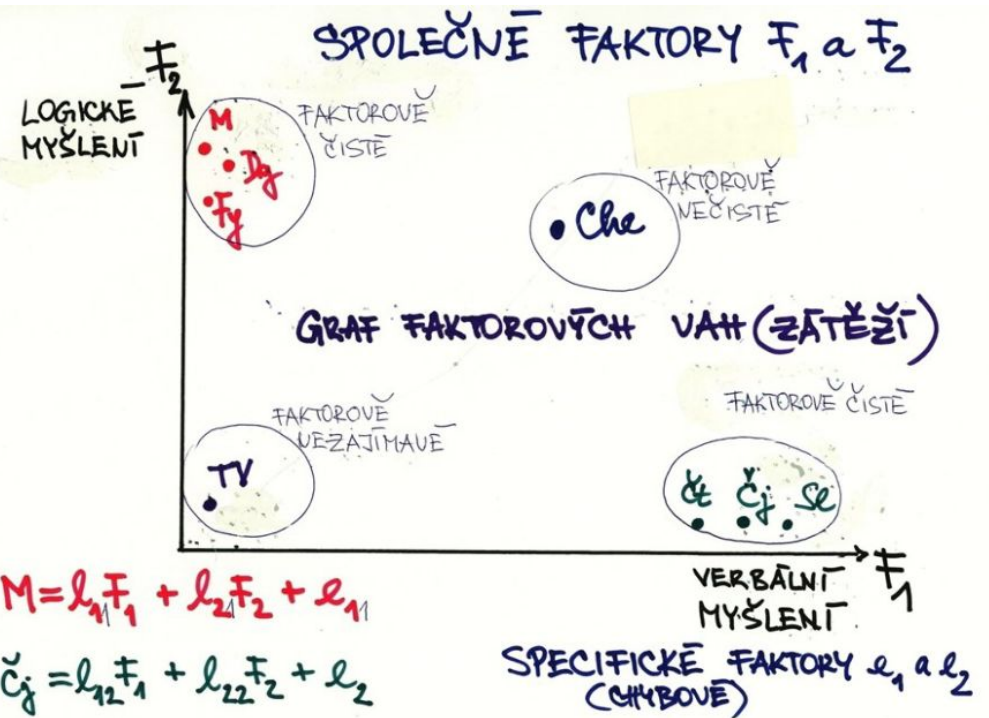
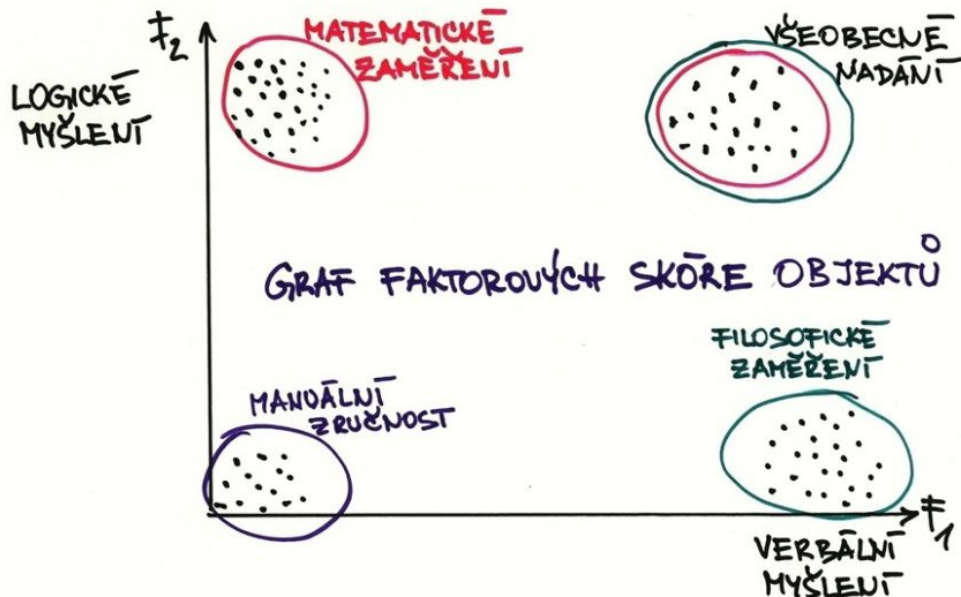


Faktorová analýza (FA)

1. Zaměření faktorové analýzy:

- 1) K vyšetření **vnitřních souvislostí** (korelace) a **odhalení struktury** datové matice.
- 2) Analýza struktury vnitřních vztahů mezi velkým počtem znaků, **zavedením** několika **faktorů**.
- 3) Po identifikaci faktorů je každému faktoru přidělen **obsahový význam**, pomocí kterého je každý znak vysvětlen.
- 4) Dva primární cíle FA:
 - a) **sumarizace a zestručnění dat**
 - b) **redukce dat.**



Formou **modelu faktorové analýzy (FA)** ve tvaru

$$x_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + e_1,$$

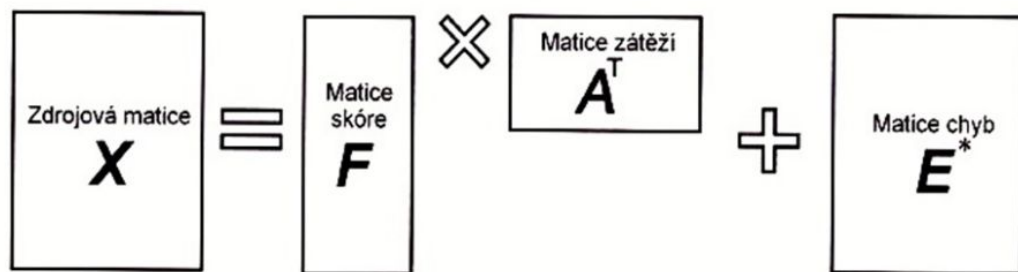
$$x_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + e_2,$$

.....

$$x_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + e_p,$$

kde $\blacklozenge F_1, F_2, \dots, F_m$ jsou vybrané **společné faktory** o počtu m , které vyvolávají korelace mezi P původními znaky.

\blacklozenge Faktory **mají** nulovou střední hodnotu a jednotkový rozptyl.



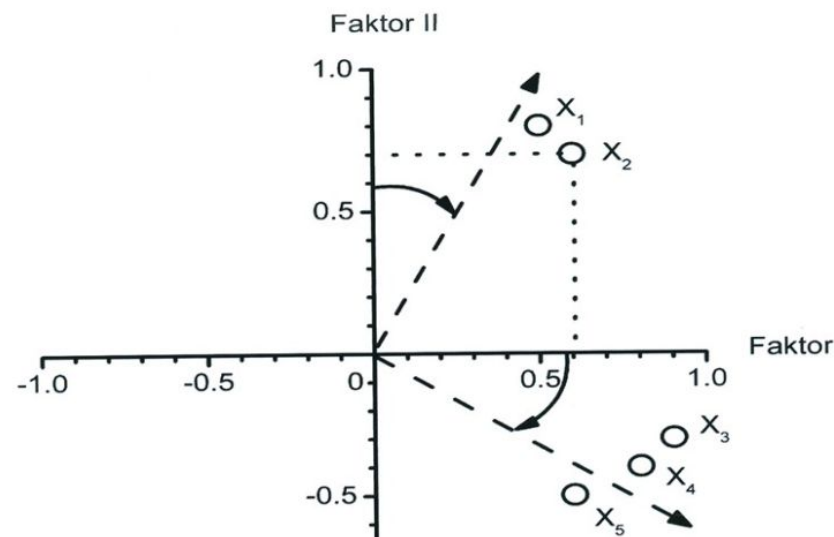
Faktorový model rozdělí:

- jednak každý znak x , na dvě části,
- jednak rozptyl každého znaku na dvě části.

(d) Pojmenování faktorů:

- ◆ Je třeba pojmenovat, vložit obsahový význam do faktorových zátěží.
- ◆ Znaky s vyššími zátěžemi jsou považovány za důležitější a mají i větší vliv na pojmenování dotyčného faktoru.
- ◆ Uživatel vyšetří **podtržené** znaky v dotyčném faktoru a vloží větší důraz na proměnné s vyšší hodnotou zátěže. Provede pojmenování či významové označení dotyčného faktoru.
- ◆ Znaménko + či - značí: kladné u faktoru značí kladný vztah zatímco záporné značí, že proměnné jsou negativně vztaženy.

Otočené faktory: některé zátěže znaků dosahují hodnot blízko ± 1 a zbývající zátěže pak téměř nulových. Proto si budeme přát, aby některý znak dosahoval vysoké zátěže pouze u jediného faktoru čili byl *faktorově čistý*.



Postup FA

Postup analýzy vnitřní struktury dat obsahuje:

1. Vyčíslí se korelační matice všech znaků:

- ◆ Vyznačí se znaky, které nejsou v korelaci s ostatními.
- ◆ Vyčíslí se vhodnost faktorového modelu.
- ◆ Rozhodne se, co dělat s "dírami" ve vstupní matici dat.

2. Provede se extrakce potřebných faktorů.

- ◆ Musíme se ujistit, jak dobře FA model prokládá data.

3. Rotací faktorů se docílí jejich lepší interpretace.

4. Vyčíslí se faktorové skóre:

- ◆ Pro každý objekt se naleznou shluky podobných objektů.

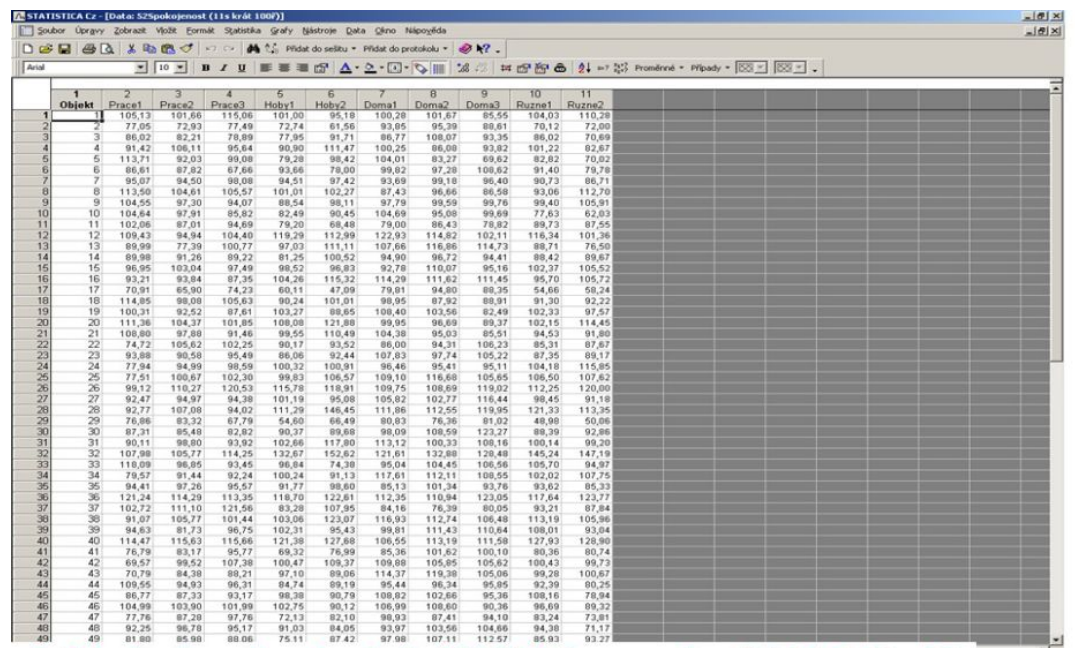
PŘÍKLAD 5.1 Sociologický průzkum spokojenosti

Dotazník o fiktivní spokojenosti v životě byl vyplněn 100 dospělými respondenty. Obsahoval 10 položek vystižení spokojenosti v v práci, se svými koníčky, spokojenost doma v rodině a v ostatních činnostech v životě.

