

LICENČNÍ STUDIUM STATISTICKÉ ZPRACOVÁNÍ DAT PŘI MANAGEMENTU JAKOSTI

PŘEDMĚT:

3.4 Počítačová analýza vícerozměrných dat

Vypracoval: Ing. Adolf Goebel, Ph.D.

V Přerově, 2. května 2001

3.4.1 DISPERGACE MLETÉHO KALCINÁTU TITANOVÉ

BĚLOBY RUTILOVÉHO TYPU

3.4.1.1 ÚVOD

Materiál pigmentových vlastností se získává kalcinací hydratovaného gelu oxidu titaničitého při teplotách kolem 950°C. Výpad z kalcinační pece, takzvaný kalcinát, se po zchlazení nejprve mele za sucha na běžných mlýnech pro sypké materiály. Jednou z důležitých technologických operací, která vysoce zlepšuje užité vlastnosti titanové běloby, je takzvané mokré mletí. Při této operaci se vysoce koncentrovaná suspenze připravená z mletého kalcinátu mele v takzvaných mokrých mlýnech pomocí malých keramických kuliček. Nezbytným předpokladem mokrého mletí je příprava suspenze vhodných rheologických vlastností. K posouzení reprodukovatelnosti přípravy suspenze standardním postupem byly po určité době odebrány provozně mleté kalcináty, které byly analyzovány a použity k laboratorní přípravě suspenze použitelné k mokrému mletí. U této suspenze byly stanoveny rheologické vlastnosti, konkrétně viskozita a doba výtoku Fordovým kelímkem. Smyslem vyhodnocení je posoudit vzájemné vazby mezi jednotlivými parametry a najít takové charakteristiky, které umožní včasný odhad rheologických vlastností bez toho, aniž by se musela provádět pracná laboratorní dispergace, popř. nalézt kalcináty podobných vyhovujících vlastností, aby se na základě údajů o podmínkách při jejich výrobě mohlo aktivně zasahovat do výrobního procesu.

3.4.1.2 Zpracovávaná data

Použitá data jsou v příloze č. 1. Data je možné rozdělit do dvou skupin. První z nich jsou data získaná pro provozně pomletý kalcinát, z nichž většina se týká obsahu rutilové krystalické fáze a jednotlivých příměsí ze suroviny nebo přísad úmyslně přidaných v průběhu výroby k žádoucí modifikaci dějů nebo výsledných vlastností produktu. Mimo to jsou zde i data z některých speciálních zkoušek, jako např. měrná vodivost a pH vodného výluhu z kalcinátu nebo spotřeba vody, bod tečení a bod smočení, které popisují dosažení určitého chování kalcinátu při přidání minimálního množství vody.

Druhá skupina údajů, konkrétně doba výtoku Fordovým kelímkem a viskozita, se týká rheologických vlastností suspenze připravené standardním postupem dispergace mletého kalcinátu v laboratoři.

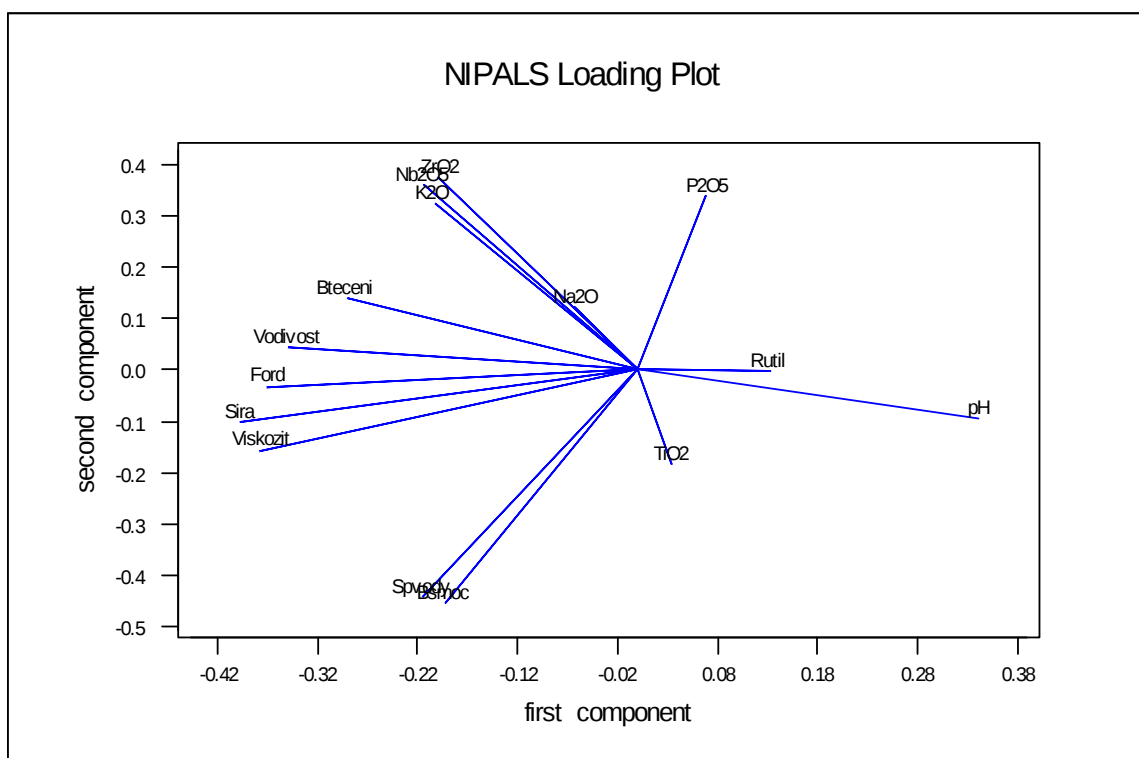
3.4.1.3 Metoda hlavních komponent

Vzhledem k problémům s během programu Statgraphics na počítači byly použity jiné statistické programy. V tomto případě SCAN 1.1. Pro metodu hlavních komponent PCA pomocí algoritmu NIPALS bylo nutno vyřadit některé sloupce, které byly indikovány jako téměř konstantní a bránily výpočtu. Konkrétně se jednalo o % SiO₂, %Fe

a % Sb_2O_3 . Jak ukazují graf komponentních vah (1.1) a dvojný graf (1.2) podobné vlastnosti mají a tím pádem spolu korelují:

- % ZrO_2 , % Nb_2O_5 , Na_2O a % K_2O
- % rutilizace a pH vodného výluhu
- spotřeba vody a bod tečení
- obsah síry, viskozita

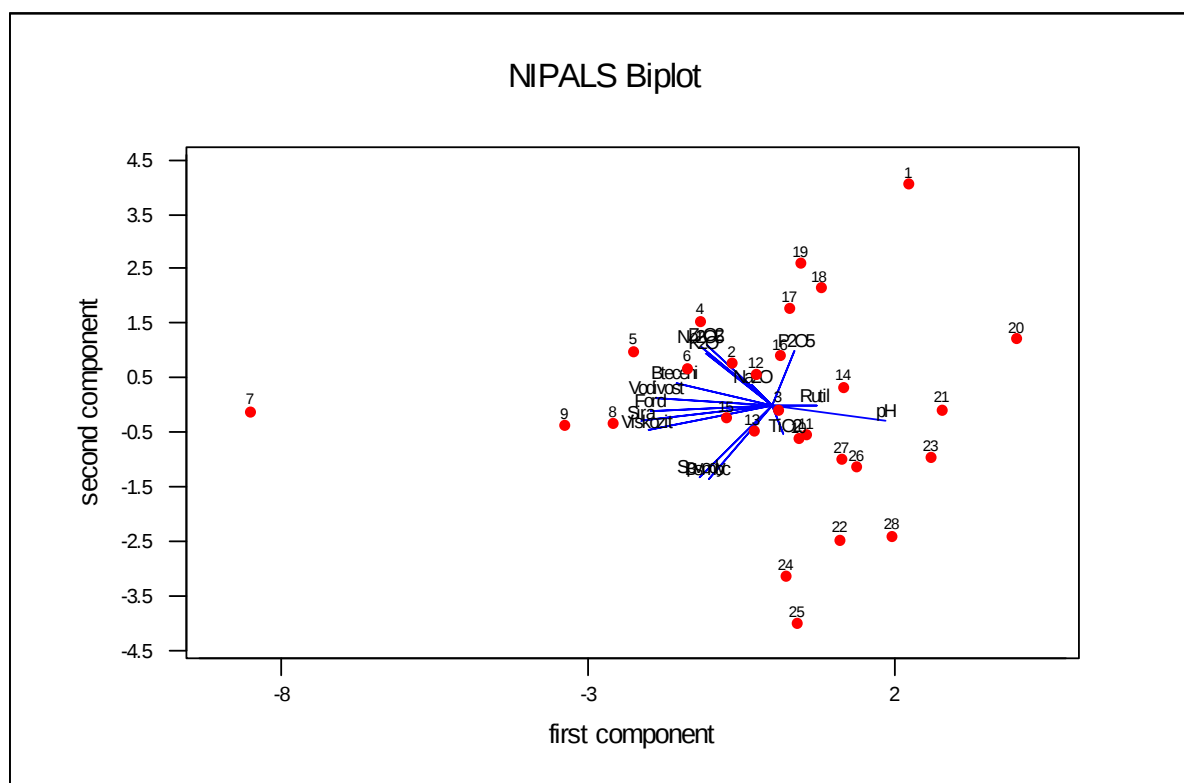
Obr. 1.1 - Graf komponentních vah



Z praktických důvodů byla z první skupiny proměnných ponechána % K_2O , protože tato složka se přidává v technologii definovaným způsobem a může podléhat změnám, zatímco obsahy ZrO_2 a Nb_2O_5 a Na_2O pocházejí ze suroviny a pomocných surovin a příliš se nemění.

Ze druhé skupiny bylo ponecháno pH vodného výluhu, které se podílí na vysvětlení vztahů významněji než rutilizace. Korelace obou parametrů je logická. S vyšší rutilizací se rozloží větší množství vázané kyseliny sírové a některých síranových komplexů, čímž se zvýší pH výluhu (jeho nižší kyselost).

Ze třetí skupiny byl, jako snáze dostupná hodnota vybrán bod tečení, ze čtvrté obsah síry.



