

**Univerzita Pardubice
Fakulta chemicko - technologická
Katedra analytické chemie**

Dvouleté licenční studium: Počítačové zpracování dat při kontrole a řízení jakosti

Počítačová analýza vícerozměrných dat

***Vyučující: Prof. RNDr. M. Meloun, DrSc
Univerzita Pardubice***

Obsah:

1. ŽELEZNICE A PŘEPRAVOVANÉ ZBOŽÍ	3
1.1 ZADÁNÍ:	3
1.2 DATA:	3
1.3 PROGRAM:	5
1.4 ŘEŠENÍ:	6
1.4.1 Korelační analýza.....	6
1.4.2 Metoda hlavních komponent.....	6
1.4.3 Faktorová analýza	13
1.4.4 Analýza shluků	15
1.4.5 Hvězdicové grafy.....	17
1.5 ZÁVĚR	19
2. ROZDĚLENÍ PŘEDMĚTŮ A PROSPĚCH ABSOLVENTŮ	20
2.1 ZADÁNÍ:	20
2.2 DATA:	20
2.3 PROGRAM:	25
2.4 ŘEŠENÍ:	25
2.4.1 Korelační analýza.....	25
2.4.2 Metoda hlavních komponent.....	26
2.4.3 Faktorová analýza	31
2.4.4 Shlukování.....	33
2.4.5 Hvězdicové grafy.....	35
2.5 ZÁVĚR	37

1. Železnice a přepravované zboží

1.1 Zadání:

V roce 1994 bylo sledováno množství přepravených tun ve veřejné dopravě podle NST druhů zboží ve vozových zásilkách. Tyto údaje byly sledovány v některých zemích Evropy, Asie a Afriky (tam, kde tyto údaje byly sledovány). Naším úkolem je zjistit pomocí metod vícerozměrné statistické analýzy (PCA, FA, ...), zda se z těchto údajů dá zjistit hospodářská úroveň daného státu - většinou totiž jde o státní železnice. Zdroj - DATIS, datové a informační služby, statistika v dopravě.

1.2 Data:

Tabulka1.1 - Zkratky a názvy přepravovaného zboží

Číslo	zkratka	název přepravovaného zboží
1	Zem	zemědělské produkty a živá zvířata
2	Pot	ostatní potraviny a krmivo
3	Pal	pevná nerostná paliva
4	Rop	ropné produkty
5	Rud	rudy a šrot pro hutní průmysl
6	Výr	hutní výroba
7	Mat	rudy surové a opracované stavební materiály
8	Hno	hnojiva
9	Che	chemické výrobky
10	Voz	vozidla, stroje polotovary i hotové výrobky

Tabulka 1.2 - Sledované železnice

Číslo	Zkratka	Země	Název
1	BR	Velká Britanie	Britské dráhy
2	CFL	Lucembursko	Lucemburské státní dráhy
3	CH	Řecko	Řecká železniční organizace
4	CIE	Irsko	Irská dopravní společnost
5	CP	Portugalsko	Portugalské dráhy
6	DBAG	Německo	Německé dráhy a.s.
7	DSB	Dánsko	Dánské státní dráhy
8	FS	Itálie	Italské státní dráhy
9	NS	Nizozemsko	Nizozemské dráhy
10	RENFE	Španělsko	Španělské státní dráhy
11	SNCB/NMBS	Belgie	Belgické státní dráhy
12	SNCF	Francie	Francouzské státní dráhy
13	CFF/SBB/FFS	Švýcarsko	Švýcarské federální spolkové dráhy
14	NSB	Norsko	Norské státní dráhy
15	OBB	Rakousko	Rakouské spolkové dráhy
16	SJ	Švédsko	Švédské státní dráhy
17	VR	Finsko	Finské dráhy
18	BC	Bělorusko	Běloruské dráhy
19	BDZ	Bulharsko	Bulharské státní dráhy
20	CD	Česká republika	České dráhy
21	CFARYM	Makedonie	Dráhy rep. Makedonie
22	MAV	Maďarsko	Maďarské státní dráhy
23	PKP	Polsko	Polské státní dráhy
24	SZ	Slovinsko	Slovinské železnice
25	ZSR	Slovensko	Železnice Slovenské republiky
26	ONFCM	Maroko	Státní správa železnic
27	SNCFT	Tunisko	Tuniské státní dráhy
28	CFS	Sýrie	Státní správa syrských drah
29	ISR	Izrael	Izraelské dráhy
30	TCDD	Turecko	Turecké státní dráhy

Tabulka 1.3 - zjištěné údaje (v tisících tun)

ZEM	POT	PAL	ROP	RUD	VYR	MAT	HNO	CHE	VOZ	ZKR
445	1343	41962	7674	9791	7062	18829	904	1653	7679	BR
77	48	1163	1543	2636	5735	1701	160	739	4061	CFL
337	81	0	0	69	68	483	0	105	176	CH
166	0	17	50	618	0	659	177	279	1049	CIE
908	126	900	0	583	159	2402	235	246	1580	CP
10440	4661	74048	28617	27413	46847	44038	7723	16368	46759	DBAG
1059	1100	213	66	218	1780	229	335	686	3178	DSB
5266	1725	460	1410	5509	13053	5668	281	3541	29887	FS
458	614	707	1043	2590	806	1664	1810	2462	4574	NS
1450	2299	1996	506	1931	3151	2786	495	2483	3733	RENFE
1408	2093	8711	2478	10987	18437	2984	864	3373	12076	SNCB/NMBS
10811	10962	6366	9382	8354	20482	21436	3609	9172	23307	SNCF
3111	2345	206	5668	2283	4963	8635	254	3086	15888	CFF/SBB/FFS
1321	295	201	0	13257	263	201	5	99	4085	NSB
9943	1741	3858	3642	6980	5419	5530	1032	4419	19940	OBB
5861	659	235	736	24947	5588	1032	18	3294	7986	SJ
13631	264	450	3635	2816	3449	1251	494	5968	6563	VR
7965	5770	3486	20055	735	7830	21298	9088	2421	4926	BC
195	737	8060	4485	2054	3337	3686	1230	961	4920	BDZ
8966	800	50228	4256	2992	11188	15107	1115	13948	1412	CD
140	50	168	131	64	502	295	85	100	384	CFARYM
4456	2894	8154	5447	4886	2538	3387	1048	3158	2809	MAV
5667	2255	114507	10911	16966	10303	19983	3881	9046	15366	PKP
2823	505	696	985	1221	2109	309	209	1570	1642	SZ
5935	678	16075	447	13818	6790	8567	1399	3301	1943	ZSR
659	577	151	1328	79	25	1895	20415	1879	95	ONCFM
575	98	105	7	7461	321	1514	808	687	51	SNCFT
536	2	104	1567	259	129	895	29	389	128	CFS
0	780	2284	56	2416	75	273	1452	232	1099	ISR
660	313	3593	301	5583	606	1283	354	8	1677	TCDD

1.3 Program:

STATGRAPHICS- vícerozměrné statistiky, EXCEL - analýza dat

1.4 Řešení:

1.4.1 Korelační analýza

Tabulka 1.4 Korelační matice

	ZEM	POT	PAL	ROP	RUD	VYR	MAT	HNO	CHE	VOZ
ZEM	1	0,543338	0,323044	0,551905	0,391256	0,554288	0,563628	0,140244	0,758028	0,56275
POT	0,543338	1	0,19711	0,634124	0,237628	0,591547	0,652503	0,275795	0,503149	0,56099
PAL	0,323044	0,19711	1	0,588152	0,550579	0,553574	0,71494	0,174896	0,700145	0,43551
ROP	0,551905	0,634124	0,588152	1	0,460658	0,786816	0,927082	0,412122	0,677527	0,67585
RUD	0,391256	0,237628	0,550579	0,460658	1	0,648913	0,548536	0,044791	0,531564	0,59091
VYR	0,554288	0,591547	0,553574	0,786816	0,648913	1	0,851573	0,227335	0,808311	0,86695
MAT	0,563628	0,652503	0,71494	0,927082	0,548536	0,851573	1	0,350685	0,79314	0,72298
HNO	0,140244	0,275795	0,174896	0,412122	0,044791	0,227335	0,350685	1	0,231808	0,14359
CHE	0,758028	0,503149	0,700145	0,677527	0,531564	0,808311	0,79314	0,231808	1	0,67311
VOZ	0,562753	0,560993	0,435512	0,675851	0,590911	0,866957	0,722983	0,143591	0,673111	

Determinant korelační matice = 0,00003

Poněvadž je determinant korelační matice velice blízký nule, dá se předpokládat multikolinearita, tj. závislost mezi proměnnými. Z korelační matice je také vidět, že hodnoty některých korelačních koeficientů jsou blízké jedné, to znamená poměrně značnou míru lineární závislosti mezi proměnnými.

1.4.2 Metoda hlavních komponent

Pomocí metody hlavních komponent se pokusíme redukovat počet sledovaných faktorů a rozseparovat sledované dráhy.

Výpočtem ve Statgraphicsu vyšlo, že první dvě hlavní komponenty popisují 71,54% variability v datech a první tři 80,82% variability (tab.1.6)

Tab.1.6

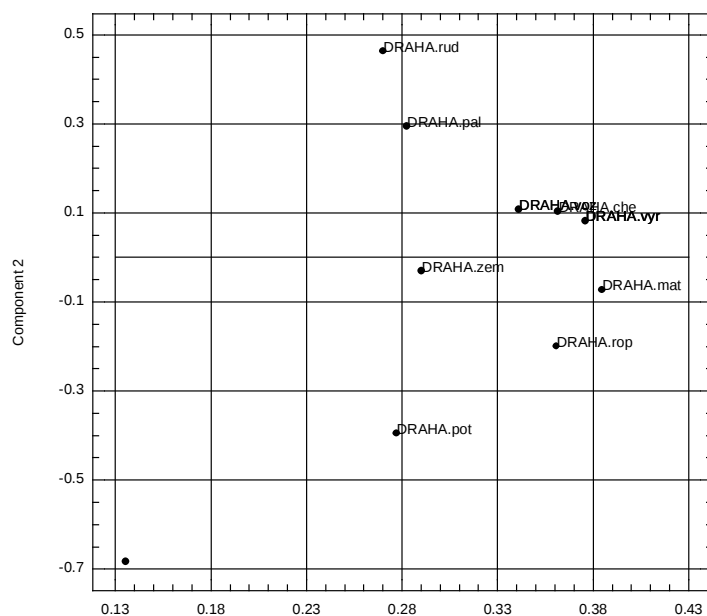
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	60.18792	60.18792
2	11.34884	71.53676
3	9.28159	80.81835
4	6.17308	86.99143
5	5.07782	92.06924
6	3.34167	95.41092
7	2.35049	97.76141
8	1.52402	99.28543
9	0.40209	99.68752
10	0.31248	100.00000

