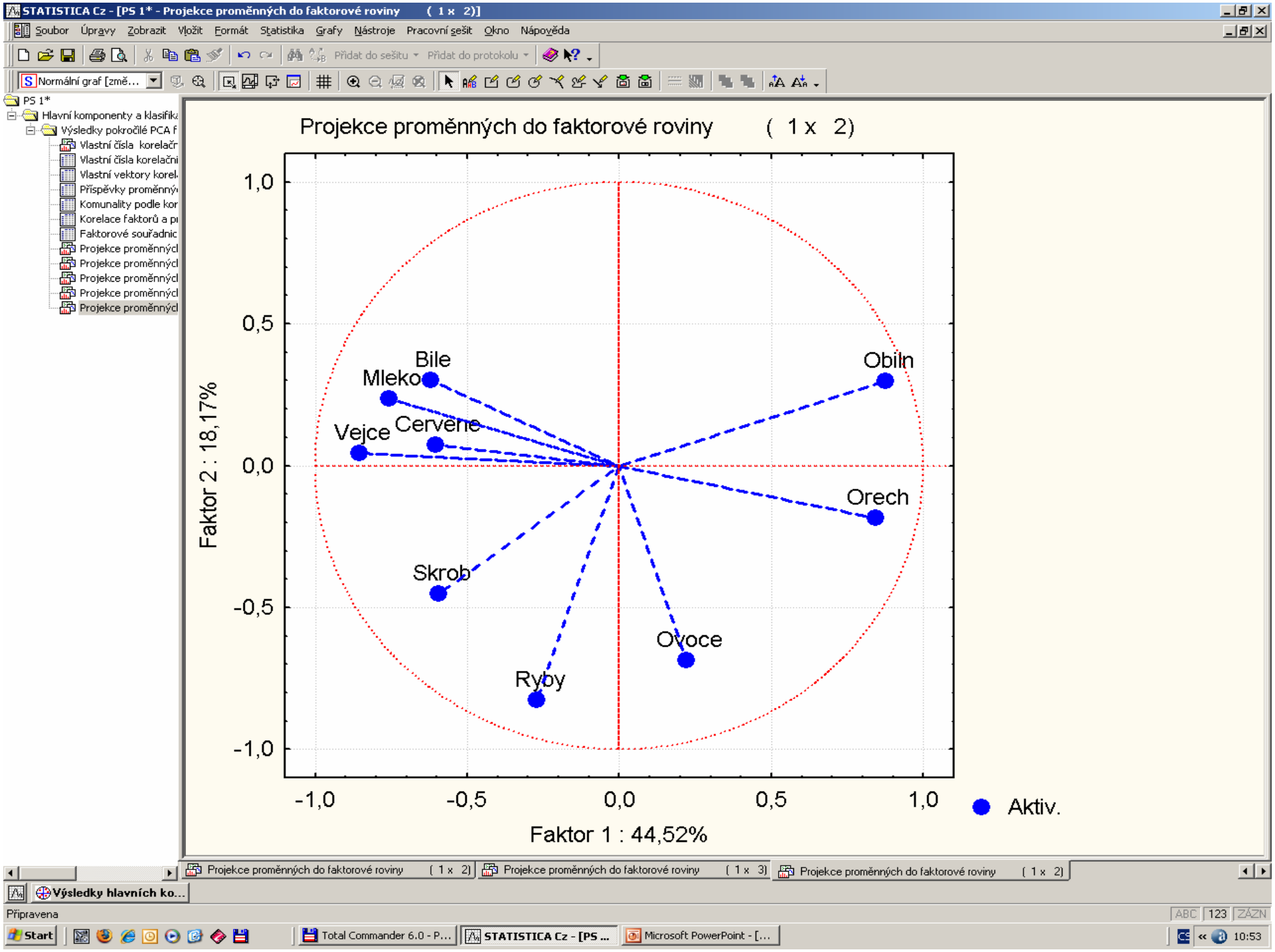
2. Graf komponentních vah: *Mléko* a *Obilniny* spolu vzhledem obsahu proteinů nekorelují. Vyjímečně si stojí i znak *Ryby*. Okolo počátku je shluk znaků, které jsou spolu v silné korelaci, jsou to *Červené* maso, *Bílé* maso, *Vejce*, *Škrob*, *Ořechy* a *Ovoce* a zelenina.

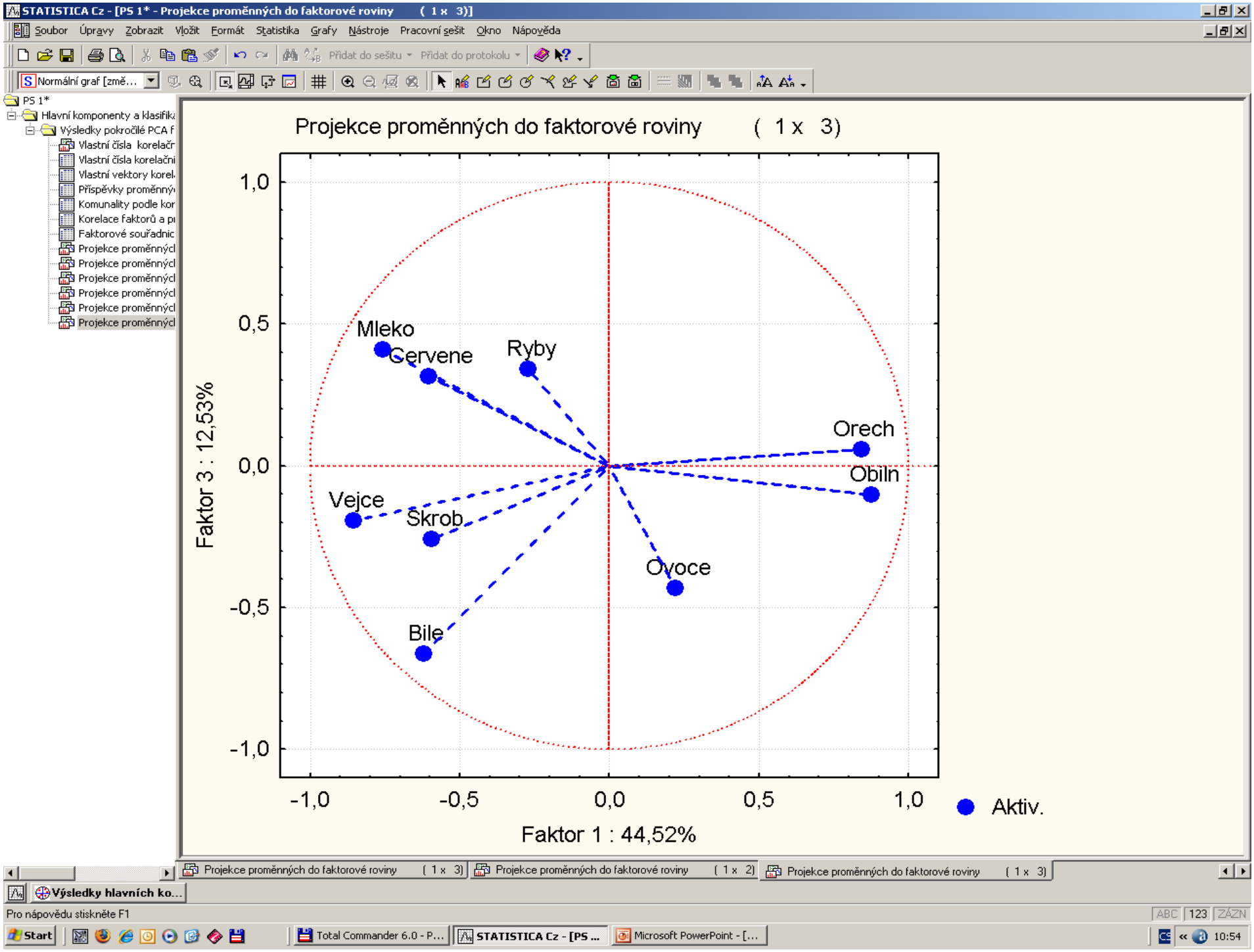


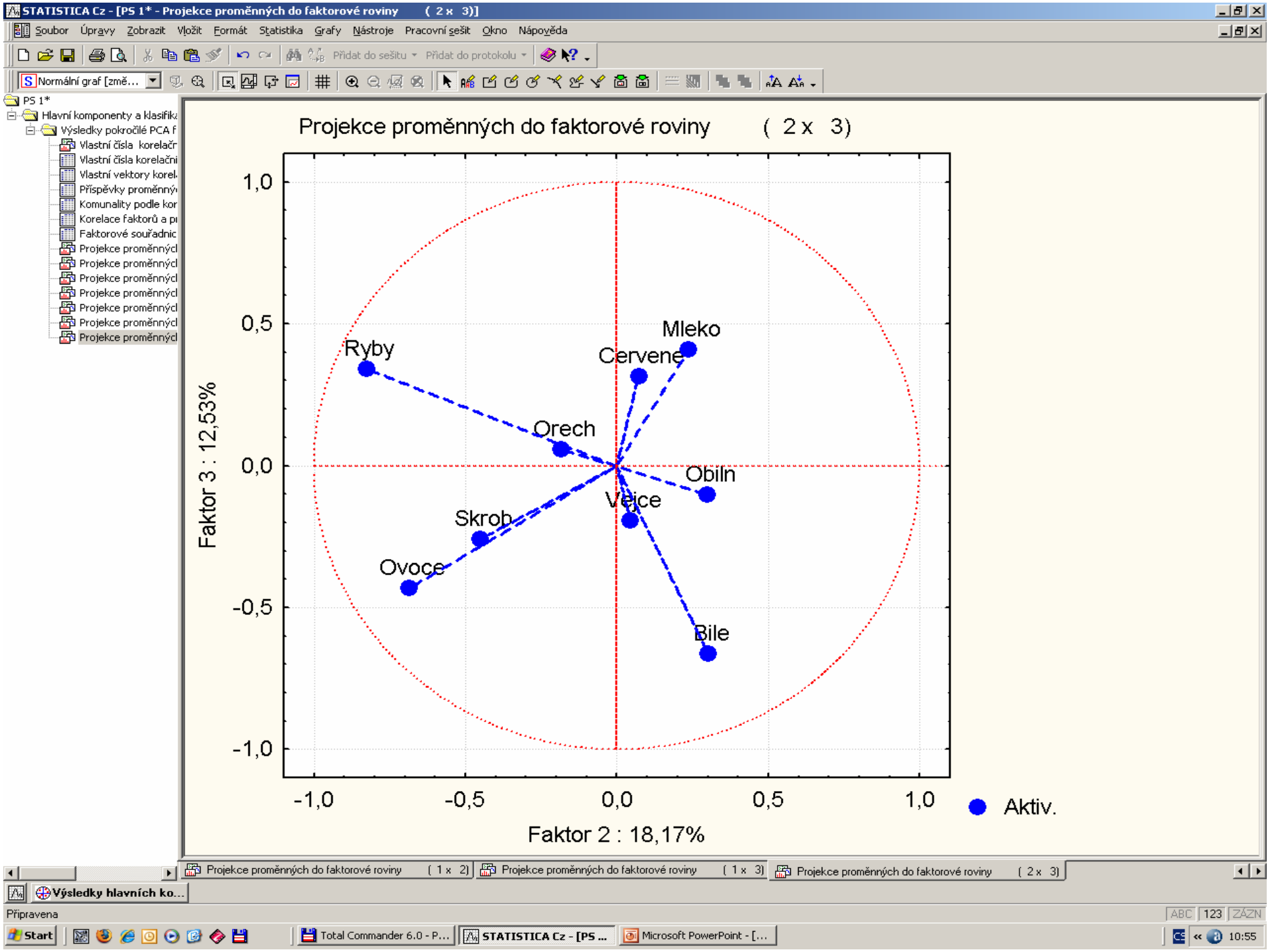
Obr. 4.16a Graf komponentních vah 1 a 2 dat *Proteiny* (STATISTICA).



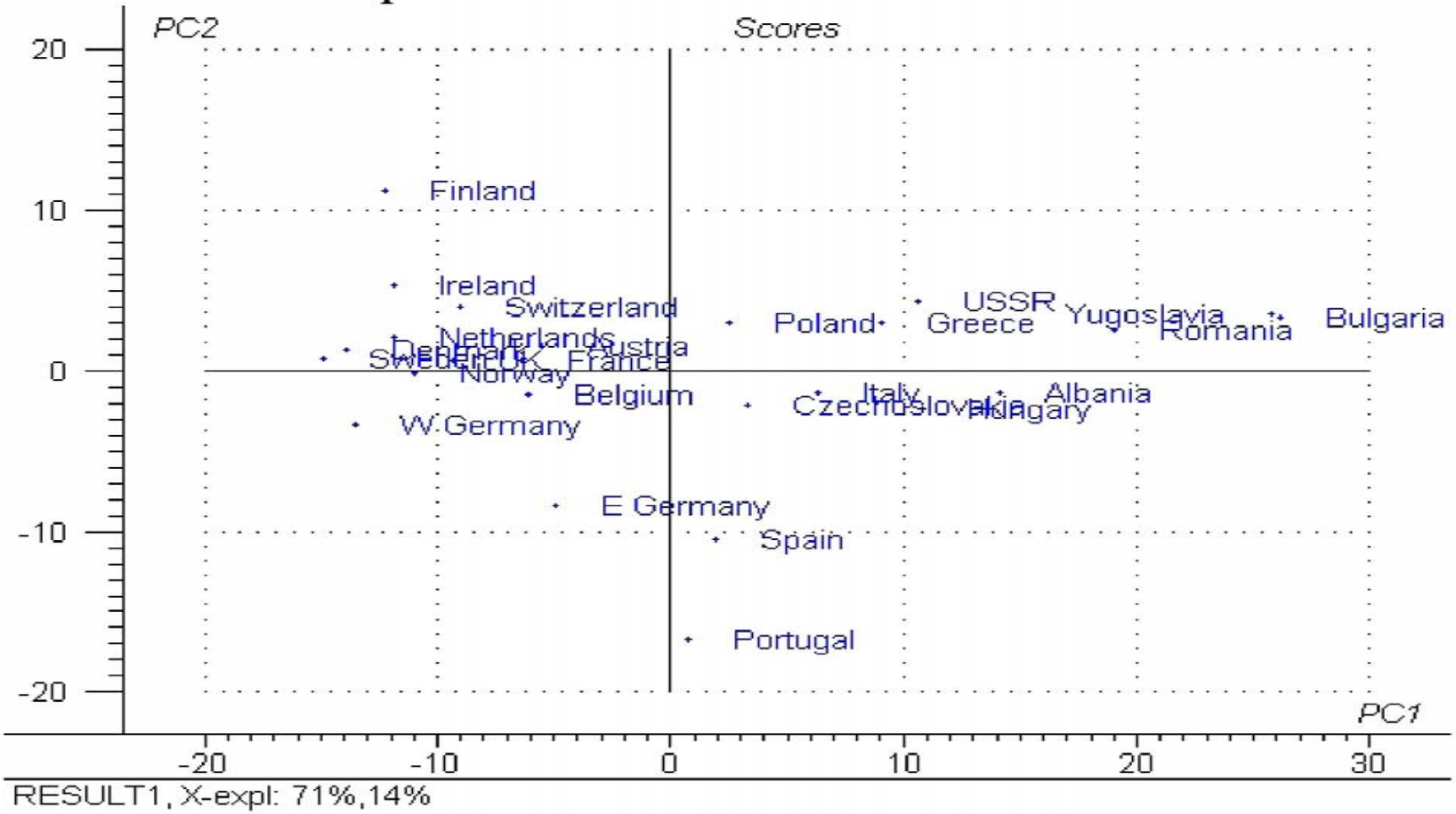
Obr. 4.16b Graf komponentních vah 1 a 2 dat *Proteiny* (UNSCRAMBLER).


STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 2)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačn
Vlastní čísla korelační
Vlastní vektory korel
Příspěvky proměnný
Komunality podle kor
Korelace faktorů a p
Faktorové souřadnic
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnných do faktorové roviny (1 x 2)
Bile
Mleko
Vejce
Cervene
Obiln
Orech
Skrob
Ovoce
Ryby
1,0
0,5
0,0
-0,5
-1,0
Faktor 2 : 18,17%
Faktor 1 : 44,52%
Aktiv.
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (1 x 2)
Výsledky hlavních ko...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
10:53

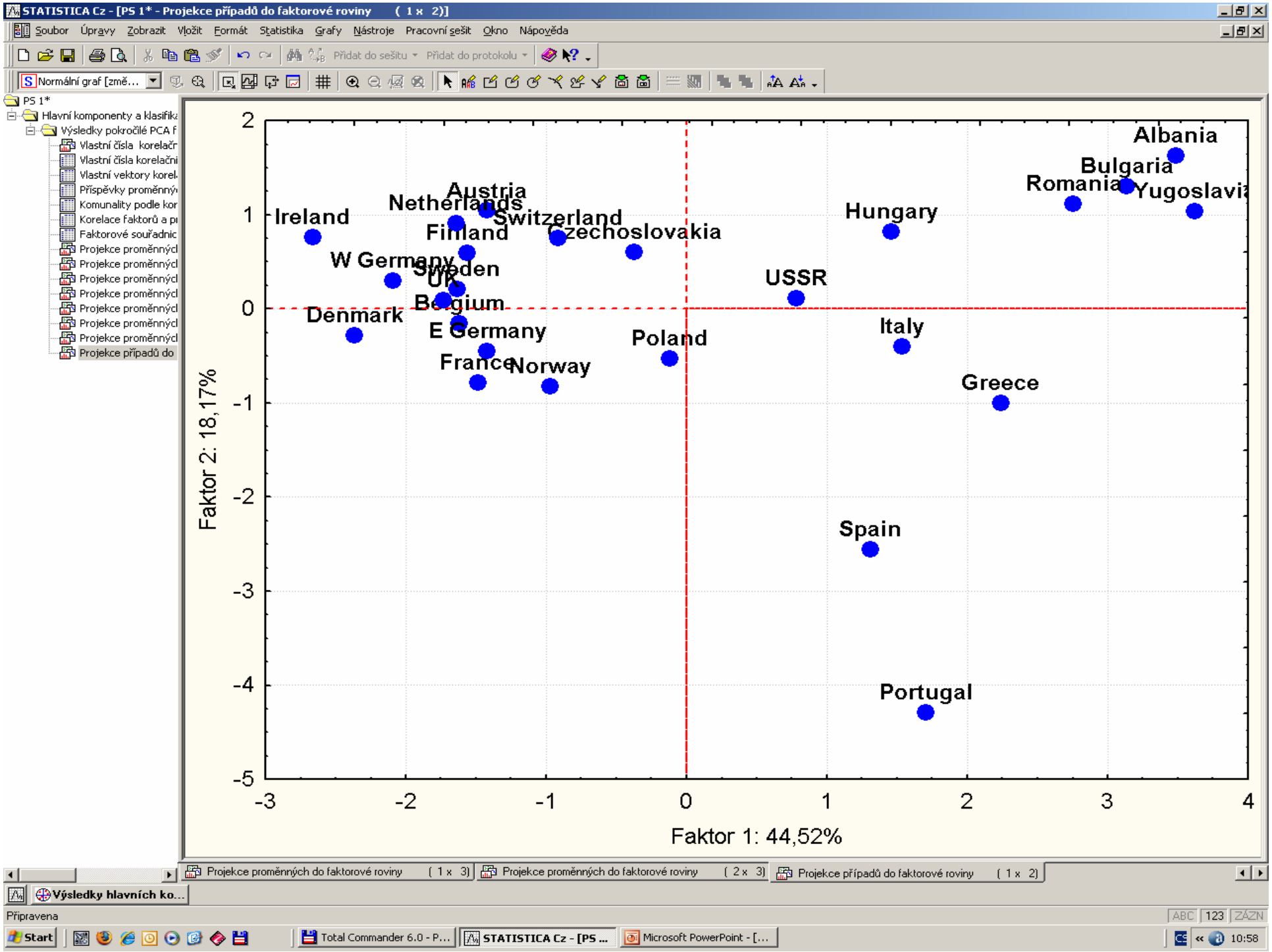

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačr
Vlastní čísla korelační
Vlastní vektory korel
Příspěvky proměnný
Komunality podle kor
Korelace faktorů a p
Faktorové souřadnic
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnných do faktorové roviny (1 x 3)
Mleko
Cervene
Ryby
Orech
Obiln
Vejce
Skrob
Ovoce
Bile
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 12,53%
Faktor 1 : 44,52%
Aktiv.
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
10:54

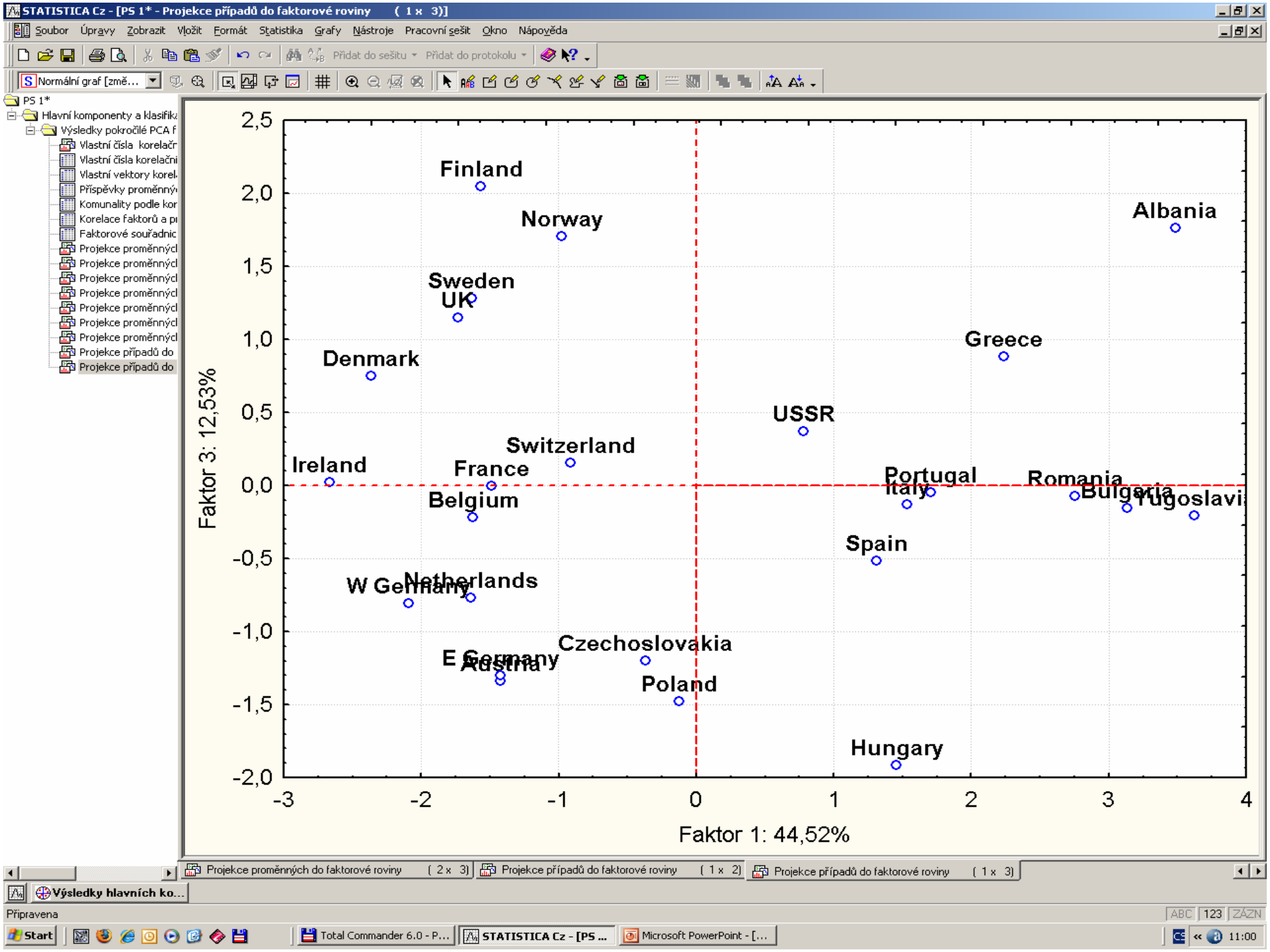

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (2 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačr
Vlastní čísla korelační
Vlastní vektory korel
Příspěvky proměnný
Komunality podle kor
Korelace faktorů a p
Faktorové souřadnic
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnných do faktorové roviny (2 x 3)
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 12,53%
Ryby
Mleko
Cervene
Orech
Obiln
Veice
Skrob
Ovoce
Bile
-1,0
-0,5
0,0
0,5
1,0
Faktor 2 : 18,17%
Aktiv.
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Výsledky hlavních ko...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
10:55

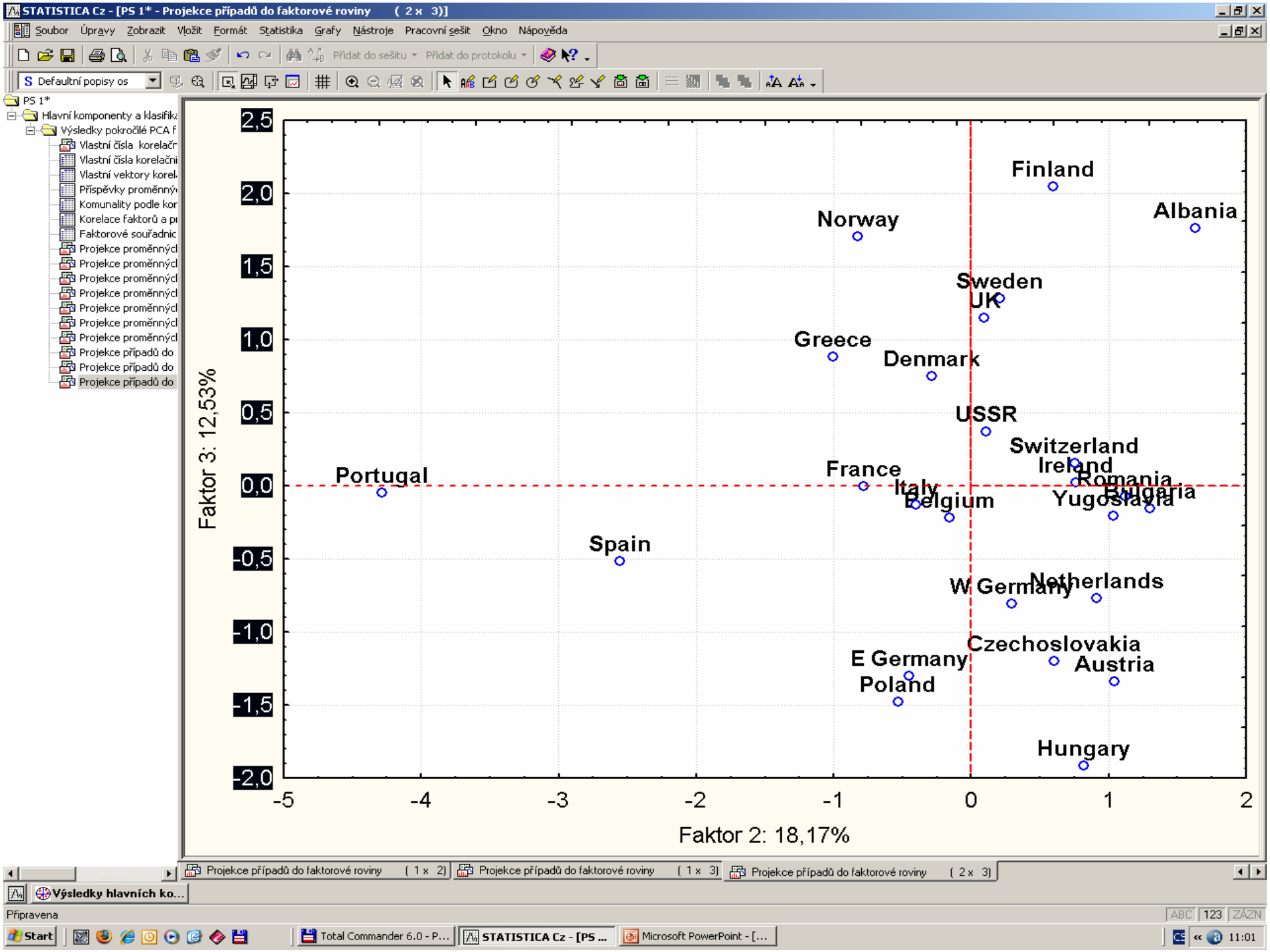
3. Rozptylový diagram komponentního skóre: roztřídil státy dle spotřeby proteinů do shluků: shluk balkánských zemí (Bulharsko, Rumunsko, Albanie, Jugoslavie), shluk s zemí Polsko, Řecko, SSSR, Československo, Itálie a Maďarsko. Španělsko koreluje s Portugalskem a Východním Německem. Velký shluk obsahuje státy západní Evropy, ze kterých vybočuje Finsko a částečně i Západní Německo.



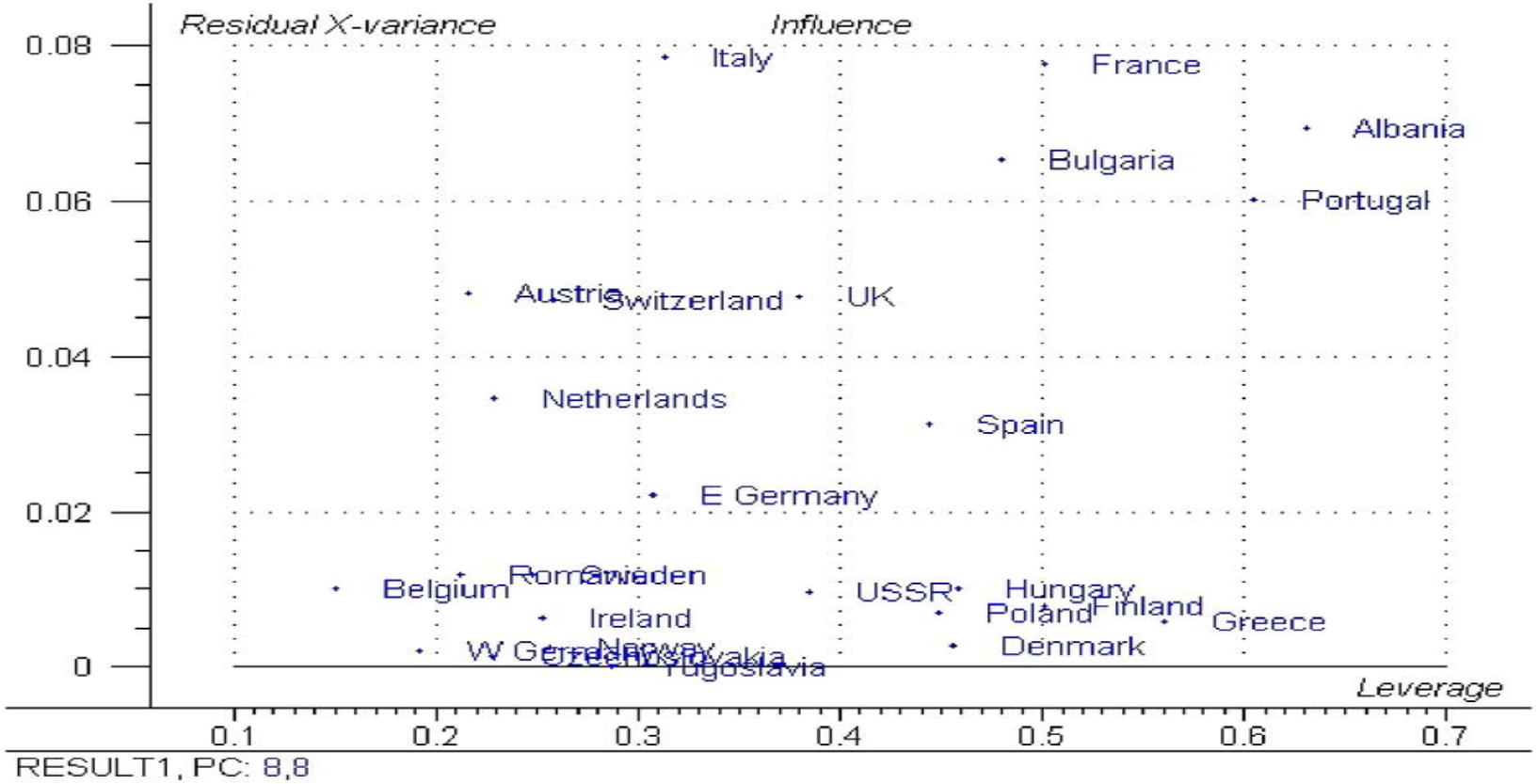
Obr. 4.17 Rozptylový diagram komponentního skóre dat *Proteiny* (UNSCRAMBLER).


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 2)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačr
Vlastní čísla korelační
Vlastní vektory korel
Příspěvky proměnný
Komunality podle kor
Korelace faktorů a p
Faktorové souřadnic
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce případů do
Albania
Bulgaria
Romania
Yugoslavi
Austria
Netherlands
Switzerland
Ireland
Finland
Czechoslovakia
Hungary
W Germany
Sweden
UK
USSR
Belgium
Denmark
E Germany
Poland
Italy
France
Norway
Greece
Spain
Portugal
Faktor 2: 18,17%
Faktor 1: 44,52%
2
1
0
-1
-2
-3
-4
-5
-3
-2
-1
0
1
2
3
4
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Projekce případů do faktorové roviny (1 x 2)
Výsledky hlavních ko...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
10:58


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačn
Vlastní čísla korelační
Vlastní vektory korel
Příspěvky proměnný
Komunality podle kor
Korelace faktorů a p
Faktorové souřadnic
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce případů do
Projekce případů do
Finland
Norway
Sweden
UK
Denmark
Ireland
France
Switzerland
Belgium
W Germany
Netherlands
E Germany
Austria
Czechoslovakia
Poland
Hungary
Spain
Italy
Portugal
USSR
Greece
Romania
Bulgaria
Yugoslavi
Albania
Faktor 3: 12,53%
Faktor 1: 44,52%
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-3
-2
-1
0
1
2
3
4
Projekce proměnných do faktorové roviny (2 x 3)
Projekce případů do faktorové roviny (1 x 2)
Projekce případů do faktorové roviny (1 x 3)
Výsledky hlavních ko...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
11:00

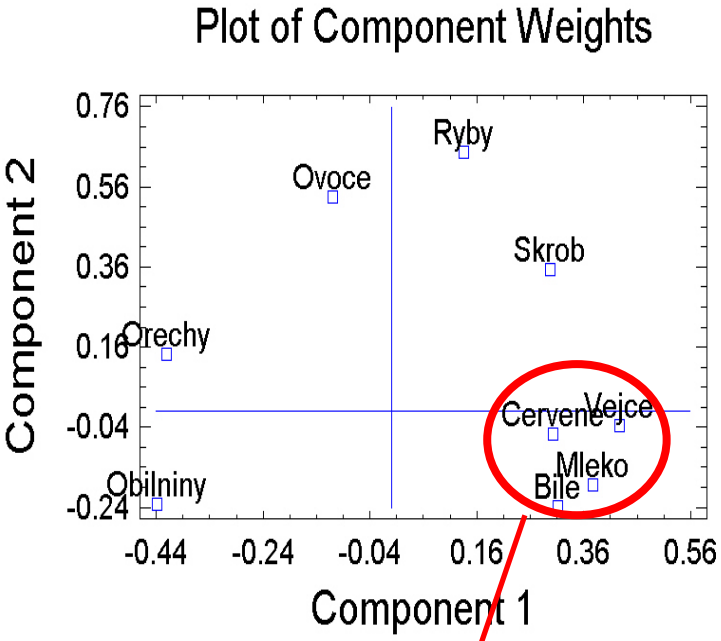

STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (2 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Defaultní popisy os
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f
Vlastní čísla korelačr
Vlastní čísla korelační
Vlastní vektory korel
Příspěvky proměnný
Komunality podle kor
Korelace faktorů a p
Faktorové souřadnic
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce proměnnýc
Projekce případů do
Projekce případů do
Projekce případů do
Finland
Albania
Norway
Sweden
UK
Greece
Denmark
USSR
Switzerland
Ireland
Romania
Bulgaria
Portugal
France
Italy
Belgium
Yugoslavia
Spain
W Germany
Netherlands
Czechoslovakia
E Germany
Austria
Poland
Hungary
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-5
-4
-3
-2
-1
0
1
2
Faktor 3: 12,53%
Faktor 2: 18,17%
Projekce případů do faktorové roviny (1 x 2)
Projekce případů do faktorové roviny (1 x 3)
Projekce případů do faktorové roviny (2 x 3)
Výsledky hlavních ko...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
11:01

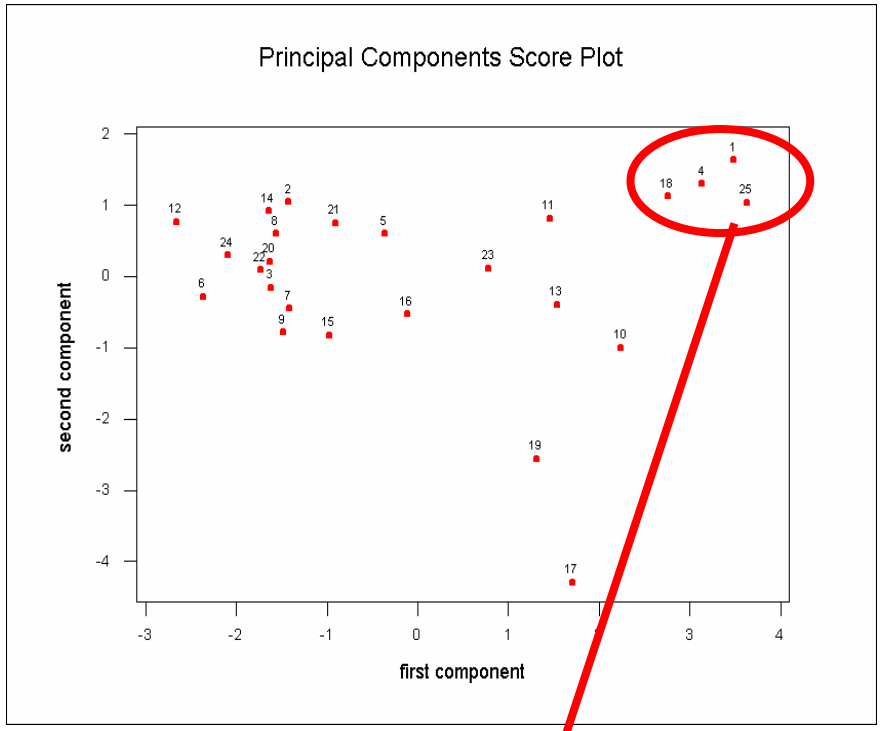
4. Graf vlivných bodů: ***vybočující objekty*** jsou země umístěné v ***horní části*** grafu (Itálie, Francie, Bulharsko, Albánie a Portugalsko) a ***extrémy*** jsou země při ***pravém okraji*** grafu jako Francie, Bulharsko, Albánie a Portugalsko, ale také Finsko a řecko.



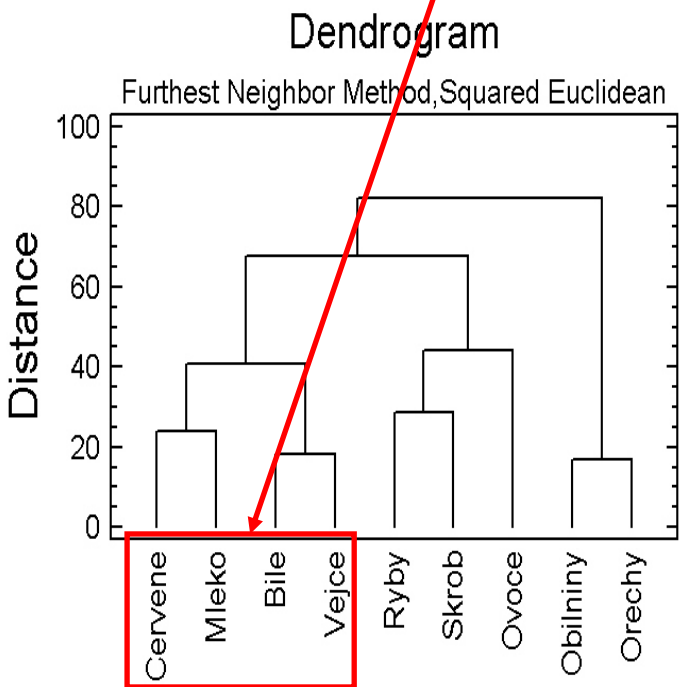
Obr. 4.18 Graf vlivných bodů statistické analýzy reziduí dat *Proteiny* (UNSCRAMBLER).

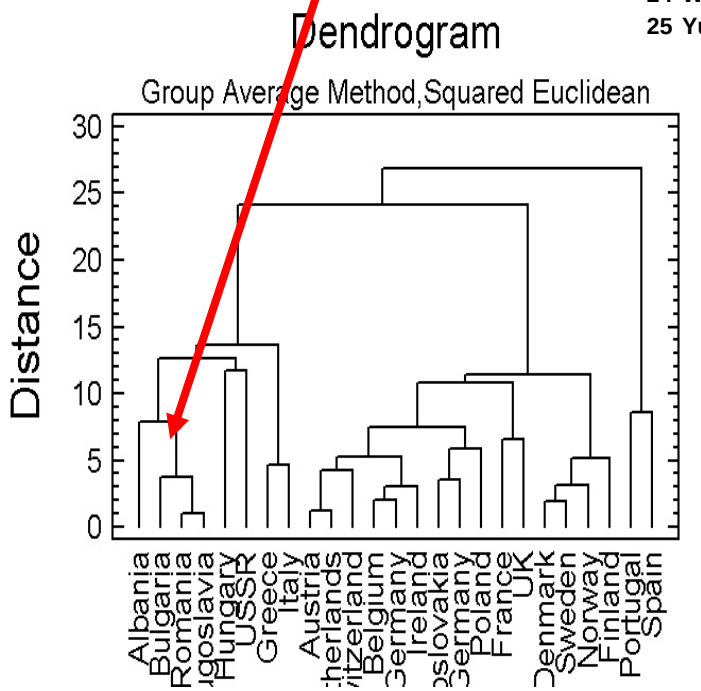
○ *Závěr:* PCA klasifikuje objekty do shluků, došlo k roztřídění zemí Evropy dle spotřeby proteinů s přihlédnutím ke 9 znakům.

