

SEMESTRÁLNÍ PRÁCE

Licenční studium

STATISTICKÉ ZPRACOVÁNÍ DAT PŘI KONTROLE A ŘÍZENÍ JAKOSTI

Předmět

POČÍTAČOVÁ ANALÝZA VÍCEROZMĚRNÝCH DAT

Ústav experimentální biofarmacie, Hradec Králové

Ing. Martina Růžicková

Příklad 1.

Zadání: Od 35 pacientů byly získány hodnoty některých biochemických a hematologických laboratorních parametrů. Na příkladu předved'te výpočet kovarianční a korelační matice, metodu hlavních komponent, faktorovou analýzu, shlukovou analýzu a grafické metody porovnání proměnných.

Data: Biochemické a hematologické parametry změřené u 35 pacientů:

Číslo pacienta	ALT [μkat/l]	AST [μkat/l]	ALP [μkat/l]	Neutrofily [%]	Eosinofily [%]	Bazofily [%]	Hemoglobin [g/l]
1	0,26	0,37	1,68	59,8	9,2	0,7	131
2	0,52	0,57	3,18	49,0	1,4	1,1	118
3	1,08	0,68	2,41	70,5	1,9	0,7	167
4	0,39	0,39	1,29	66,0	3,5	1,7	134
5	0,21	0,60	1,43	52,3	3,1	1,2	144
6	0,41	0,51	1,31	57,4	0,7	0,6	137
7	0,46	0,37	2,01	69,0	3,0	0,0	159
8	0,54	0,52	1,83	60,3	4,1	1,1	159
9	1,07	0,76	1,43	47,4	0,53	0,3	148
10	0,57	0,43	1,30	61,0	2,4	0,8	157
11	0,53	0,87	1,54	49,9	1,8	0,9	147
12	0,42	0,50	1,17	64,8	0,5	0,7	166
13	0,38	0,43	1,37	46,7	5,6	0,6	162
14	0,30	0,25	1,60	63,8	1,0	0,8	142
15	0,19	0,37	1,56	53,3	1,1	0,4	151
16	0,25	0,44	1,55	47,0	0,6	0,2	174
17	0,72	0,44	1,85	63,3	1,5	0,1	151
18	1,34	0,93	0,86	48,6	0,8	0,0	135
19	0,38	0,45	1,08	10,0	1,5	0,1	142
20	0,64	1,12	1,37	52,4	5,1	0,5	157
21	0,73	0,48	2,34	57,0	2,0	0,0	168
22	0,29	0,26	1,96	38,5	2,1	0,8	149
23	0,66	0,46	1,94	56,0	1,0	0,0	176
24	0,34	0,24	1,07	62,0	1,0	1,0	164
25	0,45	0,24	0,93	55,0	3,0	1,0	121
26	0,44	0,41	1,19	52,5	1,0	0,0	113
27	0,41	0,27	0,76	72,0	6,0	1,0	153
28	0,32	0,39	2,39	59,0	1,0	3,0	163
29	0,34	0,49	1,39	51,0	1,0	0,0	147
30	0,55	0,47	1,89	63,0	3,0	0,0	191
31	0,45	0,27	1,75	40,1	7,8	0,7	169
32	0,39	0,46	2,19	53,6	0,0	0,3	143
33	0,34	0,64	1,90	57,7	1,8	1,2	146
34	0,21	0,29	1,67	61,7	1,5	1,0	150
35	0,86	0,47	1,27	42,3	4,6	1,0	154

Program: Statgraphics, Multivariate methods

Řešení a obrázky:

1. Kovarianční analýza

Pomocí programu Statgraphics se spočítala kovarianční matice, která obsahuje kovariance pro jednotlivé objekty. V hodnotě kovariance se sice odráží variabilita jednotlivých proměnných, ale nelze bezprostředně usuzovat na míru souvislosti. Používá se především jako vstup pro další výpočty.

Tabulka 1 – Kovarianční matice

	ALT	AST	ALP	Neutro	Eosino	Bazo	Hemoglo
ALT	0,06814						
AST	0,02990	0,03947					
ALP	0,00088	0,00183	0,25305				
Neutro	40,9327	-7,85156	44,3843	$1,12 \cdot 10^6$			
Eosino	-0,05798	-0,04881	-0,13569	-176,59	-0,05798		
Bazo	-0,0462	-0,01920	0,04259	-104,053	0,17971	0,36681	
Hemoglo	0,34309	-0,06687	1,28208	15,0056	1,94176	-1,04454	276,904

2. Korelační analýza

Nevýhodou kovariance je fakt, že její hodnoty závisí na měřítku, ve kterém jsou vyjádřeny základní hodnoty. Proto se provádí standardizace, kdy z kovariance získáme párový korelační koeficient. V tabulce 2 jsou zaznamenány hodnoty jednotlivých párových korelačních koeficientů.

Korelační koeficienty jsou mírou lineární závislosti mezi dvěma náhodnými veličinami. Pokud se korelační koeficient blíží 1, mezi danými proměnnými je silný korelační vztah. Záporné znaménko u korelačního koeficientů znamená sestupnou tendenci korelačního vztahu. Je-li velikost korelačního koeficientu menší jak 0,5, nelze mezi proměnnými hovořit o závislosti.

V prvním řádku každé proměnné v tabulce 2 jsou uvedeny korelační koeficienty a v druhém řádku je uvedena hladina významnosti, tj. pravděpodobnost, že při skutečné nulové korelaci, by empiricky zjištěný korelační koeficient mohl dosáhnout nynější hodnoty. Jeli hladina významnosti menší jak 0,05, lze usoudit, že mezi příslušnými proměnnými existuje nějaká souvislost.

Z korelační matice vyplývá, že silný korelační vztah je jedině u dvojice AST – ALT.

Tabulka 2 – Korelační matice

	ALT	AST	ALP	Neutro	Eosino	Bazo	Hemoglo
ALT	1,0 0,0						
AST	0,5765 0,0003	1,0 0,0					
ALP	0,0067 0,9695	0,0183 0,9169	1,0 0,0				
Neutro	0,1477 0,3972	-0,0372 0,8319	0,0831 0,6351	1,0 0,0			
Eosino	-0,1037 0,5534	-0,1147 0,5119	-0,1259 0,47711	-0,0776 0,6576	1,0 0,0		
Bazo	-0,2922 0,0890	-0,1596 0,3598	0,1398 0,4231	-0,1618 0,3531	0,1385 0,4275	1,0 0,0	
Hemoglo	0,0790 0,6520	-0,0202 0,9082	0,1532 0,3797	0,0008 0,9961	0,0545 0,7560	-0,1036 0,5535	1,0 0,0

3. Analýza hlavních komponent

Analýza hlavních komponent na rozdíl od korelační matice, která určila závislost mezi proměnnými, je schopna určit důvod, proč spolu některé z proměnných souvisí. Pomáhá nalézt společné latentní proměnné, které se v metodě PCA nazývají hlavní komponenty. Hlavní komponenty ve velké míře popisují a vystihují variabilitu v původních datech.

Souhrnně lze říci, že metoda hlavních komponent snižuje počet původních proměnných tím, že vytvoří lineární kombinaci zdrojových proměnných, které vysvětlují největší část její variability. První hlavní komponenta je taková kombinace zdrojových proměnných, která má největší rozptyl mezi všemi lineárními kombinacemi. Podobně následuje druhá hlavní komponenta atd. Pro dostatečné vysvětlení chování zdrojových proměnných se požaduje 85 – 90 % vysvětlené variability.

V tabulce 3 jsou uvedeny výsledky analýzy hlavních komponent. Při výpočtu byla použita standardizace, neboť vstupní data nejsou uvedena ve stejném měřítku. V prvním sloupci jsou uvedena pořadová čísla jednotlivých komponent. Ve druhém sloupci jsou uvedena procenta variability vysvětlená jednotlivými hlavními komponentami (tj. podíl rozptylů jednotlivých hlavních komponent na celkové variabilitě, v pořadí od nejvyššího k nejnižšímu). Ve třetím sloupci jsou tyto podíly uvedeny kumulovaně a lze z nich určit optimální počet komponent. Jak už bylo uvedeno výše, většinou se požaduje aby součet prvních nejvyšších komponent měl 85-90 %. Jak vyplývá z tabulky 3, tento předpoklad splňuje 5 hlavních komponent.

Tabulka 3 – Metoda hlavních komponent

Principal Components Analysis		
Component Number	Percent of Variance	Cumulative Percentage
1	25,84292	25,84292
2	17,00896	42,85188
3	15,42611	58,27799
4	15,05645	73,33444
5	12,62512	85,97956
6	8,58963	94,56919
7	5,43081	100,00000

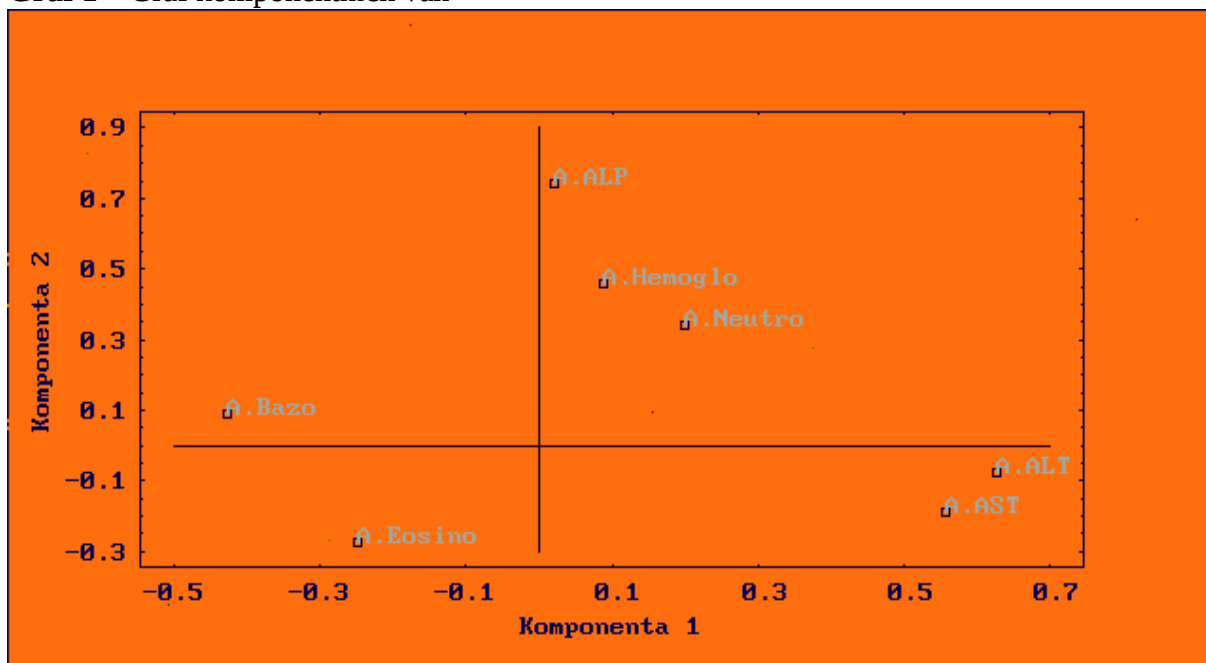
Analýza hlavních komponent nabízí i grafickou analýzu:

1. Plot Component Weigth (graf komponentních vah).

Graf zobrazuje komponentní váhy zdrojových proměnných pro první dvě hlavní komponenty. Největší přínosem pro danou komponentu mají proměnné, které se v grafu nachází co nejbližce u souřadnice dané komponenty a na číselné ose co nejdále od nuly.

Z grafu je patrné, že největší přínos pro první komponentu mají parametry ALT a AST a pro druhou komponentu parametr ALP (viz graf 1).

Graf 1 – Graf komponentních vah



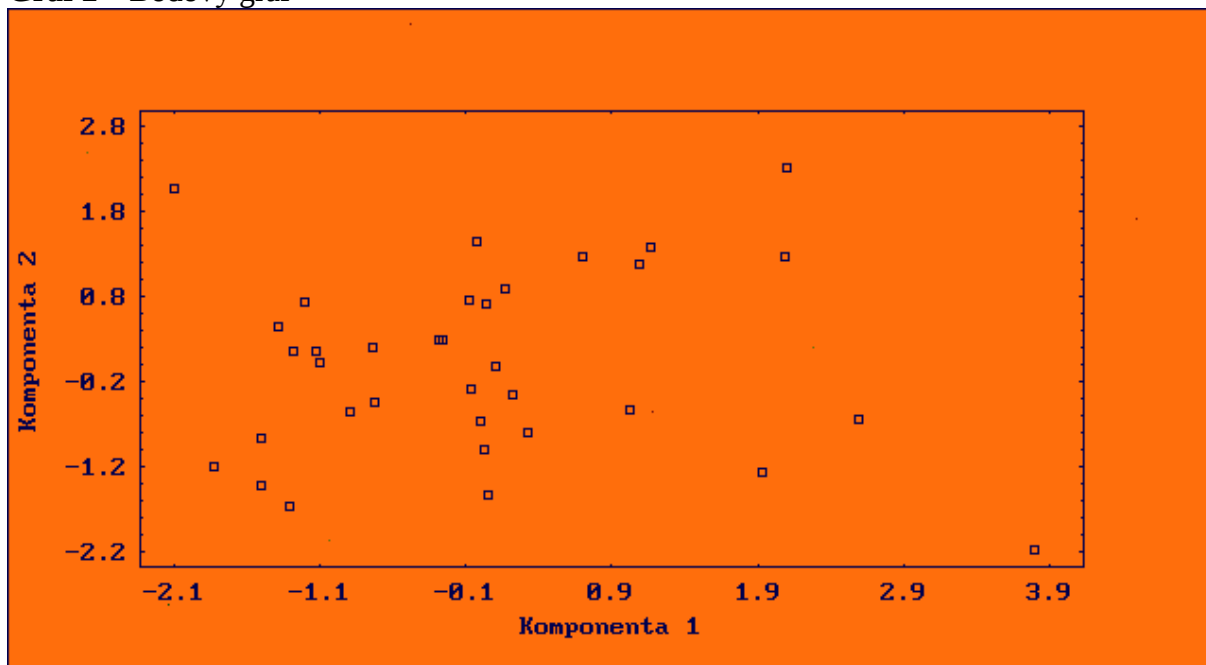
2. Scatterplot (bodový diagram).