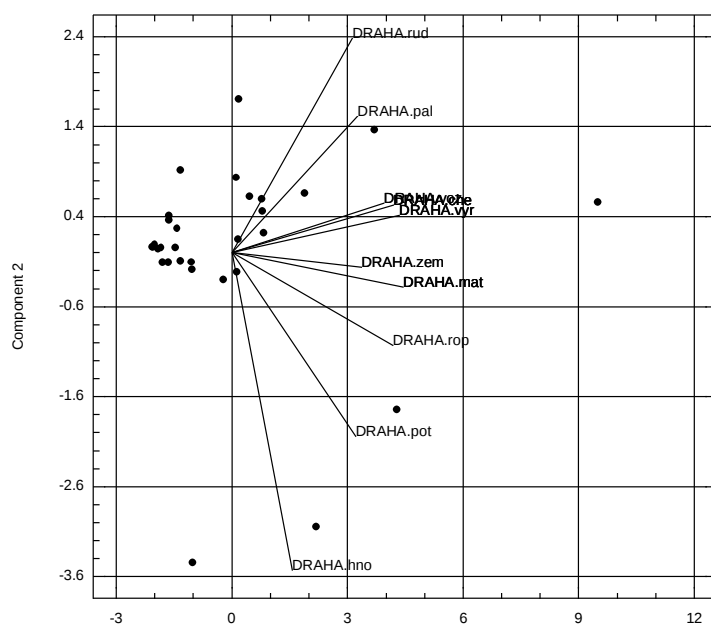
Graf komponentních vah (obr.1.1) a dvojný graf - biplot (obr.1.2) ukazují, že podobné vlastnosti mají (a tím pádem spolu korelují)

- rud, pal - rudy a šrot pro hutní průmysl, pevná nerostná paliva (přepravy důležité pro hutnictví a těžký průmysl)

- voz, che, vyr - vozidla, stroje polotovary i hotové výrobky, chemické výrobky, hutní výroba (to jsou produkty průmyslové výroby)
- mat, rop, zem - surové rudy a stavební materiály, ropné produkty, zemědělské produkty a živá zvířata



(obr.1.1)



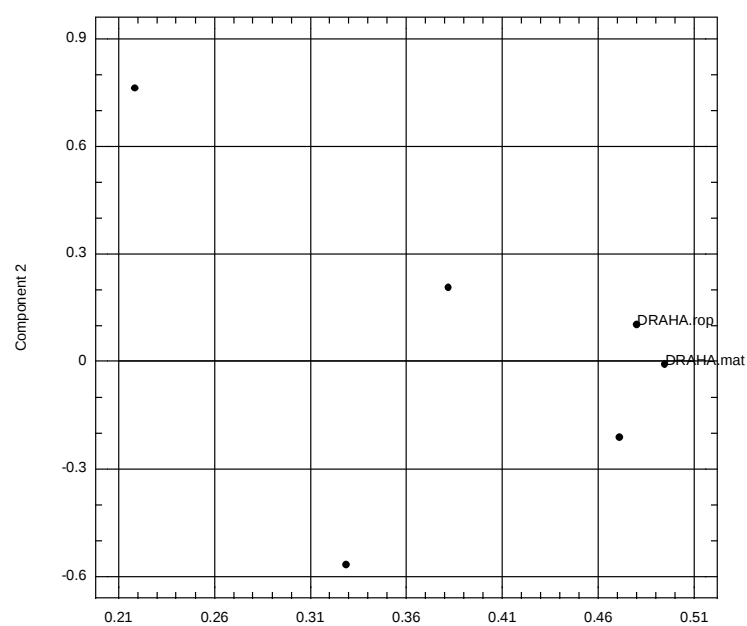
(obr.1.2)

Po odstranění pal, voz, che, zem je u zbylých proměnných vysvětleno pro první dvě hlavní komponenty 79,35% variability v datech a pro první tři 90,38% (tab.1.7).

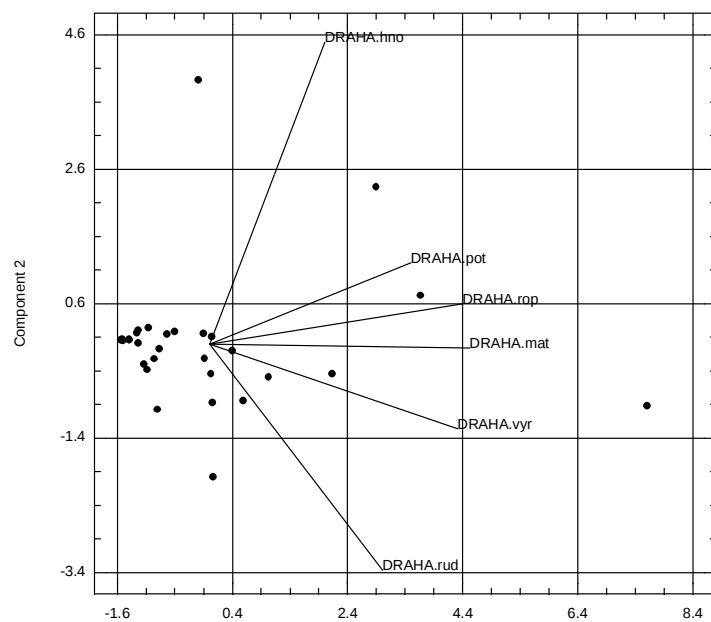
Tab.1.7
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	62.24010	62.24010
2	17.10895	79.34905
3	11.02683	90.37587
4	5.76235	96.13822
5	2.83332	98.97154
6	1.02846	100.00000

Grafy komponentních vah (obr.1.3) a dvojný (obr.1.4) ukazují, že podobné vlastnosti vykazují rop a mat. Odstraníme mat.



(obr.1.3)



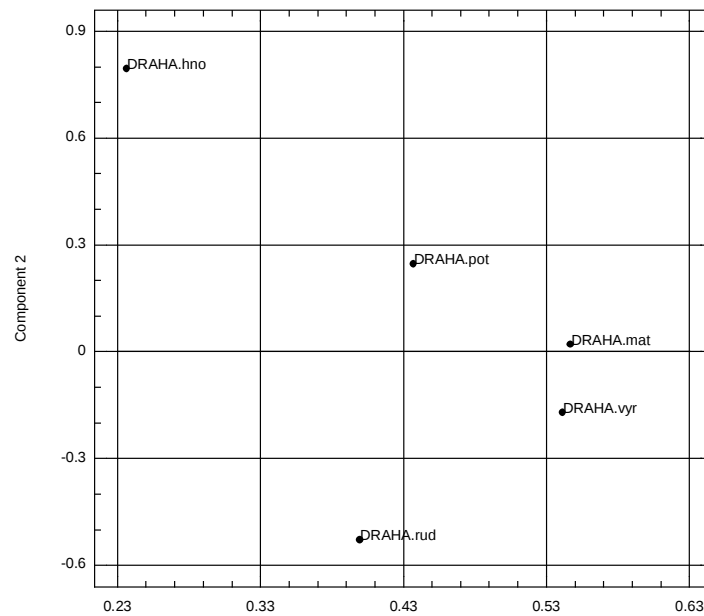
(obr.1.4)

Po opětovném provedení analýzy vyjde procento vysvětlené variability pro první dvě hlavní komponenty 78,6% a pro první tři 91,82% (tab.1.8).

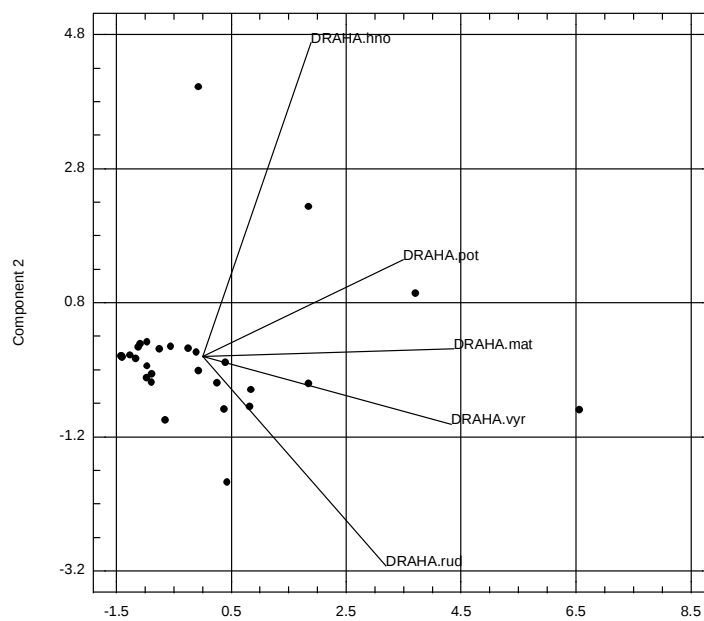
Tab.1.8
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	58.32654	58.32654
2	20.28552	78.61206
3	13.20527	91.81733
4	5.55588	97.37321
5	2.62679	100.00000

Z grafů komponentních vah (obr.1.5) a dvojného (obr.1.6) je vidět, že podobné vlastnosti vykazují mat a vyr (surové rudy a opracované stavební materiály, hutní výroba), odstraníme mat.



(obr.1.5)

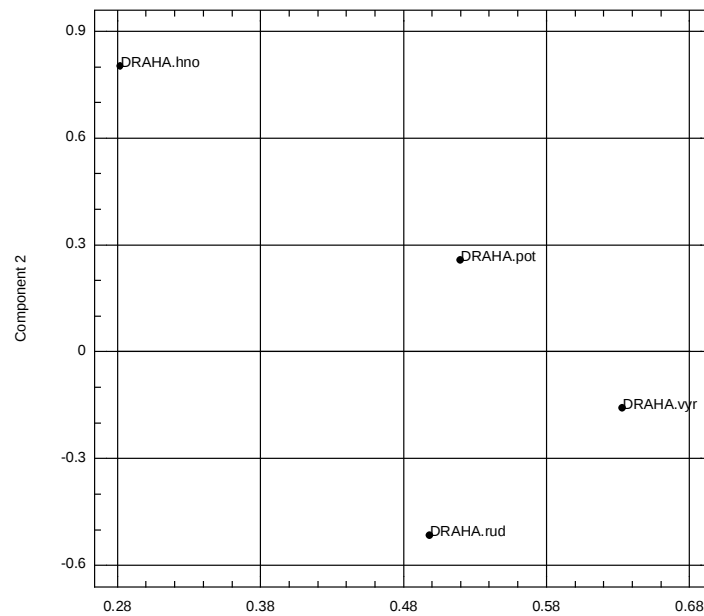


(obr.1.6)

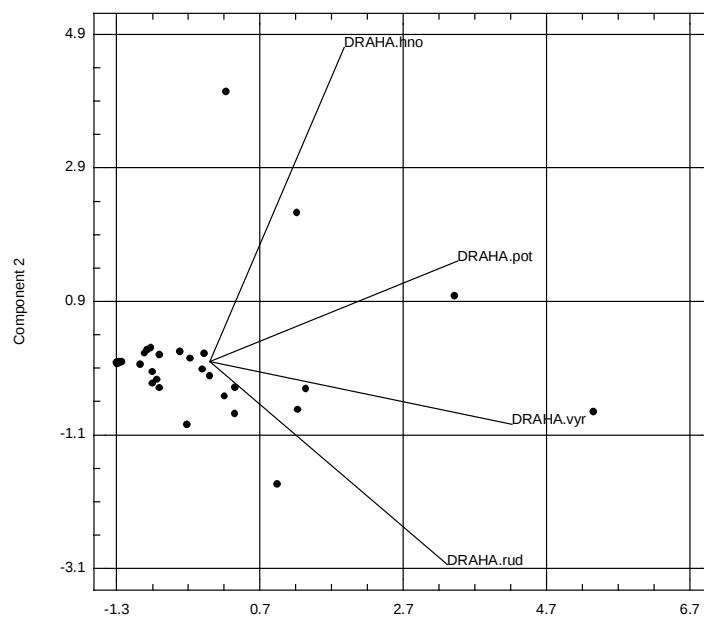
Výsledná analýza ukazuje, jak hlavní komponenty charakterizují variabilitu v datech (tab.1.9) . Z tabulky je patrné zmenšení podílu vysvětlené variability v datech, avšak grafy (obr.1.7 a obr.1.8) ukazují na nekorelovanost zbývajících veličin.