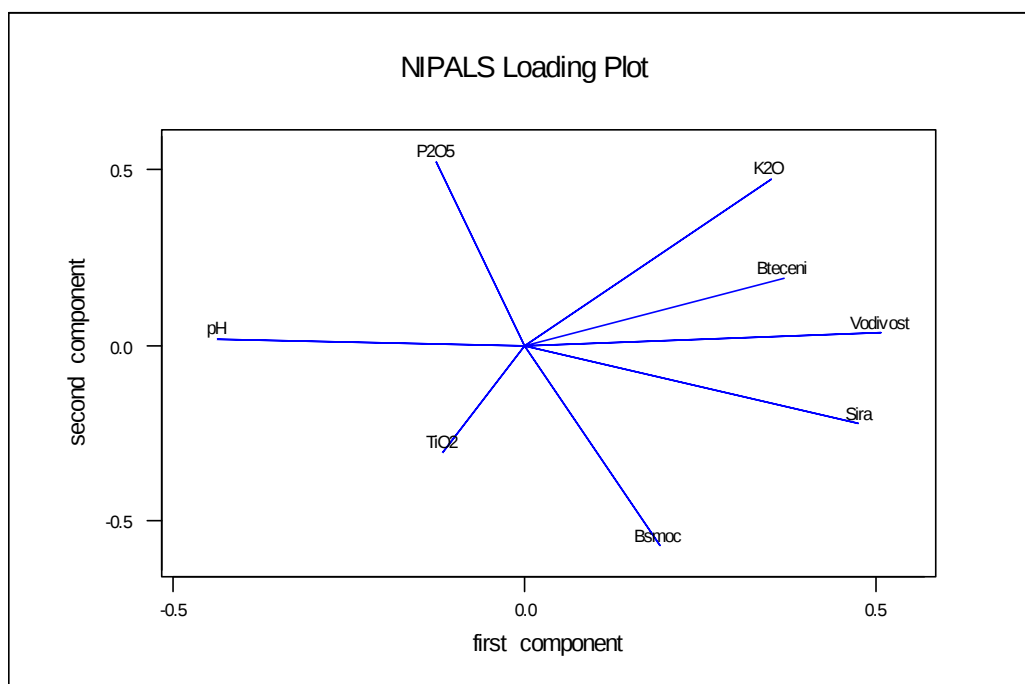
Obr. 1.2 - Dvojný graf

Před popsanou redukcí proměnných byl stav následující:

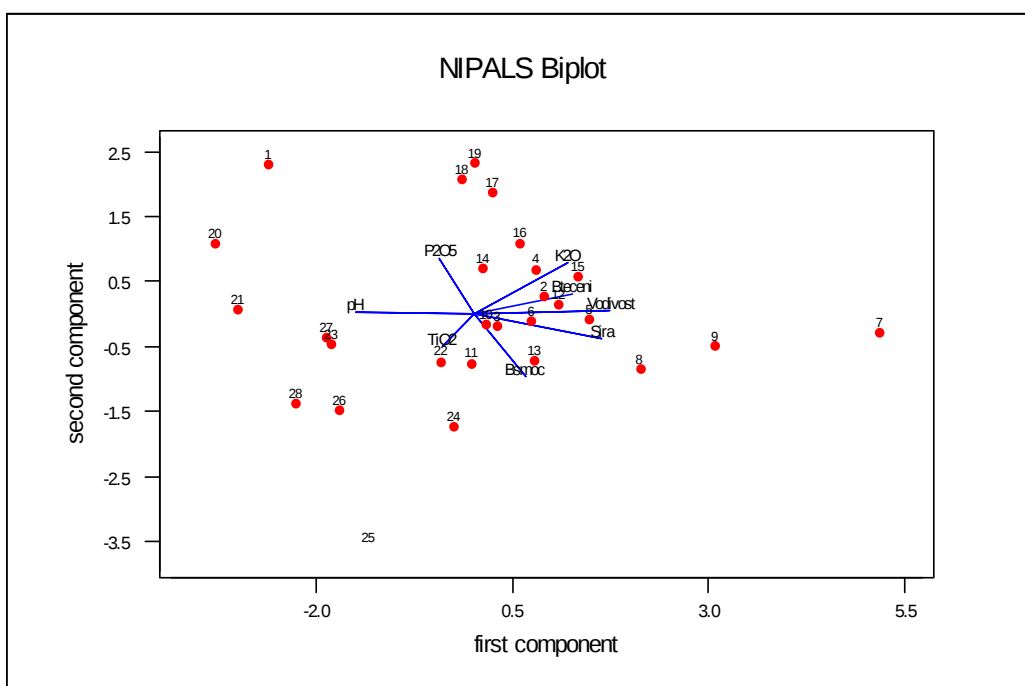
Tabulka I-1: Stav před redukcí počtu proměnných

Hlavní komponenta	Eigenvalue	Proporčně	Kumulativně
1	5,4261	0,362	0,362
2	2,9724	0,198	0,560
3	1,5190	0,101	0,661
4	1,3144	0,088	0,749
5	0,9748	0,065	0,814
6	0,8445	0,056	0,870
7	0,6663	0,044	0,914
8	0,4818	0,032	0,947
9	0,2994	0,020	0,967
10	0,2248	0,015	0,982

Po redukci počtu proměnných se situace změnila následovně:



Obr
Gra
kom
ntní
vah
prv
red
počt
pro
ých



. 1.3 -
f
pone
ch
po
ní
ukci
u
měnn

Obr. 1.4 - Dvojný graf po první redukci počtu proměnných

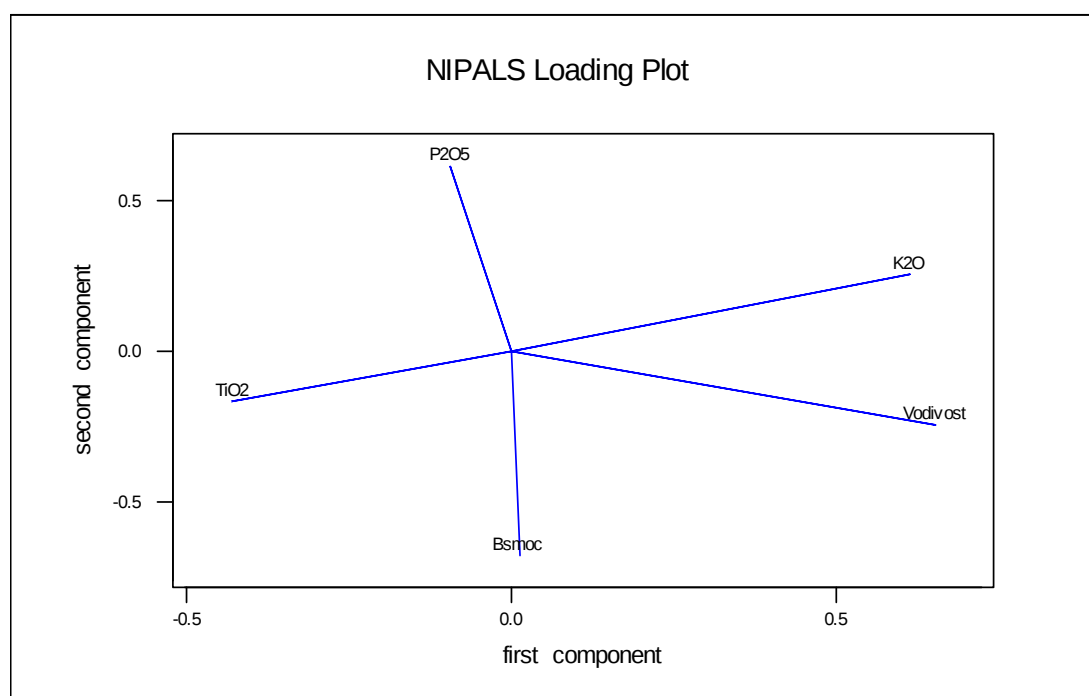
Numerické výpočty přinesly následující hodnoty:

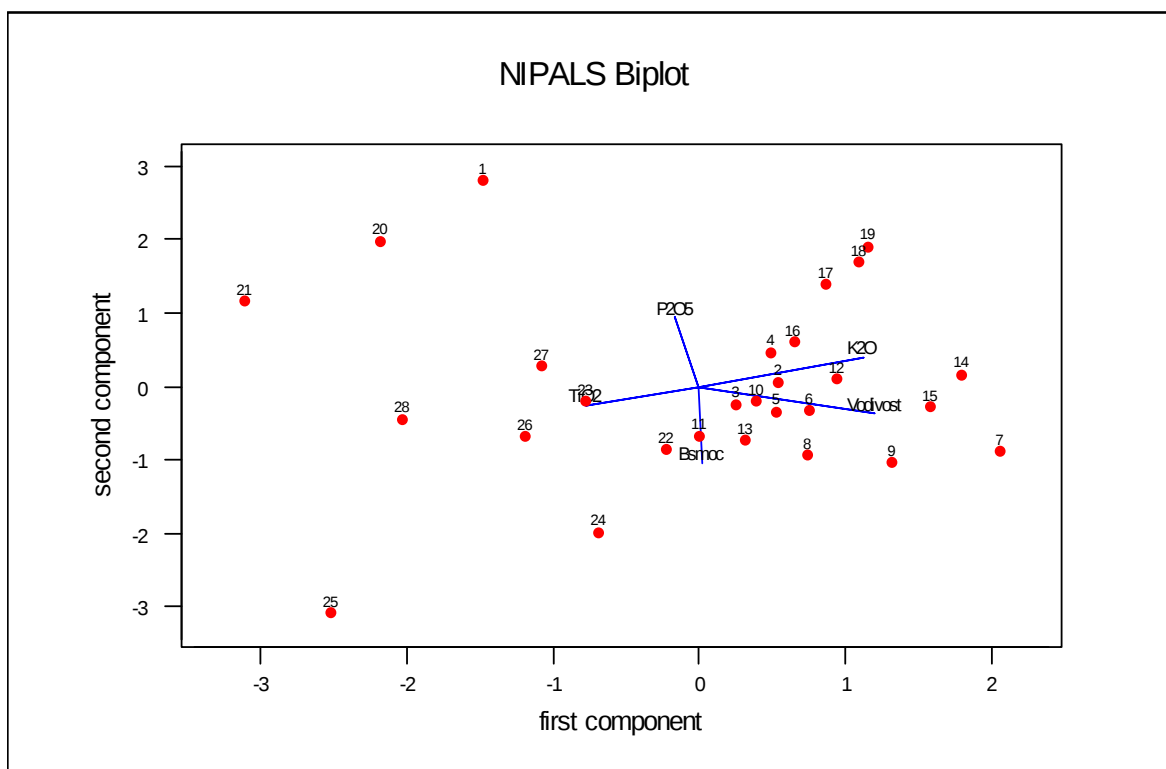
Tabulka I-2: Stav po první redukci počtu proměnných

Hlavní komponenta	Eigenvalue	Proporčně	Kumulativně
1	3,4325	0,429	0,429
2	1,6725	0,209	0,638
3	1,1548	0,144	0,782
4	0,6306	0,079	0,861
5	0,4956	0,062	0,923
6	0,3230	0,040	0,964
7	0,1695	0,021	0,985
8	0,1215	0,015	1,000

Dalším krokem bylo nahrazení bodu tečení a obsahu síry a pH vodivostí. V případě vyřazení pH bylo přihlédnuto k velmi významné korelaci mezi vodivostí a pH (korelační koeficient -0,7369). To se v příslušných grafech promítá tím, že obě veličiny leží na společné přímce a každá z nich směřuje na opačnou stranu. K dalšímu hodnocení tedy zbyly pouze % K_2O , % P_2O_5 , % TiO_2 , bod smočení a vodivost. Grafický výstup je následující:

Obr. 1.5 - Graf komponentních vah po druhé redukci počtu proměnných





Obr. 1.6 - Dvojný graf po druhé redukci počtu proměnných

Tabulka I-3: Stav po druhé redukci počtu proměnných

Hlavní komponenta	Eigenvalue	Proporčně	Kumulativně
1	1,8227	0,365	0,365
2	1,5201	0,304	0,669
3	0,9141	0,183	0,851
4	0,5560	0,111	0,963
5	0,1871	0,037	1,000

Závěr:

Jak ukazuje porovnání jednotlivých tabulek, podařilo se druhou redukcí počtu proměnných dosáhnout pro dvě latentní proměnné maxima vysvětlené variability 66,9 %. Latentní proměnné y_1 a y_2 mají následující tvar:

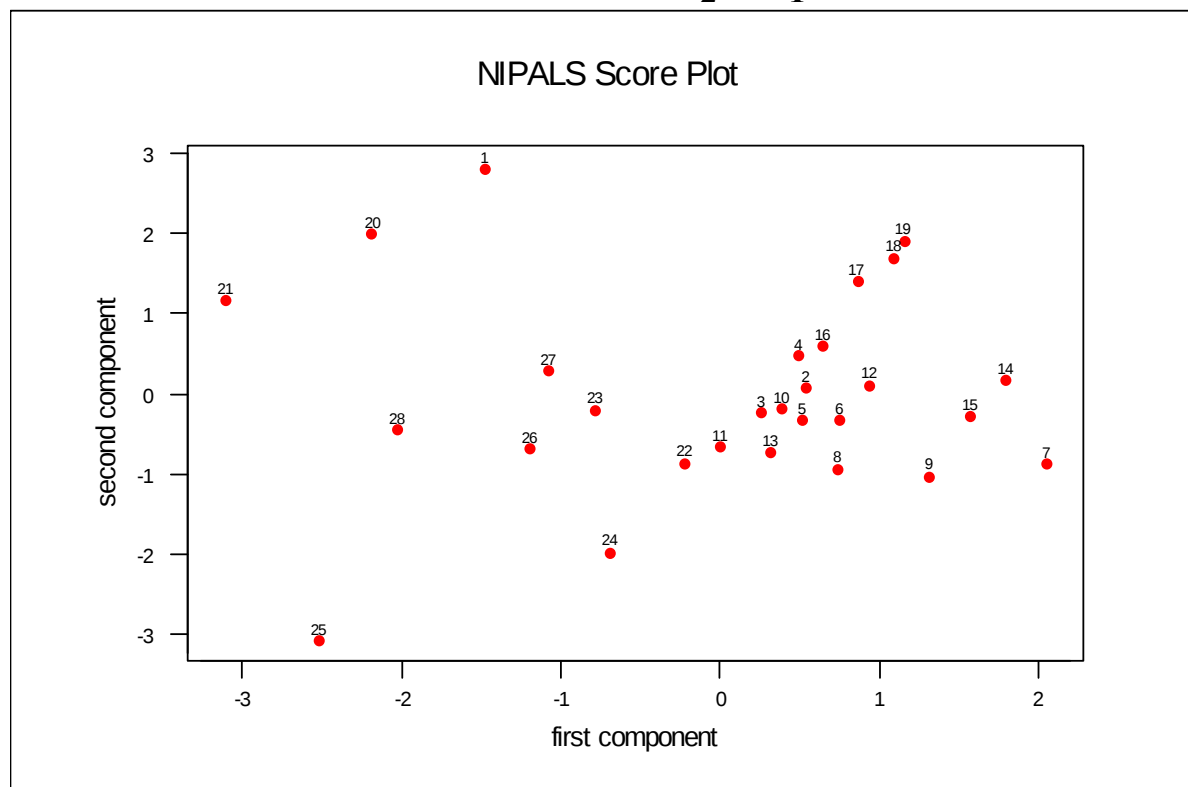
$$y_1 = 0,615 * \%K_2O - 0,095 * \%P_2O_5 - 0,431 * \%TiO_2 + 0,653 * \text{vodivost} + 0,013 * \text{bod smočení}$$

$$y_2 = 0,257 * \%K_2O + 0,619 * \%P_2O_5 - 0,167 * \%TiO_2 - 0,248 * \text{vodivost} - 0,679 * \text{bod smočení}$$

Následující obrázek ukazuje rozložení jednotlivých bodů v grafu při použití uvedených

dvou proměnných:

Obr. 1.7 - Graf znázorňující závislost y_2 na y_1 pro jednotlivé body



Jak ukazuje graf, objekty byly rozděleny do dvou až tří shluků. Pokud bychom se snažili o fyzikální vysvětlení latentních proměnných y_1 a y_2 , dalo by se říci, že první proměnná zvýrazňuje především parametry související s vyluhovatelností pigmentu vodou. Pro to svědčí nejvyšší koeficienty u % K_2O , který představuje vesměs vyluhovatelné draselné soli a u vodivosti, která je daná množstvím vyloužených solí. Obsah TiO_2 souvisí s přítomností rozpustných i nerozpustných solí, oba typy jej snižují. Málo významné jsou koeficienty pro % P_2O_5 , který je nerozpustný a pro bod smočení, který je typickou charakteristikou hydrofility (hydrofobity) povrchu.

V případě y_2 jsou nejvyšší koeficienty u obsahu P_2O_5 , pro který je známo, že má velký vliv na tvrdost a texturu pigmentu a u bodu smočení, který je, jak již bylo uvedeno, typickou charakteristikou hydrofility (hydrofobity) povrchu.

3.4.1.4 Faktorová analýza

Při faktorové analýze bylo využito redukce počtu proměnných provedené pomocí PCA. Byly vypočteny komunalita při použití pěti, čtyř, tří a dvou faktorů. Pro rotaci bylo využito metody EQUIMAX, která spojuje kritéria metody Varimax a Quartimax, což znamená, že při rotaci je maximalizován rozptyl čtverců faktorových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. Výpočet byl proveden pro volbu počtu faktorů 1 až 5 s následujícím výsledkem:

Tabulka I-4: Komunalita pro zvolený počet faktorů

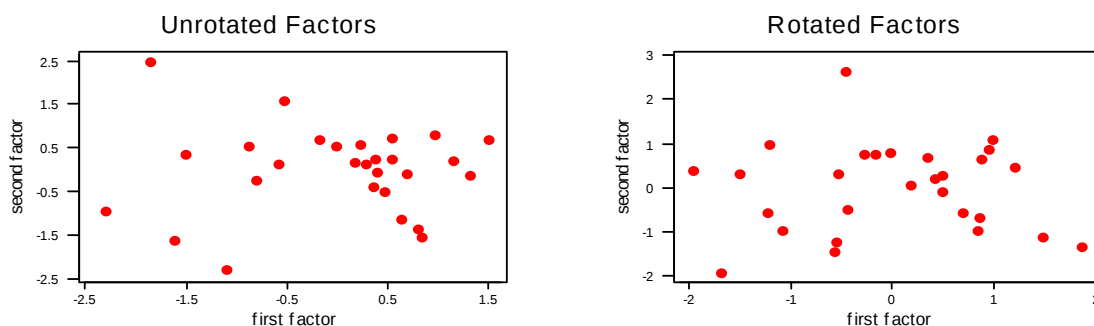
Proměnná	Komunalita pro zvolený počet vypočítaných faktorů				
	1	2	3	4	5
% K ₂ O	0,688	0,789	0,900	0,928	1,000
% P ₂ O ₅	0,017	0,599	0,803	0,991	1,000
% TiO ₂	0,339	0,382	0,853	0,998	1,000
Bod smočení	0,000	0,702	0,790	0,982	1,000
Vodivost	0,778	0,871	0,912	0,913	1,000

Tabulka I-5: Vysvětlený rozptyl pro zvolený počet faktorů

Vysvětlený rozptyl pro zvolený počet vypočítaných faktorů [%]				
1	2	3	4	5
36,5	66,9	85,1	96,3	100

Grafické znázornění výsledků faktorové analýzy před a po rotaci znázorňuje následující obrázek:

Factor Analysis Score Plot



Obr. 1.8 - Graf pro dva faktory před a po rotaci

Závěr:

Soubor dat neumožňuje separovat faktorově čisté proměnné. Pro popis souboru je možné použít dvou (66,9 % variability), popř. tří faktorů (85,1 % variability). Po rotaci EQUIMAX byly získány následující komponentní váhy:

Tabulka I-6: Komponentní váhy pro dva faktory

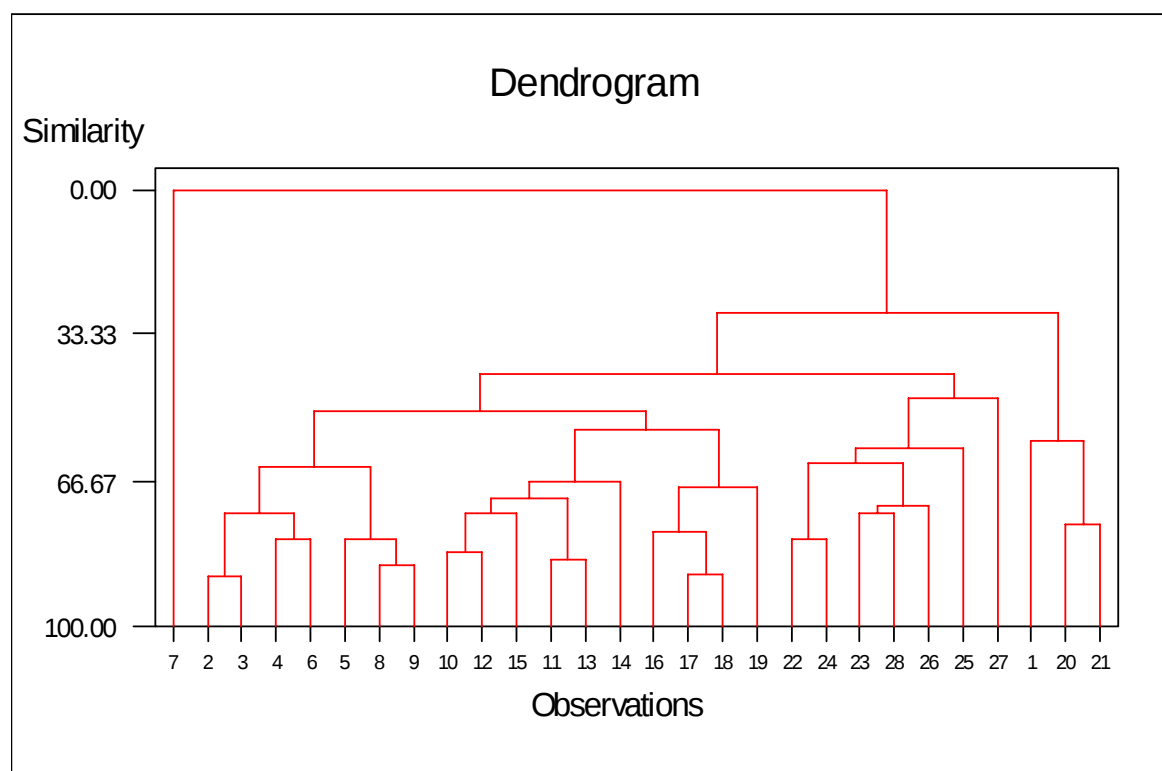
Proměnná	Faktor 1	Faktor 2
% K ₂ O	-0,869	0,185
% P ₂ O ₅	0,010	0,774
% TiO ₂	0,607	-0.113
Bod smočení	0,111	-0,830
Vodivost	-0.825	-0.438