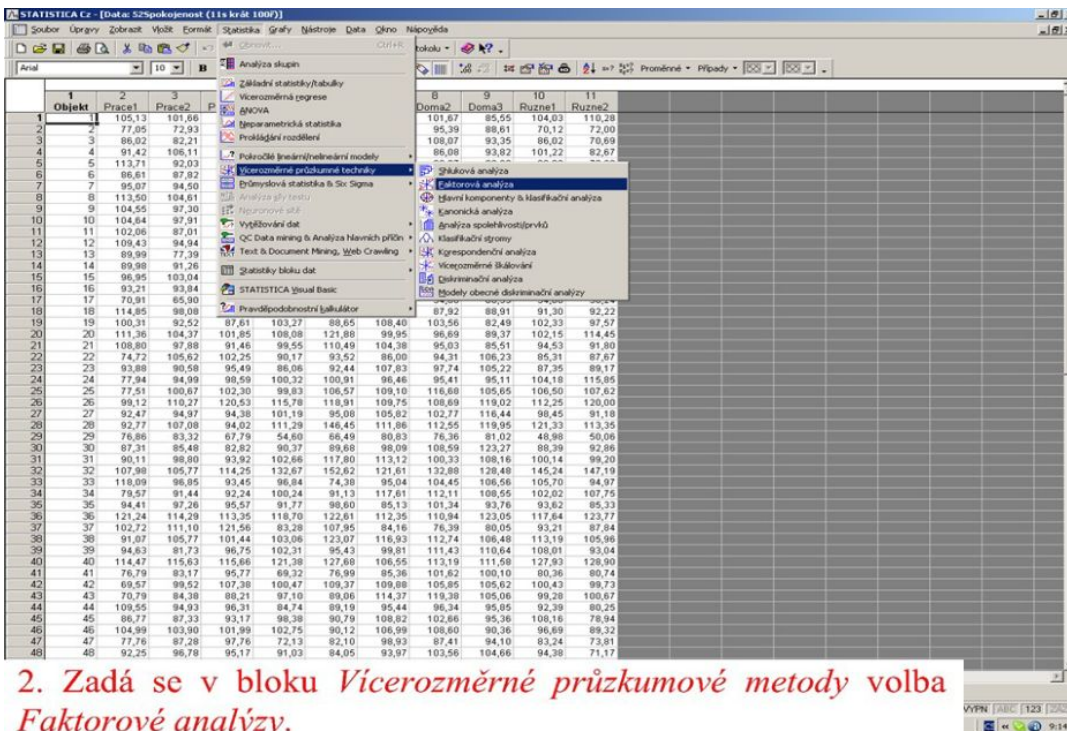
- 1) Cílem je nalezení vztahu mezi spokojenostmi v různých doménách.
- 2) Existují nějaké faktory, které jsou v pozadí těchto posuzovaných domén?
- 3) Jaký je obsah těchto faktorů a jak je pojmenujeme?

o **Data:** data obsahují 100 řádků (respondentů) a 10 sloupců znaků spokojenosti k činnosti na pracovišti (*Práce1, Práce2, Práce3*), spokojenosti s koníčky (*Hobby1, Hobby2*), spokojenosti doma v rodině (*Doma1, Doma2, Doma3*) a spokojenosti v ostatních aktivitách života (*Různé1, Různé2*). Odpovědi jsou škálovány, aby průměr znaku dosahoval hodnoty okolo 100.

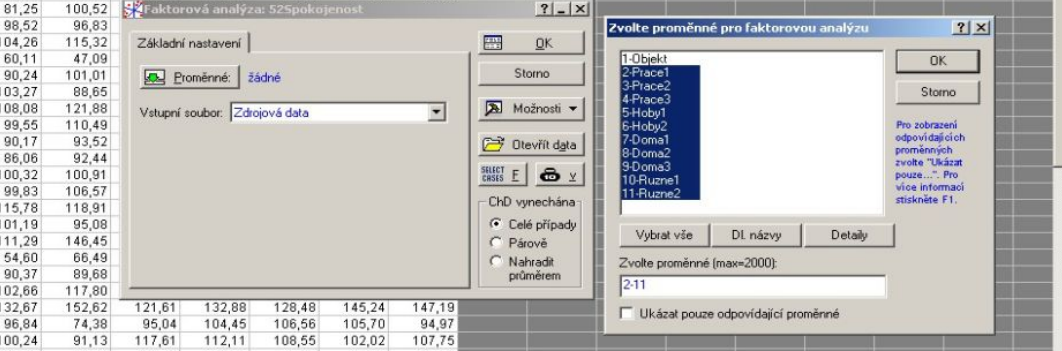
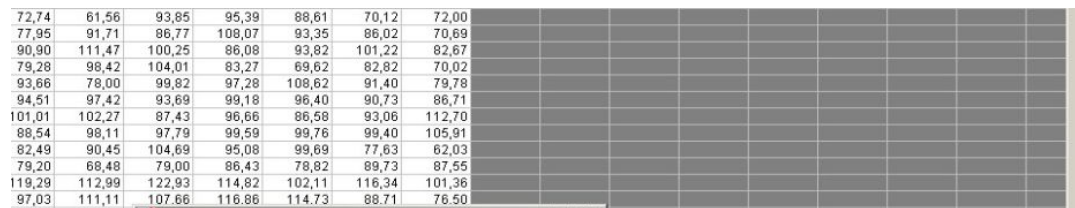
Objekt	Prace1	Prace2	Prace3	Hobby1	Hobby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
1	105.13	101.66	115.06	101.00	95.18	100.28	101.67	85.55	104.03	110.28
2	77.05	72.93	77.49	72.74	61.56	93.85	95.39	88.61	70.12	72.00
3	86.02	82.21	78.89	77.95	91.71	86.77	108.07	93.35	86.02	70.69
...
100	106.05	120.71	119.81	101.85	94.96	75.86	93.17	93.38	109.36	83.79



1. Zdrojová matice obsahuje 100 řádků respondentů (objektů) a 10 sloupců znaků s vyjádřením spokojenosti na pracovišti, s koníčky a doma.



2. Zadá se v bloku *Vícerozměrné průzkumové metody* volba *Faktorové analýzy*.



3. FA: V záložce *Základní nastavení* otevřeme blok *Proměnné* a zadáme 10 znaků.

100,47	109,37	109,00	105,05	105,02	100,43	99,73
97,10	89,06	114,37	119,38	105,06	99,28	100,67
84,74	89,19	95,44	96,34	95,85	92,39	80,25
98,38	90,79	108,82	102,66	95,36	108,16	78,94
102,75	90,12	106,99	108,60	90,36	96,69	89,32

1	101,66	115,06	101,00	95,18	100,28	101,67	85,55	104,03	110,28
15	72,93	77,49	72,74	61,56	93,85	95,39	88,61	70,12	72,00
12	82,21	78,89	77,95	91,71	86,77	108,07	93,35	86,02	70,69
12	106,11	95,64	90,90	111,47	100,25	86,08	93,82	101,22	82,67
1	92,03	99,08	79,28	98,42	104,01	83,27	69,62	82,82	70,02
1	87,82	67,66	93,66	78,00	99,82	97,28	108,62	91,40	79,78
17	94,50	98,08	94,51	97,42	93,89	99,18	96,40	90,73	86,71
10	104,61	105,57	101,01	102,27	87,43	96,66	86,58	93,06	112,70
15	97,30	94,07	88,54	98,11	97,79	99,59	99,76	99,40	105,91
14	97,91	85,82	82,49	90,45	104,69	95,08	99,69	77,63	62,03
16	87,01	94,69	79,20	68,4					
13	94,94	104,40	119,29	112,9					
19	77,39	100,77	97,03	111,1					
18	91,26	89,22	81,25	100,5					
15	103,04	97,49	98,52	96,8					
21	93,84	87,35	104,26	115,3					
1	65,90	74,23	60,11	47,0					
15	98,08	105,63	90,24	101,0					
1	92,52	87,61	103,27	88,6					
16	104,37	101,85	108,08	121,8					
10	97,88	91,46	99,55	110,4					
2	105,62	102,25	90,17	93,5					
18	90,58	95,49	86,06	92,4					
14	94,99	98,59	100,32	100,9					
1	100,67	102,30	99,83	106,5					
2	110,27	120,53	115,78	118,9					
17	94,97	94,38	101,19	95,0					
7	107,08	94,02	111,29	146,4					
16	83,32	67,79	54,60	66,4					
11	85,48	82,82	90,37	89,6					
1	98,8								
18	105,7								
19	96,8								
17	91,4								
11	97,2								
14	114,2								
2	111,1								
17	105,77								
13	81,73								
17	115,63								
9	83,17								
17	99,5								
9	84,3								
15	94,9								
7	87,33								

Metoda extrakce faktorů: 52Spokojenost

Celé případy byly vymečány u ChD

100 příp. bylo zpracováno (zvoleno)
100 platn. příp. akceptováno

Korelační matice byla spočítána pro 10 prom.

Základní nastavení | Detaily | Popisné statistiky

Max. počet faktorů:

Min. vlastní číslo:

OK | Storno | Možnosti

4. FA: V okně Metoda extrakce faktorů jsou podmínky výpočtu a tři záložky: Základní nastavení, Detaily a Popisné statistiky. Zůstaneme v záložce Základní nastavení a ve volbě Maximální počet faktorů kam zadáme číslo 2, v okénku omezení na vlastní čísla zadáme Minimální vlastní číslo a zvolíme zde Kaiserovo omezení 1.00.

1	101,66	115,06	101,00	95,18	100,28	101,67	85,55	104,03	110,28
15	72,93	77,49	72,74	61,56	93,85	95,39	88,61	70,12	72,00
12	82,21	78,89	77,95	91,71	86,77	108,07	93,35	86,02	70,69
12	106,11	95,64	90,90	111,47	100,25	86,08	93,82	101,22	82,67
1	92,03	99,08	79,28	98,42	104,01	83,27	69,62	82,82	70,02
1	87,82	67,66	93,66	78,00	99,82	97,28	108,62	91,40	79,78
17	94,50	98,08	94,51	97,42	93,89	99,18	96,40	90,73	86,71
10	104,61	105,57	101,01	102,27	87,43	96,66	86,58	93,06	112,70
15	97,30	94,07	88,54	98,11	97,79	99,59	99,76	99,40	105,91
14	97,91	85,82	82,49	90,45	104,69	95,08	99,69	77,63	62,03
16	87,01	94,69	79,20	68,4					
13	94,94	104,40	119,29	112,9					
19	77,39	100,77	97,03	111,1					
18	91,26	89,22	81,25	100,5					
15	103,04	97,49	98,52	96,8					
21	93,84	87,35	104,26	115,3					
1	65,90	74,23	60,11	47,0					
15	98,08	105,63	90,24	101,0					
1	92,52	87,61	103,27	88,6					
16	104,37	101,85	108,08	121,8					
10	97,88	91,46	99,55	110,4					
2	105,62	102,25	90,17	93,5					
18	90,58	95,49	86,06	92,4					
14	94,99	98,59	100,32	100,9					
1	100,67	102,30	99,83	106,5					
2	110,27	120,53	115,78	118,9					
17	94,97	94,38	101,19	95,0					
7	107,08	94,02	111,29	146,4					
16	83,32	67,79	54,60	66,4					
11	85,48	82,82	90,37	89,6					
1	98,8								
18	105,77								
19	96,85								
17	91,44								
11	97,26								
14	114,29								
2	111,10								
17	105,77								
13	81,73								
17	115,63								
9	83,17								
17	99,52								
9	84,38								
15	94,93								
7	87,33								

Metoda extrakce faktorů: 52Spokojenost

Celé případy byly vymečány u ChD

100 příp. bylo zpracováno (zvoleno)
100 platn. příp. akceptováno

Korelační matice byla spočítána pro 10 prom.

