



## SEMESTRÁLNÍ PRÁCE

# Matematické principy analýzy vícerozměrných dat

2005/2006

Ing. Petr Eliáš



## 1. OTÁZKA 1

### 1.1 Zadání

Najděte vlastní (charakteristická) čísla, determinat a stopu od této jednoduché matice

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 8 & 2 \\ 3 & 2 & 10 \end{pmatrix}$$

### 1.2 Řešení a závěr

- *Vlastní (charakteristická) čísla*

Pro vlastní čísla platí:

$$\det(A - \lambda E) = 0$$

V našem případě tedy:

$$\det \begin{pmatrix} 1 - \lambda & 2 & 3 \\ 2 & 8 - \lambda & 2 \\ 3 & 2 & 10 - \lambda \end{pmatrix} = 0$$

Po vyčíslení dostáváme kubickou rovnici:

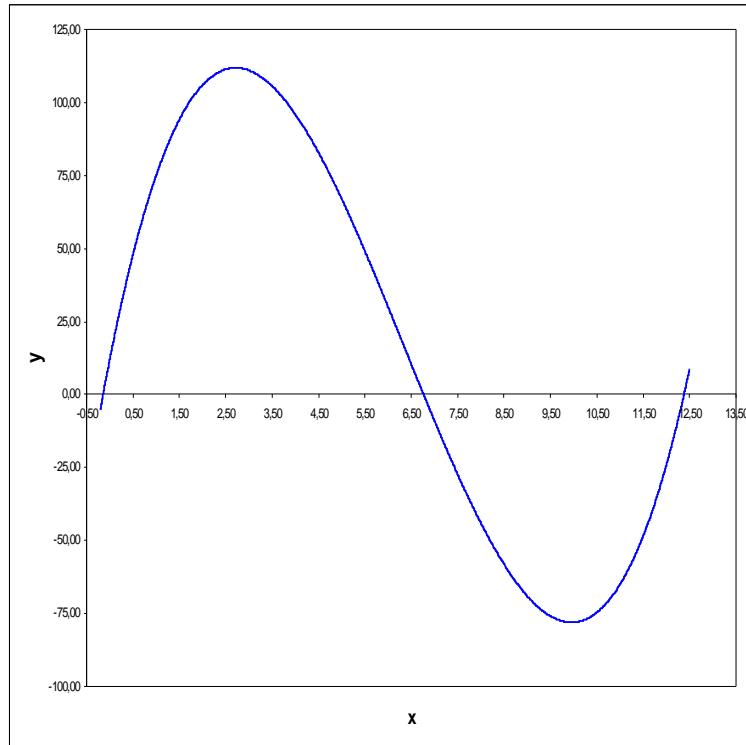
$$\lambda^3 - 19\lambda^2 + 81\lambda - 12 = 0$$

Grafickým řešením (viz. obr. 1) získáváme vlastní čísla

$$\lambda_1 = -0,143$$

$$\lambda_2 = 6,765$$

$$\lambda_3 = 12,378$$



**Obr. 1** Průběh křivky  $y = \lambda^3 - 19\lambda^2 + 81\lambda + 12$

- **Determinant**

$$\det(A) = 1 \cdot \det \begin{pmatrix} 8 & 2 \\ 2 & 10 \end{pmatrix} - 2 \cdot \det \begin{pmatrix} 2 & 2 \\ 3 & 10 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 2 & 8 \\ 3 & 2 \end{pmatrix}$$

$$\det(A) = 1 \cdot 8 \cdot 10 - 1 \cdot 2 \cdot 2 + 2 \cdot 2 \cdot 10 - 2 \cdot 2 \cdot 3 + 3 \cdot 2 \cdot 2 - 3 \cdot 8 \cdot 3$$

$$\det(A) = -12$$

- **Stopa matice**

Stopa čtvercové matice je definována jako součet diagonálních prvků, tedy

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}$$

$$\text{tr}(A) = 1 + 8 + 10$$

$$\text{tr}(A) = 19$$

## 2. OTÁZKA 2

### 2.1 Zadání

Pro typická data z vašeho pracoviště, minimálně 4 rozměrná, určete projekci do komponent, dvojný graf (resp. graf komponentních vah a rozptylový diagram komponentního skóre) a diskutujte jeho význam.

### 2.2 Úvod

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.