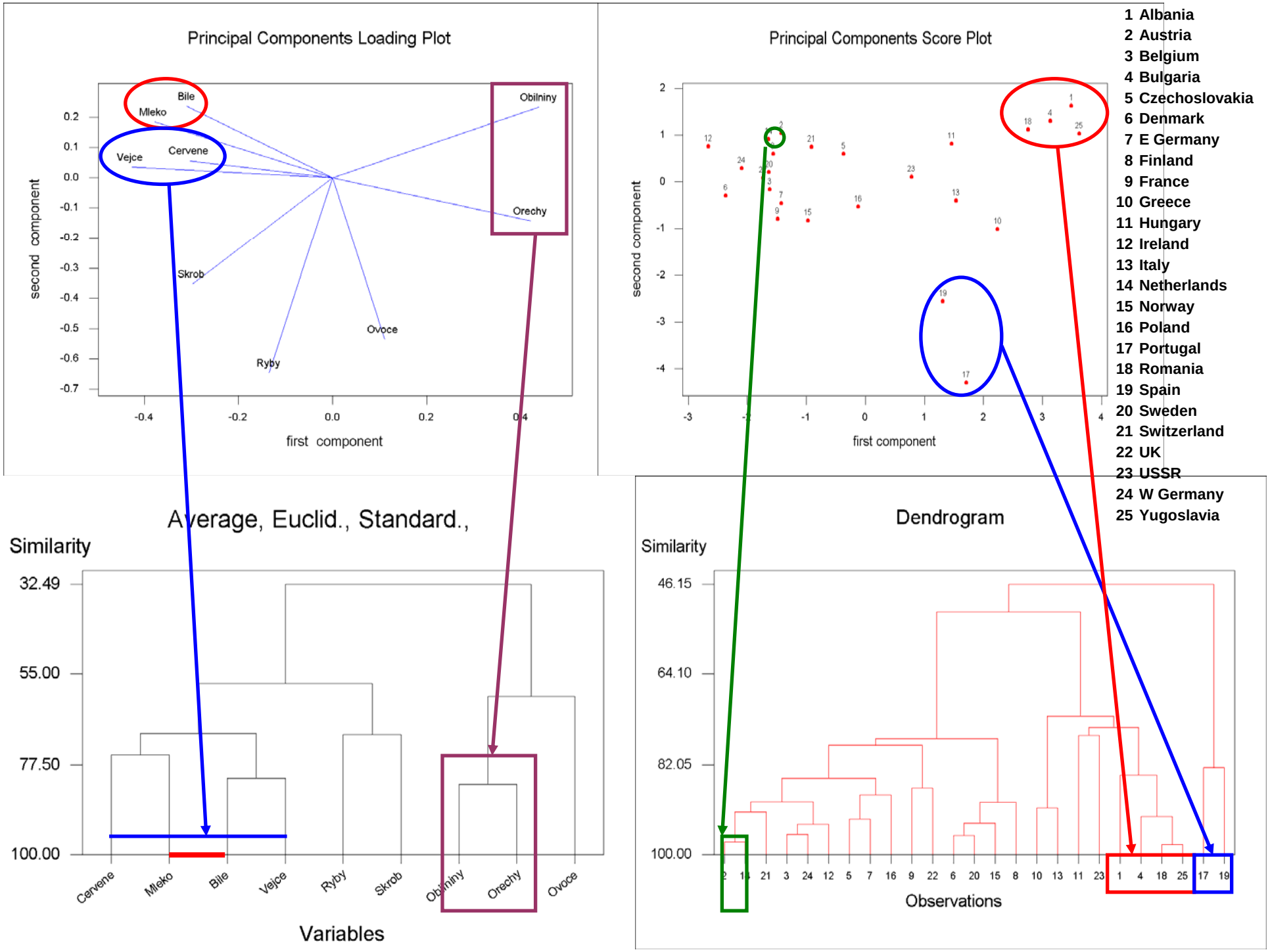




1 Albania
2 Austria
3 Belgium
4 Bulgaria
5 Czechoslovakia
6 Denmark
7 E Germany
8 Finland
9 France
10 Greece
11 Hungary
12 Ireland
13 Italy
14 Netherlands
15 Norway
16 Poland
17 Portugal
18 Romania
19 Spain
20 Sweden
21 Switzerland
22 UK
23 USSR
24 W Germany
25 Yugoslavia





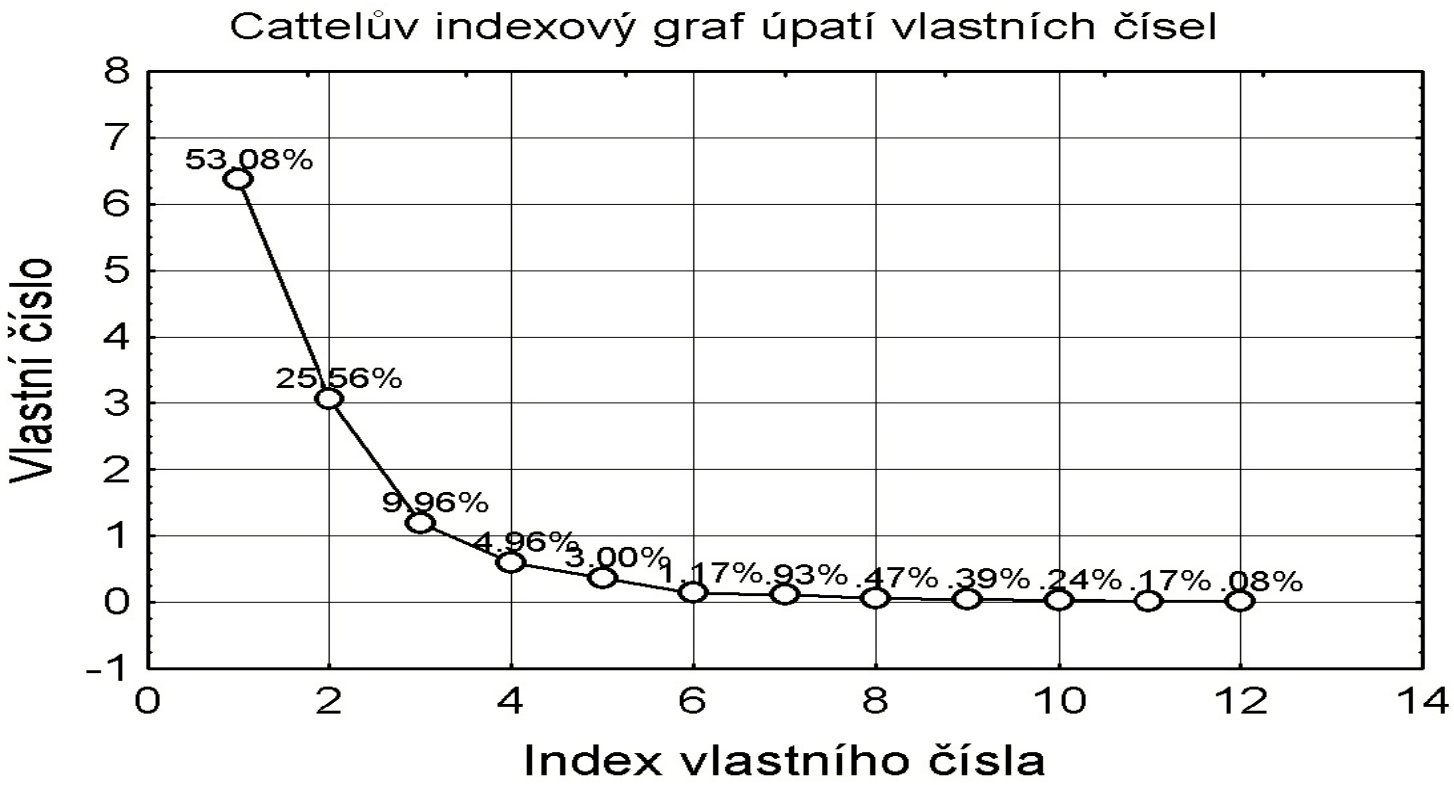

Principal Components Loading Plot
Bile
Mleko
Vejce
Cervene
Obilniny
Orechy
Skrob
Ovoce
Ryby
second component
first component
Principal Components Score Plot
second component
first component
1 Albania
2 Austria
3 Belgium
4 Bulgaria
5 Czechoslovakia
6 Denmark
7 E Germany
8 Finland
9 France
10 Greece
11 Hungary
12 Ireland
13 Italy
14 Netherlands
15 Norway
16 Poland
17 Portugal
18 Romania
19 Spain
20 Sweden
21 Switzerland
22 UK
23 USSR
24 W Germany
25 Yugoslavia
Average, Euclid., Standard.,
Similarity
32.49
55.00
77.50
100.00
Cervene
Mleko
Bile
Vejce
Ryby
Skrob
Obilniny
Orechy
Ovoce
Variables
Dendrogram
Similarity
46.15
64.10
82.05
100.00
Observations

PŘÍKLAD 4.2 ***Posouzení hrachu diagramem komponentního skóre***

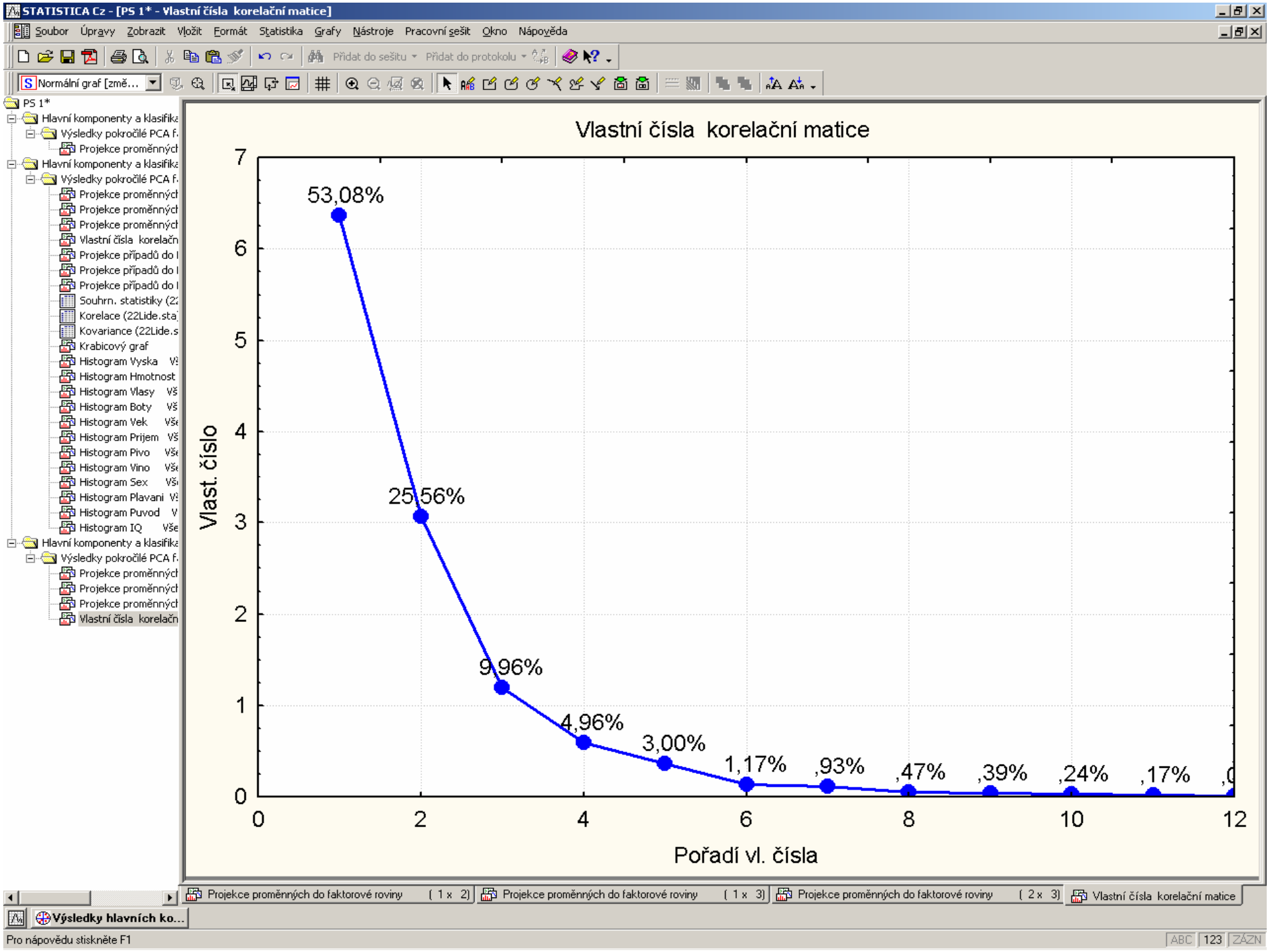
Je třeba roztřídit druhy vyšetřovaného hrachu dle smyslového posouzení hrachu člověkem, které znaky subjektivního posouzení se nejlépe hodí k popisu. Které znaky se nejlépe podílejí na popisu proměnlivosti hrachu?

○ *Řešení:*

1. Počet potřebných hlavních komponent: První hlavní komponenta popisuje 53% celkového rozptylu, druhá hlavní komponenta 25.6% a třetí hlavní komponenta 9.9%.



Obr. 4.7a Cattelův indexový graf úpatí vlastních čísel Scree Plot zdrojové matice dat *Hrách* (STATISTICA).


STATISTICA Cz - [PS 1* - Vlastní čísla korelační matice]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f.
Projekce proměnných
Vlastní čísla korelačn
Projekce případů do
Souhrn. statistiky
Korelace (22Lide.sta
Kovariance (22Lide.s
Krabicový graf
Histogram Vyska
Histogram Hmotnost
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavani
Histogram Puvod
Histogram IQ
Vlastní čísla korelační matice
Vlast. číslo
Pořadí vl. čísla
53,08%
25,56%
9,96%
4,96%
3,00%
1,17%
,93%
,47%
,39%
,24%
,17%
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Vlastní čísla korelační matice
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN

2. Graf komponentních vah: vysvětlení grafu

1) Vůně hrachu (znak *Aro*), sladkost (*Slad*) a medovost (*Med*) pozitivně korelují,

2) Tvrdost hrachu (*Tvrd)*, klasovost (*Klas*) a bezchuťovost (*Bez*) jsou rovněž pozitivně korelovány ale jsou negativně korelovány se znaky vůně hrachu (*Aro*), sladkost (*Slad*) a medovost (*Med*), protože oba shluky znaků leží na opačných stranách vůči počátku.

3) Druhá hlavní komponenta *PC*2 ukazuje, že barva 1 (*Bar1*), bělost (*Bel*) a ztráta (*Ztr*) jsou v horní části diagramu a obě jsou negativně korelovány s barvou 2 (*Bar2*) a barvou 3 (*Bar3*), které jsou umístěny v dolní části diagramu.

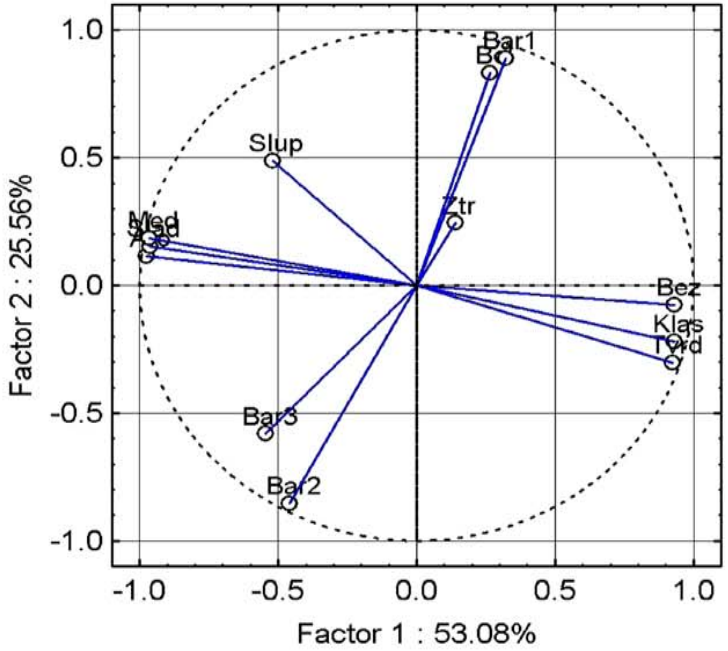
4) Vzorky hrachu nahoře diagramu jsou bělejší a vzorky v dolní části budou barevnější.

5) Slupka zrn *Slup* hrachu nekoreluje ani s bělostí (*Bel*) ani s chuťovými vlastnostmi hrachu vůně (*Aro*), sladkost (*Slad*) a medovost (*Med*).

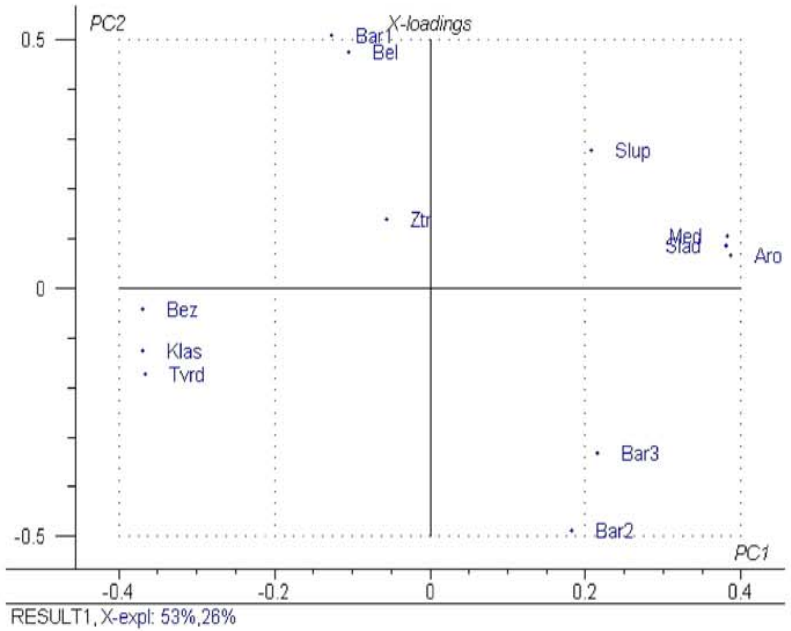
3) Druhá hlavní komponenta *PC*2 ukazuje, že barva 1 (*Bar1*), bělost (*Bel*) a ztráta (*Ztr*) jsou v horní části diagramu a obě jsou negativně korelovány s barvou 2 (*Bar2*) a barvou 3 (*Bar3*), které jsou umístěny v dolní části diagramu.

4) Vzorky hrachu nahoře diagramu jsou bělejší a vzorky v dolní části budou barevnější.

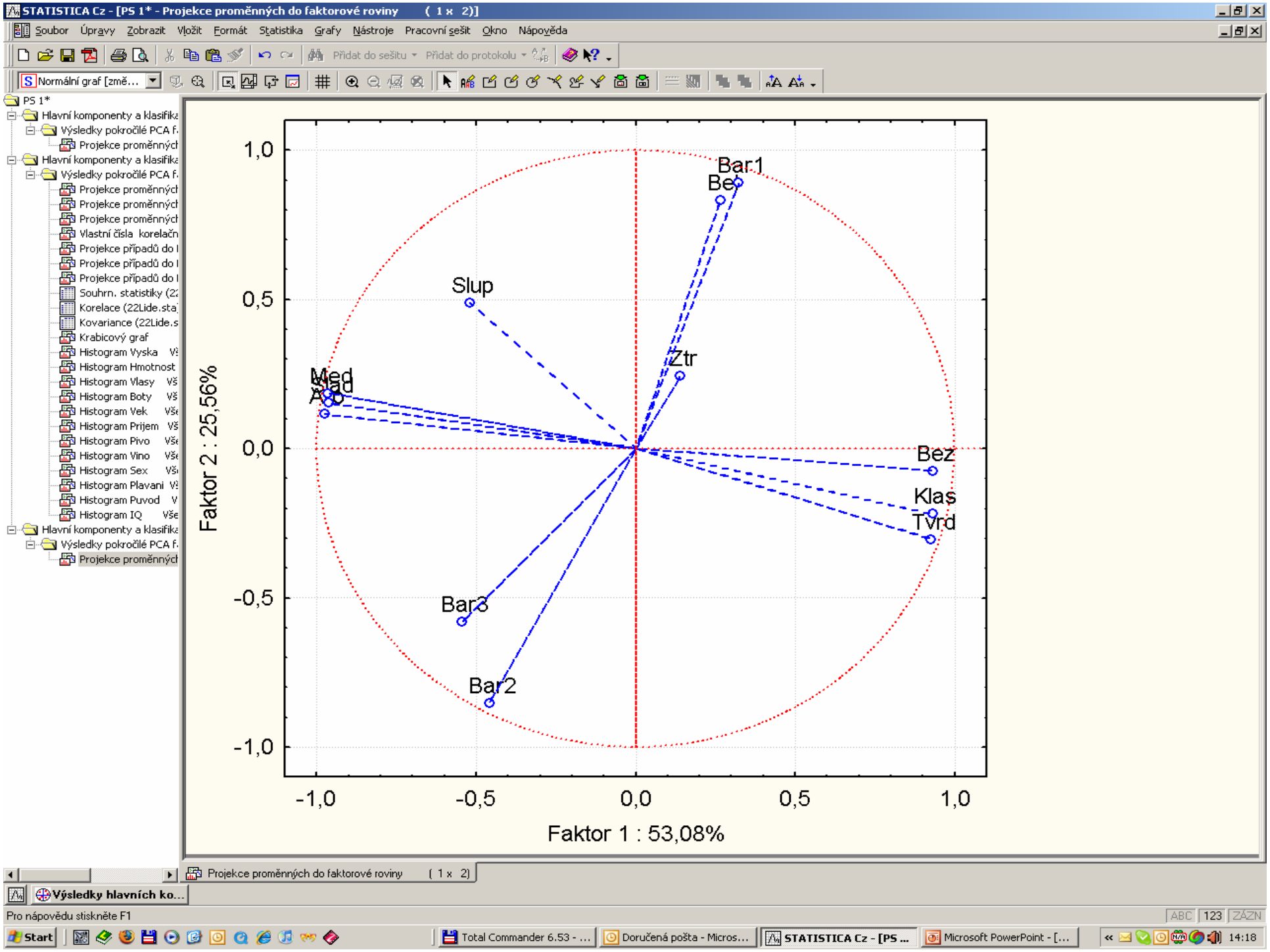
5) Slupka zrn *Slup* hrachu nekoreluje ani s bělostí (*Bel*) ani s chuťovými vlastnostmi hrachu vůně (*Aro*), sladkost (*Slad*) a medovost (*Med*).

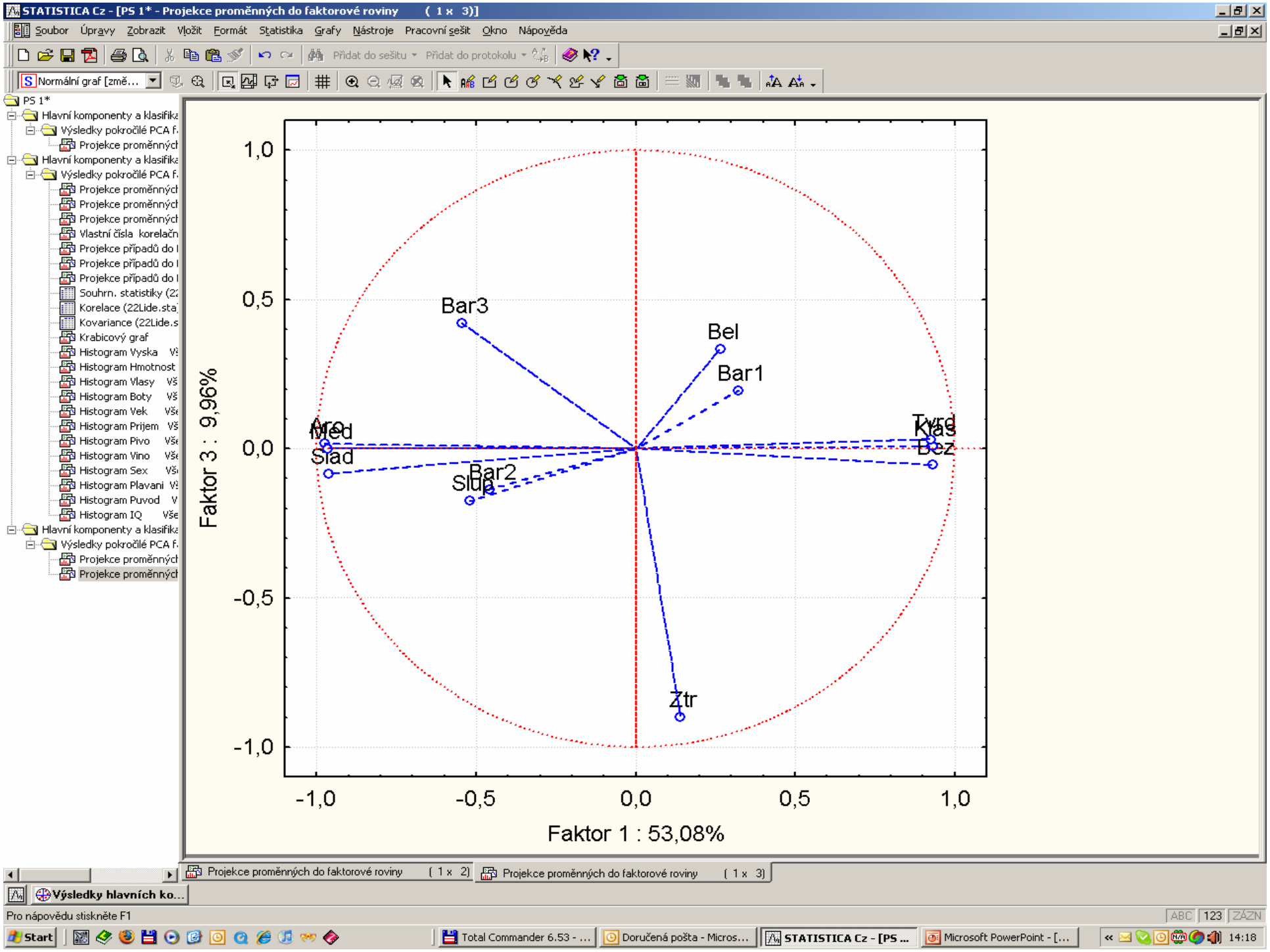


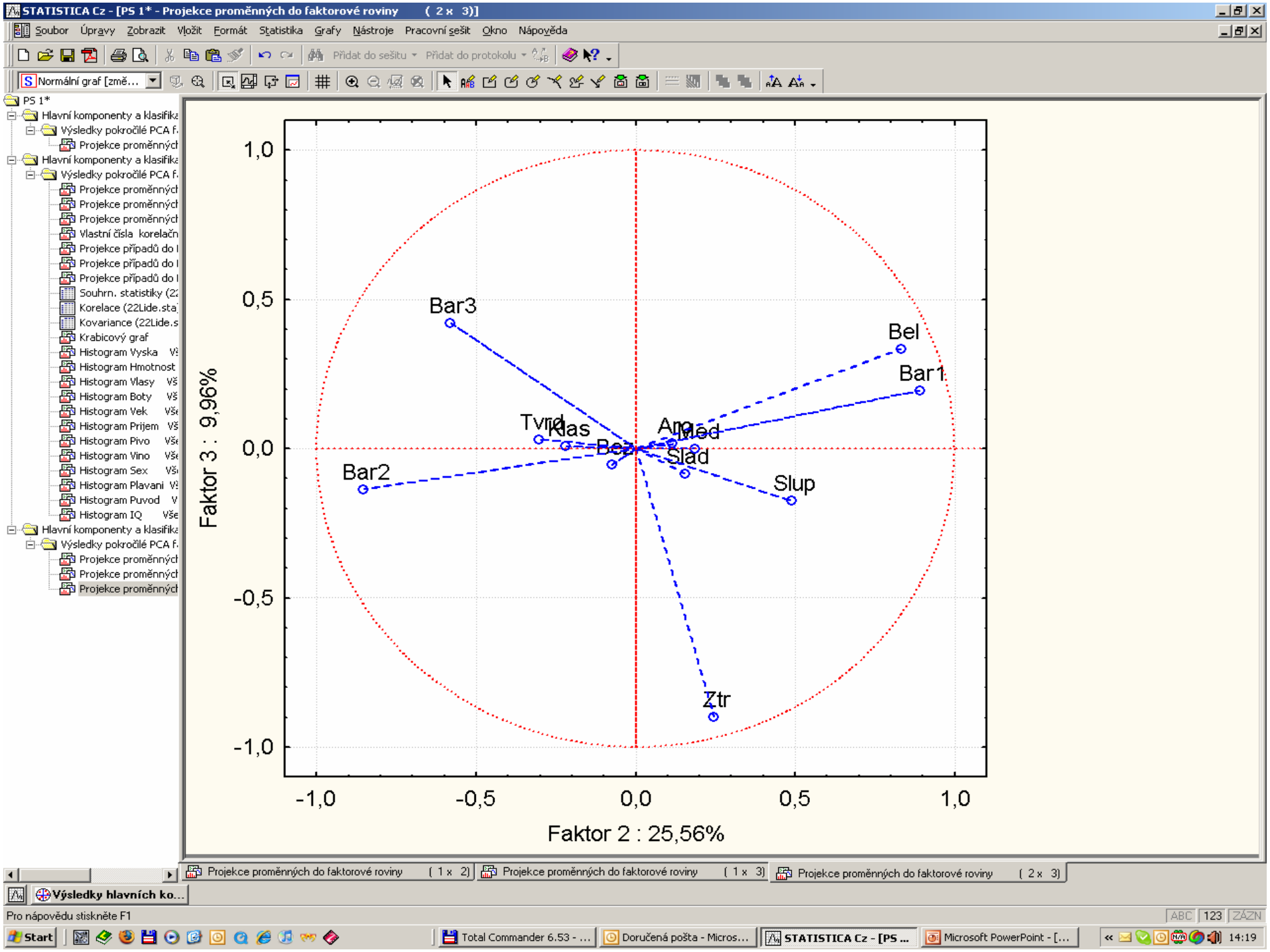
Obr. 4.8a Graf komponentních vah 1 a 2 matice dat *Hrách*.



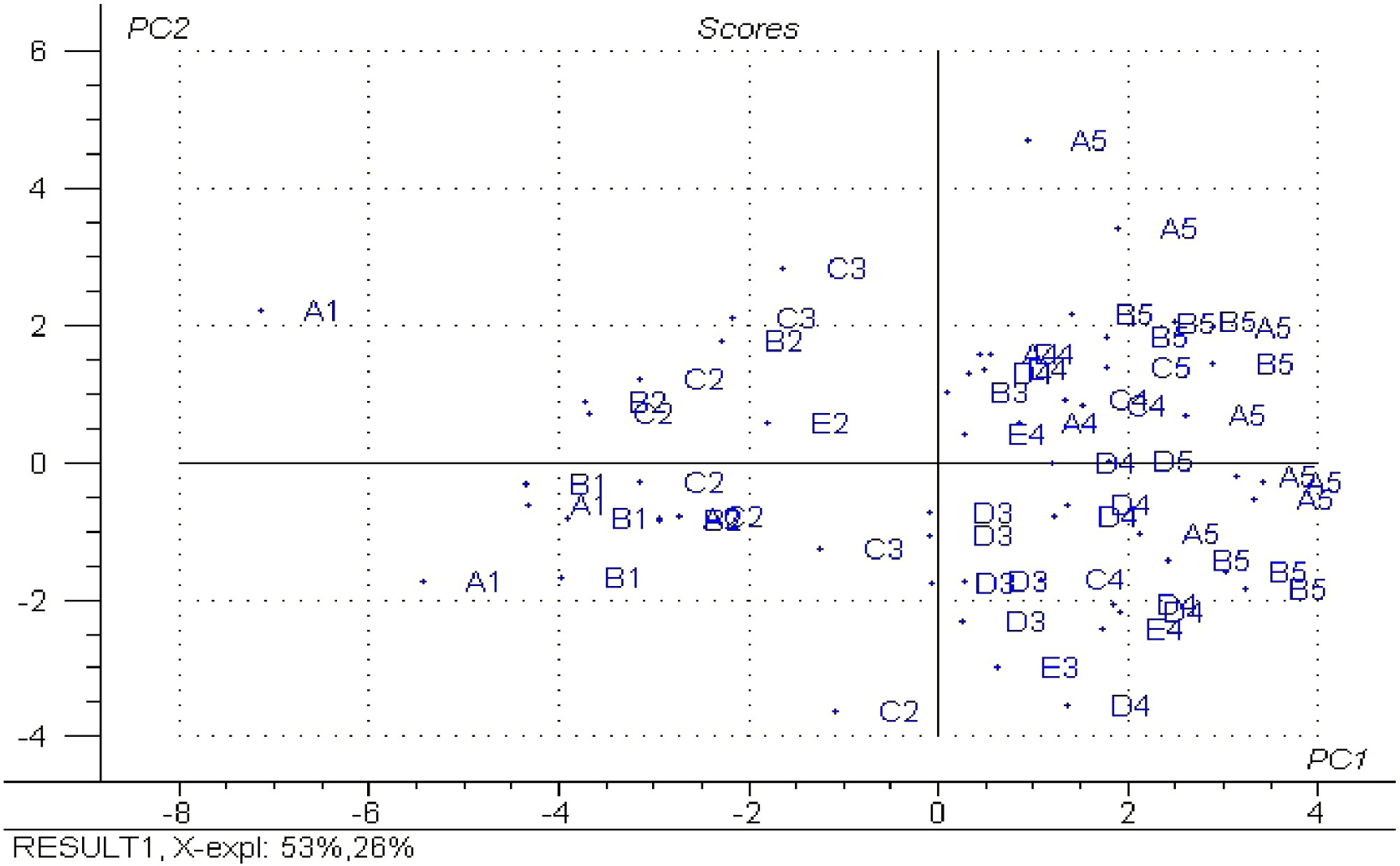
Obr. 4.8b Graf komponentních vah 1 a 2 matice dat *Hrách*.


STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 2)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f.
Projekce proměnných
Vlastní čísla korelačn
Projekce případů do
Souhrn. statistiky (2
Korelace (22Lide.sta
Kovariance (22Lide.s
Krabicový graf
Histogram Vyska
Histogram Hmotnost
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavani
Histogram Puvod
Histogram IQ
Bar1
Bar3
Bar2
Slup
Ztr
Med
Bez
Klas
Tvrd
1,0
0,5
0,0
-0,5
-1,0
Faktor 2 : 25,56%
Faktor 1 : 53,08%
Projekce proměnných do faktorové roviny (1 x 2)
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
Start
Total Commander 6.53 - ...
Doručená pošta - Micros...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
14:18

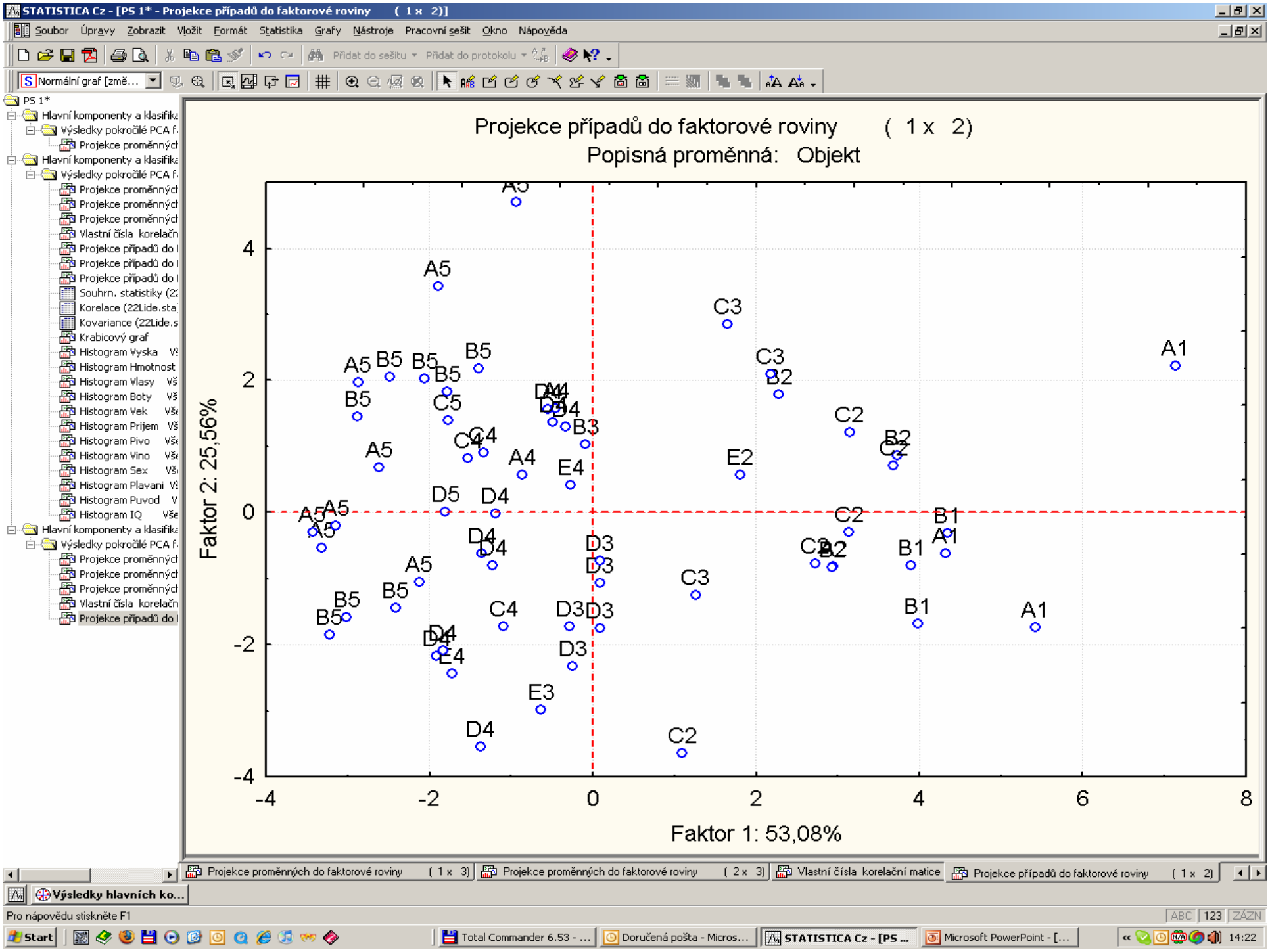

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f.
Projekce proměnných
Vlastní čísla korelačn
Projekce případů do
Souhrn. statistiky (2
Korelace (22Lide.sta
Kovariance (22Lide.s
Krabicový graf
Histogram Vyska
Histogram Hmotnost
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavani
Histogram Puvod
Histogram IQ
Bar3
Bel
Bar1
Bar2
Ztr
Faktor 3 : 9,96%
Faktor 1 : 53,08%
1,0
0,5
0,0
-0,5
-1,0
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
Start
Total Commander 6.53 - ...
Doručená pošta - Micros...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
14:18

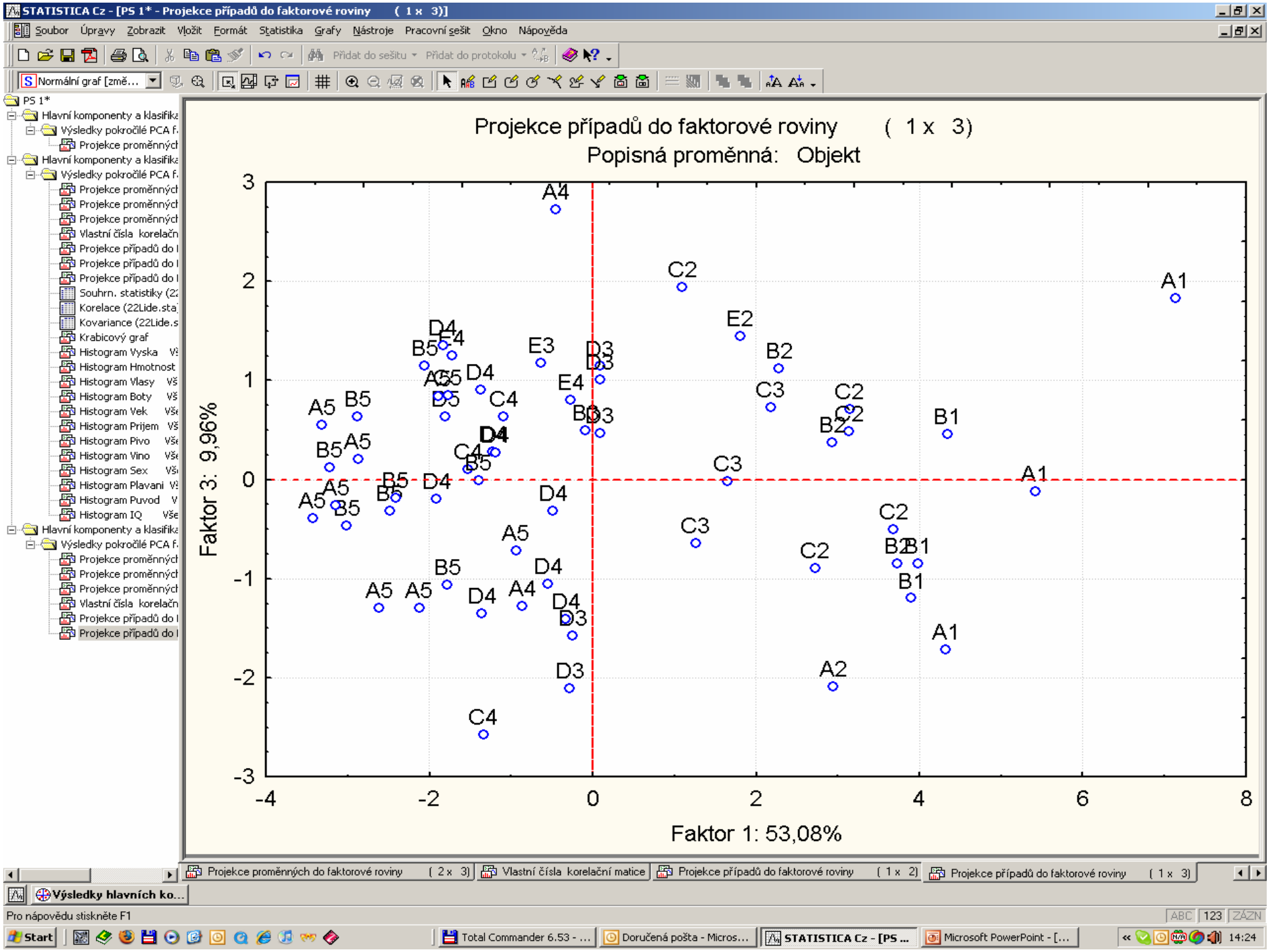

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (2 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Hlavní komponenty a klasifika
Výsledky pokročilé PCA f.
Projekce proměnných
Vlastní čísla korelačn
Projekce případů do
Souhrn. statistiky (2
Korelace (22Lide.sta
Kovariance (22Lide.s
Krabicový graf
Histogram Vyska
Histogram Hmotnost
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavani
Histogram Puvod
Histogram IQ
Bar3
Bel
Bar1
Tvrd
Klas
Arg
Mod
Bar2
Slad
Slup
Ztr
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 9,96%
Faktor 2 : 25,56%
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
Start
Total Commander 6.53 - ...
Doručená pošta - Micros...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
14:19

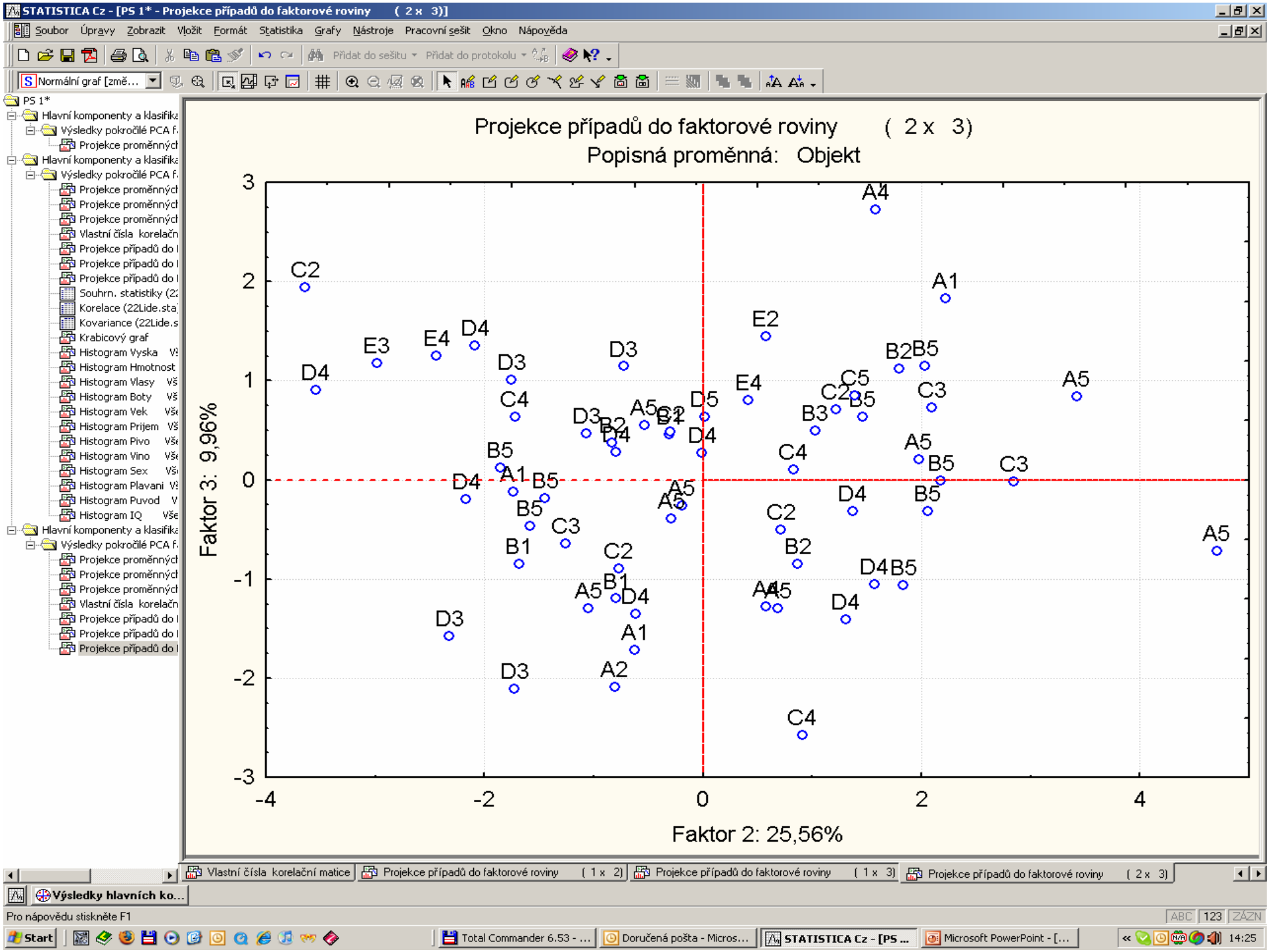
3. Rozptylový diagram komponentního skóre: Písmena *A, B, C, D* a *E* označují typ odrůdy hrachu, zatímco číslo 1, 2, 3, 4 a 5 značí čas sklizně. *PC*1 souvisí s časem sklizně.



