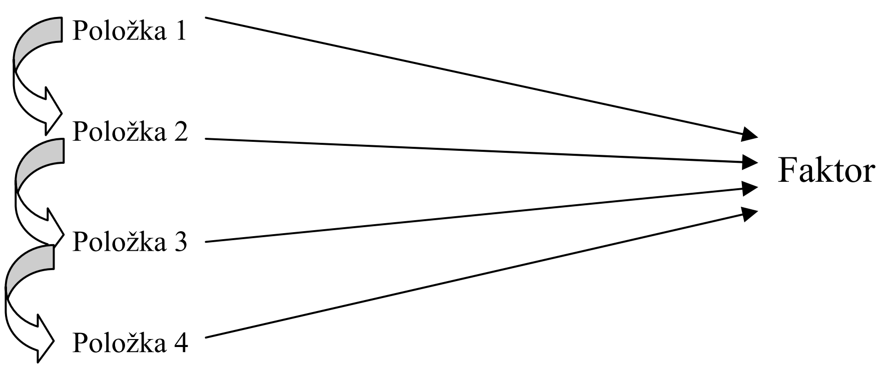
Lekce 11
EXPLORAČNÍ FAKTOROVÁ ANALÝZA

Faktorová analýza (FA) ve své **explorační** verzi (viz níže) je technika, která se odlišuje od dosud probraných postupů analýzy. Není primárně určena k meritorní analýze, to je neslouží k testování hypotéz ani k měření souvislostí mezi nezávisle a závisle proměnnou. Jejím cílem je především redukce dat – umí to, že z mnoha položek vybere ty, které statisticky „patří k sobě“, vybere ty, z nichž je možné vytvořit novou proměnnou.

Techniky faktorové analýzy je možné rozdělit do dvou skupin: (1) faktorová analýza *explorační* (*Exploratory factor analysis*), která je určena především pro konstrukci škál – používá se tudíž v počátečních fázích výzkumu; (2) faktorová analýza *konfirmační* (*Confirmatory factor analysis*), která se již používá k meritorní analýze – k ověření specificky formulovaných hypotéz. Tato procedura je poměrně složitá a nebudeme zde o ní pojednávat[1].

Explorační faktorová analýza je technika, s jejíž pomocí se snažíme nahradit vztahy mezi sadou vzájemně spjatých proměnných malým počtem ne přímo pozorovatelných znaků, faktorů (viz obr. 11.1). Tím je dána také její primární funkce – redukce dat, redukce proměnných. Poté, kdy faktor objevíme a pojmenujeme, můžeme z něj vytvořit novou proměnnou, kterou používáme v další analýze namísto původních položek. Komentovaný příklad nám nejlépe ukáže její smysl.

Obr. 11.1: Model vztahu mezi položkami a faktorem



Výzkum EVS 1999 zjišťoval, jaké jsou postoje respondenta k různým skutečnostem v oblasti ekonomické a sociálně politické. Baterie položek měla následující podobu (viz níže), měřicí stupnice měla deset stupňů, takže jednotlivé položky je možné považovat za kvazi intervalové proměnné, které jsou vhodné pro FA:

[1] Zájemce je možné odkázat na texty o strukturním modelování, v češtině Urbánek (2000), MacDonald (1991), které konfirmační faktorovou analýzu obsahují.

Teď bych byl(a) rád(a), kdybyste mi mohl(a) říci své názory na různá témata. Kam byste svůj názor umístil(a) na takovéto škále?

1	2	3	4	5	6	7	8	9	10

A) Jednotlivci by měli převzít více odpovědnosti, aby se o sebe dokázali postarat	Stát by měl převzít více odpovědnosti a zajistit, aby bylo o každého postaráno
B) Nezaměstnaní, by měli mít povinnost přijmout jakoukoli práci, která je k dispozici, nebo ztratit podporu v nezaměstnanosti	Nezaměstnaní by měli mít právo odmítnout zaměstnání, které nechtějí vykonávat
C) Konkurence je prospěšná. Podněcuje k usilovné práci a k tvorbě nových myšlenek	Konkurence je škodlivá. Vyvolává v lidech to nejhorší
D) Stát by měl dát firmám větší svobodu.	Stát by měl na firmy účinněji dohlížet.
E) Příjmy by měly být vyrovnanější	Mělo by se více podnítit úsilí jednotlivce
F) Soukromé vlastnictví obchodu a průmyslu by mělo vzrůst	Státní vlastnictví obchodu by se mělo v co nejširší míře zachovat
G) Každý člověk by měl být zodpovědný za své vlastní důchodové zabezpečení	Stát by měl být zodpovědný za důchodové zabezpečení všech občanů
H) Každý člověk by měl být zodpovědný za zajištění svého vlastního bydlení	Stát by měl být zodpovědný za zajištění bydlení pro všechny občany

Je možné se domnívat, že tyto položky jsou zřejmě určitým způsobem vzájemně spjaty a že by mohly v sobě skrývat latentní proměnnou, faktor paternalistického nebo liberálního postoje.

Cílem úlohy je tedy zjistit, zdali je možné tyto položky redukovat do několika málo faktorů (jednoho až tří) a z nich pak vytvořit nové proměnné. Použijeme k tomu faktorovou analýzy, konkrétně metodu hlavních komponent (*Principal components analysis*).

Provedení faktorové analýzy se děje ve třech krocích.