### 2.3 Data

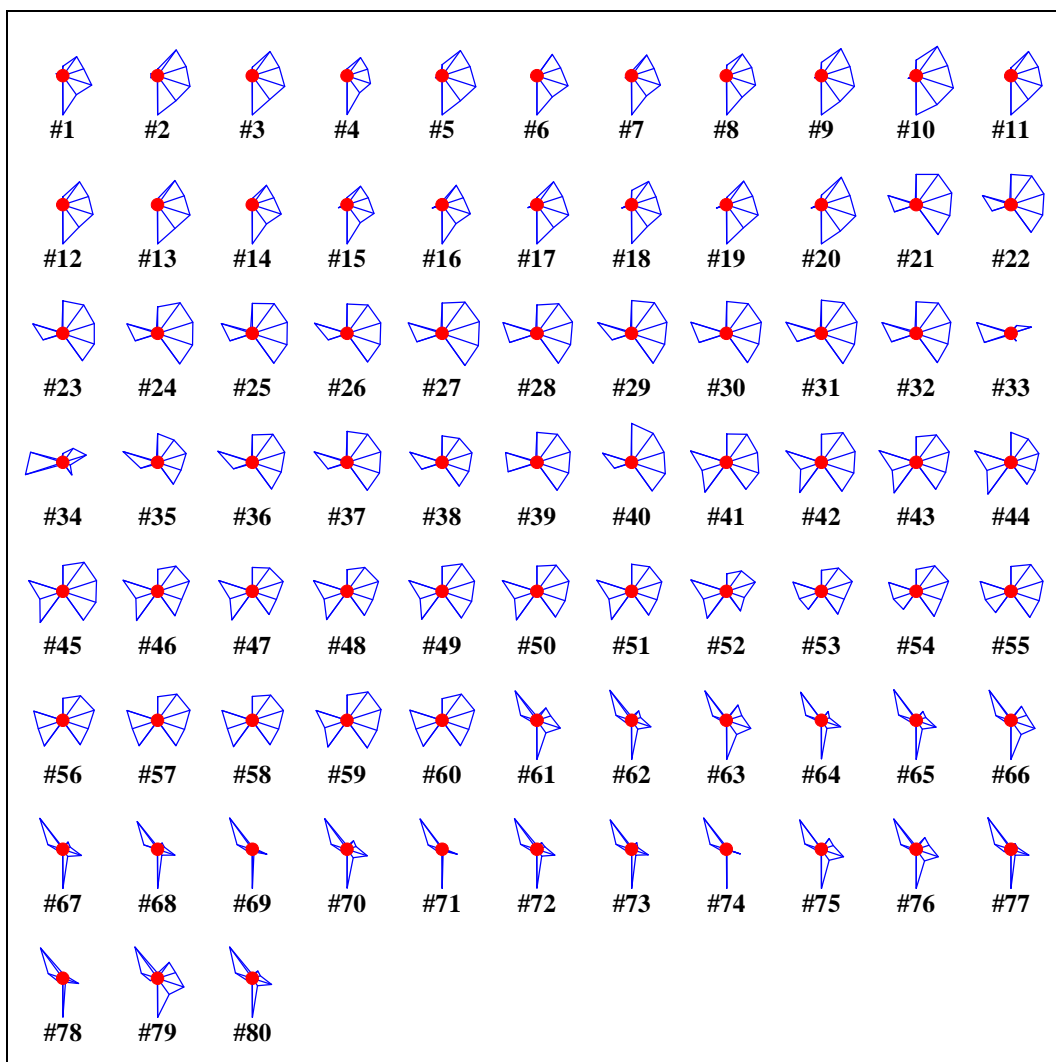
Vzorek	SFM2	ZF	MF	ČF	Bělost	Rutil	Sb <sub>2</sub> O <sub>3</sub>	K <sub>2</sub> O	P <sub>2</sub> O <sub>5</sub>	Al <sub>2</sub> O <sub>3</sub>
1	126,0	96,7	95,8	97,9	96,8	99,8	0,003	0,161	0,196	0,009
2	116,1	97,9	97,4	98,4	97,9	100,0	0,004	0,165	0,197	0,013
3	123,0	97,6	97,0	98,4	97,7	100,0	0,003	0,163	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
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

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>
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

