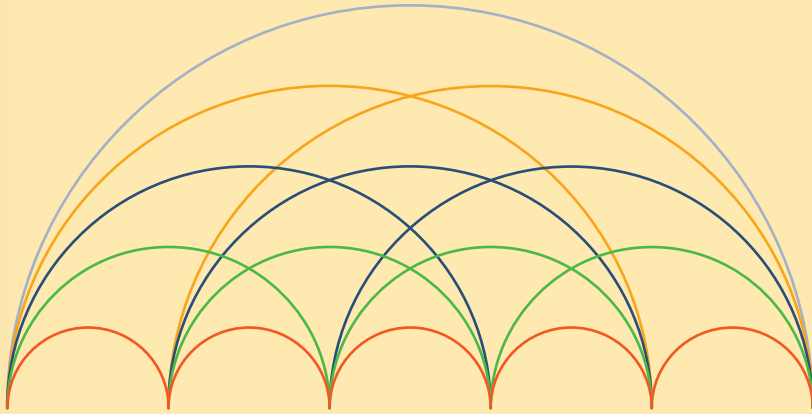
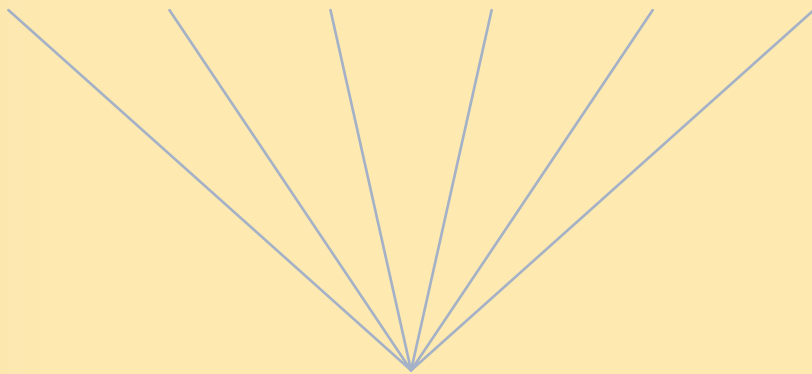


Petr Soukup



Pokročilá analýza dat v SPSS a AMOS



**MASARYKOVA
UNIVERZITA**

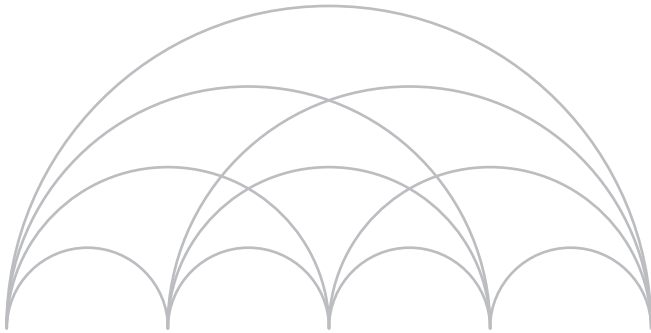
Petr Soukup

Pokročilá analýza dat v SPSS a AMOS

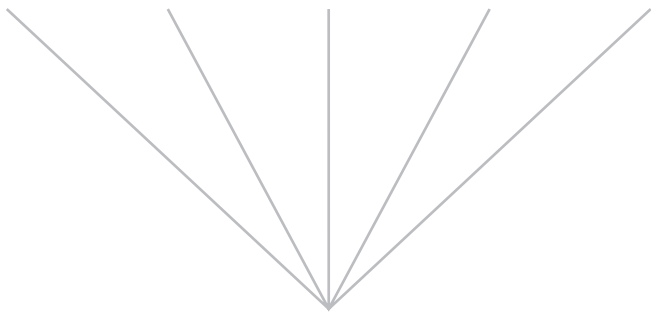
MUNI
PRESS

Učebnice vychází s laskavou podporou společnosti
IBM Česká republika, spol. s r. o.

Petr Soukup



Pokročilá analýza dat v SPSS a AMOS



**Masarykova univerzita
Brno 2022**

KATALOGIZACE V KNIZE – NÁRODNÍ KNIHOVNA ČR

Soukup, Petr, 1976-

Pokročilá analýza dat v SPSS a AMOS / Petr Soukup. -- 1. vydání. -- Brno : Masarykova univerzita, 2022. -- 285 stran

Obsahuje bibliografie, bibliografické odkazy a rejstřík

ISBN 978-80-210-9935-7 (brožováno)

* 303.7 * 519.23 * 004.9:311 * 004.42SPSS * 004.42Amos * 30 * (075.8)

- analýza dat
- statistická analýza
- statistický software
- SPSS (software)
- SPSS Amos (software)
- sociální vědy
- učebnice vysokých škol

311 - Statistika [4]

37.016 - Učební osnovy. Vyučovací předměty. Učebnice [22]

Citace

SOUKUP, Petr. Pokročilá analýza dat v SPSS a AMOS. Brno: Masarykova univerzita, 2022.

ISBN 978-80-210-9935-7.

Knihu recenzovali

prof. RNDr. Jan Hendl, CSc.

Doc. RNDr. Jan Řehák

© 2022 Petr Soukup

© 2022 Masarykova univerzita

ISBN 978-80-210-9936-4 (online ; pdf)

ISBN 978-80-210-9935-7 (brožováno)

Obsah

Úvod	9
Kapitola 1	
Úvod do strukturního modelování	11
Předehra ke strukturnímu modelování a základy práce s AMOS	12
1.1 Vysvětlení podstaty strukturního modelování a dílčích technik	13
1.2 Základní terminologie a konvence pro grafické zachycení modelů v tzv. diagrafech včetně ukázek	14
1.3 Podoby datové matice pro různé modely	19
1.3.1 Výběry a jejich velikost	19
1.3.2 Využití individuálních dat	21
1.3.3 Kovarianční (korelační) podoba dat	24
1.4 Problém identifikace modelů	27
1.5 Odhadovací techniky a typy použitých proměnných	30
1.6 Vyhodnocení modelů a parametrů	33
1.6.1 Chí-kvadrát test	35
1.6.2 Kritéria pro vyhodnocení kvality modelu	36
1.6.3 Vyhodnocení shody velikosti dvou parametrů	42
1.7 Principy budování strukturních modelů	43
Kapitola 2	
Úseková analýza	51
2.1 Úseková analýza a představení úlohy pro analýzu	51
2.2 Přímé, nepřímé a celkové efekty	52
2.3 Postup úsekové analýzy	53
2.3.1 Je náš model úsekové analýzy ve světle našich dat udržitelný?	54
2.3.2 Jak se nám daří vysvětlit jednotlivé endogenní proměnné v našem modelu?	56
2.3.3 Jsou prediktory v našem modelu statisticky a věcně významné?	57
2.3.4 Bylo by vhodné model nějak upravit?	61
Kapitola 3	
Konfirmační faktorová analýza	69
3.1 Úvod do terminologie a představení úlohy pro analýzu	69
3.2 Jednofaktorový model a jeho možná modifikace	73

3.2.1	Je náš model konfirmační faktorové analýzy ve světle našich dat udržitelný? . . .	78
3.2.2	Jak se nám daří vysvětlit jednotlivé endogenní proměnné v našem modelu? . . .	79
3.2.3	Jsou prediktory v našem modelu statisticky a věcně významné?	79
3.2.4	Bylo by vhodné model nějak upravit?	82
3.3	Dvoufaktorový model a jeho modifikace (aneb Mají být faktory vzájemně provázány?)	86
3.4	Konfirmační faktorová analýza druhého řádu (aneb Máme použít jediný faktor?)	94

Kapitola 4

Komplexní strukturní modely	109
4.1 Úvodní motivace a představení analýzy a jejího postupu	109
4.2 Dílčí modely měření	111
4.2.1 Model měření pro individuálně nedeklarovanou stránku kolektivní paměti . . .	111
4.2.2 Model měření pro veřejně deklarovanou stránku kolektivní paměti	116
4.2.3 Model měření pro individuálně deklarovanou stránku kolektivní paměti	120
4.3 Regresní model pro vysvětlení individuálně deklarované paměti (práce s průměrovými indexy)	124
4.4 Komplexní model	126
4.5 Komplexní model bez modifikací jednotlivých modelů měření	129