Graf zobrazuje komponentní hodnoty prvních dvou hlavních komponent u jednotlivých objektů. Vstupní data pochází z normálního rozdělení, protože body v grafu jsou rozmístěny v jakémsi pomyslném kruhu.

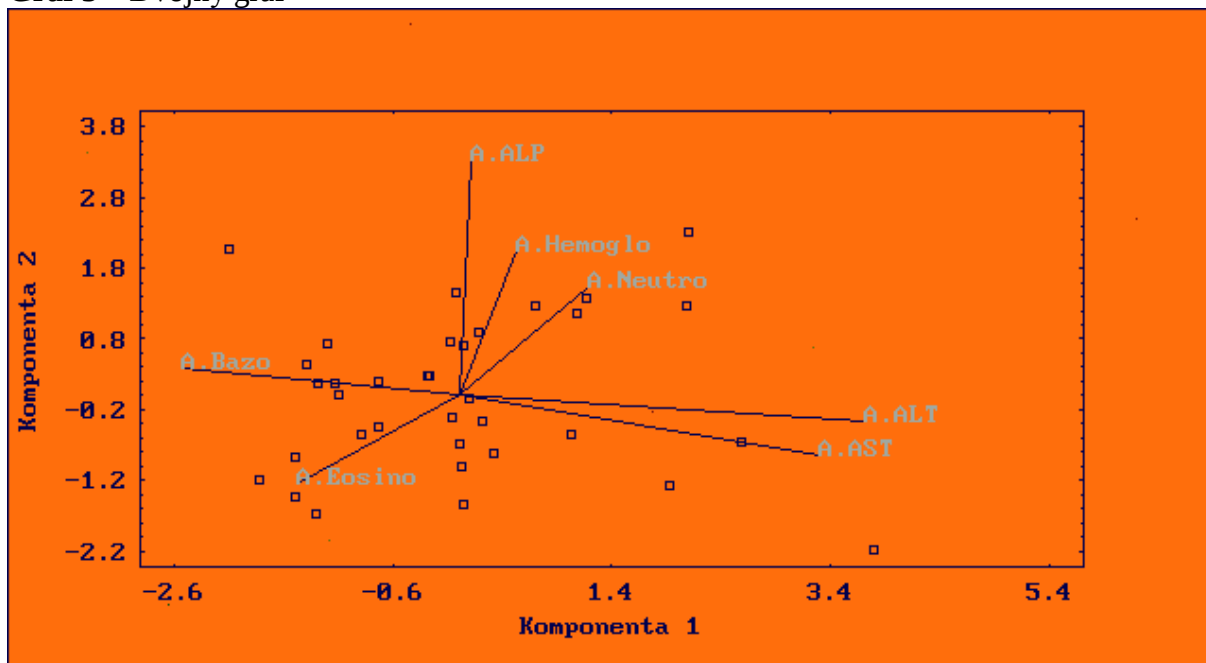
3. Biplot (dvojný graf)

Tento graf kombinuje oba předchozí grafy. Přímkami spojující se v bodě (0,0) představují původní proměnné (jejich „příspěvek“ k prvním dvěma hlavním komponentám) a body představují komponentní hodnoty jednotlivých objektů. Úhel mezi přímkami je nepřímo úměrný jejich vzájemné korelaci. Př. úhel mezi proměnnými AST a ALT je nejmenší, neboť parametry mají nejvyšší párový korelační koeficient.

Graf 2 – Bodový graf



Graf 3 – Dvojný graf



4. Faktorová analýza

Faktorová analýza patří, podobně jako Metoda hlavních komponent, mezi metody, které redukuje počet proměnných. FA předpokládá, že každou vstupující proměnnou lze vyjádřit jako lineární funkci nevelkého počtu společných latentních proměnných (faktorů) a jediného chybového faktoru. Na rozdíl od komponentní analýzy se tedy při faktorové analýze snažíme vysvětlit závislost proměnných.

Vstupní data byla standardizována. Do výpočtu byly zařazeny pouze objekty se známými hodnotami všech vstupních proměnných (*Missing Values:Listwise*). Jako typ rotace byla zvolena metoda *Equimax*, která spojuje kriteria *Varimax* a *Quartimax*, což znamená, že při rotaci je maximalizován rozptyl čtverců faktorových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. Za kritérium konvergence byla dosazena standardní hodnota 1E-5 a byl zvolen maximální počet interakcí při rotování faktoru 100. V tabulce 4 jsou uvedeny získané hodnoty faktorové analýzy. V prvním sloupci jsou jednotlivé objekty (tj. parametry). Ve druhém sloupci jsou odhady komunalit (tj. čtverce indexu korelace). Ve třetím sloupci jsou uvedena čísla faktorů. Ve čtvrtém sloupci jsou příslušné variability. Procenta variability jsou v pátém sloupci a v posledním sloupci jsou uvedena procenta variability kumulativně.

Obecně platí pravidlo, že čím má proměnná větší komunalitu, tím má i větší společné vlastnosti s ostatními. Proměnné mají poměrně nízké komunality, což značí o jejich nepodobnosti. Podobně jako u metody PCA i zde je zapotřebí k popsání alespoň 85 % variability dat 5 faktorů, které kumulativně popisují 86 % variability.

Tabulka 4 – Faktorová analýza

Variable	Communality	Factor	Eigenvalue	Percent Var	Cum Percent
ALT	0,39747	1	1,80900	25,8	25,8
AST	0,35664	2	1,19063	17,0	42,9
ALP	0,08507	3	1,07983	15,4	58,3
Neutrofily	0,07431	4	1,05395	15,1	73,3
Eosinofily	0,05881	5	0,88516	12,6	86,0
Bazofily	0,15149	6	0,60127	8,6	94,6
Hemoglobin	0,05895	7	0,38016	5,4	100,0

Dalším výstupem faktorové analýzy je tabulka matice odhadu faktorových vah (viz tabulka 5). Protože jsme v předešlé analýze nechali hodnoty zdrojových proměnných standardizovat, představují prvky matice korelační koeficienty mezi jednotlivými proměnnými (v řádcích) a společnými faktory (v sloupcích).

Tabulka 5 – Matice faktorových vah

Variable	Factor				
	1	2	3	4	5
ALT	0,84608	-0,08665	0,18720	0,04232	0,20068
AST	0,75919	-0,20772	0,42570	-0,17167	0,13251
ALP	0,02725	0,80827	0,31211	-0,23964	0,15040
Neutrofily	0,27018	0,37047	-0,66356	0,05575	0,54058
Eosinofily	-0,33522	-0,30123	0,24642	0,59109	0,56415
Bazofily	-0,57554	0,09446	0,44376	-0,34345	0,34150
Hemoglobin	0,12048	0,49976	0,26115	0,70343	-0,27863

V tabulce 6 jsou zaznamenány odhady komunalit výše uvedené faktorové matice po rotaci metodou Equimax. Oproti prvním odhadům komunalit u proměnných došlo u všech ke značnému zvýšení (v některých případech se komunality zvýšily až 10x).

Tabulka 6 – Odhady komunalit faktorové matice

Variable	Est Comunality
ALT	0,80046
AST	0,83268
ALP	0,83151
Neutrofily	0,94588
Eosinofily	0,93145
Bazofily	0,99167
Hemoglobin	0,90492

V následující tabulce jsou odhady faktorových vah pro faktorovou matici po rotaci, kterou jsme zadali ve vstupním formuláři (Equimax). Důvodem rotace je optimální natočení souřadnic tak, aby se získalo co nejvíce faktorově čistých proměnných. Při rotaci došlo ke zjednodušení sloupců (viz tabulky 5 a 7). Rotace nám umožnila s větší jistotou určit proměnné, které příslušný faktor nejvíce korelují:

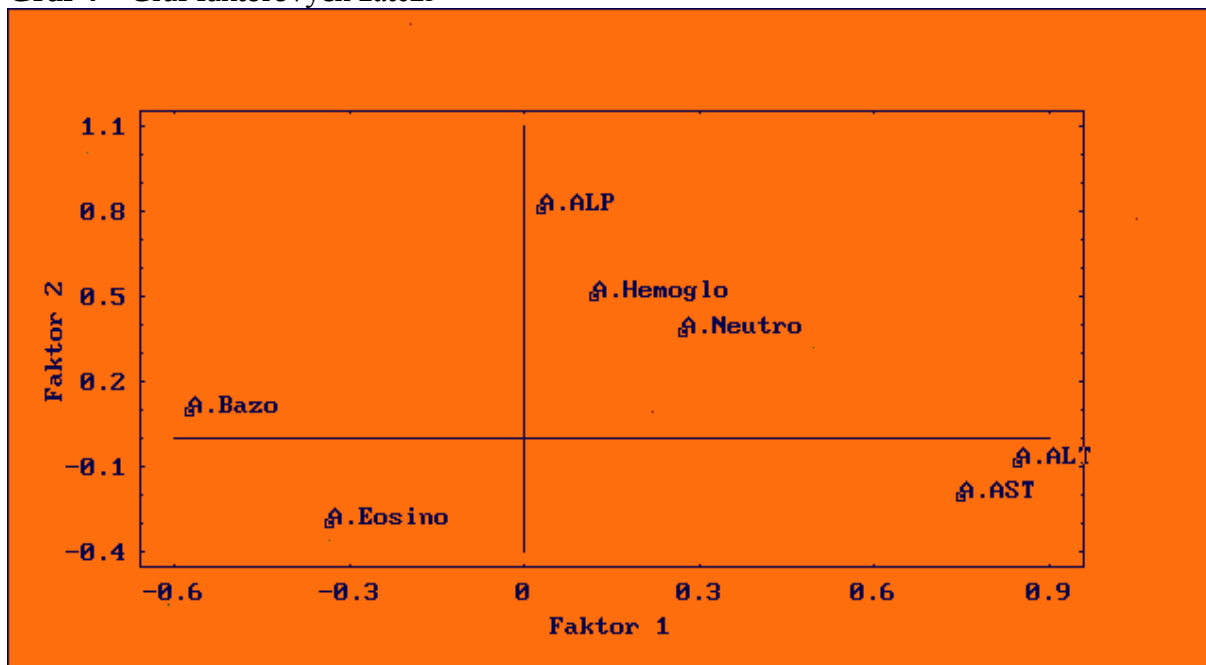
1. faktor je korelován parametry ALT a AST
2. faktor je korelován parametry ALP a Bazofily
3. faktor je korelován parametrem Hemoglobin
4. faktor je korelován parametrem Neutrofily
5. faktor je korelován parametrem Eosinofily

Tabulka 7 – Matice faktorových vah po rotaci

Variable	Factor				
	1	2	3	4	5
ALT	0,86071	-0,10043	0,11047	0,18635	-0,05121
AST	0,90099	0,01431	-0,05859	-0,10952	-0,07262
ALP	0,05636	0,83592	0,24955	0,15379	-0,20892
Neutrofily	0,00898	0,02743	-0,03280	0,97148	-0,01425
Eosinofily	-0,04097	-0,05439	0,06984	-0,01023	0,96012
Bazofily	-0,21582	0,64889	-0,34705	-0,2801	0,32421
Hemoglobin	-0,00523	0,07635	0,94465	-0,03811	0,07249

