

# Faktorová analýza jako známá neznámá (aneb metoda hlavních komponent a varimax nejsou vždy ideální postup)

PETR SOUKUP\*

Fakulta sociálních věd, Univerzita Karlova, Praha

## Factor Analysis as a Known Unknown – Principal Component Analysis with a Varimax Rotation Is Not Always the Ideal Approach

**Abstract:** This article seeks to show that, although factor analysis (mostly in its exploratory version) is a method frequently applied by social-science researchers (it is often also discussed in basic data analysis textbooks), only a very basic version of it is used, with settings that are far from optimal. However, what settings are used can have major implications, primarily in the form of conceptual problems, where the exploratory version is often used instead of the confirmatory version. Other settings used can also have an impact on the results. These are mainly partial options, which are used mainly in the exploratory version, in particular the choice of the correct correlation coefficients, the choice of method for the initial extraction of factors, the choice of the rotation method and the choice of the number of factors with which we want to work in the exploratory version. The text discusses the algorithms for ordinal variables, and the possibility of determining the number of factors through parallel analysis or MAP. The practical example discusses the advantages of the oblique rotation of factors. The article seeks to highlight good practices that best reflect the current state of the art of quantitative methodology and statistics. In addition to the general guidelines, the article contains practical advice about software and recommends a procedural schema for using factor analysis.

**Keywords:** factor analysis, parallel analysis, polychoric coefficient, oblique rotation

*Sociologický časopis / Czech Sociological Review, 2021, Vol. 57, No. 4: 455–484*  
<https://doi.org/10.13060/csr.2021.021>

Faktorová analýza je jednou z technik, které umožňují pracovat s nepřímými měřeními proměnnými, jejichž výskyt je v sociálních vědách poměrně častý (nejvíce v psychologii, pedagogice, ale řidké užití není ani v sociologii či politologii).

---

\* Veškerou korespondenci posílejte na adresu: PhDr. Ing. Petr Soukup, Ph.D., Institut sociologických studií, Fakulta sociálních věd Univerzity Karlovy, U Kříže 8, 158 00 Praha 5 – Jinonice, e-mail: petr.soukup@fsv.cuni.cz.

Základním smyslem užívání faktorové analýzy z datově analytického pohledu je náhrada mnoha měřených proměnných indikátorů několika málo faktory (minimálně jedním), odtud pramení název techniky. Formálně matematicky jde o soubor přístupů z oblasti vícerozměrné statistiky, který není jednoduchý (ruční výpočet většinou nepřichází díky iteračním postupům do úvahy), nadto i v posledních letech dochází k poměrně bouřlivému rozvoji této techniky (zejména jde o rozvoj nových odhadovacích algoritmů a softwaru). Běžně používaný komerční software však tyto novinky příliš nereflektuje a výzkumníci tak používají starší postupy, které jsou z dnešního pohledu již přežité. Cílem tohoto textu je na novější postupy poukázat a popsat ty z nich, které jsou v současné literatuře preferovány.

Text se snaží poukázat na postupy vhodné pro ordinální data, popisuje též různé odhadovací techniky a jejich použití. V závěru se pak zaměřuje i na specifické problémy explorační faktorové analýzy; určení počtu faktorů a jejich rotaci.

Před samotným představením jednotlivých komplikovanějších postupů a jejich rozbořem se krátce zaměříme na stručné představení techniky samotné, tj. faktorové analýzy. Čtenář, který má hlubší znalosti, může následující pasáž vynechat.

Jako první musíme odlišit *explorační* (převážně užívanou) a *konfirmační* (zatím stále málo užívanou) *faktorovou analýzu*.<sup>1</sup> Historicky starší model explorační faktorové analýzy (dále též jen „EFA“<sup>2</sup>) vznikl v psychologii, aby na základě jednotlivých manifestních (přímo měřených) indikátorů (proměnných, původně testových položek v testech inteligence) bylo možné dovodit latentní (skrytou) proměnnou (faktor), která způsobuje získané hodnoty jednotlivých manifestních indikátorů.<sup>3</sup> Z pohledu statistiky je faktorová analýza technika, kdy latentní proměnná (ta není v našich datech) je nezávisle proměnnou a sada jednotlivých manifestních (jsou v našich datech) proměnných je sadou závisle proměnných (srov. dále Obrázek 1). Fakticky jde tedy o sadu regresních analýz. Tyto závisle proměnné jsou vysvětlovány latentní proměnnou (faktorem), jinými slovy faktor zejména vysvětluje vztahy mezi manifestními proměnnými a ukazuje, co mají společné (například položky určitého testu/dotazníku měří dohromady inteligenci, depresivitu, proenvironmentální chování apod.). Explorační model faktorové analýzy se snaží na základě získaných manifestních proměnných dovodit, zda za nimi existuje jeden či více faktorů. Na základě našich dat skrze prozkoumání (exploraci) tak lze nabýt přesvědčení, že můžeme připustit existenci určitých

<sup>1</sup> Samozřejmě se zde liší jednotlivé disciplíny. Zejména v psychologii je užívání konfirmační faktorové analýzy poměrně časté, v jiných sociálních vědách tomu tak není. Je to jistě škoda a snad i tento text přispěje k rozšíření užívanosti této techniky.

<sup>2</sup> EFA = Exploratory Factor Analysis.

<sup>3</sup> Více se lze dočíst například v knihách Blahuše [1985], Urbánka [2000], praktický postup je v učebnici Rabušice, Soukupa a Mareše [2019: kapitola 15].

faktorů v pozadí. Z pohledu teorie a metodologie vědy slouží *explorační faktorová analýza k vytvoření teorie, resp. hypotéz* o určitých vztazích mezi proměnnými, resp. o existenci faktoru/faktorů v pozadí.

Druhou, mladší verzí faktorové analýzy je *konfirmační faktorová analýza* (dále též jen „CFA“<sup>4</sup>). Zde se nepokoušíme naše data prozkoumávat, ale vycházíme z již daných teorií, resp. hypotéz. A tyto *hypotézy o vzájemné struktuře proměnných a existenci faktoru/faktorů*, zpravidla označované jako model, *ověřujeme* (konfirmujeme). Výsledkem tak může být zamítnutí nebo nezamítnutí naší modelové představy. Z tohoto popisu je zřejmé zcela jiné načasování pro užití obou režimů faktorové analýzy ve výzkumu. Pokud se omezíme na proces vývoje a ověřování škál (velice častý zejména v pedagogice a psychologii, méně častý v sociologii), pak *explorační faktorová analýza* má své místo na počátku, aby nám naznačila možnou faktorovou strukturu (tj. existenci škály). Konfirmační faktorová analýza pak má své místo pro ověření této struktury a případně může ukázat, zda škála vyvinutá v jedné zemi (skupině) je použitelná i v jiné. Ostatně dost často toto odlišení výzkumníci nerespektují a používají pro všechny situace *explorační faktorovou analýzu*. Ta je ale značně subjektivní a umožňuje tak „prokázat“<sup>5</sup> téměř cokoli. Závěrem k základnímu představení EFA, resp. CFA uveďme, že oba přístupy vznikly původně jako procedury pro kardinální manifestní proměnné (vychází se z korelací mezi proměnnými dle Pearsona). Také faktor/faktory má/mají dle modelu faktorové analýzy charakter kardinální proměnné (tato proměnná původně sice není v datech, ale na základě EFA či CFA lze uložit její odhad, který označujeme jako faktorové skóre). S ohledem na skutečnost, že sociálněvědní data často nemají kardinální charakter (nejčastěji budou ordinální), začaly být vyvíjeny i postupy faktorové analýzy (EFA i CFA) pro ordinální či dichotomická data. Nutno podotknout, že ne všechny software tyto postupy obsahuje. Budeme se tomu více věnovat v samostatné části 3.

Pro úplnost ilustrujme rozdíl mezi *explorační* a *konfirmační* faktorovou analýzou jednoduchým Obrázkem 1 (tzv. diafragem).<sup>6</sup>

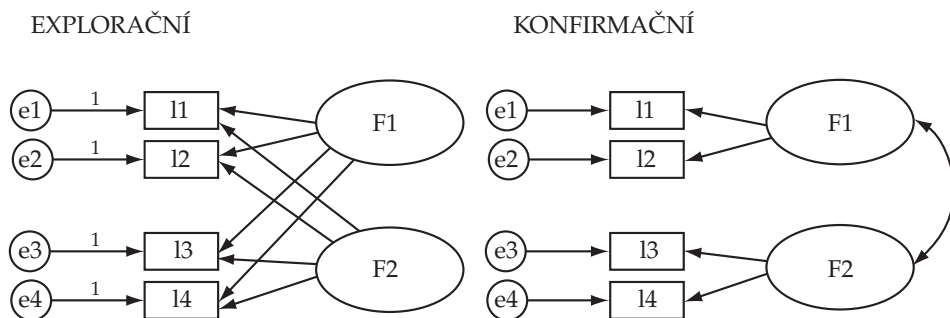
Na obrázku jsou vždy čtyři manifestní proměnné (v obdélnících, značené I1–I4) a dva faktory (značené jako F1, resp. F2). Levý panel Obrázku 1 ukazuje *explorační* režim, kdy se očekává, že mohou existovat vazby mezi všemi manifest-

<sup>4</sup> CFA = Confirmatory Factor Analysis.

<sup>5</sup> Výraz je v uvozovkách, protože samozřejmě *explorační* technika neumožňuje prokázat nic.

<sup>6</sup> Použitý obrázek je značně zjednodušený, má jen čtyři manifestní proměnné. V praxi tak jednoduchý model zpravidla nevyužijeme, jak z věcných důvodů, tak z důvodů statistických. Detailní výklad přesahuje možnosti tohoto článku, lze opět odkázat na literaturu uvedenou na počátku. Zde jen doplníme, že je vhodné používat pro každý faktor minimálně tři manifestní proměnné, a ne pouze dvě, jak je zobrazeno na zjednodušeném Obrázku 1.

**Obrázek 1. Schematické zachycení explorační a konfirmační faktorové analýzy**



ními proměnnými a všemi faktory. Připomeňme, že hodnotám těchto vazeb říkáme faktorové zátěže či faktorové náboje. Právě z hodnot těchto zátěží usuzujeme na interpretaci významu faktorů. Mohli bychom například zjistit (není zobrazeno v Obrázku 1), že díky silné vazbě I1 a I2 k prvnímu faktoru a obdobně silné vazbě I3 a I4 ke druhému faktoru mají položky I1, I2, resp. I3, I4 něco společného a tento společný rys bychom mohli pojmenovat (interpretovat).

U CFA (pravý panel) je základní situace jiná. Nic neprozkoumáváme<sup>7</sup> ani se nepokoušíme interpretovat. Dopředu formulujeme model (například na základě dřívějšího výzkumu a provedené EFA) a tento model jasně říká, že první faktor je tvořen I1 a I2, zatímco druhý je tvořen I3 a I4 (prvním faktorem může být například matematická gramotnost, měřená dvěma testovými položkami I1 a I2, a druhým gramotnost fyzikální, měřená položkami I3 a I4). Nadto model CFA na Obrázku 1 předpokládá, že naše faktory (např. matematická a fyzikální gramotnost) jsou vzájemně provázány (srov. obousměrnou šipku mezi faktory).

Po tomto stručném výkladu základních pojmů a koncepcí můžeme představit dále probíraná témata. Cílem textu je pojednat možnosti EFA a CFA, které by měl sociálněvědní výzkumník znát pro používání těchto technik v souladu se soudobými poznatky. Jednotlivé části textu formulujeme pracovně jako odpovědi na otázky:

1. Musím vždy užít faktorovou analýzu, resp. baterii otázek?
2. Mohu skrze EFA něco prokázat, otestovat, ověřit? A mohu pro stejná data použít EFA i CFA, tj. mohu na stejných datech nejdříve vytvořit teorii a poté ji ověřit?
3. Jak naložit s EFA či CFA pro data, která nejsou kardinální?

<sup>7</sup> Pro úplnost dodejme, že i zde existuje jistá možnost prozkoumávání a úprav (zejména skrze modifikační indexy a rezidua), nicméně těmito postupy se již vykokročí z čistě konfirmačního rámce.

