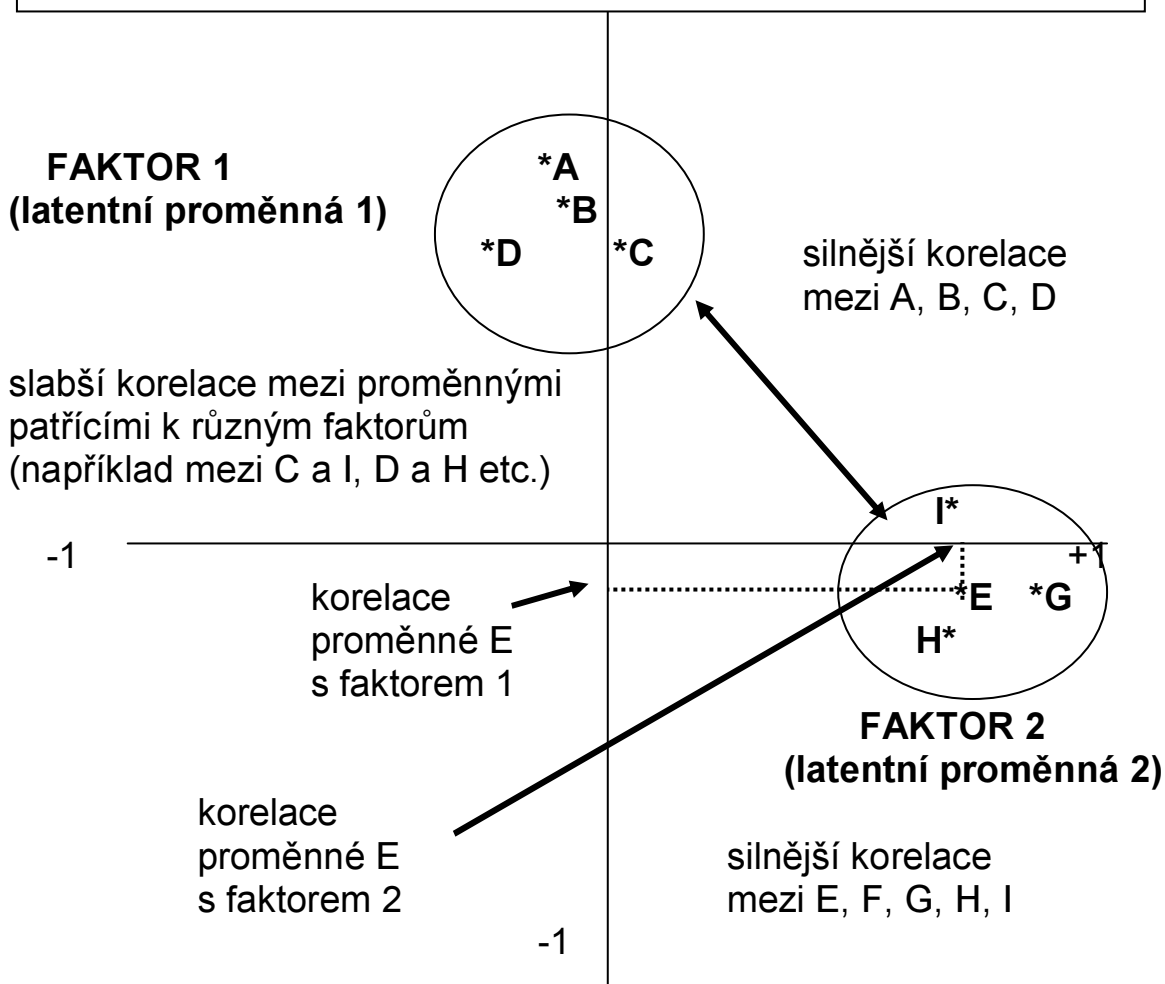


LEKCE12 FAKTOROVÁ ANALÝZA

Představuje způsob REDUKCE DAT:

- Jde o přeměnu souboru vzájemně korelovaných proměnných (matice jejich korelací)
- na menší soubor „nekorelovaných“ faktorů, tento původní soubor reprezentující

Barevně označené položky patří do společného faktoru a lze je zahrnout pod obecnější pojem



Poznámky:

- Faktory jsou statistickými útvary.
- Rozsah stupnic na osách je $\langle -1; 1 \rangle$ neboť jde o korelační koeficienty

DEMONSTRACE VÝSLEDKU FAKTOROVÉ ANALÝZY

Zde je seznam věcí, o nichž si lidé myslí, že jsou nutné pro úspěšné manželství. Prosím, řekněte mi pro každou z nich, zda ji považujete za 1 = velmi důležitou, 2 = spíše důležitou, nebo 3 = nepříliš důležitou pro úspěšné manželství.

