

# Možnosti praktické práce s daty z mezinárodních vzdělávacích studií: problémy a jejich praktická řešení<sup>1,2</sup>

Petr Soukup

Univerzita Karlova v Praze, Fakulta sociálních věd

**Abstrakt:** Česká republika se již od poloviny devadesátých let minulého století účastní několika mezinárodních vzdělávacích studií, zejména TIMSS, PIRLS, PISA, ICCS, PIAAC a ICILS. Po ukončení jednotlivých cyklů studie jsou vždy datové soubory za všechny účastnické země dostupné všem (Soukup, 2012). Nicméně čeští pedagogičtí výzkumníci dosud tato data používají velice zřídka (spíše přejímají výsledky), ačkoli jsou velice bohatá a umožňují používání složitých statistických přístupů. Komplikace při práci s nimi vyplývají zejména z těchto skutečností:

(a) data obsahují váhy, které zohledňují nestejnou pravděpodobnost výběrání jednotlivých respondentů, a tyto váhy je nutné užívat pro korektní odhady populačních parametrů; (b) data pocházejí z vícestupňových náhodných výběrů a je nutné upravovat standardní chyby odhadů pro výpočty statistických testů nebo intervalů spolehlivosti (využívají se replikační přístupy, nejčastěji jackknife); (c) mnohost dat na národní a mezinárodní úrovni (tj. reálně existují desítky datových souborů a jejich spojování není zcela triviální); (d) některé proměnné neměříme přímo (typicky výsledky v kognitivních testech) je třeba zohlednit chybu měření těchto latentních proměnných (využívá se IRT metodologie a tzv. plausible values).

Cílem textu bude na několika příkladech analytických procedur ukázat správné postupy práce s daty z mezinárodních vzdělávacích šetření a zároveň podat přehled dostupného softwaru pro zpracování těchto dat, aby je mohli čeští pedagogičtí výzkumníci zcela běžně využívat.

**Klíčová slova:** mezinárodní vzdělávací studie, spojování dat, výběrové váhy, replikační váhy, jackknife, latentní proměnná, vícenásobné odhady

## Possibilities of Practical Work with Data from International Large Scale Educational Assessments: Problems and Practical Solutions

**Abstract:** The Czech Republic participated in the last 20 years in several international large scale educational assessment, especially TIMSS, PIRLS, PISA, ICCS, PIAAC and ICILS. Data for all these surveys are always publicly available after completion of each cycle (Soukup, 2012). However, Czech educational researchers used data very rarely (mostly only commented published results), although this data is very rich, and allow the use of complex statistical approaches. Problems when working with these data arise mainly from the following factors:

(a) The data include weights, which take into account unequal probabilities of selection for individual respondents, and these weights should be used to correct estimates of population parameters; (b) the data come from a multistage sampling, and it is necessary to modify the standard error estimates for the calculation of statistical tests or confidence intervals (the use replication approaches, e.g. Jackknife); (c) the multiple data files at national and international level (i.e. there are actually dozens of datasets and their merge are not entirely trivial); (d) some variables not measured directly (typically results in cognitive tests) and should be considered a measurement error of these latent variables (IRT methodology and plausible values can be used).

<sup>1</sup> Výzkum je výstupem projektu GA ČR *Vztahy mezi dovednostmi, vzděláváním a výsledky na trhu práce: longitudinální studie* (P402/12/G130).

<sup>2</sup> Doprovodné video k článku s postupy práce v IDB Analyzer bude dostupné na Youtube kanálu Institutu sociologických studií (<https://www.youtube.com/user/ISSfsvUK>).

98 The aim of the text will be to serve several examples of analytical procedures to demonstrate the proper procedures for working with data from an international large scale educational assessment and also to provide an overview of available software for processing such data so that Czech educational researchers can routinely use.