Základní nastavení | Detaily | Popisné statistiky

Metoda extrakce

Hlavní komponenty
 Komunalita = více R2
 Iterované komun. (MINRES)
 Max. věrohodné faktory
 Centroidová metoda
 Metoda hlavní osy

Max. počet faktorů:

Min. vlastní číslo:

Iterované komunalita:
Min. změna v komunalitě:
Maximální počet iterací:

OK | Storno | Možnosti

5. EDA: Otevřeme záložku Detaily a zadáme Hlavní komponenty ve volbě Metody extrakce.

STATISTICA CZ - [PS 1* - Korelace (52Spokojenost)]

Spoubr Upravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Qřno Nápověda

Arial 10 B I U

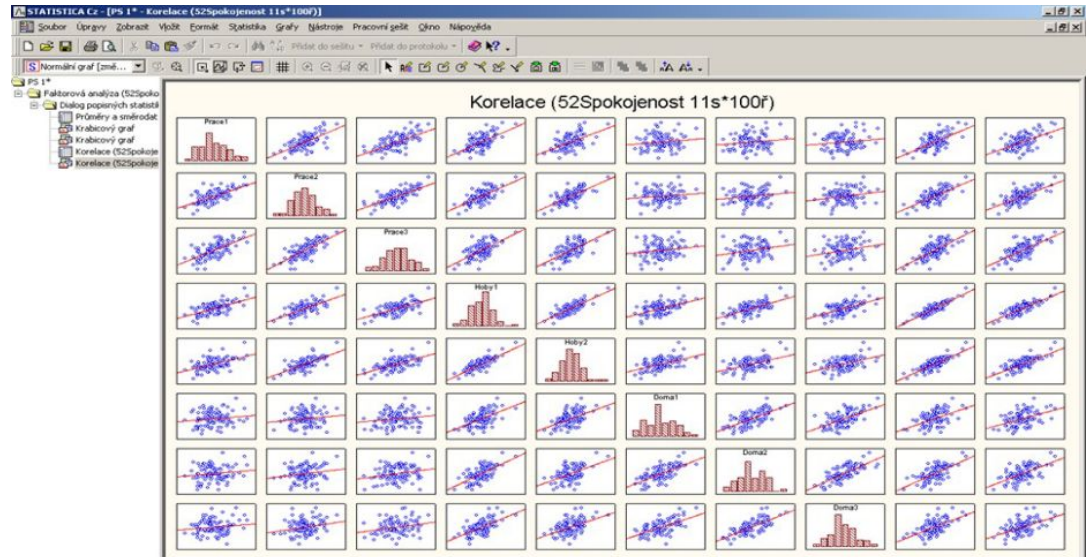
PS 1*

- Faktorová analýza (52Spoko)
- Dialog popisných statistik
- Přehled popisných statistik
- Krabčkový graf
- Krabčkový graf
- Korelace (52Spoko)

Korelace (52Spokojenost)
ChD vymečána případově
N=100

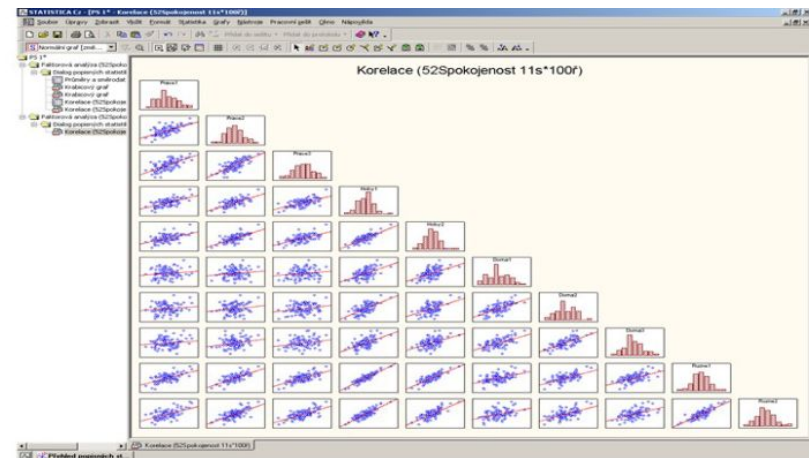
Proměnná	Prace1	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	1,00	0,65	0,65	0,60	0,52	0,14	0,15	0,14	0,61	0,55
Prace2	0,65	1,00	0,73	0,69	0,70	0,14	0,18	0,24	0,71	0,68
Prace3	0,65	0,73	1,00	0,64	0,63	0,16	0,24	0,25	0,70	0,67
Hoby1	0,60	0,69	0,64	1,00	0,80	0,54	0,63	0,58	0,90	0,84
Hoby2	0,52	0,70	0,63	0,80	1,00	0,51	0,50	0,48	0,81	0,76
Doma1	0,14	0,14	0,16	0,54	0,51	1,00	0,66	0,59	0,50	0,42
Doma2	0,15	0,18	0,24	0,63	0,50	0,66	1,00	0,73	0,64	0,59
Doma3	0,14	0,24	0,25	0,58	0,48	0,59	0,73	1,00	0,59	0,52
Ruzne1	0,61	0,71	0,70	0,90	0,81	0,50	0,64	0,59	1,00	0,84
Ruzne2	0,55	0,68	0,67	0,84	0,76	0,42	0,59	0,52	0,84	1,00

14. EDA: Okno Korelace v záložce Detaily okna Přehled popisných statistik otevře Maticový diagram všech znaků. Obsahuje Pearsonovy korelační koeficienty a ukazuje na největší (zde 0.9) a nejmenší (zde 0.14) korelaci mezi znaky. Tím zřetelně odkryvá dosud skrytou strukturu v datech. Pravidlo: Když by byla korelace všech znaků malá a statisticky nevýznamná, nelze faktorovou analýzu vůbec použít, protože v datech není žádná skrytá struktura.



15. EDA: Kliknutím na ikonu diagramu korelace se otevře Maticový diagram. V něm červené přímky rovnoběžné s osou x upozorňují na malou a nevýznamnou korelaci této právě analyzované dvojice znaků.

16. **EDA:** Po kliknutí kdekoli do plochy diagramu se otevře okno *Všeobecné možnosti*, ve kterém v záložce *Rozvržení grafu* vybereme u *Typ matice* volbu zobrazení pouze dolní poloviny diagramu pod diagonálou histogramů, a to *Pol*.



17. **EDA:** Tato podoba maticového diagramu je pak přehlednější. Soustředíme se na vyhledávání červených rovnoběžek s osou x v mraku modrých bodů. Požadovaná nenulová směrnice červené přímky ukazuje zde jednak na rozmístění bodů okolo přímky, a jednak na potřebnou korelaci u této dvojice znaků. Tato statisticky ještě významná korelace je totiž důležitým předpokladem dalšího využití faktorové analýzy.

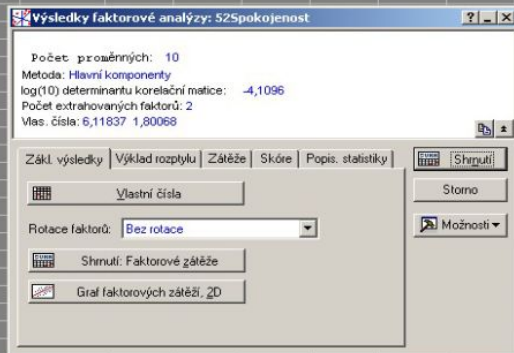
Proměnná	Prace1	Prace2	Prace3	Hoby1	Hoby2	Doma1	Doma2	Doma3	Ruzne1	Ruzne2
Prace1	1,00	0,65	0,65	0,60	0,52	0,14	0,15	0,14	0,61	0,55
Prace2	0,65	1,00	0,73	0,69	0,70	0,14	0,18	0,24	0,71	0,68
Prace3	0,65	0,73	1,00	0,64	0,63	0,16	0,24	0,25	0,70	0,67
Hoby1	0,60	0,69	0,64	1,00	0,80	0,54	0,63	0,58	0,90	0,84
Hoby2	0,52	0,70	0,63	0,80	1,00	0,51	0,50	0,48	0,81	0,76
Doma1	0,14	0,14	0,16	0,54	0,51	1,00	0,66	0,59	0,50	0,42
Doma2	0,15	0,18	0,24	0,63	0,50	0,66	1,00	0,73	0,64	0,59
Doma3	0,14	0,24	0,25	0,58	0,48	0,59	0,73	1,00	0,59	0,52
Ruzne1	0,61	0,71	0,70						1,00	
Ruzne2	0,55	0,68	0,67							1,00

20. **FA:** V záložce *Detaily* bloku *Metoda extrakce faktorů* je možné zadat *Maximální počet faktorů* roven 2 ale také jiné číslo, například původní počet znaků 10.

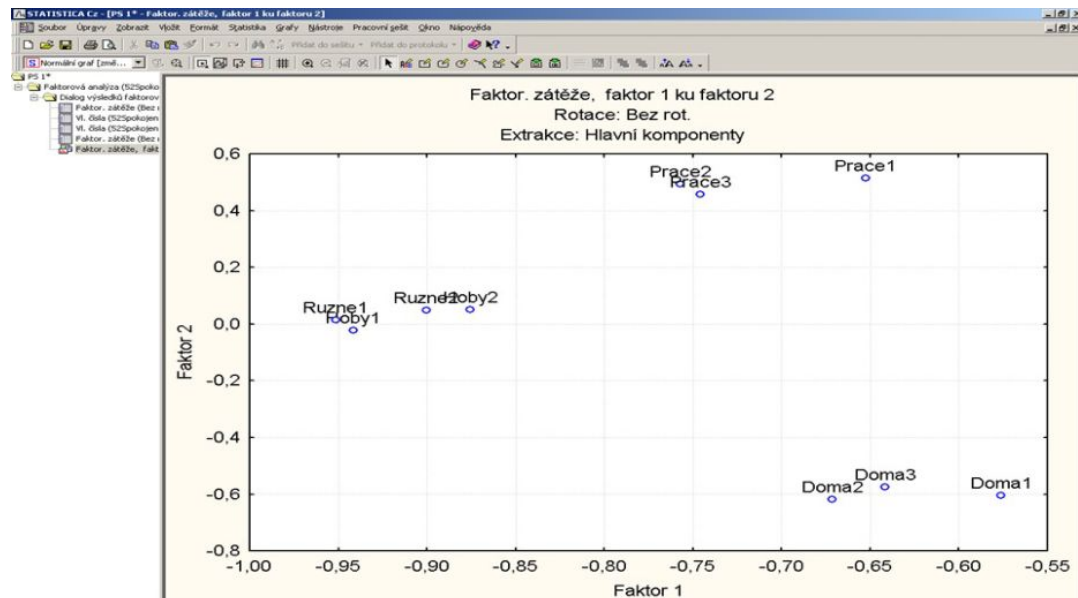
Hodn.	rozptylu	vlast. číslo	%
1	6,118369	6,118369	61,18369
2	1,800682	18,00682	7,919051

21. **FA:** Ukazuje se, že výhodnější je zadat *Maximální počet faktorů* roven 10.

