



## SEMESTRÁLNÍ PRÁCE

### Metody s latentními proměnnými

2005/2006

Ing. Petr Eliáš



## 1. OTÁZKA 1

### 1.1 Zadání

Vypočtete algoritmem NIPALS 1. latentní proměnnou z níže zadané matice. Matici před zpracováním standardizujte.

$$\begin{matrix} \hat{e}_1 & 1 & 6 \\ \hat{e}_2 & 2 & 4 \\ \hat{e}_3 & 0 & 2 \end{matrix} \begin{matrix} u \\ u \\ u \end{matrix}$$

### 1.2 Řešení a závěr

První krok představuje výpočet vektorů aritmetických průměrů a vektorů směrodatných odchylek

$$\begin{aligned} \bar{x}^T &= \frac{\hat{e}_1 + \hat{e}_2 + \hat{e}_3}{3} = \frac{1 + 2 + 0}{3} = \frac{6 + 4 + 2}{3} = \frac{12}{3} = 4 \\ \bar{x}^T &= [2 \quad 1 \quad 4] \\ s^T &= [s_1 \quad s_2 \quad s_3] \\ s^T &= [1 \quad 1 \quad 2] \end{aligned}$$

Standardizovanou matici získáme, jestliže od každého prvku výběru odečteme střední hodnotu výběru a výsledek podělíme směrodatnou odchylkou. Získáme tak matici:

$$\begin{matrix} \hat{e}_1 - 4 & 0 & 1 \\ \hat{e}_2 & 0 & 0 \\ \hat{e}_3 & 1 & -1 \end{matrix} \begin{matrix} u \\ u \\ u \end{matrix}$$

Variabilita všech sloupců je stejná, jako odhad hlavní komponenty lze vzít libovolný sloupec, např. první

$$t_1^T = [-1 \quad 0 \quad 1]$$

Dosazením do vzorce

$$p_1^T = (t_1^T t_1)^{-1} t_1^T A$$

a znormováním podle výrazu

$$p_1^N = (p_1^T p_1)^{-\frac{1}{2}} p_1$$

získáváme počáteční odhad vektoru vyjadřujícího míru příspěvků odhadu hlavní komponenty  $t_1$ .

Dosazením do vztahu

$$\mathbf{t}_1^T = (\mathbf{p}_1^T \mathbf{p}_1)^{-1} \mathbf{p}_1^T \mathbf{A}^T$$

se získá odhad hlavní komponenty  $t_1$ . Opakováním postupu získáme stabilní rozklad vektorů  $t_1$  a  $p_1$ . Jako konvergenční kritérium uijeme vztahu

$$\mathbf{d} = (\mathbf{t}_{\text{nové}} - \mathbf{t}_{\text{staré}})^T (\mathbf{t}_{\text{nové}} - \mathbf{t}_{\text{staré}}) (\mathbf{t}_{\text{nové}}^T \mathbf{t}_{\text{nové}})^{-1}$$

Aby byl výpočet ukončen, je potřeba dosáhnout  $d/N < 10^{-10}$ . V našem případě bylo potřeba devíti kroků, abychom získali následující stabilní rozklad vektorů.

$$\begin{aligned} \mathbf{p}_1^T &= [0,628 \quad -0,460 \quad -0,628] \\ \mathbf{t}_1^T &= [-1,256 \quad -0,460 \quad 1,716] \end{aligned}$$

## 2. OTÁZKA 2

### 2.1 Zadání

S použitím vhodných kritérií určete nezbytný počet latentních proměnných, bylo-li z dat určeno:  $\text{PRESS}(0) = S(0) = 100$ ;  $\text{PRESS}(1) = 20$ ;  $S(1) = 10$ ;  $\text{PRESS}(2) = 3,5$ ;  $S(2) = 3,4$ ;  $\text{PRESS}(3) = 3,45$ ;  $S(3) = 3,39$ .

### 2.2 Řešení a závěr

Nezbytný počet latentních proměnných určíme s využitím Woldova kritéria, které je založeno na poměru

$$\frac{\text{PRESS}(P)}{S_R(P-1)}$$

Je-li hodnota kritéria vyšší než 0,95, není žádoucí zařazovat do modelu další latentní proměnnou.

V našem případě tedy získáváme

$$\begin{aligned} P = 1: \frac{\text{PRESS}(1)}{S_R(0)} &= \frac{20}{100} = 0,20 \\ P = 2: \frac{\text{PRESS}(2)}{S_R(1)} &= \frac{3,5}{10} = 0,35 \\ P = 3: \frac{\text{PRESS}(3)}{S_R(2)} &= \frac{3,45}{3,4} = 1,01 \end{aligned}$$

Vzhledem k tomu, že pro  $P = 3$  je hodnota poměru vyšší než 0,95, není čtvrtá latentní proměnná významná.

### 3. OTÁZKA 3

#### 3.1 Zadání

Odhadněte hodnotu chybějícího prvku  $A[2,2]$ , jestliže výpočtem z nekompletní matice byly určeny vektory  $p$ : 0,541; 0,423; 0,514 a  $t$ : -1,340; -0,735; 2,076.

#### 3.2 Řešení a závěr

Prvky zdrojové matice odpovídající  $p$ -té latentní proměnné lze získat ze vztahu

$$\mathbf{A}_p^{\text{pred}} = \mathbf{t}_p \mathbf{p}_p^T$$

Pro odhad hodnoty prvku  $A[2,2]$  tedy platí:

$$A[2;2] = t[2]p[2] = -0,311$$

### 4. OTÁZKA 4

#### 4.1 Zadání

Výpočtem metodou PCA byly určeny vektory

$$\begin{aligned} \mathbf{p}_1 &= [0,012 \quad 0,458 \quad -0,352 \quad 0,987] \\ \mathbf{p}_2 &= [0,926 \quad -0,238 \quad 0,872 \quad -0,115] \end{aligned}$$

Vypočtete komunalitu a vyberte sloupec, který nejlépe charakterizuje celou matici.

#### 4.2 Řešení a závěr

$p$  ... počet latentních proměnných (zde 2)  
 $i$  ... index sloupce ( $M = 4$ )

Komunalitou je určena míra rozptylu vyjádřená příslušnou latentní proměnnou. Jsou-li známy vektory zátěží, spočtou se prve normované zátěže s využitím výrazu

$$\mathbf{P}_{ip}^N = \frac{\mathbf{P}_{ip}}{\sqrt{\sum_{i=1}^M \mathbf{P}_{ip}^2}}$$

Komunalita se poté vyjádří ze vztahu

$$\mathbf{h}_i^2 = \sum_{p=1}^P \mathbf{P}_{ip}^N$$

Tzn. pro

$$\mathbf{P}_{11}^N = \frac{0,012}{\sqrt{(0,012^2 + 0,458^2 + 0,352^2 + 0,987^2)}}$$

$$P_{11}^N = 0,0092$$

nebo pro

$$P_{12}^N = \frac{0,926}{(0,926^2 + 0,238^2 + 0,872^2 + 0,115^2)}$$

$$P_{12}^N = 0,549$$

Hodnota komunality

$$h_1^2 = (P_{11}^N)^2 + (P_{12}^N)^2$$

$$h_1^2 = 0,0092^2 + 0,549^2$$

$$h_1^2 = 0,297$$

Obdobně vyčíslíme i další komunality

$$h_2^2 = 0,143$$

$$h_3^2 = 0,340$$

$$h_4^2 = 0,574$$

Platí, že čím vyšší je komunalita, tím lépe sloupec charakterizuje zdrojovou matici. V našem případě tedy zdrojovou matici nejlépe charakterizuje sloupec č. 4.

## 5. OTÁZKA 5

### 5.1 Zadání

Vysvětlete, proč vysvětlená variabilita je při výpočtu metodou FA vždy nižší, než při výpočtu metodou PCA.

### 5.2 Řešení a závěr

U metody PCA je celková variabilita shodná s vysvětlenou variabilitou při použití všech sloupců zdrojové matice. U metody FA je celková variabilita rozkládána na část vysvětlenou – komunalitu a část nevysvětlenou – jedinečnost. Metoda FA tedy nepopíše 100 % variability.