**Keywords:** international large scale assessment, data merge, sampling weights, replication weights, jackknife, latent variable, plausible value

**DOI:** 10.14712/23363177.2016.15

Cílem tohoto textu je navázat na dosud česky publikované práce, které podávají přehled mezinárodních vzdělávacích studií (Soukup, 2012; Basl, 2014), a ukázat reálné možnosti práce s daty z těchto studií. Článek se zaměřuje na jednodušší statistické techniky (výpočty průměrů a statistické testování odlišnosti pro jednotlivé skupiny, tvorbu kontingenčních tabulek, korelační analýzu a regresní analýzu), konkrétně je využíván software IDB Analyzer, který spolupracuje se statistickým programem SPSS. Práce s daty z mezinárodních vzdělávacích studií může být samozřejmě mnohem bohatší, nicméně v jednom článku nelze tuto tematiku úplně popsat. Proto je text jen úvodem do této problematiky a v závěru obsahuje podrobnější přehled možností jednotlivých statistických softwarů pro složitější statistické analýzy dat z mezinárodních vzdělávacích studií. Praktické zaměření textu vyústí v přípravu videoukázky z reálné práce s daty s jedním šetřením, konkrétně TIMSS 2011, nicméně s ohledem na analogičnost zpracování dat z jiných mezinárodních šetření není toto na překážku obecnosti použití textu. S ohledem na již publikované texty (Soukup, 2012; Basl, 2014) není opakována materie v těchto textech obsažená, tj. není rozebíráno zaměření jednotlivých vzdělávacích studií, jejich opakování apod. Obdobně není cílem textu získávat zajímavé věcné výsledky, které jsou pravidelně publikovány v mezinárodních a národních zprávách z mezinárodních vzdělávacích šetření (odkazy na aktuální lze nalézt zejména v Soukup, 2012; Basl, 2014). Ukázky mají demonstrovat problémy při zpracování dat z mezinárodních vzdělávacích studií a jejich řešení. Čtenář, který se zajímá o detailní návody na zpracování dat z mezinárodních vzdělávacích studií softwarem IDB Analyzer, může využít anglicky psaných technických manuálů publikovaných k jednotlivým vlnám těchto studií (např. Foy, Arora, & Stanco, 2013, s. 5–38). Problém, na který je nutno v této souvislosti upozornit, je skutečnost, že tyto manuály bývají publikovány výrazně později, než je výzkumníkům umožněn přístup k datům z jednotlivých studií, a kvůli tomu se může stát následná analýza neaktuální. Čtenářům lze též doporučit semináře k práci s daty z mezinárodních vzdělávacích studií (mezinárodně organizuje pravidelně IEA DPC v Hamburku, u nás pak Česká školní inspekce, naposledy k šetření ICILS 2013).

Nelze než vyslovit přání, aby v budoucnu vzniklo mnoho originálních českých textů založených na analýzách mezinárodních vzdělávacích studií. Díky tomu se využije bohatství, které data skrývají. Nadto nelze pominout ani skutečnost, že při zpracování těchto dat a pronikání do jejich metodologických detailů se český výzkumník může mnohému přiučit a vylepší se standard tuzemských kvantitativně

laděných vzdělávacích studií a obecně vzroste kvantitativní gramotnost českých výzkumníků.

## 1 Proč nelze využívat běžné statistické nástroje?

Na počátku je nutné dát jasnou odpověď na otázku, která se nabízí. Proč nemohu vzít data z mezinárodních vzdělávacích studií a použít běžné softwarové vybavení, tj. v pedagogickém výzkumu nejčastěji MS Excel či SPSS?

Odpověď je poměrně komplexní, obecně lze ale říci, že při využití těchto prostředků bez dalšího (tj. jejich základních procedur) bychom získali nesprávné výsledky v dvojím slova smyslu:

1. Výsledky ve formě průměrů a procent by byly nesprávné, protože data mezinárodních vzdělávacích studií nepocházejí z prostého náhodného výběru (srov. dále popis vážení) a nadto se studie nezúčastní všichni původně vybraní jedinci.<sup>3</sup>

2. Výsledky statistických testů (pro srovnávání průměrů ve skupinách, souvislosti proměnných apod.) by byly nesprávné (častěji bychom nesprávně našli průkazný rozdíl či souvislost), protože data jsou běžně získávána vícestupňovým vybíráním a vzorce pro výpočty standardních chyb, které se používají pro výpočty statistických testů, musí být modifikovány.

