

Faktorová analýza

Nematematický úvod do studia problému

JIŘÍ LINHART

Fakulta sociálních věd a publicistiky UK, Praha

DALIBOR HOLDA

Ústav sociálních výzkumů mládeže a výchovného
poradenství, Praha

I. Úvodní poznámky

1. *Faktorová analýza* je matematicko-statistická metoda umožňující analýzu vnitřní struktury matice interkorelací (kovariancí). Svými tvůrci, Ch. Spearmanem, K. Pearsonem, L. I. Thurstonem, M. Hotellingem, G. M. Thomsonem a C. Burtem, byla původně určena k analýze dat z oblasti psychologie. V současné době se však poměrně často užívá mimo rámec této vědní disciplíny, především v sociologii, ekonomii, biologii a medicíně.

Faktorová analýza patří do komplexu tzv. multivariačních postupů, kam spolu s ní patří též regresní analýza, path analýza, diskriminační analýza, analýza variance a analýza hlavních komponent. Vztahy mezi faktorovou analýzou a metodou hlavních komponent jsou tak těsné, že většina autorů pojednává zároveň o metodě hlavních komponent i o faktorové analýze. Obě metody se však liší jak svým matematickým aparátem, tak i logikou analýzy a obě metody samozřejmě vedou i k poněkud odchylným interpretacím studovaných vztahů.

2. Situace, ve které se badatel rozhodne použít metody faktorové analýzy ve své výzkumné práci, je přibližně následující: badatel má určitý počet proměnných (testů) $x_1, x_2 \dots X_m$. Těmito proměnnými (testy) mohou být jednak určité psychologické testy (první data, na která byla faktorová analýza hromadněji použita, jsou právě výsledky psychologických testů z oblasti měření inteligence) nebo např. údaje o velikosti příjmu, stupni vzdělání, složitosti práce, úrovni spotřeby. Tyto proměnné má badatel vyjádřeny buď intervalovými, ordinálními nebo dichotomickými nominálními znaky. Badatel stanoví vzájemné korelace proměnných a sestaví je do korelační tabulky s tolika řádky a sloupce, kolik je proměnných. Zpravidla předpokládá, že mu tyto proměnné (testy)

charakterizují a operacionalizují určitý obecný problém (např. „inteligenci“ respondentů nebo jejich socioekonomický status), nemusí to však být podmínkou. V rámci těchto dat si může badatel klást řadu otázek.

Pro lepší pochopení logiky faktorové analýzy si uvedeme několik jednodušších úvah nad materiálem; tyto úvahy mají bezprostřední souvislost s dále vyloženým matematickým aparátem faktorové analýzy.

3. Zakladatelem faktorové analýzy je Ch. Spearman, který se pokusil objasnit interkorelace testů inteligence. Předpokládal, že každá baterie testů je vysvětlitelná dvěma typy faktorů:

1. faktoru, společného pro všechny testy zahrnuté do matice (obecný faktor),
2. faktorů specifických, jichž je tolik, kolik je testů (proměnných) zahrnutých do korelační matice, a jež jsou specifickou určitěho testu.

Tuto úvahu si ozřejmíme na elementárním vymyšleném příkladu. Předpokládejme, že jsme korelační tabulku (tabulka 1) získali experimentálně.

Tabulka 1

Test	1	2	3	4
I.	—	0,00	0,00	0,00
III.	0,00	0,00	—	0,00
IV.	0,00	0,00	0,00	—

Jestliže se pak pokusíme na tuto korelační tabulku aplikovat výše uvedenou úvahu, zjistíme, že v dané korelační matici není obsažen žádný společný faktor; všechny faktory jsou specifické.

Naproti tomu korelační tabulka 2 by obsahovala pouze faktor společný a žádné faktory specifické. Jednodušeji řečeno: testy v tabulce 1 neměří žádnou společnou

vlastnost; matice proměnných obsahuje tolik vlastností, kolik je proměnných.

Tabulka 2

Test	1	2	3	4
I.	—	1,00	1,00	1,00
II.	1,00	—	1,00	1,00
III.	1,00	1,00	—	1,00
IV.	1,00	1,00	1,00	—

Naproti tomu proměnné obsažené v tabulce 2 měří (a všechny se stejnou přesností či nepřesností) pouze jednu vlastnost a jsou navzájem zcela zaměnitelné. Celá baterie testů je zaměnitelná libovolnou proměnnou obsaženou v této baterii.

Tabulka 3

Test	1	2	3	4
I.	—	0,91	0,01	0,10
II.	0,91	—	0,11	0,15
III.	0,01	0,10	—	0,71
IV.	0,10	0,15	0,71	—

V tabulce 3 je znázorněna poněkud složitější situace. Korelace mezi testy 1 a 2 je na jedné straně sice velmi vysoká, právě tak jako korelace mezi testy 3 a 4, není však rovna 1, právě tak jako korelace mezi testy 1 a 2 na jedné straně a testy 3 a 4 na straně druhé není rovna 0, ale je o něco vyšší.

Znamená to, že dvojice testů 1 a 2, resp. 3 a 4 měří sice vždy něco společného, zároveň však měří i něco, co v druhém testu není obsaženo a co můžeme považovat buď za šum způsobený chybami měření nebo za příměs nějaké jiné, další vlastnosti. Tato baterie testů tedy obsahuje (samozřejmě v různé míře) oba typy spearmanovských faktorů: faktor společný pro všechny testy a faktory specifické a typické pouze pro jednotlivé proměnné zahrnuté do matice. Na tomto příkladu také vidíme i zjednodušení typické pro původní koncepci faktorové analýzy. Vidíme, že proměnné zahrnuté do matice se vlastně rozpadají do dvou skupin:

1. skupinu testů (proměnné 1 a 2),
2. skupinu testů (proměnné 3 a 4).

Uvnitř těchto skupin existují vysoké korelace, avšak korelace meziskupinové jsou velmi nízké. Tato skutečnost samozřejmě ovlivňuje odhad společného faktoru, a to tak, že snižuje odhad váhy společného faktoru v baterii testů.

Modernější přístupy faktorové analýzy používají k analýze baterie testů ještě dalšího pojmu — pojmu skupinového faktoru. Skupinový faktor zahrnuje tu část celkové variance proměnné, která je společná alespoň dvěma testům.

II. Matematická formulace

1. Pokusme se nyní formulovat úlohu faktorové analýzy exaktněji. [1] Všimněme si dvou odlišných přístupů k výše uvedeným problémům.

Cílem faktorové analýzy je vyřešit základní rovnici

$$x_i = \sum_{r=1}^k 1_{ir} f_r + e_i (i = 1, 2, \dots, m) \quad \dots [1],$$

kde x_i je i -tá proměnná analyzované baterie testů,

f_r je společný faktor

k je počet zjištěných faktorů a

e_i je zbytek (neobjasněná korelace, chyba), který lze připsat na vrub pouze určité proměnné.

Koeficient 1_{ir} se chápe zpravidla jako kvantifikace vlivu r -tého faktoru na i -tou proměnnou zahrnutou do interkorelační matice (tzv. factor loading, česky nejlépe „faktorová zátěž“). Smysl výše uvedené rovnice je zhruba v následující úvaze: Při pohledu na korelační matici si může badatel položit otázku, zdali existuje taková náhodná proměnná, jejíž odečtení z matice interkorelací by znamenalo, že celá matice korelací by byla vyplněna pouze nulovými korelacemi.

Jinými slovy řečeno, kdybychom odstranili z matice vliv této hypotetické proměnné f_r , korelace mezi proměnnými by byly rovny nule; mezi proměnnými by nebyl žádný jiný vztah než ten, který zprostředkovává hypotetická proměnná f_r . Tato situace zpravidla nenastává. Po „odečtení“ vlivu hypotetické proměnné f_r zůstává v korelační matici ještě celá řada korelací, které se nule neblíží. Je tedy nutno z této matice odečíst další hypotetickou proměnnou (náhodnou veličinu f_2 , případně f_3 , až $\dots f_k$).

K řešení výše naznačené úlohy se pak používá řady specifických metod a technik výpočtu, z nichž nejčastější je metoda centroidní a metoda maxima věrohodnosti. Technickými detaily se zde nezabýváme, zájemce odkazujeme na literaturu (Lawley, Maxwell 1963, Thurstone 1947, Harman 1965, Horst 1967). Počet faktorů, který různými metodami získáme, je vždy menší než počet proměnných zahrnutých do korelační (kovarianční) matice.

Nejobecněji řečeno, cílem faktorové analýzy je rozložit kovarianci proměnných na:

- a) faktory společné,
- b) faktory specifické, charakteristické pro jednotlivé proměnné,
- c) faktory náhodné, interpretované zpravidla jako chyby v měření.

2. Metoda hlavních komponent se naproti tomu neorientuje na analýzu korelací (kovariancí), ale na analýzu rozptylu určité proměnné, kterou rozkládáme na „p“ nezávislých komponent.

Základní rovnice vypadá takto:

$$x_i = \sum_{r=1}^m w_{ir} z_r \quad | i, r = 1, 2 \dots m | \dots | 2 |,$$

kde z_r označuje r -tou komponentu a w_{ir} váhu r -té komponenty v i -té proměnné.