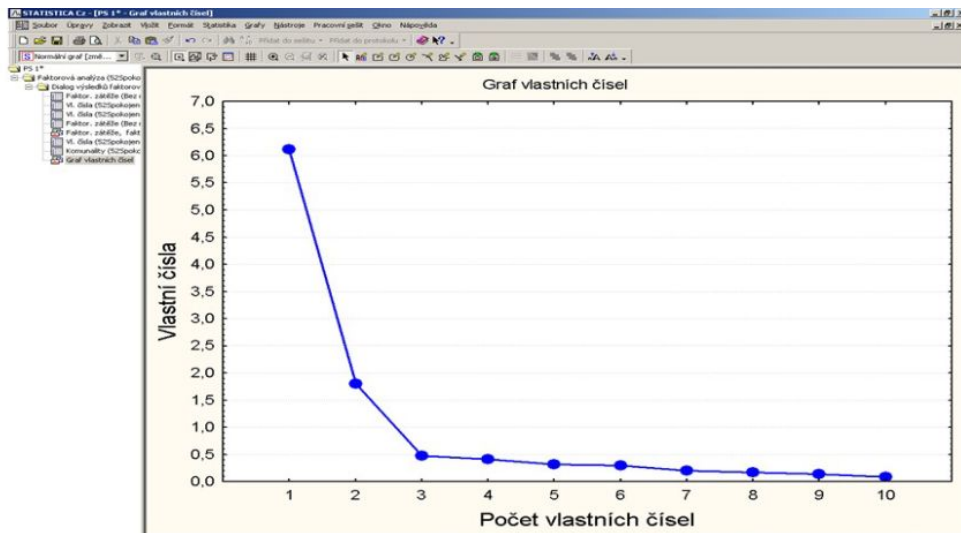
Hodn.	rozptylu	vlast. číslo	%
1	6,118369	61,18369	61,18369
2	1,800682	18,00682	7,919051



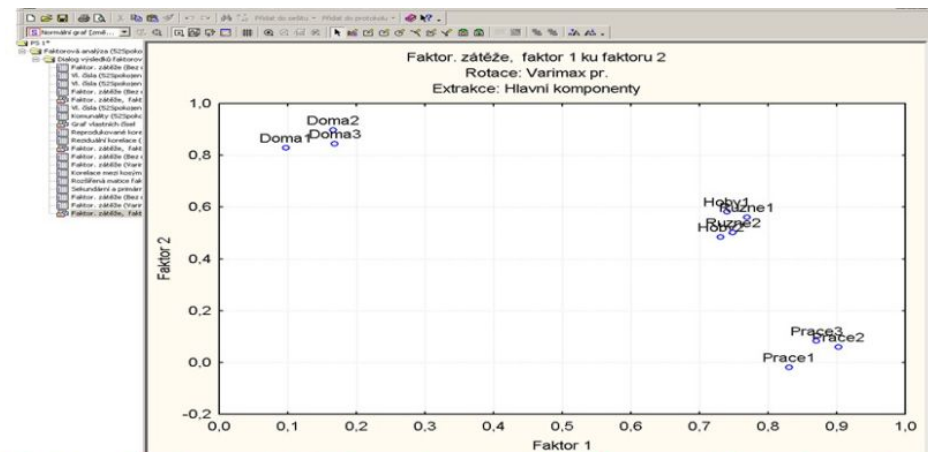
22. FA: Otevře se okno *Výsledky faktorové analýzy*, ve kterém v záložce *Základní výsledky* otevřeme *Vlastní čísla* a v okénku *Rotace faktorů* ponecháme zatím ještě: *Bez rotace*.



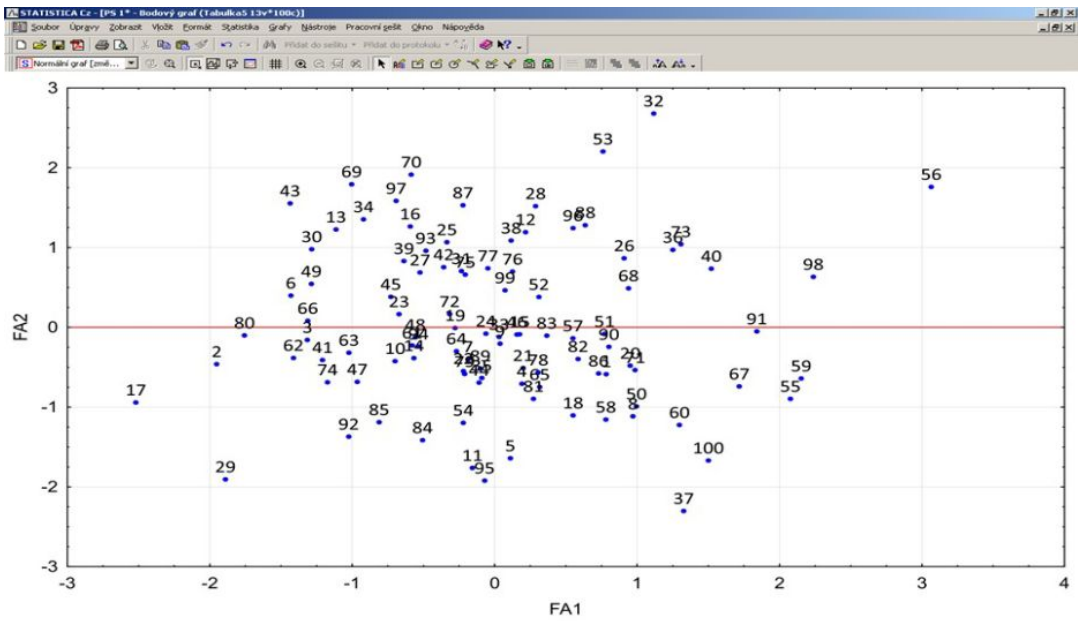
25. FA: *Graf faktorových vah*: graf třídí znaky v rovině. Blízko sebe umístěné znaky jsou si vzájemně podobné a silně spolu korelují. Vzdáleně umístěné nikoliv.



27. FA: *Cattelův indexový diagram vlastních čísel*: podle Kaiserova kritéria 1.0 pouze první dva faktory dosahují větších vlastních čísel než tato limita 1.0. Grafická analýza tohoto grafu však ukazuje na dva závažné zlomy (čili kolena) na křivce pro 2 a 3 faktory.



31. FA: *Faktorové čisté znaky v grafu faktorových vah po Varimax rotaci*: **první faktor** se zde týká znaků spokojenosti v práci, zatímco znaky spokojenosti doma v něm dosahují nejmenších vah. Ostatní znaky pak leží někde uprostřed. **Druhý faktor** vykazuje nejvyšší váhy znaků spokojenosti doma a nejnižší váhy pro znaky spokojenosti v práci. Ostatní znaky spokojenosti leží opět někde uprostřed.



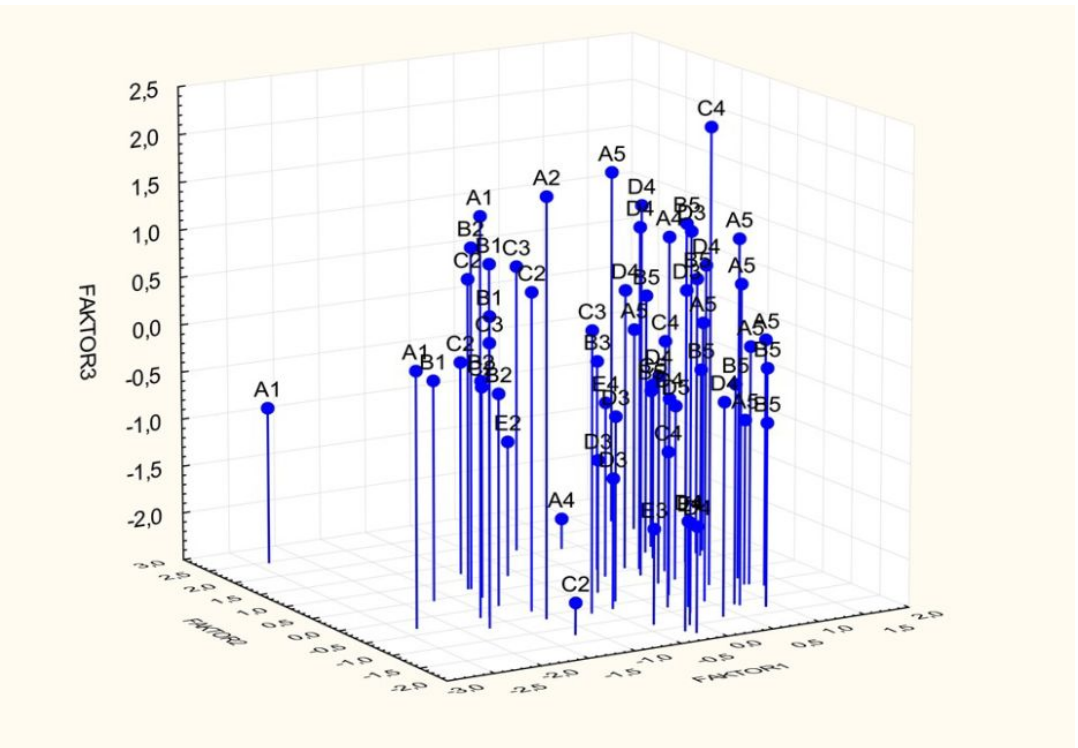
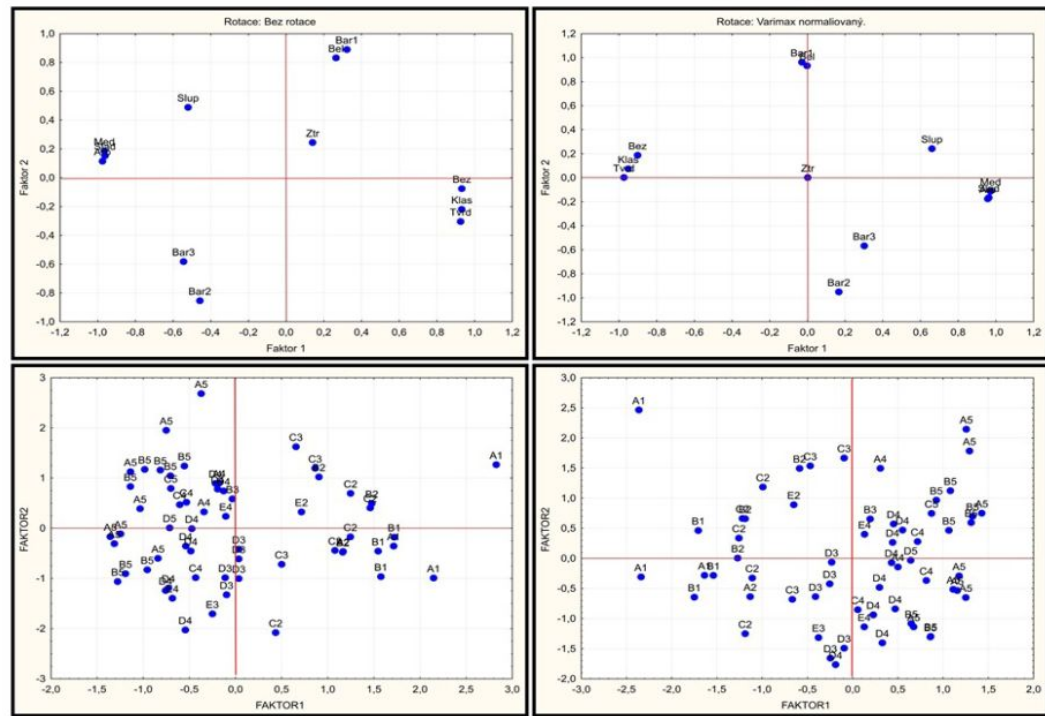
34. FA: Graf faktorového skóre objektů (zde respondentů): graf třídí a klasifikuje respondenty v rovině prvních dvou faktorů.

