Faktorová analýza

Osnova

Motivace

Formulace modelu faktorové analýzy

Vhodnost použití modelu faktorové analýzy

Odhad faktorové matice a její rotace

Volba počtu společných faktorů

Odhad faktorového skóre

Informace o testu studijních předpokladů

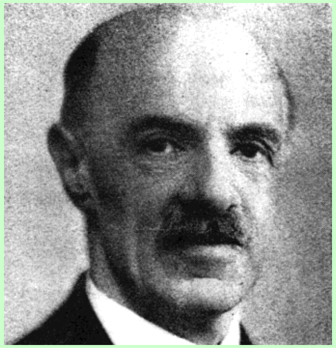
Výsledky faktorové analýzy a jejich psychologická interpretace

Závěr

Upozornění: Přednáška vychází z článku:

Budíková, Marie - Koutková, Helena - Dan, Jiří. Faktorová analýza testu studijních předpokladů na Masarykově univerzitě. In 6th International Conference Aplimat. Bratislava 2007.

Motivace: Za zakladatele faktorové analýzy je považován britský psycholog Ch. E. Spearman, o její další rozvoj se pak zasloužil americký psycholog L. L. Thurstone.



Charles Edward Spearman (1863 – 1945): Britský psycholog a statistik

Louis Leon Thurstone (1887 – 1955): Americký psycholog, jeden ze zakladatelů psychometrie

Faktorová analýza se původně používala jen v psychologii, s rozvojem výpočetní techniky však postupně pronikla i do dalších oborů.

Faktorová analýza má podobné cíle jako analýza hlavních komponent. Vycházíme opět z p-rozměrného datového souboru $\begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \cdots & \cdots & \cdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix}$. Např. předpokládáme, že každá z n osob je podrobena p různým testům. Výsledek, kterého i-tá osoba dosáhla v j-tém testu, ozn. y_{ij}. (Toto číslo se nazývá skór.) Předpokládáme, že výsledky testů lze vysvětlit pomocí určitého, nepříliš velkého počtu faktorů. Tyto faktory odpovídají nějakým lidským schopnostem, jako je např. paměť, schopnost soustředění, vytrvalost, nadání apod. Tyto faktory však nemůžeme přímo měřit.

Základní princip faktorové analýzy spočívá v tom, že každá z pozorovaných náhodných veličin Y_j ($j = 1, \dots, p$) může být vyjádřena jako součet lineární kombinace menšího počtu m nepozorovatelných (hypotetických) náhodných veličin $F_1, \dots, F_m$ - tzv. společných faktorů a dalšího zdroje variability E_j ($j = 1, \dots, p$) - tzv. specifické (reziduální) složky.

Faktorová analýza má tři cíle:

a) Analyzovat korelace většího počtu proměnných tím, že se více proměnných seskupí tak, aby většina proměnných v jednom shluku spolu silně korelovala, zatímco proměnné z různých shluků spolu nekorelují buď vůbec nebo jen velmi slabě.
b) Interpretovat faktory podle toho, jaké proměnné obsahuje příslušný shluk.
c) Shrnout variabilitu proměnných pomocí několika málo faktorů

Faktorová analýza musí zodpovědět čtyři otázky:

a) Kolik různých faktorů je zapotřebí k vysvětlení závislostí mezi proměnnými (tj. ke kvalitní reprodukci korelační matice)?
b) V jakém vztahu jsou faktory k jednotlivým proměnným?
c) Jak dobře faktory vysvětlují původní proměnné?
d) Jakou část jedinečné variability a jakou část náhodné variability obsahují původní proměnné?

Formulace modelu faktorové analýzy

Přepokládáme, že pro náhodný vektor $\mathbf{Y} = (Y_1, \dots, Y_p)'$ platí model

$$\mathbf{Y} = \mathbf{\Lambda F} + \mathbf{E}$$

kde $\mathbf{F} = (F_1, \dots, F_m)'$, $\mathbf{E} = (E_1, \dots, E_p)'$ a $\mathbf{\Lambda} = (\lambda_{jk})$ je neznámá reálná matice typu p x m.

Matice $\mathbf{\Lambda}$ se nazývá faktorová matice.

E_j je náhodná odchylka od přesného modelu příslušná k j-té veličině, j = 1,…, p.

Prvek λ_{jk} faktorové matice je faktorová váha (zátěž) k-tého společného faktoru příslušná k j-té veličině, k = 1, …, m.