4. Je pro stanovení počtu faktorů ideální Kaiserovo pravidlo, případně Cattellův suťový graf (scree plot)?
5. Jaké jsou vhodné postupy pro prvotní řešení a pro rotované řešení v rámci EFA?

Při rozboru většiny otázek často vystačíme se slovními vyjádřeními, nicméně pro otázky 3–5 využijeme i výpočetní ilustrace, aby došlo k jasné ukázce problémů, které působí, když nedodržíme postupy doporučované v soudobé literatuře. Mnohé odpovědi na otázky 1–3 již naznačily úvodní výklady, zde bude tedy rozbor velice stručný.

Než přikročíme k odpovědím na výše uvedené otázky, budeme prezentovat menší výzkumnou sondu. Tato sonda ukáže, jak se reálně uplatňují postupy v rámci explorační a konfirmační faktorové analýzy v publikovaných empirických statích. Výzkumná sonda byla zaměřena na dva přední české časopisy: *Československou psychologii* a *Sociologický časopis / Czech Sociological Review* (dále jen *Sociologický časopis*). S ohledem na výsledky obsahových analýz těchto dvou časopisů provedených Soukupem [2019] byly v *Sociologickém časopise* hledány aplikace explorační faktorové analýzy a v *Československé psychologii* aplikace konfirmační faktorové analýzy. Již tento rozdíl může být zářezující, tj. sociologové téměř výhradně užívají explorační model<sup>8</sup>, psychologové konfirmační. Bylo nalezeno posledních pět použití těchto technik v obou časopisech (stav k roku 2019), s ohledem na mnohem častější užívání faktorové analýzy psychology pocházejí aplikace z posledních let (2016–2019), v případě *Sociologického časopisu* pak bylo třeba jít zpětně až k roku 2008. Kategorie, které byly sledovány u explorační verze faktorové analýzy, byly následující: použité stupnice vstupních proměnných, kritérium pro stanovení počtu faktorů, použitá odhadovací technika pro prvotní řešení a použitá rotace. V případě konfirmační faktorové analýzy šlo o analogické kategorie samozřejmě přizpůsobené této technice, konkrétně: použité stupnice vstupních proměnných, typ použité korelace pro výpočet, použitá odhadovací technika a zdůvodnění modelu (provázanost faktorů a často též pro vazby mezi chybovými složkami).

Následující dvě tabulky (Tabulka 1 a 2) souhrnně popisují výsledek obsahové analýzy článků dle uvedených kategorií, v dalším textu je pak na tyto výsledky opakovaně odkazováno.

### Ad 1) Musím vždy užít faktorovou analýzu, resp. baterii otázek?

Občas se diskutuje o tom, že zejména explorační faktorová analýza je nadužívaná a dotazníky jsou nesmírně dlouhé, protože obsahují nekonečné baterie otázek (nejsou výjimkou ani dotazníky o 400 položkách). V mnoha případech totiž panuje přesvědčení, že musíme za každou cenu použít baterii otázek v dotazníku

<sup>8</sup> Navíc v posledních letech tyto analýzy jako výsledky nepublikují ve statích.

**Tabulka 1. Příklady užití explorační faktorové analýzy na stránkách Sociologického časopisu a popis používaných postupů**

Rok publikace	Autor	Technika extrakce	Technika rotace	Zdůvodnění počtu faktorů	Další uplatněné postupy
2008	Hamplová	PCA	varimax	počet faktorů nezdůvodněn	potlačování zátěží
2009	Vráblíková	PCA	oblmin	počet faktorů nezdůvodněn	binární položky
2011	Luthar, Koprivnik	PA	oblmin	počet faktorů částečně zdůvodněn, neužito žádného kritéria	
2011	Bernard	PCA	varimax	počet faktorů nezdůvodněn	
2012	Vidovičová, Petrová, Kafková	PCA	varimax	počet faktorů nezdůvodněn	potlačování zátěží, ordinální stupnice proměnných

Poznámka: PCA – Principal Component Analysis; PA – Principal Axis Analysis.

**Tabulka 2. Příklady užití konfirmační faktorové analýzy na stránkách Československé psychologie a popis používaných postupů**

Rok publikace	Autor	Odhadovací technika	Typ korelací	Stupnice proměnných (software)
2016	Miklošová, Makulajová, Jakubek	neuvedeno	neuvedeno	dichotomické položky (SW AMOS)
2017	Gocet Tekin	neuvedeno	neuvedeno	4bodové stupnice (SW AMOS)
2018	Tišanská, Kožený	WLSMV	polychor	1–5bodové stupnice (Mplus)
2019	Köverová	neuvedeno	neuvedeno	5bodové stupnice (SW AMOS)
2019	Boberová, Ropovik, Kolarčík, Madarasová Gecková, Paakkari	WLSMV	polychor	4bodové stupnice (R-lavaan)

Poznámka: WLSMV – Weighted Least Square Mean and Variance Adjusted.

pro zjišťování fenoménů, které sledujeme. Poté, co takovou baterií disponujeme, máme užít faktorovou analýzu, abychom ověřovali, že naše baterie měří náš fenomén (více v rozboru otázky č. 2). Problémem nejsou jen dotazníky plné baterií otázek, ale často bezhlavé přejímání těchto baterií bez ověření, ideálně provedeného skrze CFA. Problémem bývá též následné zpracovávání skrze několik exploračních faktorových analýz, často bez dalších analytických postupů. Nyní je třeba nalézt odpověď na první položenou otázku. Uvedme, že je krajně nevhodné, abychom užívali bez dalšího námi ad hoc vytvořených baterií otázek, stejně tak bez dalšího přejímali nevyzkoušené nástroje ze zahraničí. Pokud vymyslíme vlastní baterii a nepoužijeme standardní postupy známé z psychometrie, resp. edukometrie, je téměř jisté, že námi vytvořený nástroj bude špatný. Nepomůže mu ani následně provedená explorační faktorová analýza (její správné místo v procesu vývoje škál jsme popsali na počátku článku, detailně je popisuje např. DeVellis [2003: 133–135]). Obdobně nevhodné je přeložit zahraniční baterii a slepě doufat, že bude v našich poměrech fungovat stejně jako v zahraničí. Vývoj nové baterie (škály), případně její národní adaptace, je složitý proces a zabere v lepším případě několik měsíců. V mnoha situacích je proto vhodné, pokud potřebujeme rychle zjišťovat nějaký fenomén, vytvořit pouze jednu otázku (či sadu nesouvisejících otázek), u které (kterých) nám je poměrně jasné, co měří. Získáme sice nástroj, jehož validita a reliabilita nejsou nijak vysoké<sup>9</sup>, ale bude to postup poměrně rychlý, nadto i pro analytické zpracování jednoduchý. Samozřejmě si uvědomujeme, že některé fenomény měřit přímo nelze a výše uvedený postup nelze užít. Ještě jeden argument svědčí ve prospěch preference jednotlivých otázek oproti baterii otázek. Dotazník by kromě zahrnutí více fenoménů měl být též poměrně krátký (délka dotazníku ovlivňuje jak odmítnutí jeho vyplnění, tak také výpadek v průběhu samotného dotazování). Pokud budeme pro mnoho fenoménů v našem dotazníku užívat baterie, stane se dotazování nerealisticky dlouhým. Je proto namísto užít baterie (s výše uvedenými doporučeními) jen pro ty fenomény, které jsou pro nás v rámci výzkumu nejdůležitější. Pro ostatní fenomény pak máme užít jednotlivé otázky, případně velice krátké baterie.

**Ad 2) Mohu skrze EFA něco prokázat, otestovat, ověřit? A mohu pro stejná data použít EFA i CFA, tj. mohu na stejných datech nejdříve vytvořit teorii a poté ji ověřit?**

Velice často se při posuzování článků, ale i při čtení již publikovaných studií setkáváme s představou, že explorační faktorovou analýzou je možné ověřit škálu, prokázat, že proměnné tvoří jednu či více škál, apod. Obdobně není ojedinělé, že autoři na zcela stejných datech užijí EFA i CFA. Dodejme, že zejména první část otázky navazuje na tvrzení, která se v minulosti objevovala i v mnoha učeb-

<sup>9</sup> Reliabilitu nadto neumíme na rozdíl od škál ani odhadnout.



nicích. Z pohledu dnešních poznatků psychometrie a edukometrie je tento postup ale nepřijatelný. Náhledem do standardních knih [např. Brown 2006: 1–2; Netemeyer, Bearden, Sharma 2003: 36] a na požadavky mnoha časopisů je vše zcela jasné. Jak bylo uvedeno v exkurzu na počátku, explorační faktorová analýza slouží k exploraci, vytvoření představy o souvislostech mezi proměnnými v našich datech. Pokud chceme tuto představu ověřit, užijeme konfirmační variantu faktorové analýzy. Proti užití explorační faktorové analýzy pro ověřování svědčí i několik dalších okolností. V rámci explorační faktorové analýzy je možné provádět různé subjektivní volby (metody extrakce faktorů, rotační techniky, počtu faktorů apod., srov. dále). Díky těmto subjektivním volbám je možné ovlivnit výsledky a dokázat v zásadě cokoli (srov. zejména dále otázky č. 4 a 5). Bohužel bez formálního testu či ověřitelného kritéria, protože těmi explorační faktorová analýza na rozdíl od konfirmační varianty nedisponuje. Samozřejmě je namístě uvést, že zřejmě nejoblíbenější sociálněvědní software SPSS konfirmační faktorovou analýzu nenabízí (explorační naopak nabízí), pro výzkumníka je tak poměrně obtížné tento postup aplikovat. Lze zde doporučit například freeware JASP, který je SPSS velice podobný a CFA (kromě EFA) nabízí skrze přehledné nabídky.

Zmínku si ještě zaslouží postup, kdy pro stejná data užijeme explorační a poté konfirmační faktorovou analýzu. Tento postup je nutné odmítnout v základním režimu, tj. na zcela stejné případy uplatníme EFA a CFA. I když jde z pohledu analytika o postup více než vhodný, protože pokud explorační analýza naznačí vhodné řešení, je téměř jisté, že jej konfirmační postup potvrdí. To ale není empirický důkaz, jde o jinou verzi tautologie. Doplňme, že takto ověřený nástroj zpravidla nebude dobrý a další výzkumy jeho fungování nepotvrdí. Můžeme opět odkázat na již citovanou knihu DeVellis [2003: 133–135], která jasně popisuje roli explorační a konfirmační faktorové analýzy. Explorační analýza má své místo na počátku při vývoji škály (ať už se jedná o vývoj zcela nové škály nebo adaptaci stávající užívané škály), konfirmační faktorová analýza má pak místo později při jejím ověřování.

Doporučit lze ovšem postupy, kdy rozdělíme náš soubor na dvě části (např. poloviny) a na jedné spočítáme EFA a toto řešení ověříme na druhé polovině souboru skrze CFA. Analogicky lze v případě, že náš soubor je menší, používat postupů z oblasti resamplingu, například bootstrap.<sup>10</sup> Poměrně zajímavý postup zvolili autoři článku Vlčková et al. [2019], kdy na stejných datech provedli konfirmaci mezinárodně užívané baterie a kromě toho skrze explorační přístup vytvořili národní adaptaci mezinárodní baterie. Sice tak na stejných datech užili EFA a CFA, ale nikoli sériově (následně), ale paralelně (vedle sebe) se zcela odlišným cílem. Takový postup je též vhodný a lze jej považovat za jeden z možných příkladů dobré praxe.

---

<sup>10</sup> Za toto doporučení patří poděkování anonymnímu recenzentovi článku.