Nutno poznamenat, že někdy by byl dopad použití nesprávných postupů minimální, ale jindy může jít o velice odlišné výsledky. Připomeňme, že cílem vědeckého snažení je podávat objektivní informace a používat nejnovější vědecké postupy, proto je namístě zohlednit je i při analýze dat, pokud jsou známy a lze je bez větších obtíží využít v dostupných programech.

K výše popsaným problémům přistupují ještě dva poměrně speciální (tj. neplatí nezbytně pro všechny analýzy). Konkrétně se jedná o přítomnost více datových souborů, které jsou výstupem výzkumu v jednotlivé zemi (např. existuje jeden národní datový soubor pro výsledky žáků v testech, jiný pro jejich odpovědi v dotaznících, další pro odpovědi rodičů žáků a ještě jiný pro odpovědi ředitelů škol, které žáci navštěvují). Reálně pak může být výstupem jedné vzdělávací studie několik stovek datových souborů a jen samotné propojení jednotlivých souborů, se kterými by výzkumník rád pracoval (z různých zemí a od rozličných skupin), může být poměrně komplikovanou záležitostí. Proto ještě před demonstrací jednotlivých analýz bude v článku ukázána elegantní možnost spojování dat s využitím programu IDB Analyzer. Samozřejmě, pokud chce výzkumník pracovat jen s jedním konkrétním datovým souborem z jedné konkrétní země (nebývá to příliš časté), výše popsaný problém odpadá.

<sup>3</sup> Zde bychom *stricto sensu* (jak správně upozornil jeden z anonymních recenzentů) s běžným softwarem většinou vystačili, jen by musel umožňovat vážení, což běžný software na požádání umí. Nicméně, jak bude ukázáno dále, váhové struktury v datech z mezinárodních vzdělávacích studií jsou poměrně komplikované, a tak je užívání speciálních prostředků více než žádoucí.

Poslední obtíží<sup>4</sup>, na kterou narážíme, je skutečnost, že zejména znalosti či gramotnosti<sup>5</sup> (např. čtenářská, matematická) jsou měřeny nepřímou (skrže tzv. latentní proměnné), a ty bývají reprezentovány v datech několika (nejčastěji pěti) hodnotami, které dále nazýváme *vícenásobné odhady hodnot latentních proměnných* (anglicky *plausible values*<sup>6</sup>). Pro správné výpočty je nutno zohlednit všech těchto pět hodnot a to opět běžné statistické vybavení (Excel, SPSS) neumí. Shrňeme-li výše popsané obtíže, jde o následující čtyři oblasti:

1. data obsahují váhy, které zohledňují nestejnou pravděpodobnost výběru jednotlivých respondentů, případně jejich neúčast ve studii;
2. data pocházejí z víceúrovňových náhodných výběrů;
3. mnohost dat na národní a mezinárodní úrovni (tj. reálně existují desítky datových souborů a jejich spojování není zcela triviální);
4. některé proměnné neměříme přímo (typicky výsledky v kognitivních testech) a je třeba zohlednit chybu měření těchto latentních proměnných.

Stručně popíšeme detailněji jednotlivé problémy a postup, jak se s nimi lze vyrovnat. V dále uvedených příkladech analýz je pak ukázáno použití správných postupů, které tyto problémy řeší.

### 1.1 Výběrové váhy

V mezinárodních vzdělávacích šetřeních zpravidla nejsou stejné pravděpodobnosti výběru jednotlivců, v případě škol bývá nejčastější výběrání s pravděpodobnostmi úměrnými velikosti příslušné školy (typicky měřené počtem žáků). Důvodem je požadavek na dostatečné zastoupení určitých skupin (např. v Česku by to mohly být děti na víceletých gymnáziích). Můžeme například jako výzkumníci rozhodnout, že zatímco v celém ročníku v ČR je jen desetina gymnazistů, uděláme výběr škol tak, aby gymnazistů bylo cca 40%. Pokud chceme výsledky zobecňovat na populaci, musíme tuto disproporci v našem výběrovém souboru zohlednit skrze využití vah (tj. v našem případě budou muset váhy snížit vliv gymnazistů a naopak navýšit vliv dětí ze základních škol). Konkrétní technický postup tvorby vah a jejich implementaci ve statistických výpočtech může čtenář nalézt například v učebnici Mareš, Rabušic a Soukup (2015, s. 197–201). I pokud by výzkumník nezanášel do svých dat tyto disproporce designem svého výzkumu, tj. měl například ve výběru stejně zastoupené děti ze základních škol i víceletých gymnázií, běžně nastane situace, že osoby z jedné skupiny se častěji nezúčastní výzkumu než osoby z jiné skupiny (např. děti z víceletých gymnázií přijde do školy v den testování 90%, ale u dětí ze základních