(1) Musíme rozhodnout, zdali jsou naše data vhodná pro faktorovou analýzu;

(2) musíme se rozhodnout, s jakým počtem faktorů budeme pracovat;

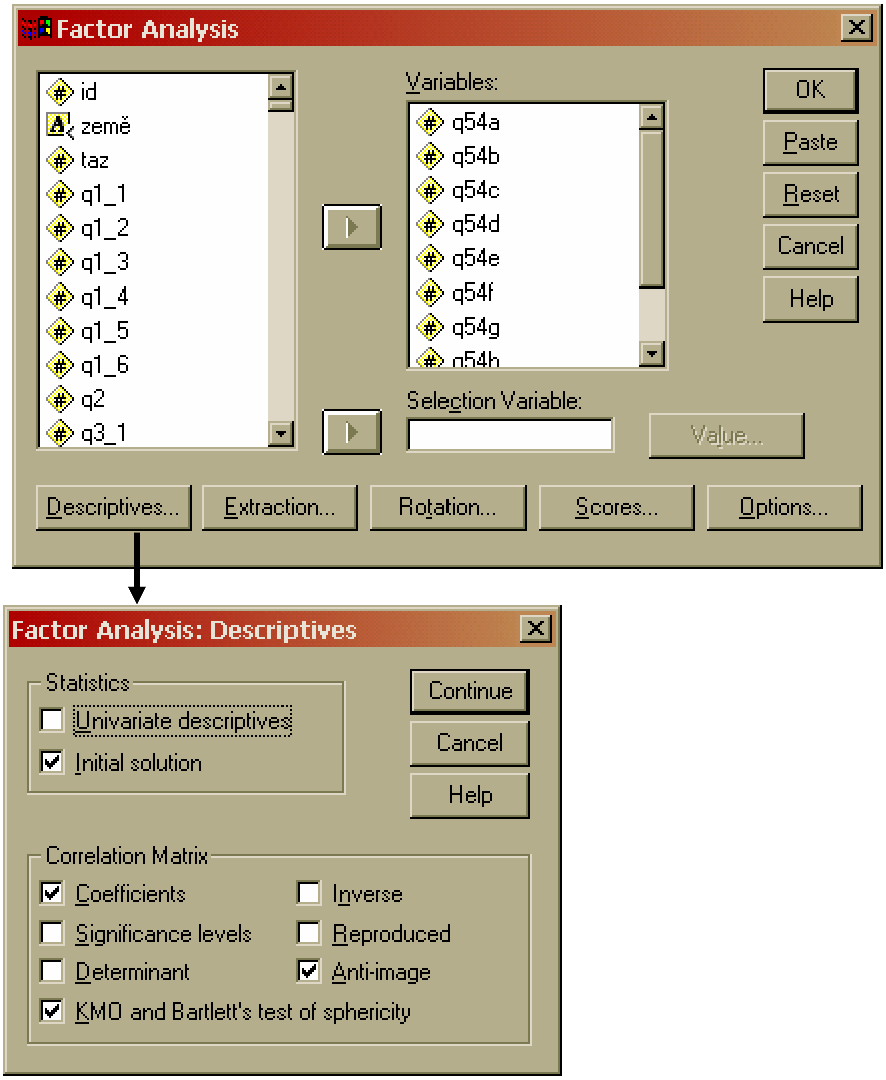
(3) musíme vypočítat faktorové zátěže a jednotlivé faktory pojmenovat.

Ad 1. Nejdříve musíme zjistit, zdali položky, které chceme faktorovat, jsou pro FA vůbec vhodné
Proměnné, které vstupují do FA, musí být měřeny přinejmenším na ordinální úrovni a měly by mít delší stupnice. Dichotomické proměnné jsou pro aplikaci FA nevhodné[2]. Proměnné by navíc mezi sebou neměly být příčinně vztaženy. Položky musí mezi sebou korelovat takovým způsobem, že těsnost korelace by měla být vyšší než 0,3. Zatímco první dvě podmínky můžeme odhalit pouhou úvahou nad povahou dat – v našem případě není ani jedna z podmínek porušena – těsnost korelace si musíme nechat samozřejmě spočítat. Vzájemné korelace mezi proměnnými ukáže matice vzájemných korelací (viz tab. 11.1), kterou dostaneme již jako první část výstupu po zadání faktorové analýzy:

[2] Alternativou zde může být použití shlukové analýzy.

Analyze – Data reduction – Factor (vepište příslušné proměnné) – *Descriptives* (viz obr. 11.2 – *OK*

Obr. 11.2: Zadání výpočtu faktorové analýzy



Výstup z *Descriptives* přinese především matici vzájemných korelací (viz tab. 11. 1)[3]. Vidíme v ní, že některé korelace nejsou příliš vysoké (jsou vysvíceny žlutě). Dále přinese tzv. anti-image matici (viz tab. 11.2), v níž v její dolní polovině jsou na diagonále (vysvíceny žlutě) uvedeny hodnoty Kaiser-Meyer-Olkinovy míry (KMO míra) vhodnosti položek pro faktorovou analýzu. Tato míra by měla být vyšší než 0,6, vhodnější ale je, pokud dosahuje hodnoty 0,7 a více. Hodnoty KMO jsou v našem případě u všech položek dobré. Konečně výstup z tohoto zadání přinese také celkovou hodnotu KMO (pro všechny položky a také Bartlettův test (viz tab. 11.3). Oba tyto údaje nám pomohou dále zhodnotit, zdali naše položky, které chceme faktorovat, jsou pro FA vůbec vhodné.

[3] Poznamenejme, že výpočet faktorové analýzy je založena právě na matici vzájemných korelací, a tudíž pro získání řešení není třeba mít původní data, ale postačí mít pouze korelační matici. Způsob, jak tímto způsobem zadávat faktorovou analýzu přes syntaxi je popsán v anglicky psaných manuálech k SPSS.

Tab. 12.1: Matice vzájemných korelací zkoumaných položek

Correlation Matrix

	Q54A	Q54B	Q54C	Q54D	Q54E	Q54F	Q54G	Q54H
Q54A Odpovědnost za jednotlivce	1,00							
Q54B Nezaměstnaní	,176	1,00						
Q54C Konkurence je:	,366	,175	1,00					
Q54D Stát by měl:	,322	,068	,264	1,00				
Q54E Příjmy by měly být:	-,243	,028	-,131	-,252	1,00			
Q54F Soukromé vlastnictví	,412	,041	,407	,403	-,208	1,00		
Q54G Za důchodové zabezp. odpov.	,397	,146	,169	,319	-,186	,301	1,00	
Q54H Za bydlení zodpovědný:	,452	,178	,243	,289	-,204	,303	,481	1,00