### Ad 3) Jak naložit s EFA či CFA pro data, která nejsou kardinální?

Občas se při četbě textů o faktorové analýze dozvíme, že tyto techniky jsou určeny jen pro kardinální proměnné. To by samozřejmě při extrémním výkladu znamenalo, že v sociálních vědách nebude mít faktorová analýza místo. Opačným extrémem je často praxe výzkumníků, kteří užívají EFA či CFA v zásadě pro jakákoli data (vč. dichotomických), tj. bez zohlednění charakteru dat. Ostatně v Tabulce 1 a 2 lze nalézt příklady těchto aplikací. Oba extrémní názory nejsou zcela správné. Zkusme uvést základní důvody a vše zpřesnit. Před dalším rozbořením třetí otázky poznamenejme, že základní logika výpočtu faktorové analýzy (nerozhoduje, zda explorační, či konfirmační) je založena na tom, že se nejdříve spočítají korelace (či kovariance) mezi proměnnými a z nich se poté vypočte řešení faktorové analýzy. Právě proto, že základním korelačním koeficientem je Pearsonův korelační koeficient, převažuje názor, že jiné proměnné než kardinální se pro tyto procedury nehodí. Nadto je ještě někdy vyžadována normalita našich proměnných, její požadavek plyne z užívání metody maximální věrohodnosti pro odhad faktorového řešení (zejména v konfirmačním režimu, není ale vyloučena v explorační verzi, srov. Tabulku 3). Zde je namístě několik vysvětlení. Pokud chceme pracovat s běžným nastavením faktorové analýzy (typicky založené na korelacích dle Pearsona), pak bychom skutečně měli mít data kardinální, případně delší ordinální stupnice (simulační studie ukazují, že již cca pětibodové stupnice se svým chováním blíží kardinálním proměnným [Rhetmulla, Brosseau-Liard, Savalei 2012]). Pokud ale naše data mají krátké ordinální stupnice nebo jsou dichotomická, není používání Pearsonova korelačního koeficientu, případně odhadu skrze metodu maximální věrohodnosti zcela vhodné. Probereme dopady, které toto použití může mít. Pokud užijeme Pearsonovu korelaci pro kratší ordinální stupnice, dojde často k podhodnocení odhadu korelace (srov. dále popis tetrachorických korelací a předpokladu jejich užití, tj. ordinální proměnná je vnímána jako proměnná vzniklá „rozsekáním“ původně kardinální proměnné). Pokud užijeme Pearsonovu korelaci pro ordinální data, dobrovolně se často vzdáváme „lepších“ výsledků a dosahujeme výsledků „horších“. To u EFA znamená, že faktorové zátěže budou menší, míra vysvětleného rozptylu bude menší, počet faktorů bude nejspíše větší než správný. U CFA jsou dopady obdobné s tím, že celkové hodnocení modelu často vyhodnocované skrze souhrnné charakteristiky, jako je CFI, TLI, AGFI apod., bude svědčit o tom, že model není vhodný, a my budeme mít tendenci ho zamítat a hledat jiný, typicky po modifikaci složitější. Obojí dopady jsou nežádoucí a zkusme tedy popsat možné alternativy vhodné pro různé typy proměnných. Situaci přehledně<sup>11</sup> shrnuje Tabulka 3.

<sup>11</sup> Bez dalšího výkladu musíme přiznat, že tabulka (i když vypadá složitě) je značným zjednodušením. Odhadovacích technik je mnohem více, byly tedy vybrány ty nejužívanější, vhodné pro jednotlivé typy dat. Dále je ignorováno, že různé techniky se lépe či hůře hodí pro různé vzorce chybějících dat či pro různé složité faktorové modely. Poslední, co

**Tabulka 3. Přehled typu proměnných a příslušných koeficientů a metod odhadu EFA a CFA**

Typ proměnných	Užívaný korelační koeficient	Vhodná metoda odhadu EFA	Vhodná metoda odhadu CFA
Kardinální, dlouhé ordinální stupnice	Pearsonův	PCA, PA, ML*	ML*, ADF (WLS)
Kratší ordinální stupnice	polychorický	PCA, PA	WLS, WLSMV
Dichotomické proměnné	tetrachorický	PCA, PA	WLS, WLSMV

*Zdroj: Vlastní přehled založený na článcích Baglin [2014], Flora, Curran [2004] a knize Brown [2006].*

Poznámka: PCA – Principal Component Analysis (metoda hlavních komponent); PA – Principal Axis Analysis (metoda hlavních os); ML – Maximum Likelihood (metoda maximální věrohodnosti); ADF (WLS) – Asymptotic Distribution Free (asymptotická metoda bez předpokladu normálního rozdělení); WLSMV – Diagonally Weighted Least Squares (diagonálně vážená metoda nejmenších čtverců).

\* Vhodná, pokud naše proměnné mají normální rozdělení.

Stručně tedy okomentujme Tabulku 3. Je z ní patrné, že pokud mají naše proměnné kratší ordinální stupnice, máme místo Pearsonových korelací spíše užívat korelace polychorické, pro binární data pak korelace tetrachorické. Dodejme, že tyto koeficienty nejsou českému čtenáři neznámé, tetrachorický pojednává například kniha Chrásky [2006: 147]. Logika polychorického koeficientu je následující. Jeho použití předpokládá, že naše ordinální stupnice vznikla (umělym) rozdělením spojitě normálně rozdělené veličiny do několika málo kategorií. Polychorický korelační koeficient se pak snaží o odhad vztahu původních spojitých veličin, z nichž dle předpokladu vznikly naše dvě ordinální proměnné. Za jinak stejných okolností (tj. pro stejná data, která navíc nejsou výrazně sešikmená) platí, že polychorický koeficient bude vyšší než Pearsonův korelační koeficient. Připomeňme, že jeho nepoužitím se tak připravujeme o výsledky, které budou „lepší“. Dopředu již ale vyslovme varování, že výpočty polychorických korelací jsou někdy problematické (srov. dále), a tak nelze brát doporučení pro jejich užívání zcela absolutně.

Doporučené metody odhadu pro EFA či CFA, prezentované v Tabulce 1, vycházejí z textů (přehled literatury je pod tabulkou), které provádějí simula-

---

přehled v tabulce ignoruje, je existence bayesovských odhadů, které jsou velice slibné a jejich vlastnosti jsou přinejmenším srovnatelné s nejlepšími klasickými technikami. Nicméně jejich rozšíření v softwaru je zatím poměrně malé a znalost bayesovské statistiky nízká, jejich zařazení jsme tedy nepovažovali za vhodné. Zájemce o vhléd do tématu lze v češtině odkázat na článek Mazance a Dvořákové [2017].

ce pro různé typy proměnných. Více se odhadovacím metodám EFA věnujeme v páté části, proto je zde ponecháme bez výkladu. Zaměříme se krátce na srovnání odhadovacích technik pro CFA s ordinálními daty. Několik simulačních studií se snažilo zjistit, které algoritmy jsou v této situaci nejlepší. Tým Rhetmulla, Brosseau-Liard, Savalej [2012] ukázal, že cca od úrovně dichotomií do úrovně pětibodových stupnic je kategoriální odhadovací algoritmus lepší než algoritmus ML, konkrétně WLSMV poskytuje lepší odhady faktorových zátěží i jejich standardních chyb. V textu týmu Rhetmully [ibid.: 371] je zejména zdůrazněno, že ML není jako odhadovací technika vhodná pro dichotomická data. Li [2016a] při porovnání ML a WLSMV (v jeho textu označované jako DWLS) jasně prokázal, že WLSMV je pro ordinální data vhodnější.<sup>12</sup> Dle výsledků Li [2016a] má ML technika zejména sklon podhodnocovat hodnoty faktorových zátěží a její fungování v menších souborech (v řádech stovek jednotek) má i další negativní dopady. Li [2016b] provedl také srovnání algoritmu WLSMV s algoritmem označovaným jako MLR, tj. metodou maximální věrohodnosti s robustním odhadem standardních chyb (tato technika se někdy doporučuje pro nenormálně rozdělené veličiny). I zde Li [ibid.] prokázal, že WLSMV předčí algoritmus MLR, zejména lépe odhaduje hodnoty faktorových zátěží. Poměrně zajímavým výsledkem srovnání je též skutečnost, že MLR pro menší soubory a nenormálně rozdělené latentní proměnné lépe odhaduje korelace mezi těmito latentními proměnnými pro model CFA s více faktory<sup>13</sup> [ibid.: 947]. Ostatně tento výsledek prezentoval Li už ve svém předchozím textu [Li 2016a: 384], kde se objevuje doporučení, aby v případě, kdy výzkumníky nezajímají jednotlivé části CFA modelu, ale jen souvislost (vazba) mezi faktory a nemají software, který umožňuje užít odhad skrze WLSMV, užívali robustní verzi ML algoritmu (MLR). Nutno podotknout, že ale většinou buď software nabízí jak MLR, tak WLSMV, nebo ani jednu možnost. Tato rada je tedy diskutabilní. Lze ji ale aplikovat, pokud narazíme na výpočetní komplikace při užití WLSMV (srov. dále ilustraci algoritmu WLSMV). Uvedme, že náhledem do Tabulky 2 snadno zjistíme, že publikační praxe výše uvedeným doporučením ne vždy odpovídá, velmi oblíbený software AMOS totiž odhadovací procedury pro kratší stupnice nenabízí, a přesto je v těchto situacích pro výpočty využíván (naopak R v balíčku lavaan či Mplus tyto možnosti poskytují).

Namísto je několik praktických poznámek k možným výpočtům. I když čtenář snadno najde vzorce pro tetrachorické či polychorické korelace, nebude zřejmě chtít tyto koeficienty ručně počítat. Je tedy třeba krátce pojednat možnosti softwaru pro EFA (či CFA) pro různé typy proměnných. Začneme s EFA, jde o převládající užití faktorové analýzy. V SPSS (zřejmě nejužívanější platformě) přímé procedury pro ordinální či dichotomické proměnné nenajdeme (existují

<sup>12</sup> Li [2016a] navíc ukazuje, že techniku ML předčí i zcela jednoduchá technika nevážených nejmenších čtverců, označovaná jako ULS.

<sup>13</sup> Li [2016b] ve svých simulacích užívá data se dvěma faktory.

jejich alternativy, srov. dále). Proto máme dvě možnosti: použít více procedur v SPSS postupně nebo užít jiný software. Jak by tedy měl vypadat postup v SPSS:

1. Výpočet polychorických korelačních koeficientů.<sup>14</sup> Výpočet je možný buď ruční, nebo (lépe) skrze využití propojení SPSS s R<sup>15</sup> v rámci procedury *Analyze-Correlate-Heterogenous Correlation*.
2. Výsledné korelační koeficienty musíme vložit do datové matice ve formě korelační matice (prostým nakopírováním celé výstupní tabulky).
3. Z této korelační matice skrze příkaz (nelze přes SPSS menu) vypočítat faktorovou analýzu dle našich požadavků (typicky buď PCA, nebo PA metodou).

Jinou možností je užívat software, který tyto postupy umí přímo. Z volně dostupných bude vše umět software R, který má ale komplikované ovládání skrze specifický příkazový jazyk. Jako vhodná alternativa se jeví software FACTOR [Baglin 2014], který má snadné ovládání a je zdarma. Samozřejmě existuje mnoho komerčních produktů, které toto zvládnou.

Pro konfirmační režim faktorové analýzy bude třeba užít některý z produktů, který umí pracovat s ordinálními či dichotomickými daty. Patří mezi ně komerční EQS, LISREL a již zmíněný MPlus. Zájemce o bezplatné řešení pak sáhne nejspíše k R (zřejmě nejlepší je již zmíněný balíček lavaan). Zajímavou možností je ovládání R skrze freeware JASP (srov. jeho ukázkou v části čtvrté).<sup>16</sup>

Namísto je ovšem i několik praktických varování. Pro odhad tetrachorických koeficientů se užívají různé algoritmy, a tak mohou dva výzkumníci užívající například dva odlišné balíčky v R získat jiné výsledky. Dalším praktickým problémem je výpočet tetrachorických korelačních koeficientů pro mnoho proměnných. V řádu několika málo desítek proměnných již často dochází k tomu, že výsledná korelační matice není pozitivně definitní<sup>17</sup>, a proto nezískáme řešení našeho problému. Jako možné řešení (plyne z opakované vlastní zkušenosti autora) je zřejmě nejlepší vyřadit některé z položek (typicky ty s malými hodnotami korelací s ostatními položkami). Existují i výpočetní postupy, které pomáhají tento problém řešit, například již zmíněný software FACTOR je nabízí, ale bohužel ani tento postup nepomůže ve všech situacích.