Tabulka I-7: Komponentní váhy pro tři faktory

Proměnná	Faktor 1	Faktor 2	Faktor 3
% K ₂ O	0,886	-0,295	0,166
% P ₂ O ₅	0,069	-0,871	-0,197
% TiO ₂	-0,219	-0,077	-0,894
Bod smočení	0,161	0,711	-0,509
Vodivost	0,886	0,340	0,105

3.4.1.5 Analýza shluků

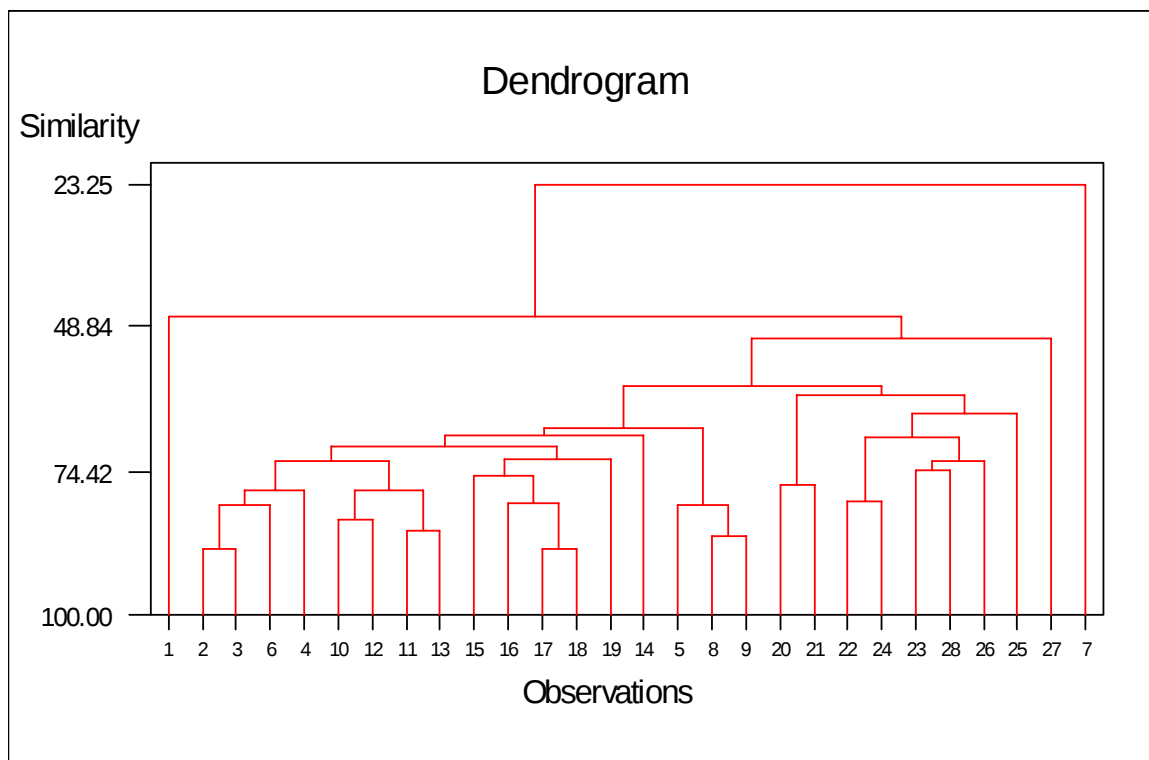
Bylo použito hierarchické klastrování podle Eukleidovské vzdálenosti. Na základě informace programu bylo nutno vyřadit některé sloupce, které byly indikovány jako téměř konstantní a bránily výpočtu. Konkrétně se jednalo o % SiO₂, %Fe a % Sb₂O₃. Všechny ostatní veličiny byly do klastrování zahrnuty. Dendrogram ukazuje následující obrázek:



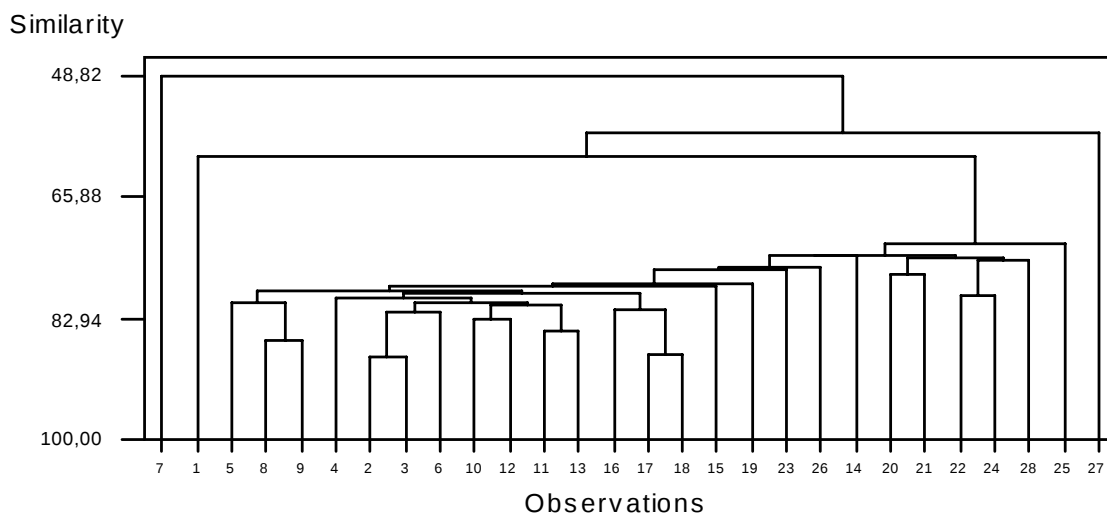
br. 1.9 - Dendrogram - Eukleidovská vzdálenost

O

V dalším hodnocení bylo použito průměru a Eukleidovské vzdálenosti a téhož se standardizací:



Obr. 1.10 - Dendrogram - Průměr a Eukleidovská vzdálenost



Obr. 1.11 - Dendrogram - Průměr a Eukleidovská vzdálenost, standardizováno

Závěr:

Objekty je možné klasifikovat různými způsoby. Ukazuje se velmi výrazná odlišnost některých objektů (obvykle jeden až dva) od ostatních. Jinak jsou však objekty velmi snadno seřaditelné do několika skupin s vysokou podobností.

3.4.1.6 Korelační analýza

Pro výpočet korelačních koeficientů bylo použito programu QCExpert.

Korelační analýza

Název úlohy : Sheet1

Párová kor.

Sloupec

Ford [s] - Viskozita [mPas]	0,9016109777
Ford [s] - pH	-0,5397026345
Ford [s] - Vodivost [mScm]	0,5592781191
Ford [s] - Rutil [%]	-0,2843553926
Ford [s] - Fe [%]	-0,01457583665
Ford [s] - Sb ₂ O ₃ [%]	0,3282946937
Ford [s] - K ₂ O [%]	0,3113967046
Ford [s] - Obsah S [%]	0,8514780065
Ford [s] - P ₂ O ₅ [%]	-0,07493763536
Ford [s] - SiO ₂ [%]	0,186477877
Ford [s] - Nb₂O₅ [%]	0,4326324552
Ford [s] - ZrO ₂ [%]	0,3347346661
Ford [s] - TiO ₂ [%]	0,06713226368
Ford [s] - Na ₂ O [%]	0,0002337717955
Ford [s] - Spotřeba vody [g/100g]	0,4054723011
Ford [s] - Bod smočení [g/100 g]	0,3469676349
Ford [s] - Bod tečení [g/100g]	0,5778302468
Viskozita [mPas] - pH	-0,6053059041
Viskozita [mPas] - Vodivost [mScm]	0,5839139925
Viskozita [mPas] - Rutil [%]	-0,3367318735
Viskozita [mPas] - Fe [%]	0,1330873116
Viskozita [mPas] - Sb ₂ O ₃ [%]	0,2831865447
Viskozita [mPas] - K ₂ O [%]	0,1230572952
Viskozita [mPas] - Obsah S [%]	0,8900758262
Viskozita [mPas] - P ₂ O ₅ [%]	-0,1978654143
Viskozita [mPas] - SiO ₂ [%]	0,2570362956
Viskozita [mPas] - Nb ₂ O ₅ [%]	0,3292641241
Viskozita [mPas] - ZrO ₂ [%]	0,2671269809
Viskozita [mPas] - TiO ₂ [%]	0,06308452848
Viskozita [mPas] - Na ₂ O [%]	0,01080055817

Viskozita [mPas] - Spotřeba vody [g/100g]	0,5939320638
Viskozita [mPas] - Bod smočení [g/100 g]	0,5608667519
Viskozita [mPas] - Bod tečení [g/100g]	0,4499419439
pH - Vodivost [mScm]	-0,7369238762
pH - Rutil [%]	0,1987977171
pH - Fe [%]	0,002183704645
pH - Sb2O3 [%]	-0,3877701027
pH - K2O [%]	-0,3868080035
pH - Obsah S [%]	-0,7194239708
pH - P2O5 [%]	0,1760851866
pH - SiO2 [%]	-0,1179711874
pH - Nb2O5 [%]	-0,492827669
pH - ZrO2 [%]	-0,4921562773
pH - TiO2 [%]	0,1561508613
pH - Na2O [%]	-0,1496098268
pH - Spotřeba vody [g/100g]	-0,256859732
pH - Bod smočení [g/100 g]	-0,1577700897
pH - Bod tečení [g/100g]	-0,3889534479
Vodivost [mScm] - Rutil [%]	-0,1318414485
Vodivost [mScm] - Fe [%]	-0,1002241823
Vodivost [mScm] - Sb2O3 [%]	0,1554039976
Vodivost [mScm] - K2O [%]	0,6292171373
Vodivost [mScm] - Obsah S [%]	0,7658272983
Vodivost [mScm] - P2O5 [%]	-0,245260445
Vodivost [mScm] - SiO2 [%]	0,08567481024
Vodivost [mScm] - Nb2O5 [%]	0,2068850837
Vodivost [mScm] - ZrO2 [%]	0,2737421647
Vodivost [mScm] - TiO2 [%]	-0,2870323603
Vodivost [mScm] - Na2O [%]	0,273690249
Vodivost [mScm] - Spotřeba vody [g/100g]	0,2939877743
Vodivost [mScm] - Bod smočení [g/100 g]	0,2746110059
Vodivost [mScm] - Bod tečení [g/100g]	0,56232779
Rutil [%] - Fe [%]	-0,6262584678
Rutil [%] - Sb2O3 [%]	-0,3881129896
Rutil [%] - K2O [%]	-0,09295238454
Rutil [%] - Obsah S [%]	-0,2360760163
Rutil [%] - P2O5 [%]	-0,2942222103
Rutil [%] - SiO2 [%]	-0,08677769338
Rutil [%] - Nb2O5 [%]	-0,1158715106
Rutil [%] - ZrO2 [%]	-0,1562680368
Rutil [%] - TiO2 [%]	0,130628866
Rutil [%] - Na2O [%]	-0,1873970983
Rutil [%] - Spotřeba vody [g/100g]	-0,1877296745
Rutil [%] - Bod smočení [g/100 g]	-0,2631442529
Rutil [%] - Bod tečení [g/100g]	0,04449135855