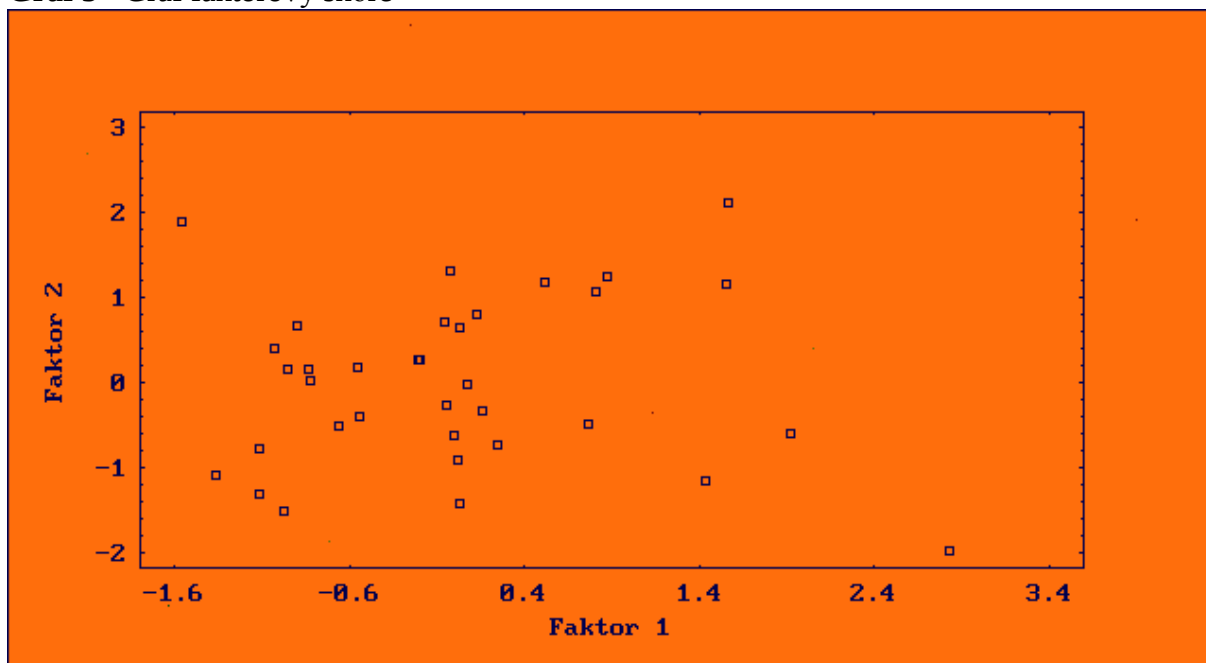
Faktorová analýza stejně jako metoda hlavních komponent poskytuje i grafické hodnocení. Prvním grafem, který FA nabízí je graf **Plot Factor Weights** (graf faktorových zátěží). Graf zobrazí hodnoty faktorových zátěží vstupujících proměnných v grafu, kde souřadnice tvoří první dva společné faktory (viz graf 4).

Graf 4 – Graf faktorových zátěží



Dalším grafem, který faktorová analýza nabízí je Plot Factor Scores (graf faktorových skóre), ve kterém jsou znázorněny hodnoty vypočtených faktorových skóre pro první dva společné faktory u jednotlivých vstupujících pozorování (objektů).

Graf 5 - Graf faktorový skóre



5. Shluková analýza

Shluková analýza patří mezi metody zabývající se zkoumáním podobnosti vícerozměrných objektů a jejich rozříděním do skupin. Statgraphics nabízí pro shlukovou analýzu několik metod tzv. hierarchické analýzy (metoda průměrová, centroidní metoda a metoda nejbližšího a nejvzdálenějšího souseda) a jeden postup nehierarchické analýzy (metoda typických bodů).

Jednou z nevýhod shlukové analýzy je to, že předem musíme určit do kolika shluků chceme proměnné seřadit a druhou nevýhodou je to, že shluková analýza neposkytuje jako výstup grafický dendrogram.

Provedla jsem shlukovou analýzu všemi metodami, které program Statgraphics nabízí. Výstupem každé metody je tabulka a grafické znázornění. Všechny metody poskytují shodné výsledky, proto si zde uvedeme pouze výstup jedné z nich, např. průměrové metody (Average). Při analýze jsem postupovala od tvorby 20 shluků po 2 shluky. Optimální počet shluků mi připadal 10. Zastoupení bodů v jednotlivých sluchách jsou uvedena v tabulce 8, grafické znázornění je v grafu 6.

Tabulka 8 – Shluková analýza

Shluk	Pořadová čísla pacientů patřících do příslušných shluků	Absolutní četnost	Procentuální četnost
1	1	1	2,8571
2	2	1	2,8571
3	3,6,7,10,12,14,15,16,19,21,22,23,24,26,29,30,32,34	18	51,4286
4	4,25	2	5,7143
5	5,33	2	5,7143
6	8,13,27,31,35	5	14,2857
7	9,18	2	5,7143
8	11,20	2	5,7143
9	17	1	2,5871
10	28	1	2,5871

6. Grafické metody zkoumání podobnosti objektů

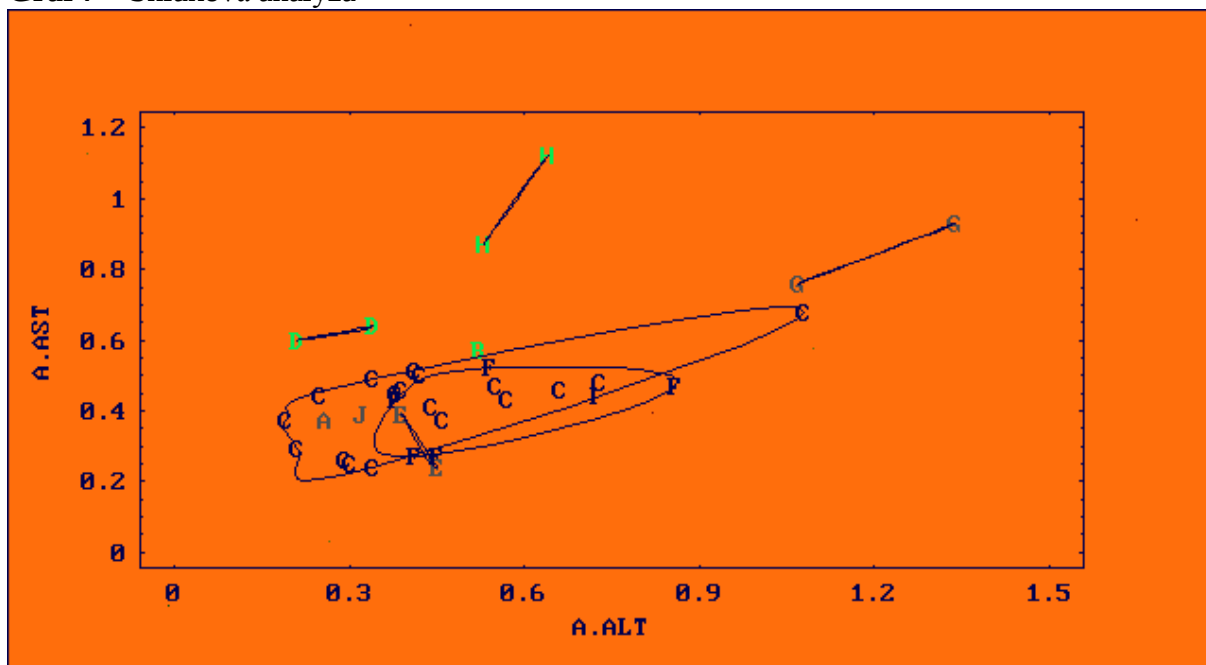
Tyto metody slouží k vizuálnímu porovnávání vícerozměrných objektů. Program Statgraphics používá dvou metod *Sun Ray Plot* a *Star Symbol Plot*

1. Sun Ray Plot

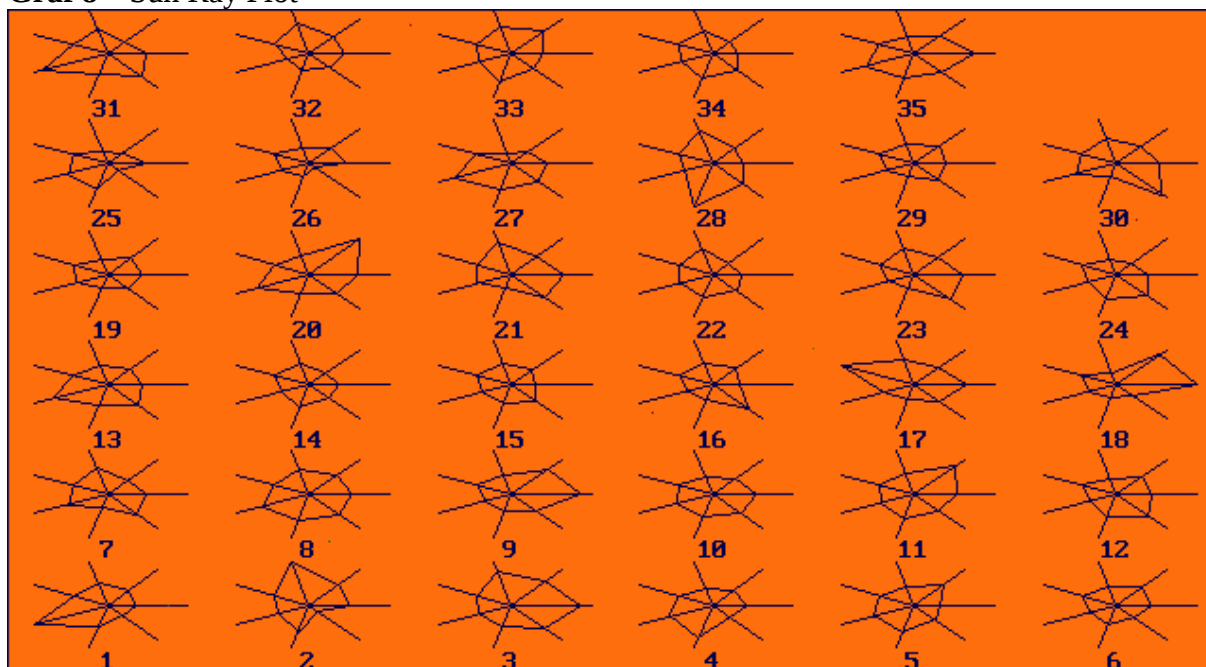
Tato procedura slouží k vizuálnímu porovnávání různých vícerozměrných objektů z hlediska jejich porovnání s průměrným objektem. Program nakreslí pro každý objekt hvězdicu, která je složena s paprsků spojených do jednoho bodu a úseček mezi body, které tvoří polygon. Počet paprsků odpovídá počtu proměnných (v našem případě 7), střed paprsku představuje průměr odpovídající proměnné a délka paprsku 2n násobek směrodatné odchylky.

Graf Sun Ray Plot nám poskytl podobné hodnocení jako shluková analýza. Za zcela odlišné objekty můžeme podle vizuálního porovnání označit pacienty s pořadovým číslem 1, 2, 7, 28. Podobné jsou objekty (pacienti) s pořadovými čísly 5-33, 11-20, 27-31-35.