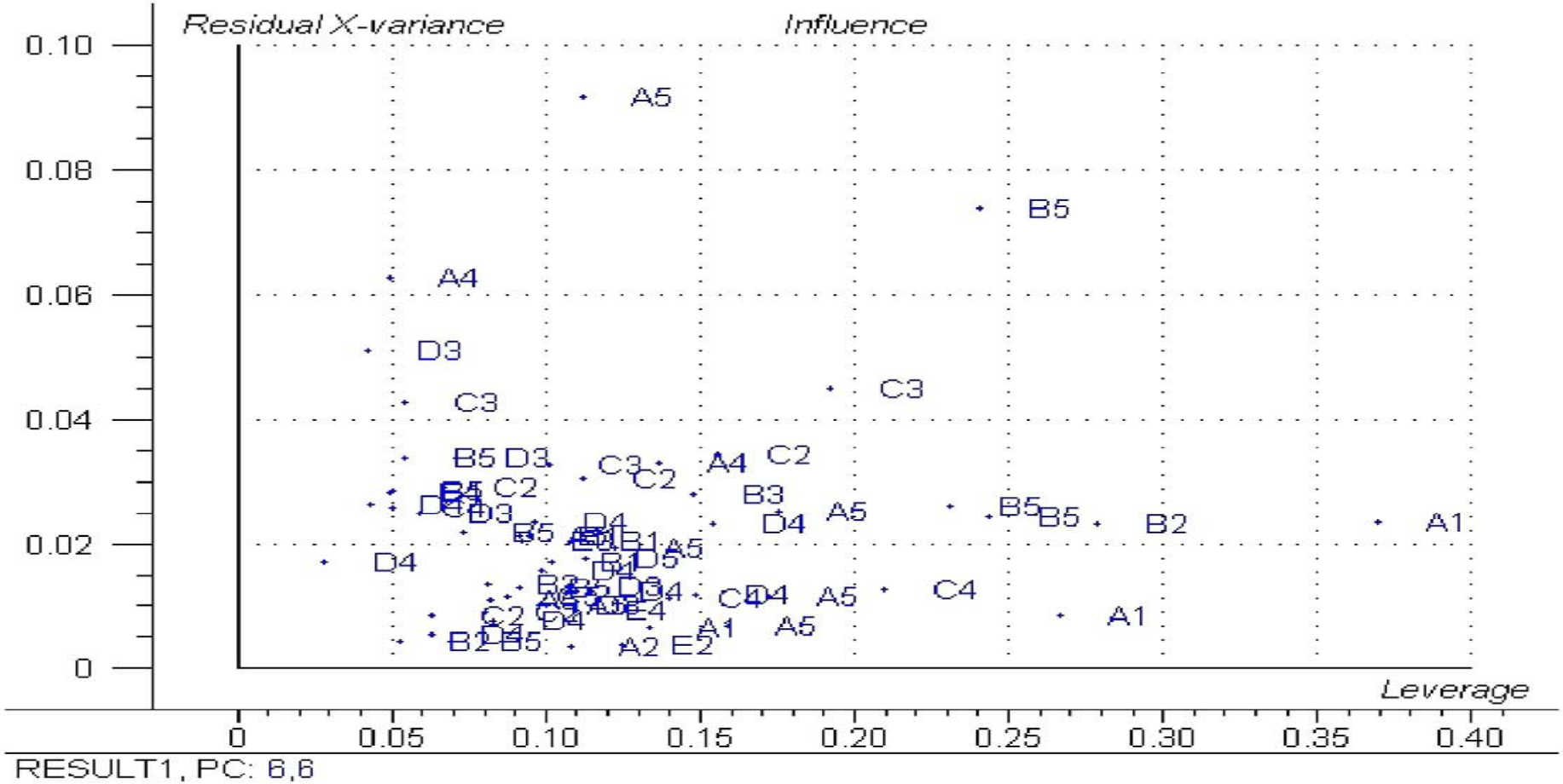
Obr. 4.9 Rozptylový diagram komponentního skóre dat *Hrách*.


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 2)]
Projekce případů do faktorové roviny (1 x 2)
Popisná proměnná: Objekt
Faktor 2: 25,56%
Faktor 1: 53,08%
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Vlastní čísla korelační matice
Projekce případů do faktorové roviny (1 x 2)
Výsledky hlavních ko...
Pro nápovědu stiskněte F1


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Projekce případů do faktorové roviny (1 x 3)
Popisná proměnná: Objekt
Faktor 3: 9,96%
Faktor 1: 53,08%
Projekce proměnných do faktorové roviny (2 x 3)
Vlastní čísla korelační matice
Projekce případů do faktorové roviny (1 x 2)
Projekce případů do faktorové roviny (1 x 3)
Pro nápovědu stiskněte F1


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (2 x 3)]
Projekce případů do faktorové roviny (2 x 3)
Popisná proměnná: Objekt
Faktor 3: 9,96%
Faktor 2: 25,56%
Vlastní čísla korelační matice
Projekce případů do faktorové roviny (1 x 2)
Projekce případů do faktorové roviny (1 x 3)
Projekce případů do faktorové roviny (2 x 3)
Výsledky hlavních ko...
Pro nápovědu stiskněte F1

4. Graf vlivných bodů: objekty které nejsou dostatečně popsány PCA modelem jsou umístěné při horním okraji grafu.



Obr. 4.10 Graf vlivných bodů statistické analýzy reziduí objektů dat *Hrách*.

○ *Závěr:* byl posouzen graf komponentního skóre k roztřídění odrůd hrachu dle svých dvou dominantních vlastností, dle času sklizně a dle svých odrůd.

PŘÍKLAD 9.13 *Klasifikace vlastností rozličných druhů kávy*

Byl získán výběr 43 vzorků kávy, pocházejících ze 30 zemí. U každého druhu kávy byly změřeny jeho chemické a fyzikální vlastnosti. Splňují data požadavky na homogenitu a je možné indikovat dvě či více rozličných kategorií? Vytvořte dendrogram klasifikovaných druhů kávy.

○ ***Data:*** Soubor dat *Kava* obsahuje 2 druhy kávy, Robusta a Arabica ve 43 vzorcích ze 30 zemí a popsaných 13 fyzikálně-chemickými znaky:

i značí index kávy,
Objekt značí původ kávy,
Voda značí obsah vody x_1,
Zrno značí hmotnost zrn x_2,
Extrakt značí extrakt x_3,
pH značí hodnotu pH x_4,
Acidita značí hodnotu volné acidity x_5,
Mineral značí obsah minerálů x_6,
Tuky značí obsah tuků x_7,
Kofein značí obsah kofeinu x_8,
Trinonelin značí obsah trinonelinu x_9,
Kchlorogen značí obsah kyseliny chlorogenikové x_{10},
Kneochlor značí obsah kyseliny neochlorogenikové x_{11},
Kisochlor značí obsah kyseliny isochlorogenikové x_{12},
Sumakys značí sumu kyselin chlorogenikových x_{13}.

i	*Objekt*	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	Mexico 1	8.9	156.6	33.5	5.8	32.7	3.8	15.2	1.1	1.0	5.4	0.4	0.8	6.6
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
43	Hawai	9.7	191.2	35.1	5.6	34.6	4.2	14.2	1.1	0.9	0.7	0.5	0.3	6.5

U 43 vzorků kávy ze 30 zemí byly změřeny chemické a fyzikální vlastnosti. Nalezněte shluky podobných vlastností a shluky podobných prvků.

Data: 13 proměnných (sloupce): **i** index kávy, **j** je původ kávy, **x1** obsah vody, **x2** hmotnost zrn, **x3** extrakt, **x4** pH, **x5** volná acidita, **x6** obsah minerálů, **x7** tuky, **x8** kofein, **x9** trinonelin, **x10** kyselina chlorogeniková, **x11** kyselina neochlorogeniková, **x12** kyseliny isochlorogeniková, **x13** suma kyselin chlorogenikových.

i	ii	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
1	Mexico 1	8.9	156.6	33.5	5.8	32.7	3.8	15.2	1.1	1	5.4	0.4	0.8	6.6
2	Mexico 2	7.4	157.3	32.1	5.8	30.8	3.7	15	1.3	1	5.1	0.3	1	6.4
3	Guatemala	9.7	152.9	33.1	5.3	36.7	4.2	16.1	1.2	1	5.9	0.2	0.8	6.9
4	Honduras	10.4	174	31.5	5.6	34.2	3.9	15.8	1.1	0.9	5.9	0.4	0.6	6.8
5	Salvador 1	10.5	145.1	35.2	5.8	31.8	4.1	15.2	1.1	1	5.1	0.5	0.7	6.3
6	Salvador 2	10	156.4	34.5	5.8	32.6	3.9	15.4	1.2	0.8	5.3	0.4	0.7	6.4
7	Salvador 3	8.2	155.2	32.4	5.6	29.7	3.8	15.6	1.3	1.2	4.8	0.3	0.7	5.9
8	Nicaragua 1	9.2	167.8	30.6	5.9	28.9	3.8	15.1	1.3	1	5	0.3	0.7	5.9
9	Nicaragua 2	9.3	165.4	35.3	5.8	32.6	4.2	14.3	1.2	1	5.5	0.4	0.8	6.7
10	Costa Rica 1	7.1	180.3	33	5.8	29.3	4	15.1	1.3	1	5.1	0.3	0.7	6.1
11	Costa Rica 2	7.6	153.2	36	5.9	30.5	3.9	16.8	1.4	1.1	5.3	0.3	0.7	6.3
12	Costa Rica 3	7.3	159.6	35	5.8	29.9	3.7	16.5	1.2	1.2	5.5	0.3	0.7	6.5
13	Panama	9.3	161.8	32.4	5.8	31	3.7	15.5	1.3	1.2	5.6	0.3	0.6	6.6
14	Haiti	8.3	160.8	35.7	5.9	30	4.4	13	1.3	1	6.1	0.6	0.8	7.5
15	Dominica	11.6	174.8	32.5	5.4	35.2	3.7	14.5	1	1	5.7	0.3	0.5	6.5
16	Venezuela 1	9.7	169.1	34	5.8	31.6	4	15.7	1.3	1.3	5.1	0.3	0.3	6.2
17	Venezuela 2	10.6	163.7	35	5.8	35	3.8	15.8	1.2	1.1	6.1	0.3	0.9	7.3
18	Columbia 1	12	178.8	32.9	5.3	36.2	4.4	15.6	1.3	1	5.6	0.4	0.7	6.7
19	Columbia 2	10.6	169.1	33	5.3	37.5	4.4	15.1	1.2	1	6.1	0.1	0.6	6.9
20	Ecuador	11.6	148.5	34.6	5.3	39.4	4.2	14.6	1	1.1	5.7	0.5	0.4	6.6
21	Peru	10.1	153.7	34.5	6	28.4	3.7	15.9	1.3	1.1	6.1	0.4	0.8	7.3
22	Brasil 1	10.7	134.5	29.8	5.4	34.1	3.7	15.8	1.2	0.9	5.4	0.4	0.6	6.4
23	Brasil 2	9.7	160.7	33.8	5.3	37.2	4.2	15.2	1.1	0.9	5.4	0.3	0.5	6.2
24	Brasil 3	10.8	133.2	35	5.2	34.7	4.5	15.1	1.2	1.4	5	0.5	0.5	6
25	Brasil 4	11.1	131.7	29.8	5.4	33	4.1	15.8	1.1	1.2	5.1	0.5	0.5	6
26	Brasil 5	10.1	121.6	33.6	5.4	34.7	3.5	15.4	1.1	0.9	5.5	0.4	0.6	6.5
27	Cotedivoir	8	141.8	33.7	5.8	41.9	4.2	11	2	0.5	6.4	0.6	1.5	8.5
28	Togo	9	144.6	29.9	5.6	38	3.9	7.5	1.9	0.3	5.4	0.8	0.9	7.1
29	Cameroon	10.3	119.2	35.5	6.1	41.7	4.1	9.8	1.8	0.8	6	0.5	1.1	7.6
30	Congo	10	143.2	31.7	6.1	29.3	4.1	17	1.2	0.6	5.4	0.3	0.7	6.4
31	Angola 1	9.2	150.4	31.5	5.7	36.4	4.2	8.5	1.9	0.6	5.9	0.6	1.4	7.9
32	Angola 2	9.6	136.6	33.9	5.6	38.2	4	7.2	2.2	0.5	6.2	0.4	1.6	8.3
33	Angola 3	9.5	136.5	32	5.8	31.2	3.8	14.6	1.3	1	5.2	0.4	0.8	6.4
34	Ethiopie	9.3	124.2	35.6	5.8	31.8	3.8	15.7	0.9	0.9	5.5	0.2	0.8	6.5
35	Uganda 1	10.5	132.9	36.2	5.4	36.7	4	15.6	1	1	5.9	0.4	0.6	6.9
36	Uganda 2	10.7	181.2	33.1	5.8	30.7	3.9	15.8	1.3	1.1	5.3	0.3	0.6	6.2
37	Kenya	10.5	159.1	30.3	5.6	31.5	3.7	15.2	1.3	0.9	5.1	0.3	0.7	6
38	Tanganika	9.9	169.4	29	5.6	30.2	3.7	16.5	1.3	0.9	5	0.2	0.7	5.9
39	Madagascar	5	152	30.6	5.3	40.5	3.9	9.6	1.6	0.7	5.3	0.6	0.8	6.7
40	India	11.5	156.8	30.8	5.5	37.5	3.9	14.3	1.2	1	5.8	0.4	0.4	6.6
41	Sumatra	8.4	110.8	31.6	5.7	43.4	4.5	10.1	1.7	0.8	6.3	0.7	0.9	7.9
42	Java	5.6	163.1	34.5	5.5	33.3	4	16	1.2	1.1	5.1	0.3	0.8	6.3
43	Hawai	9.7	191.2	35.1	5.6	34.6	4.2	14.2	1.1	0.9	0.7	0.5	0.3	6.5

○ ***Řešení:*** Graf komponentních vah znaků odhaluje především korelaci znaků. Je-li úhel mezi průvodiči dvou znaků malý, jsou dva znaky v silné korelaci.

První shluk obsahuje znaky *Voda, pH, Kchlorogen, Sumakysel*, *Mineral, Kofein, Trinonelin, Kneochlor, Kizochlor,* a *Tuky*.

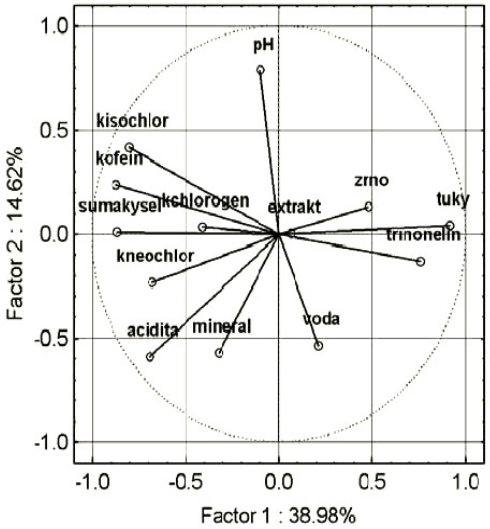
Druhý shluk obsahuje dva znaky, *Extrakt* a *Acidita.*

Vznik shluků druhů kávy lze sledovat na grafu komponentního skóre objektů a na dendrogramu objektů.

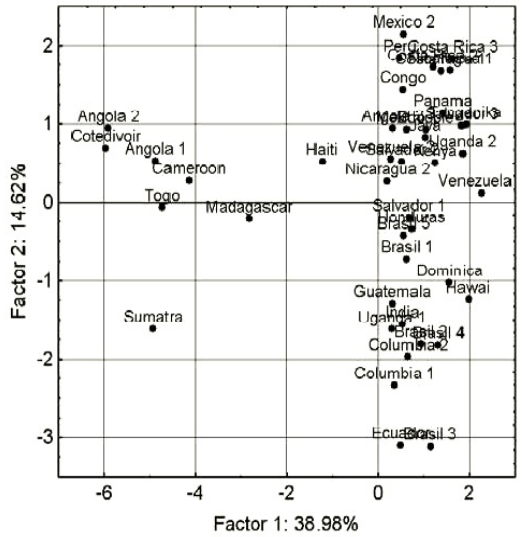
Graf komponentního skóre ukazuje, že 43 objektů čili druhů kávy v datovém souboru *Kava* nejsou dostatečně homogenní.

Objekty zde lze rozdělit do dvou shluků, v prvním vlevo je 7 objektů a ve druhém svislém shluku vpravo je zbývajících 36 objektů. Klasifikace do těchto shluků je především vlivem znaků *Tuky, Kofein*, *Trinonelin* a *Sumakys*.

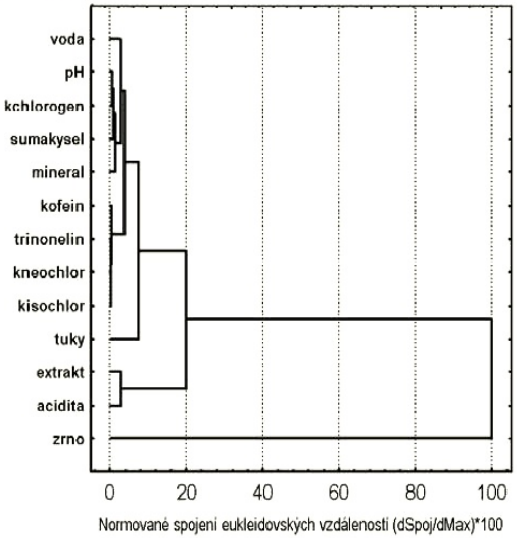
Na **dendrogramu objektů** při postupu zprava doleva je zřejmé, že druhy kávy lze rozdělit do dvou velkých shluků. Větší shluk nazvaný Arabica lze dále rozdělit na dva menší shluky Arabica A a Arabica B a jeden odlehlý objekt. Ve spodní části obrázku zůstává jeden větší shluk 13 druhů kávy, patřících zřejmě do druhu Robusta.



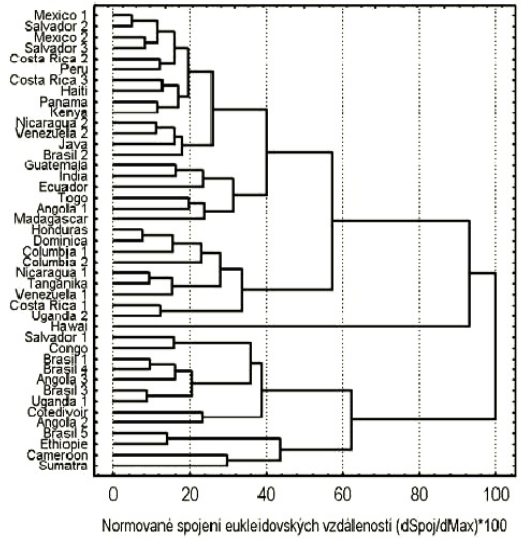
Graf komponentních vah znaků matice dat *Kava*, (STATISTICA).



Graf komponentního skóre objektů matice dat *Kava*

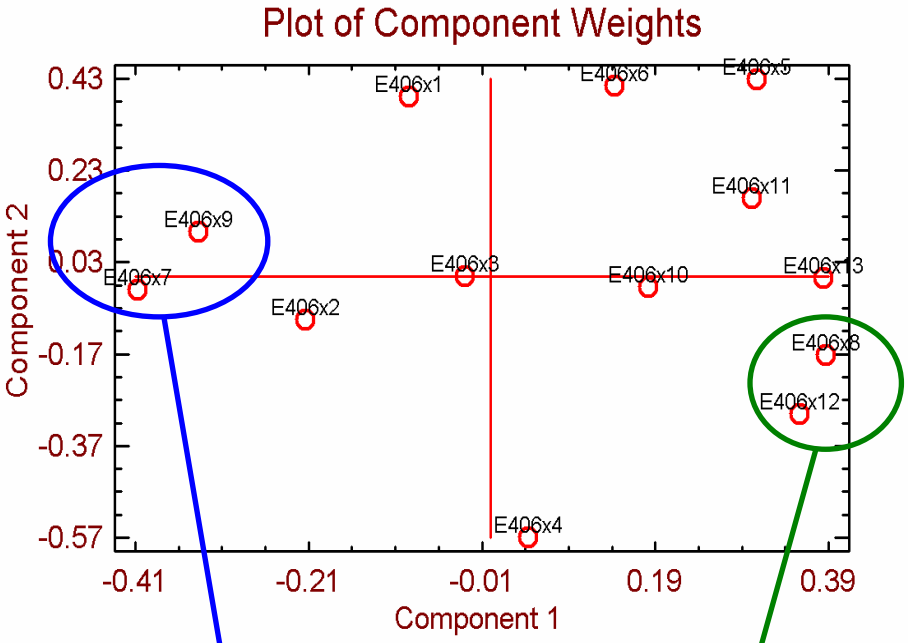


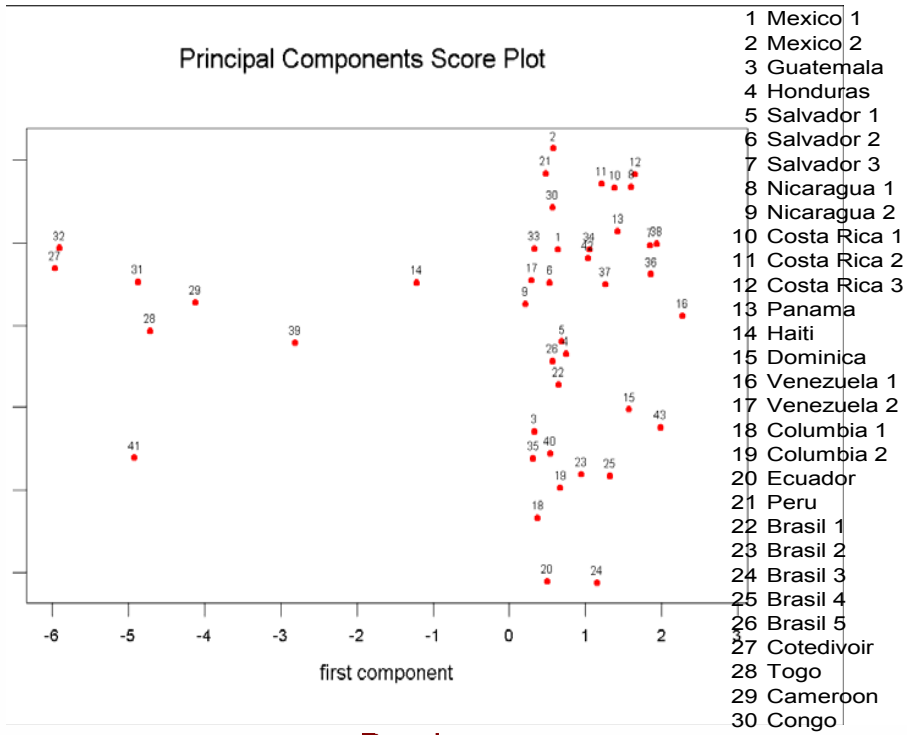
Dendrogram znaků matice dat *Kava* (STATISTICA).



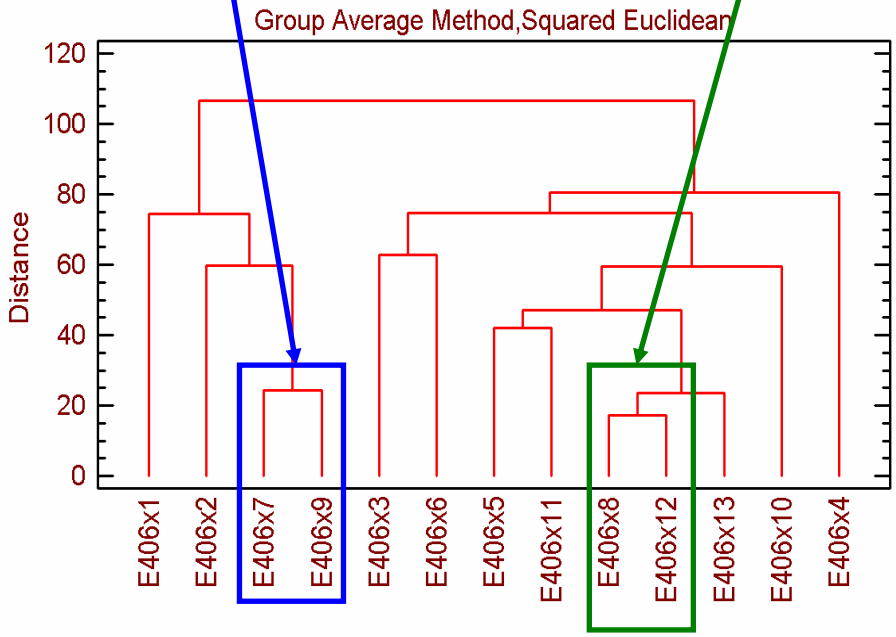
Dendrogram objektů matice dat *Kava*, (STATISTICA).

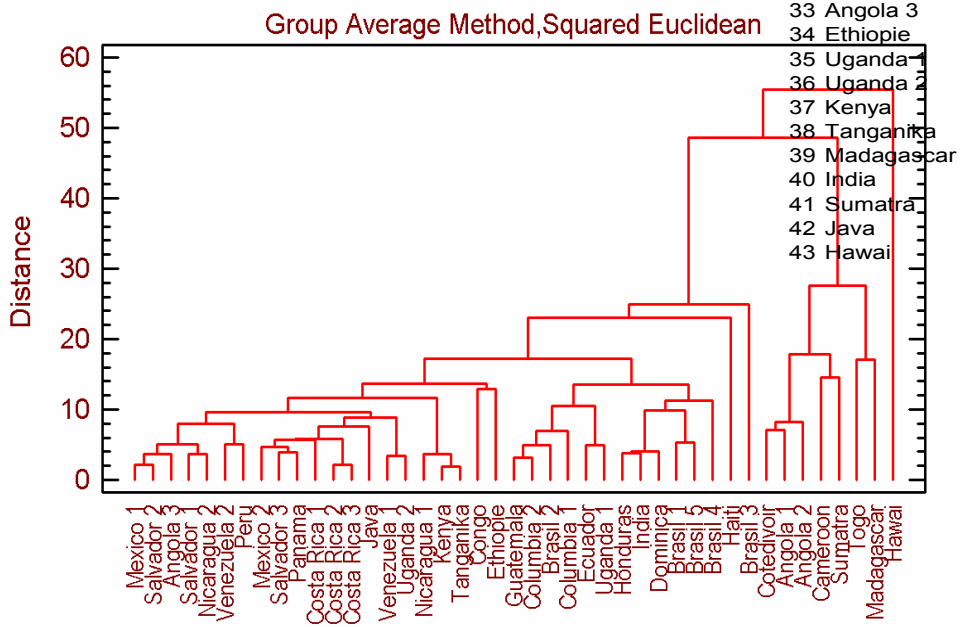
○ ***Závěr:*** Dendrogram znaků ukazuje shluky podobných vlastností kávy, zatímco dendrogram objektů klasifikuje podobné druhy kávy do shluků.

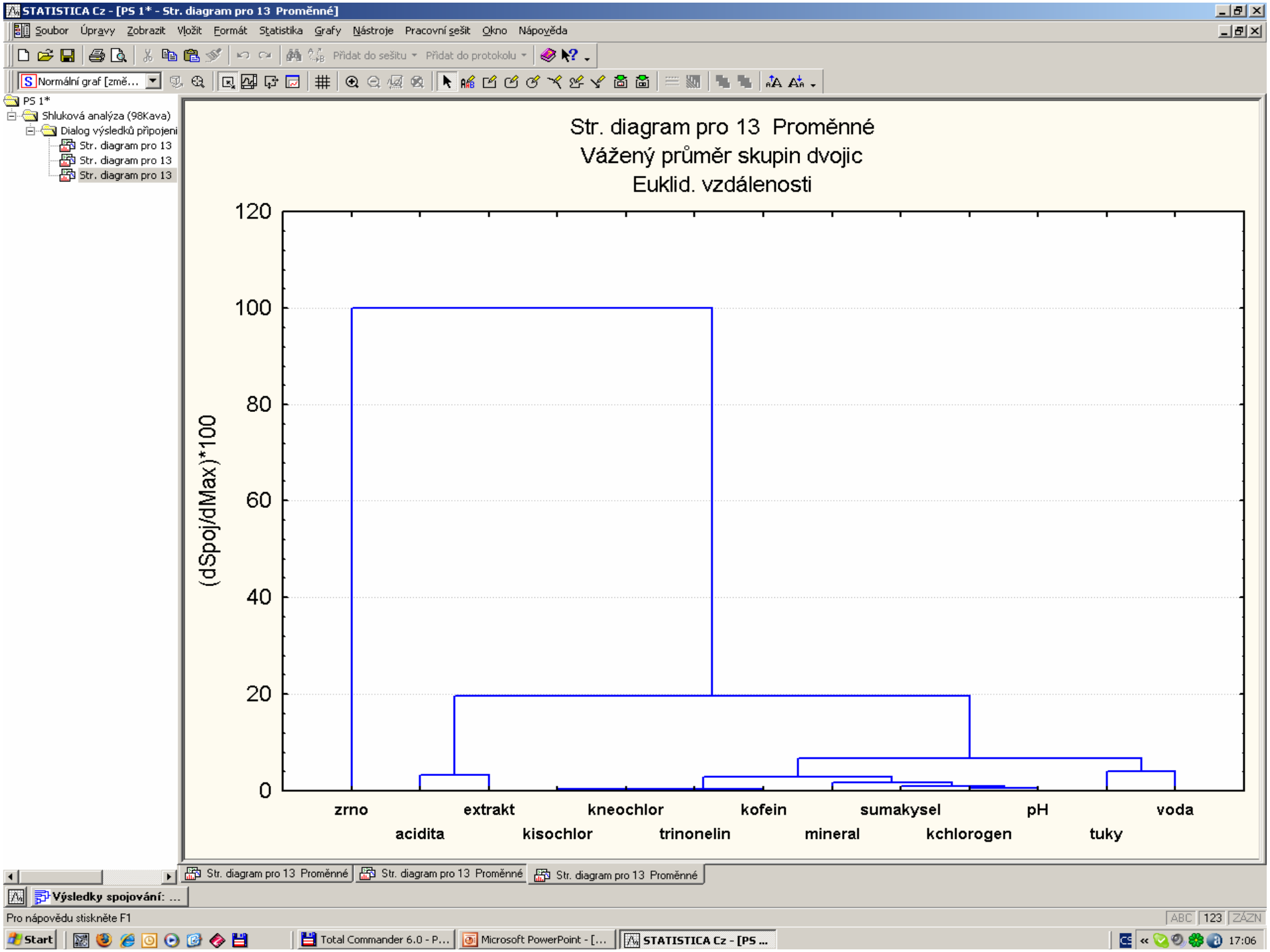


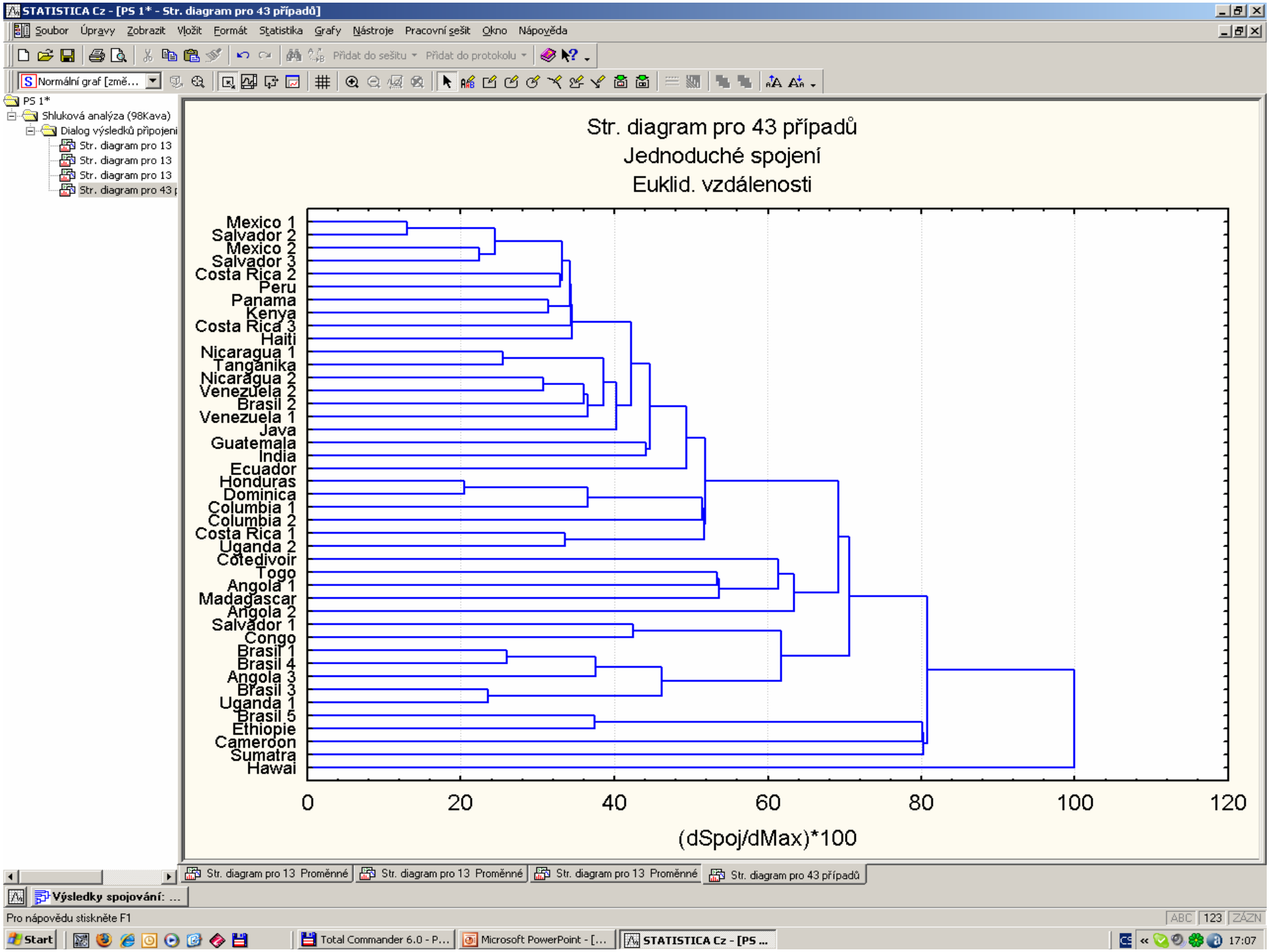


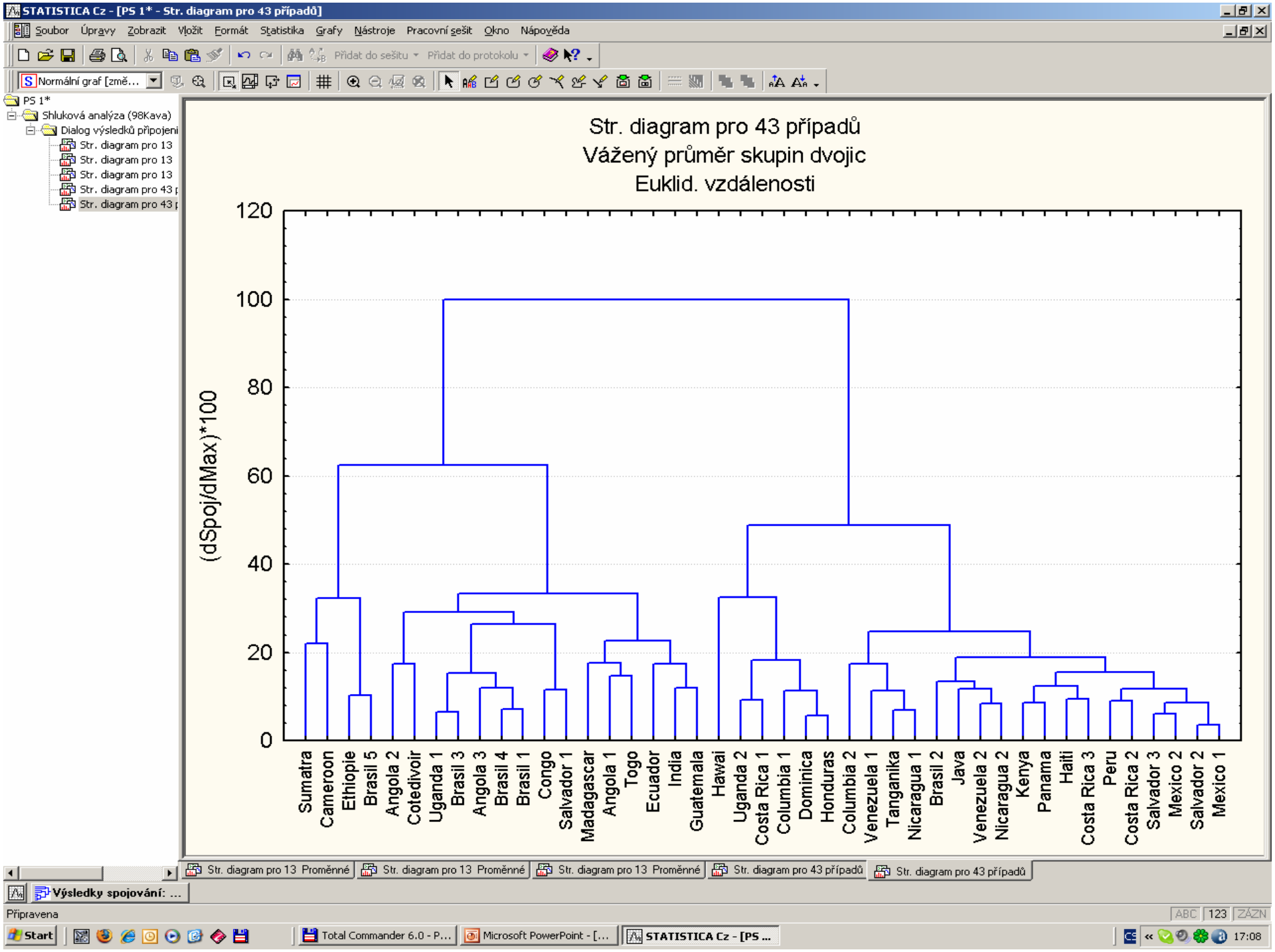
1 Mexico 1
2 Mexico 2
3 Guatemala
4 Honduras
5 Salvador 1
6 Salvador 2
7 Salvador 3
8 Nicaragua 1
9 Nicaragua 2
10 Costa Rica 1
11 Costa Rica 2
12 Costa Rica 3
13 Panama
14 Haiti
15 Dominica
16 Venezuela 1
17 Venezuela 2
18 Columbia 1
19 Columbia 2
20 Ecuador
21 Peru
22 Brasil 1
23 Brasil 2
24 Brasil 3
25 Brasil 4
26 Brasil 5
27 Cotedivoir
28 Togo
29 Cameroon
30 Congo
31 Angola 1
32 Angola 2
33 Angola 3
34 Ethiopie
35 Uganda 1
36 Uganda 2
37 Kenya
38 Tanganika
39 Madagascar
40 India
41 Sumatra
42 Java
43 Hawai






STATISTICA Cz - [PS 1* - Str. diagram pro 13 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojeni
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13 Proměnné
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
zrno
acidita
extrakt
kisochlor
kneochlor
trinonelin
kofein
mineral
sumakysel
kchlorogen
pH
tuky
voda
Str. diagram pro 13 Proměnné
Str. diagram pro 13 Proměnné
Str. diagram pro 13 Proměnné
Výsledky spojování: ...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:06


STATISTICA Cz - [PS 1* - Str. diagram pro 43 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43 p
Str. diagram pro 43 případů
Jednoduché spojení
Euklid. vzdálenosti
Mexico 1
Salvador 2
Mexico 2
Salvador 3
Costa Rica 2
Peru
Panama
Kenya
Costa Rica 3
Haiti
Nicaragua 1
Tanganika
Nicaragua 2
Venezuela 2
Brasil 2
Venezuela 1
Java
Guatemala
India
Ecuador
Honduras
Dominica
Columbia 1
Columbia 2
Costa Rica 1
Uganda 2
Cotedivoir
Togo
Angola 1
Madagascar
Angola 2
Salvador 1
Congo
Brasil 1
Brasil 4
Angola 3
Brasil 3
Uganda 1
Brasil 5
Ethiopie
Cameroon
Sumatra
Hawai
0
20
40
60
80
100
120
(dSpoj/dMax)*100
Str. diagram pro 13 Proměnné
Str. diagram pro 13 Proměnné
Str. diagram pro 13 Proměnné
Str. diagram pro 43 případů
Výsledky spojování: ...
Pro nápovědu stiskněte F1
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:07


STATISTICA Cz - [PS 1* - Str. diagram pro 43 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43 případů
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
Sumatra
Cameroon
Ethiopie
Brasil 5
Angola 2
Cotedivoir
Uganda 1
Brasil 3
Angola 3
Brasil 4
Brasil 1
Congo
Salvador 1
Madagascar
Angola 1
Togo
Ecuador
India
Guatemala
Hawai
Uganda 2
Costa Rica 1
Columbia 1
Dominica
Honduras
Columbia 2
Venezuela 1
Tanganika
Nicaragua 1
Brasil 2
Java
Venezuela 2
Nicaragua 2
Kenya
Panama
Haiti
Costa Rica 3
Peru
Costa Rica 2
Salvador 3
Mexico 2
Salvador 2
Mexico 1
Str. diagram pro 13 Proměnné
Str. diagram pro 13 Proměnné
Str. diagram pro 13 Proměnné
Str. diagram pro 43 případů
Str. diagram pro 43 případů
Výsledky spojování: ...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:08

PŘÍKLAD 9.14 *Klasifikace vzorků italských vín*

Pro 90 vzorků italských vín bylo naměřeno 8 fyzikálně-chemických vlastností. Ve vínech jsou obsaženy tři kultury, a to Nebbiolo ve víně Barolo, Grignolino a Barbera ve vínech stejného jména, a to každá ve 30 vzorcích. Kolik faktorů rozliší tři kategorie vín? Do kolika shluků lze vína roztřídit? Souvisí počet shluků se zadanými druhy vín?