Kapitola 5

Víceúrovňové lineární modely	135
5.1 Úvodem	135
5.1.1 Využití statistického softwaru pro víceúrovňové modelování	136
5.1.2 Základní myšlenky a zdůvodnění víceúrovňových modelů	136
5.1.3 Podrobnější statistická motivace hierarchických modelů	138
5.2 Vnitrotřídní koeficient korelace a jeho výpočet (nulový model)	142
5.3 Model s jednou vysvětlující proměnnou na první úrovni	147
5.4 Zavedení proměnných na druhé úrovni (model 4)	151
5.5 Model s náhodnou konstantou a proměnnými na první i druhé úrovni (model 5)	154
5.6 Model s náhodnou konstantou i směrnici a proměnnými na první i druhé úrovni (model 6)	155
5.7 Stručné shrnutí pravidel pro budování víceúrovňových modelů a kritéria k porovnání modelů	158
5.8 Stručně o výběrech a jejich velikostech	159
5.9 Varování aneb Nepodlehňte kouzlu víceúrovňového modelování	159
5.10 Smířlivé zakončení aneb V čem spočívají přednosti víceúrovňových modelů?	160
5.11 Přehled základní učební literatury o víceúrovňovém modelování a přehled řešených problémů	161

Kapitola 6

Víceúrovňové modely pro kategoriální závisle proměnné	165
6.1 Úvodem	165
6.1.1 Typické proměnné pro modelování v sociálních vědách	166
6.1.2 Přehled přístupů pro kategoriální data a jejich základní logika	166
6.2 Model pro kategoriální závisle proměnnou – jednoúrovňový přístup	169
6.2.1 Model pro školní doučování s žákovskými prediktory	172
6.2.2 Modely pro školní doučování obsahující školní prediktory	182
6.3 Model pro kategoriální závisle proměnnou – víceúrovňový přístup	184
6.3.1 Nulový model	184
6.3.2 Víceúrovňový model s prediktory na první úrovni (žakovská)	188
6.3.3 Víceúrovňový model s prediktory na druhé úrovni (školní)	194
6.3.4 Víceúrovňový model s prediktory na první i druhé úrovni (žakovská a školní) ..	196
6.3.5 Víceúrovňový model se standardizovanými prediktory na první i druhé úrovni (žakovská a školní)	198
6.4 Víceúrovňový model pro nominální závisle proměnnou: modelování místa, kde žák navštěvuje doučování	201

Kapitola 7

Korespondenční analýza	209
7.1 Úvod	209
7.2 Základní logika korespondenční analýzy pro dvě proměnné	210
7.2.1 Exkurz o logice výpočtu korespondenční analýzy a základní terminologii ..	214
7.3 Příklad analýzy vztahu dvou proměnných: volená strana a vzdělání	218
7.3.1 Základní zadání a interpretace výstupů	218
7.3.2 Normalizace a její dopady	228
7.3.3 Využití doplňkové kategorie	231
7.3.4 Exkurz: Jak použít korespondenční analýzu pro publikovanou kontingenční tabulku?	235
7.4 Vícerozměrná korespondenční analýza (příklad analýzy vztahu tří proměnných) ..	240
7.5 Shrnutí a doporučení pro publikaci výsledků korespondenční analýzy	248

Kapitola 8

Analýza latentních tříd	251
8.1 Úvod	251
8.2 Základní myšlenky lazarsfeldovské analýzy latentních struktur	251
8.3 Analýza latentních tříd	252
8.3.1 Srovnání LCA s LSA	252
8.3.2 Porovnání LCA s faktorovou analýzou	253
8.4 Aplikace LCA na reálných datech (typologie žáků 4. ročníku podle jejich vztahu k matematice)	255
8.4.1 Použitá data a proměnné	255

8.4.2	Typologie za pomoci LCA	256
8.4.3	Shrnutí postupu explorační LCA	261
8.4.4	Modely se dvěma, třemi a čtyřmi třídami	261
8.4.5	Uložení třídní příslušnosti do dat a následné analýzy	268
8.5	Detailní rozbor analýzy latentních tříd a její další možnosti	272
8.5.1	Počet tříd a nepodmíněné pravděpodobnosti	272
8.5.2	Podmíněné pravděpodobnosti náležení do třídy při odpovědích na jednotlivé výroky	273
8.5.3	Explorační a konfirmační verze LCA	273
8.6	Explorační analýza latentních tříd (možnosti a doporučení)	273
8.6.1	Postup a odhadovací techniky	273
8.6.2	Vyhodnocení kvality modelu	274
8.6.3	Zařazení respondenta do latentní třídy pro další práci s daty	274
8.7	Výhody a nevýhody LCA	275
8.8	Počítačové programy pro analýzu latentních tříd	276
8.9	Shrnutí	277
	Rejstřík	283
	Písmena řecké abecedy	285

Úvod

V roce 2015 a podruhé v roce 2019 vyšla v nakladatelství Masarykovy univerzity učebnice věnovaná analýze dat v SPSS od autorského kolektivu dílem z FSS MU (Ladislav Rabušic a Petr Mareš), dílem z FSV UK (Petr Soukup). Učebnice byla čtenáři velice dobře přijata, získala dokonce Cenu Jaroslava Jirsy za nejlepší učebnici v oborech přírodovědných a matematicko-fyzikálních, udělovanou Univerzitou Karlovou. To je pro nás autory samozřejmě velký závazek. S ohledem na vývoj softwaru i statistických přístupů bude třeba brzy začít věnovat se přípravě dalšího vydání této knihy. Kniha pokrývá základní postupy užívané v sociálněvědní analýze dat a končí výklady o některých mnohorozměrných přístupech (regresní analýze, explorační faktorové analýze a seskupovací analýze). Čtenáři logicky požadovali, aby vzniklo pokračování, zaměřené na stále populárnější modelovací přístupy (zejména strukturní a víceúrovňové). Z původního tříčlenného autorského kolektivu se jeho přípravy ujal Petr Soukup, který se zmíněným postupům pravidelně věnuje při výuce i při řešení různých výzkumných projektů. Svoje zkušenosti se rozhodl zúročit v nyní vydávané knize nazvané *Pokročilá analýza dat v SPSS a AMOS*. Kromě teoretických výkladů obsahuje kniha též mnoho řešených úloh, v nichž se čtenář detailně seznámí se zadáním a výstupy příslušné analýzy. Hlavní důraz je přitom kladen na interpretaci výsledků reálných analýz. Formálně je kniha rozdělena do tří větších bloků – strukturní modelování (kapitoly 1–4), víceúrovňové modelování (kapitoly 5–6) a postupy pro práci s kategoriálními daty (kapitoly 7–8). Kniha seznamuje čtenáře s postupy v oblíbených programech od IBM SPSS Statistics (kapitoly 5–8) a SPSS AMOS (kapitoly 1–4). Zájemci mohou využít také doprovodný web i ukázky práce v jiných softwarových nástrojích včetně freeware, konkrétně v JASP, Mplus a R. Doprovodný web nabízí také veškeré materiály používané v knize (datové soubory, skripty aj.).

Těmito dvěma učebnicemi by měla být tematicky pokryta většina procedur a typů úloh, které sociální vědci mohou řešit. To samozřejmě neznamená, že v nich autoři vyložili všechna relevantní témata. V první knize ani v jejím pokračování, které má čtenář právě před sebou, nenalezne podrobné výklady o komplexnějších regresních a příbuzných přístupech: o Coxově regresi, tobitové regresi, regresních a klasifikačních stromech či o oblasti zobecněných modelů (GLM).^{*} Velmi stručně jsou však tyto

^{*} Trochu se s těmito modely seznámí čtenář v kapitole 5.

modely představeny v kapitole 5. Stejně tak se čtenář v předkládané učebnici nedočte nic o složitějších škálovacích postupech – mnohorozměrném škálování, rozprostírání (unfolding) škál nebo teorii odpovědi na položku (item response theory). Ani jedna kniha nerozebírá detailně otázky spojené s volbou velikosti výzkumného souboru, řešení práce s daty, která nepocházejí z randomizovaných experimentů či náhodných výběrů.

Je tedy zřejmé, že i nadále zůstává otevřený prostor pro vznik nejméně jedné další publikace, která by českému zájemci o sociálněvědní analýzu dat zprostředkovala současný stav poznání. Autor cítí v tomto ohledu dluh a slibuje čtenářům, že pokračováním první učebnice celý cyklus nekončí. Ostatně původní autorský kolektiv nedávno přepracoval základní učebnici analýzy dat, aby byla použitelná v prostředí R (zejména s využitím tzv. R commanderu), a kniha bude díky nakladatelství Masarykovy univerzity českému čtenáři brzy k dispozici.