<sup>4</sup> Problémů je jistě mnohem více, zde popisujeme ty nejběžněji řešené, se kterými si lze analyticky poradit.

<sup>5</sup> V šetření TIMSS, které je použito pro ilustraci v tomto článku, se mluví o *znalostech* (jejich prokazování a využívání, dále o uvažování), v nejznámějším šetření PISA se užívá označení *gramotnosti*. Děkuji za upozornění na tyto odlišnosti Dominiku Dvořákoví.

<sup>6</sup> Autorovi textu není znám český překlad tohoto pojmu (jediný, na který narazil, zněl „možné hodnoty“ a to není vůbec výstižné), proto užívá vlastní překlad, který se snaží vyjádřit podstatu, tj. skutečnost, že pro zachycení chyby měření generujeme (nejčastěji metodou maximální věrohodnosti) několik odhadů a s nimi pracujeme najednou.

škol činí tento podíl jen 70 %). I v tomto případě vznikne disproporce mezi složením výběrového souboru v našich datech a celé populace, a je opět zapotřebí napravit to vážením.

Jak plyne z výše uvedených příkladů, ve vzdělávacích studiích může být potřeba vážení vyvolána nejrůznějšími situacemi, a tak i váhy obsažené v datových souborech z těchto studií jsou mnohé<sup>7</sup> (kromě výše popsaných situací můžeme ještě zmínit specifické váhy pro jednotlivé školy pro data získaná od ředitelů, váhy učitelské pro data získaná od učitelů). V datech tak máme často několik vah (a nadto v rozličných souborech z jedné studie ještě váhy různé). Pro analytika je pak poměrně komplikované rozhodování, kterou váhu (či jejich kombinaci) má pro konkrétní analýzu použít. Zde výrazně pomůže užití programu IDB Analyzer, který užije správnou váhu s ohledem na to, jaká data využíváme.<sup>8</sup> Pokud používáme jiný software, nezbyvá než si důkladně přečíst technickou zprávu k příslušné studii a správné váhy vybrat a použít. S ohledem na skutečnost, že o vahách se hovoří i v další části, dodejme, že váhy výše popsané budeme dále označovat též jako *výběrové váhy*, protože napravují nestejně pravděpodobnosti vybírání jednotek.

## 1.2 Vícestupňové výběry

Data z mezinárodních vzdělávacích studií (ale platí to i např. pro mezinárodní sociologické či politologické studie) jsou zpravidla sbírána skrze postupy náhodných (pravděpodobnostních) výběrů. Nicméně s ohledem na skutečnost, že výzkumníci běžně nemají dostupné seznamy vybíraných jednotek (typicky žáků), nejde v praxi o prostý náhodný výběr jednotek, ale běžně se používá vícestupňový náhodný výběr, případně kombinace vícestupňového vybírání a skupinkového (cluster) vybírání. Typicky jsou v prvním stupni vybírány náhodně školy ze seznamu všech škol a poté v druhém kroku žáci ve vybraných školách (alternativně jsou náhodně voleny celé třídy ve vybrané škole). Protože neužíváme prostý náhodný výběr, ale vícestupňový, není korektní používat standardní procedury pro výpočty statistických testů či intervalů spolehlivosti (tj. běžných nástrojů pro zobecňování z výběru na populaci). Konkrétně dochází k tomu, že kvůli nesprávným výpočtům budou testy vést častěji k zamítní nulových hypotéz (častěji prokážeme, byť nesprávně, odlišnosti či souvislosti) a budeme docházet k užším intervalům spolehlivosti. V datech z mezinárodních vzdělávacích studií jsou pro tuto situaci připraveny *replikační váhy* (neplést s výše popsanými vahami, které se používají pro zajištění reprezentativity našich výsledků). Těchto vah bývá několik desítek (nejčastěji 75 či 80)<sup>9</sup> a skrze replikační postupy (srov. dále) se zajistí korektní výpočet statistických testů či interval spolehlivosti.

