

## LEKCE12 PŘÍKLAD

Faktorová analýza (FA) je technika s jejíž pomocí se snažíme nahradit vztahy mezi sadou vzájemně spjatých proměnných malým počtem nezávislých, avšak ne přímo pozorovatelných znaků. Tím je dána také její primární funkce – redukce dat, redukce proměnných.<sup>1</sup> Komentovaný příklad nám nejlépe ukáže její smysl.

Výzkum EVS 1999 zjišťoval, jaké skutečnosti jsou podle názorů veřejnosti důležité pro úspěšné manželství. Otázka byla formulována následovně: *Zde je seznam věcí, o nichž si někteří lidé myslí, že jsou nutné pro úspěšné manželství. Prosím, řekněte mi pro každou z nich, zda ji považujete za velmi důležitou, spíše důležitou nebo nepříliš důležitou pro úspěšné manželství.*

Položky byly:

- A Věrnost
- B Přiměřený příjem
- C Stejný sociální původ
- D Vzájemná úcta a uznání
- E Společné náboženské přesvědčení
- F Dobré bydlení
- G Shoda názorů na politiku
- H Porozumění a snášenlivost
- I Oddělené bydlení od rodičů partnera
- J Dobré sexuální soužití
- K Dělbá domácích prací
- L Děti
- M Ochota diskutovat o problémech, které mezi manželi vyvstávají
- N Společné trávení co nejvíce času
- O Časté debaty o společných zájmech
- P Stejný etnický (národnostní) původ

Postoj k jejich důležitosti byl měřen na stupnici: 1 Velmi důležité, 2 Spíše důležité, 3 Nepříliš důležité.<sup>2</sup> Je zřejmé, že tyto položky jsou určitým způsobem vzájemně spjaty – vždyť jsou považovány za jednu z možných příčin úspěšného manželství.

Vzájemnou souvislost těchto proměnných lze zjistit např. tak, že si necháme vypočítat matici interkorelací (viz tab. 1). Jenže tím získáme velké množství údajů, které nám příliš nepomohou v tom, abychom se ve vzájemných souvislostech dobře orientovali. Proto na tato data nasadíme faktorovou analýzu, metodu hlavních komponent (*Principle component analysis*), která toto zjednodušení provede – to je zjistí, kolik je v pozadí faktorů, do nichž je možné tyto položky redukovat.

*Analyze – Data reduction – Factor* (vepište příslušné proměnné) – *Descriptives* (viz obr. 1) – OK

<sup>1</sup> Abychom byli úplně přesní, tento výrok platí o jednom typu faktorové analýzy, o tzv. explorační faktorové analýze (*Exploratory factor analysis*). Kromě toho ještě existuje tzv. konfirmační faktorová analýza (*Confirmatory factor analysis*), jejímž cílem je testování hypotéz.

<sup>2</sup> Takto krátká stupnice (jen se třemi variantami) není pro faktorovou analýzu úplně nejvhodnější. Bylo by lepší, kdyby postoj k položkám byl měřen alespoň s pomocí pětistupňové škály.