Na závěr je namísto poděkovat právě nakladatelství Masarykovy univerzity za přízeň, kterou knihám o sociálněvědní analýze poskytuje. Jmenovitě patří poděkování zejména paní ředitelce PhDr. Aleně Mizerové a také její zástupkyni PhDr. Lei Novotné. Velký dík zaslouží též vždy skvělá korektorka Mgr. Hana Borovská, Ph.D., a grafička Hana Balášová.

Velmi děkuji i oběma recenzentům, prof. Janu Hendlovi a doc. Janu Řehákovi, za jejich cenné připomínky a komentáře.

Speciální poděkování patří též Míše Spurné, která text přečetla a jako zkušená analytička doporučila mnohé věcné změny.

Za pomoc vděčíme i českému zastoupení firmy IBM, bez jejíhož finančního přispění by bylo vydání knihy jen obtížně představitelné.

Rodině se bohužel většinou neděkuje, tam lze jen prosit za odpuštění.

Petr Soukup, FSV UK

Kapitola 1

Úvod do strukturního modelování

Strukturní modelování je souborem technik, které slouží pro komplexní modelování vztahů mezi proměnnými. Výhodou strukturních modelů je skutečnost, že kromě práce s běžně užívanými proměnnými, které se vyskytují v datové matici (budeme jim říkat manifestní), umožňují i práci s proměnnými, které se v datové matici nevyskytují (budeme jim říkat latentní). Do třídy strukturních modelů patří i některé základní postupy vykládané v běžných kurzech statistiky, jako je korelační analýza a regresní analýza. Strukturní modelování je do jisté míry zobecněním těchto a dalších jednodušších přístupů. V současnosti je strukturní modelování svébytnou subdisciplínou statistiky. Je popsáno v mnoha knihách a nové postupy se většinou objevují na stránkách oborového časopisu Structural Equation Modeling. Aspirací prvních čtyř kapitol této knihy je zasvětit čtenáře do základních postupů v oblasti strukturního modelování, přičemž může čtenář využít i odkazy na pokročilejší texty s komplikovanějšími aplikacemi, které najde na mnoha místech knihy. Prostudování prvních čtyř kapitol této knihy by nicméně mělo být dostatečnou přípravou k tomu, aby čtenář bez větších problémů zvládl formulaci a analýzu komplexních strukturních modelů a byl schopen tyto modely vyhodnotit. Úvodní kapitoly knihy mohou samozřejmě sloužit i zájemcům, kteří nebudou modely používat aktivně, ale chtějí jen porozumět modelům publikovaným jinými autory.

Před meritorním výkladem jednotlivých strukturních modelů je třeba zavést některé pojmy a podat některá vysvětlení tak, aby se nemusela stále opakovat. Půjde o obecné předpoklady pro užívání široké třídy statistických modelů, kterým se většinou obecně říká strukturní modely (SEM = structural equation modeling)¹. Každý, kdo používá statistiku, se s těmito modely v zásadě setkal, často aniž by to tušil, protože, jak již bylo uvedeno, mezi nejjednodušší strukturní modely patří korelace a jednoduchá či vícenásobná regrese. Právě z těchto jednoduchých (a známých) modelů budeme na počátku vycházet a ukážeme na nich principy, které posléze využijeme pro modely složitější.

¹ V ekonometrii bývá zvykem označovat tuto techniku jako simultánní rovnice, v této knize se však podržíme označení strukturní modely

Předehra ke strukturnímu modelování a základy práce s AMOS

Strukturní modelování je poměrně komplexní soubor statistických technik. Tento text je pouhým úvodem a rozhodně si neklade za cíl vyčerpat všechna témata. Pro užívání reálných modelů bude čtenář potřebovat minimálně software (v naší knize budeme využívat zejména AMOS)² a dokumentaci k softwaru. Pro korektní aplikace je také více než vhodné číst výsledky analýz jiných autorů v prestižních časopisech a inspirovat se jimi. Je třeba též pamatovat na to, že méně někdy znamená více, tj. ne vždy musíme užívat komplikované strukturní modely. A i ve strukturním modelování se musíme snažit, aby naše modely byly co nejjednodušší a srozumitelné pro běžné čtenáře. Dříve než začneme s výkladem jednotlivých dílčích technik, je třeba věnovat se některým obecným rysům, se kterými budeme pracovat v následujících třech kapitolách. Počáteční výklad zahrnuje tato témata:

1. Vysvětlení podstaty strukturního modelování a dílčích technik
2. Základní terminologie a konvence pro grafické zachycení modelů v tzv. diagramech včetně ukázek
3. Podoby datové matice pro různé modely
4. Problém identifikace modelů
5. Odhadovací techniky a typy použitých proměnných
6. Vyhodnocení modelů a jejich parametrů včetně ukázky nejjednodušších modelů – lineární regrese
7. Principy budování strukturních modelů

Uvedeným bodům bude v následujících podkapitolách věnován pouze stručný základní výklad. V komplexních textech o strukturním modelování (např. Kline, 2010; Bollen, 1989; Kaplan, 2000) je každému z těchto témat vždy věnováno minimálně sto stran, v detailech lze tedy odkázat na první kapitoly těchto klasických knih.

² Více se o možnostech softwaru čtenář dozví v online dodatku k této knize. AMOS je používán zejména s ohledem na jeho velice snadné ovládání a intuitivnost. Pro jednotlivé kapitoly jsou připraveny online dodatky, které ukazují zadání analýz pro jiné běžně užívané softwary.

1.1 Vysvětlení podstaty strukturního modelování a dílčích technik

Strukturní modely jsou v podstatě komplexním souborem statistických technik. Základem jsou regresní přístupy a faktorová analýza (srov. detailnější vysvětlení dále). K nejjednodušším strukturním modelům patří běžně používané korelace či lineární regresní modely. Nejsložitější modely pak kombinují několik technik v jednom modelu, často navíc využívají data z opakovaných měření či data, která mají víceúrovňový charakter (srov. kapitoly 5 a 6).

Klasicky (a tato kniha není výjimkou) bývají představeny jednotlivé modely postupně v tomto sledu:

- a) úseková analýza (path analysis) v kapitole 2,
- b) konfirmační faktorová analýza (confirmatory factor analysis) v kapitole 3 a
- c) komplexní strukturní modely v kapitole 4.

V této knize jim ještě z pedagogických důvodů předřadíme dva jednodušší modely (korelaci a lineární regresi, viz bod 6, respektive přílohu 2 této kapitoly), na což naváže výklad úsekové analýzy v další kapitole. Stručně zde již nyní vysvětlíme jednotlivé techniky.

Ad a) Úseková analýza je soubor několika (minimálně dvou) na sebe navazujících regresních analýz, tj. závisle proměnná z jedné regresní analýzy se může stát nezávisle proměnnou v další regresní analýze. V rámci strukturních modelů provádíme výpočet všech těchto regresních modelů dohromady a jsme schopni ověřit i celkovou kvalitu modelu (tj. několika regresních analýz najednou). Původcem tohoto přístupu je Wright, který začal úsekovou analýzu rozvíjet již ve 20. letech minulého století (Wright, 1920). Základní verze úsekové analýzy předpokládá, že máme několik měřených spojitých proměnných a teoretickou představu o jejich vzájemném působení, a tuto představu analýzou ověřujeme.

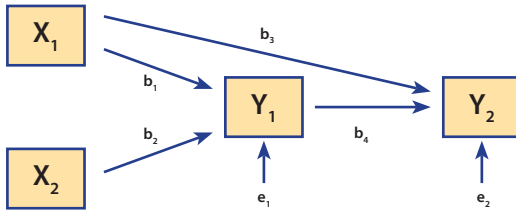
Ad b) Konfirmační faktorová analýza je obdobou explorační faktorové analýzy³ (viz např. Rabušic a kol., 2019, kapitola 15) s tím rozdílem, že na základě předchozích výzkumů či teoretických očekávání předpokládáme, že vazby mezi některými faktory (proměnnými, které nejsou přímo pozorovány, nejsou přímo měřeny) a proměnnými měřenými (indikátory) neexistují, a tuto předpokládanou strukturu ověřujeme (proto má technika přívlastek konfirmační). Někdy se o konfirmační faktorové analýze hovoří v kontextu strukturních modelů příznačně jako o modelu měření, protože tato technika ověřuje, zda lze určité fenomény (např. inteligenci, anomii, depresi, sociální status) měřit určitými indikátory (testovými položkami či otázkami z dotazníku). Historie konfirmační faktorové analýzy sahá do přelomu 60. a 70. let minulého století, základní texty i software (Lisrel) vytvořili Jöreskog a Sörbom.

