

Příklad 2: Obsah PCB v játrech zemřelých lidí

Zadání:

V rámci Monitoringu zdraví byly měřeny koncentrace polychlorovaných bifenyků v játrech lidí zemřelých náhodnou smrtí ve věku 40 let a více. Sedm vybraných kongenerů PCB zatěžuje organismus svým účinkem nejvíce. Máme vyhodnotit výsledky analýz pilotní studie a určit, které kongenery zatěžují organismus nejvíce. Koncentrace PCB je uvedena v $\mu\text{g/kg}$ jater.

Data:

PCB028	PCB052	PCB101	PCB118	PCB138	PCB153	PCB180
0.89	1.06	3.14	0.38	48.5	79.2	60.5
0.93	9.85	30.04	20.72	130.3	309.7	286.9
2.33	1.34	0.68	1.88	4.4	8.1	6.4
0.29	0.67	0.14	0.61	3.1	4.9	3.5
0.15	1.05	0.35	1.90	8.4	17.8	12.7
4.11	1.59	6.70	12.12	67.4	171.4	155.1
0.12	0.74	0.23	1.02	14.0	11.7	8.2
1.81	2.40	2.77	5.88	97.5	55.1	44.8
0.40	0.40	0.40	0.40	17.6	49.1	45.3
0.37	0.10	0.10	0.21	1.7	3.6	3.8
0.29	0.27	0.11	0.11	2.5	11.9	11.3
0.33	0.67	0.33	1.46	25.4	22.8	18.1
2.05	0.32	0.32	1.34	20.5	34.8	25.9
1.76	0.24	10.74	20.77	29.7	47.6	36.3
0.13	0.13	0.13	0.50	4.2	6.6	6.2
0.21	0.21	0.62	0.86	9.6	16.9	17.8
0.07	0.07	0.62	0.17	1.8	1.6	1.2
0.19	0.19	0.19	0.19	5.2	3.8	3.7
0.21	0.21	0.48	0.43	6.9	9.3	9.4
0.08	0.08	0.08	0.30	2.3	2.2	1.8
0.15	0.15	0.52	0.15	19.6	8.0	6.1
0.18	0.91	0.18	0.66	7.9	13.0	9.8
1.00	1.00	1.00	1.00	28.5	37.7	39.9
11.24	0.24	0.24	0.24	4.4	5.5	4.2
5.21	20.50	10.23	4.67	30.7	56.0	41.1