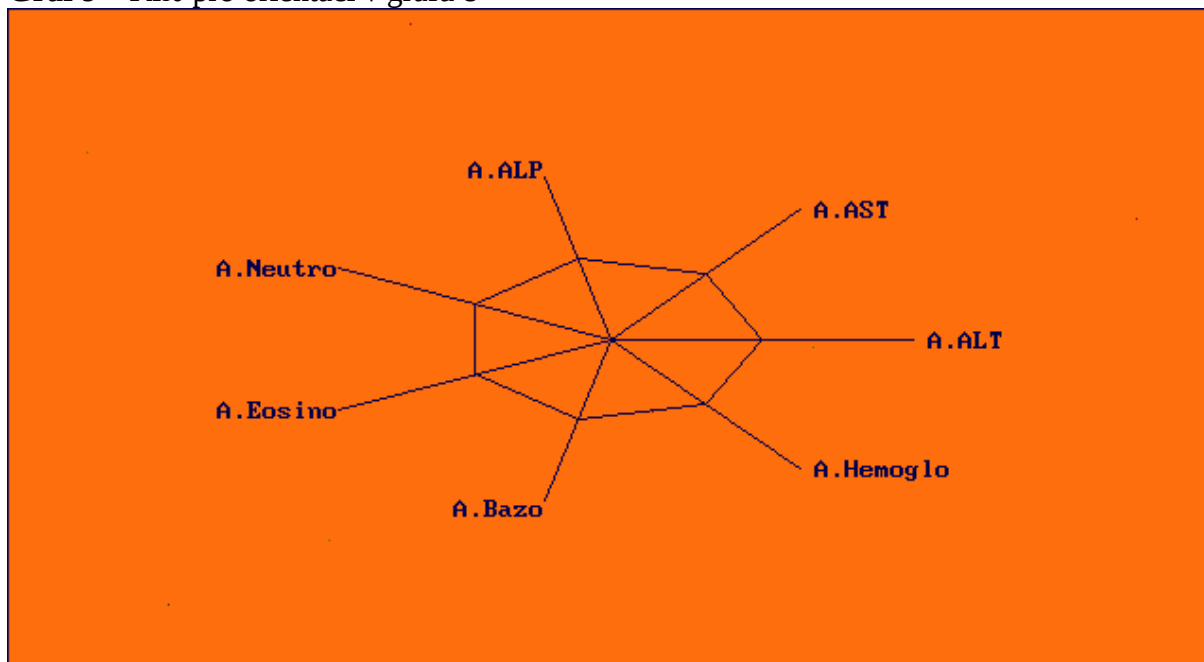
Graf 7 – Shluková analýza



Graf 8 – Sun Ray Plot



Graf 9 – Klíč pro orientaci v grafu 8

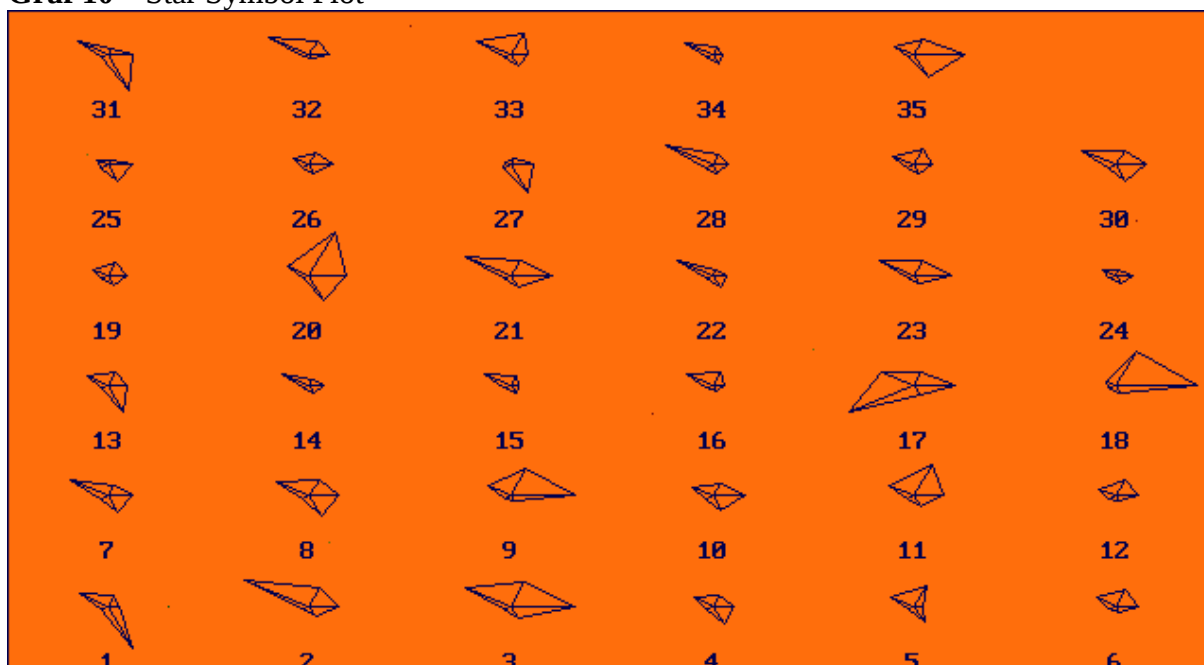


2. Star Symbol Plot

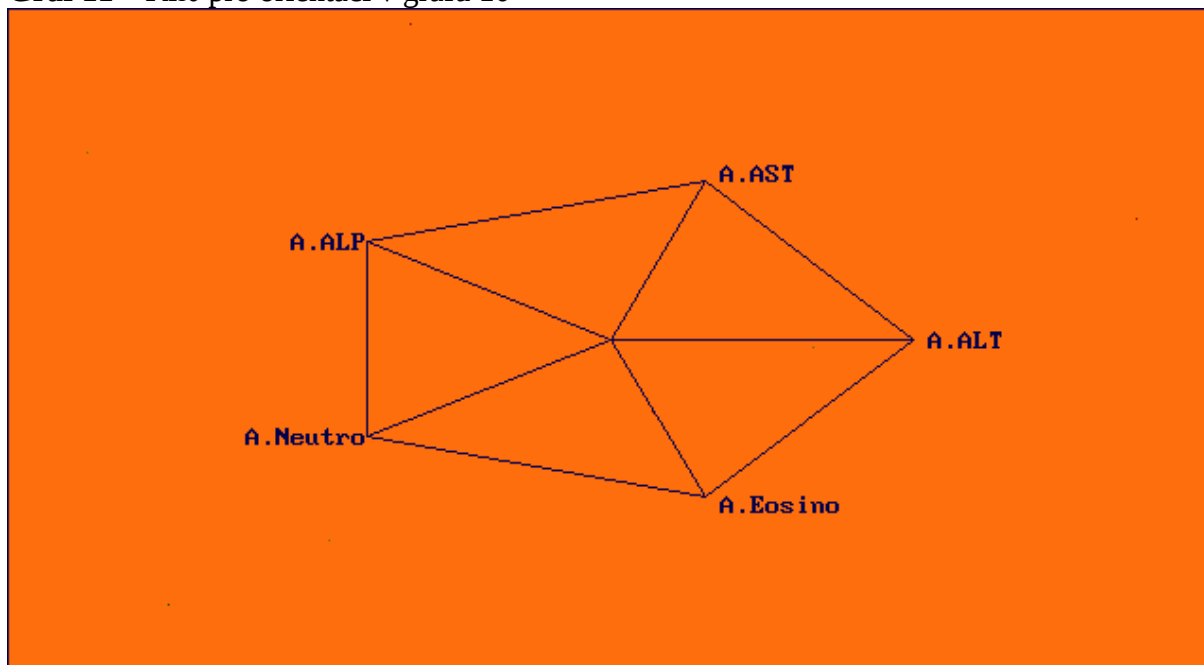
Podobně jako předchozí graf slouží k vizuálnímu zobrazení podobnosti vícerozměrných objektů. Graf je opět složen z paprsků, které se spojují v jednom centrálním bodě.

Obrazce tvořené ze 7 komponent nejsou přehledné. Proto k tvorbě grafu použijeme pouze 5 hlavních komponent, které vysvětlují požadovanou variabilitu souboru (86%). Grafické znázornění podává shodné závěry jako předešlý graf.

Graf 10 – Star Symbol Plot



Graf 11 – Klíč pro orientaci v grafu 10



7. Závěr

Soubor získaných biochemických a hematologických parametrů od 35 pacientů byl zpracován statistickým softwarem Statgraphics s využitím metod:

- kovarianční analýza,
- korelační analýza,
- metoda hlavních komponent,
- faktorová analýza,
- shluková analýza,
- grafické metody.

Výpočet **kovarianční matice** byl proveden jako vstup pro další výpočty. Pro porovnání podobnosti objektů má větší váhu **korelační matice**, která vznikne standardizací kovarianční matice (standardizací se zamezí vlivu rozdílného měřítka vstupních dat). Korelační koeficienty jsou mírou lineární závislosti mezi dvěma náhodnými veličinami. Pokud se korelační koeficient blíží 1, mezi danými proměnnými je silný korelační vztah. Záporné znaménko u korelačního koeficientů znamená sestupnou tendenci korelačního vztahu. Je-li velikost korelačního koeficientu menší jak 0,5, nelze mezi proměnnými hovořit o závislosti. Z korelační matice v našem případě vyplývá, že silný korelační vztah je jedině u dvojce AST – ALT ($r=0,5765$), v ostatních případech jsou párové korelační koeficienty velmi malé (nižší jak 0,3).

Další metodou, které byly vstupní data podrobeny, je **Metoda hlavních komponent**. Tato metoda určuje společné latentní proměnné (hlavní komponenty), které vystihují 85-90 % vysvětlené variability v zdrojových datech. Metoda určila 5 hlavních komponent, kterými bylo vysvětleno 86 % variability zdrojových dat.

Maketa Analýza hlavních komponent v softwaru Statgraphics nabízí i grafickou analýzu, kterou jsou 3 grafy:

Graf komponentních vah zobrazuje komponentní váhy zdrojových proměnných pro první dvě hlavní komponenty. Největší přínosem pro danou komponentu mají proměnné, které se v grafu nachází co nejbližší u souřadnice dané komponenty a na číselné ose co nejdále od nuly. Z grafu je patrné, že největší přínos pro první komponentu mají parametry ALT a AST a pro druhou komponentu parametr ALP.

Bodový graf zobrazuje komponentní hodnoty prvních dvou hlavních komponent u jednotlivých objektů. Z grafu je patrné, že vstupní data pochází z normálního rozdělení, protože body v grafu jsou rozmístěny v jakémsi pomyslném kruhu.

Dvojný graf kombinuje oba předchozí grafy. Přímkami spojující se v bodě (0,0) představují původní proměnné (jejich „příspěvek“ k prvním dvěma hlavním komponentám) a body představují komponentní hodnoty jednotlivých objektů. Úhel mezi přímkami je nepřímo úměrný jejich vzájemné korelaci. Př. úhel mezi proměnnými AST a ALT je nejmenší, neboť parametry mají nejvyšší párový korelační koeficient.

Další metoda, která byla použita a která rovněž určuje společné latentní proměnné je **Faktorová analýza**. Tato metoda předpokládá, že každou vstupující proměnnou lze vyjádřit jako lineární funkci nevelkého počtu společných latentních proměnných a jediného chybového faktoru. Podobně jako metodou PCA bylo nalezeno 5 společných latentních proměnných, které vysvětlují 86 % variability vstupních dat. V této metodě jsou latentní proměnné označovány jako faktory.

Při výpočtu byla použita rotace metodou Equimax, což znamenalo, že při rotaci byl maximalizován rozptyl čtverců faktorových vah a zároveň byl maximalizován součet čtvrtých mocnin faktorových zátěží. Jedním z následků rotace bylo to, že oproti původním hodnotám komunalit před rotací, došlo po rotaci k značnému zvýšení hodnot komunalit (v některém případě i 10x). Obecně platí pravidlo, že čím je hodnota komunality vyšší, tím je objekt podobnější s ostatními objekty. Druhým následkem rotace bylo zjednodušení tabulky Matice faktorových vah. Optimálním natočením souřadnic se získalo co nejvíce faktorově čistých proměnných. S větší jistotou můžeme určit proměnné, které příslušný faktor nejvíce korelují: 1. faktor (ALT, AST), 2. faktor (ALP, Bazofily), 3. faktor (Hemoglobin), 4. faktor (Neutrofily), 5. faktor (Eosinofily).

Faktorová analýza stejně jako metoda hlavních komponent poskytuje i grafické hodnocení. Prvním grafem, který FA nabízí je *graf faktorových zátěží*, který zobrazuje hodnoty faktorových zátěží vstupujících proměnných v grafu, kde souřadnice tvoří první dva společné faktory. Z grafu je patrné, že největší přínos pro první společný faktor mají parametry ALT a AST a pro druhý společný faktor parametry ALP a Bazofily. Dalším grafem, který faktorová analýza nabízí je *graf faktorových skóre*, ve kterém jsou znázorněny hodnoty vypočtených faktorových skóre pro první dva společné faktory u jednotlivých vstupujících pozorování (objektů).

Shluková analýza patří mezi metody zabývající se zkoumáním podobnosti vícerozměrných objektů a jejich roztríděním do skupin. Její nevýhodou je to, že předem musíme určit do kolika shluků chceme proměnné seřadit a že neposkytuje jako grafický výstup dendrogram. Za optimální bylo určeno seskupení vstupních dat do 10 shluků.

Posledními použitými metodami byly **grafické metody**, které jsou založeny na vizuálním porovnání objektů. Byly použity metody Sun Ray Plot a Start Symbol Plot. Obě grafické metody poskytly podobné hodnocení jako shluková analýza. Za zcela odlišné objekty můžeme podle vizuálního porovnání označit pacienty s pořadovým číslem 1, 2, 7, 28. Podobné jsou objekty (pacienti) s pořadovými čísly 5-33, 11-20, 27-31-35 a další.

Příklad 2.

Zadání: Ve firmě Life a.s. Hradec Králové proběhla v roce 1999 klinická studie ursoxyly. Pacientům byl podán přípravek a v časových intervalech po podání byla měřena koncentrace přípravku v krevní plazmě pacientů. Na příkladu předved'te výpočet kovarianční a korelační matice, metodu hlavních komponent, faktorovou analýzu, shlukovou analýzu a grafické metody porovnání proměnných.

Data: Koncentrace přípravku v plazmě pacientů [ng/ml]

Iniciály pacienta	Čas [hod]									
	0,5	1	1,5	2	2,5	3	4	5	6	8
YS	308	670	575	552	445	377	225	173	140	69
JS	691	717	688	539	457	372	296	168	157	90
JK	486	976	788	718	568	496	336	231	156	83
JD	576	586	608	540	373	434	309	311	213	82
RV	120	187	288	309	443	586	449	292	357	190
MV	321	500	887	885	744	617	430	295	180	97
IN	658	743	752	705	620	551	404	321	220	138
MR	548	1069	872	715	571	452	336	229	181	134
PS	210	571	879	932	848	684	413	322	211	152
MD	378	581	606	600	478	360	247	189	147	99

Program: Statgraphics, Multivariate methods

Řešení a obrázky:

1. Kovarianční analýza

Pomocí programu Statgraphics se spočítala kovarianční matice, která obsahuje kovariance pro jednotlivé objekty. V hodnotě kovariance se sice odráží variabilita jednotlivých proměnných, ale nelze bezprostředně usuzovat na míru souvislosti. Používá se především jako vstup pro další výpočty.