## 6. OTÁZKA 6

### 6.1 Zadání

Výpočtem metodou kanonických korelací bylo zjištěno:

$$0,297X_1 + 0,298X_2 + 0,050X_3 + 0,256X_4 = 0,493Y_1 - 0,213Y_2$$

$$r_1 = 0,830$$

$$0,006X_1 - 0,115X_2 + 0,950X_3 + 0,056X_4 = 0,493Y_1 + 0,213Y_2$$

$$r_1 = 0,512$$

Vypočtete skupinový korelační koeficient a interpretujte výsledky.

## 6.2 Řešení a závěr

Skupinový korelační koeficient vypočteme dosazením hodnot do výrazu

$$R = \left[1 - (1 - r_1^2)(1 - r_2^2)\right]^{\frac{1}{2}}$$
$$R = \left[1 - (1 - 0,830^2)(1 - 0,512^2)\right]^{\frac{1}{2}}$$
$$R = 0,878$$
$$R^2 = 0,770$$

Skupinový korelační koeficient má hodnotu 0,770, tzn., že je popsáno 77 % variability v datech.

## 7. OTÁZKA 7

### 7.1 Zadání

Uveďte nějaký konkrétní příklad vhodný pro zpracování metodou PLS.

### 7.2 Řešení a závěr

U vybraných šarží titanové běloby typu RGU se pokuste popsat vztah mezi chemickým složením a dalšími sledovanými parametry (spotřeba oleje, barevné souřadnice  $L^*$  a  $b^*$ )

Šarže	Spotřeba oleje	$L^*$	$b^*$	$Al_2O_3$	$SiO_2$	C organický
1	21,8	97,49	1,57	2,783	0,771	0,280
2	22,0	97,39	1,37	2,741	0,835	0,260
3	21,8	97,59	1,27	2,831	0,939	0,230
4	21,8	97,69	1,17	3,052	1,007	0,220
5	22,0	97,59	1,07	3,130	0,997	0,250
6	22,2	98,09	1,37	3,212	1,006	0,230
7	22,3	97,79	1,37	3,287	1,009	0,214
8	22,2	97,89	1,37	3,348	1,006	0,216
9	23,0	97,19	1,27	3,332	1,000	0,244
10	23,1	97,19	1,37	2,969	0,818	0,222
11	23,6	97,29	1,37	3,083	0,941	0,189
12	24,0	97,09	1,37	3,146	0,991	0,201
13	23,2	97,09	1,17	3,258	1,068	0,203
14	23,3	96,99	1,17	3,232	1,096	0,219
15	23,5	97,09	1,27	3,213	1,110	0,216
16	23,6	97,09	1,17	3,114	1,110	0,251
17	23,2	97,09	1,17	3,042	1,078	0,281
18	23,0	97,19	1,17	3,016	1,105	0,238
19	23,4	96,79	1,27	3,005	0,877	0,422
20	23,1	97,59	1,37	3,011	1,012	0,230
21	23,2	97,29	1,07	2,915	1,013	0,289
22	22,4	97,39	1,17	2,857	0,994	0,231
23	23,2	97,49	1,17	2,939	1,043	0,216
24	22,9	97,39	1,07	2,918	1,048	0,243
25	23,5	97,49	1,07	2,919	1,042	0,231
26	23,6	97,59	1,27	2,945	1,036	0,248
27	23,9	97,29	1,47	2,958	1,023	0,253
28	23,5	97,59	1,57	2,993	1,027	0,228
29	21,9	97,39	1,07	2,662	0,656	0,235
30	21,6	97,39	1,17	2,862	0,729	0,230
31	21,2	97,49	1,17	2,738	0,717	0,186
32	20,1	97,29	1,07	2,545	0,748	0,241
33	20,2	97,39	1,17	2,557	0,735	0,233
34	22,0	97,39	1,07	2,727	0,634	0,281

Sarže	Spotřeba oleje	L*	b*	Al <sub>2</sub> O <sub>3</sub>	SiO <sub>2</sub>	C organický
35	21,5	97,29	1,17	2,735	0,534	0,269
36	21,8	97,39	1,17	2,757	0,530	0,282
37	21,7	97,39	1,27	2,863	0,592	0,274
38	21,5	97,59	1,17	2,823	0,597	0,314
39	21,7	97,49	1,07	2,740	0,611	0,341
40	21,1	97,49	1,07	2,716	0,643	0,288
41	21,0	97,39	0,97	2,726	0,675	0,258
42	19,7	97,39	1,07	2,698	0,697	0,265
43	21,4	97,39	1,17	2,762	0,726	0,247
44	21,4	97,39	1,27	2,744	0,725	0,279
45	20,7	97,09	0,97	2,980	0,534	0,306
46	21,1	97,39	1,17	2,885	0,620	0,295
47	21,4	96,99	1,07	2,822	0,658	0,231
48	20,9	97,39	1,07	2,918	0,731	0,352
49	21,0	97,39	1,07	2,868	0,755	0,315
50	21,0	97,19	1,17	2,806	0,750	0,304
51	20,6	97,39	1,17	2,770	0,764	0,300
52	21,3	97,59	1,17	2,853	0,726	0,204
53	20,8	97,79	1,17	2,701	0,752	0,281
54	21,1	96,99	0,87	2,845	0,750	0,420
55	21,0	97,39	1,07	2,732	0,763	0,261
56	21,0	97,19	1,07	2,781	0,773	0,282
57	21,4	96,89	0,87	2,909	0,758	0,344
58	20,6	97,59	1,17	2,694	0,777	0,309
59	21,2	96,99	0,97	2,840	0,767	0,310
60	21,2	97,39	1,17	2,849	0,776	0,292
61	21,5	97,09	0,87	2,913	0,777	0,331
62	21,3	97,29	1,07	3,087	0,769	0,310
63	21,2	96,79	0,87	2,900	0,783	0,323
64	21,1	97,09	1,07	2,857	0,771	0,303
65	22,0	97,19	1,27	3,103	1,128	0,224
66	22,0	97,09	1,17	2,901	1,047	0,227
67	22,0	97,09	1,17	2,953	0,777	0,255
68	22,0	97,19	1,17	2,874	0,922	0,342
69	21,7	97,59	1,27	2,932	0,766	0,311
70	21,7	97,29	1,27	2,882	0,775	0,315
71	21,7	97,19	1,17	2,962	0,885	0,321
72	21,6	97,09	1,07	2,816	0,764	0,306
73	21,8	97,49	1,17	2,901	0,837	0,274
74	20,4	97,19	1,27	2,729	0,738	0,307
75	20,8	96,99	1,17	2,804	0,850	0,275
76	20,7	97,19	1,37	2,698	0,798	0,257
77	20,7	96,69	1,37	2,673	0,743	0,287
78	20,7	96,69	1,37	2,561	0,757	0,303
79	20,5	97,19	1,37	2,513	0,689	0,257
80	20,8	96,49	1,37	2,507	0,703	0,330
81	20,5	96,89	1,37	2,463	0,679	0,266
82	20,4	96,89	1,37	2,400	0,698	0,297
83	20,6	97,09	1,37	2,489	0,739	0,319
84	20,8	96,99	1,27	2,452	0,778	0,330
85	21,0	97,09	1,17	2,518	0,557	0,269
86	21,5	97,29	1,17	2,724	0,664	0,333
87	21,2	97,09	1,07	2,739	0,671	0,345
88	21,7	96,99	1,17	2,734	0,705	0,344
89	21,5	96,99	1,07	2,797	0,695	0,345
90	21,5	96,99	1,07	2,797	0,695	0,345
91	21,3	97,29	1,17	2,788	0,667	0,302
92	21,5	97,19	1,07	2,848	0,682	0,386
93	22,0	97,39	1,27	2,851	0,696	0,252
94	21,4	97,49	1,27	2,761	0,686	0,253
95	21,3	97,49	1,17	2,804	0,752	0,299
96	21,0	97,49	1,17	2,796	0,788	0,298
97	21,1	97,39	1,17	2,849	0,825	0,332
98	21,8	97,09	1,07	2,892	0,999	0,298
99	21,3	97,29	1,07	2,878	0,829	0,340
100	21,3	97,59	1,07	2,838	0,856	0,313

Soubor s daty: RGU\_sarze.xls

## 8. Otázka 8

### 8.1 Zadání

Jeden objekt je charakterizován metrickými znaky (2,10), druhý (3,8), třetí (4,9), čtvrtý (10,4) a pátý (11,5). Vypočtete matici vzdáleností v Eukleidově metrice a dokumentujte výpočet shlukování některou z používaných metod. Výsledky interpretujte graficky.