Tab. 12.2: Anti-image matice

Anti-image Matrices

		Q54A Odpovědnost za jednotlivce má:	Q54B Nezaměstnaní nabízené zaměstnání:	Q54C Konkurence je:	Q54D Stát by měl:	Q54E Příjmy by měly být:	Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	Q54G Za důchodové zabezp. má být zodpovědný:	Q54H Za bydlení má být zodpovědný:
Anti-image Covariance	Q54A Odpovědnost za jednotlivce má:	,643	-6,76E-02	-,130	-4,5E-02	8,38E-02	-,121	-,109	-,155
	Q54B Nezaměstnaní nabízené zaměstnání:	-6,763E-02	,927	-,115	-2,8E-03	-7,9E-02	6,8E-02	-5,0E-02	-7,E-02
	Q54C Konkurence je:	-,130	-,115	,763	-6,0E-02	6,47E-03	-,197	4,7E-02	-3,E-02
	Q54D Stát by měl:	-4,492E-02	-2,84E-03	-6,E-02	,757	,118	-,170	-,101	-4,E-02
	Q54E Příjmy by měly být:	8,378E-02	-7,94E-02	6,5E-03	,118	,890	3,8E-02	2,9E-02	5,1E-02
	Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	-,121	6,796E-02	-,197	-,170	3,81E-02	,678	-6,2E-02	-3,E-02
	Q54G Za důchodové zabezp. má být zodpovědný:	-,109	-5,05E-02	4,7E-02	-,101	2,94E-02	-6,2E-02	,697	-,225
	Q54H Za bydlení má být zodpovědný:	-,155	-7,16E-02	-3,E-02	-4,3E-02	5,10E-02	-3,4E-02	-,225	,668
Anti-image Correlation	Q54A Odpovědnost za jednotlivce má:	,831[a]	-8,76E-02	-,185	-6,4E-02	,111	-,184	-,163	-,236
	Q54B Nezaměstnaní nabízené zaměstnání:	-8,763E-02	,695[a]	-,136	-3,4E-03	-8,7E-02	8,6E-02	-6,3E-02	-9,E-02
	Q54C Konkurence je:	-,185	-,136	,782[a]	-7,8E-02	7,85E-03	-,274	6,5E-02	-5,E-02
	Q54D Stát by měl:	-6,443E-02	-3,39E-03	-8,E-02	,842[a]	,144	-,238	-,140	-6,E-02
	Q54E Příjmy by měly být:	,111	-8,74E-02	7,8E-03	,144	,842[a]	4,9E-02	3,7E-02	6,6E-02
	Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	-,184	8,573E-02	-,274	-,238	4,90E-02	,796[a]	-9,0E-02	-5,E-02
	Q54G Za důchodové zabezp. má být zodpovědný:	-,163	-6,28E-02	6,5E-02	-,140	3,73E-02	-9,0E-02	,795[a]	-,330
	Q54H Za bydlení má být zodpovědný:	-,236	-9,10E-02	-5,E-02	-6,1E-02	6,61E-02	-5,1E-02	-,330	,801[a]

a. Measures of Sampling Adequacy(MSA)

Tab. 12.3: Míry KMO a Bartlettův test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,807
Bartlett's Test of Sphericity	Approx. Chi-Square	2560,806
	df	28
	Sig.	,000

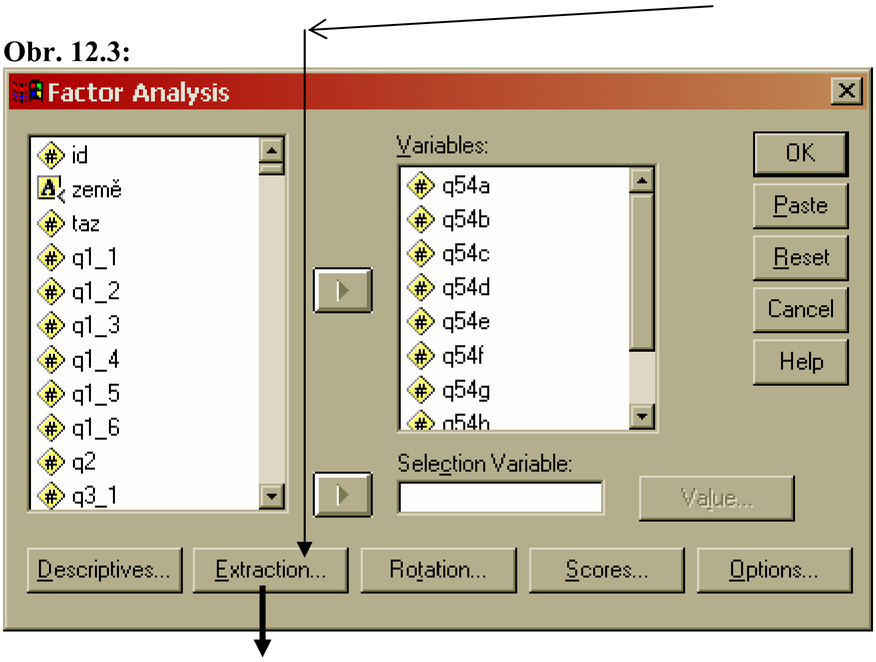
KMO by měla být vyšší než 0,7 (což je) a signifikance Bartlettova testu by měla být významná minimálně na úrovni 0,05 (což je). Nulová hypotéza u Bartlettova testu je, že proměnné na sobě v základním souboru nezávisejí. Zdá se tedy, byť ne všechny korelační vztahy jsou na požadované úrovni těsnosti, že naše data jsou pro FA vhodná. Můžeme tedy přistoupit k druhému kroku FA, a tou je extrakce faktorů