Tabulka 1 – Kovarianční matice

	YS	JS	JK	JD	RV	MV	IN	MR	PS	MD
YS	41694,0									
JS	42040,4	56466,1								
JK	58464,1	62134,8	83415,3							
JD	31245,4	40069,3	45649,0	31390,4						
RV	-799,5	-9724,6	-4069,3	-4144,9	20041,0					
MV	47147,7	42423,8	62006,6	34781,3	13074,2	79422,3				
IN	42870,6	51033,8	61295,7	38266,7	-1868,8	51423,1	50665,1			
MR	61175,8	67309,5	88120,5	48105,6	-9066,3	60771,0	63845,4	94769,8		
PS	49507,9	38507,4	63821,2	31788,2	16807,4	83069,3	50141,0	61768,1	90368,4	
MD	38575,3	40981,6	53794,8	30435,0	-1434,4	47761,6	41836,2	56171,1	48801,8	37331,4

2. Korelační analýza

Nevýhodou kovariance je fakt, že její hodnoty závisí na měřítku, ve kterém jsou vyjádřeny základní hodnoty. Proto se provádí standardizace, kdy z kovariance získáme párový korelační koeficient. V tabulce 2 jsou zaznamenány hodnoty jednotlivých párových korelačních koeficientů.

Korelační koeficienty jsou mírou lineární závislosti mezi dvěma náhodnými veličinami. Pokud se korelační koeficient blíží 1, mezi danými proměnnými je silný korelační vztah. Záporné znaménko u korelačního koeficientů znamená sestupnou tendenci korelačního vztahu. Je-li velikost korelačního koeficientu menší jak 0,5, nelze mezi proměnnými hovořit o závislosti.

V prvním řádku každé proměnné v tabulce 2 jsou uvedeny korelační koeficienty a v druhém řádku je uvedena hladina významnosti, tj. pravděpodobnost, že při skutečné nulové korelaci, by empiricky zjištěný korelační koeficient mohl dosáhnout nynější hodnoty. Jeli hladina významnosti menší jak 0,05, lze usoudit, že mezi příslušnými proměnnými existuje nějaká souvislost.

Z korelační matice vyplývá, že silný korelační vztah je např. mezi pacienty YS-MR, YS-MD, JK-MR a slabý korelační vztah je např. mezi pacienty RV-MD, RV-YS (viz tabulka 2).

Tabulka 2 – Korelační matice

	YS	JS	JK	JD	RV	MV	IN	MR	PS	MD
YS	1,0 0,0									
JS	0,8664 0,0012	1,0 0,0								
JK	0,9914 0,0000	0,9054 0,003	1,0 0,0							
JD	0,8637 0,0013	0,9517 0,0000	0,8921 0,0005	1,0 0,0						
RV	-0,0277 0,9395	-0,2891 0,4179	-0,0995 0,7844	-0,1653 0,6482	1,0 0,0					
MV	0,8193 0,0037	0,6335 0,0492	0,7618 0,0104	0,6966 0,0252	0,3277 0,3553	1,0 0,0				
IN	0,9328 0,0001	0,9541 0,0000	0,9429 0,0001	0,9595 0,0000	-0,0586 0,8721	0,8106 0,0044	1,0 0,0			
MR	0,9732 0,0000	0,9201 0,0002	0,9911 0,0000	0,8820 0,0007	-0,2080 0,5641	0,7005 0,0241	0,9214 0,002	1,0 0,0		
PS	0,8065 0,0048	0,5391 0,1078	0,7351 0,0154	0,5968 0,0685	0,3949 0,2587	0,9805 0,0000	0,7410 0,0142	0,6675 0,0350	1,0 0,0	
MD	0,9778 0,0000	0,8926 0,0005	0,9640 0,0000	0,8891 0,0006	-0,0524 0,8856	0,8771 0,0009	0,9620 0,0000	0,9444 0,0000	0,8402 0,0023	1,0 0,0

3. Analýza hlavních komponent

Metoda hlavních komponent na rozdíl od korelační matice, která určila závislost mezi proměnnými, je schopna určit důvod, proč spolu některé z proměnných souvisí. Pomáhá nalézt společné latentní proměnné, které se v metodě PCA nazývají hlavní komponenty. Hlavní komponenty ve velké míře popisují a vystihují variabilitu v původních datech.

Souhrnně lze říci, že metoda hlavních komponent snižuje počet původních proměnných tím, že vytvoří lineární kombinaci zdrojových proměnných, které vysvětlují největší část její variability. První hlavní komponenta je takový kombinace zdrojových proměnných, která má největší rozptyl mezi všemi lineárními kombinacemi. Podobně následuje druhá hlavní

komponenta atd. Pro dostatečné vysvětlení chování zdrojových proměnných se požaduje 85 – 90 % vysvětlené variability.

V tabulce 3 jsou uvedeny výsledky analýzy hlavních komponent. Při výpočtu nebyla použita standardizace, neboť vstupní data jsou uvedena ve stejném měřítku (tj. v ng léčiva / ml plazmy). V prvním sloupci jsou uvedena pořadová čísla jednotlivých komponent. Ve druhém sloupci jsou uvedena procenta variability vysvětlená jednotlivými hlavními komponentami (tj. podíl rozptylů jednotlivých hlavních komponent na celkové variabilita, v pořadí od nejvyššího k nejnižšímu).

Ve třetím sloupci jsou tytéž podíly uvedeny kumulovaně a lze z nich určit optimální počet komponent. Jak už bylo uvedeno výše, většinou se požaduje aby součet prvních nejvyšších komponent měl 85-90 %. Jak vyplývá z tabulky 3, tento předpoklad splňují 2 hlavní komponenty.

Tabulka 3 – Metoda hlavních komponent

Principal Components Analysis		
Component Number	Percent of Variance	Cumulative Percentage
1	82,98865	82,98865
2	12,86992	95,85856
3	2,20019	98,05875
4	1,48432	99,54308
5	0,27396	99,81703
6	0,12624	99,94327
7	0,03570	99,97897
8	0,01863	99,99761
9	0,00234	99,99995
10	0,00005	100,00000

Analýza hlavních komponent nabízí i grafickou analýzu:

1. *Plot Component Weigth* (graf komponentních vah).

Graf zobrazuje komponentní váhy zdrojových proměnných pro první dvě hlavní komponenty. Největší přínosem pro danou komponentu mají proměnné, které se v grafu nachází co nejbližší u souřadnice dané komponenty a na číselné ose co nejdále od nuly.

Z grafu je patrné, že největší přínos pro první komponentu mají pacienti JK a MR a pro druhou komponentu pacient RV (viz graf 1).

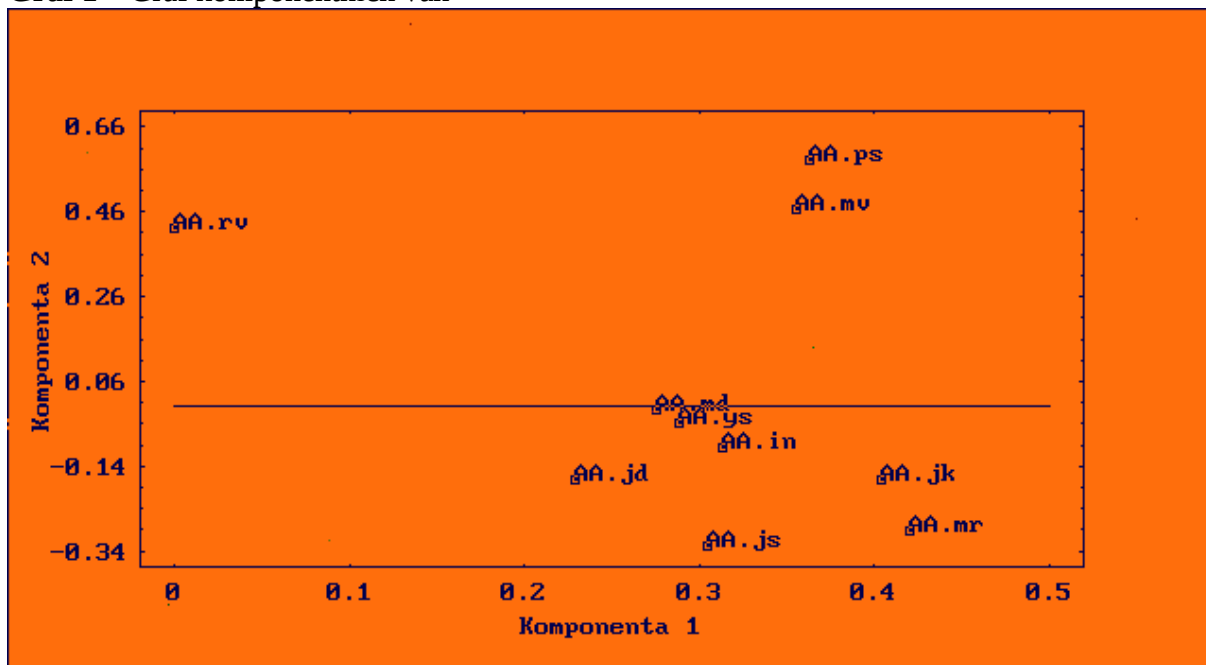
2. *Scatterplot* (bodový diagram).

Graf zobrazuje komponentní hodnoty prvních dvou hlavních komponent u jednotlivých objektů. Vstupní data pochází z normálního rozdělení, protože body v grafu jsou rozmístěny v jakémsi pomyslném kruhu.

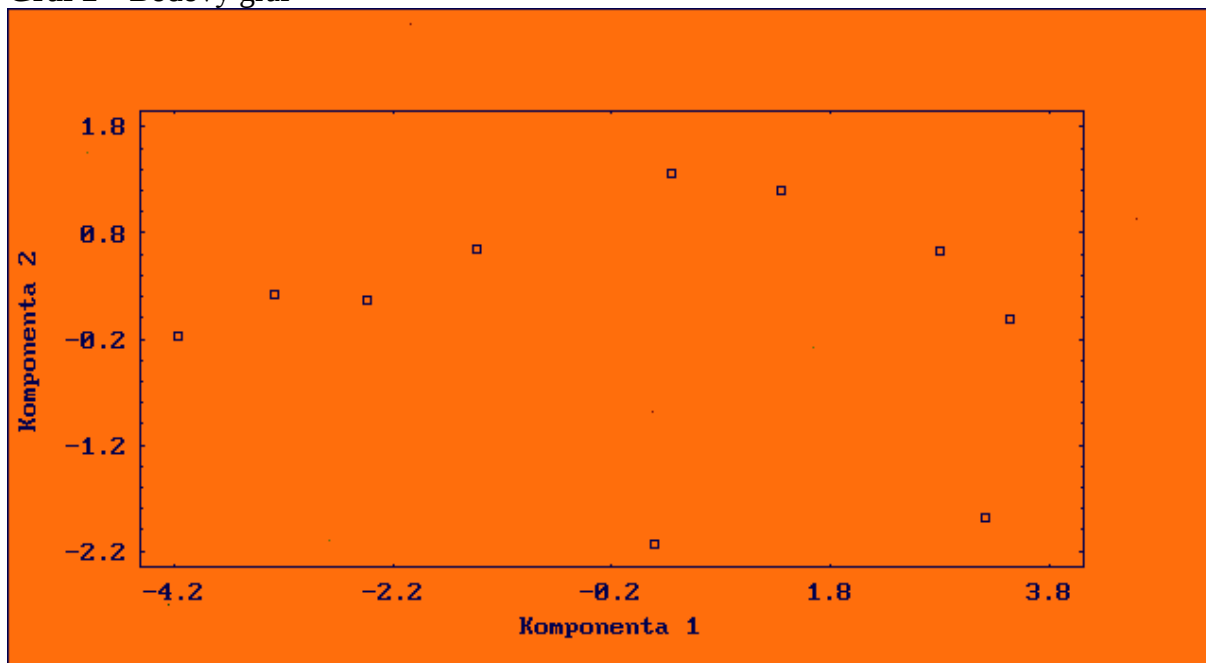
3. *Biplot* (dvojný graf)

Tento graf kombinuje oba předchozí grafy. Přímkami spojující se v bodě (0,0) představují původní proměnné (jejich „příspěvek“ k prvním dvěma hlavním komponentám) a body představují komponentní hodnoty jednotlivých objektů. Úhel mezi přímkami je nepřímo úměrný jejich vzájemné korelaci. Př. úhel mezi proměnnými MD a YS je velmi malý $\Rightarrow$ obě proměnné spolu velice korelují (jak je patrné i z tabulky 2 - párový korelační koeficient je 0,9778).