### 8.2 Řešení a závěr

Eukleidovu vzdálenost spočteme podle vzorce

$$d_E(\mathbf{X}_k, \mathbf{X}_l) = \sqrt{\sum_{p=1}^P (x_{kp} - x_{lp})^2}$$

Pro vzdálenosti prvků 2 a 1 tedy dostáváme

$$d_{2,1} = \sqrt{(2 - 3)^2 + (10 - 8)^2}$$
$$d_{2,1} = 2,24$$

Aplikací vzorce na kombinaci všech prvků získáme matici vzdáleností, která má tvar

1	0,00				
2	2,24	0,00			
3	2,24	1,41	0,00		
4	10,00	9,75	7,81	0,00	
5	10,30	8,54	8,06	1,41	0,00

Z této matice je patrné, že nejmenší vzdálenosti mají prvky 3-2 a 5-4, které vytvoří první shluky. Spočte se těžiště nových shluků a též odpovídající matice vzdáleností.

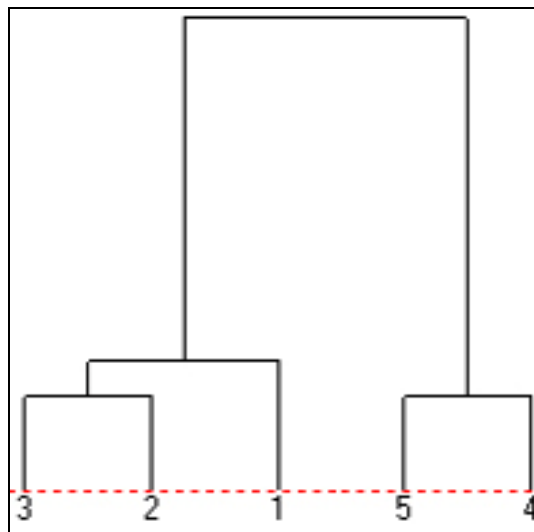
1	0,00		
(2 - 3)	2,01	0,00	
(4 - 5)	10,12	8,06	0,00

Z výše uvedené matice vyplývá, že nejmenší vzdálenost má prvek 1 a shluk (2-3). Spojením do nového shluku, vypočtením těžiště a vzdáleností získáváme matici

(1 - 2 - 3)	0,00	
(4 - 5)	8,75	0,00

Výše uvedený postup lze znázornit pomocí dendrogramu (obr. 1).





Obr. 1 Dendrogram

## 9. OTÁZKA 9

### 9.1 Zadání

Popište slovně postup aplikace metod s latentními proměnnými nebo klasifikačních metod na nějakém konkrétním příkladu ze své praxe.

### 9.2 Řešení

Ve výrobním programu podniku Precheza a.s. se nachází několik typů titanových bělob v rutilové i anatasové modifikaci. Jedním z polotovarů vznikajících při jejich výrobě jsou různé druhy kalcinátů, jejichž vlastnosti se liší v závislosti na typu vyráběného konečného produktu.

Proveďte analýzu vícerozměrných dat metodou hlavních komponent a pokuste se objasnit vazby mezi jednotlivými veličinami.

### 9.3 Data

Vzorek	SFM2	ZF	MF	ČF	Bělost	Rutil	Sb <sub>2</sub> O <sub>3</sub>	K <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>	Al <sub>2</sub> O <sub>3</sub>
1	126,0	96,7	95,8	97,9	96,8	99,8	0,003	0,161	0,196	0,009
2	116,1	97,9	97,4	98,4	97,9	100,0	0,004	0,165	0,197	0,013
3	123,0	97,6	97,0	98,4	97,7	100,0	0,003	0,163	0,193	0,018
4	117,9	96,6	95,7	97,1	96,5	99,6	0,003	0,162	0,188	0,011
5	130,7	97,8	97,3	98,6	97,9	99,9	0,003	0,165	0,187	0,014
6	94,6	97,1	96,4	98,2	97,2	100,0	0,001	0,163	0,187	0,018
7	109,0	97,0	96,1	97,9	97,0	100,0	0,004	0,165	0,185	0,010
8	128,8	97,2	96,7	97,9	97,3	99,9	0,003	0,157	0,181	0,006
9	138,3	98,1	97,4	98,5	98,0	99,9	0,002	0,165	0,179	0,011
10	145,4	98,5	98,0	99,1	98,5	99,7	0,003	0,170	0,181	0,013
11	119,7	97,6	97,1	98,2	97,6	100,0	0,003	0,164	0,176	0,008
12	120,9	97,4	96,8	98,1	97,4	100,0	0,002	0,156	0,176	0,011
13	113,7	97,6	97,0	98,4	97,7	99,9	0,003	0,164	0,180	0,012
14	115,0	96,8	95,9	97,9	96,9	100,0	0,003	0,162	0,179	0,018
15	117,5	96,6	95,8	97,6	96,7	99,9	0,003	0,172	0,178	0,009
16	105,8	96,8	95,9	97,8	96,8	100,0	0,004	0,175	0,180	0,011
17	118,1	97,4	96,8	98,3	97,5	99,9	0,003	0,173	0,177	0,020
18	140,5	97,5	96,5	98,0	97,3	99,5	0,003	0,175	0,180	0,009
19	118,7	97,5	96,9	98,3	97,6	100,0	0,003	0,176	0,179	0,010
20	125,9	98,2	97,4	98,7	98,1	99,8	0,002	0,176	0,181	0,013