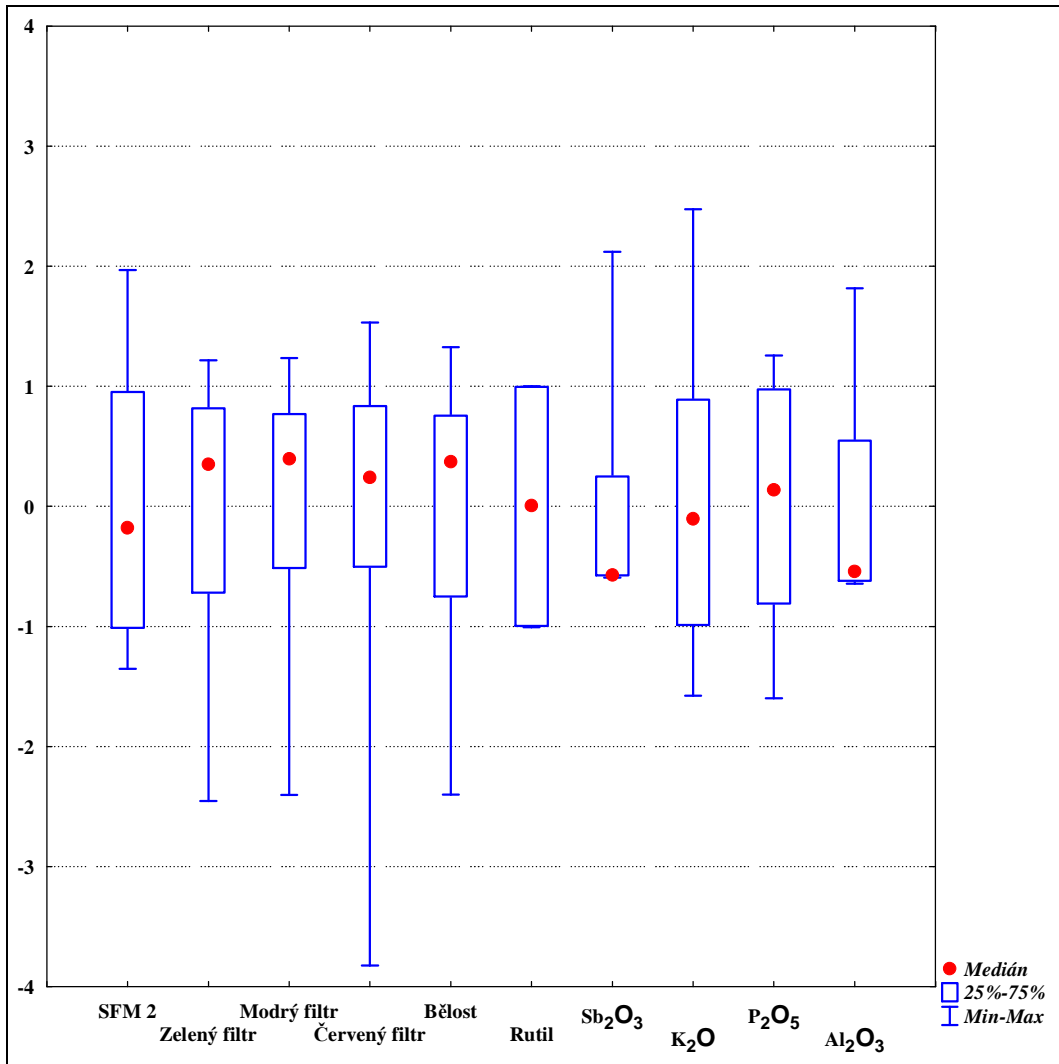
Soubor s daty:           Kalcinyaty.xls  
Použitý software:       STATISTICA v7.1

## 2.4 Analýzy variability znakov

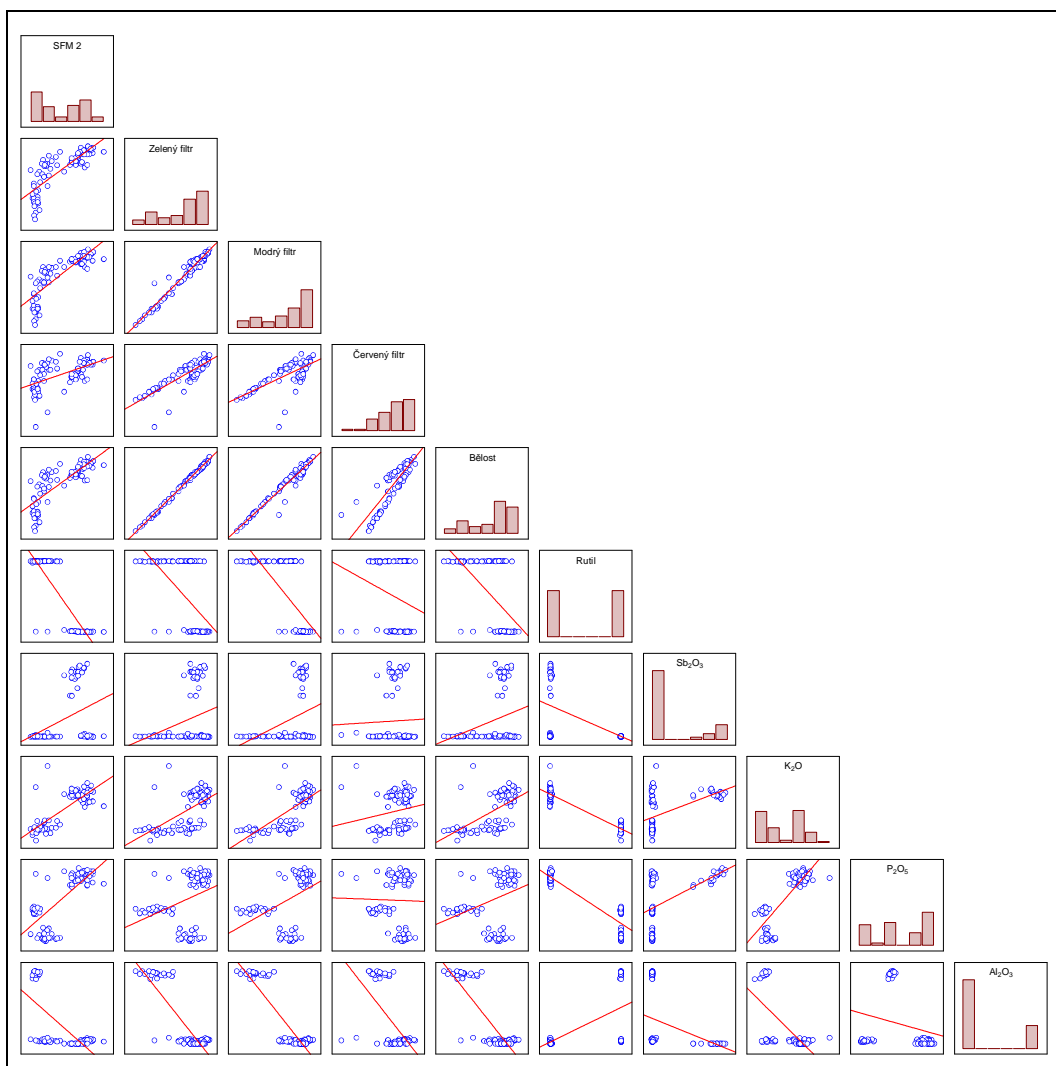
### 2.4.1 Vybrané grafy grafického výstupu programu STATISTICA