Předpokládáme, že složky náhodného vektoru $\mathbf{Y}$ jsou standardizované, čímž je eliminován vliv jednotek.

Dále předpokládáme, že

náhodné vektory **F** a **E** mají nulové vektory středních hodnot,

náhodné vektory **F** a **E** jsou nekorelované,

$$\mathrm{var}(\mathbf{E}) = \begin{pmatrix} u_1^2 & 0 & \dots & 0 \\ 0 & u_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & u_p^2 \end{pmatrix}$$ (tj. náhodné odchylky jsou nekorelované),

$\mathrm{var}(\mathbf{F}) = \mathbf{I}$, kde $\mathbf{I}$ je jednotková matice řádu m (tj. faktory jsou nekorelované a mají jednotkové rozptyly).

V tomto případě hovoříme o ortogonálním faktorovém modelu.

Číslo u_j^2 se nazývá specifický rozptyl (unicita) j-té veličiny a číslo $h_j^2 = 1 - u_j^2$ se nazývá komunalita j-té veličiny, j = 1, …, p.

Vzhledem k předpokladům o ortogonálním faktorovém modelu platí:

$$\lambda_{jk} = R(Y_j, F_k) = C(Y_j, F_k), \quad j = 1, \dots, p, \; k = 1, \dots, m$$

$$\mathbf{\Lambda \Lambda}' = \mathrm{cor}(\mathbf{Y}) - \mathrm{var}(\mathbf{E}) = \mathrm{var}(\mathbf{Y}) - \mathrm{var}(\mathbf{E})$$

Matice $\mathbf{\Lambda \Lambda}'$ se nazývá redukovaná korelační matice. Tato matice je až na hlavní diagonálu shodná s maticí cor($\mathbf{Y}$). Její diagonální prvky jsou komunality h_j^2 veličiny Y_j, j = 1, …, p, přičemž $\sum_{k=1}^{m} \lambda_{jk}^2 = h_j^2$. Komunalita h_j^2 veličiny Y_j udává část rozptylu veličiny Y_j, která je vysvětlena působením společných faktorů, z čehož na k-tý faktor F_k připadá λ_{jk}^2.

Vhodnost použití modelu faktorové analýzy

K posouzení, zda vůbec má smysl provádět faktorovou analýzu (tj. zda korelace mezi veličinami $Y_1, \ldots, Y_p$ jsou vysvětlitelné pomocí jiných veličin $F_1, \ldots, F_m$) slouží Kaiserova – Meierova – Olkinova statistika (KMO statistika), která je založena na výběrových korelačních a parciálních korelačních koeficientech veličin $Y_1, \ldots, Y_p$. KMO nabývá hodnot mezi 0 a 1. Pro posouzení, zda má smysl provést faktorovou analýzu, můžeme použít následující tabulku:

KMO statistika	použití faktorové analýzy
0,90 – 1,00	vynikající
0,80 – 0,89	chvályhodné
0,70 – 0,79	středně užitečné
0,60 – 0,69	průměrné
0,50 – 0,59	špatné
0,00 – 0,49	nepřijatelné

Vedle KMO statistiky můžeme rovněž použít Bartlettův test sféricity, kde nulová hypotéza tvrdí, že výběrová korelační matice je matice jednotková. Testová statistika je dána vzorcem

$$\chi^2 = \frac{11 + 2p - 6n}{6} \ln|\mathbf{R}|.$$

Platí-li nulová hypotéza, testová statistika se asymptoticky řídí rozložením $\chi^2(p(p-1)/2)$. Nulovou hypotézu tedy zamítáme na asymptotické hladině významnosti α, když $\chi^2 \geq \chi^2_{1-\alpha}(p(p-1)/2)$. Nezamítneme-li nulovou hypotézu, neměli bychom faktorovou analýzu vůbec provádět.

Odhad faktorové matice a její rotace

Je-li formulován faktorový model, je zapotřebí pro daný počet faktorů m odhadnout faktorovou matici $\mathbf{\Lambda}$.

Z celé řady metod odhadu $\mathbf{\Lambda}$ zde naznačíme modifikovanou metodu hlavních komponent, která spočívá v aplikaci metody hlavních komponent na odhad redukované korelační matice $\mathbf{\Lambda\Lambda}'$.

