Faktorová analýza (FA)

1. Zaměření faktorové analýzy:

1) K vyšetření **vnitřních souvislostí** (korelace) a **odhalení struktury** datové matice.

2) Analýza struktury vnitřních vztahů mezi velkým počtem znaků, **zavedením** několika ***faktorů***.

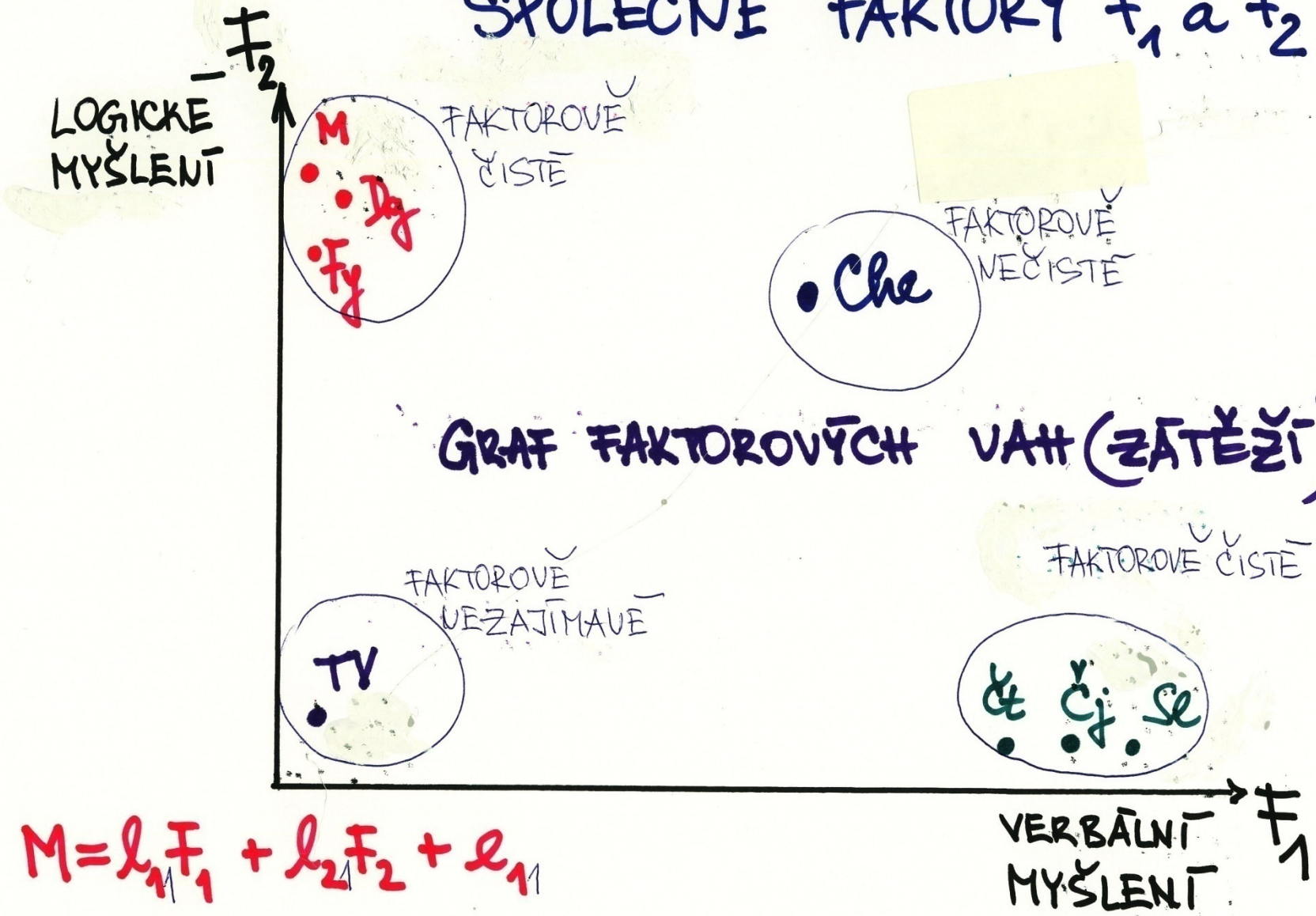
3) Po identifikaci faktorů je každému faktoru přidělen **obsahový význam,** pomocí kterého je každý znak vysvětlen.

4) **Dva primární cíle FA**:

   a) sumarizace a **zestručnění dat**

   b) **redukce dat**.

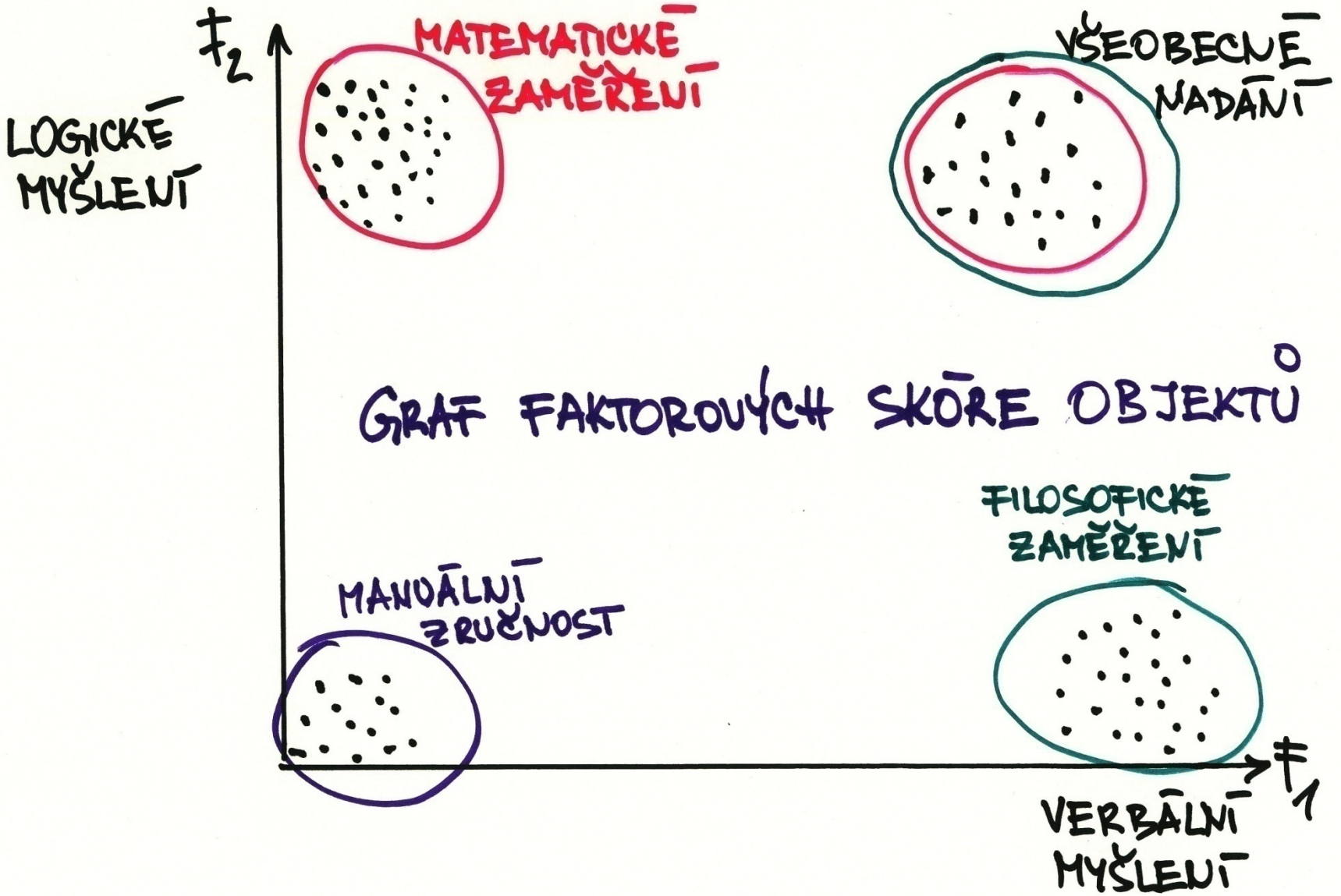
SPOLEČNÉ FAKTORY F_1 a F_2



$$M = l_{11}F_1 + l_{21}F_2 + e_{11}$$

$$Čj = l_{12}F_1 + l_{22}F_2 + e_2$$

SPECIFICKÉ FAKTORY e_1 a e_2
(CHYBOVÉ)


f_2
LOGICKÉ MYŠLENÍ
MATEMATICKÉ ZAMĚŘENÍ
VŠEOBECNÉ NADÁNÍ
GRAF FAKTOROVÝCH SKÓRE OBJEKTŮ
FILOSOFICKÉ ZAMĚŘENÍ
MANUÁLNÍ ZRUČNOST
f_1
VERBÁLNÍ MYŠLENÍ

2. Podstata faktorové analýzy

Matice standardizovaných dat $x_1 = (X_i - \bar{X}_i)/ s_i$ obsahuje P sloupců znaků $X_1, X_2, ..., X_P$ a N řádků objektů.

Standardizovaná data mají u znaků střední hodnotu 0 a rozptyl 1 a jejich kovariance jsou vlastně korelační koeficienty.

$\boldsymbol{x}^{\mathbf{T}} = (x_1, x_2, ..., x_P)^{\mathrm{T}}$ značí jeden i-tý objekt o P znacích s korelační maticí **R**.

Objekty představíme jako lineární kombinaci faktorů $F_1, F_2, ..., F_m$ plus specifický faktor e_i

Formou ***modelu faktorové analýzy (FA)*** ve tvaru

$$x_1 = l_{11}F_1 + l_{12}F_2 + \dots l_{1m}F_m + e_1 ,$$

$$x_2 = l_{21}F_1 + l_{22}F_2 + \dots l_{2m}F_m + e_2 ,$$

$$\dots\dots\dots\dots$$

$$x_P = l_{P1}F_1 + l_{P2}F_2 + \dots l_{Pm}F_m + e_P ,$$

kde ◆ F_1, F_2, ..., F_m jsou *vybrané společné faktory o počtu m*, které vyvolávají korelace mezi P původními znaky.

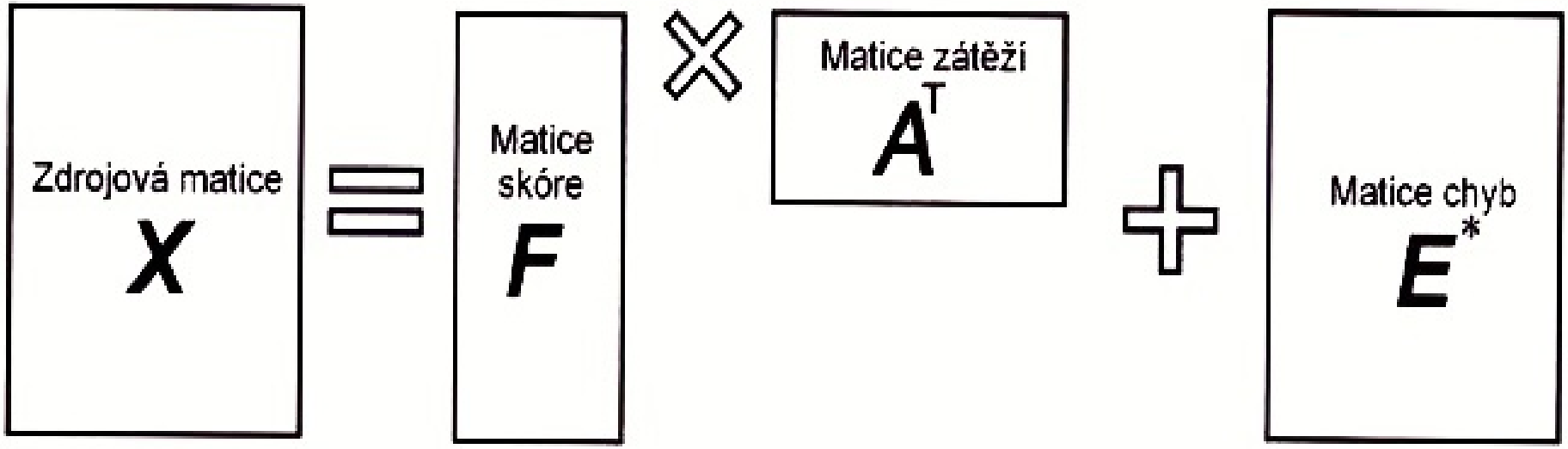
◆ Faktory mají nulovou střední hodnotu a jednotkový rozptyl.

◆ Faktory mají nulovou střední hodnotu a jednotkový rozptyl.

◆ $e_1, e_2, ..., e_P$ jsou *specifické (chybové) faktory*, které přispívají k rozptylu jednotlivých proměnných.

◆ Koeficienty l_{ij} nazýváme *faktorové zátěže* i-tého znaku na j-tém společném faktoru F_j a představují prvky matice faktorových zátěží.

◆ Výběr N objektů lze vyjádřit pouhým přidáním indexu ke každému x_i nebo F_j a e_i, a tak vystihnout každý objekt.



Faktorový model rozdělí:

jednak každý znak x_i na dvě části,

jednak rozptyl každého znaku na dvě části.

Rozptyl $s_i^2 = 1$ i-tého znaku $x_i = l_{i1}F_1 + l_{i2}F_2 + \dots l_{im}F_m + e_i$ je složen ze dvou částí:

(1) ***Komunalita*** h_i^2 je část rozptylu, který vznikl v důsledku použití společných faktorů, a

(2) ***Specificita*** u_i^2 je část rozptylu, který vznikl v důsledku použití specifického faktoru,

a platí vztah $$s_i^2 = h_i^2 + u_i^2 = 1.$$

Cílem FA je vyčíslit ***faktorové zátěže*** l_{ij} a také ***komunality*** h_i^2.

Nalezenému řešení se říká *počáteční faktorová extrakce.*

Původní znaky se staly znaky standardizovanými a faktorové zátěže l_{ij} vyjadřují korelaci mezi x_i a F_j.

Matice faktorových zátěží se také nazývá *matice faktorové struktury.*

Komunalita proměnné x_i je dána vztahem

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 .$$

Součet komunalit je roven kumulativnímu součtu rozptylu, objasněného společnými faktory.

Suma specifit je rovna rozdílu celkového rozptylu a komunalit, je to zbývající část rozptylu, která zůstala neobjasněna společnými faktory.

- *Komunalita h* vyjadřuje míru proměnlivosti a je vahou, s jakou jednotlivé faktory přispívají do rozptylu odpovídající původní proměnné.

- *Čtverec komunality* h^2 je suma faktorových zátěží faktorů.

- *Jedinečnost* Γ^2 lze dále rozdělit na část *specifity* Γ^2_s a část *nespolehlivosti* Γ^2_n.

  Specifita Γ^2_s představuje část proměnlivosti, kterou nelze vysvětlit ani chybou experimentu, ani společnými faktory.

  Nespolehlivost Γ^2_n představuje experimentální chybu při měření faktorů.

Způsob rozkladu proměnlivosti představuje základní hledisko pro klasifikaci metod faktorové analýzy.

3. Extrakce faktorů metodou PCA

Dvěmi numerickými metodami lze získat společné faktory:

Metoda hlavních komponent PCA: nalezení prvních m hlavních komponent a jejich modifikace do faktorového modelu.

Prvních m hlavních komponent

◆ zahrnuje největší podíl rozptylu ze všech znaků, proto nejdůležitější.

◆ je *nekorelovaných*, a proto atraktivní k využití za faktory.

◆ aby byl jednotkový rozptyl faktorů, vydělí se každá hlavní komponenta její směrodatnou odchylkou a definuje se společný faktor

$$F_j = C_j \,/\, \sqrt{s_j^2},$$

kde C_j je j-tá hlavní komponenta.

Pro první hlavní komponentu platí

$$C_1 = a_{11}x_1 + a_{12}x_2 + \dots a_{1P}x_P ,$$

pro druhou hlavní komponentu

$$C_2 = a_{21}x_1 + a_{22}x_2 + \dots a_{2P}x_P ,$$

a konečně pro P-tou komponentu

$$C_P = a_{P1}x_1 + a_{P2}x_2 + \dots a_{PP}x_P ,$$

lze invertovat tak, že x_i jsou funkcemi C_j.

s výsledkem

$$x_1 = a_{11}C_1 + a_{12}C_2 + \ldots + a_{P1}C_P ,$$

$$x_2 = a_{12}C_1 + a_{22}C_2 + \ldots + a_{P2}C_P ,$$

$$\ldots\ldots\ldots\ldots$$

$$x_P = a_{1P}C_1 + a_{2P}C_2 + \ldots + a_{PP}C_P .$$

Řádky v první řadě rovnic se stanou sloupci v následném bloku invertovaných rovnic.

Protože platí vztah $F_j = C_j / \sqrt{s^2(C_j)}$,

který upravíme dále na $C_j = F_j \sqrt{s^2(C_j)}$,

bude ve vztahu

$$x_1 = a_{1i} F_1 \sqrt{s^2(C_1)} + a_{2i} F_2 \sqrt{s^2(C_2)} + \ldots + a_{Pi} F_P \sqrt{s^2(C_P)}$$

u prvních m komponent platit $l_{ij} = a_{ji} \sqrt{s^2(C_j)}$

a u posledních (P - m) komponent označíme tyto výrazy za e_i dle

$$e_i = a_{m+1,i} F_{m+1} \sqrt{s^2(C_{m+1})} + \ldots + a_{P,i} F_P \sqrt{s^2(C_P)}$$

a dostaneme $x_i = l_{i1} F_1 + l_{i2} F_2 + \ldots l_{im} F_m + e_i$,

pro $i = 1, 2, \ldots, P$ proměnných.

Závěr: Tím jsme přetransformovali hlavní komponenty na faktory.

5. Faktorová rotace

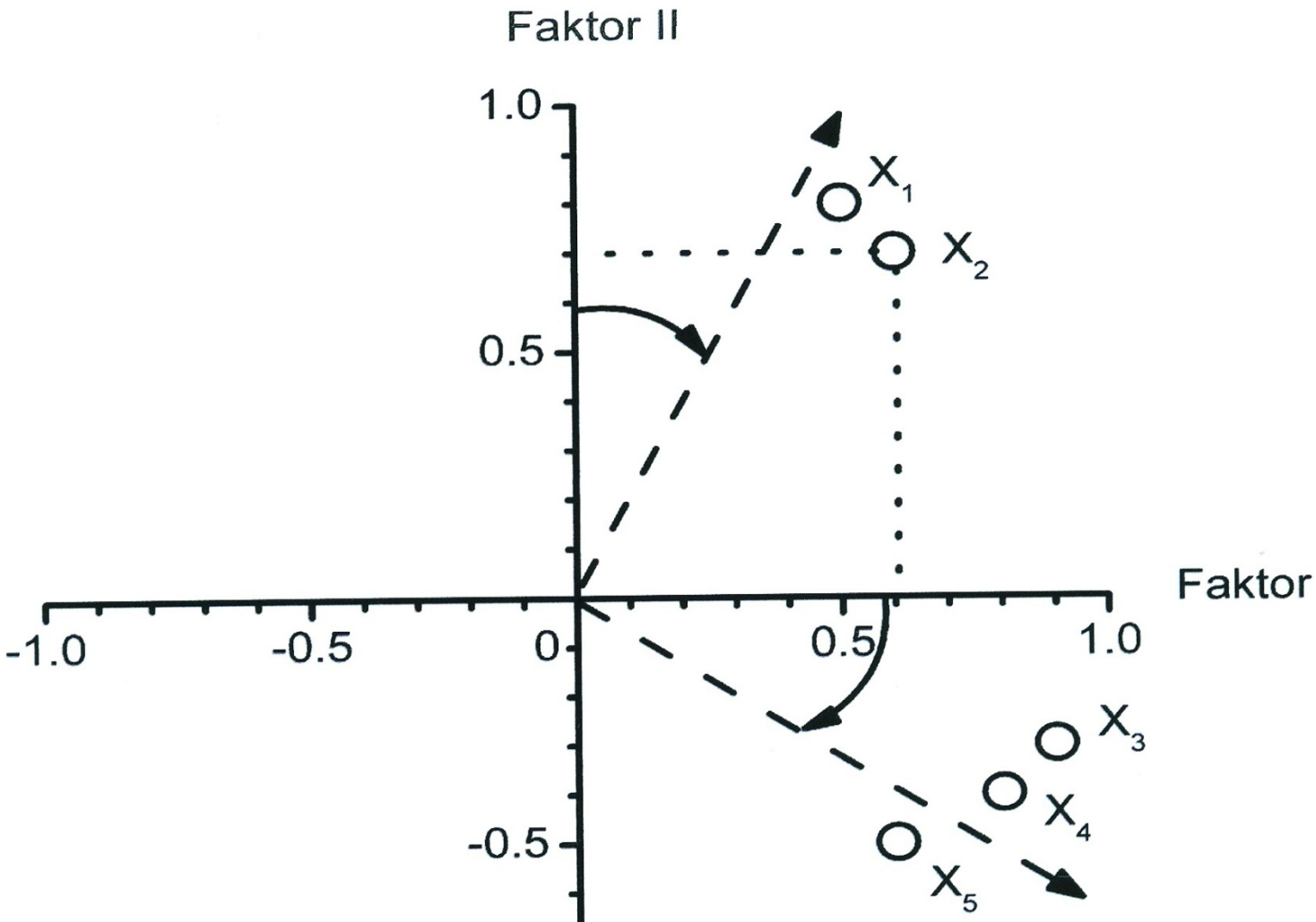
Hlavním cílem FA: odvodit z dat snadno vysvětlitelné a pojmenovatelné společné faktory.

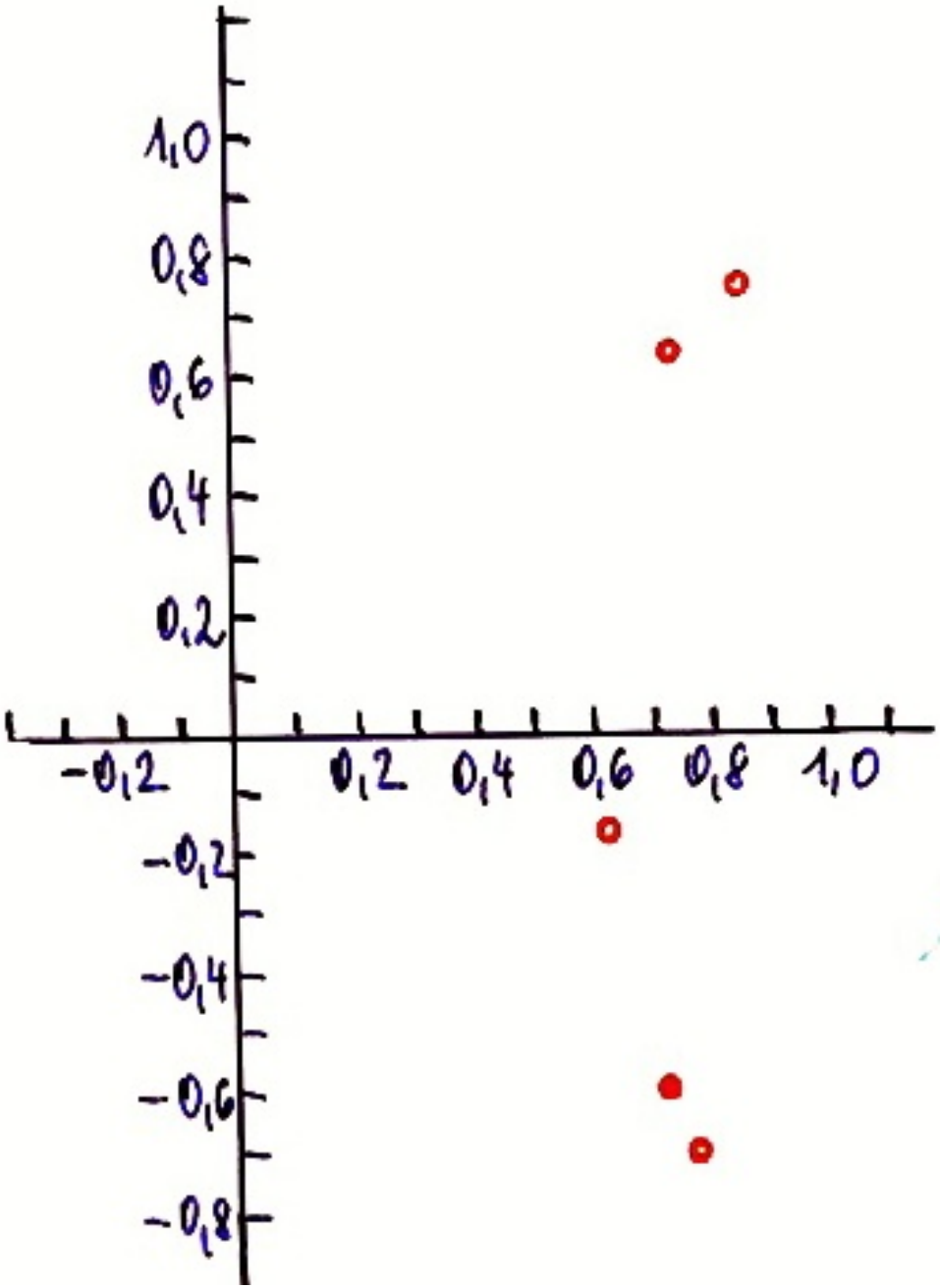
Počáteční odhady faktorů však bývají často obtížně vysvětlitelné, protože většina faktorů je korelována s mnoha znaky.

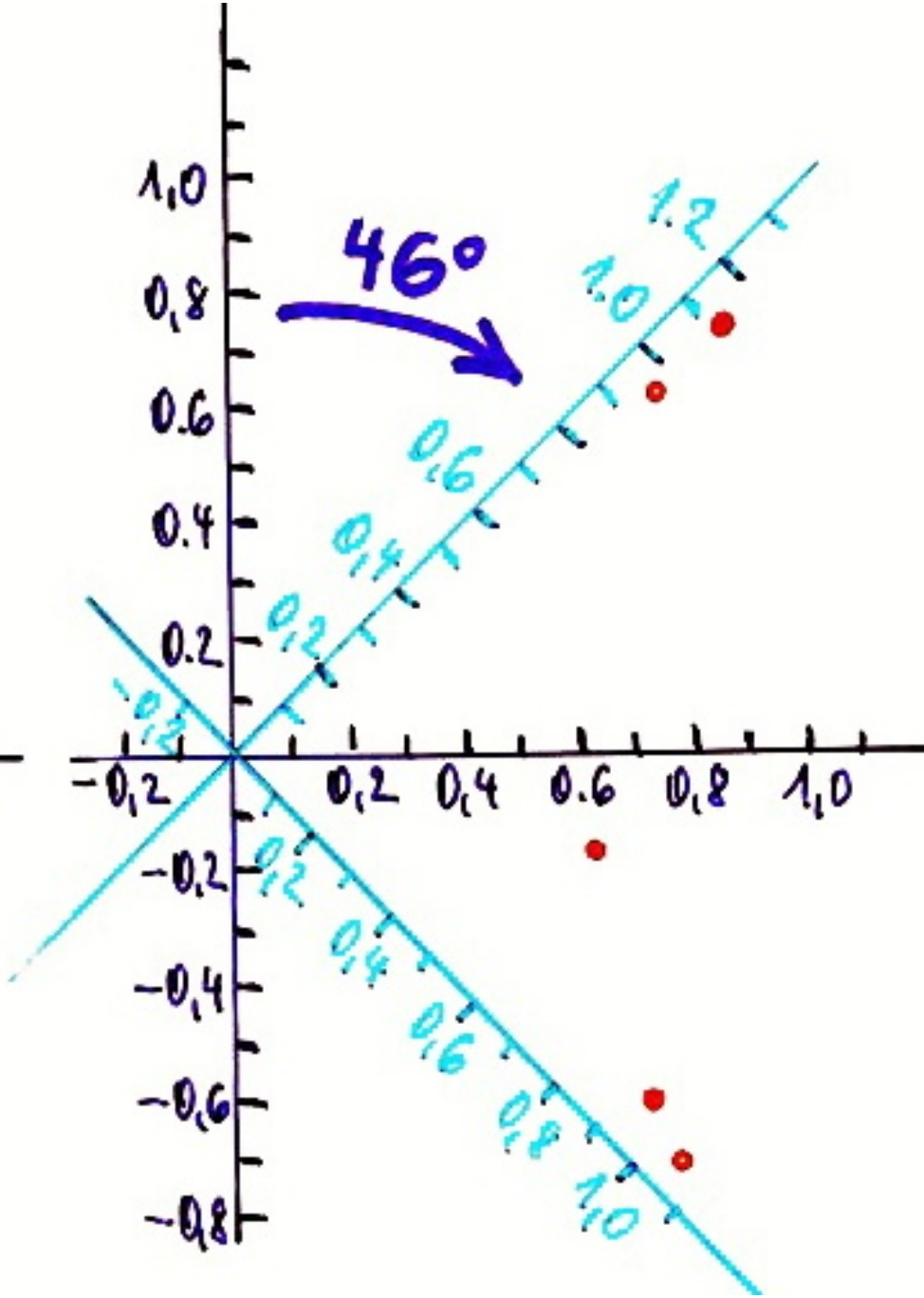
Identifikace smysluplných faktorů, které zestručňují popis původních dat vede k ***rotaci (otočení) faktorů*** (= transformace faktorů do nové podoby, která je snadnější k vysvětlení).

Otočené faktory: některé zátěže znaků dosahují hodnot blízko ±1 a zbývající zátěže pak téměř nulových.

Proto si budeme přát, aby některý znak dosahoval vysoké zátěže pouze u jediného faktoru čili byl *faktorově čistý*.




1,0
0,8
0,6
0,4
0,2
-0,2
-0,4
-0,6
-0,8
-0,2
0,2
0,4
0,6
0,8
1,0


46°
1,0
0,8
0.6
0.4
0.2
-0,2
-0,4
-0,6
-0,8
-0,2
0,2
0,4
0.6
0,8
1,0
1.2
1.0
0,6
0,4
0,2
-0,2
0,2
0,4
0,6
0,8
1,0

Matici lze zobrazit v grafu faktorových zátěží:

- Otočí-li se souřadné osy faktorů F_1 a F_2 o 90° nazýváme toto otočení *orthogonální rotace*.
- Otočí-li se souřadné osy faktorů o jiný úhel než 90° nazýváme toto otočení *neorthogonální rotace*.

Účelem otočení faktorů: získání *jednodušší struktury*, která se snadněji vysvětlí čili proměnné budou spíše faktorově čisté.

Otočením faktorů: neovlivní se těsnost proložení faktorového řešení.

Závěr: Ačkoliv se faktorová matice změní, komunality a procento vysvětleného rozptylu z celkového se nezmění.

Metody otočení faktorů

vedou k identifikaci rozdílných faktorů.

Metoda varimax minimalizuje počet znaků, které vykazují vysokou zátěž faktoru.

Metoda quartimax minimalizuje počet faktorů, potřebných k popisu původního znaku.

Metoda equimax je kombinací obou předchozích.

1) Orthogonální otočení vede k faktorům, které jsou nekorelované.
2) Neorthogonální rotace vede k podstaně smysluplnějším faktorům.

Závěr: Konečně vlivy v přírodě jsou také vesměs korelované.

6.Grafické pomůcky faktorové analýzy

jsou daleko názornější než hodnoty faktorových zátěží v tabulce.

1. Cattelův indexový graf úpatí vlastních čísel

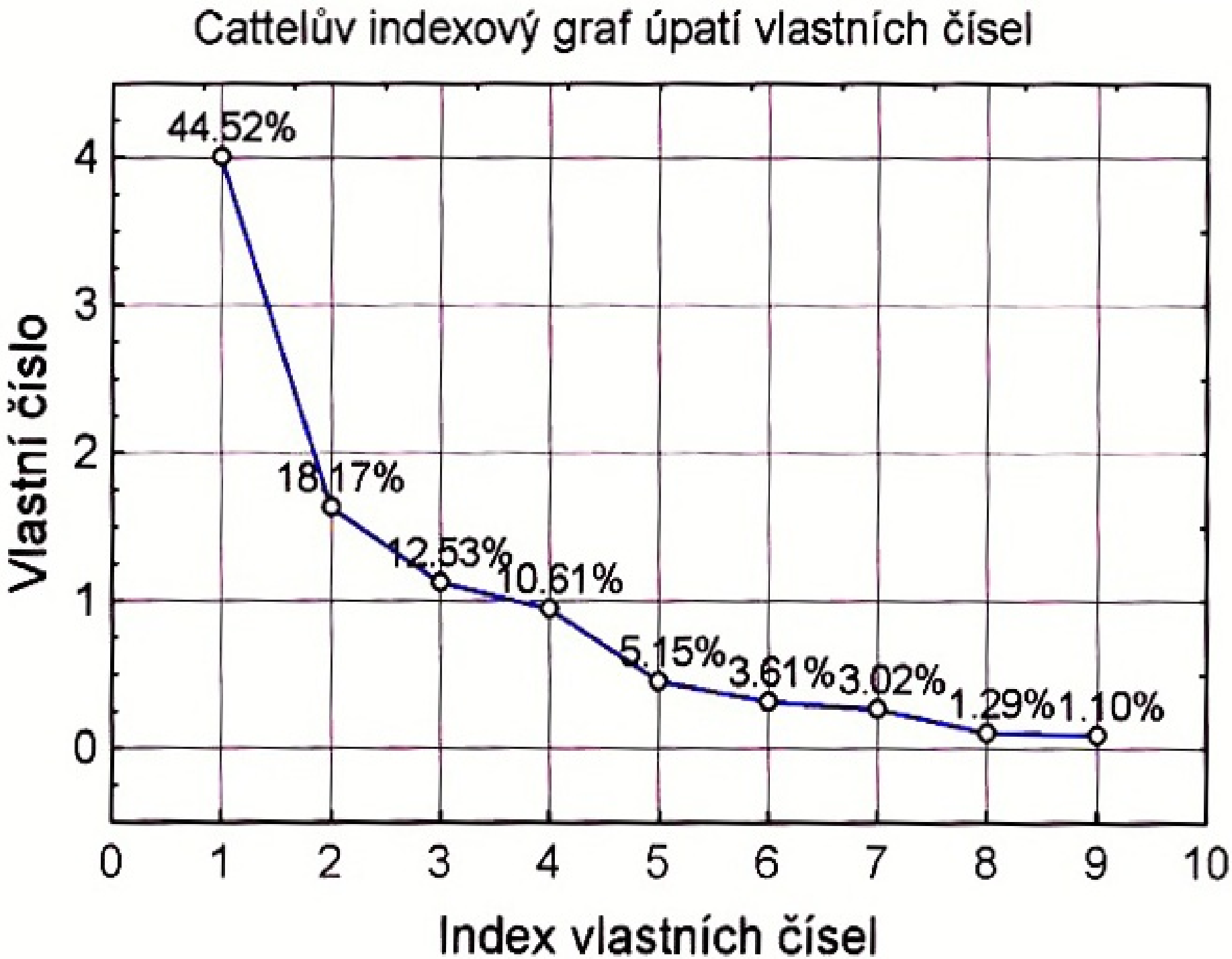
(Scree Plot): Je diagramem vlastních čísel $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ zdrojové matice dat v závislosti na indexu i a zobrazuje relativní velikost jednotlivých vlastních čísel.

Alternativou je graf úpatí pro $\ln \lambda_j$.

Pravidlo „zlomené hole“: pokud je jednotkový úsek náhodně rozdělen na m částí, je očekávaná délka k-té části rovna

$$g_k = \frac{1}{m} \sum_{i=k}^{m} \left(\frac{1}{i} \right) .$$

Zahrnují se pak ty komponenty, pro které platí, že $\lambda_k > g_k$.


Cattelův indexový graf úpatí vlastních čísel
44.52%
18.17%
12.53%
10.61%
5.15%
3.61%
3.02%
1.29%
1.10%
Vlastní číslo
Index vlastních čísel
0
1
2
3
4
5
6
7
8
9
10

Kritéria extrakce použitelných faktorů:

Pro velký počet původních znaků se postupuje:

- extrahuje se taková kombinace znaků, která objasňuje největší množství rozptylu a
- pak se hledají kombinace, které se týkají menšího a menšího rozptylu.

Pravidla:

(a) ***Kritérium vlastního čísla:*** faktory, které mají vlastní číslo větší než 1 jsou uvažovány za statisticky významné a menší než 1 jsou statisticky nevýznamné, (nejspolehlivější pro 20 až 50 původních proměnných).

(b) ***Kritérium apriori*** vychází z faktu, že počet faktorů se bere na základě věcného, obsahového většinou fyzikálního významu.

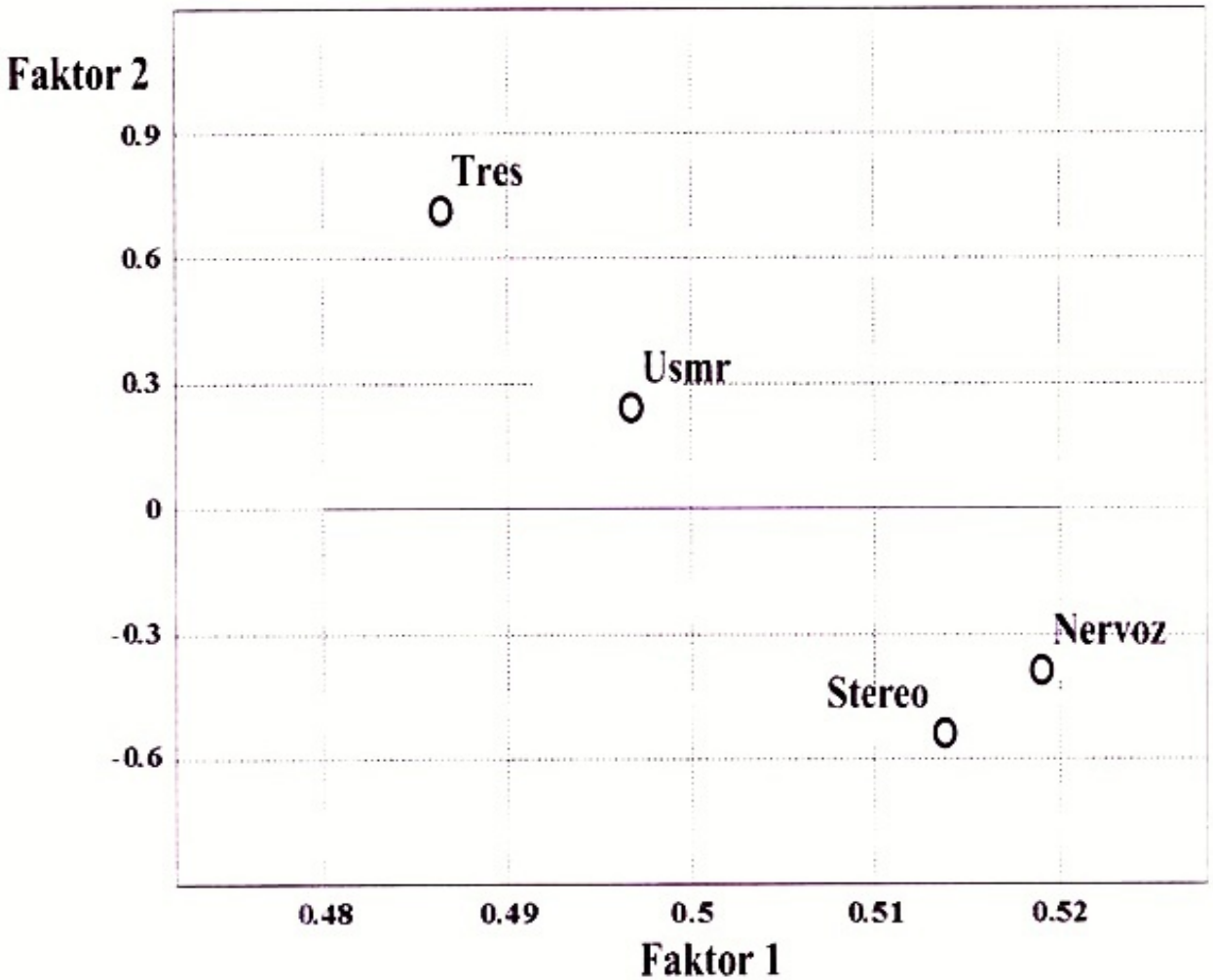
(c) ***Kritérium procenta rozptylu:*** v přírodních vědách se bere obvykle 95% pokrytého rozptylu, zatímco ve společenských vědách pouze 60%.

(d) ***Kritérium Cattelova indexového grafu úpatí vlastních čísel:*** Cattelův graf úpatí vlastních čísel proti počtu faktorů a zlom na křivce přináší obvykle o 1 až 2 faktory více.

(e) ***Přehled všech kritérií:*** v praxi se zřídka užívá pouze jediného kritéria k určování počtu extrahovaných faktorů, dává se přednost porovnání závěrů z více kritérií.

2. Graf faktorových zátěží: na souřadných osách jsou jednotlivé faktory (2D nebo 3D graf):

- shluky proměnných,
- *faktorově čistou původní proměnnou*,
- znaky blízko počátku (0, 0) mají malé zátěže obou faktorů.
- *faktorově nečisté*.



Kritéria statistické významnosti faktorových zátěží:

Které faktory se vyplatí uvažovat z hlediska praktické a statistické významnosti.

(1) Zaručení praktického významu:

- zátěže větší než ±0.30 jsou uvažovány za minimální hladinu,
- zátěže větší než ±0.40 jsou pak důležitější,
- zátěže větší než ±0.50 jsou považovány za významné.

Platí pravidlo: *čím větší absolutní velikost faktorové zátěže, tím důležitější je zátěž při výkladu faktorové matice.*

Výklad:

- Faktorové zátěže představují korelaci mezi faktorem a původním znakem.
- Čtverec faktorové zátěže je množství celkového rozptylu dotyčného znaku vysvětlené faktorem.

Například:

(a) zátěž 0.30 představuje $0.30^2 = 0.09$ čili skoro 10% vysvětleného rozptylu dotyčného původního znaku faktorem ,

(b) zátěž 0.50 představuje $0.50^2 = 0.25$ čili 25%vysvětleného rozptylu dotyčného původního znaku faktorem.

©) aby bylo pokryto 50% rozptylu původního znaku, musí faktorová zátěž tohoto znaku dosáhnout alespoň hodnoty 0.7 (tj. $0.7^2 = 0.49$).

Vysoké hodnoty zátěže, okolo 0.8 a vyšší, nejsou sice typické, přesto toto kritérium se jeví velice užitečné.

Pravidla platí pro více jak 100 objektů.

Úprava založená na počtu proměnných:

Kritéria pro významnost faktorových zátěží jsou:

1. Čím větší počet objektů, tím menší faktorová zátěž je ještě významná.

2. Čím větší počet původních znaků, tím menší faktorová zátěž je ještě významná.

3. Čím větší počet faktorů, tím větší velikost zátěží dalších faktorů, které by byly ještě statisticky významné.

Pomůcky k interpretaci faktorové matice

(a) Vyšetření matice faktorových zátěží: každý sloupec faktorové matice představuje jeden faktor. Sloupce čísel jsou faktorové zátěže pro každý znak v každém faktoru.

V neorthogonální rotaci existují dvě faktorové matice zátěží:

(1) *Matice faktorového modelu* obsahuje zátěže, které představují specifický příspěvek každého znaku do faktoru.

(2) *Matice faktorové struktury* obsahuje jednoduché korelace mezi znaky a faktory, ale tyto zátěže obsahují oba rozptyly, jak specifický mezi znaky a faktory a dále korelaci mezi faktory.

(b) Odhalení nejvyšší zátěže u každé proměnné: by mělo

- začít u první proměnné na prvním faktoru a posouvat se vodorovně zleva doprava a hledat nejvyšší faktorovou zátěž pro *i*-tý znak na libovolném faktoru,

- když je nalezena nejvyšší absolutní hodnota zátěže, její významnost se zdůrazní podtržením,

- pak se u druhé proměnné posouvá znova zleva doprava a hledá se nejvyšší faktorová zátěž pro tento znak a libovolný faktor a nejvyšší se opět podtrhne,

- postup se opakuje až jsou podtrženy nejvyšší zátěže u všech znaků,

- pro velikost výběru menší než 100 bude nejnišší faktorová zátěž brána ještě za významnou, když bude vždy větší než ±0.30,

- znak, který je propojen s několika faktory s vysokými zátěžemi je obvykle kandidátem na odstranění z dat.

(c) Stanovení komunalit proměnných:

◆ Po podtržení znaků dominantně ovlivňujících faktor se odhalí nepodtržené znaky, které nezatěžují ve faktorové matici žádný faktor.

◆ Komunality u každého znaku vyčíslují množství rozptylu, vystiženého v dotyčném faktoru.

◆ Uživatel by měl sledovat komunalitu u každého znaku a určit zda komunality mají přijatelné vysvětlení.

◆ Byla alespoň polovina rozptylu každého znaku vzata do úvahy.

◆ Takto by se měly označit znaky s komunalitami menšími než 0.50 jako "nedávají dostatečné objasnění".

◆ Znak nezatěžuje žádný faktor (komunalita je velmi nízká):

(1) tento znak ignorovat,

(2) vyhodnotit takový znak pro možné vyloučení z dat a faktorový model opravit na menší počet znaků.

(d) Pojmenování faktorů:

- Je třeba pojmenovat, vložit obsahový význam do faktorových zátěží.

- Znaky s vyššími zátěžemi jsou považovány za důležitější a mají i větší vliv na pojmenování dotyčného faktoru.

- Uživatel vyšetří podtržené znaky v dotyčném faktoru a vloží větší důraz na proměnné s vyšší hodnotou zátěže. Provede pojmenování či významové označení dotyčného faktoru.

- Znaménko + či - značí: kladné u faktoru značí kladný vztah zatímco záporné značí, že proměnné jsou negativně vztaženy.

- Pojmenování se týká všech extrahovaných a využitelných faktorů.

- Někdy však nelze přidělit jméno či označení využitelnému faktoru, pak bude označeno "nepojmenované".

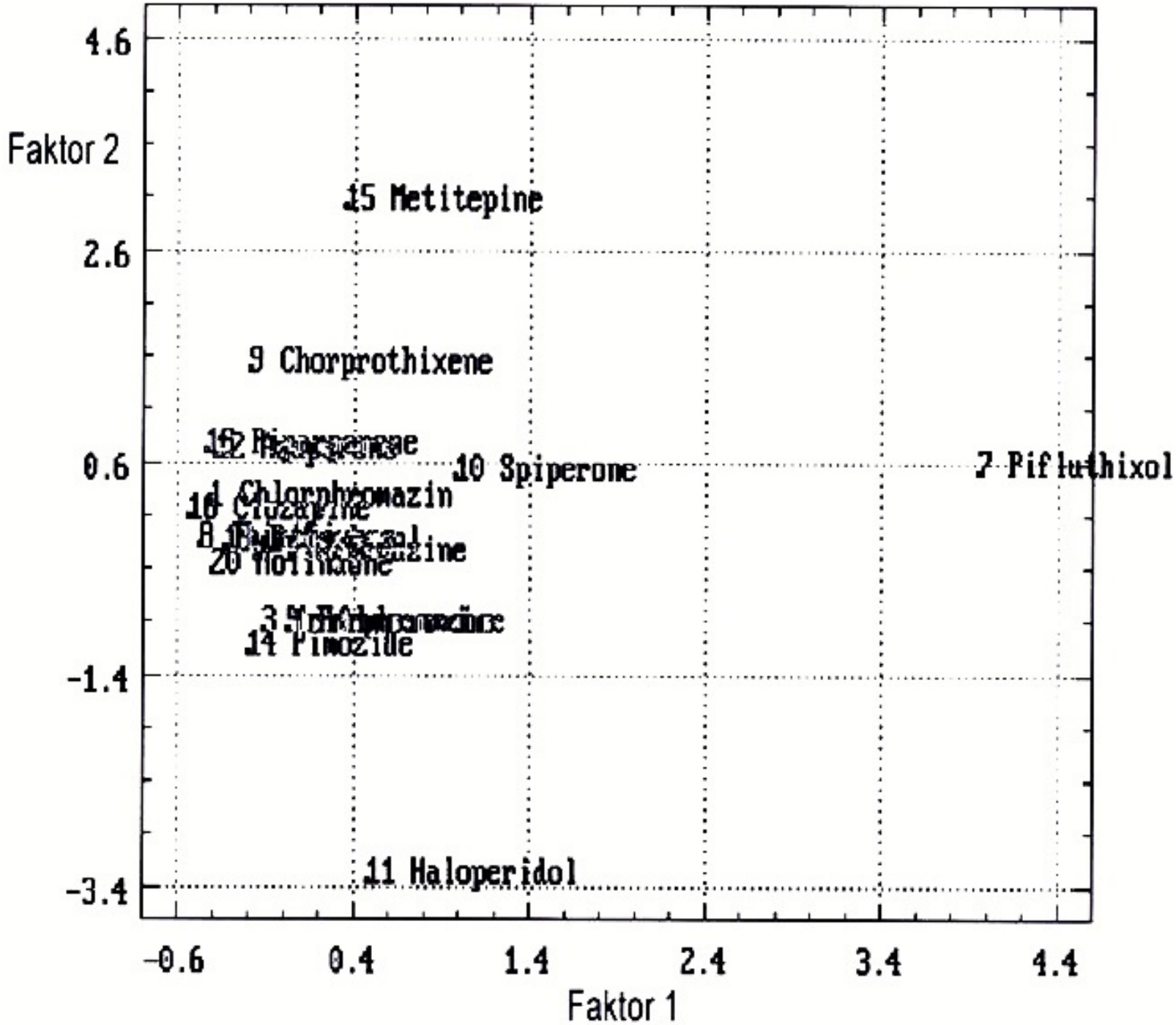
- Vysvětlují se pouze ty faktory, které jsou smysluplné a nevšímáme si nepojmenovaných faktorů.

- Volba užitečného počtu faktorů a rotační metoda jsou svázány vnitřním vztahem.

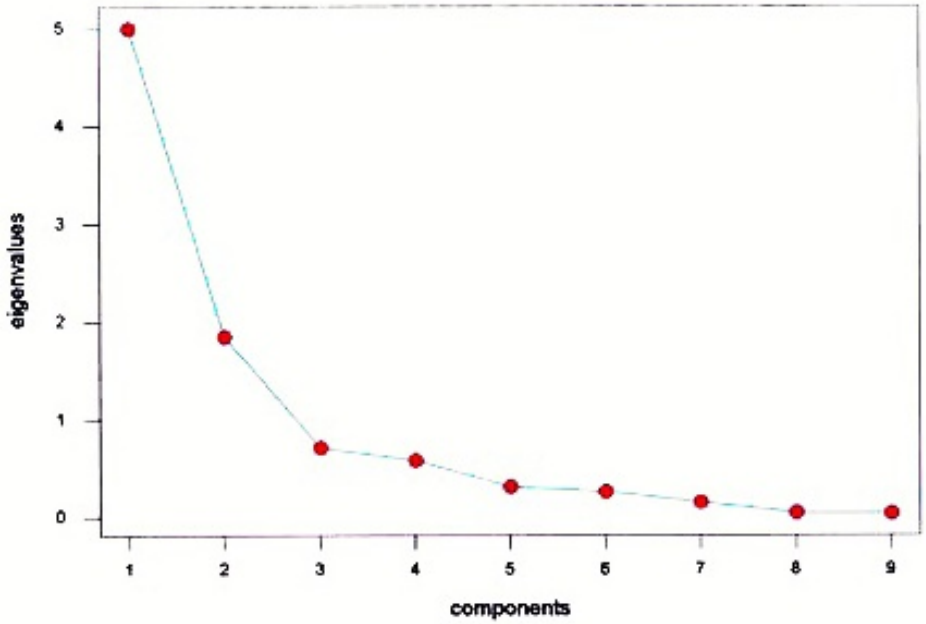
3. Graf faktorového skóre: faktorové skóre pro každý objekt pro vybraný pár faktorů je užitečný k odhalení vyjímečných objektů (odlehlých). Pro k-tý objekt je skóre pro j-tý faktor

$$\hat{F}_{jk} = \sum_{i=1}^{P} w_{ji} x_{ik},$$

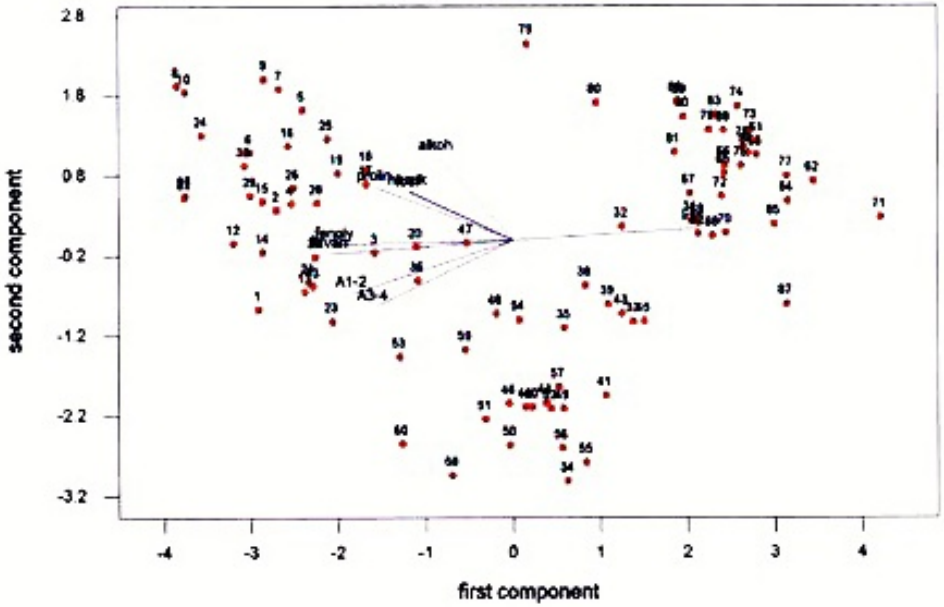
kde x_{ik} je standardizovaná hodnota i-té proměnné pro k-tý objekt a w_{ji} je koeficient faktorového skóre pro j-tý faktor a i-tou proměnnou, P je celkový počet všech původních proměnných.


Faktor 2
4.6
2.6
0.6
-1.4
-3.4
15 Metitepine
9 Chorprothixene
10 Spiperone
7 Piflufhixol
14 Pimozide
11 Haloperidol
-0.6
0.4
1.4
2.4
3.4
4.4
Faktor 1

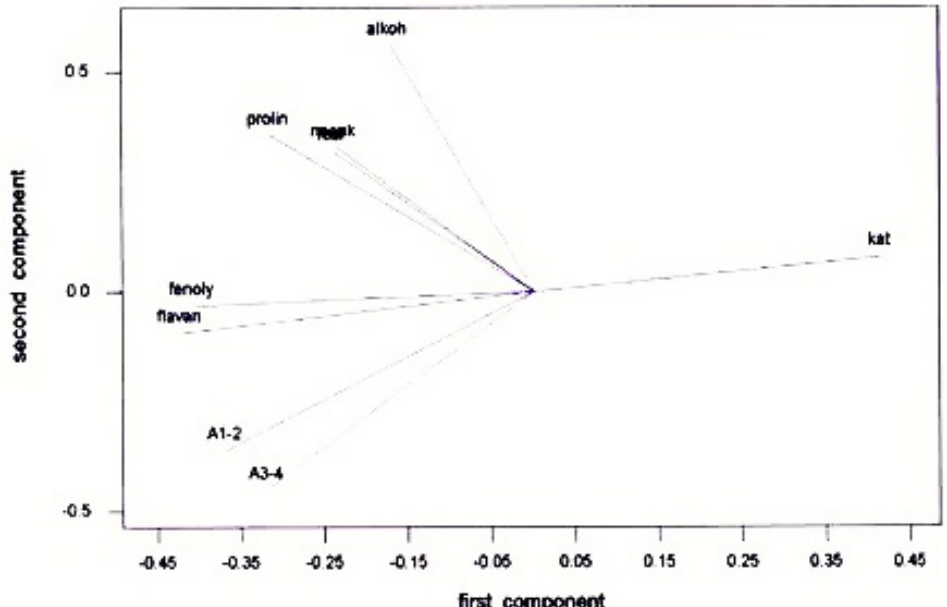
Principal Components Eigenvalue (Scree) Plot



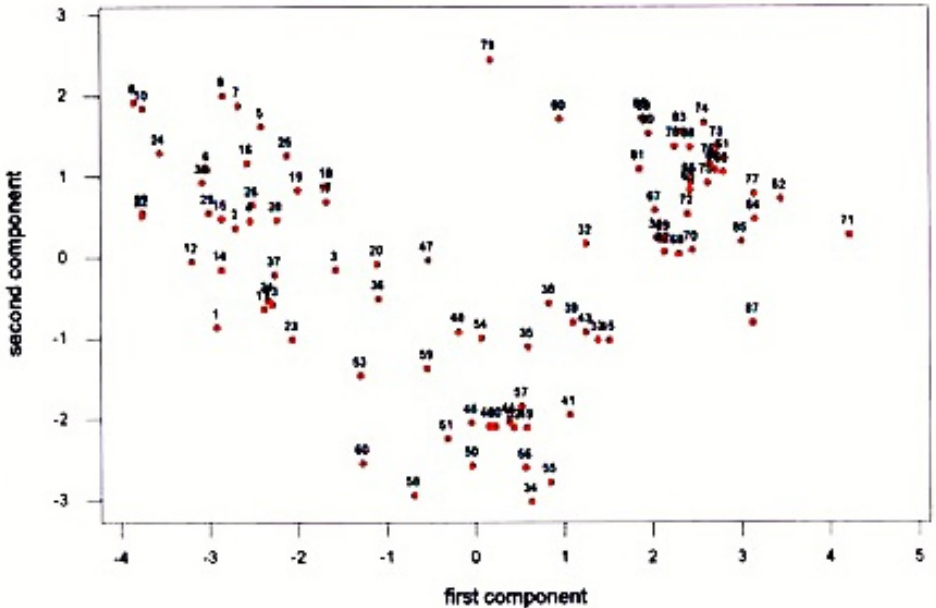
Principal Components Biplot



Principal Components Loading Plot



Principal Components Score Plot



PŘÍKLAD 5.1 *Sociologický průzkum spokojenosti*

Dotazník o fiktivní spokojenosti v životě byl vyplněn 100 dospělými respondenty. Obsahoval 10 položek vystižení spokojenosti v v práci, se svými koníčky, spokojenost doma v rodině a v ostatních činnostech v životě.

1) Cílem je nalezení vztahu mezi spokojenostmi v různých doménách.

2) Existují nějaké faktory, které jsou v pozadí těchto posuzovaných domén?

3) Jaký je obsah těchto faktorů a jak je pojmenujeme?

○ ***Data:*** data obsahují 100 řádků (respondentů) a 10 sloupců znaků spokojenosti k činnosti na pracovišti (*Práce1, Práce2, Práce3*), spokojenosti s koníčky (*Hoby1, Hoby2*), spokojenosti doma v rodině (*Doma1, Doma2, Doma3*) a spokojenost v ostatních aktivitách života (*Různé1, Různé2*). Odpovědi jsou škálovány, aby průměr znaku dosahoval hodnoty okolo 100.

Objekt	*Prace1*	*Prace2*	*Prace3*	*Hoby1*	*Hoby2*	*Doma1*	*Doma2*	*Doma3*	*Ruzne1*	*Ruzne2*
1	105.13	101.66	115.06	101.00	95.18	100.28	101.67	85.55	104.03	110.28
2	77.05	72.93	77.49	72.74	61.56	93.85	95.39	88.61	70.12	72.00
3	86.02	82.21	78.89	77.95	91.71	86.77	108.07	93.35	86.02	70.69
..	...	...	...	...	...	...	...	...	...	...
100	106.05	120.71	119.81	101.85	94.96	75.86	93.17	93.38	109.36	83.79

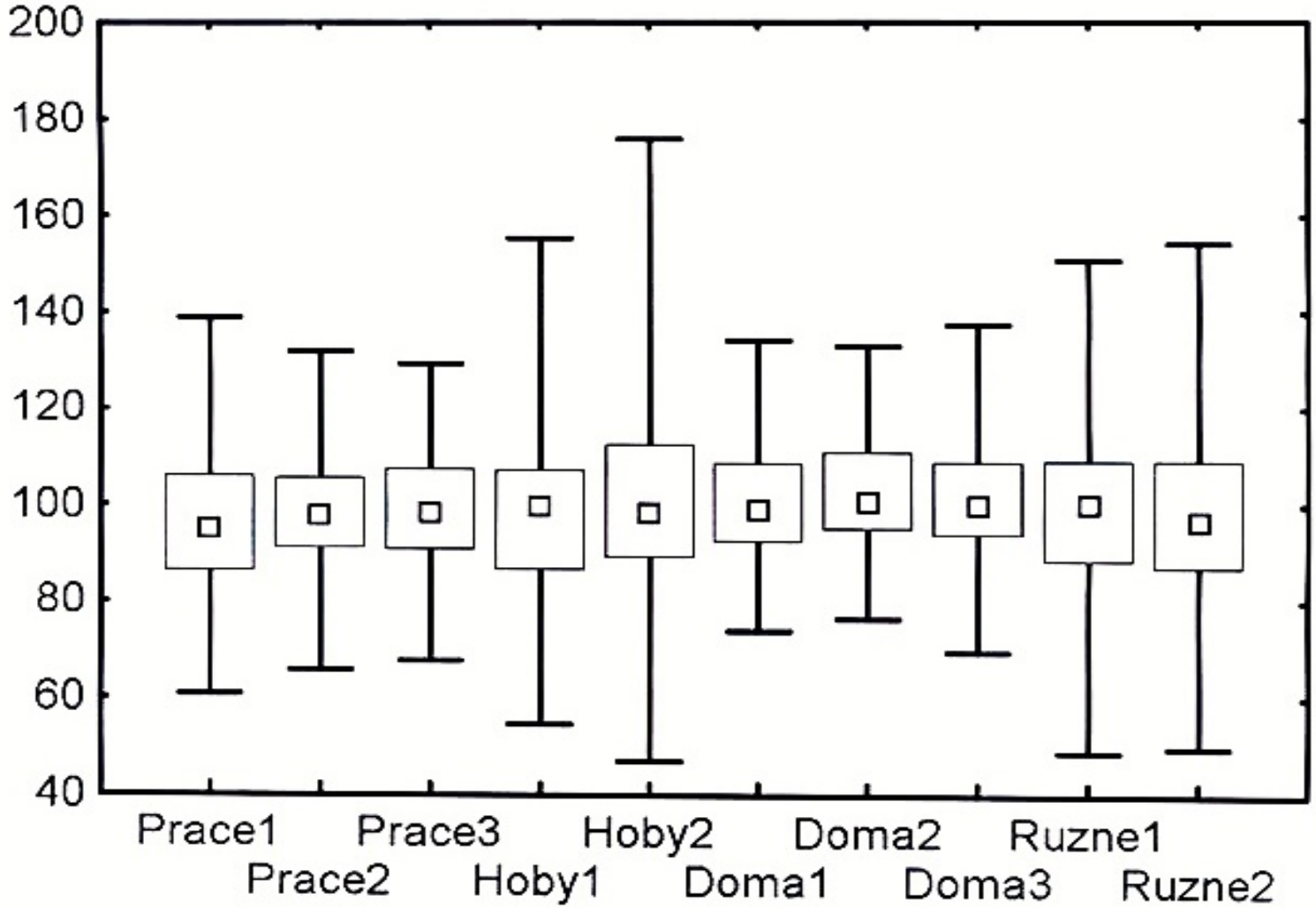
○ ***Řešení:*** použita STATISTICA, NCSS2000.