³ Pro další porozumění je vhodná minimálně základní znalost této techniky a související terminologie.

Ad c) Komplexní strukturní modely jsou spojením regrese či úsekové analýzy a konfirmační faktorové analýzy (tedy dvou předchozích modelů zmíněných v bodech a) a b)). Zjednodušeně řečeno, pod strukturním modelem můžeme tedy rozumět regresní analýzu, ve které použité nezávisle i závisle proměnné nejsou přítomné v našich datech (což je na první pohled bizarní představa). Pro čtenáře s vizuální představivostí dodejme, že jednoduché grafické ukázky úsekové analýzy, konfirmační faktorové analýzy a komplexního strukturního modelu nalezneme v další části věnované základním pojmům a grafickým vyjádřením. Reálné modely jsou pak náplní jednotlivých kapitol.

1.2 Základní terminologie a konvence pro grafické zachycení modelů v tzv. diagrafech včetně ukázek

Nejdříve vysvětlíme několik pojmů, které se týkají proměnných. V rámci strukturních modelů hraje klíčovou roli rozlišení proměnných na manifestní a latentní. Tohoto rozlišení jsme se již dotkli v předchozích odstavcích. **Manifestní** (pozorovaná) **proměnná** je taková, která je přítomná v našem datovém souboru. Typicky je to odpověď na otázku z dotazníku nebo testu, odpovídá jí tedy jeden sloupec v datové matici. **Latentní** (skrytá, nepozorovaná) **proměnná** pak je taková, která v našich datech přímo přítomná není, přesto s ní lze ve strukturních modelech pracovat (je zpravidla odvozená z několika manifestních proměnných). Latentní proměnné užíváme zejména pro fenomény, které nelze měřit přímo (např. inteligence, anomie, deprese, sociální status), nicméně lze je použít i pro fenomény, které přímo měřitelné jsou (např. levičovo-pravicové zaměření jedince). Důvodem pro užití latentních proměnných (nepřímého měření) může být snaha o komplexnější zachycení fenoménu. Z hlediska metod výzkumu lze takto dosáhnout vyšší validity i reliability. Z pohledu statistiky pak dodejme, že při použití latentní proměnné místo manifestní dokážeme vyjádřit tzv. chybu měření, která je vždy ve výzkumu přítomna. Tuto chybu pak lze zohlednit při vyhodnocení strukturního modelu a naše výsledky budou díky tomu přesnější. Proměnné dále rozlišujeme na endogenní a exogenní; toto rozlišení nyní doplní naše dosavadní poznatky z lineární regrese, kde se proměnné rozlišovaly na nezávisle a závisle proměnné. Jak již bylo uvedeno v bodu a) věnovanému úsekové analýze, jedna proměnná může mít v téže analýze roli závisle i nezávisle proměnné. Toto rozlišení tedy nepostačuje. **Exogenní** („vnější“) **proměnná** je taková, kterou **neovlivňuje žádná proměnná z našeho modelu**, **endogenní** („vnitřní“) **proměnnou pak ovlivňuje minimálně jedna proměnná z našeho modelu**. Pro snazší pochopení doplníme grafickou notaci strukturních modelů a na ní vysvětlíme rozdíl mezi endogenní a exogenní proměnnou. Použijeme jednoduchý příklad úsekové analýzy se čtyřmi manifestními proměnnými, značenými pro jednoduchost X_1 , X_2 , Y_1 a Y_2 (viz obrázek 1.1).



Obr. 1.1 Model úsekové analýzy

Z obrázku plyne, že dvě proměnné jsou exogenní (modelem neovlivněné), jsou to proměnné X_1 a X_2 . Z grafického znázornění poznáme exogenní proměnnou jednoduše podle toho, že do ní nesměruje žádná jednosměrná šipka (posléze zjistíme, že tyto šipky symbolizují vazby mezi proměnnými). Proměnné Y_1 a Y_2 jsou endogenní, neboť jsou ovlivněny jinými proměnnými v modelu, konkrétně Y_1 je přímo ovlivněna X_1 a X_2 , Y_2 je přímo ovlivněna Y_1 a X_1 (srov. šipky přímo mezi těmito proměnnými a Y_2 v obrázku), nepřímo ji pak ovlivňují X_2 a X_1 (šipky z nich nejdu do Y_2 přímo, ale přes proměnnou Y_1). Popisu přímých a nepřímých vlivů se budeme věnovat ještě později u výkladu úsekové analýzy, teď se zaměříme na další aspekty popisovaného obrázku. Obrázek respektuje konvenci, že manifestní proměnné, tedy proměnné X_1 , X_2 , Y_1 a Y_2 , kreslíme čtvercem či obdélníkem, a jak bylo uvedeno, najdeme je přímo v naší datové matici.⁴ Když se na náš obrázek podíváme detailně, lze jej rozložit na dva samostatné obrázky pro regresní analýzy se závisle proměnnou Y_1 a se závisle proměnnou Y_2 . Pokud bychom chtěli náš model vyjádřit rovnicově, použili bychom pro každou regresní analýzu jednu rovnici:

$$Y_1 = b_1X_1 + b_2X_2 + e_1$$

$$Y_2 = b_3X_1 + b_4Y_1 + e_2$$





Nyní je třeba alespoň základním způsobem porozumět vazbě mezi rovnicemi a obrázkem a ujasnit si, co vše jsou odhadované parametry našeho modelu (bude se nám to posléze hodit v části věnované identifikaci modelů). V obou rovnicích jsou kromě proměnných regresní koeficienty (označené jako „b“ s indexy 1–4) a také chybové složky (e_1 , resp. e_2). Na obrázku 1.1 jsou regresní koeficienty znázorněny šipkami mezi proměnnými (jde tedy o vazby mezi proměnnými regresního charakteru). Chybové složky jsou pak v našem obrázku vyjádřeny šipkami, které směřují do endogenních proměnných.⁵ Na tomto místě je třeba dvojí upozornění: **Endogenní proměnnou modelem vysvětlujeme a má vždy chybovou složku** (jde o jedno z pravidel budování modelů, které si vysvětlíme na konci kapitoly). **Rozptyl chybové**

⁴ Pro úplnost dodejme, že pro latentní proměnné se užívá kruhů či elips. Brzy se s nimi seznámíme detailněji.

⁵ Tato forma prezentace není jediná možná, posléze si ukážeme alternativní a možná častěji užívanou formu, která je obsažena též v softwaru AMOS, jež budeme používat pro výpočty. V prvním příkladu jsme ale pro jednoduchost zvolili jiné zobrazení.

složky (jako rozptyl jiných exogenních proměnných) **bude vždy společně s regresními koeficienty odhadovaným parametrem** a z jeho hodnoty pak dokážeme vypočítat (nebojte se, software to udělá za vás), s jakou přesností a spolehlivostí je modelem vysvětlena příslušná endogenní proměnná (bude se zde počítat R^2 , které už známe z regrese). Posledními odhadovanými parametry budou rozptyly exogenních proměnných X_1 a X_2 . Když tento popis použijeme na náš model z obrázku 1.1, lze konstatovat, že ve znázorněné úsekové analýze budeme odhadovat čtveřici regresních koeficientů (b_1 – b_4), dále rozptyly dvou chybových složek (pro e_1 , resp. e_2) a rozptyly dvou exogenních manifestních proměnných (X_1 , X_2), celkem tedy osm neznámých parametrů. Pozorný čtenář jistě namítne, že v rovnicích (a na obrázku 1.1) chybí konstanty pro jednotlivé regresní rovnice. Dodejme, že pro jednodušší strukturní modely nejsou konstanty potřebné (výpočet s nimi je komplikovanější a žádá si více vstupních informací). Proto budeme v této knize předpokládat, že konstanty v rovnicích nepoužíváme, resp. je neodhadujeme.