Program: Statgraphics – multivariate methods

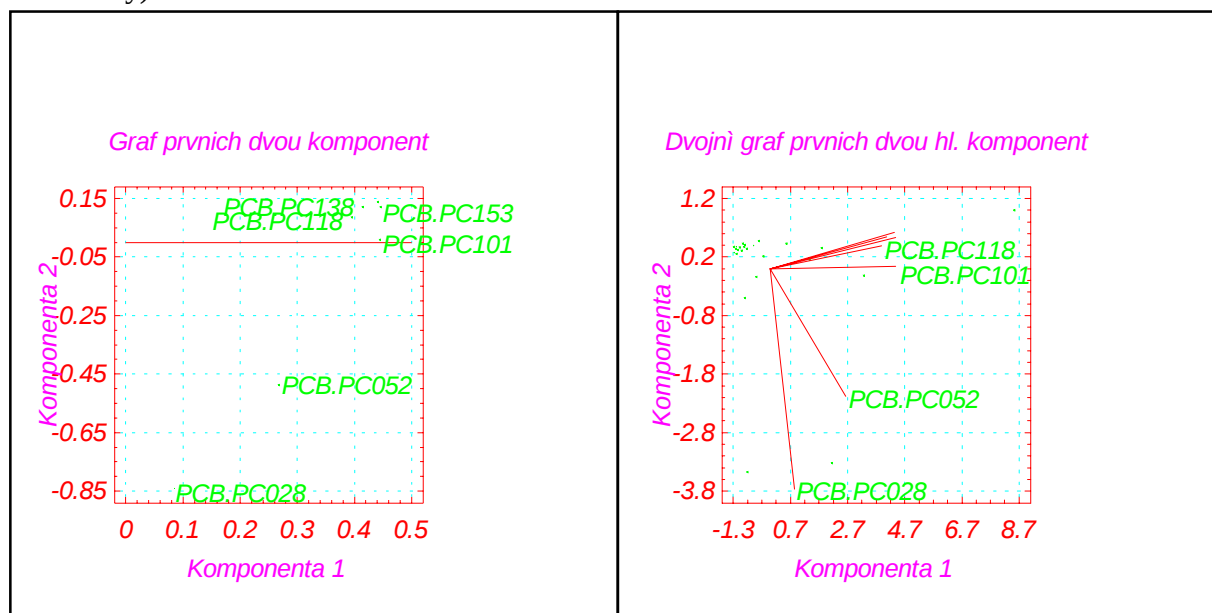
1. Analýza hlavních komponent

Metoda snižuje počet původních proměnných tím, že vytvoří lineární kombinaci zdrojových proměnných, které vysvětlují největší část jejich variability. První hlavní komponenta je taková kombinace vstupujících proměnných, která má největší rozptyl mezi všemi lineárními kombinacemi. Podobně následuje druhá hlavní komponenta atd. Pro dostatečné vysvětlení chování zdrojových proměnných požadujeme 85 – 90 % vysvětlené variability. Vstupní data byla při výpočtu standardizována (i když nemusela – mají stejný rozměr).

Tabulka vypočtených hodnot:

Komponenta č.	% variability	Celkem % variability
1	66.62	66.62
2	15.92	82.54
3	9.12	91.67
4	5.16	96.82
5	2.57	99.39
6	0.59	99.98
7	0.02	100

Jak vyplývá z uvedené tabulky, dostatečný počet hlavních komponent je 3 (91.7% vysvětlené variability).



2. Shluková analýza

Shluková analýza je metoda, která na základě podobnosti objektů umožňuje rozklad objektů do několika sourodých tříd (shluků). Podle způsobu shlukování se metody dělí na hierarchické a nehierarchické.

Podmínky výpočtu:

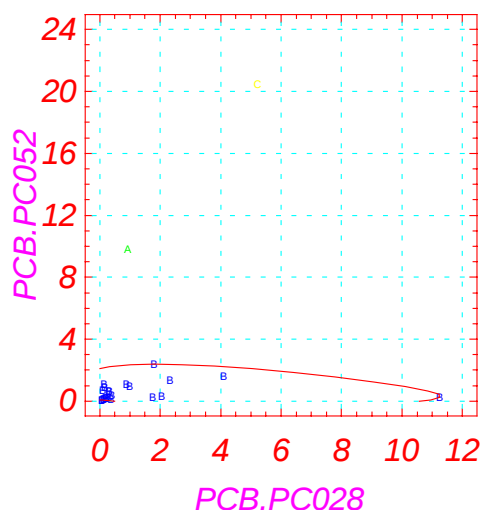
Metoda: Centroid

Konečný počet shluků: 3

Vzdálenost: Euclidean

Shluk	Četnost	Počet %
1	1(bod č. 2)	4
2	23	92
3	1(bod č. 25)	4

Shlukova analyza



3. Faktorová analýza

Patří mezi metody, které snižují počet proměnných. Každou proměnnou můžeme vyjádřit jako lineární funkci latentních proměnných. Touto metodou lze vysvětlit i závislost proměnných.

Podmínky výpočtu:

Typ rotace: Equimax

Kritérium konvergence: 1E-5

Počet iterací: 100

Výsledky:

Proměnná	Komunalita	Faktor	Char. číslo	% variability	Celkem % variability
PCB028	0.15566	1	4.66343	66.6	66.6
PCB052	0.69819	2	1.11457	15.9	82.5
PCB101	0.94216	3	0.63860	9.1	91.7
PCB118	0.83400	4	0.36096	5.2	96.8
PCB138	0.80313	5	0.17983	2.6	99.4
PCB153	0.99753	6	0.04132	0.6	100
PCB180	0.99736	7	0.00128	0	100

Na základě výsledků je počet faktorů odhadnut na 3 (91.7 % vysvětlené variability).