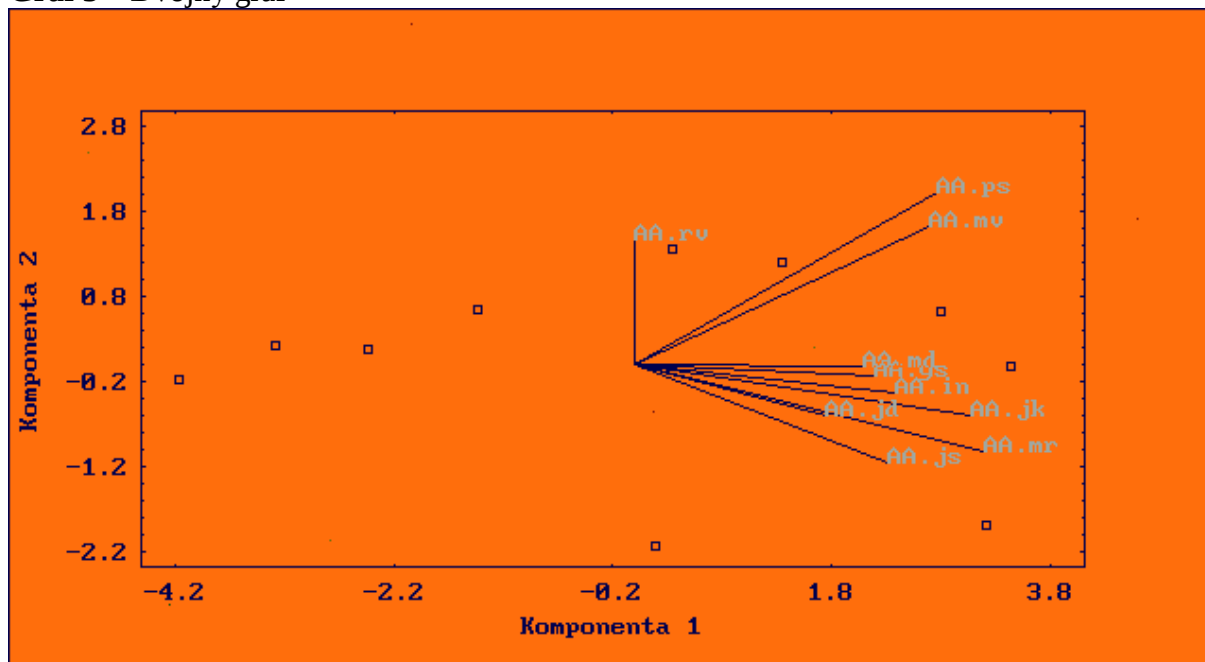
Graf 1 – Graf komponentních vah



Graf 2 – Bodový graf



Graf 3 – Dvojný graf



4. Faktorová analýza

Faktorová analýza patří, podobně jako Metoda hlavních komponent, mezi metody, které redukuji počet proměnných. FA předpokládá, že každou vstupující proměnnou lze vyjádřit jako lineární funkci nevelkého počtu společných latentních proměnných (faktorů) a jediného chybového faktoru. Na rozdíl od komponentní analýzy se tedy při faktorové analýze snažíme vysvětlit závislost proměnných.

Vstupní data byla standardizována. Do výpočtu byly zařazeny pouze objekty se známými hodnotami všech vstupních proměnných (*Missing Values:Listwise*). Jako typ rotace byla zvolena metoda *Equimax*, která spojuje kriteria *Varimax* a *Quartimax*, což znamená, že při rotaci je maximalizován rozptyl čtverců faktorových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. Za kritérium konvergence byla dosazena standardní hodnota 1E-5 a byl zvolen maximální počet interakcí při rotování faktoru 100. V tabulce 4 jsou uvedeny získané hodnoty faktorové analýzy. V prvním sloupci jsou jednotlivé objekty (tj. parametry). Ve druhém sloupci jsou odhady komunalit (tj. čtverce indexu korelace). Ve třetím sloupci jsou uvedena čísla faktorů. Ve čtvrtém sloupci jsou příslušné variability. Procenta variability jsou v pátém sloupci a v posledním sloupci jsou uvedena procenta variability kumulativně.

Všechny proměnné mají komunalitu rovnou hodnotě 1. Protože platí pravidlo, že čím má proměnná větší komunalitu, tím má i větší společné vlastnosti s ostatními, lze tvrdit, že mezi proměnnými existuje velmi silná závislost.

Podobně jako u metody PCA i zde je zapotřebí k popsání alespoň 85 % variability dat 2 faktory, které kumulativně popisují 94,4 % variability.

Tabulka 4 – Faktorová analýza

Variable	Communality	Factor	Eigenvalue	Percent Var	Cum Percent
YS	1,0000	1	7,87072	78,7	78,7
JS	1,0000	2	1,57025	15,7	94,4
JK	1,0000	3	0,29741	3,0	97,4
JD	1,0000	4	0,20438	2,0	99,4
RV	1,0000	5	0,03977	0,4	99,8
MV	1,0000	6	0,01144	0,1	99,9
IN	1,0000	7	0,00428	0,0	100,0
MR	1,0000	8	0,00162	0,0	100,0
PS	1,0000	9	0,00014	0,0	100,0
MD	1,0000	10	0,00001	0,0	100,0

Dalším výstupem faktorové analýzy je tabulka matice odhadu faktorových vah (viz tabulka 5). Protože jsme v předešlé analýze nechali hodnoty zdrojových proměnných standardizovat, představují prvky matice korelační koeficienty mezi jednotlivými proměnnými (v řádcích) a společnými faktory (v sloupcích). První faktor je silně korelován proměnnými YS, JK, IN, MD a slabě korelován proměnnou RV. Oproti tomu druhý faktor je silně korelován proměnnou RV a slabě korelován proměnnými YS, JK, IN, MD.

Tabulka 5 – Matice faktorových vah

Variable	Factor	
	1	2
YS	0,98052	0,00918
JS	0,91793	-0,32158
JK	0,97677	-0,08977
JD	0,92384	-0,20049
RV	-0,03625	0,94183
MV	0,85769	0,45217
IN	0,98011	-0,05812
MR	0,95670	-0,19724
PS	0,81230	0,53329
MD	0,99321	0,01802

V tabulce 6 jsou zaznamenány odhady komunalit výše uvedené faktorové matice po rotaci metodou Equimax. Oproti prvním odhadům komunalit u proměnných došlo u všech ke snížení.

Tabulka 6 – Odhady komunalit faktorové matice

Variable	Est Comunalit
YS	0,96150
JS	0,94600
JK	0,96214
JD	0,89368
RV	0,88835
MV	0,94009
IN	0,96400
MR	0,95419
PS	0,94422
MD	0,98679

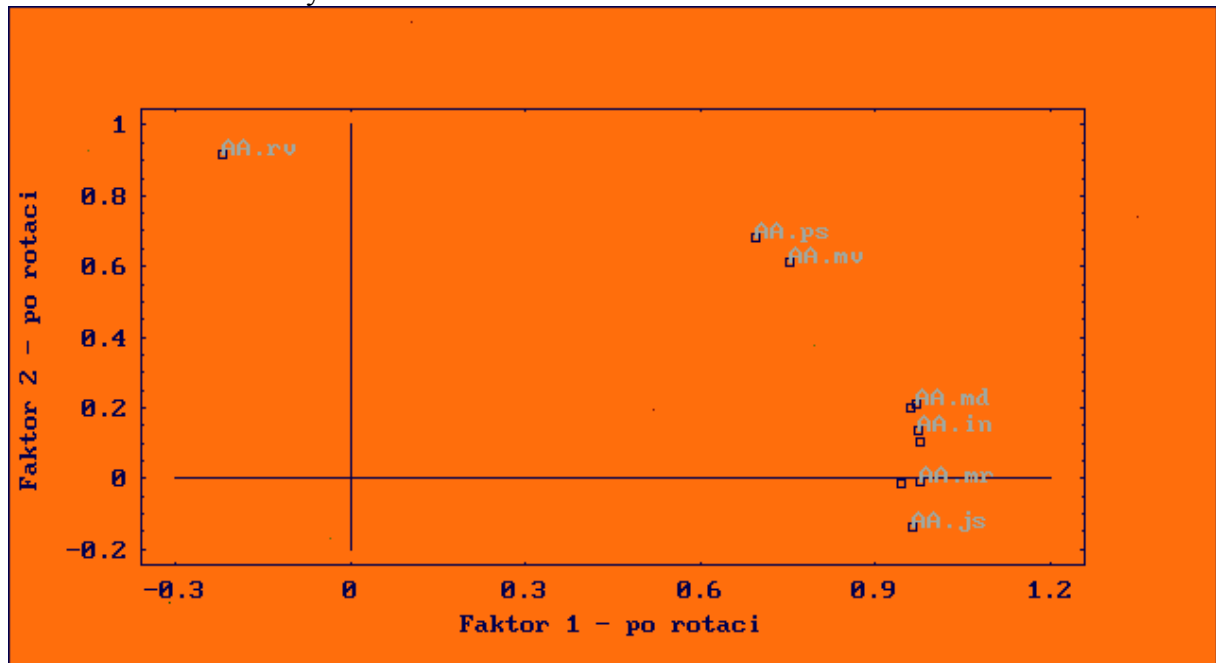
V následující tabulce jsou odhady faktorových vah pro faktorovou matici po rotaci, kterou jsme zadali ve vstupním formuláři (Equimax). Důvodem rotace je optimální natočení souřadnic tak, aby se získalo co nejvíce faktorově čistých proměnných. První faktor je korelován proměnnými: YS, JS, JK, JD, IN, MR, MD. Druhý faktor je korelován proměnnou RV. Proměnné MV a PS významně nekorelují žádný z faktorů. Celkové hodnocení rotace bude provedeno v kapitole 7 – Závěr.

Tabulka 7 – Matice faktorových vah po rotaci

Variable	Factor	
	1	2
YS	0,96016	0,19896
JS	0,96284	-0,13766
JK	0,97566	0,10116
JD	0,94518	-0,01771
RV	-0,21803	0,91696
MV	0,75384	0,60976
IN	0,97280	0,13286
MR	0,97679	-0,00816
PS	0,69360	0,68055
MD	0,97090	0,21010

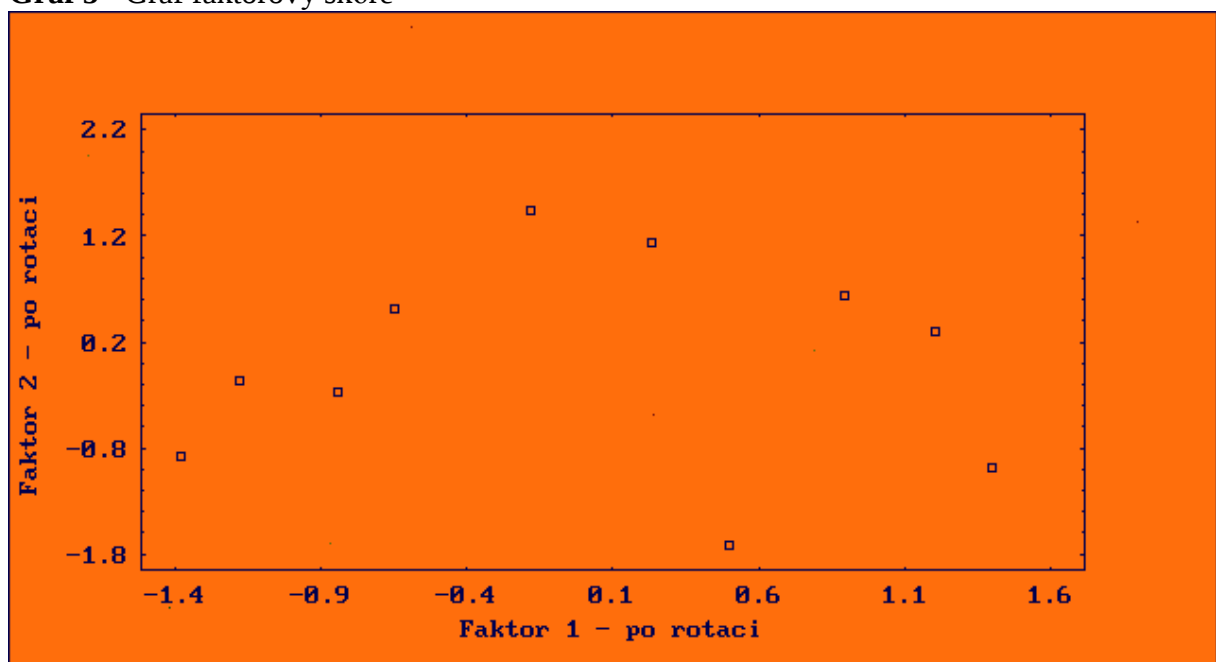
Faktorová analýza stejně jako metoda hlavních komponent poskytuje i grafické hodnocení. Prvním grafem, který FA nabízí je graf **Plot Factor Weights** (graf faktorových zátěží). Graf zobrazí hodnoty faktorových zátěží vstupujících proměnných v grafu, kde souřadnice tvoří první dva společné faktory (viz graf 4). Protože body v grafu jsou velice blízko u sebe, nejsou všechny označeny názvem proměnné. Poblíž bodu MD je bod YS, poblíž bodu IN je JK a u MR je bod JD.

Graf 4 – Graf faktorových zátěží



Dalším grafem, který faktorová analýza nabízí je Plot Factor Scores (graf faktorových skóre), ve kterém jsou znázorněny hodnoty vypočtených faktorových skóre pro první dva společné faktory u jednotlivých vstupujících pozorování (objektů).