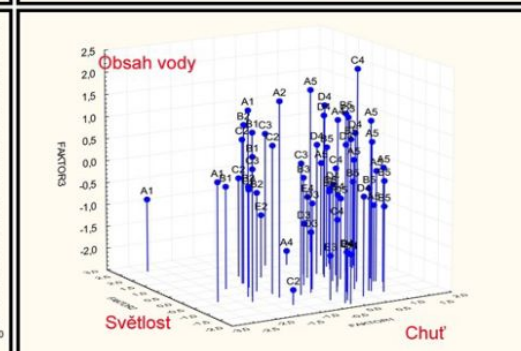
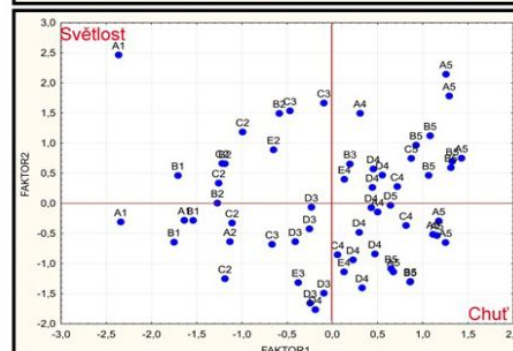
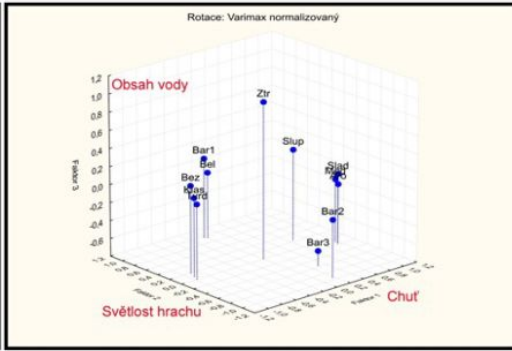
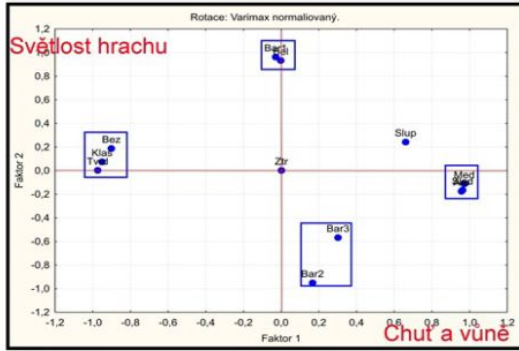
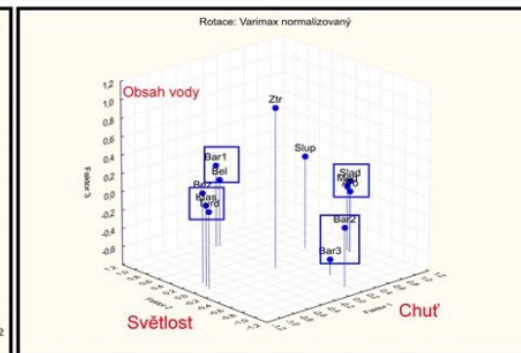
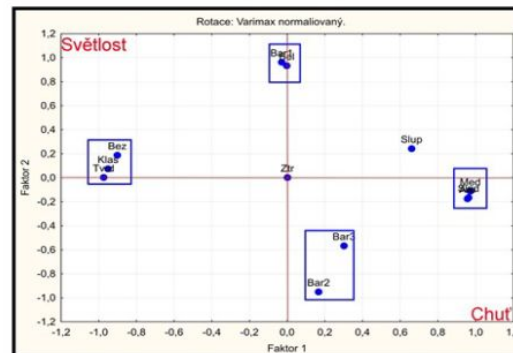
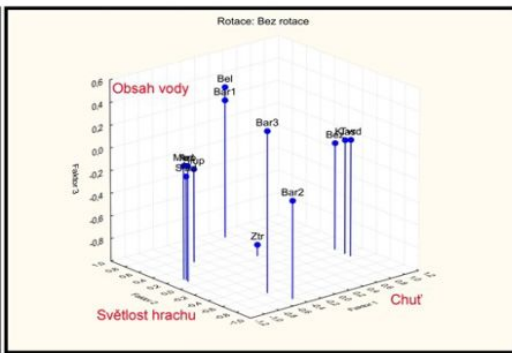
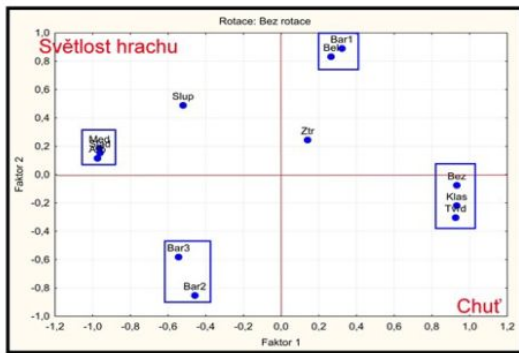
PŘÍKLAD 1.1 Popisné statistiky jednorozměrné analýzy zdrojové matice dat Hrách

Zdrojová matice dat Hrách obsahuje znaky smyslového posouzení charakteristik rozličných odrůd hrachu. Objekty zde představují vzorky pěti různých odrůd hrachu A až E, které byly sklizeny v pěti rozličných obdobích 1 až 5. Výsledná zdrojová matice o 12 znacích převážně smyslových charakteristik obsahuje 60 vzorků hrachu. Posouzení každého objektu hrachu bylo provedeno 10 porotci dvojím odhadem tak, že smyslové charakteristiky byly bodovány ve stupnici od 1 (nejhorší) do 9 (nejlepší). Tak bylo získáno 1200 řádků (objektů) postupem: 60 vzorků × 2 hodnocení × 10 porotců. V praxi se data obvykle průměrují, aby se kompenzovaly rozdíly v subjektivní škále přisnosti jednotlivých porotců. Výsledkem je pro každý z šedesáti objektů průměrná hodnota senzoričského hodnocení. Cílem úlohy je: 1. průměrovat data, 2. vynést původní data do grafu a 3. vypočítat popisné jednorozměrné statistiky.

Data: Zdrojová matice dat $n = 1200$, $m = 12$ byla průměrována a výsledkem byla matice 60×12 . Obsahovala průměrné hodnoty senzoričského hodnocení pro znaky ve sloupcích: Aro je aroma, Slad je sladkost, Med je medovost, Bez je bezchuťovost, Klas je klasovost, Tvrd je tvrdost, Bel je bělost, Bar1 je barva1, Bar2 je barva2, Bar3 je barva3, Slup je slupka, Ztr je ztráta.

Objekt	Aro	Slad	Med	Bez	Klas	Tvrd	Bel	Bar1	Bar2
B5	6.48	6.66	4.56	2.2	2.91	3.47	4.72	5.59	5.73
C4	5.75	6.09	3.81	2.32	4.03	3.77	4.17	5.73	5.75
B2	3.94	4.12	2.44	3.63	5.77	5.39	4.77	6.67	5.11



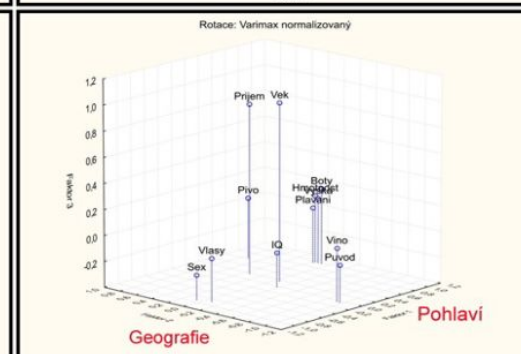
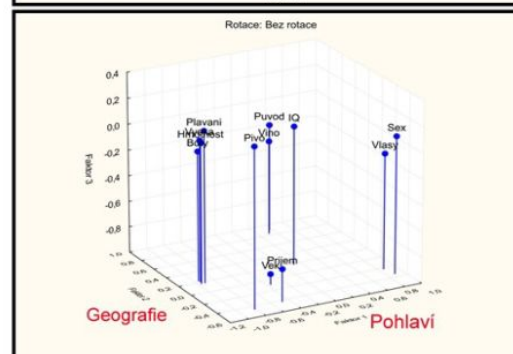
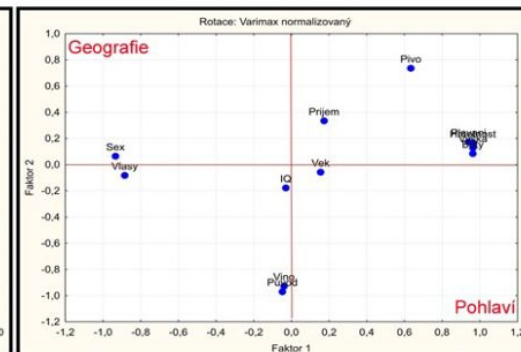
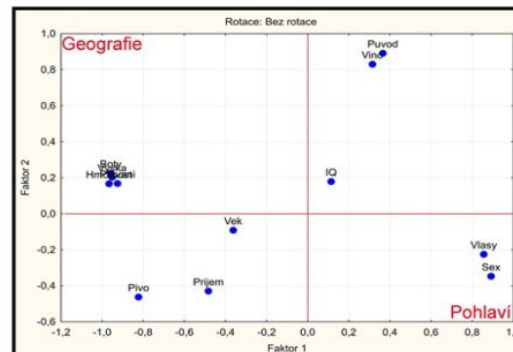