Faktory vybíráme tak, aby vysvětlily co největší část celkového rozptylu $D(Y_1) + \ldots + D(Y_p) = p$ veličin $Y_1, \ldots, Y_p$.

Odhad matice $\mathbf{\Lambda}$ se pak provádí iteračním postupem:

Za počáteční odhad $\mathbf{T}_0$ matice $\mathbf{\Lambda\Lambda}'$ volíme realizaci výběrové korelační matice $\mathbf{R}$ vektoru $\mathbf{Y}$, tj. za počáteční odhad komunalit h_j^2 volíme jedničky.

Stanovíme m největších vlastních čísel $l_{1,0} \geq \ldots \geq l_{m,0}$ a jim odpovídajících jednotkových vlastních vektorů $\mathbf{v}_{1,0}, \ldots, \mathbf{v}_{m,0}$ matice $\mathbf{T}_0$. Za odhad $\hat{\mathbf{\Lambda}}_0$ matice $\mathbf{\Lambda}$ pak volíme matici $\hat{\mathbf{\Lambda}}_0 = (v_{1,0}\sqrt{l_{1,0}}, \ldots, v_{m,0}\sqrt{l_{m,0}})$.

Prvky hlavní diagonály matice $\hat{\mathbf{\Lambda}}_0\hat{\mathbf{\Lambda}}_0'$ jsou nové odhady komunalit, které v dalším kroku iteračního postupu dosadíme na diagonálu realizace matice $\mathbf{R}$, čímž zpřesníme odhad redukované korelační matice.

Přistoupíme k novému výpočtu vlastních čísel a vlastních vektorů.

Iterační postup zastavíme, pokud euklidovská norma rozdílu dvou po sobě následujících odhadů komunalit klesne pod předem zvolené malé kladné číslo, např. 0,001. Odhad $\hat{\mathbf{\Lambda}} = (\hat{\lambda}_{jk})$ faktorové matice $\mathbf{\Lambda} = (\lambda_{jk})$ pak vyjádříme ve tvaru

$$\hat{\mathbf{\Lambda}} = (\mathbf{v}_1\sqrt{l_1}, \ldots, \mathbf{v}_m\sqrt{l_m})$$

kde $l_1, \ldots, l_m$ jsou vlastní čísla redukované korelační matice získané v posledním kroku iteračního postupu seřazená sestupně a $\mathbf{v}_1, \ldots, \mathbf{v}_m$ jsou jim odpovídající jednotkové vlastní vektory. Potom

- $\hat{\lambda}_{jk}$ je odhad $R(Y_j, F_k)$,
- $\hat{\lambda}_{jk}$ je odhad příspěvku k-tého společného faktoru k rozptylu veličiny Y_j,
- součet čtverců prvků v j-tém řádku matice $\hat{\Lambda}$ je odhad komunality h_j^2, tj. odhaduje tu část rozptylu veličiny Y_j, kterou lze vysvětlit působením společných faktorů,
- součet čtverců prvků v k-tém sloupci matice $\hat{\Lambda}$ je roven vlastnímu číslu l_k matice $\hat{\Lambda}\hat{}$ a odhaduje příspěvek k-tého faktoru k celkovému rozptylu p. To znamená, že k-tý faktor vyčerpává $(l_k / p) \cdot 100\,\%$ celkového rozptylu, z čehož na j-tou veličinu připadá $(\lambda_{jk} / l_k) \cdot 100\,\%$.

Je-li počet faktorů m > 1 a současně m < p, není faktorová matice $\mathbf{\Lambda}$ vztahem $\mathbf{\Lambda}\,\mathbf{\Lambda}' = \mathrm{cor}(\mathbf{Y}) - \mathrm{var}(\mathbf{E})$ určena jednoznačně. Splňuje–li matice **U** řádu m podmínku $\mathbf{U}\mathbf{U}' = \mathbf{I}$, pak matice $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{U}$ rovněž tomuto vztahu vyhovuje. Říkáme, že matice $\mathbf{\Lambda}^*$ se získala rotací matice $\mathbf{\Lambda}$ a nazýváme ji rotovanou faktorovou maticí.

Jelikož odhadnuté faktorové váhy musí být snadno interpretovatelné, je cílem rotace získat co nejpřesvědčivější interpretaci faktorů. Požadujeme, aby každá z veličin $Y_1, \dots, Y_p$ měla vysoké faktorové váhy u co nejmenšího počtu faktorů a nízké či středně vysoké váhy u zbývajících faktorů.

