

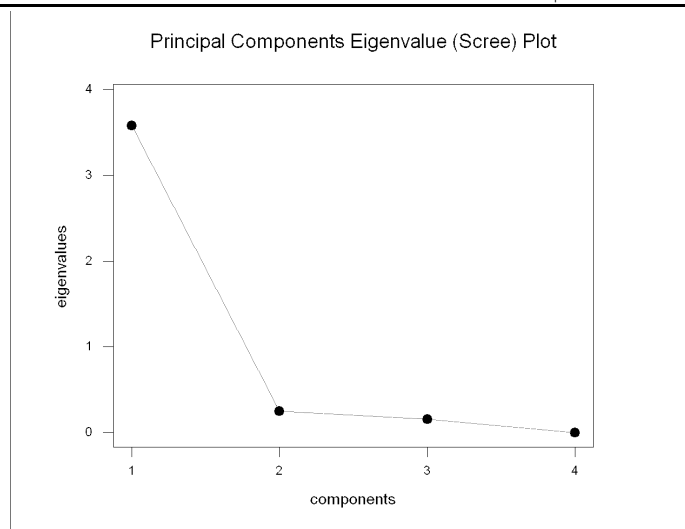
Vzorová úloha 4.2 Postup metody hlavních komponent

Na úloze **B4.02 Účinky neuroleptik při tlumení rozličných psychóz** si ukážeme pomůcky vícerozměrné analýzy dat. K analýze uijeme také škálovaná data.

Řešení: k analýze byl použit program NCSS2000. Výstup metody PCA programu NCSS2000 pro nestandardizovaná data úlohy B402 obsahuje:

1. Vyšetření indexového grafu úpatí vlastních čísel:

Index	Vlastní číslo λ_i	Individuální procento	Kumulativní procento	Kumulativní čárový graf úpatí
1	3394.339	92.62	92.62	
2	252.286	6.88	99.50	
3	15.8825	0.43	99.94	
4	2.295	0.06	100.00	



Obr. 4.7 Indexový graf úpatí vlastních čísel (Scree Plot) pro 20 objektů a 4 proměnné $B402X1$, $B402X2$, $B402X3$, $B402X4$ ze standardizovaných dat B402, SCAN.

Vlastní čísla slouží k určení počtu A "využitelných" hlavních komponent, jež si zvolíme v analýze k dalšímu užívání. Procento a kumulativní procento popisuje proměnlivost v původních proměnných, popsanou dotyčnou hlavní komponentou. K dalšímu popisu proměnlivosti bereme obvykle tolik hlavních komponent, aby jimi bylo popsáno 90 až 99 % celkové proměnlivosti. V tomto případě stačí užít první dvě. **Indexový graf úpatí vlastních čísel** je vlastně sloupcový diagram velikosti vlastních čísel proti stoupající hodnotě indexu, pořadového čísla. Zobrazuje relativní velikost jednotlivých vlastních čísel. Užitečné komponenty jsou tak odděleny zřetelným zlomovým místem a souřadnice x tohoto zlomu je hledanou hodnotou indexu.

2. (a) Vlastní vektory pro hlavní komponenty:

Proměnná	y_1	y_2	y_3	y_4
$B402X1$	-0.5684	0.4339	0.532	-0.4534
$B402X2$	-0.5688	0.4042	-0.584	0.4148
$B402X3$	-0.0758	-0.0286	0.6125	0.7863
$B402X4$	-0.5896	-0.8047	-0.0282	-0.0642

Vlastní vektory jsou váhy, jež umožňují kombinovat komponentní proměnné, které byly předem normovány vzorcem $(x_i - \bar{x})/F_i$. Např. první hlavní komponenta y_1 je *vážený průměr normovaných proměnných*, kdy váha každé proměnné je dána odpovídajícím prvkem prvního vlastního vektoru

$$y_1 = \leq_{11} x_1 \% \leq_{12} x_2 \% \dots \% \leq_{1m} x_m$$

Koeficienty v této rovnici vystihují relativní důležitost každé proměnné při tvorbě hlavní komponenty. Vlastní vektory bývají často normovány, takže rozptyl komponentního skóre je roven jedné.