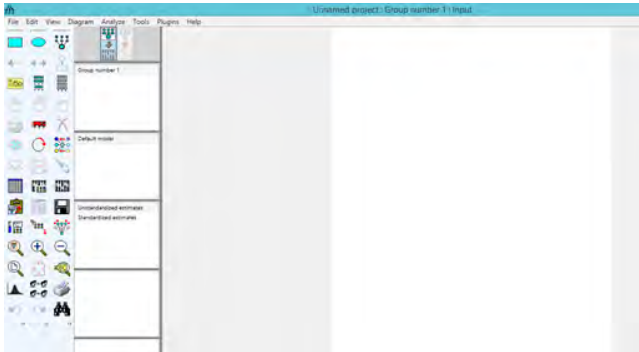
Po poměrně hutném příkladu se nyní seznámíme s přehledem grafických symbolů, které se běžně pro zadání i prezentaci strukturních modelů používají.⁶

-  čtverec/obdélník pro manifestní proměnné
-  kruh/elipsa pro latentní proměnné; typicky je užíváme pro faktory ve faktorové analýze a pro chybové složky u endogenních proměnných
-  šipka jednosměrná – regresní vazba, šipka směřuje od nezávisle proměnné k závisle proměnné
-  šipka obousměrná – kovarianční a korelační vazba

Kromě symbolů je vhodné seznámit se i se základní konvencí pro značení proměnných v rámci strukturního modelování. Pro pozorované exogenní a endogenní proměnné užíváme nejčastěji písmena X, Y, případně Z. Pro značení latentních proměnných ve faktorové analýze užíváme nejčastěji písmene F (factor) a pro značení chybových složek (mají opět latentní charakter) napříč modely je nejčastější písmeno E (error).⁷ Dobrá zpráva je, že minimálně značení latentních proměnných je v softwaru AMOS automatické, tj. software tyto proměnné označí za nás (více v kapitole 3 Konfirmační faktorová analýza). Nicméně již nyní si ukážeme základy práce s AMOS, aby bylo posléze možné již počítat s reálnými daty. Po spuštění AMOS (modulu Graphics) získáme prázdnou plochu jako na obrázku 1.2.

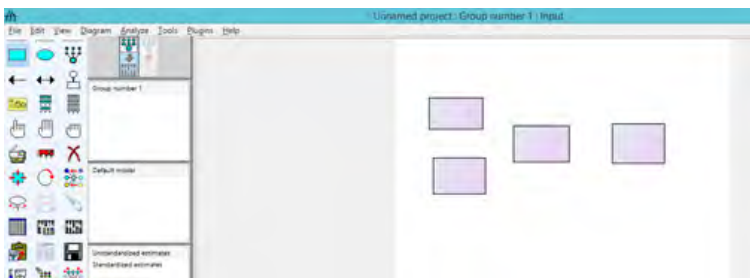
⁶ Máme pro čtenáře první dobrou zprávu: Pro zadání modelu ani jeho prezentaci nemusíte ovládat rovnice. Stačí základy kreslení, a software udělá vše za vás. Nicméně je vhodné rovnicím alespoň základním způsobem rozumět, jinak analytik klouže jen po povrchu a netuší, co se děje pod ním. V knize se budeme snažit omezovat rovnicové vyjádření na nezbytném minimum, doporučujeme mu ale věnovat pozornost.

⁷ Poměrně často je užíváno též písmeno U (U = uniqueness).



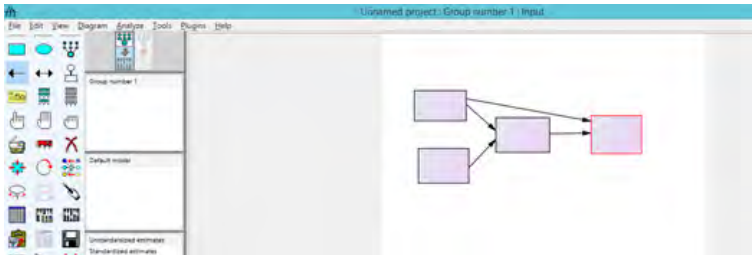
Obr. 1.2 Plocha pro zadání modelu v softwaru AMOS

Vlevo a nahoře máme základní volby v podobě ikon a menu. Celý zbytek obrazovky vyplňuje plocha určená k nakreslení modelu. Právě tuto plochu musíme naplnit naší modelovou představou. Pokud bychom chtěli například nakreslit model z obrázku 1.1, bude zapotřebí nakreslit nejprve obdélníky pro jednotlivé manifestní proměnné. To technicky uděláme tak, že z ikon vlevo vybereme první nahoře (obdélník) a po kliknutí na levé tlačítko myši můžeme začít kreslit (vždy tak, že na ploše začínáme od levého horního rohu, poté držíme levé tlačítko myši, které pustíme, až ukotvíme pravý dolní roh obdélníku). Takto můžeme nakreslit čtyři obdélníky pro čtyři manifestní proměnné (obrázek 1.3). Poté musíme opět kliknout na ikonu s obdélníkem („vrátit obdélníky na své místo“). Tak AMOS zjistí, že nyní chceme užít jiné grafické prvky.



Obr. 1.3 Manifestní proměnné našeho modelu v softwaru AMOS

Další část modelu na obrázku 1.1 tvoří vazby mezi proměnnými, musíme tedy vybrat jednosměrnou šipku (ikona vlevo ve druhé řadě) a obrázek doplnit (obrázek 1.4). Šipku vždy kreslíme od proměnné, od níž směřuje, a končíme u proměnné, ke které směřuje (provedeme to opět držetím levého tlačítka myši).

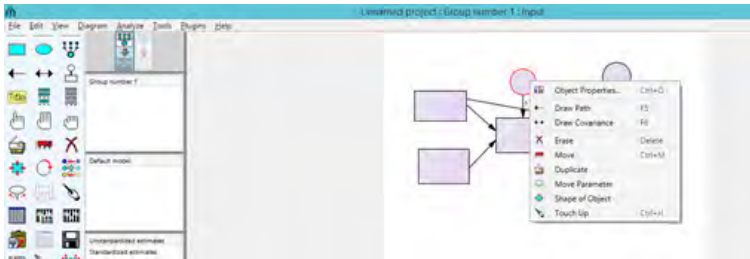


Obr. 1.4 Doplnění vazeb mezi manifestními proměnnými v našem modelu v softwaru AMOS

Poslední prvky, které jsou v modelu na obrázku 1.1 znázorněny pouze schematic-ky, jsou chybové složky pro proměnné uprostřed a vpravo (jde o endogenní, modelem určené proměnné, které musí mít vždy chybovou složku). AMOS nevystačí s šipkovou konvencí, která byla představena na obrázku 1.1, ale požaduje, abychom nakreslili latentní proměnnou navázanou na proměnnou endogenní, a navíc velikost vazby určili (typicky na hodnotu 1). To by znamenalo provést postupně 3 kroky, AMOS nám ale pomůže ikonkou, která vše udělá najednou (druhá řada vpravo, ikona). Získáme tak obrázek 1.5. Všimněte si, že AMOS automaticky kreslí chybové složky vždy nad endogenní proměnné (pohyb je možný skrze ikonu s kamionem v páté řadě uprostřed). Poslední, co si nyní předvedeme, bude pojmenování chybových složek (pojmenujeme je E_1 , resp. E_2). Stačí vždy kliknout pravým tlačítkem myši na příslušný kruh (obrázek 1.6) a z nabídky vybrat první volbu *Object Properties*. Zde do pole *Variable name* napíšeme: E_1 , resp. E_2 (obrázek 1.7). Samozřejmě bychom mohli pojmenovat zbylé proměnné tvořené obdélníky, ale to v praxi neděláme (k ničemu by nám to nepomohlo). Jde o pozorované proměnné (obsažené v našich datech), a proto tyto proměnné načteme s názvy přímo z datové matice (viz další výklad). Nyní tedy umíte první krok práce s AMOS, specifikaci modelu diafragem.