<sup>7</sup> Reálně jsou v těchto datech kromě vah ještě korekční váhové koeficienty, detaily čtenář nalezne v technických zprávách k jednotlivým studiím, pro pochopení dalšího textu a analytické používání popsaných postupů není výklad tohoto fenoménu zde nezbytný.

<sup>8</sup> Nicméně pro některé úlohy musíme automatický výběr váhy zprostředkovaný IDB Analyzerem změnit, vše odvisí od naší výzkumné úlohy, tj. na jakou populaci chceme výsledky zobecnit.

<sup>9</sup> Alternativně je v datovém souboru dvojice proměnných, které popisují replikační zóny a v rámci nich rozdělují jednotky na dvě poloviny (srov. dále popsaný příklad v obr. 1).

Jak tedy konkrétně replikace probíhá? Náš datový soubor (např. data o žácích ze 160 škol) je rozdělen do tzv. replikačních zón. Při počtu 160 škol je praktické vždy dvě podobné školy sloučit do jedné replikační zóny. V těchto zónách jsou pak žáci navíc rozděleni na dvě poloviny (např. na žáky z první a druhé školy v rámci zóny). A poté je nejběžnější postup (Jackknife Repeated Replication) následovný: Analýza je spočítána poprvé tak, že v rámci první replikační zóny je vynechána druhá škola a vliv žáků z první školy je zdvojnásoben, u žáků ze škol z ostatních replikačních zón se nic zvláštního neděje a tito jsou využiti ve výpočtu všichni. Analogicky je podruhé analýza počítána s vynecháním jedné ze škol z druhé replikační zóny atd. (viz ukázkou a popis dále uvedený). Výpočet je takto 75krát opakován, respektive replikován<sup>10</sup> (odtud plyne název replikace) a poté je z něj vypočtena standardní chyba odhadu, která bude korektní v rámci testování statistických hypotéz či výpočtů intervalu spolehlivosti.

Tabulka 1 naznačuje technické provádění výpočtu.<sup>11</sup> V prvním sloupci je identifikace 4 různých škol (SCHOOL), dále je uveden identifikátor 8 žáků<sup>12</sup> (STUDENT) a poté jsou uvedeny výsledky kognitivního testu (TEST). Dále následují dvě proměnné charakterizující replikační zónu (JKZONE) a postup replikací (JKREP). Ve sloupcích REPL 1–4 je následně uveden výsledek jednotlivých replikací výsledku testů jednotlivých žáků. Konkrétně v první replikaci (sloupec REPLIKACE 1 ve schématu 1) se z první replikační zóny (hodnotu „1“ v JKZONE mají školy 101 a 102) vezme dvojnásobek výsledku žáků v testu ze školy 101 (je u nich hodnota JKREP „1“) a žáci ze školy 102 se vynechají (hodnota jejich výsledku je nulová, neboť je u nich hodnota JKREP „0“). Výsledky žáků z ostatních škol se zkopírují. Analogicky pro druhou replikaci (sloupec REPL 2 v tab. 1) se provede operace pro školy 103 a 104 (pro žáky ze školy 103 bude výsledek v replikaci nulový a pro žáky ze školy 104 dvojnásobný).