Ad 2. Extrakce faktorů

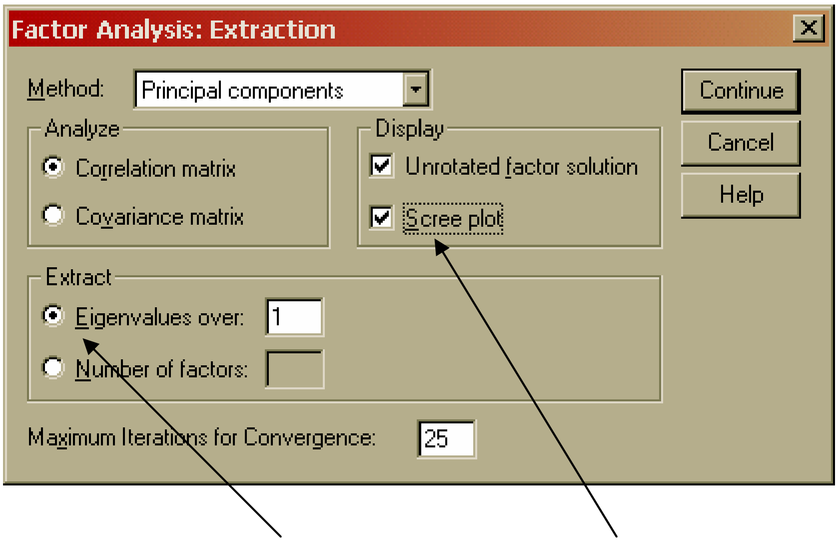
Pro nalezení faktorů (ve statistické literatuře se tomu kroku říká „extrakce") existují různé možnosti, velmi častým postupem je již dříve zmíněná tzv. metoda hlavních komponent. Nejdříve se musíme rozhodnout, s kolika faktory chceme vlastně pracovat – to zjistíme právě na základě postupu, s nímž jednotlivé faktory (komponenty) analyzované baterie položek vyextrahujeme.

Extrakci faktorů provedeme tak, že klikneme na tlačítko Extraction.

Obr. 12.3:



V něm pak zadáme operace, které nám napoví, s kolika faktory máme pracovat. Předně to je výpočet tzv. vlastních čísel (*eigenvalues*) a dále sestrojení Cattelova suťového grafu (*Sree plot*):



Na základě výpočtu *eigenvalues* a jejich grafu (*Scree plot*) se pak rozhodneme, s kolika faktory budeme pracovat. Pro toto rozhodnutí platí dva návody, které ne vždy dávají stejné výsledky:

(1) volíme takový počet faktorů (komponent), které mají hodnotu eigenvalue vyšší než 1. To je Kaiserovo pravidlo[4];

(2) volíme takový počet faktorů, které v grafu (*Scree plot*) jsou nad prolomením křivky.

Ukažme si to.

Tab. 11.4: Hodnoty eigenvalue

Total Variance Explained

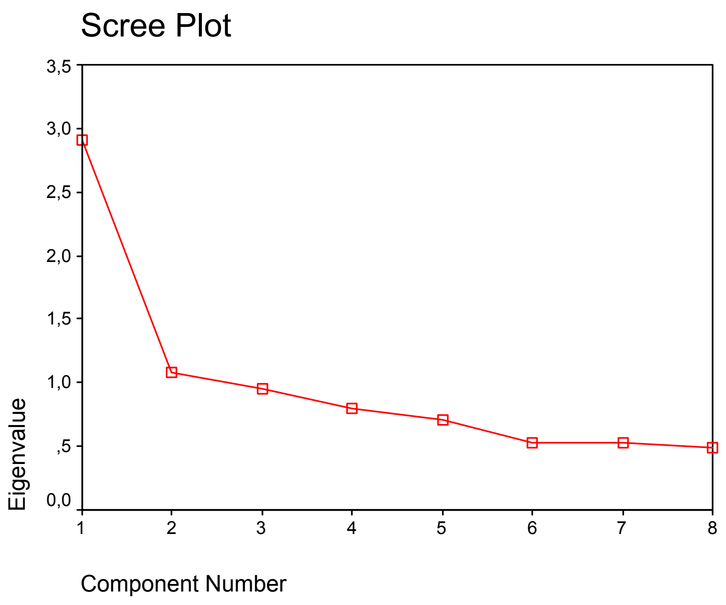
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**2,907**	**36,336**	36,336	2,907	36,336	36,336
2	**1,083**	**13,544**	**49,880**	1,083	13,544	49,880
3	,956	11,948	61,828			
4	,800	9,994	71,821			
5	,705	8,810	80,631			
6	,533	6,668	87,300			
7	,524	6,550	93,849			
8	,492	6,151	100,000			

Extraction Method: Principal Component Analysis.

Vidíme (v tab. 11.4), že pouze u dvou komponent jsou hodnoty eigenvalues větší než 1. První komponenta (faktor) vyčerpává 36 % variance v položkách, druhá komponenta vyčerpává 13 % variance, oba faktory dohromady pak 50 %. Samozřejmě, čím více variance je vysvětleno, tím lépe můžeme redukovat původní položky.

Graf vypadá následovně:

[4] Logika tohoto pravidla je následující: bereme jen faktory, které jsou z hlediska vysvětleného rozptylu lepší než původní proměnné (to jsou ty, které mají vlastní číslo větší než jedna).

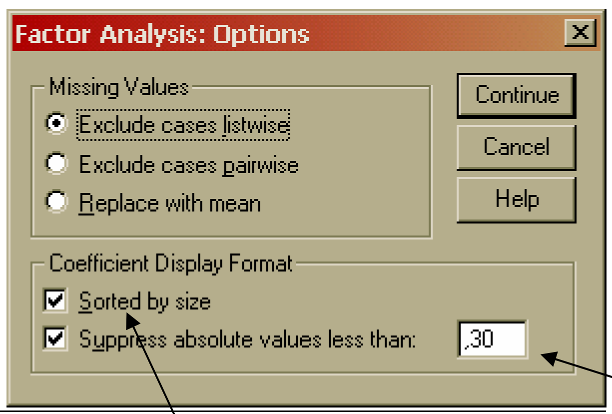


Podle tohoto obrázku bychom měli extrahovat pouze jeden faktor, neboť u druhého se křivka zřetelně láme – a pravidlo říká, že máme vzít takový počet faktorů, který je nad zlomem křivky. Jelikož ale víme, že jeden faktor vyčerpává pouze 36 % variance, rozhodneme se v tomto případě pro dvoufaktorové řešení.