Tab.1.9
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	52.44410	52.44410
2	25.34337	77.78746
3	16.45296	94.24042
4	5.75958	100.00000



(obr.1.7)

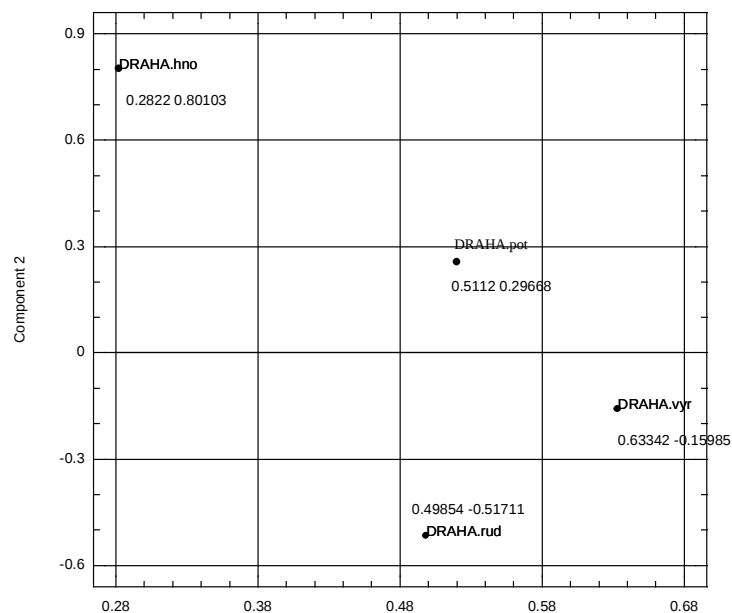


(obr.1.8)

Pokud v grafu komponentních vah (obr.1.9) vezmeme souřadnice bodů, tak nám tyto souřadnice poskytují míru příspěvku dané proměnné do příslušné hlavní komponenty:

$$y_1 = 0,282 \cdot hno + 0,511 \cdot pot + 0,499 \cdot rud + 0,633 \cdot vyr$$

$$y_2 = 0,801 \cdot hno + 0,3 \cdot pot - 0,52 \cdot rud - 0,16 \cdot vyr$$

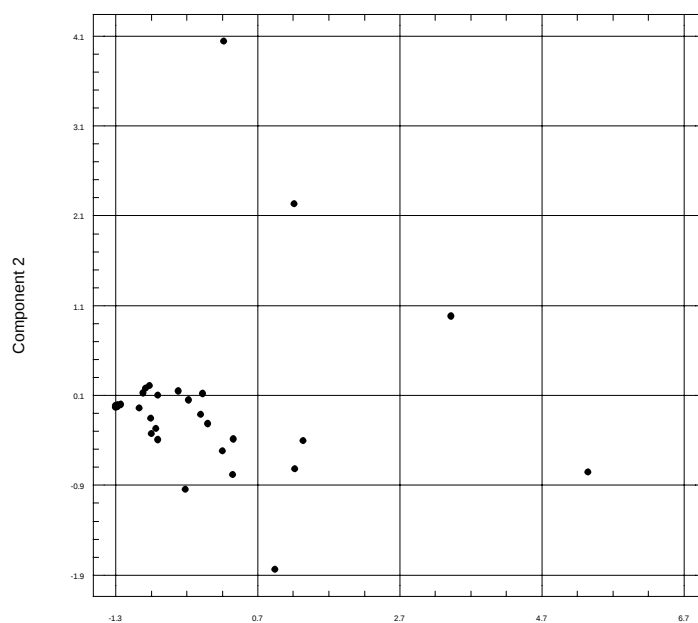


(obr.1.9)

Celkem zůstaly proměnné:

- rud - rudy a šrot pro hutní průmysl
- vyr - hutní výroba
- hno - hnojiva
- pot - potraviny a krmivo

Vezmeme-li v úvahu bodový graf (scatterplot; obr.1.10), dá se předpokládat, že data budou vytvářet 2-3 shluky.



(obr.1.10)

1.4.3 Faktorová analýza

Použijeme-li u faktorové analýzy dva faktory, vyjde nám, že vysvětlují 100% variability v datech (tab.1.10)

Tab.1.10

Variable	Factor	Eigenvalue	Percent Var	Cum Percent
DRAHA.pot	1	1.59499	84.3	84.3
DRAHA.rud	2	0.29782	15.7	100.0
DRAHA.vyr	3	-0.08174	0.0	100.0
DRAHA.hno	4	-0.22886	0.0	100.0

Tab.1.10a

Variable	Communality
DRAHA.pot	0.40118
DRAHA.rud	0.45947
DRAHA.vyr	0.62996
DRAHA.hno	0.09162

Z tabulky 1.10a je vidět, že komunalita (tj. část rozptylu proměnné, která je vysvětlena působením společných faktorů) je u hno velmi malá.

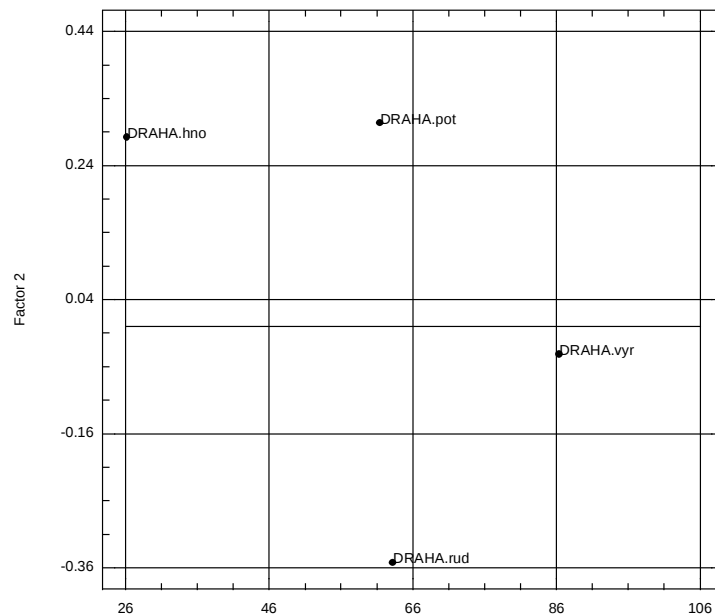
Faktorová matice (tab.1.11) a graf faktorových vah (obr.1.11) před rotací a především faktorová matice po rotaci metodou Varimax a graf faktorových vah po rotaci (obr.1.12) ukazují, že faktorově čisté jsou :

- rud - pro 1.faktor
 - hno, pot - pro 2. Faktor
- a faktorově smíšené jsou vyr.

Tab.1.11

Factor Matrix

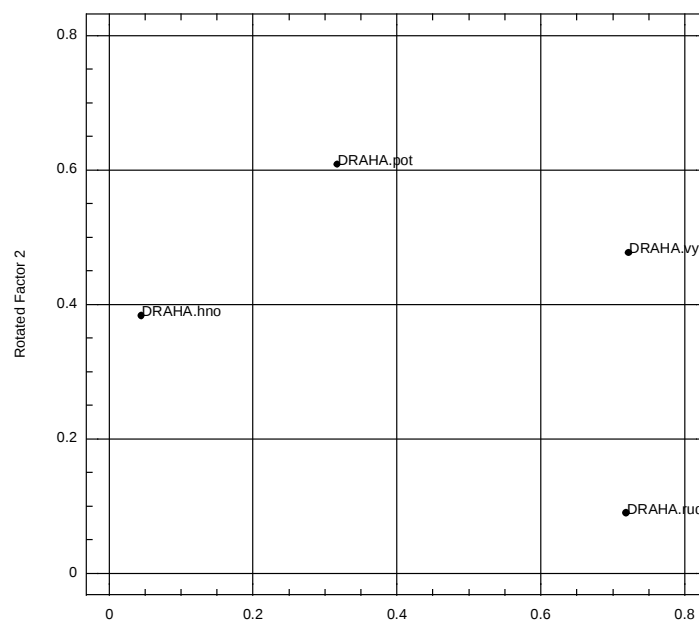
Variable/Factor	1	2
DRAHA.pot	0.61496	0.30303
DRAHA.rud	0.63303	-0.35325
DRAHA.vyr	0.86443	-0.04242
DRAHA.hno	0.26239	0.28180



(obr.11)

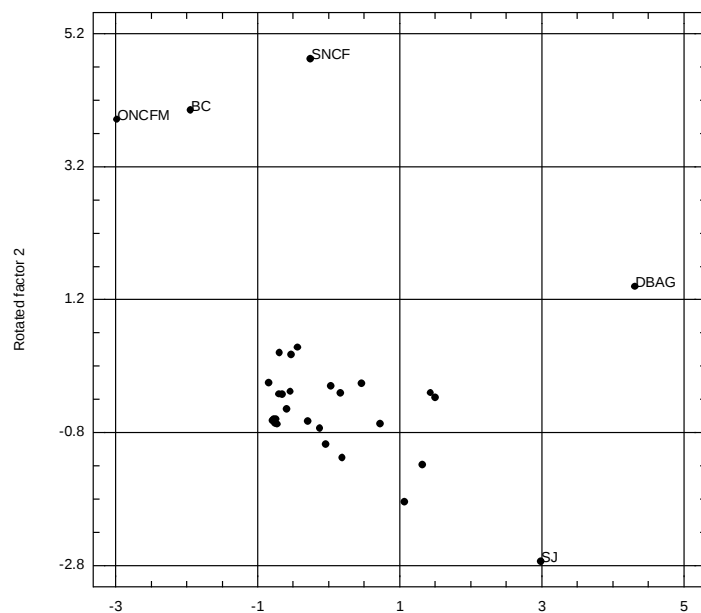
Tab.1.12
VARIMAX ROTATED FACTOR MATRIX

Variable/Factor	1	2
DRAHA.pot	0.31751	0.60761
DRAHA.rud	0.71950	0.08854
DRAHA.vyr	0.72280	0.47603
DRAHA.hno	0.04546	0.38235



(obr.1.12)

Graf faktorových score ukazuje na existenci asi tří shluků (obr.1.13). V jednom bude DBAG, v druhém SNCF, BC a ONCFM a ve třetím ostatní dráhy.