Obr. 2 Ikonový graf (hvězdy)



Obr. 3 Krabicový graf pro standardizovaná data



**Obr. 4** Graf korelace

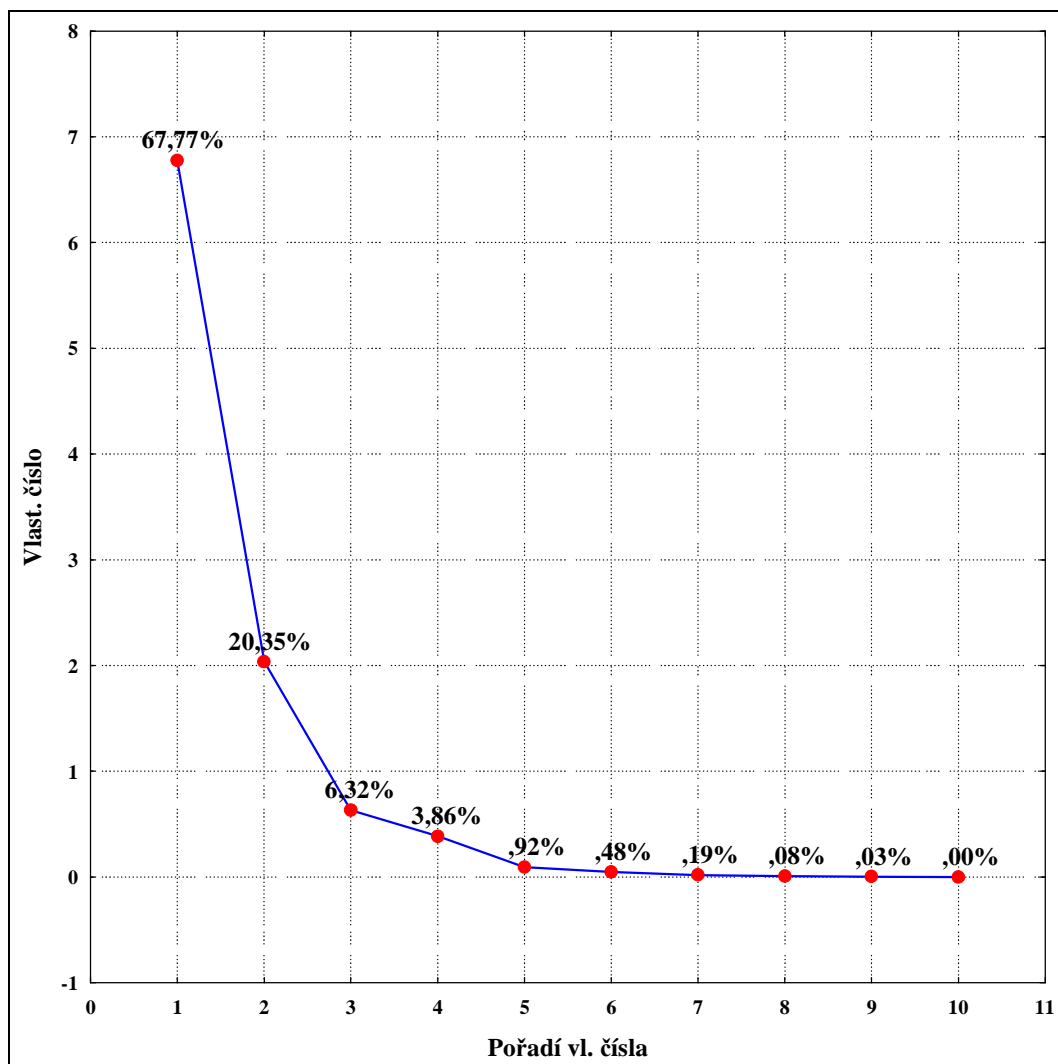
#### 2.4.2 *Komentář k vybraným grafům grafického výstupu programu STATISTICA*