1. Popisné statistiky měr polohy a rozptýlení:

Znak	Četnost	Průměr	Směrodatná odchylka	Kommunalita
Prace1	100	97.03	15.52	0.57
Prace2	100	98.17	11.27	0.79
Prace3	100	98.94	12.49	0.69
Hoby1	100	98.03	15.94	0.89
Hoby2	100	100.11	19.94	0.73
Doma1	100	99.51	11.98	0.55
Doma2	100	101.60	11.06	0.82
Doma3	100	101.37	12.73	0.63
Ruzne1	100	99.16	16.97	0.92
Ruzne2	100	98.22	19.11	0.79

Komunalita ukazuje, jak dobře je tento znak predikován vybranými faktory.

Míru proměnlivosti v datech ilustruje krabicový graf na obrázku 5.3a. Je zřejmé, že znaky *Hoby1*, *Hoby2*, *Různé1* a *Různé2* dosahují největší proměnlivosti v datech.

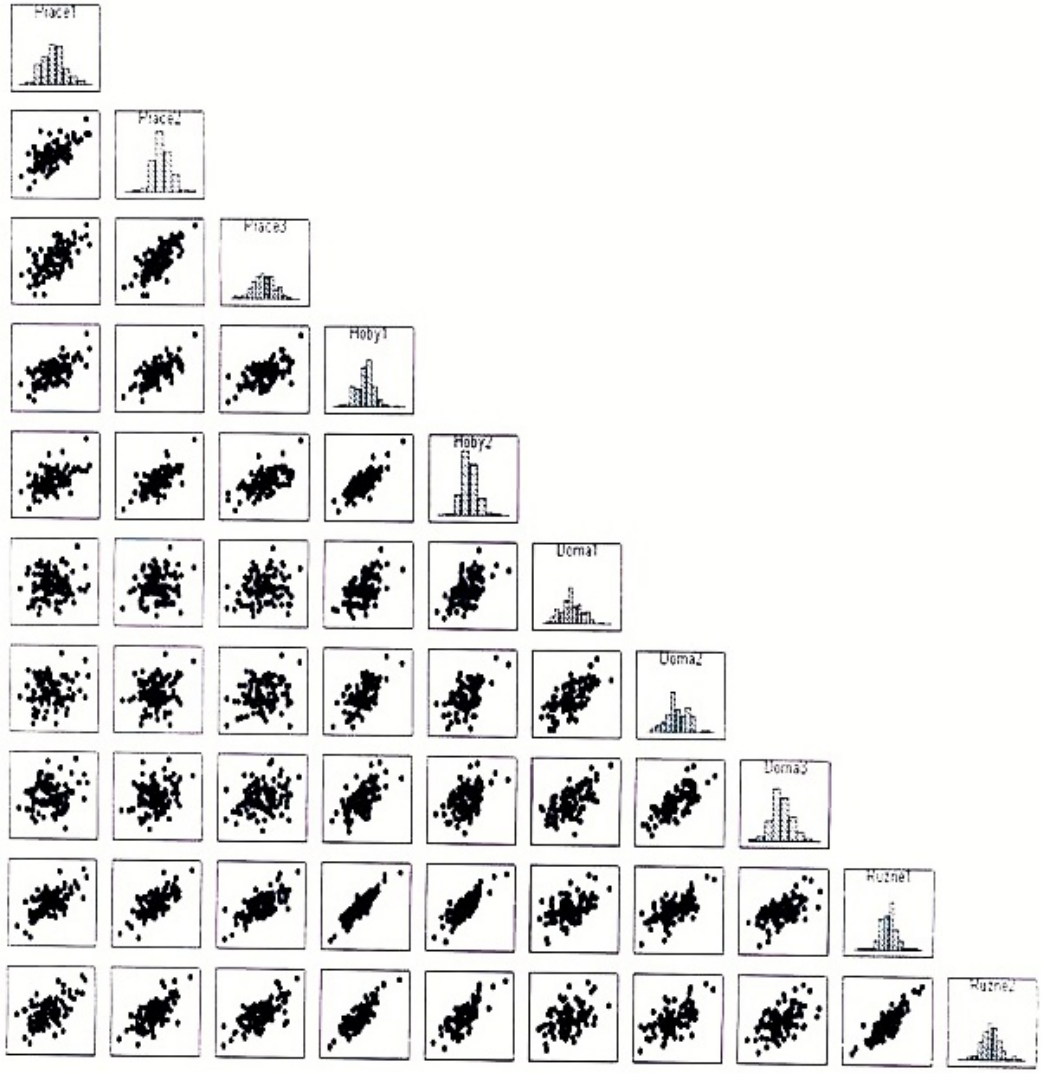


Obr. 5.3a Krabicové grafy proměnlivosti 10 znaků matice zdrojových dat *Spokojenost* pro 100 respondentů (STATISTICA).

Maticový diagram ukazuje největší a nejmenší korelaci dvojic znaků.

■ Příkladně *Hoby1* a *Různé1* jsou silně korelovány $r = 0.90$, jiné *Prace1* a *Doma1* dosahují nízké nevýznamné hodnoty $r = 0.14$.

■ **Korelační matice proto indikuje jakousi zřetelnou strukturu v datech.**



Obr. 5.3b Maticový diagram dat *Spokojenost* pro 100 respondentů a 10 znaků (STATISTICA).

2. Korelační matice:

a) Čárový diagram absolutních hodnot korelační matice

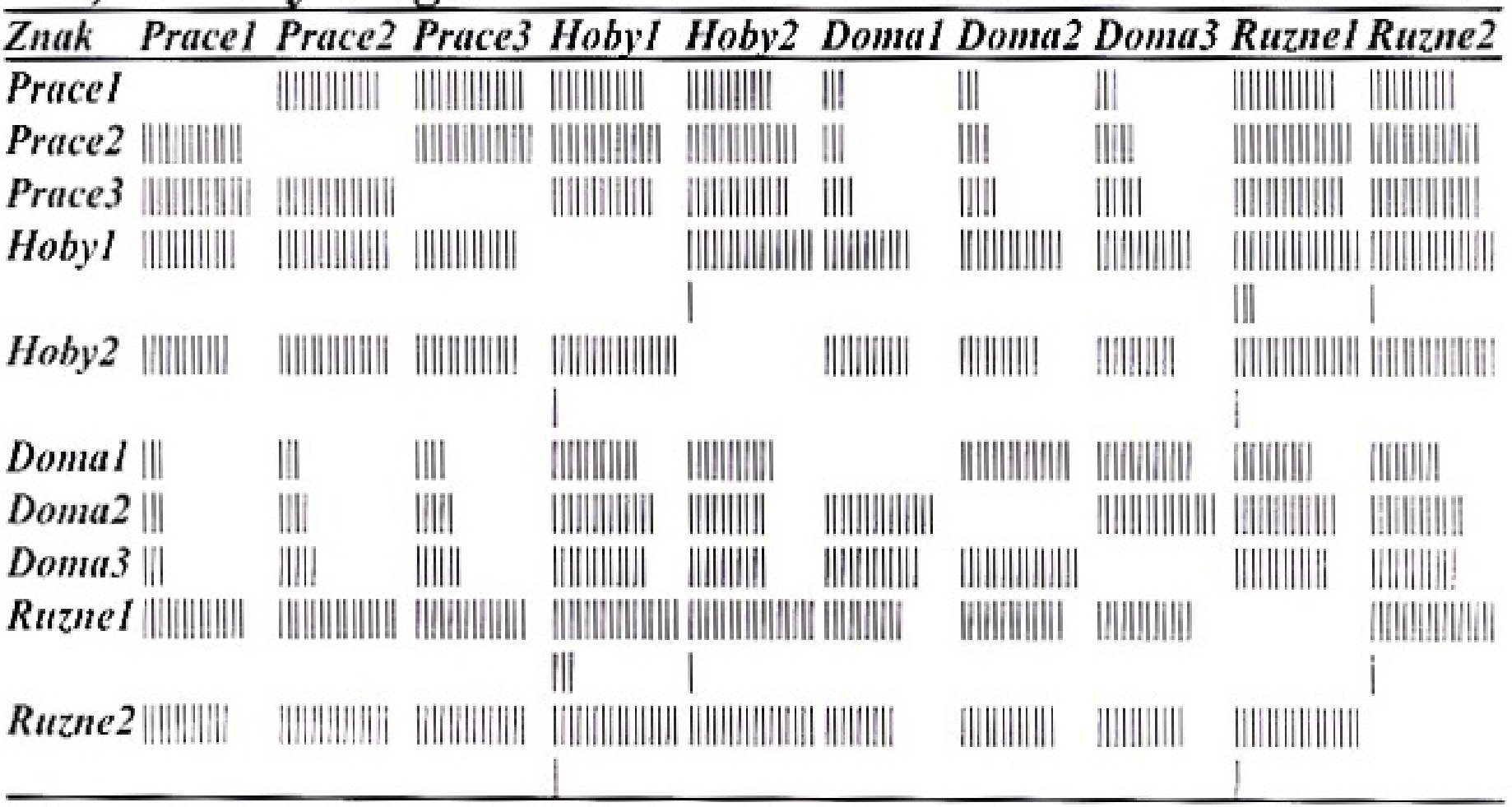
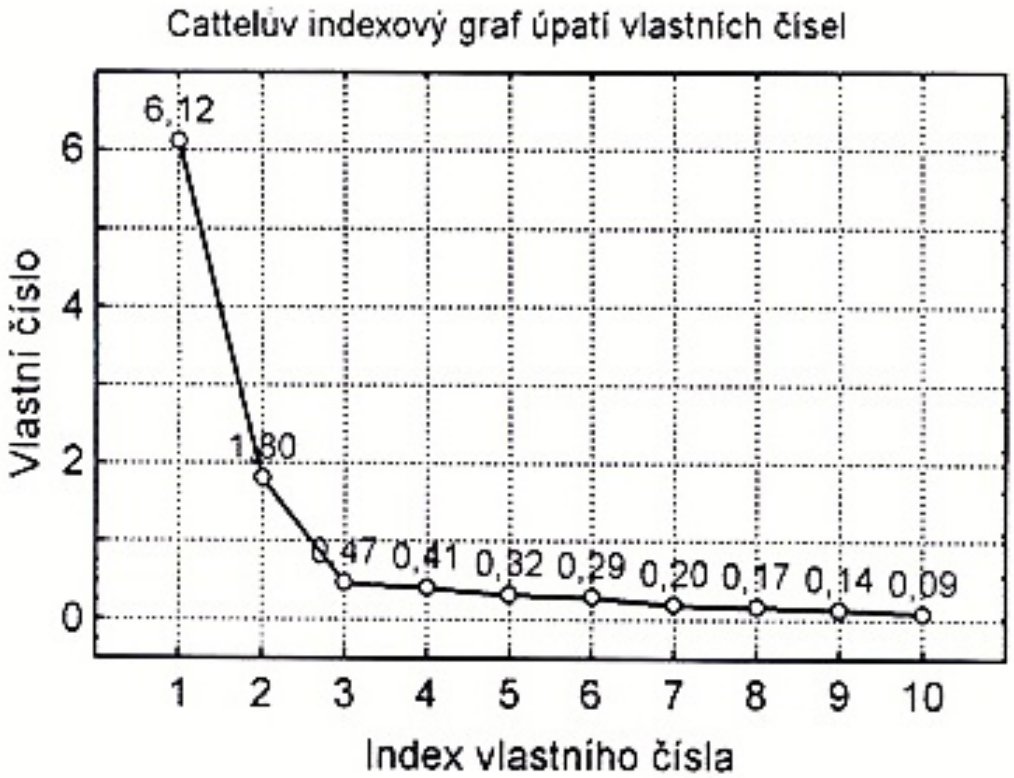


Diagram zobrazuje absolutní hodnoty korelací a ukazuje největší a nejmenší korelaci dvojice znaků.

3. Vyšetření Cattelova indexového grafu úpatí vlastních čísel (Scree Plot)

Index	Vlastní číslo λ_i	Procento celkového rozptylu	Kumulativní vlastní číslo	Kumulativní procento
1	6.118	61.184	6.118	61.184
2	1.801	18.007	7.919	79.191
3	0.473	4.729	8.392	83.919
4	0.408	4.080	8.800	87.999
5	0.317	3.172	9.117	91.172
6	0.293	2.933	9.410	94.105
7	0.196	1.958	9.606	96.063
8	0.170	1.704	9.777	97.767
9	0.138	1.380	9.915	99.147
10	0.085	0.853	10.000	100.000

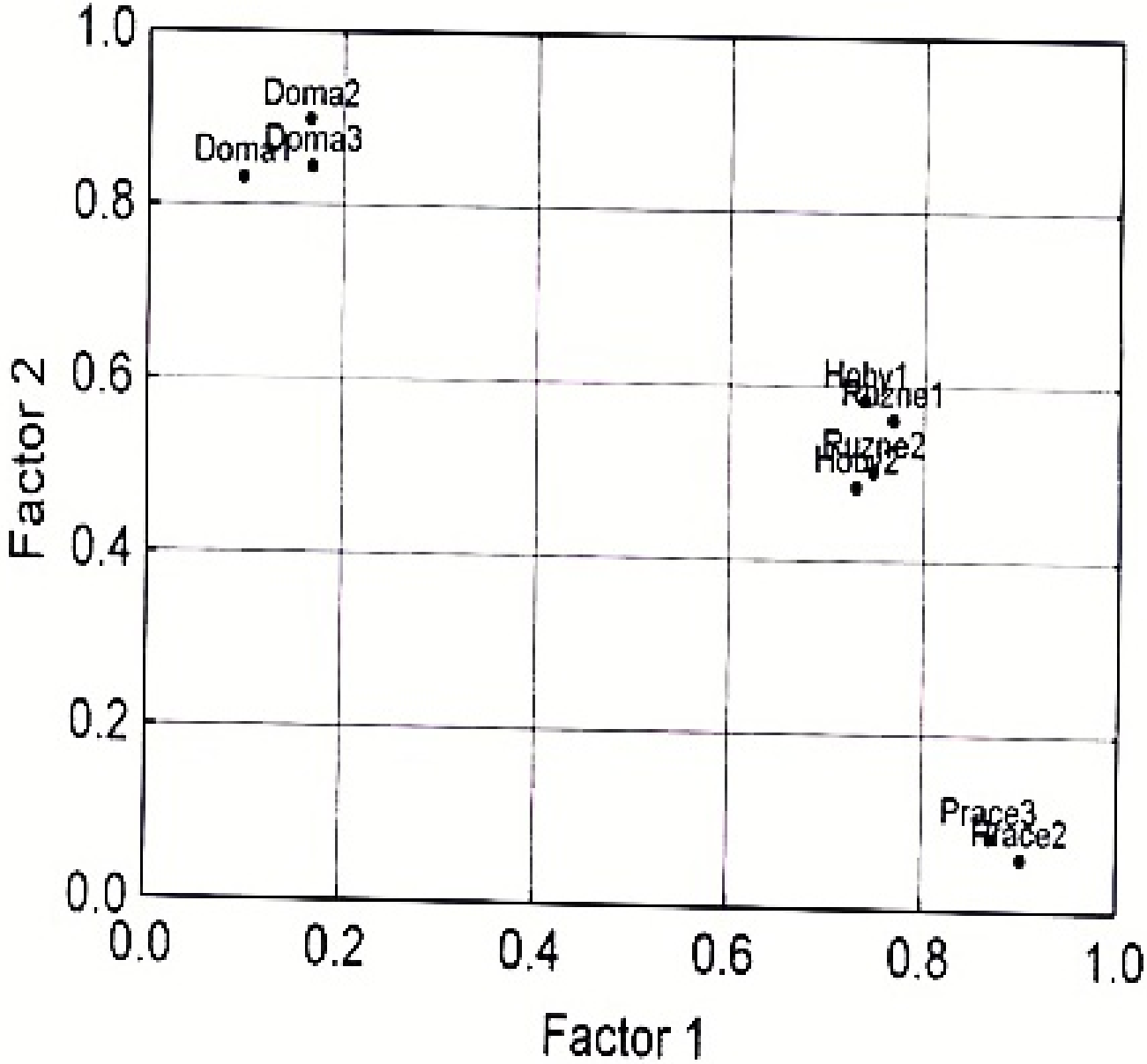


Obr. 5.4a Cattelův indexový graf úpatí vlastních čísel k určení počtu faktorů pro 100 respondentů a 10 znaků zdrojové matice dat *Spokojenost*. Hodnoty v grafu jsou vlastní čísla zdrojové matice dat (STATISTICA).

- První faktor popisuje 61.2% proměnlivosti v datech a druhý faktor 18.0%.
- Užívá se těch faktorů, jejichž vlastní čísla jsou větší než 1.
- Suma vlastních čísel je rovna počtu znaků.
- *Procento celkového rozptylu* přináší procento celkové proměnlivosti ve znacích, vystižené tímto faktorem.

4. Faktorové váhy pro jednotlivé znaky po rotaci:

Znak	Faktor1	Faktor2
Prace1	-0.653	0.514
Prace2	-0.757	0.495
Prace3	-0.746	0.457
Hoby1	-0.942	-0.022
Hoby2	-0.876	0.052
Doma1	-0.576	-0.605
Doma2	-0.671	-0.618
Doma3	-0.642	-0.574
Ruzne1	-0.952	0.014
Ruzne2	-0.900	0.048



Obr. 5.4b Graf faktorových vah pro 100 respondentů a 10 znaků zdrojové matice dat *Spokojenost* (STATISTICA).

5. Otáčení faktorů

- Otočení faktorů ve faktorovém prostoru je libovolné.
- Otočení lze provést tak, aby interpretace byla snažší a jednodušší.
- Cílem je aby některé znaky dosahovaly maximálních hodnot jednoho faktoru a jiné znaky minimálních hodnot čili byly "faktorově čisté".
- Pokusíme se otočit první tři faktory metodou Varimax.

První faktor se týká všech znaků spokojenosti v práci, zatímco znaky spokojenosti doma v něm dosahují nejmenších vah.
Druhý faktor vykazuje nejvyšší váhy pro znaky spokojenosti doma a nejnižší váhy jsou pro znaky spokojenosti v práci.

První faktor vystihuje spokojenost v práci,
Druhý faktor spokojenost doma.

Znaky *Prace2* a *Prace3* leží v diagramu blízko sebe, a proto silně korelují. Znaky *Prace* a *Doma* jsou dál od sebe, proto méně korelují. Méně korelují se zbývajícími dvěma znaky *Hoby* a *Ruzne*.

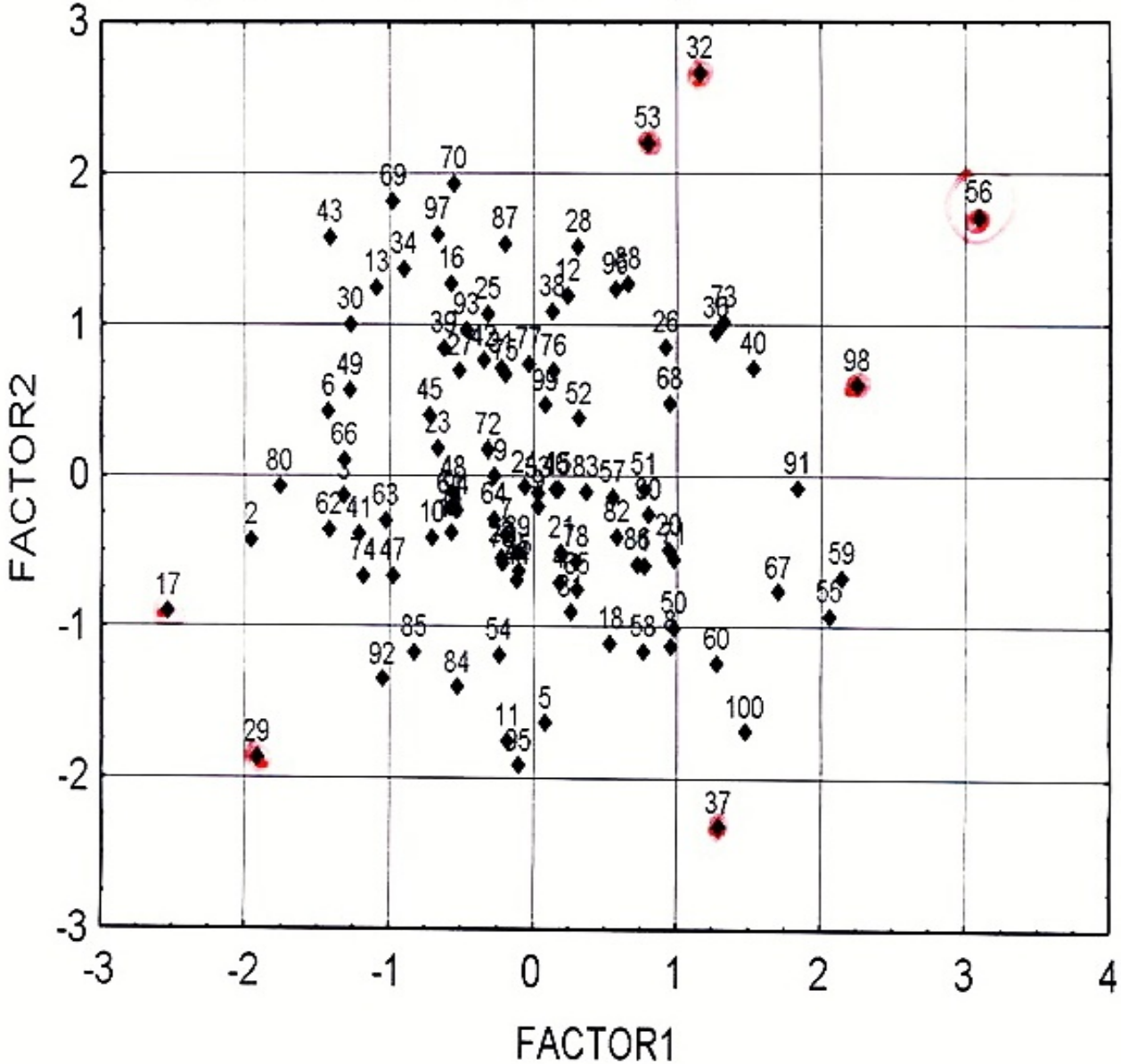
6. Příspěvky faktoru do komunality po Varimax rotaci:

Znak	Faktor1	Faktor2	Kommunalita
Prace1	0.426	0.690	0.560
Prace2	0.573	0.818	0.735
Prace3	0.556	0.765	0.654
Hoby1	0.887	0.887	0.866
Hoby2	0.767	0.769	0.739
Doma1	0.332	0.698	0.541
Doma2	0.451	0.833	0.739
Doma3	0.412	0.741	0.584
Ruzne1	0.905	0.906	0.884
Ruzne2	0.811	0.813	0.779

- Komunalita představuje podíl proměnlivosti znaku vyjádřené dotyčnými faktory, kterou dostaneme, když budeme původní znaky vysvětlovat pomocí regrese vybranými faktory.

- Tabulka obsahuje příspěvek prvního a druhého faktoru do komunality.

7. Faktorová skóre jednotlivých objektů v rozptylovém diagramu po Varimax rotaci



Obr. 5.4c Rozptylový diagram komponentního skóre zdrojové matice dat *Spokojenost* pro 100 respondentů a 10 znaků (STATISTICA).

- Kromě podezřelých objektů 17, 29, 32, 53, 55, 98, 37 jsou zbývající objekty v rozptylovém diagramu komponentního skóre na obr. 5.4c umístěny v jediném shluku.

Vysloveně odlehlý bode budeme chápat objekt 56.

○ *Závěr:*

■ Rozptylový diagram faktorového skóre ukazuje na polohu objektů v prostoru faktorů.

■ Prvních *k* faktorů (kde *k* je počet největších vlastních čísel) ukazuje na hlavní strukturu, která byla nalezena v datech.

■ Zbytek faktorů ukazuje odlehlé hodnoty a lineární závislosti.

■ U *komunalit* si je třeba uvědomit, že komunalita znaků je ta část rozptylu, která může být reprodukována z vybraného počtu faktorů a rotace faktorů nemá na komunality žádný vliv.

■ Velmi nízké hodnoty komunalit pro jeden nebo dva znaky v řadě ostatních indikují, že tyto znaky nejsou dobře započteny navrženým faktorovým modelem.

Diagnostikování problémů FA

Uvažuje se lineární model: každý znak je lineární funkcí společných faktorů plus specifický chybový faktor.

Musí existovat alespoň mírná korelace mezi původními znaky.

Problémy:

1. Výběr zobrazí také odlehlé body a prověří linearitu mezi znaky. Může být potřebná transformace znaků k linearizaci vztahů mezi znaky.

2. Při interpretaci faktorů se mohou výsledky drasticky změnit při jiném počtu faktorů. Volba počtu faktorů m závisí na nejsnadněji vysvětlitelných. Použije se několik počtů m k vyšetření, který počet je nejlepší.

3. Ideálně užijeme alespoň 2 faktory s nenulovými zátěžemi u každého faktoru. Má-li však faktor pouze jediný znak, může se považovat za specifický faktor a pak je možné užívat místo faktorů původní znaky.

4. Očekává-li uživatel korelované faktory, je třeba použít neorthogonální rotaci faktorů.

5. Výsledky FA jsou za účelem interpretace dat než statistického testování.

6. FA je pomůcka exploratorní analýzy **EDA** a je užitečná pro svou pragmatičnost.

FA je vnitřně závislá technika

1. Znaky (proměnné) nejsou rozlišovány na závislé a nezávislé.

2. Všechny znaky jsou uvažovány souběžně.

3. Každý znak je vztažen ke všem ostatním lineárním vztahem všech původních znaků.

4. Každý znak je predikován ze všech ostatních.

5. Na každý faktor se lze dívat jako na závisle proměnnou, která je funkcí celého souboru původních znaků.

5. Na každý faktor se lze dívat jako na závisle proměnnou, která je funkcí celého souboru původních znaků.

6. Analogie ilustruje rozdíl v účelu mezi technikami závislými (predikce) a vnitřně závislými (identifikace struktury).

7. FA považujeme za exploratorní techniku, která je užitečná v hledání struktury v souboru znaků nebo slouží k redukci dat.

8. Při testování hypotézy, které znaky by se mohly seskupovat v jednom faktoru nebo ve skupině faktorů. Pak FA provádí konfirmatorní analýzu čili přiděluje datům očekávanou strukturu.

Postup FA

Postup analýzy vnitřní struktury dat obsahuje:

1. Vyčíslí se korelační matice všech znaků:

- Vyznačí se znaky, které nejsou v korelaci s ostatními.
- Vyčíslí se vhodnost faktorového modelu.
- Rozhodne se, co dělat s "dírami" ve vstupní matici dat.

2. Provede se extrakce potřebných faktorů.

- Musíme se ujistit, jak dobře FA model prokládá data.

3. Rotací faktorů se docílí jejich lepší interpretace.

4. Vyčíslí se faktorové skóre:

- Pro každý objekt se naleznou shluky podobných objektů.

1. krok: Cíle faktorové analýzy

Cílem: nalézt způsob *zestručnění* informace,
- obsažené ve značném počtu původních proměnných,
- do menšího souboru nových, latentních faktorů,
- s minimální ztrátou informace.

Jedná se: o vyhledání základní konstrukce latentních faktorů,

Identifikace struktury zestručněním dat:
- mezi znaky (ve sloupcích) nebo mezi objekty (v řádcích),
- vyšetřením korelací mezi proměnnými nebo mezi objekty,

R-faktorová analýza: zestručnění popisných charakteristik znaků, analýza korelační matice znaků

Q-faktorová analýza: korelační matici objektů zhušťuje velký počet objektů do různě velkých skupin (lepší je však analýza shluků).

2. krok: Formulace úlohy faktorové analýzy

Formulace úlohy FA:

(a) korelační matice dat, která vystihuje spojení znaků a objektů,

(b) návrh studie v pojmech počtu faktorů, měřitelných vlastností znaků a typů využitelných znaků,

(c) návrh velikosti matice objektů, které jsou funkcí využitelných znaků.

Korelace v proměnných a v objektech: lze vyčíslit

(a) Korelační matice znaků (R-faktorová analýza): shlukování znaků,

(b) Korelační matici objektů (Q-faktorová analýza): shlukování objektů,

Volba proměnných a problém naměření dat: před FA je třeba odpovědět na otázky

(1) jak jsou znaky měřeny?,

(2) kolik znaků by se mělo použít ve faktorové analýze?

Znaky v FA jsou charakteru *metrického* nebo *nemetrického* (0 či 1).

Klíčovým znakům říkáme ***markery*** nebo ***indikanty***, které odrážejí obsah latentních proměnných - faktorů.

Velikost výběru dat: obvykle 50 objektů, lépe však 100 a více.

Pravidlo: na jeden znak by mělo být alespoň 5 objektů, lépe však 10.

Empirická zkušenost: na každý znak by mělo být alespoň 20 objektů.

3. krok: Předpoklady faktorové analýzy

Předpoklady: normalita, homoskedasticita a linearita neovlivní FA.

Určitá multikolinearita: žádoucí, protože cílem je identifikovat vnitřně svázaný čili korelovaný soubor znaků.

Úvodní test: zda zdrojová matice dat obsahuje dostatečné korelace, aby se vůbec mohla použít faktorová analýza.

Korelace mezi proměnnými

lze vyšetřit vyčíslením parciálních korelací mezi proměnnými , když neexistuje dostatečný počet korelací nad hodnotu 0.30, je faktorová analýza neúčinná.

1. **Bartlettův test sféricity:** ověřuje korelace mezi znaky a testuje

H_0: korelační matice je [redacted] NULOVÁ maticí, která má diagonální prvky jedničky a mimodiagonální prvky nuly.

Testační statistika χ^2 transformace determinantu korelační matice:

(a) Je-li *velká*, je proto nepravděpodobné, že korelační matice je [redacted] NULOVÁ.

(b) Je-li *malá* měli bychom si rozmyslet, zda takový faktorový model vůbec užijeme.

2. **Parciální korelační koeficienty**: indikátorem síly vztahu mezi znaky.

- Sdílejí-li znaky společné faktory, jsou parciální korelační koeficienty mezi páry původních znaků malé.
- Specifické faktory jsou předpokládány vzájemně nekorelované.
- Negativní hodnoty parciálních korelačních koeficientů se nazývají "anti-image" korelace.
- Je-li podíl velkých hodnot koeficientů vysoký, měli bychom zvážit použití faktorové analýzy.

3. **Anti-image korelační matice** je matice parciálních korelací mezi původními znaky po provedené faktorové analýze.

- Představuje stupeň jak faktory vysvětlují povahu znaků.
- Na diagonále jsou hodnoty MSA pro každý znak a mimodiagonální prvky jsou pak parciální korelace mezi znaky.

4. **Kaiser-Meyer-Olkinova míra (KMO):** index k porovnání velikostí experimentálních korelačních koeficientů vůči velikostem parciálních korelačních koeficientů. Vypočte se dle

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2},$$

kde r_{ij} je jednoduchý korelační koeficient mezi i-tým a j-tým znakem a a_{ij} je parciální korelační koeficient mezi i-tým a j-tým znakem.

1. Je-li součet čtverců parciálních korelačních koeficientů mezi všemi páry znaků malý ve srovnání se součtem čtverců jednoduchých korelačních koeficientů, vychází míra *KMO* blízká 1.
2. Malé hodnoty *KMO* indikují, že faktorová analýza těchto původních znaků nebude dobrou myšlenkou, protože korelaci mezi páry znaků nelze vysvětlit pomocí ostatních znaků.

5. **Čtverec vícenásobného korelačního koeficientu R^2** mezi jedným znakem a všemi ostatními představuje kritérium síly lineárního vztahu.

Malá hodnota R^2 značí: oba znaky by měly být z výběru původních znaků spíše odstraněny.

V datech musí být určitá struktura, povaha a charakter, aby byla data vhodná k aplikaci faktorové analýzy. Určitou strukturou je zde myšlena *pouze korelace*, žádná jiná závislost.

4. krok: Nalezené řešení a dosažená těsnost proložení

Po zadání původních znaků rozhodneme

(1) zda použít FA nebo PCA,

(2) počet faktorů nebo hlavních komponent, které *dostatečně* popíší zdrojovou matici dat.

FA versus PCA:

PCA soustředí maximum původní informace do minimálního počtu faktorů k nejlepší predikci,

FA je k určení faktorů, odrážejících původní znaky do společné proměnné, faktoru.

FA rozlišuje tři druhy rozptylů:

(1) ***Společný rozptyl***: rozptylem v proměnné je chápán rozptyl, který je společný se všemi ostatními proměnnými.

(2) ***Specifický rozptyl*** je roztyl vlastní pouze jednomu jedinému původnímu znaku.

(3) ***Chybový rozptyl*** je rozptyl, způsobený nespolehlivostí datového procesu, měřící chyby, nebo náhodnou komponentou proměřovaného jevu.

Těsnost proložení faktorového modelu

slouží k prokázání adekvátnosti nalezených m faktorů v modelu.

Pro velké výběry vedou statistiky testu těsnosti proložení k rozdělení χ^2.

Počet faktorů se zvyšuje dokud není dosaženo rozumného proložení, dokud vypočtená hladina významnosti není malá.

Hodnota χ^2 v testu těsnosti proložení je přímo úměrná velikosti výběru.

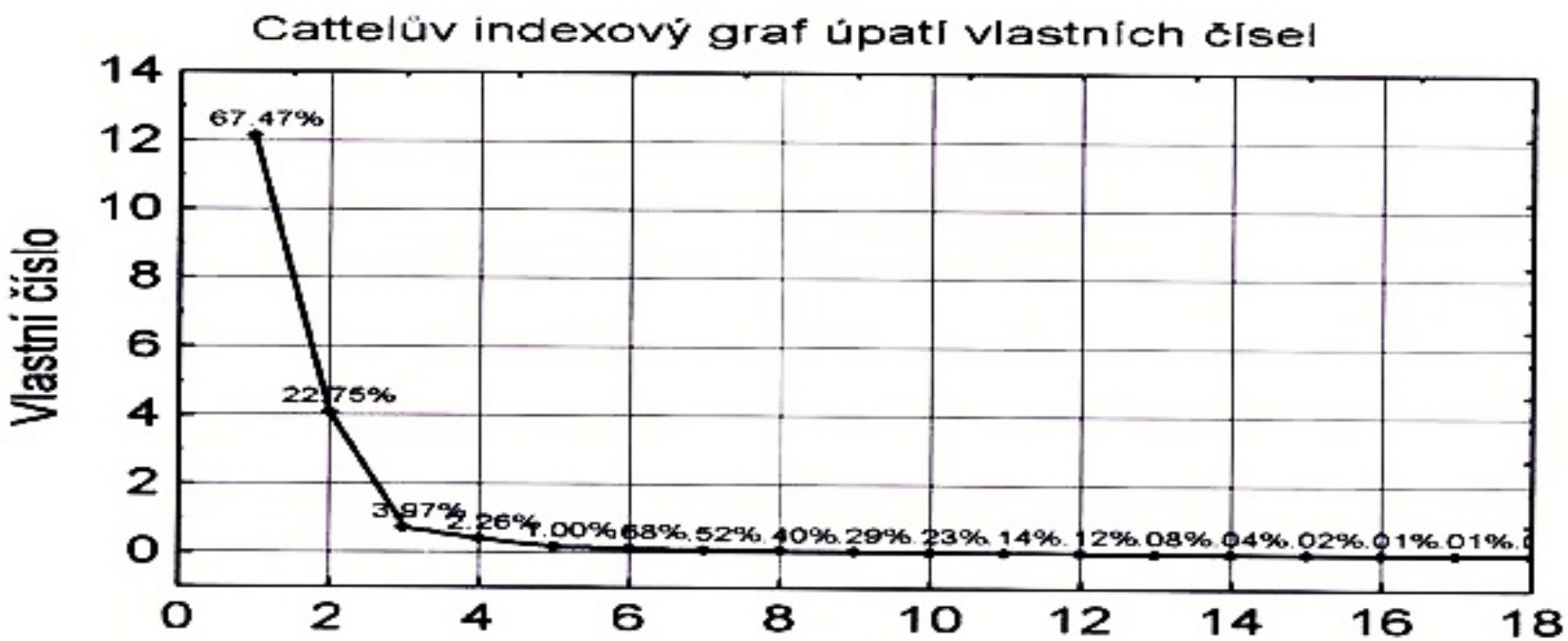
Počet stupňů volnosti je funkcí *pouze* počtu společných faktorů a počtu znaků.

Pravidlo: Aby statistika χ^2 měla kladný počet stupňů volnosti, nesmí počet společných faktorů překročit největší celé číslo m,

$$m < 0.5\ (2P + 1 - \sqrt{8P + 1}),$$

kde m je počet extrahovaných společných faktorů a P je počet původních proměnných.

Počet společných faktorů m je vhodné stanovovat na základě **Cattelova indexového grafu úpatí vlastních čísel.**



5. krok: Interpretace výsledků

Existují tři kroky v interpretaci faktorů:

1. Neotočená faktorová matice slouží:

- k získání počátečního počtu faktorů, které je třeba extrahovat z původních dat.
- obsahuje faktorové zátěže pro každý znak v každém faktoru.
- představuje nejlepší lineární kombinace původních znaků,

První faktor je ojedinělý a představuje nejlepší součet lineárních vztahů, nalezený v původních datech.

Druhý faktor je druhý nejlepší součet lineárních vztahů původních dat, má však omezovací podmínku, musí totiž být orthogonální k prvnímu faktoru.

Ostatní faktory jsou vypočteny podobným způsobem.

2. Faktorová zátěž: představuje vysvětlení role každého původního znaku při definování dotyčného faktoru:

- Jde o korelační koeficient mezi každým znakem a faktorem.
- Rotace faktorů zjednoduší faktorovou strukturu a faktory budou také smysluplnější a dostatečně "faktorově čisté".
- Úprava faktorového modelu z hlediska
  (a) vypuštění nadbytečných znaků z faktorové analýzy,
  (b) zařazení různých rotačních metod ke zlepšení interpretace faktorů,
  (c) potřeby extrahovat různý počet faktorů,
  (d) potřeby změny jedné extrakční metody v jinou.

3. Otáčení faktorů:

Osy faktorů jsou otočeny okolo počátku k dosažení lepší polohy.

Neotočené faktory extrahují proměnlivost dle své důležitosti:

- První faktor je vždy obecným faktorem, ve kterém je zastoupena většina znaků a je proto pokryt co největším objemem rozptylu.
- Druhý a všechny ostatní faktory jsou pak pokryty zbylým rozptylem.
- Otáčení obou faktorů může být *orthogonální* čili o 90° nebo *neorthogonální* o libovolný úhel a osy nemusí být na sebe kolmé.
- Každá z faktorových os je pak těsnější objektům v rovině.
- Vhodnou rotací by mělo dojít ke zjednodušení faktorové struktury, více objektů by mělo být "faktorově čistších".

Orthogonální rotační metody: cílem je zjednodušení faktorové interpretace. Zjednodušením řádků je co největší počet prvků v řádku byl blízký nule.

Prvky budou "faktorově čisté":

Metoda Quartimax: zjednoduší *řádky* faktorové matice, kdy znaky nabijí vysoko jeden faktor a nízko ostatní faktory.

Metoda Varimax: zjednoduší sloupce (na jedničky a nuly). Interpretace je snadnější, když korelace mezi faktorem a původní proměnnou je

(a) blízká buď +1 nebo -1, čímž indikuje těsný vztah pozitivní nebo negativní mezi znakem a faktorem,

(b) blízko nule indikuje naprostý nedostatek tohoto vztahu. Varimax dává lepší rozlišení faktorů než Quartimax.

Metoda Equimax: kompromis mezi Quartimaxem a Varimaxem. Pokouší se zjednodušit oboje, řádky i sloupce.

Výběr metody se provádí zkusmo.

Cvičení v programu STATISTICA

STATISTICA Cz - [Data: 52Spokojenost (11s krát 100ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

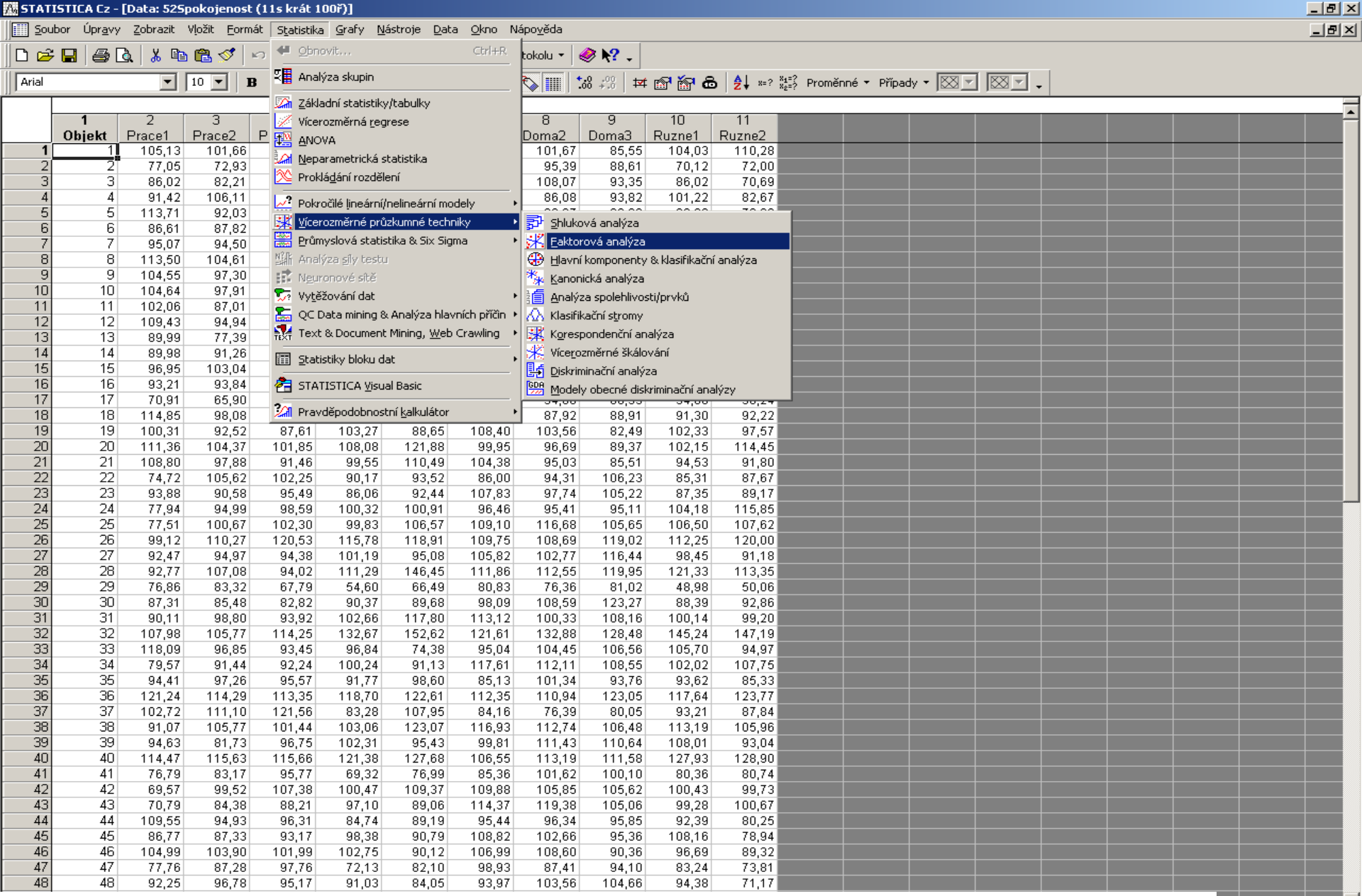
Arial 10 Proměnné Případy

	1 Objekt	2 Prace1	3 Prace2	4 Prace3	5 Hoby1	6 Hoby2	7 Doma1	8 Doma2	9 Doma3	10 Ruzne1	11 Ruzne2
1	1	105,13	101,66	115,06	101,00	95,18	100,28	101,67	85,55	104,03	110,28
2	2	77,05	72,93	77,49	72,74	61,56	93,85	95,39	88,61	70,12	72,00
3	3	86,02	82,21	78,89	77,95	91,71	86,77	108,07	93,35	86,02	70,69
4	4	91,42	106,11	95,64	90,90	111,47	100,25	86,08	93,82	101,22	82,67
5	5	113,71	92,03	99,08	79,28	98,42	104,01	83,27	69,62	82,82	70,02
6	6	86,61	87,82	67,66	93,66	78,00	99,82	97,28	108,62	91,40	79,78
7	7	95,07	94,50	98,08	94,51	97,42	93,69	99,18	96,40	90,73	86,71
8	8	113,50	104,61	105,57	101,01	102,27	87,43	96,66	86,58	93,06	112,70
9	9	104,55	97,30	94,07	88,54	98,11	97,79	99,59	99,76	99,40	105,91
10	10	104,64	97,91	85,82	82,49	90,45	104,69	95,08	99,69	77,63	62,03
11	11	102,06	87,01	94,69	79,20	68,48	79,00	86,43	78,82	89,73	87,55
12	12	109,43	94,94	104,40	119,29	112,99	122,93	114,82	102,11	116,34	101,36
13	13	89,99	77,39	100,77	97,03	111,11	107,66	116,86	114,73	88,71	76,50
14	14	89,98	91,26	89,22	81,25	100,52	94,90	96,72	94,41	88,42	89,67
15	15	96,95	103,04	97,49	98,52	96,83	92,78	110,07	95,16	102,37	105,52
16	16	93,21	93,84	87,35	104,26	115,32	114,29	111,62	111,45	95,70	105,72
17	17	70,91	65,90	74,23	60,11	47,09	79,81	94,80	88,35	54,66	58,24
18	18	114,85	98,08	105,63	90,24	101,01	98,95	87,92	88,91	91,30	92,22
19	19	100,31	92,52	87,61	103,27	88,65	108,40	103,56	82,49	102,33	97,57
20	20	111,36	104,37	101,85	108,08	121,88	99,95	96,69	89,37	102,15	114,45
21	21	108,80	97,88	91,46	99,55	110,49	104,38	95,03	85,51	94,53	91,80
22	22	74,72	105,62	102,25	90,17	93,52	86,00	94,31	106,23	85,31	87,67
23	23	93,88	90,58	95,49	86,06	92,44	107,83	97,74	105,22	87,35	89,17
24	24	77,94	94,99	98,59	100,32	100,91	96,46	95,41	95,11	104,18	115,85
25	25	77,51	100,67	102,30	99,83	106,57	109,10	116,68	105,65	106,50	107,62
26	26	99,12	110,27	120,53	115,78	118,91	109,75	108,69	119,02	112,25	120,00
27	27	92,47	94,97	94,38	101,19	95,08	105,82	102,77	116,44	98,45	91,18
28	28	92,77	107,08	94,02	111,29	146,45	111,86	112,55	119,95	121,33	113,35
29	29	76,86	83,32	67,79	54,60	66,49	80,83	76,36	81,02	48,98	50,06
30	30	87,31	85,48	82,82	90,37	89,68	98,09	108,59	123,27	88,39	92,86
31	31	90,11	98,80	93,92	102,66	117,80	113,12	100,33	108,16	100,14	99,20
32	32	107,98	105,77	114,25	132,67	152,62	121,61	132,88	128,48	145,24	147,19
33	33	118,09	96,85	93,45	96,84	74,38	95,04	104,45	106,56	105,70	94,97
34	34	79,57	91,44	92,24	100,24	91,13	117,61	112,11	108,55	102,02	107,75
35	35	94,41	97,26	95,57	91,77	98,60	85,13	101,34	93,76	93,62	85,33
36	36	121,24	114,29	113,35	118,70	122,61	112,35	110,94	123,05	117,64	123,77
37	37	102,72	111,10	121,56	83,28	107,95	84,16	76,39	80,05	93,21	87,84
38	38	91,07	105,77	101,44	103,06	123,07	116,93	112,74	106,48	113,19	105,96
39	39	94,63	81,73	96,75	102,31	95,43	99,81	111,43	110,64	108,01	93,04
40	40	114,47	115,63	115,66	121,38	127,68	106,55	113,19	111,58	127,93	128,90
41	41	76,79	83,17	95,77	69,32	76,99	85,36	101,62	100,10	80,36	80,74
42	42	69,57	99,52	107,38	100,47	109,37	109,88	105,85	105,62	100,43	99,73
43	43	70,79	84,38	88,21	97,10	89,06	114,37	119,38	105,06	99,28	100,67
44	44	109,55	94,93	96,31	84,74	89,19	95,44	96,34	95,85	92,39	80,25
45	45	86,77	87,33	93,17	98,38	90,79	108,82	102,66	95,36	108,16	78,94
46	46	104,99	103,90	101,99	102,75	90,12	106,99	108,60	90,36	96,69	89,32
47	47	77,76	87,28	97,76	72,13	82,10	98,93	87,41	94,10	83,24	73,81
48	48	92,25	96,78	95,17	91,03	84,05	93,97	103,56	104,66	94,38	71,17
49	49	81,80	85,98	88,06	75,11	87,42	97,98	107,11	112,57	85,93	93,27

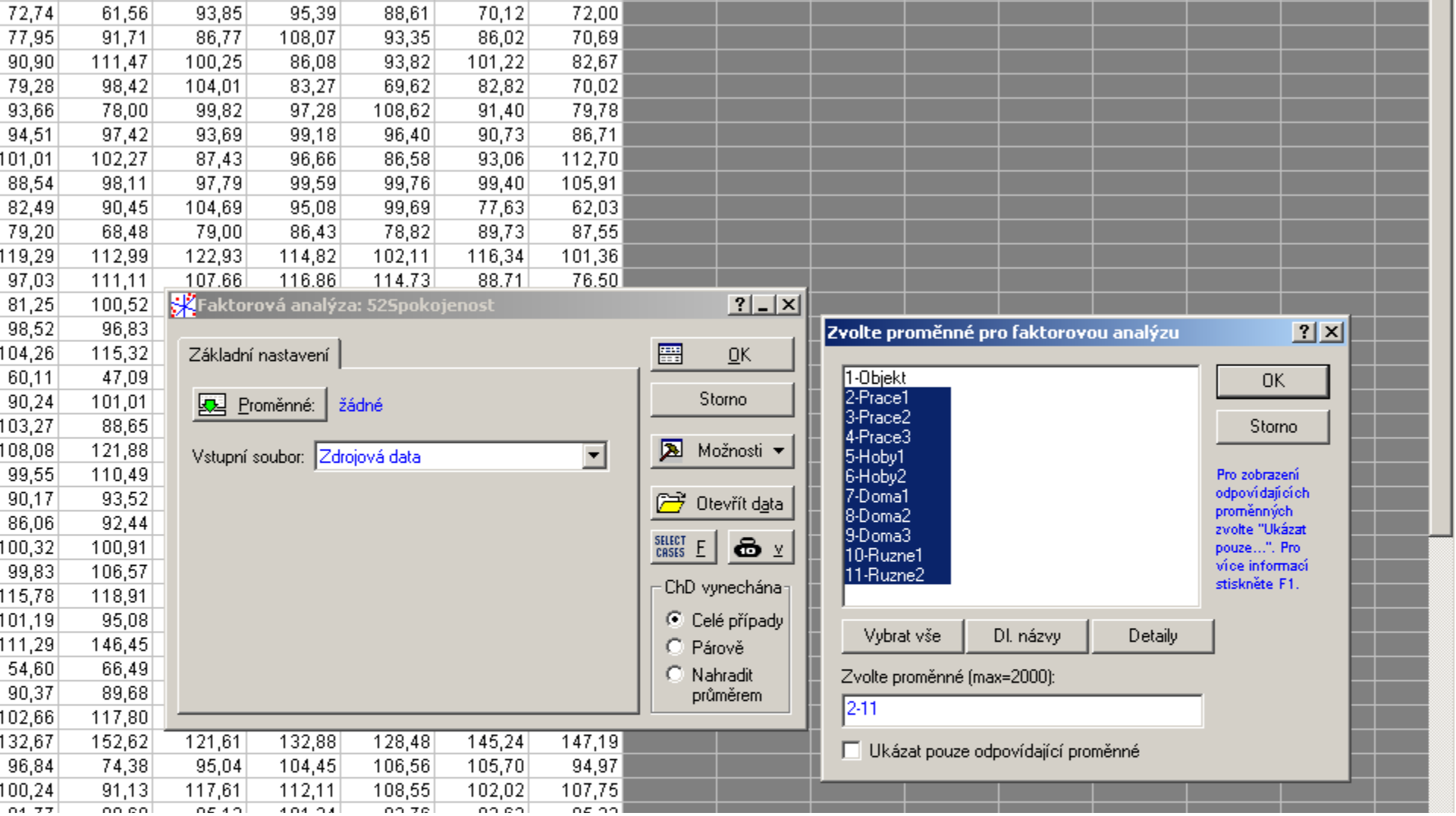
ABC 123 ZÁZN

9:12

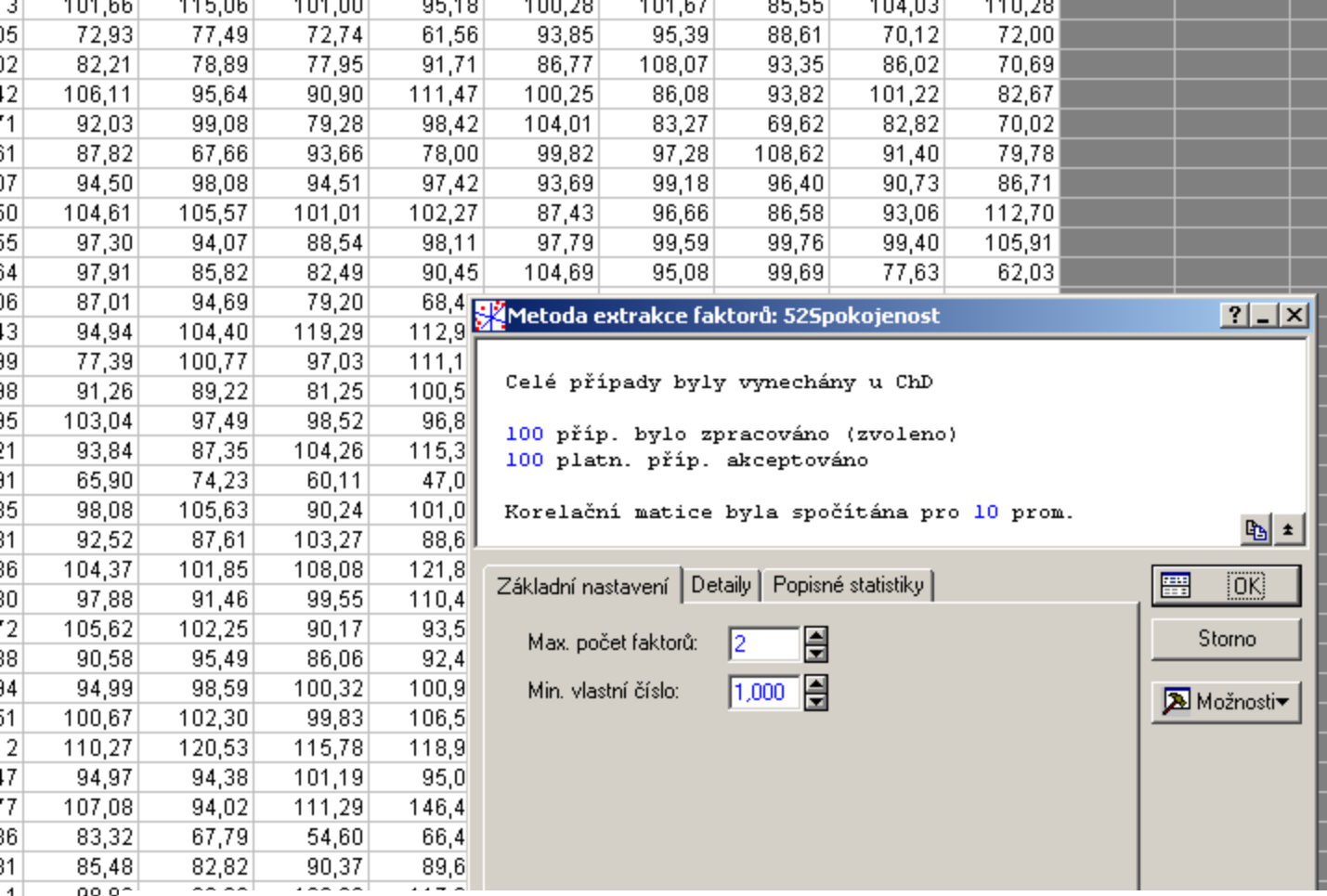
1. Zdrojová matice obsahuje 100 řádků respondentů (objektů) a 10 sloupců znaků s vyjádřením spokojenosti na pracovišti, s koníčky a doma.



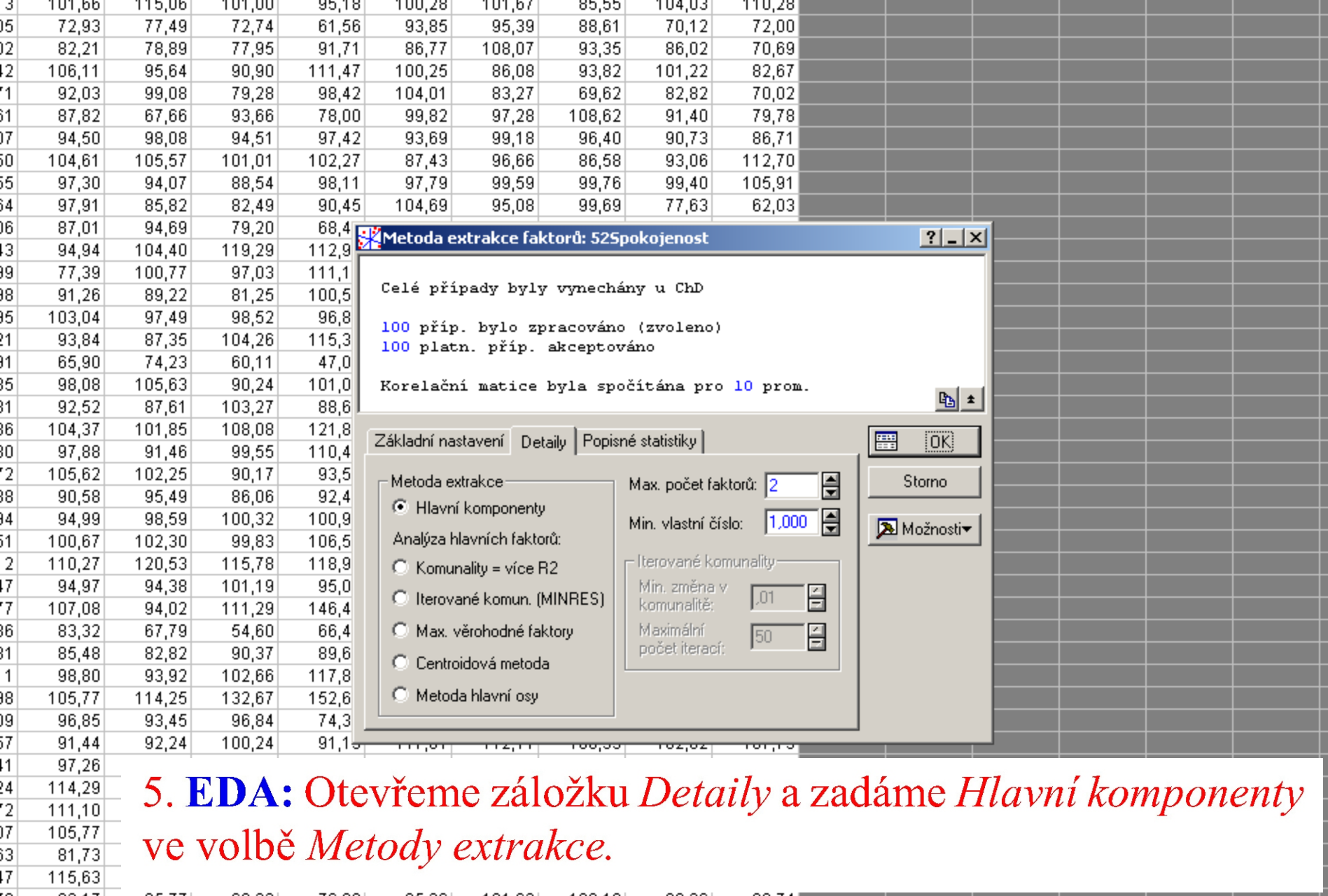
2. Zadá se v bloku *Vícerozměrné průzkumové metody* volba *Faktorové analýzy.*



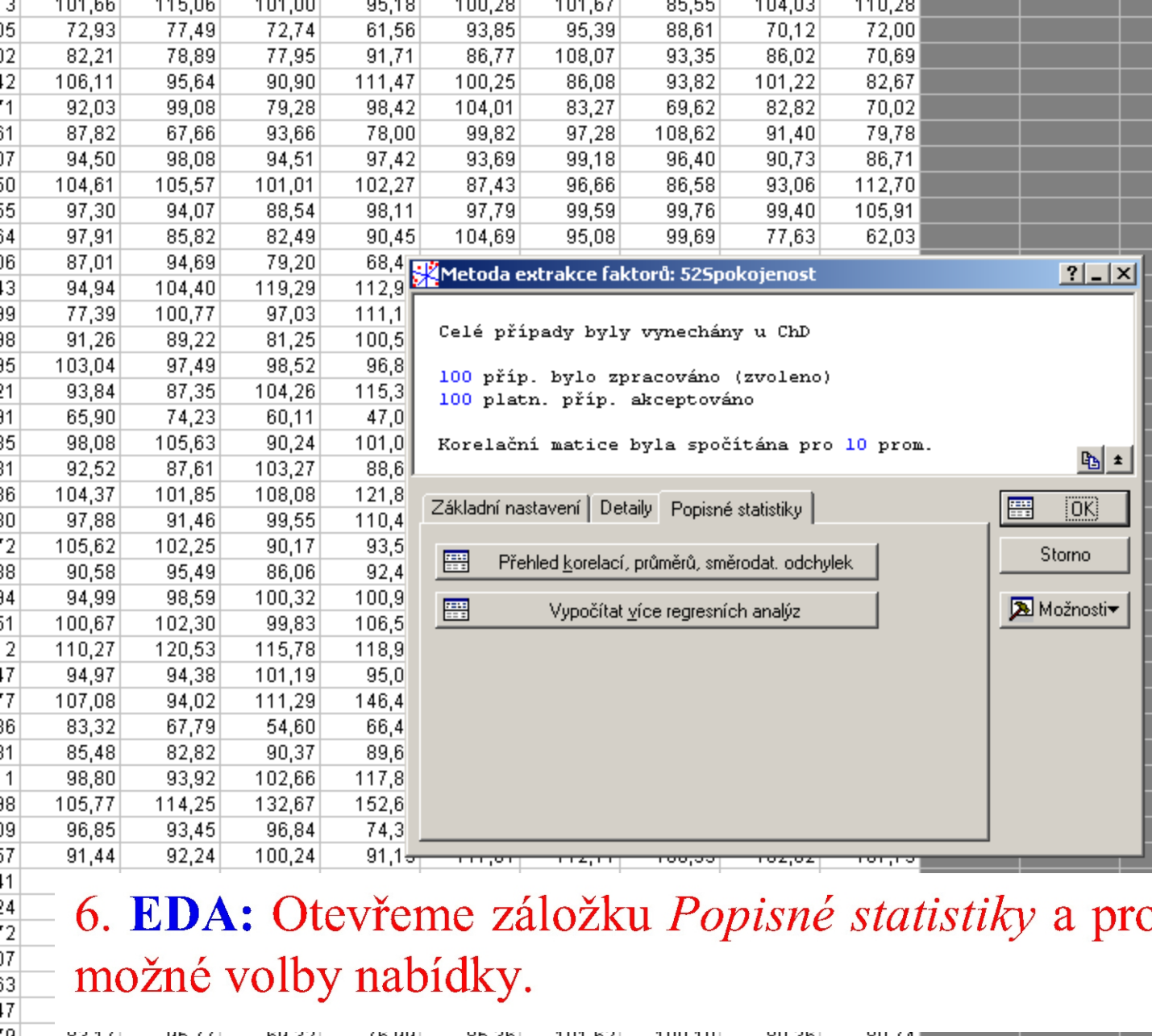
3. **FA:** V záložce *Základní nastavení* otevřeme blok *Proměnné* a zadáme 10 znaků.