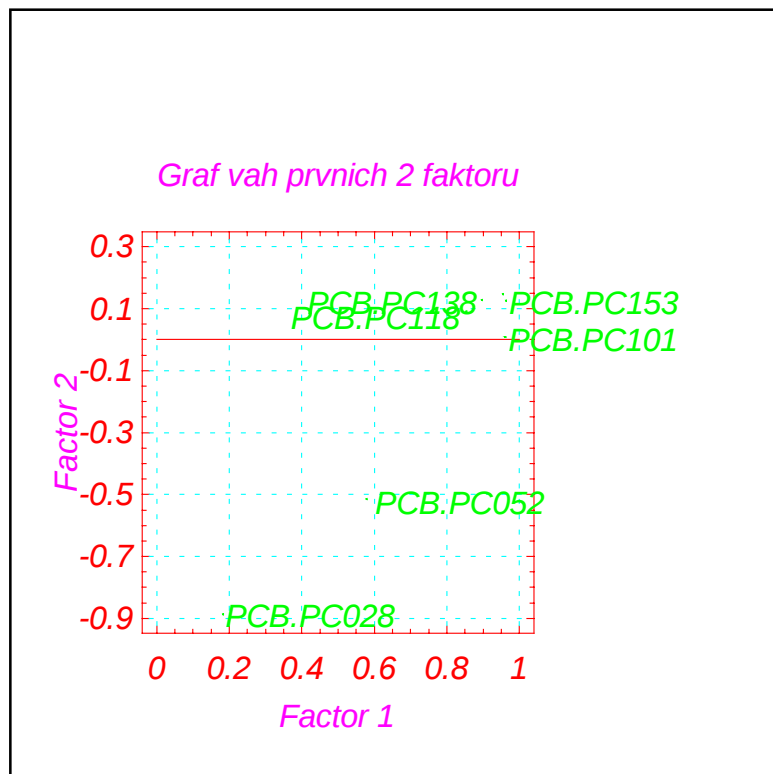
Matice faktorových vah:

Proměnná/Faktor	1	2	3
PCB028	0.18387	-0.88767	0.41968
PCB052	0.57916	-0.51457	-0.62888
PCB101	0.96079	0.00879	-0.10762
PCB118	0.85410	0.09098	0.18753
PCB138	0.89592	0.12768	0.08659
PCB153	0.96328	0.12618	0.07486
PCB180	0.95446	0.14578	0.08445

Odhady komunalit faktorové matice:

Proměnná	Odhad komunity
PCB028	0.99789
PCB052	0.99571
PCB101	0.93478
PCB118	0.77294
PCB138	0.82649
PCB153	0.94943
PCB180	0.93938

Graf faktorových zátěží zobrazuje hodnoty faktorových zátěží vstupujících proměnných pro první dva faktory:



Rotace faktorové matice:

Proměnná/Faktor	1	2	3
PCB028	0.01725	0.12581	0.99084
PCB052	0.12609	0.97404	0.17623
PCB101	0.81196	0.52301	0.04429
PCB118	0.84633	0.20830	0.11523
PCB138	0.86038	0.29137	0.03657
PCB153	0.91517	0.33219	0.03931
PCB180	0.91708	0.31245	0.02665

Faktor 1 koreluje s PCB101 – 180, faktor 2 s PCB052, faktor3 s PCB028. V tomto případě je popis pomocí 3 faktorů velmi zjednodušující..

4. Korelační matice

Používají se pro popis souvislostí mezi proměnnými. Nabývají hodnot od -1 do 1 . Pokud se hodnota korelačního koeficientu blíží těmto hodnotám, lze mluvit buď o kladné nebo záporné korelaci. V tabulce jsou v prvním řádku hodnoty korelačních koeficientů, pod nimi hladina významnosti. Pro $p < 0.05$ lze předpokládat korelaci mezi proměnnými (v tabulce jsou zvýrazněny kurzívou).

Jak je patrné z následující tabulky, není žádná korelace mezi PCB028 a ostatními PCB, PCB052 nekoreluje s PCB118. Ostatní korelační koeficienty jsou velmi vysoké a signalizují, že vstupují do organismu společně.