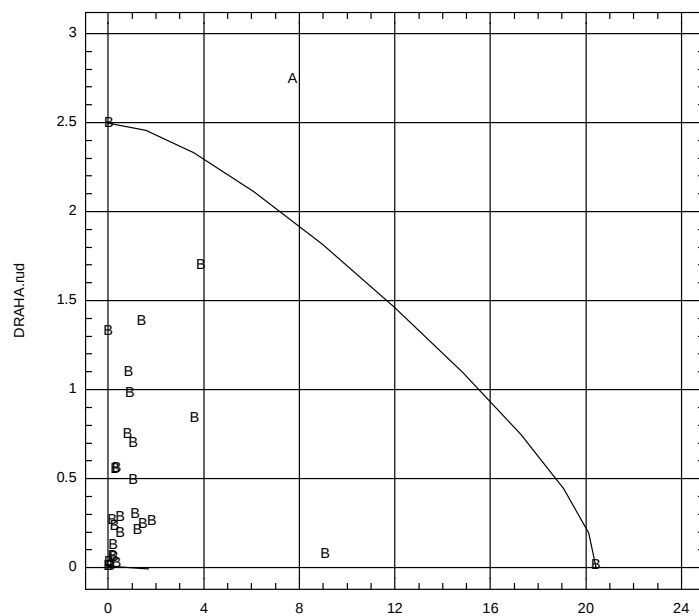


(obr.1.13)

1.4.4 Analýza shluků

Po shlukování průměrem (2 shluky) se vyčlení dráha DBAG (Německo) a ostatní dráhy zůstávají v jednom shluku (obr.1.14)

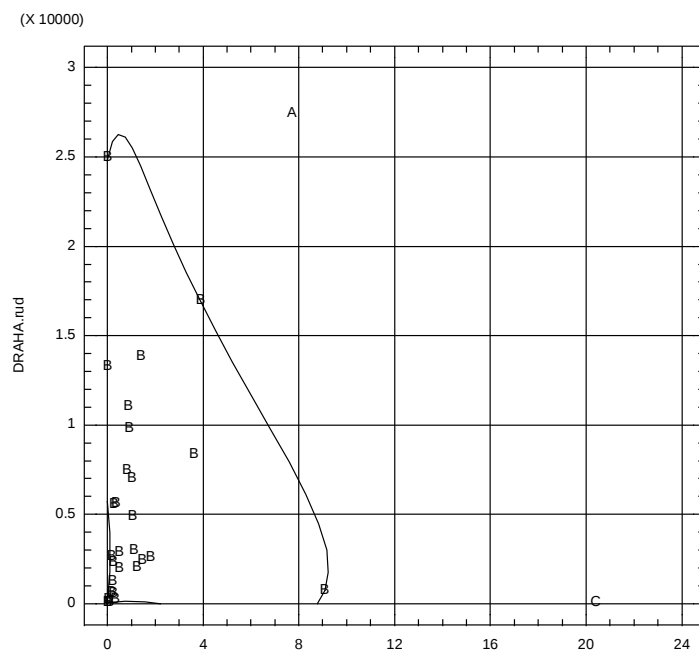
(X 10000)



(obr.1.14)

Použijeme-li při shlukování tři shluky (obr.1.15), vyjde:

1. DBAG (Německo)
2. ostatní
3. ONCFM (Maroko)



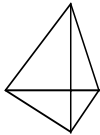
(obr.1.15)

Pro 4 shluky:

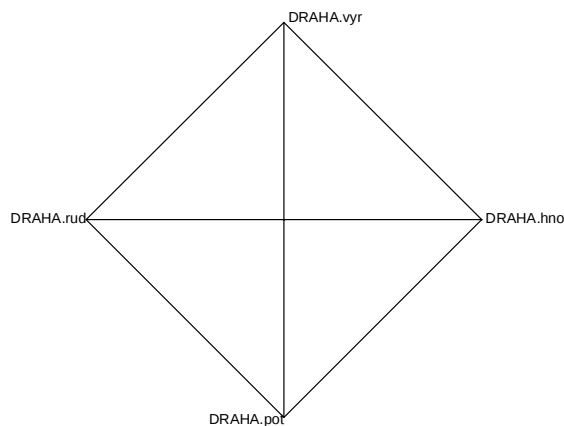
1. DBAG
2. SNCF (Francie)
3. ONCFM
4. ostatní

A až po deset shluků je to podobné, vždy se vydělí jedna železnice.

1.4.5 Hvězdicové grafy

					
ZSR	ONCFM	SNOFT	CFS	ISR	TODD
					
BDZ	CD	CFARYM	MAV	PKP	SZ
					
CFF/SBB/FFS	NSB	OBB	SJ	VR	BC
					
DSB	FS	NS	RENFE	SNCB/NMBS	SNCF
					
BR	CFL	CH	CIE	CP	DBAG

(obr.1.16)



(obr.1.17)

Pomocí hvězdicových grafů (obr.1.16) a jejich porovnáním se vzorovým hvězdicovým grafem (obr.1.17) můžeme rozdělit železnice do následujících skupin:

- CFS, ISR, BDZ, CFARYM, SZ, VR, DSB, NS, CH, CIE, CP (u této skupiny je vliv všech čtyř přepravovaných skupin materiálů přibližně stejný)
- ZSR, SNCFT, TCDD, PKP, NSB, SJ, BR (u této skupiny je silnější vliv skupiny rud)
- ONCFM (jediný představitel s velmi silnou pozicí hnojiv)
- MAV, CFF/SBB/FFS, BC, SNCF (druhá největší železnice), RENFE (silný vliv pot)
- OBB, SNCB/NMBS, FS, CFL, DBAG (tato železnice je nejmasivnější, převládá největší množství nákladů - to však vyplývá z pozice Německa jako jedné z největších a průmyslově nejvyspělejších evropských zemí s velmi rozsáhlou železniční sítí) (zde je silnější vliv rud a vyr).

1.5 Závěr

Metoda hlavních komponent nám pomohla zredukovat počet proměnných. Výsledkem použití této metody je, že zbyly čtyři skupiny přepravovaného zboží:

- potraviny a krmivo
 - rudy a šrot pro hutní průmysl
 - hnojiva
 - hutní výroba.
- Ke skupině potraviny a krmivo můžeme ještě přiřadit zemědělské produkty a živá zvířata (párový korelační koeficient $r = 0,54$) - jedná se o přepravované zboží velmi podobného charakteru.
 - Ke skupině rudy a šrot pro hutní výrobu můžeme přiřadit pevná nerostná paliva ($r = 0,55$). Opět zde existuje logická souvislost mezi hutnictvím (tj. výrobou kovů) a množstvím používaného (tj. i přepravovaného) nerostného paliva.
 - Ke skupině hutní výroba můžeme přiřadit chemické výrobky ($r = 0,81$), vozidla, stroje polotovary i hotové výrobky ($r = 0,87$), ropné produkty ($r = 0,93$) a rudy surové a opracované stavební materiály ($r = 0,85$). U těchto proměnných vypadá redukovaná korelační matice následovně:

	vyr	che	voz	mat	rop
vyr	1	0,81	0,87	0,85	0,79
che	0,81	1	0,67	0,79	0,68
voz	0,87	0,67	1	0,72	0,68
mat	0,85	0,79	0,72	1	0,93
rop	0,79	0,68	0,68	0,93	1

Všechny korelační koeficienty jsou poměrně vysoké, proto se dá předpokládat lineární závislost mezi jednotlivými proměnnými. Logická souvislost mezi jednotlivými skupinami výrobků se zde také dá vysledovat, jedná o zboží, které má spotřební charakter, nejsou to suroviny.

- V poslední skupině se nachází pouze hnojiva, která sice nemají (z hlediska PCA a korelační analýzy) vazbu na ostatní přepravované skupiny, ale dá se předpokládat vazba na chemické výrobky i na zemědělské produkty.

Faktorová analýza poskytuje dva faktory (dvě vlastnosti). První, sem patří rud, je odrazem úrovně a rozvoje hutnictví a těžby (surovin) jako takové. S tím souvisí i úroveň těžkého a strojírenského průmyslu. Druhý, sem patří hno a pot, je odrazem úrovně zemědělství a do jisté míry i úrovně chemického průmyslu (výroba hnojiv). Vyr, (che, voz, mat, rop) jsou faktorově smíšené. Tento závěr je odrazem souvislosti mezi danými druhy přepravovaného zboží a úrovní průmyslu a zemědělství v daném státě. (I když se jedná o určité zjednodušení, poněvadž tento rozbor opomíjí ostatní druhy přeprav.)

V analýze shluků asi nejlepší obraz dává použití tří shluků, a to:

- DBAG(Německo) - největší množství přepravovaného zboží, velmi vysoká úroveň průmyslu i zemědělství
- ONCFM(Maroko) - jednostranně zaměřená přeprava, a to na hnojiva (chemický průmysl ?)
- Ostatní dráhy - proporcionálně podobná skladba přepravovaného zboží.

Použijeme-li hvězdicové grafy, tak vyjde pět skupin, i když je zde opět patrná význačná pozice DBAG (Německé dráhy) a i rozdílná skladba přepravovaného zboží u ONCFM (Maroko). U ostatních železnic je skladba přepravovaného zboží poměrně podobná, rozdíly nejsou příliš velké. Rozdíly jsou v množství přepravovaného zboží, to ovšem ovlivňuje i rozloha státu, jeho poloha, počet kilometrů kolejí a řada dalších faktorů.

2. Rozdělení předmětů a prospěch absolventů

2.1 Zadání:

Na VDA byly v průběhu několika sledovány výsledky absolventů oboru logistika a management. Naším úkolem je redukovat nadbytečný počet předmětů, definovat charakter předmětů (tj. dovednosti, které tento předmět rozvíjí a roztrždit studenty podle těchto vlastností.

2.2 Data:

Tab.2.1 Zkratky a názvy předmětů

Zkratka	Název	Stručná charakteristika
MAT	Matematika	základy dif. a integ. počtu - základní matematické postupy a znalosti
STA	Statistika	základní pojmy a dovednosti z pravděpodobnosti a statistiky
KOM	Komunikace	prohlubuje jazykové dovednosti a rozvíjí stylistické a komunikativní dovednosti
CIJ1	Cizí jazyk	jeden z jazyků - ANJ, NEJ, RUJ (uváděn jako hlavní jazyk)
CIJ2	Cizí jazyk	jeden z jazyků - ANJ, NEJ, RUJ
MAM1	Materiálový management	plánování, řízení a provádění materiálového toku v oblasti od dodavatele až po finální výrobek
MFD	Management fyzické distribuce	materiálové toky - hledání vhodných modelů pomocí metod aplikované matematiky
TEC	Technologie	základní znalosti o vlastnostech a zkouškách různých materiálů, základní měřicí metody
MEC	Mechanika	základní technické znalosti a dovednosti (statika, kinematika, dynamika, ..), základy technického kreslení
DPP	Dopravní prostředky	rozdělení a princip silničních a kolejových vozidel, znalost konstrukce, jízdních a opravárenských vlastností, dopravní systémy
MEP	Mechanizace podniku	manipulace s materiálem uvnitř a vně podniku, skladové systémy, vnitropodniková doprava a přepravní prostředky
NEZ	Nebezpečné zboží	základní chemické znalosti, předpisy o přepravě nebezpečných látek
EKO	Obecná ekonomika	základní pojmy ekonomické teorie
EKP1	Ekonomika podniku	obecná charakteristika podniku a základní pojmy a metody používané v podnikové ekonomice
EKD	Ekonomika dopravy	základní pojmy, znalosti a metody používané při řešení ekonomických problémů v různých druzích doprav
MAR	Marketing	základní pojmy - trh, tržní mechanismy, zákazník, marketingové řízení,...
INF	Informatika	znalosti hardwarových prostředků a nejrozšířenějších softwarových produktů
AZP	Administrativní základy přepravy	přehled o právních a tarifních základech všech oborů doprav v ČR , vnitrostátní i mezinárodní

Tab.2.2 Studijní výsledky studentů

(Číslo je katalogové číslo studenta, klasifikace je prováděna pomocí známek 1-3)