(b) Čárový diagram absolutních hodnot vlastních vektorů:

Proměnná	y_1	y_2	y_3	y_4
$B402X1$				
$B402X2$				
$B402X3$				
$B402X4$				

Diagram absolutních hodnot vlastních vektorů vysvětluje strukturu a ukazuje, která proměnná silně koreluje s hlavní komponentou.

3. (a) Komponentní váhy:

Proměnná	y_1	y_2	y_3	y_4
B402X1	-0.9769	0.2033	0.0625	-0.0203
B402X2	-0.9793	0.1897	-0.0688	0.01857
B402X3	-0.8486	-0.0874	0.4688	0.2288
B402X4	-0.9372	-0.3487	-0.0031	-0.0026

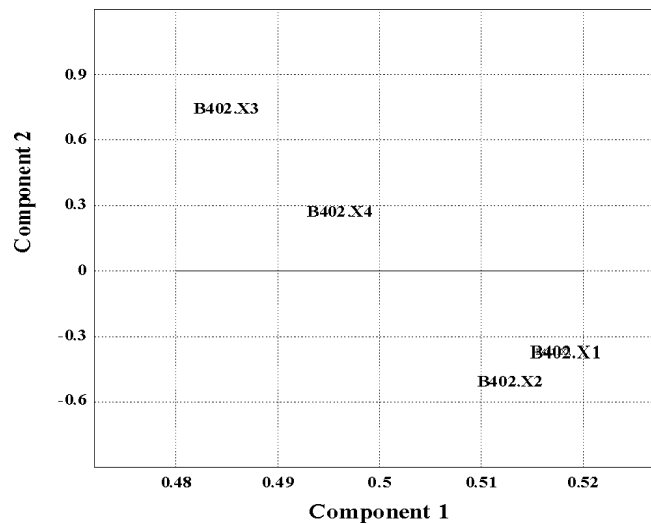
Ukazuje korelace mezi původními proměnnými a hlavními komponentami.

(b) Čárový diagram absolutních hodnot komponentních vah:

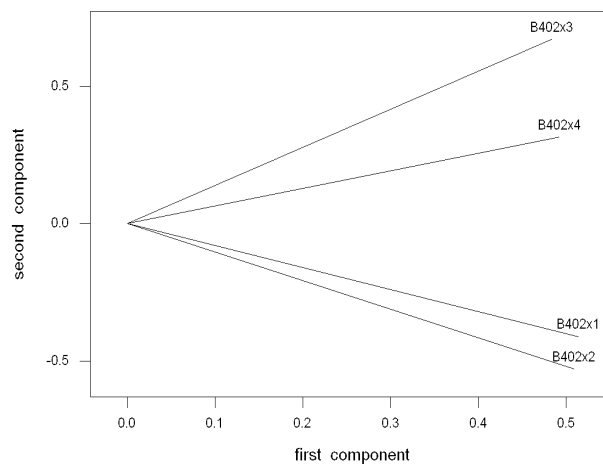
Proměnná	y_1	y_2	y_3	y_4
B402X1				
B402X2				
B402X3				
B402X4				

Diagram zobrazuje absolutní hodnoty komponentních vah a vysvětluje korelační strukturu, tj. která původní proměnná silně koreluje s hlavní komponentou. Je zřejmé, že 1. hlavní komponenta silně koreluje se všemi proměnnými, zatímco 3. nebo 4. hlavní komponenta koreluje pouze s B402X3.