Časté debaty o společných zájmech	1	2	3
Dělba domácích prací	1	2	3
Společné trávení co nejvíce času	1	2	3
Věrnost	1	2	3
Shoda názorů na politiku	1	2	3
Společné náboženské přesvědčení	1	2	3
Stejný etnický původ	1	2	3
Stejný sociální původ	1	2	3
Děti	1	2	3
Dobré bydlení	1	2	3
Přiměřený příjem	1	2	3
Ochota diskutovat o problémech	1	2	3
Porozumění a snášenlivost	1	2	3
Vzájemná úcta a uznání	1	2	3
Dobré sexuální soužití	1	2	3
Oddělené bydlení od rodičů partnera	1	2	3

Zde jsou použité škály seřazeny tak, jak se seskupily při FAKTOROVÉ ANALÝZE v jednotlivých faktorech (nejsou tedy zde prezentovány vlastní výstupy z faktorové analýzy). Příklad uvádíme i s vědomím, že pouze třístupňové škály nejsou pro použití faktorové analýzy nejideálnější.

POUŽITÍ FAKTOROVÉ ANALÝZY

- Vytvoření faktorů a jejich věcná interpretace dle významu proměnných ve faktoru seskupených (vytvořit a pojmenovat).
- Identifikace relevantních ukazatelů (vyřazení otázek, které nepracují tak, jak se od nich očekávalo - nepatří do faktoru jak se původně předpokládalo).
- Vytvoření sumačních indexů uvnitř faktorů, nebo použití tzv. faktorových skóre jako nových proměnných a jejich další analýza např. pomocí procedur jako t-test, analýza variancí, regrese, diskriminační analýza.

Příklad:

Otázka: Prosím, řekněte o každé z následujících skutečností, jak je ve Vašem životě důležitá:

	Velmi důležitá	Dost důležitá	Ne příliš důležitá	Vůbec ne důležitá
Q1_1 Práce	1	2	3	4
Q1_2 Rodina	1	2	3	4
Q1_3 Přátelé a známí	1	2	3	4
Q1_4 Volný čas	1	2	3	4
Q1_5 Politika	1	2	3	4
Q1_6 Náboženství	1	2	3	4

Poznámka: Pro demonstraci volíme menší počet položek, ale ve skutečnosti má smysl položky faktorovat **až od určitého většího počtu**.

POSTUP FAKTOROVÉ ANALÝZY

1. Musíme rozhodnout co s případy, které mají "missing value" pro některé z použitých proměnných.

2. Provedeme výpočet korelační nebo kovarianční matice (někdy tato matice může představovat vstupní data).

Korelační matice je základním vstupem dat pro faktorovou analýzu, která je de facto právě analýzou korelační matice.

3. Zjišťujeme jak dobře odpovídá zvolený model faktorové analýzy našim datům a také zda jsou všechny proměnné vhodné (zda některé z nich neoslabují řešení).

Mají-li proměnné velmi slabé korelace s ostatními proměnnými, měly by být v dalším běhu z analýzy vyloučeny (zjišťujeme, zda je model pro soubor o tyto proměnné oproštěný vhodnější).

4. Provedeme faktorovou extrakci (factor extraction) neboli současně identifikaci faktorů a určení příslušnosti jednotlivých položek k faktorům.

Je možné zvolit z několika metod extrakce (doporučuje se začít s metodou hlavních komponentů - principle components).

Provádí se též a odhad faktorových zátěží (factor loadings) neboli korelací mezi faktorem a jeho prvky (položkami k faktoru příslušujícími).

5. Rotace řešení. Má přispět k lepší interpretovatelnosti faktorové zátěže. Rotace faktorovou má za cíl dále zvýšit sílu silných zátěží a snížit sílu nízkých zátěží. Máme možnost volit z několika metod rotace.
6. Výpočet faktorových skóre pro každý případ a jejich přiřazení těmto případům (nové proměnné).