○ ***Data:*** Soubor dat ***Vina*** je popsán:

i značí index vzorku vína,

Objekt značí jméno vzorku vína,

Kateg značí kategorie vzorku vína a 90 druhů vín v řádcích se týká tří kategorií 1. Barolo, 2. Grignolino a 3. Barbera, popsaných 8 následujícími vlastnostmi čili znaky ve sloupcích:

Alkohol značí obsah alkoholu x_1,

Necuk značí necukerný extrakt x_2,

Fosfaty značí obsah fosfátů x_3,

Fenoly značí obsah celkových fenolů x_4,

Flavan značí obsah flavanoidů x_5,

PomerA1 značí naměřený poměr absorbancí při 280 a 315 nm pro naředěné víno x_6,

PomerA2 značí naměřený poměr absorbancí při 280 a 315 nm pro určení flavanoidů x_7,

Prolin značí obsah prolinu x_8.

i	*Objekt*	*Kateg*	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	Olo0171	1	14.23	24.82	320	2.80	3.06	3.92	4.77	1065
..	..	..	..	..	..	..	..	..	..	..
90	Era2878	3	13.17	23.45	534	1.65	0.68	1.62	2.05	840

Pro 38 vzorků vín bylo nalezeno 24 analytických obsahů stopových prvků a charakteristických fyzikálně-chemických vlastností. Utvořte shluky podobných vlastností a dále shluky podobných vín.

Index	Cd	Mo	Mn	Ni	Cu	Al	Ba	Cr	Sr	Pb	B	Mg	Si	Na	Ca	P	K	Arom	Clar	Body	Flavor	Oakn	Quality	Reg
1	0.005	0.044	1.51	0.122	0.83	0.982	0.387	0.029	1.23	0.561	2.63	128	17.3	66.8	80.5	150	1130	3.3	1	2.8	3.1	4.1	9.8	1
2	0.055	0.16	1.16	0.149	0.066	1.02	0.312	0.038	0.975	0.697	6.21	193	19.7	53.3	75	118	1010	4.4	1	4.9	3.5	3.9	12.6	1
3	0.056	0.146	1.1	0.088	0.643	1.29	0.308	0.035	1.14	0.73	3.05	127	15.8	35.4	91	161	1160	3.9	1	5.3	4.8	4.7	11.9	1
4	0.063	0.191	0.959	0.38	0.133	1.05	0.165	0.036	0.927	0.796	2.57	112	13.4	27.5	93.6	120	924	3.9	1	2.6	3.1	3.6	11.1	1
5	0.011	0.363	1.38	0.16	0.051	1.32	0.38	0.059	1.13	1.73	3.07	138	16.7	76.6	84.6	164	1090	5.6	1	5.1	5.5	5.1	13.3	1
6	0.05	0.106	1.25	0.114	0.055	1.27	0.275	0.019	1.05	0.491	6.56	172	18.7	15.7	112	137	1290	4.6	1	4.7	5	4.1	12.8	1
7	0.025	0.479	1.07	0.168	0.753	0.715	0.164	0.062	0.823	2.06	4.57	179	17.8	98.5	122	184	1170	4.8	1	4.8	4.8	3.3	12.8	1
8	0.024	0.234	0.906	0.466	0.102	0.811	0.271	0.044	0.963	1.09	3.18	145	14.3	10.5	91.9	187	1020	5.3	1	4.5	4.3	5.2	12	1
9	0.009	0.058	1.84	0.042	0.17	1.8	0.225	0.022	1.13	0.048	6.13	113	13	54.4	70.2	158	1240	4.3	1	4.3	3.9	2.9	13.6	3
10	0.033	0.074	1.28	0.098	0.053	1.35	0.329	0.03	1.07	0.552	3.3	140	16.3	70.5	74.7	159	1100	4.3	1	3.9	4.7	3.9	13.9	1
11	0.039	0.071	1.19	0.043	0.163	0.971	0.105	0.028	0.491	0.31	6.56	103	9.5	45.3	67.9	133	1090	5.1	1	4.3	4.5	3.6	14.4	3
12	0.045	0.147	2.76	0.071	0.074	0.483	0.301	0.087	2.14	0.546	3.5	199	9.2	80.4	66.3	212	1470	3.3	0.5	5.4	4.3	3.6	12.3	2
13	0.06	0.116	1.15	0.055	0.18	0.912	0.166	0.041	0.578	0.518	6.43	111	11.1	59.7	83.8	139	1120	5.9	0.8	5.7	7	4.1	16.1	3
14	0.067	0.166	1.53	0.041	0.043	0.512	0.132	0.026	0.229	0.699	7.27	107	6	55.2	44.9	148	854	7.7	0.7	6.6	6.7	3.7	16.1	3
15	0.077	0.261	1.65	0.073	0.285	0.596	0.078	0.063	0.156	1.02	5.04	94.6	6.3	10.4	54.9	132	899	7.1	1	4.4	5.8	4.1	15.5	3
16	0.064	0.191	1.78	0.067	0.552	0.633	0.085	0.063	0.192	0.777	5.56	110	7	13.6	64.1	167	976	5.5	0.9	5.6	5.6	4.4	15.5	3
17	0.025	0.009	1.57	0.041	0.081	0.655	0.072	0.021	0.172	0.232	3.79	75.9	6.4	11.6	48.1	132	995	6.3	1	5.4	4.8	4.6	13.8	3
18	0.02	0.027	1.74	0.046	0.153	1.15	0.094	0.021	0.358	0.025	4.24	80.9	7.9	38.9	57.6	136	876	5	1	5.5	5.5	4.1	13.8	3
19	0.034	0.05	1.15	0.058	0.058	1.35	0.294	0.006	1.12	0.206	2.71	120	14.7	68.1	64.8	133	1050	4.6	1	4.1	4.3	3.1	11.3	1
20	0.013	0.03	2.82	0.058	0.05	0.623	0.349	0.082	2.91	0.171	3.54	208	9.3	79.2	66.4	266	1430	3.4	0.9	5	3.4	3.4	7.9	2
21	0.043	0.268	2.32	0.066	0.314	0.627	0.099	0.045	0.36	1.28	5.68	98.4	9.1	19.5	64.3	176	945	6.4	0.9	5.4	6.6	4.8	15.1	3
22	0.061	0.245	1.61	0.07	0.172	2.07	0.071	0.053	0.186	1.19	4.42	87.6	7.6	11.6	70.6	156	820	5.5	1	5.3	5.3	3.8	13.5	3
23	0.047	0.161	1.47	0.154	0.082	0.546	0.181	0.06	0.898	0.747	8.11	160	19.3	12.5	82.1	218	1220	4.7	0.7	4.1	5	3.7	10.8	2
24	0.048	0.146	1.85	0.092	0.09	0.889	0.328	0.1	1.32	0.604	6.42	134	19.3	125	83.2	173	1810	4.1	0.7	4	4.1	4	9.5	2
25	0.049	0.155	1.73	0.051	0.158	0.653	0.081	0.037	0.164	0.767	4.91	86.5	6.5	11.5	53.9	172	1020	6	1	5.4	5.7	4.7	12.7	3
26	0.042	0.126	1.7	0.112	0.21	0.508	0.299	0.054	0.995	0.686	6.94	129	43.6	45	85.9	165	1330	4.3	1	4.6	4.7	4.9	11.6	2
27	0.058	0.184	1.28	0.095	0.058	1.3	0.346	0.037	1.17	1.28	3.29	145	16.7	65.8	72.8	175	1140	3.9	1	4	5.1	5.1	11.7	1
28	0.065	0.211	1.65	0.102	0.055	0.308	0.206	0.028	0.72	1.02	6.12	99.3	27.1	20.5	95.2	194	1260	5.1	1	4.9	5	5.1	11.9	2
29	0.065	0.129	1.56	0.166	0.151	0.373	0.281	0.034	0.889	0.638	7.28	139	22.2	13.3	84.2	164	1200	3.9	1	4.4	5	4.4	10.8	2
30	0.068	0.166	3.14	0.104	0.053	0.368	0.292	0.039	1.11	0.831	4.71	125	17.6	13.9	59.5	141	1030	4.5	1	3.7	2.9	3.9	8.5	2
31	0.067	0.199	1.65	0.119	0.163	0.447	0.292	0.058	0.927	1.02	6.97	131	38.3	42.9	85.9	164	1390	5.2	1	4.3	5	6	10.7	2
32	0.084	0.266	1.28	0.087	0.071	1.14	0.158	0.049	0.794	1.3	3.77	143	19.7	39.1	128	146	1230	4.2	0.8	3.8	3	4.7	9.1	1
33	0.069	0.183	1.94	0.07	0.095	0.465	0.225	0.037	1.19	0.915	2	123	4.6	7.5	69.4	123	943	3.3	1	3.5	4.3	4.5	12.1	1
34	0.087	0.208	1.76	0.061	0.099	0.683	0.087	0.042	0.168	1.33	5.04	92.9	7	12	56.3	157	949	6.8	1	5	6	5.2	14.9	3
35	0.074	0.142	2.44	0.051	0.052	0.737	0.408	0.022	1.16	0.745	3.94	143	6.8	36.8	67.6	82	1170	5	0.8	5.7	5.5	4.8	13.5	1
36	0.084	0.171	1.85	0.088	0.038	1.21	0.263	0.072	1.35	0.899	2.38	130	6.2	101	64.4	99	1070	3.5	0.8	4.7	4.2	3.3	12.2	1
37	0.106	0.307	1.15	0.063	0.051	0.643	0.29	0.031	0.885	1.61	4.4	151	17.4	7.3	103	177	1100	4.3	0.8	5.5	3.5	5.8	10.3	1
38	0.102	0.342	4.08	0.065	0.077	0.752	0.366	0.048	1.08	1.77	3.37	145	5.3	33.1	58.3	117	1010	5.2	0.8	4.8	5.7	3.5	13.2	1

○ ***Řešení:*** **Graf komponentních vah znaků** odhaluje především korelaci znaků. Blízké průvodiče znaků s malým úhlem indikují silnou pozitivní korelaci znaků.

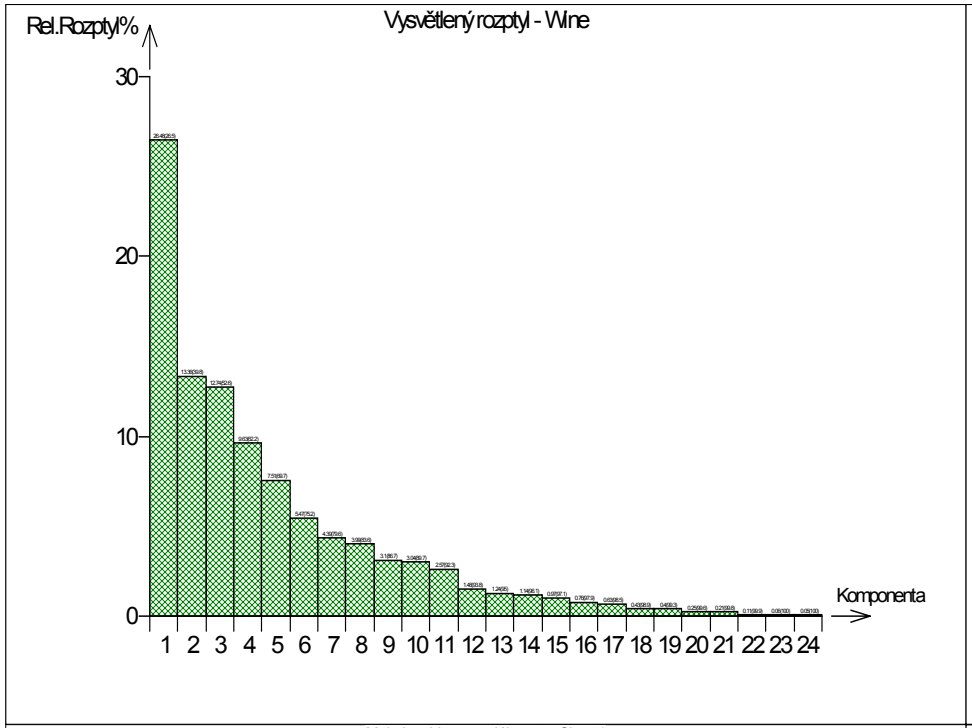
Dendrogram znaků ukazuje na první shluk podobných znaků *Fenoly, Flavan, PomerA1, PomerA2, Alkohol, Necuk* a také *Kateg*.

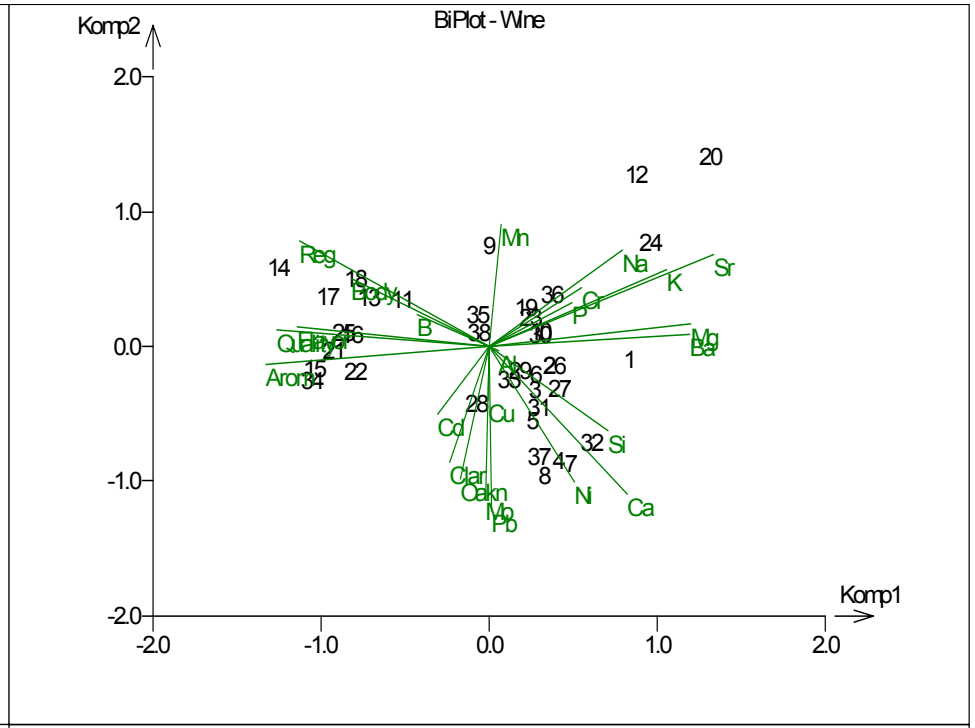
K tomuto shluku se pojí již podstatně méně podobný znak *Fosfaty*.

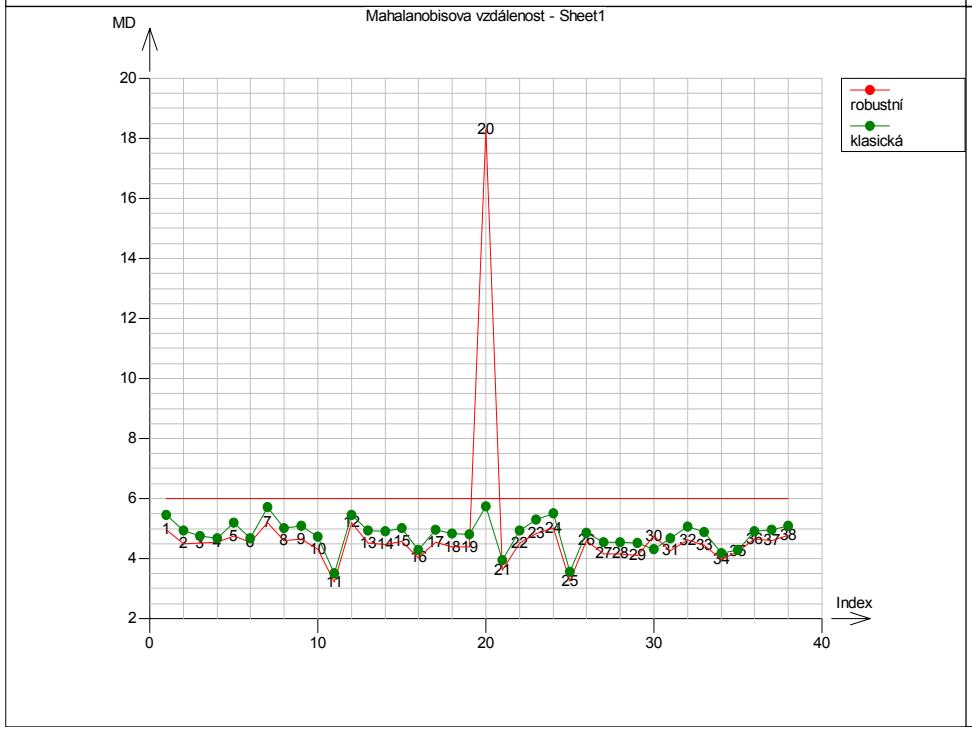
Znak *prolin* je zcela nepodobný ostatním a v dendrogramu je indikován jako odlehlý znak.

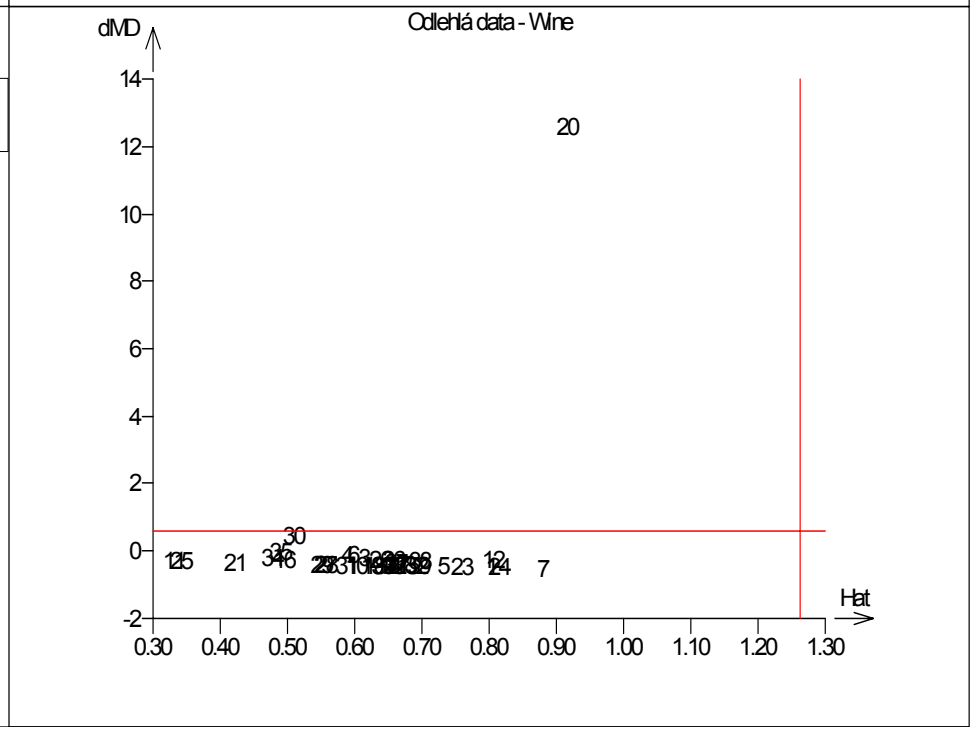
Graf komponentního skóre objektů vykazuje tři větší shluky vín ve shodě s jejich kategoriemi *Barolo* ve zkratce *Olo,* Barbera ve zkratce *Era* a konečně *Grignolino* ve zkratce *Gri.*

Dendrogram objektů rovněž ukazuje na tři shluky, zhora první shluk *Olo*, uprostřed grafu shluk *Era* a v dolní části grafu pak shluk *Gri*.

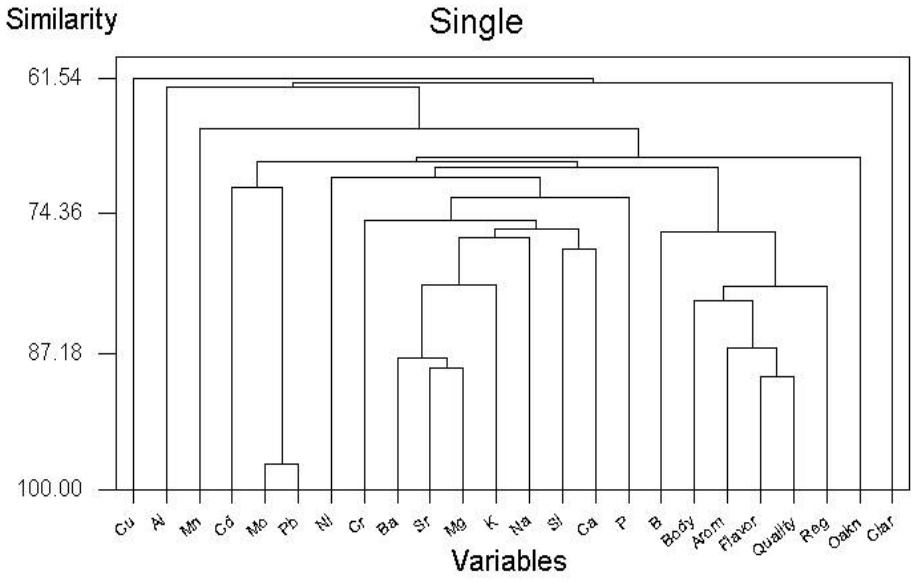

Vysvětlený rozptyl - Wine
Rel.Rozptyl%
Komponenta

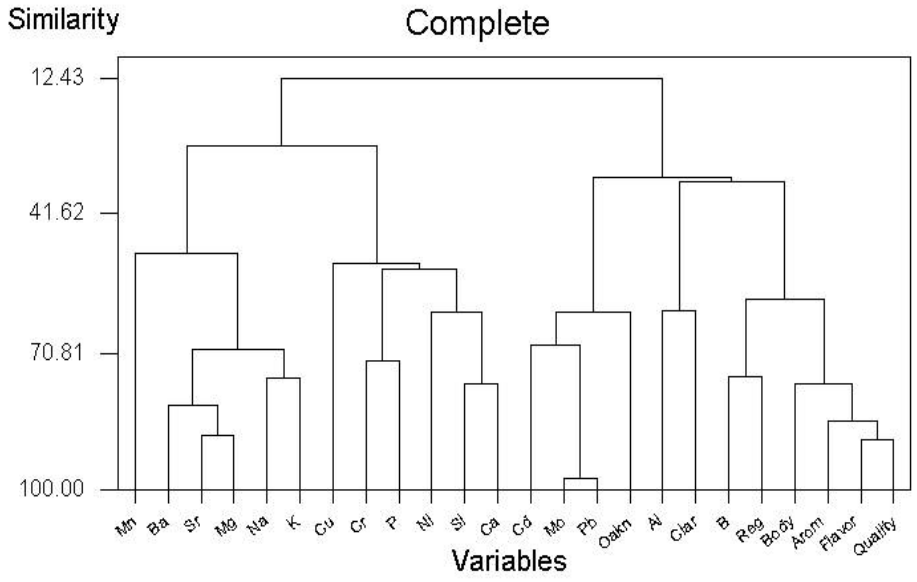

BiPlot - Wine
Komp2
Komp1
Mn
Na
Sr
K
Cl
P
Mg
Ba
Si
Ca
N
Pb
Cu
Cd
Reg
B
Aron

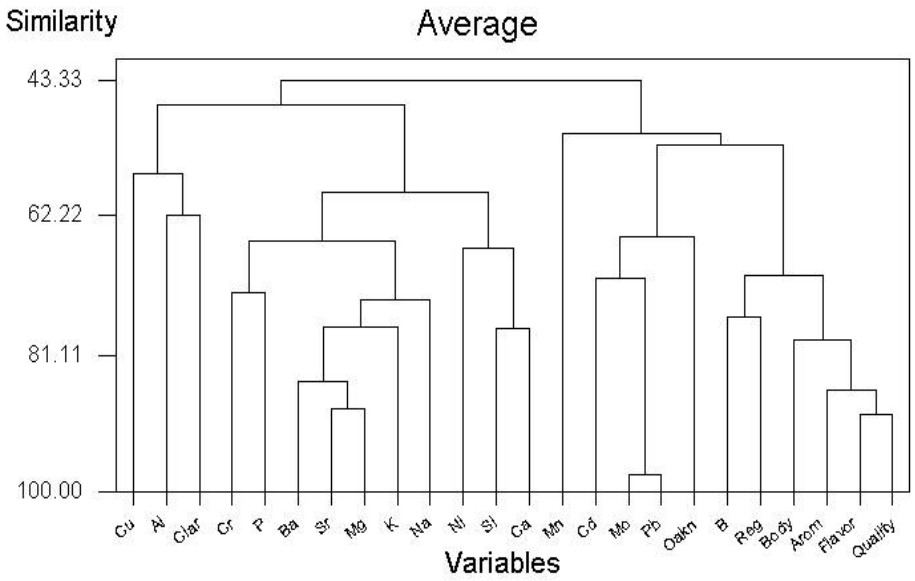

Mahalanobisova vzdálenost - Sheet1
MD
Index
robustní
klasická

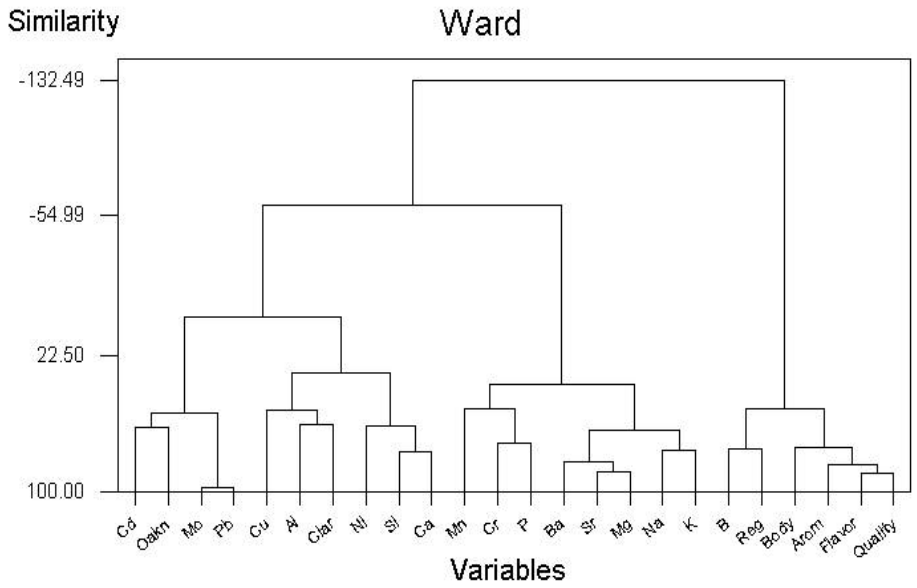

Odlehlá data - Wine
dMD
Hat

Hledání nejlepší metody dle kofenetického korelačního koeficientu a kritéria delta.

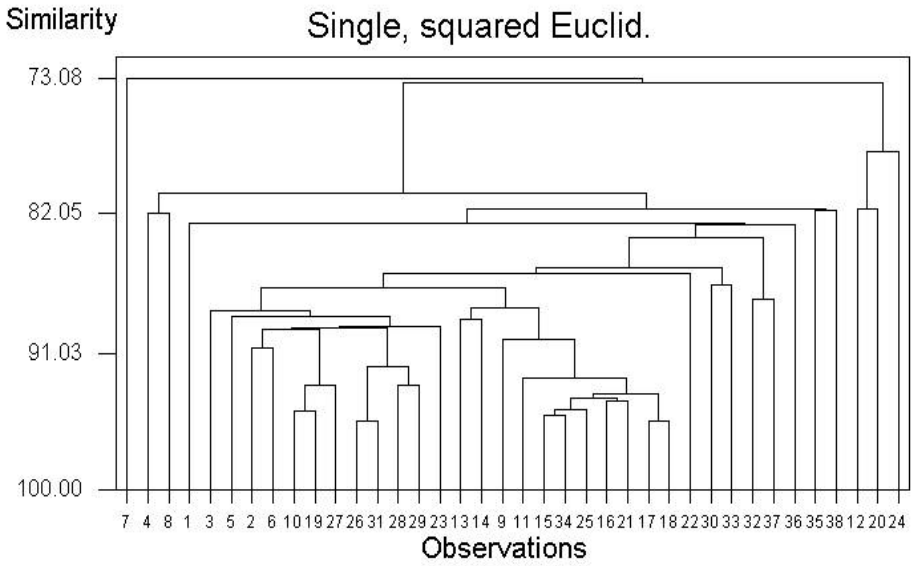


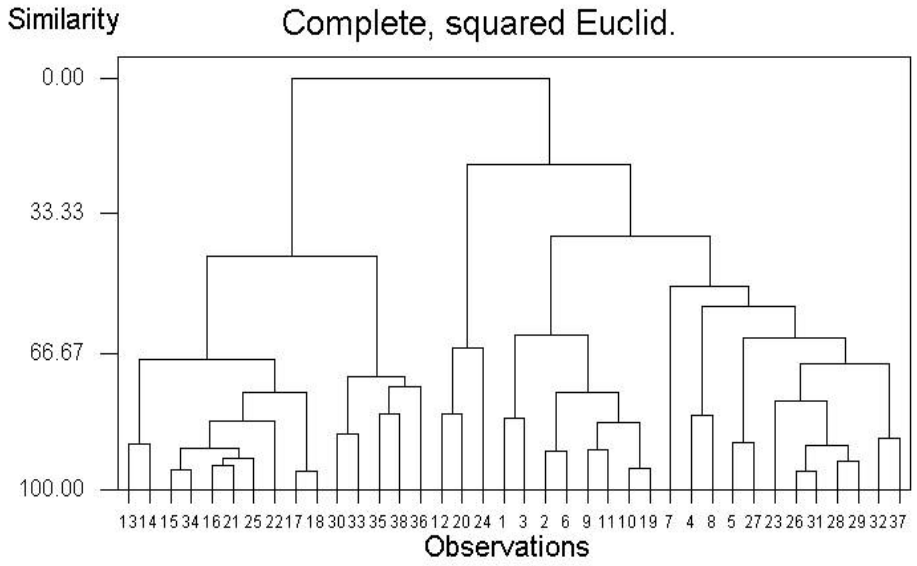


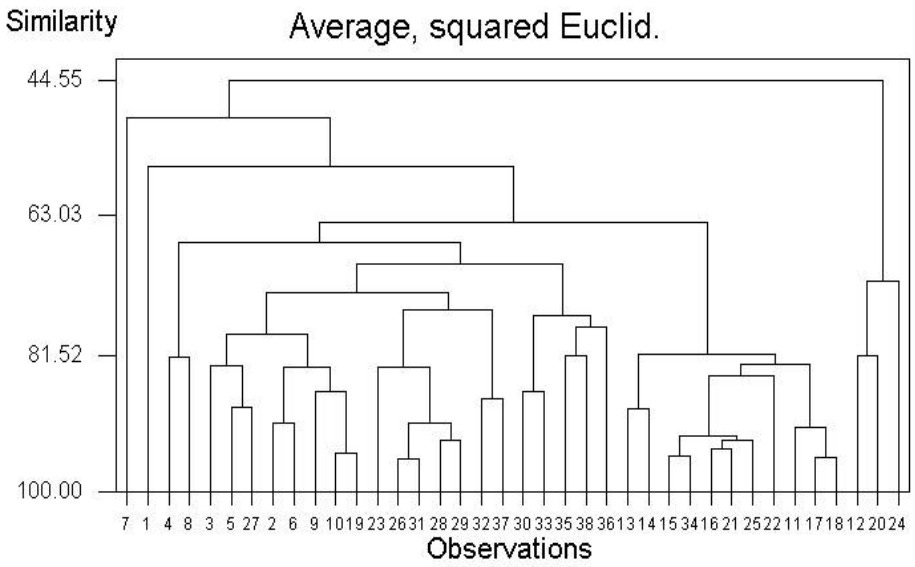


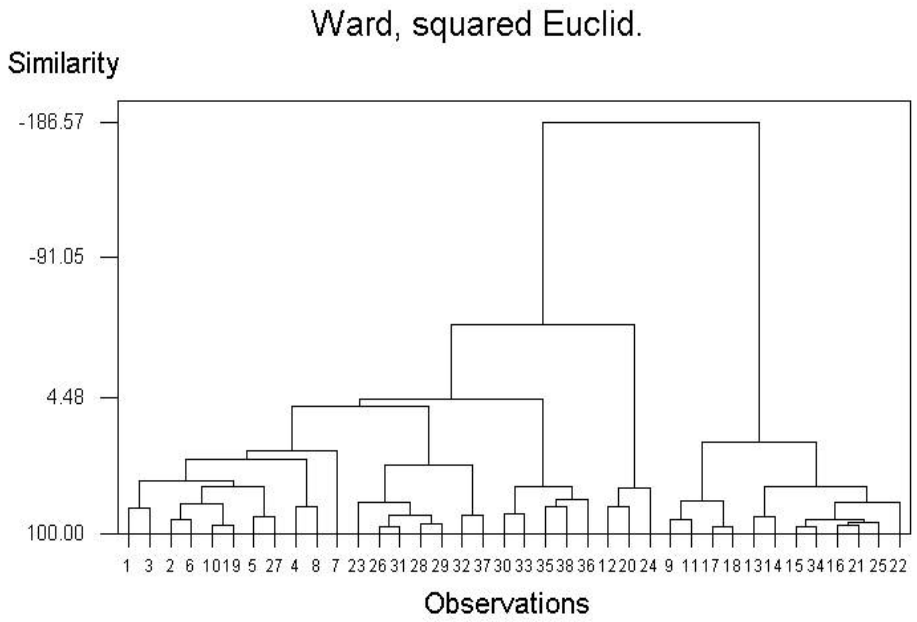


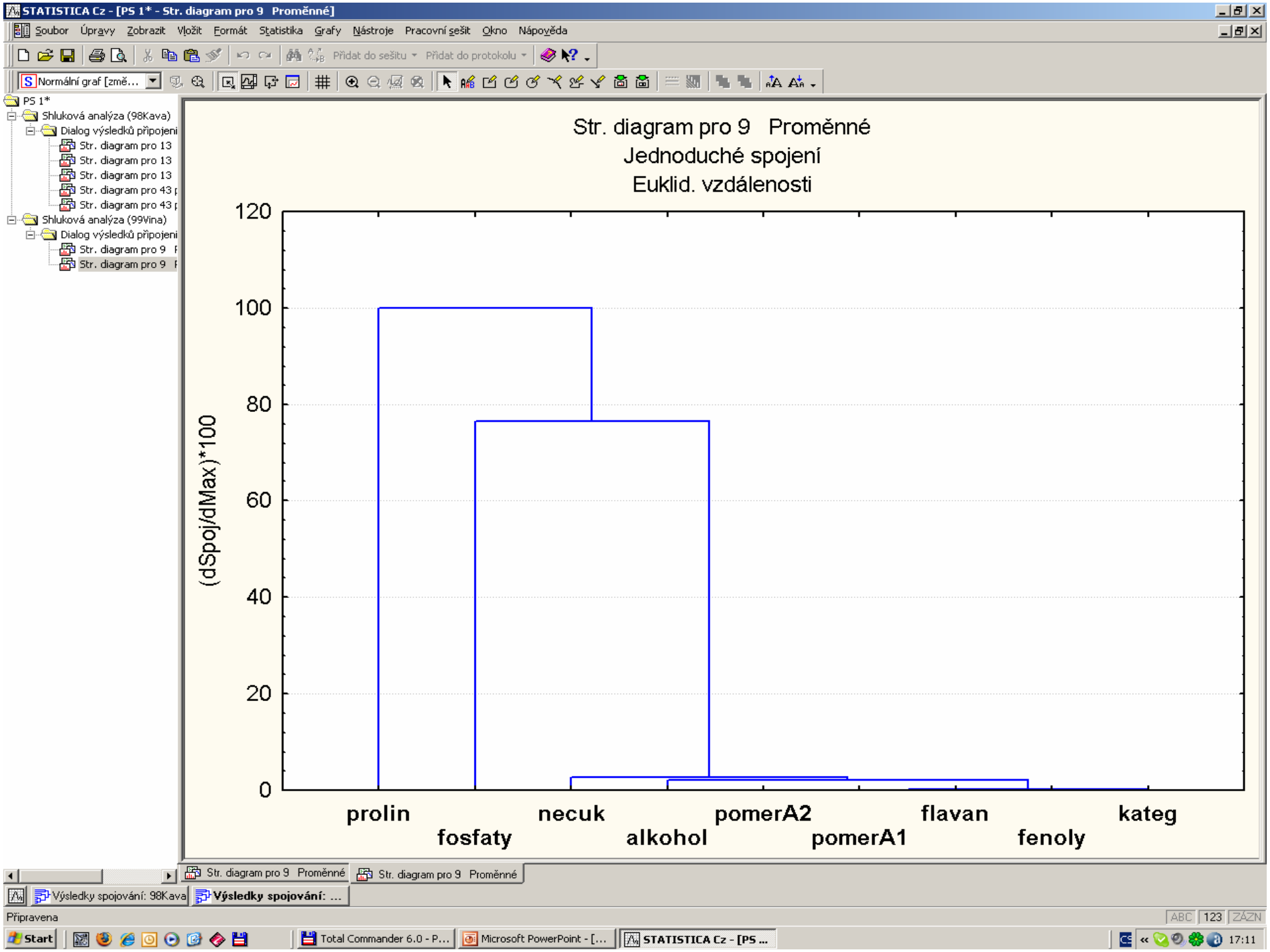
Hledání nejlepší metody dle kofenetického korelačního koeficientu a kritéria delta.

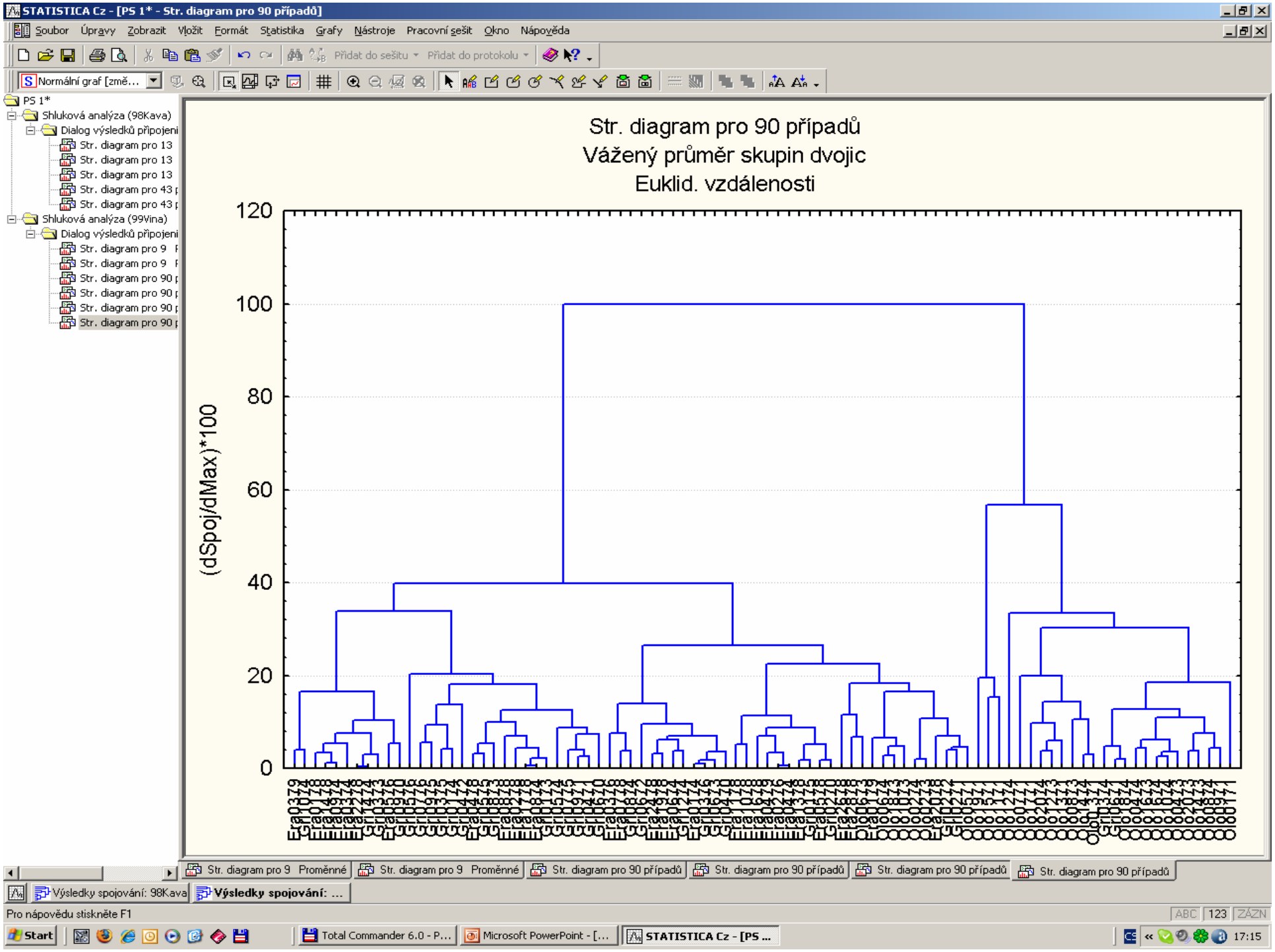


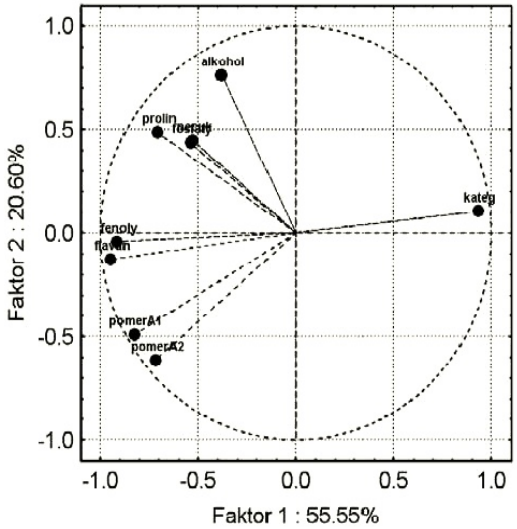




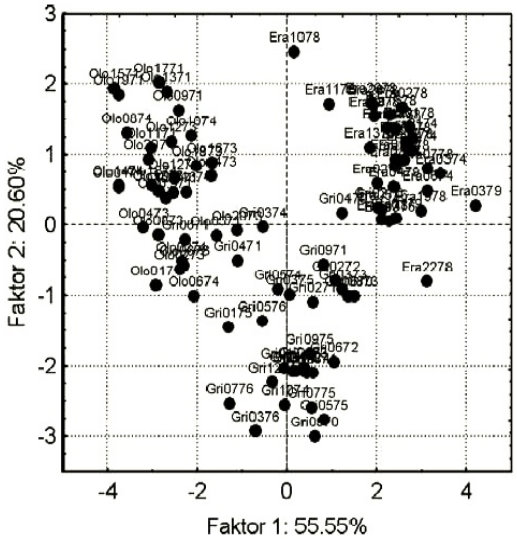



STATISTICA Cz - [PS 1* - Str. diagram pro 9 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 9 Proměnné
Jednoduché spojení
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
prolin
fosfaty
necuk
alkohol
pomerA2
pomerA1
flavan
fenoly
kateg
Str. diagram pro 9 Proměnné
Str. diagram pro 9 Proměnné
Výsledky spojování: 98Kava
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:11

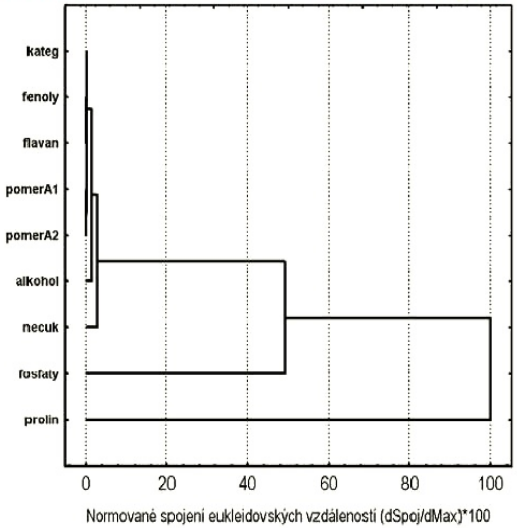

STATISTICA Cz - [PS 1* - Str. diagram pro 90 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojeni
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojeni
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90 případů
(dSpoj/dMax)*100
0
20
40
60
80
100
120
Str. diagram pro 9 Proměnné
Str. diagram pro 9 Proměnné
Str. diagram pro 90 případů
Str. diagram pro 90 případů
Str. diagram pro 90 případů
Str. diagram pro 90 případů
Výsledky spojování: 98Kava
Výsledky spojování: ...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:15



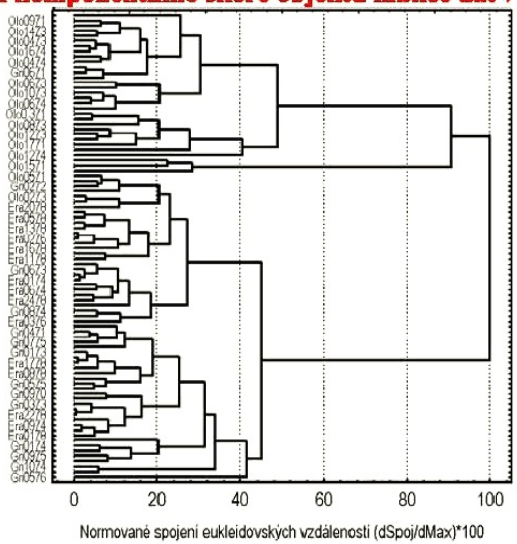
Graf komponentních vah znaků matice dat *Vina*, (STATISTICA).