Fe [%] - Sb ₂ O ₃ [%]	-0,06196354256
Fe [%] - K ₂ O [%]	-0,2690628541
Fe [%] - Obsah S [%]	0,0006972635082
Fe [%] - P ₂ O ₅ [%]	0,1335701121
Fe [%] - SiO₂ [%]	0,5621437529
Fe [%] - Nb ₂ O ₅ [%]	-0,1631930578
Fe [%] - ZrO ₂ [%]	-0,2811680808
Fe [%] - TiO ₂ [%]	0,07315719878
Fe [%] - Na ₂ O [%]	0,05013248971
Fe [%] - Spotřeba vody [g/100g]	0,348897181
Fe [%] - Bod smočení [g/100 g]	0,3392697064
Fe [%] - Bod tečení [g/100g]	-0,403181886
Sb ₂ O ₃ [%] - K ₂ O [%]	0,2467354545
Sb ₂ O ₃ [%] - Obsah S [%]	0,2796892222
Sb ₂ O ₃ [%] - P ₂ O ₅ [%]	0,00998823508
Sb ₂ O ₃ [%] - SiO ₂ [%]	-0,2348316425
Sb₂O₃ [%] - Nb₂O₅ [%]	0,5296212461
Sb₂O₃ [%] - ZrO₂ [%]	0,4871391509
Sb ₂ O ₃ [%] - TiO ₂ [%]	0,005852484
Sb ₂ O ₃ [%] - Na ₂ O [%]	0,2157093356
Sb ₂ O ₃ [%] - Spotřeba vody [g/100g]	0,1297973729
Sb ₂ O ₃ [%] - Bod smočení [g/100 g]	0,03924324535
Sb ₂ O ₃ [%] - Bod tečení [g/100g]	0,2833463861
K₂O [%] - Obsah S [%]	0,3949996194
K ₂ O [%] - P ₂ O ₅ [%]	0,1869893763
K ₂ O [%] - SiO ₂ [%]	-0,1377414855
K₂O [%] - Nb₂O₅ [%]	0,3851713661
K ₂ O [%] - ZrO ₂ [%]	0,2905993633
K ₂ O [%] - TiO ₂ [%]	-0,2662124299
K ₂ O [%] - Na ₂ O [%]	0,08986054682
K ₂ O [%] - Spotřeba vody [g/100g]	-0,2440372549
K ₂ O [%] - Bod smočení [g/100 g]	-0,1903904771
K₂O [%] - Bod tečení [g/100g]	0,5591979194
Obsah S [%] - P ₂ O ₅ [%]	-0,3274671316
Obsah S [%] - SiO ₂ [%]	0,2365440751
Obsah S [%] - Nb ₂ O ₅ [%]	0,3358001772
Obsah S [%] - ZrO ₂ [%]	0,2298015869
Obsah S [%] - TiO ₂ [%]	-0,01480404648
Obsah S [%] - Na ₂ O [%]	0,08639279199
Obsah S [%] - Spotřeba vody [g/100g]	0,4987506685
Obsah S [%] - Bod smočení [g/100 g]	0,4801172116
Obsah S [%] - Bod tečení [g/100g]	0,4777803232
P ₂ O ₅ [%] - SiO ₂ [%]	-0,0625257587
P ₂ O ₅ [%] - Nb ₂ O ₅ [%]	0,3030403693
P ₂ O ₅ [%] - ZrO ₂ [%]	0,2404033413

P2O5 [%] - TiO2 [%]	0,06674003429
P2O5 [%] - Na2O [%]	0,0822844614
P2O5 [%] - Spotřeba vody [g/100g]	-0,4481775015
P2O5 [%] - Bod smočení [g/100 g]	-0,3301222107
P2O5 [%] - Bod tečení [g/100g]	0,1468361545
SiO2 [%] - Nb2O5 [%]	-0,03340271928
SiO2 [%] - ZrO2 [%]	-0,1857412419
SiO2 [%] - TiO2 [%]	-0,1386050025
SiO2 [%] - Na2O [%]	-0,0424855917
SiO2 [%] - Spotřeba vody [g/100g]	0,2513193065
SiO2 [%] - Bod smočení [g/100 g]	0,1486711836
SiO2 [%] - Bod tečení [g/100g]	-0,2461159389
Nb2O5 [%] - ZrO2 [%]	0,7971650115
Nb2O5 [%] - TiO2 [%]	0,03675239844
Nb2O5 [%] - Na2O [%]	0,1075329696
Nb2O5 [%] - Spotřeba vody [g/100g]	-0,1386453143
Nb2O5 [%] - Bod smočení [g/100 g]	-0,2252096236
Nb2O5 [%] - Bod tečení [g/100g]	0,4188422691
ZrO2 [%] - TiO2 [%]	-0,1888275776
ZrO2 [%] - Na2O [%]	0,1080466858
ZrO2 [%] - Spotřeba vody [g/100g]	-0,1450769367
ZrO2 [%] - Bod smočení [g/100 g]	-0,2591273306
ZrO2 [%] - Bod tečení [g/100g]	0,4619657451
TiO2 [%] - Na2O [%]	-0,209171562
TiO2 [%] - Spotřeba vody [g/100g]	0,2031725226
TiO2 [%] - Bod smočení [g/100 g]	0,1923021059
TiO2 [%] - Bod tečení [g/100g]	-0,00239445289
Na2O [%] - Spotřeba vody [g/100g]	-0,02172855669
Na2O [%] - Bod smočení [g/100 g]	-0,06400548922
Na2O [%] - Bod tečení [g/100g]	0,07559216625
Spotřeba vody [g/100g] - Bod smočení [g/100 g]	0,9189417321
Spotřeba vody [g/100g] - Bod tečení [g/100g]	0,2596532001
Bod smočení [g/100 g] - Bod tečení [g/100g]	0,211300464

Významné vztahy jsou vyznačeny tučně.

Závěr:

Analýza korelací je v dobrém souladu se závěry získanými při analýze dat metodou hlavních komponent a faktorovou analýzou.

3.4.2 VLASTNOSTI PLANETÁRNÍCH TĚLES SLUNEČNÍ SOUSTAVY

3.4.2.1 ÚVOD

Sluneční soustava, vezmeme-li v úvahu pouze kompaktní tělesa rozumné velikosti, je mimo centrální hvězdy tvořena planetami a jejich měsíci, planetkami a kometami. Kosmická technika umožnila získat poměrně přesné údaje jak o planetách, tak i o planetkách. Údaje o kometách jsou méně dostupné, přesněji vzato jsou v potřebném plném rozsahu k dispozici jen pro ojedinělá tělesa, jejichž hmotnost a často i tvar a rozměry se výrazně snižují každým přiblížením ke slunci, kdy se velká část materiálu vypaří. Vzhledem k malé hmotnosti komet dochází rovněž často ke změnám dráhy např. vlivem gravitace velkých planet. Z uvedených důvodů nebyly komety do hodnocení zahrnuty. Použitá data jsou v příloze č. 2.

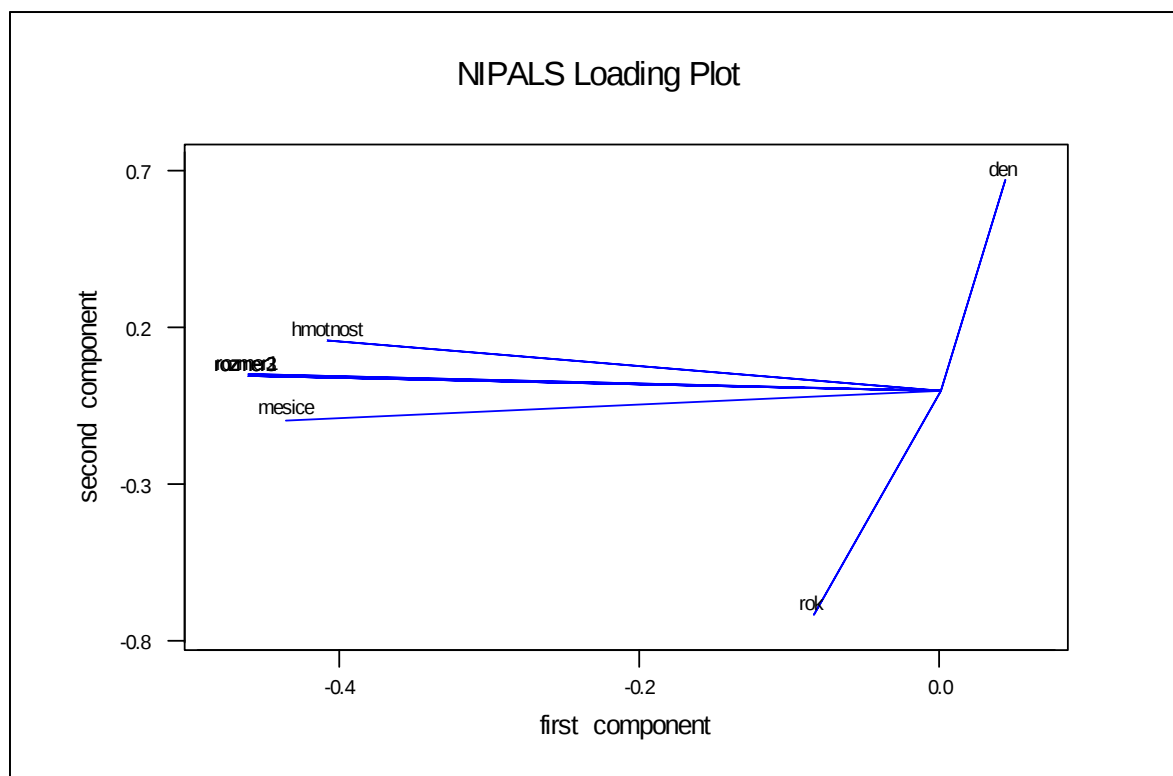
3.4.2.2 Zpracovávaná data

Data, uvedená v příloze č. 2, byla získána z webových stránek NASA. Pojmem rozměr 1, rozměr 2 jsou míněny průměr na rovníku (u rychle rotujících planet bývá nejvyšší vlivem snížení vlivu gravitace odstředivou silou) a průměr tělesa na pólech. V případě planetek, které jsou vesměs tvořeny nepravidelnými kusy hmoty, je uváděn i třetí rozměr takového tělesa, který je u planet se zploštěním na pólech nahrazen údajem pro průměr tělesa na pólech. Údaj o hmotnosti a počtu měsíců není třeba komentovat. Pojmem délka dne se rozumí doba jedné otáčky tělesa kolem jeho osy rotace odchýlené nejméně od kolmice směrem ke Slunci, pojmem délka roku doba jednoho oběhu tělesa kolem Slunce.