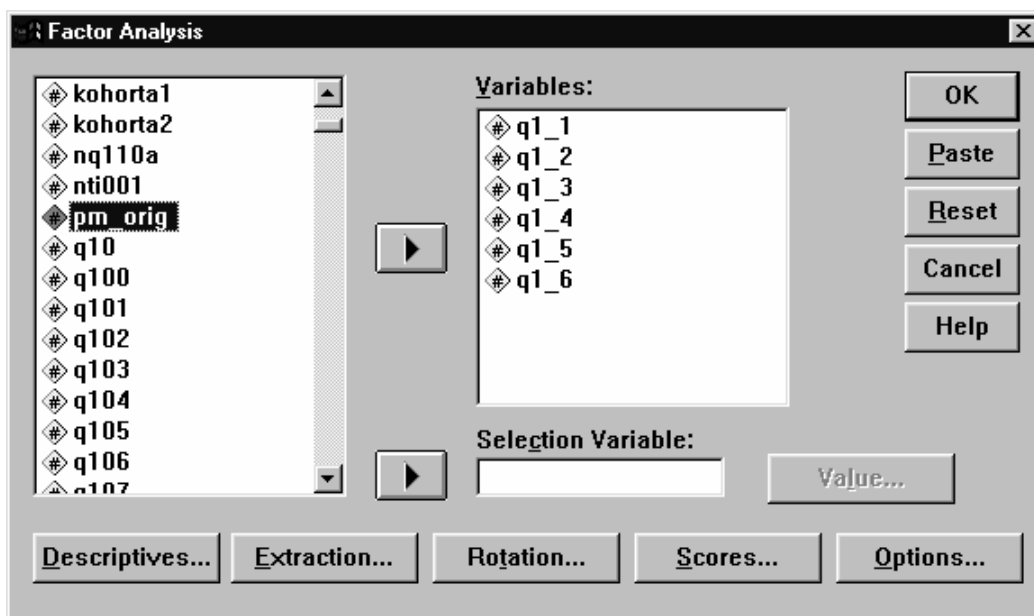
Skóre může sloužit také k identifikaci odlehlých případů, formulaci strategie zacházení s nimi nebo pro další analýzy.

7. Eventuální uložení faktorových skóre pro další výpočty.

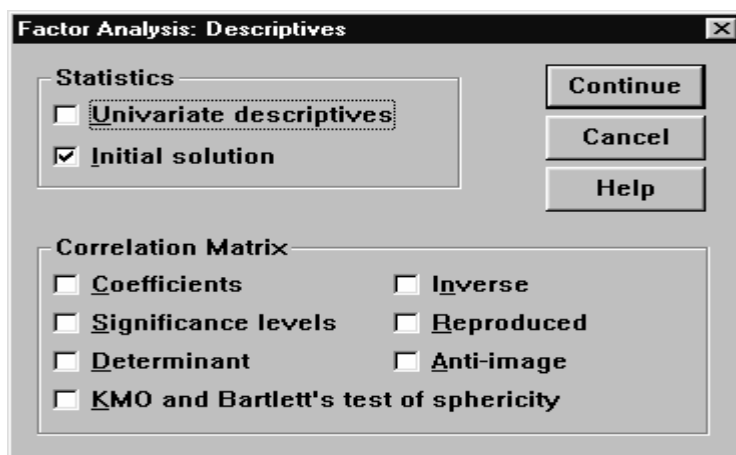
ANALYZE**Data Reduction****Factor****Variables:** *(uvedou se proměnné)*

- **DESCRIPTIVES**
 - STATISTICS**
 - Univariate descriptives
 - Initial Solution
 - CORRELATION MATRIX**
 - Coefficients
 - KMO and Bartlett's test of Sphericity
- **EXTRACTION**
 - METHOD**
 - Principal Components ...
 - ANALYZE**
 - Correlation Matrix ...
 - DISPLAY**
 - Unrotated Factor Solution
 - Scree Plot
- **ROTATION**
 - METHOD**
 - Varimax, ...
 - DISPLAY**
 - Rotated Solution
 - Loading Plots
- **SCORES**
 - SAVE AS VARIABLES**
 - DISPLAY FACTOR SCORE COEFFICIENT MATRIX**
- **OPTIONS**
 - MISSING VALUE**
 - COEFFICIENT DISPLAY FORMAT**
 - Sorted by Size
 - Suppress absolute values less than: 0.20