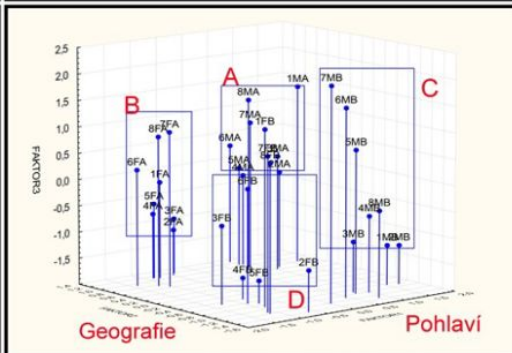
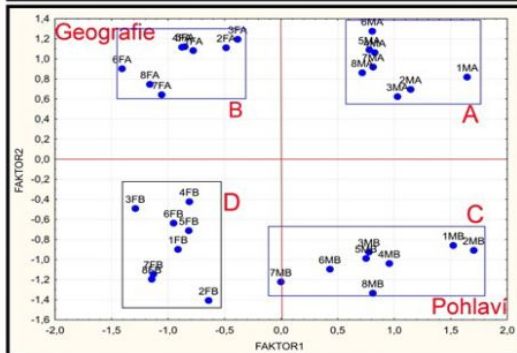
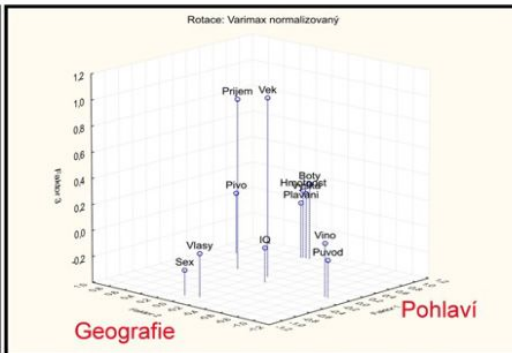
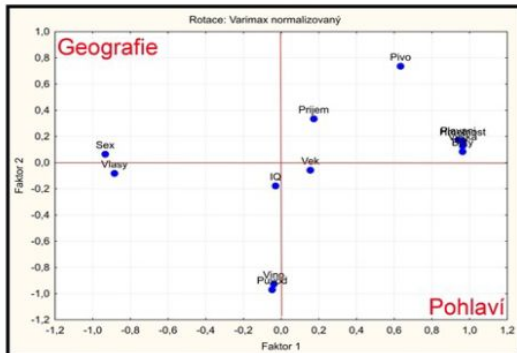
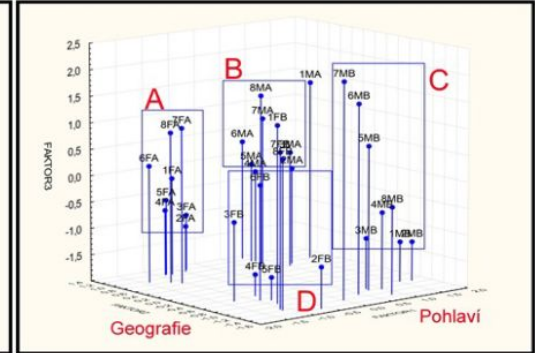
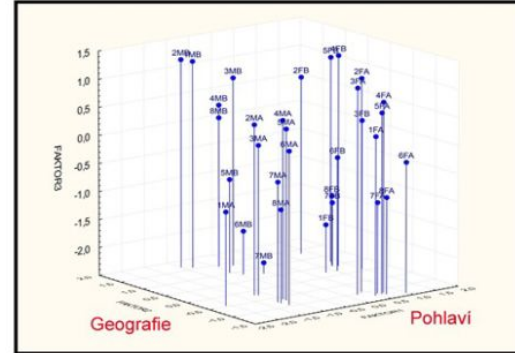
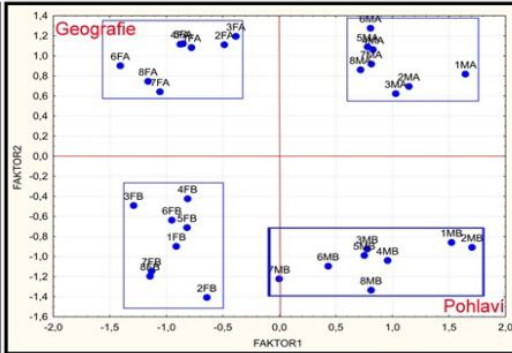
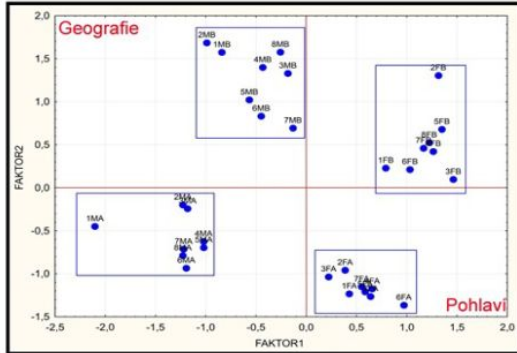
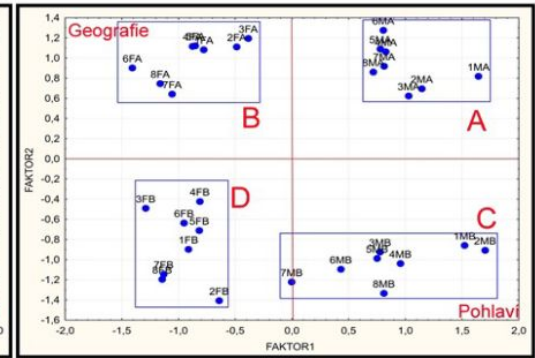
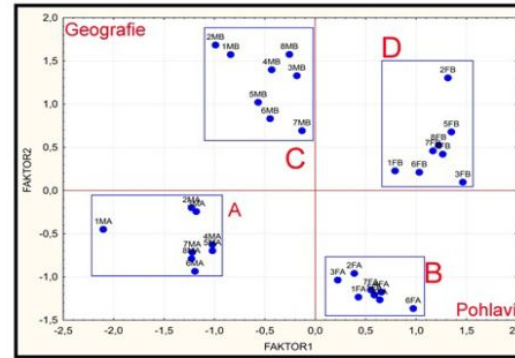
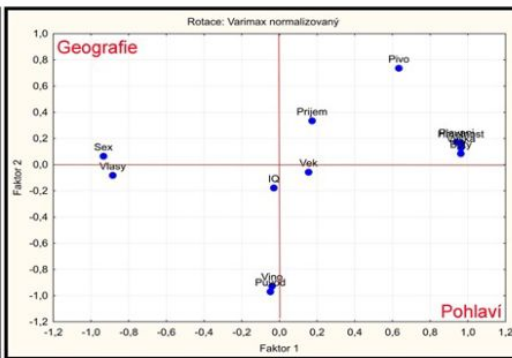
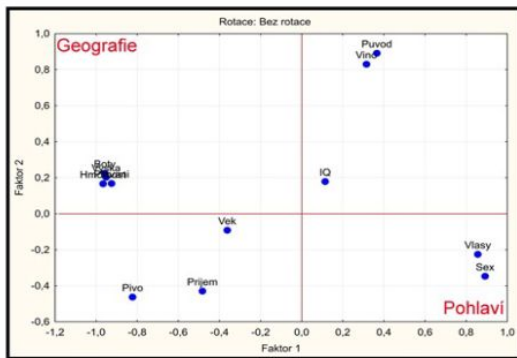
PŘÍKLAD 2.2 Průzkumová analýza zdrojové matice dat demografického souboru Lidé

Vyšetřete grafickými diagnostikami průzkumové analýzy vícerozměrných dat, které ze 12 znaků demografického souboru dat *Lidé* jsou nejvýhodnější k charakterizaci osob a které znaky mají největší míru rozptýlení. Matice obsahuje data pro $n = 32$ osob a $m = 12$ znaků, kde 16 osob bylo vybráno ze Skandinávie (kód A) a 16 osob ze Středomoří (kód B), 16 osob jsou muži (kód M) a 16 osob jsou ženy (kód F).

Data: Znaky obsahují u každé osoby výšku [cm], hmotnost [kg], délku vlasů [krátká: -1, dlouhá: +1], velikost boty [evropský standard], věk [roky], příjem [Euro], spotřeba piva [litry na rok], spotřeba vína [litry na rok], pohlaví [muž: -1, žena: +1], schopnost plavat [naměřený čas na uplavání 500 m], původ [A: -1 Skandinávie, B: +1 Středomoří], inteligentní kvocient IQ [evropský standardizovaný test IQ]. Mezi znaky jsou tři dichotomické, binární proměnné, a to pohlaví, délka vlasů a původ a ostatních 9 znaků nabývá kvantitativních hodnot.

Osoba	Výška	Hmotnost	Vlasy	Boty	Věk	Příjem	Pivo	Víno	Sex
MA	198	92	-1	48	48	45000	420	115	-1
	98	-1	100						
MA	184	84	-1	44	33	33000	350	102	-1
	92	-1	130						
MA	183	83	-1	44	37	34000	320	98	-1
	91	-1	127						
FA	166	47	-1	36	32	28000	270	78	1
	75	-1	112						



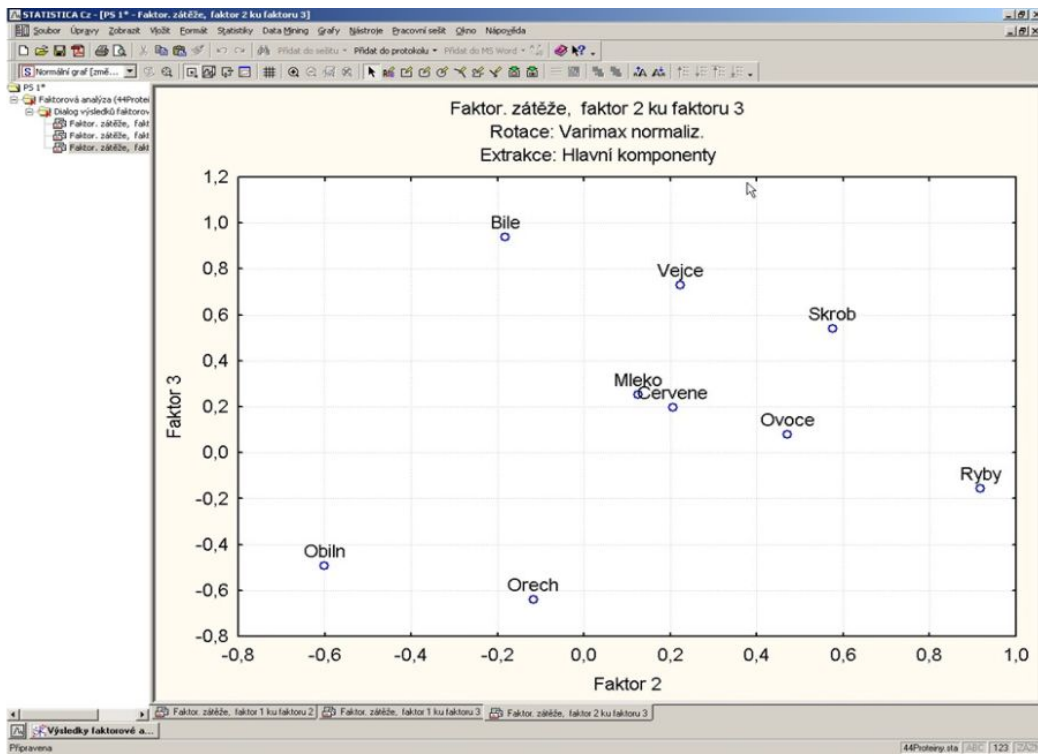
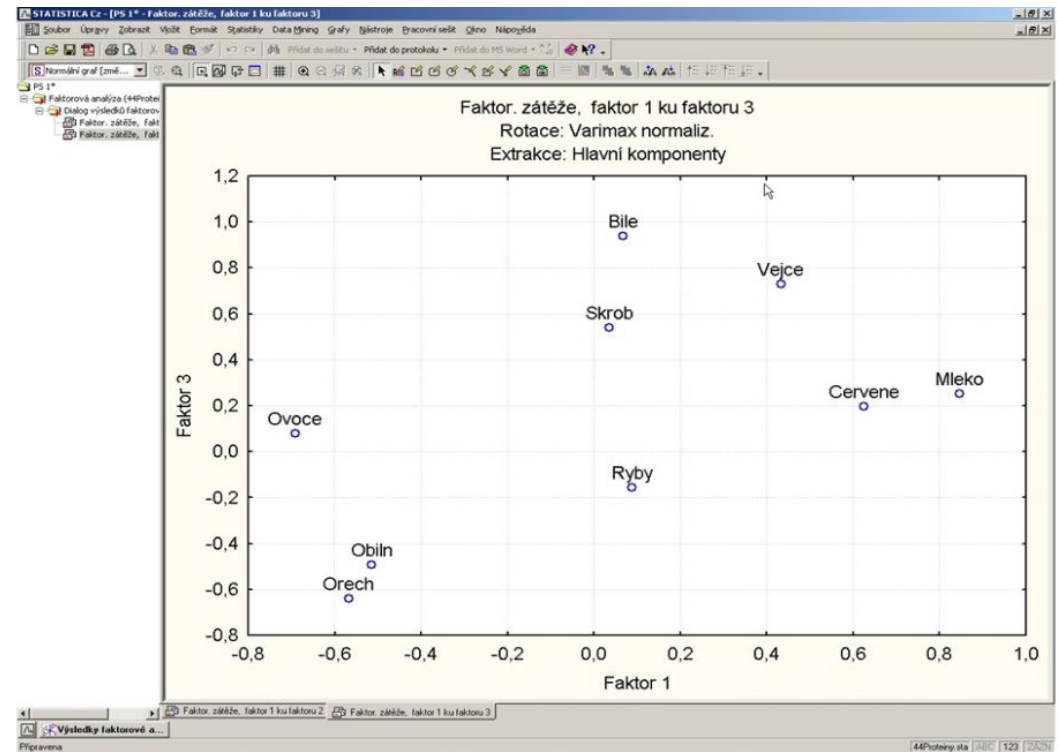
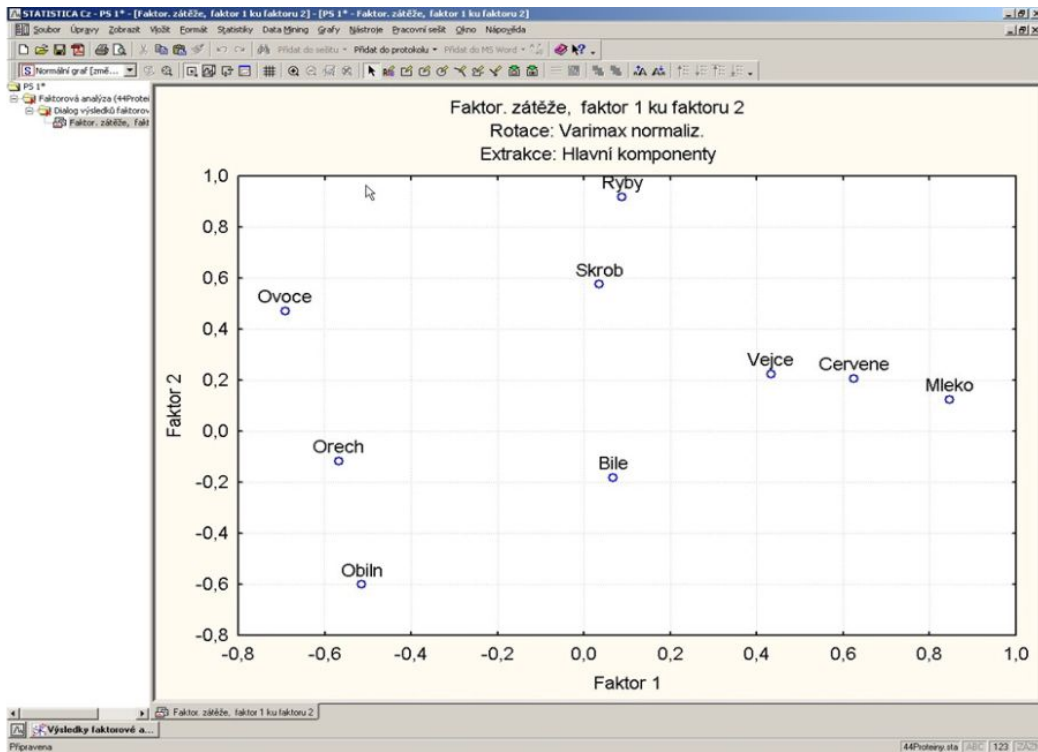