Obr. 1.5 Doplnění chybových složek do našeho modelu v softwaru AMOS



Obr. 1.6 Možnost změn u jednotlivých proměnných v softwaru AMOS



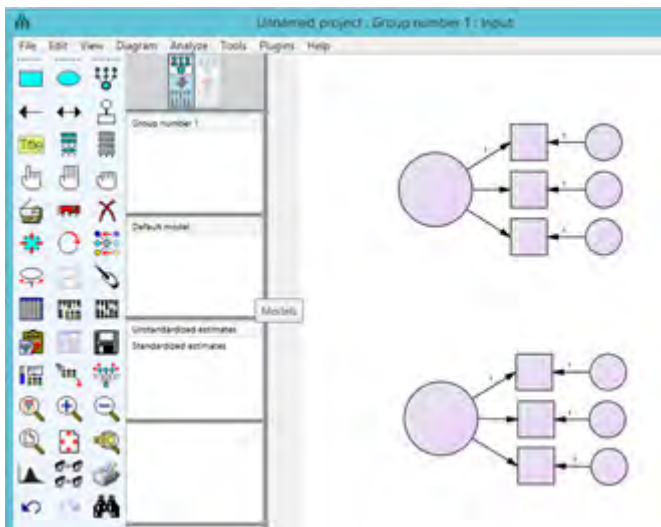
Obr. 1.7 Vepsání názvu latentní proměnné (Variable name) v softwaru AMOS

1.3 Podoby datové matice pro různé modely

1.3.1 Výběry a jejich velikost

Před výkladem o využití individuálních dat a jejich načtení v AMOS je vhodné krátce zmínit problematiku výběrů a jejich velikostí. Je nutné si uvědomit, že strukturní modely běžně budujeme na základě výběrových dat a naším cílem je výsledky zobecnit na celý základní soubor (cílovou skupinu). Není zde prostor podrobně pojednávat o všech detailech ohledně výběrů, můžeme čtenáře odkázat na publikace zabývající se tímto tématem (Kalton, 1983; Kish, 2014; Thompson, 2002), případně na český přehledový článek Soukupa a Kočvarové (2016). Výběry můžeme na základní úrovni rozlišovat na náhodné (pravděpodobnostní) a záměrné. Pro náhodné výběry jsou vyvinuty postupy statistické indukce (statistické testy, případně intervaly spolehlivosti), které umožňují zobecňování z výběru na celou populaci. Pro záměrné výběry takové postupy k dispozici nemáme a ani je vytvořit nelze. Z toho plyne poměrně jednoduché doporučení: chceme-li pracovat se strukturními modely a používat v nich statistické testy (pokud tedy modely neslouží k pouhému popisu vztahů v našich datech), měli bychom mít k dispozici data z náhodného výběru.

S výběrem, resp. jeho typem, souvisí i otázka velikosti výběrového souboru, tj. počtu jednotek, které soubor obsahuje. Detaily lze opět nalézt v odborných knihách (Cohen, 1988; Kraemer & Blasey, 2016), některá základní doporučení i v již citovaném článku Soukupa a Kočvarové (2016). Zaměříme se teď na několik jednodušších doporučení ohledně velikosti výběrového souboru pro strukturní modelování a poté uvedeme odkaz na složitější postupy, které využívají koncept síly testu. Před konkrétními doporučeními uvedme obecný princip. Velikost výběrového souboru by měla být stanovena před samotnou realizací výzkumu, tj. ve stadiu, kdy uvažujeme o designu našeho výzkumného šetření. Již v této fázi bychom měli vědět, že pro naše zkoumané hypotézy bude vhodné využít strukturního modelování, měli bychom zhruba vědět, jakou podobu bude mít náš model, a nadto i umět orientačně odhadnout, jaké výsledky nám poskytně.⁸ Jednoduché pravidlo, které se používá v učebnicích regresní či faktorové analýzy, říká, že minimální velikost výběrového souboru by měla být 10násobkem počtu odhadovaných parametrů (co jsou odhadované parametry, si řekneme v dalších částech knihy). Pro ilustraci trochu předběhneme výklad a podíváme se na model konfirmační faktorové analýzy se šesti indikátory a dvěma faktory (obrázek 1.8).



Obr. 1.8 Model konfirmační faktorové analýzy s šesti indikátory

V tomto modelu je třeba odhadnout **čtyři jednosměrné vazby** (proč jsou dvě z nich nastaveny na jedničku, si povíme v kapitole 3 Konfirmační faktorová analýza), dále pak **rozptyly šesti chybových složek** a **dvou latentních proměnných**. Celkem **4 + 6 + 2**, tj. **12 parametrů**, **minimální velikost výběru** by tedy byla **120 jednotek**


⁸ Jsme si vědomi toho, že to čtenáři může připadat příliš náročné, nezbývá však než se s tím smířit. Druhou možností je rezignace na používání složitějších analytických postupů.

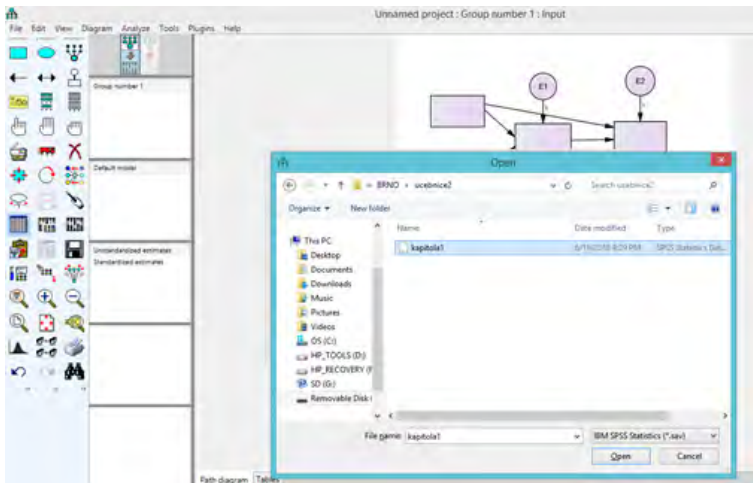
(respondentů). V literatuře lze nalézt i jiná zjednodušující doporučení, nejčastěji se týkají faktorové analýzy a jsou absolutní. Objevuje se například doporučení, že výběr by měl obsahovat minimálně 200 jednotek. S ohledem na detailní zkoumání tématu se domníváme, že tato absolutní doporučení jsou nevhodná a že relativní pohled, který jsme představili v předchozích odstavcích (založený na počtu odhadovaných parametrů), je výhodnější.

Ještě lepší přístup je založen na přesném výpočtu. Tento výpočet vychází z tzv. síly testu, tj. pravděpodobnosti zamítnutí nulové hypotézy, pokud tato hypotéza neplatí (více viz např. Hendl, 2015). Cohen (1988) a mnozí další totiž doporučují, že síla námi prováděných statistických testů by měla dosahovat minimálně hodnoty 0,8, a z této velikosti je podle nich třeba odvozovat velikost výběrového souboru. U jednodušších statistických technik to lze provést pomocí vzorců (Cohen, 1988; Kraemer & Blasey, 2016), u strukturních modelů ale takové vzorce neexistují a velikost výběru se musí spočítat pomocí Monte Carlo simulace. U strukturních modelů musíme mít jasnou představu o podobě našeho modelu, předpokládaných výsledcích a zadat cílovou sílu testu. Po provedení simulací (typicky se simulují řádově tisíce až desetitisíce výběrů odpovídajících našemu modelu) software uvede vhodnou velikost výběru. Poznamenejme, že AMOS užívaný v této knize toto bohužel neumí, lze odkázat např. na software Mplus, případně balíčky v R.

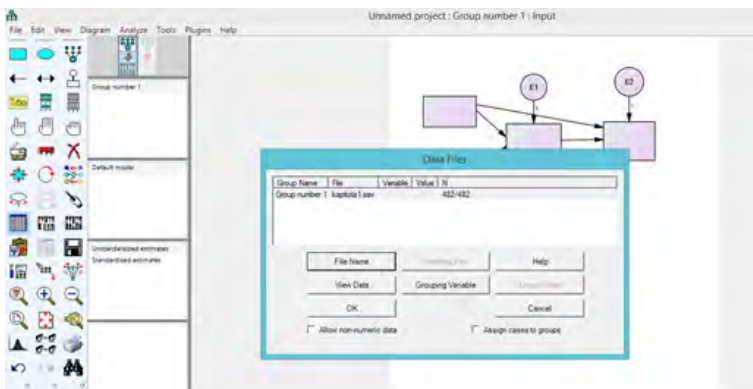
1.3.2 Využití individuálních dat

Nejběžnější podobou datové matice jsou individuální data, tj. data zaznamenaná od matice, kde jednotlivé řádky odpovídají případům (v sociálních vědách typicky respondentům výzkumu) a sloupce jednotlivým proměnným (víme, že těm, kterým říkáme manifestní). Různé softwary pro strukturní modely mají různé požadavky na vstupní formát těchto individuálních dat. AMOS je jeden z nejflexibilnějších a umí pracovat s daty pro Excel (*.xls), pro databázové systémy Foxpro či dBase (*.dbf), dnes již nepoužívaného předchůdce Excelu Lotus (*.wk1-*, wk4) a CSV formát. Pro uživatele SPSS je určitě velkou výhodou, že AMOS umí pracovat též s daty pro SPSS (*.sav). Poté, co jsme specifikovali model (viz kapitola 1.2), je vhodné určit datový soubor, se kterým chceme pracovat, a z něj vybrat proměnné pro analýzu (vyplnit jimi prázdné obdélníky v našem grafu). Budeme i nadále sledovat náš modelový obrázek 1.1. Pro tyto účely potřebujete mít k dispozici soubor kapitola1.sav, který je uložen na doprovodném webu k této knize.