Existuje několik metod rotace, často používaná je metoda varimax:

Odhady faktorových vah $\hat{\lambda}_{j1}, \dots, \hat{\lambda}_{jm}$ příslušné k j-té veličině chápeme nyní jako souřadnice bodu v R^m. Hlavní myšlenka této metody spočívá v tom, že se provádí transformace souřadných os v R^m, dokud souřadnice $\hat{\lambda}_{j1}, \dots, \hat{\lambda}_{jm}$ každé náhodné veličiny Y_j nejsou blízké buď 0 nebo ±1.

Volba počtu společných faktorů

Zbývá se vrátit k problematice volby počtu m společných faktorů $F_1, \dots, F_m$. V explorativní faktorové analýze nemusíme mít o tomto počtu jasnou představu. Existují různá doporučení, jak m zvolit. Pokud nelze přepokládat p-rozměrné normální rozložení náhodného vektoru **Y**, nemůžeme pro volbu m použít běžné statistické testy. Můžeme se ale opřít o několik různých kritérií:

- Kaiserovo kritérium - za m volíme počet těch vlastních čísel matice **R**, která jsou větší než 1.
- Sutinový test (scree test) – grafická metoda, která spočívá v subjektivním posouzení vzhledu sutinového grafu (scree plot), tj. grafu znázorňujícího velikosti sestupně uspořádaných vlastních čísel matice **R**. Objeví-li se v grafu určité zploštění, pak za m vezmeme to pořadové číslo, kde se zploštění projevilo.
- Kritérium založené na součtu prvních m největších vlastních čísel matice **R** – požadujeme, aby tento součet byl „přibližně“ p, tj. aby byl podíl celkového rozptylu pozorovaných veličin dostatečně vysvětlen příslušným faktorovým modelem.
- Kritérium založené na reziduální korelační matici po extrakci m společných faktorů, tj. matici $\mathbf{R} - \hat{\Lambda}\hat{\Lambda}'$. Požadujeme, aby všechny reziduální korelace (tj. mimodiagonální prvky reziduální korelační matice) byly „malé“, např. v absolutní hodnotě menší než 0,1. Pak m faktorů vysvětluje korelační matici pozorovaného vektoru dobře.

Odhad faktorového skóre

Faktorovou analýzu je vhodné doplnit o odhady hodnot faktorů F_1, …, F_m u jednotlivých výběrových objektů, které tyto objekty charakterizují – tzv. odhad faktorového skóre. Můžeme např. použít regresní metodu, kdy hledáme lineární regresní odhad $\widehat{F}_k$ veličiny F_k založený na vektoru **Y**. Lze ukázat, že

$$\hat{\boldsymbol{F}} = \boldsymbol{\Lambda}'\mathbf{R}^{-}\boldsymbol{Y}$$

kde $\hat{\boldsymbol{F}} = (\widehat{F}_1, \dots, \widehat{F}_m)'$ a $\boldsymbol{\Lambda}$ se prakticky nahradí odhadem $\hat{\boldsymbol{\Lambda}}$. Označme dále $\mathbf{y}_i = (y_{i1},\dots,y_{ip})'$ standardizovaný vektor pozorování vztahující se k i-tému objektu (i = 1,…, n). Potom pro vektor $\hat{\boldsymbol{f}}_i = (\widehat{F}_{i1}, \dots, \widehat{F}_{im})'$ odhadů faktorových skóre m společných faktorů dostáváme