Z korelační matice jsou komponenty extrahovány v pořadí, v jakém vyčerpávají množství variance z proměnných, tj. nejprve je extrahována ta komponenta, která vyčerpává nejvíce variance, pak ta,

kteřá vyčerpává nejvíce variance po komponentě č. 1, atd. Množství vyčerpané variance se vyjadřuje v procentech.

Jestliže tedy ve faktorové analýze je faktory vysvětlena pouze ta část variance určité proměnné, která je vysvětlitelná z korelací, tj. faktory (společná variance), a zbytek je vysvětlitelný pouze z proměnné samé (specifická variance) nebo hypoteticky nějakými jinými proměnnými, jež nejsou v matici zahrnuty, pak při metodě hlavních komponent je vysvětlena variance z c e l a beze zbytku.

Procento vyčerpání variance lze přiřknout určité proměnné, pokud se badatelé podaří dešifrovat vztah mezi proměnnou a komponentou. Variance vyčerpání určitou komponentou zahrnuje jak společnou, tak i specifickou varianci (ve smyslu faktorové analýzy).

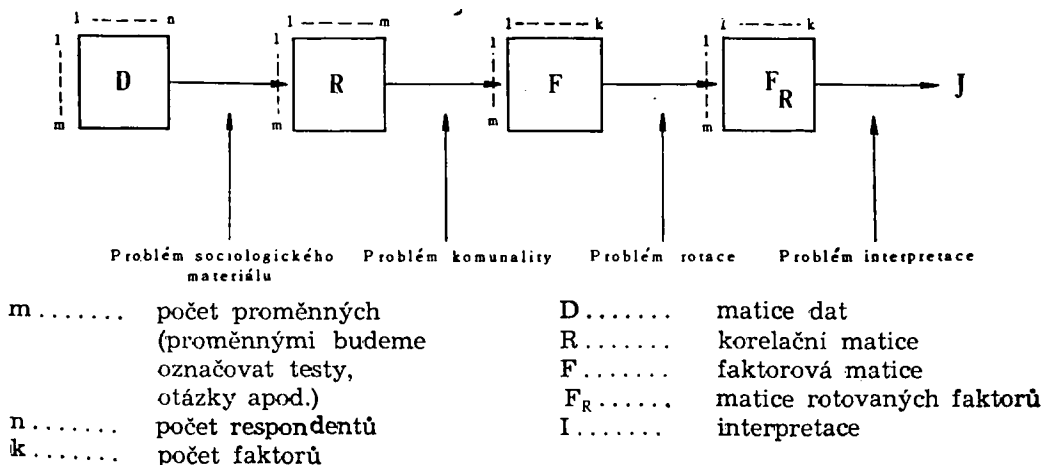
V případě, že na diagonálu korelační matice dosadíme místo jedniček nejvyšší korelace příslušného řádku matice (nebo jinak odhadnutou komunalitu), získáme pomocí matematického aparátu metody hlavních komponent výsledky blízké faktorovému šetření.

V dalším výkladu se budeme zabývat pouze faktorovou analýzou, nikoliv metodou hlavních komponent.

III. Základní etapy aplikace faktorové analýzy

1. Uvedená základní formulace problému faktorové analýzy však nepokrývá všechny etapy aplikace faktorové analýzy na sociologická data. Aplikační schéma je v komplexnosti obsaženo v obrázku 1.

Obrázek 1



Z obrázku 1 je patrné, že praktická aplikace faktorové analýzy na sociologické problémy má tyto etapy:

1. shromáždění dat
2. sestavení korelační matice
3. odhad komunalit
4. výpočet tzv. faktorové matice (výpočet faktorů)
5. rotace faktorů.

2. Charakteristika dat:

Na tomto místě není možno se touto otázkou zabývat podrobně. Podotýkáme však, že výchozí data musejí být validní, použité testy (stupnice) jednoznačně interpretované a s vysokým stupněm reliability. Použité stupnice by měly mít intervalový charakter a normální rozložení. Lze však použít i stupnic ordinálních, popř. dichotomických dat nominálních. Co se týká charakteristiky populace, není možno zde stanovit žádné přesnější instrukce nebo kritéria. Výběr populace je zcela určen charakterem řešeného problému.

3. Sestavení korelační matice:

V podstatě jde o rutinní záležitost. Předpoklady faktorové analýzy jsou nejlépe vystiženy Pearsonovým koeficientem lineární korelace, lze však v podstatě použít jakéhokoli koeficientu lineární korelace, a to např. Spearmanova koeficientu pořadové korelace, Kendallova (τ) koeficientu pořadové korelace, tetrachorické korelace, nejruznějších koeficientů asociace pro tetrachorické tabulky apod. Za určitých podmínek lze použít např. i seriálních koeficientů korelace v kombinaci s nějakým jiným korelačním koeficientem, zpravidla však je matice zaplněna korelačními koeficienty stejného typu. Vzhledem k tomu, že se při faktorové analýze nepřihlíží k výběrové chybě koeficientů korelace, je třeba, aby korelační koeficienty byly získány z co možno největšího výběrového souboru (alespoň ze souboru 100 lidí), pokud používáme R-techniky faktorové analýzy.

4. Odhad komunalit:

Většina metod výpočtu faktorů používá tzv. odhadu komunalit. Pojetí komunalit je patrné z následující rovnice:

Označíme si tu část rozptylu proměnné, která je pokryta společnými faktory jako

h_i^2 , a budeme ji nazývat komunalitou. Zbytek rozptylu vyčerpaný faktory specifickými a náhodnými označíme u_i^2 a nazveme jej unicitou.

Potom: $h_i^2 + u_i^2 = 1 \dots [3]$

Před provedením vlastního výpočtu faktorů neznáme velikost komunalit a můžeme ji pouze s větší či menší přesností odhadnout. Způsobů, jak odhadnout komunalitu, je několik. Uvedeme zde jen dva nejpoužívanější.

Nejjednodušším odhadem komunalit je dosazení nejvyšší korelace z příslušného řádku, resp. sloupce, na diagonálu korelační matice. V případě, že nejvyšší korelační koeficient v příslušném řádku (sloupci) má záporné znaménko, do diagonály matice jej zapíšeme se znaménkem kladným.

Poměrně nejpřesnější metodou odhadu je tzv. metoda malého centroidu. Při použití této metody postupujeme podle následujícího algoritmu:

1. pro každou proměnnou sestavíme submatici o příslušné proměnné a třech dalších proměnných, které s touto proměnnou mají nejvyšší korelace;

2. na diagonálu této submatice dosadíme nejvyšší korelaci z řádku, resp. sloupce této submatice;

3. sečteme všechny korelace včetně hodnot diagonály; výsledek označíme N;

4. sečteme pro proměnnou, jejíž komunalitu odhadujeme, hodnoty všech korelací v příslušném řádku submatice. Výsledek označíme R;

5. odhad komunalit vypočteme pomocí vzorce:

$$h_i^2 = \frac{R}{N} \dots [4]$$

Kromě těchto metod existuje řada dalších metod odhadu; např. metoda výpočtu průměrné korelace, metoda využívající vícenásobné korelace apod.

O těchto metodách se může čtenář poučit nejlépe v Harman 1960, Thurstone 1947.

Praxe ukazuje, že všechny metody dávají dosti podobné odhady komunalit. V praxi se obvykle neukazuje potřeba nějakým způsobem tyto odhady zpřesňovat. Řešitel si však musí uvědomit, že na odhadu komunalit závisí přesnost konečných výsledků.

5. Výpočet faktorové matice

Problematikou vlastního výpočtu faktorů se zde nemůžeme zabývat vzhledem k omezenému rozsahu článku. Není to nutné ani vzhledem k tomu, že pro většinu používaných metod výpočtu (např. pro centroidní metodu, pro metodu hlavních komponent atd.) existují spolehlivé programy. Tyto programy pracují v podstatě bez zásahu badatele. Činnost badatele, pokud používá nejobvyklejších postupů výpočtu, se omezuje na tyto předběžné úkony:

Předání matice korelačních koeficientů

V některých programech, které jsou k dispozici ve výpočetních střediscích, je vlastní program pro výpočet faktorové analýzy zařazen do bloku programů, přičemž součástí tohoto bloku je i vlastní výpočet korelační matice. V některých případech je jako vstupu použito korelačních tabulek, popřípadě neutříděných dat. Praxe však ukazuje, že nejvhodnější je přímé zadání interkorelační matice.

Určení metody odhadu komunalit

V řadě případů je vhodné dodat matici již s diagonálou obsazenou odhady komunalit.

- Určení a) počtu faktorů, které chceme z matice abstrahovat,
b) metody odhadu statistické průkaznosti faktorů.

Počet faktorů udáváme tehdy, když faktorovou analýzu používáme jako verifikující techniku striktně formulovaných teoretických předpokladů nebo při komparativních výzkumech. Počet je vždy relativně nízký.