Graf komponentního skóre objektů matice dat *Vina*.



Dendrogram znaků matice dat *Vina* (STATISTICA).



Dendrogram objektů matice dat *Vina*, (STATISTICA).

○ ***Závěr:*** Graf komponentního skóre objektů a dendrogram objektů shodně vykazují tři větší shluky vín ve shodě s jejich kategoriemi ***Olo, Era*** a ***Gri***.

PŘÍKLAD 4.5

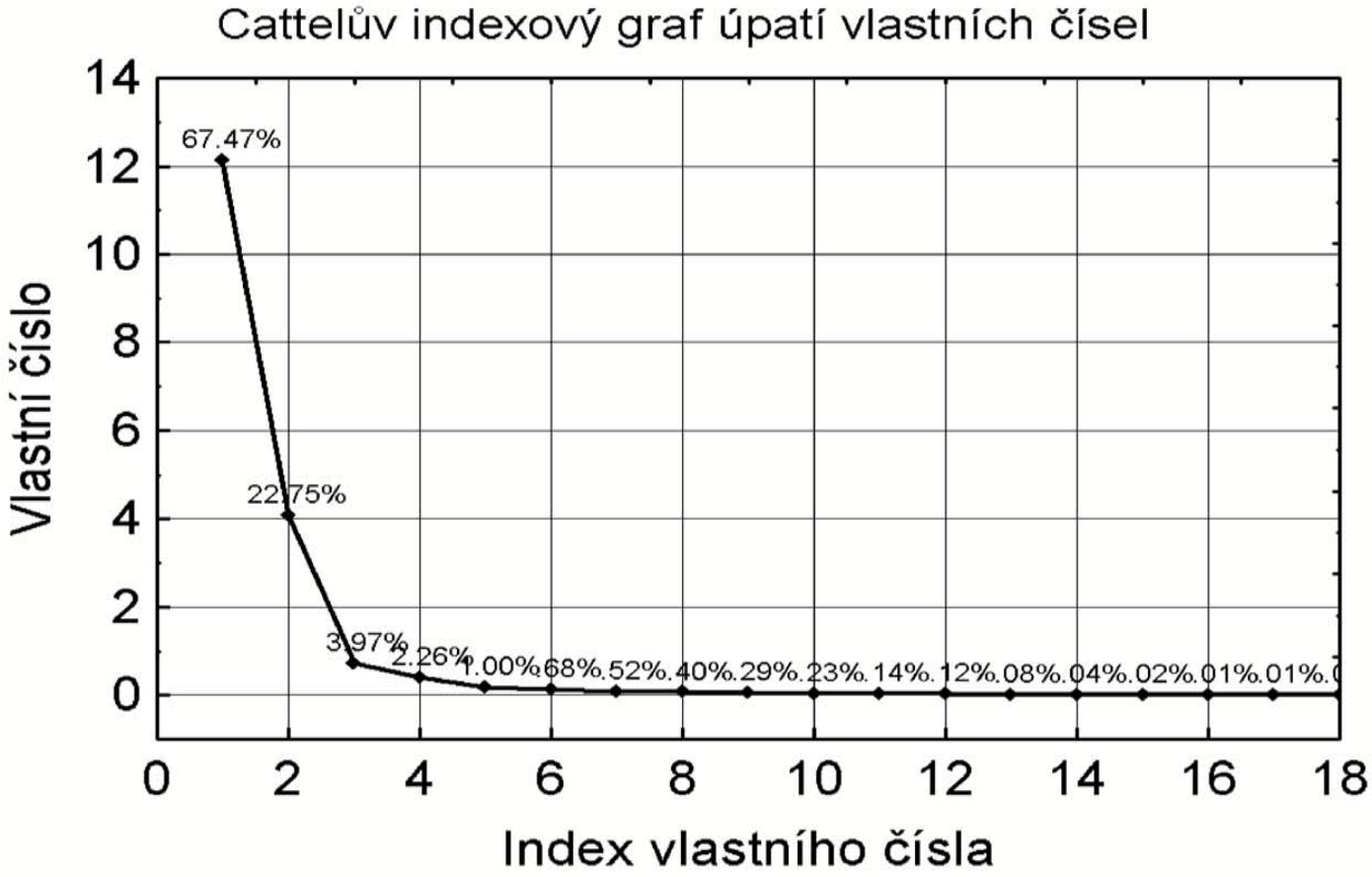
Chromatografická analýza farmakologických sloučenin

Byly měřeny hodnoty R_F pro 20 sloučenin s 18 eluenty. Žádné eluční činidlo však neprovedlo úplné rozdělení. Cílem je nalézt minimální výběr elučních činidel, které by daly dostatek informace pro kvalitativní analýzu.

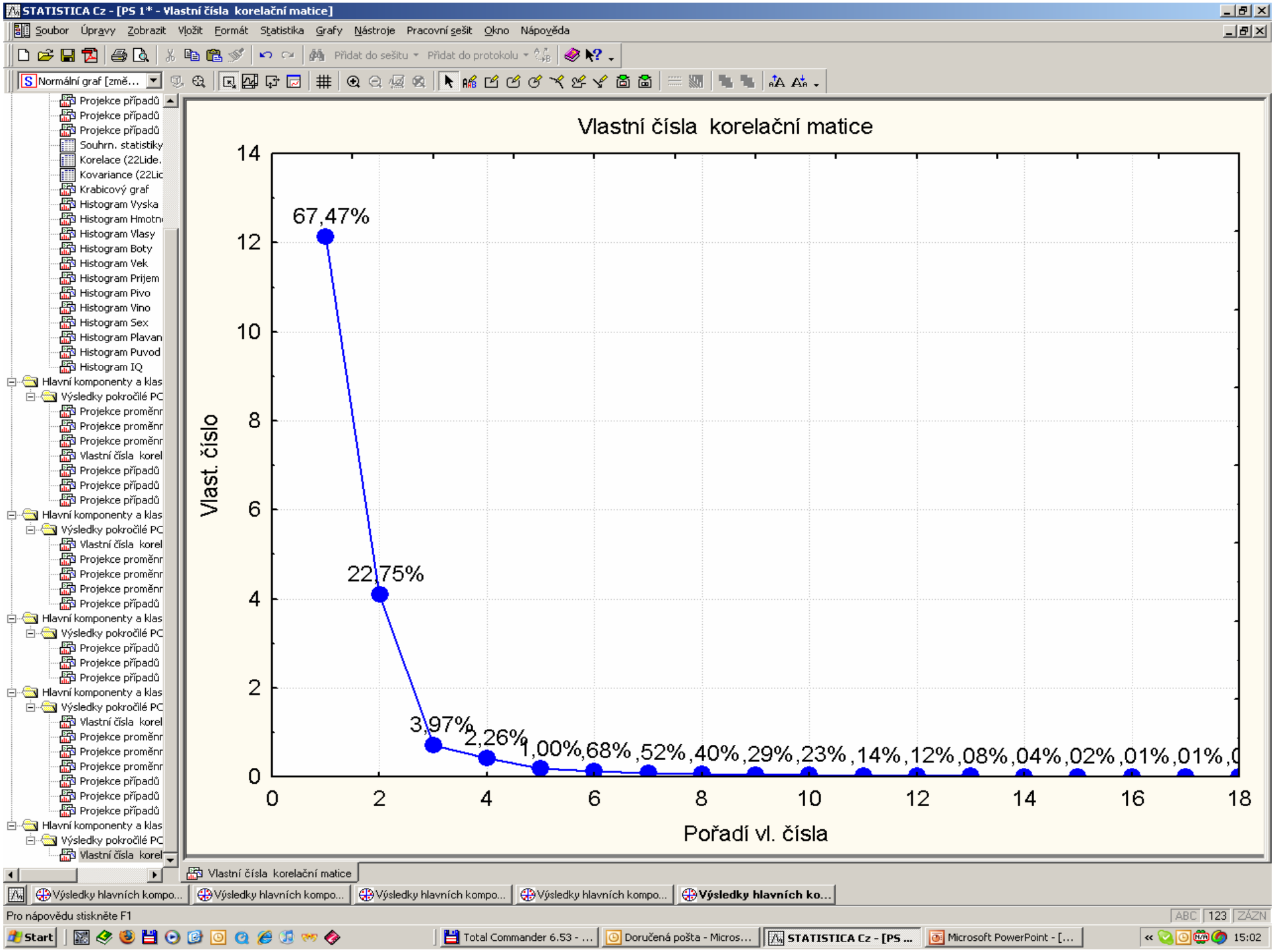
○ ***Data***: Datový soubor GIUSEPPE obsahuje 100 × R_F pro 20 sloučenin (v řádcích byla jména zkrácena na maximálně 8 písmen) a ve sloupcích je 18 elučních činidel představujících zde znaky: i vzorek, x_1 směs toluen : aceton : ethanol : 30 % amoniak = 45 : 45 : 7 : 3, x_2 směs ethylacetát : benzen : methanol : 30 % amoniak = 60 : 35 : 6.5 : 2.5, x_3 směs benzen : dioxan : ethanol : 30 % amoniak = 50 : 40 : 7.5 : 2.5, x_4 směs methanol : 30 % amoniak = 100 : 1.5, x_5 směs benzen : 2-propanol : methanol : 30 % amoniak = 70 : 30 : 20 : 5, x_6 směs ethylacetát : methanol : 30 % amoniak = 85 : 10 : 5, x_7 směs cyklohexan : toluen : diethylamin = 65 : 25 : 10, x_8 směs cyklohexan : toluen : diethylamin = 75 : 15 : 10, x_9 směs cyklohexan : benzen : metanol : diethylamin = 70 : 20 : 10 : 5, x_{10} směs chloroform : aceton : diethylamin = 50 : 40 : 10, x_{11} směs cyklohexan : chloroform : diethylamin = 50 : 40 : 10, x_{12} směs benzen : ethylacetát : diethylamin = 50 : 40 : 10, x_{13} směs xylen : methylethylketon : methanol : diethylamin = 40 : 40 : 6 : 2, x_{14} směs diethylether : diethylamin = 95 : 5, x_{15} směs ethylacetát : chloroform = 50 : 50, x_{16} směs ethylacetát : chloroform [A] = 50 : 50, x_{17} směs butanol : methanol = 40 : 60, x_{18} směs butanol : methanol [A] = 40 : 60, kde [A] značí, že byl užit 0.1M methanolát draselný.

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
Atropine	20	16	29	23	62	33	4	2	13	47	25	42	18	12	0	0	5	8
Biperide	91	90	87	68	92	87	73	72	64	85	81	86	68	94	11	40	40	65
Caffeine	55	42	52	68	77	54	8	5	13	60	30	51	41	20	13	12	54	57
Cocaine	81	82	81	71	87	82	46	41	38	81	72	80	52	72	6	24	30	57
Codeine	38	31	44	39	71	43	12	9	16	49	29	36	22	14	0	0	15	21
Cyclizin	71	72	80	64	85	80	49	47	40	75	71	73	39	59	2	9	34	54
Diazepam	76	79	80	78	85	80	28	21	29	80	61	75	72	54	54	50	85	87
Ketamine	77	79	80	76	86	79	71	33	32	81	66	76	67	66	27	37	66	79
Lignocaine	77	79	80	73	86	80	35	30	28	84	73	77	66	64	25	54	68	84
Lorazepam	47	34	53	77	73	46	2	0	12	52	7	22	47	8	28	15	85	79
Mebeveri	85	90	90	65	90	85	43	33	38	88	70	87	62	76	5	29	29	53
Methadon	85	84	88	48	89	83	63	64	48	85	73	86	38	86	1	10	13	29
Morphine	18	9	15	39	56	20	2	0	5	20	2	8	13	3	0	1	16	18
Naloxone	48	40	62	75	79	48	15	11	22	52	26	40	60	21	18	21	67	77
Papaverine	68	66	76	79	88	71	12	7	18	78	56	62	54	30	28	41	76	80
Pentazoc	72	66	81	65	87	76	22	18	26	69	41	54	39	44	2	11	32	59
Phenacet	64	58	62	79	87	66	4	1	13	68	18	41	58	24	41	40	86	84
Phenazon	66	53	70	83	86	65	30	22	24	77	60	66	45	54	15	21	68	74
Prazepam	81	83	86	83	88	81	41	31	35	83	66	82	75	74	65	67	86	88
Procaine	64	60	70	65	82	73	8	5	16	66	24	54	37	50	1	11	29	53

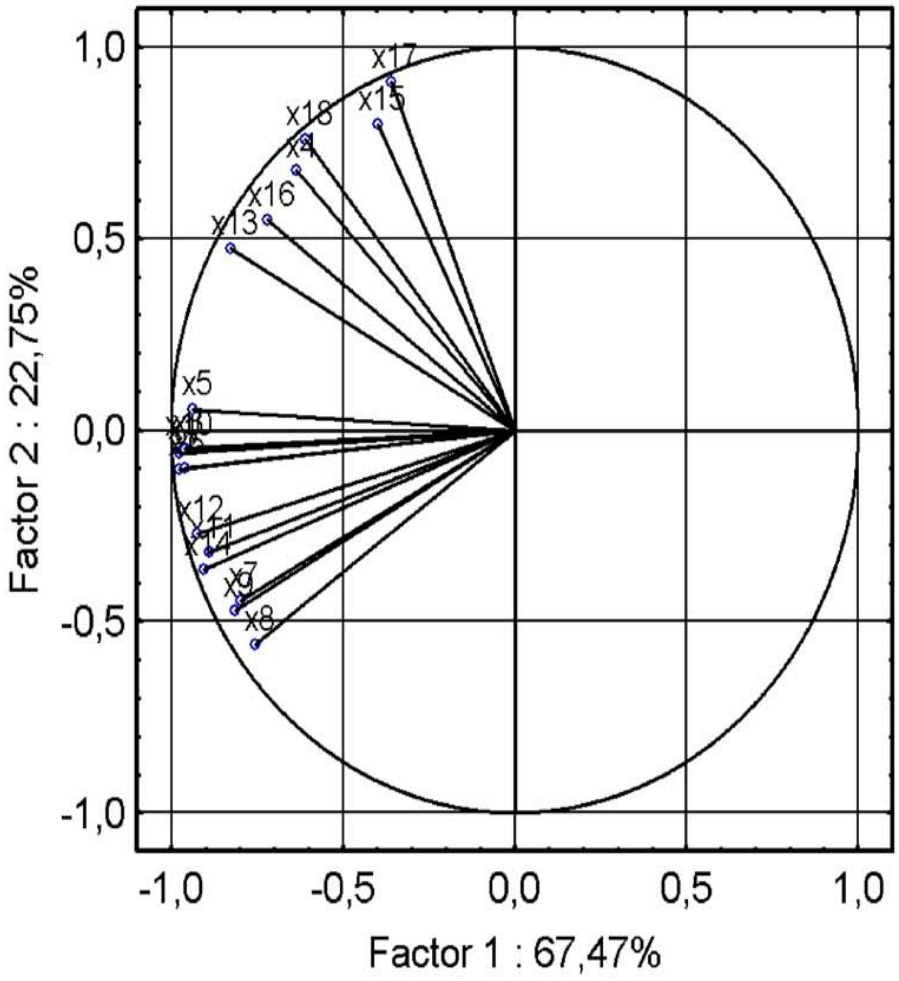
1. Cattelův indexový graf úpatí vlastních čísel: první dvě hlavní komponenty z 90% popisují data, lze snížit rozměrnost dat z 18 znaků na 2 proměnné *PC1* a *PC2*.



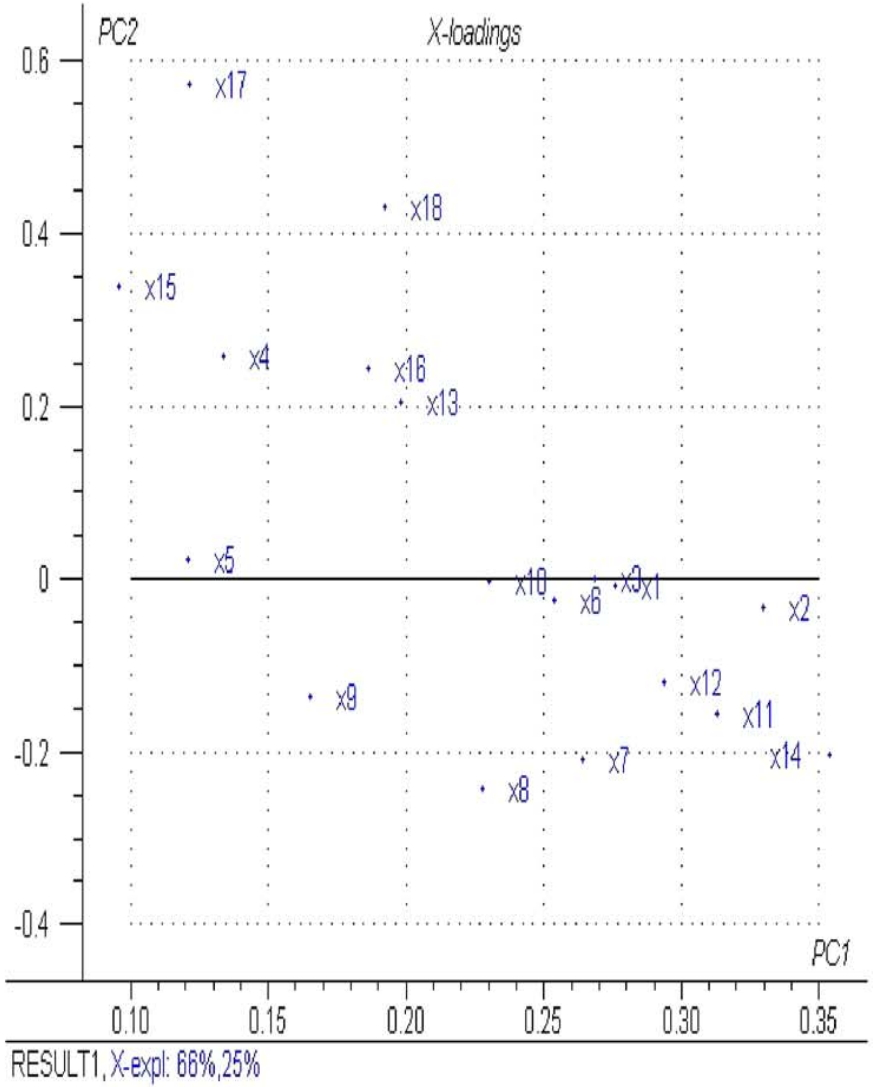
Obr. 4.19 Cattelův indexový graf úpatí vlastních čísel Scree Plot dat *Guiseppe* (STATISTICA).


STATISTICA Cz - [PS 1* - Vlastní čísla korelační matice]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Vlastní čísla korelační matice
Vlast. číslo
Pořadí vl. čísla
67,47%
22,75%
3,97%
2,26%
1,00%
,68%
,52%
,40%
,29%
,23%
,14%
,12%
,08%
,04%
,02%
,01%
,01%
Pro nápovědu stiskněte F1

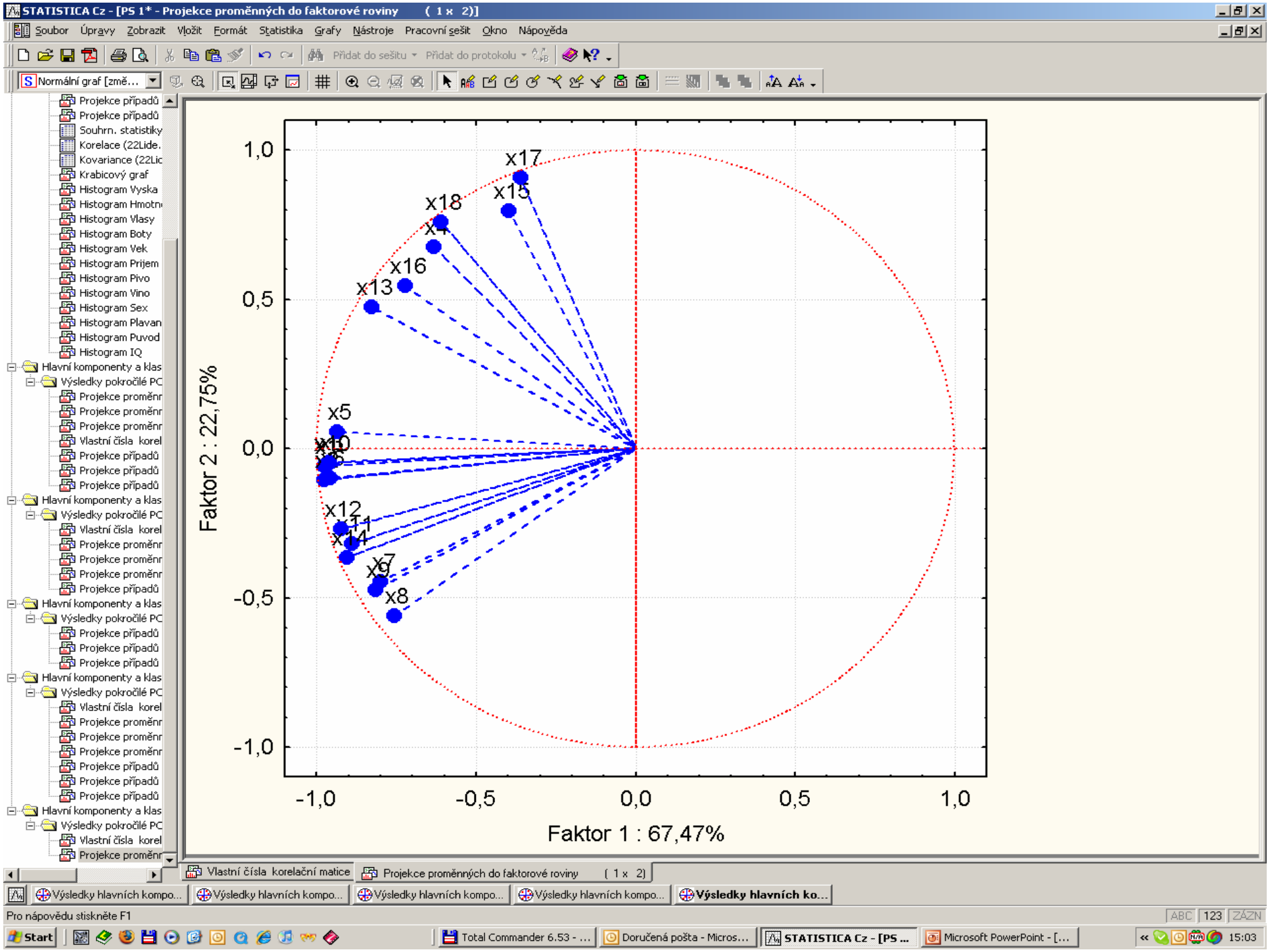
2. Graf komponentních vah: které směsi elučních činidel jsou si podobné a spolu korelují. Existují eluční činidla, která se silně odlišují, která spolu nekorelují.

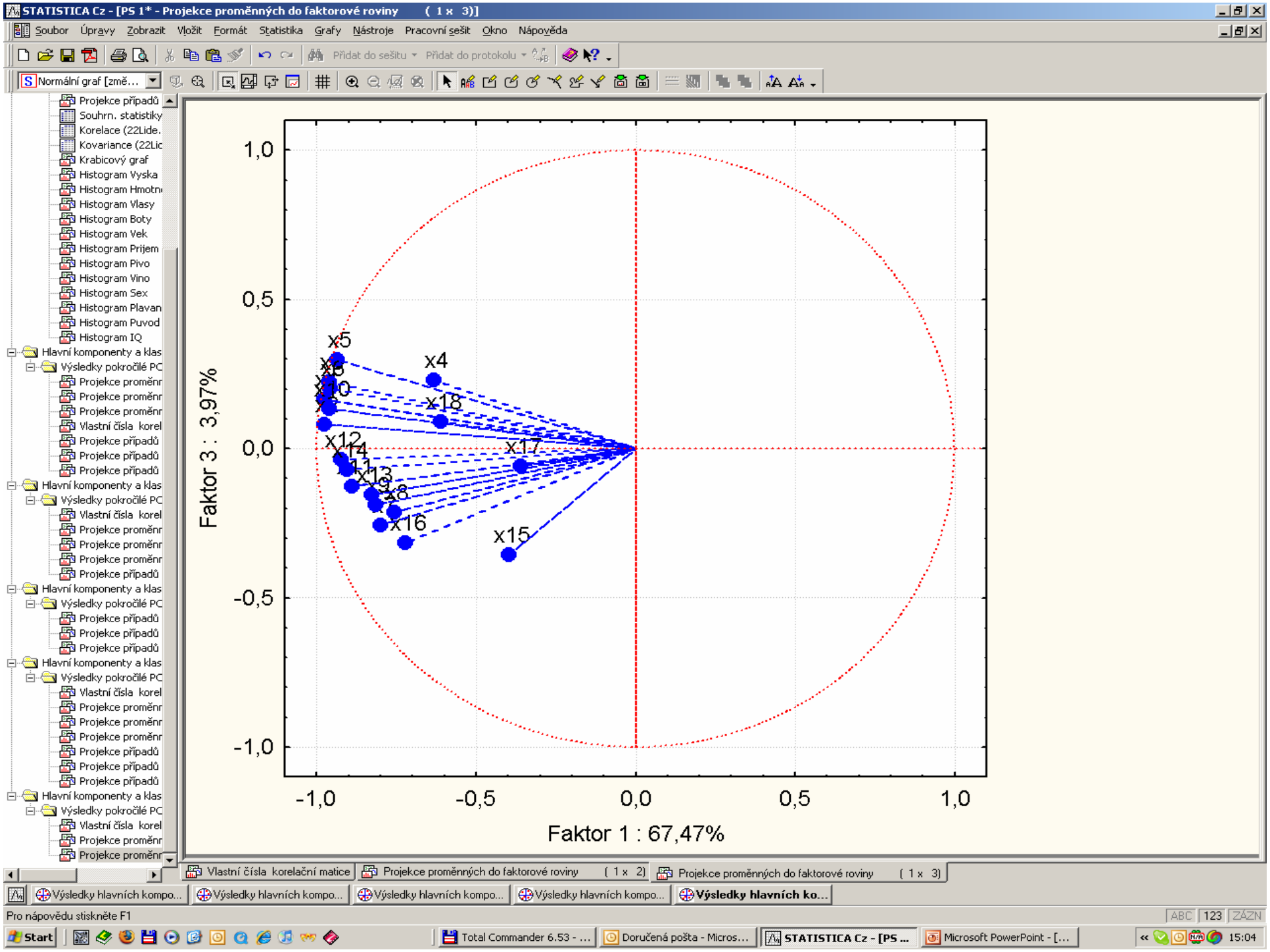


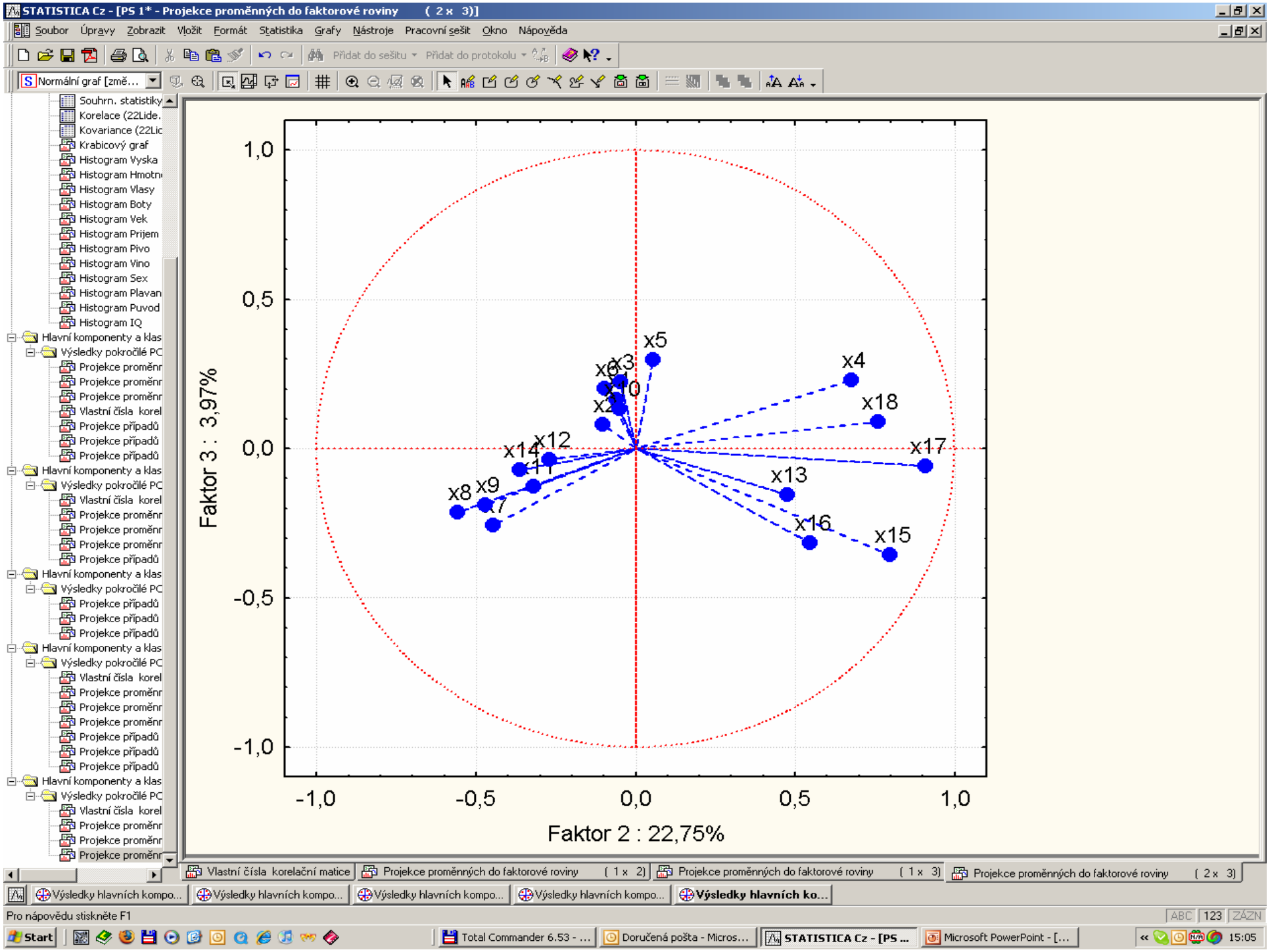
Obr. 4.20a Graf komponentních vah 1 a 2 matice dat *Guiseppe* (STATISTICA).



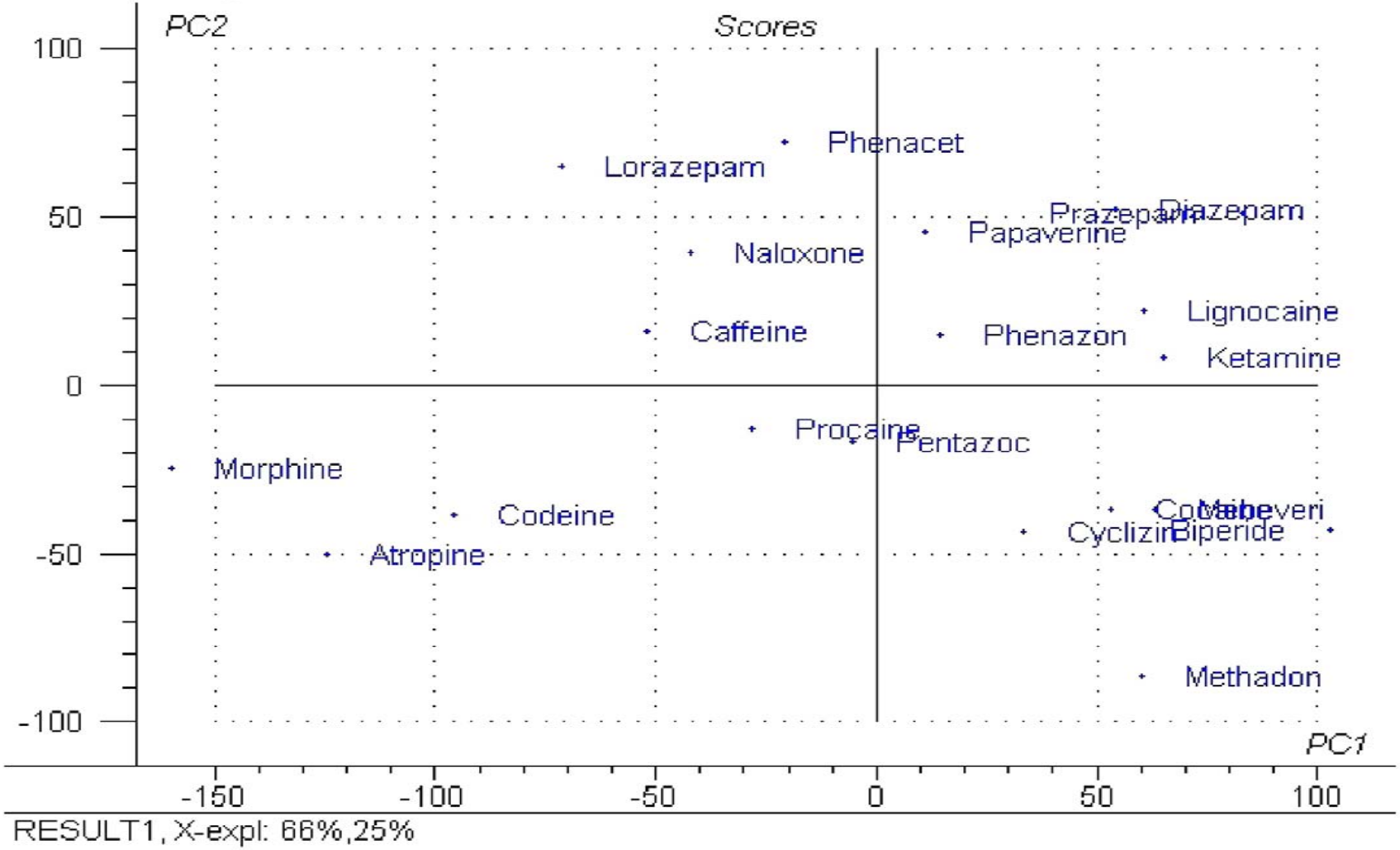
Obr. 4.20 Graf komponentních vah 1 a 2 dat *Guiseppe* (UNSCRAMBLER).


STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 2)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
Projekce případů
Souhrn. statistiky
Korelace (22Lide.
Kovariance (22Lic
Krabicový graf
Histogram Vyska
Histogram Hmotn
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavan
Histogram Puvod
Histogram IQ
Hlavní komponenty a klas
Výsledky pokročilé PC
Projekce proměnn
Vlastní čísla korel
Projekce případů
x17
x15
x18
x4
x16
x13
x5
x10
x12
x11
x14
x7
x9
x8
1,0
0,5
0,0
-0,5
-1,0
Faktor 2 : 22,75%
Faktor 1 : 67,47%
Vlastní čísla korelační matice
Projekce proměnných do faktorové roviny (1 x 2)
Výsledky hlavních kompo...
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.53 - ...
Doručená pošta - Micros...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
15:03

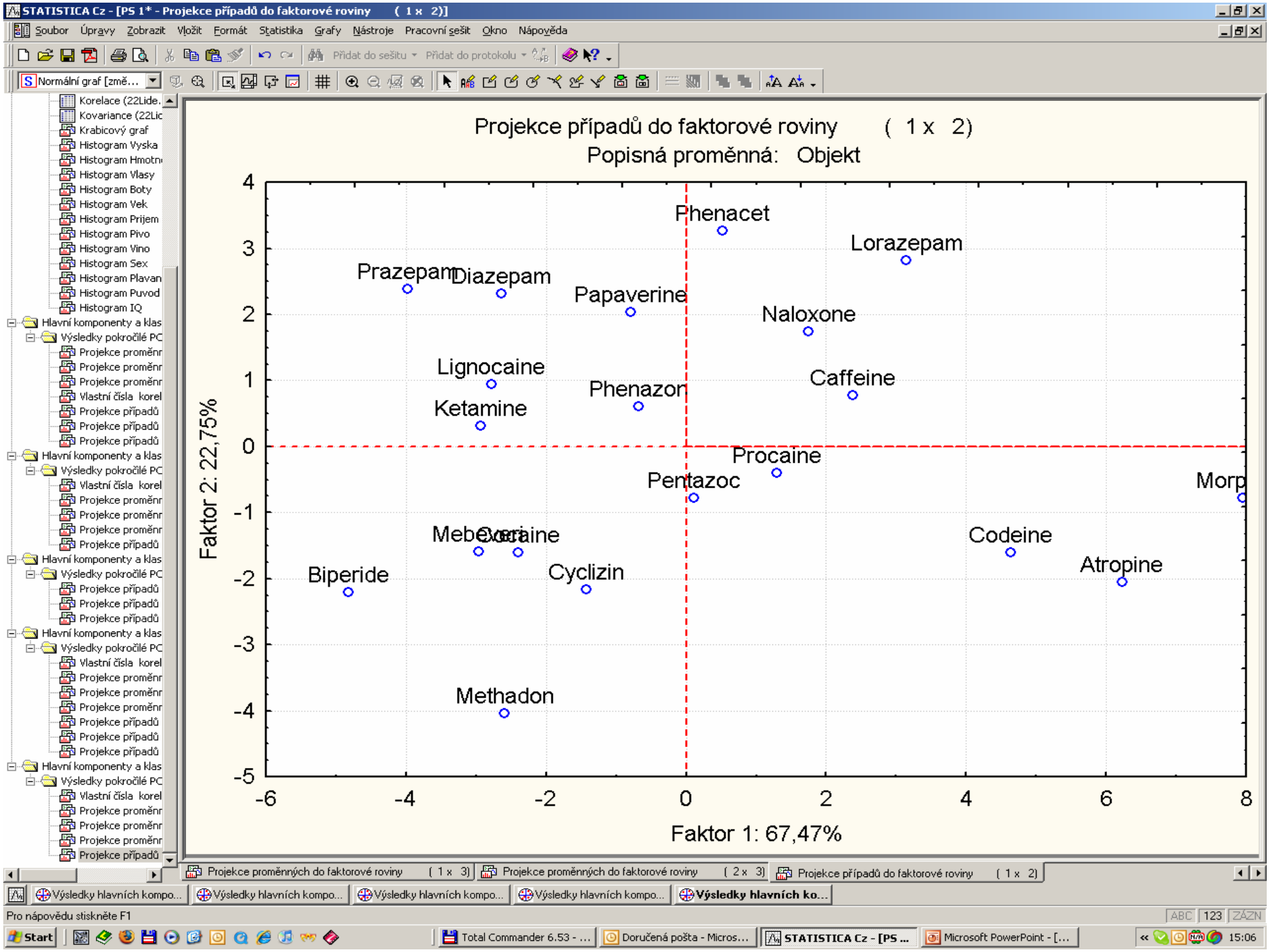

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (1 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
Projekce případů
Souhrn. statistiky
Korelace (22Lide.
Kovariance (22Lic
Krabicový graf
Histogram Vyska
Histogram Hmotn
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavan
Histogram Puvod
Histogram IQ
Hlavní komponenty a klas
Výsledky pokročilé PC
Projekce proměnn
Vlastní čísla korel
Projekce případů
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 3,97%
x5
x4
x18
x17
x12
x14
x11
x13
x9
x8
x16
x15
-1,0
-0,5
0,0
0,5
1,0
Faktor 1 : 67,47%
Vlastní čísla korelační matice
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Výsledky hlavních kompo...
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.53 - ...
Doručená pošta - Micros...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
15:04

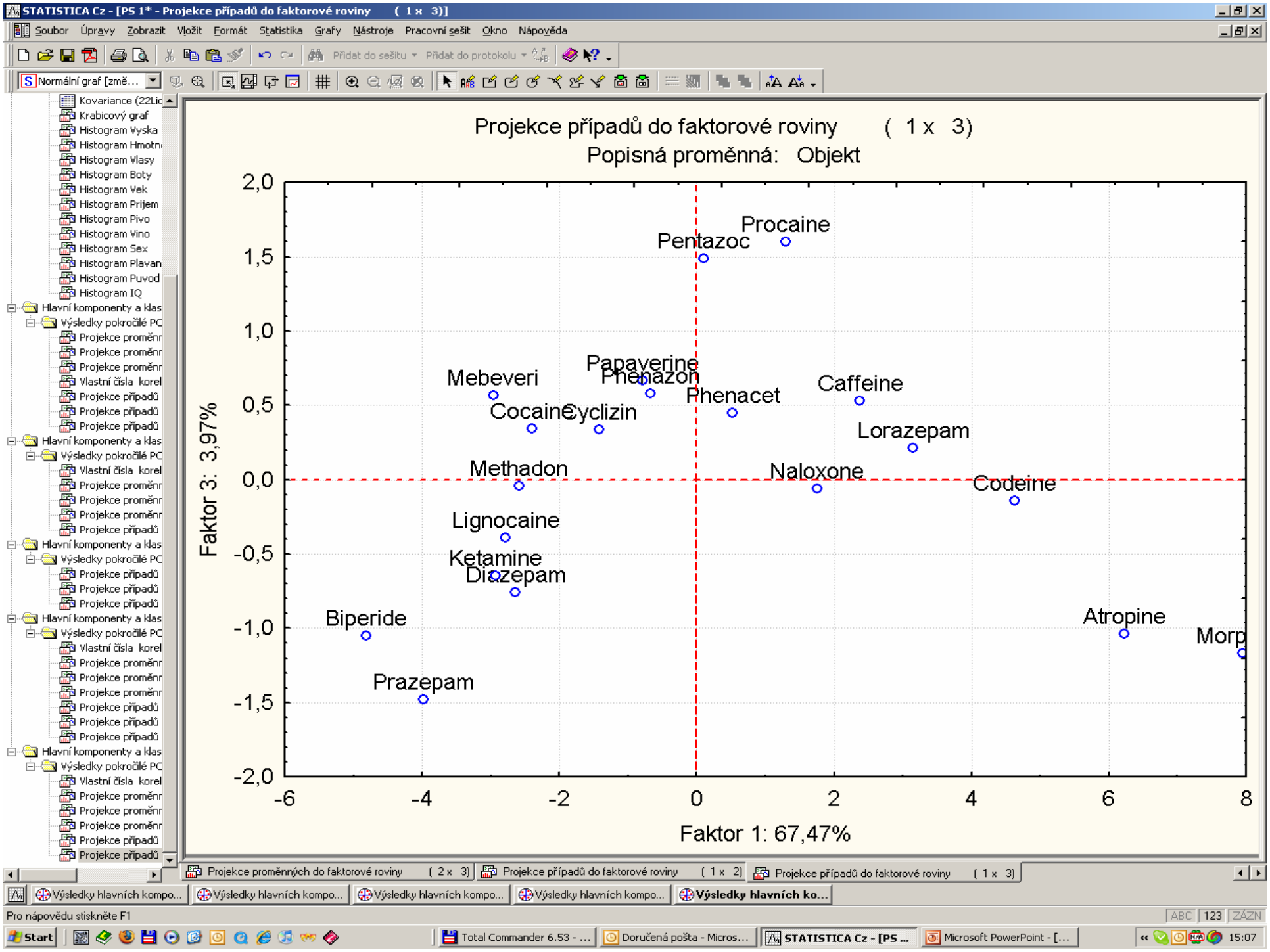

STATISTICA Cz - [PS 1* - Projekce proměnných do faktorové roviny (2 x 3)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
Souhrn. statistiky
Korelace (22Lide.
Kovariance (22Lic
Krabicový graf
Histogram Vyska
Histogram Hmotn
Histogram Vlasy
Histogram Boty
Histogram Vek
Histogram Prijem
Histogram Pivo
Histogram Vino
Histogram Sex
Histogram Plavan
Histogram Puvod
Histogram IQ
Hlavní komponenty a klas
Výsledky pokročilé PC
Projekce proměnn
Vlastní čísla korel
Projekce případů
1,0
0,5
0,0
-0,5
-1,0
Faktor 3 : 3,97%
Faktor 2 : 22,75%
x5
x4
x18
x17
x13
x16
x15
x12
x14
x11
x8
x9
x7
x3
x10
Vlastní čísla korelační matice
Projekce proměnných do faktorové roviny (1 x 2)
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Výsledky hlavních kompo...
Výsledky hlavních ko...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.53 - ...
Doručená pošta - Micros...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
15:05

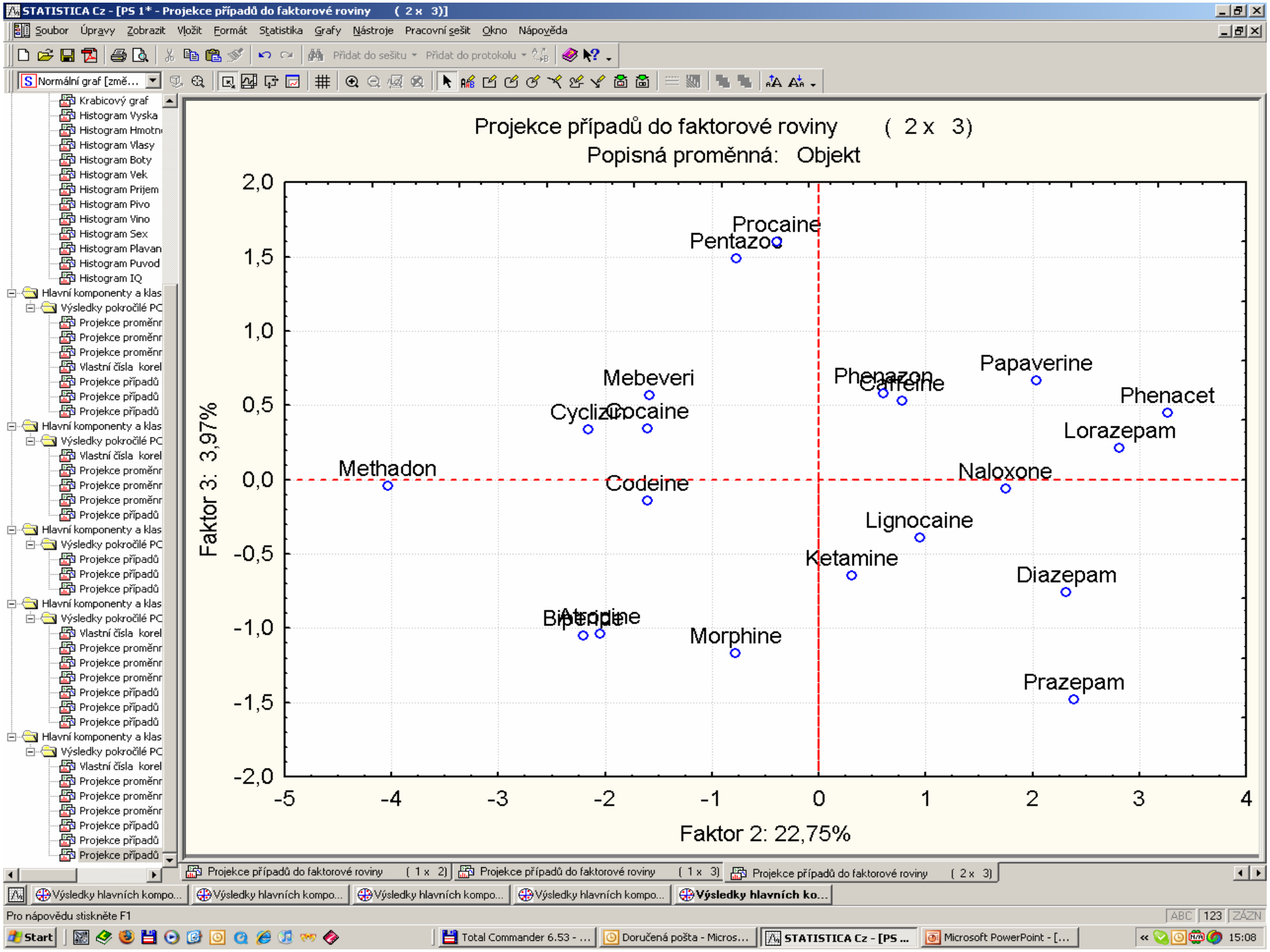
3. Rozptylový diagram komponentního skóre: roztřídil 20 sloučenin do shluků. Sloučeniny blízko sebe jsou si z hlediska chromatografického dělení značně podobné.