$$\hat{\boldsymbol{f}}_i = \boldsymbol{\Lambda}'\mathbf{R}^{-}\boldsymbol{y}_i$$

Informace o testu studijních předpokladů

Od roku 2002 se na některých fakultách MU v Brně používá v přijímacím řízení test studijních předpokladů (dále jen TSP). Tento test zkoumá předpoklady uchazeče úspěšně studovat na MU. Je tvořen 80 otázkami, které mají podobu tzv. položek nucené volby. Ke každé z nich existuje jedna správná odpověď a čtyři chybné odpovědi, tzv. distraktory. Na řešení celého testu má uchazeč 80 minut. Počet správných odpovědí nabývá hodnot od 0 do 80. Položky jsou dle obsahu seřazeny do 8 subtestů po 10 položkách. Každý subtest je zaměřen na jinou oblast lidského myšlení. Zjišťuje se úroveň myšlení verbálního, numerického, symbolického, analytického, kritického a vědeckého. Položky jednoho ze subtestů slouží k posouzení prostorové představivosti, dalšího pak k ověření schopnosti usuzování.
Cílem faktorové analýzy je prozkoumat korelační strukturu vztahů mezi proměnnými, které obsahují počty správných odpovědí v jednotlivých subtestech a případně zvážit možnou redukci počtu subtestů.
Datový soubor obsahuje údaje o 2619 uchazečích o studium na Přírodovědecké fakultě MU v Brně v roce 2005, z toho je 1026 mužů, 1593 žen.
Pozorované náhodné veličiny jsou:

SBT1 – počet správných odpovědí v subtestu prostorová představivost,

SBT2 - počet správných odpovědí v subtestu symbolické myšlení,

SBT3 - počet správných odpovědí v subtestu verbální myšlení,

SBT4 - počet správných odpovědí v subtestu kritické myšlení,

SBT5 - počet správných odpovědí v subtestu numerické myšlení,

SBT6 - počet správných odpovědí v subtestu kulturní přehled,

SBT7 - počet správných odpovědí v subtestu úsudky,

SBT8 - počet správných odpovědí v subtestu analytické myšlení.

Výsledky faktorové analýzy a jejich psychologická interpretace

Při výpočtech byly použity statistické programové systémy SPSS a STATISTICA. Prvotní informace o korelační struktuře datového souboru získáme z realizace výběrové korelační matice uvažovaných osmi proměnných. Tu získáme např. takto: Statistiky – Vícerozměrné průzkumné techniky – Faktorová analýza – Proměnné SBT1 až SBT8 – OK – OK. Na záložce Popisné statistiky zvolíme Přehled korelací, průměrů, směrodatných odchylek – Korelace.

Proměnná	SBT1	SBT2	SBT3	SBT4	SBT5	SBT6	SBT7	SBT8
SBT1	1,00	0,28	0,26	0,14	0,29	0,16	0,21	0,22
SBT2	0,28	1,00	0,28	0,14	0,25	0,15	0,20	0,20
SBT3	0,26	0,28	1,00	0,24	0,27	0,29	0,34	0,30
SBT4	0,14	0,14	0,24	1,00	0,14	0,37	0,16	0,20
SBT5	0,29	0,25	0,27	0,14	1,00	0,19	0,31	0,29
SBT6	0,16	0,15	0,29	0,37	0,19	1,00	0,23	0,26
SBT7	0,21	0,20	0,34	0,16	0,31	0,23	1,00	0,36
SBT8	0,22	0,20	0,30	0,20	0,29	0,26	0,36	1,00

Vidíme, že korelace kolísají od 0,14 až po 0,37. Všechny korelace jsou poměrně nízké, i když některé jsou statisticky významné na hladině významnosti 0,05. Nad 0,3 (hranice pro středně silnou korelaci) jsou 4 korelační koeficienty. V korelační matici jsou tedy jakási 4 „těžiště“, můžeme tudíž očekávat, že k vysvětlení její struktury budeme potřebovat 4 faktory.

Pro posouzení, zda má vůbec smysl provádět faktorovou analýzu, použijeme KMO statistiku a provedeme Bartlettův test. Provedení testu v systému SPSS:
Analyze – Data Reduction – Factor - Variables sbt1 až sbt8 – Descriptives – zaškrtneme KMO and Bartlett's test of sphericity.

KMO and Bartlett's Test

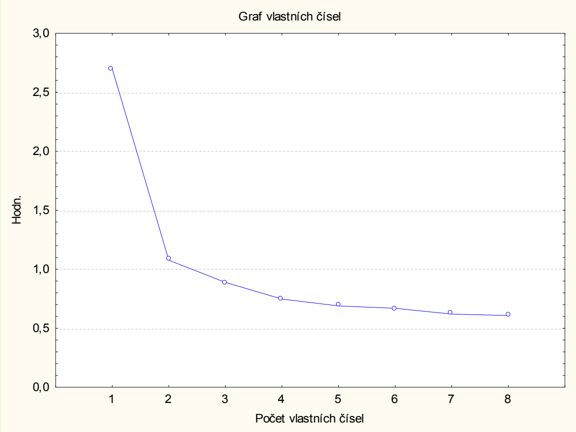
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,811
Bartlett's Test of Sphericity	Approx. Chi-Square	2851,192
	df	28
	Sig.	,000

Hodnota KMO statistiky je 0,81, tedy provedení faktorové analýzy se jeví jako chvályhodné. Testová statistka Bartlettova testu sféricity nabývá hodnoty 2851,19, počet stupňů volnosti je 28, odpovídající p-hodnota je velmi blízká 0, tedy hypotézu, že realizace výběrové korelační matice 8 uvažovaných proměnných je jednotková, zamítáme na asymptotické hladině významnosti 0,05.