Známe-li počet proměnných a chceme-li odhadnout počet faktorů, který v nejlepším případě můžeme z nich extrahovat a spolehlivě identifikovat, použijeme vzorce:

$$r = \lfloor \sqrt{2m+1} - \sqrt{8m+1} \rfloor \cdot \frac{1}{2} \dots \lfloor 5 \rfloor$$

Naopak, známe-li určitý počet faktorů a snažíme-li se získat odhad minimálního počtu proměnných, postupujeme podle vzorce:

$$m = \lfloor 2r+1 \rfloor + \lfloor \sqrt{8r+1} \rfloor \cdot \frac{1}{2} \dots \lfloor 6 \rfloor$$

Vztah mezi počtem faktorů a počtem proměnných pro „m“ a „r“ je znázorněn v tabulce 4.

Z důvodů stability řešení se doporučuje volit m dvakrát až třikrát větší, než je uvedeno v tabulce.

Matematické metody odhadu statistické průkaznosti nejsou obecné a používané. Pro některé metody výpočtu faktorů vůbec chybí. Zpravidla se stanoví nejmenší množství vyčerpané variance, kterou faktor musí vyčerpat, aby měl pro badatele nějaký význam. Toto množství variance se stanoví buď odhadem nebo tak, že badatel nejprve nechá vyčerpat veškerou společnou varianci (komunalitu) a získá tak maximálně možný počet faktorů, a pak na základě této informace rozhodne, které faktory potřebuje k další analýze dat.²

Specifické nároky nejrůznějších postupů faktorizace interkorelační matice nalezne zájemce v tabulce 5 a v citované literatuře.³

Splní-li badatel tyto předpoklady, pak již výpočet běží zcela bez zásahu badatele. Faktorová analýza nevyžaduje předběžného rozlišení proměnných na proměnné závislé a nezávislé, nevyžaduje stanovení kauzálních řetězců, nevyžaduje od badatele rozhodnutí, zda některá proměnná v baterii testů je „důležitější“ než proměnná jiná.

IV. Rotace faktorů

Rozložení vah jednotlivých faktorů vzhledem k proměnným, získané libovolnou metodou výpočtu (podotýkáme, že rozložení vah je z praktického hlediska stabilní a nezávislé na použité metodě výpočtu, přičemž ovšem dochází k nume-

Tabulka 4

Počet faktorů	r	1	2	3	4	5	6	7	8	9	10
Počet proměnných	m	3	5	6	8	9	10	12	13	14	15

Tabulka 5: Přehled metod výpočtu faktorů (nejpoužívanější způsoby)

Způsob výpočtu		Fak- torový model	Předběžné odhady		Před- běžné grupo- vání proměn- ných	Množství vyčerpané variance	Rozeznávané charakteristiky
			komu- nality	počtu faktorů			
1	Metoda hlavních komponent	2	ne	ne	ne	klesá s pořá- dovým číslem faktoru	všechny obecné faktory
2	Metoda hlavních faktorů	1	ano	ne	ne	klesá s pořá- dovým číslem faktoru	všechny obecné faktory
3	Centroidní metoda	1	ano	ne	ne	klesá s pořá- dovým číslem faktoru	všechny obecné faktory
4	Metoda maxima pravděpodobnosti	1	ano	ne	ne	klesá s pořá- dovým číslem faktoru	všechny obecné faktory
5	Minres	1	ne	ano	ne	klesá s pořá- dovým číslem faktoru	všechny obecné faktory
6	Metoda mnoho- násobných skupin	1	ano	ano	ano	všechny fak- tory jsou stejně silné	pouze skupino- vé faktory (šikmé řešení)
7	Jednofaktorové řešení	1	ne	ano	ano	všechny fak- tory jsou stejně silné	rozlišující skupiny faktorů
8	Dvoufaktorové řešení	1	ne	ano	ano	první faktor je silnější	jeden obecný faktor
9	Bifaktorové řízení	1	ne	ano	ano	první faktor je silnější, ostatní stejně	jeden obecný a m skupino- vých

rickým rozdílností) nemusí být vždy z hlediska interpretace nejvýhodnější. Badatel používající metodu faktorové analýzy má zpravidla již před použitím faktorové analýzy určitou představu o tom, jakým způsobem mají být váhy faktorů rozloženy. Tato představa je založena na teoretických předpokladech, znalostech korelací mezi testy a prostém zdravém úsudku.

Cílem rotace je tedy: ověřit matematicky korektní transformací výsledků faktorové analýzy výchozí koncepci, popř. upravit ji tak, aby shoda mezi interpretací a výsledky faktorové analýzy byla co největší. Je však třeba podotknout, že toto ověřování zpravidla nemá charakter statistické procedury (využívající testování hypotéz) a že jde spíše o pokusné porovnávání transformovaných dat a modelu.

Tabulka 6: Rozložení faktorových zátěží ve Spearmanově modelu jednoho obecného a jedinečných faktorů

Číslo proměnné	Typ faktoru								
	obecný G	skupinový			jedinečný				
		I.	II.	III.	1	2	3	4	5 5 2
1	×	nejsou extrahovány			×				
2	×	nejsou extrahovány				×			
3	×	nejsou extrahovány					×		
4	×	nejsou extrahovány						×	
5	×	nejsou extrahovány							×
6	×	nejsou extrahovány							×

Tabulka 7: Rozložení faktorových zátěží v Holzingerově bifaktorové teorii

Číslo proměnné	Typ faktoru											
	obecný G	skupinový			jedinečný							
		I	II	III	1	2	3	4	5	6	7	8
1	×	×			×							
2	×	×				×						
3	×	×					×					
4	×			×				×				
5	×		×						×			
6	×									×		
7	×			×							×	
8	×			×								×

Pro zajímavost zde uvádíme historicky nejvlivnější schémata hypotéz o struktuře proměnných, verifikovaných jak přímým výpočtem faktorů (Spearmanova koncepce jednoho společného faktoru a zbývajících jedinečných faktorů), tak metodami rotace.

V současné době se nejvíce používají modely založené na principu prosté struktury. Princip prosté struktury je zpravidla dobře teoreticky použitelný a vede k přehledným výsledkům. Také technická realizace (díky standardním programům) nečiní obvykle potíže.

Princip prosté struktury definoval jako první L. L. Thurstone, který vycházel z přesvědčení, že každá proměnná v baterii testů je sycena pouze malým počtem společných faktorů. Rotovaná matice faktorů odpovídající tomuto principu by měla vypadat takto: [4]

a) Každý řádek faktorové matice (rozměň redukované faktorové matice) by měl obsahovat alespoň jednu nulu.

b) Je-li „m“ počet společných faktorů, pak by měl každý sloupec faktorové matice obsahovat alespoň „m“ nul.

c) Pro každou dvojici sloupců faktorové matice by mělo existovat několik proměnných, jejichž faktorové koeficienty jsou rovny nule v jednom sloupci a jsou nenulové v druhém sloupci.

d) Pro každou dvojici sloupců faktorové matice by měla mít větší část proměnných nulové koeficienty v obou sloupcích (pro „m“).

e) Pro každou dvojici sloupců faktorové matice by mělo mít jen málo proměnných nenulové koeficienty v obou sloupcích. Ideálem prosté struktury je taková faktorová matice, která obsahuje po rotaci

co největší počet nulových prvků. Před provedením rotace ovšem neznáme místa, na kterých se tyto nulové prvky budou nacházet.

Výše uvedená pravidla je třeba chápat jako návod pro posouzení konečných výsledků získaných transformací, nikoli jako výpočtový algoritmus. Čtenář si musí uvědomit, že princip prosté struktury je pouze jedním z možných modelů konfigurace proměnných a že nadto může být chápán různě. Např. řešení, které vyhovuje principu prosté struktury, může být získáno jak za předpokladu, že faktory zůstávají na sobě nezávislé (ortogonální), tak i za předpokladu, že spolu svírají určitý úhel (menší než 90 stupňů), tj. jsou korelované. Toto různé pojetí vede také k odlišení dvou typů rotace: tzv. *rotace ortogonální* (faktory zůstávají na sobě nezávislé) a *rotace šikmé* (faktory jsou korelovány). Programy, které realizují rotaci, jsou pak konstruovány tak, aby výpočetní algoritmus optimalizoval přiblížení ke zvolenému modelu, tj. např. aby se výsledky maximálně přibližovaly modelu prosté struktury. Rotaci faktorů můžeme realizovat jak pomocí analytických metod (a samočinných počítačů), tak také řadou postupů využívajících tužky a papíru.

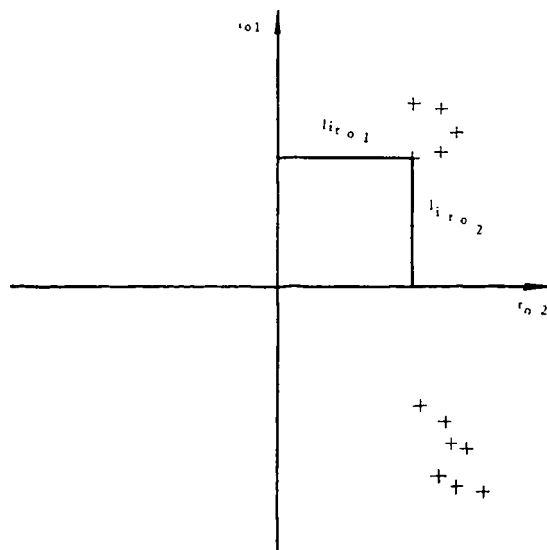
Představíme-li si proměnné jako konfiguraci bodů v systému určeném faktory, dá se potom význam jednoduché struktury popsat jako vztah mezi konfigurací proměnných a osami tohoto systému. Jednoduchou strukturu získáme nalezením takového kritéria pro rotaci systému, aby co největší počet vektorů proměnných se dostal přímo do roviny os anebo do jejich nejtěsnější blízkosti. Předpokládá se, že takový systém je možný.