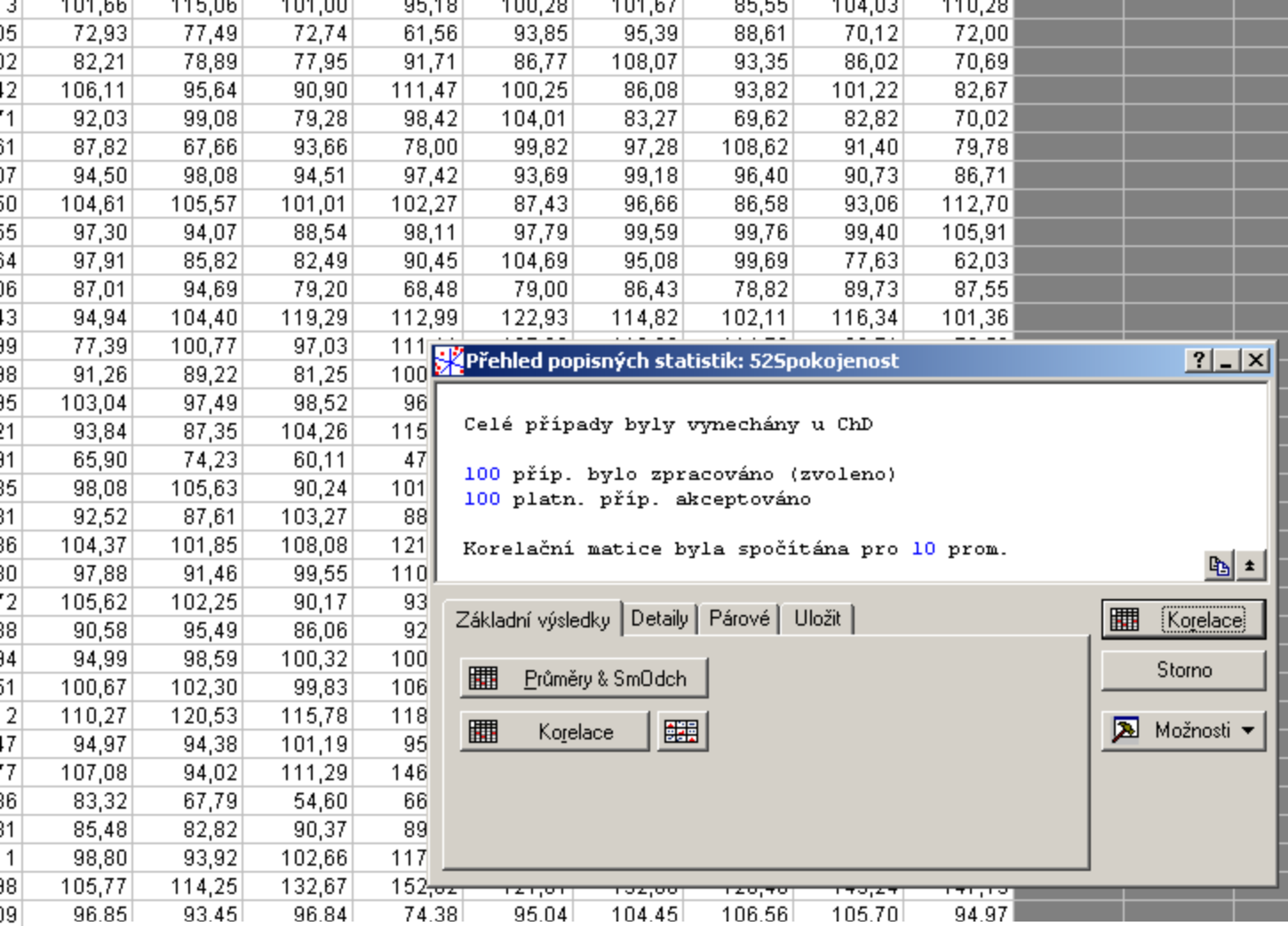
4. **FA:** V okně *Metoda extrakce faktorů* jsou podmínky výpočtu a tři záložky: *Základní nastavení, Detaily* a *Popisné statistiky*. Zůstaneme v záložce *Základní nastavení* a ve volbě *Maximální počet faktorů* kam zadáme číslo 2, v okénku omezení na vlastní čísla zadáme *Minimální vlastní číslo* a zvolíme zde Kaiserovo omezení 1.00.



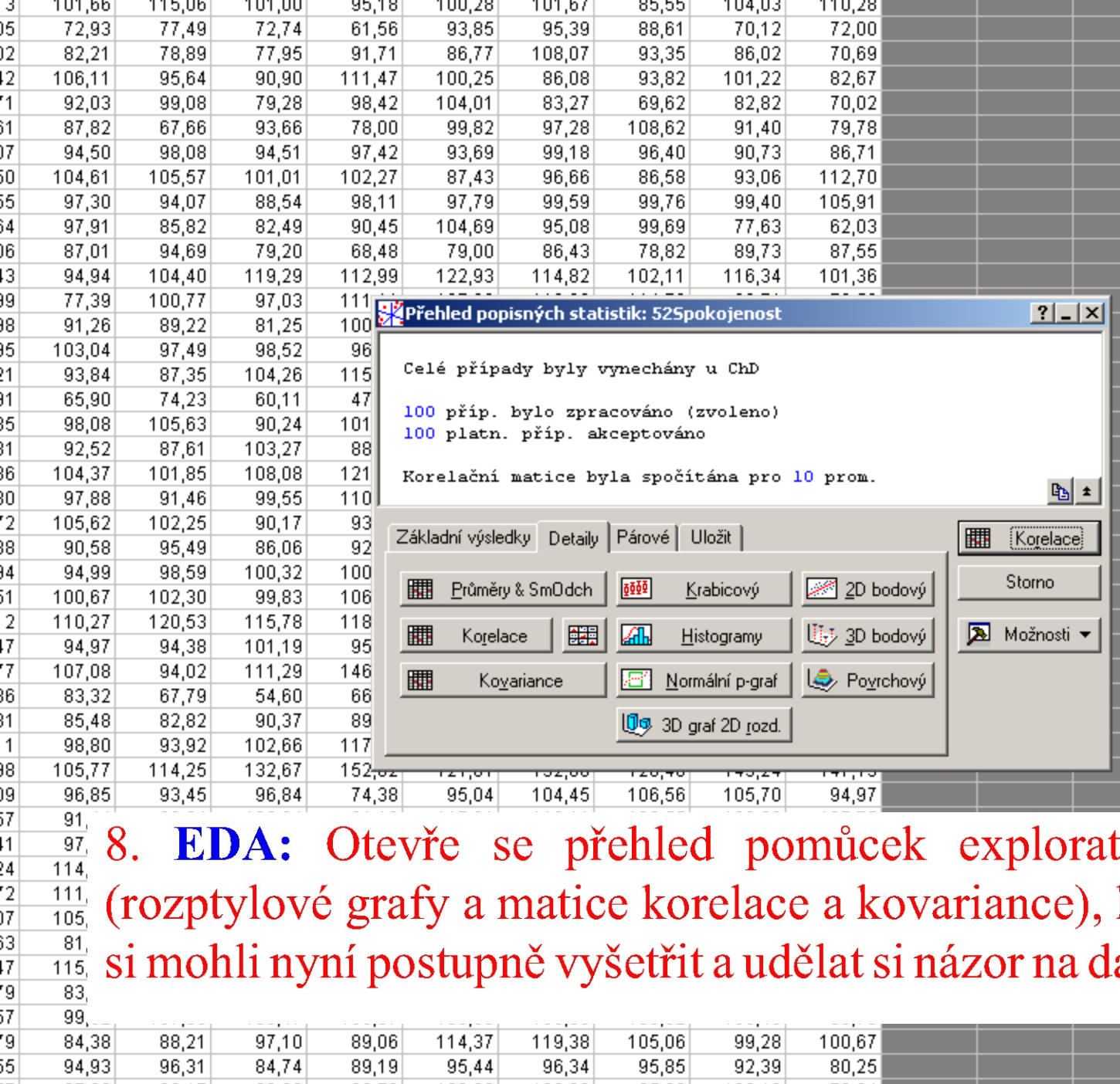
5. **EDA:** Otevřeme záložku *Detaily* a zadáme *Hlavní komponenty* ve volbě *Metody extrakce.*



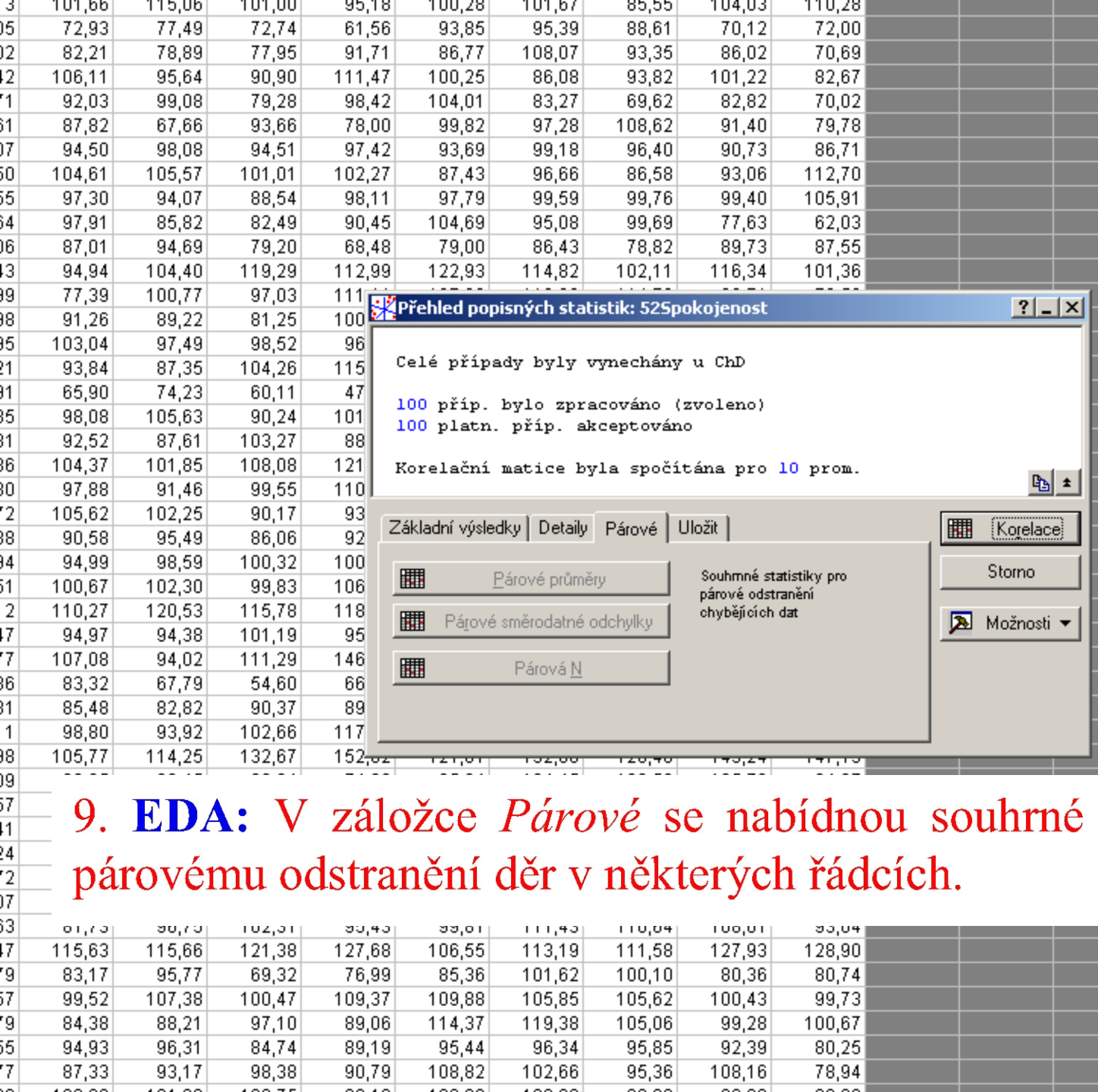
6. **EDA:** Otevřeme záložku *Popisné statistiky* a prohlédneme si možné volby nabídky.



7. **EDA:** Po otevření okna *Popisné statistiky* má okno hlavičku *Přehled popisných statistik* a následují čtyři volby: *Základní výsledky, Detaily, Párové* a *Uložit*. Klikneme pak na *Detaily*.



8. **EDA:** Otevře se přehled pomůcek exploratorní analýzy (rozptylové grafy a matice korelace a kovariance), které bychom si mohli nyní postupně vyšetřit a udělat si názor na data této úlohy.



9. **EDA:** V záložce *Párové* se nabídnou souhrné statistiky k párovému odstranění děr v některých řádcích.

STATISTICA Cz - [PS 1* - Průměry a směrodat. odchylky (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

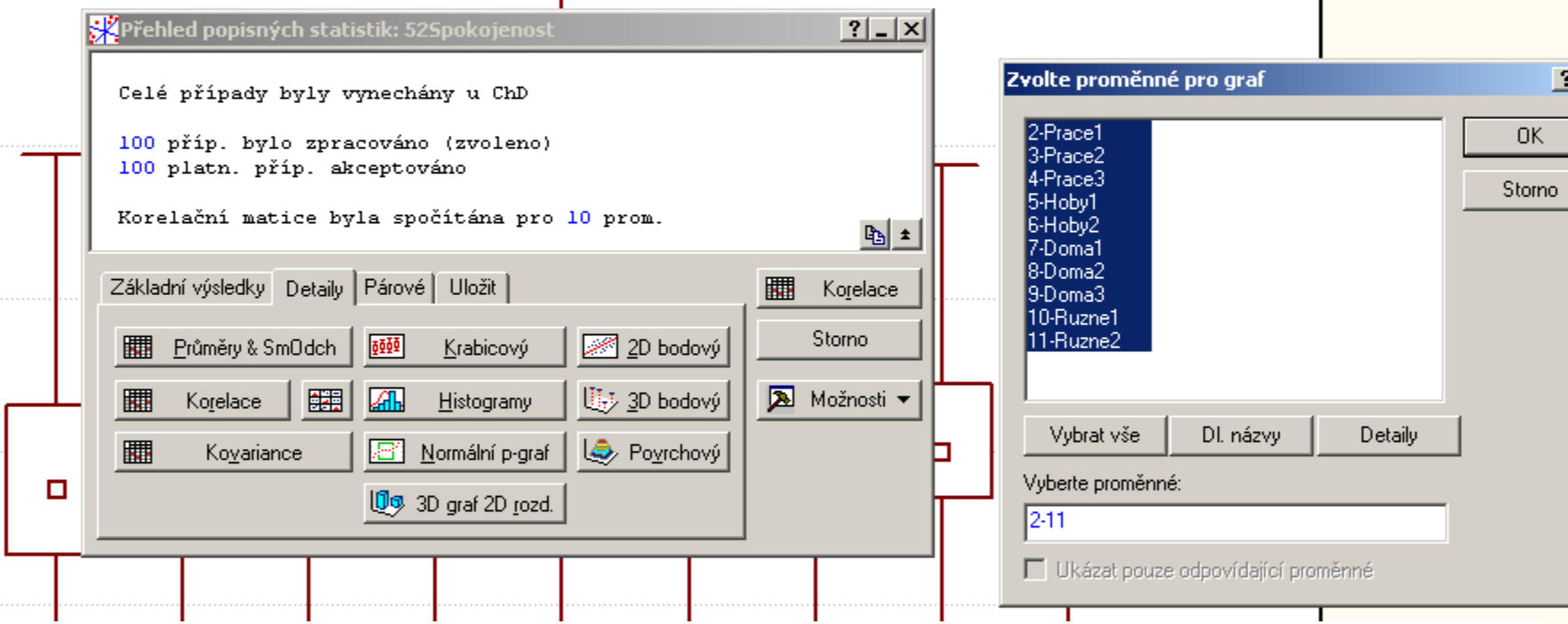
PS 1*
- Faktorová analýza (52Spoko
  - Dialog popisných statistik
    - Průměry a směrodat

Průměry a směrodat. odchylky (52Spokojenost)
ChD vynechána případově
N=100

Proměnná	Průměry	Sm.Odch.
Prace1	97,0296	15,52089
Prace2	98,1689	11,26746
Prace3	98,9430	12,48538
Hoby1	98,0252	15,93673
Hoby2	100,1081	19,94109
Doma1	99,5083	11,97813
Doma2	101,6029	11,06155
Doma3	101,3717	12,73475
Ruzne1	99,1629	16,96749
Ruzne2	98,2238	19,11151

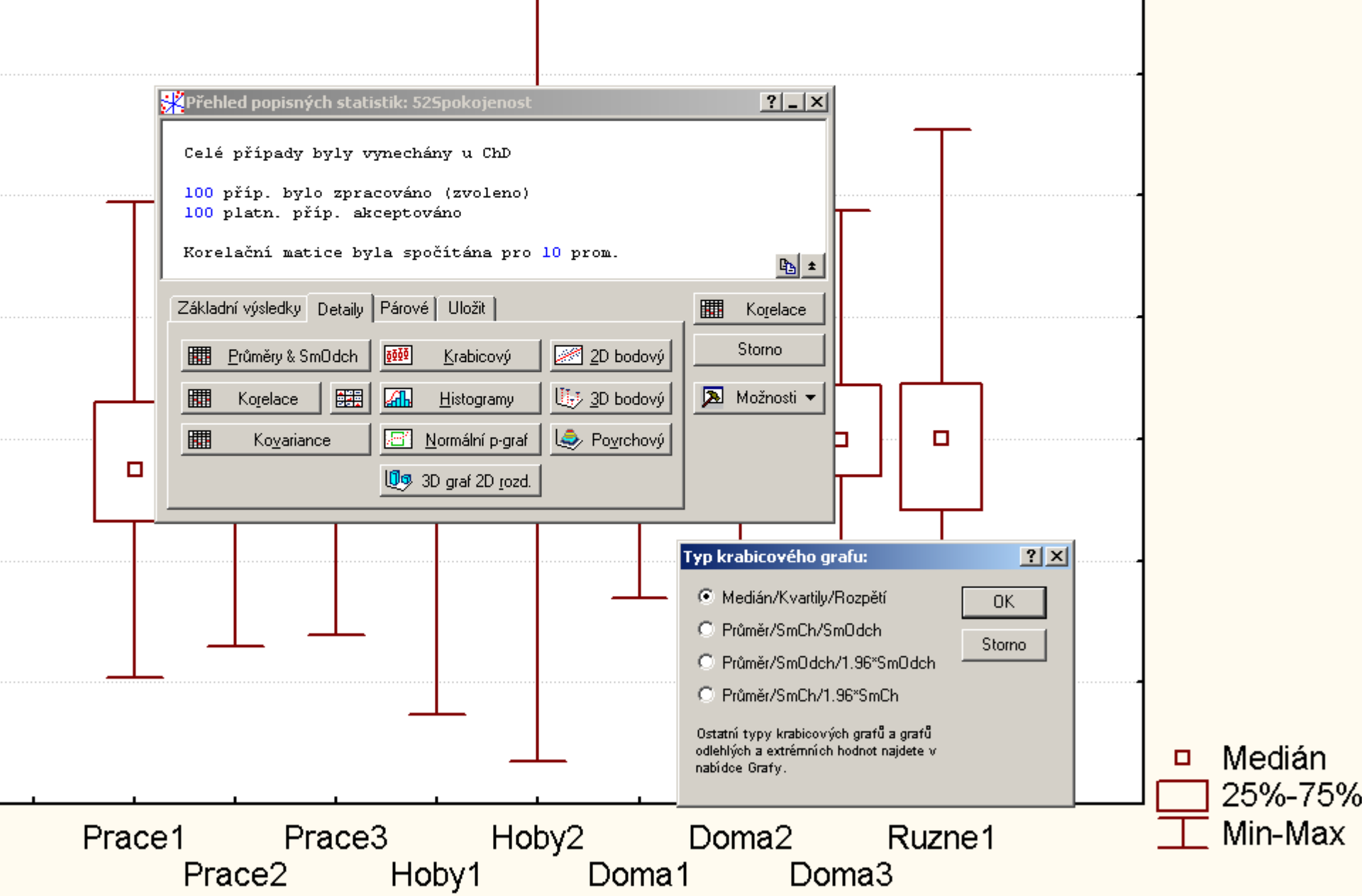
10. **EDA:** Otevření záložky *Průměry a směrodatné odchylky* v záložce *Detaily* obdržíme tabulku těchto dvou měr polohy a rozptýlení.

Krabicový graf

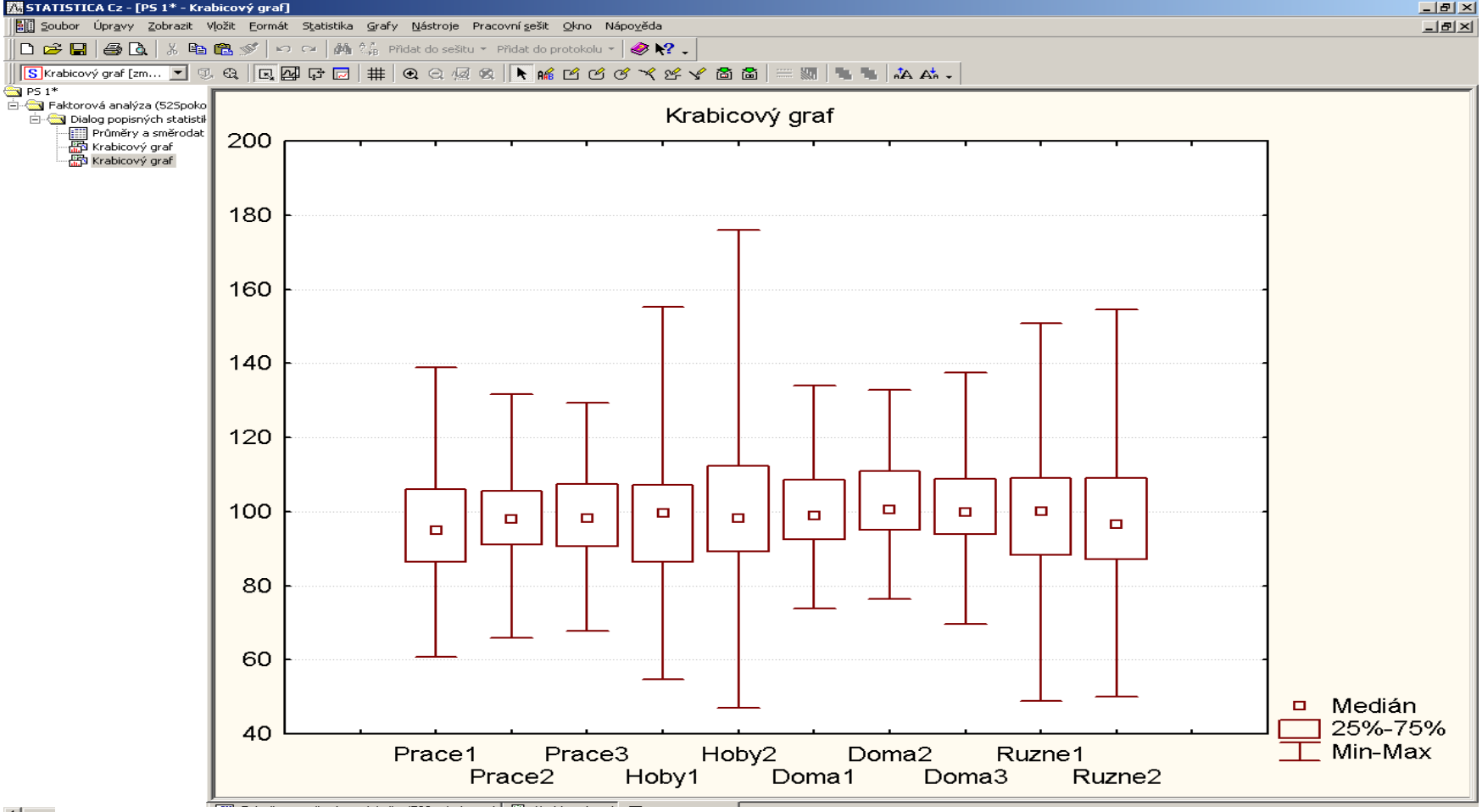


11. **EDA:** zvolíme *Krabicový graf* a v jeho okně zadáme data formou užívaných znaků, určených zde k vyšetření.

Medián



12. **EDA:** U typu krabicového grafu jsou nabídnuty čtyři možnosti měr polohy a rozptýlení této grafické diagnostiky.



13. **EDA:** Krabicový graf odkrývá velikost proměnlivosti u všech znaků a robustních měr polohy, rozptýlení a tvaru. Totiž, zde proměnné s malou či nejmenší proměnlivostí (zde Doma1, Doma2) jsou obvykle určeny k vyškrtnutí z další faktorové analýzy, protože v analýze nepomohou ale chovají se spíše jako numerický šum.

STATISTICA Cz - [PS 1* - Korelace (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

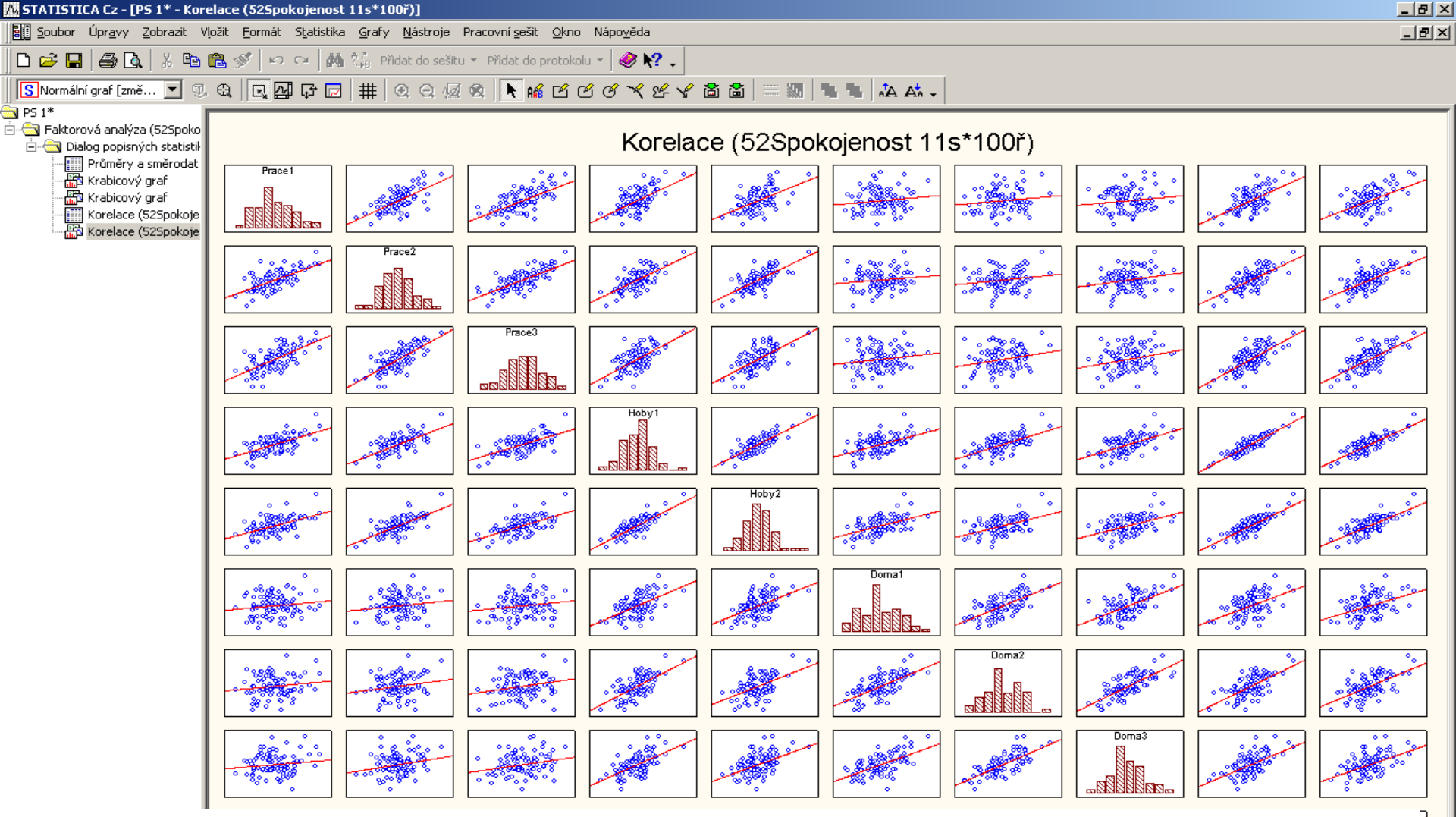
PS 1*
- Faktorová analýza (52Spoko
  - Dialog popisných statistik
    - Průměry a směrodat
    - Krabicový graf
    - Krabicový graf
    - Korelace (52Spokoje

Korelace (52Spokojenost)
ChD vynechána případově
N=100

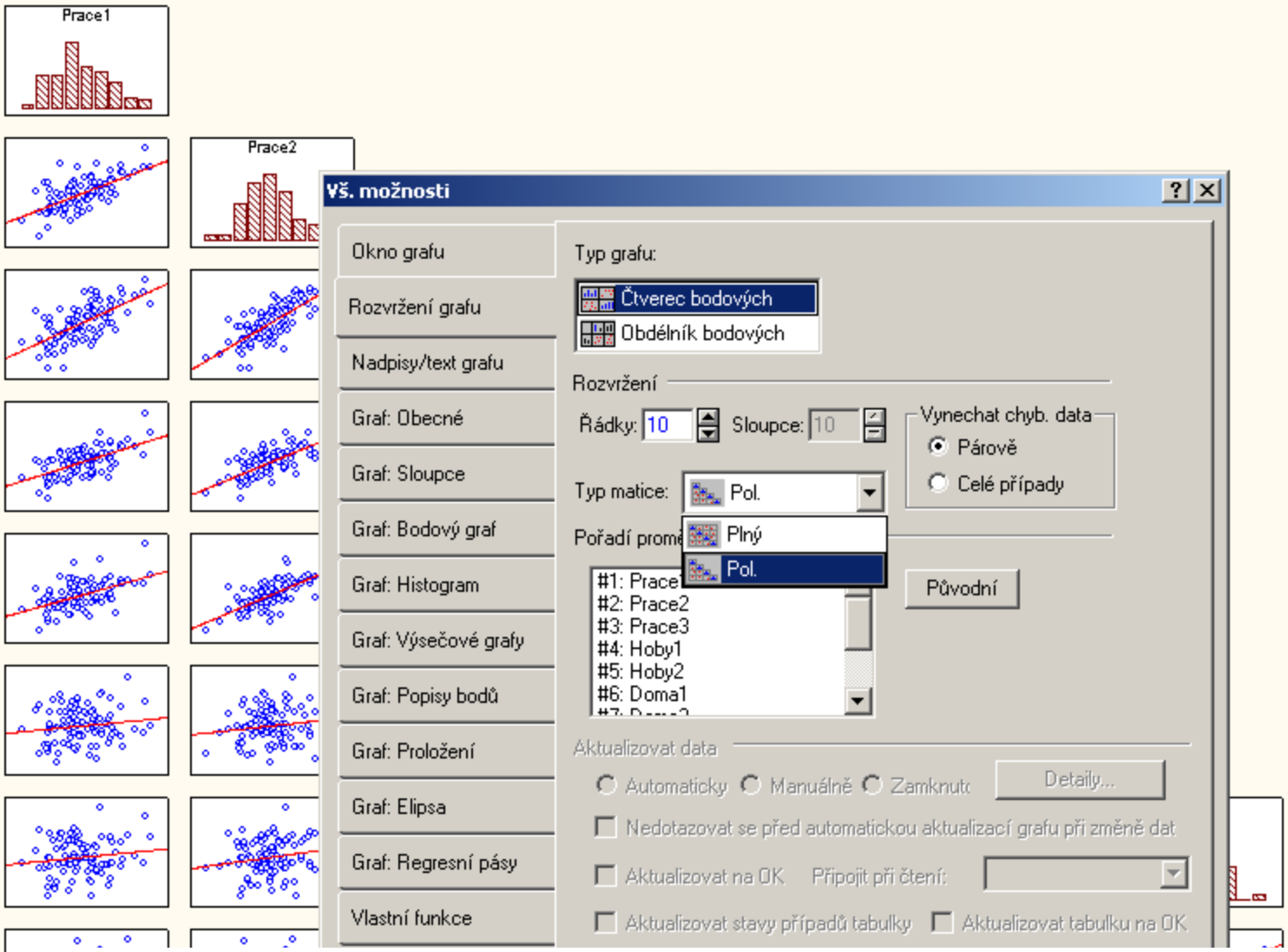
Proměnná	Prace1	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	1,00	0,65	0,65	0,60	0,52	0,14	0,15	0,14	0,61	0,55
Prace2	0,65	1,00	0,73	0,69	0,70	0,14	0,18	0,24	0,71	0,68
Prace3	0,65	0,73	1,00	0,64	0,63	0,16	0,24	0,25	0,70	0,67
Hoby1	0,60	0,69	0,64	1,00	0,80	0,54	0,63	0,58	0,90	0,84
Hoby2	0,52	0,70	0,63	0,80	1,00	0,51	0,50	0,48	0,81	0,76
Doma1	0,14	0,14	0,16	0,54	0,51	1,00	0,66	0,59	0,50	0,42
Doma2	0,15	0,18	0,24	0,63	0,50	0,66	1,00	0,73	0,64	0,59
Doma3	0,14	0,24	0,25	0,58	0,48	0,59	0,73	1,00	0,59	0,52
Ruzne1	0,61	0,71	0,70	0,90	0,81	0,50	0,64	0,59	1,00	0,84
Ruzne2	0,55	0,68	0,67	0,84	0,76	0,42	0,59	0,52	0,84	1,00

14. **EDA:** Okno *Korelace* v záložce *Detaily* okna *Přehled popisných statistik* otevře *Maticový diagram* všech znaků. Obsahuje Pearsonovy korelační koeficienty a ukazuje na největší (zde 0.9) a nejmenší (zde 0.14) korelaci mezi znaky. Tím zřetelně odkrývá dosud skrytou strukturu v datech.

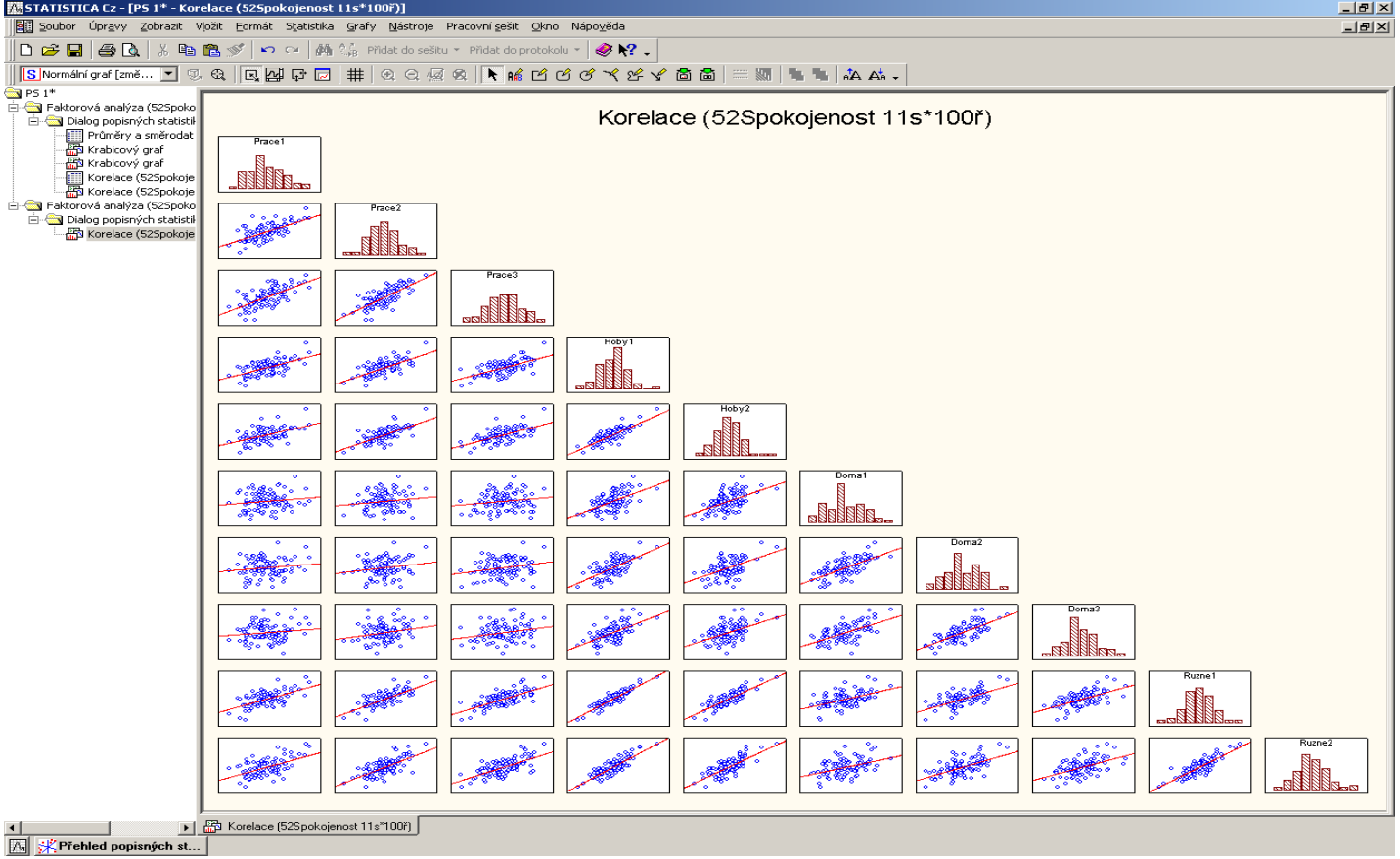
Pravidlo: ***Když by byla korelace všech znaků malá a statisticky nevýznamná, nelze faktorovou analýzu vůbec použít, protože v datech není žádná skrytá struktura.***



15. **EDA:** Kliknutím na ikonu diagramu korelace se otevře *Maticový diagram*. V něm červené přímky rovnoběžné s osou *x* upozorňují na malou a nevýznamnou korelaci této právě analyzované dvojice znaků.



16. **EDA:** Po kliknutí kdekoliv do plochy diagramu se otevře okno *Všeobecné možnosti*, ve kterém v záložce *Rozvržení grafu* vybereme u *Typ matice* volbu zobrazení pouze dolní poloviny diagramu pod diagonálou histogramů, a to *Pol.*



17. **EDA:** Tato podoba maticového diagramu je pak přehlednější. Soustředíme se na vyhledávání červených rovnoběžek s osou x v mraku modrých bodů. Požadovaná nenulová směrnice červené přímky ukazuje zde jednak na rozmístění bodů okolo přímky, a jednak na potřebnou korelaci u této dvojice znaků. Tato statisticky ještě významná korelace je totiž důležitým předpokladem dalšího využití faktorové analýzy.

STATISTICA Cz - [PS 1* - Kovariance (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

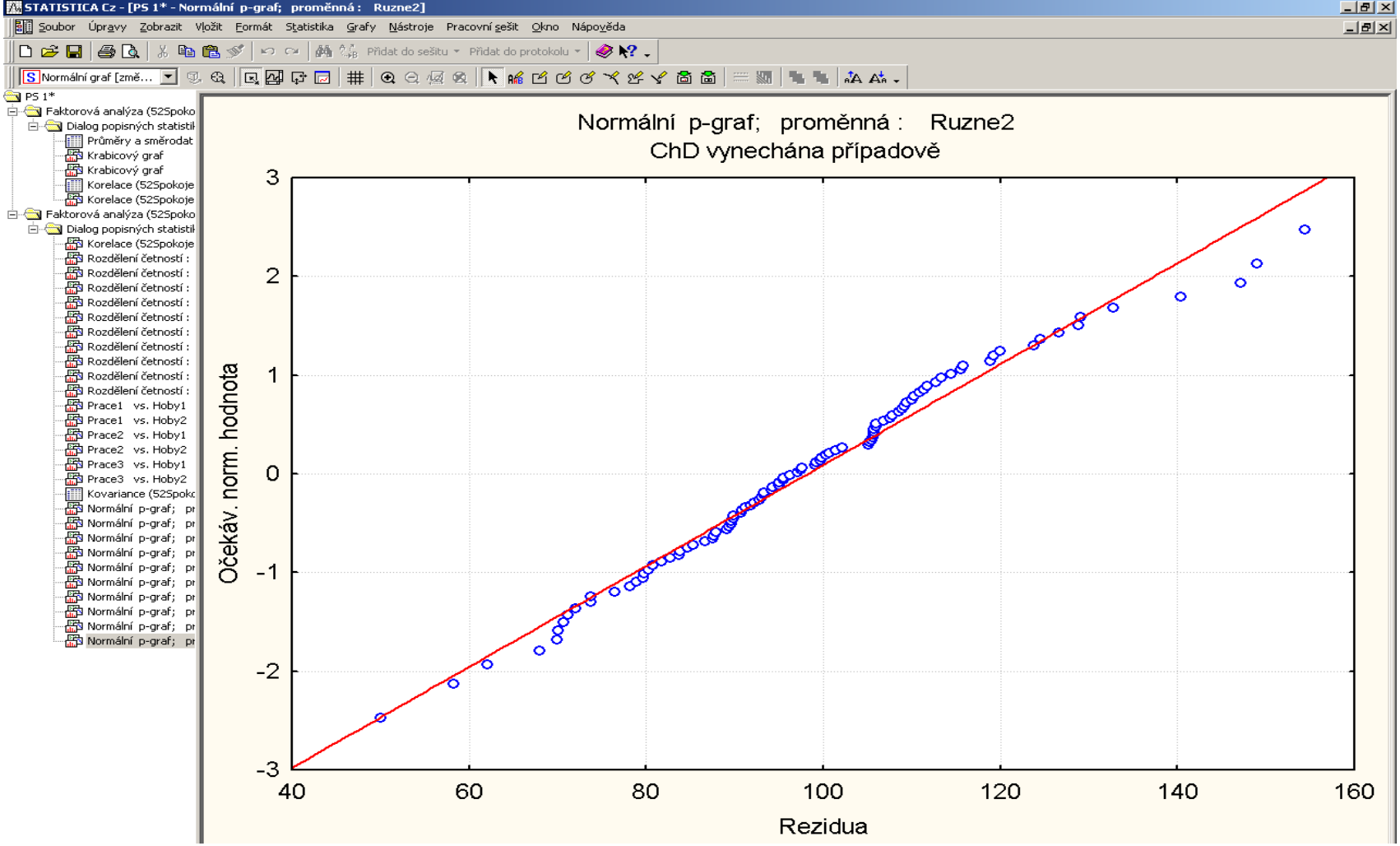
Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog popisných statistik
    - Průměry a směrodat
    - Krabicový graf
    - Krabicový graf
    - Korelace (52Spokoje
    - Korelace (52Spokoje
- Faktorová analýza (52Spoko
  - Dialog popisných statistik
    - Korelace (52Spokoje
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Rozdělení četností :
    - Prace1 vs. Hoby1
    - Prace1 vs. Hoby2
    - Prace2 vs. Hoby1
    - Prac
    - Prac
    - Prac
    - Prac
    - Kova

Kovariance (52Spokojenost)
ChD vynechána případově
N=100

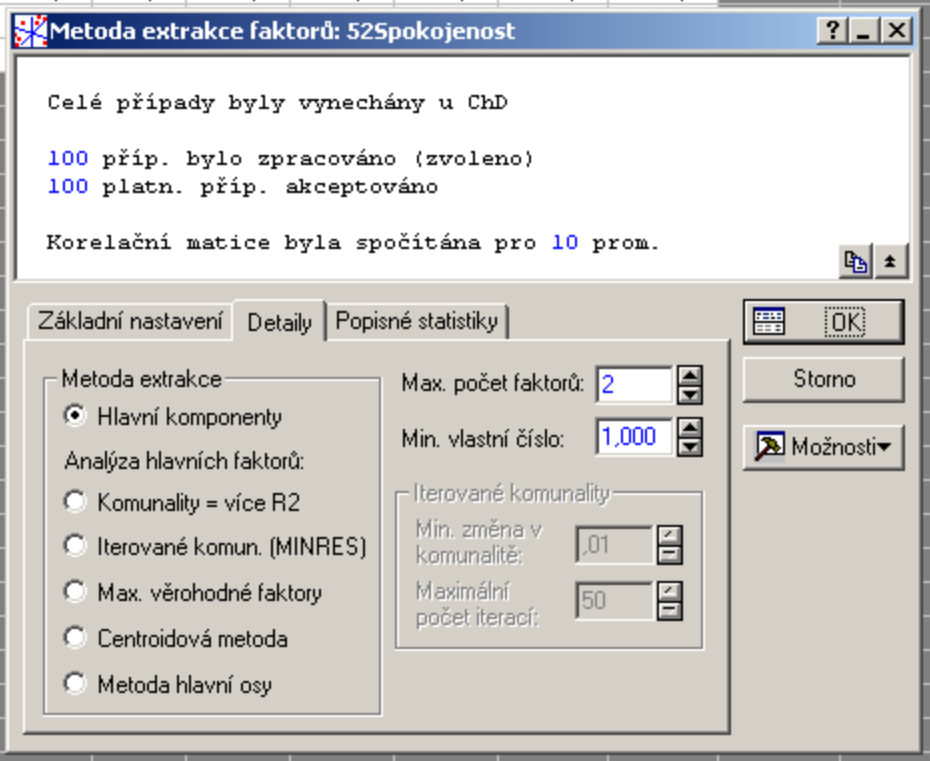
Proměnná	**Prace1**	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	241	113	126	148	161	27	25	27	161	163
Prace2	113	127	103	124	157	19	23	34	135	147
Prace3	126	103	156	127	157	24	33	40	148	160
Hoby1	148	124	127	254	256	102	112	118	245	257
Hoby2	161	157	157	256	398	121	109	122	274	288
Doma1	27	19	24	102	121	143	87	90	101	97
Doma2	25	23	33	112	109	87	122	103	121	125
Doma3	27	34	40	118	122	90	103	162	127	126
Ruzne1	161	135	148	245	274	101	121	127	288	273
Ruzne2	163	147	160	257	288	97	125	126	273	365

18. **EDA:** Kovarianční matice znaků není k našemu vyšetřování natolik přehledná a užitečná jako její normovaná podoba zvaná korelační matice. Kovarianční matice totiž obsahuje hodnoty kovariance v původních jednotkách směsných rozměrů. Můžeme zde posoudit jenom výskyt malých a velkých hodnot kovariance.



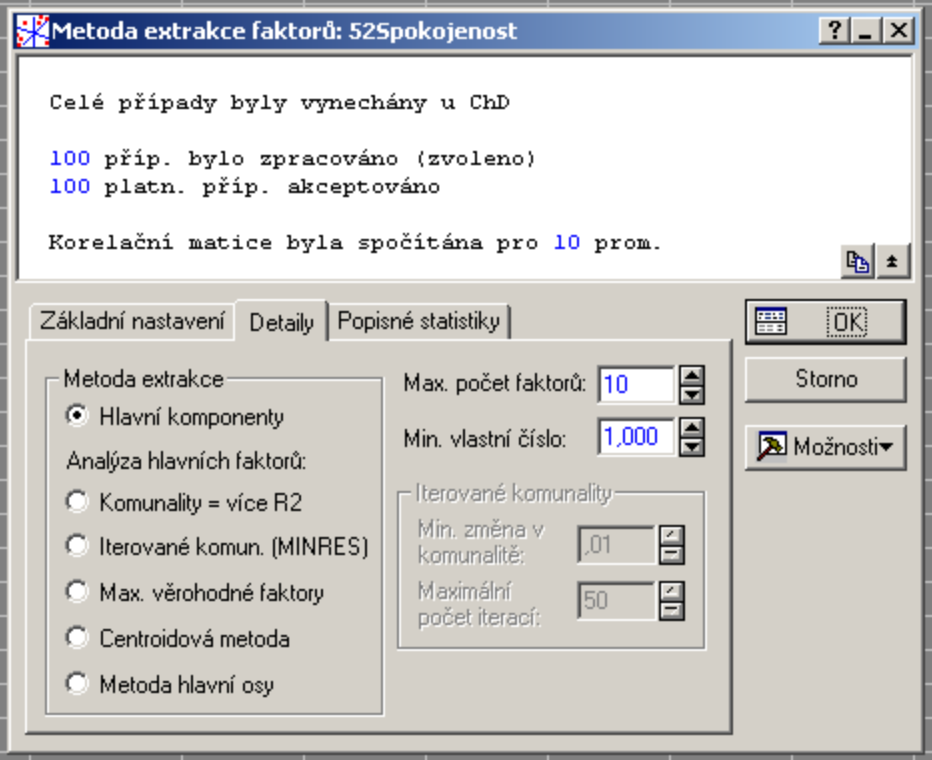
19. **EDA:** Protože je těsnost proložení použitým faktorovým modelem mírou věrohodnosti navrženého FA modelu, podrobíme rezidua u všech znaků detailní statistické analýze. Užijeme například rankitový Q-Q graf, který vyšetřuje normalitu, vybočující body a symetrii reziduí.

Proměnná	Prace1	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	1,00	0,65	0,65	0,60	0,52	0,14	0,15	0,14	0,61	0,55
Prace2	0,65	1,00	0,73	0,69	0,70	0,14	0,18	0,24	0,71	0,68
Prace3	0,65	0,73	1,00	0,64	0,63	0,16	0,24	0,25	0,70	0,67
Hoby1	0,60	0,69	0,64	1,00	0,80	0,54	0,63	0,58	0,90	0,84
Hoby2	0,52	0,70	0,63	0,80	1,00	0,51	0,50	0,48	0,81	0,76
Doma1	0,14	0,14	0,16	0,54	0,51	1,00	0,66	0,59	0,50	0,42
Doma2	0,15	0,18	0,24	0,63	0,50	0,66	1,00	0,73	0,64	0,59
Doma3	0,14	0,24	0,25	0,58	0,48	0,59	0,73	1,00	0,59	0,52
Ruzne1	0,61	0,71	0,70							
Ruzne2	0,55	0,68	0,67							



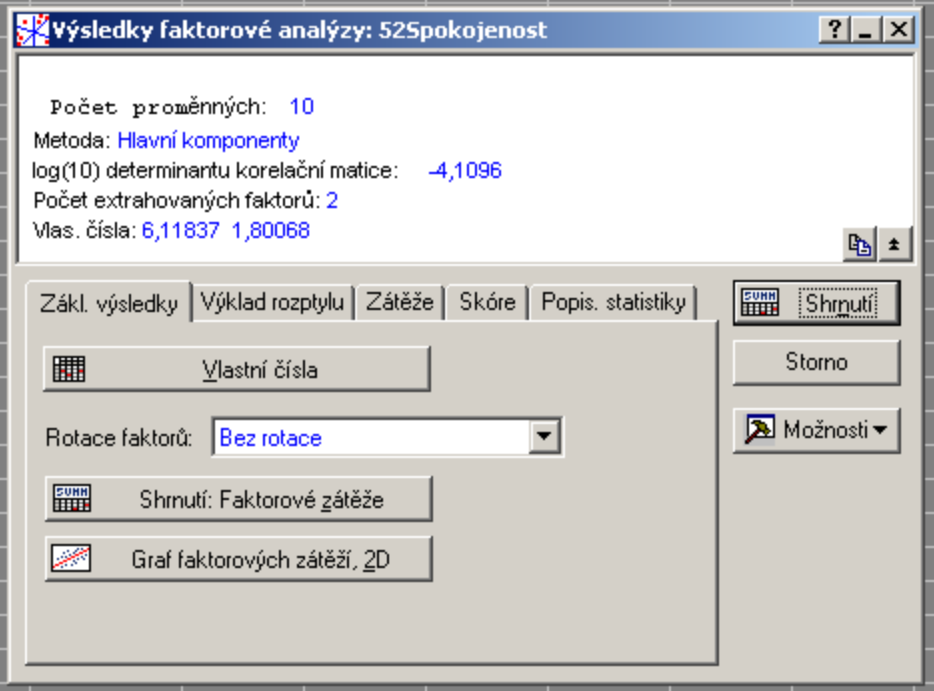
20. **FA:** V záložce *Detaily* bloku *Metoda extrakce faktorů* je možné zadat *Maximální počet faktorů* roven 2 ale také jiné číslo, například původní počet znaků 10.

Hodn.		rozptylu	vlast. číslo	%
1	6,118369	61,18369	6,118369	61,18369
2	1,800682	18,00682	7,919051	79,19051



21. **FA:** Ukazuje se, že výhodnější je zadat *Maximálný počet faktorů* roven 10.

Hodn.		rozptylu	vlast. číslo	%
1	6,118369	61,18369	6,118369	61,18369
2	1,800682	18,00682	7,919051	79,19051



22. **FA:** Otevře se okno *Výsledky faktorové analýzy*, ve kterém v záložce *Základní výsledky* otevřeme *Vlastní čísla* a v okénku *Rotace faktorů* ponecháme zatím ještě: *Bez rotace.*

STATISTICA Cz - [PS 1* - Vl. čísla (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen

Vl. čísla (52Spokojenost)
Extrakce: Hlavní komponenty

Hodn.	vl. číslo	% celk. rozptylu	Kumulativ. vlast. číslo	Kumulativ. %
1	6,118369	61,18369	6,118369	61,18369
2	1,800682	18,00682	7,919051	79,19051

23. **FA:** *Vlastní čísla:* Tabelární podoba Cattelova indexového diagramu vlastních čísel ukazuje na zvolený počet faktorů pomocí zlomu v diagramu vlastních čísel, a tím i na velikost měr vystižené datové proměnlivosti formou zvoleného počtu faktorů.

STATISTICA Cz - [PS 1* - Faktor. zátěže (Bez rot.) (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

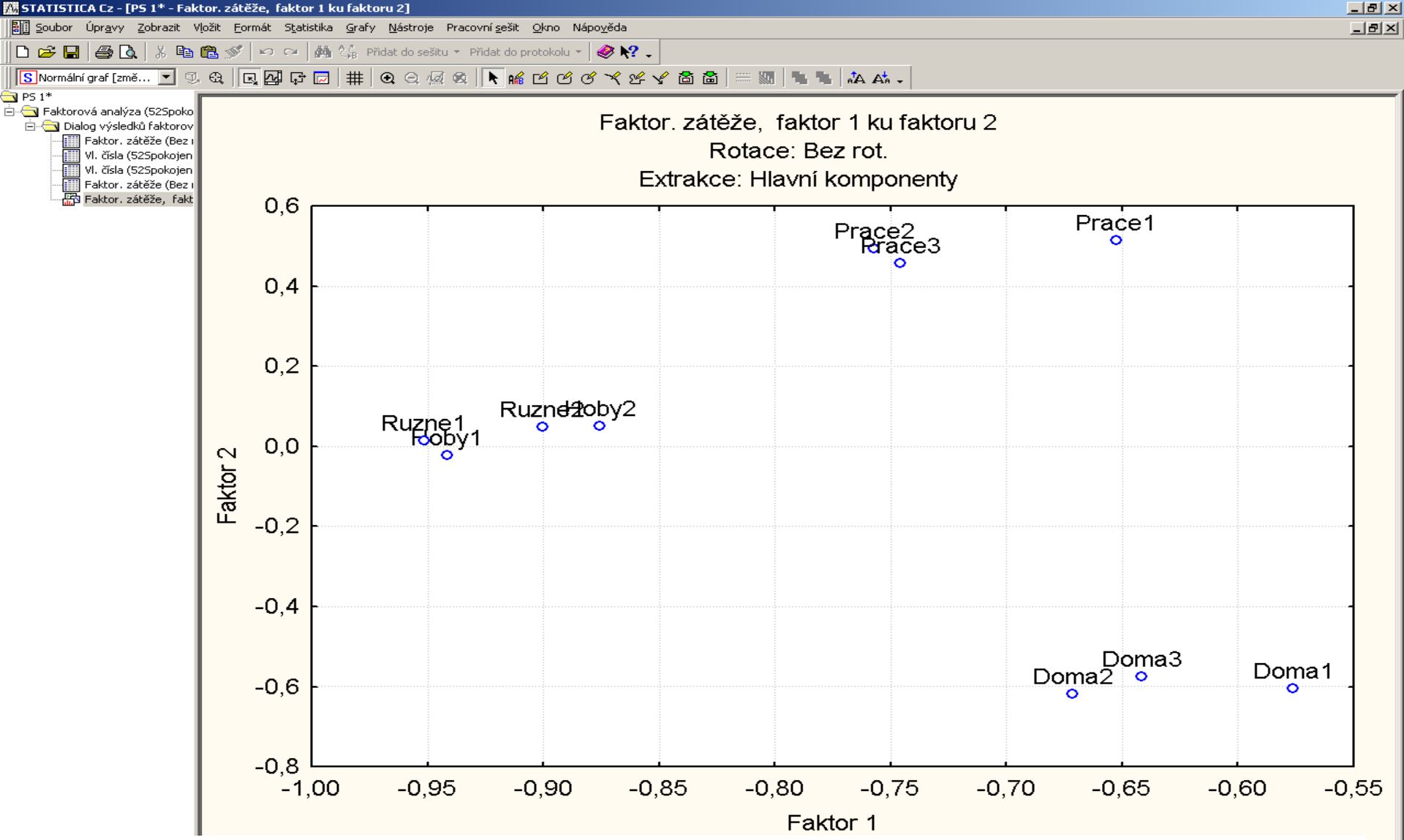
Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez

Faktor. zátěže (Bez rot.) (52Spokojenost)
Extrakce: Hlavní komponenty
(Označené zatěže jsou >,700000)

Proměnná	**Faktor 1**	Faktor 2
Prace1	-0,652601	0,514217
Prace2	-0,756976	0,494770
Prace3	-0,745706	0,456680
Hoby1	-0,941630	-0,021835
Hoby2	-0,875615	0,051643
Doma1	-0,576062	-0,604977
Doma2	-0,671289	-0,617962
Doma3	-0,641532	-0,573925
Ruzne1	-0,951516	0,013513
Ruzne2	-0,900333	0,048154
Výkl.roz	6,118369	1,800682
Prp.celk	0,611837	0,180068

24. **FA:** *Faktorové váhy (zátěže)*: od prvních dvou faktorů ukazují faktorové váhy, jak se jednotlivé znaky podílejí na vytvoření dotyčného faktoru. **Červeně** jsou vyznačeny ještě statisticky významné hodnoty, **černě** už nevýznamné.



25. **FA:** *Graf faktorových vah:* graf třídí znaky v rovině. Blízko sebe umístěné znaky jsou si vzájemně podobné a silně spolu korelují. Vzdáleně umístěné nikoliv.

STATISTICA Cz - [PS 1* - Komunality (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

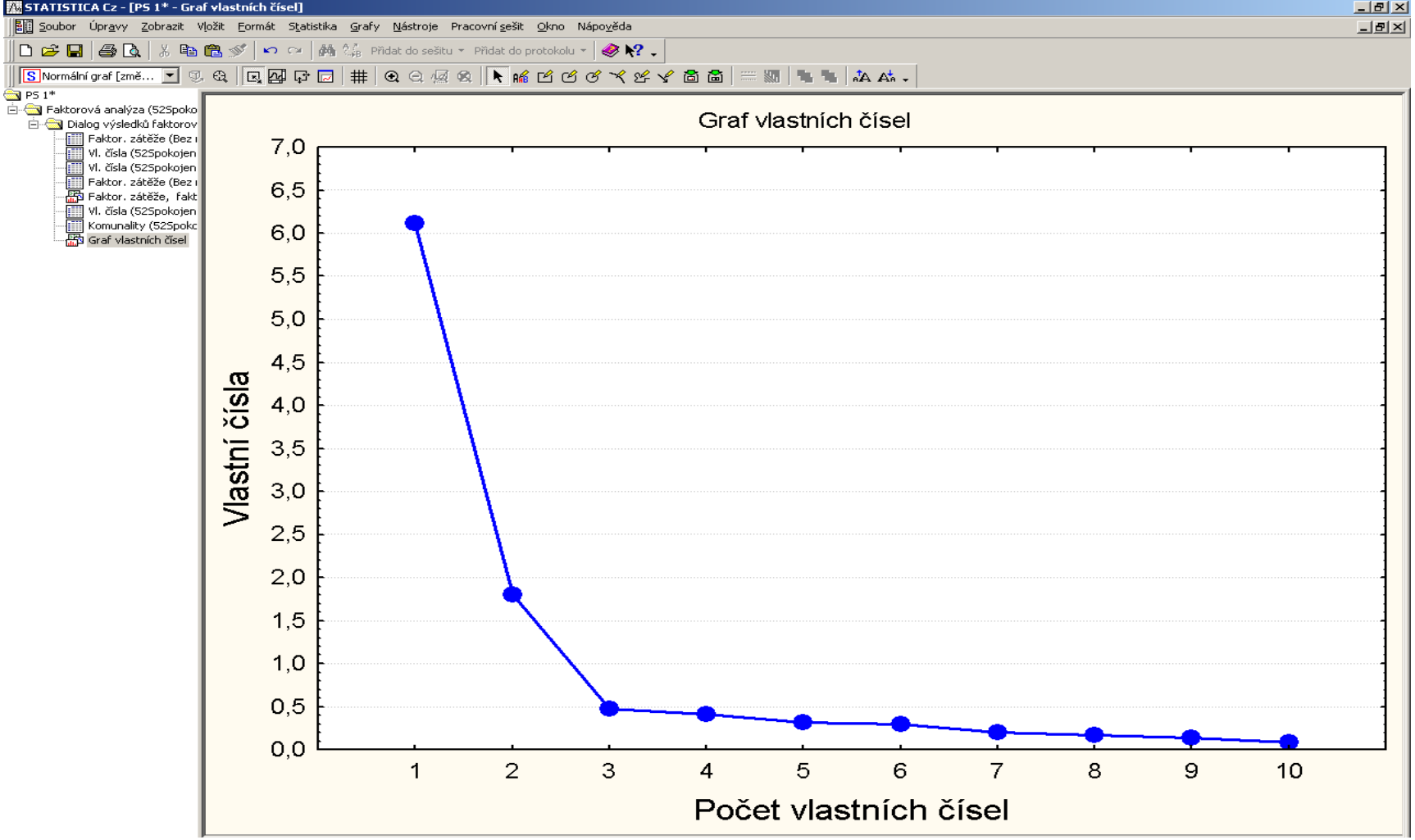
Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez
    - Faktor. zátěže, fakt
    - Vl. čísla (52Spokojen
    - Komunality (52Spokc

Komunality (52Spokojenost)
Extrakce: Hlavní komponenty
Rotace: Bez rot.