Vzorek	SFM2	ZF	MF	ČF	Bělost	Rutil	Sb <sub>2</sub> O <sub>3</sub>	K <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>	Al <sub>2</sub> O <sub>3</sub>
21	192,9	98,7	98,9	98,6	98,7	1,1	0,002	0,210	0,264	0,013
22	192,9	98,4	98,3	98,3	98,3	0,7	0,000	0,202	0,265	0,018
23	201,5	98,4	98,0	98,3	98,2	0,8	0,004	0,204	0,269	0,011
24	181,6	98,8	98,8	98,8	98,8	0,8	0,004	0,212	0,271	0,016
25	192,2	98,6	98,7	98,7	98,7	0,7	0,003	0,212	0,273	0,024
26	191,1	98,4	98,4	98,5	98,4	0,5	0,003	0,202	0,278	0,010
27	193,5	98,9	99,2	99,0	99,0	0,5	0,001	0,221	0,281	0,017
28	187,0	98,6	98,9	98,6	98,7	0,6	0,003	0,226	0,282	0,015
29	202,1	98,8	98,8	98,5	98,7	0,5	0,001	0,204	0,282	0,018
30	199,4	98,6	98,7	98,4	98,6	0,3	0,003	0,227	0,287	0,010
31	202,6	98,8	98,9	98,5	98,7	0,5	0,001	0,218	0,287	0,017
32	198,2	98,7	98,7	98,3	98,6	0,5	0,002	0,219	0,283	0,020
33	103,0	94,8	95,7	93,7	94,7	0,7	0,013	0,225	0,283	0,014
34	123,7	95,9	96,3	94,8	95,7	0,5	0,026	0,256	0,276	0,025
35	188,2	97,3	97,4	97,1	97,3	0,4	0,020	0,196	0,283	0,015
36	184,8	98,3	98,2	97,9	98,1	0,9	0,014	0,202	0,284	0,012
37	195,9	98,3	98,4	98,4	98,4	0,2	0,010	0,203	0,279	0,014
38	181,3	97,6	97,7	97,6	97,6	0,4	0,009	0,206	0,276	0,009
39	192,6	98,3	98,4	98,3	98,3	0,5	0,007	0,231	0,273	0,015
40	221,1	98,5	98,2	98,6	98,4	0,2	0,005	0,197	0,267	0,009
41	187,3	98,3	97,9	98,4	98,2	0,9	0,404	0,218	0,288	0,000
42	186,6	98,6	98,5	98,4	98,5	0,9	0,409	0,207	0,287	0,000
43	177,8	98,2	98,3	98,0	98,2	1,1	0,414	0,217	0,292	0,000
44	192,8	97,5	97,2	97,2	97,3	0,9	0,433	0,217	0,289	0,000
45	179,4	98,5	98,5	98,5	98,5	0,7	0,423	0,214	0,282	0,000
46	166,3	97,8	98,1	97,2	97,7	1,0	0,399	0,210	0,284	0,000
47	172,7	97,5	97,9	97,2	97,5	0,7	0,386	0,212	0,281	0,000
48	165,1	97,6	98,0	97,4	97,7	1,0	0,385	0,215	0,281	0,000
49	176,0	98,2	98,2	98,0	98,1	0,7	0,380	0,211	0,280	0,000
50	177,3	98,0	98,1	97,8	98,0	0,9	0,387	0,212	0,283	0,000
51	183,3	97,7	97,7	97,4	97,6	0,6	0,383	0,213	0,285	0,000
52	152,4	97,0	97,4	96,3	96,9	3,0	0,365	0,215	0,287	0,000
53	162,9	97,5	97,9	97,0	97,5	1,1	0,245	0,211	0,265	0,000
54	166,2	97,9	98,2	97,4	97,8	1,1	0,243	0,213	0,262	0,000
55	175,4	98,1	98,1	97,8	98,0	0,3	0,290	0,222	0,272	0,000
56	165,9	97,7	98,0	97,4	97,7	1,5	0,348	0,219	0,266	0,000
57	171,1	98,0	98,2	97,6	97,9	1,4	0,345	0,221	0,271	0,000
58	172,8	97,8	97,9	97,5	97,7	1,5	0,357	0,220	0,271	0,000
59	176,5	98,4	98,5	98,0	98,3	0,9	0,364	0,213	0,273	0,000
60	165,5	98,0	98,2	97,6	97,9	1,1	0,351	0,223	0,276	0,000
61	107,6	95,6	94,3	97,1	95,7	99,9	0,003	0,160	0,227	0,295
62	106,4	95,1	93,3	96,5	95,0	99,4	0,003	0,164	0,225	0,301
63	100,0	96,1	94,6	97,2	96,0	99,7	0,003	0,157	0,219	0,293
64	97,6	95,0	93,1	96,5	94,9	98,0	0,002	0,154	0,225	0,283
65	99,9	95,0	93,4	96,6	95,0	99,9	0,005	0,161	0,231	0,306
66	100,0	95,9	94,7	97,2	95,9	100,0	0,004	0,161	0,223	0,300
67	99,6	94,7	93,0	96,4	94,7	99,7	0,003	0,159	0,223	0,312
68	103,5	94,5	92,8	96,2	94,5	99,9	0,003	0,146	0,219	0,280
69	103,1	93,9	92,0	95,8	93,9	99,7	0,003	0,161	0,224	0,314
70	106,5	94,9	93,3	96,6	94,9	99,3	0,004	0,162	0,226	0,306
71	100,3	93,7	91,8	95,9	93,8	99,9	0,003	0,161	0,220	0,302
72	107,6	94,6	93,1	96,3	94,7	99,2	0,003	0,161	0,227	0,306
73	102,1	94,6	92,7	96,1	94,5	99,7	0,002	0,161	0,225	0,296
74	101,9	93,4	91,4	95,7	93,5	99,7	0,004	0,160	0,227	0,314
75	97,6	95,3	93,9	96,9	95,4	99,7	0,004	0,158	0,228	0,296
76	104,3	95,5	94,1	96,9	95,5	99,5	0,002	0,157	0,224	0,305
77	101,2	94,4	92,7	96,3	94,5	99,5	0,004	0,166	0,227	0,315
78	109,9	94,1	92,4	96,0	94,2	99,1	0,003	0,163	0,221	0,302
79	111,1	96,2	95,2	97,5	96,3	99,4	0,002	0,167	0,227	0,312
80	106,7	94,7	93,1	96,5	94,8	99,4	0,004	0,162	0,229	0,311

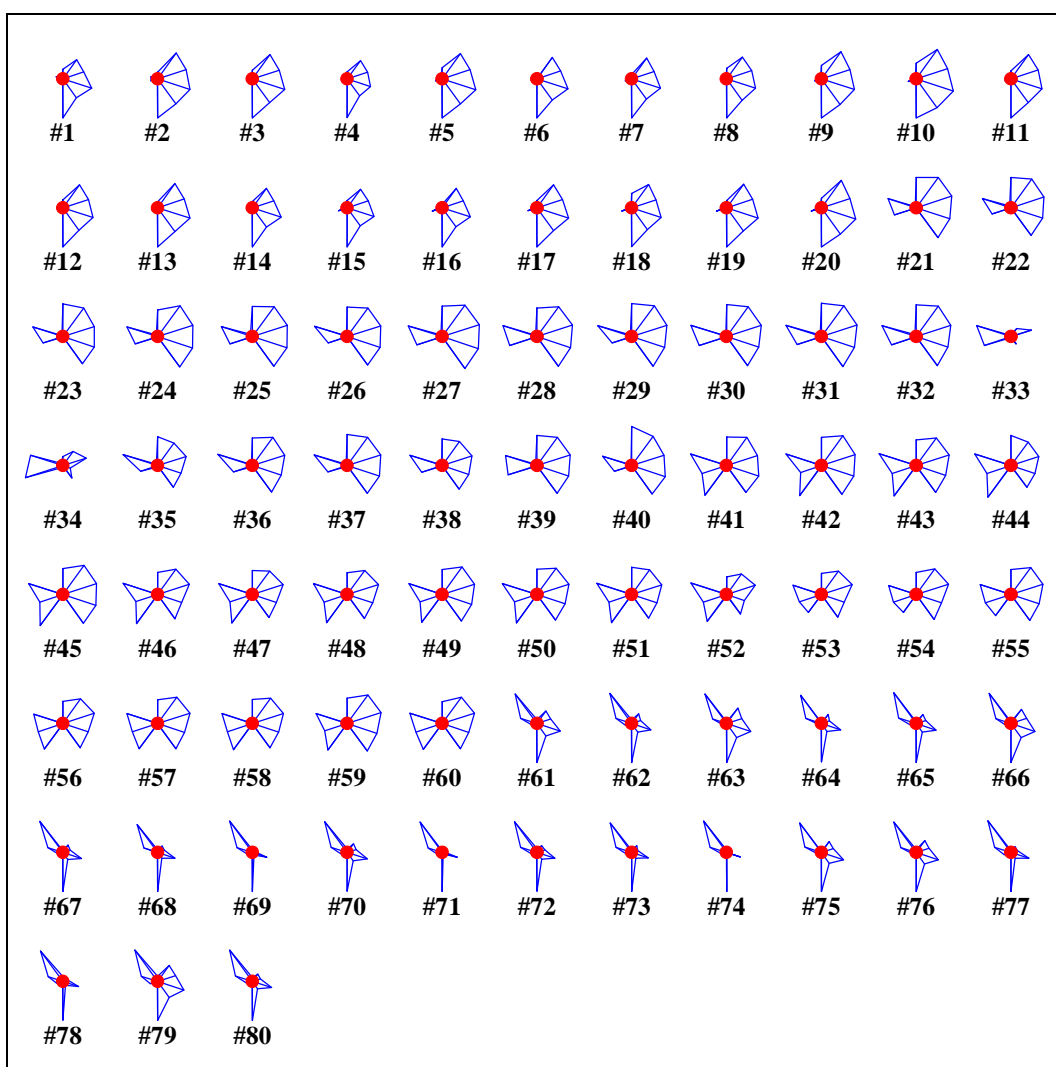
Soubor s daty: Kalcinaty.xls  
Použitý software: STATISTICA v7.1