Zadání položek, které budou faktorovány:



Zadání popisných statistik a měř adekvátnosti použití faktorové analýzy



JE FAKTOROVÁ ANALÝZA VHODNÝM MODELEM PRO DATA? (míry adekvátnosti použití faktorové analýzy)

- Bartlettův test sféricity (Bartlett's test of sphericity).
- KMO: Kaiser-Mayer-Olkinova míra adekvátnosti výběru (Kaiser-Mayer-Olkin Measure of Sampling Adequacy)
- Matice negativních hodnot parciálních koeficientů (anti-image correlation matrix)

BARTLETTŮV TEST SFÉRICITY (Bartlett's Test of Sphericity).

Testování nulové hypotézy, že *korelační matice je maticí jednotkovou (identity matrix) neboli, že v korelační matici jsou na diagonále hodnoty 1 a v ostatních polích hodnoty 0 neboli, že vztahy mezi kombinacemi všech dvojic proměnných neexistují.*

KMO - Kaiser-Mayer-Olkinova míra adekvátnosti výběru (Kaiser-Mayer-Olkin Measure of Sampling Adequacy)

Index porovnávající velikosti pozorovaných korelačních koeficientů a velikost koeficientů parciální korelace.

Malé hodnoty KMO indikují, nevhodnost použití faktorové analýzy (korelace mezi páry proměnných nemohou být vysvětleny jinými proměnnými).

Velikost KMO a použitelnost FA:

0,90 skvělá
0,80 velmi dobrá
0,70 střední
0,60 slabší
0,50 bídná
pod 0,5 nevyhovující

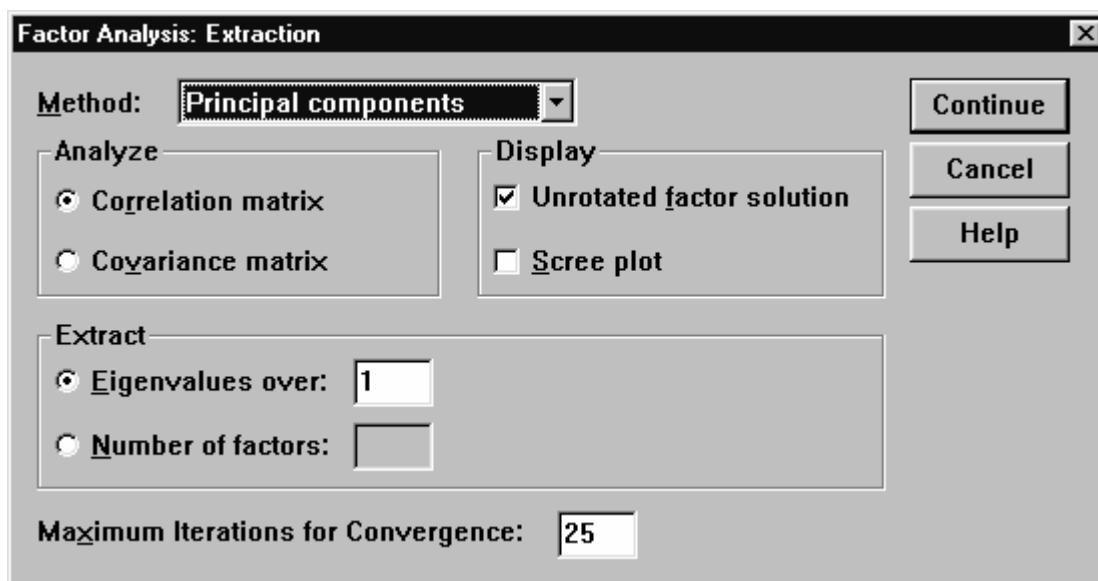
Příklad:

KMO and Bartlett's Test

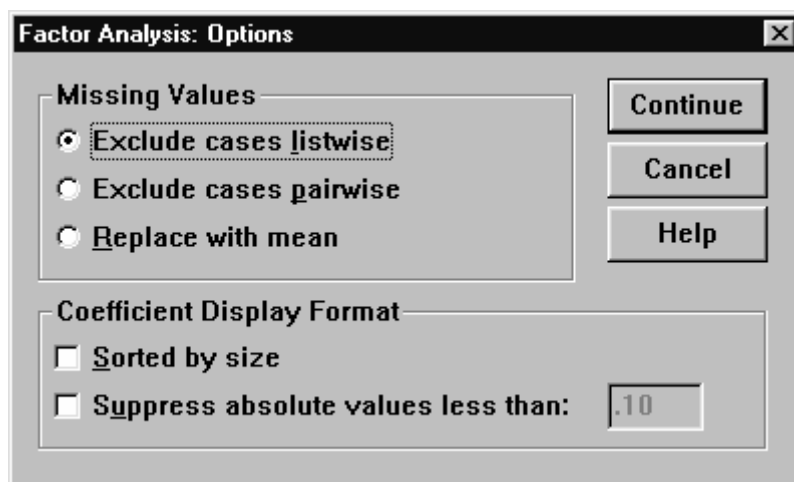
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.565
Bartlett's Test of Sphericity	Approx. Chi-Square	455.950
	df	15
	Sig.	.000

Jak je vidět, použití faktorové analýzy není ideální (nepředstavuje vhodný model pro data).

Zadání extrakce faktorů a metody této extrakce (zde Principal components):



Rozhodnutí o podobě výstupu a o způsobu zacházení s missing values



Označte při zadávání „sorted by size“, dostanete nejčitelnější výstup!

Zacházení s missing values:

- Listwise: Z analýzy jsou vyloučeny všechny případy, jež mají v některé z faktorovaných položek „missing value“.
- Pairwise: Z analýzy jsou vyloučeny případy s „missing values“ jen pro ty konkrétní případy (položky), v nichž „missing value“ existuje.

TOTAL VARIANCE EXPLAINED

Tabulka obsahuje ve sloupci "Eigenvalue" celkovou variance vysvětlenou každým faktorem. EIGENVALUES (také „characteristic roots“ nebo „latent roots“) jsou variance komponent. Jejich velikosti popisuje rozptyl nebo tvar shluku bodů v multivariačním prostoru, který má osu pro každou proměnnou, zahrnutou do analýzy.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.550	25.827	25.827	1.550	25.827	25.827	1.312	21.864	21.864
2	1.175	19.586	45.414	1.175	19.586	45.414	1.272	21.206	43.070
3	1.011	16.842	62.256	1.011	16.842	62.256	1.151	19.186	62.256
4	.855	14.244	76.499						
5	.719	11.975	88.474						
6	.692	11.526	100.000						

Extraction Method: Principal Component Analysis.

variance vysvětlitelné jednotlivými faktory

variance vysvětlená všemi 3 faktory

2. faktor:
 $(1,175/6) \cdot 100 = 19,59$

1. faktor:
 $(1,550/6) \cdot 100 = 25,83$

VŠIMNĚTE SI, ŽE V TOMTO PŘÍPADĚ JDE O DVĚ TABULKY:

FAKTOROVÉ ZÁTĚŽE

Při extrakci faktorů jsou pro jednotlivé položky počítány faktorové zátěže.

A_{ji} čili koeficienty zvané „THE FACTOR LOADINGS“ neboli FAKTOROVÉ ZÁTĚŽE indikují jaká váha je přisouzena každému faktoru. Představují analogii standardizovaných regresních koeficientů v mnohonásobné regresi s:

- původní proměnnou jako závisle proměnnou a
- faktory jako nezávislymi proměnnými.

Faktory s vysokými loadings (v absolutní hodnotě) pro konkrétní proměnné jsou s těmito proměnnými úzce asociovány.

INDIVIDUÁLNÍ FAKTOROVÉ ZÁTĚŽE (FACTOR LOADINGS) A_{ji} PRO KAŽDOU PROMĚNNOU BY MĚLY BÝT BUĎ VELMI VYSOKÉ NEBO VELMI NÍZKÉ.

Pak je proměnná asociována s minimálním počtem faktorů a je možné vysvětlit pozorované korelace s minimálním počtem faktorů.

Prvotní **extrakce faktorů** (initial factor extraction) obvykle neposkytuje příliš interpretovatelné faktory. Provádíme tedy rotaci faktorů.