- *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 Sb<sub>2</sub>O<sub>3</sub>.
- *Graf korelace (obr. 4)*  
Z předmětného grafu je patrná silná lineární závislost mezi znaky SFM 2, bělost, zelený, modrý a červený filtr.

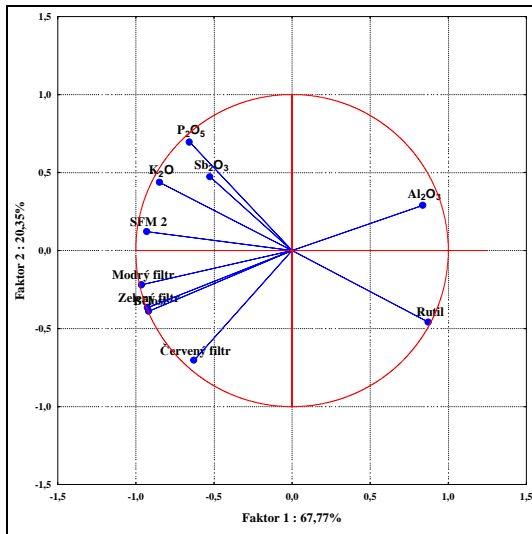


## 2.5 Metoda hlavních komponent (PCA)

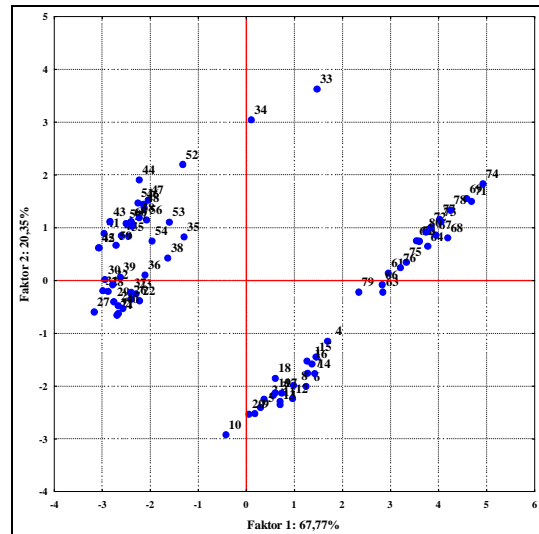
### 2.5.1 Vybrané grafy grafického výstupu programu STATISTICA (před redukcí počtu znaků)



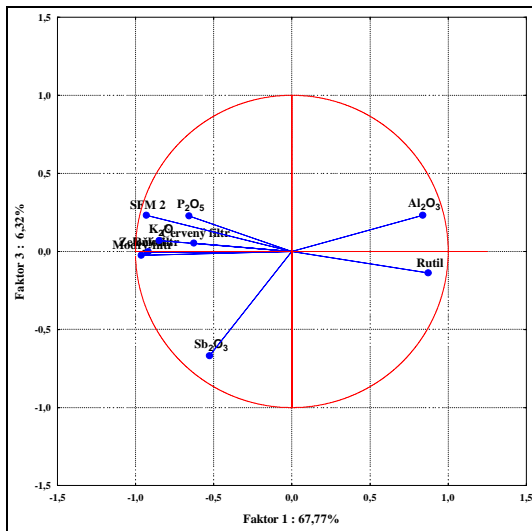
Obr. 5 Cattelův indexový graf úpatí vlastních čísel