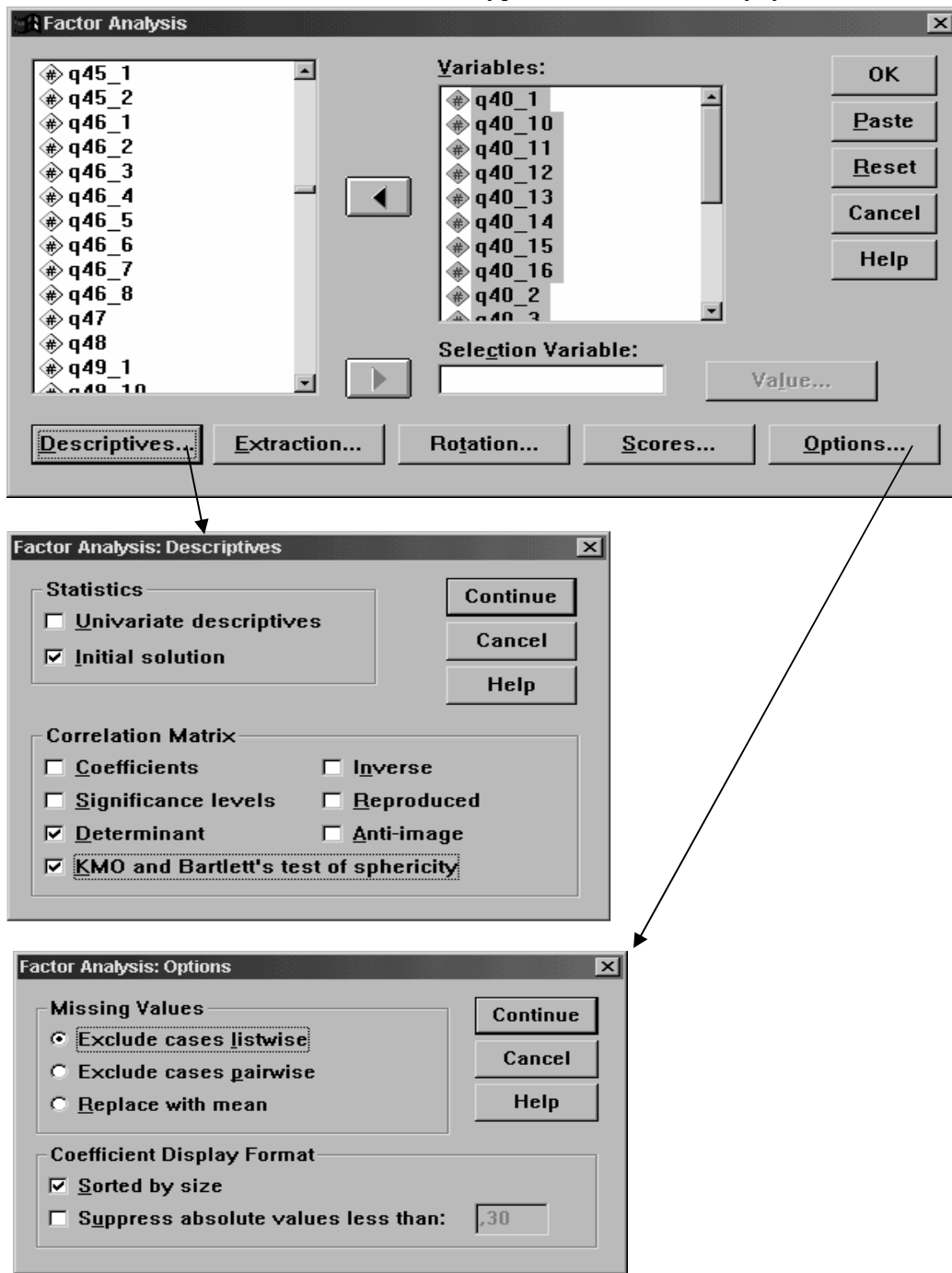


Tab. 1: Matice korelací položek důležitých pro úspěšné manželství (Spearmanův koef.)

Correlations

	Q40_1 Věrnost v manželství	Q40_10 Sexuální soužití manželů	Q40_11 Dělbba dom. práce mezi manželi	Q40_12 Děti v manželství	Q40_13 Ochota manželů diskutovat o problémech	Q40_14 Společné trávení volného času manželů	Q40_15 Debaty o společných zájmech manželů	Q40_16 Stejný etnický původ manželů	Q40_2 Přiměřený příjem manželů	Q40_3 Stejný soc. původ	Q40_4 Úcta a uznání manželů	Q40_5 Společné náboženství manželů	Q40_6 Dobré bydlení manželů	Q40_7 Shoda manželů v názorech na politiku	Q40_8 Porozumění manželů	
Věrnost v manželství	1,000															
Sexuální soužití manželů	,083	1,000														
Dělbba dom. práce mezi manželi	,108	,212	1,000													
Děti v manželství	,148	,126	,150	1,000												
Ochota diskutovat o problémech	,093	,210	,170	,124	1,000											
Společné trávení volného času	,217	,205	,215	,196	,254	1,000										
Debaty o společných zájmech	,162	,171	,216	,133	,296	,489	1,000									
Stejný etnický původ manželů	,066	,051	,063	,019	-,043	,120	,117	1,000								
Přiměřený příjem manželů	,051	,176	,121	,126	,015	,081	,075	,085	1,000							
Stejný soc. původ	,095	,082	,117	,051	-,008	,078	,108	,396	,211	1,000						
Úcta a uznání manželů	,125	,100	,100	,138	,260	,166	,176	-,035	,109	,028	1,000					
Společné náboženství manželů	,100	,006	,084	,024	,023	,070	,078	,207	,033	,279	,079	1,000				
Dobré bydlení manželů	,092	,244	,159	,173	,057	,211	,153	,092	,377	,173	,062	,045	1,000			
Shoda v názorech na politiku	,053	,063	,115	,066	,003	,118	,194	,278	,082	,249	,064	,251	,142	1,000		
Porozumění manželů	,093	,190	,112	,116	,328	,186	,179	-,014	-,006	-,004	,320	,068	,071	,062	1,000	
Oddělené bydlení od rodičů	-,003	,209	,130	,068	,094	,070	,101	,072	,142	,069	,069	-,013	,175	,052	,045	1,000