	PCB028	PCB052	PCB101	PCB118	PCB138	PCB153	PCB180
PCB028	1	0.2978	0.1205	0.1335	0.0933	0.1051	0.0910
		p=0.148	p=0.566	p=0.525	p=0.657	p=0.617	p=0.665
PCB052	0.2978	1	0.6076	0.3398	0.4102	0.4421	0.4193
	p=0.148		p=0.001	p=0.097	p=0.042	p=0.027	p=0.037
PCB101	0.1205	0.6076	1	0.8549	0.7821	0.8972	0.8915
	p=0.566	p=0.001		p=0.000	p=0.000	p=0.000	p=0.000
PCB118	0.1335	0.3398	0.8549	1	0.7066	0.7636	0.7531
	p=0.525	p=0.097	p=0.000		p=0.000	p=0.000	p=0.000
PCB138	0.0933	0.4102	0.7821	0.7066	1	0.8790	0.8671
	p=0.657	p=0.042	p=0.000	p=0.000		p=0.000	p=0.000
PCB153	0.1051	0.4421	0.8972	0.7636	0.8790	1	0.9980
	p=0.617	p=0.027	p=0.000	p=0.000	p=0.000		p=0.000
PCB180	0.0910	0.4193	0.8915	0.7531	0.8671	0.9980	1
	p=0.665	p=0.037	p=0.000	p=0.000	p=0.000	p=0.000	