PŘÍKLAD 4.4 Sledování spotřeby proteinů v zemích Evropy

Byla sledována spotřeba proteinů v 25 zemích Evropy formou spotřeby 9 druhů potravin. Greenacre (1984) převzal ve své učebnici data *Proteiny od Webersa z díla Agrarpolitik im Spannungsfeld der Internationalen Ernährungspolitik, Kiel Universität, 1973. Cílem úlohy je především odhalit, zda existuje korelace mezi znaky, tj druhy potravin? Lze odhalit nějaké interakce mezi druhy potravin a zeměmi?*

Data: V datech *Proteiny* jsou uvedeny znaky: Červené značí spotřebu červeného masa, Bílé značí spotřebu bílého masa, Vejce značí spotřebu vajec, Mléko se týká spotřeby mléka, Ryby značí spotřebu ryb, Obiln značí spotřebu obilnin, Škrob značí spotřebu škrobu, Ořechy značí spotřebu ořechů, Ovoce značí spotřebu ovoce a zeleniny.

Země	Červené	Bílé	Vejce	Mléko	Ryby	Obiln	Škrob	Ořechy	Ovoce
Albánie	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Rakousko	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgie	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulharsko	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Československo	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
Dánsko	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4

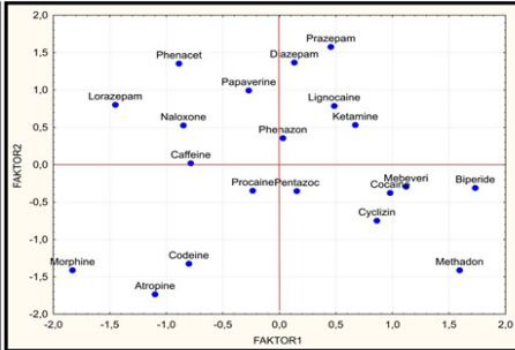
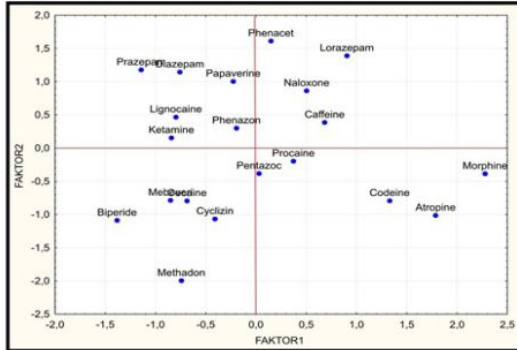
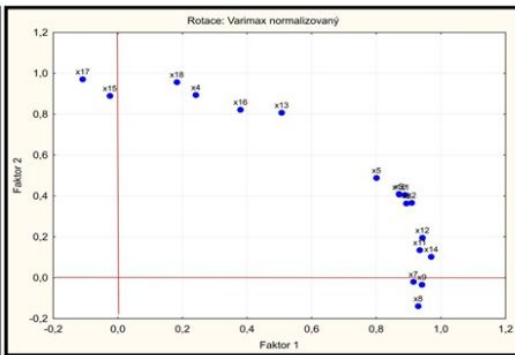
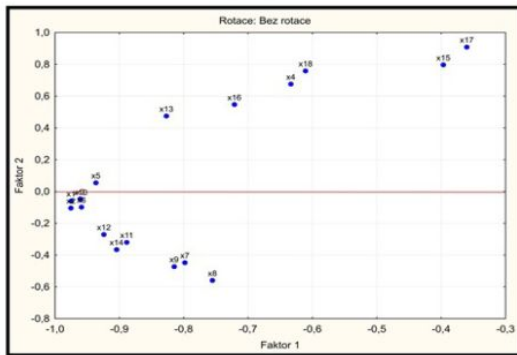


PŘÍKLAD 4.5 Chromatografická analýza farmakologických sloučenin

Byly měřeny hodnoty R_F pro 20 sloučenin s 18 eluenty, úloha B401 [100]. Žádné eluční činidlo však neprovedlo úplné rozdělení. Cílem je nalézt minimální výběr elučních činidel, které by daly dostatek informace pro kvalitativní analýzu.

Data: Datový soubor GIUSEPPE obsahuje $100 \times R_F$ pro 20 sloučenin (v řádcích byla jména zkrácena na maximálně 8 písmen) a ve sloupcích je 18 elučních činidel představujících zde znaky - i vzorek, x_1 směs toluen : aceton : etanol : 30% amoniak = 45 : 45 : 7 : 3, x_2 směs etylacetát : benzen : metanol : 30% amoniak = 60 : 35 : 6.5 : 2.5, x_3 směs benzen : dioxan : etanol : 30% amoniak = 50 : 40 : 7.5 : 2.5, x_4 směs metanol : 30% amoniak = 100 : 1.5, x_5 směs benzen : 2-propanol : metanol : 30% amoniak = 70 : 30 : 20 : 5, x_6 směs etylacetát : metanol : 30% amoniak = 85 : 10 : 5, x_7 směs cyklohexan : toluen : dietylamin = 65 : 25 : 10, x_8 směs cyklohexan : toluen : dietylamin = 75 : 15 : 10, x_9 směs cyklohexan : benzen : metanol : dietylamin = 70 : 20 : 10 : 5, x_{10} směs chloroform : aceton : dietylamin = 50 : 40 : 10, x_{11} směs cyklohexan : chloroform : dietylamin = 50 : 40 : 10, x_{12} směs benzen : etylacetát : dietylamin = 50 : 40 : 10, x_{13} směs xylen : metyletylketon : metanol : dietylamin = 40 : 40 : 6 : 2, x_{14} směs dietyleter : dietylamin = 95 : 5, x_{15} směs etylacetát : chloroform = 50 : 50, x_{16} směs etylacetát : chloroform [A] = 50 : 50, x_{17} směs butanol : metanol = 40 : 60, x_{18} směs butanol : metanol [A] = 40 : 60, kde [A] značí, že byl užit 0.1M metanolát draselný.

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Atropine	20	16	29	23	62	33	4	2	13
	47	25	42	18	12	0	0	5	8
Biperide	91	90	87	68	92	87	73	72	64
	85	81	86	68	94	11	40	40	65
Caffeine	55	42	52	68	77	54	8	5	13
	60	30	51	41	20	13	12	54	57