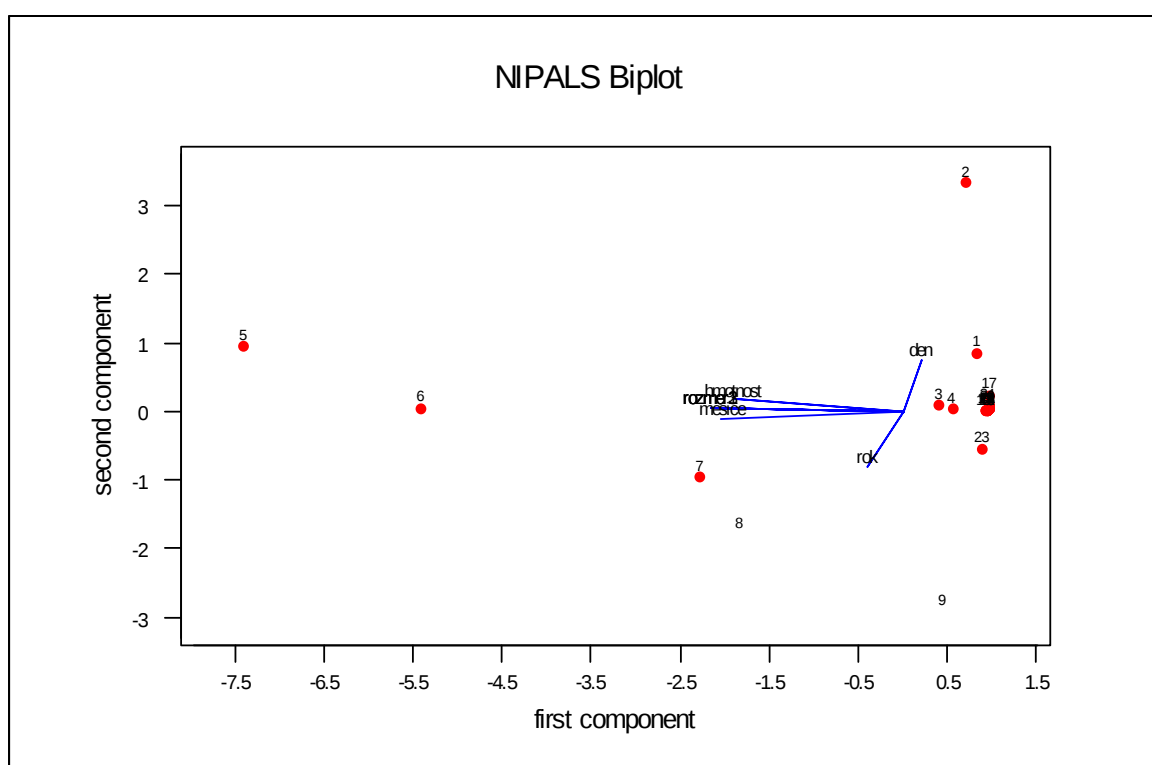
3.4.2.3 Metoda hlavních komponent

Vzhledem k problémům s během programu Statgraphics na počítači byly použity jiné statistické programy. V tomto případě SCAN 1.1. Jak ukazují graf komponentních vah (2.1) a dvojný graf (2.2) podobné vlastnosti mají a tím pádem spolu korelují:

- Rozměr 1, rozměr 2 a rozměr 3 (většinou se, až na některé planetky, jedná o velmi pravidelné kulovité útvary, které lze popsat jediným rozměrem - průměrem)
- Poměrně blízko k rozměrům je rovněž hmotnost, která logicky roste s rozměrem. Velké “plynové” planety - Jupiter, Saturn, Uran a Neptun mají mnohem menší hustotu než tzv. terestrické planety (Merkur, Venuše, Země, Mars) a planetky. To je důvodem, proč se vektory odklánějí.
- Blízko rozměrům a hmotnosti je rovněž počet měsíců, což je rovněž logické, protože velké a hmotné útvary mají větší gravitaci a tím větší možnost přitáhnout na svoji oběžnou dráhu tělesa procházející kolem i ve větších vzdálenostech.



Obr. 2.1 - Graf komponentních vah



Obr. 2.2 - Dvojný graf

Tabulka II-1: Stav před redukcí počtu proměnných

Hlavní komponenta	Eigenvalue	Proporčně	Kumulativně
1	4,6903	0,670	0,670
2	1,1183	0,160	0,830
3	0,8991	0,128	0,958
4	0,2746	0,039	0,997
5	0,0173	0,002	1,000

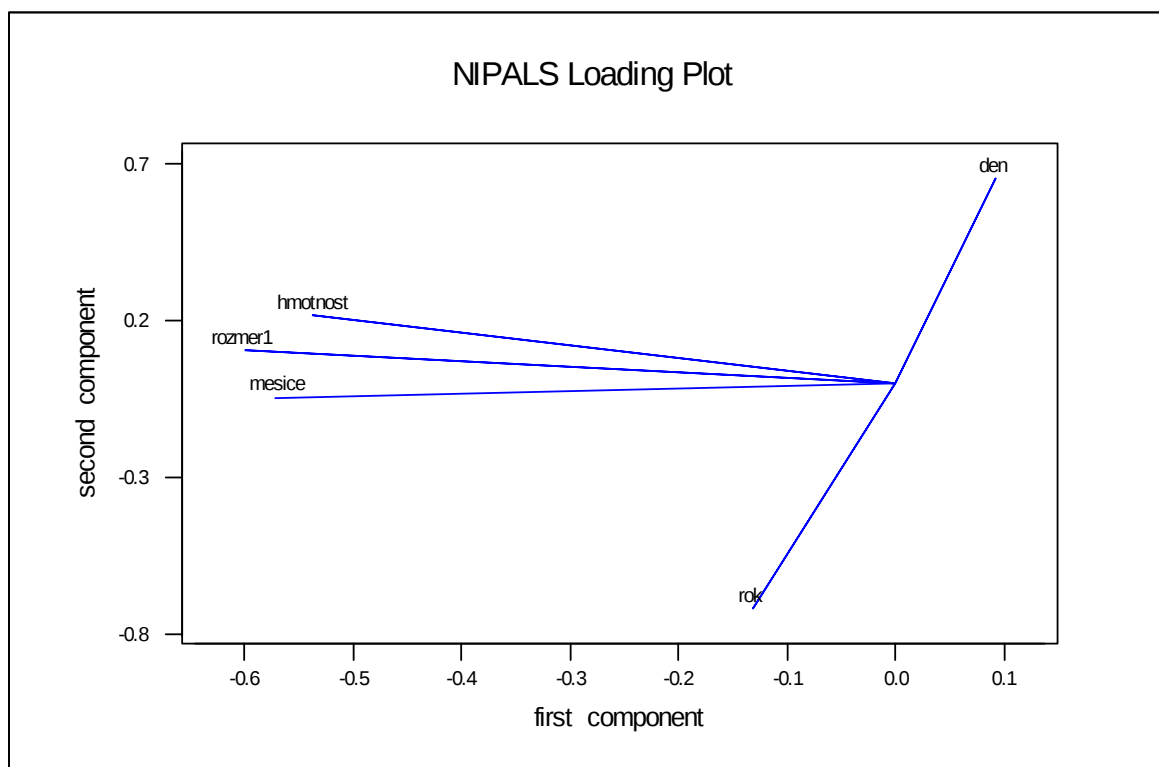
Jak je zřejmé z tabulky, i před redukcí proměnných by první dvě komponenty byly schopny popsat 83 % variability. Na základě dříve popsanych skutečností byla provedena redukce proměnných - místo rozměru 2, rozměru 3 bylo použito rozměru 1; hmotnost a počet měsíců byly ponechány.

Ve druhém kroku tedy bylo využito proměnných rozměr 1, hmotnost, počet měsíců a délka dne a délka roku.

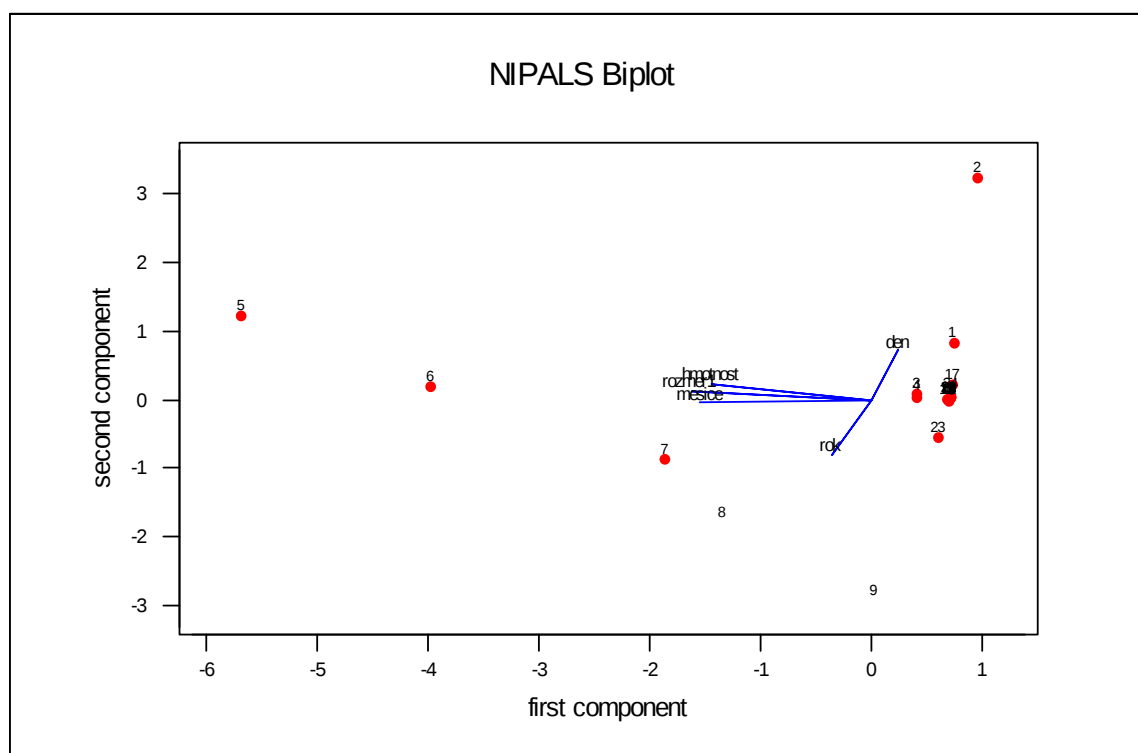
Tabulka II-2: Stav po první redukcí počtu proměnných

Hlavní komponenta	Eigenvalue	Proporčně	Kumulativně
1	2,7116	0,542	0,542
2	1,1075	0,222	0,764
3	0,8975	0,180	0,943
4	0,2718	0,054	0,998
5	0,0115	0,002	1,000

Vzhledem k tomu, že rozměry 1, 2 a 3 se prakticky nelišily, projevilo se jejich oprávněné vyřazení logicky mírným poklesem vysvětleného rozptylu. Nicméně, pomocí dvou latentních proměnných lze popsat 76,4 % rozptylu a pomocí tří dokonce 94,3 %. Situaci dobře popisují graf komponentních vah a dvojný graf 2-3 a 2-4.