Snadno můžeme nahlédnout, že výsledné průměry jednotlivých replikací (viz předposlední řádek schématu 1) jsou navzájem odlišné a pro tyto jednotlivé průměry bude vypočtena směrodatná odchylka, která bude rovna standardní chybě odhadu využívané pro statistické testy. Konkrétně je výpočet proveden tak, že se napočítají rozdíly druhých mocnin průměrů z jednotlivých replikací a skutečného průměru (tyto hodnoty jsou v posledním řádku s pomocnými výpočty) a tento součet se dělí počtem replikací, v našem případě dvěma. Výsledek je rozptyl a pro získání standardní chyby odhadu je třeba jej odmocnit. V našem příkladu bude výsledná hodnota standardní chyby odhadu pro testové výsledky vypočtená skrze replikace 17,9. Pokud bychom počítali klasicky<sup>13</sup>, tj. předstírali, že naše data tvoří prostý náhodný výběr, byl by vý-

<sup>10</sup> Studie administrované IEA (TIMSS, PIRLS, ICILS, ICCS) využívají 75 zón, studie PISA užívá 80 zón. Počet replikací obecně může být i jiný. Z počtu 75, respektive 80, zón plyne i požadavek mít ve výběru cca 150, respektive 160, škol, tj. dvakrát více, než je replikačních zón.

<sup>11</sup> Na tomto místě je nutné upozornit, že použití dvou replikací a osmi studentů je pro reálné výpočty zcela nevhodné, jde jen o ukázkou logiky výpočtu na velice omezeném prostoru.

<sup>12</sup> Pro zjednodušení ilustrace jsou v každé škole jen dva žáci (srov. sloupec student), výsledek kognitivního testu je ve sloupci TEST.

<sup>13</sup> Formálně jej spočítáme jako odmocninu z výběrového rozptylu poděleného počtem pozorování, tento vzorec uvádí každá běžná učebnice základů statistiky.

sledek 14,5, tj. nesprávně o čtvrtinu nižší.<sup>14</sup> Dodejme, že v praxi mohou být hodnoty i výrazněji odlišné (srov. například výsledek prezentovaný v tabulce 2 a obrázku 5 a komentář tam uvedený). Situace je nadto v reálných šetřeních ještě komplikována použitím výběrových vah (těmi bychom museli výsledky testů pronásobit), pro jednoduchost schématu však tuto komplikaci zde nezobrazujeme. Obecně ovšem platí, že výsledek skrze replikace správně zohledňuje design studie (typicky vícestupňový náhodný výběr) a umožňuje korektní odhad standardních chyb, které jsou užívány pro konstrukce intervalů spolehlivosti a testy statistické významnosti.<sup>15</sup>

**Tabulka 1** Grafické zobrazení replikačního postupu pro 8 škol a 4 replikační zóny

SCHOOL	STUDENT	TEST	JKZONE	JKREP	REPL 1	REPL 2	
101	10101	510	1	1	1020	510	
101	10102	515	1	1	1030	515	
102	10201	503	1	0	0	503	
102	10202	511	1	0	0	511	
103	10301	540	2	0	540	0	
103	10302	592	2	0	592	0	
104	10401	460	2	1	460	920	
104	10402	470	2	1	470	940	
Průměr:					Průměr:	Průměr:	
512,625					514,000	487,375	
					Pomocné výpočty:	1,890	637,563

Zdroj: vlastní výpočty

Opět platí, že využití replikací nebývá v běžných softwarech (např. Excel či SPSS) dostupné, a tak je nutné používat speciální software (viz přehled na konci článku) nebo již zmíněný IDB Analyzer, který ve spolupráci s SPSS umí tyto replikační postupy využívat a je přímo připraven pro mezinárodní vzdělávací studie (analytikovi tedy odpadá potřeba zjišťovat, která proměnná udává replikační zóny a replikace, případně kde jsou replikační váhy umístěny v datech<sup>16</sup>).

<sup>14</sup> Podíl správně spočtené standardní chyby odhadu a charakteristiky spočtené za předpokladu prostého náhodného výběru se nazývá efekt designu (v našem případě je cca 1,25) a udává, kolikrát větší výběr bychom potřebovali provést skrze vícestupňový náhodný výběr, pokud bychom chtěli v našich datech zachovat přesnost stejnou jako v datech z prostého náhodného výběru.

<sup>15</sup> On-line verze článku publikovaná na stránkách časopisu obsahuje barevné verze tabulek a obrázků uváděných níže.