Proměnná	Z 1 faktoru	Z 2 faktorů	Více R^2
Prace1	0,425888	0,690307	0,559765
Prace2	0,573012	0,817809	0,734614
Prace3	0,556077	0,764634	0,653962
Hoby1	0,886667	0,887143	0,865921
Hoby2	0,766702	0,769369	0,739078
Doma1	0,331847	0,697844	0,541367
Doma2	0,450630	0,832506	0,738620
Doma3	0,411564	0,740954	0,584202
Ruzne1	0,905382	0,905565	0,884012
Ruzne2	0,810599	0,812918	0,779456

26. **FA:** *Příspěvky daného faktoru do komunality:* komunalita ve třetím sloupci představuje podíl proměnlivosti znaku, vyjádřený zvolenými faktory. V tabulce je uveden příspěvek prvního a druhého faktoru do této proměnlivosti čili do komunality.



27. **FA:** *Cattelův indexový diagram vlastních čísel:* podle Kaiserova kritéria 1.0 pouze první dva faktory dosahují větších vlastních čísel než tato limita 1.0. Grafická analýza tohoto grafu však ukazuje na dva závažné zlomy (čili kolena) na křivce pro 2 a 3 faktory.

STATISTICA Cz - [PS 1* - Reziduální korelace (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

- PS 1*
  - Faktorová analýza (52Spoko
    - Dialog výsledků faktorov
      - Faktor. zátěže (Bez
      - Vl. čísla (52Spokojen
      - Vl. čísla (52Spokojen
      - Faktor. zátěže (Bez
      - Faktor. zátěže, fakt
      - Vl. čísla (52Spokojen
      - Komunality (52Spokc
      - Graf vlastních čísel
      - Reprodukované kore
      - Reziduální korelace (

Reziduální korelace (52Spokojenost)
Extrakce: Hlavní komponenty
(Označená rezidua jsou > ,100000)

Proměnná	Prace1	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	0,31	-0,10	-0,07	-0,01	-0,08	0,08	0,02	0,01	-0,02	-0,06
Prace2	-0,10	0,18	-0,06	-0,01	0,01	0,01	-0,02	0,03	-0,02	-0,02
Prace3	-0,07	-0,06	0,24	-0,06	-0,05	0,01	0,02	0,04	-0,02	-0,02
Hoby1	-0,01	-0,01	-0,06	0,11	-0,02	-0,02	-0,01	-0,03	0,01	-0,00
Hoby2	-0,08	0,01	-0,05	-0,02	0,23	0,03	-0,06	-0,05	-0,02	-0,04
Doma1	0,08	0,01	0,01	-0,02	0,03	0,30	-0,10	-0,13	-0,04	-0,06
Doma2	0,02	-0,02	0,02	-0,01	-0,06	-0,10	0,17	-0,05	0,01	0,02
Doma3	0,01	0,03	0,04	-0,03	-0,05	-0,13	-0,05	0,26	-0,02	-0,03
Ruzne1	-0,02	-0,02	-0,02	0,01	-0,02	-0,04	0,01	-0,02	0,09	-0,02
Ruzne2	-0,06	-0,02	-0,02	-0,00	-0,04	-0,06	0,02	-0,03	-0,02	0,19

28. **FA:** *Reziduální korelace:* diagonální prvky korelační matice reziduí obsahují směrodatnou odchylku, která jinak nemůže být vyčíslena přímo. Hodnoty mimodiagonálních prvků matice **nejsou** větší než +0.1 a menší než -0.1 a přitom 79% celkové proměnlivosti je vystiženo prvními dvěma faktory.

STATISTICA Cz - [PS 1* - Faktor. zátěže (Bez rot.) (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez
    - Faktor. zátěže, fakt
    - Vl. čísla (52Spokojen
    - Komunality (52Spokc
    - Graf vlastních čísel
    - Reprodukované kore
    - Reziduální korelace (
    - Faktor. zátěže, fakt
    - Faktor. zátěže (Bez

Faktor. zátěže (Bez rot.) (52Spokojenost)
Extrakce: Hlavní komponenty
(Označené zatěže jsou >,700000)

Proměnná	**Faktor 1**	Faktor 2
Prace1	-0,652601	0,514217
Prace2	-0,756976	0,494770
Prace3	-0,745706	0,456680
Hoby1	-0,941630	-0,021835
Hoby2	-0,875615	0,051643
Doma1	-0,576062	-0,604977
Doma2	-0,671289	-0,617962
Doma3	-0,641532	-0,573925
Ruzne1	-0,951516	0,013513
Ruzne2	-0,900333	0,048154
Výkl.roz	6,118369	1,800682
Prp.celk	0,611837	0,180068

29. **FA:** *Faktorové váhy bez rotace:* faktorové váhy (zátěže) lze vysvětlit jako korelace mezi faktory a znaky. **Červeně** jsou vyznačeny ještě statisticky významné a **černě** již statisticky nevýznamné. Vyšší faktory vždy popisují stále méně a méně popisované proměnlivosti v datech. Nejvíce proměnlivosti popisuje první faktor, o něco méně pak druhý faktor. Proto se v prvních dvou faktorech nalézají největší faktorové váhy.

STATISTICA Cz - [PS 1* - Faktor. zátěže (Varimax pr.) (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

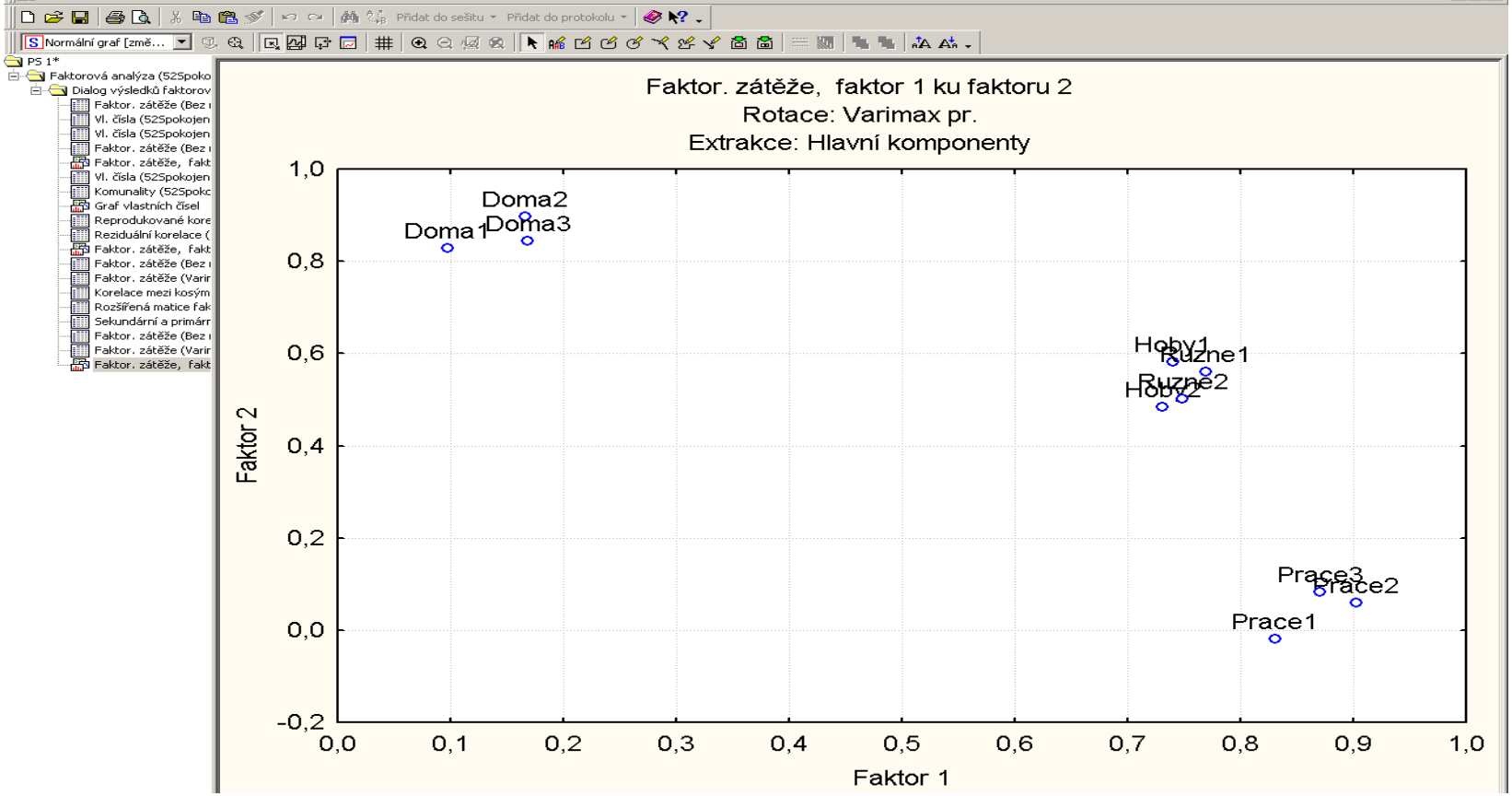
Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez
    - Faktor. zátěže, fakt
    - Vl. čísla (52Spokojen
    - Komunality (52Spokc
    - Graf vlastních čísel
    - Reprodukované kore
    - Reziduální korelace (
    - Faktor. zátěže, fakt
    - Faktor. zátěže (Bez
    - Faktor. zátěže (Varir
    - Korelace mezi kosým
    - Rozšířená matice fak
    - Sekundární a primárr
    - Faktor. zátěže (Bez

Faktor. zátěže (Varimax pr.) (52Spokojenost)
Extrakce: Hlavní komponenty
(Označené zatěže jsou >,700000)

Proměnná	Faktor 1	Faktor 2
Prace1	0,830623	-0,019320
Prace2	0,902408	0,058905
Prace3	0,870524	0,082595
Hoby1	0,739857	0,582885
Hoby2	0,731191	0,484489
Doma1	0,097371	0,829676
Doma2	0,165722	0,897242
Doma3	0,168370	0,844159
Ruzne1	0,768988	0,560555
Ruzne2	0,748861	0,502121
Výkl.roz	4,561544	3,357507
Prp.celk	0,456154	0,335751

30. **FA:** *Faktorové váhy po Varimax rotaci:* otočení faktorů ve faktorovém prostoru je libovolné. Cílem je co nejlépe reprodukovat korelace mezi faktory a znaky, aby vysvětlení bylo snažší a jednodušší a aby bylo dosaženo “faktorově čistých znaků”. U faktorově čistých znaků je dosaženo maximálních hodnot pouze u jednoho faktoru osy, zatímco u druhého faktoru pak minimálních hodnot na ose.



31. **FA:** *Faktorově čisté znaky v grafu faktorových vah po Varimax rotaci:* **první faktor** se zde týká znaků spokojenosti v práci, zatímco znaky spokojenosti doma v něm dosahují nejmenších vah. Ostatní znaky pak leží někde uprostřed. **Druhý faktor** vykazuje nejvyšší váhy znaků spokojenosti doma a nejnižší váhy pro znaky spokojenosti v práci. Ostatní znaky spokojenosti leží opět někde uprostřed.

STATISTICA Cz - [PS 1* - Koeficienty faktor. skóre (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez
    - Faktor. zátěže, fakt
    - Vl. čísla (52Spokojen
    - Komunality (52Spokc
    - Graf vlastních čísel
    - Reprodukované kore
    - Reziduální korelace (
    - Faktor. zátěže, fakt
    - Faktor. zátěže (Bez
    - Faktor. zátěže (Varir
    - Korelace mezi kosým

Koeficienty faktor. skóre (52Spokojenost)
Rotace: Varimax pr.
Extrakce: Hlavní komponenty

Proměnná	**Faktor 1**	Faktor 2
Prace1	0,256768	-0,164304
Prace2	0,263925	-0,145425
Prace3	0,249750	-0,129616
Hoby1	0,115785	0,102111
Hoby2	0,131660	0,063002
Doma1	-0,126453	0,325194
Doma2	-0,118337	0,340306
Doma3	-0,107542	0,317830
Ruzne1	0,128865	0,087384
Ruzne2	0,133727	0,066977

32. **FA:** *Koeficienty faktorového skóre po Varimax rotaci:* užijí se k vyčíslení faktorového skóre. Tyto koeficienty představují váhy, které se užijí při výpočtu faktorového skóre ze znaků. Tato matice koeficientů představuje však pouhý mezivýsledek a nemá proto pro uživatele valný význam.

STATISTICA Cz - [PS 1* - Faktor. skóre (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

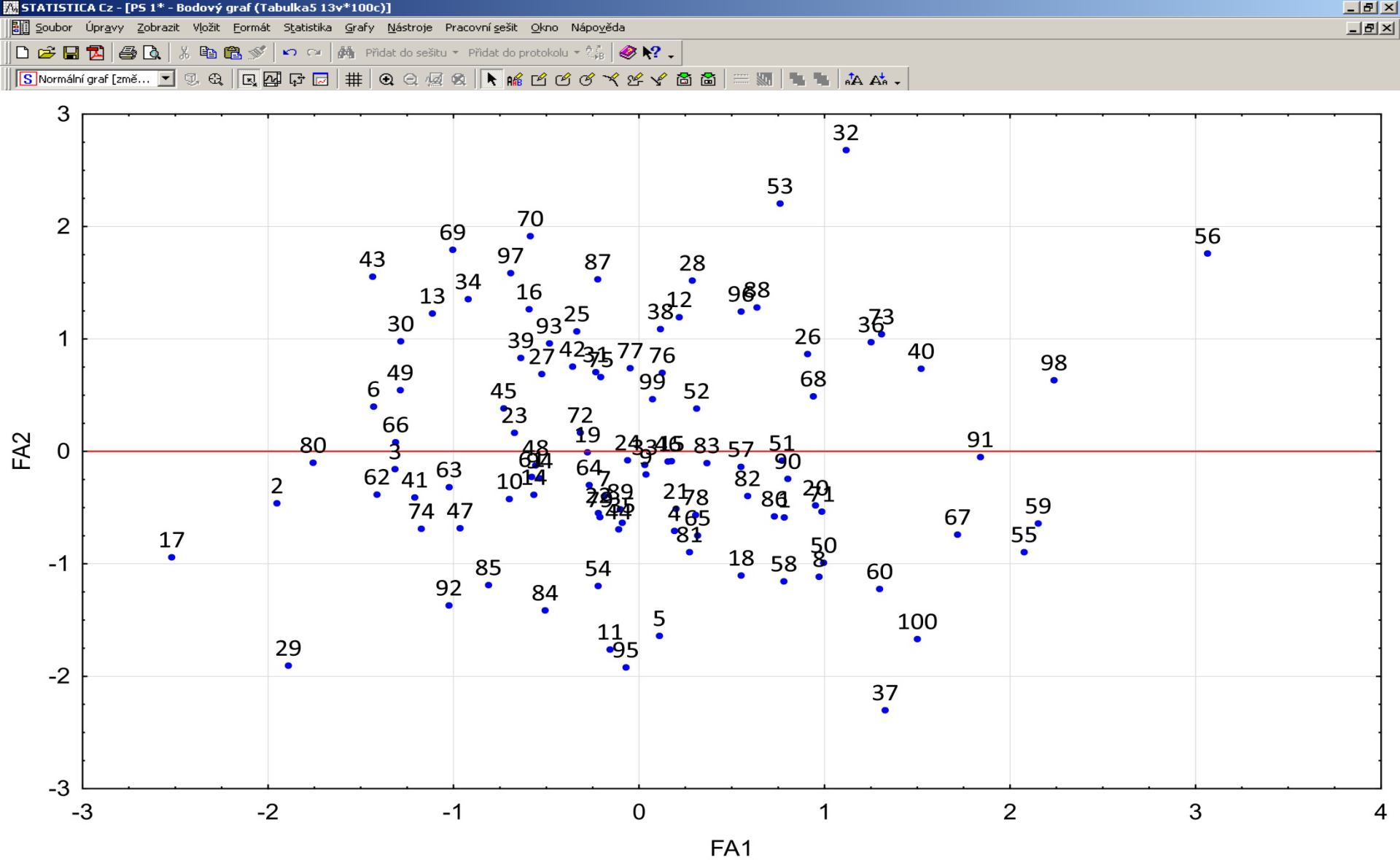
Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez
    - Faktor. zátěže, fakt
    - Vl. čísla (52Spokojen
    - Komunality (52Spoko
    - Graf vlastních čísel
    - Reprodukované kore
    - Reziduální korelace (
    - Faktor. zátěže, fakt
    - Faktor. zátěže (Bez
    - Faktor. zátěže (Varir
    - Korelace mezi kosým
    - Rozšířená matice fak
    - Sekundární a primárr
    - Faktor. zátěže (Bez
    - Faktor. zátěže (Varir
    - Faktor. zátěže, fakt
    - Faktor. zátěže (Varir
    - Korelace mezi kosým
    - Rozšířená matice fak
    - Sekundární a primárr
    - Koeficienty faktor. sl
    - Fa

Faktor. skóre (52Spokojenost)
Rotace: Varimax pr.
Extrakce: Hlavní komponenty

Případ	Faktor 1	Faktor 2
1	0,77326	-0,59909
2	-1,95924	-0,42839
3	-1,31803	-0,13560
4	0,17915	-0,70837
5	0,08277	-1,64135
6	-1,42460	0,42254
7	-0,19411	-0,39425
8	0,95212	-1,13020
9	0,03346	-0,20582
10	-0,70690	-0,41079
11	-0,18579	-1,75809
12	0,23559	1,19109
13	-1,09461	1,24608
14	-0,57400	-0,37563
15	0,17399	-0,08925
16	-0,57290	1,27404
17	-2,53492	-0,89944
18	0,53181	-1,11260
19	-0,27819	-0,00231
20	0,94282	-0,49513
21	0,19051	0,51291

33. **FA:** *Faktorová skóre objektů (*zde *respondentů):* vyvrcholením celé faktorové analýzy je právě tato matice, která třídí objekty (respondenty) ve faktorovém prostoru v tabulkové ale lépe také v grafické podobě. Tabulku lze uložit a potom znázornit graficky.



34. **FA:** *Graf faktorového skóre objektů (*zde *respondentů)*: graf třídí a klasifikuje respondenty v rovině prvních dvou faktorů.

STATISTICA Cz - [PS 1* - Reziduální korelace (52Spokojenost)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

PS 1*
- Faktorová analýza (52Spoko
  - Dialog výsledků faktorov
    - Faktor. zátěže (Bez
    - Vl. čísla (52Spokojen
    - Vl. čísla (52Spokojen
    - Faktor. zátěže (Bez
    - Faktor. zátěže, fakt
    - Vl. čísla (52Spokojen
    - Komunality (52Spokc
    - Graf vlastních čísel
    - Reprodukované kore
    - Reziduální korelace (
    - Faktor. zátěže, fakt
    - Faktor. zátěže (Bez
    - Faktor. zátěže (Varir
    - Korelace mezi kosým

Reziduální korelace (52Spokojenost)
Extrakce: Hlavní komponenty
(Označená rezidua jsou > ,100000)

Proměnná	Prace1	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	0,31	-0,10	-0,07	-0,01	-0,08	0,08	0,02	0,01	-0,02	-0,06
Prace2	-0,10	0,18	-0,06	-0,01	0,01	0,01	-0,02	0,03	-0,02	-0,02
Prace3	-0,07	-0,06	0,24	-0,06	-0,05	0,01	0,02	0,04	-0,02	-0,02
Hoby1	-0,01	-0,01	-0,06	0,11	-0,02	-0,02	-0,01	-0,03	0,01	-0,00
Hoby2	-0,08	0,01	-0,05	-0,02	0,23	0,03	-0,06	-0,05	-0,02	-0,04
Doma1	0,08	0,01	0,01	-0,02	0,03	0,30	-0,10	-0,13	-0,04	-0,06
Doma2	0,02	-0,02	0,02	-0,01	-0,06	-0,10	0,17	-0,05	0,01	0,02
Doma3	0,01	0,03	0,04	-0,03	-0,05	-0,13	-0,05	0,26	-0,02	-0,03
Ruzne1	-0,02	-0,02	-0,02	0,01	-0,02	-0,04	0,01	-0,02	0,09	-0,02
Ruzne2	-0,06	-0,02	-0,02	-0,00	-0,04	-0,06	0,02	-0,03	-0,02	0,19

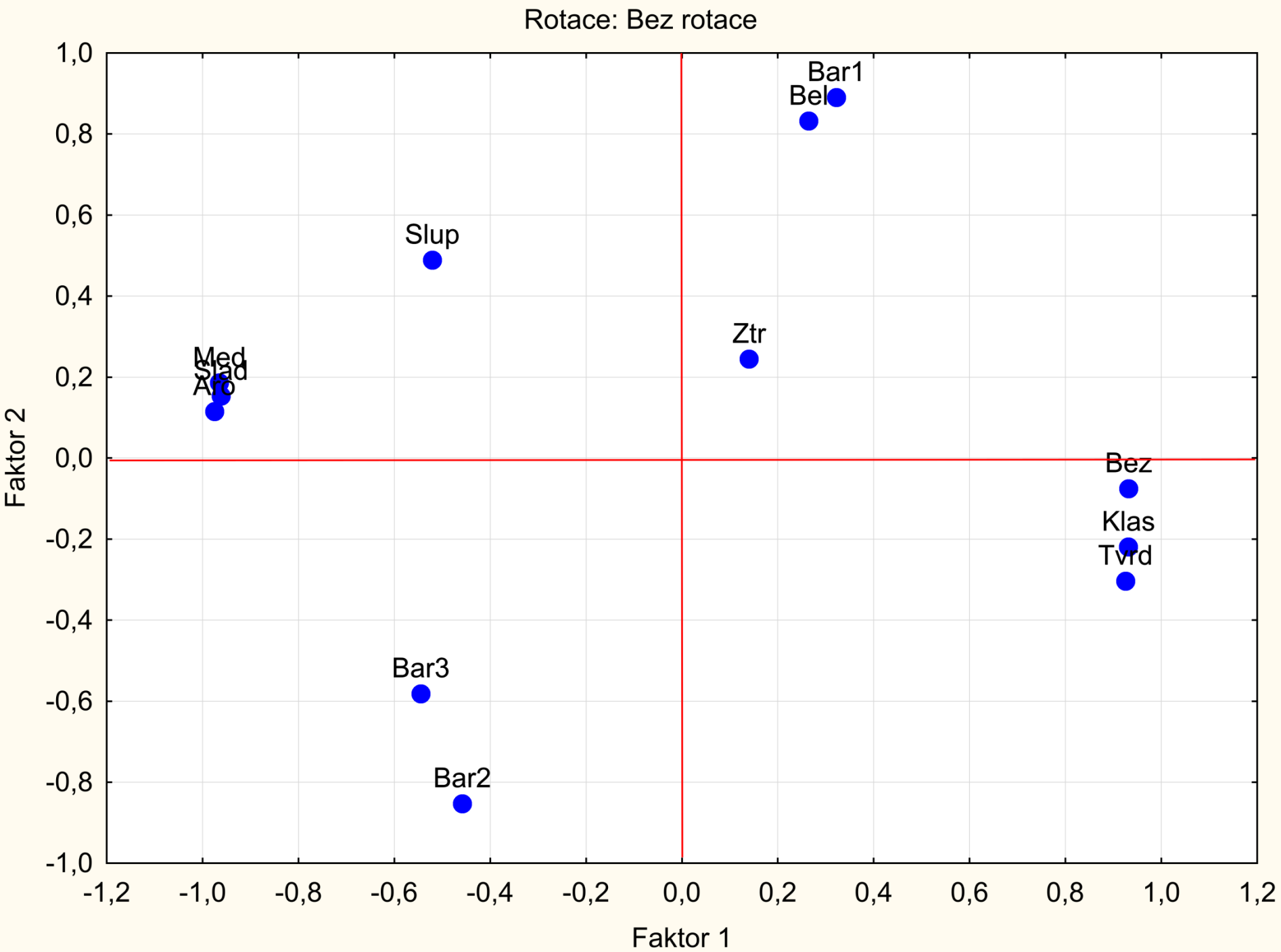
35. **FA:** *Korelační matice reziduí po Varimax rotaci:* po pečlivém prohlédnutí vidíme, že hodnoty leží v intervalu od +0.1 do -0.1 a došlo i ke zmenšení hodnot (čili zlepšení) oproti této dřívější matici.

PŘÍKLAD 1.1 ***Popisné statistiky jednorozměrné analýzy zdrojové matice dat*** *Hrách*

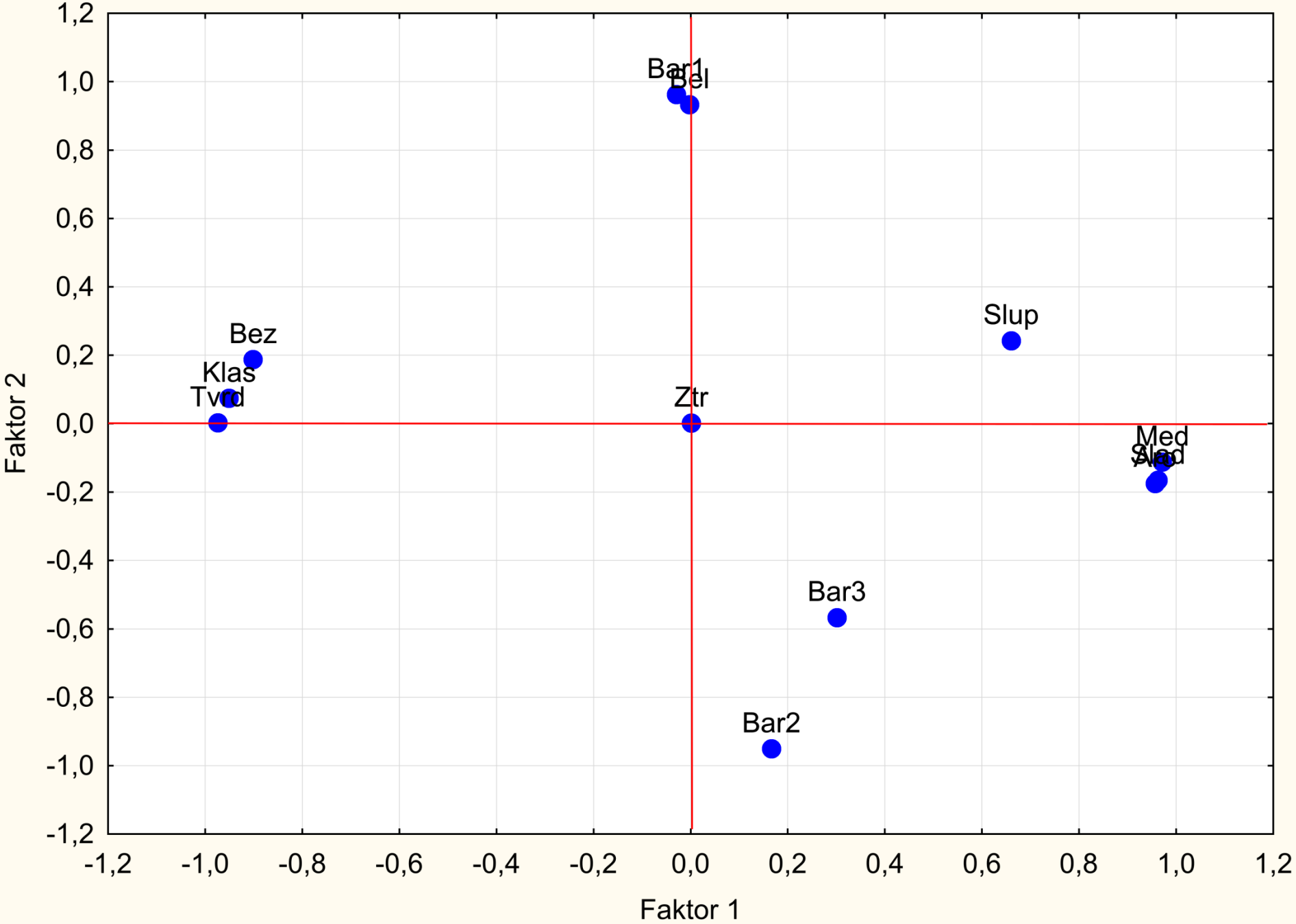
Zdrojová matice dat *Hrách obsahuje znaky smyslového posouzení charakteristik rozličných odrůd hrachu. Objekty zde představují vzorky pěti různých odrůd hrachu A až E, které byly sklízeny v pěti rozličných obdobích 1 až 5. Výsledná zdrojová matice o 12 znacích převážně smyslových charakteristik obsahuje 60 vzorků hrachu. Posouzení každého objektu hrachu bylo provedeno 10 porotci dvojím odhadem tak, že smyslové charakteristiky byly bodovány ve stupnici od 1 (nejhorší) do 9 (nejlepší). Tak bylo získáno 1200 řádků (objektů) postupem: 60 vzorků × 2 hodnocení × 10 porotců. V praxi se data obvykle průměrují, aby se kompenzovaly rozdíly v subjektivní škále přísnosti jednotlivých porotců. Výsledkem je pro každý z šedesáti objektů průměrná hodnota senzorického hodnocení. Cílem úlohy je: 1*. průměrovat data, 2. vynést původní data do grafu a 3. vypočítat popisné jednorozměrné statistiky.

Data: *Zdrojová matice dat n = 1200, m = 12 byla průměrována a výsledkem byla matice 60 × 12. Obsahovala průměrné hodnoty senzorického hodnocení pro znaky ve sloupcích: Aro je aroma, Slad je sladkost, Med je medovost, Bez je bezchuťovost, Klas je klasovost, Tvrd je tvrdost, Bel je bělost, Bar1 je barva1, Bar2 je barva2, Bar3 je barva3, Slup je slupka, Ztr je ztráta.*

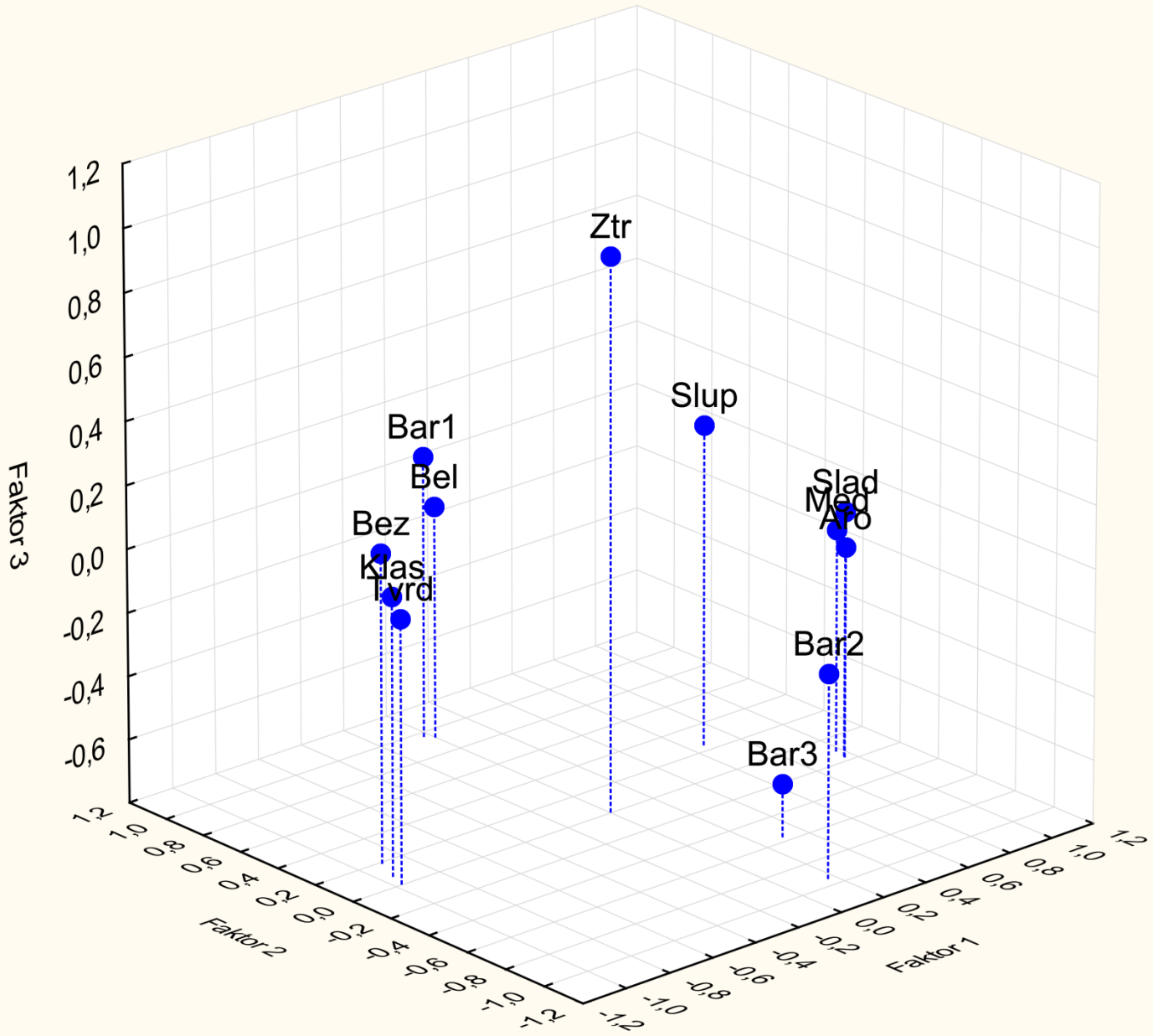
Objekt	*Aro*	*Slad*	*Med*	*Bez*	*Klas*	*Tvrd*	*Bel*	*Bar1*	*Bar2*
	Bar3	*Slup*	*Ztr*						
B5	6.48	6.66	4.56	2.2	2.91	3.47	4.72	5.59	5.73
	5.99	4.26	3.25						
C4	5.75	6.09	3.81	2.32	4.03	3.77	4.17	5.73	5.75
	5.32	3.82	3.38						
B2	3.94	4.12	2.44	3.63	5.77	5.39	4.77	6.67	5.11
	4.60	3.5	3.03						

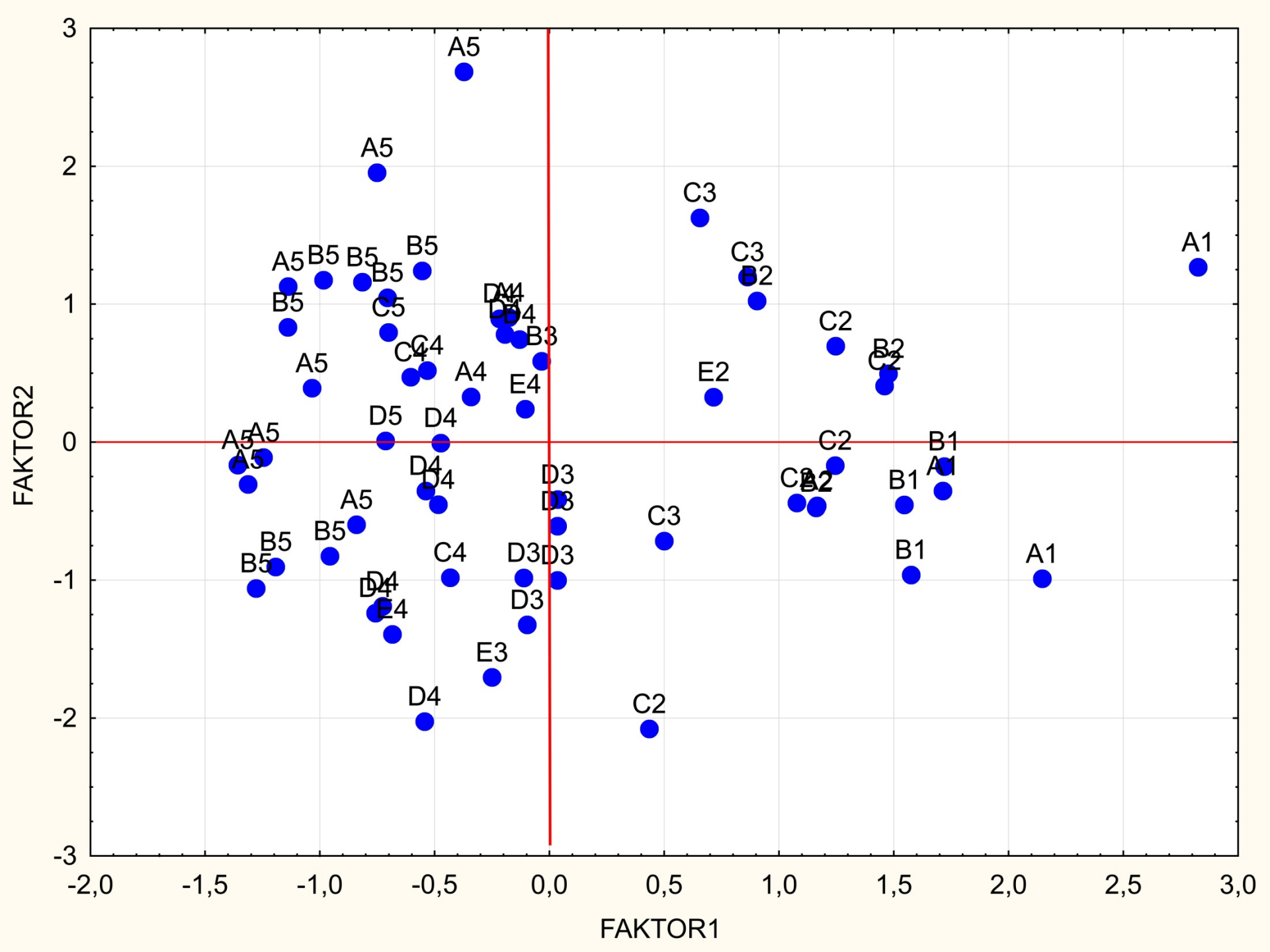

Rotace: Bez rotace
Bar1
Bel
Slup
Ztr
Med
Slad
Aro
Bez
Klas
Tvrd
Bar3
Bar2
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-1,2
-1,0
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0
1,2
Faktor 2
Faktor 1

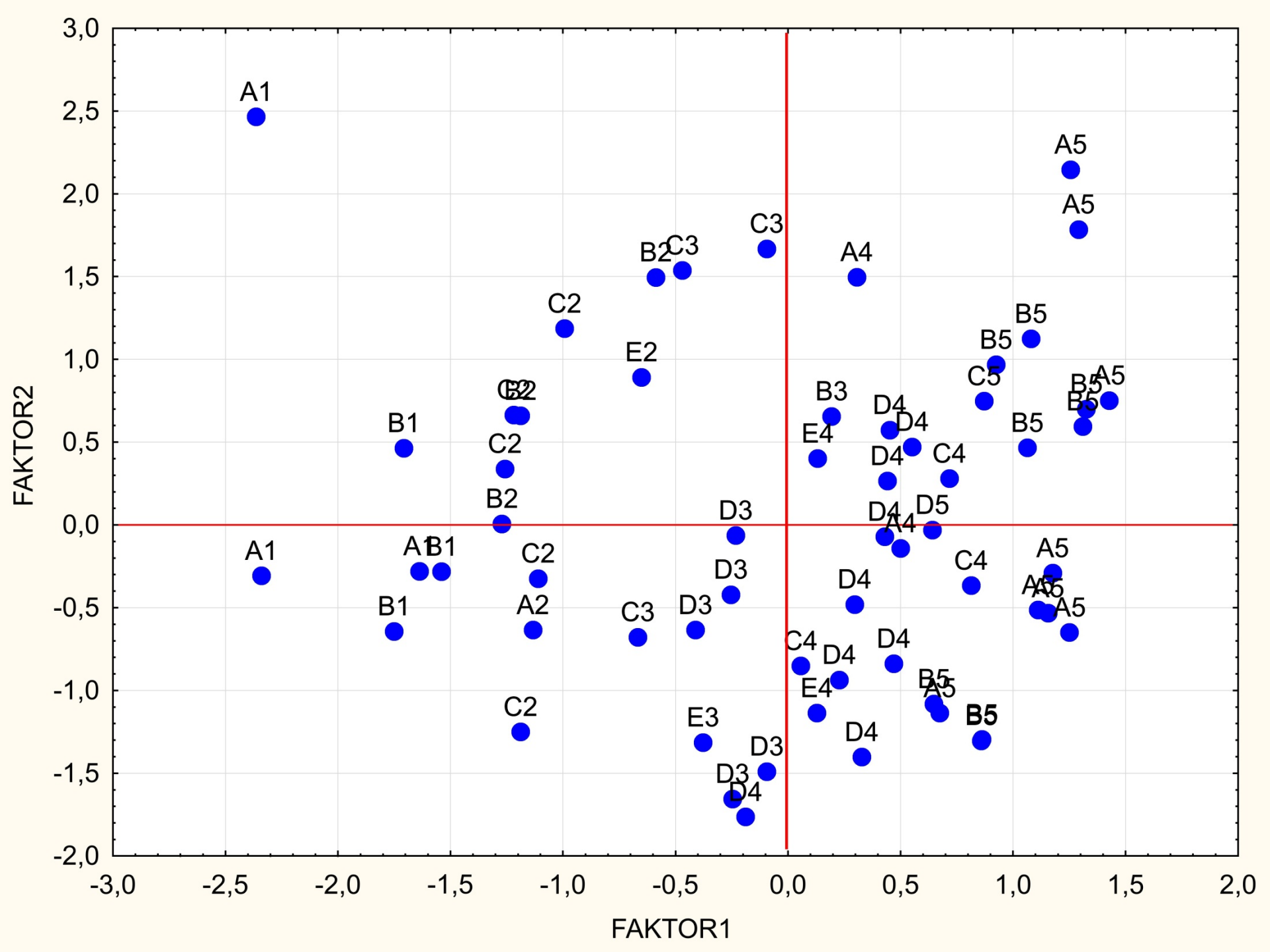
Rotace: Varimax normaliovaný.

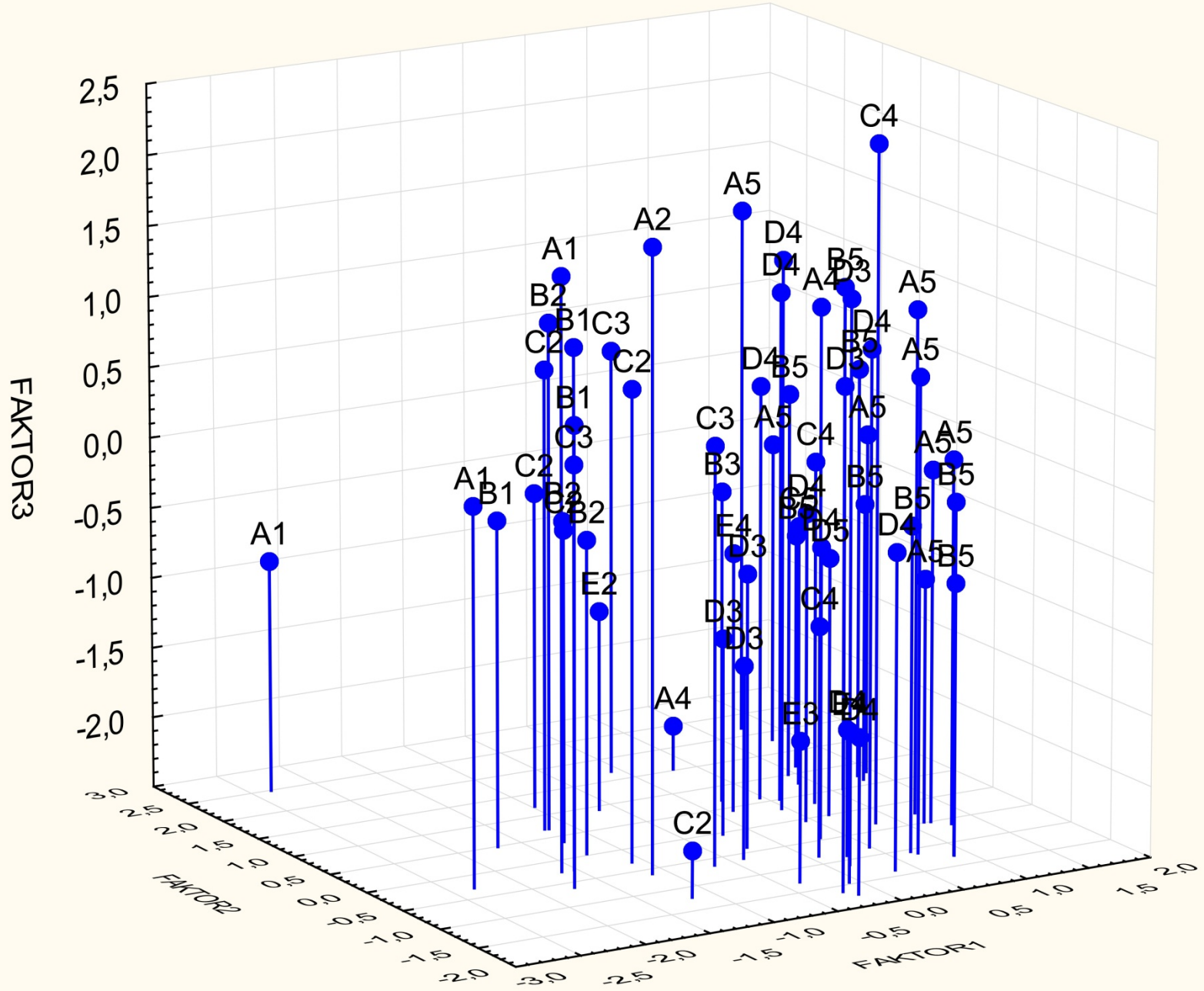


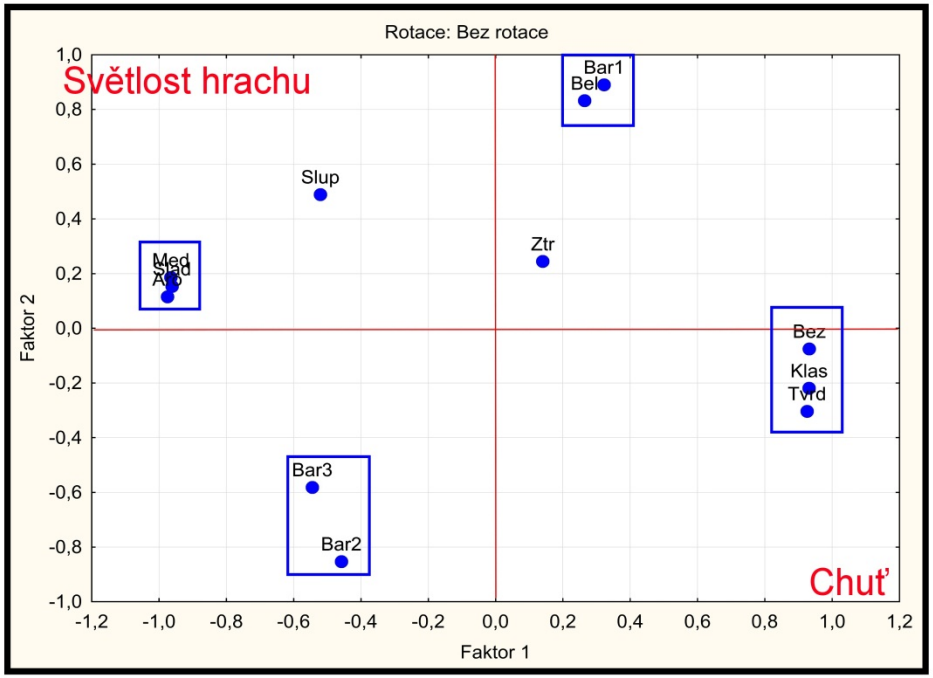
Rotace: Varimax normalizovaný

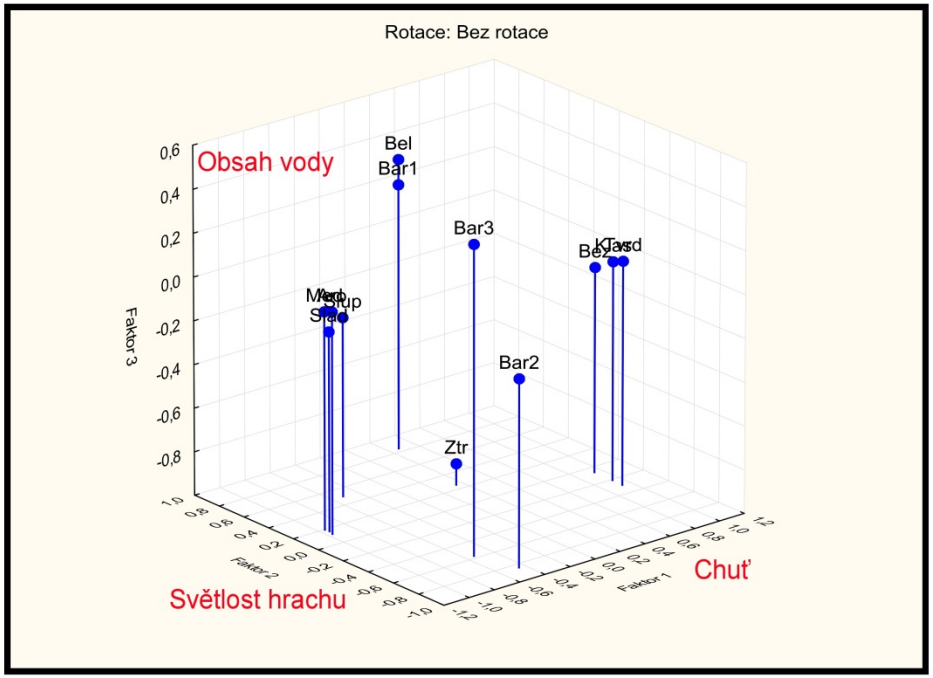


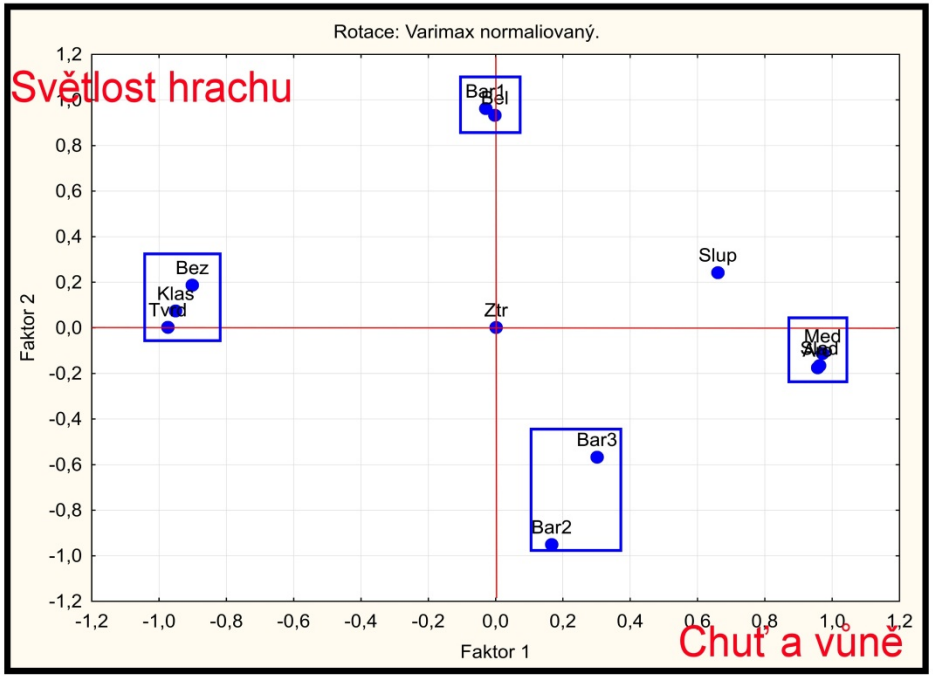

FAKTOR2
FAKTOR1
3
2
1
0
-1
-2
-3
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
2,5
3,0
A5
A5
C3
A1
A5
B5
B5
B5
C3
B2
B5
B5
C5
C2
B3
A5
C4
A4
E2
B2
C2
E4
D5
D4
A5
A5
A5
C2
B1
A1
D4
D4
D3
B1
A5
D3
C3
B5
B5
B5
C4
D3
D3
B1
A1
D4
D4
E4
D3
E3
D4
C2

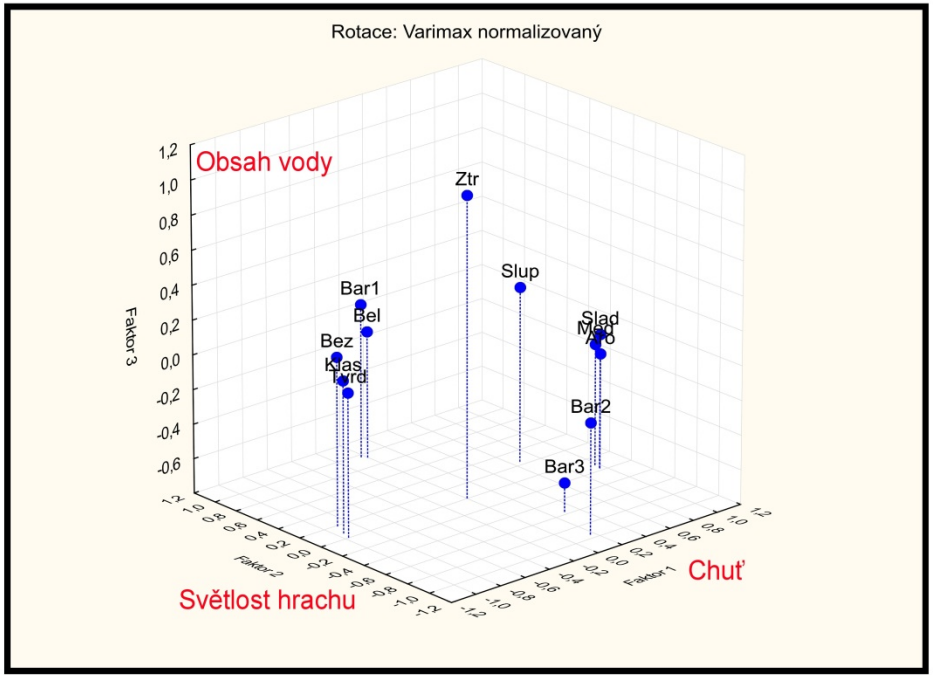

FAKTOR2
FAKTOR1
3,0
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-3,0
-2,5
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
A1
A5
A5
C3
B2
C3
A4
C2
B5
B5
E2
C5
A5
B3
D4
B1
D4
B5
E4
C2
D4
C4
B2
D3
D4
D5
A4
A1
A1
B1
C2
C4
A5
D3
D4
B1
A2
C3
D3
A5
C4
D4
D4
E4
B5
A5
C2
B5
E3
D4
D3
D3
D4

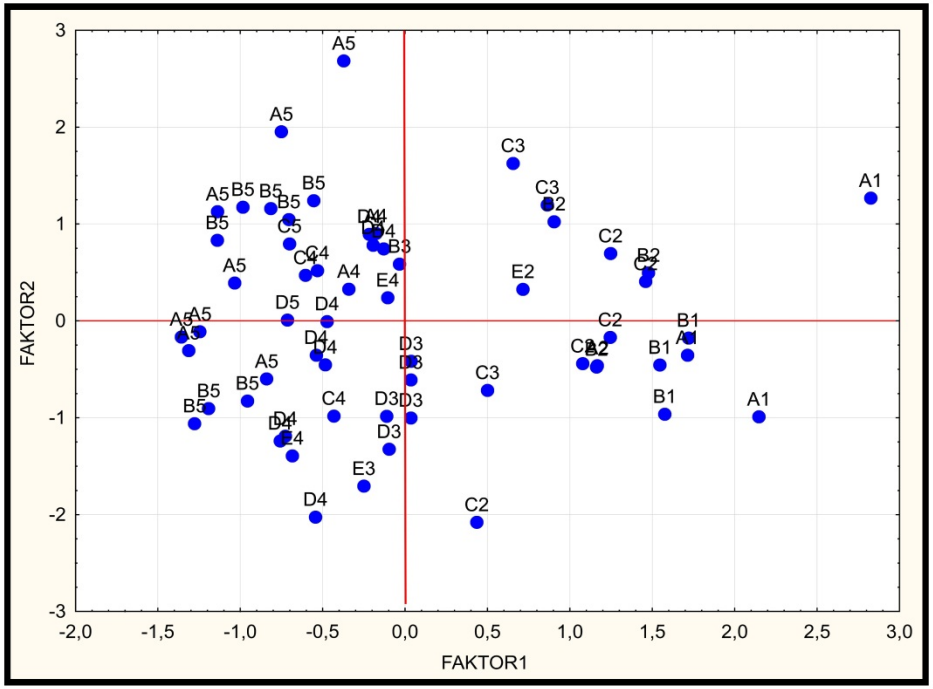

FAKTOR3
FAKTOR2
FAKTOR1
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
3,0
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-3,0
-2,5
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0

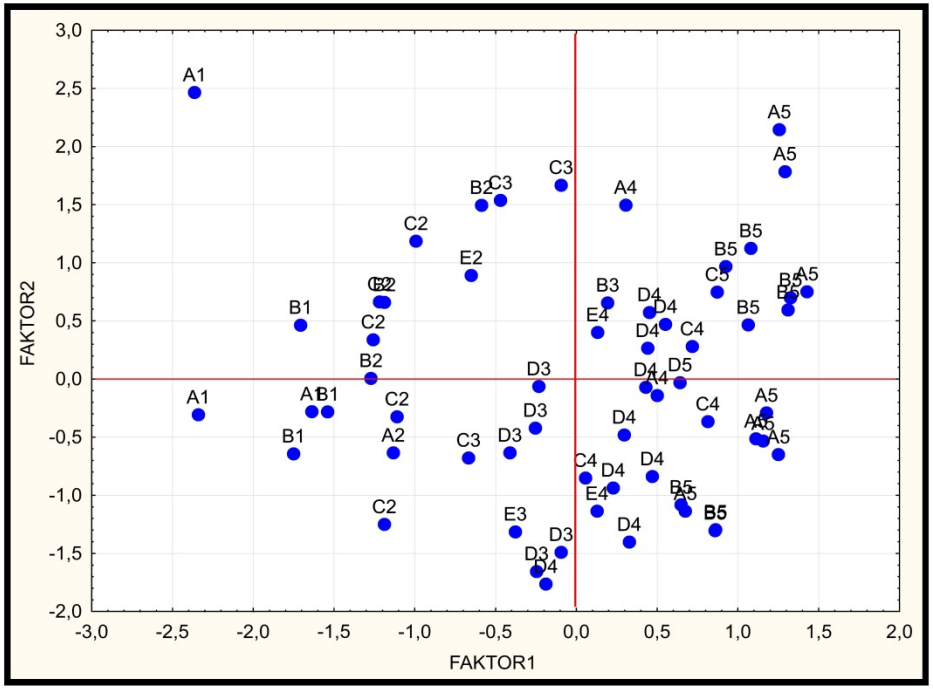

Rotace: Bez rotace
Světlost hrachu
Chuť
Faktor 1
Faktor 2
Bar1
Bel
Slup
Ztr
Med
Slad
Bez
Klas
Tvrd
Bar3
Bar2

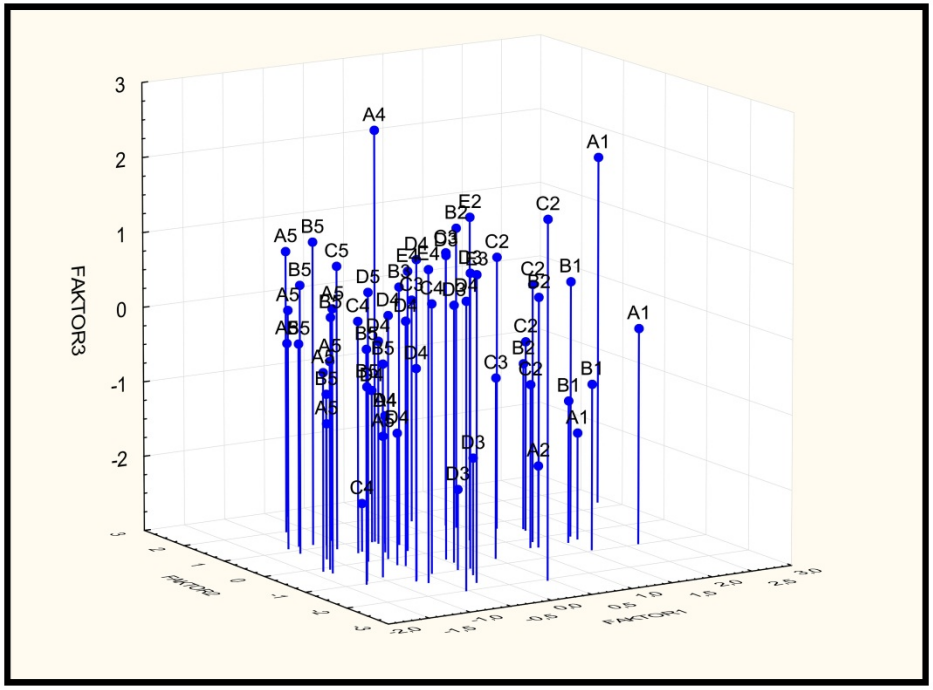

Rotace: Bez rotace
Obsah vody
Světlost hrachu
Chuť
Faktor 1
Faktor 2
Faktor 3
Bel
Bar1
Bar3
Bar2
Ztr