Ad 3. Pojmenování faktorů

Nyní již tedy zbývá poslední krok – nalézt, které položky spadají do kterého faktoru, nebo řečeno jinak, nalézt, kterými položkami jsou jednotlivé faktory syceny. Zjistíme to prostřednictvím výpočtu faktorových zátěží (*factor loadings*). Tyto zátěže jsou korelace mezi faktorem a příslušnou položkou. Čím vyšší je tato korelace, tím více je faktor touto položkou sycen. Doporučuje se, aby tato korelace byla vyšší než 0,30.

Obr. 11. 4. Nastavení *Options* ve FA



Před tím, než výpočet faktorů spustíme (nastavení, které jsme udělali na obr. 11.3 by bylo již pro výpočet postačující), zadáme si *Options* tak, aby do tabulky nebyly zobrazeny hodnoty se zátěží nižší než 0,3 a aby položky byly seřazeny podle velikosti.

<u>Výstup:</u>

Tab. 11.5: Faktorové zátěže

Component Matrix[a]

	Component	
	1	2
Q54A Odpovědnost za jednotlivce má:	,744	
Q54H Za bydlení má být zodpovědný:	,690	
Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	,683	
Q54G Za důchodové zabezp. má být zodpovědný:	,655	
Q54D Stát by měl:	,630	
Q54C Konkurence je:	,576	
Q54B Nezaměstnaní nabízené zaměstnání:		,806
Q54E Příjmy by měly být:	-,426	,517

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Tabulka 11.5 je hlavním výstupem z FA. Přináší hodnoty korelačních koeficientů mezi položkou a příslušným faktorem. Vidíme, že už toto řešení je docela povedené, neboť většina faktorových zátěží je silná pouze v jednom faktoru a slabá ve faktoru druhém. Výjimkou je položka Q54E, která má vysoké faktorové zátěže v obou komponentách – korelace s 1. faktorem je -0,426 a s 2. faktorem je 0,517. Abychom tuto anomálii odstranili, provedeme tzv. rotaci faktorů.

Rotace faktorů

Abychom zvýšili interpretovatelnost faktorů, necháme je rotovat. Smyslem rotace faktorů je, aby se původně rozptýlené body co nejvíce přimkly k jednomu z extrahovaných faktorů. Představme si například, že jsme v nějakém výzkumu udělali FA z deseti položek a dostali jsme na základě metody hlavních komponent dvoufaktorové řešení, které je uvedeno v tab. 11.6.

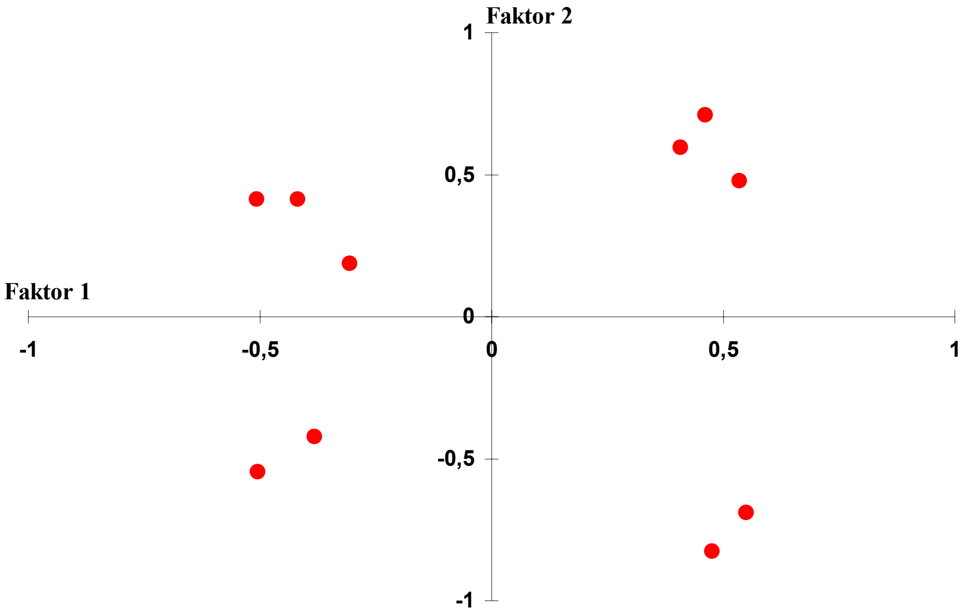
Tab. 11.6: Smyšlené faktorové zátěže deseti položek (ilustrace pro rotaci)

Položka	1. faktor	2. faktor
p1	-0,419	0,414
p2	-0,306	0,188
p3	0,476	-0,825
p4	0,461	0,711
p5	0,407	0,597
p6	0,549	-0,688
p7	0,535	0,479
p8	-0,382	-0,421
p9	-0,507	0,415
p10	-0,505	-0,545

Každou položku (p1 až p10) máme v této tabulce popsanou dvěma souřadnicemi, hodnotou korelace (faktorovou zátěží) této položky s prvním faktorem a hodnotou korelace této položky s druhým faktorem.

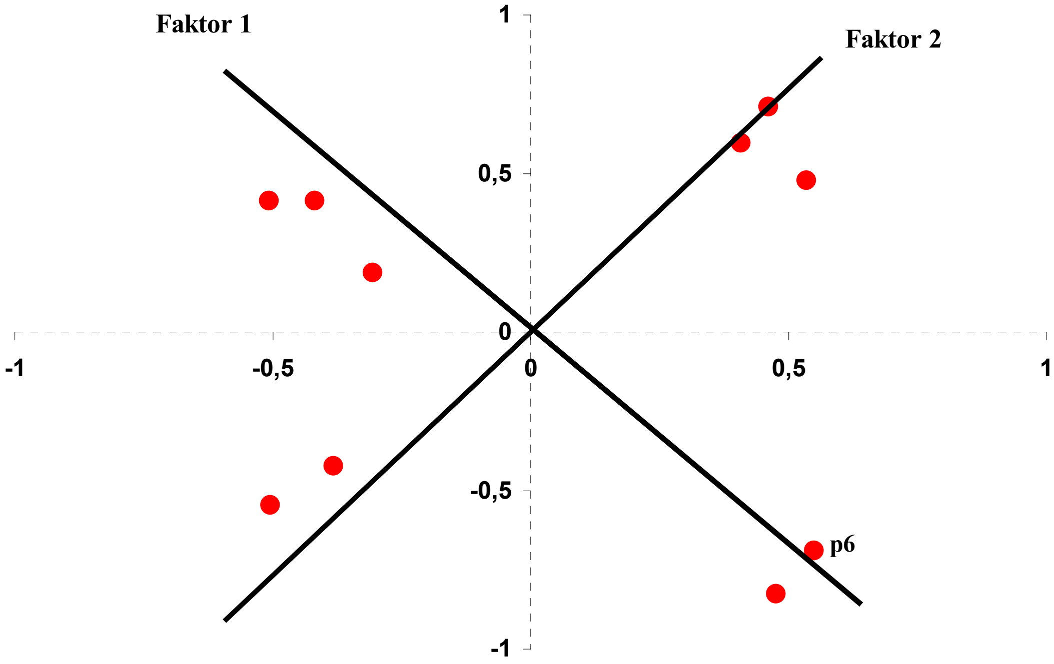
Nechejme si tyto hodnoty zobrazit do grafu (viz obr. 11.5), v němž jsme u každé položky vynášeli hodnoty faktorové zátěže prvního. faktoru na osu X a hodnoty faktorové zátěže druhého faktoru na osu Y.