CISLO	TEC	MAR	MEC	EKO	CIJ1	MAM1	NEZ	EKP1	STA	INF	MAT	CIJ2	MAM2	MEP
0192	1	1	3	3	1	3	1	2	1	2	2	2	1	2
0292	1	1	3	3	2	3	2	1	2	2	2	1	2	3
0392	1	1	1	2	3	2	2	2	2	1	3	3	1	2
0492	1	2	1	3	2	2	2	1	2	3	1	2	1	1
0892	1	3	3	3	3	3	2	3	2	3	3	3	1	2
1092	3	3	3	3	3	3	2	3	3	3	3	3	2	3
1292	2	1	2	3	1	2	3	2	3	3	3	2	1	3
1492	2	2	3	3	1	3	2	3	2	2	2	3	2	2
1692	1	1	2	3	1	3	2	2	2	1	2	1	1	1
1892	3	1	1	3	2	3	3	3	1	3	2	3	1	3
1992	2	1	2	3	2	3	3	3	3	3	2	3	3	3
2292	1	1	3	1	1	2	1	2	1	3	1	1	1	1
2392	1	1	1	3	3	2	1	1	2	2	2	2	2	2
2492	3	2	2	3	2	3	3	2	2	3	3	2	2	3
2592	1	2	3	3	2	3	2	3	3	2	3	3	2	2
2792	1	1	3	2	2	2	2	2	1	1	3	2	1	1
3492	2	1	1	3	1	2	2	3	2	2	2	2	2	2
3592	2	2	2	3	3	3	2	2	3	3	1	2	1	2
3792	2	2	2	3	3	3	3	3	3	3	3	3	2	1
3892	1	1	2	1	3	1	2	1	2	2	2	2	1	2
3992	1	3	1	3	2	2	2	1	2	3	2	3	1	2
4192	2	2	2	3	2	2	3	2	2	3	1	3	2	3
4292	1	2	2	3	3	3	3	2	2	3	2	3	1	2
4492	2	2	3	3	1	3	1	1	1	2	2	3	1	2
4792	1	2	1	3	2	2	2	1	2	2	3	3	1	2
4892	2	2	3	3	3	2	2	1	3	2	3	3	2	1
0593	2	2	2	3	3	3	2	2	2	2	2	3	1	1
0893	3	3	2	3	2	3	2	3	2	3	2	2	1	1
1193	3	1	3	2	2	2	1	2	2	2	1	3	1	1
1293	2	2	3	2	2	1	1	2	1	1	2	2	1	2
1493	3	3	2	2	3	2	1	2	2	3	1	3	1	2
1593	1	1	2	3	2	3	2	1	2	3	1	3	2	2
1993	3	2	3	3	2	3	2	2	2	3	2	2	1	2
2093	2	1	2	3	1	3	2	1	3	1	3	3	1	2
2193	3	2	1	2	2	3	3	2	3	2	1	3	1	2
2393	3	1	3	3	1	2	3	2	3	2	2	2	1	2
2993	1	2	3	2	3	2	1	1	3	1	3	2	1	2
3193	3	1	2	2	3	3	2	3	3	1	1	2	1	3
3293	3	3	1	3	2	2	1	1	1	1	3	3	1	1
3593	3	1	3	3	2	3	1	3	1	3	1	2	1	2
4493	3	2	3	3	3	3	2	3	2	2	2	3	1	3
5093	3	1	3	2	1	3	3	1	3	3	2	2	2	2
5593	3	3	3	3	2	3	3	3	2	2	2	3	1	1
5693	3	3	2	3	2	3	1	2	2	2	3	3	2	2
5993	2	1	3	3	3	3	2	1	3	3	3	3	2	2
6493	3	2	3	3	3	3	2	3	2	3	2	2	1	1
6593	2	3	2	3	2	2	1	2	1	3	2	3	2	2
6693	2	1	3	3	3	3	3	2	2	3	2	2	1	3

6793	3	3	2	3	3	3	3	2	3	1	1	3	1	2
6893	3	2	2	3	2	3	1	2	3	3	2	2	1	1
6993	2	2	1	1	2	2	1	1	1	1	1	1	1	1
0294	3	3	3	3	3	3	1	2	3	3	3	3	2	3
0394	3	1	3	3	1	3	2	1	3	3	3	2	1	2
0594	2	1	3	3	1	2	1	1	1	1	1	2	1	1
0794	2	1	3	3	3	3	2	1	3	3	2	1	1	2
0894	3	3	3	3	2	3	3	3	3	3	3	3	2	3
0994	3	2	3	3	3	3	3	2	3	3	3	2	2	3
1094	1	2	3	3	2	2	1	2	2	2	3	2	1	1
1394	2	1	3	2	3	3	2	2	2	3	2	2	1	2
1494	2	2	2	2	1	2	2	2	2	3	2	1	1	2
1594	1	1	2	1	1	2	1	1	2	2	2	1	1	1
1694	2	1	3	3	1	3	1	3	3	3	3	2	1	3
2194	3	1	3	3	3	3	2	1	2	3	2	2	1	3
2494	1	2	2	2	1	3	2	2	1	2	1	1	1	2
2594	2	1	3	3	3	3	2	2	2	3	3	2	1	2
2794	2	1	3	3	3	3	2	2	2	3	3	2	3	2
2694	3	1	3	3	2	3	3	2	2	2	2	1	1	3
2994	2	2	1	3	3	2	2	1	2	3	3	2	1	2
3294	2	2	3	3	3	3	2	1	3	3	3	2	3	2
3494	3	1	3	2	2	2	2	2	2	3	2	1	1	3
3694	2	2	3	3	1	3	2	2	3	3	3	2	1	3
3794	2	1	2	3	2	2	3	1	3	2	3	1	1	2
3894	2	1	2	3	2	3	2	2	1	2	3	2	2	2
3994	3	3	3	3	3	3	2	3	3	3	3	3	1	2
4094	3	1	2	2	1	3	1	1	1	3	2	1	1	1
4194	1	3	3	2	1	3	2	1	3	3	1	2	1	2
4394	2	1	3	3	3	3	2	1	3	2	3	3	1	1
4494	2	1	3	3	1	3	1	2	2	3	3	1	1	2
4594	2	2	3	3	3	3	2	2	2	2	2	2	1	1
4994	3	1	2	3	1	3	2	2	2	2	2	2	1	3
5094	2	2	3	3	3	3	2	3	3	3	3	2	3	2
5194	3	1	2	3	1	2	3	1	2	2	2	1	2	3
5394	2	2	3	2	2	3	1	3	2	3	1	2	2	2
5594	2	1	1	3	2	3	2	3	2	2	2	2	2	2
5694	3	2	3	3	3	2	2	3	3	3	2	2	2	3
5994	3	2	3	3	3	3	2	2	3	3	3	1	1	3
6094	3	1	2	3	1	3	2	2	2	3	2	1	2	3
6194	2	1	3	3	2	2	1	1	2	2	3	2	1	1
6294	3	2	3	3	2	3	1	3	2	3	3	2	1	2
6394	3	3	3	3	3	3	3	3	2	3	2	3	1	3
6894	3	3	2	2	3	3	3	1	2	3	2	3	3	3
6994	3	3	3	3	3	3	2	2	2	3	3	1	1	2
7094	1	1	3	3	1	3	1	2	1	1	2	1	1	1
7194	3	1	3	3	3	3	2	2	1	3	2	2	1	2
7494	3	1	3	3	3	3	2	2	2	2	2	2	1	2
7594	3	1	3	2	1	3	2	1	3	2	3	2	1	2
8094	3	3	3	3	3	3	1	2	3	1	3	3	1	2

Tab.2.2 - pokračování

CISLO	EKD	KOM	MFD	DPP	EKP2	AZP
0192	2	1	1	2	1	1
0292	1	1	3	2	2	1
0392	2	2	2	2	1	2
0492	2	1	2	2	1	1
0892	3	1	2	3	2	3
1092	3	3	3	2	3	3
1292	3	1	3	3	1	3
1492	3	2	3	2	2	3
1692	2	1	2	1	2	2
1892	3	2	3	2	3	2
1992	3	2	3	1	2	3
2292	2	1	2	2	1	1
2392	1	2	1	1	2	1
2492	2	2	3	2	2	2
2592	3	2	3	3	2	3
2792	3	1	3	1	1	3
3492	1	1	2	3	1	2
3592	3	1	2	2	2	1
3792	2	2	3	3	1	2
3892	2	1	1	2	1	1
3992	2	1	1	2	2	1
4192	3	1	3	2	2	3
4292	2	2	2	1	2	2
4492	2	1	1	2	1	1
4792	3	1	3	2	2	2
4892	1	1	1	1	1	1
0593	3	1	3	3	3	3
0893	3	2	3	3	2	2
1193	1	1	2	1	1	2
1293	1	3	1	1	1	1
1493	3	2	2	2	2	3
1593	3	1	3	2	2	2
1993	2	2	2	2	1	2
2093	2	2	2	2	1	3
2193	2	1	3	2	2	2
2393	2	1	2	3	2	2
2993	1	1	1	2	2	2
3193	3	1	3	3	3	3
3293	1	2	1	2	1	2
3593	3	1	2	3	2	2
4493	2	3	3	3	1	2
5093	2	2	3	3	2	2
5593	3	1	2	2	2	1
5693	2	2	2	2	2	2
5993	3	2	3	3	3	3
6493	1	1	3	2	2	1
6593	3	2	3	3	2	1

6693	3	2	3	3	1	2
6793	3	1	2	3	2	3
6893	1	1	2	2	1	1
6993	1	1	2	1	1	1
0294	2	2	3	2	3	3
0394	2	2	3	3	2	1
0594	1	1	3	1	1	1
0794	3	1	3	3	2	1
0894	3	3	3	1	3	3
0994	2	2	3	3	3	2
1094	1	2	3	2	1	1
1394	1	2	3	2	2	1
1494	2	1	2	1	2	1
1594	1	2	1	2	1	2
1694	3	1	3	2	2	3
2194	2	2	2	1	2	2
2494	1	1	3	3	2	1
2594	2	2	3	2	2	1
2794	2	2	3	3	3	2
2694	2	1	3	3	1	2
2994	1	1	2	1	2	1
3294	3	2	3	3	3	3
3494	3	1	2	3	1	3
3694	3	1	3	3	3	3
3794	1	1	2	3	1	2
3894	1	1	3	3	2	1
3994	3	3	3	3	2	3
4094	1	1	2	2	2	1
4194	1	1	2	2	2	2
4394	3	2	3	2	2	1
4494	1	1	3	3	2	3
4594	2	1	3	2	3	2
4994	1	2	3	3	2	1
5094	3	2	3	2	1	1
5194	3	1	3	2	2	2
5394	2	3	3	3	3	3
5594	2	1	3	3	2	1
5694	1	2	3	2	2	1
5994	1	1	3	3	2	1
6094	2	1	3	3	1	1
6194	2	1	2	1	1	2
6294	3	1	3	3	2	3
6394	3	2	3	3	3	3
6894	3	1	3	3	2	2
6994	3	2	3	1	2	2
7094	1	3	2	2	2	2
7194	3	1	3	3	2	2
7494	3	1	2	3	2	1
7594	3	3	3	2	2	2
8094	3	2	3	2	2	2

2.3 Program:

STATGRAPHICS- vícerozměrné statistiky, EXCEL - analýza dat

2.4 Řešení:

2.4.1 Korelační analýza

Tab.2.3 Korelační matice

	TEC	MAR	MEC	EKO	CIJ1	MAM1	NEZ	EKP1	STA	INF	MAT	CIJ2	MAM2
TEC	1,000	0,189	0,169	0,186	0,135	0,302	0,199	0,262	0,184	0,220	0,027	0,088	0,023
MAR	0,189	1,000	-0,025	0,116	0,306	0,073	0,005	0,221	0,136	0,143	0,064	0,450	0,066
MEC	0,169	-0,025	1,000	0,105	0,085	0,301	-0,113	0,163	0,176	0,161	0,207	-0,108	-0,011
EKO	0,186	0,116	0,105	1,000	0,141	0,397	0,210	0,216	0,186	0,184	0,354	0,254	0,174
CIJ1	0,135	0,306	0,085	0,141	1,000	0,091	0,148	0,123	0,235	0,149	0,160	0,338	0,137
MAM1	0,302	0,073	0,301	0,397	0,091	1,000	0,206	0,320	0,222	0,287	0,092	0,093	0,148
NEZ	0,199	0,005	-0,113	0,210	0,148	0,206	1,000	0,134	0,350	0,220	0,029	0,147	0,212
EKP1	0,262	0,221	0,163	0,216	0,123	0,320	0,134	1,000	0,045	0,194	-0,008	0,199	0,109
STA	0,184	0,136	0,176	0,186	0,235	0,222	0,350	0,045	1,000	0,188	0,319	0,193	0,209
INF	0,220	0,143	0,161	0,184	0,149	0,287	0,220	0,194	0,188	1,000	0,039	-0,012	0,250
MAT	0,027	0,064	0,207	0,354	0,160	0,092	0,029	-0,008	0,319	0,039	1,000	0,109	0,159
CIJ2	0,088	0,450	-0,108	0,254	0,338	0,093	0,147	0,199	0,193	-0,012	0,109	1,000	0,186
MAM2	0,023	0,066	-0,011	0,174	0,137	0,148	0,212	0,109	0,209	0,250	0,159	0,186	1,000
MEP	0,314	-0,003	0,077	0,191	0,085	0,202	0,415	0,206	0,268	0,327	0,090	0,031	0,282
EKD	0,199	0,185	0,138	0,210	0,206	0,288	0,331	0,294	0,213	0,253	0,068	0,368	0,134
KOM	0,172	0,211	0,214	0,059	0,170	0,161	-0,018	0,234	0,137	0,057	0,275	0,279	0,217
MFD	0,298	-0,023	0,222	0,285	0,097	0,502	0,355	0,362	0,235	0,325	0,153	0,023	0,297
DPP	0,229	-0,026	0,114	0,168	0,030	0,355	0,219	0,248	0,167	0,233	0,056	-0,036	0,099
EKP2	0,208	0,245	0,131	0,210	0,262	0,468	0,187	0,215	0,246	0,294	0,076	0,189	0,256
AZP	0,138	0,176	0,160	0,083	0,040	0,199	0,181	0,311	0,302	0,058	0,172	0,327	0,127

Tab.2.3 - pokračování

	MEP	EKD	KOM	MFD	DPP	EKP2	AZP
TEC	0,314	0,199	0,172	0,298	0,229	0,208	0,138
MAR	-0,003	0,185	0,211	-0,023	-0,026	0,245	0,176
MEC	0,077	0,138	0,214	0,222	0,114	0,131	0,160
EKO	0,191	0,210	0,059	0,285	0,168	0,210	0,083
CIJ1	0,085	0,206	0,170	0,097	0,030	0,262	0,040
MAM1	0,202	0,288	0,161	0,502	0,355	0,468	0,199
NEZ	0,415	0,331	-0,018	0,355	0,219	0,187	0,181
EKP1	0,206	0,294	0,234	0,362	0,248	0,215	0,311
STA	0,268	0,213	0,137	0,235	0,167	0,246	0,302
INF	0,327	0,253	0,057	0,325	0,233	0,294	0,058
MAT	0,090	0,068	0,275	0,153	0,056	0,076	0,172
CIJ2	0,031	0,368	0,279	0,023	-0,036	0,189	0,327
MAM2	0,282	0,134	0,217	0,297	0,099	0,256	0,127
MEP	1,000	0,284	0,147	0,341	0,276	0,308	0,291
EKD	0,284	1,000	0,089	0,377	0,243	0,336	0,512
KOM	0,147	0,089	1,000	0,172	-0,052	0,199	0,213
MFD	0,341	0,377	0,172	1,000	0,376	0,434	0,273
DPP	0,276	0,243	-0,052	0,376	1,000	0,213	0,210
EKP2	0,308	0,336	0,199	0,434	0,213	1,000	0,356
AZP	0,291	0,512	0,213	0,273	0,210	0,356	1,000

determinant = 0,005

Hodnota determinantu je velmi malá, proto se dá předpokládat multikolinearita, tj. lineární závislost mezi proměnnými, i když samotné hodnoty párových korelačních koeficientů jsou poměrně malé.

2.4.2 Metoda hlavních komponent

Poněvadž Statgraphics může najednou pracovat pouze se třinácti proměnnými, zadáme tyto předměty (jsou uvedeny pouze ve zkratkách):

TEC, MAR, MEC, EKO, CIJ1, MAM1, NEZ, EKP1, STA, INF, MAT, CIJ2, MAM2.

Po provedení výpočty v metodě hlavních komponent pro daných třináct proměnných, vychází, že první dvě latentní proměnné popisují 35% variability v datech a první tři 45% (tab.2.4).

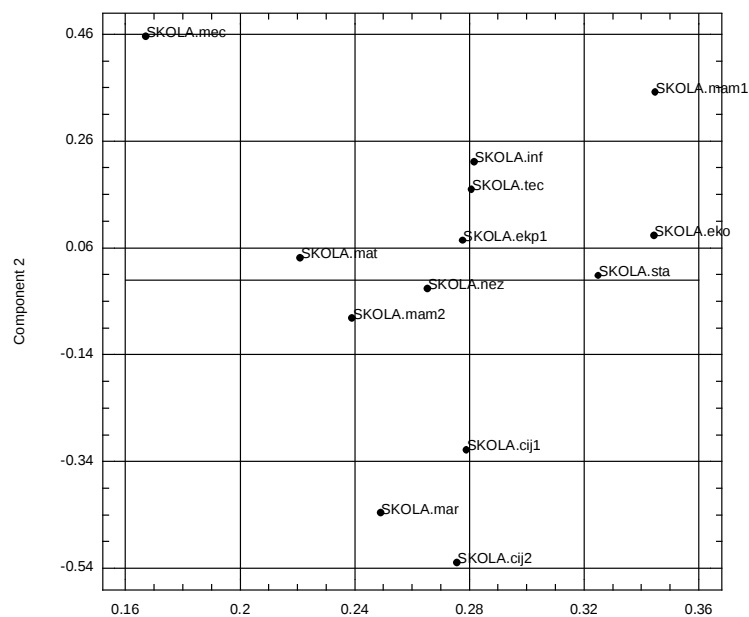
Tab.2.4

Principal Components Analysis

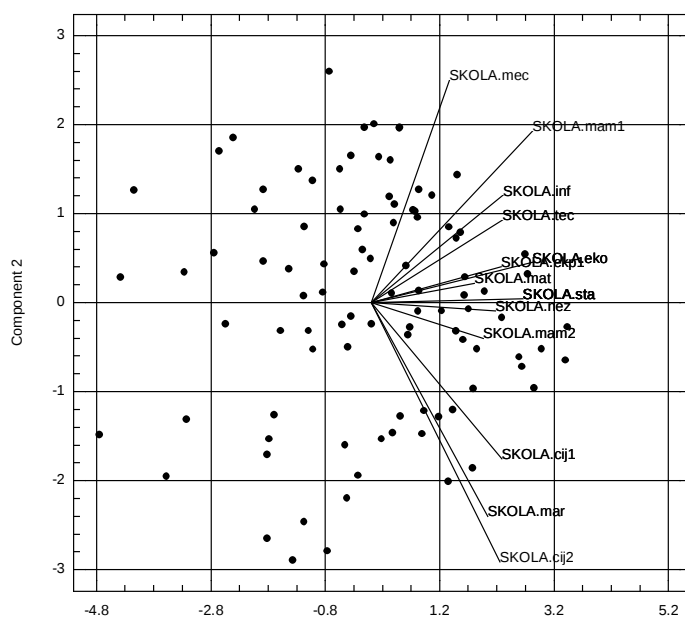
Component Number	Percent of Variance	Cumulative Percentage
1	23.16701	23.16701
2	11.80889	34.97591
3	10.10154	45.07744
4	9.39085	54.46830
5	7.56751	62.03581
6	7.13700	69.17280
7	5.63728	74.81008
8	5.34754	80.15762
9	4.95212	85.10974
10	4.80044	89.91018
11	3.64752	93.55771
12	3.36818	96.92589
13	3.07411	100.00000

Graf komponentních vah (obr.2.1) a dvojný graf (obr.2.2) ukazují, že podobné vlastnosti (koreluji spolu) mají :

- MAM1, INF, TEC
- EKP1, EKO, MAT
- STA, NEZ, MAM2
- CIJ1, CIJ2, MAR



(obr.2.1)



(obr.2.2)

Odstraníme - INF, TEC, EKP1, MAT, NEZ, MAM2, CIJ1, MAR.

Zůstávají předměty - MEC, EKO, MAM1, STA, CIJ2; a k nim přidáme zbývající předměty - MEP, EKD, KOM, MFD, DPP, EKP2, AZP. Poté opět řešíme pomocí metody hlavních komponent ve Statgraphicsu a vyjde, že první dvě latentní proměnné popisují 41,9% variability v datech a první tři 51,33% (tab.2.5).

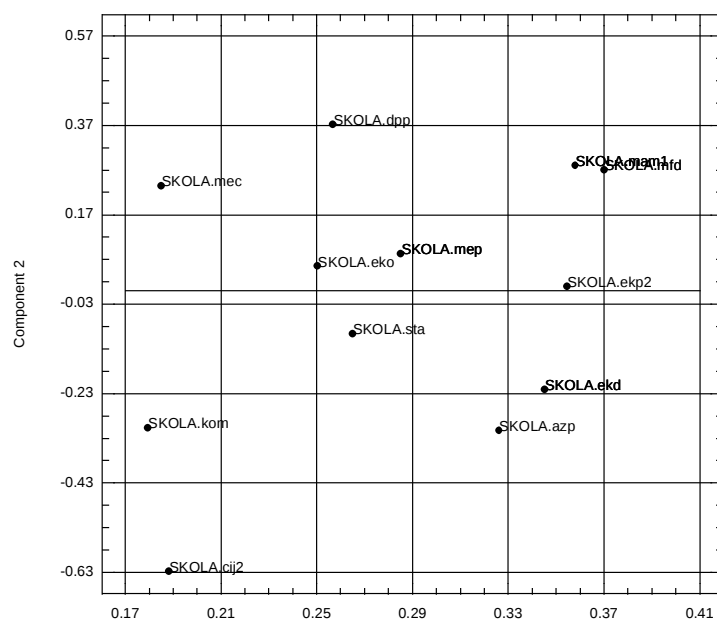
Tab.2.5
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	29.87405	29.87405
2	12.01078	41.88483