Zadání rotace faktorů:

The screenshot shows the 'Factor Analysis: Rotation' dialog box. Under the 'Method' section, the 'None' radio button is selected. Other options include 'Quartimax', 'Varimax', 'Equamax', 'Direct Oblimin', and 'Promax'. The 'Delta' input field contains the value '0' and the 'Kappa' input field contains the value '4'. In the 'Display' section, the 'Rotated solution' checkbox is checked, while the 'Loading plot(s)' checkbox is unchecked. At the bottom, the 'Maximum Iterations for Convergence' input field contains the value '25'. On the right side of the dialog, there are three buttons: 'Continue', 'Cancel', and 'Help'.

Všimněte si, že zde máme možnost zadat rotované řešení (důležité, neboť nerotované řešení není optimální) a způsob rotace, ale také grafické vyjádření výsledku.

ROTACE

Představuje snahu zvýšit hodnotu vyšších faktorových zátěží proměnných a snížit hodnotu menších.

- ORTHOGONÁLNÍ ROTACE. Osy stále svírají pravý úhel.
 - VARIMAX. Je pokusem minimalizovat počet proměnných, které mají vysokou faktorovou zátěž s faktorem.
 - QUARTIMAX. Klade důraz na jednoduchou interpretaci proměnných, minimalizuje počet faktorů potřebných k vysvětlení proměnné.
 - EQUAMAX. je kombinací obou předchozích.
- OBLIQUE ROTACE. Osy nemusí v konečném řešení svírat pravý úhel.
 - OBLIMIN. Parametr delta (δ) kontroluje rozsah šikmosti. Když je roven 0, dosáhlo se nejvyšší šikmosti. Pro záporné hodnoty delta se faktory stávají méně šikmé jak se delta stává zápornější.

ROTATED COMPONENT MATRIX

Rotated Component Matrix ^a

	Component		
	1	2	3
Q1_4 Volný čas	.816	-.047	.029
Q1_3 Přátelé a známí	.777	.104	.046
Q1_1 Práce	.031	.773	.103
Q1_2 Rodina	.019	.757	.013
Q1_6 Náboženství	-.074	-.089	.869
Q1_5 Politika	.186	.285	.618

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

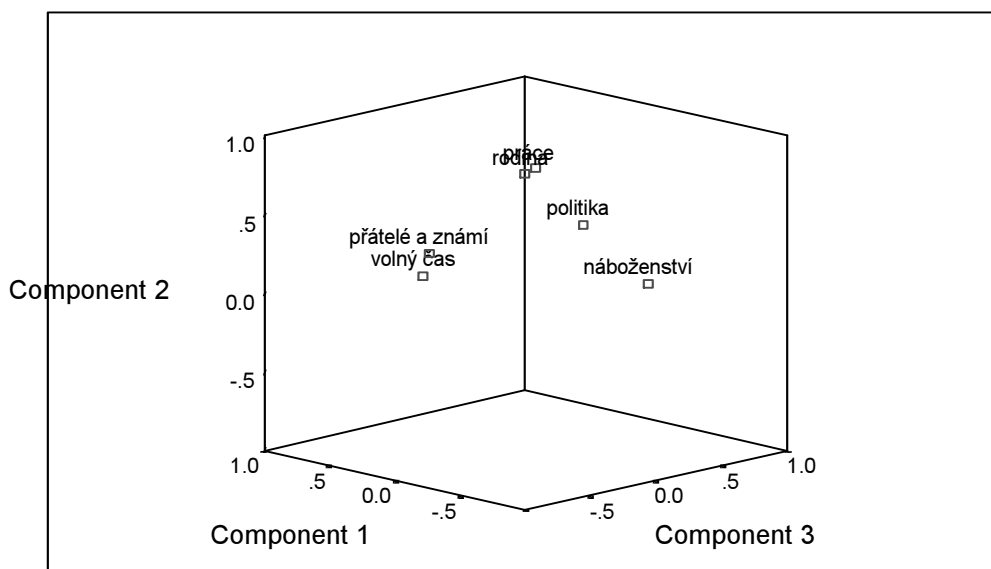
a. Rotation converged in 4 iterations.

Všimněte si, jak se Vám, díky použití „sorted by size“, položky přehledně seřadily (konfrontujte výstup, který byste dostali bez použití tohoto příkazu).

GRAFICKÉ VYJÁDŘENÍ

Vytvoření grafu proměnných, který používá faktorové zátěže (faktorová sycení - factor loadings) jako souřadnice. Zadáte ho v menu „rotation“ (loading plot).