Obr. 11.5: Faktorové zátěže 10 položek ve dvoufaktorovém řešení



Je zřejmé, že toto nerotované řešení nemá dobrou interpretaci, neboť jednotlivé body jsou od os (faktorů) poměrně daleko. Proto se rozhodneme pro rotaci faktorů, v daném případě pro rotaci ortogonální, to je pravoúhlou, kdy dodržíme skutečnost, že i po rotaci budou osy (faktory) svírat pravý úhel a že tedy budou na sobě nezávislé,budou nekorelované (viz obr. 11.6).

Obr. 11.6: Ukázka rotace faktorů:

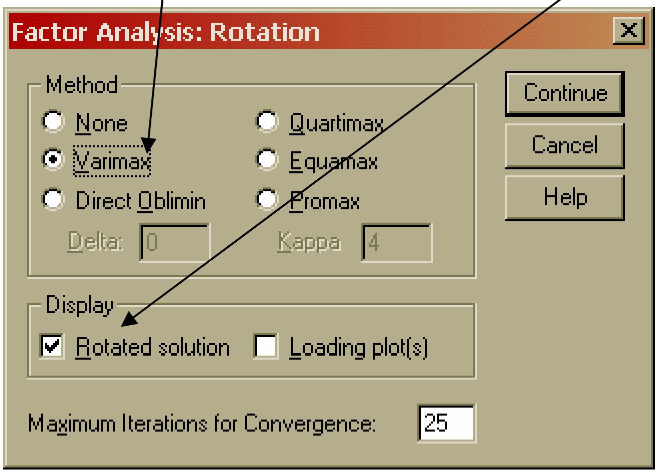


Rotací se jednotlivé souřadnice změnily. Tak např. položka p6 (viz v pravém dolním kvadrantu obrázku 11.6) měla v nerotovaném řešení souřadnice 0,549 a –0,688, zatímco v řešení rotovaném se souřadnice změnily na (řečeno přibližně) 0,7 a 0,04. To má velký význam pro interpretaci: zatímco v nerotovaném řešení, jsme nevěděli, zda přiřadit položku *p6* k prvnímu, nebo ke druhému faktoru, v řešení rotovaném je jasné, že tato položka jasně spadá pod faktor 1, neboť s tímto faktorem je silně korelována (0,7), zatímco s faktorem 2 je korelace nulová (0,04).

Kromě pravoúhlé rotace existuje i rotace šikmá (velmi často se používá postupu *Oblimin*). Ta spočívá v tom, že předpokládáme, že extrahované faktory nejsou nekorelované, ale že naopak spolu souvisejí. Rotace pak probíhá stejně, jako u rotace ortogonální, pouze s tím rozdílem, že faktory (osy) spolu při rotaci udržují úhel menší než 90 stupňů[5].

Vrátíme-li se opět k našemu příkladu, je zřejmé, že je třeba volit rotaci faktorů. Držme se rotace ortogonální, která se provádí nejčastěji metodou *Varimax*[6]. Rotaci zadáme tím způsobem, že v dialogovém okně faktorové analýzy (viz obr. 11.2) klikneme na tlačítko *Rotation* a v jeho dialogovém okně zaškrtneme metodu Varimax a budeme požadovat zobrazení rotovaného řešení.

Obr. 11. 2: Zadání rotace



Rotované řešení uvádí tabulka 11.7.

---

[5] Pěkná ukázka použití šikmé rotace je v článku Matějů, Kreidl (1999).

[6] Více informací o možnostech rotace a jejich dopadech je náplní magisterského kurzu a samozřejmě mnohých textů, česky např. Hebák a kol. (3. díl, 2005), Blahuš (1984).

Tab. 11.7: Faktorové zátěže po rotaci varimax

Rotated Component Matrix[a]

	Component	
	1	2
Q54A Odpovědnost za jednotlivce má:	,747	
Q54H Za bydlení má být zodpovědný:	,707	
Q54G Za důchodové zabezp. má být zodpovědný:	,663	
Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	,646	-,302
Q54D Stát by měl:	,587	-,344
Q54C Konkurence je:	,586	
Q54B Nezaměstnaní nabízené zaměstnání:	,379	,760
Q54E Příjmy by měly být:	-,349	,572

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Toto rotované řešení nám bohužel nijak nepomůže v interpretaci. Naopak přibylo položek, které mají poměrně silnou korelaci v obou položkách, takže pojmenování faktorů by bylo obtížné. Zkusme ještě šikmou rotaci *obliminovou* (viz obr. 11.8*)*:

Tab. 11.8: Faktorové zátěže po šikmé rotaci oblimin

Structure Matrix

	Component	
	1	2
Q54A Odpovědnost za jednotlivce má:	,744	
Q54H Za bydlení má být zodpovědný:	,691	
Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	,681	
Q54G Za důchodové zabezp. má být zodpovědný:	,656	
Q54D Stát by měl:	,628	
Q54C Konkurence je:	,577	
Q54B Nezaměstnaní nabízené zaměstnání:		,791
Q54E Příjmy by měly být:	-,422	,538

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Toto řešení se blíží nerotovanému postupu v metodě hlavních komponent. Můžeme tedy faktorovou analýzu uzavřít s tím, že se ukazuje, že v českém prostředí se položky rozkládají přinejmenším do dvou faktorů. Jelikož ale položka Q54E spadá do obou faktorů, je jasným kandidátem na to, aby byla z dalších analýz vyloučena. Pak už nám ale zůstane pouze položka Q54B tvořící samostatný faktor, což nemá příliš valného smyslu – i tu můžeme z analýzy vyloučit. [7]

Nyní proveďme FA znovu, a to pouze se zbylými šesti položkami – výsledkem je přehledná jednofaktorová struktura (viz tab. 11.9).