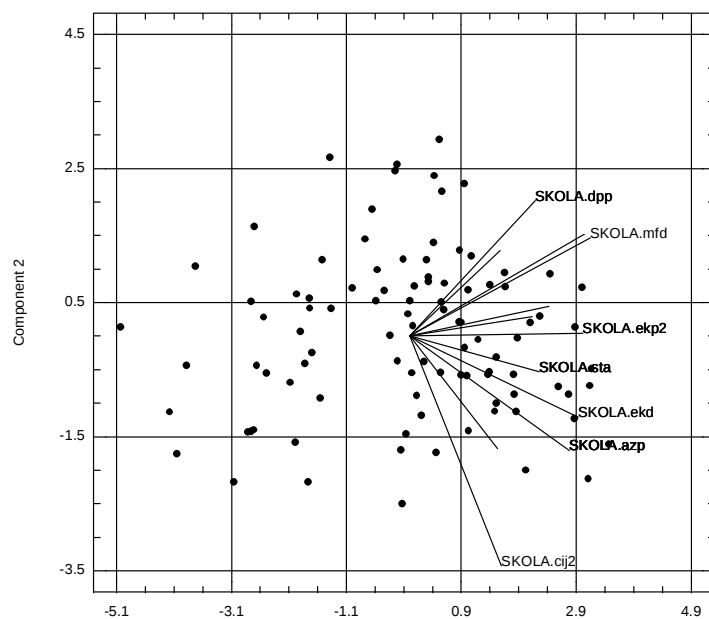
3	9.44750	51.33233
4	8.74703	60.07937
5	7.17752	67.25689
6	7.00090	74.25779
7	5.89418	80.15197
8	5.55188	85.70385
9	4.41694	90.12079
10	3.71420	93.83499
11	3.28726	97.12225
12	2.87775	100.00000

Graf komponentních vah (obr.2.3) a dvojný graf (obr.2.4) ukazují na podobnost (korelovanost) následujících předmětů :

- MEC, DPP
- MAM1, MFD
- MEP, EKO, EKP2, STA
- EKD, AZP
- KOM, CIJ2



(obr.2.3)



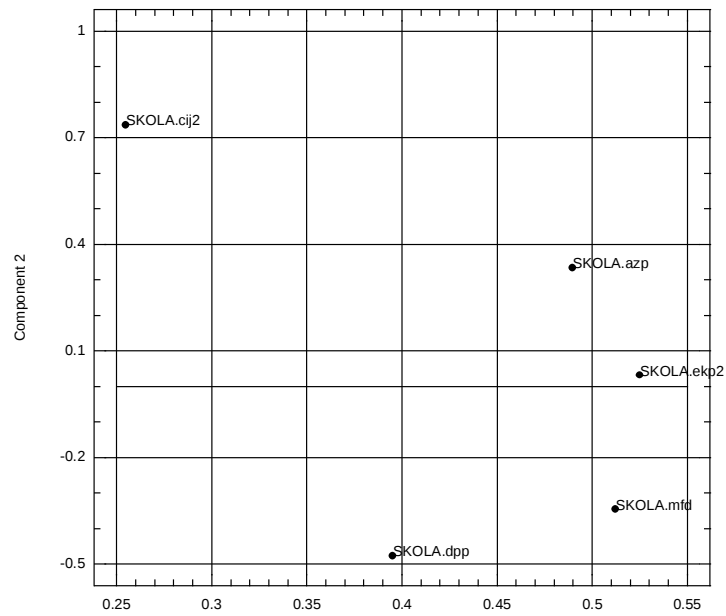
(obr.2.4)

Odstraníme - MEC, MAM1, EKO, STA, EKD, KOM; a zůstanou předměty - DPP (dopravní prostředky), MFD (management fyzické distribuce), EKP2 (ekonomika podniku 2), CIJ2 (cizí jazyk 2). Opět řešíme metodou PCA (metoda hlavních komponent) a dostáváme, že první dvě latentní proměnné vysvětlují 63,53% variability v datech a první tři 78,15% (tab.2.6)

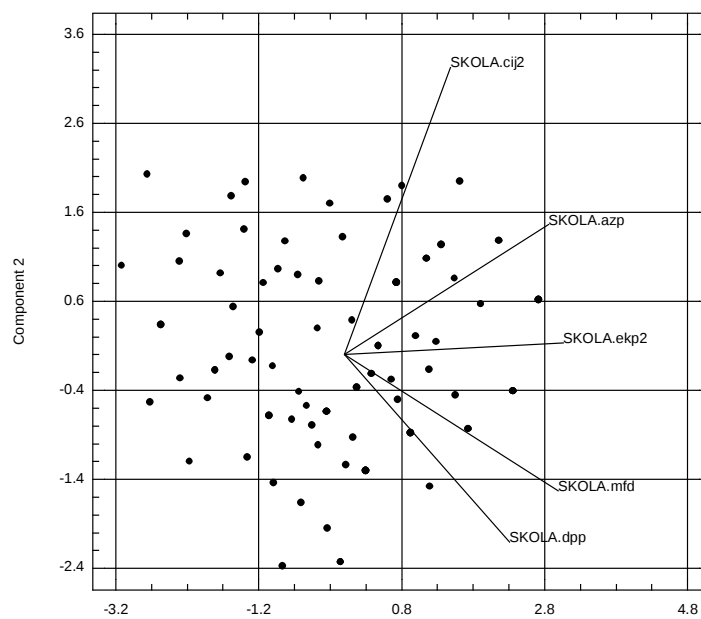
Tab.2.6
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	40.15874	40.15874
2	23.37439	63.53313
3	14.61297	78.14609
4	11.73268	89.87877
5	10.12123	100.00000

Rozvrstvení proměnných ukazují graf komponentních vah (obr.2.5) a dvojný graf (obr.2.6). Zbylé proměnné jsou nekorelované (žádný z bodů ani žádná z úseček se nenacházejí blízko sebe). Graf komponentních vah (obr.2.5) ukazuje, že CIJ2 se výrazně liší od ostatních zbylých předmětů, bude tedy charakterizovat jinou vlastnost.

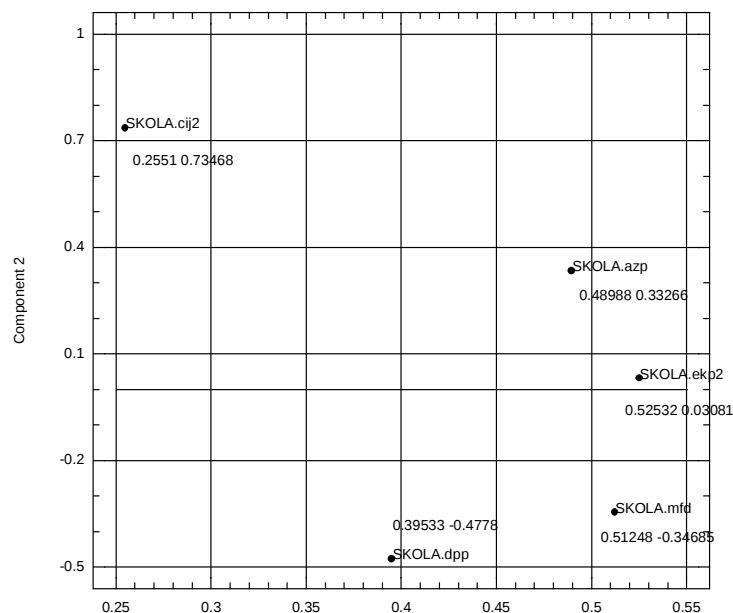


(obr.2.5)



(obr.2.6)

Chceme-li znát příspěvek daného předmětu do latentních proměnných, musíme vyznačit v grafu komponentních vah souřadnice bodů (obr.2.7).



(obr.2.7)

Vyjde následující vztah

$$y_1 = 0,26*CIJ2 + 0,49*AZP + 0,4*DPP + 0,51*MFD + 0,53*EKP2$$

$$y_2 = 0,73*CIJ2 + 0,33*AZP - 0,48*DPP - 0,35*MFD + 0,03*EKP2$$

2.4.3 Faktorová analýza

Při použití algoritmů faktorové analýzy pro dva faktory vyjdou poměrně nízké hodnoty komunalit (tab.2.7), a to především u CIJ2 a DPP.

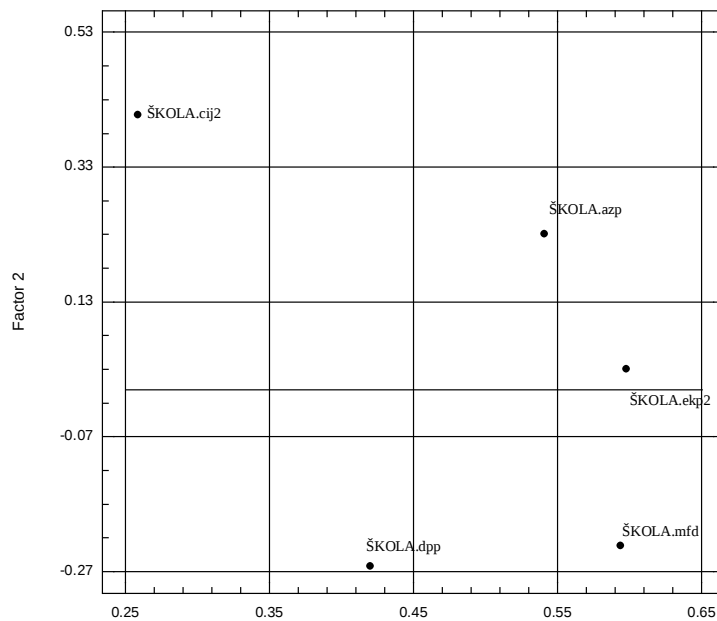
Tab.2.7

Variable	Communality
SKOLA.cij2	0.13276
SKOLA.mfd	0.28440
SKOLA.dpp	0.16301
SKOLA.ekp2	0.26129
SKOLA.azp	0.23313

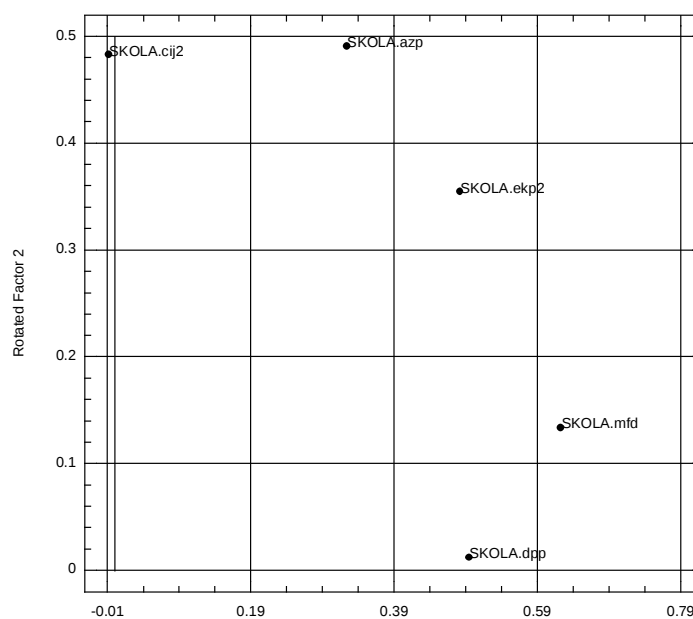
Na základě údajů ve faktorové matici před rotací (tab.2.8) a především po rotaci (tab.2.9) a grafů faktorových vah před rotací (obr.2.8) a po rotaci (obr.2.9) můžeme říci, že faktorově čisté vyjdou :

- DPP a MFD pro první faktor
- CIJ2 pro druhý faktor.

Faktorově smíšené jsou EKP2 a AZP.