## 9.4 Obecný postup

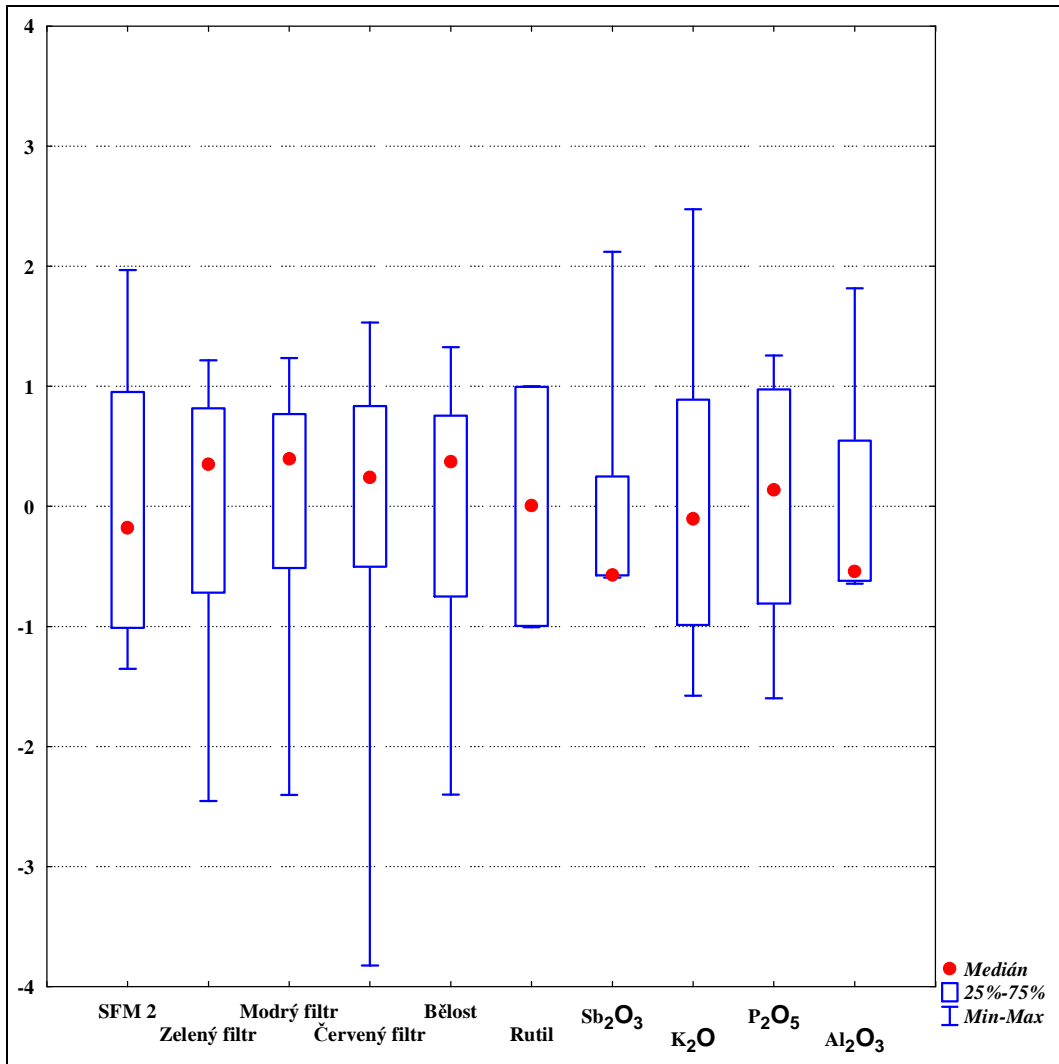
- Průzkumová analýza
- Analýza korelační matice
- Volba metody, zde PCA
- Určení počtu latentních proměnných
- Výpočet latentních proměnných
- Analýza matice zátěží a matice latentních proměnných
- Interpretace výsledků

## 9.5 Analýzy variability znaků

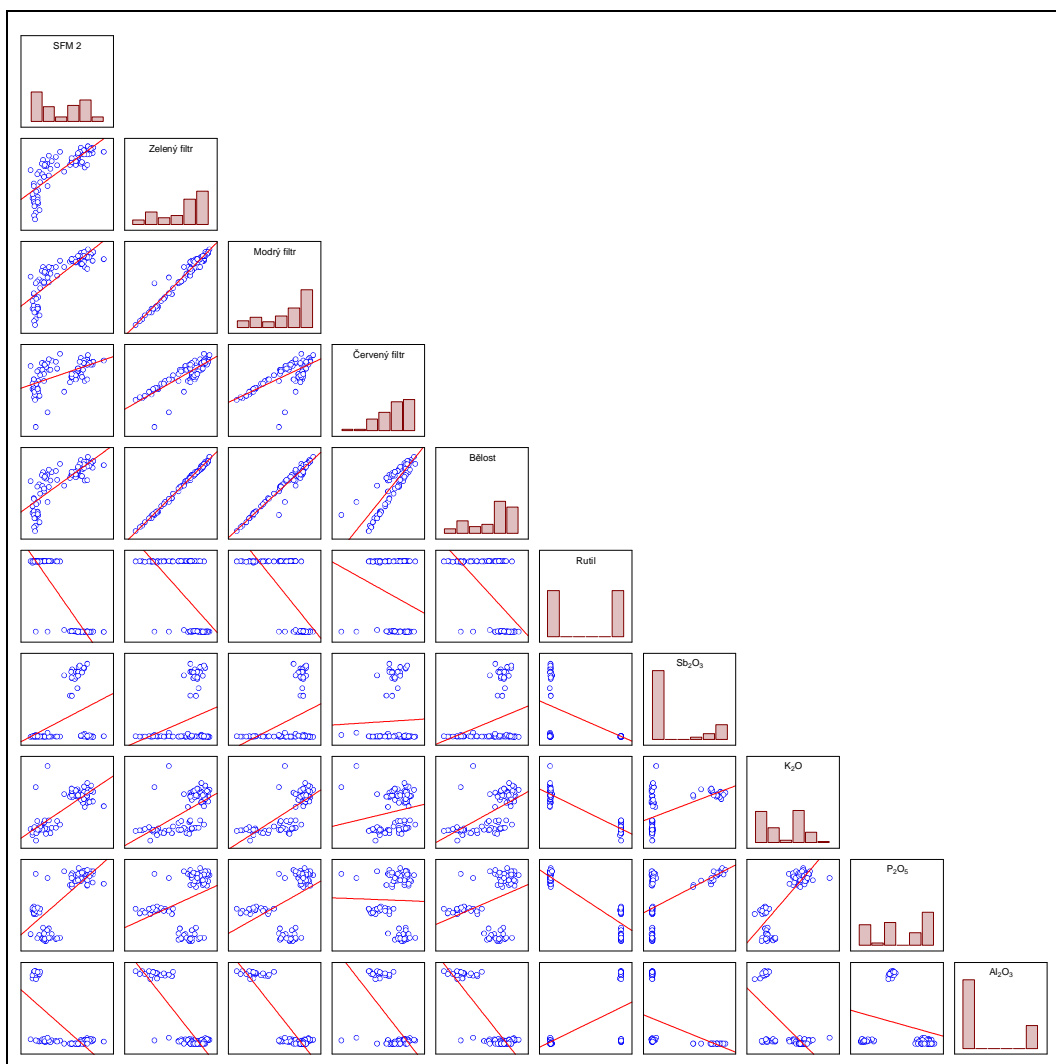
### 9.5.1 Vybrané grafy grafického výstupu programu STATISTICA



Obr. 2 Ikonový graf (hvězdy)