Component Plot in Rotated Space

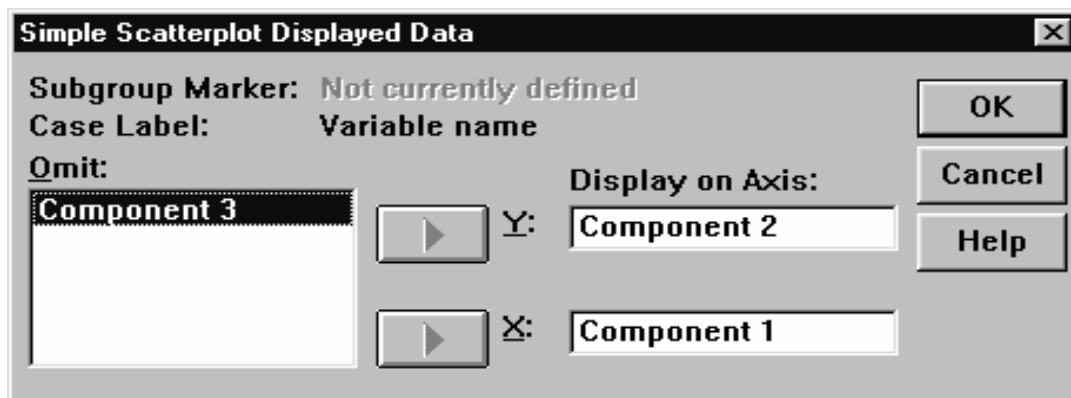


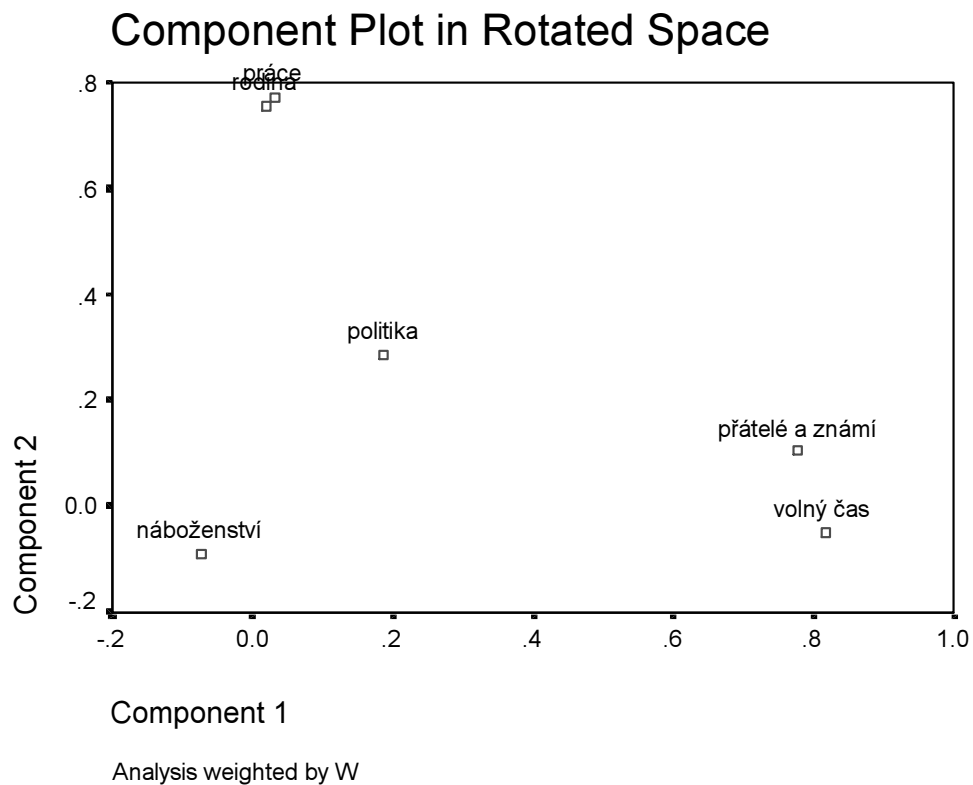
Analysis weighted by W

Trojrozměrný graf lze převést na několik dvourozměrných grafů. Klikneme v Output 2x na graf a editujeme ho.

Edit → Gallery → Scatter → Simple →

(replace) → (omitt component ...):





LEKCE12
SHLUKOVÁ ANALÝZA