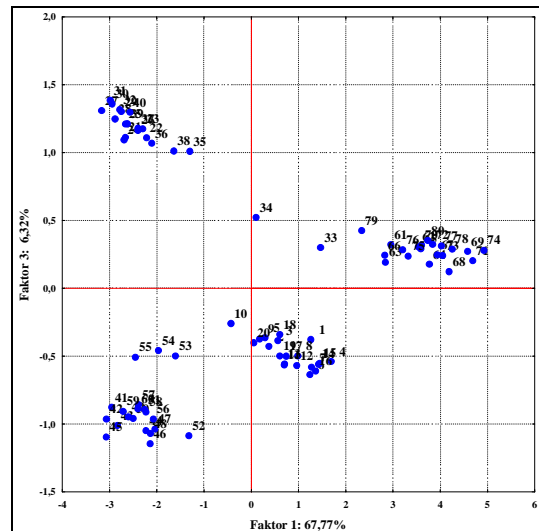
Obr. 6 Graf komp. vah 1 a 2



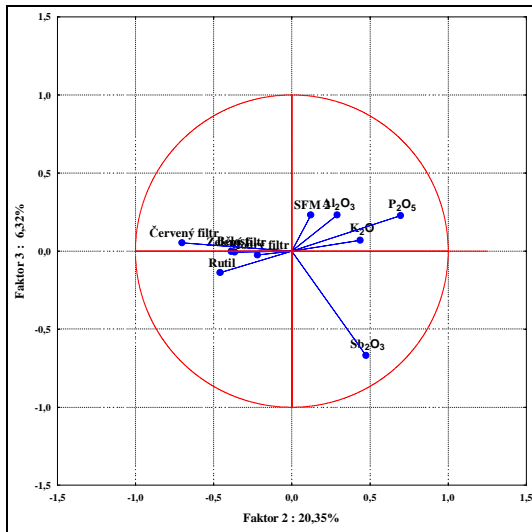
Obr. 7 Rozpt. diagram komp. skóre 1 a 2



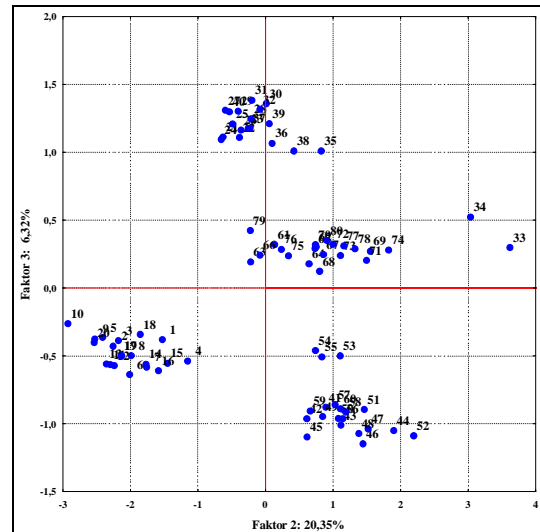
Obr. 8 Graf komp. vah 1 a 3



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



Obr. 10 Graf komp. vah 2 a 3

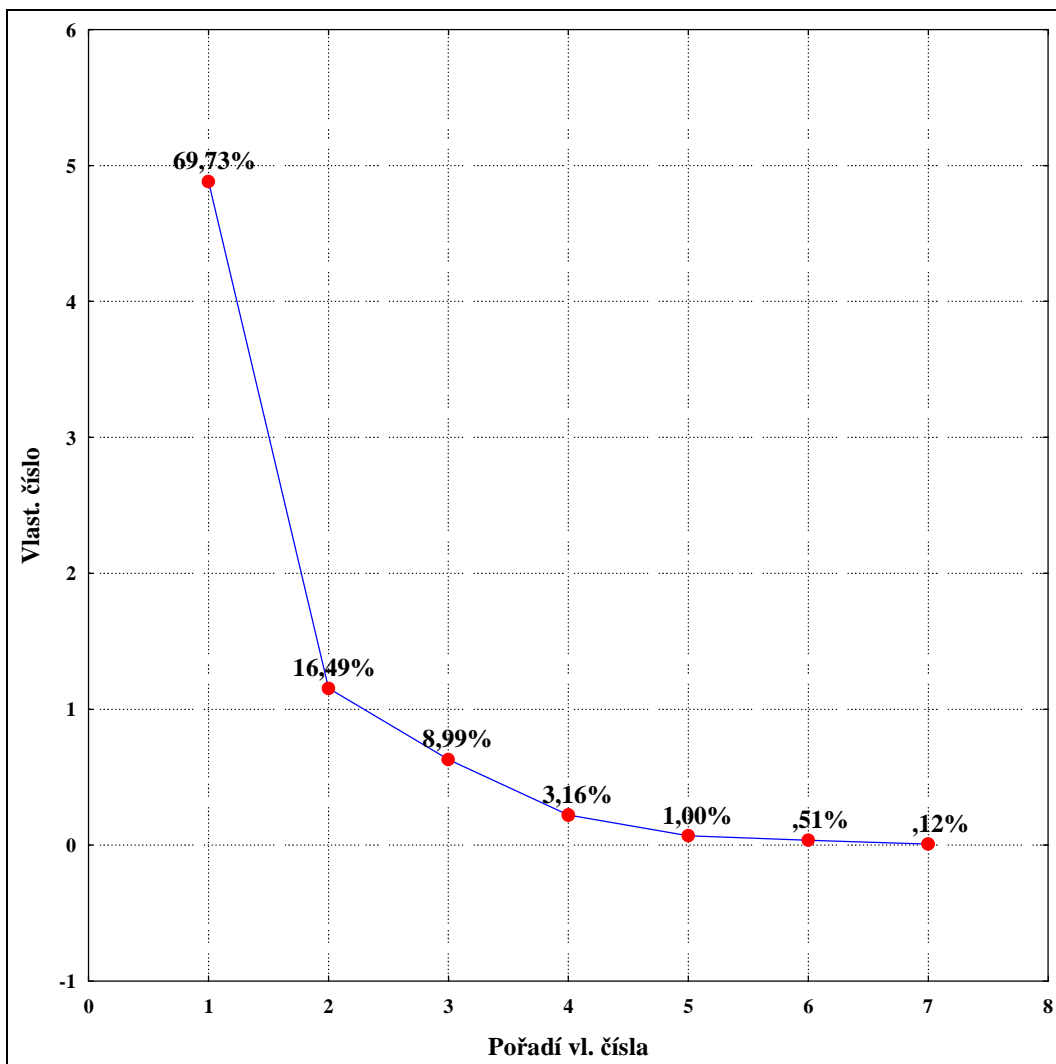