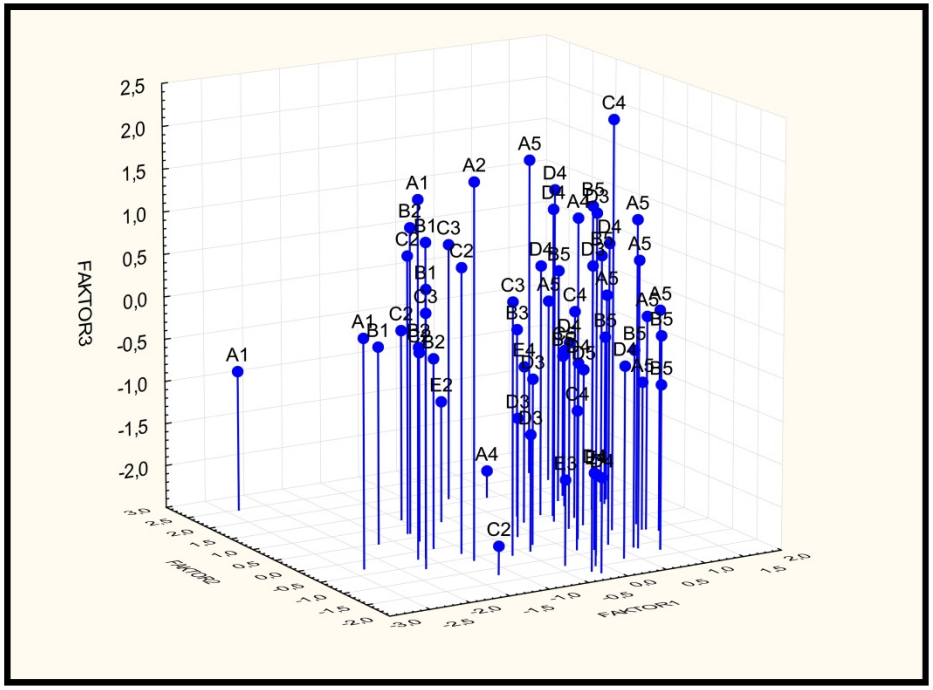

Rotace: Varimax normaliovaný.
Světlost hrachu
Chuť a vůně
Faktor 1
Faktor 2
Bar1
Bel
Bez
Klas
Tvrd
Ztr
Slup
Med
Slad
Bar3
Bar2

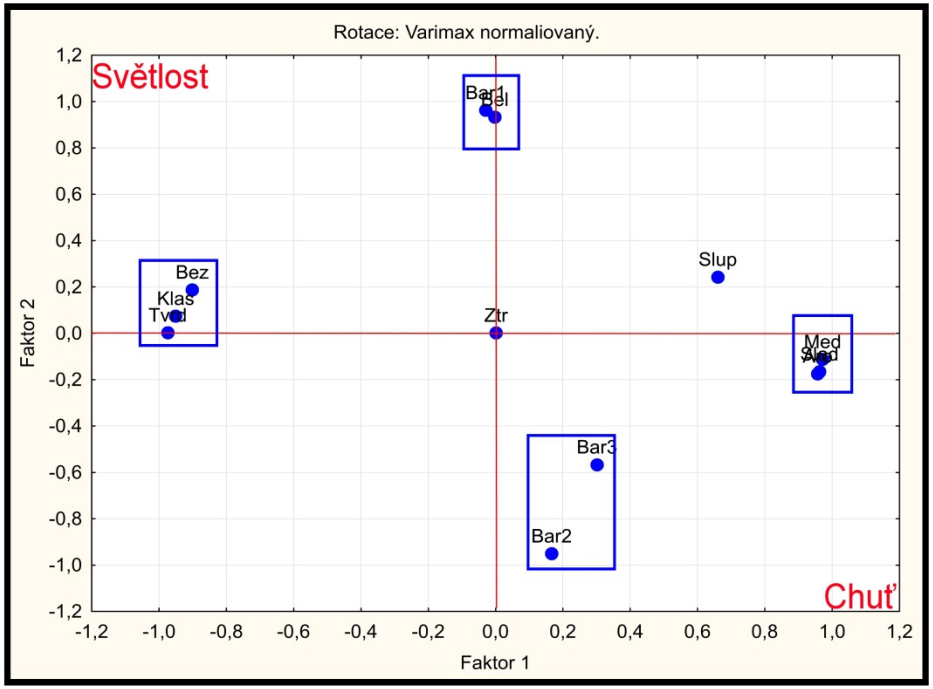

Rotace: Varimax normalizovaný
Obsah vody
Světlost hrachu
Chuť
Faktor 1
Faktor 2
Faktor 3
Ztr
Slup
Bar1
Bel
Bez
Klas
Tvrd
Slad
Med
Bar2
Bar3

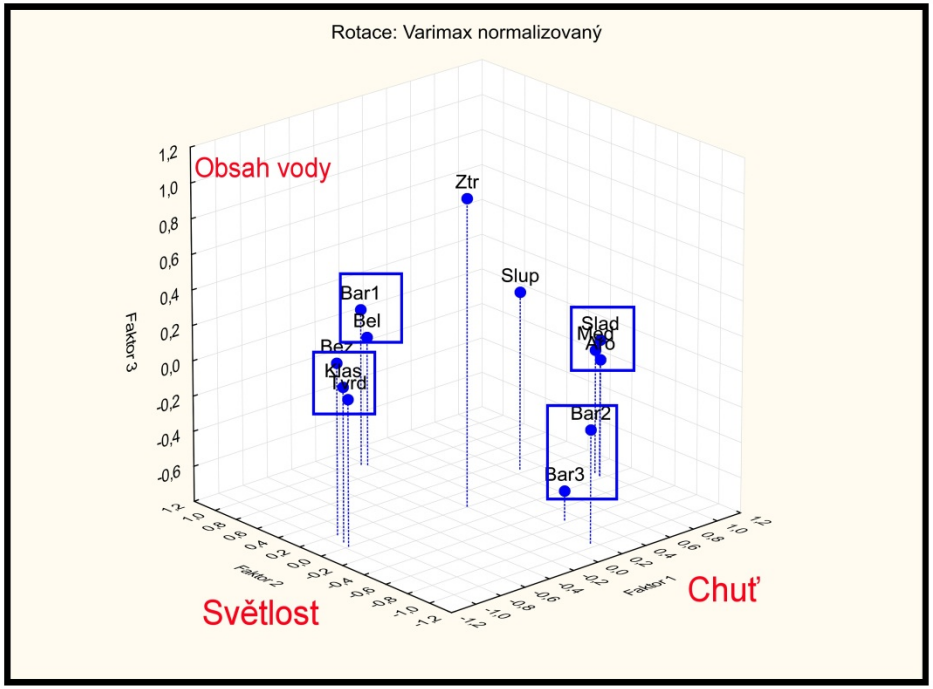

FAKTOR2
FAKTOR1

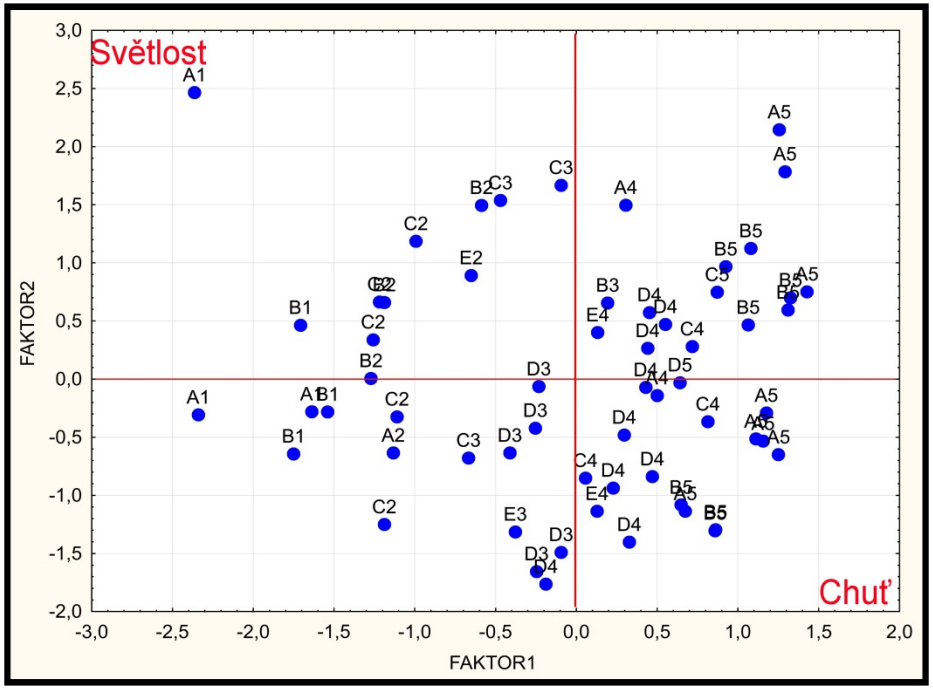

FAKTOR2
FAKTOR1

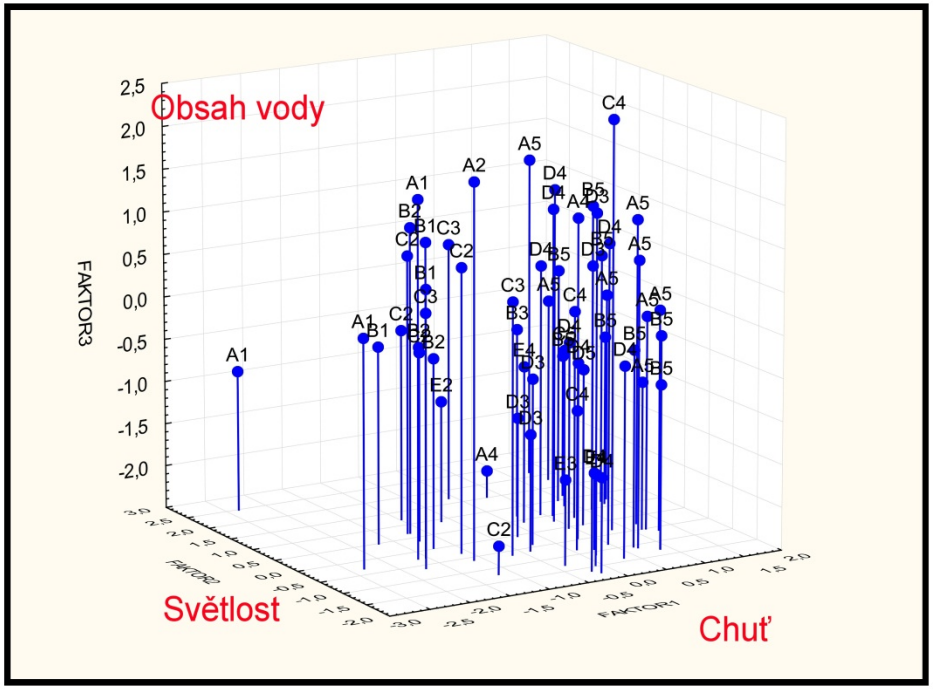

FAKTOR3
FAKTOR2
FAKTOR1

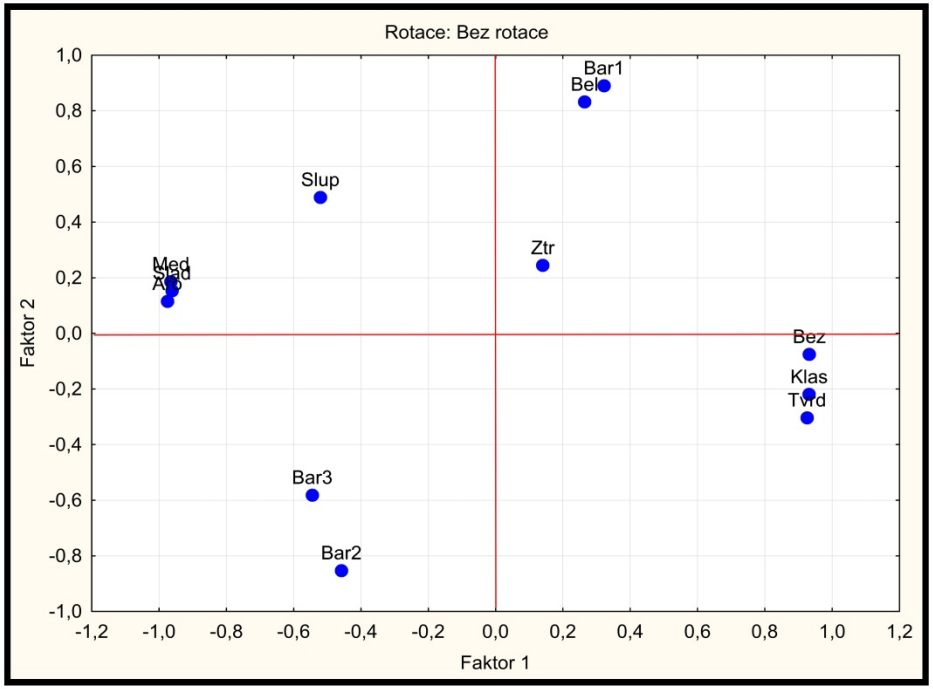

FAKTOR3
FAKTOR2
FAKTOR1

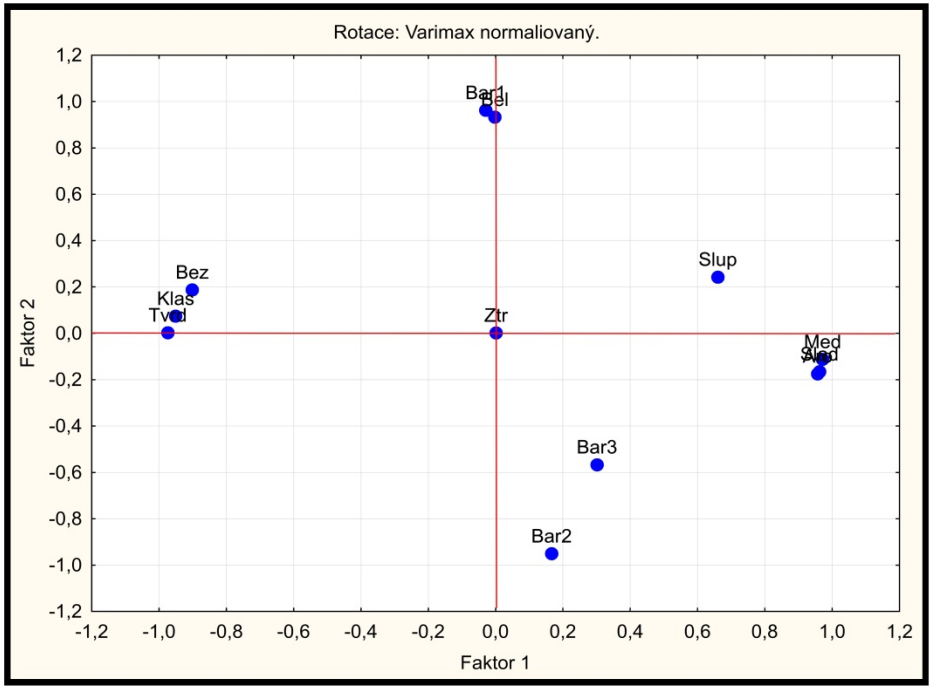

Rotace: Varimax normaliovaný.
Světlost
Chuť
Faktor 1
Faktor 2
Bar1
Bel
Bez
Klas
Tvrd
Ztr
Slup
Med
Sled
Bar3
Bar2

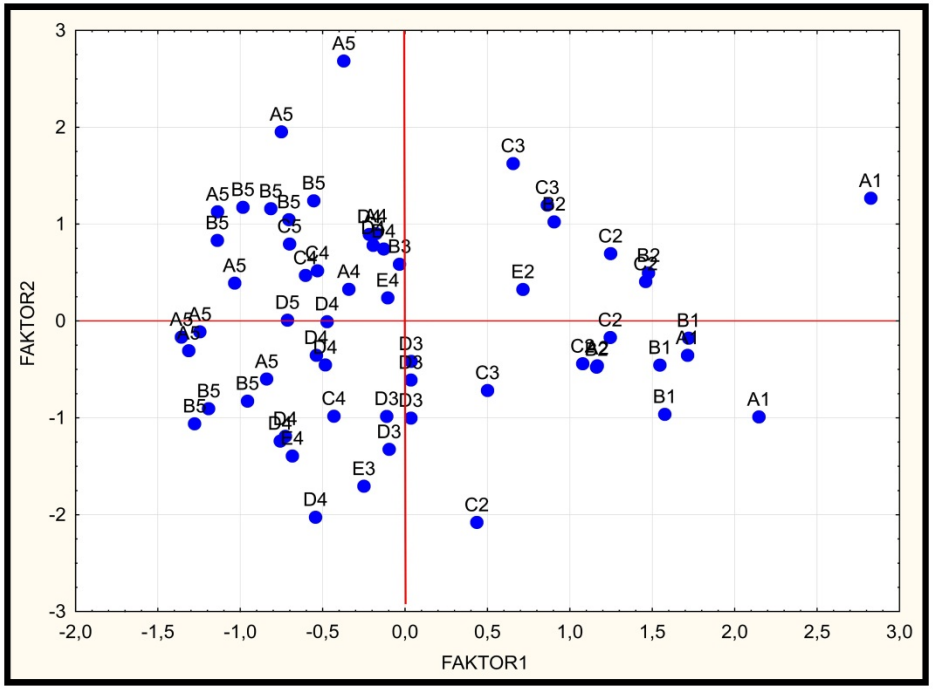

Rotace: Varimax normalizovaný
Obsah vody
Světlost
Chuť
Faktor 1
Faktor 2
Faktor 3
Ztr
Slup
Bar1
Bel
Bez
Klas
Tvrd
Sled
Med
Bar2
Bar3

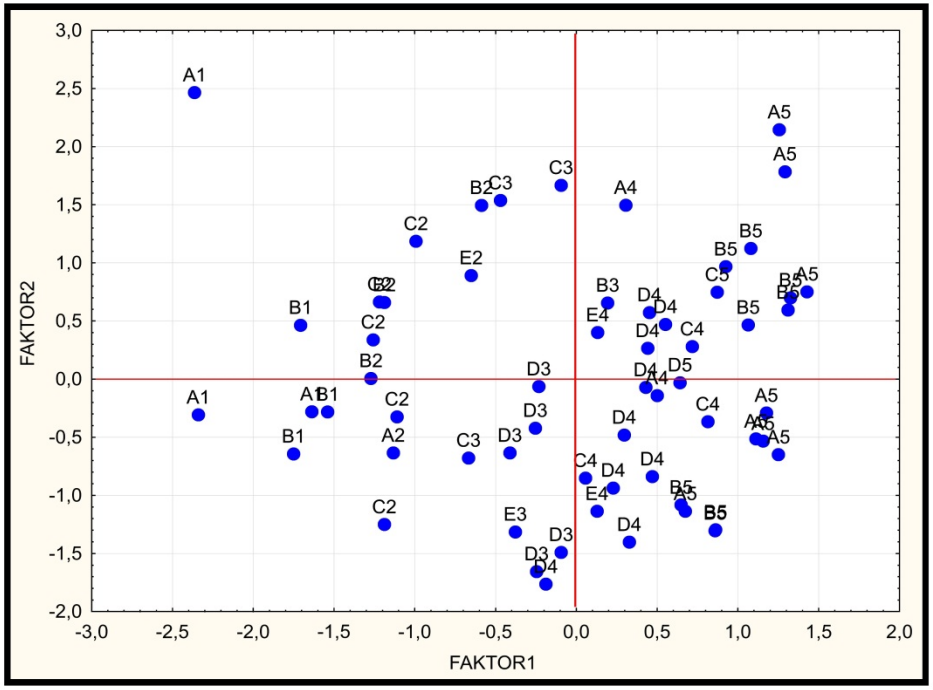

Světlost
Chuť
FAKTOR1
FAKTOR2


Obsah vody
Světlost
Chuť
FAKTOR1
FAKTOR2
FAKTOR3


Rotace: Bez rotace
Faktor 2
Faktor 1
Bar1
Bel
Slup
Ztr
Med
Bez
Klas
Tvrd
Bar3
Bar2


Rotace: Varimax normaliovaný.
Faktor 2
Faktor 1
Bar1
Bel
Bez
Klas
Tvrd
Ztr
Slup
Med
Bar3
Bar2


FAKTOR2
FAKTOR1

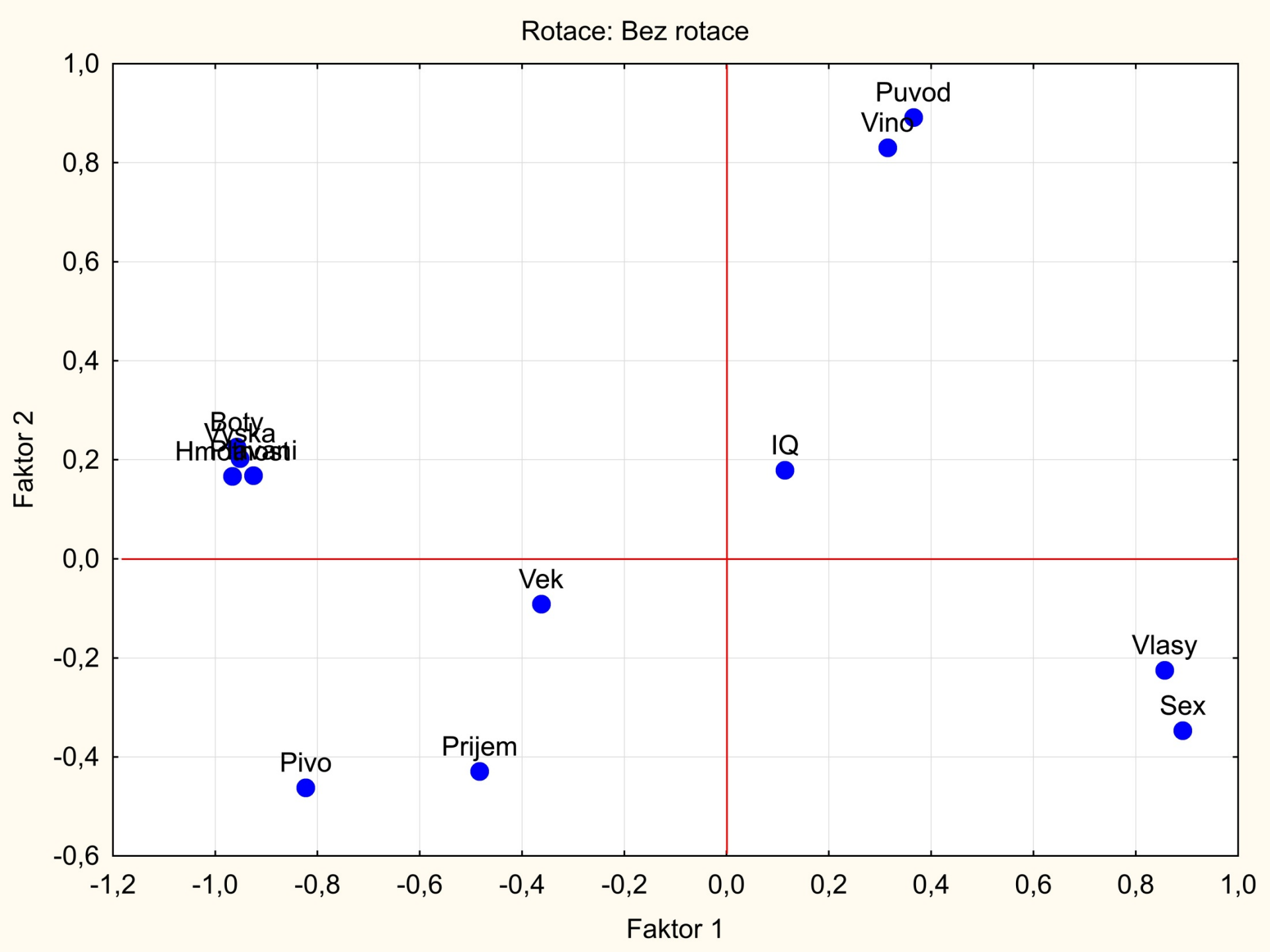

FAKTOR2
FAKTOR1

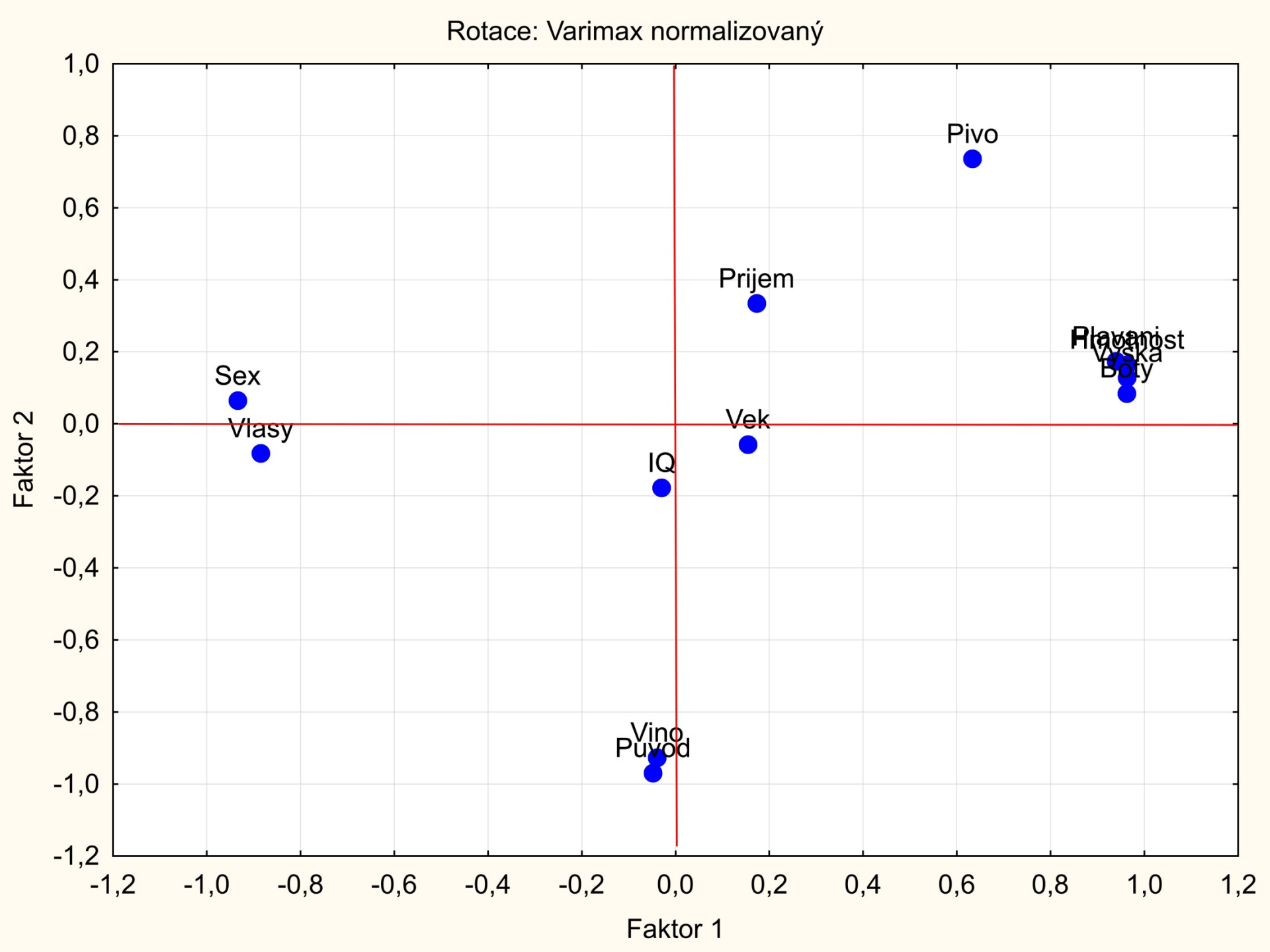
PŘÍKLAD 2.2 ***Průzkumová analýza zdrojové matice dat demografického souboru Lidé***

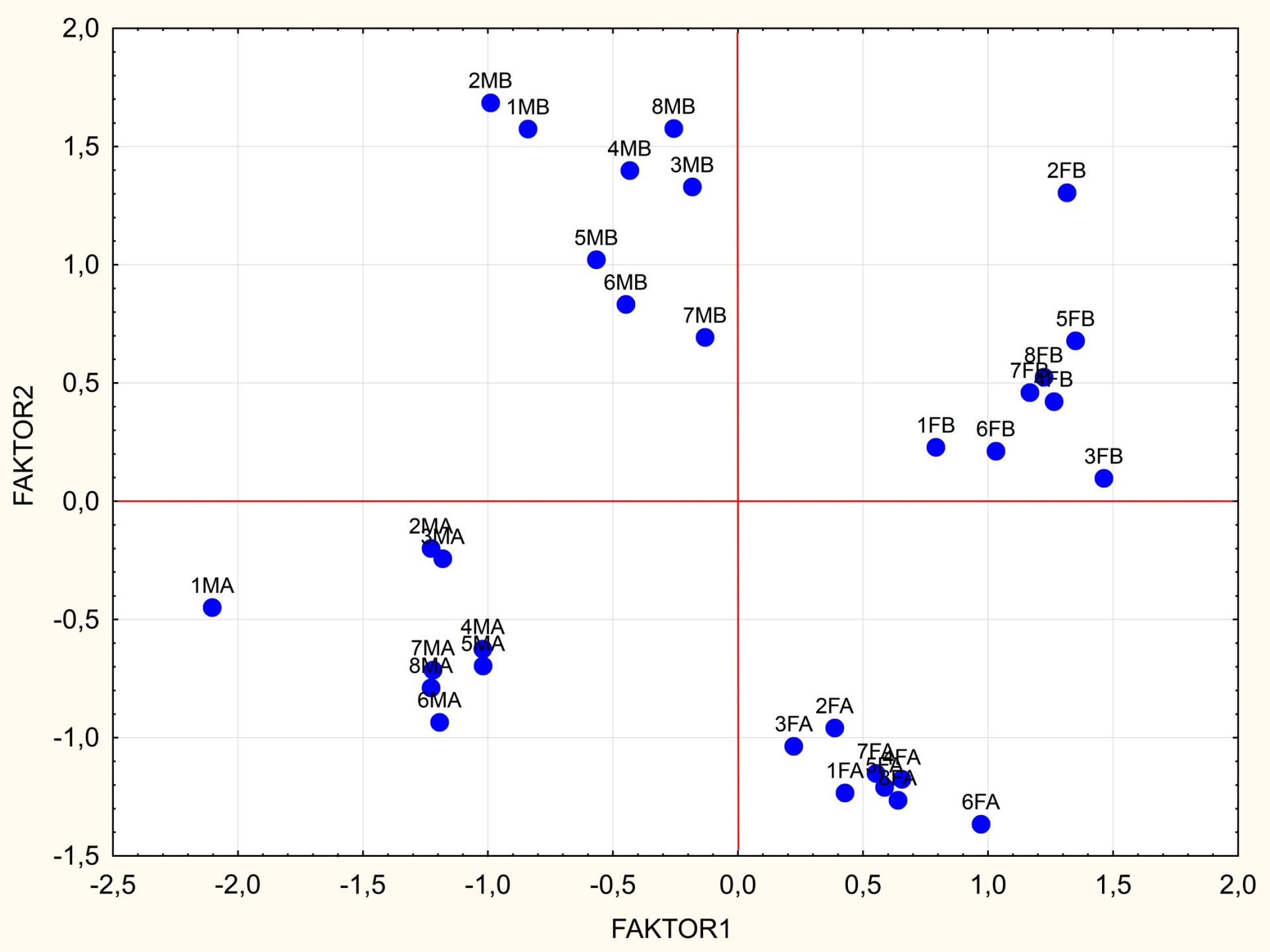
Vyšetřete grafickými diagnostikami průzkumové analýzy vícerozměrných dat, které ze 12 znaků demografického souboru dat *Lidé jsou nejvýhodnější k charakterizaci osob a které znaky mají největší míru rozptýlení. Matice obsahuje data pro n = 32 osob a m = 12 znaků, kde 16 osob bylo vybráno ze Skandinávie (kód A) a 16 osob ze Středomoří (kód B), 16 osob jsou muži (kód M) a 16 osob jsou ženy (kód F).*

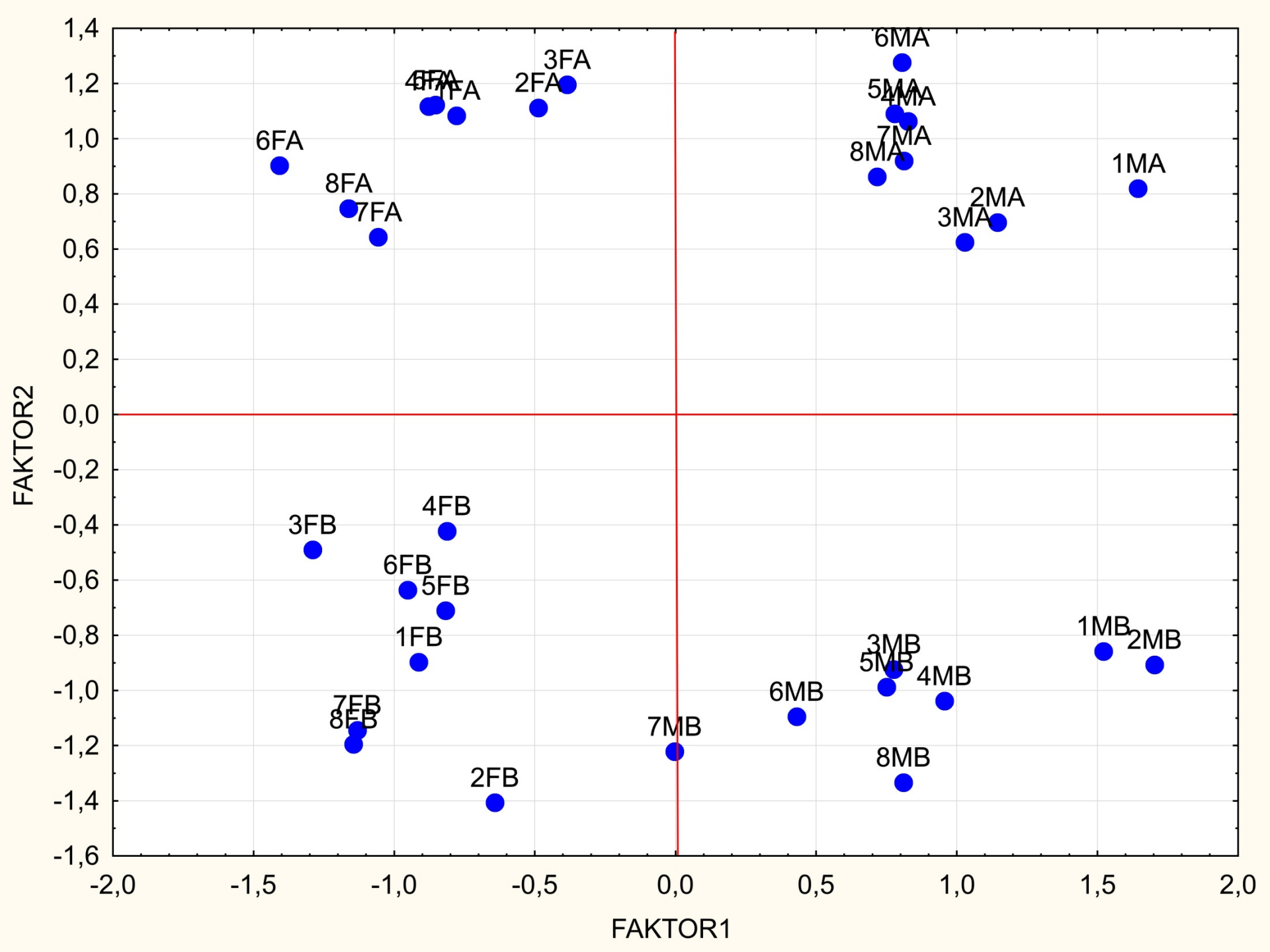
Data: *Znaky obsahují u každé osoby výšku [cm], hmotnost [kg], délku vlasů [krátká: −1, dlouhá: +1], velikost boty [evropský standard], věk [roky], příjem [Euro], spotřeba piva [litry na rok], spotřeba vína [litry na rok], pohlaví [muž: −1, žena: +1], schopnost plavat [naměřený čas na uplavání 500 m], původ [A: −1 Skandinávie, B: +1 Středomoří], inteligenční kvocient IQ [evropský standardizovaný test IQ]. Mezi znaky jsou tři dichotomické, binární proměnné, a to pohlaví, délka vlasů a původ a ostatních 9 znaků nabývá kvantitativních hodnot.*

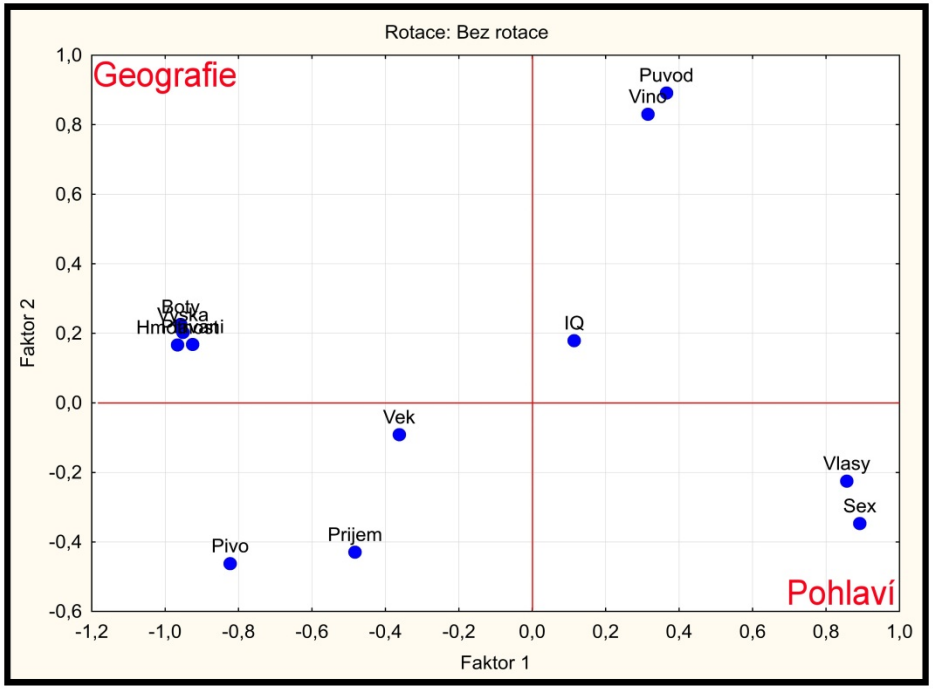
Osoba	Výška	Hmotnost	Vlasy	Boty	Věk	Příjem	Pivo	Víno	Sex
	Plavání	Původ	IQ						
MA	198	92	−1	48	48	45000	420	115	−1
	98	−1	100						
MA	184	84	−1	44	33	33000	350	102	−1
	92	−1	130						
MA	183	83	−1	44	37	34000	320	98	−1
	91	−1	127						
FA	166	47	−1	36	32	28000	270	78	1
	75	−1	112						

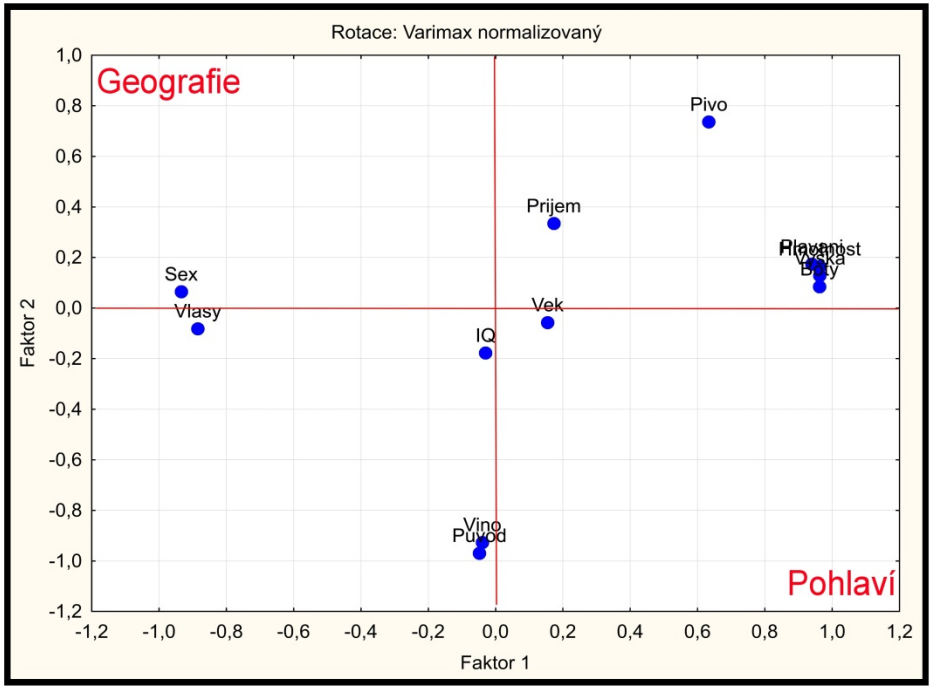

Rotace: Bez rotace
Puvod
Vino
IQ
Boty
Vyska
Hmotnost
Vek
Vlasy
Sex
Prijem
Pivo
Faktor 2
Faktor 1
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-1,2
-1,0
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0

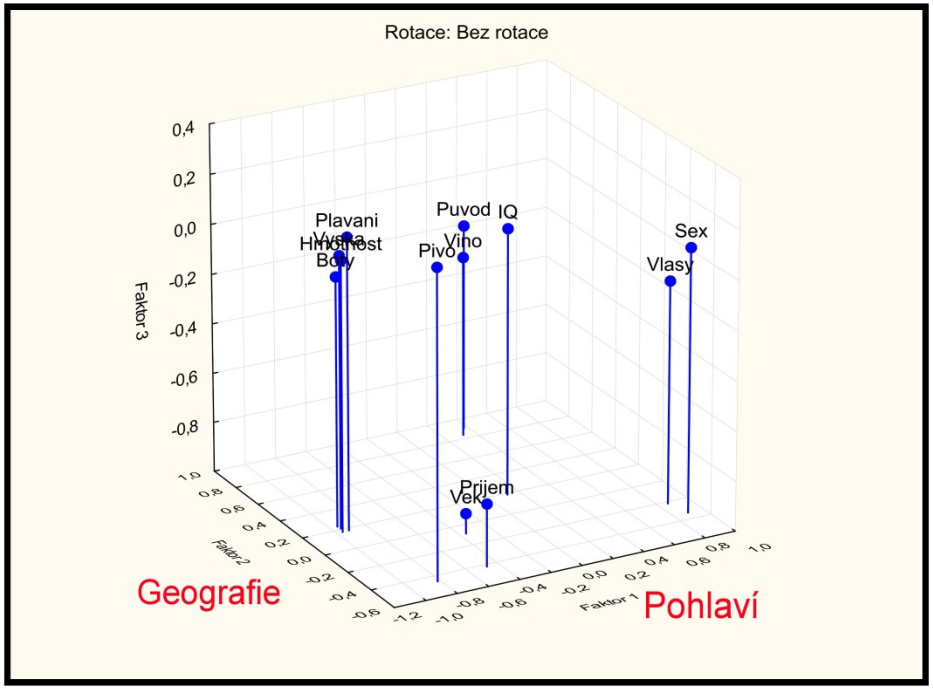

Rotace: Varimax normalizovaný
Pivo
Prijem
Plavani
Hrubost
Vaska
Boty
Sex
Vlasy
Vek
IQ
Vino
Puvod
Faktor 2
Faktor 1
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-1,2
-1,2
-1,0
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0
1,2

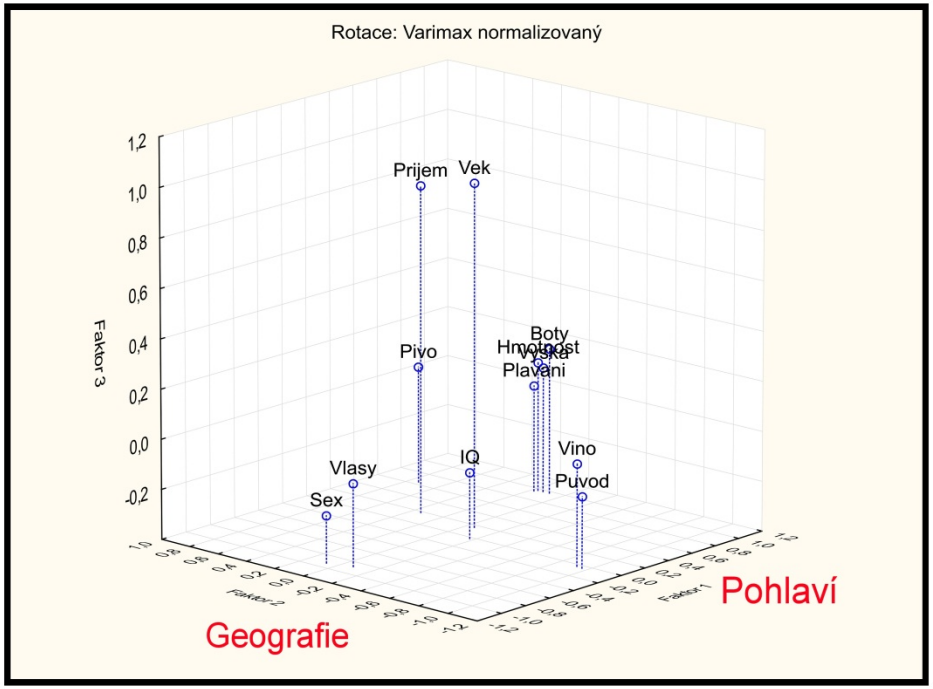

2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,5
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
FAKTOR2
FAKTOR1
2MB
1MB
8MB
4MB
3MB
5MB
6MB
7MB
2FB
5FB
8FB
7FB
4FB
1FB
6FB
3FB
2MA
3MA
1MA
4MA
5MA
7MA
8MA
6MA
2FA
3FA
7FA
5FA
4FA
1FA
8FA
6FA

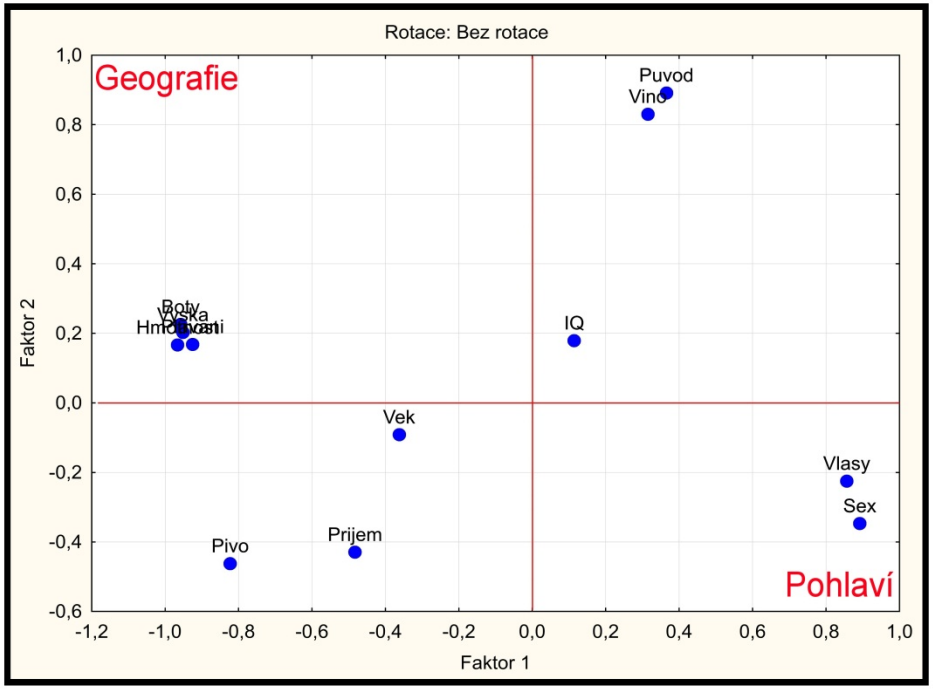

FAKTOR2
FAKTOR1
1,4
1,2
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-1,2
-1,4
-1,6
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
6MA
3FA
2FA
5MA
4MA
7MA
8MA
1MA
2MA
3MA
6FA
8FA
7FA
4FB
3FB
6FB
5FB
1FB
1MB
2MB
3MB
5MB
4MB
6MB
7MB
8MB
7FB
8FB
2FB

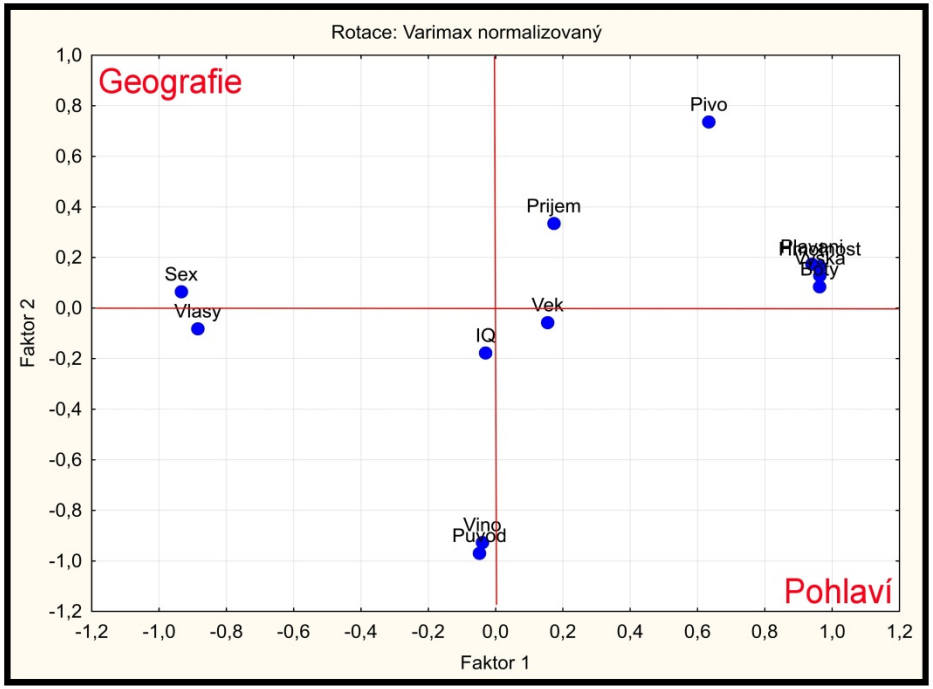

Rotace: Bez rotace
Geografie
Pohlaví
Faktor 1
Faktor 2
Puvod
Vino
IQ
Vek
Vlasy
Sex
Pivo
Prijem
Boty
Vyska
Hmotnost
Plavani

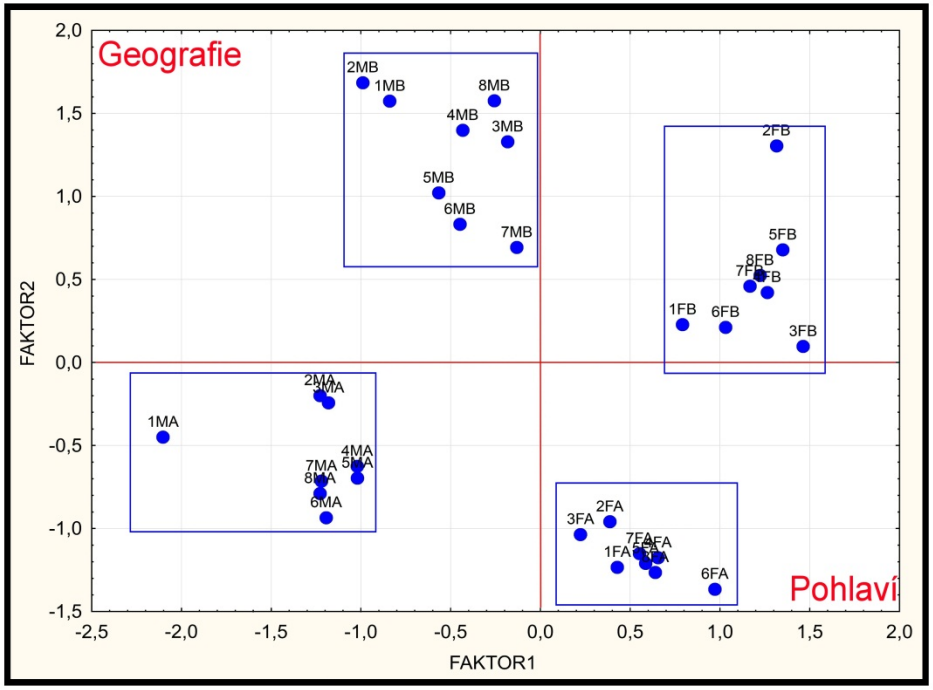

Rotace: Varimax normalizovaný
Geografie
Pohlaví
Faktor 1
Faktor 2
Pivo
Prijem
Sex
Vlasy
Vek
IQ
Vino
Puvod
Plavani
Hmotnost
Vyska
Boty

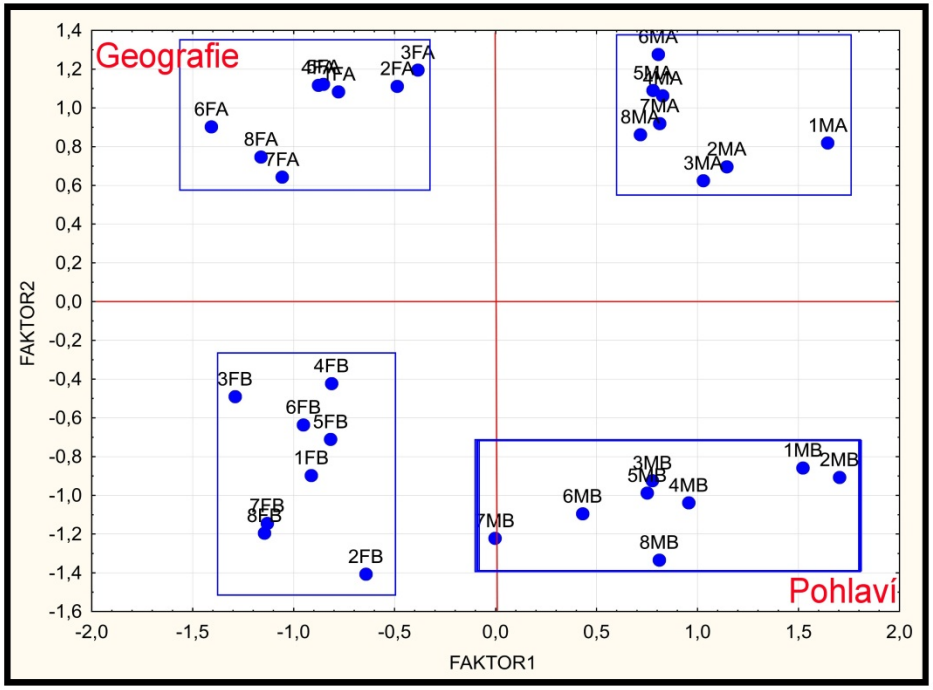

Rotace: Bez rotace
Geografie
Pohlaví
Faktor 1
Faktor 2
Faktor 3
Plavani
Vyska
Hmotnost
Boty
Puvod
IQ
Pivo
Vino
Sex
Vlasy
Prijem
Vek

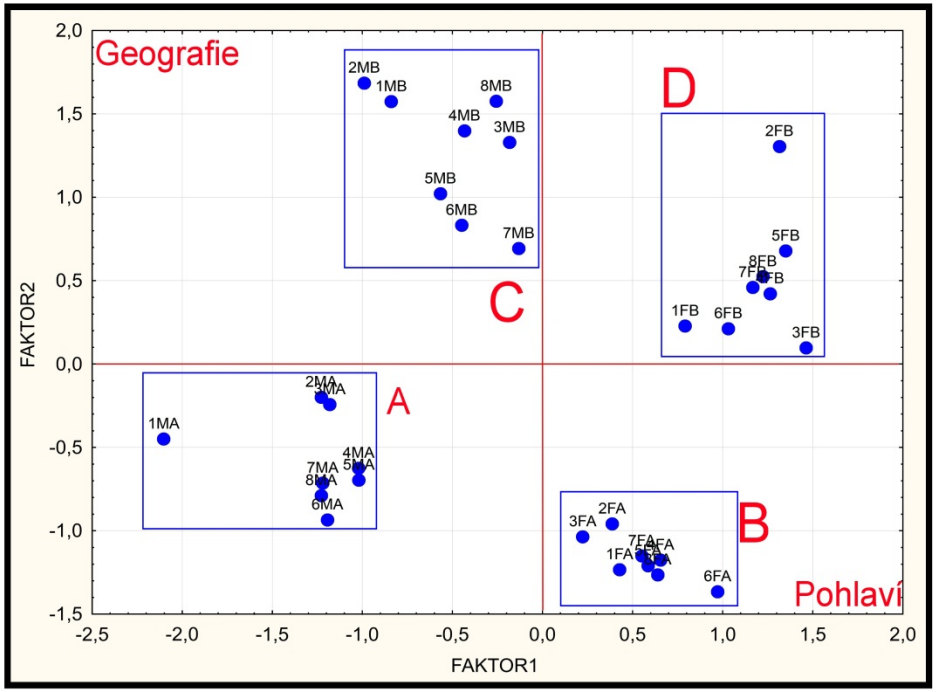

Rotace: Varimax normalizovaný
Geografie
Pohlaví
Faktor 1
Faktor 2
Faktor 3
Prijem
Vek
Pivo
Boty
Hmotnost
Vyska
Plavani
Vino
Puvod
IQ
Vlasy
Sex

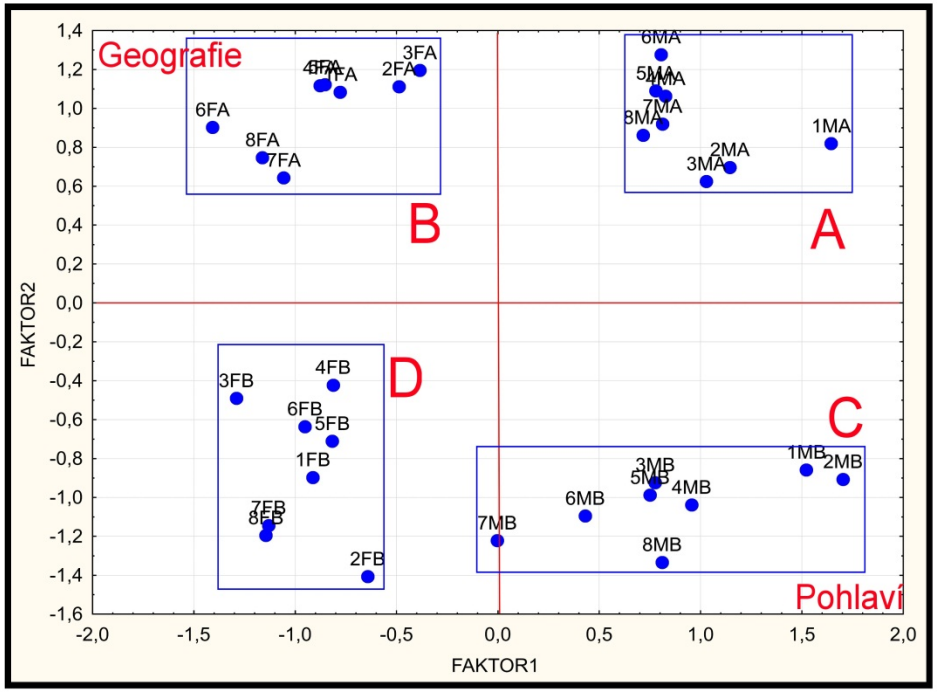

Rotace: Bez rotace
Geografie
Pohlaví
Faktor 1
Faktor 2
Puvod
Vino
IQ
Vek
Vlasy
Sex
Prijem
Pivo
Boty
Výška
Hmotnost
Plavení

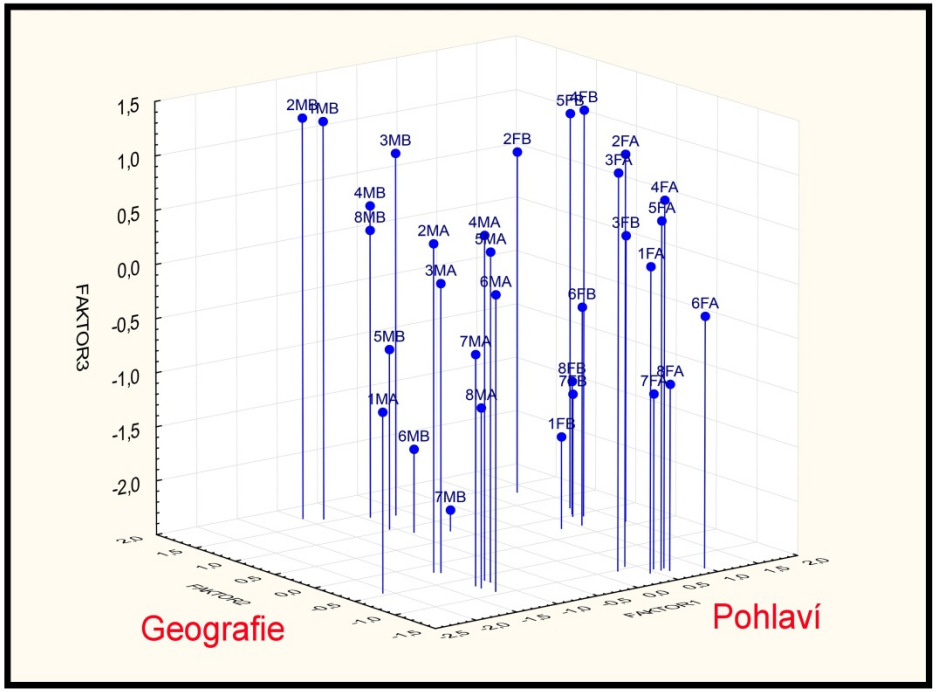

Rotace: Varimax normalizovaný
Geografie
Pohlaví
Faktor 1
Faktor 2
Pivo
Prijem
Sex
Vlasy
Vek
IQ
Vino
Puvod
Plavení
Hmotnost
Výška
Boty