4. Vyšetření grafu komponentních vah: grafická analýza předešlého kroku.

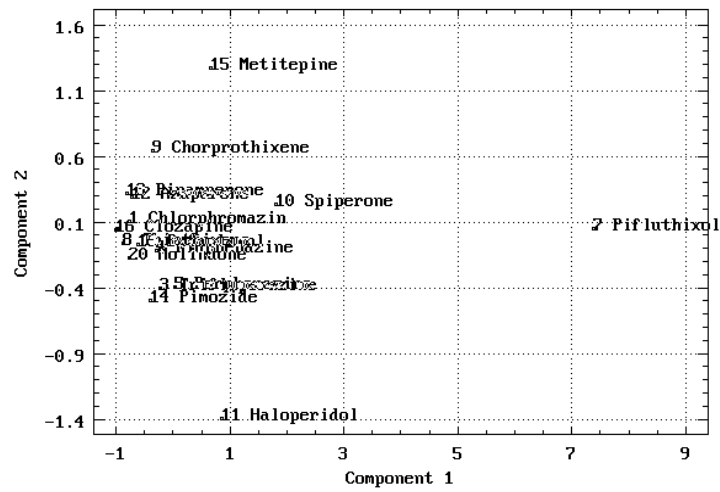


Obr. 4.8a Graf komponentních vah (Components Weights Plot) pro 20 objektů a 4 proměnné B402X1, B402X2, B402X3, B402X4 ze standardizovaných dat B402, STATGRAPHICS. Proměnné B402X1 a B402X2 leží v diagramu blízko sebe a proto korelují. Proměnné B402X3 a B402X4 jsou dále od sebe, proto méně korelují. Méně korelují rovněž dvě proměnnými B402X1 a B402X2, nachází se daleko od nich.

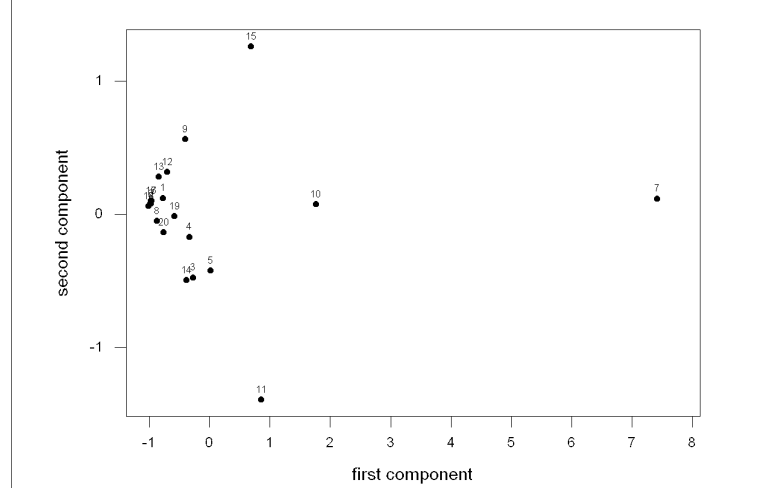


Obr. 4.8b Graf komponentních vah (Components Weights Plot) pro 20 objektů a 4 proměnné B402X1, B402X2, B402X3, B402X4 ze standardizovaných dat B402. Výklad je stejný jako u obr. 4.8a SCAN.

5. Vyšetření rozptylového diagramu komponentního skóre: nejdůležitější diagram metody hlavních komponent ukazuje celou vyšetřovanou strukturu objektů, tzn. shluky objektů, izolované objekty, odlehle objekty, anomálie atd. Objekty mohou být označeny textovým popisem nebo číselně, indexem.