Obr. 11 Rozpt. diagram komp. skóre 2 a 3

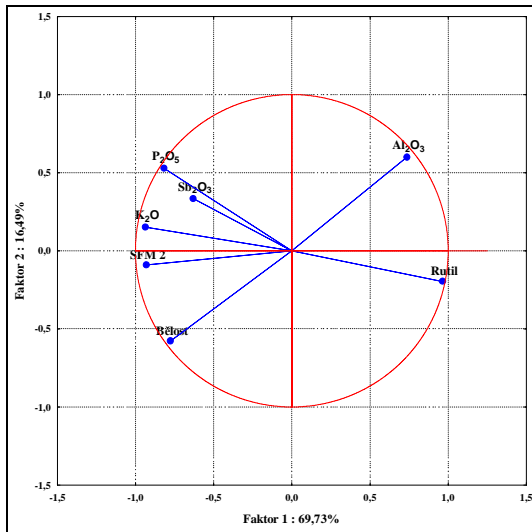
### 2.5.2 Komentář k vybraným grafům grafického výstupu programu STATISTICA (před redukcí počtu znaků)

- 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).

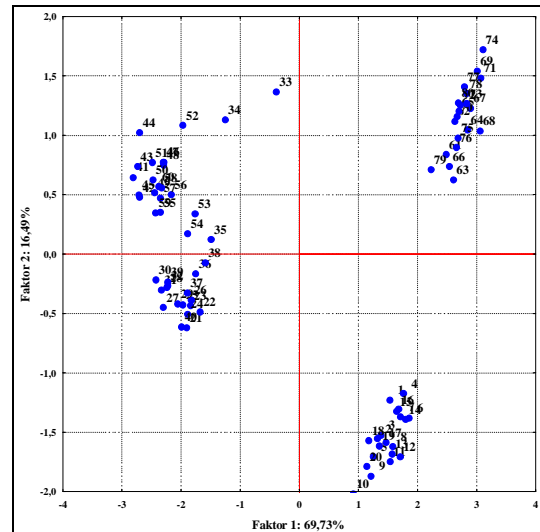
2.5.3 Vybrané grafy grafického výstupu programu STATISTICA (po 1. redukci počtu znaků)