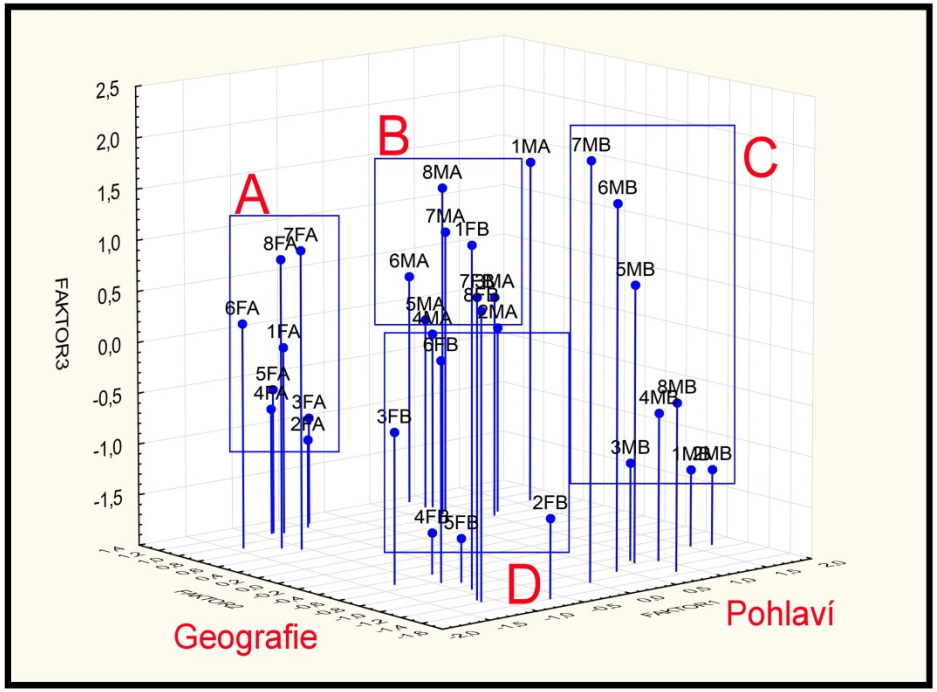

Geografie
Pohlaví
FAKTOR1
FAKTOR2
2MB
1MB
8MB
4MB
3MB
5MB
6MB
7MB
2FB
5FB
8FB
7FB
1FB
6FB
3FB
2MA
3MA
1MA
4MA
5MA
7MA
8MA
6MA
2FA
3FA
7FA
1FA
6FA

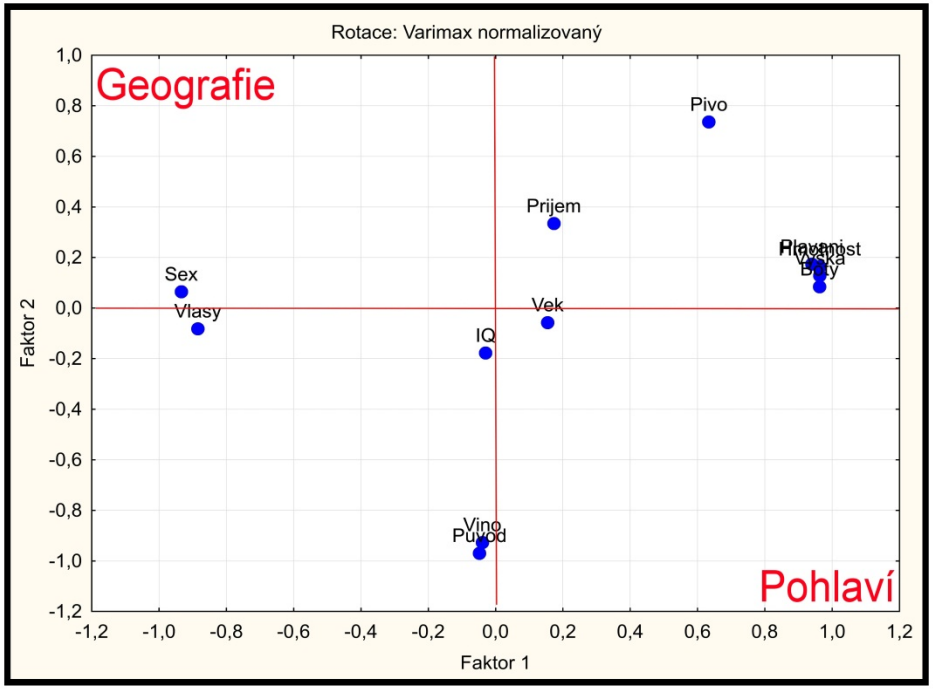

Geografie
Pohlaví
FAKTOR1
FAKTOR2
3FA
2FA
6FA
8FA
7FA
6MA
5MA
4MA
7MA
8MA
1MA
2MA
3MA
4FB
3FB
6FB
5FB
1FB
7FB
8FB
2FB
1MB
2MB
3MB
5MB
4MB
6MB
7MB
8MB

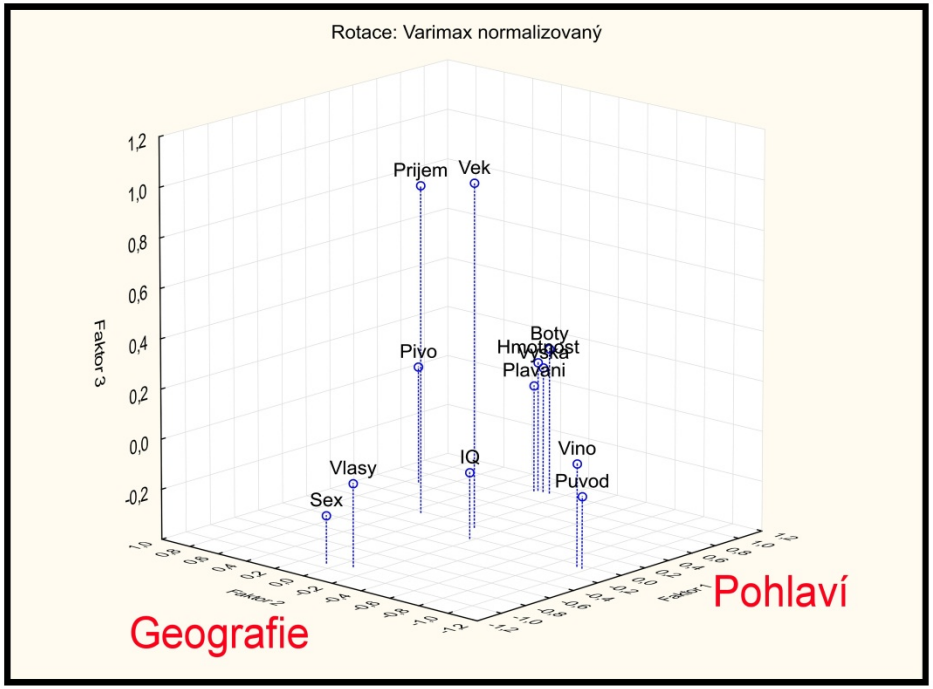

Geografie
Pohlaví
A
B
C
D
FAKTOR1
FAKTOR2

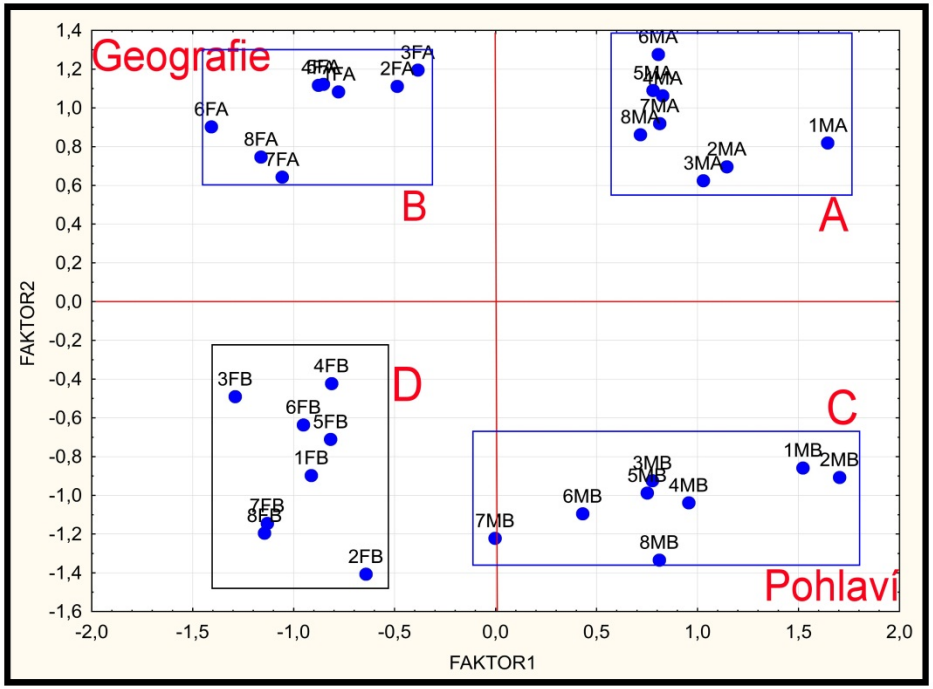

Geografie
Pohlaví
A
B
C
D
FAKTOR1
FAKTOR2

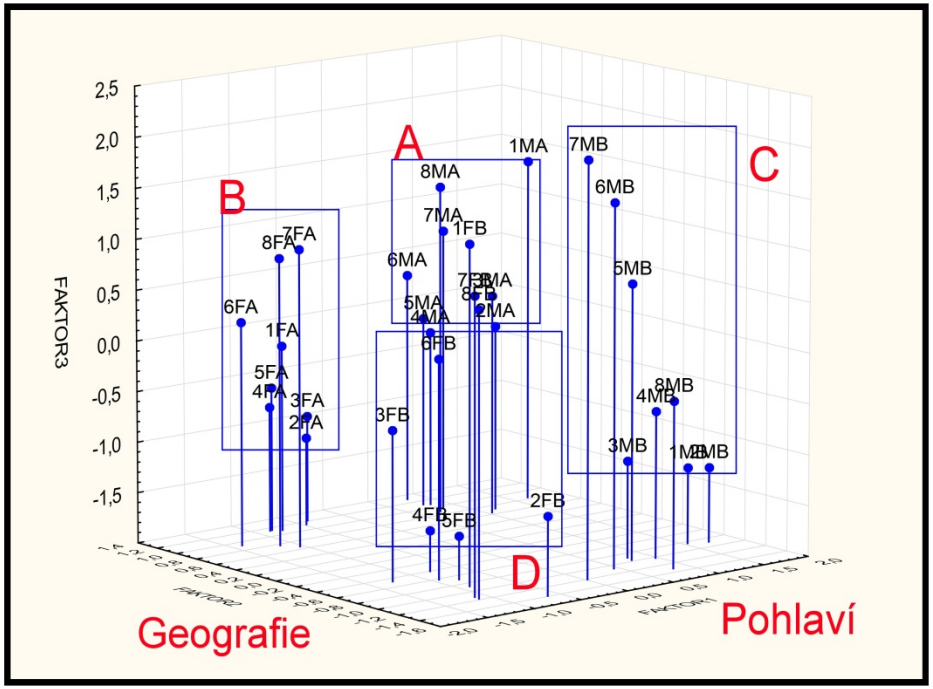

Geografie
Pohlaví
FAKTOR3


Geografie
Pohlaví
A
B
C
D
FAKTOR3


Rotace: Varimax normalizovaný
Geografie
Pohlaví
Pivo
Prijem
Plavani
Hmotnost
Vyska
Boty
Sex
Vlasy
Vek
IQ
Vino
Puvod
Faktor 1
Faktor 2


Rotace: Varimax normalizovaný
Prijem
Vek
Pivo
Boty
Hmotnost
Vyska
Plavani
Vino
Puvod
IQ
Vlasy
Sex
Faktor 3
Pohlaví
Geografie


Geografie
Pohlaví
A
B
C
D
3FA
2FA
6FA
8FA
7FA
6MA
5MA
7MA
8MA
1MA
2MA
3MA
3FB
4FB
6FB
5FB
1FB
7FB
8FB
2FB
1MB
2MB
3MB
5MB
4MB
6MB
7MB
8MB
FAKTOR1
FAKTOR2

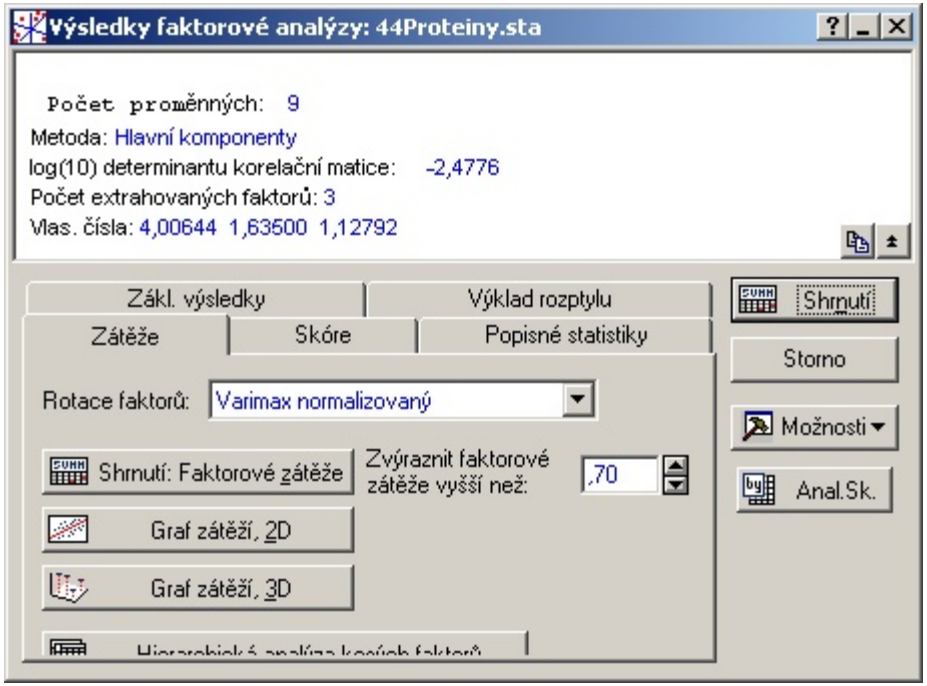

A
B
C
D
1MA
8MA
7MA
1FB
6MA
5MA
4MA
2MA
6FB
7FB
8FB
3FB
2FB
4FB
5FB
7FA
8FA
6FA
1FA
5FA
4FA
3FA
2FA
7MB
6MB
5MB
8MB
4MB
3MB
1MB
2MB
FAKTOR3
Geografie
Pohlaví

PŘÍKLAD 4.4 *Sledování spotřeby proteinů v zemích Evropy*

Byla sledována spotřeba proteinů v 25 zemích Evropy formou spotřeby 9 druhů potravin. Greenacre (1984) převzal ve své učebnici data *Proteiny od Webera z díla Agrarpolitik im Spannungsfeld der Internationalen Ernährungspolitik, Kiel Universität, 1973. Cílem úlohy je především odhalit, zda existuje korelace mezi znaky, tj druhy potravin? Lze odhalit nějaké interakce mezi druhy potravin a zeměmi?*

Data: *V datech Proteiny jsou uvedeny znaky: Červené značí spotřebu červeného masa, Bílé značí spotřebu bílého masa, Vejce značí spotřebu vajec, Mléko se týká spotřeby mléka, Ryby značí spotřebu ryb, Obiln značí spotřebu obilnin, Škrob značí spotřebu škrobu, Ořechy značí spotřebu ořechů, Ovoce značí spotřebu ovoce a zeleniny.*

Země	*Červené*	*Bílé*	*Vejce*	*Mléko*	*Ryby*	*Obiln*	*Škrob*	*Ořechy*	*Ovoce*
Albánie	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Rakousko	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgie	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulharsko	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Československo	4	9.7	11.4	2.8	12.5	2	34.3	5	1.1
Dánsko	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4


Výsledky faktorové analýzy: 44Proteiny.sta
Počet proměnných: 9
Metoda: Hlavní komponenty
log(10) determinantu korelační matice: -2,4776
Počet extrahovaných faktorů: 3
Vlas. čísla: 4,00644 1,63500 1,12792
Zákl. výsledky
Výklad rozptylu
Zátěže
Skóre
Popisné statistiky
Rotace faktorů:
Varimax normalizovaný
Shrnutí: Faktorové zátěže
Zvýraznit faktorové zátěže vyšší než:
,70
Graf zátěží, 2D
Graf zátěží, 3D
Shrnutí
Storno
Možnosti
Anal.Sk.

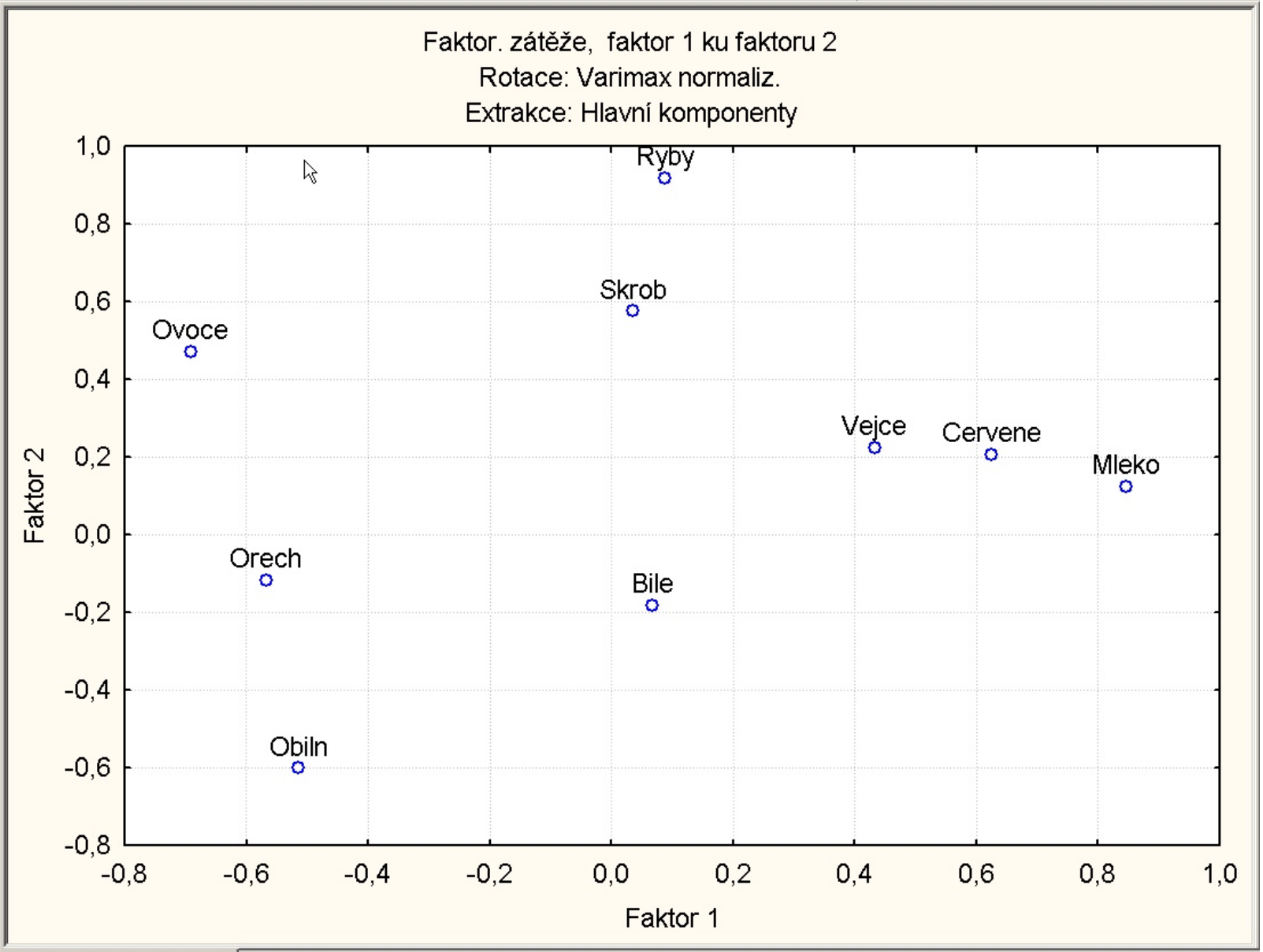
STATISTICA Cz - PS 1* - [Faktor. zátěže, faktor 1 ku faktoru 2] - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 2]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (44Protei
Dialog výsledků faktorov
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2

Výsledky faktorové a...

Připravena

44Proteiny.sta ABC 123 ZÁZN

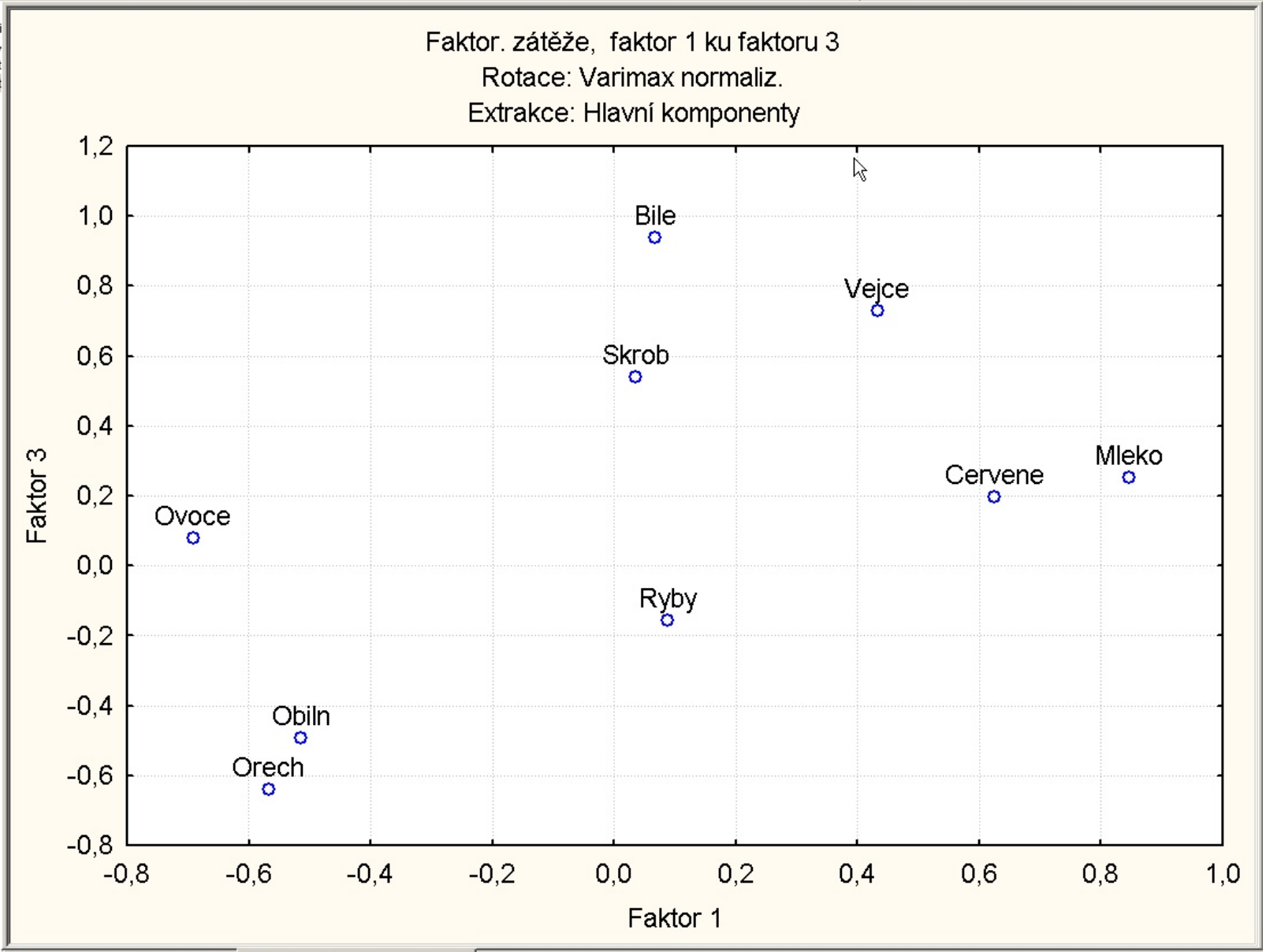
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (44Protei
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3

Výsledky faktorové a...

Připravena 44Proteiny.sta ABC 123 ZÁZN

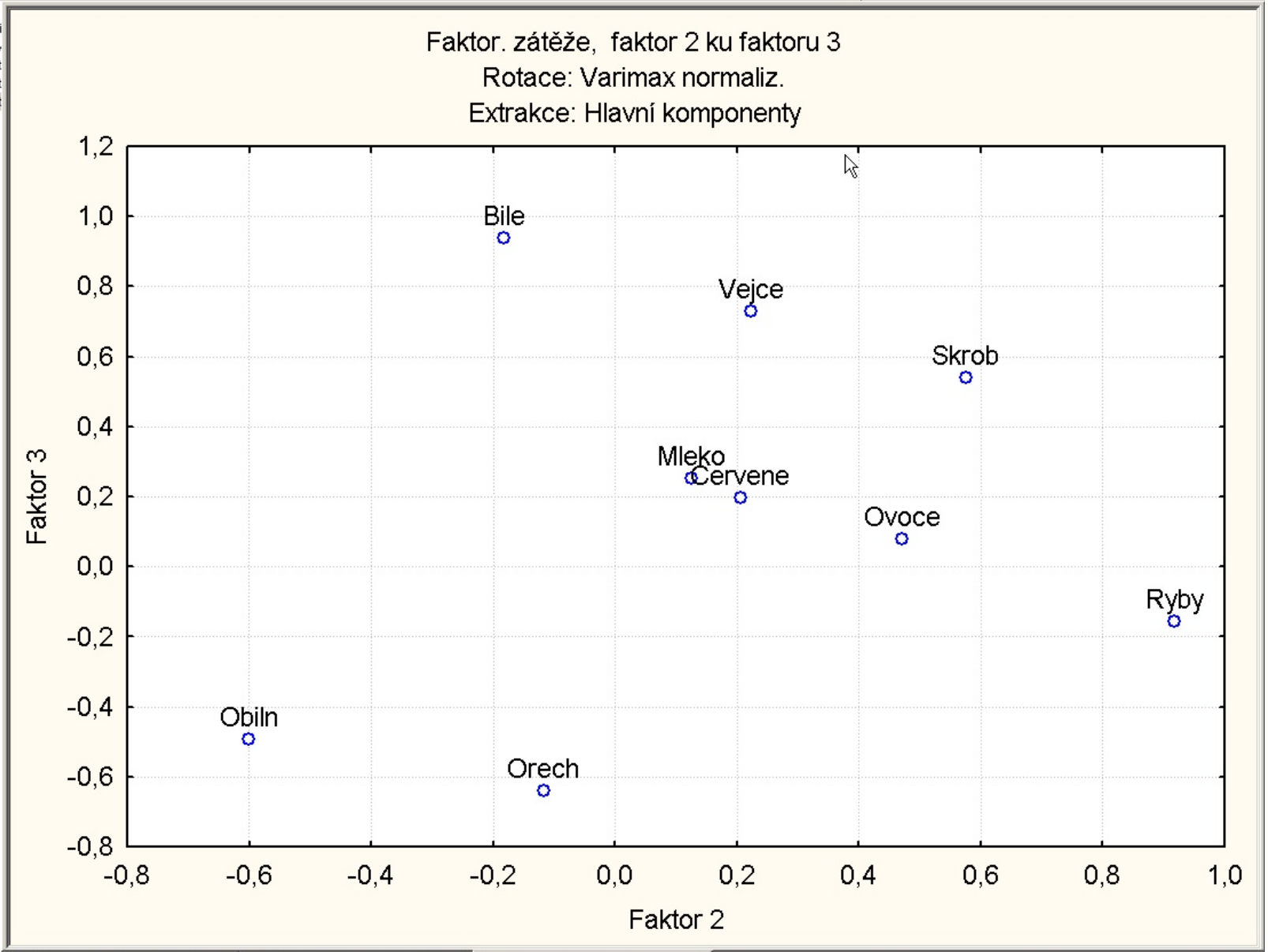
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 2 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (44Protei
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3 | Faktor. zátěže, faktor 2 ku faktoru 3

Výsledky faktorové a...

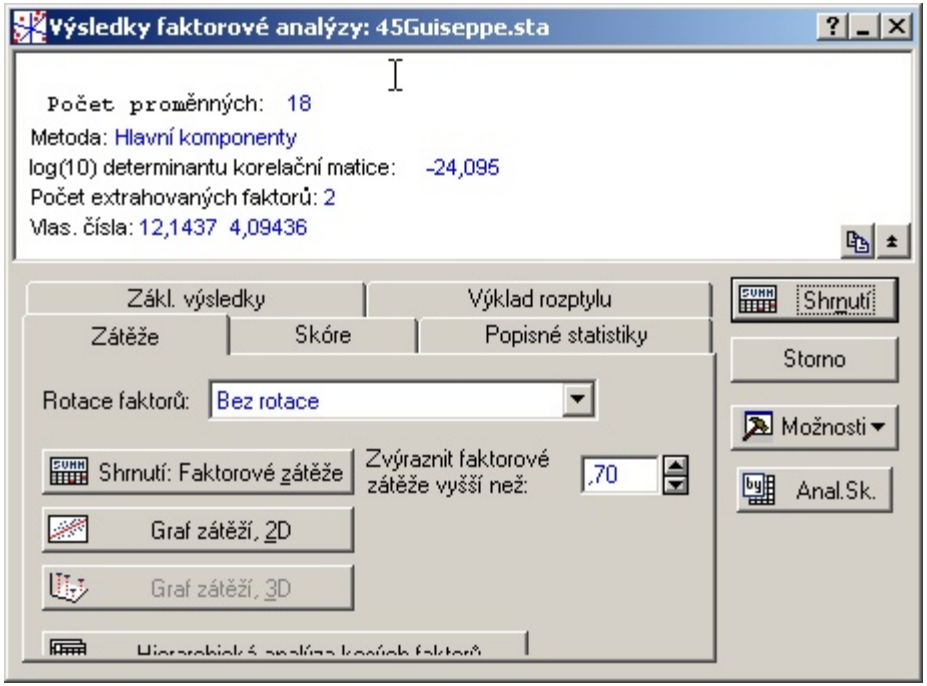
Připravena 44Proteiny.sta ABC 123 ZÁZN

PŘÍKLAD 4.5 *Chromatografická analýza farmakologických sloučenin*

Byly měřeny hodnoty R_F pro 20 sloučenin s 18 eluenty, úloha B401 [100]. 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 : etanol : 30% amoniak = 45 : 45 : 7 : 3, x_2 směs etylacetát : benzen : metanol : 30% amoniak = 60 : 35 : 6.5 : 2.5, x_3 směs benzen : dioxan : etanol : 30% amoniak = 50 : 40 : 7.5 : 2.5, x_4 směs metanol : 30% amoniak = 100 : 1.5, x_5 směs benzen : 2-propanol : metanol : 30% amoniak = 70 : 30 : 20 : 5, x_6 směs etylacetát : metanol : 30% amoniak = 85 : 10 : 5, x_7 směs cyklohexan : toluen : dietylamin = 65 : 25 : 10, x_8 směs cyklohexan : toluen : dietylamin = 75 : 15 : 10, x_9 směs cyklohexan : benzen : metanol : dietylamin = 70 : 20 : 10 : 5, x_{10} směs chloroform : aceton : dietylamin = 50 : 40 : 10, x_{11} směs cyklohexan : chloroform : dietylamin = 50 : 40 : 10, x_{12} směs benzen : etylacetát : dietylamin = 50 : 40 : 10, x_{13} směs xylen : metyletylketon : metanol : dietylamin = 40 : 40 : 6 : 2, x_{14} směs dietyleter : dietylamin = 95 : 5, x_{15} směs etylacetát : chloroform = 50 : 50, x_{16} směs etylacetát : chloroform [A] = 50 : 50, x_{17} směs butanol : metanol = 40 : 60, x_{18} směs butanol : metanol [A] = 40 : 60, kde [A] značí, že byl užit 0.1M metanolát 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


Výsledky faktorové analýzy: 45Guiseppe.sta
Počet proměnných: 18
Metoda: Hlavní komponenty
log(10) determinantu korelační matice: -24,095
Počet extrahovaných faktorů: 2
Vlas. čísla: 12,1437 4,09436
Zákl. výsledky
Výklad rozptylu
Zátěže
Skóre
Popisné statistiky
Rotace faktorů:
Bez rotace
Shrnutí: Faktorové zátěže
Zvýraznit faktorové zátěže vyšší než:
,70
Graf zátěží, 2D
Graf zátěží, 3D
Shrnutí
Storno
Možnosti
Anal.Sk.

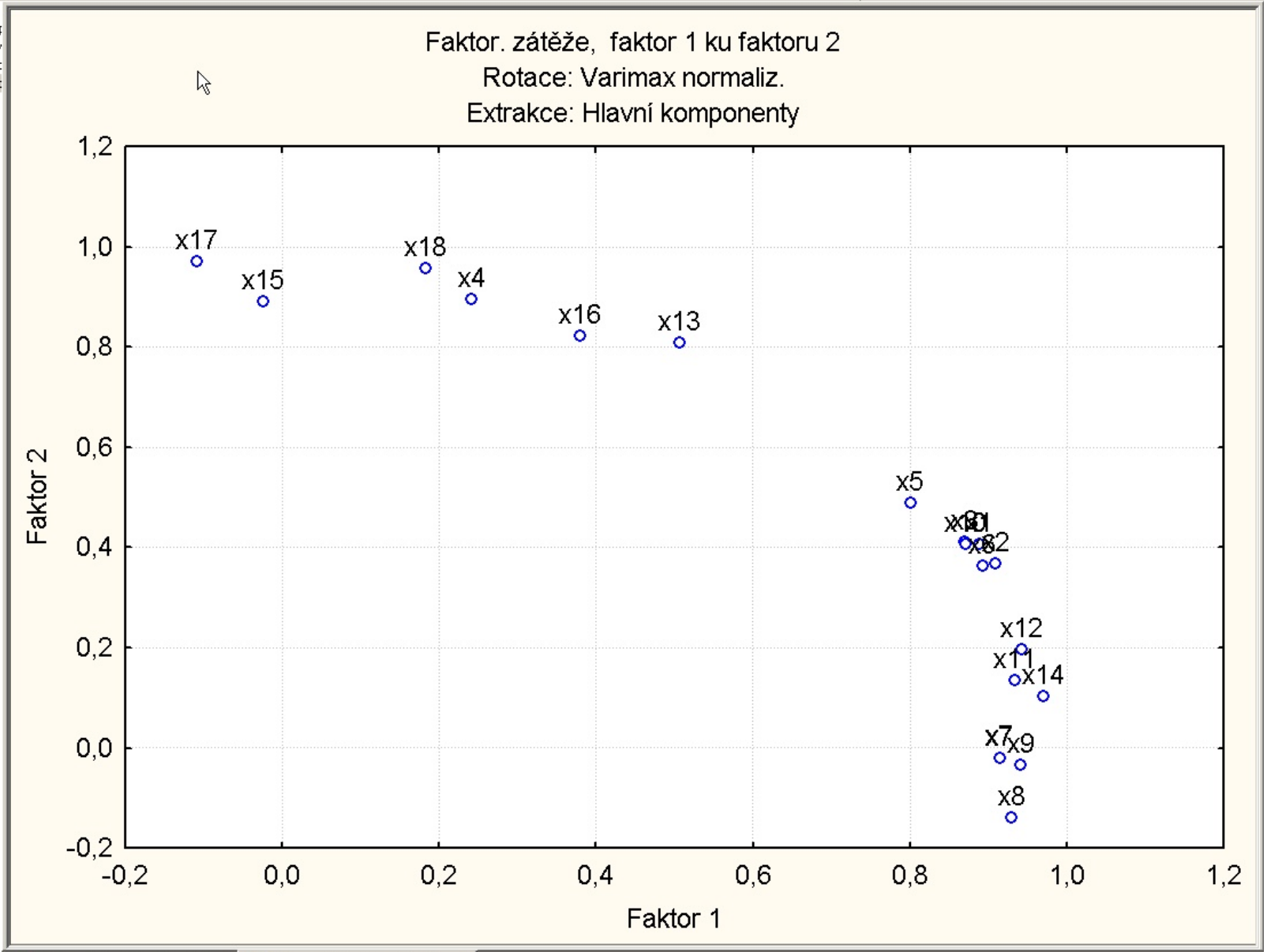
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 2]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (45Guiseppe
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt

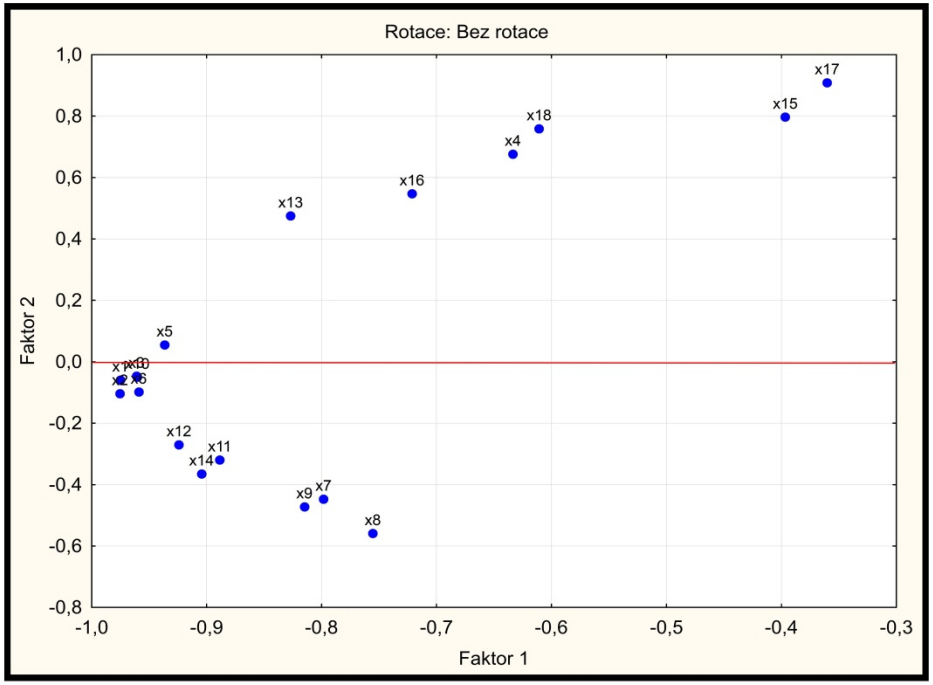


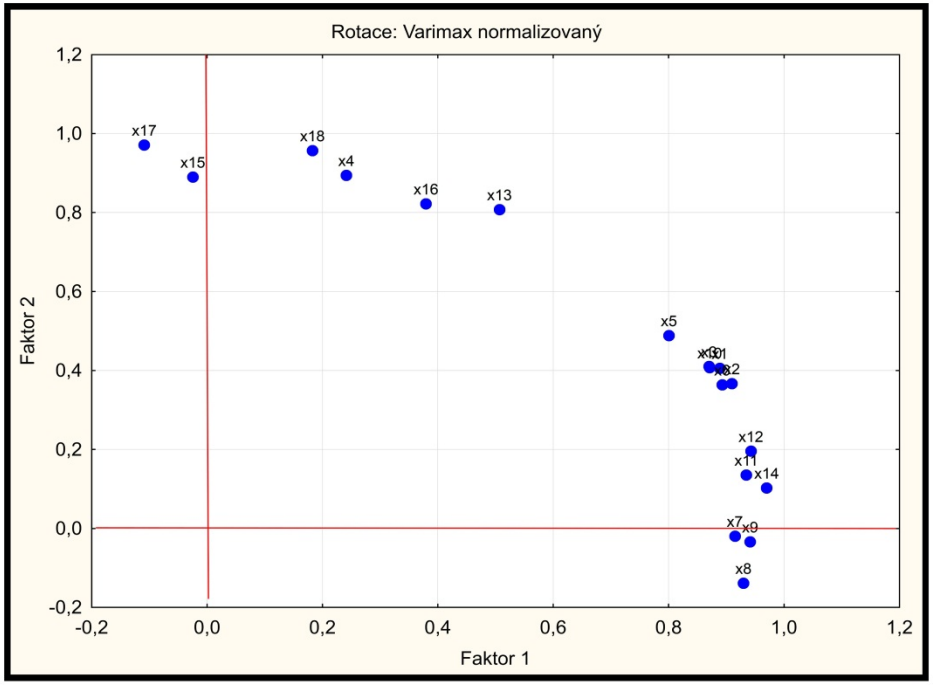
Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 2

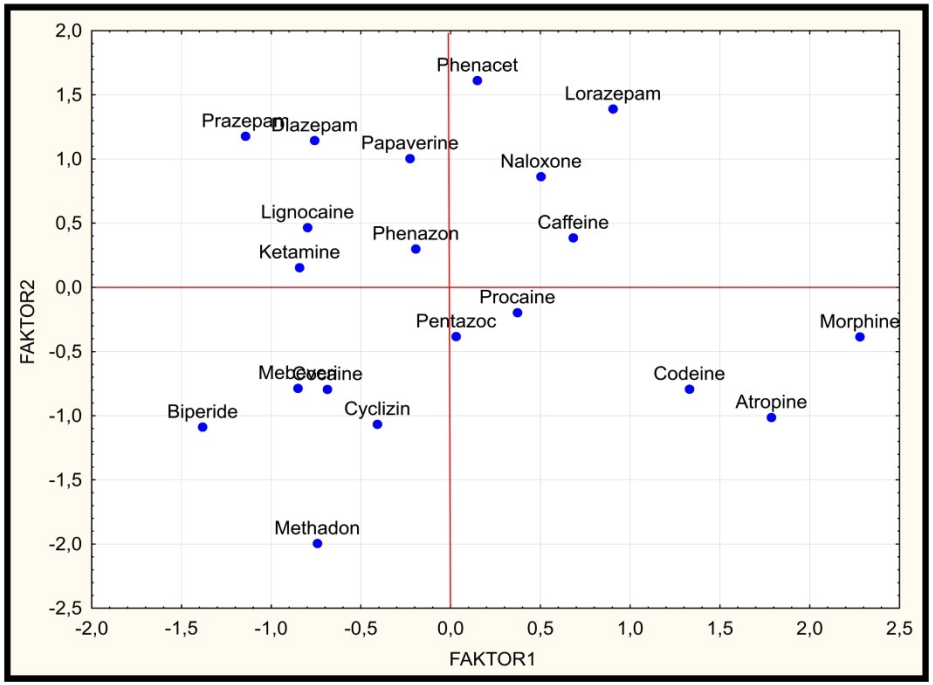
Výsledky faktorové a...

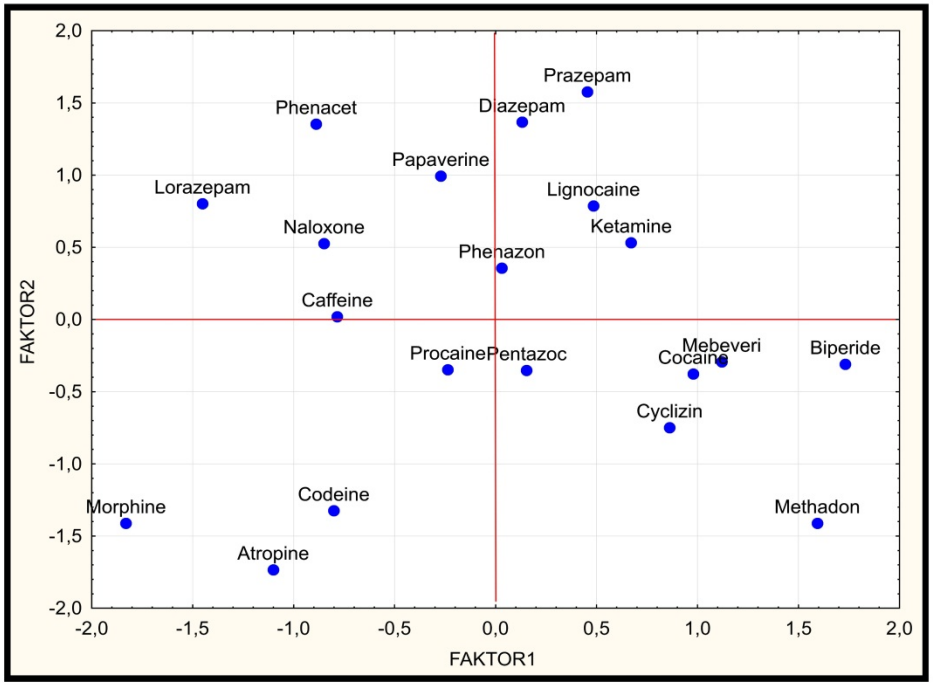
Připravena

45Guiseppe.st ABC 123 ZÁZN


Rotace: Bez rotace
Faktor 2
Faktor 1
x17
x15
x18
x4
x16
x13
x5
x12
x11
x14
x9
x7
x8


Rotace: Varimax normalizovaný
Faktor 2
Faktor 1
x17
x15
x18
x4
x16
x13
x5
x12
x11
x14
x7
x9
x8


FAKTOR2
FAKTOR1
Phenacet
Lorazepam
Prazepam
Diazepam
Papaverine
Naloxone
Lignocaine
Caffeine
Phenazon
Ketamine
Procaine
Pentazoc
Morphine
Codeine
Atropine
Biperide
Cyclizin
Methadon

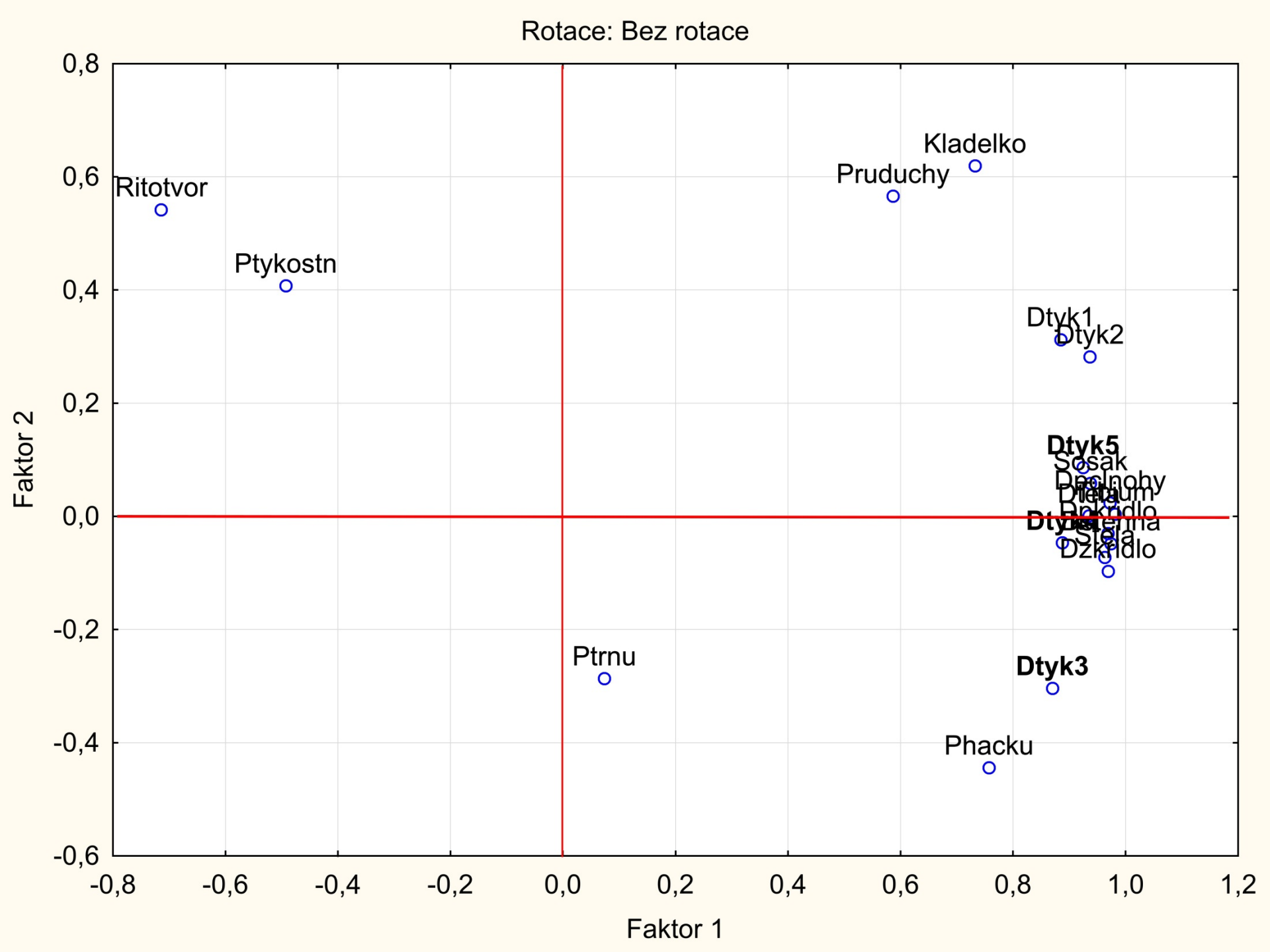

FAKTOR2
FAKTOR1
Prazepam
Phenacet
Diazepam
Papaverine
Lorazepam
Lignocaine
Naloxone
Ketamine
Phenazon
Caffeine
Procaine
Pentazoc
Mebeveri
Cocaine
Biperide
Cyclizin
Codeine
Morphine
Methadon
Atropine

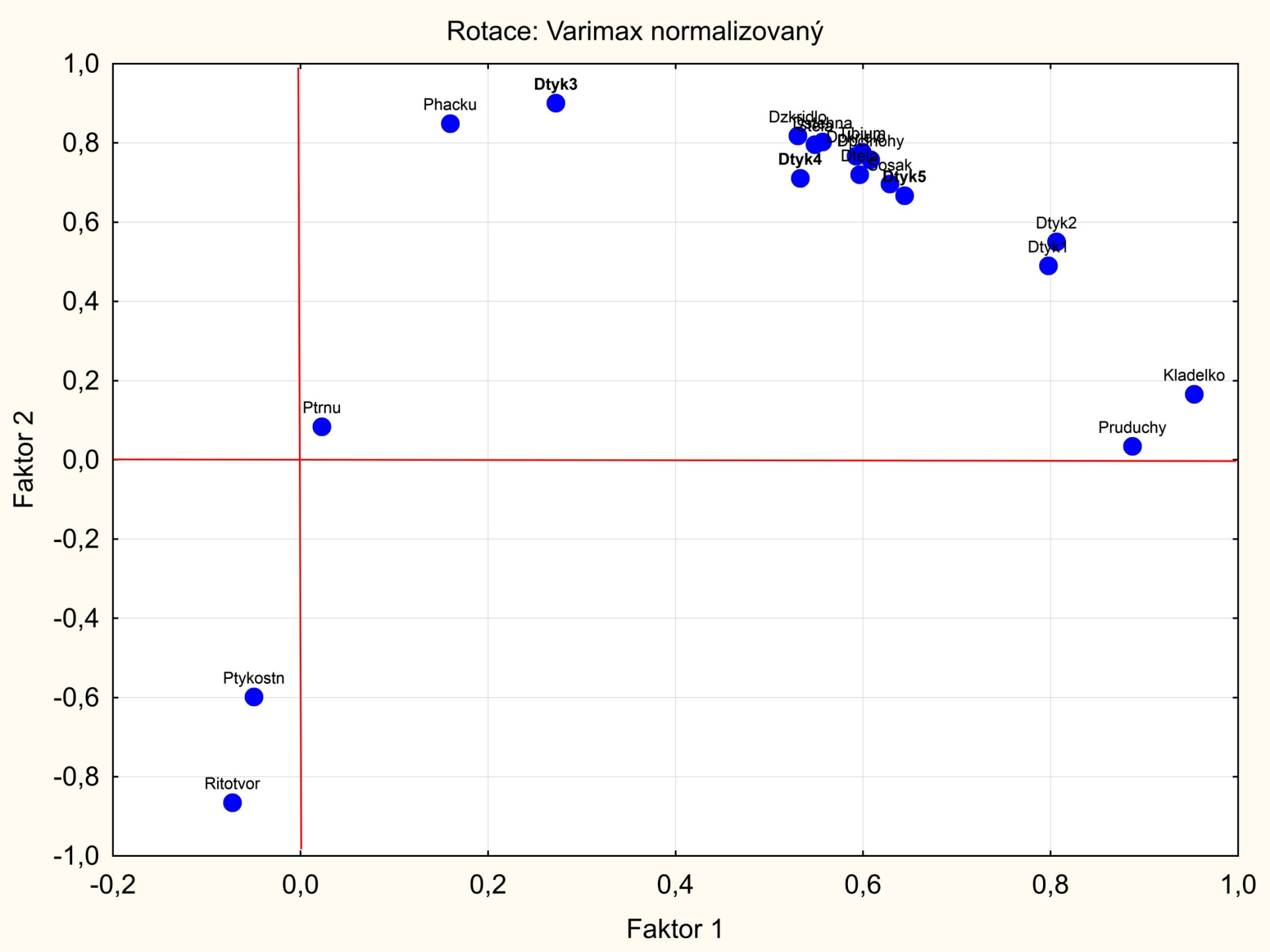
PŘÍKLAD 4.6 *Popis a třídění polétavých mšic*

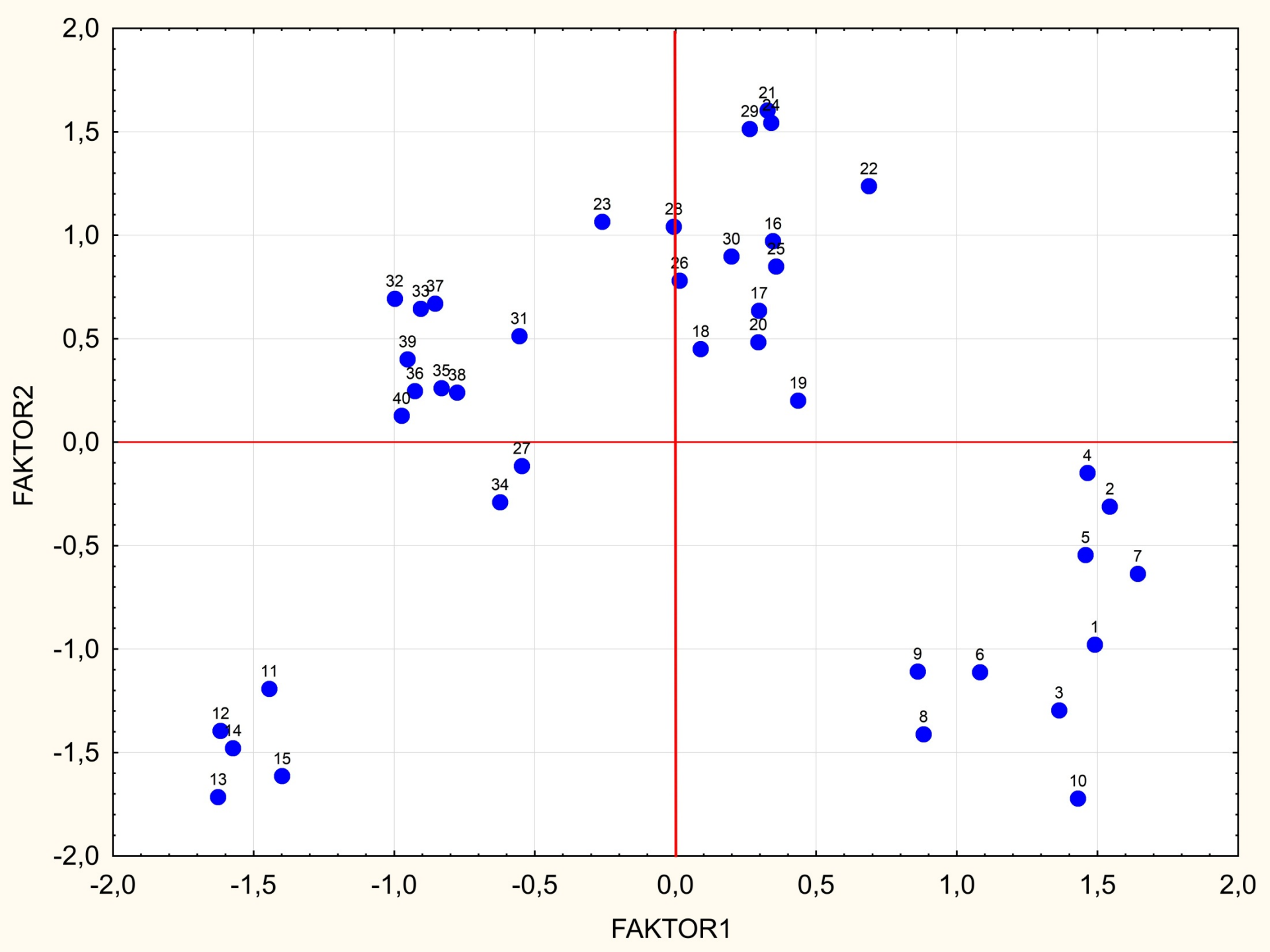
Jeffers (1967) studoval 40 jedinců polétavých mšic *(Alate adelges) za pomoci světelné pasti. U mšic bylo změřeno 19 znaků k rozlišení druhů tohoto hmyzu souboru Mšice: 14 znaků se týká délky nebo šířky, 4 znaky se týkají počtu a 1 znak je binární vyjadřující přítomnost či absenci. Mšice se totiž obtížně rozlišují dle běžných taxonometrických klíčů a obvykle je nutné detailní vyšetření dat, hledání společných rysů a následná klasifikace do větších celků. Užitím metody hlavních komponent PCA je třeba snížit původní počet 19 znaků na menší počet proměnných, který by vystihl co největší množství proměnlivosti mšic. Před užitím PCA je potřeba provést standardizaci dat, protože znaky představují směs délek a počtů.*

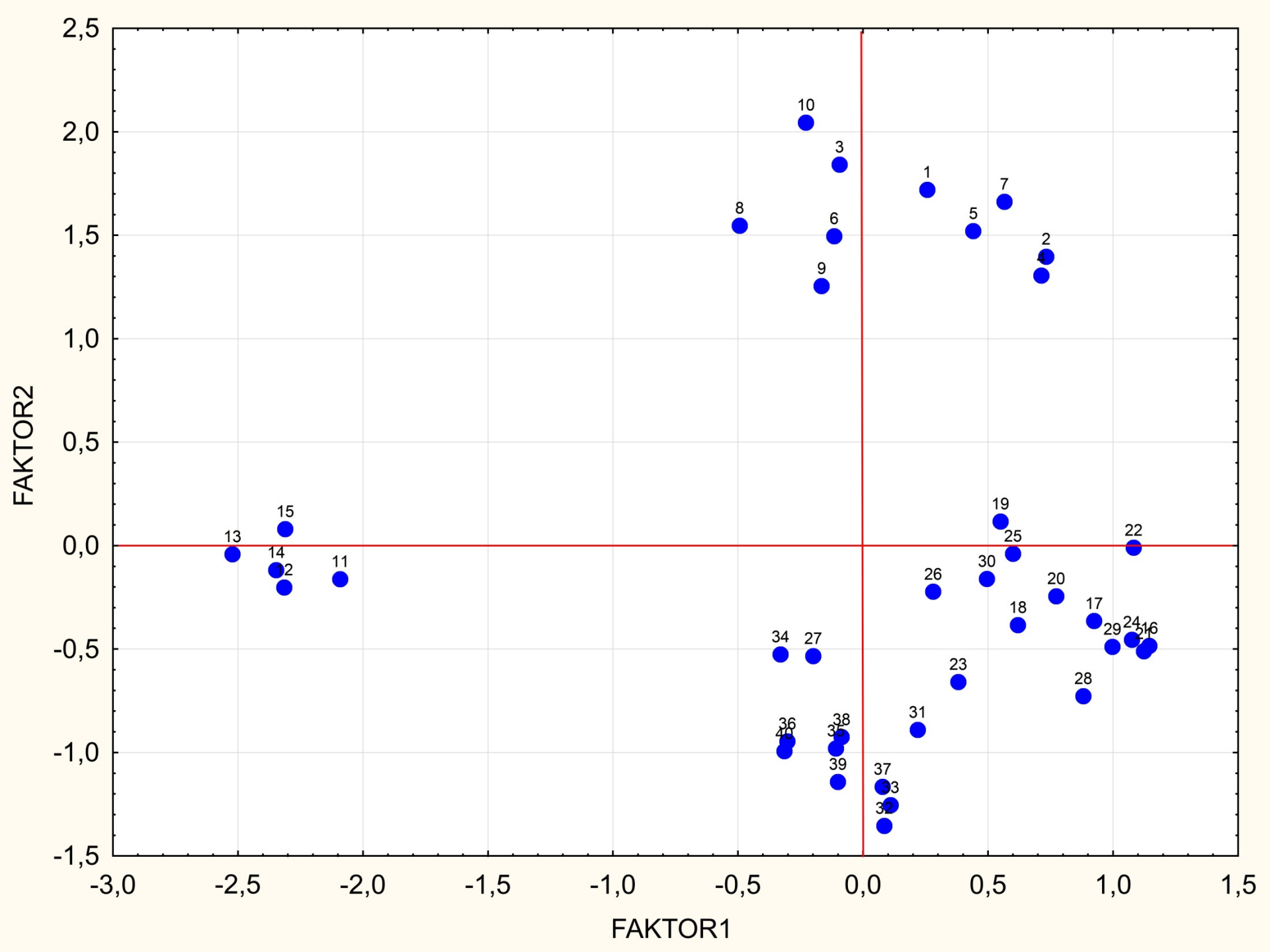
Data: *Charakter znaků − i index jedince, x_1 značí délku těla, x_2 značí šířku těla, x_3 je délka předního křídla, x_4 je délka zadního křídla, x_5 je počet průduchů, x_6 je délka tykadla I, x_7 je délka tykadla II, x_8 je délka tykadla III, x_9 je délka tykadla IV, x_{10} je délka tykadla V, x_{11} je počet tykadlových ostnů, x_{12} je délka posledního článku nohy, x_{13} je délka holeně, tibia, x_{14} je délka stehna, x_{15} je délka sosáku, x_{16} je délka kladélka, x_{17} je počet kladélkových trnů, x_{18} je řitní otvor, x_{19} je počet háčků zadních křídel.*