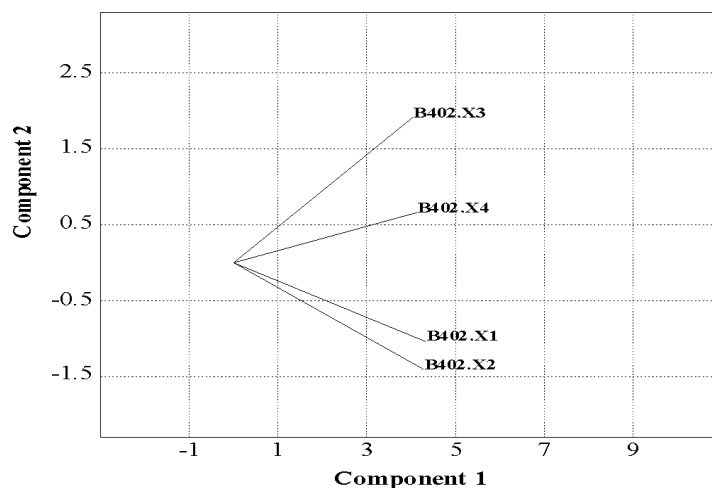


Obr. 4.9a Rozptylový diagram komponentního skóre (Scatterplot) pro 20 objektů a 4 proměnné B402X1, B402X2, B402X3, B402X4 standardizovaných dat B402, **STATGRAPHICS**. Kromě tří objektů: 7, 11 a 15 ležících v jediné shluku. Objekty 7, 11 a 15 jsou odlehle body. Nejvíce odlišný objekt od ostatních je 7, protože není odlišný na hlavní komponentě 1 popisující většinu rozptylu. První hlavní komponenta 1 popisuje rozdíl mezi Pifluthixolem a ostatními objekty. Na druhé straně objekty 11 a 15 jsou extrémní na druhé hlavní komponentě a udávají její směr. Ostatní objekty tvoří v rovině prvních dvou hlavních komponent homogenní shluk.

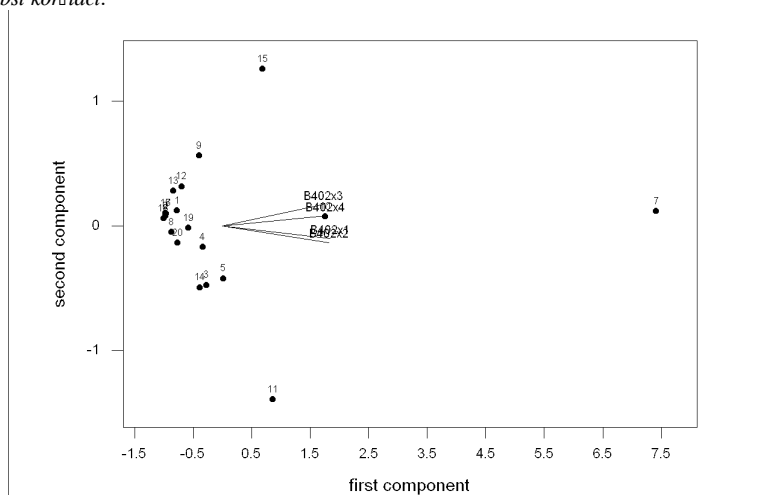


Obr. 4.9b Rozptylový diagram komponentního skóre (Scatterplot) pro 20 objektů a 4 proměnné B402X1, B402X2, B402X3, B402X4 standardizovaných dat B402. Výklad je stejný jako u obr. 4.9a, **SCAN**.

6. Vyšetření dvojného grafu: je důležité sledovat interakci objektů a proměnných. Je-li některý objekt umístěn ve dvojném grafu na stejném místě nebo alespoň poblíž místa proměnné, jsou spolu v interakci. Interakce poslouží interpretaci objektů.



Obr. 4.10a Dvojný graf (Biplot) pro 20 objektů a 4 proměnné *B402X1*, *B402X2*, *B402X3*, *B402X4* standardizovaných dat úloh B402, *STATGRAPHICS*. Úhly mezi dvěma průvodiči dvou proměnných jsou přímo úměrné velikosti korelace mezi těmito proměnnými. Malý průvodič *B402X1* a *B402X2* je malý úhel, což svědčí o silné korelaci. Úhly mezi dvěma průvodiči a průvodiči *B402X3* a *B402X4* jsou pak větší, což ukazuje na jejich slabší korelaci.



Obr. 4.10b Dvojný graf (Biplot) pro 20 objektů a 4 proměnné *B402X1*, *B402X2*, *B402X3*, *B402X4* standardizovaných dat úloh B402. Výklad je stejný jako u obr. 4.10a, *SCAN*.

7. Vyšetření grafu úpatí reziduí objektů: těsnost proložení PCA modelu danými daty se posuzuje statistickou analýzou reziduí. Rezidua by měla vykazovat dostatečnou těsnost.