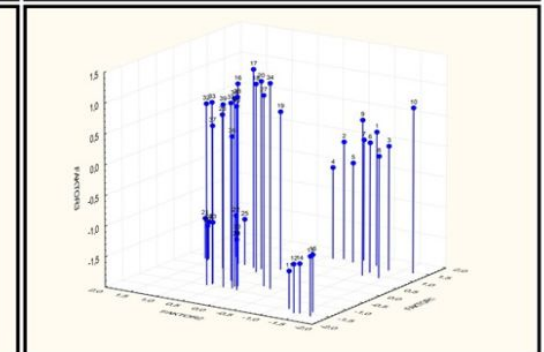
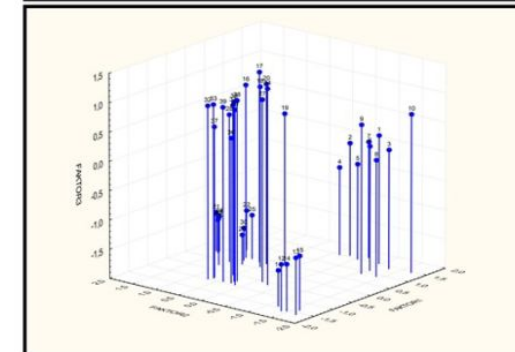
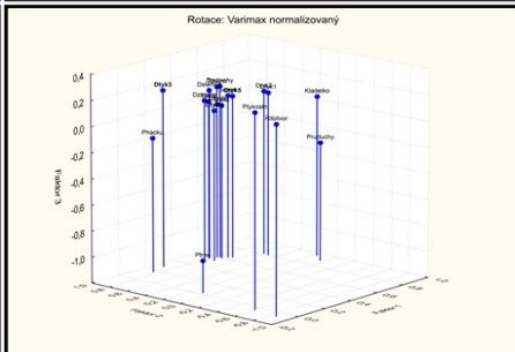
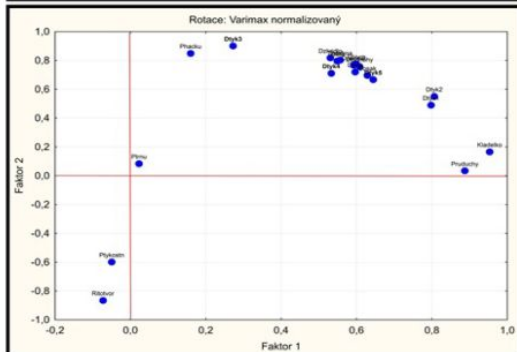
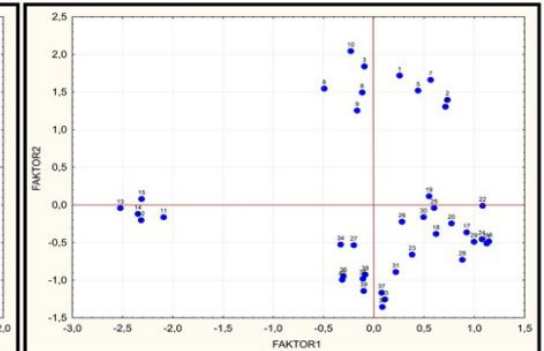
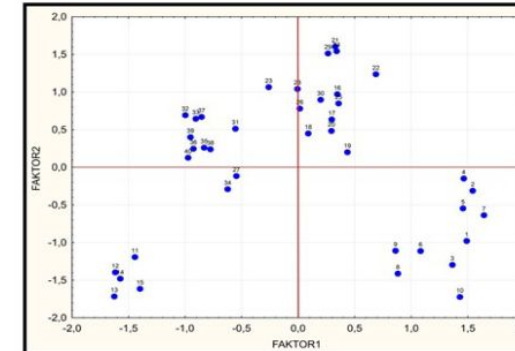
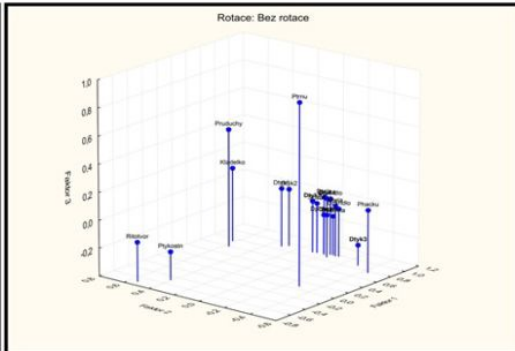
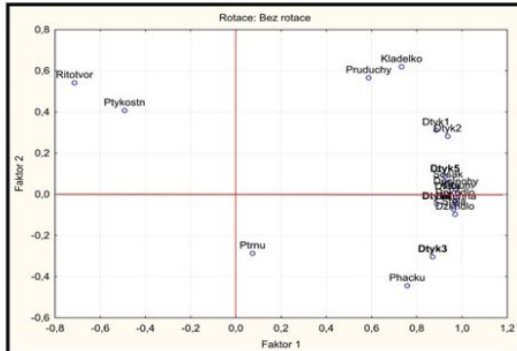


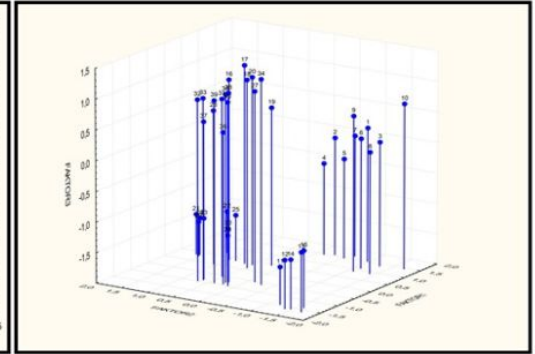
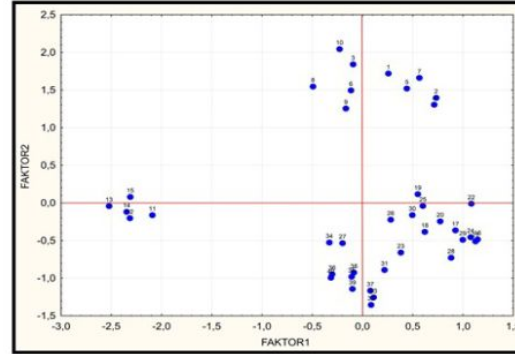
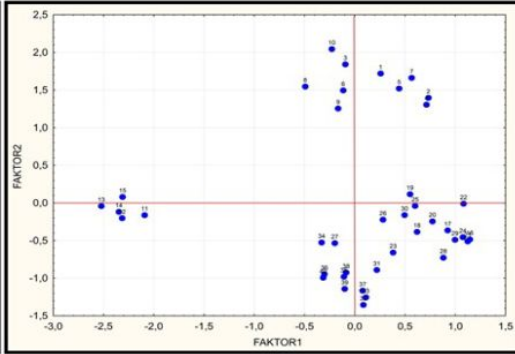
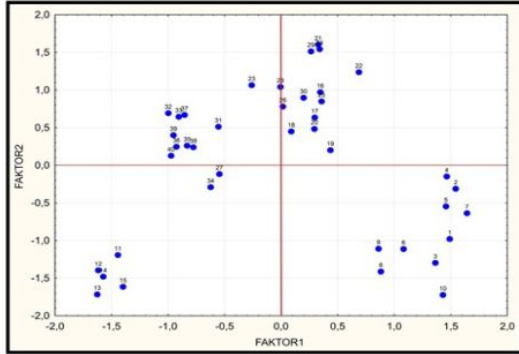
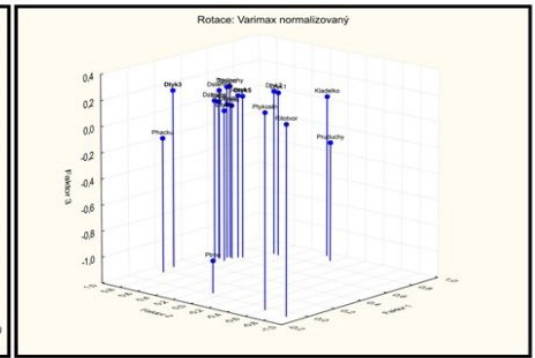
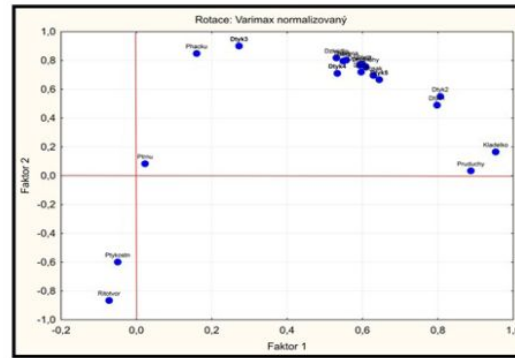
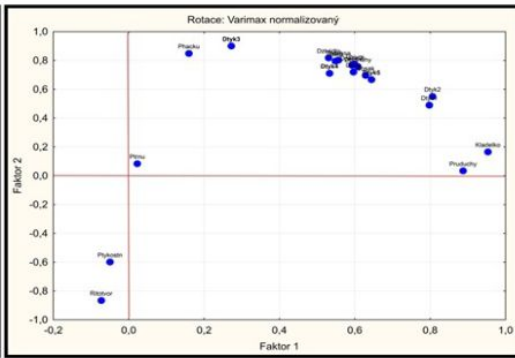
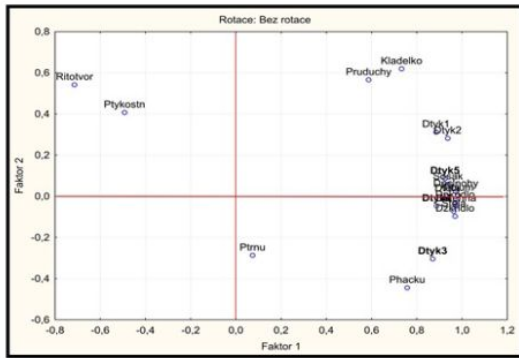
PŘÍKLAD 4.6 Popis a třídění polétavých mšic

Jeffers (1967) studoval 40 jedinců polétavých mšic (*Alate adelges*) za pomoci světelné pasti. U mšic bylo změřeno 19 znaků k rozlišení druhů tohoto hmyzu souboru Mšice: 14 znaků se týká délky nebo šířky, 4 znaky se týkají počtu a 1 znak je binární vyjadřující přítomnost či absenci. Mšice se totiž obtížně rozlišují dle běžných taxonometrických klíčů a obvykle je nutné detailní vyšetření dat, hledání společných rysů a následná klasifikace do větších celků. Užitím metody hlavních komponent PCA je třeba snížit původní počet 19 znaků na menší počet proměnných, který by vystihl co největší množství proměnlivosti mšic. Před užitím PCA je potřeba provést standardizaci dat, protože znaky představují směs délek a počtů.

Data: Charakter znaků – i index jedince, x_1 značí délku těla, x_2 značí šířku těla, x_3 je délka předního křídla, x_4 je délka zadního křídla, x_5 je počet průduchů, x_6 je délka tykadla I, x_7 je délka tykadla II, x_8 je délka tykadla III, x_9 je délka tykadla IV, x_{10} je délka tykadla V, x_{11} je počet tykadlových ostnů, x_{12} je délka posledního článku nohy, x_{13} je délka holeně, tibia, x_{14} je délka stehna, x_{15} je délka sosáku, x_{16} je délka kladélka, x_{17} je počet kladélkových trnů, x_{18} je řitní otvor, x_{19} je počet háčků zadních křídel.

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
	21.2	11.0	7.5	4.8	5.0	2.0	2.0	2.8	2.8
	3.3	3.0	4.4	4.5	3.6	7.0	4.0	8.0	0.0
2	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
	20.2	10.0	7.5	5.0	5.0	2.3	2.1	3.0	3.0
	3.2	5.0	4.2	4.5	3.5	7.6	4.2	8.0	0.0
3	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
	20.2	10.0	7.0	4.6	5.0	1.9	2.1	3.0	2.5
	3.3	1.0	4.2	4.4	3.3	7.0	4.0	6.0	0.0





PŘÍKLAD 9.15 Hledání podobnosti vlastností křupavých lupínek od různých výrobců

Tři americké firmy General Mills (G), Kellogg (K) a Quaker (Q) produkuje křupavé obilné lupínky ke snídani. U řady produktů bylo sledováno 10 znaků a vyšetřována struktura a vzájemné vazby mezi sledovanými znaky jednotlivých produktů, ale i mezi objekty. Které objekty jsou si velice podobné? Kolik shluků objektů odhaluje dendrogram podobnosti objektů optimální shlukovací procedurou? Lze nalézt i dominantní znaky a naopak redundantní znaky, které je možné vypustit?

Data: Datová matice Křupky obsahuje 55 dodavatelů obilných lupínek, u kterých bylo vyšetřováno 10 znaků: Objekt značí index obilných lupínek x_1 , i značí jednoho ze tří výrobců G, K či Q, Cal značí kalorickou hodnotu [cal] x_3 , Bilkov značí obsah bílkovin x_4 , Tuky značí obsah tuků x_5 , Na značí obsah sodných iontů x_6 , Vlaken značí obsah vlákniny x_7 , Uhlovod značí obsah uhlovodíků x_8 , Cukr značí obsah cukru x_9 , K značí obsah draselných iontů x_{10} , Skupina značí zařazení do skupiny x_{11} .

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
	x_{10}	x_{11}							
1	ACCheerios	G	110	2	2	180	1.5	10.5	
...
43	QuakerOatmeal	Q	100	5	2	0	2.7	1	
	1	3							