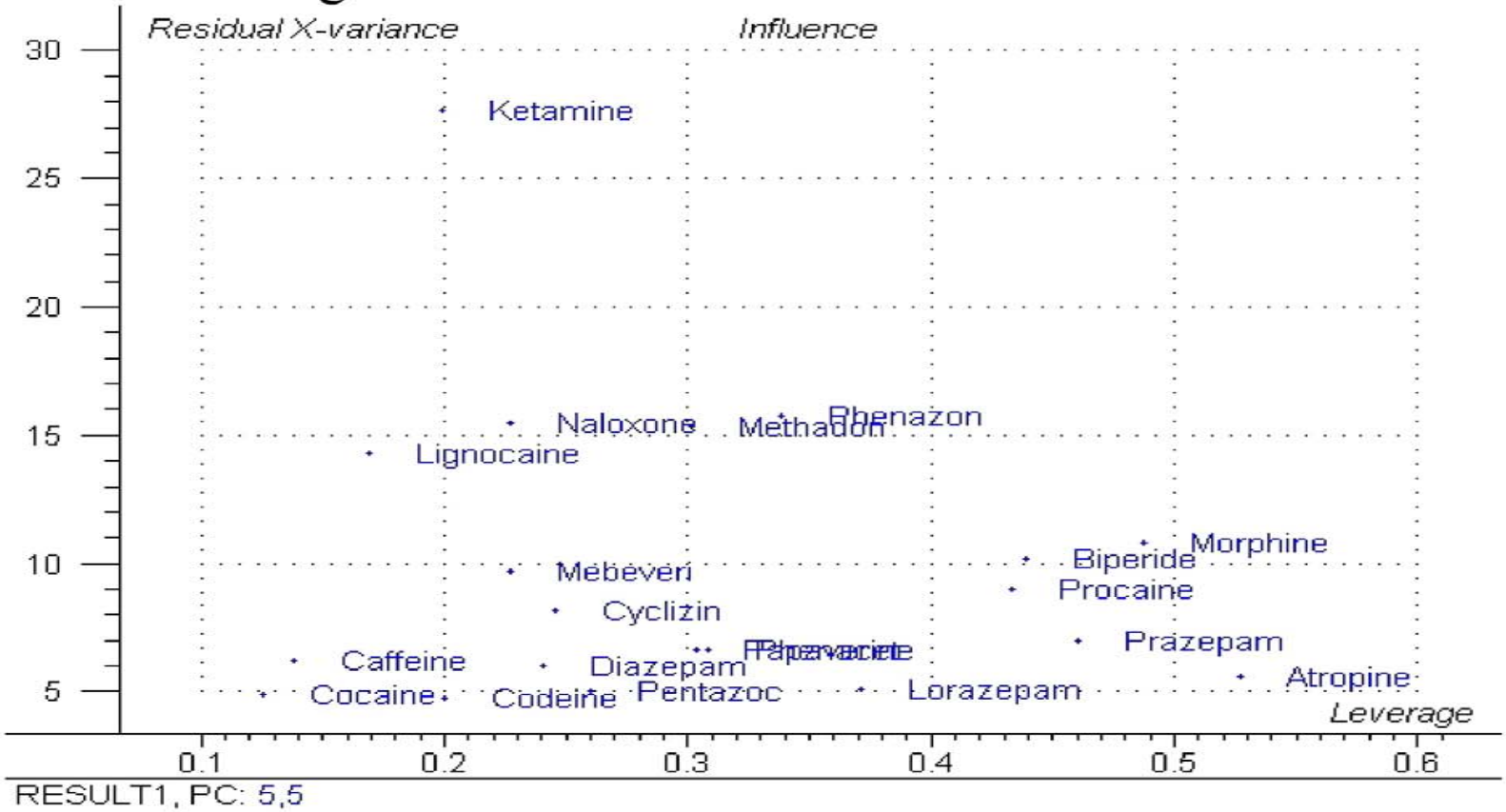
Obr. 4.21 Rozptylový diagram komponentního skóre dat *Guiseppe* (UNSCRAMBLER).


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 2)]
Projekce případů do faktorové roviny (1 x 2)
Popisná proměnná: Objekt
Phenacet
Lorazepam
Prazepam
Diazepam
Papaverine
Naloxone
Lignocaine
Caffeine
Phenazon
Ketamine
Procaine
Pentazoc
Morp
Codeine
Atropine
Biperide
Cyclizin
Methadon
Faktor 2: 22,75%
Faktor 1: 67,47%
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8
Projekce proměnných do faktorové roviny (1 x 3)
Projekce proměnných do faktorové roviny (2 x 3)
Projekce případů do faktorové roviny (1 x 2)
Pro nápovědu stiskněte F1


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (1 x 3)]
Projekce případů do faktorové roviny (1 x 3)
Popisná proměnná: Objekt
Procaine
Pentazoc
Papaverine
Phenazon
Mebeveri
Phenacet
Caffeine
Cocaine
Cyclizin
Lorazepam
Methadon
Naloxone
Codeine
Lignocaine
Ketamine
Diazepam
Biperide
Atropine
Morp
Prazepam
Faktor 3: 3,97%
Faktor 1: 67,47%
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-6
-4
-2
0
2
4
6
8


STATISTICA Cz - [PS 1* - Projekce případů do faktorové roviny (2 x 3)]
Projekce případů do faktorové roviny (2 x 3)
Popisná proměnná: Objekt
Procaine
Pentazoc
Mebeveri
Cyclizine
Cocaine
Phenazon
Caffeine
Papaverine
Phenacet
Lorazepam
Methadon
Codeine
Naloxone
Lignocaine
Ketamine
Diazepam
Atropine
Morphine
Prazepam
Faktor 3: 3,97%
Faktor 2: 22,75%
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-5
-4
-3
-2
-1
0
1
2
3
4
Projekce případů do faktorové roviny (1 x 2)
Projekce případů do faktorové roviny (1 x 3)
Projekce případů do faktorové roviny (2 x 3)
Pro nápovědu stiskněte F1

4. Indikace vlivných bodů: graf vybočujících a extrémů: většina sloučenin navrženému modelu PCA dobře vyhovuje. Výjimku tvoří *Ketamin* a dále *Lignocaine*, *Naloxone, Mathodon* a *Rhenazon*, kterým model méně vyhovuje a jsou v horní části grafu.



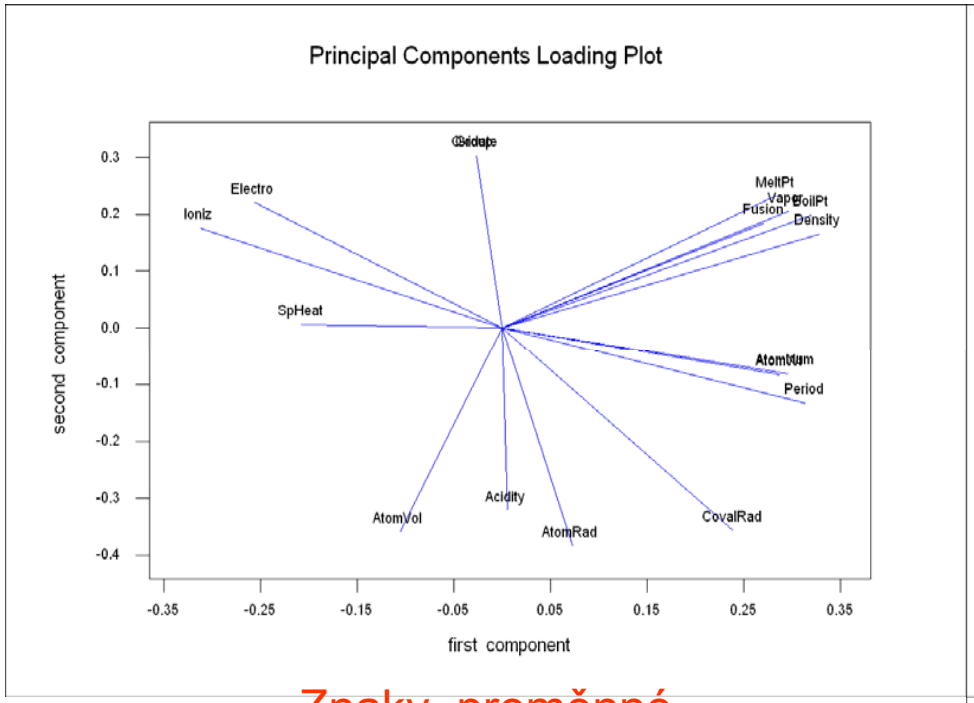
Obr. 4.22 Graf vlivných bodů statistické analýzy reziduí dat *Guiseppe* (UNSCRAMBLER).

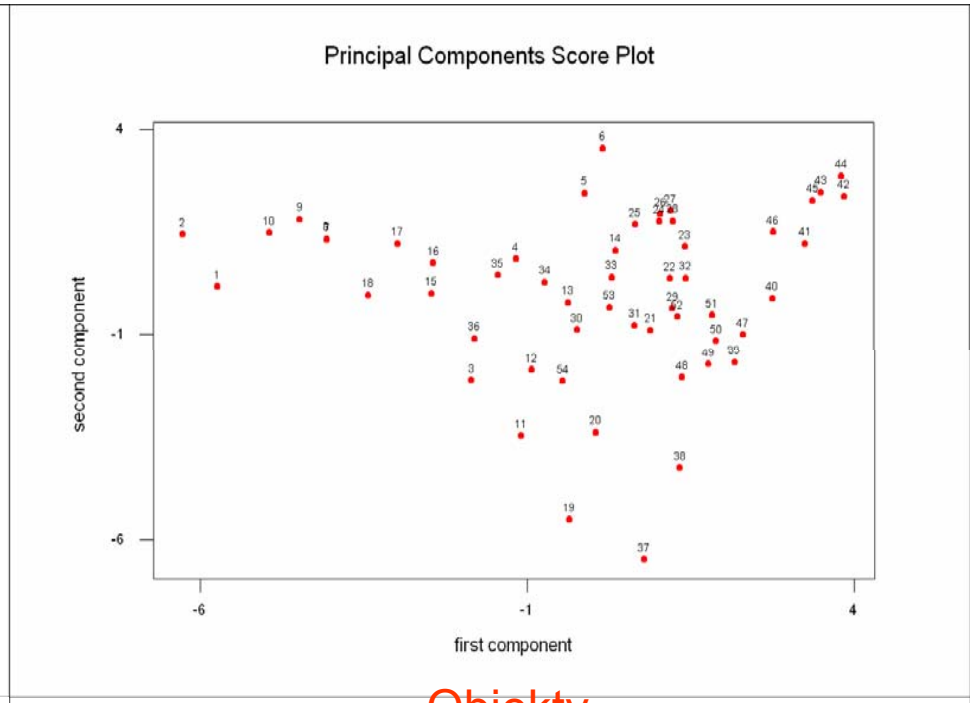
○ *Závěr:* PCA je pomůckou při chromatografickém dělení 20 sloučenin na základě 18 elučních činidel.

Úloha 5. Klasifikace prvků periodické tabulky do shluků

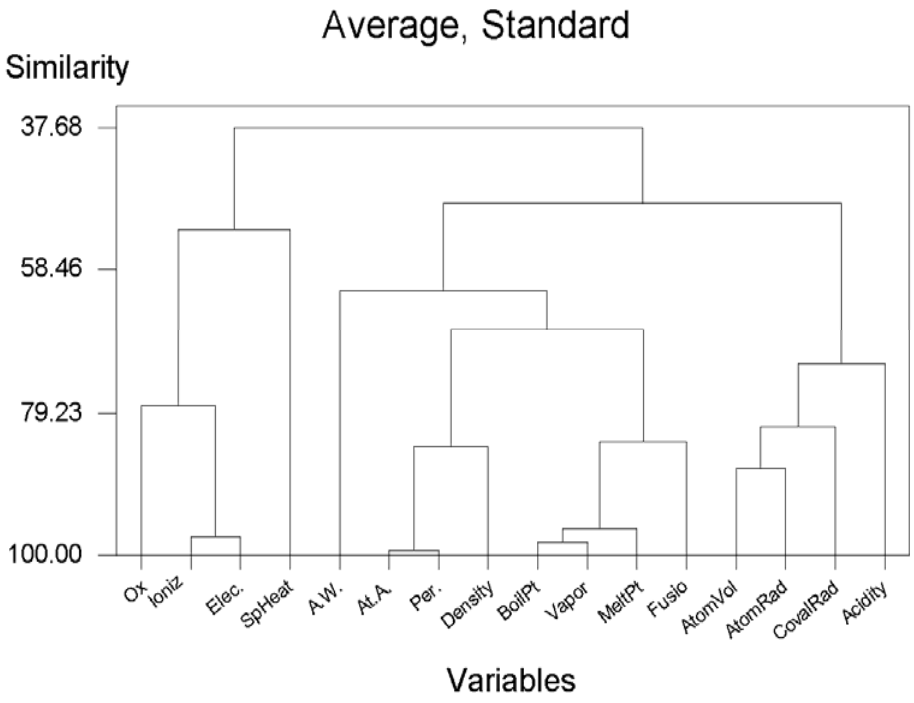
Pro 54 prvků periodické tabulky bylo použito 18 rozličných fyzikálně-chemických vlastností. Nalezněte shluky podobných vlastností a shluky podobných prvků.

Row	Elem	At.A.	Per.	A.W.	Ox	Ioniz	Elec.	BoilPt	MeltPt	Vapor	Fusio	SpHeat	AtomVol	CovalRad	AtomRad	Density	Acidity
1	H	1	1	1	1	313	2.1	20	14	0.11	0.01	3.45	14.1	0.32	0.99	0.07	3
2	He	2	1	4	8	567	5	4	3	0.02	0.01	1.25	31.8	0.93	1.3	0.13	3
3	Li	3	2	6.9	1	124	1	1603	454	32.5	0.72	0.79	13.1	1.23	1.55	0.53	5
4	Be	4	2	9	2	215	1.5	3043	1550	73.9	2.8	0.45	5	0.9	1.12	1.85	3
5	B	5	2	11	3	191	2	4198	2303	128	5.3	0.309	4.6	0.82	0.98	2.34	2
6	C	6	2	12	4	260	2.5	5103	4000	172	2.6	0.165	5.3	0.77	0.91	2.26	2
7	N	7	2	14	5	336	3	77	63	0.67	0.09	0.247	17.3	0.75	0.92	0.81	1
8	O	8	2	16	6	314	3.5	90	54	0.82	0.05	0.218	14	0.73	0.84	1.14	3
9	F	9	2	19	7	402	4	85	54	0.76	0.06	0.18	17.1	0.72	0.81	1.5	3
10	Ne	10	2	20	8	497	5	31	25	0.42	0.08	0.2	16.8	0.71	1.76	1.2	3
11	Na	11	3	23	1	119	0.9	1165	371	24.1	0.62	0.295	23.7	1.54	1.9	0.97	5
12	Mg	12	3	24	2	176	1.2	1380	923	32.5	2.14	0.25	14	1.36	1.6	1.74	5
13	Al	13	3	27	3	138	1.5	2723	933	67.9	2.55	0.215	10	1.18	1.43	2.7	3
14	Si	14	3	28	4	188	1.8	2953	1683	40.6	11.1	0.162	12.1	1.11	1.32	2.33	3
15	P	15	3	31	5	254	2.1	553	317	2.97	0.15	0.177	17	1.06	1.28	1.82	2
16	S	16	3	32	6	239	2.5	718	392	3.01	0.34	0.175	15.5	1.02	1.27	2.07	1
17	Cl	17	3	36	7	300	3	238	172	2.44	0.77	0.116	18.7	0.99	1.09	1.56	1
18	Ar	18	3	40	8	363	4	87	84	1.56	0.28	0.125	24.2	0.98	2.11	1.4	3
19	K	19	4	39	1	100	0.8	1033	337	18.9	0.55	0.177	45.3	2.03	2.35	0.86	5
20	Ca	20	4	40	2	141	1	1713	1111	36.7	2.1	0.149	29.9	1.74	1.97	1.55	5
21	Sc	21	4	45	3	151	1.3	3003	1812	81	3.8	0.13	15	1.44	1.62	3	4
22	Ti	22	4	48	4	158	1.5	3533	1941	107	3.7	0.126	10.6	1.32	1.47	4.51	3
23	V	23	4	51	5	156	1.6	3723	2173	106	4.2	0.12	8.4	1.22	1.34	6.1	3
24	Cr	24	4	52	6	156	1.6	2938	2148	73	3.3	0.11	7.2	1.18	1.3	7.19	1
25	Mn	25	4	55	7	171	1.5	2423	1518	53.7	3.5	0.115	7.4	1.17	1.35	7.43	1
26	Fe	26	4	56	8	182	1.8	3273	1809	84.6	3.67	0.11	7.1	1.17	1.26	7.86	3
27	Co	27	4	59	8	181	1.8	3173	1768	93	3.64	0.099	6.7	1.16	1.25	8.9	3
28	Ni	28	4	59	8	176	1.8	3003	1726	91	4.21	0.105	6.6	1.15	1.24	8.9	4
29	Cu	29	5	64	1	178	1.9	2868	1356	72.8	3.11	0.092	7.1	1.17	1.28	8.96	4
30	Zn	30	5	65	2	216	1.6	1179	693	27.4	1.76	0.91	9.2	1.25	1.38	7.14	3
31	Ga	31	5	70	3	138	1.6	2510	303	70.7	1.34	0.079	11.8	1.26	1.41	5.91	3
32	Ge	32	5	73	4	187	1.8	3103	1211	68	7.6	0.073	13.6	1.22	1.37	5.32	3
33	As	33	5	75	5	231	2	886	1090	7.75	6.62	0.083	13.1	1.2	1.39	5.72	2
34	Se	34	5	79	6	225	2.4	958	490	3.34	1.25	0.084	16.5	1.16	1.4	4.79	1
35	Br	35	5	80	7	273	2.8	331	266	3.58	1.26	0.07	23.5	1.14	1.24	3.12	1
36	Kr	36	5	84	8	323	3	121	116	2.16	0.39	0.08	32.2	1.12	2.16	2.6	3
37	Rb	37	6	86	1	96	0.8	961	312	18.1	0.55	0.08	55.9	2.16	2.48	1.53	5
38	Sr	38	6	88	2	131	1	1653	1041	33.8	2.1	0.055	33.7	1.91	2.15	2.6	5
39	Y	39	6	89	3	152	1.3	3200	1782	93	2.7	0.071	19.4	1.62	1.78	4.47	4
40	Zr	40	6	91	4	160	1.4	3853	2125	120	4	0.066	14.1	1.45	1.6	6.49	3
41	Nb	41	6	93	5	156	1.6	3573	2741	125	6.4	0.065	10.8	1.34	1.46	8.4	2
42	Mo	42	6	96	6	166	1.8	5833	2883	128	6.6	0.061	9.4	1.3	1.39	10.2	1
43	Tc	43	6	98	7	167	1.9	5273	2413	120	5.5	0.06	9	1.27	1.36	11.5	1
44	Ru	44	6	1	1.1	8	173	2.2	5173	2773	148	6.1	0.057	8.3	1.25	1.34	12.2	2
45	Rh	45	6	1	2.9	8	178	2.2	4773	2239	127	5.2	0.059	8.3	1.25	1.34	12.4	3
46	Pd	46	6	1	6.4	8	192	2.2	4253	1825	90	4	0.058	8.9	1.28	1.37	12	4
47	Ag	47	7	1	7.9	1	175	1.9	2483	1234	60.7	2.7	0.056	10.3	1.34	1.44	10.5	3
48	Cd	48	7	1	12	2	207	1.7	1038	594	23.9	1.46	0.055	13.1	1.48	1.54	8.65	4
49	In	49	7	1	15	3	133	1.7	2273	429	53.7	0.78	0.057	15.7	1.44	1.66	7.31	3
50	Sn	50	7	1	19	4	169	1.8	2543	505	70	1.72	0.054	16.3	1.41	1.62	7.3	3
51	Sb	51	7	1	22	5	199	1.9	1653	904	46.6	4.74	0.049	18.4	1.4	1.59	6.62	2
52	Te	52	7	1	28	6	208	2.1	1263	723	11.9	4.28	0.047	20.5	1.36	1.6	6.24	2
53	I	53	7	1	27	7	241	2.5	456	387	5.2	1.87	0.052	25.7	1.33	1.44	4.94	1
54	Xe	54	7	1	31	8	280	3	165	161	3.02	0.55	0.05	42.9	1.31	2.27	3.06	3

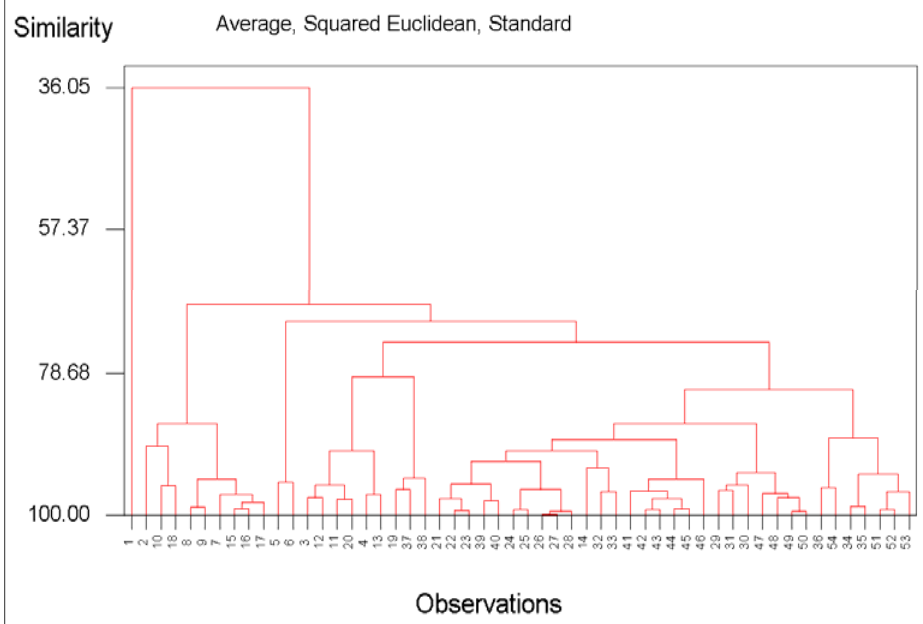




Znaky, proměnné



Objekty



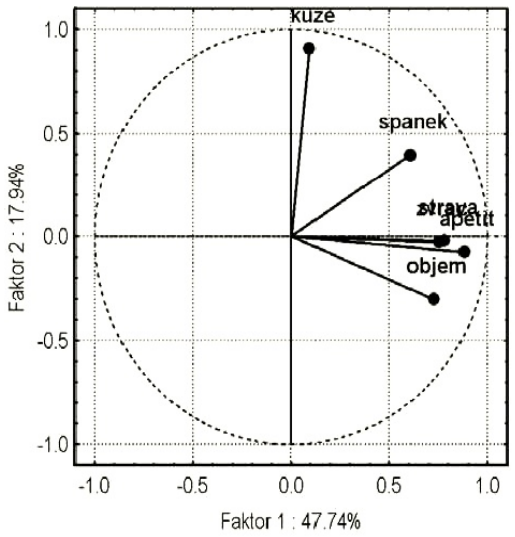
PŘÍKLAD 9.11 *Výstavba shluků u radioterapeutického léčení vybraných pacientů*

U 98 pacientů byl sledováno radioterapeutické léčení. Do kolika shluků se roztřídí 98 pacientů?

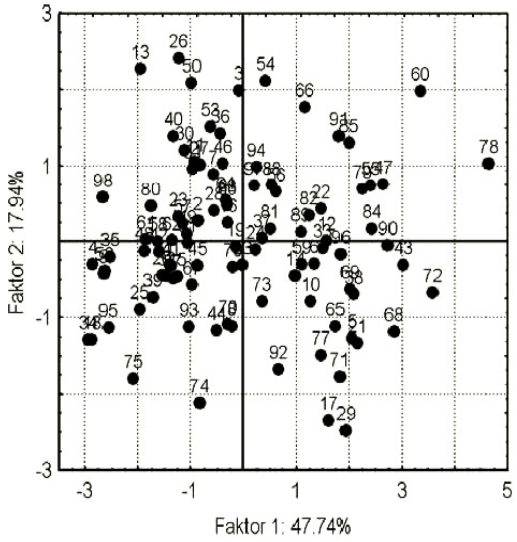
○ ***Data*:** Data *Radioterapie* obsahuje 98 pacientů 6 sledovaných znaků:
Pacient je index pacienta,
Zvrac počet symptomů jako je pálení žáhy, zvracení atd.,
Objem značí objem provedených činností ve stupnici 1 až 5,
Spanek značí objem spánku ve stupnici 1 až 5,
Strava značí množství zkonzumované stravy,
Apetit značí apetit ve stupnici 1 až 5,
Kuze značí podrážděnost kůže ve stupnici 0 až 3.

Pacient	*Zvrac*	*Objem*	*Spanek*	*Strava*	*Apetit*	*Kuze*
1	0.889	1.389	1.555	2.222	1.945	1
...	...	...	...	...		...
98	0.889	1	1	2	1	2

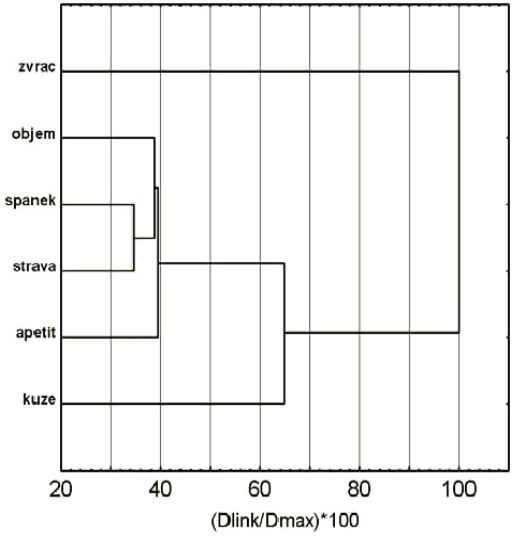
○ ***Řešení:*** Graf komponentních vah znaků ukazuje silnou korelaci znaků ***Objem, Apetit, Strava*** a ***Zvrac***, protože tyto čtyři znaky jsou v grafu představeny téměř totožnými průvodiči.



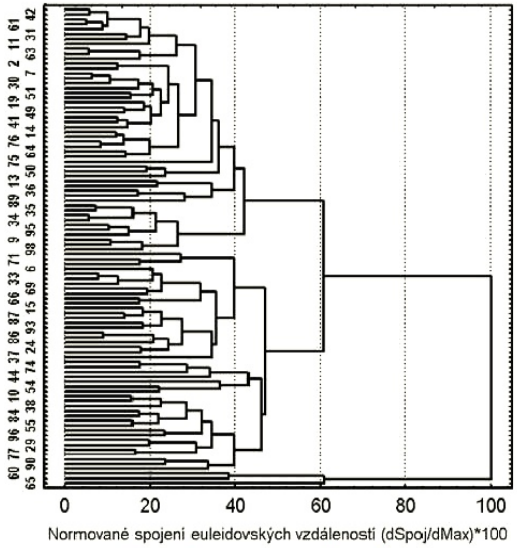
Graf komponentních vah znaků.



Graf komponentního skóre 98 pacientů

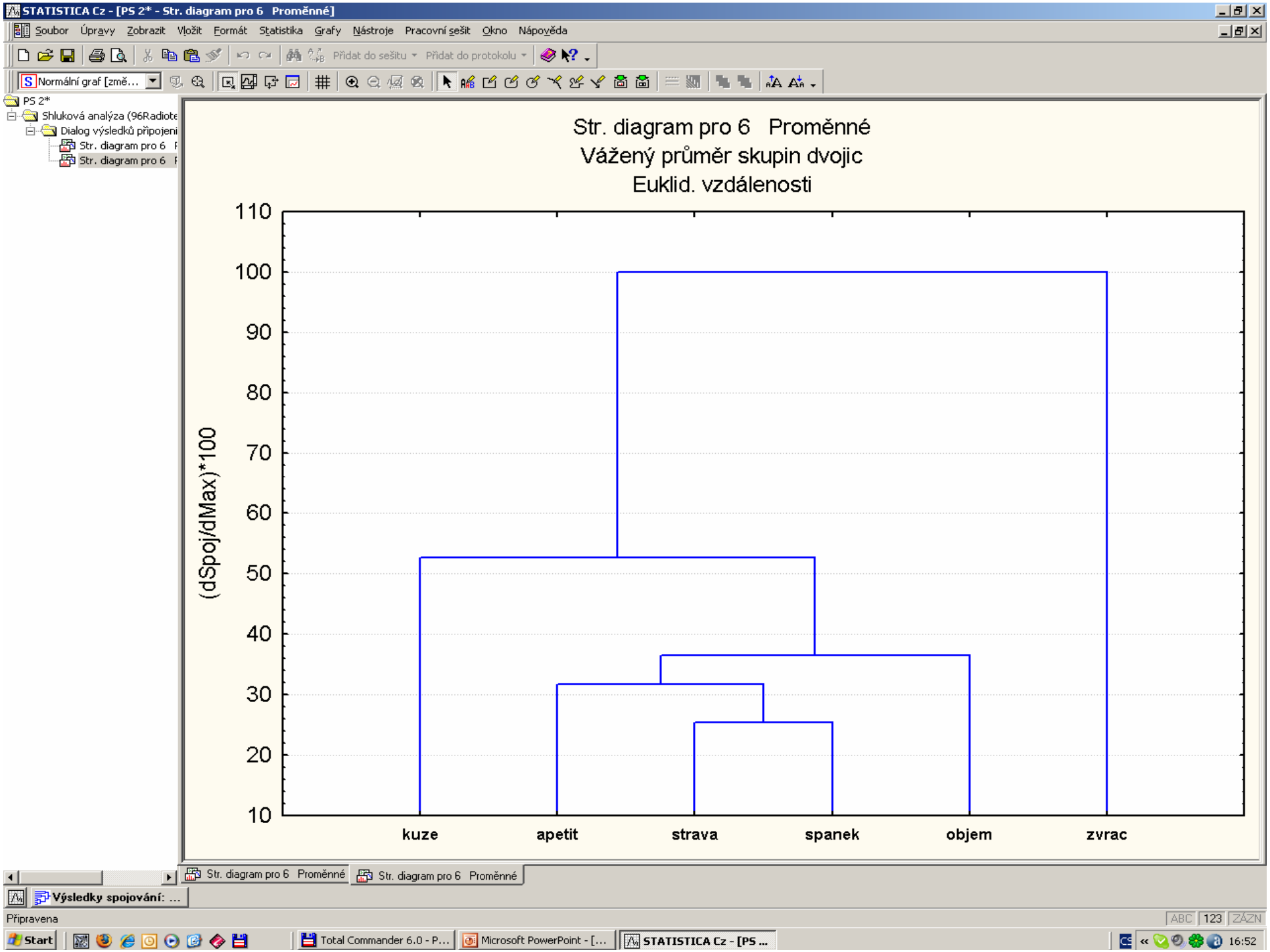


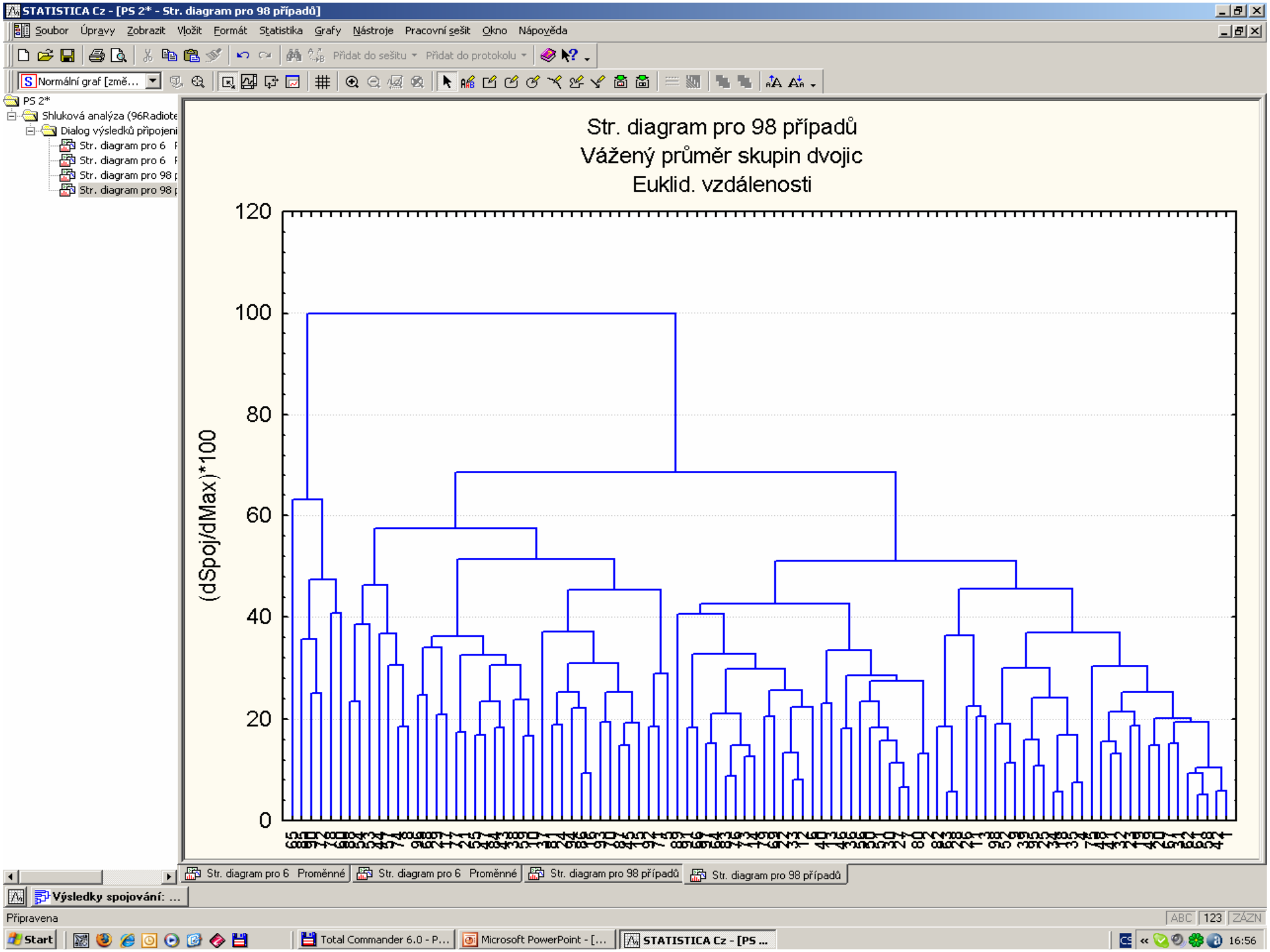
Dendrogram znaků



Dendrogram pacientů

Závěr: Dendrogram objektů klasifikuje 98 pacientů do několika shluků a 3 pacienti jsou odlišní.


STATISTICA Cz - [PS 2* - Str. diagram pro 6 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 2*
Shluková analýza (96Radiote
Dialog výsledků připojeni
Str. diagram pro 6 P
Str. diagram pro 6 P
Str. diagram pro 6 Proměnné
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
110
100
90
80
70
60
50
40
30
20
10
kuze
apetit
strava
spanek
objem
zvrac
Str. diagram pro 6 Proměnné
Str. diagram pro 6 Proměnné
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
16:52


STATISTICA Cz - [PS 2* - Str. diagram pro 98 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 2*
Shluková analýza (96Radiote
Dialog výsledků připojeni
Str. diagram pro 6
Str. diagram pro 6
Str. diagram pro 98
Str. diagram pro 98
Str. diagram pro 98 případů
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
Str. diagram pro 6 Proměnné
Str. diagram pro 6 Proměnné
Str. diagram pro 98 případů
Str. diagram pro 98 případů
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
16:56

PŘÍKLAD 9.12 *Dendrogram úbytku kostní hmoty starších žen po cvičeních a po dietách*

Zkoumáno, zda cvičení nebo doplňky vhodné diety zpomalí úbytek kostní hmoty u žen. Obsah minerálů v kostech byl měřen absorpční fotometrií ve třech kostech na dominantní a ve třech na vedlejší straně. Při klasifikaci je třeba sestrojit dendrogram blízkých znaků a dendrogram vzniklých shluků pacientů.

○ ***Data*:** Data *Kost* obsahuje 25 pacientů obsah minerálů v 6 vyšetřovaných znacích:
Pacient je index pacienta,
Domin značí poloměr u dominantní kosti,
Vedlej značí poloměr u vedlejší kosti,
Dopaze značí dominantní část kosti pažní,
Vepaze značí vedlejší část kosti pažní,
Doloket značí dominantní část kosti loketní a
Veloket značí vedlejší část kosti loketní.