Graf 5 - Graf faktorový skóre



5. Shluková analýza

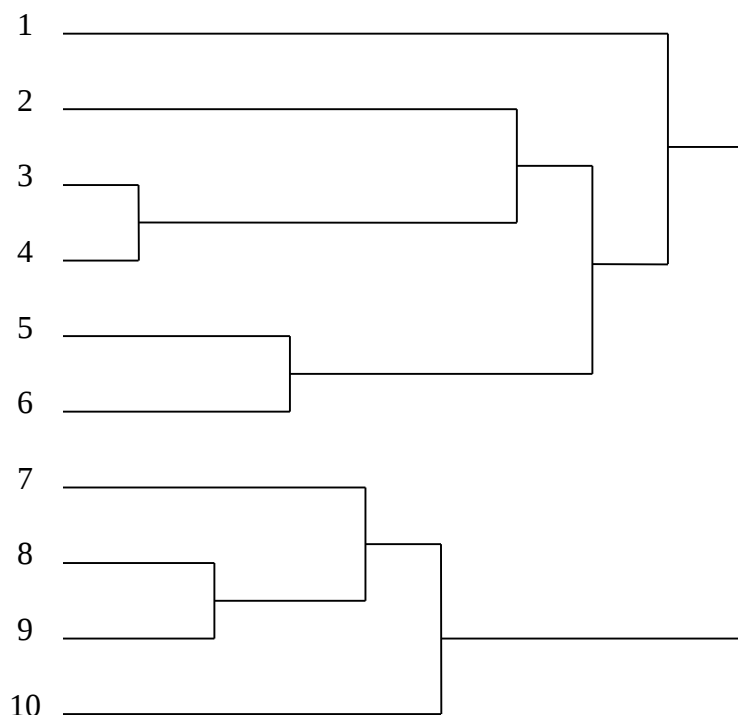
Shluková analýza patří mezi metody zabývající se zkoumáním podobnosti vícerozměrných objektů a jejich roztříděním do skupin. Statgraphics nabízí pro shlukovou analýzu několik metod tzv. hierarchické analýzy (metoda průměrová, centroidní metoda a metoda nejbližšího a nejvzdálenějšího souseda) a jeden postup nehierarchické analýzy (metoda typických bodů).

Jednou z nevýhod shlukové analýzy je to, že předem musíme určit do kolika shluků chceme proměnné seřadit a druhou nevýhodou je to, že shluková analýza neposkytuje jako výstup grafický dendrogram.

Provedla jsem shlukovou analýzu všemi metodami, které program Statgraphics nabízí. Výstupem každé metody je tabulka a grafické znázornění. Všechny metody poskytují shodné výsledky, proto si zde uvedeme pouze výstup jedné z nich, např. průměrové metody (Average). Abych získala optimální počet shluků a dendrogram, postupně jsem v zadání snižovala počet shluků od 9 do 1. Podle mého názoru je optimální počet shluků 4. Zcela individuální je koncentrace sledované látky u všech pacientů je v čase 0,5 hod, u ostatních časových intervalů je koncentrace podobná: 2. shluk (1-1,5-2hod), 3. shluk (2,5-3hod) a 4. shluk (4-5-6-8hod).

Dendrogram je uveden v grafu 6, zastoupení bodů v jednotlivých sluchách jsou uvedena v tabulce 8, grafické znázornění shlukové analýzy je v grafu 7.

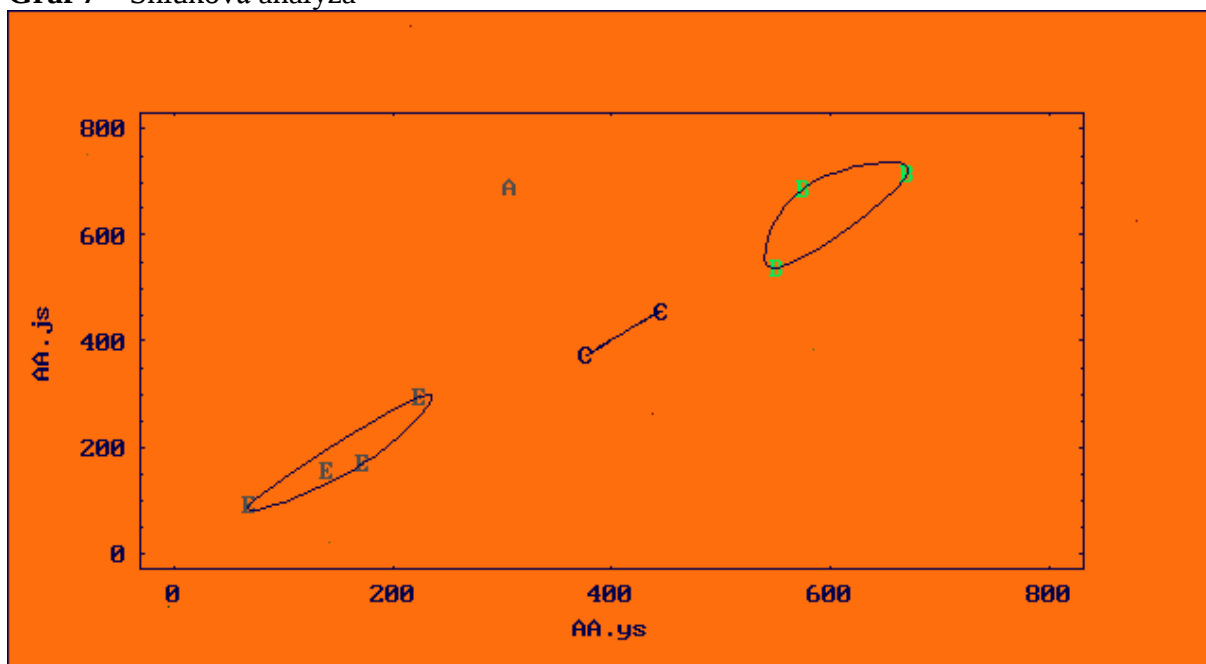
Graf 6 – Dendrogram



Tabulka 8 – Shluková analýza (metoda Average, počet shluků 4)

Pořadové číslo objektu	Časové intervaly	Shluk	Shluk	Absolutní četnost	Procentuální četnost
1	0,5	1	1	1	10
2	1	2	2	3	30
3	1,5	2	3	2	20
4	2	2	4	4	40
5	2,5	3			
6	3	3			
7	4	4			
8	5	4			
9	6	4			
10	8	4			

Graf 7 – Shluková analýza



6. Grafické metody zkoumání podobnosti objektů

Tyto metody slouží k vizuálnímu porovnávání vícerozměrných objektů. Program Statgraphcis používá dvou metod *Sun Ray Plot* a *Start Symbol Plot*

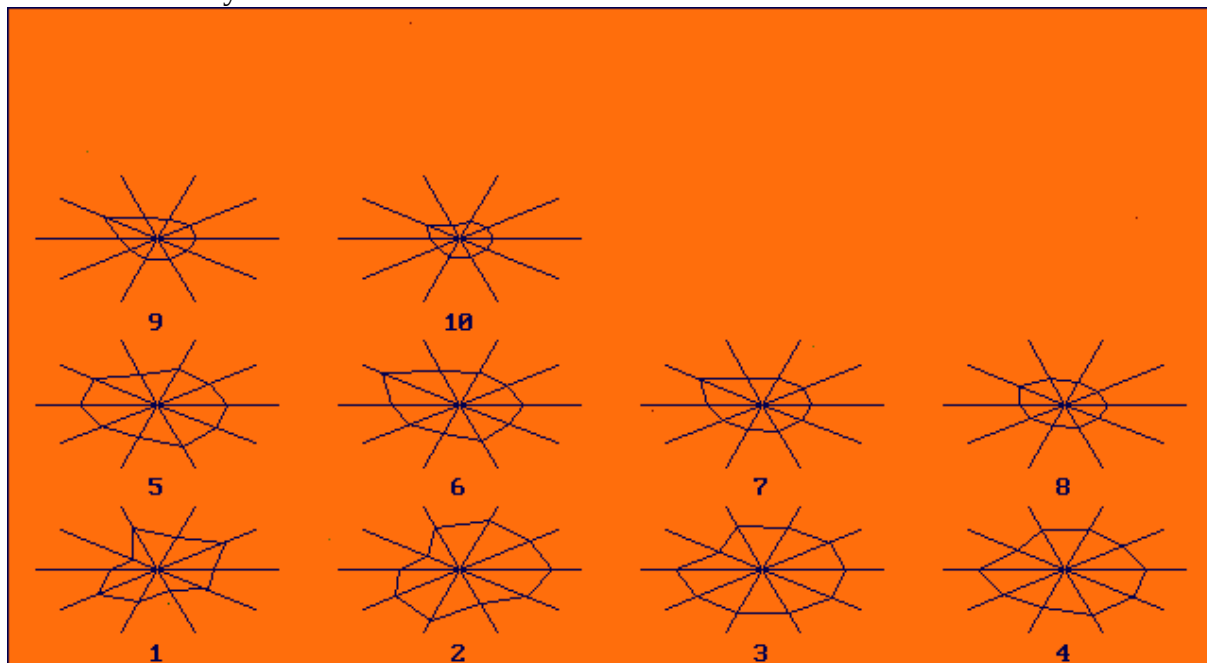
1. Sun Ray Plot

Tato procedura slouží k vizuálnímu porovnávání různých vícerozměrných objektů z hlediska jejich porovnání s průměrným objektem. Program nakreslí pro každý objekt hvězdičky, která je složena z paprsků spojených do jednoho bodu a úseček mezi body, které

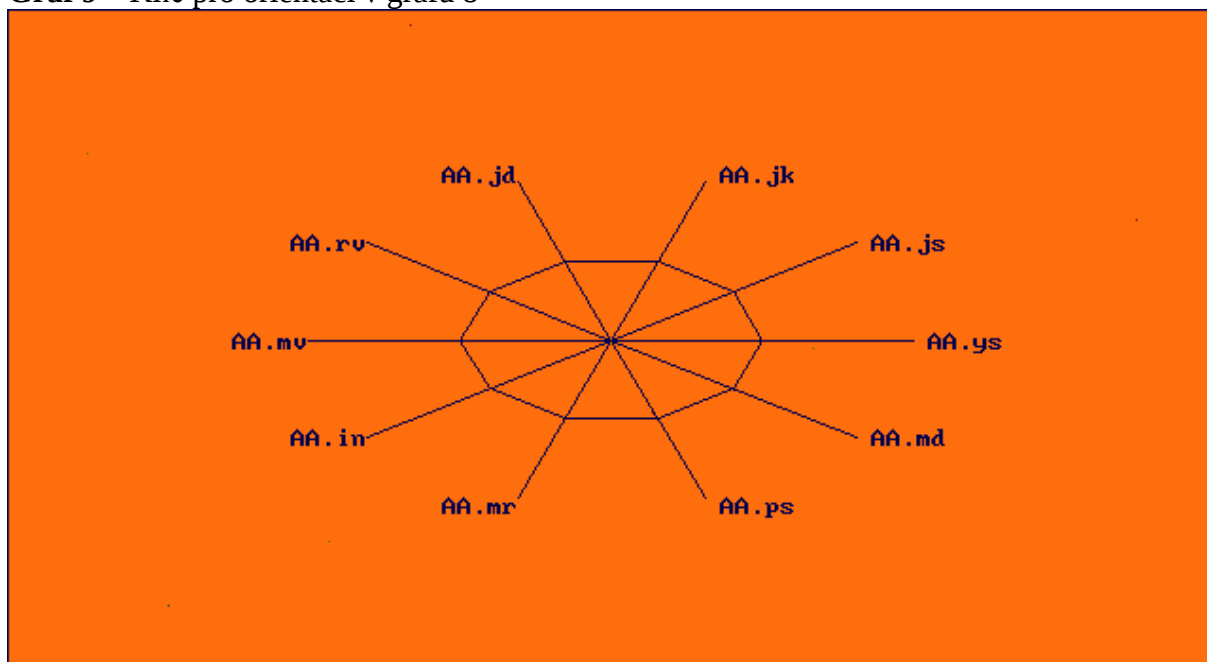
tvoří polygon. Počet paprsku odpovídá počtu proměnných (v našem případě 7), střed paprsku představuje průměr odpovídající proměnné a délka paprsku 2n násobek směrodatné odchylky.

Graf Sun Ray Plot nám poskytl podobné hodnocení jako shluková analýza. Za zcela odlišný objekt můžeme podle vizuálního porovnání označit objekt 1 (koncentrace v čase 0,5 hod). Za podobné můžeme označit koncentrace v časových intervalech: 1-1,5-2hod, 2,5-3hod a 4-5-6-8hod.