Podívejme se nejprve na vlastní čísla realizované výběrové korelační matice **R** a na procento vysvětleného rozptylu:
Na záložce Základní nastavení změníme Max. počet faktorů na 8 a Min. vlastní číslo na 0,1 – OK – na záložce Výklad rozptylu zvolíme Vlastní čísla.

	Vl. čísla (TSP.sta) Extrakce: Hlavní komponenty			
Hodn.	vl. číslo	% celk. rozptylu	Kumulativ. vlast. číslo	Kumulativ. %
1	2,699881	33,74851	2,699881	33,7485
2	1,077192	13,46490	3,777073	47,2134
3	0,889012	11,11265	4,666085	58,3261
4	0,747268	9,34085	5,413353	67,6669
5	0,689322	8,61653	6,102675	76,2834
6	0,669118	8,36397	6,771793	84,6474
7	0,620088	7,75110	7,391881	92,3985
8	0,608119	7,60149	8,000000	100,0000

První faktor tedy vysvětluje 33,75% variability obsažené v osmi sledovaných proměnných, druhý 13,46% atd.
Vykreslíme sutinový graf:
Na záložce Výklad rozptylu zvolíme Sutinový graf.



Počet m společných faktorů zvolíme čtyři na základě sutinového grafu a na základě vysvětleného rozptylu, i když jsou pouze dvě vlastní čísla realizace matice **R** větší než 1. Celkové procento variability vysvětlené prvními čtyřmi faktory je 67,67%.

Pro extrakci faktorů zvolíme metodu hlavních komponent. Rotaci faktorů provedeme metodou varimax. Zajímají nás odhady komunalit.
Na záloze Základní nastavení zvolíme Max. počet faktorů 4 – OK. Na záložce Zákl. výsledky zvolíme Rotace faktorů Varimax prostý. Na záložce Výklad rozptylu zvolíme Komunality.

	Komunality (TSP.sta) Extrakce: Hlavní komponenty Rotace: Varimax pr.				
Proměnná	Z 1 faktoru	Z 2 faktorů	Z 3 faktorů	Z 4 faktorů	Více R^2
SBT1	0,054562	0,066543	0,067018	0,777656	0,160671
SBT2	0,805778	0,807290	0,810812	0,846872	0,148954
SBT3	0,281255	0,357992	0,572528	0,573581	0,242152
SBT4	0,003442	0,721249	0,721533	0,732025	0,162803
SBT5	0,001114	0,002672	0,200540	0,618109	0,200016
SBT6	0,006258	0,580718	0,635721	0,638016	0,207026
SBT7	0,017406	0,021387	0,656718	0,664304	0,227784
SBT8	0,002194	0,038095	0,534949	0,562789	0,217256

Např. odhad komunality proměnné SBT1 je 0,778, což lze interpretovat tak, že 77,8% variability proměnné SBT1 lze vysvětlit působením čtyř společných faktorů.

Nyní získáme odhad matice rotovaných faktorových zátěží: na záložce Zátěže zvolíme Shrnutí: Faktorové zátěže

Proměnná	Faktor. zátěže (Varimax pr.) (TSP.sta) Extrakce: Hlavní komponenty (Označené zatěže jsou >,500000) Faktor 1	Faktor 2	Faktor 3	Faktor 4
SBT1	0,233586	0,109455	0,021789	0,842994
SBT2	0,897652	0,038874	0,059354	0,189894
SBT3	0,530335	0,277013	0,463181	0,032454
SBT4	0,058672	0,847235	0,016849	0,102430
SBT5	0,033379	0,039475	0,444823	0,646196
SBT6	0,079106	0,757932	0,234526	0,047908
SBT7	0,131932	0,063094	0,797077	0,087099
SBT8	0,046837	0,189475	0,704879	0,166852
Výkl.roz	1,172010	1,423935	1,603874	1,213533
Prp.celk	0,146501	0,177992	0,200484	0,151692