Tabulka 8: Rozložení nenulových faktorových zátěží po provedení rotace (ortogonální nebo šikmé) do prosté struktury

Číslo pro- měnné	Typ faktoru											
	obecný G	skupinový			jedinečný							
		I	II	III	1	2	3	4	5	6	7	8
1	neuvažuje se	×			×							
2	neuvažuje se	×				×						
3	neuvažuje se	×					×					
4	neuvažuje se		×					×				
5	neuvažuje se		×						×			
6	neuvažuje se			×						×		
7	neuvažuje se			×							×	
8	neuvažuje se			×								×

Předcházející popis doplníme grafickým znázorněním: mějme dva faktory r_{01} , r_{02} a následující konfiguraci proměnných:

Obrázek 2



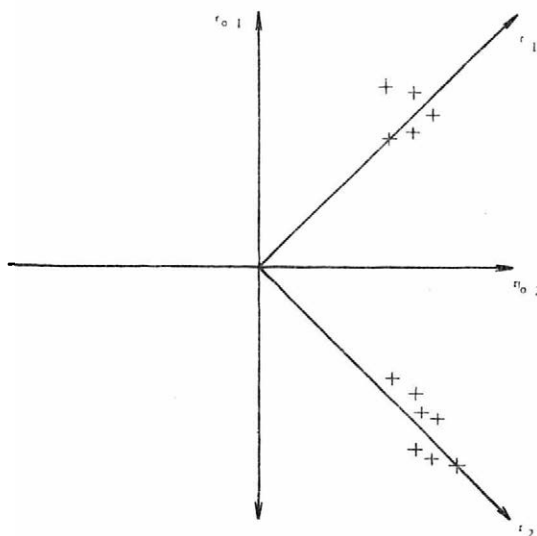
+ proměnné

$l_{ir_{01}}$ faktorová zátěž proměnné i v nerotovaném faktoru r_{01}

$l_{ir_{02}}$ faktorová zátěž proměnné i v nerotovaném faktoru r_{02}

Z obr. 2 je zřejmé, že můžeme předpokládat určité vztahy v takto uspořádané konfiguraci proměnných. Není ale splněna podmínka jednoduché struktury, kde co největší počet vektorů proměnných se má dostat do rovin os nebo do jejich blízkosti. Provede se proto rotace (viz obrázek).

Obrázek 3

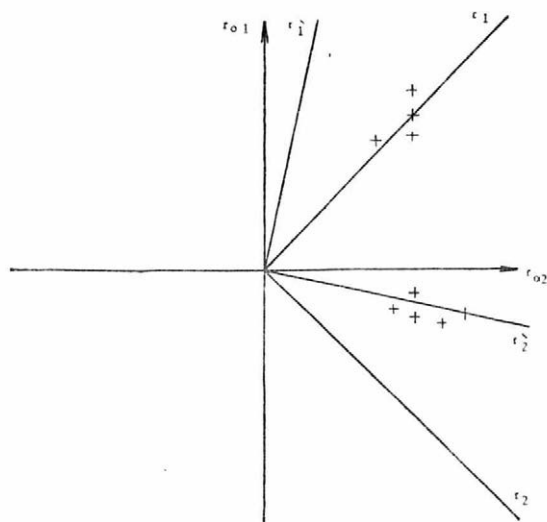


Jde tedy o ortogonální rotaci, neboť původní i rotované osy svírají pravý úhel (podmínka nekorelovanosti získaných faktorů). Získání výsledků, které splňují striktní požadavky prosté struktury ve všech bodech, je spíše výjimkou než pravidlem.

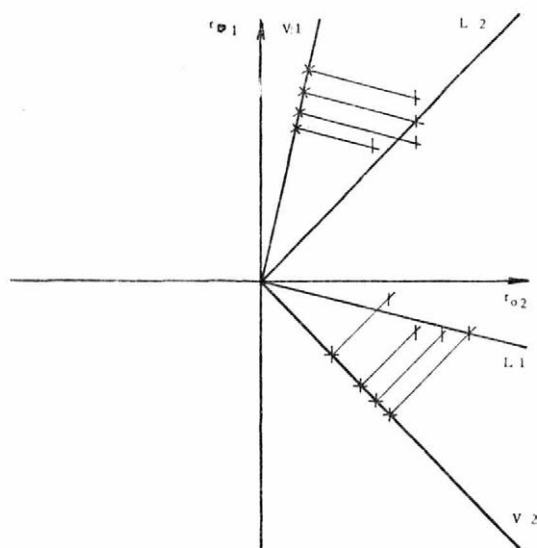
V obr. 4 je ukázán příklad, kdy nestačí ani první ani druhá rotace (ortogonální). Potom musíme použít kosoúhlou jednoduchou strukturu. V takovém modelu pracujeme s tzv. primárními faktory L_1 , L_2 a referenčními vektory V_1 , V_2 . Mějme tedy původní ortogonální systém V_{01} , V_{02} a konfiguraci proměnných jako na obr. 8.

Ukážeme si nyní princip kosoúhlé rotace, kterou máme najít kosoúhlou jednoduchou strukturu, na obr. 5.

Obrázek 4



Obrázek 5



V_1 , L_1 a V_2 , L_2 nám tvoří dva ortogonální systémy. Je obtížné rozhodnout, které řešení použít — zda primární faktory nebo referenční faktory. Thurstone vycházel z faktorů L_1 , L_2 (proto primární faktory) a pokračoval v řešení použitím referenčních vektorů. V praktické práci se dá většinou používat pouze referenčních vektorů V_1 , V_2 . Problém rotace vede k nejednoznačnému řešení, protože výsledné váhy faktorů závisí na kritériu, které použijeme pro rotaci. Na stanovení kritéria existují ovšem programy pro samočinné počítače. Pro ortogonální rotaci

např. VARIMAX, pro kosouhlou rotaci program OBLIMAX nebo OBLIMIN, BIQUARTIMIN atd.

4. Krátce upozorňujeme ještě na jednu možnost použití výsledků rotace. Uvedli jsme již, že při tzv. šikmé rotaci je úhel mezi faktory nižší nebo rovný 90 stupňům. Pomocí jednoduchého postupu lze úhel, který svírají osy roviny vymezené faktory, vyjádřit v hodnotách korelačního koeficientu. Za předpokladu, že z původní matice korelací jsme získali větší počet faktorů, získáme novou matici: matici korelací faktorů. Tuto matici můžeme opět známým způsobem faktorovat a získáme tzv. faktory vyšších řádů II, III i vyšších. Avšak vzhledem k tomu, že již faktory I. řádu jsou abstrakce, jejichž interpretace není vždy jednoduchá, a vzhledem k tomu, že matice interkorelací faktorů je výrazně ovlivněna metodou rotace, nejsou faktory vyšších řádů příliš používány a nejsou ani doporučovány teoretiky faktorové analýzy. Někteří teoretikové (např. J. P. Guilford) se dokonce k faktorům vyšších řádů staví výrazně zamítavě.

V. Interpretace výsledků faktorové analýzy

Dříve než se pokusíme nastínit určitý technický návod ke čtení výsledků faktorové analýzy, domníváme se, že je třeba si znovu uvědomit smysl faktorové analýzy. Smyslem faktorové analýzy je v podstatě zhuštěný popis struktury matice korelací. V souladu s matematickou formulací problému lze konstatovat, že cílem faktorové analýzy je nalézt určitý malý počet teoretických proměnných, které vysvětlují svými korelacemi s testy všechny (nebo podstatnou část) korelace mezi testy. Známe-li matici faktorových zátěží, můžeme reprodukovat výchozí matici interkorelací.⁶

Vydeme-li z takto formulovaného cíle faktorové analýzy, pak můžeme postulovat zhruba tyto požadavky na faktorový popis:

1. Co nejmenší počet faktorů extrahovaných z matice interkorelací by měl vysvětlit co největší množství celkové variance proměnných. Prakticky to znamená, že testy vložené do analyzované baterie by měly být co nejvíce zkorelovány. Faktorový popis získaný z matice, kde korelace se pohybují těsně kolem nuly, by

byl sice technicky možný, ale interpretačně by téměř nic nepřinesl, protože faktory získané z této matice by nevysvětlovaly proměnné baterie.

2. Totéž můžeme aplikovat i na vztah faktorů a jednotlivých proměnných. Faktory by měly vysvětlovat podstatnou část každé proměnné zahrnuté do baterie testů. I pro každou jednotlivou proměnnou by mělo platit, že množství vyčerpané variance je co největší. V extrémním případě, kdy do baterie testů začleníme proměnnou, která s ostatními prakticky vůbec nekoreluje (a tudíž i faktorové zatížení je u této proměnné téměř nulové), nemají získané faktory pro vysvětlení této proměnné žádný význam a tuto proměnnou je vlastně třeba chápat jako novou, na ostatních faktorech nezávislou „faktor“.