<sup>14</sup> SPSS neumí počítat tetrachorické koeficienty, nicméně i pro dichotomie lze užít koeficienty polychorické (jde totiž o speciální případ polychorického koeficientu). Doplňme, že podmínkou pro výpočet polychorických koeficientů v SPSS je správné nastavení měřících stupnic jednotlivých proměnných, tj. typicky je nastavíme jako ordinální.

<sup>15</sup> Pro využívání propojení SPSS s R je třeba doinstalovat doplněk SPSS Extension to R (zdarma) a také příslušnou verzi R (zdarma). Vše o tomto lze nalézt na webu IBM. Případní zájemci se mohou obrátit na autora článku v případech potíží.

<sup>16</sup> Implementace balíčku lavaan v R skrze JASP bohužel zatím umí jen práci s kardinálními proměnnými (verze 14.1).

<sup>17</sup> Jde o matematickou vlastnost korelační matice a její naplnění znamená, že bude možné provést výpočet faktorové analýzy.

Obdobně v rámci CFA působí často výpočetní potíže algoritmus WLSMV užívaný pro ordinální data. Typicky jde o situace, kdy využíváme víceúrovňové modely CFA. Tyto modely jsou výpočetně velmi komplikované, i při zcela běžných modelech dochází k nemožnosti odhadu těchto modelů. Vše si můžeme ukázat na analýze provedené týmem Straková, Simonová, Soukup [2020], který zkoumal koncept akademické marnosti skrze dvouúrovňový model CFA. Model akademické marnosti byl na úrovni žáka (1. úroveň) i třídy (2. úroveň) konstruován jako model se třemi faktory (marnost ve vztahu k žákovi, ke spolužákům a k učiteli) s tím, že faktory byly vzájemně provázány. Faktorové zátěže na první i druhé proměnné pro stejné vazby byly zafixovány na stejnou hodnotu (jde o běžný postup ve víceúrovňových CFA). Při využití polychorických korelací a odhadovacím algoritmu WLSMV, které zohlednily ordinalitu měřicích stupnic manifestních proměnných (čtyřbodové stupnice souhlasu) došlo v softwaru MPlus ke kolapsu algoritmu s tím, že šestidimenzionální integrování vyžaduje řádově  $10^8$  integračních bodů a na tento výpočet běžně dostupná výpočetní technika nestačí (je zapotřebí obrovská paměť počítače). Pro řešení této úlohy bylo možné použít alternativní postupy. Nabízí se jednak možnost bayesovských odhadů (těm se v článku nevěnujeme) nebo užití některého z odhadovacích algoritmů, který není tak výpočetně náročný. V této situaci s ohledem na předchozí doporučení bylo možné užít buď algoritmus WLS, nebo dle Liho [2016a: 384] robustní verzi metody maximální věrohodnosti (MLR). Pro maximální srovnatelnost s výsledky jednoúrovňového modelu (zde byl užit algoritmus WLSMV pro nejlepší vystižení charakteru dat) byl zvolen algoritmus WLS, který již vedl k získání odhadů a umožnil přibližné srovnání výsledků s jednoúrovňovým modelem.

Před opuštěním rozboru typu proměnných ve faktorové analýze je vhodné krátce pojednat o technikách, které jsou blízké logice EFA a jsou určeny primárně pro ordinální či binární data. Zřejmě nejbližší technikou je kategoriální analýza hlavních komponent (CATPCA), z hlediska analytických cílů bude blízká též analýza latentních tříd či složitější seskupovací algoritmy (zejména dvoustupňové seskupování). Dodejme, že jak kategoriální analýza hlavních komponent, tak dvoustupňové seskupování jsou přímo dostupné v SPSS, analýza latentních tříd je pak dostupná v SPSS opět skrze propojení s R. Slabinou těchto procedur je poměrně výrazný nedostatek literatury i dokumentace k samotnému SPSS. Uživatel tedy nemá podporu v míře, na kterou je zvyklý z jiných technik. Nejsnazší je zřejmě užití dvoustupňového seskupování, protože jde o analogii k běžně užívanému hierarchickému seskupování. Je namístě ještě upozornit na blízkost mezi konfirmační faktorovou analýzou a teorií odpovědí na položku (IRT). Je známo [viz např. Brown 2006: 396–404], že jednodušší modely IRT (jedno- a dvouparametrické) jsou přímo převoditelné na jednofaktorové modely konfirmační faktorové analýzy.

#### **Ad 4) Je pro stanovení počtu faktorů ideální Kaiserovo pravidlo, případně Cattelův suťový graf (scree plot)?**

Náhledem do Tabulky 1 lze zjistit, že v textech publikovaných na stránkách Sociologického časopisu se buď počet faktorů nezdůvodňuje nijak (čtyři případy z pěti), nebo je zdůvodněn jinými než statistickými kritérii.

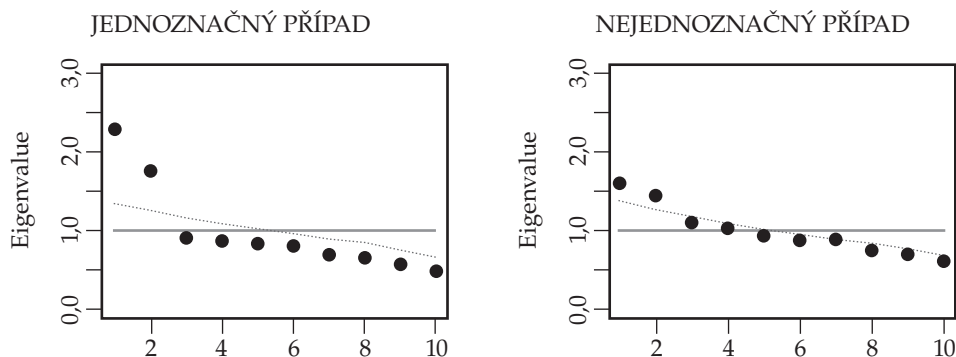
V případě explorační faktorové analýzy je před samotným výpočtem kromě dalších nastavení (metoda odhadu a rotace, viz další část) stanovit i počet faktorů, které budou získány. Nejvýhodnější strategií je samozřejmě předchozí znalost, která nás vede k požadavku určitého počtu faktorů. Připomeňme ale, že pokud již máme konkrétní znalost o počtu faktorů a jejich struktuře, pak bychom měli místo EFA užívat spíše CFA. A otázkou pak samozřejmě zůstává, jak stanovit vhodný počet faktorů, pokud předchodí znalost nemáme. Pro tyto situace vyvinuli statistici několik různých pravidel, z nichž nejužívanější je Kaiserovo<sup>18</sup> pravidlo, případně Cattelovo pravidlo odvozené ze suťového grafu (scree plot). Na velmi častou užívanost těchto kritérií upozorňují v oblasti organizačního výzkumu Hayton, Allen a Scarpell [2004], když konstatují, že v devadesátých letech bylo užito těchto pravidel zhruba v polovině článků ve dvou nejprestižnějších časopisech, naopak ani jednou nebylo užito sofistikovanějších postupů, o kterých bude dále pojednáno. Dodejme, že popsaná pravidla byla odvozena v šedesátých letech minulého století [Kaiser 1960; Cattel 1966] a již tehdy byla zpochybňována a byla nabízena alternativní pravidla či doporučení. Nicméně tato dvě pravidla byla velmi jednoduchá (bylo je možné použít pro ruční výpočty), byla implementována do většiny softwaru a jsou tak užívána (zpravidla bez pochybností) i více než 50 let od svého vzniku. Doplňme, že obě pravidla jsou přímo spojena s výpočtem faktorové analýzy (jsou založena na vlastních číslech korelační nebo kovarianční matice) a díky tomu byla preferována při ručních výpočtech (jednoduše se vypočítají). Jak prokázaly četné simulační studie, obě nejužívanější pravidla obvykle vedou k více faktorům, než je třeba [Zwick, Velicer 1986; Pallant 2005]. Připomeňme, že faktorová analýza je technika, jejímž cílem je nahradit větší počet proměnných výrazně menším počtem nových proměnných (faktorů). Není pak zřejmě vhodné pro rozhodnutí o jejich počtu užívat kritéria, která vedou k vyššímu počtu faktorů, než je potřebné. Kromě tohoto problému je třeba zmínit skutečnost, že interpretace suťového grafu je značně subjektivní, často nevede k jasnému určení počtu faktorů (srov. Obrázek 2). Jako další kritérium, které je často užíváno, musíme ještě zmínit požadavek na minimální celkový vysvětlený rozptyl. V literatuře variuje doporučovaná výše minimálního vysvětleného rozptylu od cca 50 % do 80 %. Nutno zde zdůraznit, že použít faktorové řešení, které vysvětluje méně než 50 % rozptylu původních proměnných je ve většině situací

---

<sup>18</sup> Správný název by měl znít Kaiserovo-Guttmanovo, ale na druhého autora se běžně zapomíná.



**Obrázek 2. Příklad jednoznačného a nejednoznačného určení počtu faktorů ze suťového grafu**



Zdroj: Ruscio, Roche [2012: 283, obrázek 1].

diskutabilní<sup>19</sup>, nicméně kritérium minimálního vysvětleného rozptylu je nutno brát jen jako doplňkové, prokazující, že naše řešení není zcela špatné s ohledem na charakter našich dat. Je tedy třeba se zaměřit na vhodnější postupy pro určení počtu faktorů v rámci EFA. První nebude ryze statistický, poté představíme dva postupy, které vyvinuli statistici jako reakci na Kaiserovo kritérium a Cattellův suťový graf.

Jako první je vhodné uvést, že při užívání explorační faktorové analýzy bychom měli vždy dbát na vhodnost věcné interpretace, tj. zejména na jasnou a srozumitelnou interpretaci faktorů [Gavora 2012: 38]. Vždy bychom tedy měli projít více řešení a zvažovat volbu takového řešení, které bude interpretačně jednoduché.

Nyní nahlédneme dva postupy, které vedou k určení doporučeného počtu faktorů a mají lepší vlastnosti než Kaiserovo pravidlo či suťový graf. Jde o Velicerův test MAP a tzv. paralelní analýzu. Zejména druhý uvedený postup je stále dostupnější v softwaru a jeho užití je velice jednoduché (de facto jde o rozšíření suťového grafu, pokud využijeme grafického řešení). Slabinou těchto postupů je jejich absence v běžně užívaných softwarech (zejména v SPSS a SAS). Nicméně díky některým pomůckám, jak si ukážeme dále, není jejich užití obtížné. S ohledem na malou aplikační rozšířenost představíme tyto postupy detailněji.

<sup>19</sup> Někdy je i řešení s méně než polovinou vysvětleného rozptylu přijatelné. Konkrétně půjde o situace, kdy v rámci faktorového řešení budeme mít více proměnných, které s většinou ostatních nesouvisí, dále o situace, kdy máme měřených proměnných mnoho (desítky či více) a požadujeme jen zachycení několika základních dimenzí. Děkuji za toto upozornění jednomu z anonymních recenzentů článku.



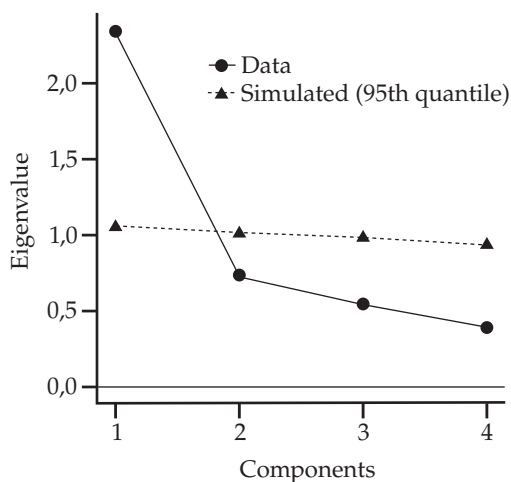
Velicerův MAP test (Minimum Average Partial Test [Velicer 1976]) je poměrně sofistikovaný postup. Jeho logika je založena na následujících výpočtech. Nejdříve je využit první získaný faktor a jsou spočteny dílčí (parciální) korelace mezi proměnnými vstupujícími do analýzy při kontrole prvního faktoru. Z těchto dílčích korelací se spočte průměrná hodnota druhé mocniny. Totéž se následně provede s prvním a druhým faktorem, opět se spočtou dílčí korelace vstupních proměnných a kontroluje se vliv prvního a druhého faktoru zároveň. Postup se opakuje tolikrát, kolik máme vstupujících proměnných (tolik faktorů lze maximálně získat). Poté se seřadí hodnoty průměrů druhých mocnin dílčích korelačních koeficientů a počet faktorů odpovídá kroku, pro který byla tato hodnota nejnižší. Dále dle Velicera platí, že pokud průměrná hodnota druhé mocniny párových korelačních koeficientů našich proměnných je menší než příslušné hodnoty získané z výše popsaných parciálních koeficientů, pak nemá smysl faktorovou analýzu vůbec provádět.