Tab. 11.9: Výsledná matice faktorových zátěží

Component Matrix[a]

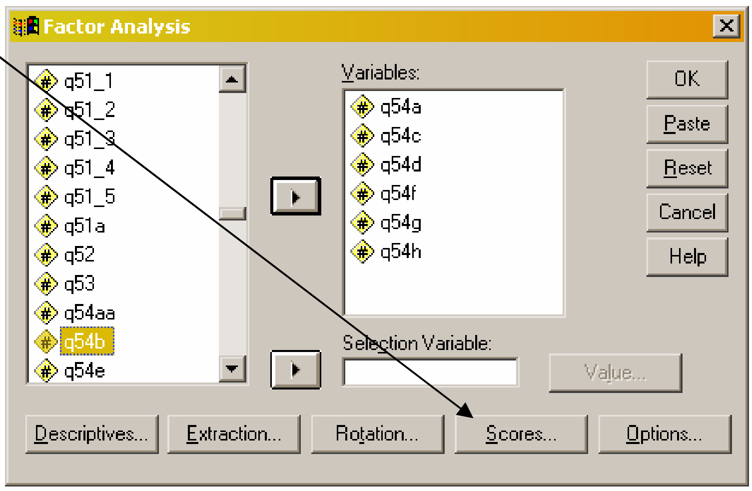
	Component
	1
Q54A Odpovědnost za jednotlivce má:	,746
Q54C Konkurence je:	,580
Q54D Stát by měl:	,636
Q54F Rozšiřovat/ zachovat by se mělo obchod. a prům.:	,706
Q54G Za důchodové zabezp. má být zodpovědný:	,671
Q54H Za bydlení má být zodpovědný:	,700

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

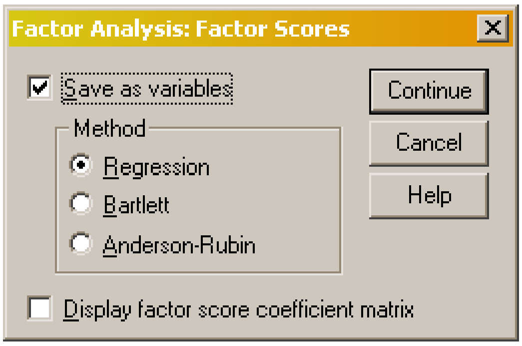
V naší úloze jsme tedy zjistili, že z osmi položek, o nichž jsme se domnívali, že měří dimenzi paternalismus–liberalismu, můžeme smysluplně pro vytvoření nové proměnné použít pouze položek šest. Nyní tedy zbývá poslední krok, a to konstrukce nové proměnné: *paternalismus*. K dispozici máme dva způsoby, jak ji vytvořit.

1. Jednoduše tak, že pomocí nám již známé procedury *Compute* sečteme hodnoty těchto šesti extrahovaných položek, čímž u každého respondenta vytvoříme skóre paternalismu-liberalismu. Pozor ale, při sečítání proměnných si musíme být jisti, že stupnice, na nichž jsou tyto proměnné měřeny, mají stejnou orientaci. Pokud by tomu tak u některých nebylo, museli bychom je patřičně otočit. U proměnných, které jsme extrahovali, je orientace identická a jde od liberalismu (odpovědnost jedince) k paternalismu (odpovědnost státu). Minimální hodnota této nové proměnné bude 6 (6 x 1 = 6), což je nejvyšší liberalismus, maximální pak 60 (6 x 10), což je nejvyšší paternalismus.
2. Vytvoříme novou proměnnou přímo s pomocí faktorové analýzy. V proceduře FA klikneme na tlačítko *Scores* a v novém dialogovém okně zaškrtneme příkaz *Save as variables*[8].



[7] Zkuste si pro zajímavost udělat FA pro sedm položek, to je včetně položkyQ54B. Uvidíte, že i tato úloha vede k tomu, že je třeba položku Q54B z analýzy vyloučit.

[8] Tyto nově vytvořené proměnné (faktorové skóry) mají tu příjemnou vlastnost, že se blíží normovanému normálnímu rozdělení, tedy rozdělení s průměrem 0 a směrodatnou odchylkou 1 a lze je bez problémů využít ve většině statistických metod.



Jelikož výše uvedený příklad byl poněkud atypický, neboť vedl k extrakci pouze jednoho faktoru, ukažme si ještě jeden příklad použití FA. Je rovněž z výzkumu EVS, z části, která sledovala, jaké položky jsou v percepci respondentů důležité pro spokojené manželství. Tabulka 11.10 již přináší rotované řešení:

Tab. 11.10: Faktorové zátěže po rotaci varimax s vynechanými hodnotami pro zátěže menší než 0,3