Obr. 1: Zadání výpočtu faktorové analýzy



V Options si zadáváme, že korelace s faktory mají být seřazeny podle velikosti (sorted by size). Také bychom požadoval, aby se netiskly hodnoty menší než určitá zvolená výše (suppress absolute values less than ...). Těto možnosti prozatím nevyužijeme.

Výstup:

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,761
Bartlett's Test of Sphericity	Approx. Chi-Square	3274,535
	df	120
	Sig.	,000

Tento výstup je důležitý. Testuje totiž, zda je nasazení faktorové analýzy vůbec smysluplné. Musí být splněny dva požadavky:

1. KMO (Kaiser-Meyer-Olkinova míra) musí být větší než 0,7. Pokud je menší než 0,5, není možné FA použít.
2. Bartlettův test by měl být signifikantní (to je Sig. by měla vyjít nižší než 0,05).

V našem případě jsou oba dva požadavky naplněny, takže použití FA je možné.

Další důležitou tabulkou výstupu z FA je tabulka *Total variance explained*, která říká, kolik faktorů bylo extrahováno a jak velký podíl variance jednotlivé faktory vysvětlují. Bylo nalezeno celkem pět faktorů: vidíme, že v tabulce ve sloupci *Component* je celkem 16 položek, tedy přesně tolik, kolik je položek v naší úloze. Teoreticky by totiž mohla být každá položka samostatným faktorem, ale to bychom náš úkol redukovat položky do několika málo faktorů nesplnili. Obecně pro SPSS platí, že každá komponenta, jejíž Eigenvalue je větší než 1, je extrahována jako jeden faktor.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	<b>2,948</b>	18,427	18,427	2,948	18,427	18,427
2	<b>1,777</b>	11,106	29,533	1,777	11,106	29,533
3	<b>1,343</b>	8,396	37,929	1,343	8,396	37,929
4	<b>1,055</b>	6,594	44,523	1,055	6,594	44,523
5	<b>1,033</b>	6,455	50,978	1,033	6,455	<b>50,978</b>
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.

FA extrahovala pět faktorů, které společně vysvětlují asi 51 % variance. Není to úplně nejvíce, dobré faktorové řešení přináší vysvětlenou varianci ze 60 – 70 %, ale není to ani řešení nejhorší.

Tab. 2: Korelace mezi faktorem a položkou:

Component Matrix<sup>a</sup>

	Component				
	1	2	3	4	5
Q40_14 Společné trávení volného času manželů	,619	-,214	-,094	-,199	-,371
Q40_15 Debaty o společných zájmech manželů	,606	-,188	-,217	-,300	-,290
Q40_11 Dělna dom. práce mezi manželi	,476	-,025	,076	-,130	-,105
Q40_10 Sexuální soužití manželů	,461	-,111	,338	-,287	,212
Q40_3 Stejný soc. původ	,407	,597	-,100	,046	,081
Q40_16 Stejný etnický původ manželů	,349	,588	-,215	-,188	,052
Q40_13 Ochota manželů diskutovat o problémech	,435	-,479	-,184	-,142	,152
Q40_8 Porozumění manželů	,380	-,421	-,183	,152	,419
Q40_7 Shoda manželů v názorech na politiku	,407	,415	-,260	-,071	,115
Q40_2 Přiměřený příjem manželů	,365	,245	,567	,320	-,007
Q40_6 Dobré bydlení manželů	,488	,159	,503	,178	-,086
Q40_5 Společné náboženství manželů	,304	,401	-,403	,249	,208
Q40_4 Úcta a uznání manželů	,365	-,348	-,172	,428	,367
Q40_12 Děti v manželství	,379	-,154	,164	,394	-,223
Q40_1 Věrnost v manželství	,378	-,103	-,196	,271	-,389
Q40_9 Oddělené bydlení manželů od rodičů	,297	-,014	,381	-,364	,384

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Tato tabulka je hlavním výstupem z FA. Přináší hodnoty korelačních koeficientů mezi položkou a příslušným faktorem. Vidíme např., že položka „společné trávení volného času manželů“ má hodnotu korelace s 1. faktorem 0,619, s 2. faktorem -0,214 atd. Cílem je vybrat ty položky, které korelují s každým faktorem silně, to je jejichž korelace je vyšší než 0,30 a pak pro tuto sadu položek najít – podle jejich věcného obsahu – společný název. To většinou není v prvním kroku faktorové analýzy možné, neboť v prvním kroku se většinou žádá nerotované řešení, které obvykle obsahuje několik vyšších korelací mezi jednou položkou a několika faktory, takže my nevíme, ke kterému faktoru danou položku přiřadit. Např. u položky Q40\_3 Stejný soc. původ je korelace s 1. faktorem 0,41 a s druhým faktorem 0,60 (0,597). Obě korelace jsou slušně vysoké, takže nevíme, zdali tato položka patří do prvního, nebo do druhého faktoru.

### Rotace faktorů

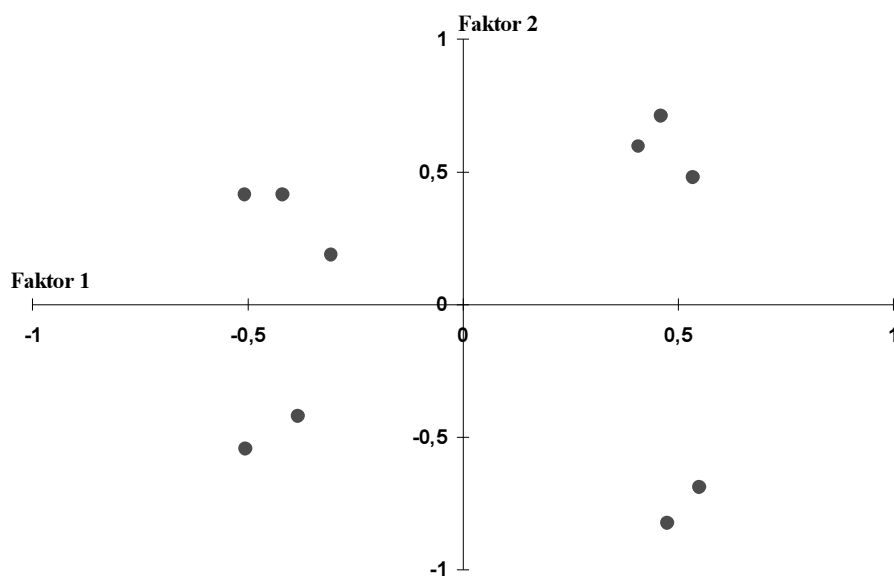
Abychom zvýšili interpretovatelnost faktorů, necháme je rotovat. Smyslem rotace faktorů je, aby se původně rozptýlené body co nejvíce přimkly k jednomu z extrahovaných faktorů. Představme si například, že jsme v nějakém výzkumu udělali FA z deseti položek a dostali jsme na základě metody hlavních komponent dvoufaktorové řešení, které je uvedeno v tab. 4.