Obr. 3 Krabicový graf pro standardizovaná data



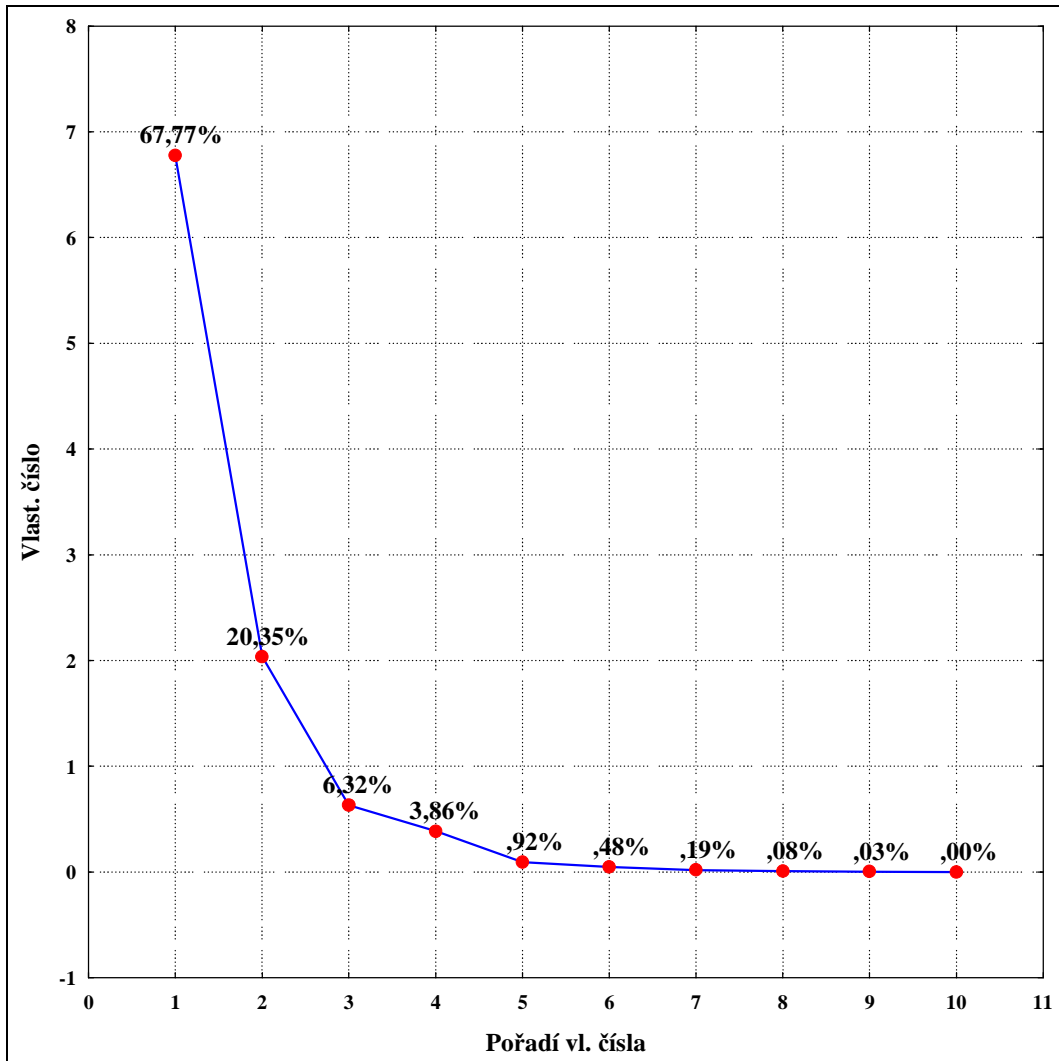
**Obr. 4** Graf korelace

### 9.5.2 *Komentář k vybraným grafům grafického výstupu programu STATISTICA*

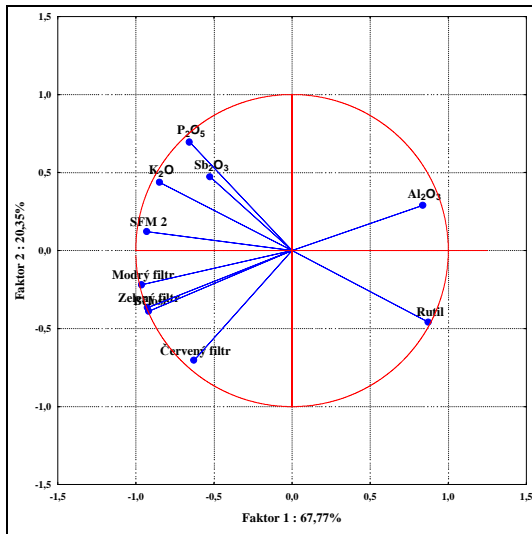
- *Ikonový graf (obr. 2)*  
Obecně umožňuje nalezení podobných objektů charakterizovaných stejnými hodnotami příslušných znaků. U konkrétního grafu na obr. 1 je možno rozlišit 4 skupiny podobných objektů, které odpovídají 4 druhům kalcinátů.
- *Krabicový graf (obr. 3)*  
Odlišné měřítko pro různé znaky u neupravených dat by neumožnilo posoudit proměnlivost každého znaku v příslušných objektech. Proto byla data standardizována. Nejmenší proměnlivost vykazuje  $\text{Sb}_2\text{O}_3$ .
- *Graf korelace (obr. 4)*  
Z předmětného grafu je patrná silná lineární závislost mezi znaky SFM 2, bělost, zelený, modrý a červený filtr.

## 9.6 Metoda hlavních komponent (PCA)

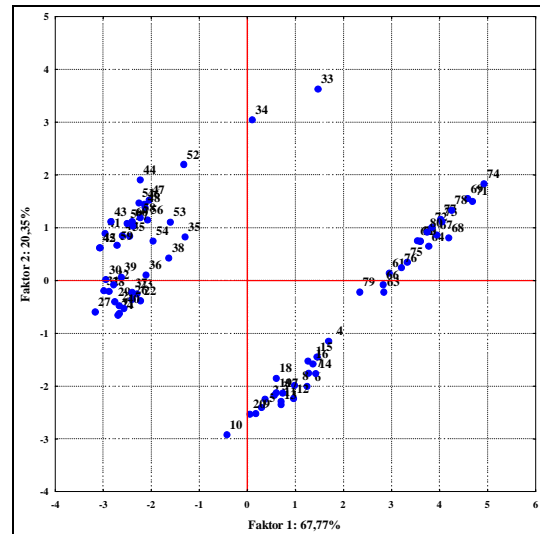
### 9.6.1 Vybrané grafy grafického výstupu programu STATISTICA (před redukcí počtu znaků)



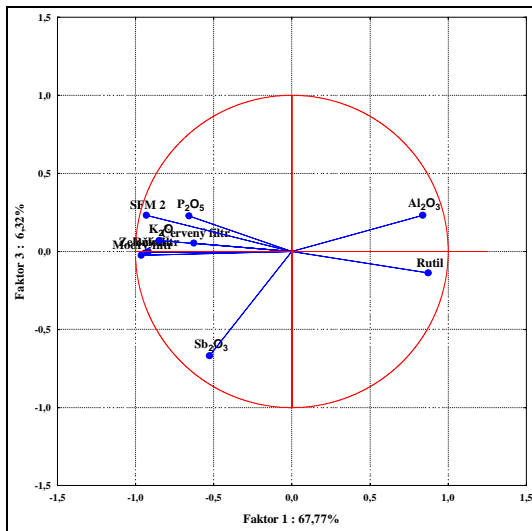
Obr. 5 Cattelův indexový graf úpatí vlastních čísel