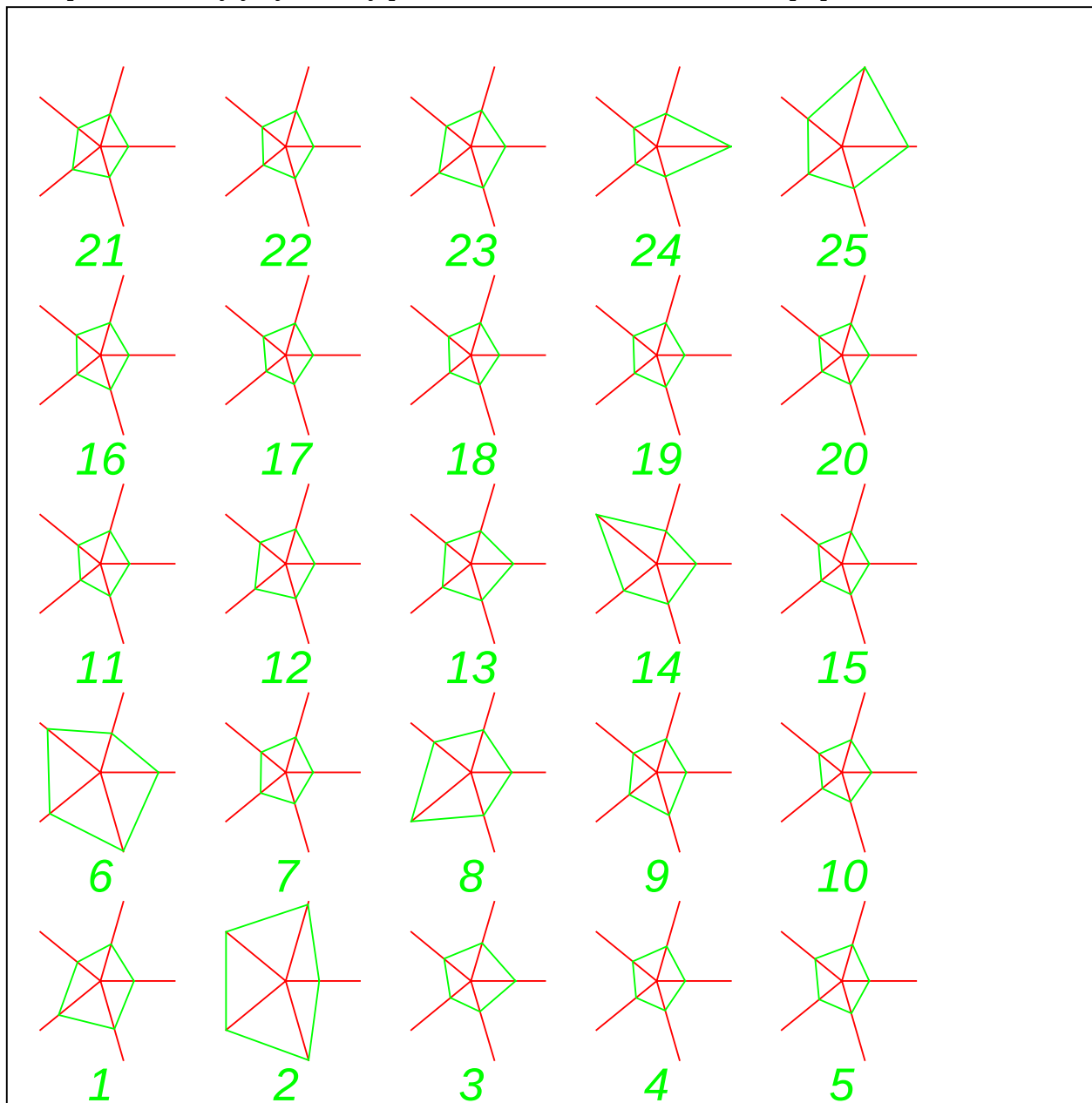
5. Kovarianční matice

Kovariance lze použít k testování závislosti několika sad dat. Závislost znamená, že velké hodnoty v jedné sadě odpovídají velkým hodnotám ve druhé sadě (kladná kovariance) nebo že velké hodnoty v jedné sadě odpovídají malým hodnotám ve druhé sadě (záporná kovariance). Pokud jsou hodnoty v obou množinách nezávislé, bude kovariance blízká nule. Z velikosti kovariance nelze ale odvodit souvislost mezi proměnnými.

	PCB028	PCB052	PCB101	PCB118	PCB138	PCB153	PCB180
PCB028	5.932213						
PCB052	3.157056	18.94456					
PCB101	1.884693	16.99046	41.26915				
PCB118	1.919913	8.733911	32.43088	34.86826			
PCB138	7.235767	56.8311	159.9525	132.8229	1013.415		
PCB153	17.16647	129.0114	386.39	302.2794	1875.899	4494.389	
PCB180	13.65685	112.4778	352.941	274.0308	1700.98	4122.901	3797.591

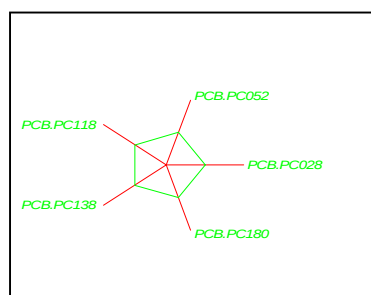
6. Graf Sluneční paprsky:

Používá se k vizuálnímu srovnání s průměrným objektem. Paprsky jsou spojeny v jednom bodě, který představuje průměr odpovídající proměnné a délka $2n$ násobek směrodatné odchylky. Pro větší přehlednost byly vynechány proměnné PCB101 a 153. V našem případě $n = 2$.



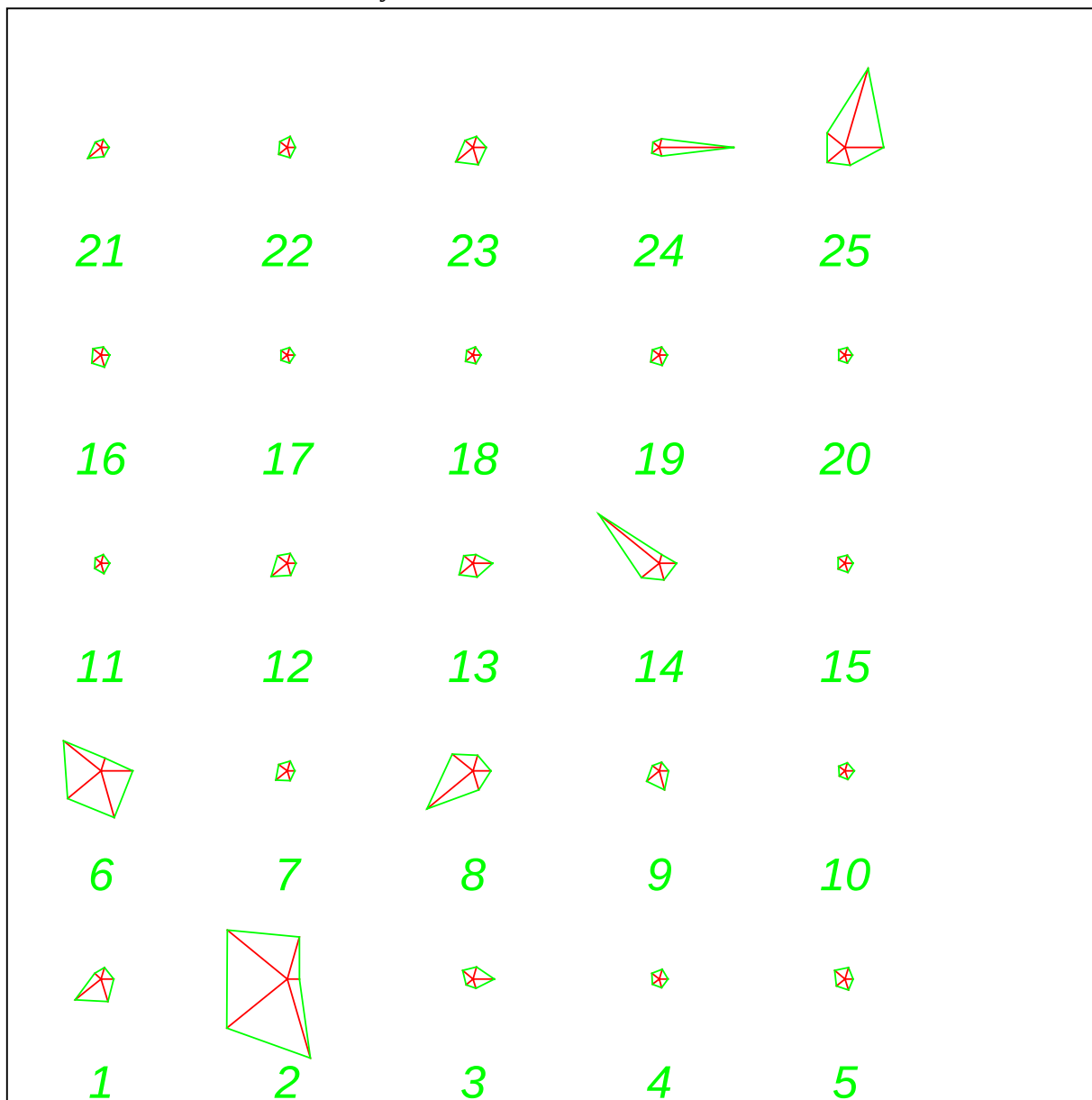
Jak je vidět ze znázorněných polygonů, jsou vzorky č. 2, 6, 8, 14 a 25 odlišné od ostatních.