3. Až doposud jsme si všimli zcela abstraktně nároků, které by měl splňovat faktorový popis. Upozorňujeme zde, že posouzení shody reálných výsledků získaných faktorovou analýzou a výše jmenovaných požadavků není (nebo je jen v malé míře) otázkou statistického testování nebo jinak založeného exaktního posuzování. Posouzení, zda faktorový popis je vhodný k popisu vztahů mezi proměnnými určité baterie, nebo není, vyplývá z charakteru úlohy a je v podstatě otázkou prosté logické úvahy a teoretického rozboru. Nyní si všimněme spíše technické stránky věci:

Badatel má k dispozici za předpokladu, že výpočet probíhá strojově, tyto údaje:

1. původní matici interkorelací,
2. redukovanou faktorovou matici koeficienty l_{ir} , tj. matici faktorových zátěží,
3. rotovanou faktorovou matici obsahující transformované koeficienty „ l'_{ir} “. Rotovaná matice faktorů bývá obvykle založena na některé z metod vycházejících z principu prosté struktury. Konkrétní příklad je uveden v tabulce 9. Můžeme uvést základní vzorce, které nám umožňují odhad komunality pro jednotlivé proměnné, resp. množství variance vyčerpané určitým faktorem. V obou případech vyjadřujeme toto množství v procentuálním vyjádření.

$$h_i^2 = 100 \cdot \sum_{r=1}^k l_{ir}^2 \quad \dots [8]$$

$$V_r = 100 \cdot \sum_{i=1}^m l'_{ir}^2 \quad \dots [9]$$

- l_{ir} faktorová zátěž r-tého faktoru v i-té proměnné před rotací;
- l'_{ir} faktorová zátěž r-tého faktoru v i-té proměnné po rotaci;
- h_i^2 komunalita i-té proměnné před rotací;
- $h_i'^2$ komunalita i-té proměnné po rotaci;
- V_r množství vyčerpané variance faktorem r před rotací;
- V_r' množství vyčerpané variance r-tým faktorem po rotaci.

Upozorňujeme, že platí:

$$h_i^2 = h_i'^2 \quad \dots [10]$$

$$100 \sum_{r=1}^k \sum_{i=1}^m l_{ir}^2 = 100 \sum_{r=1}^k \sum_{i=1}^m l'_{ir}^2;$$

$$\sum_{r=1}^k V_r = \sum_{r=1}^k V_r' \quad \dots [11]$$

Máme-li pro všechny faktory a pro všechny proměnné provedeny tyto výpočty, můžeme rozhodnout, do jaké míry splňují výpočty naše očekávání. Pak můžeme přistoupit k nejobtížnějšímu momentu faktorové analýzy: k identifikaci faktorů. Faktor je určitá matematická abstrakce, kterou není možno ztotožnit s některou konkrétní proměnnou zahrnutou do baterie testů. O tom, co je vlastně obsahem faktorů, nevíme nic určitého. Vlastnosti, obsah a smysl faktoru můžeme pouze odhadovat na základě znalosti faktorových zátěží, tj. nepřesně řečeno, na základě korelací faktoru s proměnnými. Čím více faktor „koreluje“ s určitou proměnnou, tím více určuje obsah této proměnné charakter, pojmenování a interpretaci příslušného faktoru, a naopak. Podotýkáme zde na okraj, že faktorové zátěže hodnotíme zhruba stejně jako korelační koeficienty. Výsledkem tohoto posuzování je v řadě případů určitý syntetický název pro faktor, který verbálně spojuje vlastnosti všech proměnných, jež jsou s faktorem ve vysoké „korelaci“. Tak např. faktor, který vysoce „koreluje“ se všemi proměnnými baterie inteligenčních testů, dostává název „faktor obecné inteligence“; faktor, který vysoce koreluje se všemi položkami baterie proměnných charakterizujících sociálně ekonomické postavení člověka ve společnosti, nese název „faktor socioekonomického statusu“; fak-

tor, který silně a pozitivně koreluje s testy měřícími výsledky v humanitních oborech a silně, leč záporně s testy, které měří výsledky v přírodovědných oborech, se nazývá „faktor humanitně přírodovědné orientace“ (tento typ faktorů označujeme jako faktory bipolární).

Tento přístup je v podstatě zachován jak při interpretaci rotovaných, tak i nerotovaných faktorů. Pro posouzení výsledků rotace je třeba vycházet z toho, že rotované faktory není možno ztotožnit s faktory před rotací, I. faktor před rotací se nekryje (obsahově ani formálně) s I. faktorem po rotaci.

Často se dostaneme do situace, kdy systém faktorových zátěží je natolik neurčitý a proměnně korelující s určitým faktorem jsou natolik různorodé, že neumožňují jednoduchou verbální syntetizaci. Neexistují předpisy, které by umožnily nějak tuto obtíž řešit. Osvědčuje se však tato strategie: do matice korelací, tj. do baterie testů zavedeme tu proměnnou, která by podle názoru badatele mohla vyjasnit obsah natolik, aby se mohla provést verbální identifikace obsahu faktoru. Velmi často je třeba tento postup neustále opakovat, zkoušet a hledat. Jistota, že tyto postupy povedou k vyjasnění situace, však není nikdy příliš vysoká.

VI. Konkrétní příklad interpretace

Konkrétní příklad použití faktorové analýzy jsme převzali z archivu výzkumu *Sociální diferenciacce a mobilita*. Faktorová analýza byla aplikována na soubor 11 proměnných, které podle mínění badatelů určitým způsobem konkretizovaly sociální a ekonomické postavení respondenta ve společnosti. Na tomto místě abstrahujeme od skutečnosti, do jaké míry je výběr proměnných vhodný. Meritorní význam výsledků faktorové analýzy je však zcela závislý na vhodnosti, či nevhodnosti vybraných proměnných.

K charakteristice postavení respondenta ve společnosti byly použity tyto proměnné:

1. Členství respondenta v KSČ (ano — ne).
2. Počet funkcí vyplývajících z členství v masových organizacích, politických stranách apod. (Počet funkcí byl kategorizován do šesti stupňů).

3. Významnost funkcí ve společenských organizacích a politických stranách (bez funkce, funkce v místě či na závodě, funkce na okresní úrovni, krajské, celostátní).
4. Vzdělání respondenta (neukončené základní, základní, vyučen, nižší odborné, střední všeobecné, střední odborné, vysokoškolské vzdělání).
5. Složitost práce (celkem rozlišeno 6 skupin od nejméně složité, stereotypní práce až po práci vysoce kvalifikovanou, tvůrčí).
6. Mocenské postavení v zaměstnání (šestistupňový index rozlišuje osoby zcela podřízené — nemají ani jednoho podřízeného — osoby na střední úrovni řízení, např. mistry a vedoucí závodů, provozů apod. Index přihlíží k „váze“ instituce, k její důležitosti, počtu zaměstnanců atd.).
7. Příjem respondentů kategorizovaný do šesti skupin.
8. Průměrný příjem respondentovy rodiny (příjem na hlavu — rovněž kategorizovaný do šesti skupin).
9. Kvalita bytu respondenta (byt I., II., III., IV. kategorie, závadné byty, bez vlastního bytu).
10. Index vybavenosti bytu respondenta (index sumarizoval tyto předměty: pračku, ždímačku, televizi, auto, mixer, ledničku; výsledné skóre bylo kategorizováno do 6 stupňů).
11. Index participace na kulturním životě (index sumarizoval tyto položky: vlastnictví televizoru, knihovny s více než 100 svazky, návštěva galerie v uplynulém roce, „koníčky“, sledování kritik, recenzí a informací o kulturním dění; celkové skóre bylo opět kategorizováno do 6 skupin).