<sup>16</sup> Na rozdíl od TIMSS (a ostatních studií administrovaných IEA) se v šetření PISA přímo v datovém souboru setkáme s 80 sloupci, které obsahují replikační váhy. Postup jejich využití je podobný výše představenému schématu, jen s tím rozdílem, že jednotlivé replikace jsou výsledkem pronásobení jednotlivých replikačních vah a příslušné charakteristiky, pro kterou chceme počítat standardní chybu odhadu.

### 1.3 Spojování souborů

Problémy při zpracování dat z mezinárodních studií působí i skutečnost, že výzkumník má k dispozici stovky jednotlivých datových souborů, protože výsledkem šetření na národní úrovni je několik datových souborů a celosvětově se účastní desítky zemí. Pro analýzu je proto často potřeba pracovat s několika soubory najednou.

V úvahu přichází dvě základní situace. Pro komplexnější národní analýzy spojujeme několik souborů či všechny národní soubory. Pro vysvětlení žákovských kompetencí skrze charakteristiky učitelů žáků, školní prostředí či rodinné zázemí pak připojujeme k žákovským datům data za učitele, ředitele či rodiče. Nadto někdy můžeme chtít provádět mezinárodní komparaci společně s vysvětlením skrze kontextové informace o škole či rodině žáka a problém a vše je ještě složitější. Při spojování je nutné zvolit správný identifikátor, dle kterého provedeme spojení (typicky identifikaci třídy, školy či žáka), ale musíme si též rozmyslet, které váhy napravující nerovné pravděpodobnosti jednotek ve výběru (viz bod 1 výše) máme do spojeného začlenit a také jaké replikační váhy (viz bod 2 výše) budeme potřebovat.<sup>17</sup> Opět je nutné buď detailně nastudovat technickou dokumentaci k příslušné studii, nebo se spolehnout na IDB Analyzer, který správně rozpoznává, co spojit lze, a v případě spojení skrze SPSS správně uloží do spojeného datového souboru výběrové i replikační váhy (případně proměnnou s replikačními zónami a skupinami žáků uvnitř těchto zón).

### 1.4 Práce s vícenásobnými odhady latentních proměnných

Posledním speciálním postupem, který se využívá při analýze dat z mezinárodních vzdělávacích studií, je využití latentních proměnných<sup>18</sup>, které jsou typicky v datech reprezentovány pěti hodnotami ve formě vícenásobných odhadů (plausible values)<sup>19</sup>. Běžně se tento postup používá pro jednotlivé výsledky kognitivních testů, které jsou nepřímo měřeny skrze teorii odpovědi na položku (IRT, Urbánek & Šimeček, 2001) mnoha testovými položkami. Aby bylo možné při výpočtech s těmito proměnnými zohlednit chybu jejich měření (tj. skutečnost, že každý test má chybu měření), není příslušný výsledek kognitivního testu v datech zastoupen jedinou hodnotou, ale běžně pěti hodnotami, které reprezentují rozdělení vlastnosti měřené u příslušného jedince.

Jak analyzovat data, když nemám fenomén v jediné proměnné, ale v pěti proměnných? Správný postup vypadá tak, že výpočet se provede pětkrát, kdy pokaždé se použije jedna ze sady pěti vícenásobných odhadů latentní proměnné a výsledky se poté průměrují (u standardních chyb odhadu je výpočet složitější a autor jej zde

<sup>17</sup> Spojování některých souborů nadto není někdy s ohledem na design studie možné (např. není zachycena vazba učitele a žáka).

<sup>18</sup> Tedy nepřímo měřených veličin.

<sup>19</sup> Teoreticky by mohlo být těchto hodnot více (např. 10) i méně (např. 3). Datovými simulacemi se došlo k zjištění, že pět hodnot latentních proměnných je pro vyčíslení chyby měření ve výpočtech dostačující.



neuvádí, protože pro pochopení textu není nezbytný<sup>20</sup>). Pokud vezmeme v potaz, že pro každou hodnotu vícenásobného odhadu máme provádět opakovaně výpočet (např. regresní analýzu), a víme nadto, že pro zohlednění designu studií (viz odstavce 1.2 výše) se má výpočet provádět 75krát pro jednotlivé replikační zóny, musíme tak analýzu počítat 375krát! Opět to umí snadno provést SPSS pomocí IDB Analyzery (viz poslední ukázkový příklad), nebo lze využít jiný speciální software (viz přehled na konci článku), ten ale již vyžaduje, aby uživatel přesně specifikoval názvy proměnných, které tvoří jednotlivé vícenásobné odhady pro latentní proměnnou.