Pacient	*Domin*	*Vedlej*	*Dopaze*	*Vepaze*	*Doloket*	*Veloket*
1	1.103	1.052	2.139	2.238	0.873	0.872
...	...	...	...	...		...
25	0.915	0.936	1.971	1.869	0.869	0.868

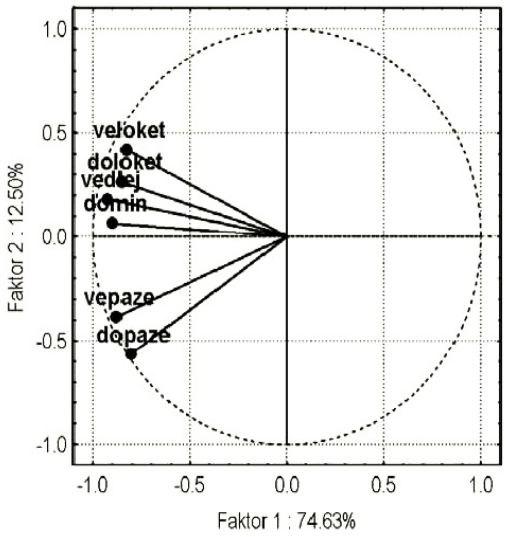
○ ***Řešení:*** **Graf komponentních vah znaků** ukazuje silnou korelaci a podobnost dvojic znaků *Domin-Vedlej*, dále *Doloket-Veloket* a konečně také *Dopaze-Vepaze.*

Dvě dvojice *Domin-Vedlej* a *Doloket-Veloket* spolu rovněž korelují a dle polohy v grafu jsou si také podobné.

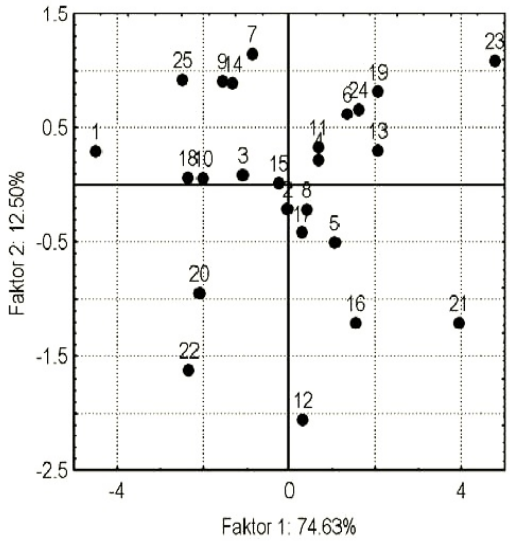
Dendrogram znaků ukazuje ve shodě s předešlým grafem na vznik dvou blízkých shluků, první obsahuje znaky *Domin* a *Vedlej* a druhý shluk obsahuje *Doloket* a *Veloket*, který je méně podobný třetímu shluku, který obsahuje dvojici *Dopaze* a *Vepaze.*

Umístění pacientů na grafu komponentního skóre objektů je vcelku ve shodě s dendrogramem pacientů.

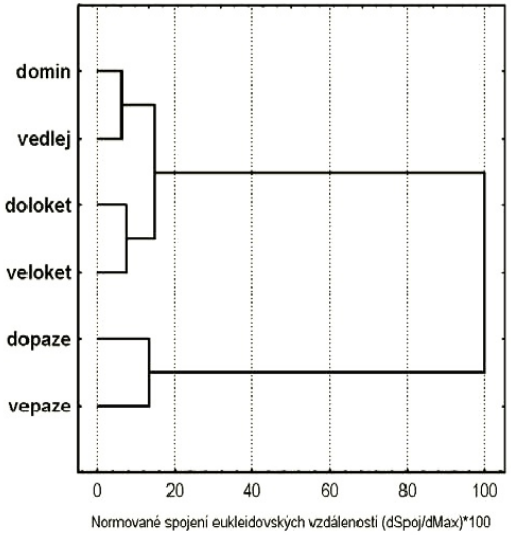
Lze indikovat tři shluky:
První obsahuje objekty 1, 20, 22, 10, 18, 25 a 12.
Druhý velký shluk obsahuje 2, 5, 8, 16, 17, 4, 11, 3, 9, 14, 7 a 15.
Třetí shluk obsahuje objekty 6, 13, 24, 19, 21.
Objekt 23 je odlehlý, nepodobný všem ostatním.



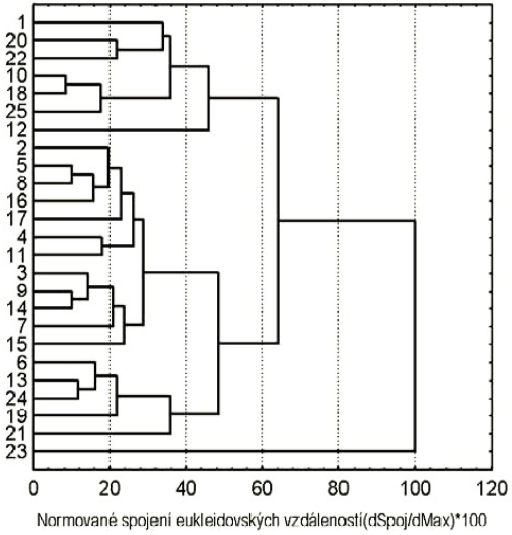
Graf komponentních vah znaků matice dat *Kost*, (STATISTICA).



Graf komponentního skóre 25 pacientů matice dat *Kost*

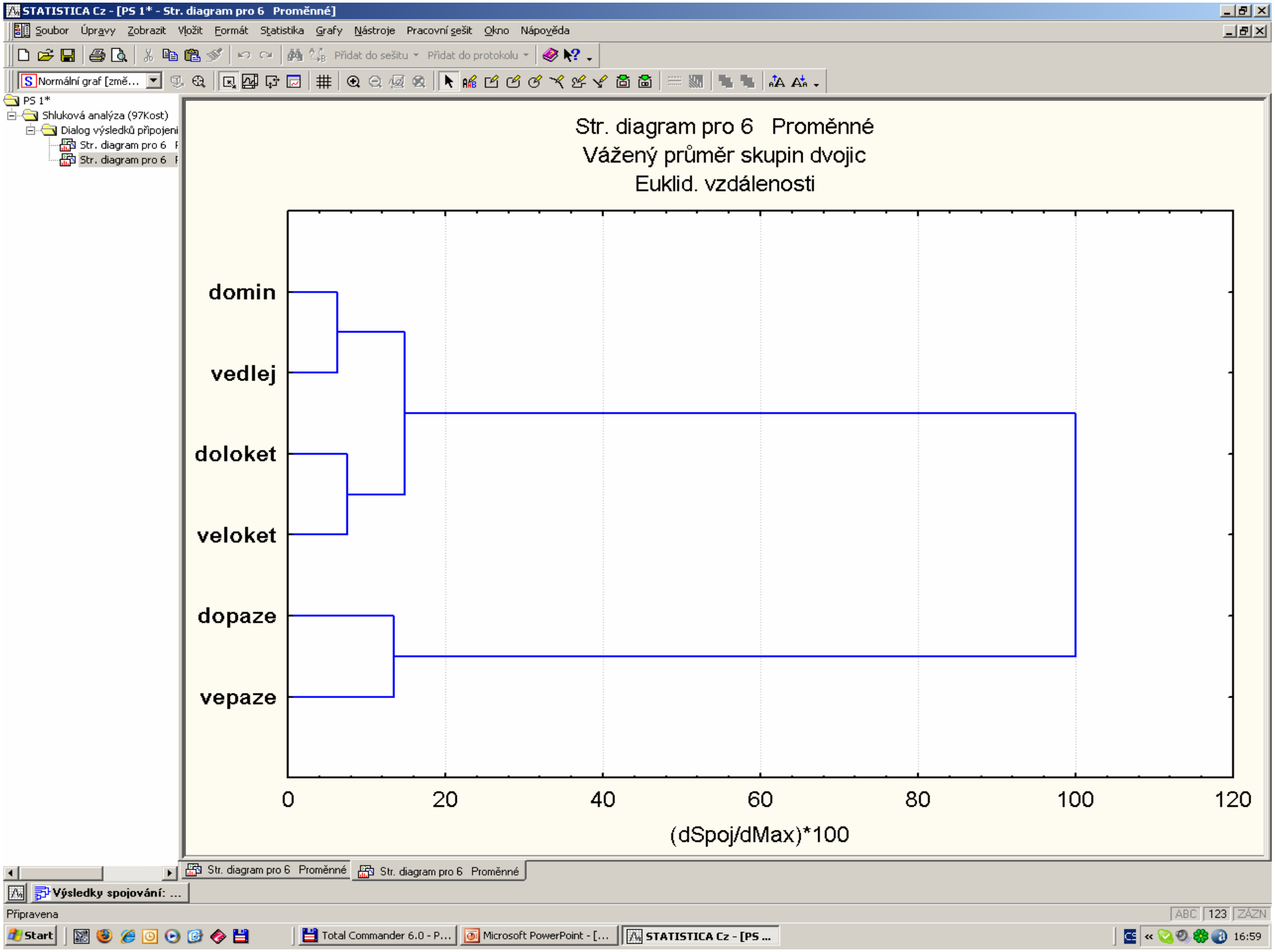


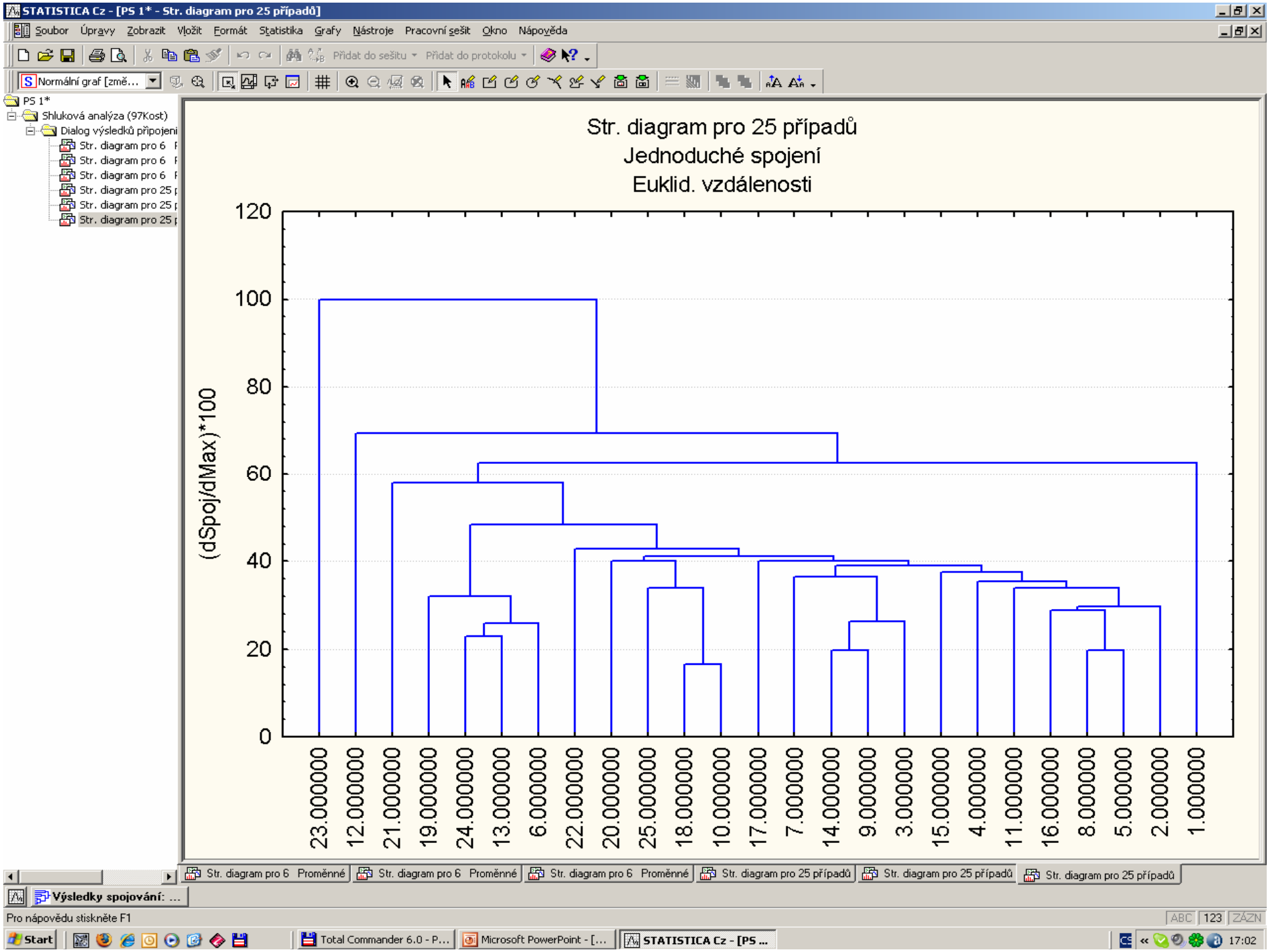
Dendrogram znaků matice dat *Kost* (STATISTICA).



Dendrogram 25 pacientů matice dat *Kost*, (STATISTICA).

○ ***Závěr:*** Pacienti byli roztříděni do třech shluků. Ostatní je třeba považovat za odlehlé.


STATISTICA Cz - [PS 1* - Str. diagram pro 6 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (97Kost)
Dialog výsledků připojeni
Str. diagram pro 6 P
Str. diagram pro 6 P
Str. diagram pro 6 Proměnné
Vážený průměr skupin dvojic
Euklid. vzdálenosti
domin
vedlej
doloket
veloket
dopaze
vepaze
0
20
40
60
80
100
120
(dSpoj/dMax)*100
Str. diagram pro 6 Proměnné
Str. diagram pro 6 Proměnné
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
16:59


STATISTICA Cz - [PS 1* - Str. diagram pro 25 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (97Kost)
Dialog výsledků připojení
Str. diagram pro 6
Str. diagram pro 25
Str. diagram pro 25 případů
Jednoduché spojení
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
23.00000
12.00000
21.00000
19.00000
24.00000
13.00000
6.00000
22.00000
20.00000
25.00000
18.00000
10.00000
17.00000
7.00000
14.00000
9.00000
3.00000
15.00000
4.00000
11.00000
16.00000
8.00000
5.00000
2.00000
1.00000
Str. diagram pro 6 Proměnné
Str. diagram pro 25 případů
Výsledky spojování: ...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:02

PŘÍKLAD 9.15 *Hledání podobnosti vlastností křupavých lupínků od různých výrobců*

Tři americké firmy General Mills (G), Kellogg (K) a Quaker (Q) produkují křupavé obilné lupínky a bylo sledováno 10 znaků. Byla vyšetřována struktura a vzájemné vazby mezi sledovanými znaky jednotlivých produktů, ale i mezi objekty. Které objekty jsou si velice podobné?

Data: Datová matice *Krupky* obsahuje 55 dodavatelů a vyšetřováno 10 znaků:

Objekt značí index obilných lupínků x_1,
i značí jednoho ze tří výrobců G, K či Q x_2,
Cal značí kalorickou hodnotu [cal] x_3,
Bilkov značí obsah bílkovin x_4,
Tuky značí obsah tuků x_5,
Na značí obsah sodných iontů x_6,
Vlakn značí obsah vlákniny x_7,
Uhlovod značí obsah uhlovodíků x_8,
Cukr značí obsah cukru x_9,
K značí obsah draselných iontů x_{10},
Skupina značí zařazení do skupiny x_{11}.

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
1	ACCheerios	G	110	2	2	180	1.5	10.5	10	70	1
...	...	...	...	...	...	...	...	...	...	...	..
55	QuakerOatmeal	Q	100	5	2	0	2.7	1	1	110	3

○ ***Řešení:*** Korelaci znaků indikuje graf komponentních vah znaků.

Tři znaky *Na, Cal, Cukr* jsou v silné korelaci, protože jsou v grafu blízko sebe a úhel mezi jejich průvodiči je velice malý.

Druhý shluk obsahuje čtyři znaky *Tuky, K, Vlakn, Bilkov*, které jsou vzájemně rovněž silně korelovány.

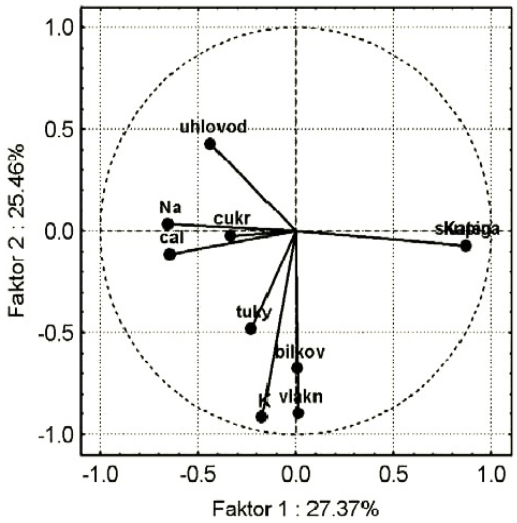
***Skupina* a *Kategorie* korelují**, protože označují stejnou věc.

Uhlovod je vybočující znak, který slabě či vůbec nekoreluje s ostatními znaky.

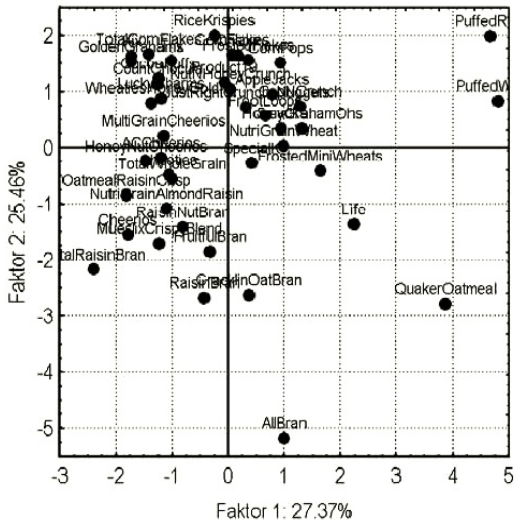
Dendrogram znaků ukazuje dva shluky a dva zcela odlehlé znaky:
První shluk obsahuje 6 vzájemně velice podobných znaků *Bilkov, Vlakn, Tuky, Skupina, Cukr* a *Uhlovod.*
Druhý shluk obsahuje 2 znaky *Kateg* a *Cal.* K nim se připojuje osamocený znak *K.*

Naprosto nepodobný znak vůči všem ostatním znakům je *Na.*

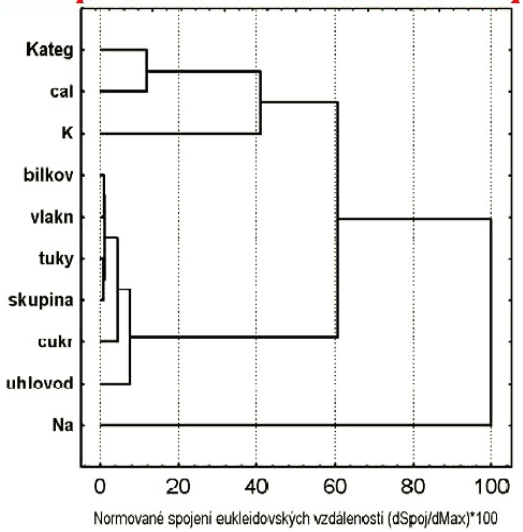
Graf komponentního skóre objektů naznačuje několik shluků objektů, které jsou v souladu se shluky určenými na základě eukleidovské vzdálenosti v dendrogramu. Zcela nepodobný objekt se všemi ostatními se jeví *AllBran*. Také další tři objekty se jeví silně odlišné od ostatních.



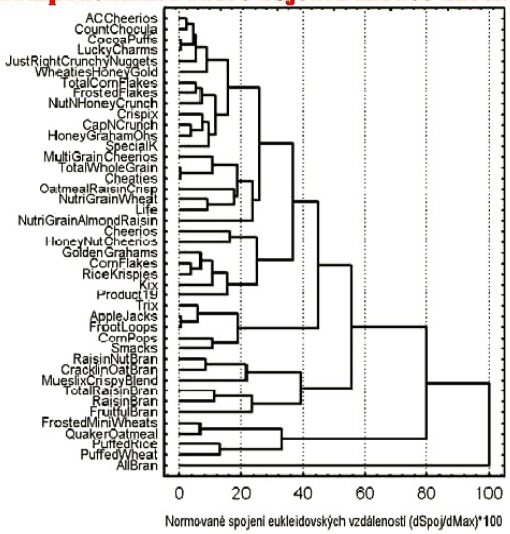
Graf komponentních vah znaků matice dat *Krupky*



Graf komponentního skóre objektů matice dat *Krupky*

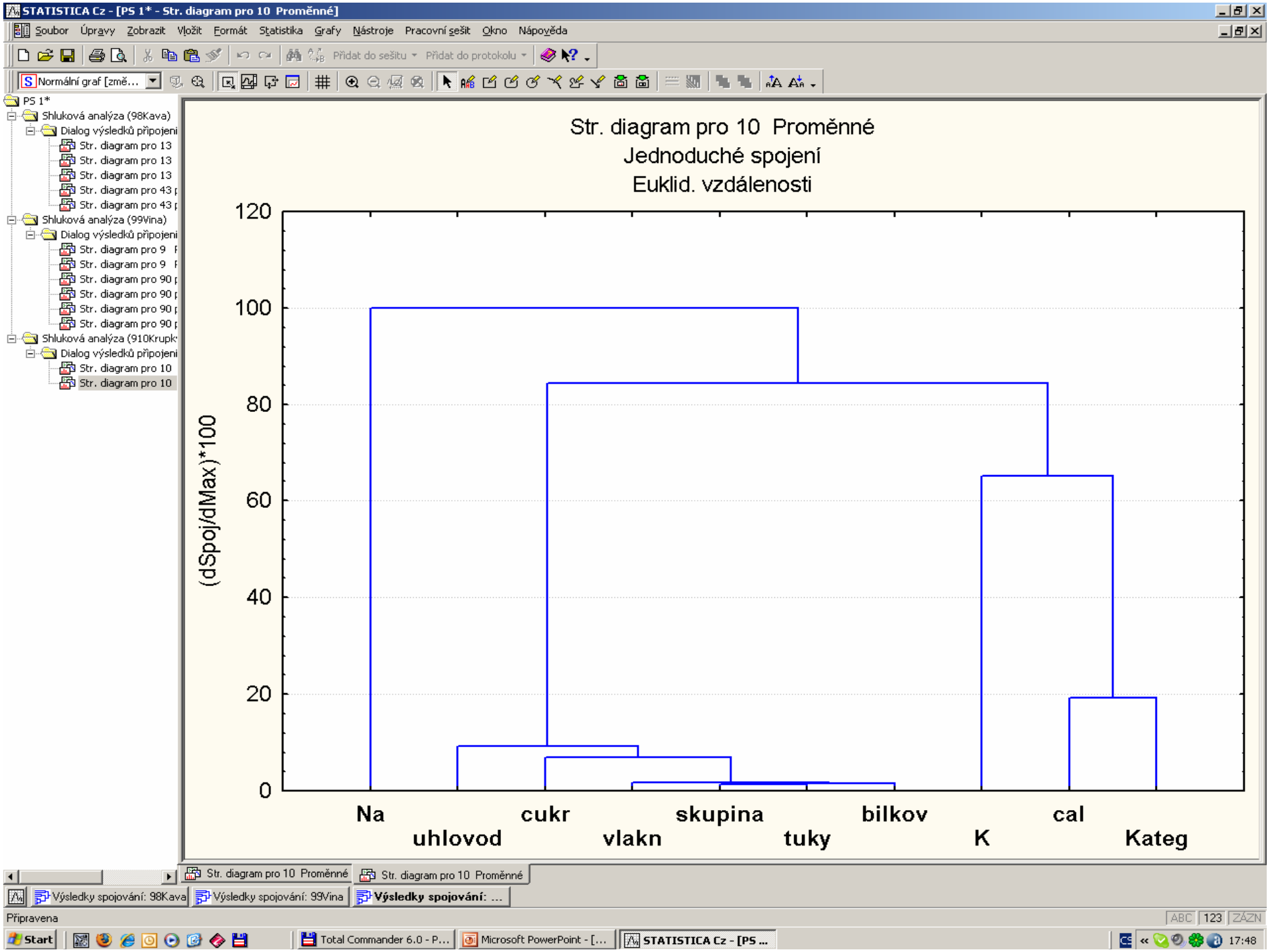


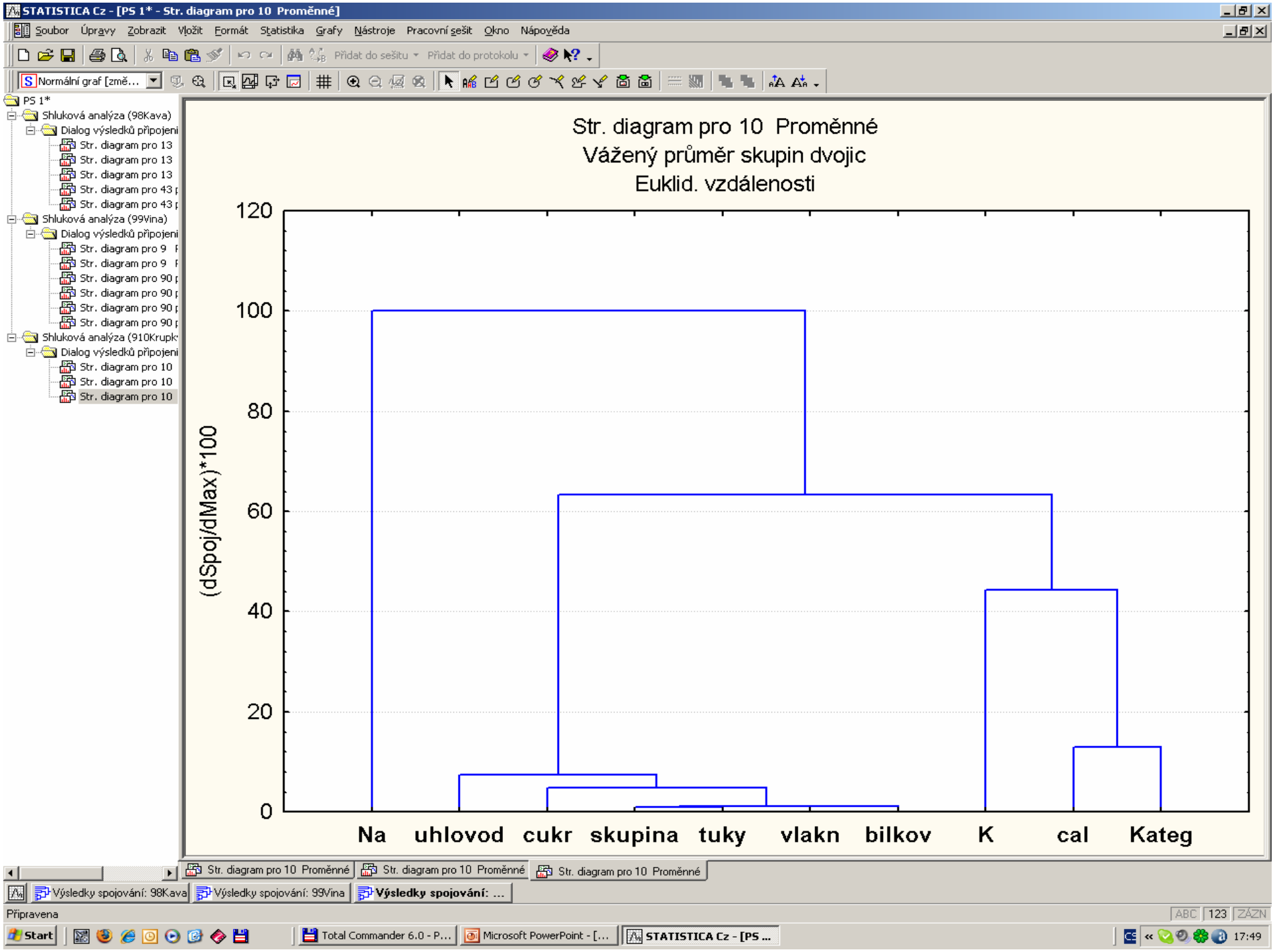
Dendrogram znaků matice dat *Krupky*, (STATISTICA).

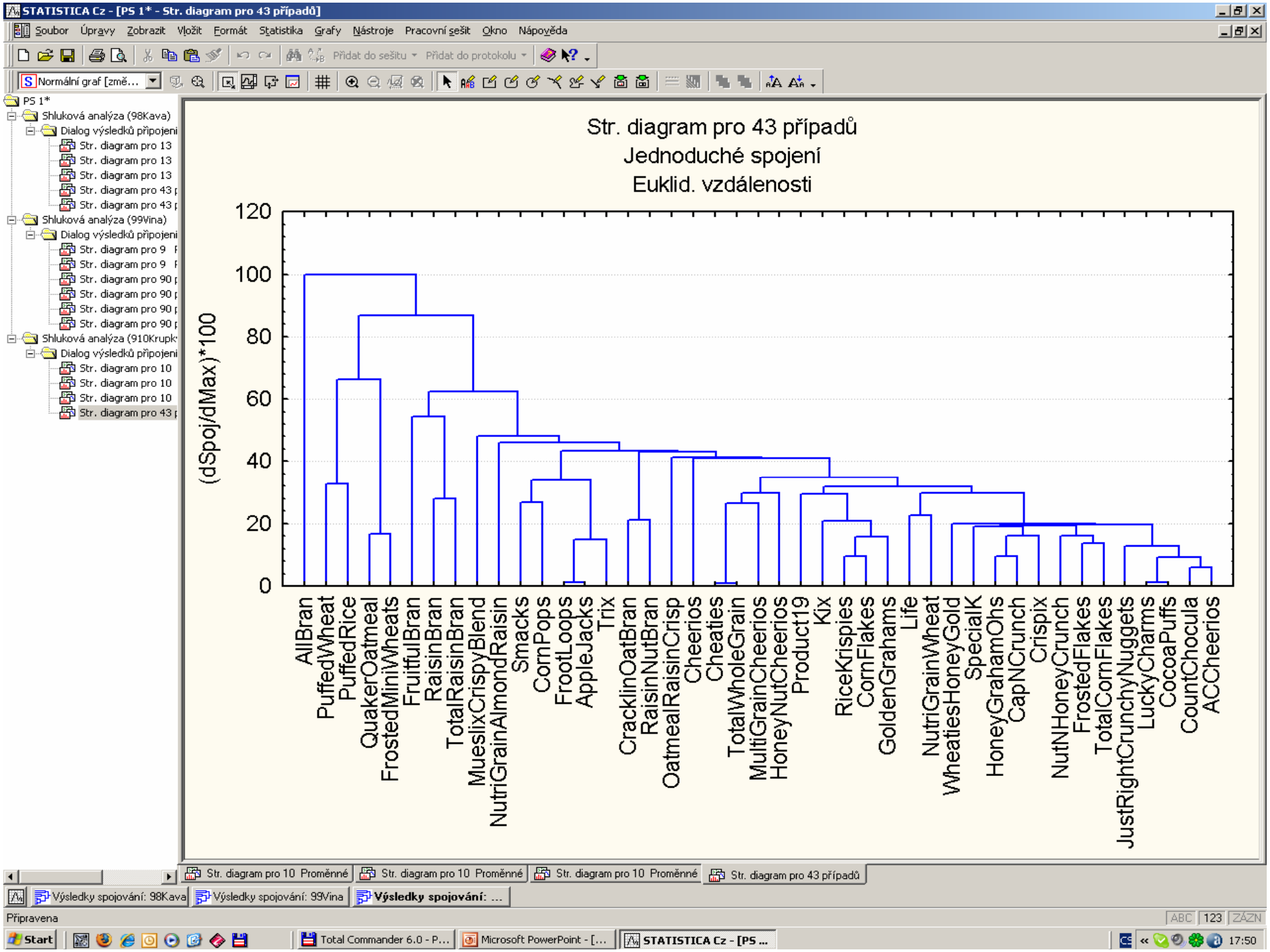


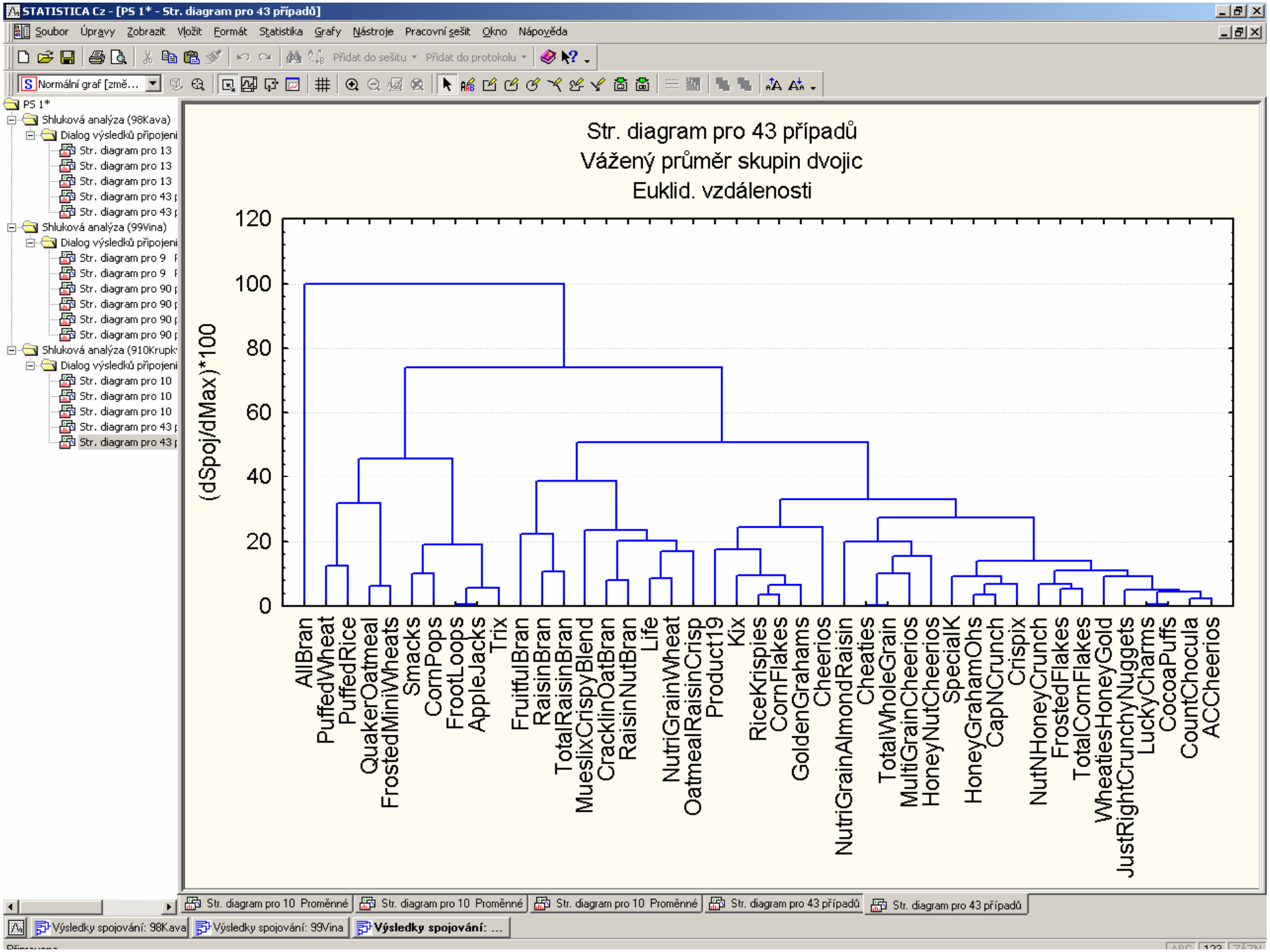
Dendrogram objektů matice dat *Krupky*, (STATISTICA).

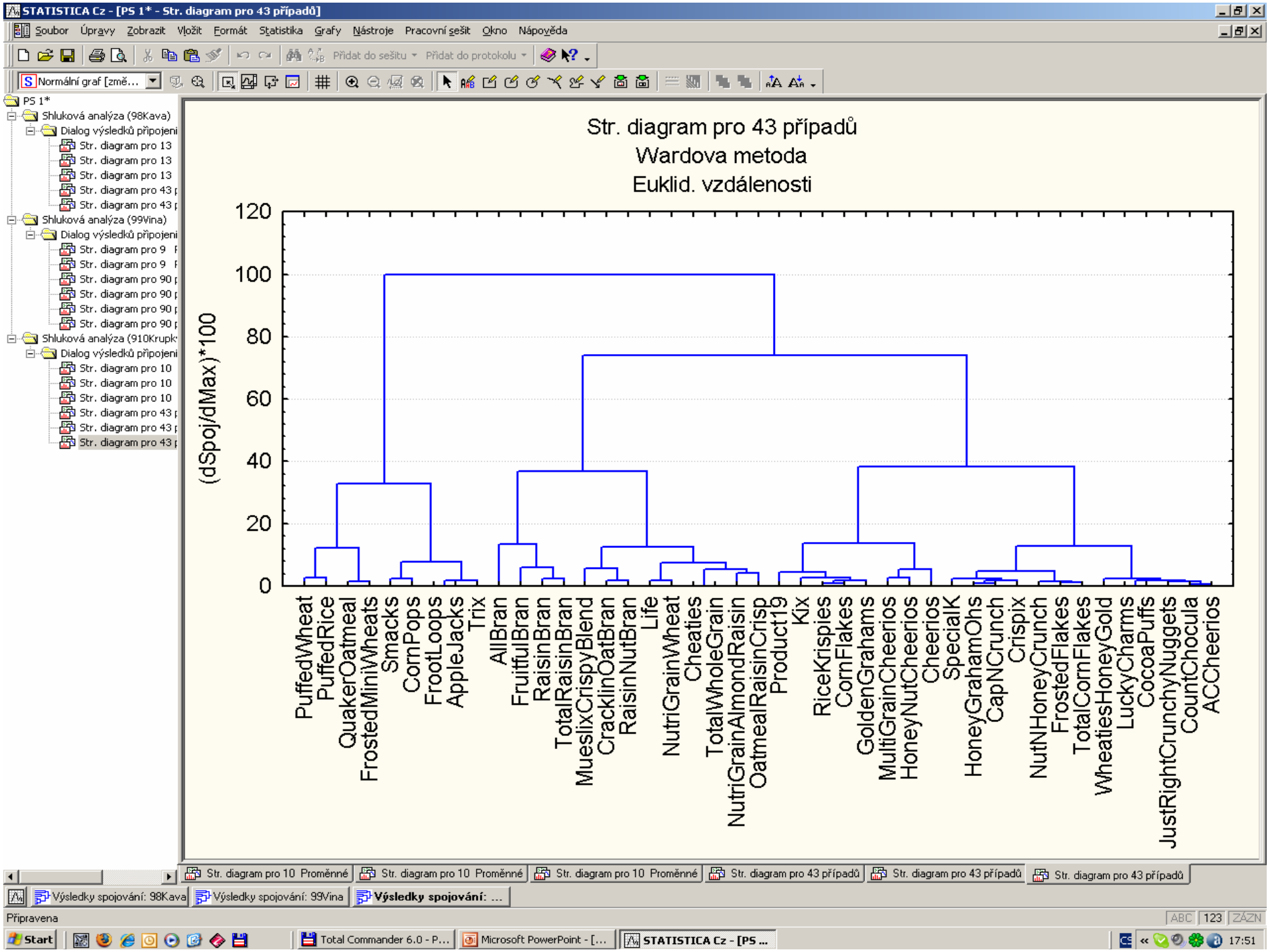
○ ***Závěr:*** Shlukováním metodou skupinového průměru se podařilo najít několik druhů křupavých lupínků, které jsou zcela nepodobné ostatním.


STATISTICA Cz - [PS 1* - Str. diagram pro 10 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10 Proměnné
Jednoduché spojení
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
Na
uhlovod
cukr
vlakn
skupina
tuky
bilkov
K
cal
Kateg
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:48


STATISTICA Cz - [PS 1* - Str. diagram pro 10 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10 Proměnné
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
Na
uhlovod
cukr
skupina
tuky
vlakn
bilkov
K
cal
Kateg
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:49


STATISTICA Cz - [PS 1* - Str. diagram pro 43 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43 případů
Jednoduché spojení
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
AllBran
PuffedWheat
PuffedRice
QuakerOatmeal
FrostedMiniWheats
FruitfulBran
RaisinBran
TotalRaisinBran
MueslixCrispyBlend
NutriGrainAlmondRaisin
Smacks
CornPops
FrootLoops
AppleJacks
Trix
CracklinOatBran
RaisinNutBran
OatmealRaisinCrisp
Cheerios
Cheaties
TotalWholeGrain
MultiGrainCheerios
HoneyNutCheerios
Product19
Kix
RiceKrispies
CornFlakes
GoldenGrahams
Life
NutriGrainWheat
WheatiesHoneyGold
SpecialK
HoneyGrahamOhs
CapNCrunch
Crispix
NutNHoneyCrunch
FrostedFlakes
TotalCornFlakes
JustRightCrunchyNuggets
LuckyCharms
CocoaPuffs
CountChocula
ACCheerios
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Str. diagram pro 43 případů
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:50


STATISTICA Cz - [PS 1* - Str. diagram pro 43 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojeni
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojeni
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojeni
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43 případů
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
AllBran
PuffedWheat
PuffedRice
QuakerOatmeal
FrostedMiniWheats
Smacks
CornPops
FrootLoops
AppleJacks
Trix
FruitfulBran
RaisinBran
TotalRaisinBran
MueslixCrispyBlend
CracklinOatBran
RaisinNutBran
Life
NutriGrainWheat
OatmealRaisinCrisp
Product19
Kix
RiceKrispies
CornFlakes
GoldenGrahams
Cheerios
NutriGrainAlmondRaisin
Cheaties
TotalWholeGrain
MultiGrainCheerios
HoneyNutCheerios
SpecialK
HoneyGrahamOhs
CapNCrunch
Crispix
NutNHoneyCrunch
FrostedFlakes
TotalCornFlakes
WheatiesHoneyGold
JustRightCrunchyNuggets
LuckyCharms
CocoaPuffs
CountChocula
ACCheerios
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Str. diagram pro 10 Proměnné
Str. diagram pro 43 případů
Str. diagram pro 43 případů
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: ...


STATISTICA Cz - [PS 1* - Str. diagram pro 43 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43 případů
Str. diagram pro 43 případů
Str. diagram pro 43 případů
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: ...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:51
Str. diagram pro 43 případů
Wardova metoda
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
PuffedWheat
PuffedRice
QuakerOatmeal
FrostedMiniWheats
Smacks
CornPops
FrootLoops
AppleJacks
Trix
AllBran
FruitfulBran
RaisinBran
TotalRaisinBran
MueslixCrispyBlend
CracklinOatBran
RaisinNutBran
Life
NutriGrainWheat
Cheaties
TotalWholeGrain
NutriGrainAlmondRaisin
OatmealRaisinCrisp
Product19
Kix
RiceKrispies
CornFlakes
GoldenGrahams
MultiGrainCheerios
HoneyNutCheerios
Cheerios
SpecialK
HoneyGrahamOhs
CapNCrunch
Crispix
NutNHoneyCrunch
FrostedFlakes
TotalCornFlakes
WheatiesHoneyGold
LuckyCharms
CocoaPuffs
JustRightCrunchyNuggets
CountChocula
ACCheerios

PŘÍKLAD 9.16 *Posouzení podobnosti kvality masa mladých býků dendrogramem*

U 76 býků mladších dvou let byly sledovány vlastnosti, determinující kvalitu masa. Podaří se v grafech nalézt tři skupiny plemene býků?

○ ***Data:*** Datová matice ***Byci*** se týká 76 býků a 9 znaků:

i je index býka,

Plemeno značí plemeno x_1 (1 značí Angus, 5 značí Hereford, 8 značí Simental),

Cena je prodejní cena x_2 [US $],

Vyska značí výšku dobytčete v kohoutku u prvním roce stáří x_3 [palce],

Hmotn značí hmotnost těla bez tuku x_4 [libry],

Maso značí procento hmoty masa bez tuku x_5 [%],

Velikost značí velikost býka ve stupnici 1 (malý) až 8 (velký) x_6,

Tuk značí tlouštku hřbetního tuku x_7 [palce],

Kohout značí výšku býka v kohoutku při prodeji x_8 [palce],

Hmotnost je hmotnost býka x_9 [libry].

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	1	2200	51	1128	70.9	7	0.25	54.8	1720
...	...	...	...	...	...	...	...	...	...
76	8	1500	51.7	992	70.6	7	0.15	55.1	1458

○ ***Řešení:*** **Graf komponentních vah znaků** odhaluje především korelaci 4 znaků *Hmotn, Vyska, Kohout, Velikost.*

Znaky *Plemeno* a *Maso* také značně korelují.