Vidíme, že první faktor má vysoké korelace s proměnnými SBT2 a SBT3. Druhý faktor vysoce koreluje s proměnnými SBT4 a SBT6. U třetího faktoru pozorujeme vysoké korelace se SBT7 a SBT8. Čtvrtý faktor má vysoké korelace se SBT1 a SBT5. Po rotaci připadá na první faktor 1,172 celkového rozptylu, což je 14,7%. Na druhý faktor připadá 1,424 celkového rozptylu, tj. 17,8% atd.

Interpretace faktorů:

1. faktor … založen na symbolickém a verbálním myšlení

(Symbolické myšlení je založeno na operacích se symboly. Symbolem se rozumíme slovní, číselné nebo obrazné vyjádření určitého objektu nebo jevu.

Verbální myšlení je schopnost pochopit strukturu a zákonitosti výstavby jazyka jako celku a využívat ho při komunikaci s jinými členy dané jazykové skupiny).

2. faktor … založen na kritickém myšlení a kulturním přehledu.

(Kritické myšlení je schopnost posoudit správnost tvrzení a závěrů týkajících se poznatků obecné vzdělanosti.

Kulturní přehled je schopnost orientovat se v historických, politických a kulturních reáliích.)

3. faktor … založen na úsudcích a analytickém myšlení.

(Úsudek je schopnost vyvozování logických závěrů na základě daných faktů.

Analytické myšlení je schopnost myšlenkově rozkládat struktury na jejich konstitutivní prvky, schopnost vyvozovat logické spojitosti, schopnost nacházet strukturu ve zdánlivě chaotickém informačním poli, a to na základě faktů.)

4. faktor … založen na prostorové představivosti a numerickém myšlení.

(Prostorová představivost je schopnost myšlenkové orientace v prostoru.

Numerické myšlení je založeno na numerické představivosti, která se opírá zejména o schopnost vybavit si strukturu uspořádání množiny reálných čísel. Je to schopnost nacházet zákonitosti ve skupinách čísel.)

Kvalitu získaného faktorového modelu posoudíme též pomocí odhadnuté korelační a reziduální korelační matice. Na záložce Výklad rozptylu vybereme Reprod./rezid. korelace.

Proměnná	SBT1	SBT2	SBT3	SBT4	SBT5	SBT6	SBT7	SBT8
SBT1	0,78	0,38	0,19	0,19	0,57	0,15	0,13	0,19
SBT2	0,38	0,85	0,52	0,11	0,18	0,12	0,18	0,12
SBT3	0,19	0,52	0,57	0,28	0,26	0,36	0,46	0,41
SBT4	0,19	0,11	0,28	0,73	0,11	0,66	0,08	0,19
SBT5	0,57	0,18	0,26	0,11	0,62	0,17	0,42	0,43
SBT6	0,15	0,12	0,36	0,66	0,17	0,64	0,25	0,32
SBT7	0,13	0,18	0,46	0,08	0,42	0,25	0,66	0,59
SBT8	0,19	0,12	0,41	0,19	0,43	0,32	0,59	0,56

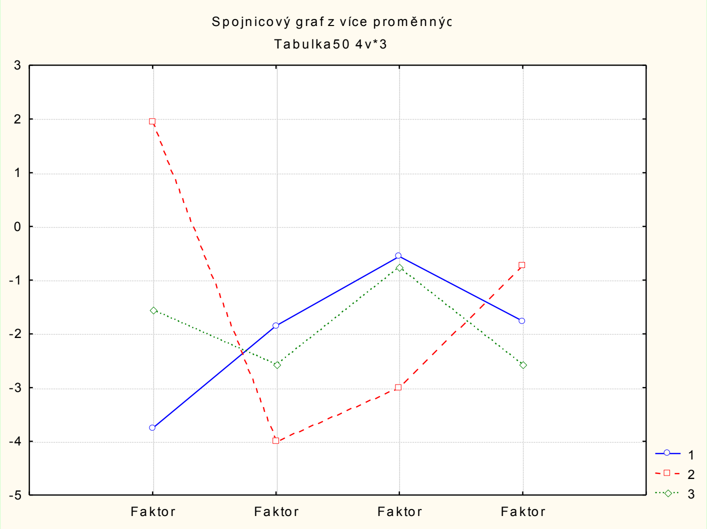
Proměnná	SBT1	SBT2	SBT3	SBT4	SBT5	SBT6	SBT7	SBT8
SBT1		-0,10	0,07	-0,06	-0,28	0,01	0,09	0,03
SBT2	-0,10		-0,24	0,03	0,07	0,03	0,01	0,08
SBT3	0,07	-0,24		-0,04	0,01	-0,07	-0,12	-0,11
SBT4	-0,06	0,03	-0,04		0,03	-0,29	0,07	0,01
SBT5	-0,28	0,07	0,01	0,03		0,02	-0,10	-0,14
SBT6	0,01	0,03	-0,07	-0,29	0,02		-0,02	-0,06
SBT7	0,09	0,01	-0,12	0,07	-0,10	-0,02		-0,23
SBT8	0,03	0,08	-0,11	0,01	-0,14	-0,06	-0,23	