Obr. 2.3 - Graf komponentných vah



Obr. 2.4 - Dvojný graf

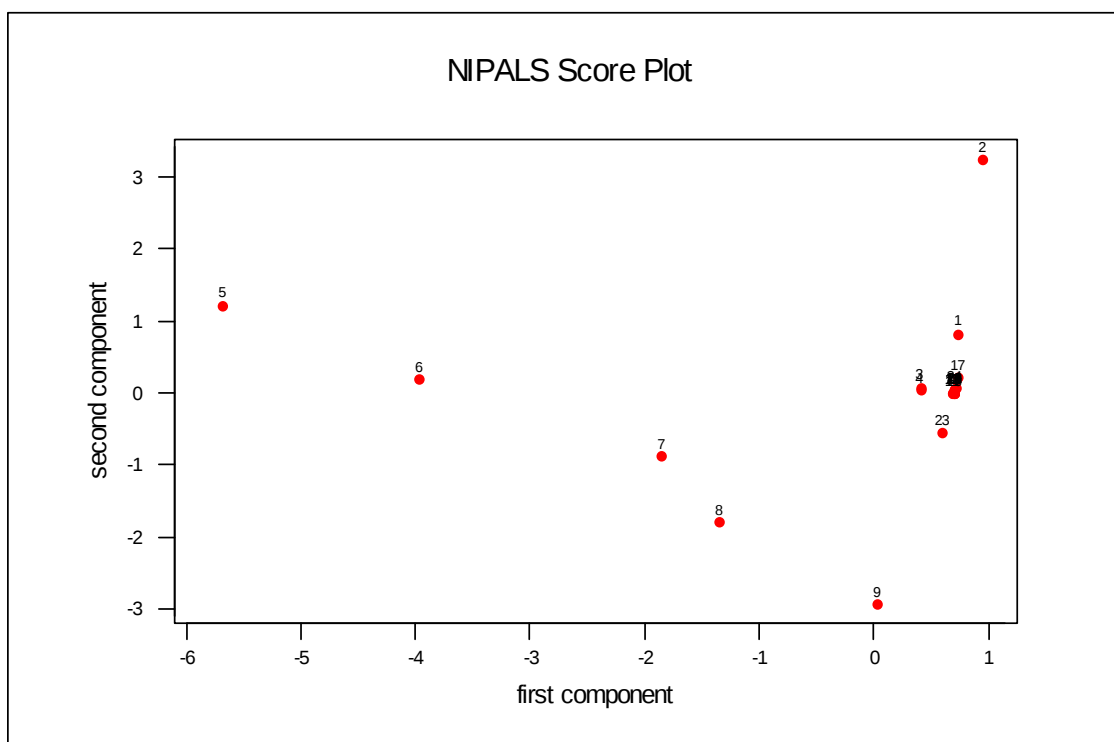
Závěr:

Jak ukazuje porovnání jednotlivých tabulek, podařilo se redukcí počtu proměnných dosáhnout pro dvě latentní proměnné maxima vysvětlené variability 76,4 %. Latentní proměnné y_1 a y_2 mají následující tvar:

$$y_1 = -0,600 * \text{rozměr 1} - 0,537 * \text{hmotnost} + 0,092 * \text{délka dne} - 0,132 * \text{délka roku} - 0,571 * \text{počet měsíců}$$

$$y_2 = 0,106 * \text{rozměr 1} + 0,215 * \text{hmotnost} + 0,653 * \text{délka dne} - 0,717 * \text{délka roku} - 0,044 * \text{počet měsíců}$$

Následující obrázek ukazuje rozložení jednotlivých bodů v grafu při použití uvedených dvou proměnných:



Obr. 2.5 - Graf znázorňující závislost y_2 na y_1 pro jednotlivé body

Jak ukazuje graf, objekty byly rozděleny do dvou až tří shluků. Pokud bychom se snažili o fyzikální vysvětlení latentních proměnných y_1 a y_2 , dalo by se říci, že první proměnná zvýrazňuje především parametry související s hmotností těles a jejich gravitací. Pro to svědčí nejvyšší koeficienty u rozměru 1, hmotnosti a počtu měsíců. Druhá latentní proměnná y_2 souvisí především s polohou planety a její rotací (délka roku a dne). Je zřejmé, že planety terestrického typu, t. j. Merkur (1), Země (3), Mars (4) i planetky (nad 10), které se nacházejí mezi Marsem a Jupiterem, se velmi podobají a tvoří společný shluk. Od nich se výrazně liší Venuše (2) což je dáno především tím, že její den je delší než rok. Planeta Pluto (9) se hmotností a

rozměry blíží k planetkám, výrazně se však odlišuje délkou roku (velká vzdálenost od Slunce). Proto se ze shluku vyděluje. Velké planety s nízkou hustotou (Jupiter, Saturn, Uran, Neptun, t.j. 5 - 8) mají vysokou hmotnost, hodně měsíců, dlouhý rok a krátký den, který je důsledkem zákona zachování hybnosti při jejich smrštění vlivem vlastní gravitace. Proto se dají považovat za samostatný shluk.

3.4.2.4 Faktorová analýza

Při faktorové analýze bylo využito redukce počtu proměnných provedené pomocí PCA. Byly vypočteny komunality při použití pěti, čtyř, tří a dvou faktorů. Pro rotaci bylo využito metody EQUIMAX, která spojuje kritéria metody Varimax a Quartimax, což znamená, že při rotaci je maximalizován rozptyl čtverců faktorových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. Výpočet byl proveden pro volbu počtu faktorů 1 až 5 s následujícím výsledkem:

Tabulka II-3: Komunalita pro zvolený počet faktorů

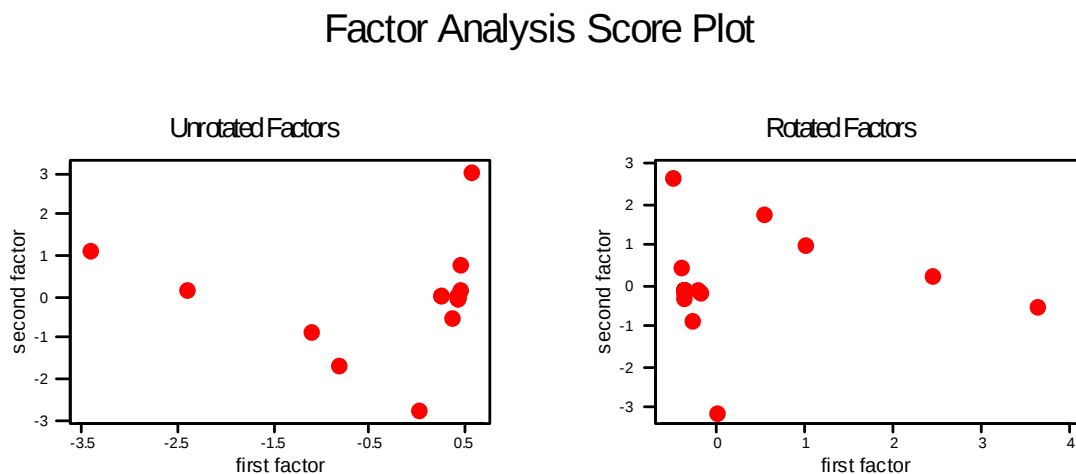
Proměnná	Komunalita pro zvolený počet vypočítaných faktorů				
	1	2	3	4	5
rozměr 1	0,976	0,988	0,989	0,993	1,000
hmotnost	0,783	0,834	0,853	0,999	1,000
délka dne	0,023	0,495	1,000	1,000	1,000
délka roku	0,047	0,616	0,984	1,000	1,000
Počet měsíců	0,883	0,885	0,890	0,997	1,000

Tabulka II-4: Vysvětlený rozptyl pro zvolený počet faktorů

Vysvětlený rozptyl pro zvolený počet vypočítaných faktorů [%]				
1	2	3	4	5
22,9	44,5	64,9	85,0	100

Grafické znázornění výsledků faktorové analýzy před a po rotaci znázorňuje následující obrázek:

Obr. 2.6 - Graf pro dva faktory před a po rotaci



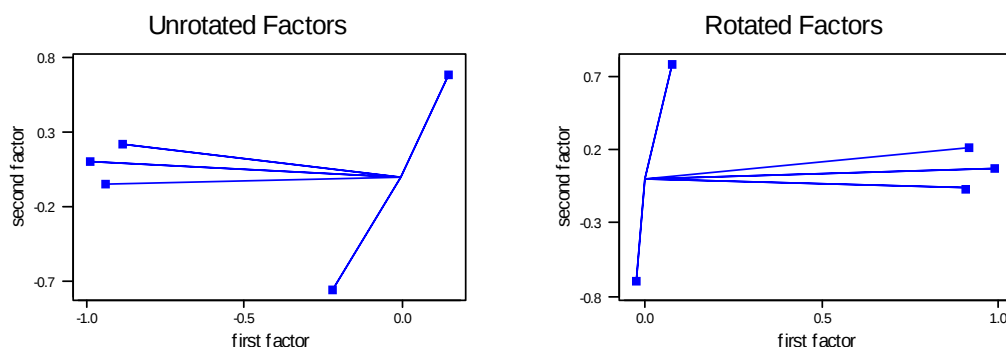
Závěr:

Soubor dat umožňuje separovat dosti faktorově čisté proměnné. Pro popis souboru je možné použít dvou (44,5 % variability), popř. tří faktorů (64,9 % variability). Po rotaci EQUIMAX byly získány následující komponentní váhy:

Tabulka II-5: Komponentní váhy pro dva faktory

Proměnná	Faktor 1	Faktor 2
rozměr 1	0,992	0,067
hmotnost	0,911	-0,064
délka dne	-0,025	-0,703
délka roku	0,078	0,781
počet měsíců	0,916	0,214