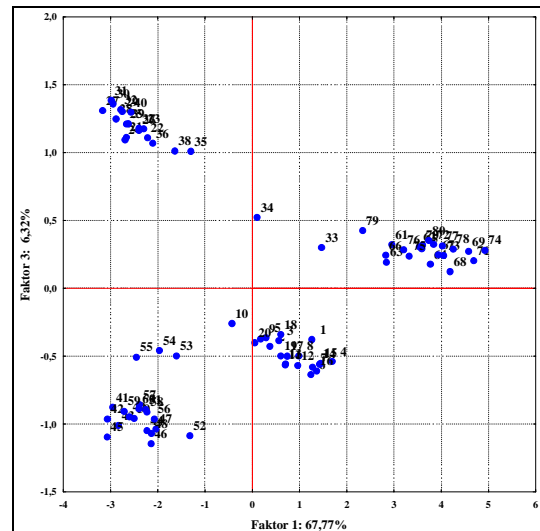
Obr. 6 Graf komp. vah 1 a 2



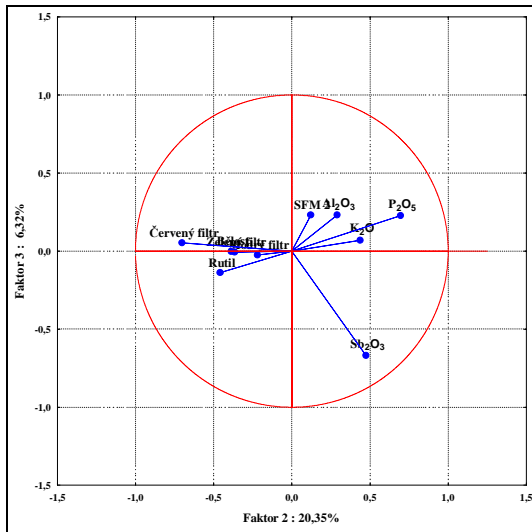
Obr. 7 Rozpt. diagram komp. skóre 1 a 2



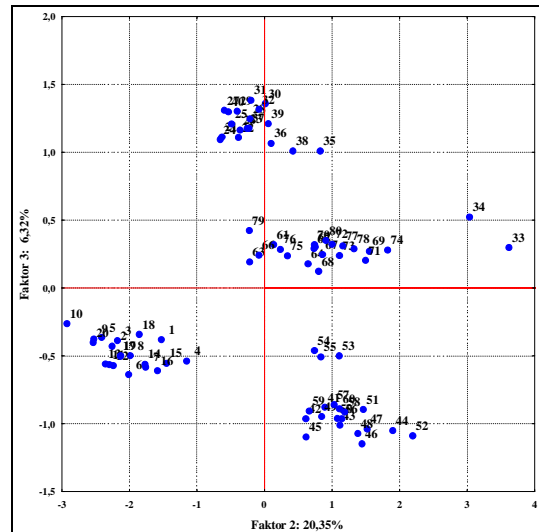
Obr. 8 Graf komp. vah 1 a 3



Obr. 9 Rozpt. diagram komp. skóre 1 a 3



Obr. 10 Graf komp. vah 2 a 3



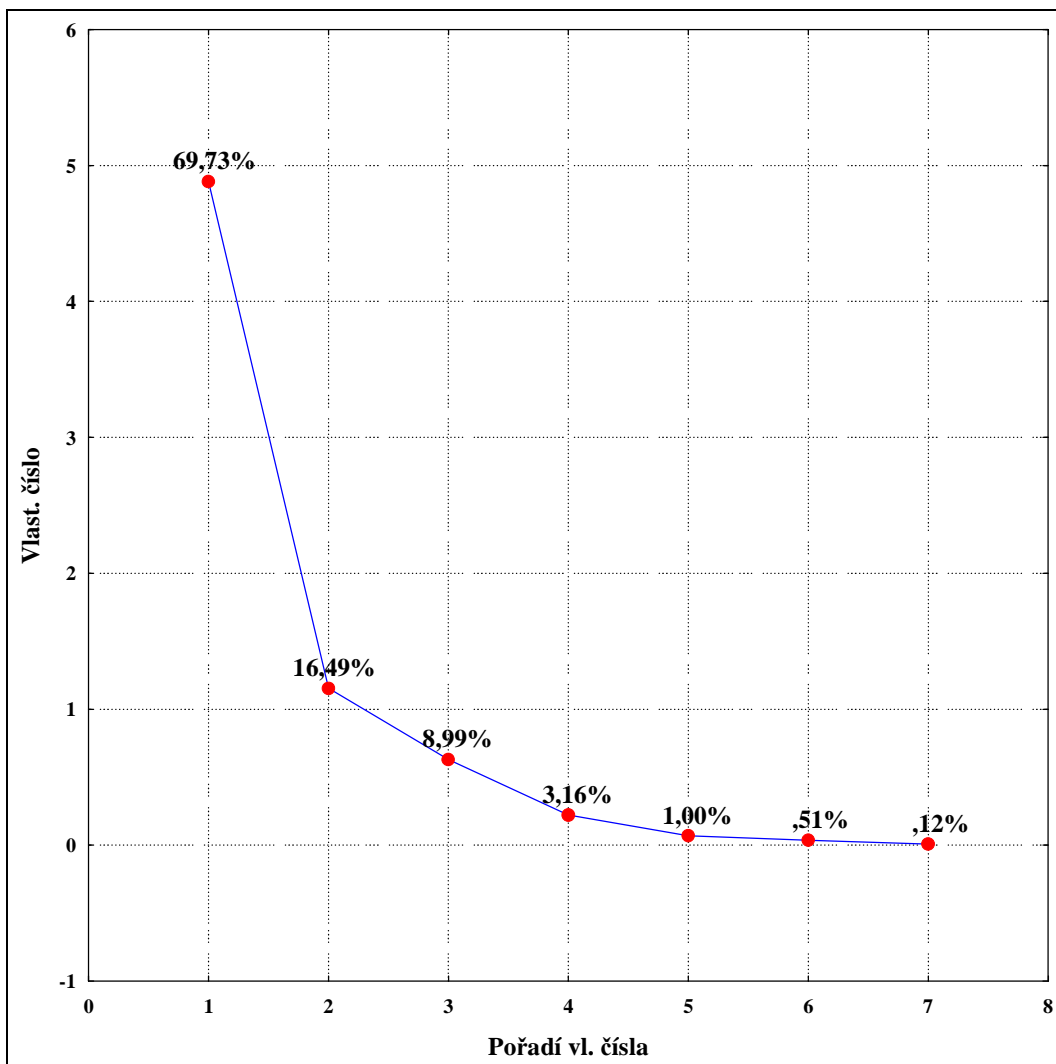
Obr. 11 Rozpt. diagram komp. skóre 2 a 3

### 9.6.2 Komentář k vybraným grafům grafického výstupu programu STATISTICA (před redukcí počtu znaků)

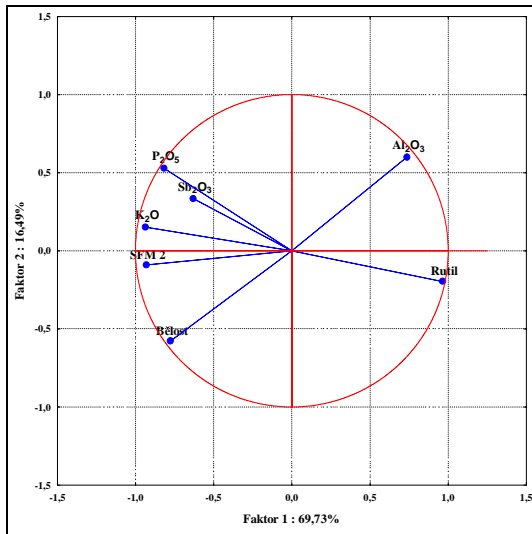
- Cattelův indexový graf úpatí vlastních čísel (obr. 5)*  
 Z tohoto grafu vyplývá, že postačí použít první 2, příp. 3 komponenty, čímž bude pokryto 88,12, resp. 94,44 % celkové proměnlivosti v datech.
- Graf komponentních vah a graf komponentního skóre pro faktory 1 a 2, 1 a 3, 2 a 3 (obr. 6, 7, 8, 9, 10, 11)*  
 Z grafu komponentních vah je patrné, že „proti sobě“ míří znaky Rutil na jedné straně a  $P_2O_5$ ,  $K_2O$ ,  $Sb_2O_3$  a částečně i SFM 2 na straně druhé, obdobně je tomu v relaci  $Al_2O_3 \times$  Bělóst. To je pochopitelné, poněvadž koncentrace uvedených prvků jsou specifické pro ten který druh kalcinátu. Hodnoty koncentrace rutilu a rovněž hodnoty SFM 2 umožní rozlišit anatasové a rutilové typy. Stejnou vypovídací schopnost mají barevné filtry a znak bělost  
 V grafu komponentního skóre je možno spatřit 4 shluky bodů odpovídající 4 základním druhům kalcinátu.  
 Obdobné poznatky je možno učinit z grafů pro dvojice faktorů 1 a 3, příp. 2 a 3, i když proměnlivost v datech vysvětlená těmito faktory bude nižší.  
 Na základě výše uvedených poznatků bude zredukován počet znaků (vypuštěny budou barevné filtry).



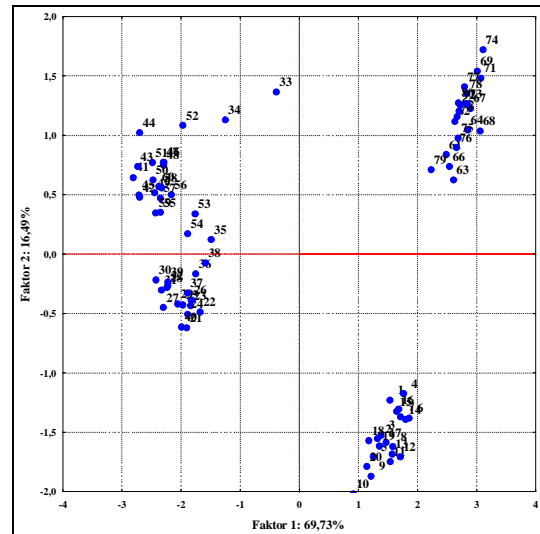
9.6.3 Vybrané grafy grafického výstupu programu STATISTICA (po 1. redukci počtu znaků)