Hornova paralelní analýza (parallel analysis [Horn 1965]) je vylepšením Kaisera pravidla. Kaiser předpokládá, že pokud by naše proměnné byly zcela nezávislé, byla by všechna vlastní čísla našeho řešení jednotková, a tak má smysl zvolit takový počet faktorů, kolik jich má vlastní číslo větší než jedna. Paralelní analýza tuto myšlenku modifikuje blíže pravdě a snaží se ukázat, že i při úplné nezávislosti proměnných budou (v důsledku působení náhody) některá vlastní čísla o něco vyšší než jedna a některá menší než jedna (jejich součet pro všechny potenciální faktory musí být vždy roven počtu proměnných vstupujících do analýzy). Logika paralelní analýzy je následující. Opakovaně nasimulujeme pro stejný počet proměnných, jako máme v EFA, vzájemně nezávislé proměnné a spočítáme pro ně vlastní čísla. Z našich simulací (běžně v řádu tisíců) pak určíme průměrnou hodnotu jednotlivých vlastních čísel a také některý jejich vyšší kvantil (typicky 95%). Poté srovnáme vlastní čísla z naší analýzy s kvantily získanými z paralelní analýzy a použijeme tolik faktorů, kolikrát platí, že naše vlastní číslo je větší než kvantil získaný z paralelní analýzy. Vše je nejsnazší ukázat na obrázku (Obrázek 3), který je rozšířením Cattelova suťového grafu, jen je zde doplněna mírně sestupující linie, která je produktem paralelní analýzy.

Z obrázku je zřejmé, jak naložit s výsledkem paralelní analýzy. Sledujeme průsečík vlastních čísel našeho řešení (z počátku strmě padá, posléze minimálně, legenda označuje jako „Data“) a simulovaných (mírně klesající linie, legenda označuje jako „Simulated (95th quantile)“). Doporučený počet faktorů je dán počtem faktorů vlevo před průsečíkem (tj. pro nás jeden faktor).

Po rozboru Velicerova MAP testu a paralelní analýzy je vhodné uvést, že simulační studie [Zwick, Velicer 1986; Ruscio, Roche 2011; Pearson, Mundfrom, Piccone 2013] ukazují, že tyto techniky k určení počtu faktorů jasně vítězí nad klasickým Kaiserovým pravidlem či Cattelovým suťovým grafem bez modifikace dle Horna. Platí, že paralelní analýza předčí Velicerův MAP test, její užití je navíc jednodušší. Uvedme proto základní možnosti, kde tyto postupy využívat. Oba postupy je možné využít v komerčním softwaru Stata, Mplus umí též výpočet

Obrázek 3. Paralelní analýza



Zdroj: Vlastní výpočet v softwaru JASP.

paralelní analýzy. Ve volně šiřitelném R jsou též oba postupy, paralelní analýza je například v balíčku *hornpa*, Velicerův MAP test i paralelní analýza pak v balíčku *paramap* (autorem je dále zmíněný O'Connor). Poměrně slibné možnosti nabízejí online nástroje, pro paralelní analýzu lze doporučit webovou stránku <https://analytics.gonzaga.edu/parallellengine/>.

Jako slibné lze hodnotit programy připravené pro SPSS a SAS od O'Connora [2000]. Jde o sled příkazů, který vypočte jak Velicerův MAP test, tak paralelní analýzu v prostředí SPSS nebo SAS<sup>20</sup>. Pro užívání těchto programů je potřebná znalost příkazového jazyka SPSS (resp. SAS), protože pro konkrétní data a počítač je potřebné příkazy na některých místech upravit.<sup>21</sup> Uživatelům SPSS neznalých příkazového jazyka je možné doporučit užívání volně šiřitelného softwaru JASP<sup>22</sup>, který v rámci faktorové analýzy obsahuje též možnost výpočtu paralelní analýzy, viz Obrázek 3). Další možností je součást softwaru ViSta<sup>23</sup> označovaná jako ViStaPARAN, která umí také skrze jednoduché klikací rozhraní určit počet

<sup>20</sup> Skripty lze stáhnout na adrese <https://people.ok.ubc.ca/briocconn/nfactors/nfactors.html>.

<sup>21</sup> Česká verze příkazů pro SPSS s pokyny pro úpravu je součástí online přílohy dostupné na <https://doi.org/10.13060/csr.2021.021>.

<sup>22</sup> JASP umí přímo načítat data ve formátu SPSS a jeho nabídky jsou velice podobné nabídkám SPSS.

<sup>23</sup> Jde o freeware určený zejména pro explorační analýzu dat.

faktorů prostřednictvím paralelní analýzy [Ledesma, Valero-Mora 2007]. Je patrné, že zejména v rámci volně dostupného softwaru je již nabídka dostačující.

Na závěr doplníme, že některé zahraniční časopisy již mají striktní požadavek na užití paralelní analýzy pro určení počtu faktorů, Pallant [2005: 175–176] zmiňuje například *Educational and Psychological Measurement* nebo *Journal of Personality Assessment*. Paralelní analýza se tak stává de facto povinnou. Je samozřejmě diskutabilní, zda je to správná cesta.

Důvody k opatrnosti jsou následovné. Paralelní analýza je typicky v softwaru vyvinuta jen pro některé odhadovací techniky (srov. například výše uvedenou online aplikaci a její nastavení), nejčastěji pro metodu hlavních komponent, někdy též pro hlavní osy. Pokud zvolíme jinou metodu odhadu, nemáme zpravidla možnost paralelní analýzu využít. Problematictějším rysem je situace, kdy naše faktory jsou závislé (jak ukáže další část, je to situace zcela běžná). Pro zkorelované faktory, jak ukazují simulace [Caron 2019], nejsou ani paralelní analýza, ani MAP test příliš dobré (nicméně stále jsou lepší než Kaiserovo pravidlo či Cattellův suťový graf). Na druhou stranu musíme zmínit, že jiné simulace prokázaly [Dinno 2009], že paralelní analýza je dostatečně robustní, lze ji tedy použít bez ohledu na rozdělení dat.

Aby vše bylo ještě komplikovanější, je vhodné uvést, že existuje poměrně mnoho postupů, jak počítat paralelní analýzu. Dosud jsme představili jen „nejjednodušší“, založený na 95% kvantilu, nicméně těchto postupů lze nalézt v literatuře (i softwaru) více. Zřejmě nejdiskutovanější je postup revidované paralelní analýzy (R-PA), vyvinuté týmem vedeným Greenem [2012], a také postup navržený v článku Ruscia a Rocheho [2012], označovaný jako comparison data method (CD). Green se svým týmem provedl dvojí srovnání revidované a klasické paralelní analýzy, jedno pro spojitá data [Green et al. 2015], jedno pro dichotomické položky<sup>24</sup> [Green et al. 2016]. V obou případech dospívají k závěru, že revidovaná verze je lepší. Tento závěr je samozřejmě diskutabilní, protože hodnocení nástroje provádějí autoři, kteří jej vyvinuli. Na základě této skutečnosti byly v nedávné době provedeny další dvě studie, první Auerswaldem a Moshagenem [2019] a druhá Limem a Jahngem [2019]. Auerswald s Moshagenem se soustředili na srovnání klasické paralelní analýzy a postupu CD a na svých simulovaných datech konstatují, že výkon obou těchto procedur je v zásadě srovnatelný s tím, že PA typicky vykazuje dobré vlastnosti při určení počtu faktorů pro řešení s jedním faktorem a pro vícefaktorová řešení s kolmými faktory. Pro zkorelované faktory je naopak dle studie Auerswalda s Moshagenem [2019] lepší postup označovaný jako comparison data (CD). Auerswald s Moshagenem [ibid.: 486] také explicitně

<sup>24</sup> Mimo jiné upozorňují ve své studii i na skutečnost, že v případě dichotomických položek je krajně nevhodné nejen pro získání faktorového řešení, ale i pro určení počtu faktorů užívat Pearsonovy korelace, a zdůrazňují vhodnost polychorických, resp. tetrachorických korelací [Green et al. 2016: 9]. Opět se tedy potvrzuje (byť z jiných než dosud uvedených důvodů) doporučení ohledně alternativních korelací pro data ordinální povahy.

doporučují na základě simulační studie užívat právě 95% kvantil jako nejvhodnější variantu paralelní analýzy.

Nová studie Lima a Jahnga [2019] reaguje na závěry studií publikovaných kolektivem vedeným Greenem [2012, 2015 a 2016]. Jejich závěry do velké míry vyvrací a ukazují na přednosti klasické paralelní analýzy. Alternativy ve formě R-PA či CD byly dle této studie lepší než klasická PA jen v případech, kdy korelace mezi faktory byly velmi silné. Autoři v závěru své studie doporučují užívat klasickou PA. S ohledem na skutečnost, že alternativní postupy jsou zatím pro běžné uživatele v zásadě nedostupné, lze toto doporučení podpořit i praktickými důvody.

S ohledem na dostupnost v softwaru a dosavadní rozporuplné výsledky publikovaných studií můžeme doporučit užívání klasické paralelní analýzy, která se jeví jako velice dobrý postup pro určení počtu faktorů. Naopak Kaiserovo pravidlo či Cattellovo určení počtu faktorů ze suťového grafu již patří k překonaným postupům a jejich užití doporučit nelze.

#### **Ad 5) Jaké jsou vhodné postupy pro získání prvotního řešení a pro rotaci v rámci EFA?**

Pátá otázka vychází ze skutečnosti, že mezi uživateli EFA je velice oblíbená metoda hlavních komponent (PCA) pro získání prvotního řešení a následně je užita kolmá rotace skrze postup označovaný jako varimax. Ostatně i Tabulka 1 tuto informaci v českém prostředí potvrzuje. Z velice detailní zahraniční přehledové studie Hensona a Robertse [2006] víme, že i v elitních časopisech (*Educational and Psychological Measurement, Journal of Educational Psychology, Personality and Individual Differences a Psychological Assessment*) není praxe uspokojivá. Henson a Roberts [ibid.: 403–405] ukazují, že PCA je užíváno v 57 % jimi zkoumaných článků (PA ve 22 %), zároveň vysvětlují, že hlavním důvodem bude skutečnost, že jde o defaultní nastavení ve většině softwaru. U rotace většinou (62 % analyzovaných článků) autoři nijak nezdůvodňují její volbu, většinou pak volí kolmou rotaci varimax (52 %).<sup>25</sup> Jde obecně o jeden z největších problémů v rámci užívání EFA, často ještě v kombinaci s užíváním EFA jako nástroje pro konfirmaci (srov. druhou část textu). V EFA lze odhadovat parametry mnoha různými metodami (část jsme již zmínili v komentáři k Tabulce 1), nadto je výsledné řešení možné upravit pro snazší interpretaci procesem rotace. Nebudeme zde do detailu vysvětlovat nuance jednotlivých metod odhadu, u rotace jen připomeneme, že může být buď kolmá (předpokládáme nezávislost faktorů), nebo šikmá (očekáváme provázanost faktorů). Základní popis nejčastěji používaných metod odhadu v EFA nabízí Tabulka 4.

<sup>25</sup> Dodejme, že v některých textech není rotace užita vůbec, podíl užívání varimax lze tedy odhadovat na cca 70–80 %.