Prvním krokem je nalezení datového souboru v našem počítači. K tomu využijeme ikonu vlevo v sedmé řadě (symbol tabulky ). Poté vybereme *File Name* a najdeme a vybereme soubor kapitola1.sav a stiskneme *Open* (obrázek 1.9). Náš datový soubor má 482 případů, což nám AMOS na další obrazovce správně určí (obrázek 1.10).

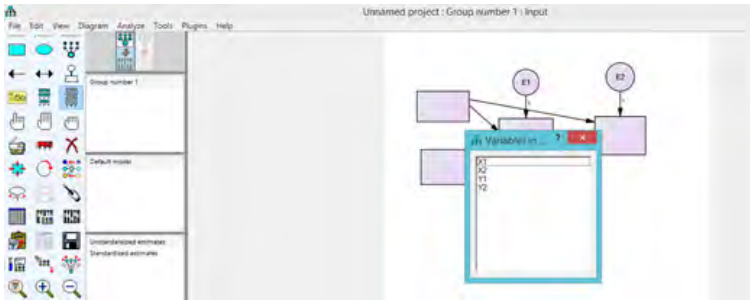


Obr. 1.9 Otevírání datového souboru v AMOS



Obr. 1.10 Vizualizace v AMOS po otevření datového souboru


Stiskneme OK a AMOS je připraven pracovat s daty, která jsme vybrali. Nyní je třeba vyplnit do obdélníků manifestní proměnné z našich dat (bystrý čtenář již jistě v SPSS zjistil, že náš datový soubor má právě čtyři proměnné, označené X_1 , X_2 , Y_1 a Y_2). Pro „nahrání“ proměnných do diagramu využijeme ikonu ve třetí řadě zcela vpravo (bílý seznam proměnných ). Po jejím stisknutí získáme na kreslicí ploše seznam proměnných v našich datech (obrázek 1.11). Proměnné teď již za pomoci levého tlačítka myši posuneme na svá místa, abychom dostali výsledný model, inspirovaný obrázkem 1.1 (obrázek 1.12). Na konci této práce zavřeme dialog s názvy proměnných.



Obr. 1.11 Zobrazení seznamu proměnných v datech

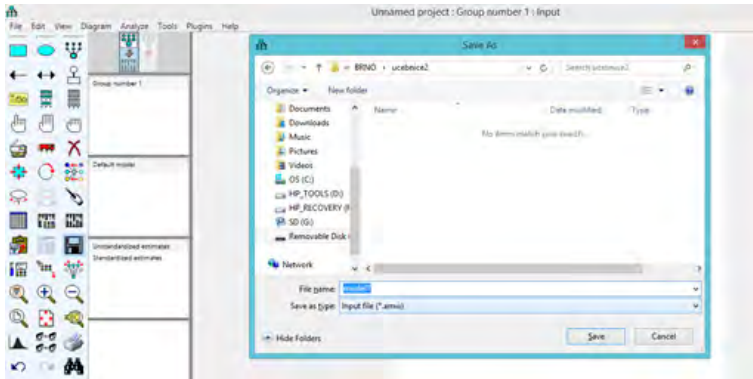


Obr. 1.12 Model po vložení proměnných z dat do obdélníků pro manifestní proměnné

Nyní tedy máme definovaný model a z dat určeny všechny manifestní proměnné (latentní proměnné musí mít určené názvy – viz dřívější E_1 , resp. E_2). Posledním krokem je uložení výsledného modelu včetně vazby k naší datové matici.⁹ Uložení provedeme nejspíše pomocí ikony s disketou (sedmá řada zcela vpravo ). AMOS po nás požaduje zadání názvu souboru (např. model1), automaticky pak doplní koncovku *.amw. Výhodou je, že po otevření souboru model1.amw bude již definován model i data a bude možné počítat výsledky.¹⁰ My si ale ještě na výsledky počkáme a řekneme si něco více o datech (při prvním čtení knihy doporučujeme následující pasáž o kovarianční matici přeskočit) a dalších krocích výpočtu (při prvním čtení doporučujeme čtenáři pokračovat studiem Kapitoly 4 Komplexní strukturní modely).

⁹ Není to nezbytně nutné, ale doporučujeme uložit model (soubor *.amw) do stejného adresáře, kde je uložený datový soubor.

¹⁰ Pokud jsme výsledky již jednou počítali, pak si je AMOS už uložil a výsledky je možné otevřít přímo, aniž bychom museli provádět opětovné výpočty. To je velká přednost AMOS.



Obr. 1.13 Uložení modelu včetně propojení na data v AMOS

1.3.3 Kovarianční (korelační) podoba dat

Kromě individuální datové matice (viz předchozí příklad) je ve strukturním modelování poměrně obvyklé, že užíváme dat ve formě korelační nebo kovarianční matice. To je poměrně dobrá zpráva, protože pro analýzu nemusíme mít originální individuální data, ale postačí matice, které dost často najdeme v publikovaných článcích či knihách.

Než si ukážeme praktický příklad, je třeba několik vysvětlení. Cílem strukturního modelování obecně je vysvětlit vztahy mezi proměnnými (víme již, že proměnné mohou mít manifestní či latentní charakter). Zdrojem pro výpočty strukturních modelů je kovarianční matice, tj. tabulka, která ukazuje provázanost mezi jednotlivými manifestními proměnnými (měřené kovariance) a obsahuje též rozptyly jednotlivých manifestních proměnných. Právě z této vstupní informace a naší modelové představy (ztělesněné diafragem či rovnicemi) se odhadují jednotlivé parametry (koeficienty). S jejich pomocí se pak spočte tzv. reprodukováná kovarianční matice (tj. matice založená na našem modelu a odhadu parametrů). Cílem samozřejmě je, aby kovarianční matice, která je založená na našich datech (vstupní informace), byla co nejpodobnější reprodukováné kovarianční matici (výstupu našeho modelování). K posouzení slouží různé dále uvedené testy a kritéria (viz kap. 1.6 Vyhodnocení modelů a parametrů). Nyní je namístě připomenout si koncept kovariance a poté ukázat, jak lze kovarianční matici prakticky využít v AMOS.

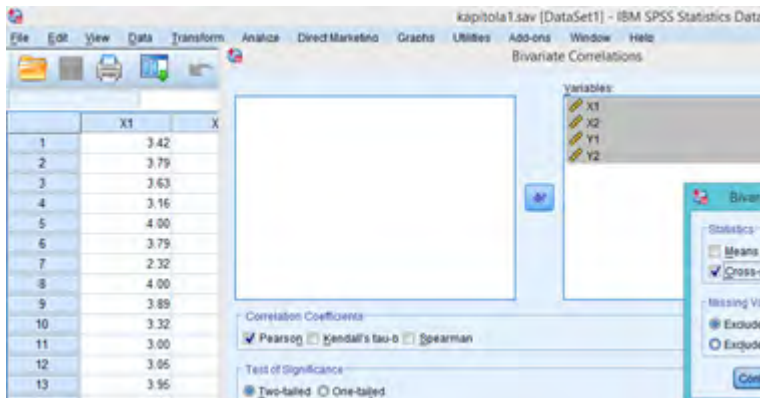
Kovariance je obecná míra lineární souvislosti pro dvě proměnné.¹¹ Praktický výpočet je založen na násobení odchylky od průměru jedné proměnné a odchylky od průměru druhé proměnné, přičemž tyto odchylky sečteme a podělíme velikostí souboru zmenšenou o jednotku ($n - 1$):

¹¹ Pro jistotu doplníme, že vzhledem k tomu, že kovariance využívá ve výpočtu průměrů, jde o míru pro kardinální proměnné. Simulace nicméně ukazují, že ji lze využít s jistou mírou nepřesnosti i pro proměnné ordinálního charakteru s delšími stupnicemi (min. 5 bodů).

$$\text{cov}(x, y) = \Sigma \quad (1.1)$$

Kovarianci tedy můžeme nahlížet jako společný rozptyl, tj. míru toho, co mají dvě proměnné společného. Teoreticky může nabývat hodnot od $-\infty$ do $+\infty$ s tím, že kladné hodnoty značí pozitivní souvislost (při zvýšení hodnot jedné proměnné má druhá tendenci též růst a vice versa), záporné hodnoty pak negativní souvislost (při zvýšení hodnot jedné proměnné má druhá tendenci klesat).