Rezidua jsou vcelku malá, kolísají od –0,29 po 0,09. Osm z nich (tj. 29%) je v absolutní hodnotě větší než 0,1.

Pro několik prvních uchazečů uvedeme ještě odhad faktorového skóre. Na záložce Skóre vybereme Faktorová skóre.

	Faktor. skóre (TSP.sta) Rotace: Varimax pr. Extrakce: Hlavní komponenty			
Případ	Faktor 1	Faktor 2	Faktor 3	Faktor 4
1	-3,75715	-1,84968	-0,55849	-1,76290
2	1,93701	-3,99452	-3,00446	-0,73139
3	-1,55294	-2,57296	-0,74519	-2,56388
4	...	...	....	....

Na odhady faktorových skóre pro daný objekt můžeme pohlížet jako na souřadnice tohoto objektu v *m*-rozměrném prostoru. Spojnicový graf pro první tři uchazeče získáme takto: v pracovním sešitě v tabulce faktorových skóre vezmeme do bloku faktorová skóre prvních tří uchazečů. Klikneme pravým tlačítkem myši – Grafy bloku dat – Vlastní graf bloku podle sloupce – Spojnicové grafy (Profily případů) – OK.



Vidíme, že první uchazeč má vysoké skóre u třetího faktoru a naopak velmi nízké u prvního faktoru. Druhý uchazeč se vyznačuje vysokým skóre u prvního faktoru a velmi nízkým u druhého faktoru. Třetí uchazeč má vcelku vyrovnaná skóre u všech čtyř faktorů.

Závěr

Posoudíme kvalitu faktorového modelu z několika různých hledisek:

- KMO statistika je 0,81 (provedení faktorové analýzy se jeví jako chvályhodné).
- Testová statistika Bartlettova testu sféricity je 2851,19, počet stupňů volnosti je 28, p-hodnota velmi blízká 0 (korelační matice sledovaných 8 proměnných je s rizikem omylu nejvýše 5% různá od jednotkové matice).
- Odhad modelu založený na modifikované metodě hlavních komponent vysvětluje 67,67% variability obsažené v datovém souboru.
- Komunality (tj. podíly variability jednotlivých proměnných vysvětlené působením čtyř společných faktorů) kolísají od 56,3% u proměnné STB8 až po 84,7% u proměnné STB2.
- Reziduální korelační matice má malé prvky, osm z nich jsou v absolutní hodnotě větší než 0,1.

Zhodnotíme nalezené společné faktory:

Metoda hlavních komponent nalezla čtyři společné faktory stojící v pozadí osmi sledovaných proměnných. Všechny faktory jsou velmi dobře určeny lineárními kombinacemi vždy dvou proměnných.

Literatura

[1] Anděl, J.: Matematická statistika. SNTL/ALFA Praha, 1978.
[2] Hebák, P., Hustopecký, J.: Vícerozměrné statistické metody s aplikacemi. SNTL/Alfa, Praha 1987.
[3] Johnson, R. A., Wichern, D. W.: Applied Multivariate Statistical Analysis. Prentice Hall International, Inc. 1992.
[4] McDonald, R. P.: Faktorová analýza a příbuzné metody v psychologii. Academia Praha, 1991.
[5] SPSS/PC+ Statistics™ 4.0. SPSS Inc. 1990.
[6] STATISTICA for Windows. StatSoft, Inc. 2000.