**Tabulka 4. Přehled metod odhadu EFA a jejich stručný popis**

Metoda odhadu	Anglický název (zkratka)	Stručný popis; výhody a nevýhody
Maximální věrohodnost	Maximum Likelihood (ML)	Hledá odhady, pro které má nejvyšší hodnotu věrohodnostní funkce, výhodou je formální test počtu faktorů, nevýhodou je předpoklad normálního rozdělení vstupujících proměnných, vhodné pouze pro normálně rozdělené kardinální proměnné, v praxi tedy minimálně použitelné.
Hlavní komponenty	Principal Components (PCA)	Maximalizuje vysvětlený rozptyl proměnných, nej- užívanější, nevýhodou může být malá míra vysvětlení vztahů mezi proměnnými. Vhodné i pro delší ordinální stupnice, pokud chceme vysvětlit co nejvíce rozptylu původních proměnných.
Hlavní osy	Principal Axes (PA)	Obdoba PCA, jen jedničky na diagonále nahrazují hodnotu komunality (vysvětleného rozptylu jednotlivé proměnné), výsledky většinou obdobné PCA. Doporučuje se užívat místo PCA.
Nevážené nejmenší čtverce	Unweighted Least Squares (ULS)	Obdobně jako v regresní analýze minimalizuje rozdíl mezi čtverci odlišností napozorované korelační matice a faktorovou analýzou vytvořené korelační matice (bez diagonály), nevýhodou je citlivost na velké rozdíly mezi vstupními korelacemi, často kvůli tomu nevhodná technika.
Vážené nejmenší čtverce	Generalized Least Squares (GLS)	Postup obdobný jako ULS s tím, že čtvercové rozdíly se váží, proměnné s vyšší hodnotou komunality mají vyšší váhu a vice versa, částečně eliminuje nevýhody ULS. Potenciální alternativa k PA či PCA.
Alfa faktorová analýza	Alpha Factoring (AF)	Odlišná od ostatních technik, alfa faktorová analýza předpokládá, že naše proměnné jsou jen výběrem ze všech potenciálních proměnných a hledá podobnosti takto náhodně vybraných proměnných.

*Zdroj: SPSS a popis metod.*

K Tabulce 4 doplňme, že nejužívanější metoda hlavních komponent je v zásadě přijatelná, i když simulace ukazují, že minimálně stejně dobrých výsledků lze dosáhnout i při užití hlavních os. Již víme, že pokud mají naše data normální rozdělení, měli bychom užívat metodu maximální věrohodnosti. Obecně platí, že PCA či PA by měly být preferovány, pokud nemáme zvláštní důvod užít jinou techniku.

Nyní se budeme věnovat problému rotace. Připomeňme, že hlavní motivací pro rotaci je získat snáze interpretovatelné řešení. Protože u rotace s pouhým slovním rozbořem nevystačíme, začneme motivačním příkladem.

### Příklad 1: Hodnocení výuky (kolmá, či šikmá rotace?)

Začneme s příkladem na reálných datech. Využijeme data o hodnocení výuky na Fakultě humanitních studií Univerzity Tomáše Bati ve Zlíně<sup>26</sup> z roku 2017. Ukážeme, jaké může mít dopady rozhodnutí o rotaci (kolmá, nebo šikmá?) v rámci EFA. Abychom zároveň prokázali, které ze dvou řešení je s ohledem na strukturu našich dat správné, použijeme doplňkově výsledky CFA.<sup>27</sup>

V rámci hodnocení výuky je od studentů získávána odpověď na deset hodnotících otázek (viz Tabulka 5). Explorační faktorovou analýzou bychom měli odhalit strukturu těchto odpovědí, předběžně z dřívějších výsledků očekáváme přítomnost dvou dimenzí (faktorů). Pro srovnání spočítáme řešení s kolmou rotací (nejčastěji užívaná varimax) a šikmou (často užívaná oblmin).

Výsledky ve formě tabulky faktorových zátěží shrnuje Tabulka 5, kde levý panel je věnován kolmé rotaci metodou varimax a pravý panel šikmé rotaci metodou oblmin.

Při prvotním náhledu na tabulku můžeme konstatovat, že jak řešení s kolmou rotací, tak řešení se šikmou rotací ukazují poměrně jasně na rozdělení proměnných do dvou faktorů, kde první čtyři proměnné tvoří druhý faktor a zbývajících šest faktor první. Při srovnání hodnot faktorových zátěží (korelace mezi faktory a položkami) zjistíme, že šikmo rotované řešení je celkově o něco lepší (zpravidla má hodnoty zátěží výraznější). V tabulce je navíc uvedena jedna důležitá skutečnost, a to hodnota korelace mezi faktory. U kolmé rotace jsou faktory z definice nezávislé, nepřekvapí tedy nulová hodnota (software tuto hodnotu neuvádí), u šikmé rotace je výsledkem korelovanost o hodnotě 0,5 (rozhodně tedy nezanedbatelná hodnota). Nyní je namístě citovat jeden z metodologicky orientovaných textů v části o rotacích:

Existujú dva základné druhy rotácie, ortogonálna (pravoúhla) a šikmá (kosoúhla)... K dispozícii sú rotácie varimax, quatrimax<sup>28</sup>, orthomax, equamax, promax, centroidná atď. Niketóre majú viaceré pojmenovania. Existuje čulá diskusia na tému, ktorá rotácia je najvhodnejšia pre daný výskum. Nepriniesla však jednoznačné riešenie. Renomovaný metodológ F. N. Kerlinger (1972, s. 646) diskusiu uzavrel lapidárnym výrokom – druh rotácie je otázkou vkusu. Vo všeobecnosti sa najčastejšie používa rotácia varimax, ktorá je ortogonálna. [Gavora 2012: 39]

Zejména poslední dvě věty jsou značně diskutabilní, tj. rotace je věcí vkusu a dobrým vkusem je rotovat kolmo metodou varimax. Dnes (rok 2021) musíme zejmé-

<sup>26</sup> Děkuji tímto dr. Kalendovi, prorektorovi UTB pro studium, za jejich laskavé zapůjčení.

<sup>27</sup> Tím se (záměrně) dopustíme praktiky, která byla ukázána jako nevhodná v části 2. Činíme tak vědomě, abychom mohli ukázat na problematičnost kolmé rotace.

<sup>28</sup> Zde je zřejmě v originále překlep, správné označení je quartimax.

**Tabulka 5. Faktorové zátěže pro kolmou (varimax) a šikmou (oblimin) rotaci dat z hodnocení výuky na FHS UTB**

	Kolmo (varimax) r = 0		Šikmo (oblimin) r = 0,5	
	1	2	1	2
Byl/a vyučující na výuku připraven/a?	0,041	0,842	-0,176	-0,917
Vysvětloval/a vyučující učivo jasně a srozumitelně?	0,455	0,626	0,347	-0,539
Objasňoval/a vyučující studentům to, co jim nebylo jasné?	0,424	0,627	0,312	-0,550
Působil/a vyučující jako odborník na vyučovanou problematiku?	0,249	0,719	0,090	-0,711
Vyvolával/a vyučující u studentů zájem o učivo a o samotný předmět?	0,690	0,329	0,689	-0,132
Umožňoval/a vyučující studentům vyjadřovat jejich názory a diskutovat?	0,626	0,116	0,674	0,082
Doporučil/a (poskytl/a) vyučující studentům vhodné studijní materiály?	0,657	0,172	0,694	0,031
Měl předmět jasnou a ucelenou koncepci?	0,573	0,494	0,514	-0,353
Byly jasně definovány požadavky pro úspěšné zakončení předmětu?	0,574	0,135	0,610	0,044
Obohatil Vás předmět o nové znalosti, dovednosti, postoje?	0,658	0,298	0,662	-0,108

Zdroj: Vlastní výpočty z dat hodnocení výuky na FHS UTB, N = 4 240.

na odmítnout odkaz na Kerlingera (kniha byla vydána v USA poprvé v roce 1964) jako autoritu v oblasti faktorové analýzy.<sup>29</sup> K vyvrácení uvedeného citátu lze užít současné učebnice nebo některé články [Costello, Osborne 2005]. Ostatně Osborne [2015: 6] přímo vyvrací tvrzení o vhodnosti kolmé rotace, protože šikmá rotace je obecnější řešení (obsahuje v sobě i řešení kolmé, pokud získáme  $r = 0$ <sup>30</sup>).

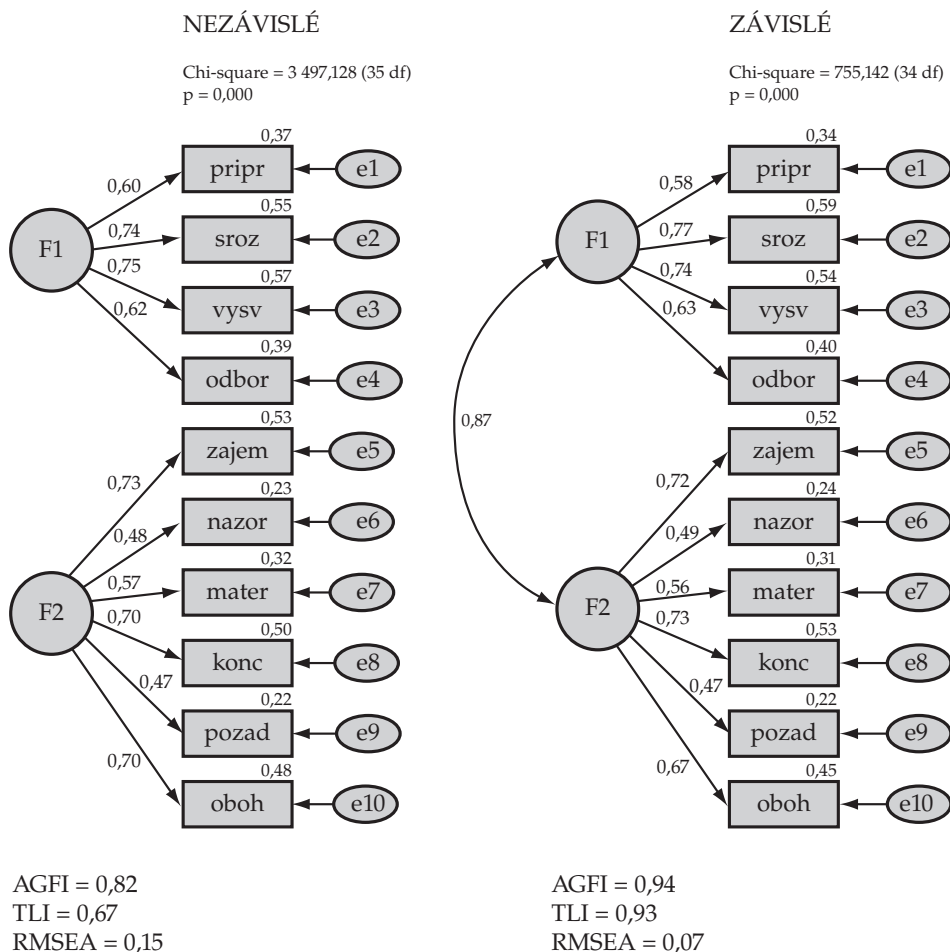
Po náznaku, že bychom měli vždy užít šikmou rotaci, případně šikmou a kolmou a jejich řešení srovnat, se věnujme dále našemu příkladu. Abychom zjistili, zda je pro naše data přílehavý model s kolmou, či šikmou rotací (nezávis-

<sup>29</sup> Kerlingerovu větu z roku 1964 lze pochopit. Tehdejší výpočet byl prováděn ručně, bylo tak vhodné užít co nejjednodušší postup, tím byla typicky kolmá rotace. V době dnešní výpočetní techniky je ale takový přístup již velmi zastaralý.

<sup>30</sup> To bude ale v praxi zpravidla nemožné, perfektní nekorelovanost je mytická.



**Obrázek 4. Konfirmační faktorová analýza pro dva nezávislé a závislé faktory pro data z hodnocení výuky na FHS UTB**



Zdroj: Vlastní výpočty z dat hodnocení výuky na FHS UTB, N = 4 240.

lými, či závislými faktory), zkusíme pro stejná data užít konfirmační faktorovou analýzu.<sup>31</sup> Výsledky nabízí Obrázek 4.