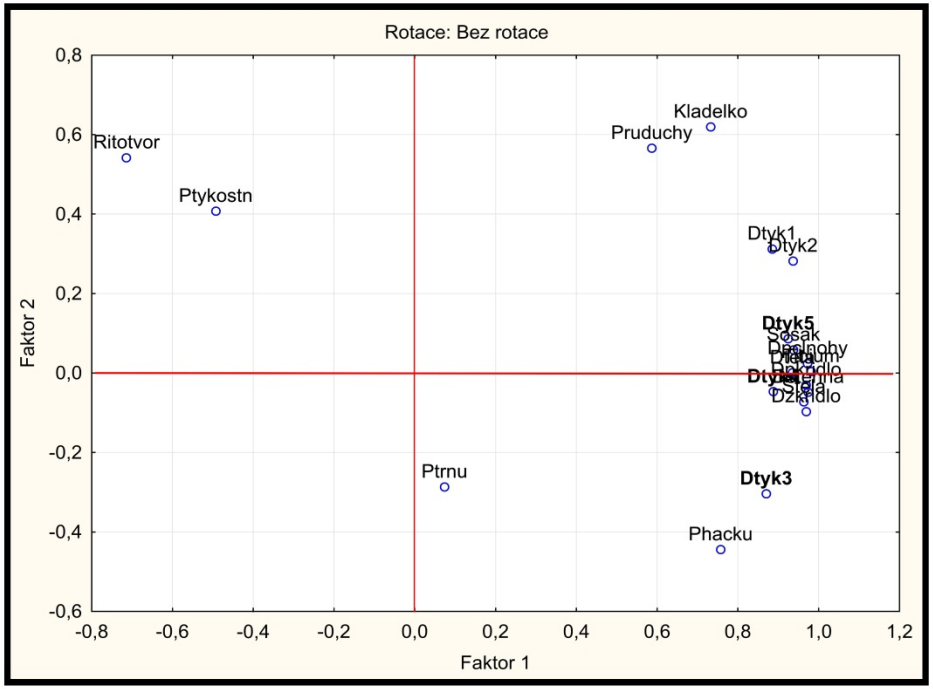
i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
	x_{19}								
1	21.2	11.0	7.5	4.8	5.0	2.0	2.0	2.8	2.8
	3.3	3.0	4.4	4.5	3.6	7.0	4.0	8.0	0.0
	3.0								
2	20.2	10.0	7.5	5.0	5.0	2.3	2.1	3.0	3.0
	3.2	5.0	4.2	4.5	3.5	7.6	4.2	8.0	0.0
	3.0								
3	20.2	10.0	7.0	4.6	5.0	1.9	2.1	3.0	2.5
	3.3	1.0	4.2	4.4	3.3	7.0	4.0	6.0	0.0
	3.0								

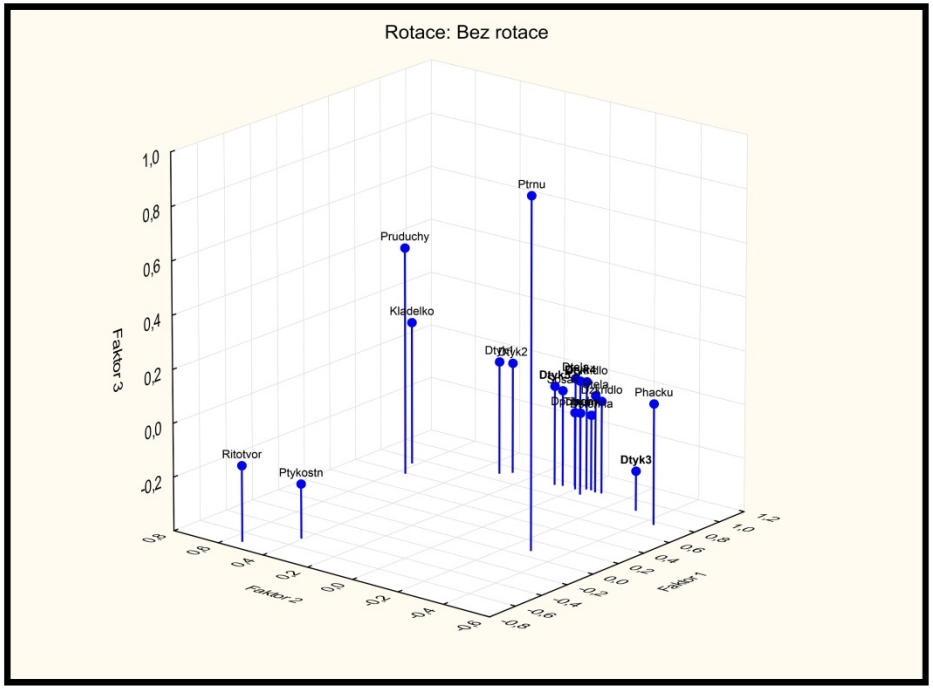

Rotace: Bez rotace
Faktor 2
Faktor 1
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0
1,2
Ritotvor
Ptykostn
Kladelko
Pruduchy
Dtyk1
Dtyk2
Dtyk5
Sosak
Dnolnohy
Dtyk3
Ptrnu
Phacku

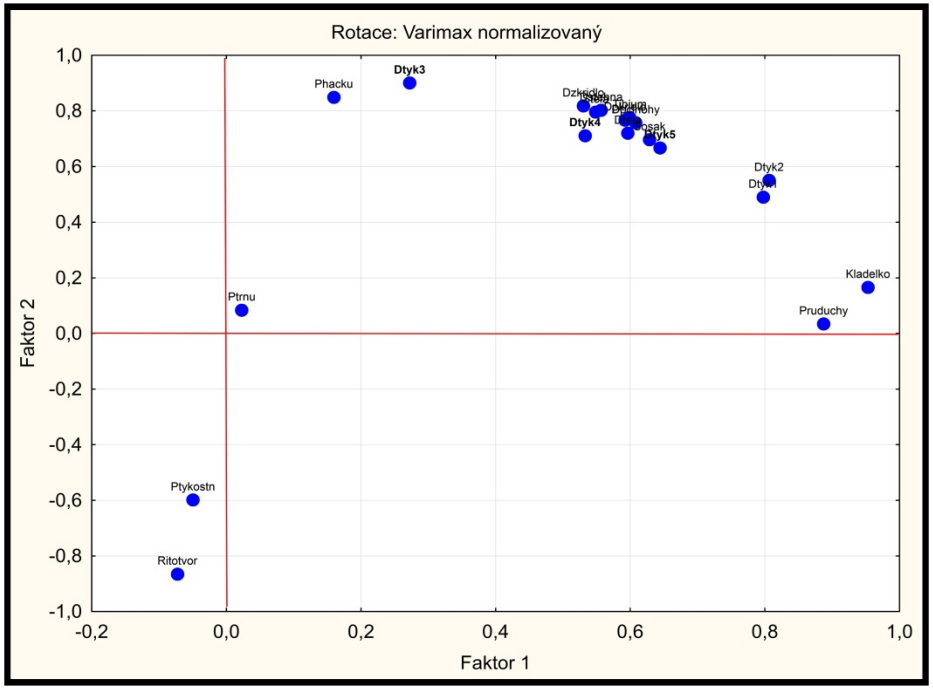

Rotace: Varimax normalizovaný
Faktor 2
Faktor 1
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-0,2
0,0
0,2
0,4
0,6
0,8
1,0
Dtyk3
Phacku
Dzkridlo
Dtyk4
Dtela
Sosak
Dtyk5
Dtyk2
Dtyk1
Kladelko
Ptrnu
Pruduchy
Ptykostn
Ritotvor

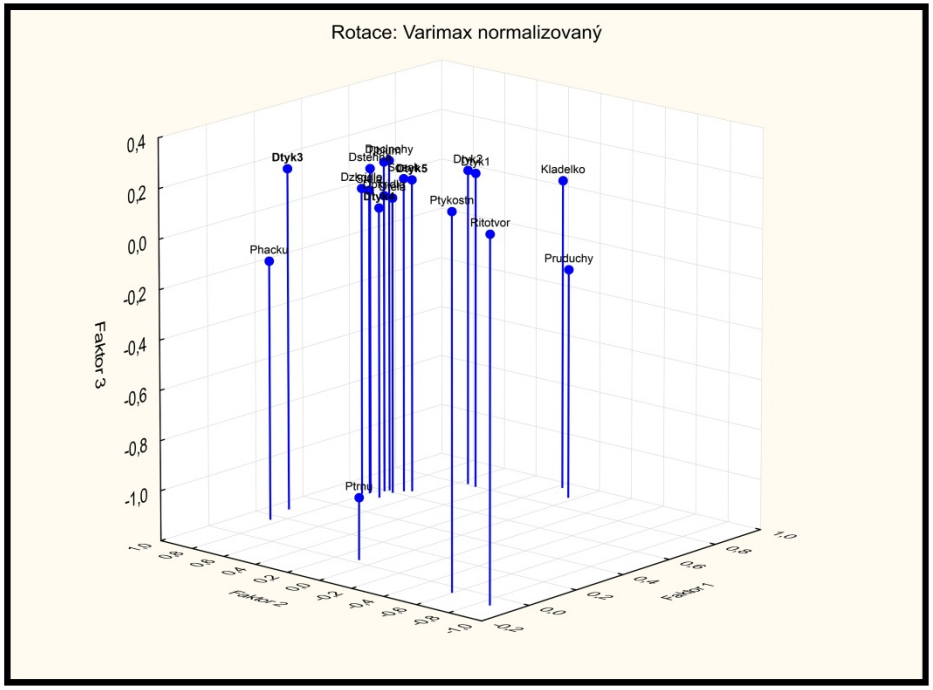

FAKTOR2
FAKTOR1
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
21
24
29
22
23
28
16
30
25
26
32
33
37
17
31
20
18
39
36
35
38
19
40
27
4
34
2
5
7
1
9
6
11
3
12
8
14
15
13
10

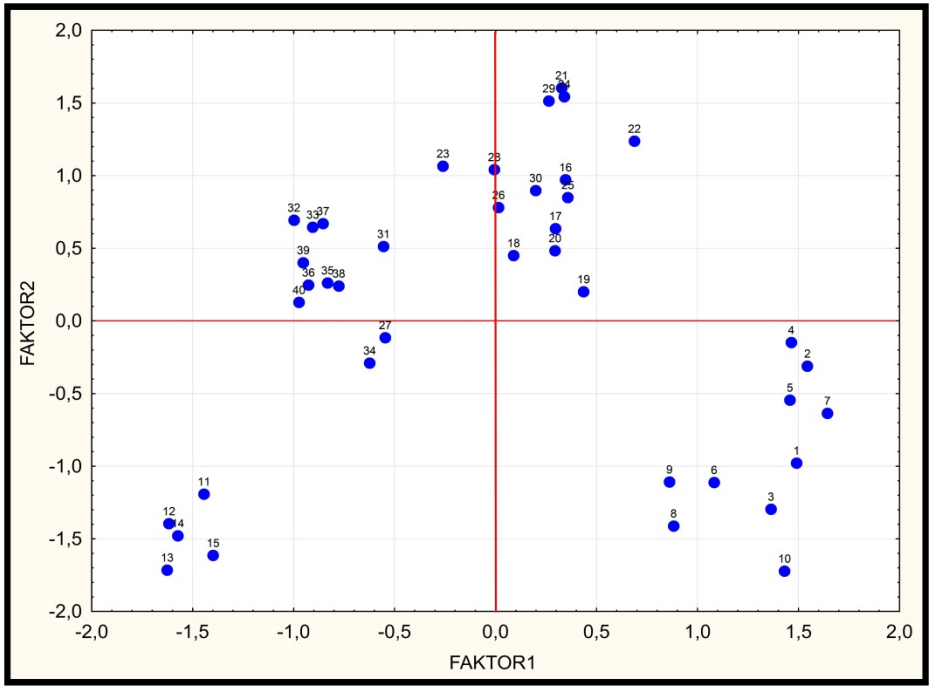

FAKTOR2
FAKTOR1
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-3,0
-2,5
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
10
3
1
7
8
6
5
2
9
15
13
14
12
11
19
25
22
30
26
20
18
17
29
24
16
34
27
23
28
31
36
38
40
35
39
37
33
32

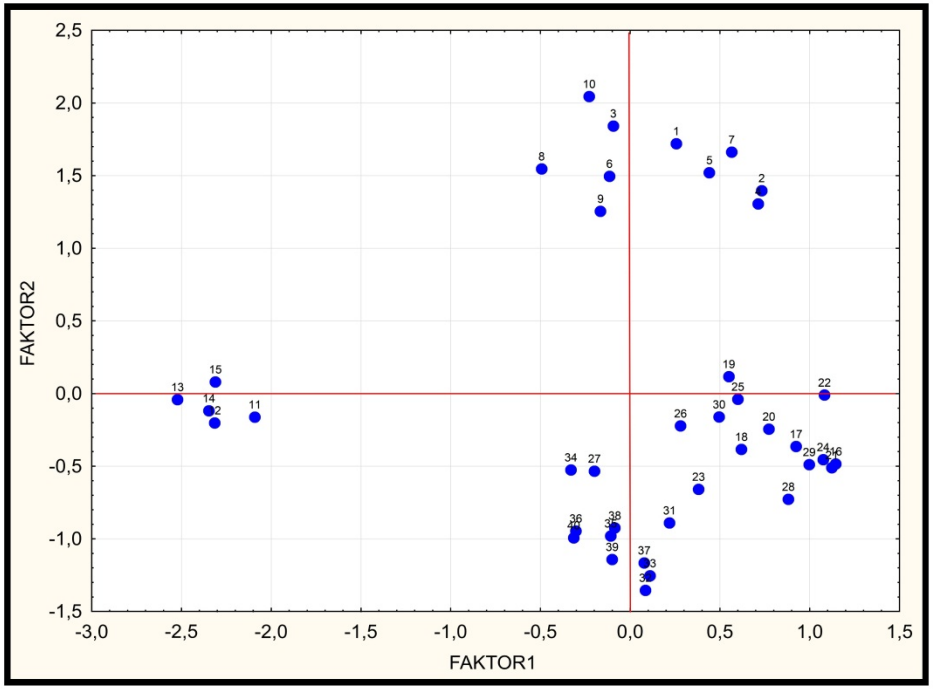

Rotace: Bez rotace
Faktor 2
Faktor 1
Ritotvor
Ptykostn
Kladelko
Pruduchy
Dtyk1
Dtyk2
Dtyk5
Ptrnu
Dtyk3
Phacku

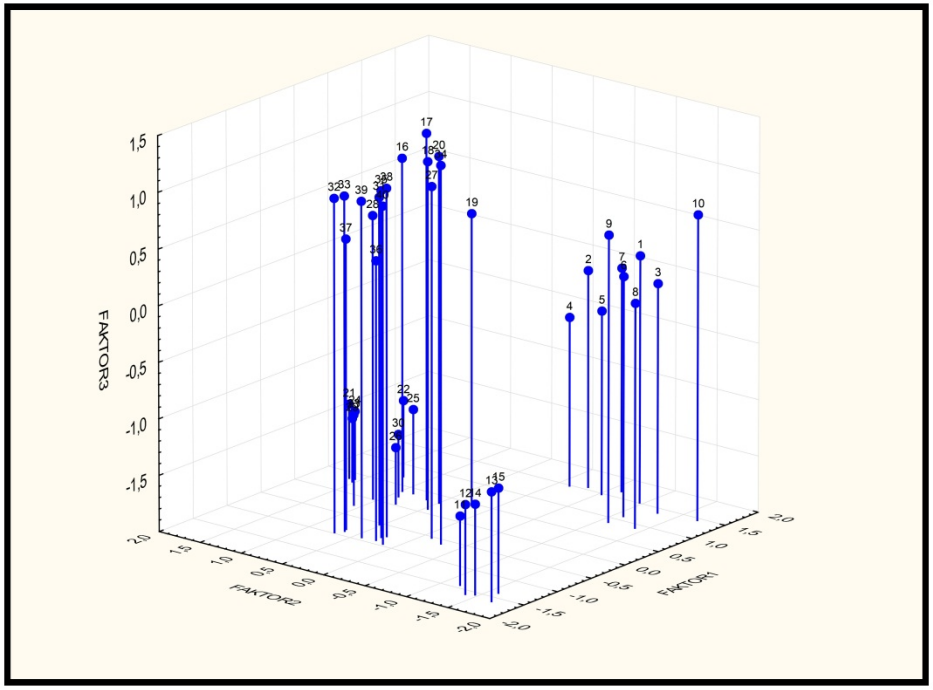

Rotace: Bez rotace
Faktor 3
Faktor 2
Faktor 1
Ptrnu
Pruduchy
Kladelko
Phacku
Dtyk3
Ritotvor
Ptykostn

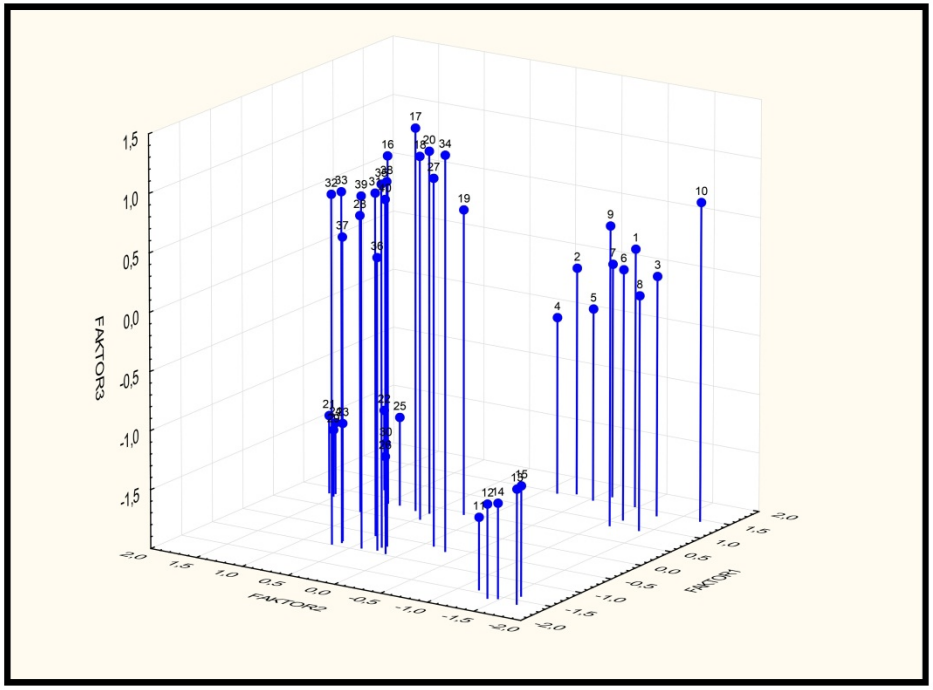

Rotace: Varimax normalizovaný
Faktor 2
Faktor 1
Phacku
Dtyk3
Dtyk4
Dtyk5
Dtyk2
Kladelko
Ptrnu
Pruduchy
Ptykostn
Ritotvor

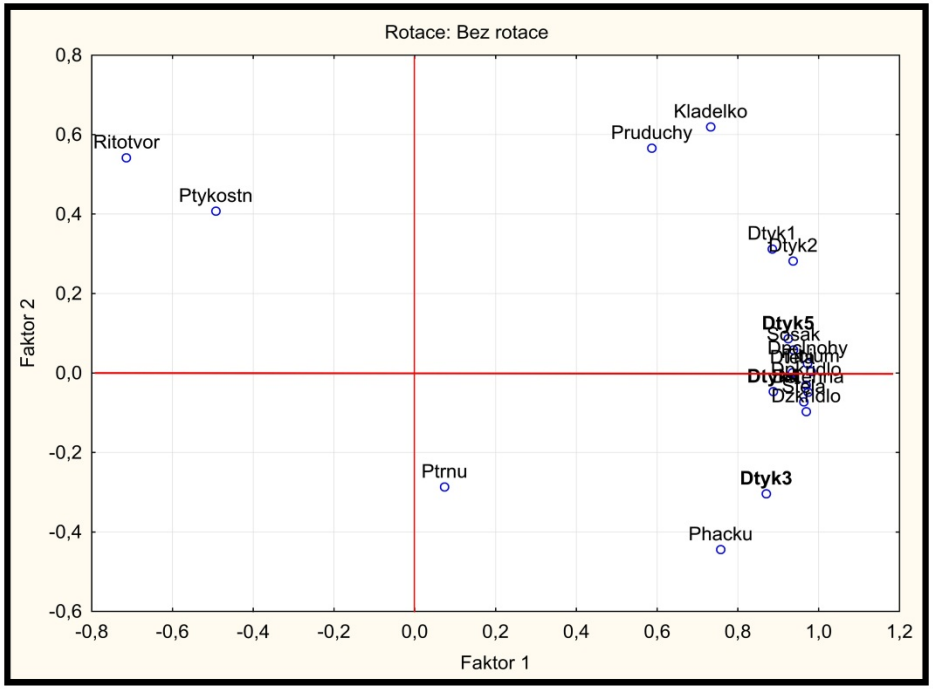

Rotace: Varimax normalizovaný
Faktor 3
Faktor 2
Faktor 1
Dtyk3
Dtyk5
Dtyk2
Dtyk1
Kladelko
Ptykostn
Ritotvor
Phacku
Pruduchy
Ptrnu

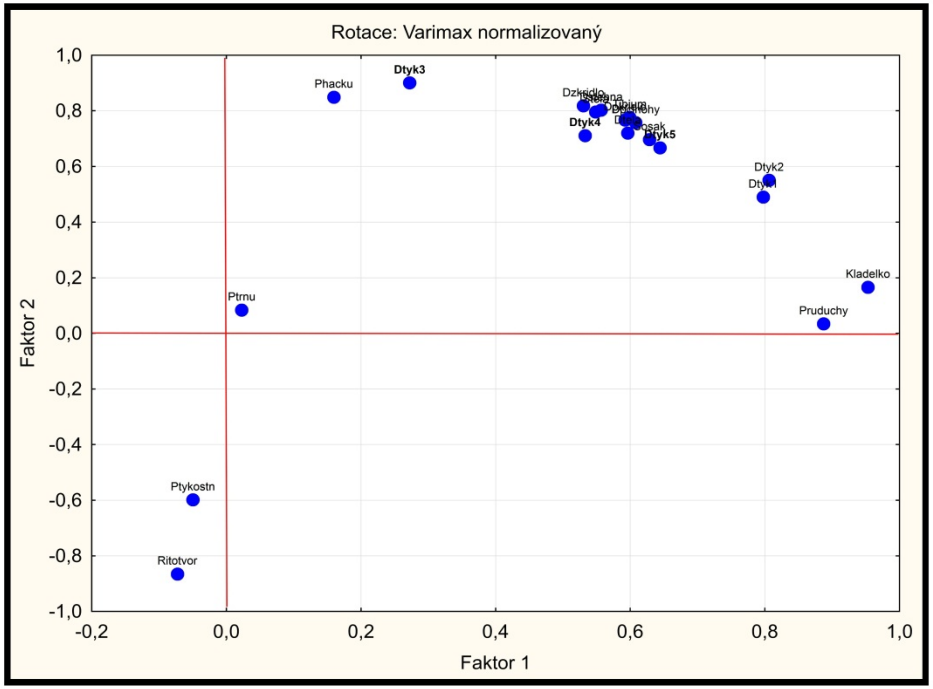

FAKTOR2
FAKTOR1

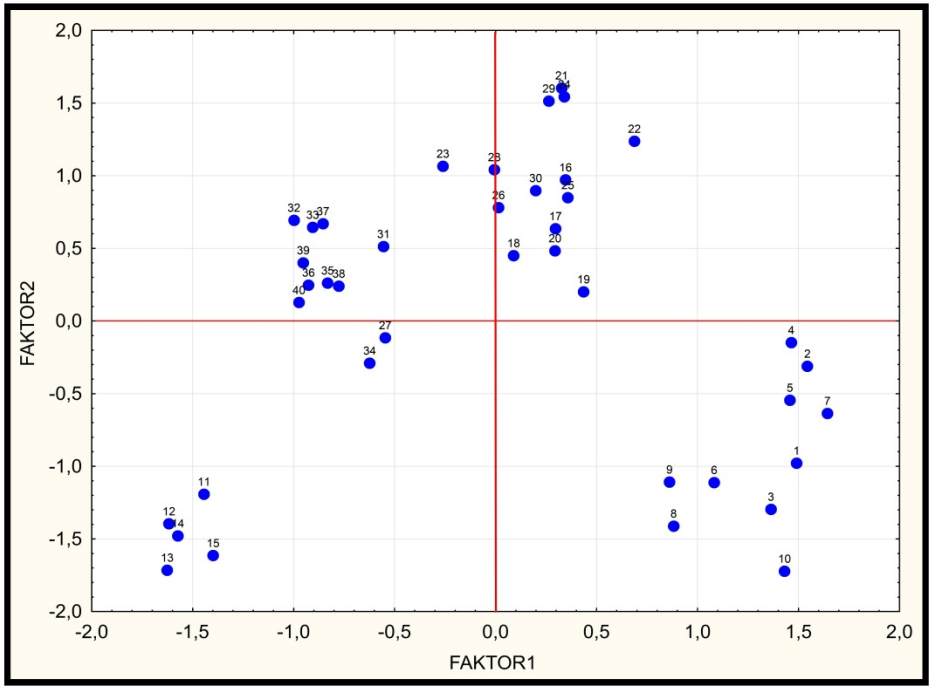

FAKTOR2
FAKTOR1

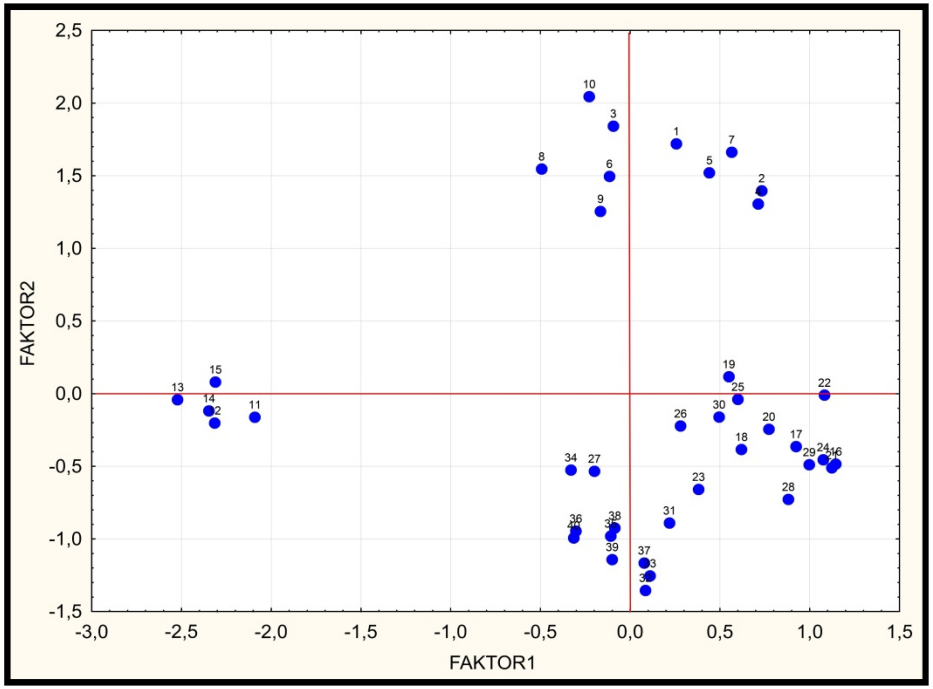

FAKTOR3
FAKTOR2
FAKTOR1

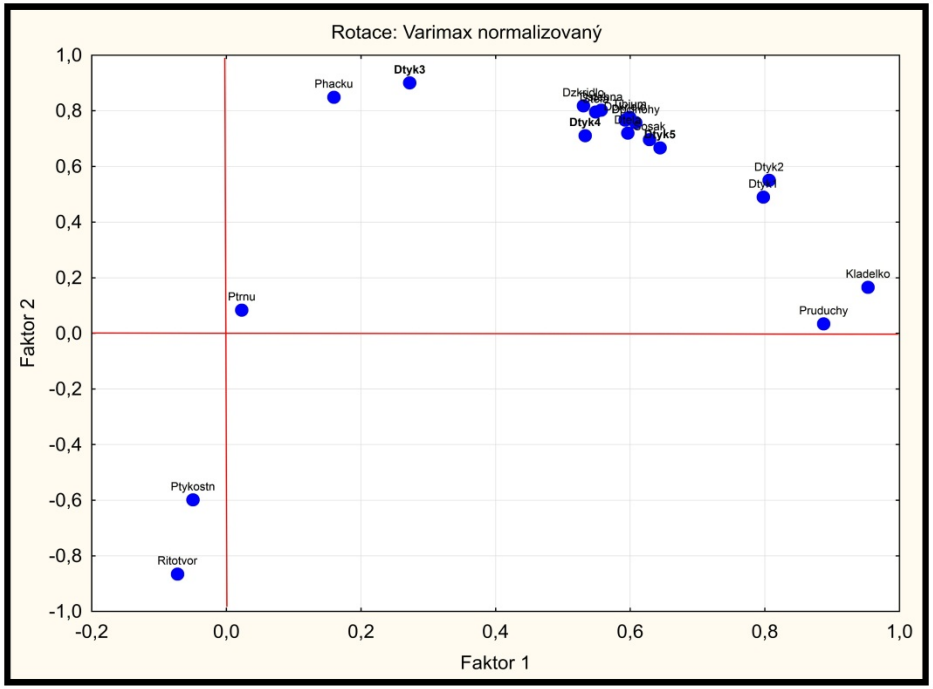

FAKTOR3
FAKTOR2
FAKTOR1

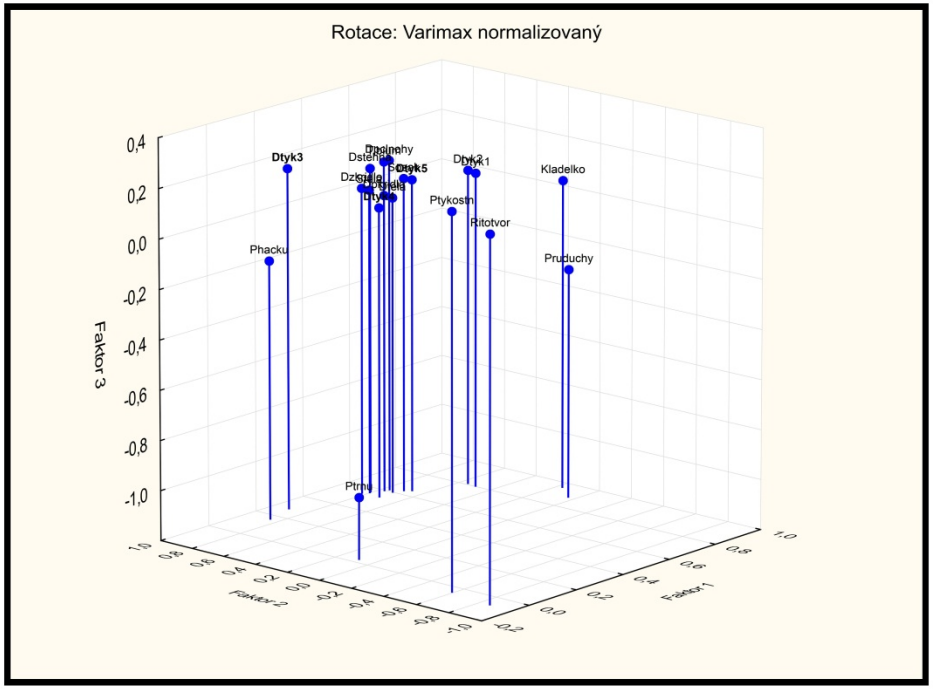

Rotace: Bez rotace
Faktor 2
Faktor 1
Ritotvor
Ptykostn
Kladelko
Pruduchy
Dtyk1
Dtyk2
Dtyk5
Ptrnu
Dtyk3
Phacku

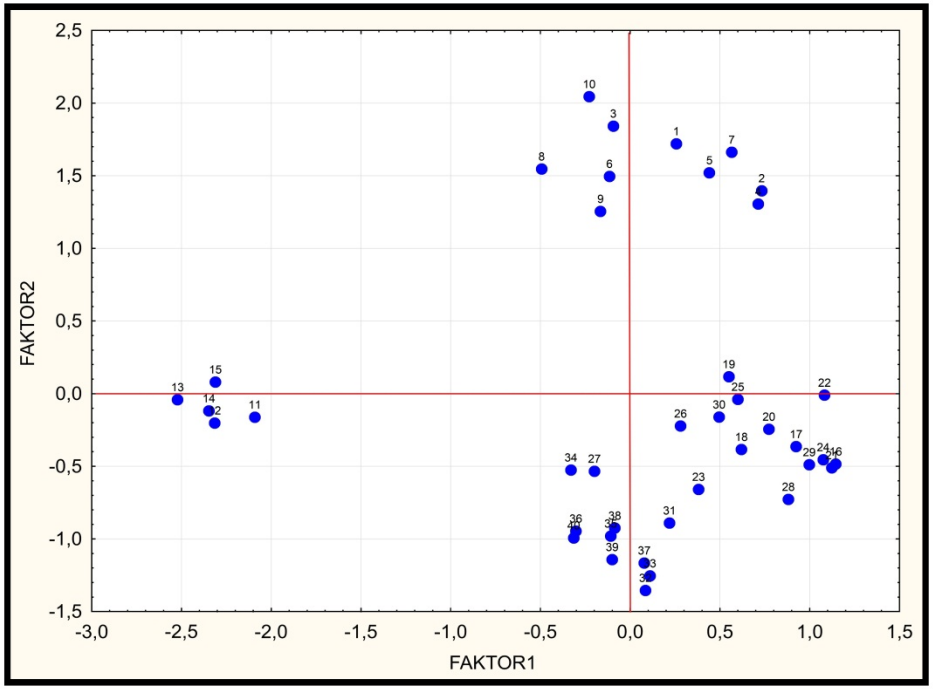

Rotace: Varimax normalizovaný
Faktor 2
Faktor 1
Phacku
Dtyk3
Dtyk4
Dtyk5
Dtyk2
Kladelko
Ptrnu
Pruduchy
Ptykostn
Ritotvor

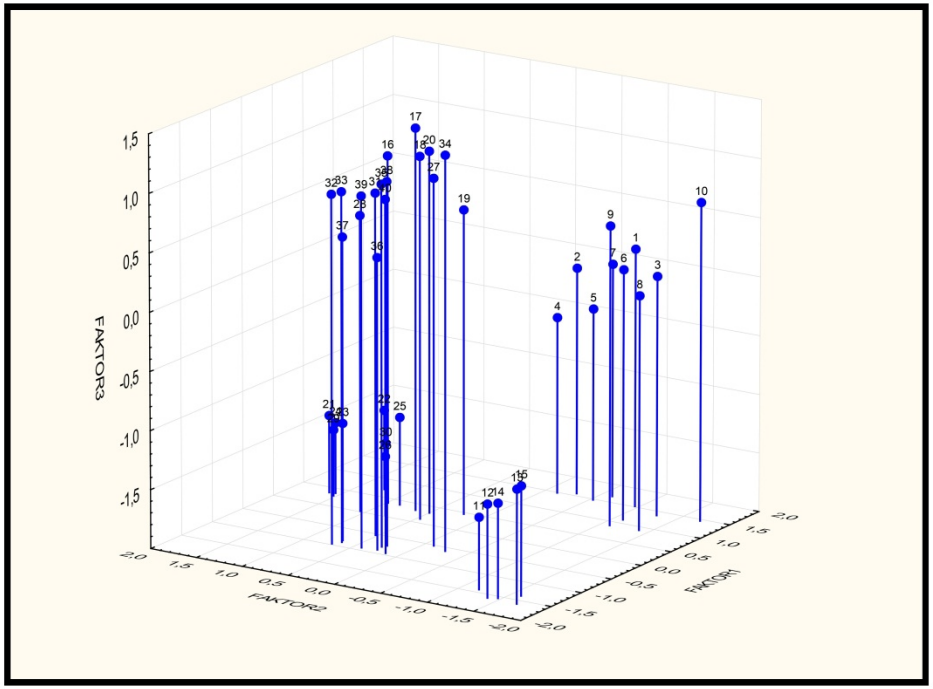

FAKTOR2
FAKTOR1


FAKTOR2
FAKTOR1


Rotace: Varimax normalizovaný
Faktor 2
Faktor 1
Phacku
Dtyk3
Dtyk4
Dtyk5
Dtyk2
Kladelko
Ptrnu
Pruduchy
Ptykostn
Ritotvor


Rotace: Varimax normalizovaný
Faktor 3
Faktor 2
Faktor 1
Dtyk3
Phacku
Kladelko
Pruduchy
Ptykostn
Ritotvor
Ptrnu


FAKTOR2
FAKTOR1


FAKTOR3
FAKTOR2
FAKTOR1

7.2 Logistická diskriminace k určení rakoviny prostaty

Logistickou diskriminaci budeme demonstrovat na úloze o aplikaci logistické diskriminační analýzy u rakoviny prostaty. Režim léčení je závislý na tom, zda je rakovina rozšířena na lymfatické uzliny. Rozhodující metodou vyšetření je laparotomie. Brownův postup vyšetření pěti diskriminantů u 53 pacientů by měl do jisté míry nahradit právě toto obtížnější laparotomické vyšetření.

Data: *Brown ve své studii použil databázi Prostata: Věk značí věk pacienta, Fosfa je hladina kyselé sérové fosfatázy v Kingových−Armstrongových jednotkách, Rtg je výsledek rentgenového vyšetření dle klíče, že 0 značí nepřítomnost a 1 značí přítomnost tumoru, Rekt vyjadřuje velikost tumoru rektálním vyšetřením dle stupnice, že 0 = malý, 1 = velký, Biopsie přináší závěr patologického bodování z biopsie dle klíče, že 0 je méně vážný a 1 je velmi vážný, je-li u proměnné Laparo výsledek laparotomického vyšetření 0, značí to nepřítomnost, a je-li roven 1, značí to přítomnost nodálního rozšíření rakoviny.*

Věk	**Fosfa**	**Rtg**	**Rekt**	**Biopsie**	**Laparo**
66	0.48	0	0	0	0
68	0.56	0	0	0	0
66	0.50	0	0	0	0
56	0.52	0	0	0	0
58	0.50	0	0	0	0

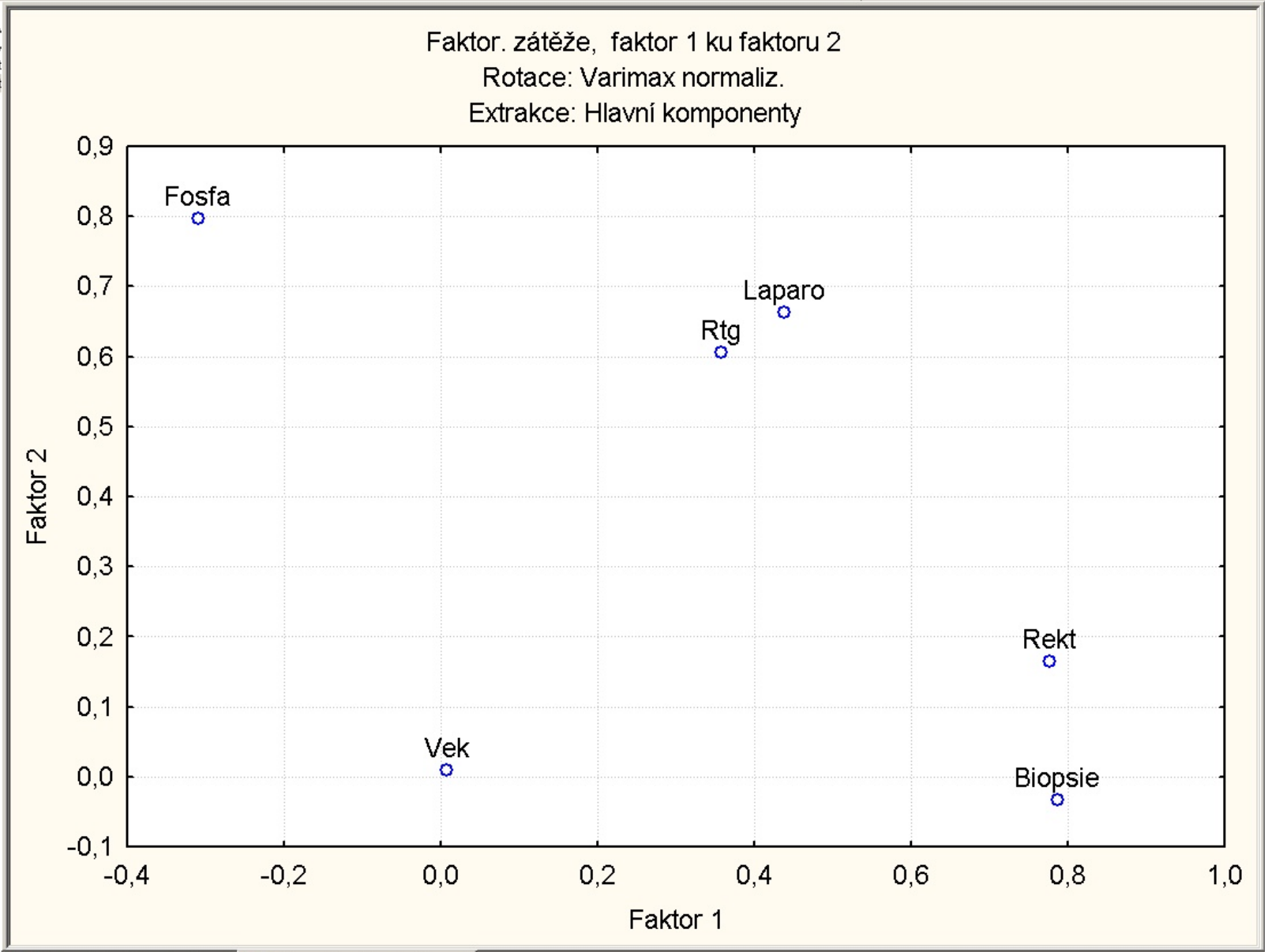
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 2]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (72Rakov
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 2

Výsledky faktorové a...

PS 1\Faktorová analýza (72Rakovina.sta)\Dialog výsledků faktorové analýzy\Faktor. zátěže, faktor 1 ku faktoru 2

Připravena 72Rakovina.st ABC 123 ZÁZN

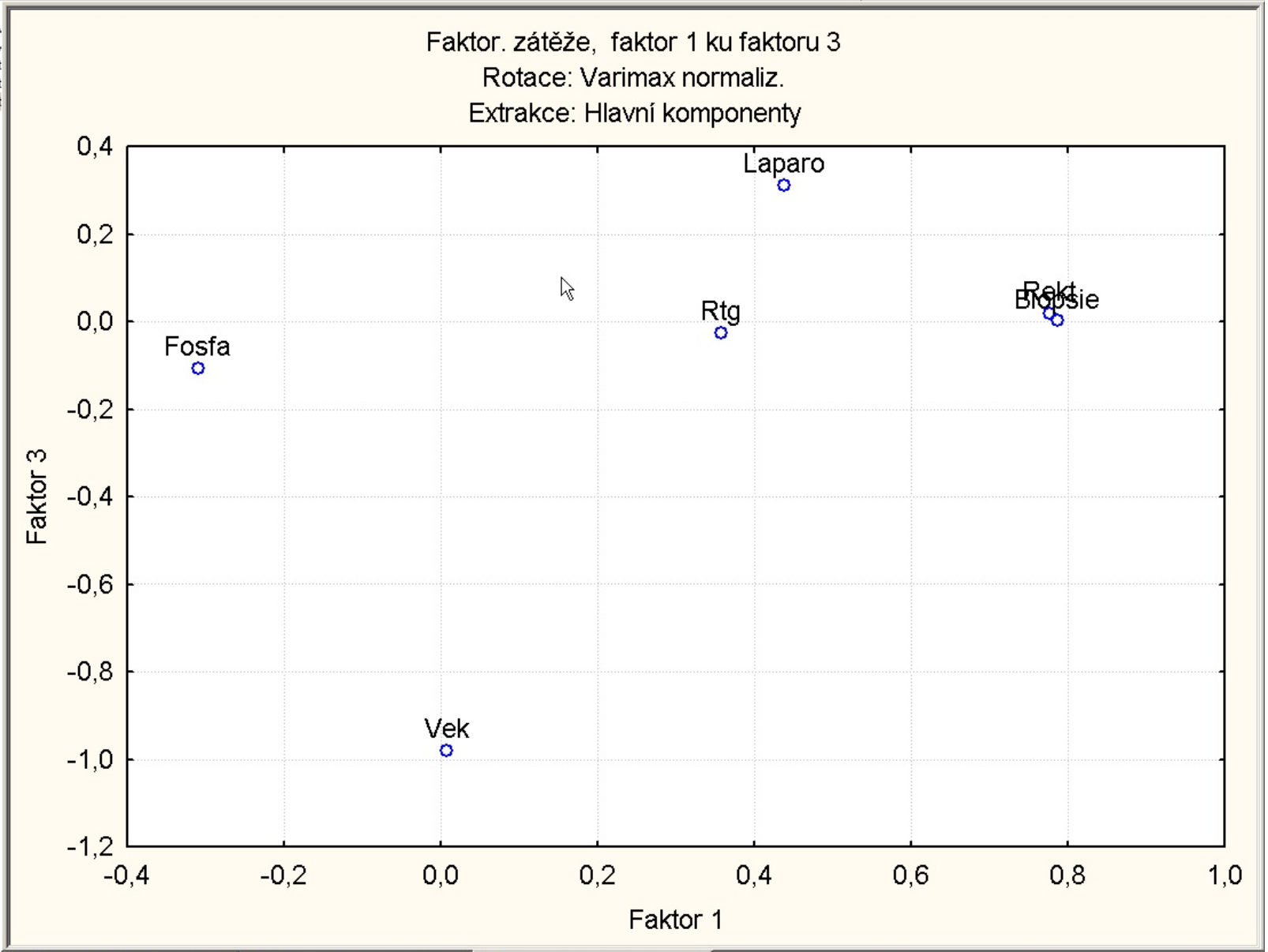
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (72Rakov
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3

Výsledky faktorové a...

Připravena 72Rakovina.sl ABC 123 ZÁZN

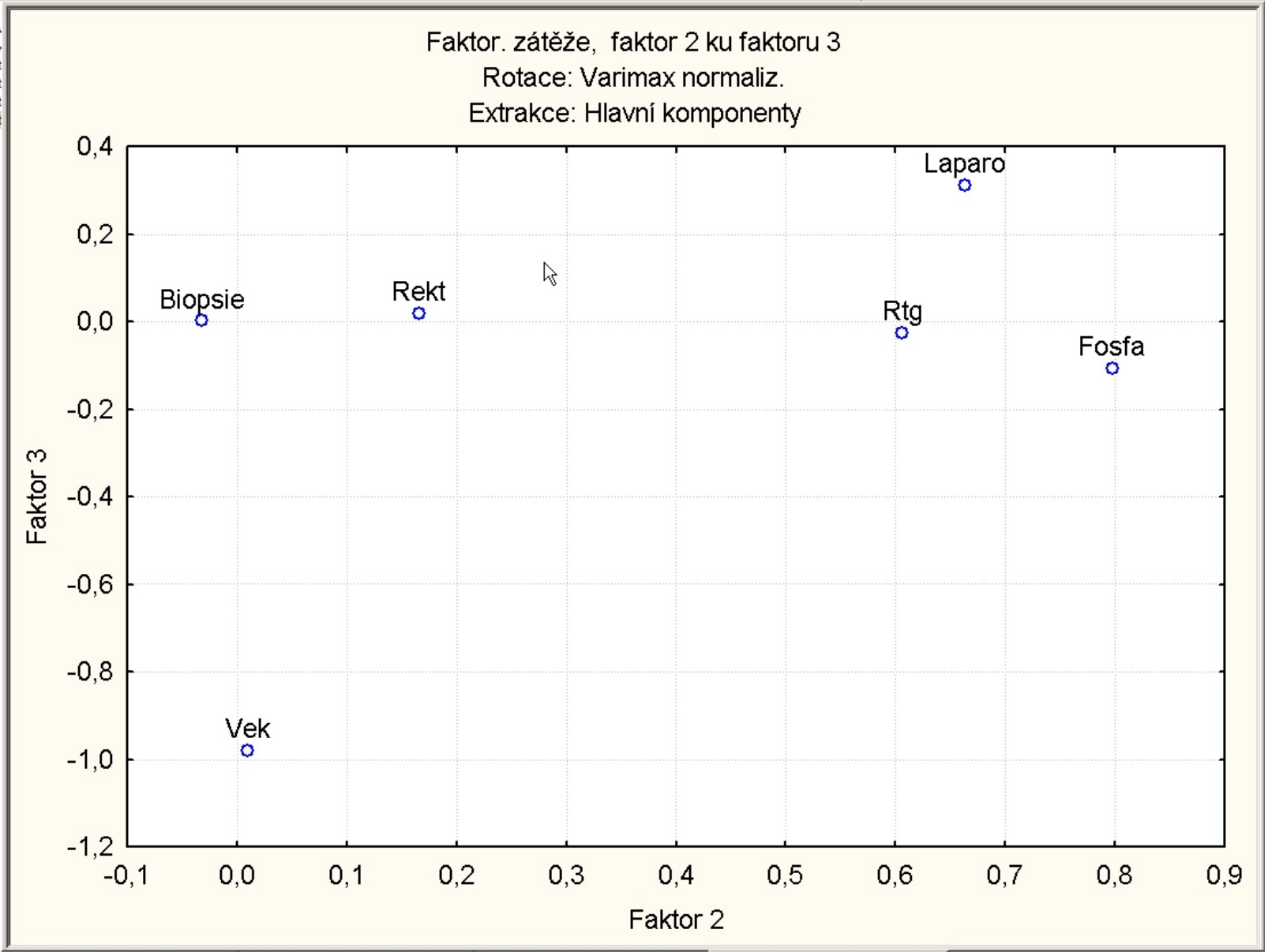
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 2 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (72Rakov
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3 | Faktor. zátěže, faktor 2 ku faktoru 3

Výsledky faktorové a...

Připravena 72Rakovina.sl ABC 123 ZÁZN

PŘÍKLAD 8.2 *Významnost sledovaných znaků ovlivňujících ústup leukemie*

Lee (1980) publikoval data o leukemii pacientů a o ovlivnění jejího ústupu. Cílem příkladu je nalézt znaky čili nezávisle proměnné, které jsou v navrženém logistickém regresním modelu statisticky významné při ovlivnění ústupu leukemie.

Data: *Počet pacientů v řádcích n = 27 a počet znaků ve sloupcích m = 7 souboru Leukemie. Závisle proměnnou je binární znak REMISS o tom, zda se objeví (1) či neobjeví (0) ústup leukemie. Nezávisle proměnnými jsou u pacientů naměřené hodnoty těchto znaků:*

CELL značí celuliritu, buněčnost sraženiny kostní dřeně,
SMEAR značí skvrnu diferenčního procenta napadení,
INFIL značí procento infiltrátu kostní dřeně buňkou leukemie,
LI je procento označeného indexu leukemických buněk kostní dřeně,
BLAST je absolutní počet napadení v periferní krvi,
TEMP značí nejvyšší naměřenou teplotu před začátkem léčby.

Index	*REMISS*	*CELL*	*SMEAR*	*INFIL*	*LI*	*BLAST*	*TEMP*
1	1	0.8	0.83	0.66	1.9	1.1	0.996
..	...	...	...	...	...	...	...
27	0	1	0.73	0.73	0.7	0.398	0.986

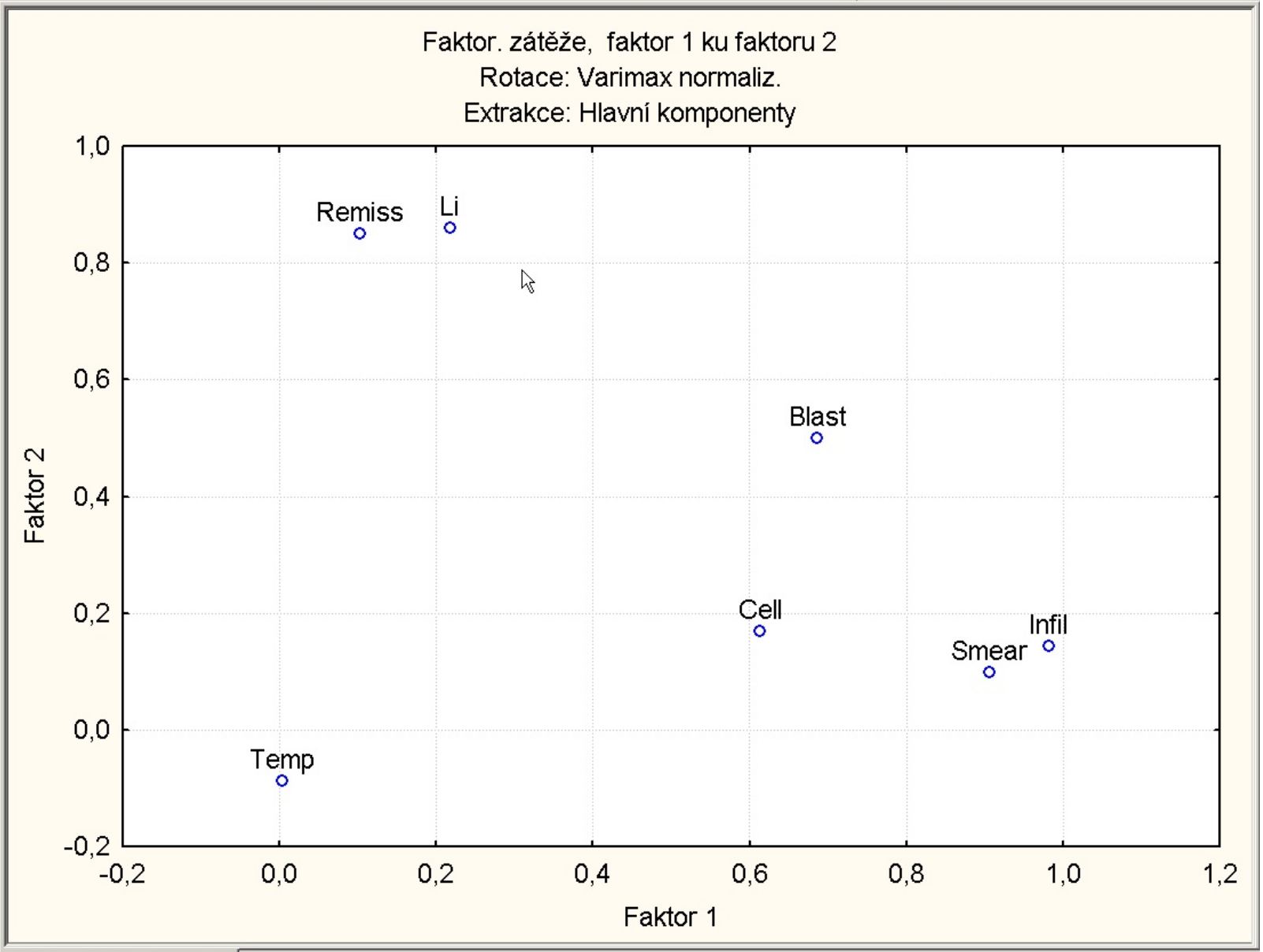
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 2]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (82Leuke
Dialog výsledků faktorov
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2

Výsledky faktorové a...

Připravena

82Leukemie.sl ABC 123 ZÁZN

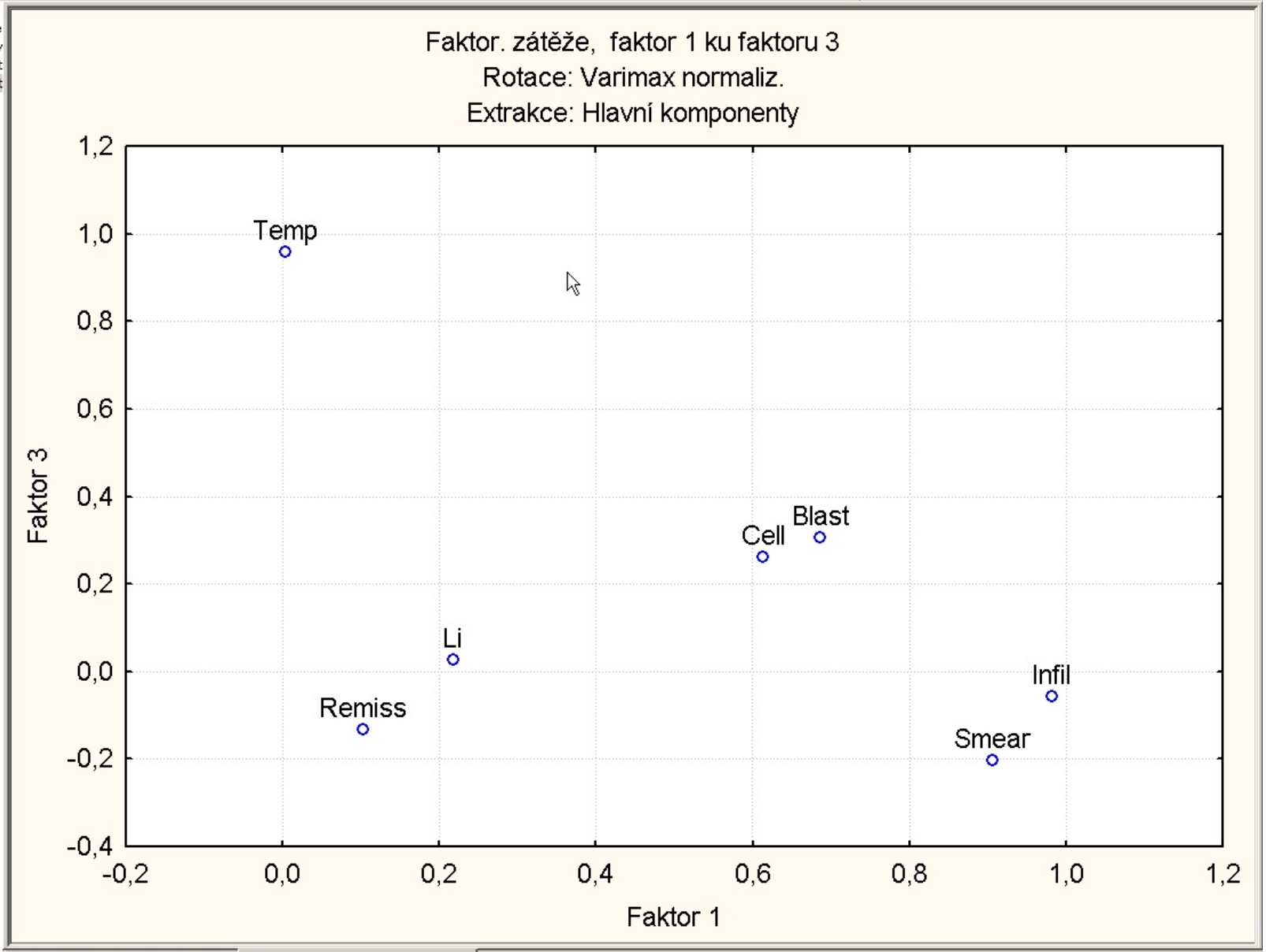
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (82Leuke
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3

Výsledky faktorové a...

Připravena 82Leukemie.sl ABC 123 ZÁZN

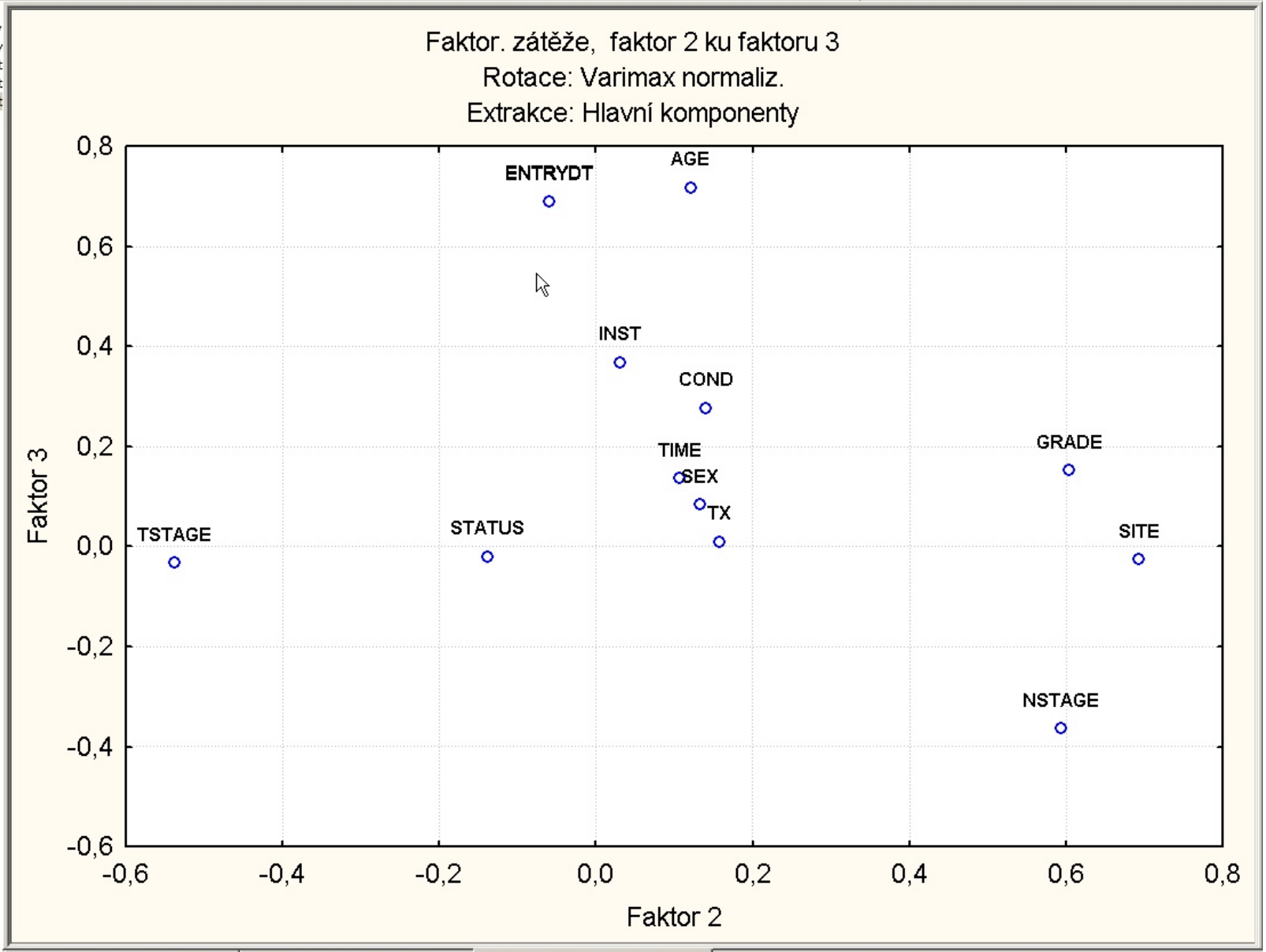
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 2 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
- Faktorová analýza (84Phyry
  - Dialog výsledků faktorov
    - Faktor. zátěže, fakt
    - Faktor. zátěže, fakt
    - Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3 | Faktor. zátěže, faktor 2 ku faktoru 3

Výsledky faktorové a...