Graf 8 – Sun Ray Plot



Graf 9 – Klíč pro orientaci v grafu 8

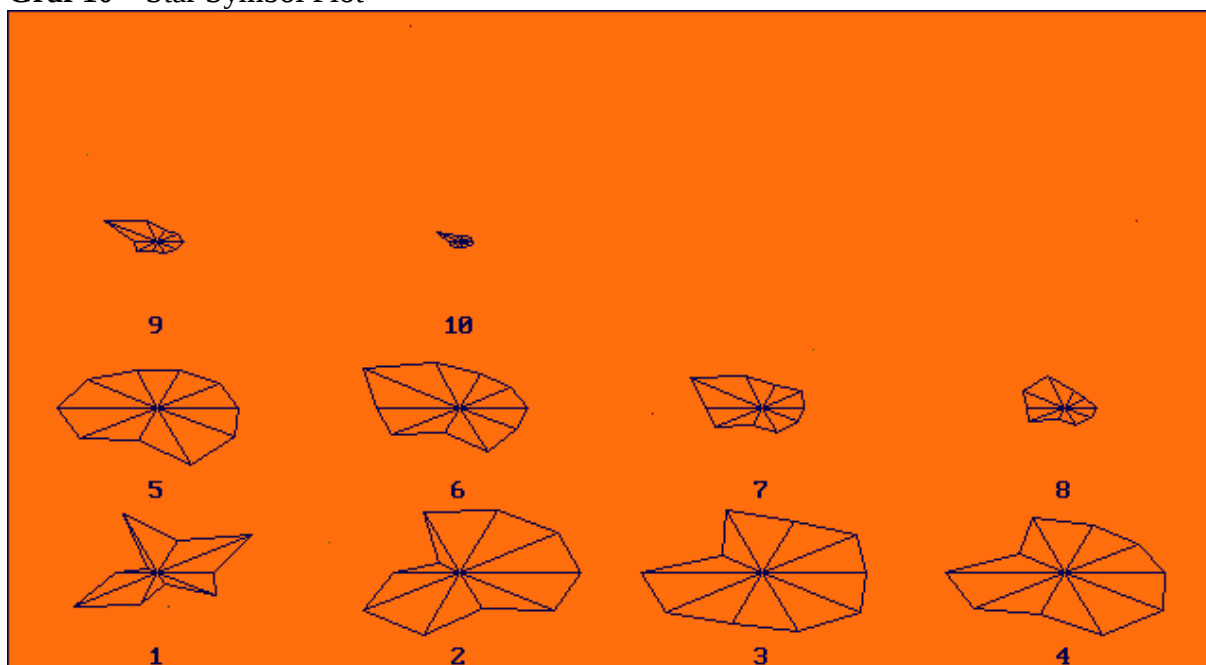


2. Star Symbol Plot

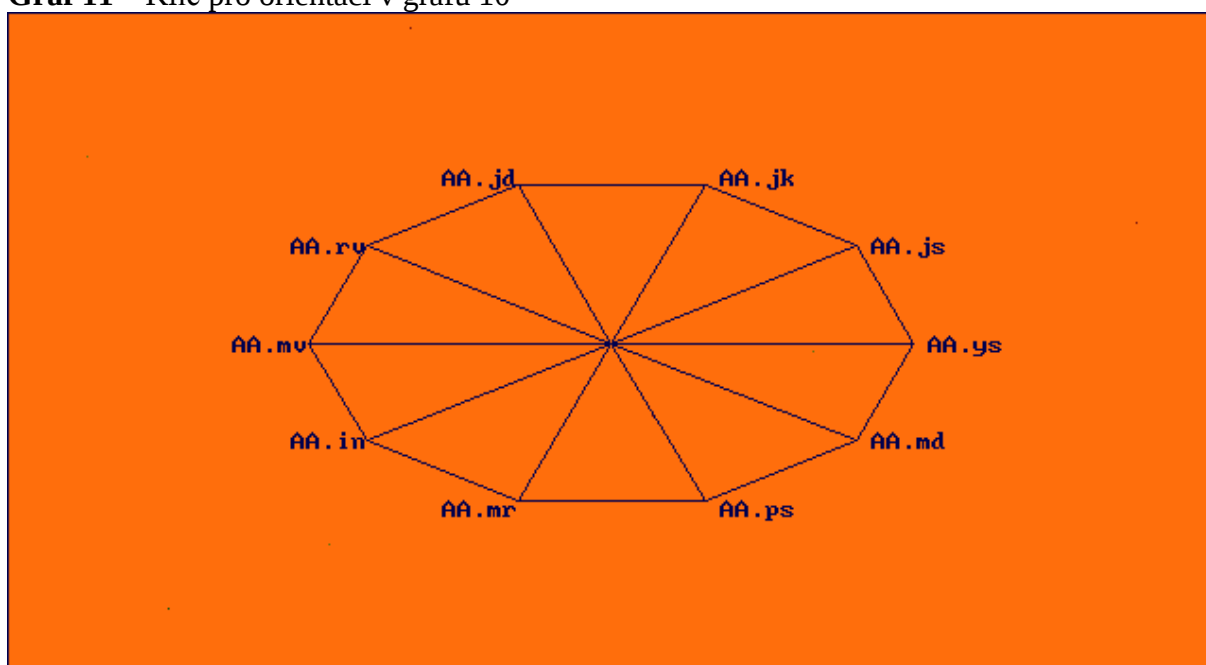
Podobně jako předchozí graf slouží k vizuálnímu zobrazení podobnosti vícerozměrných objektů. Graf je opět složen z paprsků, které se spojují v jednom centrálním bodě.

Grafické znázornění podává shodné závěry jako předešlý graf. Za zcela odlišnou můžeme označit koncentraci v čase 0,5 hod a za podobné můžeme označit koncentrace v časových intervalech: 1-1,5-2hod, 2,5-3hod a 4-5-6-8hod.

Graf 10 – Star Symbol Plot



Graf 11 – Klíč pro orientaci v grafu 10



7. Závěr

Soubor získaných koncentrací ursoxyly měřených v různých časových intervalech u 10 pacientů byl zpracován statistickým softwarem Statgraphics s využitím metod:

- kovarianční analýza,
- korelační analýza,
- metoda hlavních komponent,
- faktorová analýza,
- shluková analýza,
- grafické metody.

Výpočet **kovarianční matice** byl proveden jako vstup pro další výpočty. Pro porovnání podobnosti objektů má větší váhu **korelační matice**, která vznikne standardizací kovarianční matice (standardizací se zamezí vlivu rozdílného měřítka vstupních dat). Korelační koeficienty jsou mírou lineární závislosti mezi dvěma náhodnými veličinami. Pokud se korelační koeficient blíží 1, mezi danými proměnnými je silný korelační vztah. Záporné znaménko u korelačního koeficientu znamená sestupnou tendenci korelačního vztahu. Je-li velikost korelačního koeficientu menší jak 0,5, nelze mezi proměnnými hovořit o závislosti.

Z korelační matice v našem případě vyplývá, všichni pacienti mají mezi sebou velmi silný korelační vztah. Výjimkou je pouze pacient s iniciály RV, který nekoreluje asi s jedním z ostatních 11 pacientů a dvojce pacientů PS-JS, kteří spolu rovněž nekorelují.

Další metodou, které byly vstupní data podrobeny, je **Metoda hlavních komponent**. Tato metoda určuje společné latentní proměnné (hlavní komponenty), které vystihují 85-90 % vysvětlené variability v zdrojových datech. Metoda určila 2 hlavní komponenty, kterými bylo vysvětleno 95,9 % variability zdrojových dat.

Maketa Analýza hlavních komponent v softwaru Statgraphics nabízí i grafickou analýzu, kterou jsou 3 grafy:

Graf komponentních vah zobrazuje komponentní váhy zdrojových proměnných pro první dvě hlavní komponenty. Největší přínosem pro danou komponentu mají proměnné, které se v grafu nachází co nejbližší u souřadnice dané komponenty a na číselné ose co nejdále od nuly. Z grafu je patrné, že největší přínos pro první komponentu mají pacienti JK a MR a pro druhou komponentu pacient RV.

Bodový graf zobrazuje komponentní hodnoty prvních dvou hlavních komponent u jednotlivých objektů. Z grafu je patrné, že vstupní data pochází z normálního rozdělení, protože body v grafu jsou rozmístěny v jakémsi pomyslném kruhu.

Dvojný graf kombinuje oba předchozí grafy. Přímkami spojující se v bodě (0,0) představují původní proměnné (jejich „příspěvek“ k prvním dvěma hlavním komponentám) a body představují komponentní hodnoty jednotlivých objektů. Úhel mezi přímkami je nepřímě úměrný jejich vzájemné korelaci. Jak už bylo patrné z korelační matice, objekty spolu silně korelují, proto úhly mezi jednotlivými přímkami jsou malé. Výjimkou je opět přímka označující pacienta RV, která svírá s ostatními přímkami úhly větší. Zajímavá je poloha přímek MV a PS, ke které se později vrátíme u hodnocení faktorové analýzy.

Další metoda, která byla použita a která rovněž určuje společné latentní proměnné je **Faktorová analýza**. Tato metoda předpokládá, že každou vstupní proměnnou lze vyjádřit jako lineární funkci nevelkého počtu společných latentních proměnných a jediného chybového faktoru. Podobně jako metodou PCA byly nalezeny 2 společné latentní proměnné, které vysvětlují 95,9 % variability vstupních dat. V této metodě jsou latentní proměnné označovány jako faktory.

Při výpočtu byla použita rotace metodou Equimax, což znamenalo, že při rotaci byl maximalizován rozptyl čtverců faktorových vah a zároveň byl maximalizován součet čtvrtých mocnin faktorových zátěží. Jedním z následků rotace bylo to, že oproti původním hodnotám komunalit před rotací, došlo po rotaci k snížení hodnot komunalit. Druhým následkem rotace bylo zjednodušení tabulky Matice faktorových vah. Optimálním natočením souřadnic se získalo co nejvíce faktorově čistých proměnných.

Co se týče hodnocení hodnot komunalit. Před rotací nabývaly všechny komunality hodnoty 1. Protože obecně platí pravidlo, že čím je hodnota komunality vyšší, tím je objekt podobnější s ostatními objekty, znamenal by tento závěr, že jsou všechny objekty shodné. Po rotaci jsme získali reálnější výsledky. Přestože hodnoty klesly, jsou stále vysoké, což značí velkou podobnost mezi objekty (pacienty). Nejnižší hodnota je u pacienta RV, který nekoreluje asi s jedním z ostatních pacientů.

Podle faktorových vah před rotací korelují s prvním faktorem objekty – YS, JS, JK, JD, MV, IN, MR, PS, MD a s druhým faktorem koreluje objekt RV. Po rotaci došlo ke změně. Druhý faktor je stále korelován objektem RV, ale u prvního faktoru se počet objektů snížil. Důvodem jsou objekty MV a PS, u nichž rotace zapříčinila to, že mají shodný vliv na oba faktory. Tento závěr je reálnější. Ještě jednou se podívejme na graf komponentních vah (graf 1) a dvojný graf (graf 3) u metody hlavních komponent a na graf faktorových zátěží (graf 4) u faktorové analýzy. Ze všech 3 grafických zhodnocení vyplývá, že objekty MV a PS leží uprostřed mezi oběma latentními proměnnými.

Naše teoretické závěry se pokusíme uvést do praxe. Podívejme se do tabulky vstupních dat na pacienta RV a porovnejme koncentrace ursoxyly v jeho plazmě v různých časových intervalech s ostatními pacienty. Ostatní pacienti mají rychlý nárůst koncentrace přípravku do 1 až 1,5 hodiny po podání a od této hodiny koncentrace přípravku v krevní plazmě klesá. Oproti nim má pacient RV pomalý nárůst koncentrace ursoxyly v plazmě. Maximum koncentrace je až v časovém intervalu 3 hodiny po podání, což je dvojnásobný časový interval než u ostatních pacientů. Pacienti MV a PS nebývají maxima koncentrace přípravku v 1,5 hodině po podání přípravku jako ostatní pacienti, ale pak dochází k pomalejšímu vstřebávání z krevní plazmy. Pomocí těchto závěrů můžeme pojmenovat první i druhý faktor. První faktor nazveme klasické vstřebávání přípravku do krevní plazmy a druhý faktor nazveme pomalé vstřebávání přípravku do krevní plazmy.

Shluková analýza patří mezi metody zabývající se zkoumáním podobnosti vícerozměrných objektů a jejich roztríděním do skupin. Její nevýhodou je to, že předem musíme určit do kolika shluků chceme proměnné seřadit a že neposkytuje jako grafický výstup dendrogram. Protože jsem měla na analýzu pouze 10 objektů mohla jsem si dovolit jejich postupné shlukování od 9 do 2 shluků. Pomocí toho jsem dovedla sestavit dendrogram a zjistit optimální počet shluků – 4.

Posledními použitými metodami byly **grafické metody**, které jsou založeny na vizuálním porovnání objektů. Byly použity metody Sun Ray Plot a Start Symbol Plot. Obě grafické metody poskytly shodné hodnocení jako shluková analýza. Za zcela odlišný objekt lze považovat koncentraci přípravku v čase 0,5 hodin, shodné jsou následující časové intervaly: 1-1,5-2hod, 2,5-3hod a 4-5-6-8hod. Je zcela logické, že sousední intervaly jsou podobné, protože koncentrace se mění plynule.