Pokud nahlédneme výsledky obou řešení, pak je závěr jasný. Představa nezávislosti faktorů je neudržitelná. Hodnoty kritérií užívaných pro vyhodnocení

<sup>31</sup> Znovu upozorňujeme na skutečnost, že toto je obecně nevhodná praxe, pro zcela stejná data EFA a CFA nepoužíváme, jak jsme rozebrali ve druhé části článku.

CFA jsou výrazně mimo doporučované meze [Kline 2011: 206–208], naopak řešení s provázanými faktory je přijatelné, hodnoty AGFI, resp. TLI jsou vysoké a hodnota RMSEA je naopak přijatelně nízká. Nadto vypočtená korelace mezi faktory je úctyhodných 0,87. Ignorování této souvislosti by bylo výraznou věcnou chybou. Zde tedy můžeme předběžně uzavřít odpověď na otázku ohledně rotací v EFA. Vhodný postup je jeden ze dvou následujících. *Buď provedu jen některou ze šikmých rotací, a jen pokud mi korelace mezi faktory vyjde zanedbatelná, použiji jako výslednou některou z kolmých rotací. Nebo zároveň vypočtu jak kolmo, tak šikmo rotované řešení a dle hodnoty korelace u šikmého řešení se rozhodnu, které použiji.* Pokud bude korelace zanedbatelná, preferuji kolmou rotaci, v opačném případě (téměř vždy) rotaci šikmou. Pokud tyto postupy neužíváme, dostaneme se do velice zvláštních situací, s nimiž se lze seznámit například v článku Majerčíkové s Gavorou [2013], případně opět v monografii Gavory [2012: 78–81]. V poslední uvedené publikaci je získáno pět faktorů (nejspíše skrze kolmou rotaci<sup>32</sup>). Následně jsou vytvořeny dimenze založené na řešení EFA (zřejmě součtem položek s vysokou korelací s příslušným faktorem<sup>33</sup>). A pro tyto dimenze jsou spočteny korelace [Gavora 2012: 81], kde je nejvyšší hodnota 0,77 a nejnižší 0,12. Hodnoty korelací jsou poté užity dle slov autora ke konstruktové validizaci dotazníku. Uvedený postup je nutno jasně odmítnout.<sup>34</sup> Pokud již dopředu víme, že faktory jsou provázané, musíme užít šikmou rotaci, která nám nadto míru provázanosti ukáže. Pokud to neuděláme, budujeme vnitřně rozporný analytický systém.

## Závěrečné shrnutí (schéma doporučeného postupu)

V textu jsme pojednali odpovědi na několik otázek vztahujících se k faktorové analýze. Výčet, jak uvedeme detailněji dále, nebyl zcela jistě vyčerpávající. Nicméně i tak bylo podáno mnoho informací odděleně skrze odpovědi na jednotlivé otázky. Po této analýze by měla následovat syntéza. Proto se pokusíme většinu obsahu článku vyjádřit do dvou schémat doplněných krátkým doprovodným textem. Pevně věříme, že schémata čtenáři pomohou při samostatné analýze dat a že je nadto bude možné (s případnými zjednodušeními) užívat ve výuce pokročilých statistických technik.<sup>35</sup>

Jak vyplývá ze Schémat 1 a 2, jako první rozhodnutí je třeba zvolit mezi EFA a CFA, toto rozhodnutí provádíme dle fáze, ve které se nacházíme. Při tvorbě

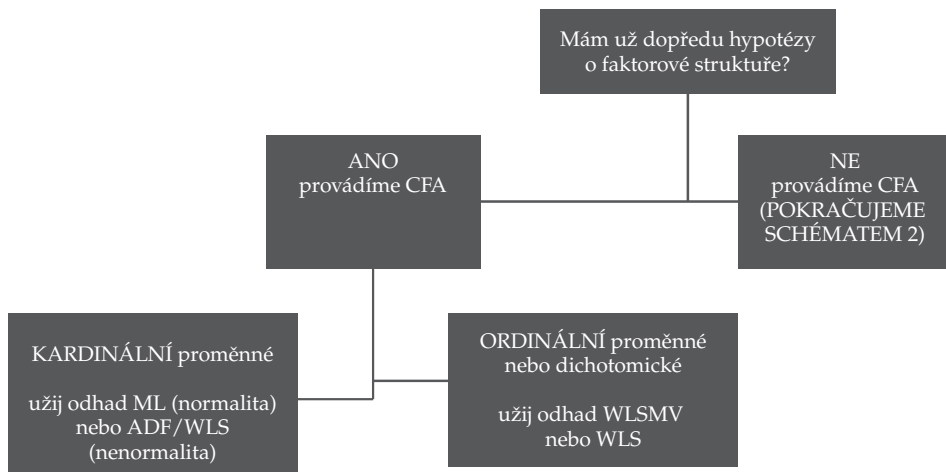
<sup>32</sup> Bohužel v textu samotném není nic uvedeno.

<sup>33</sup> Opět bohužel není uvedeno, jaký byl postup.

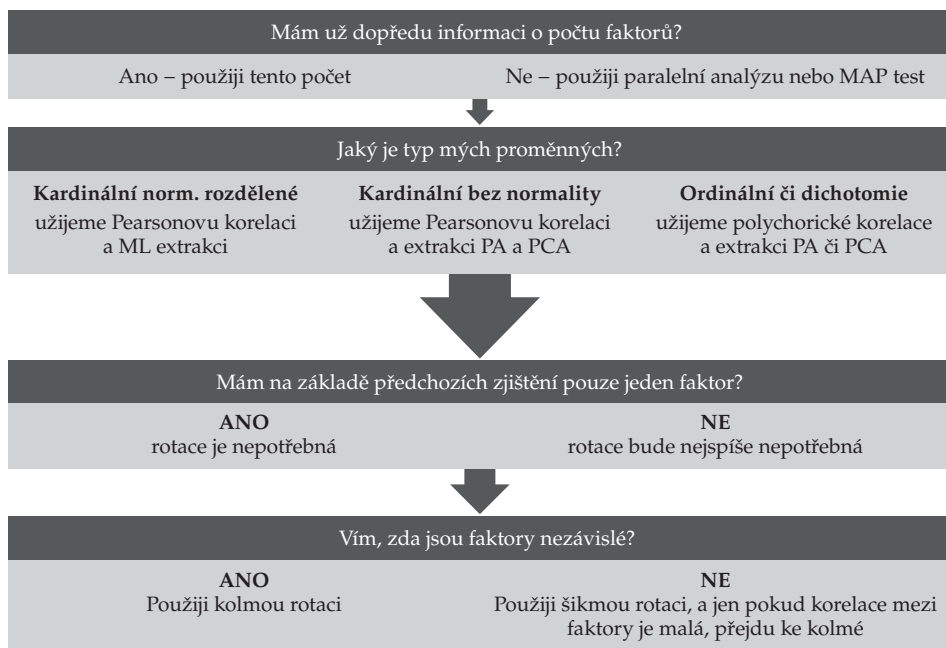
<sup>34</sup> Jako nejproblematictější lze vnímat, že celý příklad (a obdobné další v monografii) jsou v části, která nese název příklady dobré praxe. Bohužel o dobrou praxi jít nemůže, není dostatečně dokumentován postup a nadto minimálně část postupu je chybná (v rozporu se současnými poznatky).

<sup>35</sup> Je zřejmé, že schematizací se zavádí jisté zjednodušení, nicméně bez něj nelze jakákoli doporučení vůbec uvádět.

**Schéma 1. Postup při užívání faktorové analýzy (rozhodnutí mezi EFA a CFA včetně základních postupů CFA)**



**Schéma 2. Postup při užívání EFA**



škály užíváme EFA, při jejím ověřování CFA. Další postup je pak závislý na tom, jakého typu jsou proměnné, pokud máme různé typy proměnných<sup>36</sup> (pro EFA či CFA velmi řídké), dále na tom, jaký je nejnižší kvalitativní typ proměnných (nejvýše stojí kardinální normálně rozdělené proměnné, nejnižše dichotomie). Dle typu proměnných volíme typ korelací a odhadovací techniky (jiné volby provádíme u CFA, viz Schéma 1, a jiné u EFA, viz Schéma 2). U EFA musíme též vyřešit počet faktorů, pokud nemáme informace předem, je vhodné volit postupy, které fungují nejlépe, tj. paralelní analýzu nebo MAP test. Pokud máme v EFA více než jeden faktor, pak zpravidla provedeme rotaci. Bez jasnější indikace, že jsou naše faktory vzájemně nezávislé, volíme jako první rotaci šikmou a dle jejího výsledku (ukazuje nekorelovanost faktorů), můžeme přejít k rotaci kolmé. Alternativně lze spočítat najednou rotaci šikmou a kolmou. Pokud jsou jejich výsledky srovnatelné, můžeme pro interpretaci využít kolmou rotaci, v opačném případě vždy šikmou.

### **Několik poznámek závěrem**

Cílem článku bylo upozornit na některé složitější postupy ve faktorové analýze, zejména v explorační verzi, a ukázat na praxi, která odpovídá soudobým poznatkům. Čtenář se tak mohl seznámit s postupy pro ordinální či dichotomická data, s různými možnostmi stanovení počtu faktorů a odhadu faktorů a jejich rotace. Vše bylo shrnuto ve Schématech 1 a 2. Nyní je třeba ještě několik doplnění a také upozornění na limity předloženého článku.

Předně musíme upozornit, že poznatky v tomto článku mohou zastarat, nebudou tedy platné navždy. Dále je nutno přiznat, že článek neobsahuje úplný výčet potenciálních otázek a odpovědí, mnohé tak zůstává nezodpovězeno. Samozřejmě tyto mezery mohou zaplnit další texty od jiných autorů. Nezbyvá než doufat, že se tak brzy stane. Zcela jistě si zaslouží pozornost téma stanovení velikosti výběrového souboru pro faktorovou analýzu. Jde o téma, které by vydalo na samostatnou monografii. Zde jen krátce uveďme, že doporučení typu, že 100 či 200 respondentů je vhodný počet, je zavádějící, nicméně je obsaženo i v klasických textech [srov. např. Gorsuch 1983]. Nutno dodat, že snad pro žádnou jinou techniku neexistuje tak bohatá škála různých jednoduchých doporučení, vesměs špatných. Základní solidnější poučení může nalézt čtenář v článku Soukupa a Kočvarové [2016: 526], velice solidní přehledy nabízí MacCallum et al. [1999] nebo Mundfrom, Shaw a Lu Ke [2005]. Jako ideální postup lze doporučit stanovení velikosti výběru simulačními postupy ve speciálním softwaru vždy před provedením našeho výzkumu, některé časopisy již tuto praxi požadují jako publikační standard.

---

<sup>36</sup> Tedy situace, kdy některé naše proměnné jsou například kardinální a jiné krátké ordinální stupnice či dichotomické.

Na druhou stranu je nutno zmínit zajímavé studie, které se snaží ukázat, že i s velice malými výběry lze užívat faktorovou analýzu [Winter, Dodou, Wieringa 2009], nejslibnější jsou zde již zmíněné postupy využívající bayesovský přístup.

Musíme zmínit též aplikační aspekt spojený s tématy pojednávanými v textu. Na mnoha místech jsme narazili na skutečnost, že běžně užívaný software (zejména SPSS) neobsahuje procedury, které odpovídají novějším poznatkům. Lze očekávat, že tyto mezery se brzo zacelí jak vývojem softwaru, tak i novými možnostmi. Jako velice slibné se jeví propojení SPSS s R, velice slibná jsou též online prostředí pro výpočty založené na R (tzv. shiny prostředí), rozvíjející se volně šiřitelný software se snadným ovládáním (zejména zmíněný JASP). Je zřejmé, že výzkumníci se mají na co těšit, vývoj je zejména v poslední době velice rychlý.

PETR SOUKUP je sociolog a statistik. Působí na Fakultě sociálních věd Univerzity Karlovy, zaměřuje se na pokročilé techniky analýzy kvantitativních dat, badatelsky se orientuje na sociologii vzdělání.