Použité proměnné byly získány jednak na základě regulérních škálovacích postupů (např. skalogramu — viz položka 11), nebo na základě teoretické rozvahy (viz položka 5); v řadě případů pak byla stupnice dána jednoznačně povahou sledovaného jevu (viz položka 4). Všechny položky (až na položku 1) mají ordinální charakter a mají standardní tvar (6 stupňů škály). V řadě předvýzkumů byla zjištěna postačující validita a reliabilita a celkem jasná byla i konceptualita jednotlivých stupnic. Faktorová analýza byla použita k ověření hypotézy, že rozsah položek

Tabulka 9: Matice interkorelací proměnných
Mikrocensus 1967, ekonomicky aktivní úplně socializovaní muži
N = 10 800

Název a číslo proměnné	Číslo proměnné										
	1	2	3	4	5	6	7	8	9	10	11
1. Členství v KSČ	.33	.33	.30	.20	.25	.23	.22	.11	.16	.17	.22
2. Počet funkcí	.33	.59	.59	.28	.32	.30	.24	.14	.19	.22	.30
3. Významnost funkcí	.30	.59	.59	.23	.21	.21	.22	.06	.17	.18	.24
4. Vzdělání	.20	.28	.21	.65	.65	.45	.44	.27	.41	.41	.58
5. Složitost práce	.25	.32	.23	.65	.65	.55	.43	.27	.36	.35	.50
6. Mocenské postavení	.23	.30	.21	.45	.55	.55	.41	.23	.29	.28	.41
7. Příjem	.22	.24	.22	.44	.43	.41	.44	.18	.36	.36	.41
8. Prům. příjem	.11	.14	.06	.27	.27	.23	.18	.31	.22	.21	.31
9. Kvalita bytu	.16	.19	.17	.41	.36	.29	.36	.22	.43	.42	.43
10. Vybavenost bytu	.17	.22	.18	.41	.35	.28	.36	.21	.42	.47	.47
11. Participace na kulturním životě	.22	.30	.24	.58	.50	.41	.41	.31	.43	.47	.47

může být bez ztráty informace seskupen do menšího počtu faktorů, které mohou být použity k popisu postavení respondenta ve společnosti. Korelační a faktorovou analýzou byly získány tyto výsledky:

korelace v matici na diagonálu (v tab. 9 jsou odhady komunality podtrženy). Předběžnými zkouškami bylo zjištěno, že zvolená metoda výpočtu faktorů dává výsledky obdobné jako např. metoda centroidní

Matice rotovaných faktorů

proměnné	Číslo faktoru			
	I	II	III	H ²
1	.13	— .41	— .18	22,4
2	.13	— .73	— .17	59,0
3	.11	— .75	— .06	58,7
4	.50	— .14	— .57	61,1
5	.34	— .19	— .70	65,1
6	.21	— .21	— .64	50,5
7	.41	— .19	— .39	36,4
8	.31	— .04	— .23	15,7
9	.59	— .11	— .20	40,4
10	.64	— .15	— .15	45,5
11	.60	— .20	— .37	55,7
množství vyčerpané variance	17,0	13,7	15,7	

celkem 46,4 %

Matice nerotovaných faktorů

proměnné	Číslo faktoru			
	I	II	III	H ²
1	.38	— .27	— .02	22,4
2	.53	— .55	.00	59,0
3	.45	— .61	.07	58,7
4	.74	.21	— .03	61,1
5	.74	.13	— .28	65,2
6	.63	.05	— .31	50,5
7	.59	.08	— .00	36,4
8	.37	.13	.03	15,7
9	.55	.16	.25	40,4
10	.57	.14	.32	45,5
11	.71	.14	.13	55,7
množství vyčerpané variance	34,5	8,4	3,5	

Celkem 46,4 %

Korelační matice byla vyplněna Spearmanovými pořadovými korelačními koeficienty (byl použit vzorec s korekcí na stejné pořadí); [7] k vlastnímu výpočtu faktorů se použila metoda hlavních komponent. Odhad komunality se provedl dosažením nejvyšší sloupcové, resp. řádkové

(viz archiv výzkumu). Po předběžných zkouškách bylo rozhodnuto, že z matice budou extrahovány pouze ty faktory, které vyčerpávají alespoň 3⁰/₀ variance. Rotace do prosté struktury byla provedena ortogonální rotací typu VARIMAX. Podívejme se nyní na globální inter-

pretaci nerotovaných i rotovaných faktorů. Faktorovou analýzou byly získány tři faktory, které vyčerpaly necelých 50 % variance (46,4 %). Vzhledem k tomu, že do matice byly z interpretačních důvodů, tj. z důvodů identifikace faktorů, zařazeny i proměnné, jejichž zkorelovanost s ostatními byla již na první pohled malá a tyto proměnné neměly často ani charakter stupnic (viz např. členství ve straně), lze pokládat celkové množství vyčerpané variance za dostatečné a faktorový popis lze pokládat za přiměřený. Konec konců, znalost tří faktorů nám umožňuje z přibližně 50 % predikovat výsledky v jednotlivých testech, což lze pokládat za dosti dobrý výsledek. Rozbor komunalit (viz veličinu „ h^{2u} “) tento závěr pouze podtrhuje. Interpretace důležité proměnné: vzdělání, složitost práce, mocenské postavení v zaměstnání, participace na kulturním životě, množství funkcí a jejich důležitost jsou na základě znalosti těchto tří faktorů predikovatelné z 50–60 %, což lze opět pokládat za výsledek velmi dobrý. Vždyť jestliže vyjádříme tento výsledek v hodnotách vícenásobné korelace, pak komunalitě 65,2 odpovídá koeficient množné korelace o výši 0,805, což jistě není výsledek častý a v sociálních vědách obvyklý. Z interpretačně důležitých proměnných mají velmi nízké komunality pouze dvě proměnné: příjem (komunalita rovna 36,4 odpovídá hodnotě možné korelace 0,605) a příjem na osobu (komunalita 15,7 odpovídá přibližně hodnotě množné korelace 0,385). U příjmů lze tuto skutečnost vysvětlit zhruba působením dvou vlivů:

První vliv souvisí s poměrně nižší validitou a reliabilitou stupnic příjmů. Respondenti mají totiž tendenci uvádět vyšší, nebo naopak nižší příjmy. Zkušenosť ukazuje, že nelze tento nedostatek zpravidla odstranit a samozřejmě tato skutečnost má vliv i na odhady korelací mezi příjmem a ostatními proměnnými.

Druhý vliv spočívá v tom, že na příjem působí i řada dalších fenoménů, které analyzovanou baterii testů měřeny nebyly. Jde např. o takové jevy, jako je délka praxe v oboru, diferenční a preferenční přídatky za hlučnost nebo prašnost či společenskou nezbytnost vykonávané práce. Tyto fenomény měřeny nebyly a zdá se, že tyto jevy nebude možno zahrnout beze zbytku do zde již získaných a zčásti

popsaných faktorů, ale že bude třeba zavést nový faktor (v pořadí IV). Analogická situace je u průměrného příjmu. Rovněž zde (a možná ještě v daleko větší míře) je třeba připustit nižší validitu výchozího měření, avšak i zde je pravděpodobně nutno připustit vliv dalších neměřených, leč proměnnou vysvětlujících proměnných. Z dalších faktorových analýz (zde nepublikovaných) vyplývá, že množství predikované variance (komunality) se u této proměnné zvýší zavedením dalšího faktoru (hypotetický faktor V), na jehož popisu se výrazně podílí kromě průměrného příjmu takové proměnné, jako je počet příjmů v domácnosti, počet dětí a v menší míře, přesto však dosti výrazně, i některé proměnné operacionalizující životní styl či kulturnost respondenta.

Tyto úvahy jsou pak podkladem pro další plánování faktorových experimentů.

Nyní k interpretaci nerotovaných faktorů: V matici nerotovaných faktorů si snadno přečteme, že na výsledky nejsilněji působí I. faktor, který vyčerpává 34,5 % variance, což je přibližně 80 % veškeré vyčerpané variance. Z matice je dále patrné, že všechny proměnné korelují s tímto faktorem kladně a až na členství v KSČ a průměrný příjem na osobu i značně vysoko. Interpretace tohoto faktoru je celkem jednoduchá: je to faktor, který shrnuje to, co je všem proměnným společné. Pohled na faktorové zátěže může potvrdit to, co je patrné i z matice korelací: osoby, které mají výsledky v libovolné proměnné nad průměrem populace, mají vyšší pravděpodobnost, že i v dalších proměnných budou jejich výsledky nad průměrem; analogicky pak platí, že ti, kteří jsou v libovolné proměnné pod průměrem, mají i v ostatních proměnných tendenci dosahovat podprůměrných výsledků. Tento faktor lze tedy nazvat faktorem „obecného sociálního postavení“, faktorem, jehož zaměření nám umožní stanovit „průměrnou“ výši sociálního statusu příslušného měřeného jednotlivce. Interpretace druhého a třetího faktoru již není tak jednoduchá. Pozorujeme především dvě věci:

- celkové množství vyčerpané variance je v těchto faktorech podstatně nižší než u prvního faktoru,
- jde o bipolární faktory.

U druhého faktoru pozorujeme, že kladná a záporná znaménka u faktorových zátěží jsou rozmístěna, že s tímto faktorem záporně koreluji první tři proměnné, tj. členství v KSČ, počet funkcí a závažnost těchto funkcí, a že ostatní proměnné s tímto faktorem koreluji pozitivně. Třetí faktor má znaménka rozložena tak, že záporná znaménka jsou umístěna u proměnných: vzdělání, složitost práce, příjem a podíl na řízení; ostatní pak mají opět kladná znaménka. Pro interpretaci těchto faktorů je vhodné si uvědomit, že faktor II resp. III je získán z reziduální matice vzniklé odečtením I., resp. II. faktoru. Tuto reziduální matici si můžeme schematicky představit jako korelační matici, ve které kladně koreluji testy: členství v KSČ, počet funkcí a významnost funkcí na jedné straně, na straně druhé pak mezi sebou koreluji kladně testy ze zbývajících skupiny. Členství v KSČ, počet funkcí a závažnost funkcí pak koreluje s ostatními testy baterie záporně. Jinými slovy řečeno: jestliže pozorujeme výsledky testů izolovaně od průměrné výše výsledků ve všech testech, izolovaně od průměrné výše sociálního postavení, pozorujeme, že existuje následující tendence: osoby, které dosahují lepších výsledků než je průměr v testech, jež charakterizují participaci na politickém systému (členství v KSČ, počet funkcí, závažnost funkcí), mají tendenci mít výsledky spíše pod průměrem z ostatních testů, tj. v příjmu, složitosti práce, životním stylu (měřeném nejružnějším způsobem) atd.; naopak osoby, které mají nadprůměrné vzdělání, příjem, složitost práce a životní styl, mají v průměru horší než průměrné výsledky v testech, které charakterizují úroveň participace na politickém dění. Tato tendence však vzniká, a to znovu podtrhujeme, až za předpokladu, že abstrahujeme od průměrné výše socioekonomického postavení.