Obr. 12 Cattelův indexový graf úpatí vlastních čísel



Obr. 13 Graf kom. vah 1 a 2

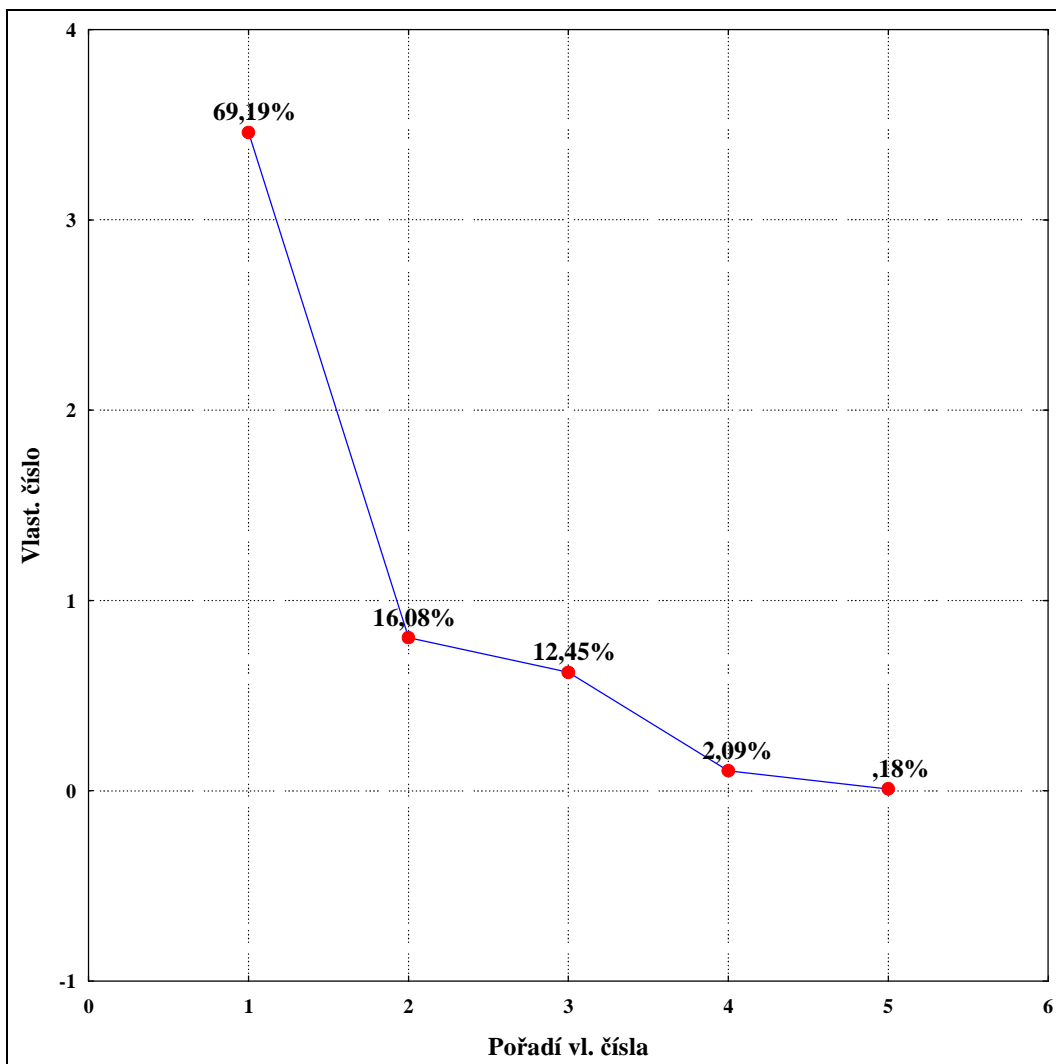


Obr. 14 Rozpt. diagram kom. skóre 1 a 2

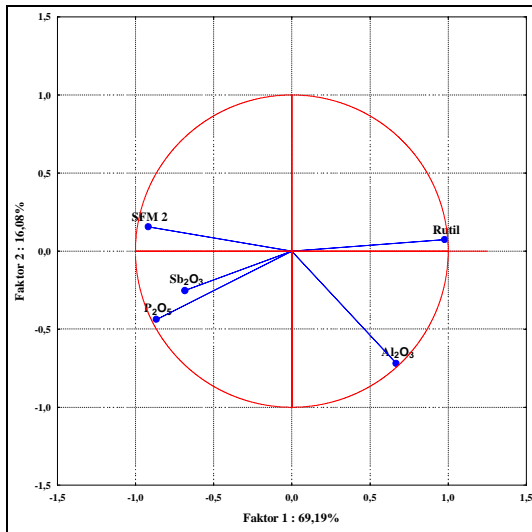
#### 2.5.4 Komentář k vybraným grafům grafického výstupu programu STATISTICA (po 1. redukci počtu znaků)

- *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).

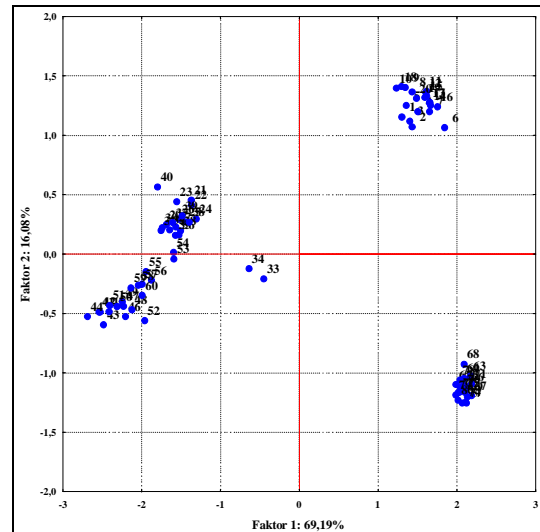
2.5.5 Vybrané grafy grafického výstupu programu STATISTICA (po 2. redukci počtu znaků)



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

### 2.5.6 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.

## 2.6 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.