Tab. 4: Smyšlené faktorové zátěže deseti položek

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

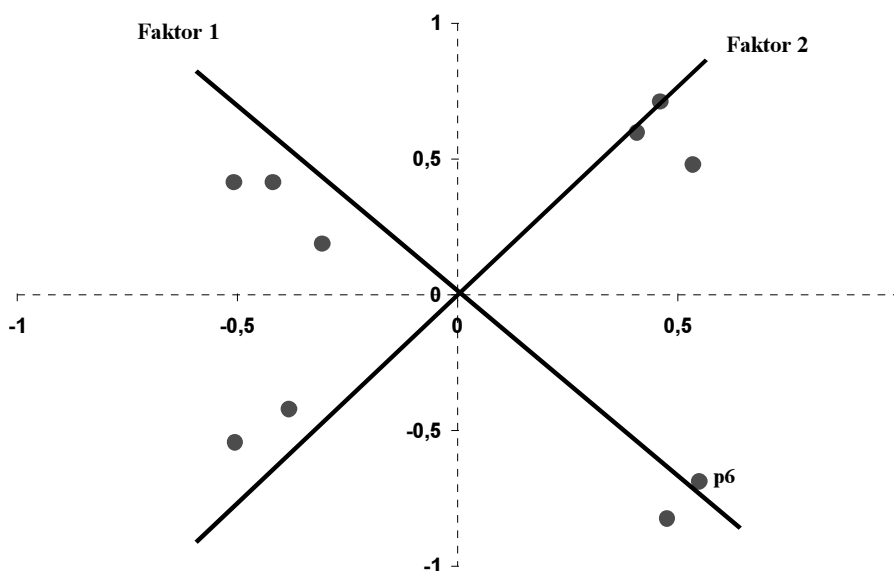
Každou položku (p1 až p10) máme v této tabulce popsanou dvěma souřadnicemi, hodnotou korelace (faktorovou zátěží) této položky s prvním faktorem a hodnotou korelace této položky s druhým faktorem. Nechejme si tyto hodnoty zobrazit do grafu (viz obr. 2), v němž jsme u každé položky vynášeli hodnoty faktorové zátěže 1. faktoru na osu X a hodnoty faktorové zátěže 2. faktoru na osu Y.

Obr. 2: Faktorové zátěže 10 položek ve dvoufaktorovém řešení



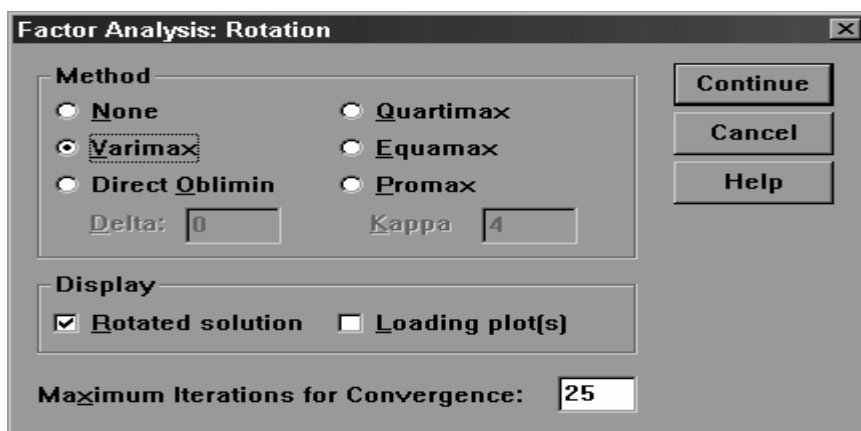
Je zřejmé, že toto nerotované řešení nemá dobrou interpretaci, neboť jednotlivé body jsou od os (faktorů) poměrně daleko. Proto se rozhodneme pro rotaci faktorů, v daném případě pro rotaci ortogonální, to je pravouhloou, kdy dodržíme skutečnost, že i po rotaci budou osy (faktory) svírat pravý úhel a že tedy budou na sobě nezávislé, budou nekorelované (viz obr. 3).

Obr. 3: Ukázka rotace faktorů:



Rotaci se jednotlivé souřadnice změnily. Tak např. položka p6 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 zda 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).