Ukažme si pro názornost, jak by vypadala kovarianční matice pro naše proměnné ze souboru kapitola1.sav. Jde o tabulku, jejíž řádky a sloupce tvoří proměnné X_1 , X_2 , Y_1 a Y_2 , a v kombinacích můžeme nalézt kovariance, resp. rozptyly. V SPSS lze výpočet kovarianční matice získat v proceduře *Analyze-Correlate-Bivariate* tím, že v rámci *Options* zaškrtneme volbu *Cross-product deviations and covariances* (viz obrázek 1.14).



Obr. 1.14 Zadání výpočtu kovariance v SPSS

Výsledná kovarianční matice (tabulku v SPSS je nutno upravit, aby obsahovala jen kovariance) vypadá následovně:

	X1	X2	Y1	Y2
X1	.195	.040	.079	-.038
X2	.040	.103	.050	-.017
Y1	.079	.050	.227	-.109
Y2	-.038	-.017	-.109	.137

Tab. 1.1 Kovarianční matice pro čtyři proměnné ze souboru kapitola1.sav

Z kovarianční matice snadno nahlédneme, že proměnné X_1 , X_2 , resp. Y_1 , jsou pro-vázány pozitivně, Y_2 pak souvisí se všemi proměnnými negativně. Všimněte si, že ko-variance je symetrická, tj. například vazba mezi X_1 a X_2 je shodná jako vazba mezi X_2

a X_1 . Toho posléze využijeme a duplicitní část získané kovarianční matice smažeme, protože ji AMOS nepotřebuje. Z hlediska rozptylů (viz hlavní diagonálu matice, tj. od levého horního rohu do pravého dolního rohu) můžeme zjistit, že největší rozptyl má proměnná Y_1 , nejmenší pak X_2 .

Aby bylo možné kovarianční matici použít v AMOS, je třeba ji transformovat do některého z formátů, který AMOS běžně používá (viz popis u individuálních dat výše), nejpraktičtější je zřejmě užít Excel (uložte vždy ve starším formátu s koncovkou *.xls). Výsledný tvar, který AMOS umí použít, bude následující:

rowtype_	varname_	X1	X2	Y1	Y2
N		482	482	482	482
cov	X1	0,195			
cov	X2	0,040	0,103		
cov	Y1	0,079	0,050	0,227	
cov	Y2	-0,038	-0,017	-0,109	0,137

Všimněte si, že kromě smazaných duplicitních kovariancí (vše nad hlavní diagonálou) je nahoře námi doplněn řádek N, kde je uveden počet respondentů, z nichž je kovarianční matice spočtena (tento údaj AMOS potřebuje k výpočtům statistických testů). Doplněn je též první sloupec, kde je kromě zmíněného N na dalších řádcích napsáno „cov“, což nese informaci pro AMOS, že naše hodnoty jsou hodnoty kovariancí. Ještě jedna změna je potřebná pro AMOS, resp. Excel. Původní desetinné tečky z SPSS je třeba nahradit desetinnými čárkami a před hodnoty, které začínají desetinnými místy, je třeba přidat před desetinnou čárku nuly.¹² Výsledný soubor má čtenář k dispozici, pro jednoduchost nese jméno kapitola1.xls. Pro úplnost uvedme, že by bylo možné použít též korelační matici.¹³ V tom případě je nutno přidat na poslední řádek ještě hodnoty směrodatných odchylek jednotlivých proměnných a místo „cov“ vepsat do jednotlivých řádků na počátek „corr“. Ukázku najde čtenář v první příloze této kapitoly. Obecně se doporučuje používat raději kovarianční matice než korelační, proto v této knize korelační matice užívat nebudeme.

¹² Tento postup lze snadno provést automaticky. Pomocí funkce nahrazování (Ctrl + H) necháme nahradit řetězec „.“ řetězcem „,“.

¹³ Korelace je standardizovaná kovariance, tj. kovarianci dělíme součinem směrodatných odchylek obou proměnných. Korelace má díky tomu hodnoty mezi -1 a $+1$. Více se lze dozvědět v základních učebnicích statistiky.

1.4 Problém identifikace modelů

Pro jednodušší statistické techniky (t-testy, analýzu rozptylu, regresní analýzu) existují vzorce, které umožňují přímo vypočítat výsledky (odhadnout neznámé parametry), nadto jsou tyto jednodušší techniky vždy vyjádřené jako jedna rovnice (lze tedy počítat i s pomocí tužky a papíru). U strukturního modelování je situace složitější, protože rovnic je zpravidla více než jedna a vzorce pro přímý výpočet většinou neexistují (potřebujeme tedy vždy výpočetní techniku). Navíc se poměrně snadno může stát, že naše modelová představa je natolik komplikovaná, že výpočet vůbec provést nelze, odborně tomu říkáme, že model není identifikován. Zde alespoň v krátkosti představíme tento koncept, ve zbytku se spolehneme na výpočetní techniku, která nám sdělí, zda v našem případě jsou jednoznačné odhady neznámých parametrů možné. Zvědavé čtenáře pak můžeme odkázat na detailnější výklady (Urbánek, 2000; Kline, 2010).

Začneme s jednoduchým motivačním příkladem. Mějme následující rovnici:

$$2A + 3B = 20$$

Najít jednoznačné řešení této rovnice (hodnoty neznámých A a B) není možné (řešení je nekonečně mnoho, např. podle volby hodnoty A vyjde vždy hodnota B). Máme k dispozici jednu rovnici, ale dvě neznámé (A, B). V matematice je ostatně známa tzv. Frobeniova věta, která říká, že aby byla soustava rovnic potenciálně řešitelná, může mít nejvýše tolik neznámých, kolik je rovnic (jde o podmínku nutnou, nikoli postačující!). Nyní tuto logiku uplatníme na strukturní modely. Roli jednotlivých rovnic a jednotlivých neznámých musíme ale trochu pozměnit. Neznámými jsou ve strukturních modelech parametry, jejichž hodnotu chceme znát. Typicky to jsou regresní či korelační koeficienty, případně rozptyly pro latentní či exogenní proměnné. Pokud například budeme chtít zjistit, kolik parametrů je třeba odhadnout na obrázku 1.1, musíme spočítat jednotlivé šipky mezi manifestními proměnnými (jsou čtyři) a přidat čtyři neznámé velikosti odhadu rozptylu exogenních proměnných (dvou chybových složek a X_1 , resp. X_2). Dojdeme tedy k celkovému počtu 8 neznámých. Zbývá vysvětlit analogii k počtu rovnic ze světa matematiky. Počet rovnic ve strukturním modelování (resp. počet prvků, které přinášejí informaci o našich datech) je dán počtem prvků kovarianční matice, které jsou jedinečné. Z tabulky 1.1 zjistíme, že máme 6 kovariancí a 4 rozptyly, celkem tedy 10 rovnic (resp. jedinečných prvků, které přinášejí informaci o našich datech). Obecně spočteme počet jedinečných prvků pro k proměnných pomocí vzorce $k*(k + 1)/2$, kde k znamená počet manifestních proměnných v modelu. Máme tedy více rovnic (deset) než neznámých (osm), tudíž existuje šance, že se nám podaří neznámé parametry odhadnout (více se o technikách odhadu dozvíte v další kapitole). Rozdíl mezi počtem jedinečných prvků kovarianční matice a počtem odhadovaných parametrů nazýváme počtem stupňů volnosti (degrees of freedom), v našem případě činí počet stupňů volnosti (df^{14}) = $10 - 8 = 2$. Nyní víme, že počet

¹⁴ Pro počet stupňů volnosti se často užívá zkratka df , z anglického výrazu degrees of freedom.