Pro nápovědu stiskněte F1

84Phyrynx.sta ABC 123 ZÁZN

8.5 *Analýza významnosti znaků intenzivní péče pacientů na JIP*

Bylo sledováno 200 pacientů na jednotce intenzivní péče − JIP, anglicky značené ICU. Cílem tohoto výzkumu bylo postavit logistický regresní model k predikci pravděpodobnosti přežití pacienta a vyšetřit statistickou významnost vybraných rizikových faktorů úmrtnosti. Cílem je soustředit se na tyto úkoly:

1. Odhadnout směrnice parametrů logistického regresního modelu technikou maximální věrohodnosti dostupným softwarem a komentovat jejich statistickou věrohodnost porovnáním výsledků testů maximální věrohodnosti a Waldova testu. Vedou vyčíslené hladiny významnosti *P ke stejným závěrům?*

2. Je třeba vyčíslit 95% interval spolehlivosti odhadnutých směrnic parametrů logistického regresního modelu a rovněž absolutního členu. Slovně je třeba vysvětlit směrnici každého parametru samostatně.

Data: *Zdrojová matice ICU obsahuje 20 znaků pro 200 pacientů na jednotce intenzivní péče JIP: ID je identifikační kód pacienta. STA je kód přežití: 0 = přežije, 1 = nepřežije. AGE vyjadřuje stáří, věk vyjádřený v letech. RACE značí rasu: 1 = bílá, 2 = černá, 3 = ostatní. SER značí obsluhu na jednotce intezivní péče JIP: 0 = běžná lékařská, 1 = operativní chirurgická. CAN udává, zda rakovina v části těla je akutní problém tohoto pacienta: 0 = není, 1 = ano je. CRN značí existenci chronické ledvinové poruchy: 0 = není, 1 = ano je. INF udává, zda pacient dostal infekci až po vstupu na jednotku intenzivní péče JIP: 0 = ne, 1 = ano. CPR značí že CPR předcházelo přijetí na jednotku intenzivní péče JIP: 0 = ne, 1 = ano. SYS udává systolický krevní tlak při přijetí na jednotku intenzivní péče JIP v mm Hg. HRA vyjadřuje krevní puls při přijetí na jednotku intenzivní péče JIP v pulsech/min. PRE značí předešlá přijetí na jednotku intenzivní péče JIP v poslední 6 měsících: 0 = ne, 1 = ano. TYP značí způsob přijetí na jednotku intenzivní péče JIP: 0 = dle volby, 1 = v mimořádné nouzi, pohotovostní. FRA značí důvod přijetí na jednotku intenzivní péče JIP jako dlouhá kost, vícenásobná, šíje, jednoduchá plocha, zlomenina kyčle: 0 = ne, 1 = ano. PO2 představuje uvolněný kyslík z krve: 0 pro PO2 ≥ 60 a 1 pro PO2 < 60. PH udává pH krve: 0 pro PH ≥ 7.25 a 1 pro PF < 7.25. PCO značí CO_2 v krvi: 0 pro PCO ≤ 45 a 1 pro PCO > 45. BIC vyjadřuje diuhličitan v krvi: 0 pro BIC ≤ 18 a 1 pro BIC > 18. CRE značí creatinin v krvi: 0 pro CRE ≤ 2.0 a 1 pro CRE > 2.0. LOC značí stav vědomí při přijetí na jednotku intenzivní péče JIP: 0 = žádné kóma,1 = hluboké omámení, 2 = kóma.*

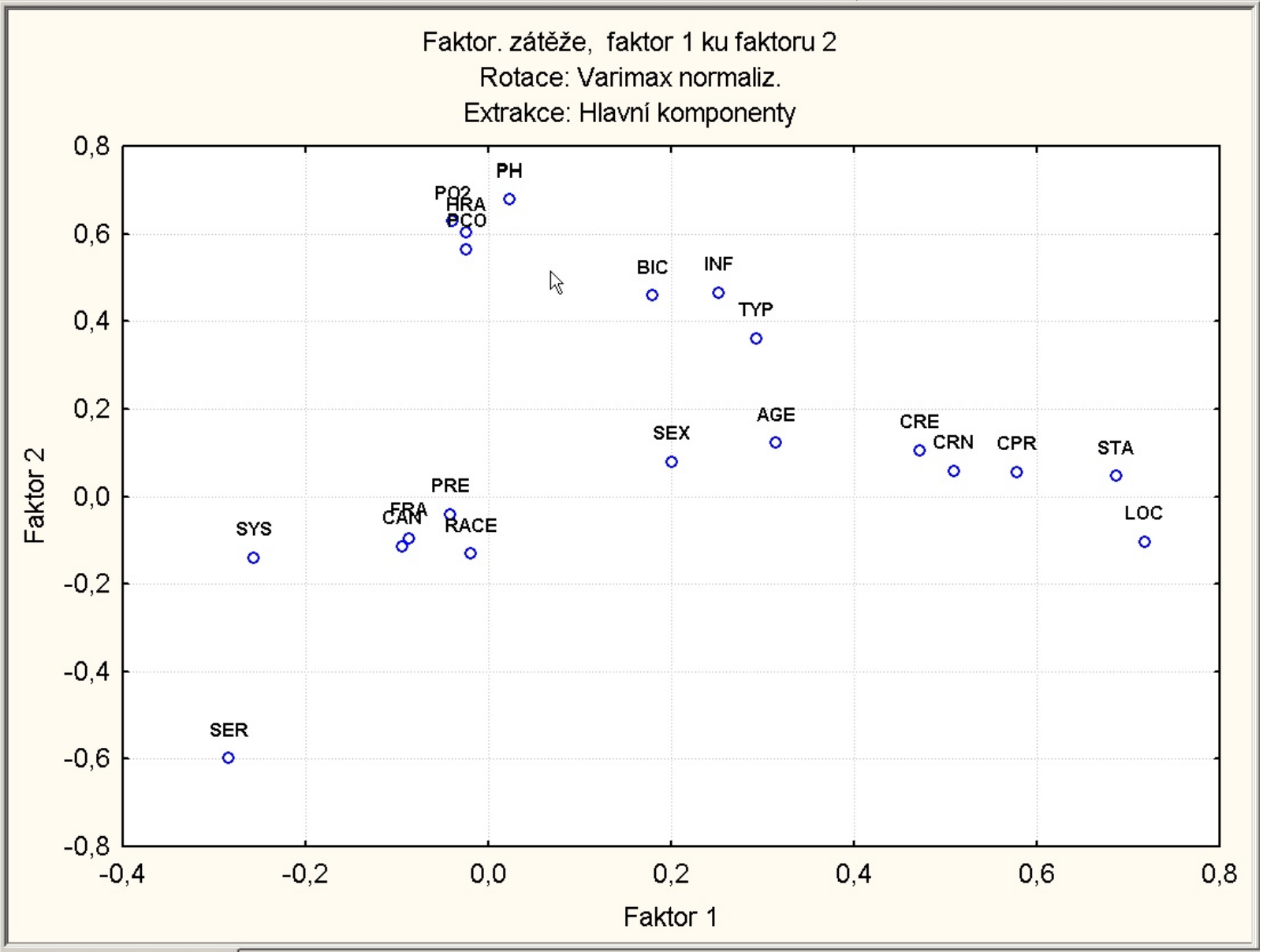
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 2]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (85ICU.st
Dialog výsledků faktorov
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2

Výsledky faktorové a...

Připravena

85ICU.sta ABC 123 ZÁZN

STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (85ICU.st
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt

Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3

Výsledky faktorové a...

Připravena 85ICU.sta ABC 123 ZÁZN

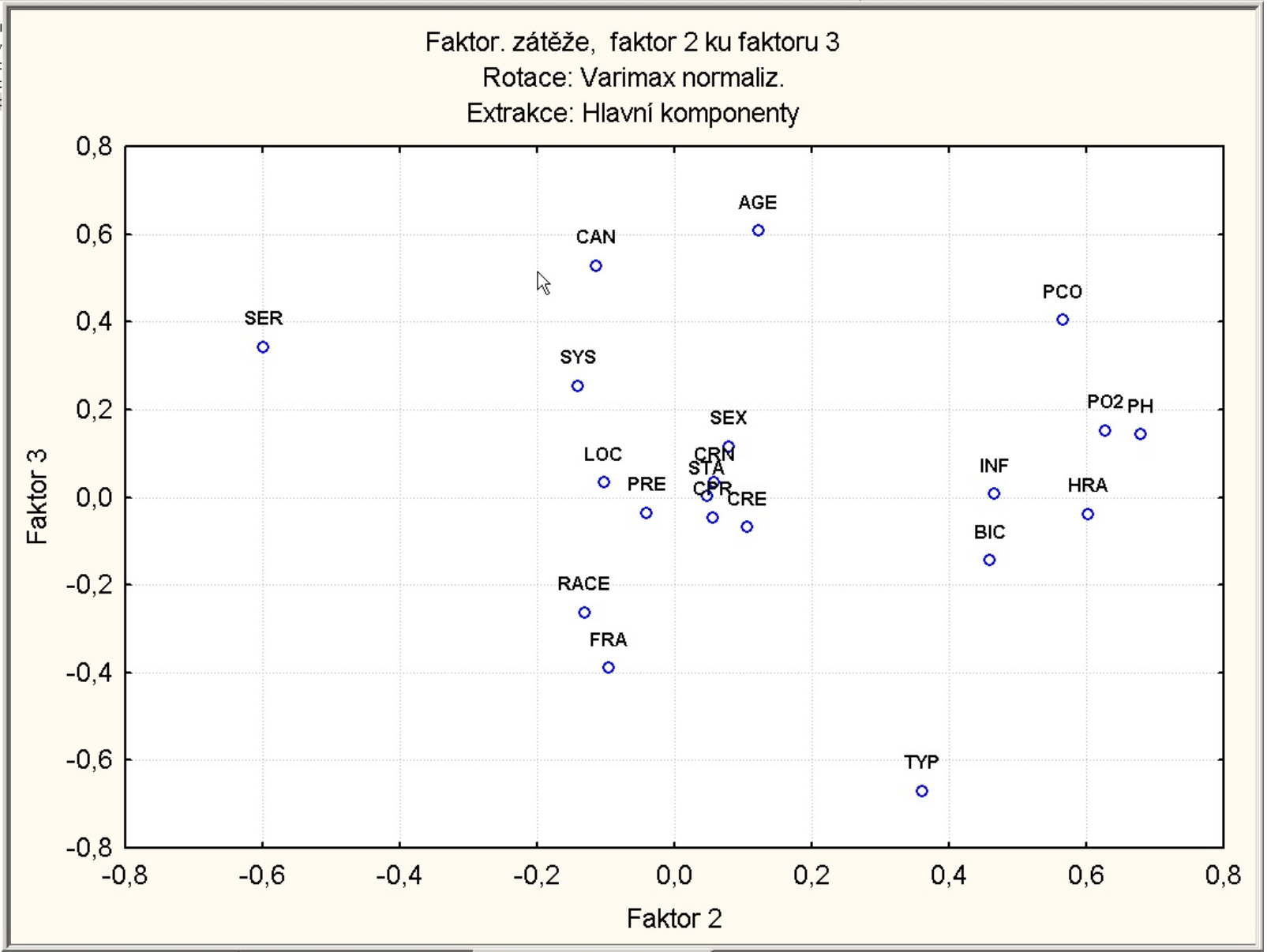
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 2 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (85ICU.st
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3 | Faktor. zátěže, faktor 2 ku faktoru 3

Výsledky faktorové a...

Připravena 85ICU.sta ABC 123 ZÁZN

PŘÍKLAD 8.6 *Odhalení rizikových faktorů nízké porodní hmotnosti*

Cílem úlohy je odhalit rizikové znaky, které jsou spojeny s nízkou porodní hmotností novorozenců, tj hmotností menší než 2500 gramů. Příčiny nízké porodní hmotnosti zajímají lékaře dlouhodobě, protože s ní souvisí i dětská úmrtnost a časté defekty novorozenců. Mezi závažné znaky ovlivňující porodní hmotnost patří i zde sledované vlastnosti a návyky, jako je u matek nevhodná dieta, kouření a nedostatečná prenatální péče v průběhu těhotenství. Je třeba vyšetřit, které z těchto znaků jsou statisticky významné a závažně ovlivňující porodní hmotnost.

Data: *Zdrojová matice Lowbwt obsahuje hmotnosti 189 novorozenců, z nichž 59 mělo nízkou porodní hmotnost a 130 jich mělo normální hmotnost. Byly především sledovány 4 znaky, které by mohly ovlivnit porodní hmotnost novorozenců: věk rodičky, hmotnost rodičky v posledním menstručním období, rasa a počet návštěv u gynekologa v prvním trimestru těhotenství. Data obsahují pro 180 novorozenců zapsaných v řádcích tyto znaky ve sloupcích:*

ID značí identifikační číslo.

LOW vyjadřuje porodní hmotnost: 0 = normální nad 2500 g, 1 = nízká pod 2500 g.

AGE udává věk rodičky v letech.

LWT značí hmotnost rodičky v librách v posledním menstruačním období před otěhotněním (1 libra = 0.543 kg).

RACE udává rasu: 1 = bílá, 2 = černá, 3 = ostatní.

SMOKE představuje kouření v průběhu těhotenství: 1 = ano, 2 = ne.

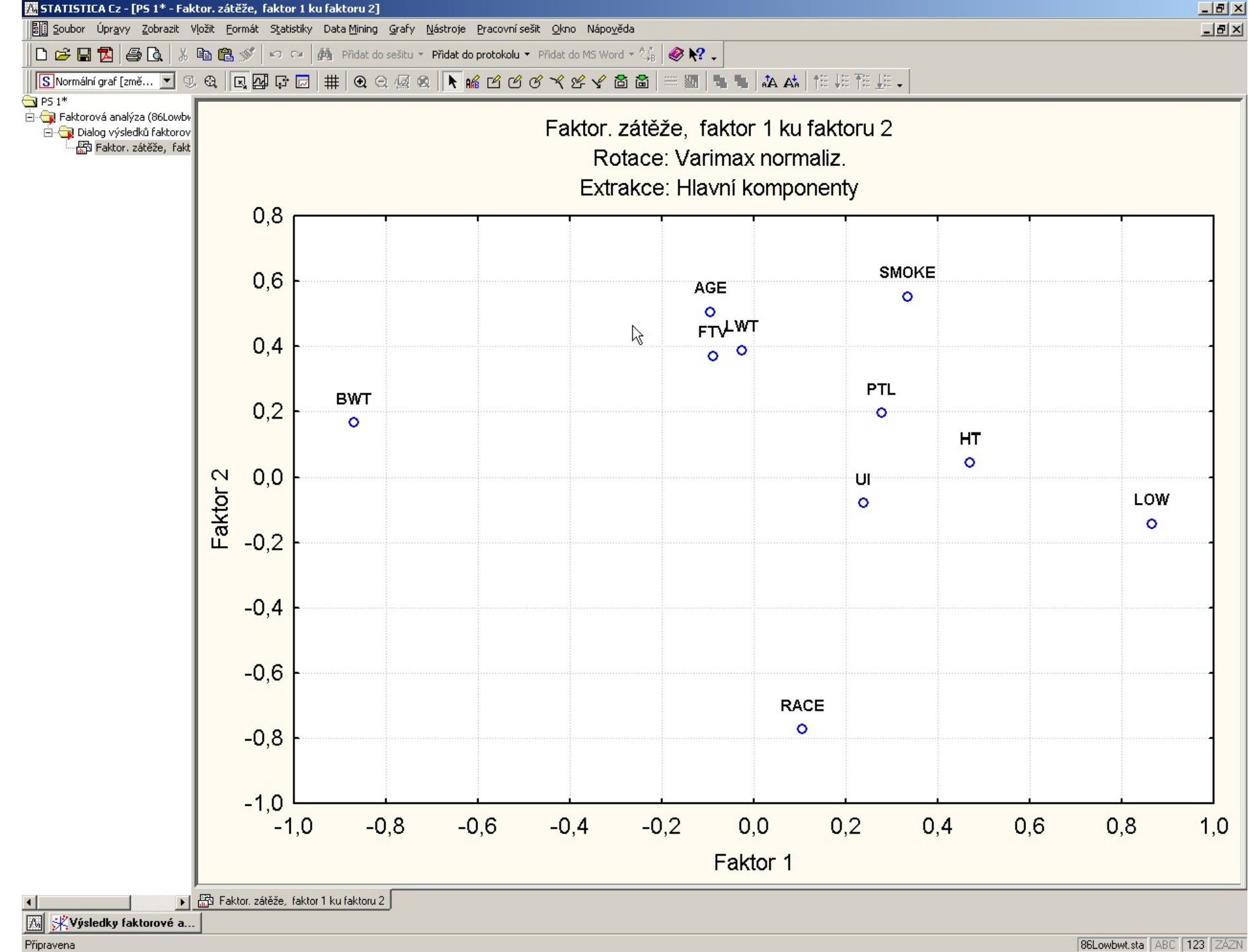
PTL značí měsíce předčasného porodu: 0 = žádný, 1 = jeden, atd.

HT udává hypertenzi v těhotenství: 1 = ano, 0 = ne.

UI značí děložní předrážděnost: 1 = ano, 0 = ne.

FTV udává počet návštěv u gynekologa v prvním trimestru těhotenství: 0 = žádná, 1 = jedna, 2 = dvě atd.

BWT vyjadřuje porodní hmotnost v gramech.


STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 2]
Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Přidat do MS Word
Normální graf [změ...
PS 1*
Faktorová analýza (86Lowbw
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, faktor 1 ku faktoru 2
Rotace: Varimax normaliz.
Extrakce: Hlavní komponenty
SMOKE
AGE
FTV
LWT
PTL
BWT
HT
UI
LOW
RACE
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
Faktor 2
-1,0
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0
Faktor 1
Faktor. zátěže, faktor 1 ku faktoru 2
Výsledky faktorové a...
Připravena
86Lowbwt.sta
ABC
123
ZÁZN

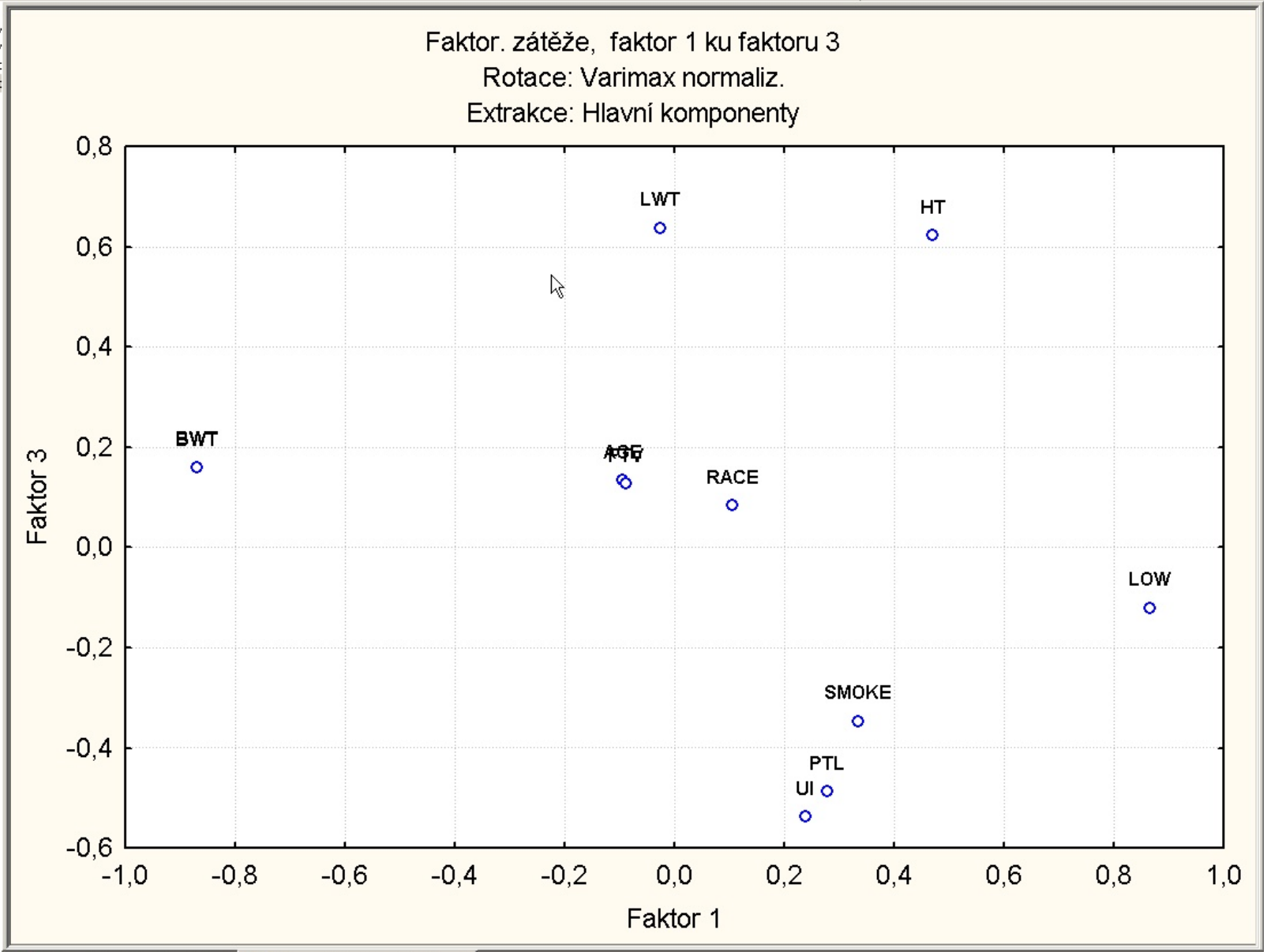
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 1 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (86Lowbw
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3

Výsledky faktorové a...

Připravena 86Lowbwt.sta ABC 123 ZÁZN

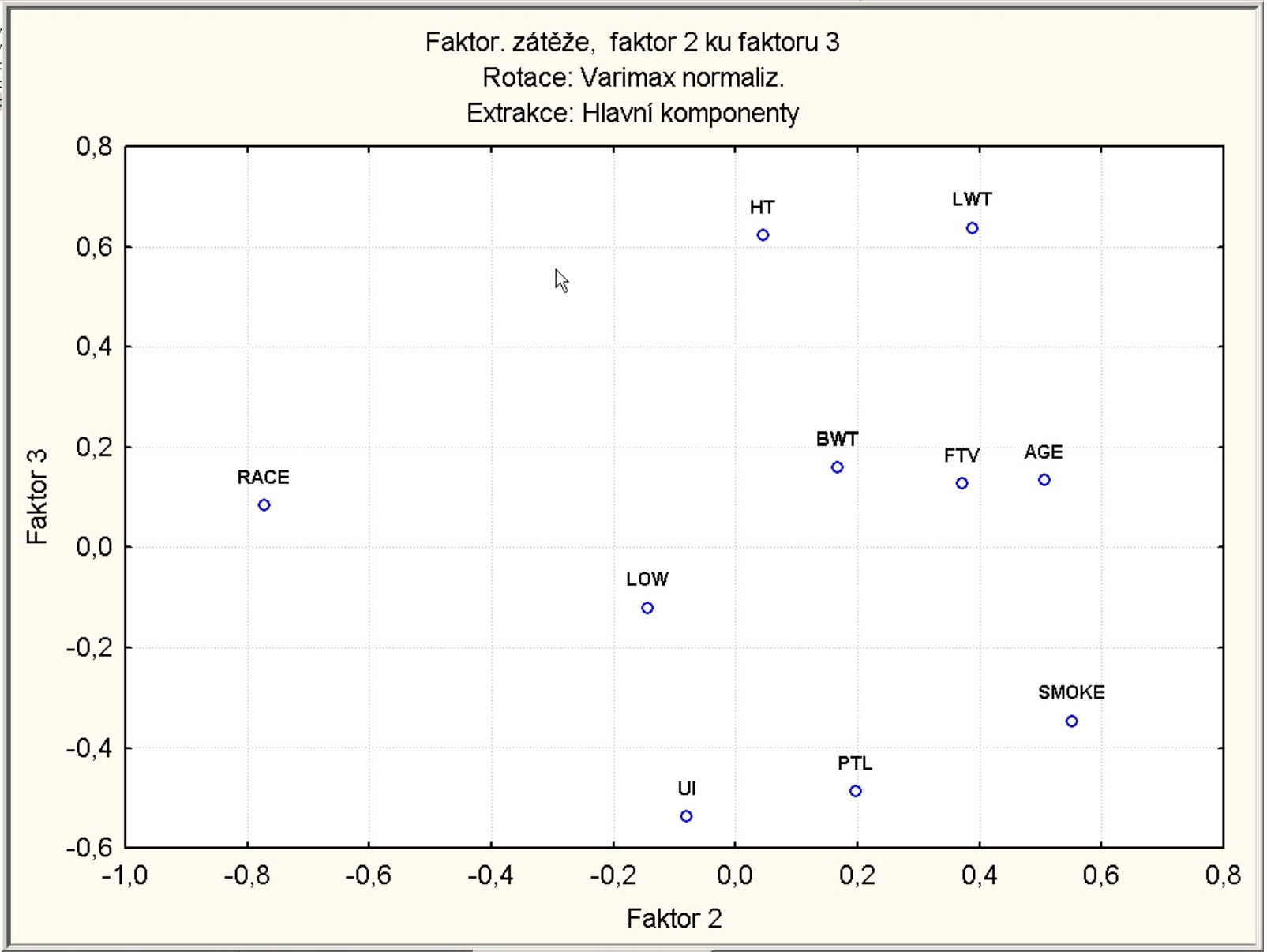
STATISTICA Cz - [PS 1* - Faktor. zátěže, faktor 2 ku faktoru 3]

Soubor Úpravy Zobrazit Vložit Formát Statistiky Data Mining Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

Normální graf [změ...

PS 1*
Faktorová analýza (86Lowbw
Dialog výsledků faktorov
Faktor. zátěže, fakt
Faktor. zátěže, fakt
Faktor. zátěže, fakt



Faktor. zátěže, faktor 1 ku faktoru 2 | Faktor. zátěže, faktor 1 ku faktoru 3 | Faktor. zátěže, faktor 2 ku faktoru 3

Výsledky faktorové a...

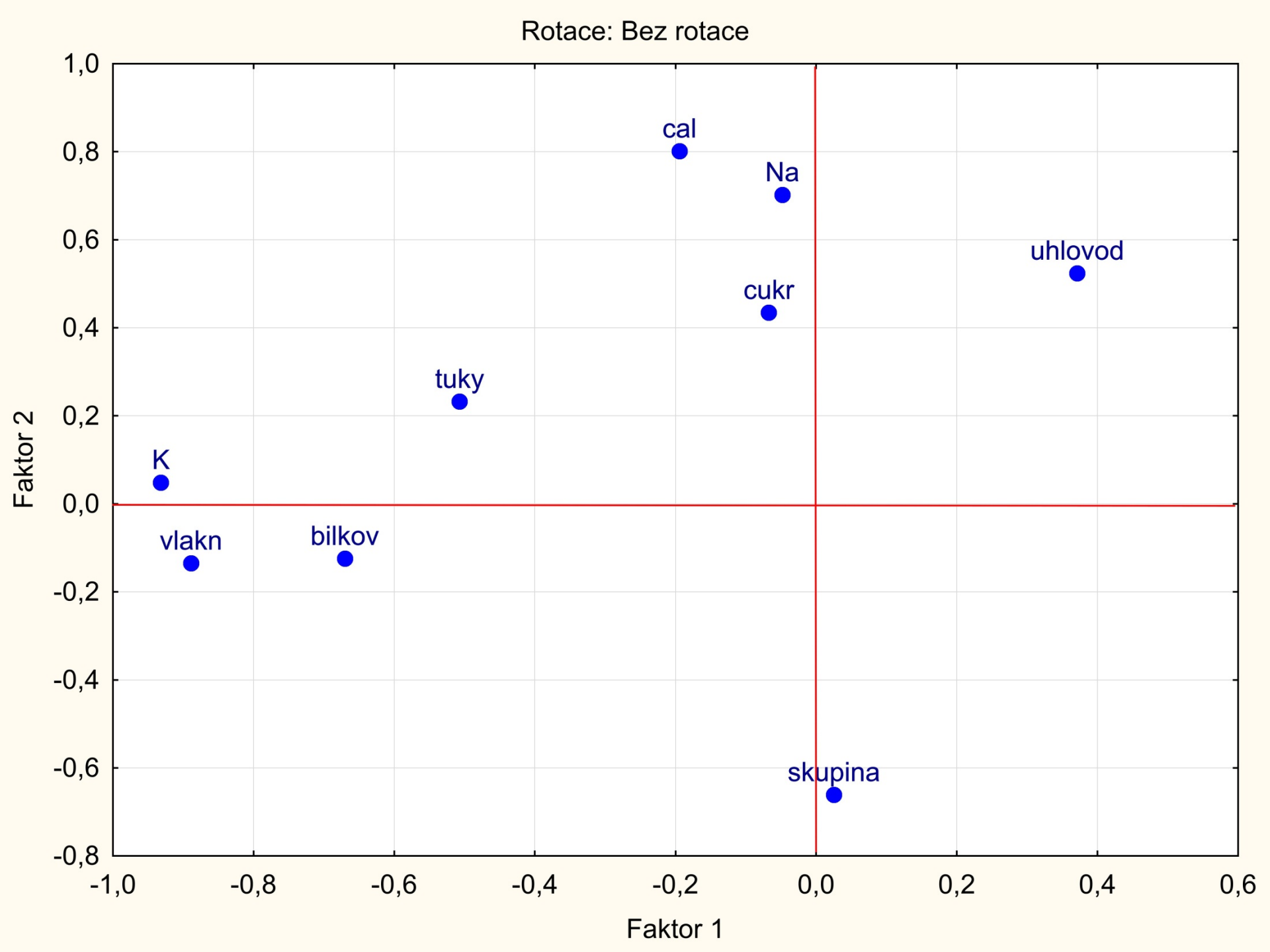
Připravena 86Lowbwt.sta ABC 123 ZÁZN

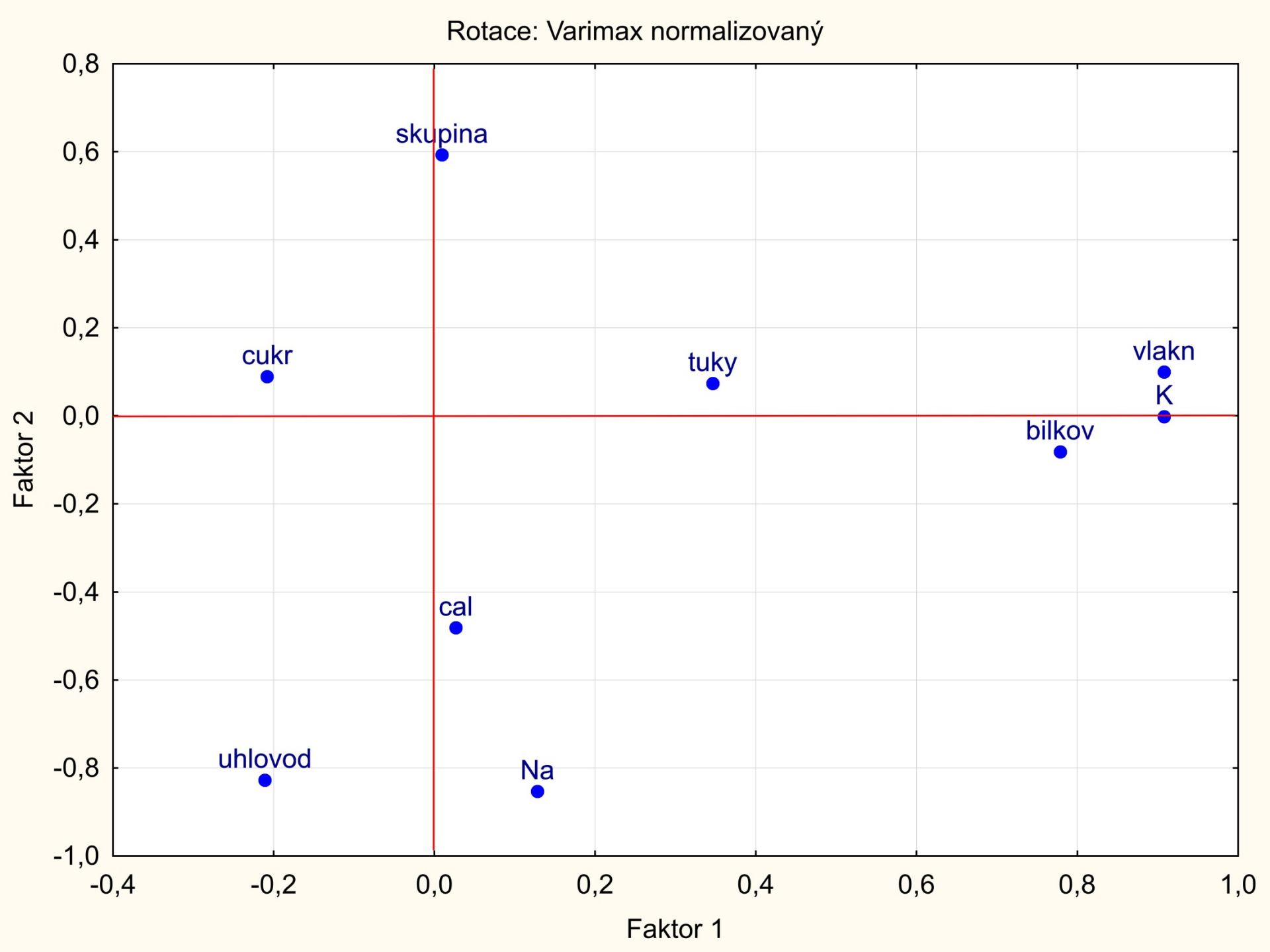
PŘÍKLAD 9.15 *Hledání podobnosti vlastností křupavých lupínků od různých výrobců*

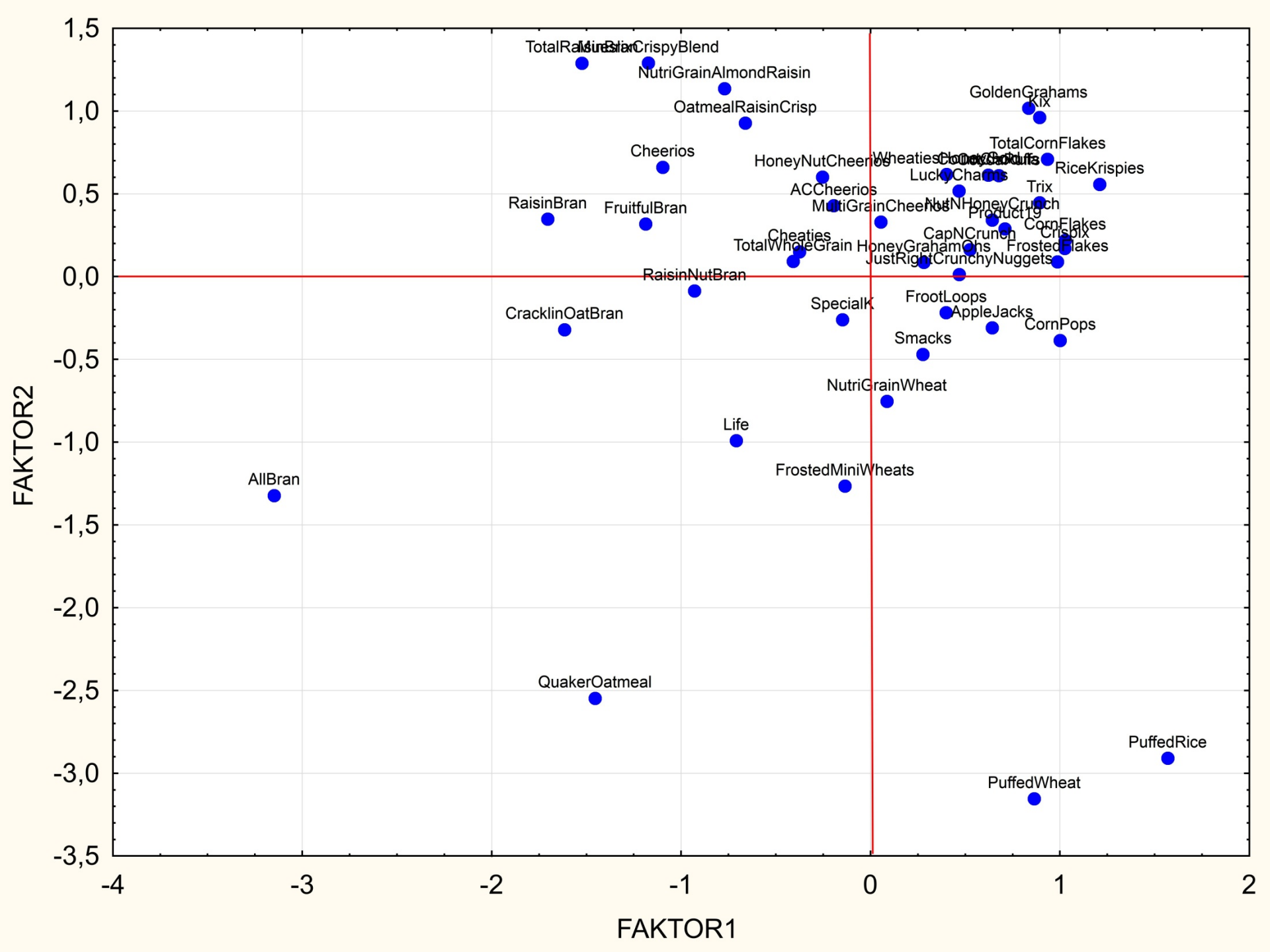
Tři americké firmy General Mills (G), Kellogg (K) a Quaker (Q) produkují křupavé obilné lupínky ke snídani. U řady produktů bylo sledováno 10 znaků a vyšetřována struktura a vzájemné vazby mezi sledovanými znaky jednotlivých produktů, ale i mezi objekty. Které objekty jsou si velice podobné? Kolik shluků objektů odhaluje dendrogram podobnosti objektů optimální shlukovací procedurou? Lze nalézt i dominantní znaky a naopak redundantní znaky, které je možné vypustit?

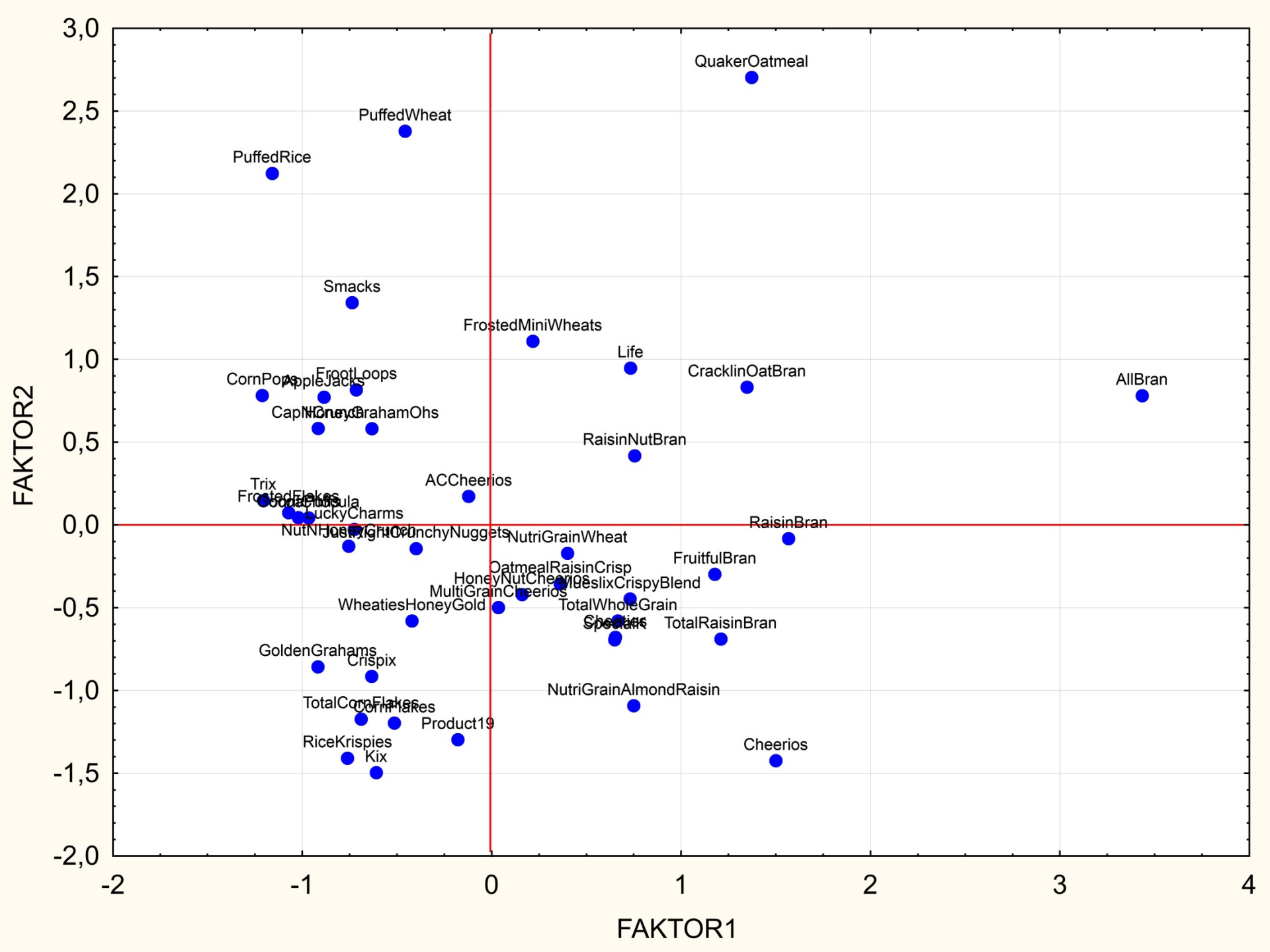
Data: *Datová matice Křupky obsahuje 55 dodavatelů obilých lupínků, u kterých bylo vyšetřováno 10 znaků: Objekt značí index obilných lupínků x_1, i značí jednoho ze tří výrobců G, K či Q x_2, Cal značí kalorickou hodnotu [cal] x_3, Bilkov značí obsah bílkovin x_4, Tuky značí obsah tuků x_5, Na značí 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}.*

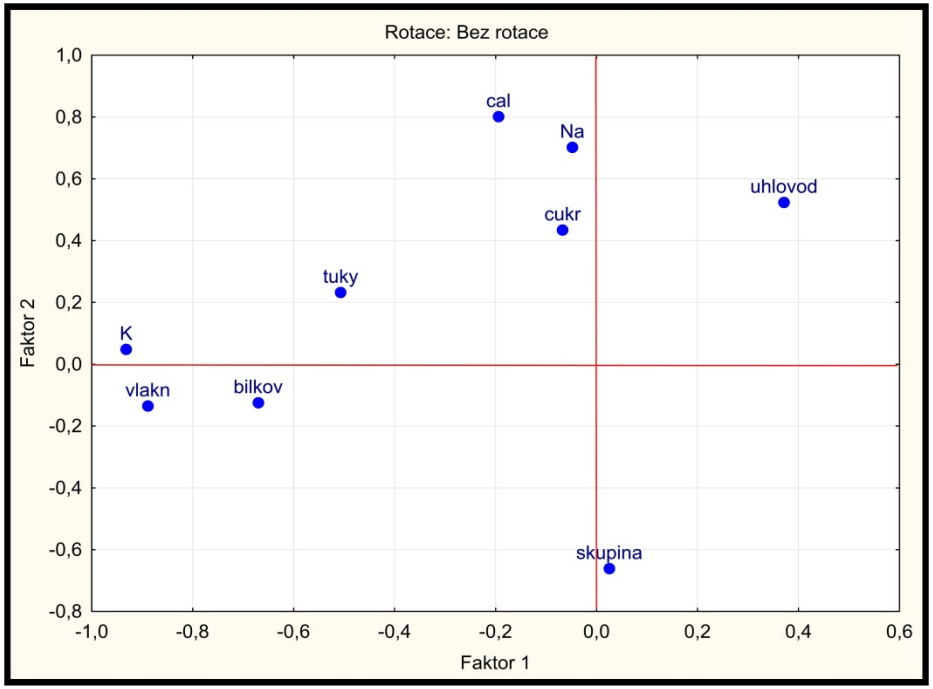
i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
	x_{10}	x_{11}							
1	ACCheerios		G	110	2	2	180	1.5	10.5
	10	70	1						
...	...	...	...	...	...	...	...	...	...
	...	..							
43	QuakerOatmeal		Q	100	5	2	0	2.7	1
	1	110	3						

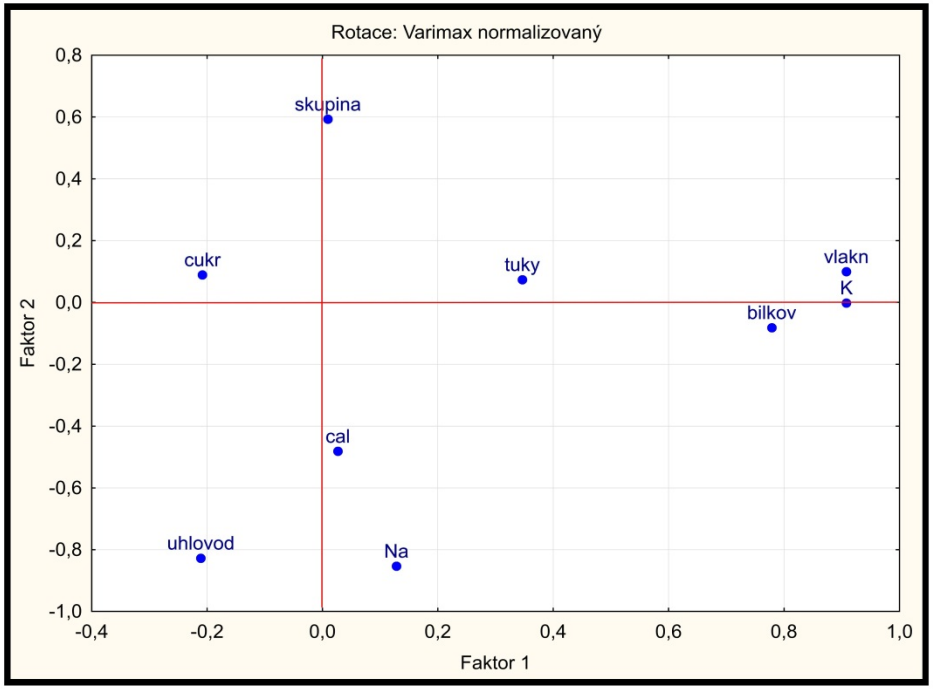

Rotace: Bez rotace
cal
Na
uhlovod
cukr
tuky
K
vlakn
bilkov
skupina
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
Faktor 2
Faktor 1

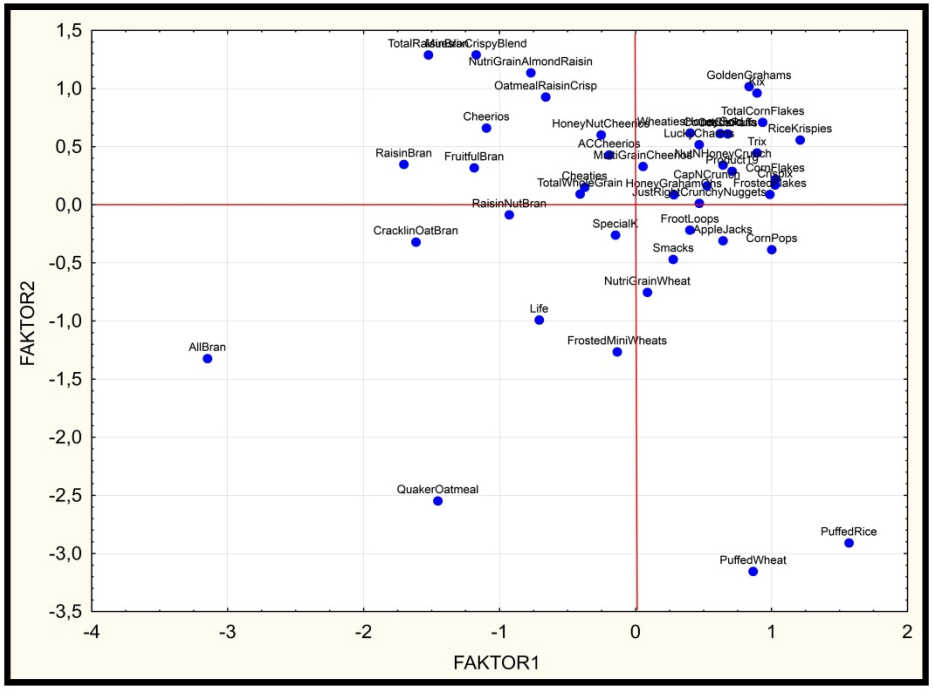

Rotace: Varimax normalizovaný
skupina
cukr
tuky
vlakn
K
bilkov
cal
uhlovod
Na
Faktor 2
Faktor 1
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0

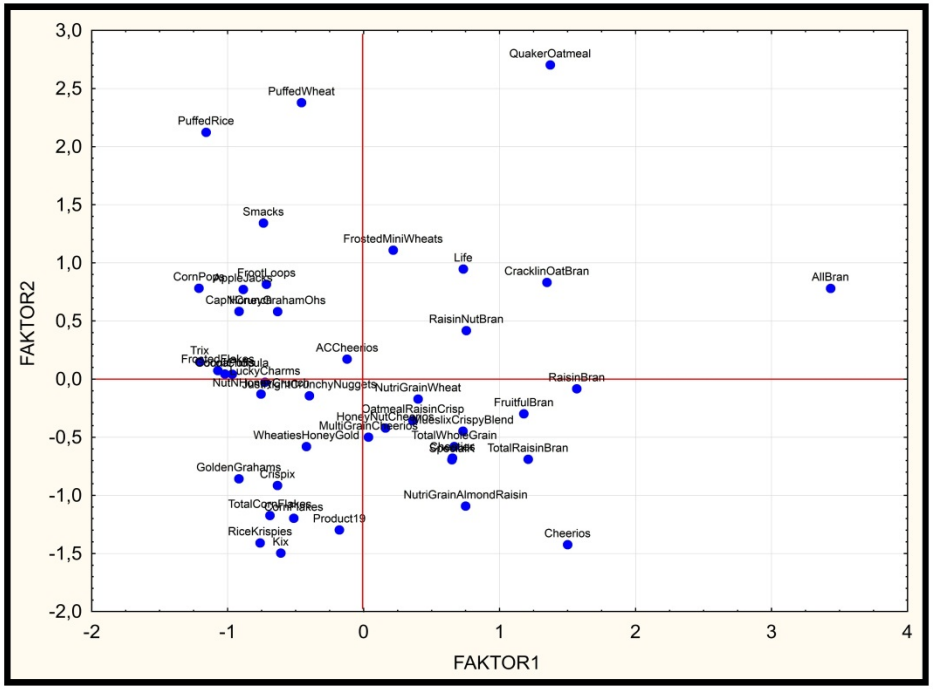

FAKTOR2
FAKTOR1
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2,5
-3,0
-3,5
-4
-3
-2
-1
0
1
2
NutriGrainAlmondRaisin
OatmealRaisinCrisp
GoldenGrahams
Kix
TotalCornFlakes
Cheerios
HoneyNutCheerios
RiceKrispies
LuckyCharms
ACCheerios
Trix
RaisinBran
FruitfulBran
MultiGrainCheerios
NutNHoneyCrunch
Product19
CornFlakes
Cheaties
CapNCrunch
Crispix
TotalWholeGrain
HoneyGrahamOhs
FrostedFlakes
JustRightCrunchyNuggets
RaisinNutBran
FrootLoops
SpecialK
AppleJacks
CracklinOatBran
CornPops
Smacks
NutriGrainWheat
Life
FrostedMiniWheats
AllBran
QuakerOatmeal
PuffedRice
PuffedWheat

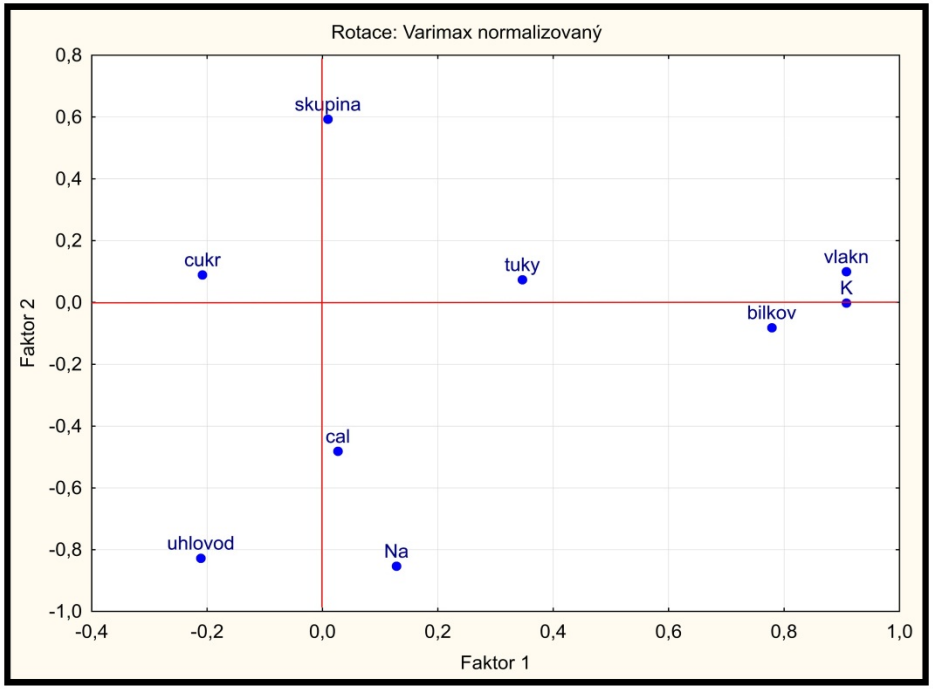

FAKTOR2
FAKTOR1
3,0
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2
-1
0
1
2
3
4
QuakerOatmeal
PuffedWheat
PuffedRice
Smacks
FrostedMiniWheats
Life
CracklinOatBran
AllBran
CornPops
AppleJacks
FrootLoops
CapNCrunch
HoneyGrahamOhs
RaisinNutBran
ACCheerios
Trix
FrostedFlakes
LuckyCharms
NutNHoney
JustRight
CrunchyNuggets
NutriGrainWheat
RaisinBran
FruitfulBran
OatmealRaisinCrisp
HoneyNutCheerios
MultiGrainCheerios
MueslixCrispyBlend
WheatiesHoneyGold
TotalWholeGrain
Cheerios
Specialk
TotalRaisinBran
GoldenGrahams
Crispix
NutriGrainAlmondRaisin
TotalCornFlakes
CornFlakes
Product19
RiceKrispies
Kix
Cheerios

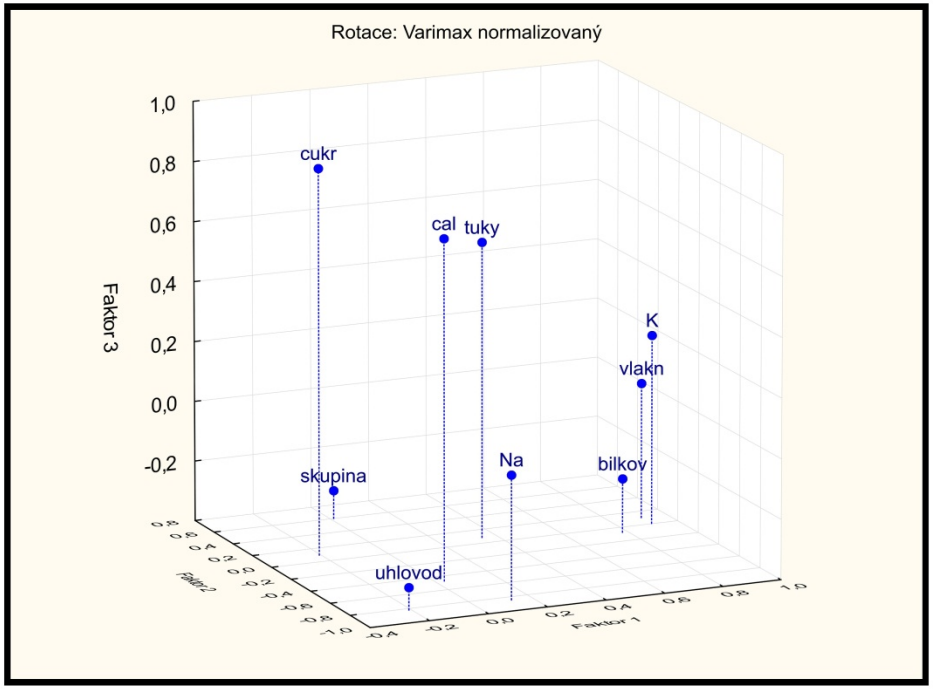

Rotace: Bez rotace
Faktor 2
Faktor 1
cal
Na
uhlovod
cukr
tuky
K
vlakn
bilkov
skupina

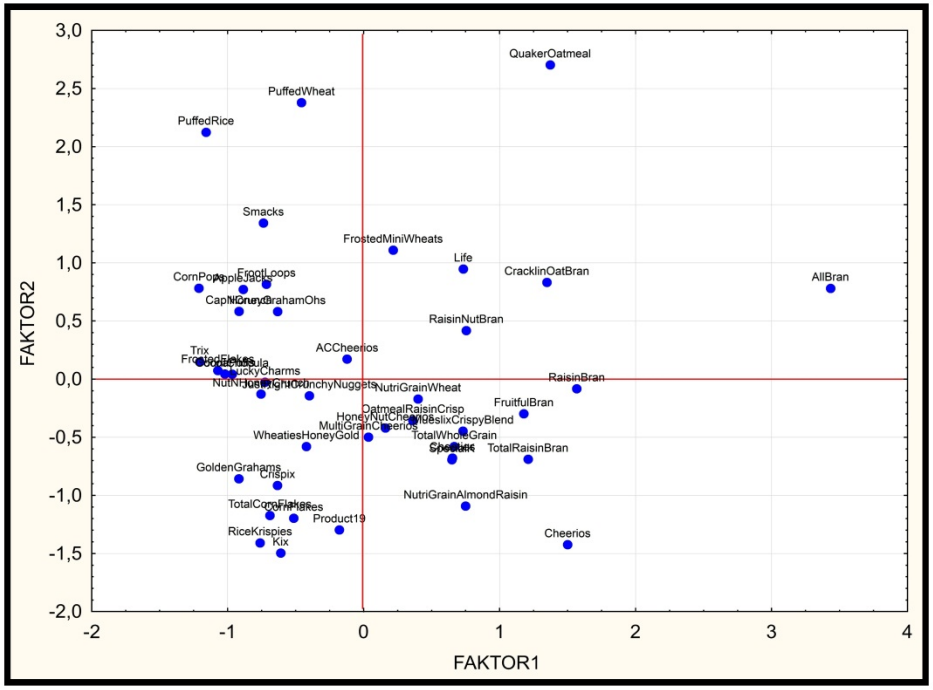

Rotace: Varimax normalizovaný
Faktor 2
Faktor 1
skupina
cukr
tuky
vlakn
K
bilkov
cal
uhlovod
Na

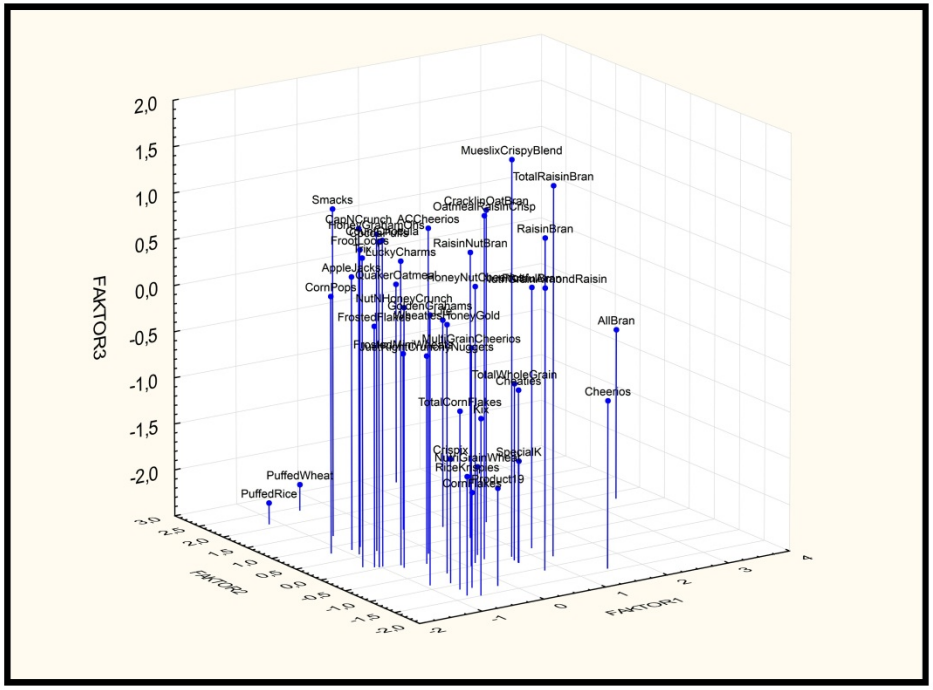

FAKTOR2
FAKTOR1
NutriGrainAlmondRaisin
OatmealRaisinCrisp
GoldenGrahams
TotalCornFlakes
Cheerios
RiceKrispies
Trix
RaisinBran
FruitfulBran
RaisinNutBran
CracklinOatBran
SpecialK
FrootLoops
AppleJacks
CornPops
Smacks
NutriGrainWheat
Life
FrostedMiniWheats
AllBran
QuakerOatmeal
PuffedRice
PuffedWheat


FAKTOR2
FAKTOR1
QuakerOatmeal
PuffedWheat
PuffedRice
Smacks
FrostedMiniWheats
Life
CracklinOatBran
AllBran
RaisinNutBran
ACCheerios
RaisinBran
FruitfulBran
NutriGrainWheat
WheatiesHoneyGold
TotalRaisinBran
GoldenGrahams
Crispix
NutriGrainAlmondRaisin
Product19
RiceKrispies
Kix
Cheerios


Rotace: Varimax normalizovaný
skupina
cukr
tuky
vlakn
K
bilkov
cal
uhlovod
Na
Faktor 2
Faktor 1


Rotace: Varimax normalizovaný
cukr
cal
tuky
K
vlakn
Na
bilkov
skupina
uhlovod
Faktor 3
Faktor 2
Faktor 1


QuakerOatmeal
PuffedWheat
PuffedRice
Smacks
FrostedMiniWheats
Life
CracklinOatBran
AllBran
RaisinNutBran
ACCheerios
RaisinBran
NutriGrainWheat
FruitfulBran
OatmealRaisinCrisp
MueslixCrispyBlend
TotalWholeGrain
TotalRaisinBran
GoldenGrahams
Crispix
NutriGrainAlmondRaisin
Product19
RiceKrispies
Kix
Cheerios
FAKTOR2
FAKTOR1


MueslixCrispyBlend
TotalRaisinBran
Smacks
RaisinBran
RaisinNutBran
AllBran
Cheerios
TotalWholeGrain
TotalCornFlakes
Kix
SpecialK
PuffedWheat
PuffedRice
Product19
FAKTOR3
FAKTOR2
FAKTOR1