Dendrogram znaků ukazuje na dva shluky.

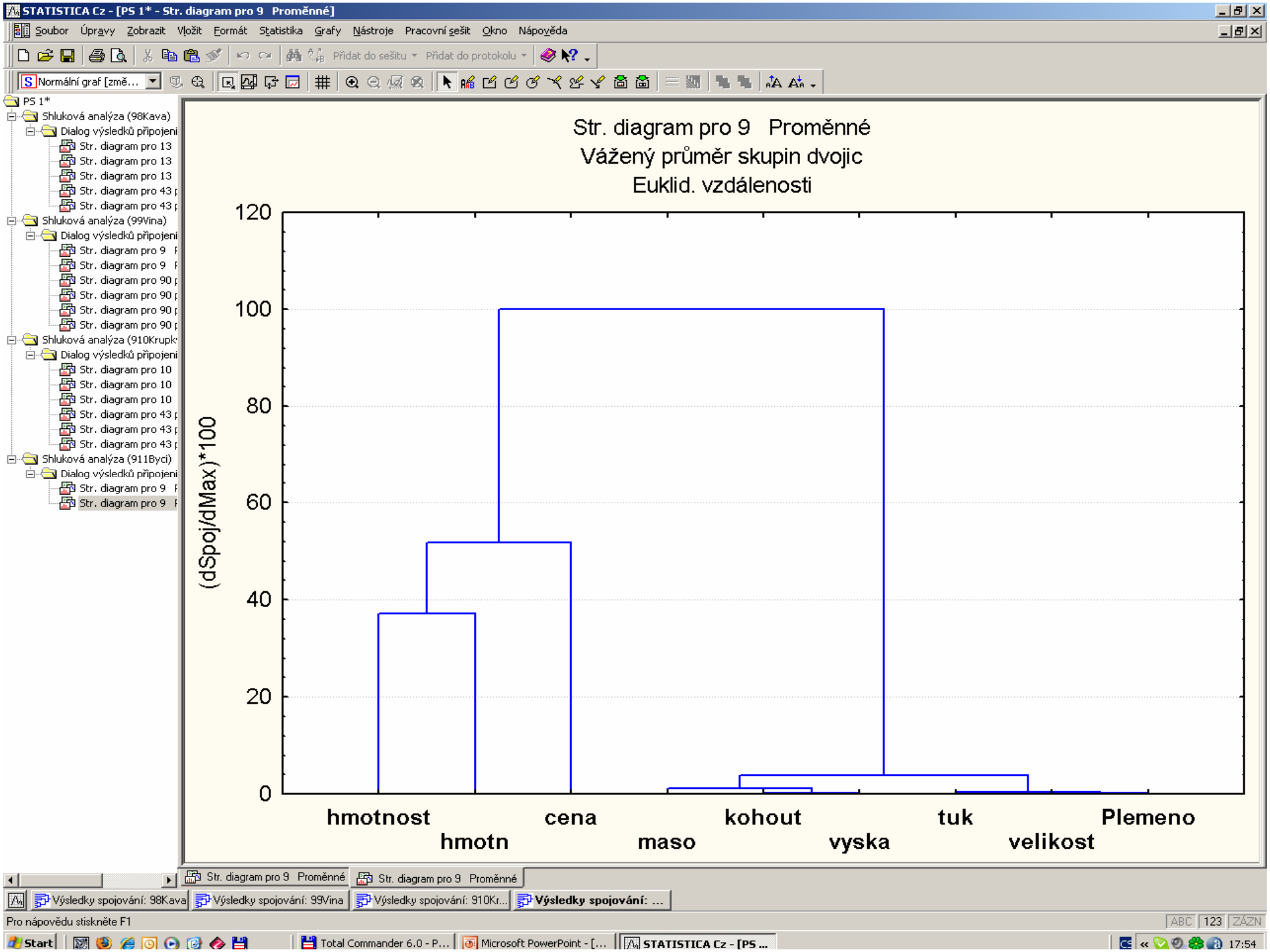
V prvním shluku je 6 znaků vzájemně velmi podobných *Plemeno, Velikost, Tuk, Vyska, Kohout* a *Maso*.

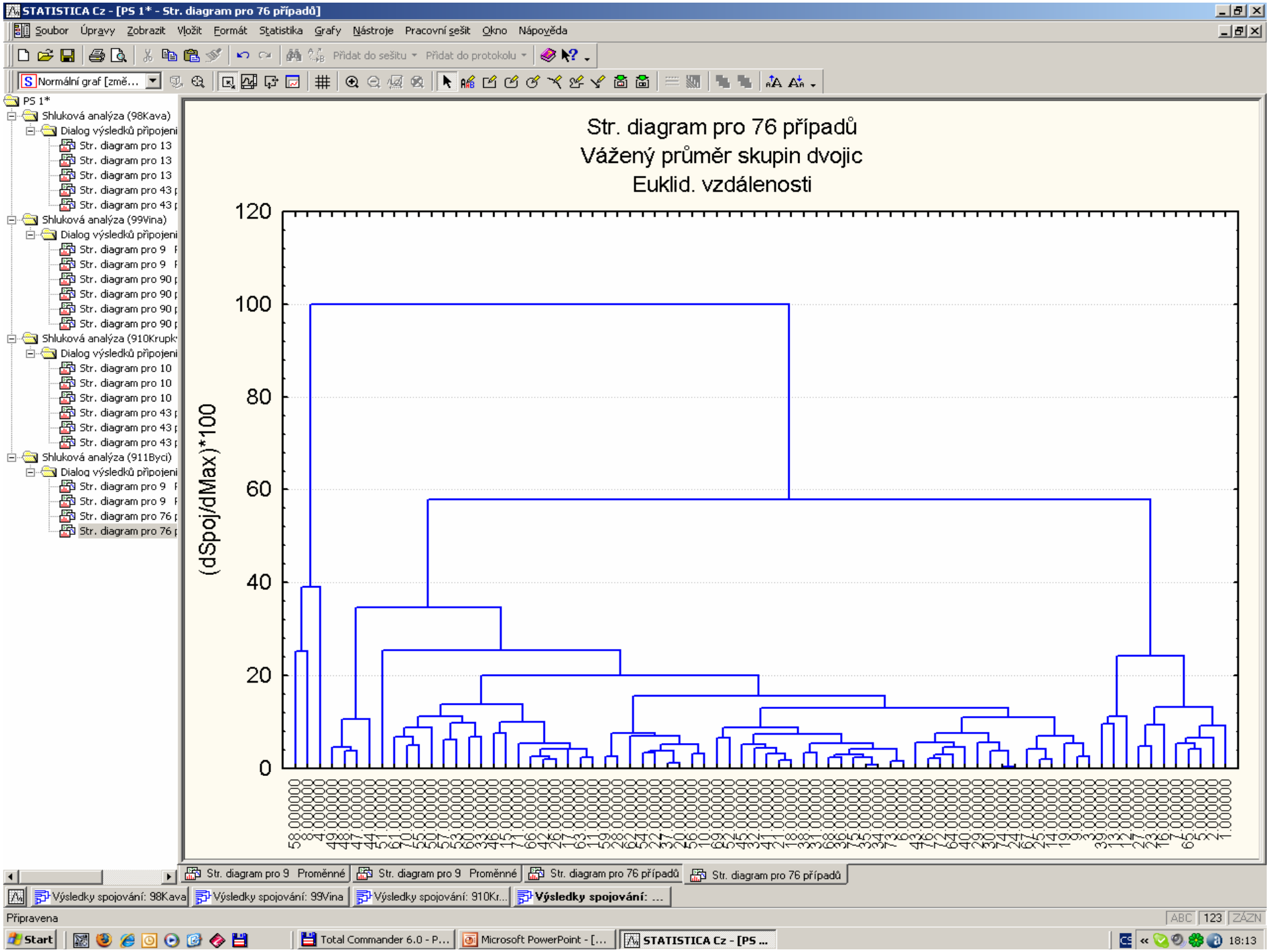
V druhém shluku jsou tři znaky vzájemně již méně podobné *Cena, Hmotn, Hmotnost.*

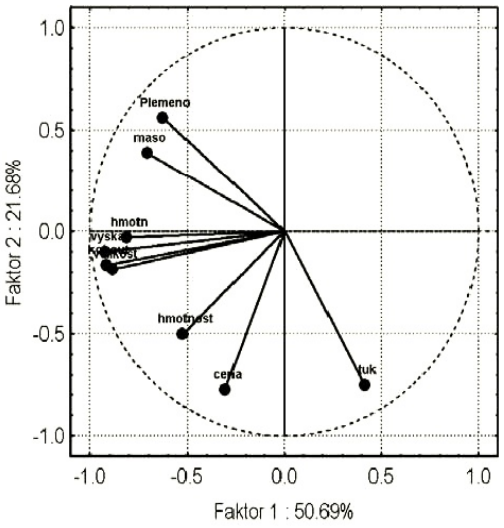
Graf komponentního skóre objektů ukazuje na dva až tři shluky a několik odlehlých objektů, málo podobných ostatním.

U hodnoty normovaného spojení eukleidovské vzdálenosti rovné 25 lze rozlišit dva větší shluky a dva menší shluky o 3 objektech.

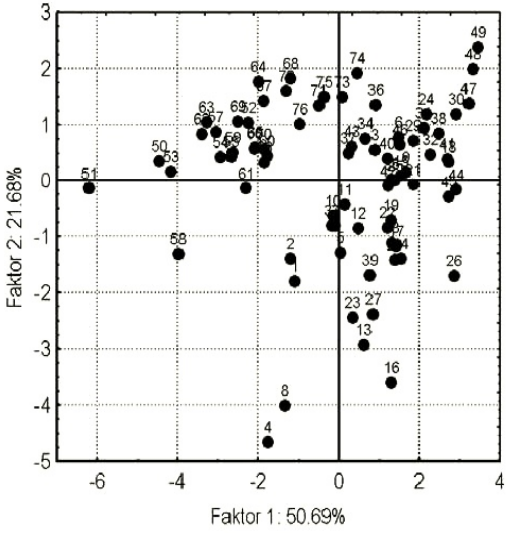
Oba menší shluky jsou málo podobné oběma větším shlukům.


STATISTICA Cz - [PS 1* - Str. diagram pro 9 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (911Byci)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 9 Proměnné
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
hmotnost
hmotn
cena
maso
kohout
vyska
tuk
velikost
Plemeno
Str. diagram pro 9 Proměnné
Str. diagram pro 9 Proměnné
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: 910Kr...
Výsledky spojování: ...
Pro nápovědu stiskněte F1
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
17:54

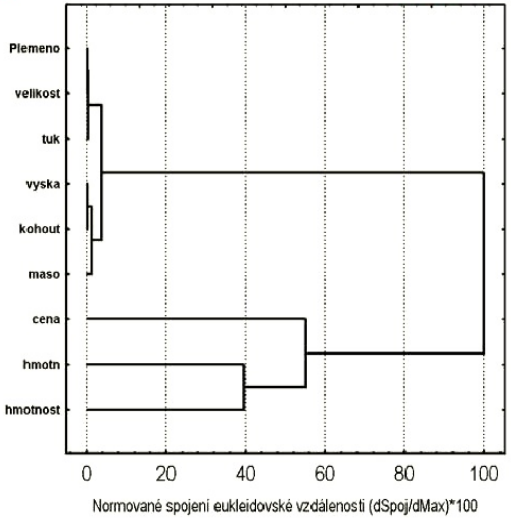

STATISTICA Cz - [PS 1* - Str. diagram pro 76 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojeni
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojeni
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojeni
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (911Byci)
Dialog výsledků připojeni
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 76
Str. diagram pro 76
Str. diagram pro 76 případů
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
Str. diagram pro 9 Proměnné
Str. diagram pro 9 Proměnné
Str. diagram pro 76 případů
Str. diagram pro 76 případů
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: 910Kr...
Výsledky spojování: ...
Připravena
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
18:13



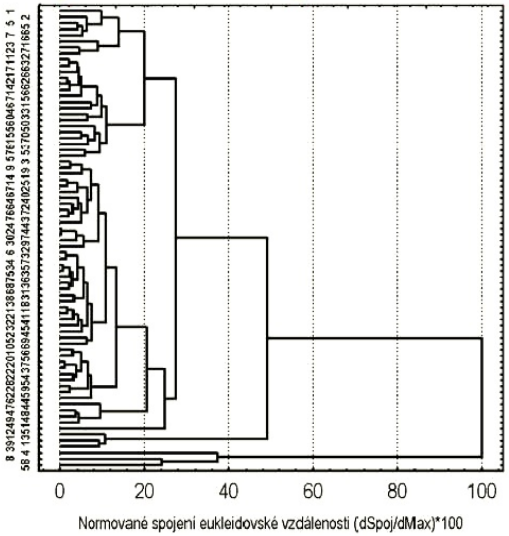
Graf komponentních vah znaků matice dat *Byci*, (STATISTICA).



Graf komponentního skóre objektů matice dat *Byci*



Dendrogram znaků matice dat *Byci*, (STATISTICA).



Dendrogram objektů matice dat *Byci*, (STATISTICA).

- ***Závěr:*** Shlukováním nalezly dominantní znaky, dle kterých se klasifikují býci do shluků. Ze 76 býků je přibližně 6 býků zcela odlišných od ostatních.

PŘÍKLAD 9.17 *Vytvoření dendrogramu traťových rekordů v lehké atletice mužů*

Byly zaznamenány národní traťové rekordy v lehké atletice mužů. Je třeba odhalit strukturu a skryté vazby mezi jednotlivými běžeckými disciplínami. Ve kterých zemích byly dosaženy podobné atletické výsledky?

○ ***Data*:** Datová matice ***Atlet*** se týká 56 zemí a 9 národních rekordů v běžeckých disciplinách:

i je index země,
100m značí běh na 100 m [s],
200m značí běh na 200 m [s],
400m značí běh na 400 m [s],
800m značí běh na 800 m [min],
1500m značí běh na 1500 m [min],
5km značí běh na 5000 m [min],
10km značí běh na 10000 m [min],
Maraton značí maraton [min],
Puvod značí zemi původu národního rekordu.

i	*100m*	*200m*	*400m*	*800m*	*1500m*	*5km*	*10km*	*Maraton*	*Puvod*
1	10.39	20.81	46.84	1.81	3.7	14.04	29.36	137.72	argentin
...	...	...	...	...	...				
55	10.82	21.86	49	2.02	4.24	16.28	34.71	161.83	wsamoa

○ ***Řešení*:** **Graf komponentních vah znaků** matice dat *Atlet* ukazuje, že první hlavní komponenta popisuje 83% a druhá 11% celkové proměnlivosti v datech.

Blízké průvodiče znaků ukazují na silnou korelaci těchto znaků.

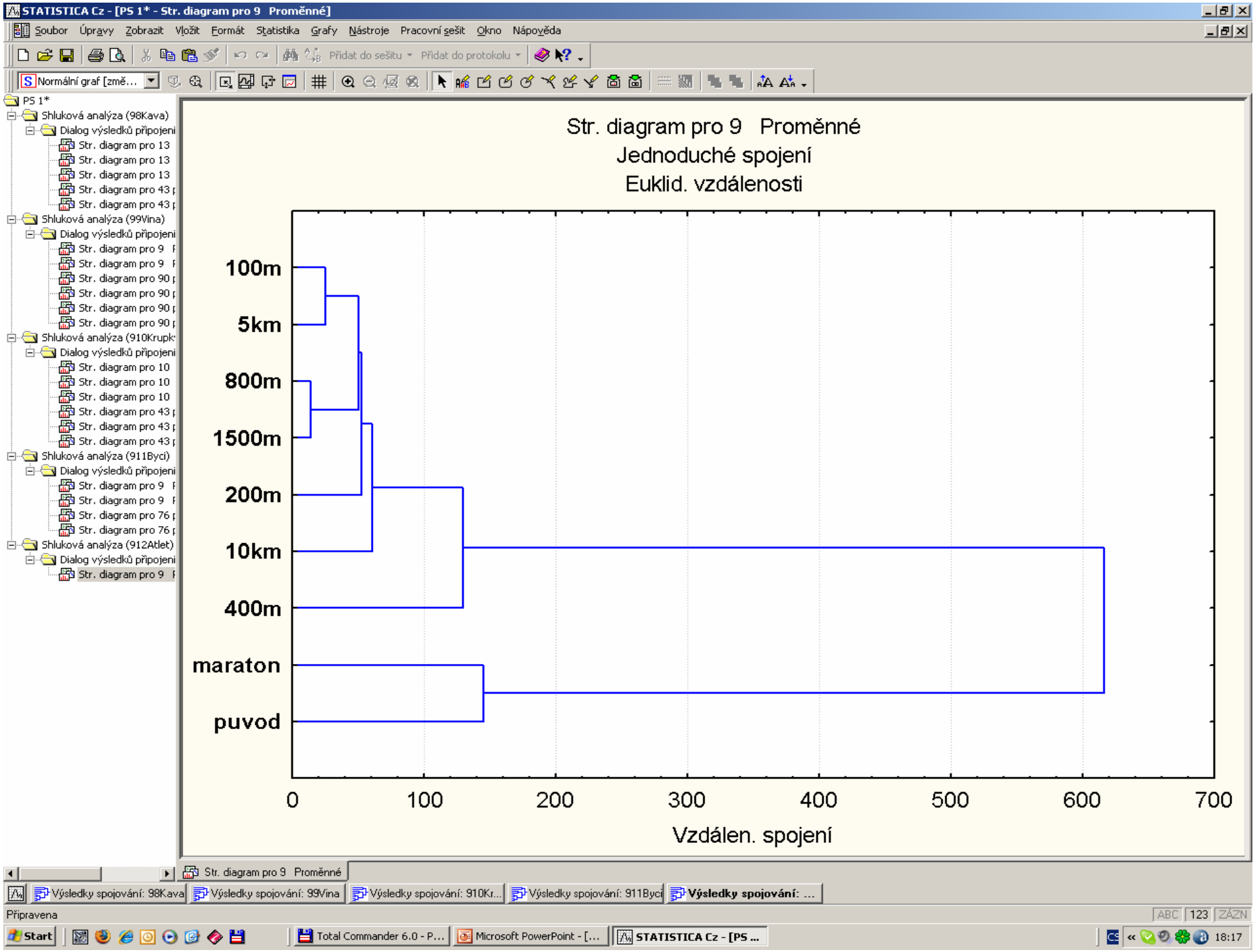
V dendrogramu znaků jsou zřejmé podobné znaky.

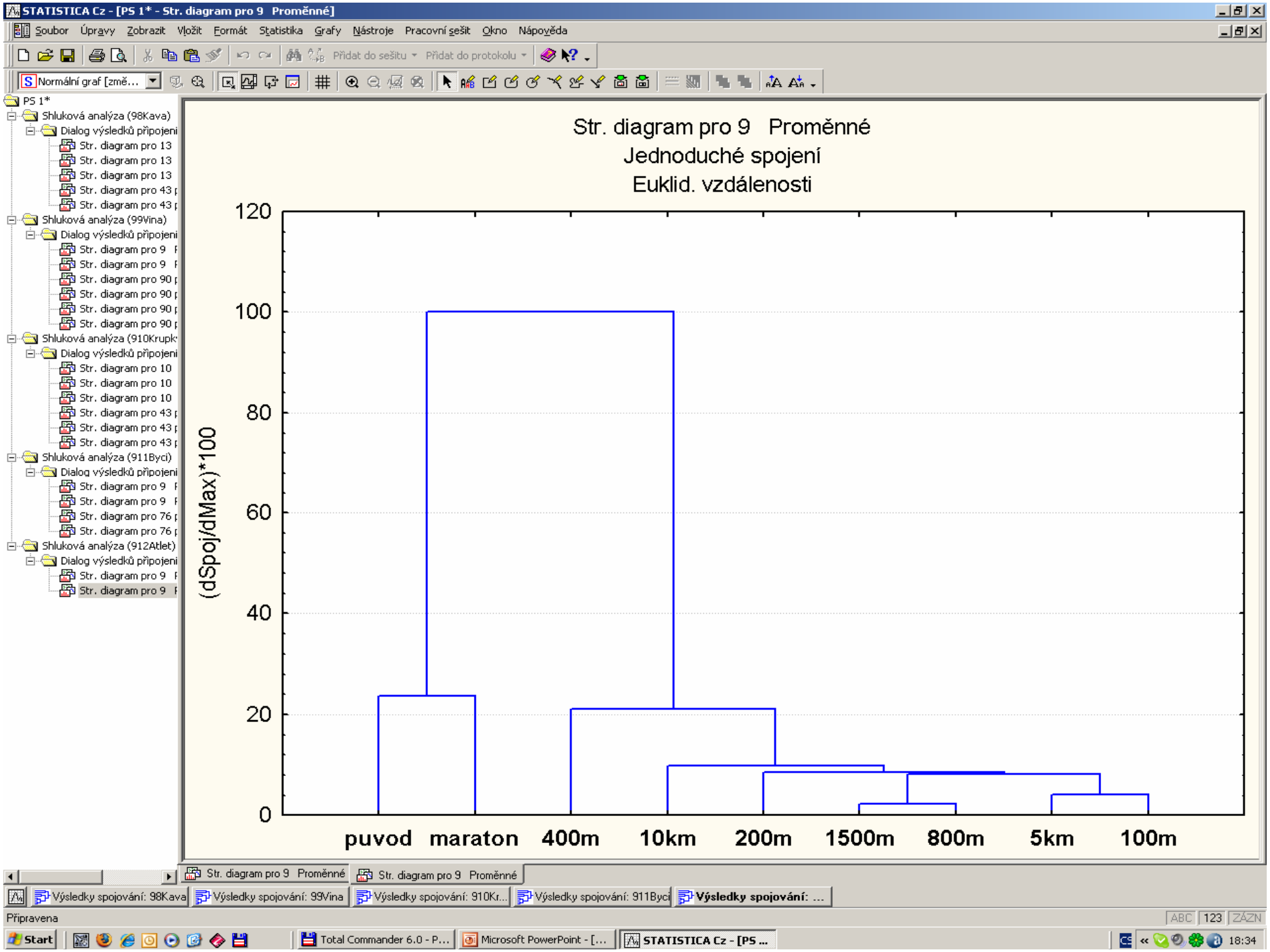
Výjimečného a značně nepodobného postavení vůči ostatním má v grafu znak *maraton.*

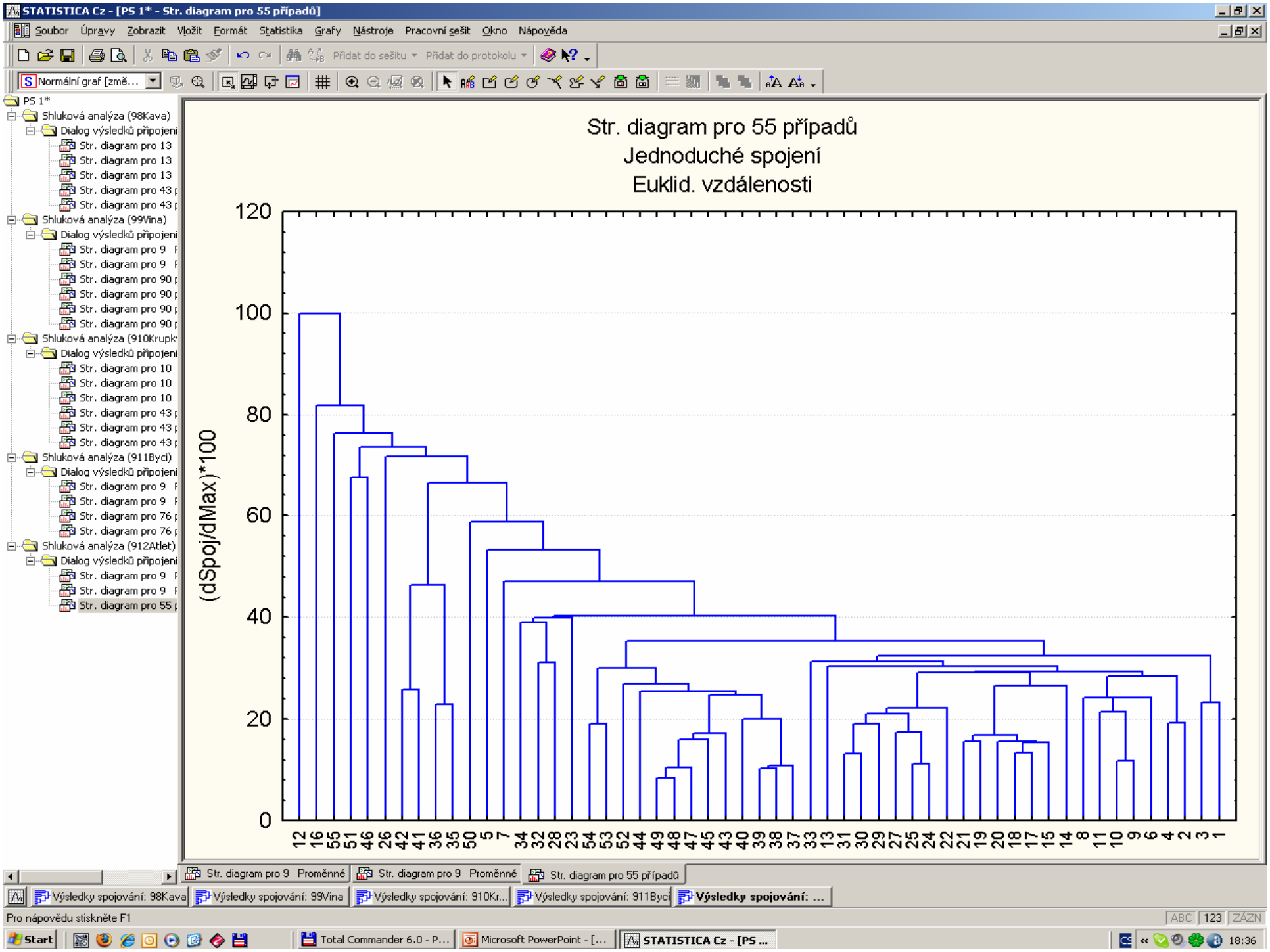
V grafu komponentního skóre objektů existuje jeden veliký shluk zemí, ve kterých byly dosaženy stejné národní běžecké rekordy.

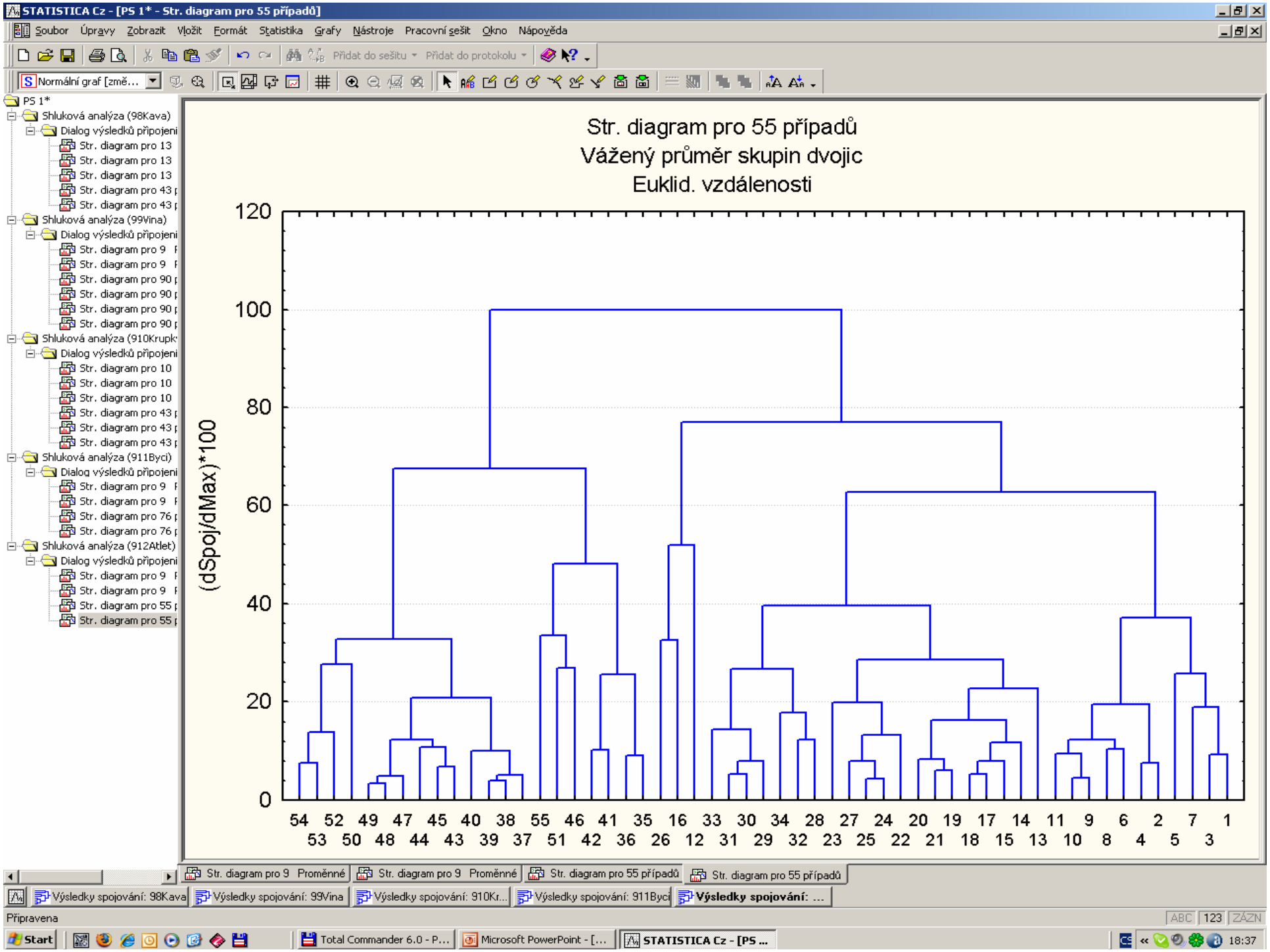
Existuje však **několik vybočujících objektů** zde zemí, které se silně odlišují svými atletickými výkony od ostatních zemí.

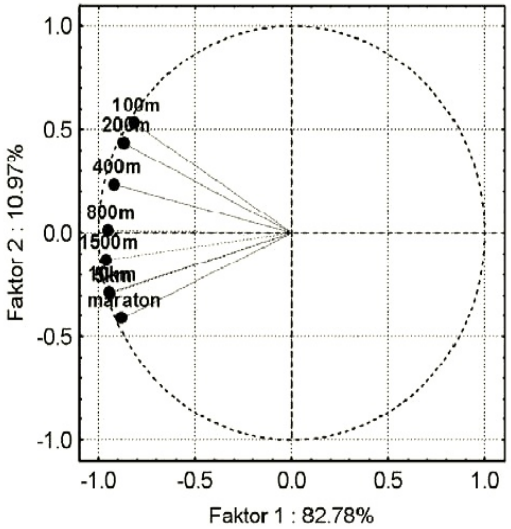
Tyto **odlehlé země** lze rovněž identifikovat i ve spodní části dendrogramu objektů.


STATISTICA Cz - [PS 1* - Str. diagram pro 9 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (911Byci)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 76
Str. diagram pro 76
Shluková analýza (912Atlet)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9 Proměnné
Jednoduché spojení
Euklid. vzdálenosti
100m
5km
800m
1500m
200m
10km
400m
maraton
puvod
0
100
200
300
400
500
600
700
Vzdálen. spojení
Str. diagram pro 9 Proměnné
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: 910Kr...
Výsledky spojování: 911Byci
Výsledky spojování: ...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
18:17

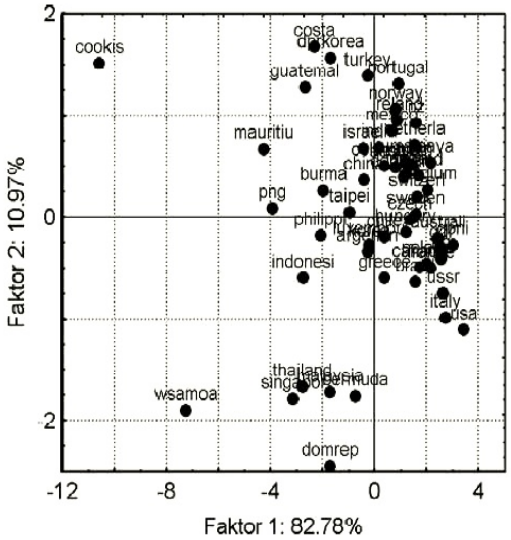

STATISTICA Cz - [PS 1* - Str. diagram pro 9 Proměnné]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Str. diagram pro 9 Proměnné
Jednoduché spojení
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
puvod
maraton
400m
10km
200m
1500m
800m
5km
100m
Připravena


STATISTICA Cz - [PS 1* - Str. diagram pro 55 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (911Byci)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 76
Str. diagram pro 76
Shluková analýza (912Atlet)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 55
Str. diagram pro 55 případů
Jednoduché spojení
Euklid. vzdálenosti
(dSpoj/dMax)*100
120
100
80
60
40
20
0
12 16 55 51 46 26 42 41 36 35 50 5 7 34 32 28 23 54 53 52 44 49 48 47 45 43 40 39 38 37 33 13 31 30 29 27 25 24 22 21 19 20 18 17 15 14 8 11 10 9 6 4 2 3 1
Str. diagram pro 9 Proměnné
Str. diagram pro 9 Proměnné
Str. diagram pro 55 případů
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: 910Kr...
Výsledky spojování: 911Byci
Výsledky spojování: ...
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
18:36

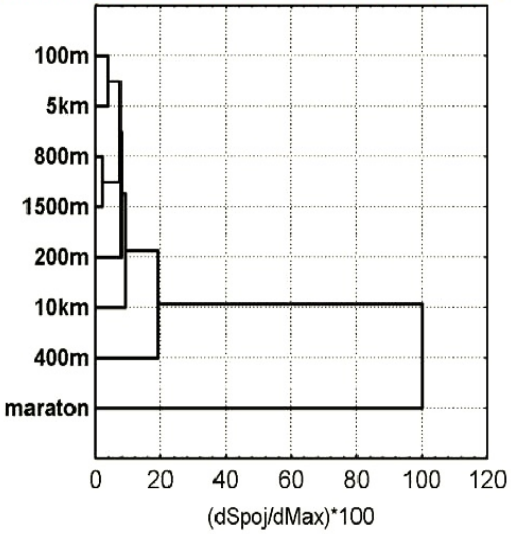

STATISTICA Cz - [PS 1* - Str. diagram pro 55 případů]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Shluková analýza (98Kava)
Dialog výsledků připojení
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 13
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (99Vina)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Str. diagram pro 90
Shluková analýza (910Krupk
Dialog výsledků připojení
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 10
Str. diagram pro 43
Str. diagram pro 43
Str. diagram pro 43
Shluková analýza (911Byci)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 76
Str. diagram pro 76
Shluková analýza (912Atlet)
Dialog výsledků připojení
Str. diagram pro 9
Str. diagram pro 9
Str. diagram pro 55
Str. diagram pro 55
Str. diagram pro 55 případů
Vážený průměr skupin dvojic
Euklid. vzdálenosti
(dSpoj/dMax)*100
0
20
40
60
80
100
120
54 52 49 47 45 40 38 55 46 41 35 16 33 30 34 28 27 24 20 19 17 14 11 9 6 2 7 1
53 50 48 44 43 39 37 51 42 36 26 12 31 29 32 23 25 22 21 18 15 13 10 8 4 5 3
Str. diagram pro 9 Proměnné
Str. diagram pro 9 Proměnné
Str. diagram pro 55 případů
Str. diagram pro 55 případů
Výsledky spojování: 98Kava
Výsledky spojování: 99Vina
Výsledky spojování: 910Kr...
Výsledky spojování: 911Byci
Výsledky spojování: ...
Připravena
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...
18:37



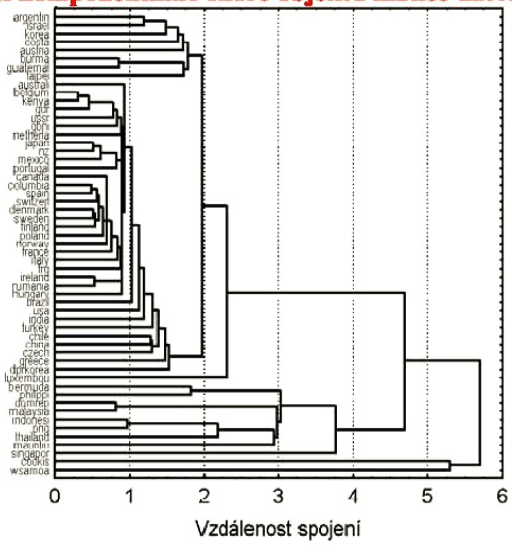
Graf komponentních vah znaků matice dat *Atlet*, (STATISTICA).



Graf komponentního skóre objektů matice dat *Atlet*



Dendrogram proměnných, znaků matice dat *Atlet*



Dendrogram objektů matice dat *Atlet*, (STATISTICA).

○ ***Závěr:*** Jako odlehlé objekty jsou detekovány země, ve kterých bylo dosaženo odlišných národních rekordů v běžeckých disciplinách než v ostatních zemích.

Doporučená literatura:

[1] Siotani M., Hayakawa T., Fujikoshi Y.: **Modern Multivariate Statistical Analysis**, A Graduate Course and Handbook. American Science Press, Columbia 1985.

[2] Johnson R.A., Wichern D.W.: **Applied Multivariate Statistical Analysis**, Prentice Hall, 1998.

[3] Meloun M., Militký J. , Forina M.: **Chemometrics for Analytical Chemistry,** Volume 1. **PC-Aided Statistical Data Analysis**, Ellis Horwood, Chichester 1992.

[4] Brereton R. G. **Multivariate Pattern Recognition in Chemometrics**, Illustrated by Case Studies, Elsevier 1992.

[5] Krzanowski W. J.: **Principles of Multivariate Analysis**, A User's Perspective, Oxford Science Publications, 1988.

[6] Meloun M. , Militký J.: **Statistické zpracování experimentálních dat**, Plus Praha 1994, Academia Praha 2004 (v tisku).

[7] Everitt B. S., Dunn G.: **Applied Multivariate Data Analysis**, Arnold, London 2001.

[8] Meloun M. , Militký J.: **Kompendium statistického zpracování dat**, Academia Praha 2002.

Je třeba studovat a na sobě pracovat!

Intenzivní týdenní kurzy (35 hodin) a

2-leté Licenční studium (280 hodin)

na Univerzitě v Pardubicích

Sledujte **http://meloun.upce.cz**

Milan Meloun - Mozilla Firefox

Soubor Úpravy Zobrazit Historie Záložky Nástroje Nápověda

Zpět Vpřed Obnovit Zastavit Domů http://meloun.upce.cz/ Google

Centrála Firefoxu Přehled zpráv

Home Personal Photo Gallery Research Lectures Papers Algorithms Data Sets WWW Visits

MILAN MELOUN

Prof. RNDr. Milan Meloun, DrSc.

professor of analytical chemistry and chemometrics

Address: Department of Analytical Chemistry
University of Pardubice
Čs. legií 565
532 10 Pardubice
Czech republic

Telephone: + 420-46 603 7026
Fax: + 420-46 603 7068

E-mail: milan.meloun@upce.cz

Naše doporučené knihy:

Základní učebnice, 2004

Nová učebnice vícerozměrných statistik, 2005

Kompendium úloh 2002 a **Nové vydání 2006**

Formy dalšího vzdělávání:

Intenzivní kurz v červnu 2007

Nové 12. licenční studium od září 2007

11. licenční studium od ledna 2006

Sledujte http://meloun.upce.cz

Hotovo

Start Doručená pošta... Total Command... meloun@melou... Milan Meloun ... LSWWW99 - Wi... Microsoft Power... 9:11

Milan Meloun - Mozilla Firefox

Soubor Úpravy Zobrazit Historie Záložky Nástroje Nápověda

Zpět Vpřed Obnovit Zastavit Domů http://meloun.upce.cz/ Google

Centrála Firefoxu Přehled zpráv

Home Personal Photo Gallery Research Lectures Papers Algorithms Data Sets WWW Visits

Univerzita Pardubice, Fakulta chemicko-technologická

nabízí týdenní intenzivní kurz na téma

Statistické zpracování dat

1. Týdenní kurz počítačové analýzy dat

Po hodině teorie následuje vždy hodina praktických úloh na počítači.

Kurz je s novou učebnicí Kompendium .....+ CD.!

Termín nejbližšího kurzu 3330/PK370002/81: **4. - 8. června 2007**

Zaplatit na číslo účtu: 37030561/0100, variabilní symbol: 3700023330

.

Termíny kurzů
Místo kurzu
Cena kurzu
Organizátor kurzu
Popis kurzu
Organizační informace
Sylabus kurzu
Obsah úloh
Objednávka kurzu

Termín: Každoročně vždy v měsících únoru, červnu a září (dotaz se na termín nejbližšího kurzu

Sledujte http://meloun.upce.cz

Milan Meloun - Mozilla Firefox

Soubor Úpravy Zobrazit Historie Záložky Nástroje Nápověda

Zpět Uložit (Ctrl+S) Obnovit Zastavit Domů http://meloun.upce.cz/ Google

Centrála Firefoxu Přehled zpráv

Home Personal Photo Gallery Research Lectures Papers Algorithms Data Sets WWW Visits

UNIVERZITA PARDUBICE,

Fakulta chemicko-technologická, Katedra analytické chemie

Projekt dvouletého licenčního studia chemometrie na téma

Statistické zpracování dat

Vedoucí licenčního studia a odborný garant: **Prof. RNDr. Milan Meloun, DrSc.**
Katedra analytické chemie, Fakulta chemicko-technologická,
Univerzita Pardubice, nám. Čs. Legií 565, 53210 Pardubice,
IČO: 216 275, **DIČ:** 248-00216275
Telefon: 466037026, **Fax:** 466037068, **E-mail:** milan.meloun@upce.cz, **http:**//meloun.upce.cz

1. soustředění nového 12. licenčního studia č. 3330/LS340005/45	**koncem měsíce září 2007**

6. soustředění nového 11. licenčního studia č. 3330/LS360001/45:	**28. května- 1. června 2007**

Přehled základních informací k licenčnímu studiu:

Charakter studia	Organizace studia	Vyučující (foto)	Plán studia	Sylaby předmětů	Přihláška do adresáře	Absolventi, Foto1, 2, Diplomky
Návody úloh	Otázky ke zkoušce	Vzory semestr. prací	Potřebná literatura	Potřebný software	Zákony Murphyho	Promoce licen. studia

Charakter studia

Přihlášky a smlouvu do léta 2007,

Zaměření licenčního studium **Statistické zpracování dat při managementu jakosti** je zaměřeno na zvyšování odborné úrovně pracovníků kontrolních laboratoří tak, aby znalosti, technika práce a především způsob zpracování výsledků chemických zkoušek, experimentálních dat byly srovnatelné s laboratořemi zemí Evropské unie.

Určení: licenční studium je určeno pro stávající i budoucí pracovníky kontrolních laboratoří OTK, OKŘJ, dále pracovníky zdravotnických, veterinárních, vodohospodářských laboratoří, potravinářské a zemědělské inspekce, chemických, potravinářských, farmaceutických a zemědělských výrob. Dále pro pracovníky laboratoří kontroly životního prostředí, všech odvětví průmyslu, energetiky a zemědělství s důrazem na využití moderní instrumentální techniky a především zpracování výsledků pomocí matematicko-statistických metod na osobním počítači a s využitím nejmodernějšího programového vybavení.

cena 37.500,- Kč.

Přihlášky do licenčního studia: mohou se přihlásit pracovníci, kteří buď již pracují v oboru anebo se v rámci rekvalifikace potřebují seznámit s

Děkuji za pozornost!

http://meloun.upce.cz