Obr. 12 Cattelův indexový graf úpatí vlastních čísel



Obr. 13 Graf kom. vah 1 a 2

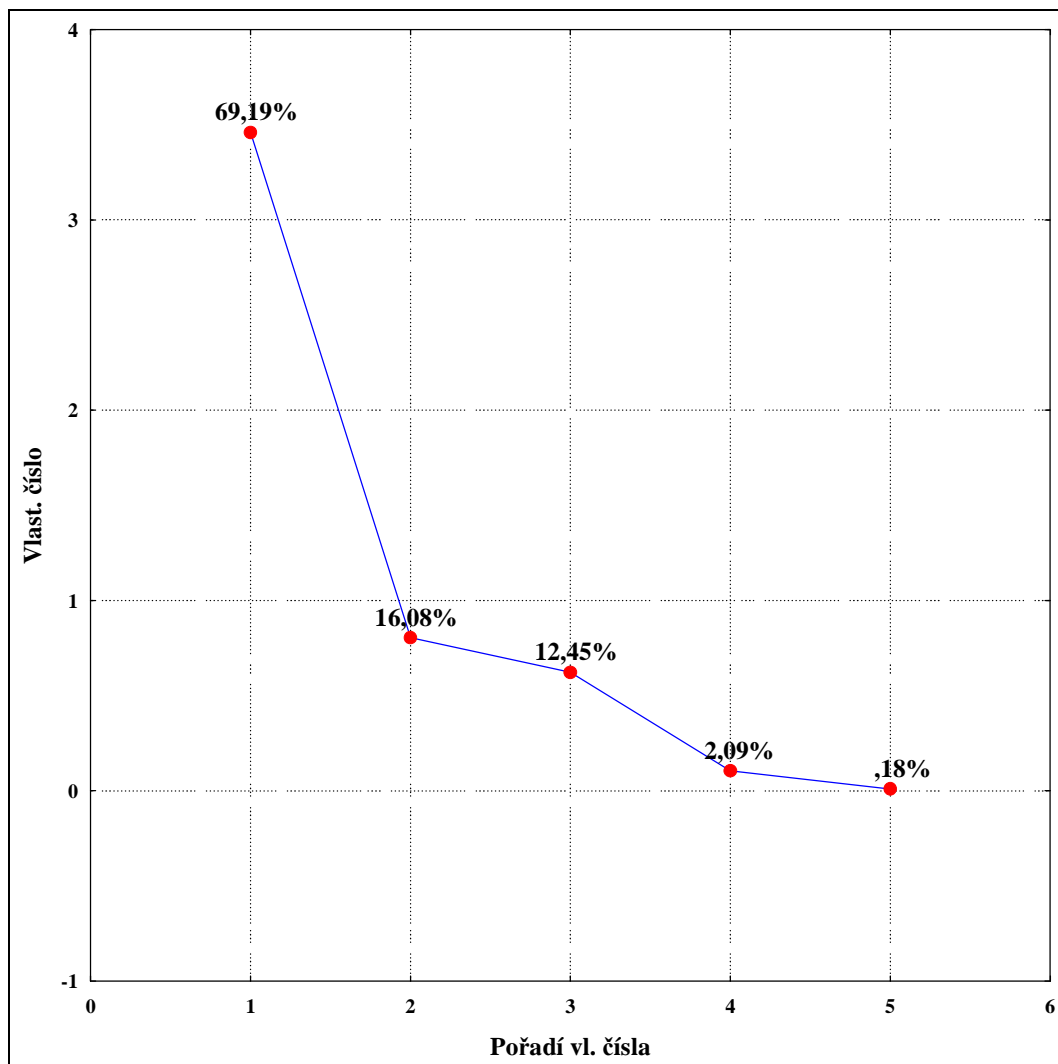


Obr. 14 Rozpt. diagram kom. skóre 1 a 2

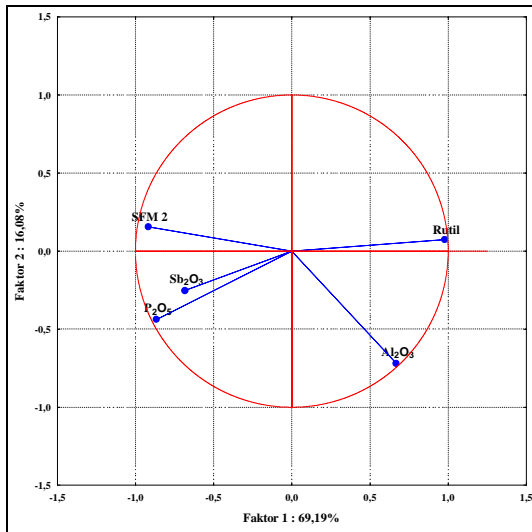
#### 9.6.4 Komentář k vybraným grafům grafického výstupu programu STATISTICA (po 1. redukci počtu znaků)

- *Cattelův indexový graf úpatí vlastních čísel (obr. 12)*  
Z tohoto grafu vyplývá, že postačí použít první 2 komponenty, čímž bude pokryto 86,22 % celkové proměnlivosti v datech.
- *Graf komponentních vah a graf komponentního skóre pro faktory 1 a 2 (obr. 13, 14)*  
Z grafu komponentních vah po redukci počtu znaků plynou podobné závěry jako před redukcí.  
V grafu komponentního skóre je možno opět spatřit 4 shluky bodů odpovídající 4 základním druhům kalcinátu.  
Na základě výše uvedených poznatků bude provedena další redukce počtu znaků (vypuštěny budou znaky Bělou a K<sub>2</sub>O).

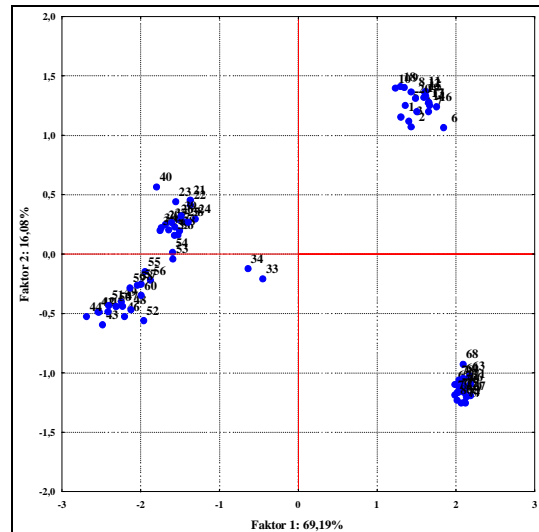
9.6.5 Vybrané grafy grafického výstupu programu STATISTICA (po 2. redukci počtu znaků)



Obr. 15 Cattelův indexový graf úpatí vlastních čísel



Obr. 16 Graf komp. vah 1 a 2



Obr. 17 Rozpt. diagram komp. skóre 1 a 2

### 9.6.6 Komentář k vybraným grafům grafického výstupu programu STATISTICA (po 2. redukci počtu znaků)

- *Cattelův indexový graf úpatí vlastních čísel (obr. 15)*  
Z tohoto grafu vyplývá, že postačí použít první 2 komponenty, čímž bude pokryto 85,27 % celkové proměnlivosti v datech. Ani 2. redukci počtu proměnných nedochází k výraznému snížení popisu proměnlivosti ve znacích.
- *Graf komponentních vah a graf komponentního skóre pro faktory 1 a 2 (obr. 16, 17)*  
Z grafu komponentních vah je patrné, že znaky  $P_2O_5$  a  $Sb_2O_3$  poskytují obdobnou informaci. Zdánlivě menší důležitost vykazuje znak  $Sb_2O_3$ , jeho odstranění by však znemožnilo od sebe rozlišit 2 druhy anatasových kalcinátů v grafu komponentního skóre.  
V grafu komponentního skóre je možno opět spatřit 4 shluky bodů odpovídající 4 základním druhům kalcinátu.

## 9.7 Závěr

Aplikací metody PCA na předložená data bylo zjištěno, že reprezentují 4 hlavní druhy kalcinátů. K popisu daného výběru postačují znaky SFM 2, rutil,  $P_2O_5$ ,  $Sb_2O_3$ ,  $Al_2O_3$ . Další redukce počtu znaků by znemožnila od sebe odlišit některé druhy kalcinátů.

Komponenta 1 vystihuje krystalovou modifikaci kalcinátu, komponenta 2 umožňuje rozlišit jednotlivé druhy rutilových a anatasových kalcinátů mezi sebou, představuje kombinaci několika znaků, a proto ji nelze jednoznačně pojmenovat.