Analogickou úvahu můžeme provést pro faktor III. Obsahová identifikace faktoru II, resp. III vychází samozřejmě z výše faktorových zátěží. U II. faktoru jsou nejvyšší faktorové zátěže umístěny u proměnné počet funkcí (a významnost vykonávaných funkcí). Ostatní faktorové zátěže jsou poměrně malé. Vyjdeme-li z této skutečnosti a z představy reziduální matice, o které jsme hovořili výše, mů-

žeme dospět k následující interpretaci tohoto faktoru: jde o faktor, který charakterizuje disproporci mezi výší politické participace a ostatními charakteristikami sociálního postavení. Můžeme pro něj vymyslet např. název „faktor politické participace“ nebo „faktor politické a profesionálně-soukromé orientace“ apod. Tyto názvy závisí v podstatě pouze na vkusu badatele; badatelé se mohou dohodnout, zda pro označení určité konfigurace faktorových zátěží (vztahujících se k určitým proměnným) budou používat ten či onen název.

Interpretace faktoru III je obdobná a i logika interpretace je zcela stejná. Opět interpretace vychází z výše a z rozložení kladných a záporných znamének u faktorových zátěží. Pohled na výši faktorových zátěží nám potvrzuje, že relativně nejsilnější jsou faktorové zátěže u proměnných: složitost práce, podíl na profesionálním řízení, vybavenost bytu. Pro tento typ faktoru lze tedy razit název „profesionální — soukromý“ nebo název „profesionální — životní styl“ či podobně. V tomto faktoru jde tedy o relativní nesoulad mezi výší skóre měřících parametry kulturní participace, vybavenosti bytu a kvality bytu na jedné straně, a výší skóre měřících stupeň profesionálních charakteristik na straně druhé.

Upozorňujeme, že v tomto případě byla interpretace nerotovaných faktorů relativně jednoduchá. Ve většině případů však lze nerotované faktory stěží takto celkem jednoznačně interpretovat, a to vzhledem k velmi nejasnému rozložení znamének i vah faktorových zátěží.

Analýzu nerotovaných faktorů lze shrnout také do tohoto resumé: v baterii testů existují kromě obecného faktoru celkem tři skupiny testů: a) testy, které měří úroveň pozic v profesionální sféře, b) testy měřící úroveň vybavení bytu, kulturní participaci a kvalitu bytu, tedy životní styl, a c) testy, které měří úroveň politické participace (viz faktor II). Mezi těmito skupinami existuje relativní nesoulad, který je však nesouladem pouze částečným, protože zde existuje obecný faktor (viz faktor I), který ukazuje, že i přes relativní odchylky výsledků v jednotlivých skupinách testů, všechny testy měří svým způsobem totéž — celkovou úroveň pozic v socioekonomické sféře. Tyto odchylky a výkyvy se

prosazují až tehdy, když srovnáváme osoby se stejnou průměrnou úrovní sociálně ekonomických pozic.

Cílem dále použité rotace do prosté struktury je ověření předpokladu, že ve zkoumané baterii skutečně existují výše uvedené tři skupiny testů. Pohled na matici rotovaných faktorů nás přesvědčuje o tom, že kritéria prosté struktury byla splněna pouze přibližně. Vidíme, že např. vzdělání a index participace na kulturním životě mají vysoké faktorové zátěže jak ve faktoru I, tak ve faktoru III, a obdobně např. i index složitosti práce. Pro tuto skutečnost je možno nalézt i meritorní vysvětlení. Např. vzdělávací instituce nepřipravují své posluchače pouze pro určité povolání, ale také je připravují pro život ve volném čase, zasvěcují je do kultury určitého národa a napomáhají jim při osvojování určitých standardů chování atd. Tento názor je tedy celkem v korespondenci s rozložením vah v obou těchto faktorech. Přesto však je možno vycházet z hypotézy, že použití tzv. šikmé rotace by pravděpodobně vedlo k lepší shodě s modelem prosté struktury. A nyní k interpretaci: celkem jednoznačně je určena interpretace II. faktoru, kde opravdu vysoké faktorové zátěže jsou pouze u prvních tří proměnných. Jde tedy o faktor, který nám vyděluje skupinu testů, jež měří úroveň politické participace. Interpretace zbývajících faktorů již není jednoduchá. V prvním faktoru jsou velmi vysoké faktorové zátěže jak u proměnné „kvalita bytu“, „participace na kulturním životě“, tak u proměnné „vybavenost domácnosti“. Tato skutečnost by nám umožňovala určit tento faktor jako faktor „životního stylu“. Avšak naše radost ze zdárné interpretace je zkalena vysokými faktorovými zátěžemi u příjmu, vzdělání a složitosti práce. Přesto však se k této interpretaci přikláníme jak z meritorních důvodů (jejich část jsme reprodukovali výše), tak i vzhledem k tomu, že ve skupině testů měřících „životní styl“ je řada testů, které mají vysoké faktorové zátěže pouze v tomto jediném faktoru (např. kvalita a vybavenost bytu). Analogická situace je u faktoru č. III, jehož interpretace je poněkud jednoznačnější. Zde máme vysoké faktorové zátěže především u proměnných: vzdělání, složitost práce, podíl na profesionálním řízení. Je tedy

možno tento faktor nazvat faktorem „profesionální účasti“, „profesionálních pozic“, „profesionálního statusu“ apod. Na jeho základě je možno vydělit skupinu testů, které měří především tuto charakteristiku respondenta. I tady jsou však určité potíže, které vyplývají například z relativně vysoké faktorové zátěže proměnné „participace na kulturním životě“ v tomto faktoru. Celkem jednoznačné umístění nejvyšší faktorové zátěže proměnné „mocenské postavení v instituci“ v tomto faktoru spolu s dalšími meritorními úvahami, které zde není možno rekapitulovat, nás vedou k tomu, že tato charakteristika faktoru je celkem oprávněná. Výsledky rotace celkem podepřely naše závěry vyplývající z interpretace nerotovaných faktorů, resp. nejsou s těmito závěry v příkrém protikladu.

Závěrem této části se pokusme ještě zrekapitulovat další možný průběh práce s uvedenými výsledky faktorové analýzy:

1. Je možno provést šikmou rotaci a pokusit se dosáhnout jednoznačnějšího ověření hypotézy prosté struktury.
2. Výsledků šikmé rotace je možno využít k další analýze pomocí faktorů vyššího řádu.
3. Zavedením nových proměnných do baterie testů je možno se pokusit změřit některé další dimenze (faktory) sociálního postavení. O konkrétním zaměření těchto pokusů jsme hovořili výše.

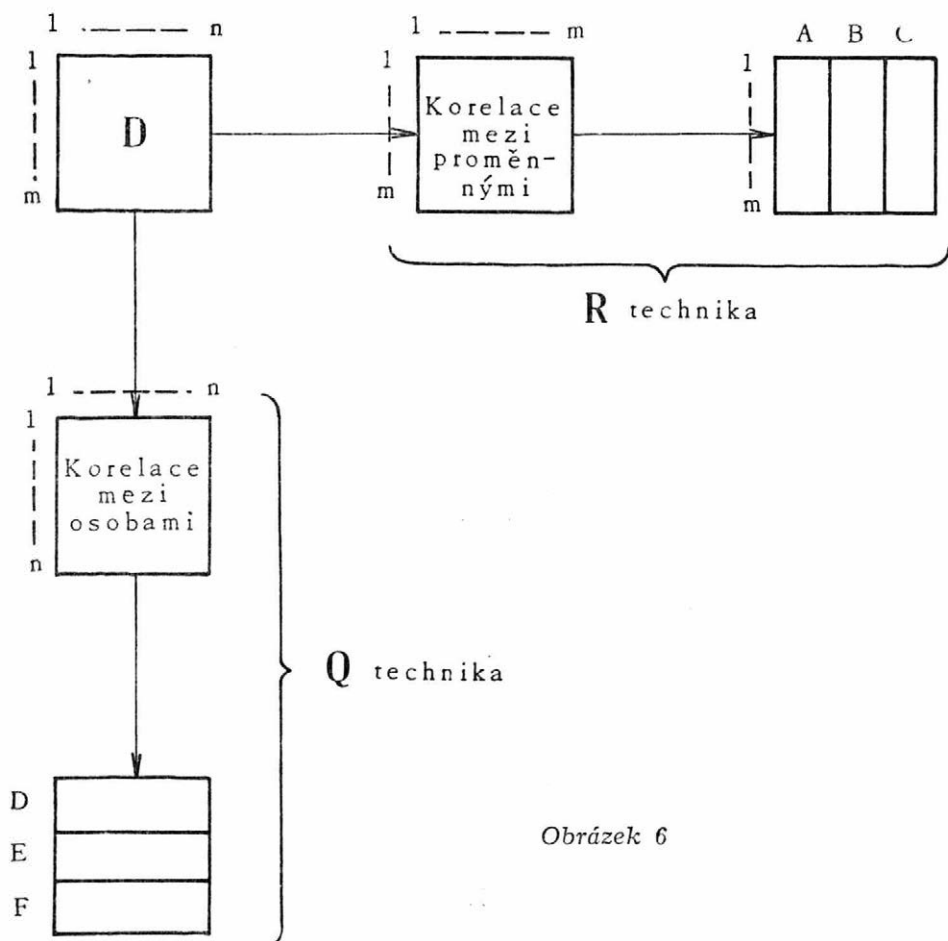
Na tomto konkrétním příkladě však je celkem jasně patrné, že výsledky faktorové analýzy sice vedou k určitým smysluplným a zajímavým výsledkům, zároveň však staví před badatele řadu dalších problémů, jejichž řešení vyžaduje značnou námahu.