Klíč ke grafu Sun Ray:

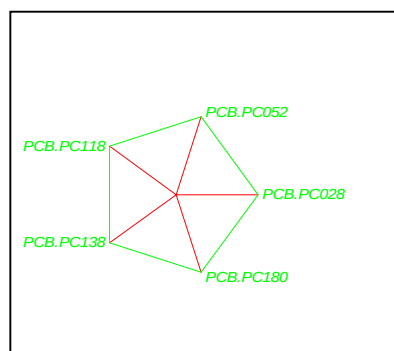


7. Graf Star Symbol

Graf je složen z paprsků, které se spojují v jednom bodě. Stejně směřující paprsky se liší svojí délkou. Nejkratší říká, že zde nabývá příslušná proměnná nejmenší hodnotu v celém souboru, obdobně se chová i nejdelší paprsek. V našem případě byly vypuštěny pro větší přehlednost PCB101 a PCB153, Shortest Ray = 0.1.



Klíč ke grafu Star Symbol:



Opět je zřetelně vidět, že body 2, 6, 8, 14 a 25 se liší od ostatních.

Závěr:

Obě grafické metody (Sun Ray a Star Symbol) shodně indikovaly body 2, 6, 8, 14 a 25 jako odlišné od ostatních. Obdobné výsledky poskytla i shluková analýza, přestože počet proměnných byl jiný (zde byly použity všechny). Metoda hlavních komponent určila 3 hlavní komponenty, metoda faktorové analýzy 3 faktory. V obou metodách je těmito vysvětleno více než 90 % variability. Korelační analýza určila PCB028 jako nekorelující s žádným dalším PCB. Nebyla nalezena korelace mezi PCB052 a PCB118. Ve všech ostatních případech jsou korelační koeficienty vysoké.

Název souboru: P2
Adresář: E:\Pom\vicerozm
Šablona: D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název: Příklad 1: Obsah PCB v játrech zemřelých lidí
Předmět:
Autor: Smid
Klíčová slova:
Komentáře:
Datum vytvoření: 07.09.97 17:14
Číslo revize: 23
Poslední uložení: 11.09.97 11:43
Uložil: Smid
Celková doba úprav: 430 min.
Poslední tisk: 15.09.00 09:57
Jako poslední úplný tisk
Počet stránek: 9
Počet slov: 1 139 (přibližně)
Počet znaků: 6 497 (přibližně)