## Literatura

- Auerswald, M., M. Moshagen. 2019. „How to Determine the Number of Factors to Retain in Exploratory Factor Analysis: A Comparison of Extraction Methods Under Realistic Conditions.“ *Psychological Methods* 24 (4): 468–491, <https://doi.org/10.1037/met0000200>.
- Baglin, J. 2014. „Improving Your Exploratory Factor Analysis for Ordinal Data: A Demonstration Using FACTOR.“ *Practical Assessment, Research & Evaluation* [online] 19 (5): 1–15. Dostupné na: <http://pareonline.net/getvn.asp?v=19&n=5>.
- Blahuš, P. 1985. *Faktorová analýza a její zobecnění*. SNTL.
- Brown, T. 2006. *Confirmatory Factor Analysis for Applied Research*. New York: Guildford.
- Caron, P. O. 2019. „Minimum Average Partial Correlation and Parallel Analysis: The Influence of Oblique Structures.“ *Communications in Statistics – Simulation and Computation* 48 (7): 2110–2117, <https://doi.org/10.1080/03610918.2018.1433843>.
- Cattell, R. 1966. „The Scree Test for the Number of Factors.“ *Multivariate Behavioral Research* 1 (2): 245–276, [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10).
- Costello, A. B., J. W. Osborne. 2005. „Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis.“ *Practical Assessment, Research & Evaluation* [online] 10 (7): 1–9. Dostupné z: <http://pareonline.net/getvn.asp?v=10&n=7>.
- DeVellis, R. F. 2003. *Scale Development: Theory and Applications*. Sage.
- Dinno, A. 2009. „Exploring the Sensitivity of Horn's Parallel Analysis to the Distributional Form of Random Data.“ *Multivariate Behavioral Research* 44 (3): 362–388, <https://doi.org/10.1080/00273170902938969>.
- Flora, D. B, P. J. Curran. 2004. „An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data.“ *Psychological Methods* 9 (4): 466–491, <https://doi.org/10.1037/1082-989X.9.4.466>.
- Gavora, P. 2012. *Tvorba výzkumného nástroja pre pedagogické bádanie*. Bratislava: SPN.

- Gorsuch, R. L. 1983. *Factor Analysis*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Green, S. B., R. Levy, M. S. Thompson, M. Lu, W. J. Lo. 2012. „A Proposed Solution to the Problem with Using Completely Random Data to Assess the Number of Factors with Parallel Analysis.“ *Educational and Psychological Measurement* 72: 357–374, <https://doi.org/10.1177/0013164411422252>.
- Green, S. B., N. Redell, M. S. Thompson, R. Levy. 2016. „Accuracy of Revised and Traditional Parallel Analyses for Assessing Dimensionality with Binary Data.“ *Educational and Psychological Measurement* 76 (1): 5–21, <https://doi.org/10.1177/0013164415581898>.
- Green, S. B., M. S. Thompson, R. Levy, W. J. Lo. 2015. „Type I and II Error Rates and Overall Accuracy of the Revised Parallel Analysis Method for Determining the Number of Factors.“ *Educational and Psychological Measurement* 75: 428–457, <https://doi.org/10.1177/0013164414546566>.
- Hayton, J. C., D. G. Allen, V. Scarpello. 2004. „Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis.“ *Organizational Research Methods* 7 (2): 191–205, <https://doi.org/10.1177/1094428104263675>.
- Henson, R. K., J. K. Roberts. 2006. „Use of Exploratory Factor Analysis in Published Research Common Errors and Some Comment on Improved Practice.“ *Educational and Psychological Measurement* 66 (3): 393–416, <https://doi.org/10.1177%2F0013164405282485>.
- Horn, J. L. 1965. „A Rationale and Test for the Number of Factors in Factor Analysis.“ *Psychometrika* 30 (2): 179–185, <https://doi.org/10.1007/BF02289447>.
- Chrástka, M. 2006. *Úvod do výzkumu v pedagogice*. 2. vyd. Olomouc.
- Kaiser, H. F. 1960. „The Application of Electronic Computers to Factor Analysis.“ *Educational and Psychological Measurement* 20 (1): 141–151, <https://doi.org/10.1177/001316446002000116>.
- Kline, R. B. 2011. *Principles and Practice of Structural Equation Modeling*. 3rd ed. Taylor & Francis.
- Ledesma, R. D., P. Valero-Mora. 2007. „Determining the Number of Factors to Retain in EFA: An Easy-to-use Computer Program for Carrying out Parallel Analysis.“ *Practical Assessment Research & Evaluation* [online] 12 (2). Dostupné na: <http://pareonline.net/getvn.asp?v=12&n=2>.
- Li, Ch. H. 2016a. „The Performance of ML, DWLS, and ULS Estimation With Robust Corrections in Structural Equation Models With Ordinal Variables.“ *Psychological Methods* 21 (3): 369–387, <https://doi.org/10.1037/met0000093>.
- Li, Ch. H. 2016b. „Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares.“ *Behavioral Research* 48: 936–949, <https://doi.org/10.3758/s13428-015-0619-7>.
- Lim, S., S. Jahng. 2019. „Determining the Number of Factors Using Parallel Analysis and Its Recent Variants.“ *Psychological Methods* 24 (4): 452–467, <https://doi.org/10.1037/met0000230>.
- MacCallum, R. C., K. F. Widaman, S. Zhang, S. Hong. 1999. „Sample Size in Factor Analysis.“ *Psychological Methods* 4 (1): 84–89, <https://doi.org/10.1037/1082-989X.4.1.84>.
- Majerčíková, J., P. Gavora. 2013. „Vnímaná zdatnosť (self-efficacy) učiteľa spolupracovať s rodičmi: konštrukcia výskumného nástroja.“ *Pedagogika* 62 (2): 128–146.
- Mazanec, J., E. Dvořáková. 2017. „Bayesovská exploratorní faktorová analýza.“ *Testforum* 9: 1–15, <https://doi.org/10.5817/TF2017-9-131>.
- Mundfrom, D. J., D. G. Shaw, T. Lu Ke. 2005. „Minimum Sample Size Recommendations for Conducting Factor Analyses.“ *International Journal of Testing* 5 (2): 159–168, [https://doi.org/10.1207/s15327574ijt0502\\_4](https://doi.org/10.1207/s15327574ijt0502_4).

- Netermeyer, R. G., W. O. Bearden, S. Sharma. 2003. *Scaling Procedures*. Sage, <https://doi.org/10.4135/9781412985772>.
- O'Connor, B. P. 2000. „SPSS and SAS Programs for Determining the Number of Components Using Parallel Analysis and Velicer's MAP Test.“ *Behavior Research Methods, Instrumentation, and Computers* 32: 396–402, <https://doi.org/10.3758/BF03200807>.
- Osborne, J. W. 2015. „What is Rotating in Exploratory Factor Analysis?“ *Practical Assessment, Research & Evaluation* [online] 20 (2): 1–7. Dostupné na: <http://pareonline.net/getvn.asp?v=20&n=2>.
- Pallant, J. 2005. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS*. 2nd ed. Allen & Unwin.
- Pearson, R., D. Mundfrom, A. Piccone. 2013. „A Comparison of Ten Methods for Determining the Number of Factors in Exploratory Factor Analysis.“ *Multiple Linear Regression Viewpoints* 39 (1): 1–15.
- Rabušic, L., P. Soukup, P. Mareš. 2019. *Statistická analýza sociálněvědních dat (nejen) v SPSS*. Masarykova univerzita.
- Rhetmulla, M., P. E. Brosseau-Liard, V. Savalei. 2012. „When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods Under Suboptimal Conditions.“ *Psychological Methods* 17 (3): 354–373, <https://doi.org/10.1037/a0029315>.
- Ruscio, J., B. Roche. 2012. „Determining the Number of Factors to Retain in an Exploratory Factor Analysis Using Comparison Data of Known Factorial Structure.“ *Psychological Assessment* 24 (2): 282–292, <https://doi.org/10.1037/a0025697>.
- Soukup, P. 2019. „P a d (Používání statistické a věcné významnosti v českých sociálních vědách).“ *Sociologický časopis / Czech Sociological Review* 55 (2): 215–254, <https://doi.org/10.13060/00380288.2019.55.2.459>.
- Soukup, P., I. Kočvarová. 2016. „Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu.“ *Pedagogická orientace* 26 (3): 512–536, <https://doi.org/10.5817/PedOr2016-3-512>.
- Straková, J., J. Simonová, P. Soukup. 2020. „Validizace konceptu akademické marnosti v českém středním vzdělávání.“ *Sociologický časopis / Czech Sociological Review* 56 (5): 599–617, <https://doi.org/10.13060/csr.2020.042>.
- Urbánek, T. 2000. *Strukturální modelování v psychologii*. Pavel Křepela.
- Velicer, W. F. 1976. „Determining the Number of Components from the Matrix of Partial correlations.“ *Psychometrika* 41 (3): 321–327, <https://doi.org/10.1007/BF02293557>.
- Vlčková, K., P. Květon, S. Ježek, J. Mareš, K. Lojdová. 2019. „Adaptace Škály managementu chování a výuky na české podmínky.“ *Studia paedagogica* 24 (1): 135–155, <https://doi.org/10.5817/SP2019-1-6>.
- Winter, J. C. F. de, D. Dodou, P. A. Wieringa. 2009. „Exploratory Factor Analysis With Small Sample Sizes.“ *Multivariate Behavioral Research* 44 (2): 147–181, <https://doi.org/10.1080/00273170902794206>.
- Zwick, W. R., W. F. Velicer. 1986. „Comparison of Five Rules for Determining the Number of Components to Retain.“ *Psychological Bulletin* 99 (3): 432–442, <https://doi.org/10.1037/0033-2909.99.3.432>.

## Články citované v tabulkách 1 a 2

- Bernard, J. 2011. „Endogenní rozvojové potenciály malých venkovských obcí – obtížné hledání a měření jejich vlivu.“ *Sociologický časopis / Czech Sociological Review* 47 (4): 745–775, <https://doi.org/10.13060/00380288.2011.47.4.06>.



- Boberová, Z., I. Ropovik, P. Kolarčík, A. Madarasová Gecková, L. Paakkari. 2019. „Štruktúra zdravotnej gramotnosti u adolescentov.“ *Československá psychologie* 63 (1): 1–12.
- Gocet Tekin, E. 2017. „Shyness and Subjective Well-being: The Mediation Role of Loneliness.“ *Československá psychologie* 61 (4): 321–330.
- Hamplová, D. 2008. „Čemu Češi věří: dimenze soudobé české religiozity.“ *Sociologický časopis / Czech Sociological Review* 44 (4): 703–723, <https://doi.org/10.13060/00380288.2008.44.4.05>.
- Kövero, M. 2018. „Konfirmačná faktorová analýza slovenkej verzie škály profesijnej kvality života (ProQOL).“ *Československá psychologie* 62 (6): 565–578.
- Luthar, B., S. Korpivnik. 2011. „Class, Cultural Capital, and the Mobile Phone.“ *Sociologický časopis / Czech Sociological Review* 47 (6): 1091–1118, <https://doi.org/10.13060/00380288.2011.47.6.01>.
- Miklošová I., M. Makulajová, M. Jakubek. 2016. „Štruktúra jazykovo-kognitívnych schopností u detí s narušeným vývinom reči predškolského veku.“ *Československá psychologie* 60 (3): 252–265.
- Tišanská L., J. Kožený. 2018. „Pohotovosť k agresívnemu chovániu u adolescentů umístěných v nápravných zařízeních.“ *Československá psychologie* 62 (6): 542–551.
- Vidovičová, L., M. Petrová Kafková. 2012. „Aktivity seniorů ve velkých městech: zdraví, prostor a subjektivní kvalita života.“ *Sociologický časopis / Czech Sociological Review* 48 (5): 939–963, <https://doi.org/10.13060/00380288.2012.48.5.06>.
- Vráblíková, K. 2009. „Politická participace a její determinanty v postkomunistických zemích.“ *Sociologický časopis / Czech Sociological Review* 45 (5): 867–897, <https://doi.org/10.13060/00380288.2009.45.5.01>.