VII. Použití faktorové analýzy

Badatelé aplikující faktorovou analýzu nejrůznějšího typu dospěli postupem doby k vytvoření určitých typických způsobů aplikace. Tyto typické formy aplikace faktorové analýzy se označují názvem techniky faktorové analýzy. V současné době se hovoří o tzv. R, Q, T, S a O technice faktorové analýzy. Většinou se pracuje však pouze s R a Q technikou. Způsob uspořádání dat při R technice a při Q technice je patrný z následujícího schématu.

V podstatě jde o toto: při tzv. technice faktorové analýzy mezi sebou navzájem koreluje jednotlivé testy. Tyto testy

Techniky typu P, T, S a O formalizují používání faktorové analýzy v časových řadách atd.



Obrázek 6

jsme aplikovali na poměrně značně velkém vzorku a výsledkem jsou faktory společné pro určité skupiny testů.

Při Q technice koreluje mezi sebou jednotlivé osoby ze zkoumané skupiny, respektive výsledky, kterých tyto osoby dosáhly v příslušné baterii testů. U každého jednotlivce známe výsledky z testů 1, 2 ... N, ($n = 15$) a pomocí korelační analýzy (prakticky vždy využívající Pearsonova koeficientu lineární korelace) zjišťujeme podobnost dosažených výsledků. Získané faktory tedy shrnují osoby s podobným a velmi podobným uspořádáním výsledků testů. Q techniku lze však s úspěchem aplikovat pouze na velmi malých vzorcích (do 50 osob). Q techniku lze nejlépe aplikovat za předpokladu, že používáme ipsativního měření.

Při technice Q se samozřejmě předpokládá (ostatně jako i u R techniky) že testy v baterii jsou v určité, ne právě zanedbatelné korelaci.

VIII. Závěr

Faktorová analýza může — za určitých podmínek — napomoci při řešení řady obtížných metodologických a metodických problémů sociologie. Nemůžeme však očekávat, že faktorová analýza vyřeší *veškeré* potíže a že je možno použít jí *vždy*. Použití faktorové analýzy tvoří vždy jen dosti malou část ve spektru nejrozličnějších metod analýzy, které bývají při výzkumu používány. Zapojení faktorové analýzy musí být vždy velmi pečlivě teoreticky zdůvodněno. Pouze za těchto podmínek

dává faktorová analýza výsledky teoreticky relevantní a prakticky použitelné. To ostatně platí i o jiných matematických postupech. Badatel usilující o využití faktorové analýzy si nadto musí uvědomit, že faktorová analýza s sebou přináší i řadu dalších problémů, které ne vždy jsou i po matematické stránce dostatečně precizně řešeny. Uvádíme zde pouze některé:

1. Závislost numerických výsledků faktorové analýzy na odhadu komunalit, použité metodě výpočtu faktorů i použité metodě rotace.

2. Praktická nemožnost porovnávat na základě objektivních statisticky formulovaných postupů výsledky faktorové analýzy pocházející z různých podsouborů.

3. Nedostatečné rozpracování faktorové analýzy jako statistické metody. Ve většině případů si badatel musí vždy napomáhat prostým úsudkem, nepodepřeným určitým statistickým kritériem.

Největší obtíže jsou však v samé podstatě interpretace. Faktorová analýza představuje vrcholně formalizovaný popis společenské reality, přičemž vztah mezi sledovanou sociální realitou a faktorovým popisem je již tak složitý, že chybí jakákoliv názornost. Tato skutečnost je permanentním zdrojem komunikačních obtíží mezi badatelem a čtenářstvem a dosti výrazně ovlivňuje možnosti použití faktorové analýzy v běžné a rutinní sociologické práci.

Domníváme se proto, že použití faktorové analýzy je v současné době vhodné především v dobře teoreticky fundovaných výzkumech, realizovaných týmem metodologicky a statisticky dobře vyzbrojených badatelů. Nelze však doporučit její použití pro běžný typ výzkumů. V každém případě však aplikace faktorové analýzy vyžaduje důkladné seznámení s matematickým aparátem použité metody a žádoucí je i průběžná konzultace se specialistou.

Literatura

- [1] Catell, R. B.: *Factor Analysis*. New York. Harper and Bros, 1952.
- [2] Hampejzová, O.: *Faktorová analýza. Matematická formulace problémů a metod.*

Československá psychologie. 6. 1962, 1. 424—445.

- [3] Harman, H. H.: *Modern Factor Analysis*. Chicago, The University of Chicago Press Chicago and London, 1960.
- [4] Horst, P.: *Factor Analysis of Data Matrices*. New York, Holt, Rinehart and Winston, 1965.
- [5] Lawley, D. N. and Maxwell, A. N.: *Factor Analysis as a Statistical Method*. London, Butterworth, 1963.
- [6] Thurstone, L. L.: *Multiple Factor Analysis*, Chicago, The University of Chicago Press, 1947.
- [7] Überla, K.: *Faktorenanalyse*. Springer — Verlag Berlin, Heidelberg, New York, 1968.

Резюме

Д. Холда - И. Лингарт: Факторный анализ

Статья дает сведения о проблемах применения факторного анализа в общественных науках, особенно в социологии.

В первой части авторы объясняют смысл факторного анализа, понятие общего, группового и единичного фактора и объясняют основную математическую модель факторного анализа и метод главных компонент.

$$x_i = \sum_{r=1}^k l_{ir}f_r + e_i (i = 1, 2, \dots, m)$$

модель факторного анализа

$$x_i = \sum_{r=1}^m w_{ir}z_r (i, r = 1, 2, \dots, m)$$

модель метода главных компонент

В статье дается объяснение различия между обоими приемами. В дальнейшем авторы сосредоточивают свое внимание только на факторном анализе и на проблеме так называемой ротации. Авторы анализируют отдельные этапы применения факторного анализа, отмечают специальные требования этого метода, уделяя одновременно большое внимание проблеме интерпретации. Кроме общего руководства по интерпретации авторы рассматривают конкретный пример факторного анализа 11 социально-экономических переменных. (Пример взят из архива исследования «Социальная дифференциация и мобильность в ЧССР 1967».)

В заключительной части статьи даются краткие сведения о факторах высшего порядка, о R, T, Q, S и P технике факторного анализа. Авторы предупреждают об опасности опрометчивого применения факторного анализа при исследовании и подчеркивают необходимость четкой, теоретической подготовки, необходимость комбинации факторного анализа с другими математическо-статистическими приемами и обращают внимание на затруднения, возникающие при записи результатов факторного анализа.

Summary

D. Holda — J. Linhart: Factor Analysis (A Non-mathematical Introduction to the Study of the Problem)

The paper deals with the problems of applying the factor analysis in social sciences and particularly in sociology.

In the first part, the authors elucidate the meaning of the factor analysis, the concepts of a common factor, a group factor and a unique factor, and explain the basic mathematical model of the factor analysis as well as the methods of the main components.

$$\mathbf{x}_i = \sum_{r=1}^k l_{ir} \mathbf{f}_r + \mathbf{e}_i (i = 1, 2, \dots, m)$$

the model of the factor analysis

$$\mathbf{x}_i = \sum_{r=1}^m w_{ir} \mathbf{z}_r (i, r = 1, 2, \dots, m)$$

the model of the method of the main components

The paper explains the difference between the two approaches. In their further explanation, the authors concentrate only on the factor analysis and on the problem of the so-called rotation. They analyse the separate stages of applying the factor analysis and respect the special claims of this method; they pay a good deal of attention to the problem of interpretation. Besides the general instruction concerning the interpretation, they analyse a concrete example of the factor analysis of eleven socio-economic variables. (The example is taken from the archives of the research into the social differentiation and mobility in Czechoslovakia, 1967.)

In conclusion, the paper shortly informs about the factors of higher orders and about the R, T, Q, S and P techniques of the factor analysis. The authors warn against a thoughtless application of the factor analysis in research and emphasize the necessity of an intensive theoretical preparation, the necessity of combining the factor analysis with other matematico-statistical procedures and point to the difficulties arising in presenting the results of the factor analysis.