Factor Analysis Loading Plot



Obr. 2.7 - Graf pro dva faktory před a po rotaci

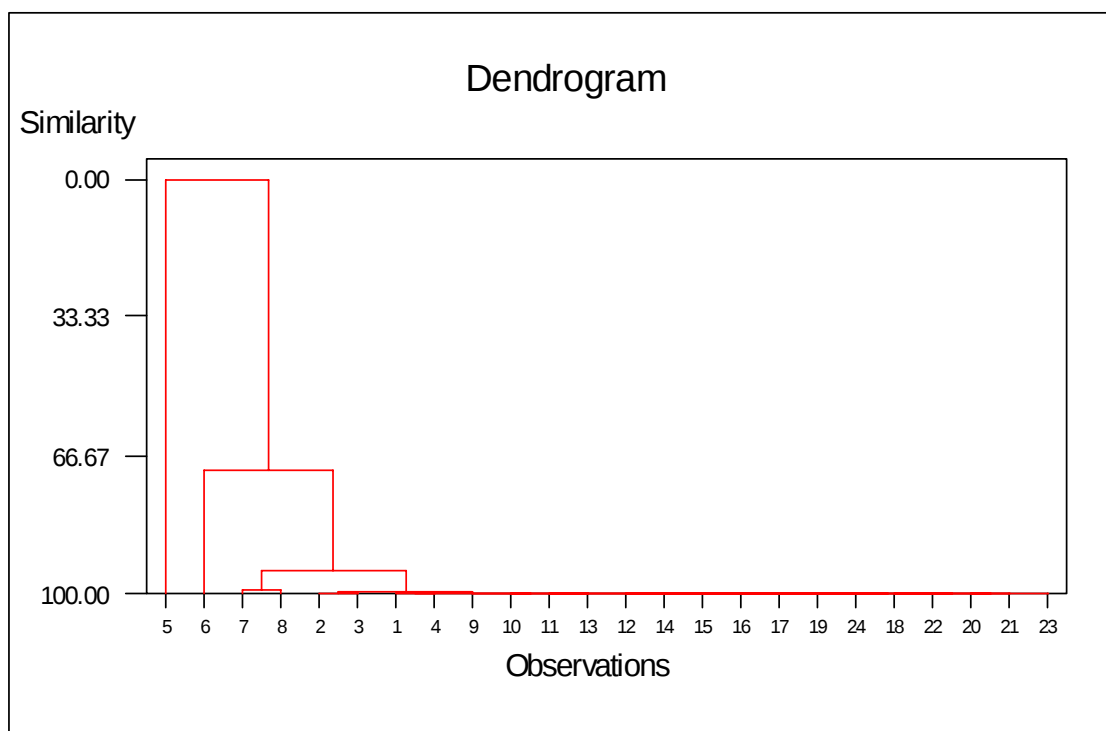
První faktor jasně zahrnuje především veličiny svázané s hmotností a gravitací (rozměr 1, hmotnost, počet měsíců). Druhý faktor je méně “čistý”, nicméně v něm jasně převládá vliv vzdálenosti od Slunce a způsobu oběhu a rotace (délka roku, délka dne), který je do jisté míry “rozmýván” počtem měsíců (vzdálené planety mimo Pluto jsou velmi hmotné a mají tedy hodně měsíců, které se tak objevují i ve druhém, především “dráhovém” faktoru).

Tabulka II-6: Komponentní váhy pro tři faktory

Proměnná	Faktor 1	Faktor 2	Faktor 3
rozměr 1	0,989	0,105	-0,009
hmotnost	0,916	-0,109	-0,051
délka dne	-0,034	-0,059	0,998
délka roku	0,046	0,989	-0,060
počet měsíců	0,910	0,237	-0,079

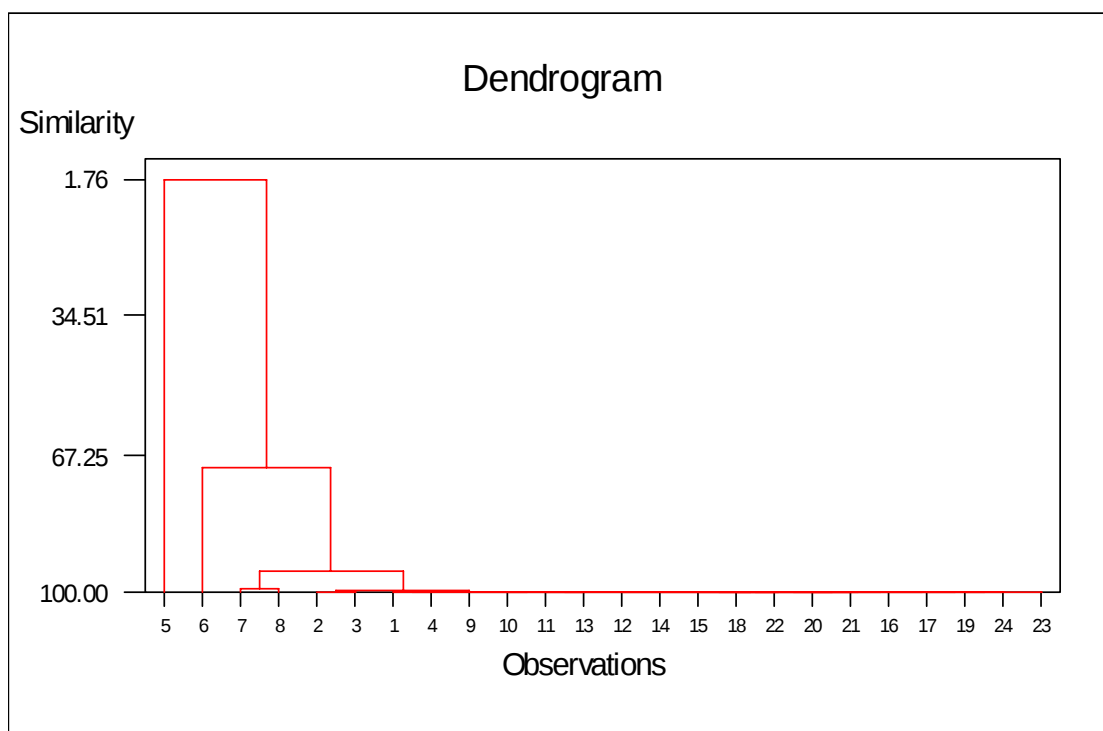
3.4.2.5 Analýza shluků

Bylo použito hierarchické klastrování podle Eukleidovské vzdálenosti. Do klastrování byly zahrnuty všechny veličiny, aby se mohly popřípadě projevit i vlivy asymetrického tvaru těles (tři rozměry). Dendrogram ukazuje následující obrázek:



Obr. 2.8 - Dendrogram - Eukleidovská vzdálenost

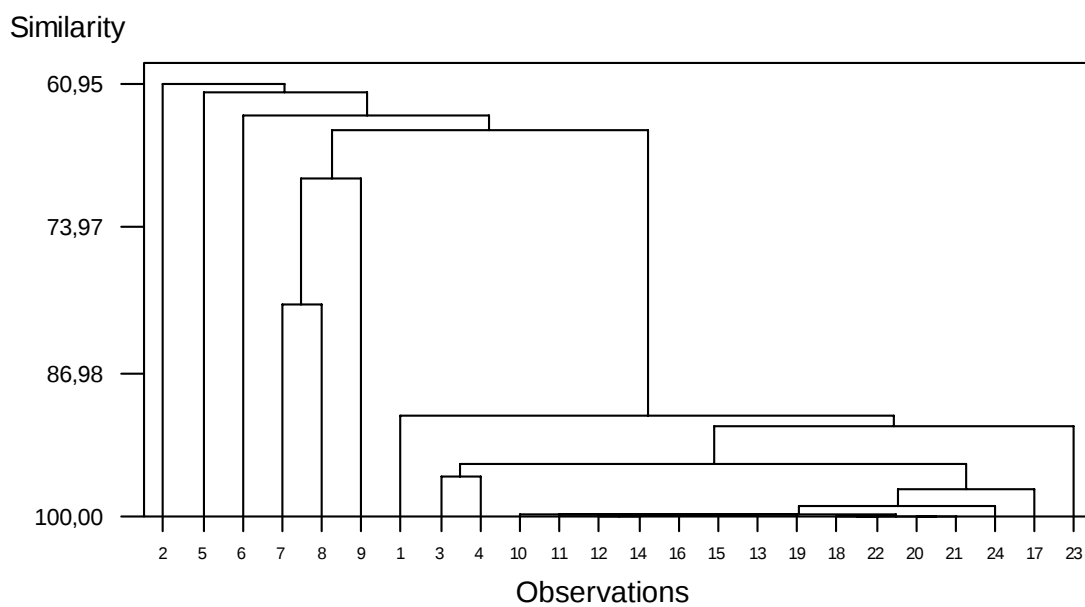
Z grafu je zřejmý především rozdíl Jupiteru (5) a méně výrazně i Saturnu (6) od ostatních planet (obří rozměr). Další velké planety (7- Uran a 8-Neptun) jsou rovněž dobře rozlišitelné od terestrických planet (1 až 4, 9). Přes problémy s měřítkem je vidět, že malé planetky (10 až 23) si jsou rovněž podobné. Byl učiněn pokus snížit rozdíly použitím pouze rozměru 1 a vyřazením rozměru 2 a 3. Třídění však proběhlo bez viditelných rozdílů v dendrogramu. Ke stejným výsledkům vedly i další zvolená kritéria klastrování, mimo postupu se standardizací (Obr. 2-10).



O

br

. 2.9 - Dendrogram - Průměr a Eukleidovská vzdálenost



Obr. 2.10 - Dendrogram - Průměr a Eukleidovská vzdálenost, standardizováno

Použití standardizace umožnilo další, jiný typ rozlišení. Vynikly atypické charakteristiky

planety Venuše (2) diskutované již v části využívající PCA, podobnost velkých a vzdálených planet (5,6,7,8), jemněji jsou rovněž roztríděny terestrické planety a planetky podobných velikostí a vlastností.

Závěr

Vzhledem k velkým číselným, víceřádkovým rozdílům jednotlivých hodnot umožňuje standardizace jemnější rozklastrování. Není-li použita, vyniknou při klastrování některé charakteristiky, které umožní jednoduché, méně komplikované a ostré třídění.

3.4.2.6 Korelační analýza

Pro výpočet korelačních koeficientů bylo použito programu QCExpert.

Korelační analýza (významné vztahy jsou vyznačeny tučně).

Párová kor.

Sloupec

rozměr 1 - rozměr 2	0,9994705177
rozměr 1 - rozměr 3	0,9994700833
rozměr 1 - hmotnost	0,8696101166
rozměr 1 - délka dne	-0,04959414624
rozměr 1 - délka roku	0,1426028228
rozměr 1 - počet měsíců	0,9398158503
rozměr 2 - rozměr 3	0,9999999695
rozměr 2 - hmotnost	0,8725746597
rozměr 2 - délka dne	-0,04555875836
rozměr 2 - délka roku	0,1513564813
rozměr 2 - počet měsíců	0,939779459
rozměr 3 - hmotnost	0,8725688073
rozměr 3 - délka dne	-0,04554309602
rozměr 3 - délka roku	0,1513739126
rozměr 3 - počet měsíců	0,9397816705
hmotnost - délka dne	-0,07510427085
hmotnost - délka roku	-0,01438447515
hmotnost - počet měsíců	0,6885842889
délka dne - délka roku	-0,1202757287
délka dne - počet měsíců	-0,1238576455
délka roku - počet měsíců	0,2397888523

Závěr:

Analýza korelací je v dobrém souladu se závěry získanými při analýze dat metodou hlavních komponent a faktorovou analýzou.