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Q40_14 Společné trávení volného času manželů	**,769**				
Q40_15 Debaty o společných zájmech manželů	**,768**				
Q40_1 Věrnost v manželství	**,441**				-,376
Q40_11 Dělba dom. práce mezi manželi	**,410**				
Q40_16 Stejný etnický původ manželů		**,704**			
Q40_3 Stejný soc. původ		**,697**			
Q40_5 Společné náboženství manželů		**,652**			
Q40_7 Shoda manželů v názorech na politiku		**,628**			
Q40_2 Přiměřený příjem manželů			**,754**		
Q40_6 Dobré bydlení manželů			**,685**		
Q40_12 Děti v manželství			**,505**		
Q40_4 Úcta a uznání manželů				**,757**	
Q40_8 Porozumění manželů				**,717**	
Q40_13 Ochota manželů diskutovat o problémech	,439			**,491**	
Q40_9 Oddělené bydlení manželů od rodičů					**,702**
Q40_10 Sexuální soužití manželů					**,586**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Ukazuje se, že baterie 16 položek se rozpadá do pěti faktorů (hodnota KMO byla 0,76 a Bartlettův test byl signifikantní). Vidíme, že první faktor je sycen čtyřmi položkami (vyznačenými žlutě) a že korelace těchto položek s ostatními faktory je ve všech případech kromě jednoho (jedná se o korelaci položku Q40_1 s pátým faktorem, ale tato korelace není příliš vysoká, takže ji můžeme ignorovat) vyšší než 0,3. Nyní je třeba se zamyslet nad sémantickým významem těchto čtyř položek a je potřeba je pojmenovat nějakým společným výrazem. Pracovně bychom mohli první faktor nazvat „faktorem společných aktivit a věrnosti". Druhý faktor by mohl být nazván „sociální homogenita manželů", třetí „materiální podmínky" atd.

Zdá se tedy, že co se týče charakteristik nutných pro úspěšné manželství, je z hlediska české veřejnosti nejsilněji působícím momentem faktor společných aktivit partnerů, druhým pak faktor sociální homogenity partnerů a třetím faktor materiálních podmínek – tyto tři faktory vyčerpávají největší podíl rozptylu (viz tab. 11.11).

Tab. 11.11: Podíl vysvětlené variance jednotlivými komponentami

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**2,948**	18,427	18,427	2,948	18,427	18,427
2	**1,777**	11,106	29,533	1,777	11,106	29,533
3	**1,343**	8,396	37,929	1,343	8,396	37,929
4	**1,055**	6,594	44,523	1,055	6,594	44,523
5	**1,033**	6,455	50,978	1,033	6,455	**50,978**
6	,882	5,513	56,491			
7	,870	5,440	61,931			
8	,847	5,294	67,225			
9	,788	4,926	72,150			
10	,771	4,816	76,967			
11	,721	4,503	81,470			
12	,673	4,209	85,679			
13	,657	4,107	89,786			
14	,582	3,636	93,422			
15	,562	3,513	96,935			
16	,490	3,065	100,000			

Extraction Method: Principal Component Analysis.

Tento meritorní (věcný) výsledek FA má ještě důležité metodologické implikace. Pokud bychom chtěli vytvořit z jednotlivých položek součtový index předpokladu úspěšného manželství, FA jasně ukazuje, že bychom těchto indexů museli vytvořit minimálně pět – tedy tolik, kolik faktorů FA extrahovala.

Tímto výpočtem ovšem řešení úlohy pomocí faktorové analýzy ještě nemusí končit. Dá se např. předpokládat, že do postojů o příčinách úspěšného manželství bude intervenovat věk, že mladá populace bude mít jiné postoje, než populace starší nebo že jiné postoje budou zaujímat respondenti svobodní a jiné respondenti ženatí či rozvedení atd. Proto bychom mohli nasadit faktorovou analýzu pro různě definované podsoubory (ty bychom vybrali např. s pomocí procedury *Select cases*) a mohli bychom srovnávat jednotlivá řešení.

* * *

Závěrečné poznámky

Faktorová analýza je mocným exploračním nástrojem analýzy dat. Používá se velmi často především jako nástroj technický – např. před vytvořením součtového indexu kontrolujeme, zdali všechny položky, které zamýšlíme sečítat do jednoho indexu, jsou extrahovány do stejného (jednoho) faktoru. Faktorová analýza ovšem přináší i výsledky věcné, které slouží pro zodpovězení příslušné výzkumné otázky. Např. v úloze o podmínkách šťastného manželství je možné jednotlivá faktorová skóre uložit jako novou proměnnou (jak na to viz obrázky na str. 12 a 13) a pak lze spočítat např. korelaci faktorových skóre s věkem. Výsledkem by byla tabulka 11.12.

Tab. 1.12: Pearsonovy korelace jednotlivých faktorových skóre s věkem respondenta

Correlations

		VEK
FAC1_1 REGR factor score 1 for analysis 1	Pearson Correlation	,023
	Sig. (2-tailed)	,342
	N	1758
FAC2_1 REGR factor score 2 for analysis 1	Pearson Correlation	-,221**
	Sig. (2-tailed)	,000
	N	1758
FAC3_1 REGR factor score 3 for analysis 1	Pearson Correlation	-,117**
	Sig. (2-tailed)	,000
	N	1758
FAC4_1 REGR factor score 4 for analysis 1	Pearson Correlation	,023
	Sig. (2-tailed)	,338
	N	1758
FAC5_1 REGR factor score 5 for analysis 1	Pearson Correlation	,121**
	Sig. (2-tailed)	,000
	N	1758

**. Correlation is significant at the 0.01 level (2-tailed).

Z korelací je patrné, že např. faktorová skóre prvního faktoru s věkem vůbec nekoreluje (r = 0,02), ale že jistá korelace se objevila u faktorových skóre druhého faktoru. Sociální homogenita manželů jakožto podmínka úspěšného manželství je tak korelována s věkem. Čím vyšší je věk respondenta, tím nižší je hodnota tohoto faktorového skóre, neboli pro starší respondenty není sociální homogenita manželů tak důležitá jako pro respondenty mladší.

Tímto konstatováním jsme se dostali k samotnému závěru kursu. Jeho poslední dvě lekce (regresní analýza a faktorová analýza) byly současně i jakýmsi úvodem k multivariačním technikám statistické analýzy, jejichž společným znakem je, že při analýze berou v úvahu paralelní působení mnoha proměnných. Jak se tyto multivariační analýzy provádějí, je obsahem kursu v magisterském studiu (kurs SOC418). Všechny z vás, jimž jsme analýzu dat naším bakalářským kursem neznechutili, v něm velmi rádi uvítáme.