(obr.2.8)



(obr.2.9)

Tab.2.8
Factor Matrix

Variable/Factor	1	2
SKOLA.cij2	0.25911	0.40679
SKOLA.mfd	0.59391	-0.23192
SKOLA.dpp	0.42007	-0.26277
SKOLA.ekp2	0.59791	0.03006
SKOLA.azp	0.54109	0.23054

Tab.2.9

VARIMAX ROTATED FACTOR MATRIX

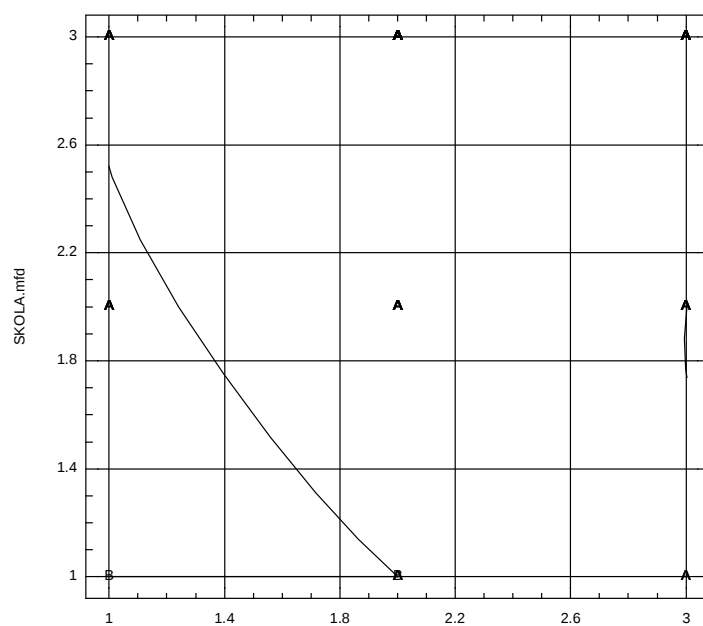
Variable/Factor	1	2
SKOLA.cij2	-0.00744	0.48224
SKOLA.mfd	0.62353	0.13310
SKOLA.dpp	0.49535	0.01170
SKOLA.ekp2	0.48274	0.35407
SKOLA.azp	0.32499	0.49021

2.4.4 Shlukování

Použitím metody shlukování pro dva shluky vyjde

- 95 studentů
- studenti s katalogovými čísly 1594 a 2993

(obr.2.10)

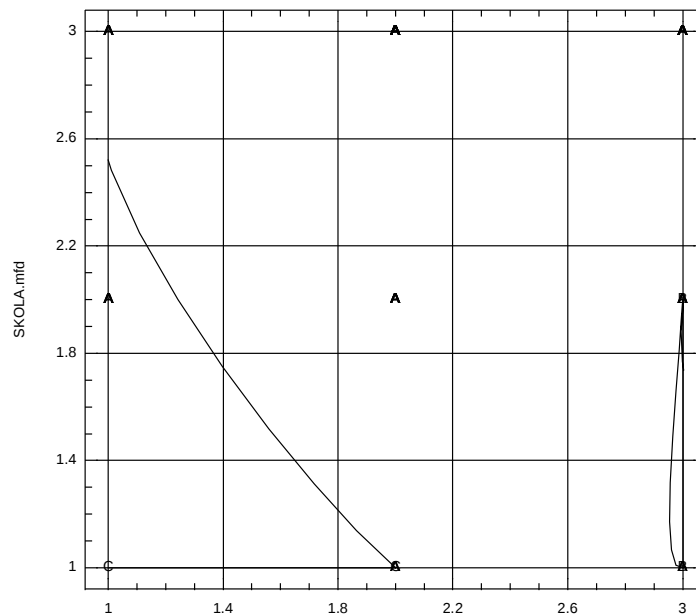


(obr.2.10)

Pro tři shluky vyjde

- 93 studentů
- 3992, 5593
- 2993, 1594

(obr.2.11)



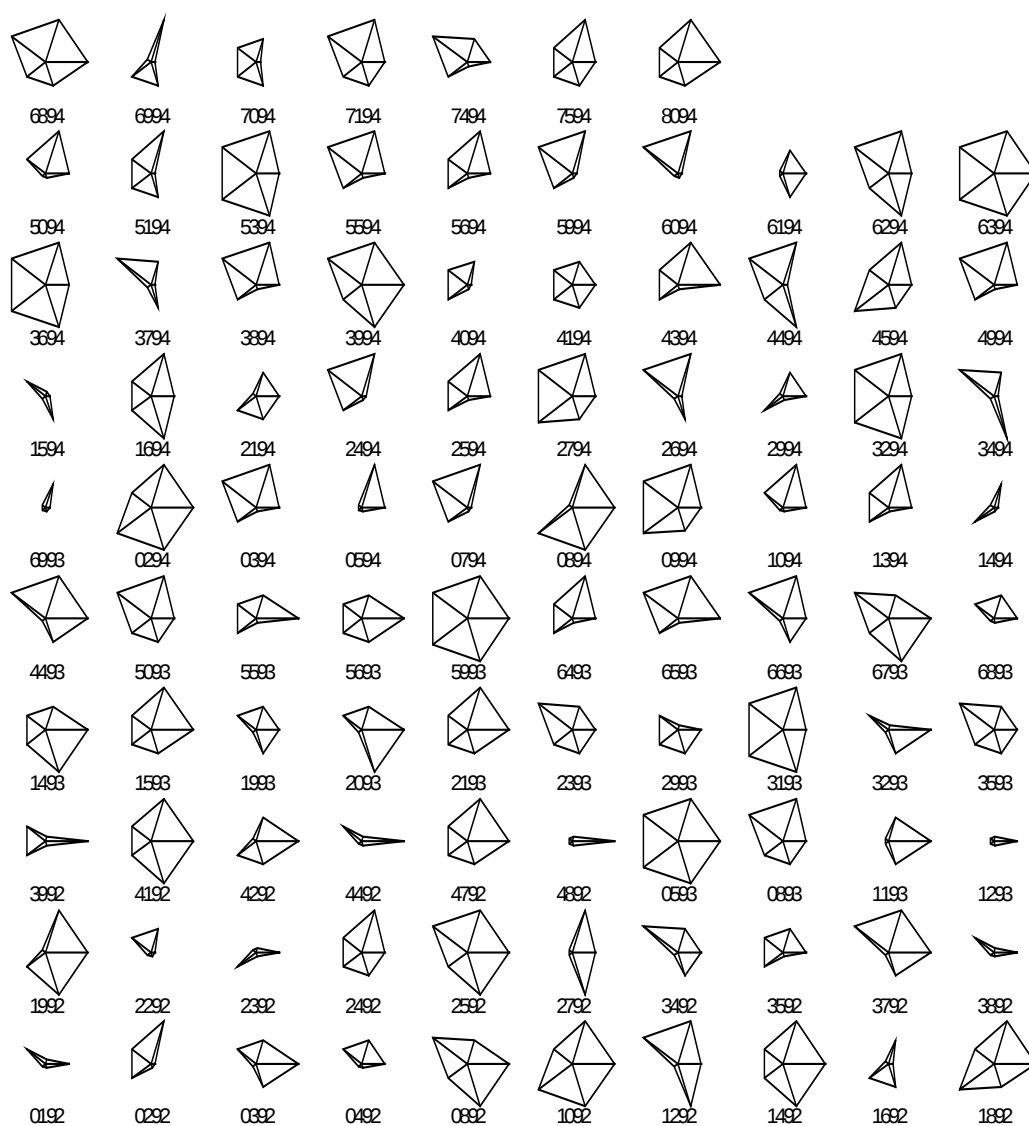
(obr.2.11)

Pro čtyři shluky vyjde:

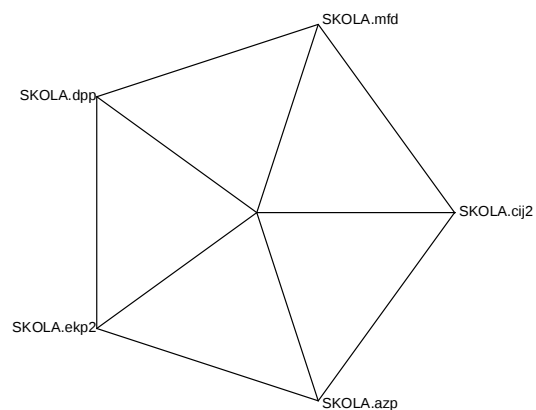
- 91 studentů
- 3992, 5593
- 1193, 6194
- 2993, 1594

Tímto způsobem bychom mohli pokračovat i dále, a vždy by se z velké (hlavní) skupiny vydělili dva nebo tři studenti. Z toho vyplývá, že výsledků studentů jsou si podobné bez velkých rozdílů.

2.4.5 Hvězdicové grafy



(obr2.12)



(obr.2.13)

Hvězdicové grafy (obr.2.12, vzor obr.2.13) nám ukazují, že většina studentů má podobné výsledky, pouze u několika je výrazné zaměření na jeden nebo dva předměty - 0192, 3892, 2392, 1293, 4892, 4492, 3992, 6993, 1494, 1594. U některých naopak jeden předmět chybí - 1292, 0392, 1992, 3492, 4292, 1993, 2093, 4493, 5593, 6493, 6593, 6693, 0394, 0894, 2594, 3694, 5694, 7094, 7494.

2.5 Závěr

Řešení tohoto příkladu metodou hlavních komponent a pomocí faktorové analýzy nám pomohlo zredukovat nadbytečný počet předmětů na pět - dopravní prostředky (DPP), management fyzické distribuce (MFD), ekonomika podniku 2(EKP2), administrativní základy přepravy (AZP) a cizí jazyk 2 (CIJ2).

Faktorově čistý je cizí jazyk - ten vyžaduje především verbální dovednosti a schopnosti - a management fyzické distribuce a dopravní prostředky - u těchto předmětů převažuje potřeba technických a matematických dovedností a schopností. Ekonomika podniku a administrativní základy přepravy jsou faktorově smíšené, to znamená, že k jejich úspěšnému zvládnutí je potřeba obojích dovedností a schopností. Tady je potřeba poznamenat, že tento typ studia je zaměřen na aplikované vědy a jejich poznatky, proto, jak se dalo předpokládat, většina předmětů je faktorově smíšených. Pouze u některých, jako např. CIJ1, CIJ2, KOM, vychází faktorová čistota.

To samé platí i u studentů (absolventů). Většina z nich má potřebné technické i verbální schopnosti (většinou ve srovnatelné míře), proto i hvězdicové grafy vycházejí podobné a shlukování neposkytuje shluky se stejnou nebo alespoň přibližně stejnou velikostí.

Tak např. verbální schopnosti převažují u studentů s čísly - 1293, 4892, 4492, 3992. Naopak technické a matematické schopnosti se dají předpokládat u studentů s čísly 6993, 1494, 6994, 0292, 1692, 5194.

Název souboru:	VÍCERO~1
Adresář:	E:\Pom\vicerozm
Šablona:	D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název:	Zadání:
Předmět:	
Autor:	Jindřich Borůvka
Klíčová slova:	
Komentáře:	
Datum vytvoření:	30.09.98 16:32
Číslo revize:	3
Poslední uložení:	30.09.98 17:06
Uložil:	Jindřich Borůvka
Celková doba úprav:	2 min.
Poslední tisk:	15.09.00 09:56
Jako poslední úplný tisk	
Počet stránek:	37
Počet slov:	5 877 (přibližně)
Počet znaků:	33 502 (přibližně)