Vrátíme-li se opět k našemu příkladu, je zřejmé, že je třeba volit rotaci faktorů. Držme se rotace ortogonální, která se v SPSS provádí metodou *varimax*. Rotaci zadáme tím způsobem, že v dialogovém okně faktorové analýzy (viz obr. 1) klikneme na tlačítko *Rotation* a v jeho dialogovém okně zaškrtneme metodu *varimax*.





Tab. 5a: Faktorové zátěže po rotaci varimax

Rotated Component Matrix<sup>a</sup>

	Component				
	1	2	3	4	5
Q40_14 Společné trávení volného času manželů	,769	,056	,122	,071	,044
Q40_15 Debaty o společných zájmech manželů	,768	,138	-,033	,092	,083
Q40_1 Věrnost v manželství	,441	,076	,234	,142	-,376
Q40_11 Dělbá dom. práce mezi manželi	,410	,132	,196	,051	,187
Q40_16 Stejný etnický původ manželů	,127	,704	-,014	-,165	,107
Q40_3 Stejný soc. původ	,041	,697	,213	-,067	,050
Q40_5 Společné náboženství manželů	-,045	,652	,023	,245	-,184
Q40_7 Shoda manželů v názorech na politiku	,153	,628	,005	,044	,065
Q40_2 Přiměřený příjem manželů	-,059	,120	,754	-,030	,172
Q40_6 Dobré bydlení manželů	,155	,114	,685	-,026	,221
Q40_12 Děti v manželství	,250	-,071	,505	,210	-,179
Q40_4 Úcta a uznání manželů	,031	,039	,152	,757	-,056
Q40_8 Porozumění manželů	,134	,005	-,025	,717	,144
Q40_13 Ochota manželů diskutovat o problémech	,439	-,077	-,126	,491	,202
Q40_9 Oddělené bydlení manželů od rodičů	,039	,049	,105	,070	,702
Q40_10 Sexuální soužití manželů	,253	,024	,205	,128	,586

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Abychom se v tabulce lépe orientovali, nechejme si ji vypočítat ještě jednou s tím, že si necháme zobrazit pouze hodnoty zátěží, které jsou vyšší než 0,3. V *Options* (viz obr. 1) klikneme na tlačítko *Supress absolute vlues less than* a do políčka vepíšeme hodnotu 0,3. Tab. 5b přináší výsledek této operace.

Tab. 4b: Faktorové zátěže po rotaci varimax s vynechanými hodnotami pro zátěže menší než 0,3

Rotated Component Matrix<sup>a</sup>

	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 Dělbá 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.

<sup>a</sup>. Rotation converged in 7 iterations.

Tento výsledek je už interpretovatelný poměrně slušně. Vidíme, že první faktor je syčen 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) nižší 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 faktorem faktor společných aktivit partnerů, druhým pak faktor sociální homogenity partnerů a třetím faktor materiálních podmínek – tyto tři faktory vyčerpávají největší podíl variance (viz tab. 2).

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 příčinám 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 s pomocí procedury *Select cases*) a srovnávat jednotlivá řešení.

#### Závěrečné poznámky

Faktorová analýza je mocným exploračním nástrojem analýzy dat. Používá se velmi často jako nástroje technického – 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.

V tomto komentovaném příkladu jsme si ukázali pouhé základní vlastnosti práce s faktorovou analýzou. Nehovořili jsme například o šikmé rotaci (metoda *oblimin*), která vychází z toho, že faktory mohou být v sociálním světě navzájem korelovány. Neukázali jsme si, že je možné faktorová skóre uložit jako novou proměnnou,<sup>3</sup> a tím každému respondentovi (nebo obecně každému případu) přiřadit hodnotu jeho faktorového skóre. To nám může posloužit pro další analýzu. Např. můžeme nechat spočítat korelaci faktorových skóre s věkem. Výsledkem by byla tabulka 5.

---

<sup>3</sup> Faktorová skóre připišeme každému respondentovi tak, že v dialogovém okně faktorové analýzy (viz obr. 1) klikneme na tlačítko *Scores* a v dialogovém okně *Scores* zaklikneme *Save as variables*.

Tab. 5: Pearsonovy korelace jednotlivých faktorových skóre s věkem respondenta

Correlations

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

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

1. 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.