### 1.5 Shrnutí

Shrneme-li předchozí výklad, lze jej uzavřít následovně. Pro korektní výpočty s daty z mezinárodních studií musíme zohlednit nestejně pravděpodobnosti vybírání jednotek (skrze výběrové váhy), složitější výběrové designy (skrze replikační váhy), někdy správně spojit jednotlivé datové soubory k analýze a jindy též opakovaně počítat s vícenásobnými odhady pro nepřímo měřené fenomény (typicky různé znalosti či gramotnosti). Běžné procedury ve statistickém softwaru (např. SPSS) to přímo neumožňují, a proto je potřebné užívat speciální programy. Pro počítání jednodušších analýz (výpočty průměrů a statistické testování odlišnosti pro jednotlivé skupiny, tvorbu kontingenčních tabulek, korelační analýzu a regresní analýzu) a spojování dat lze využít IDB Analyzer, který spolupracuje s SPSS a je uživatelsky velmi jednoduchý. Pro složitější analýzy je třeba využít specializovaný software, jehož přehled nalezne čtenář na konci tohoto článku.

## 2 Ukázky praktické práce s daty z mezinárodních vzdělávacích studií v IDB Analyzery

Po přehledu metodologických problémů, které se týkají dat z mezinárodních vzdělávacích studií, budou představeny praktické postupy, jak se s těmito problémy vyrovnat. S ohledem na charakter textu (vedení do problematiky) i jeho možný rozsah jsou představeny jen základní analytické techniky (výpočty průměrů a statistické testování odlišnosti pro jednotlivé skupiny, tvorbu kontingenčních tabulek, korelační analýzu a regresní analýzu). Před jednotlivými technikami je nadto ještě uvedena možnost spojování dat. Pro ilustraci je zvolena studie TIMSS 2011 (zaměřená na znalosti z matematiky a přírodních věd ve 4. ročníku<sup>21</sup>), nicméně dále uvedené postupy jsou univerzální a uplatnitelné pro všechny mezinárodní vzdělávací studie, které zaštiťuje OECD nebo IEA.

<sup>20</sup> Potřebné vzorce lze nalézt například v knize Petruška (2015, s. 73–74). Petrušek sice využívá vzorce pro vícenásobné imputace chybějících hodnot, nicméně postup v případě vícenásobných odhadů v IRT modelech je stejný.

<sup>21</sup> V rámci studie mohla příslušná země zkoumat též 8. ročníky, nicméně v ČR nebyla tato opce využita, proto v ilustracích využíváme 4. ročník.

## 2.1 Základní informace o šetření TIMSS 2011

Pro snazší pochopení dále popsaných úloh zde uvedeme několik informací o datech ze šetření TIMSS 2011. Zaměříme se pouze na data ze šetření žáků ve 4. ročníku, protože v ČR jiné neprobíhalo.<sup>22</sup> V rámci šetření TIMSS vznikly v každé zemi tyto datové soubory (v závorce jsou vždy uvedeny obecné názvy těchto souborů):

- soubor s odpověďmi ředitelů (ACG●●●B1<sup>23</sup>);
- soubor s výsledky testů žáka v oblasti matematiky a přírodních věd (ASA●●●B1);
- soubor s odpověďmi žáků na dotazník o sobě a vztahu k přírodním vědám<sup>24</sup> (ASG●●●B1);
- soubor s odpověďmi žáků na dotazník škoře a rodině (ASH●●●B1);
- soubor pro možnost propojení dat žáků a učitelů (AST●●●B1);
- soubor s odpověďmi učitelů (ATG●●●B1).

Jen pro ČR (4. ročník) máme tedy k dispozici 6 datových souborů, které můžeme získat (viz dále) v různých datových formátech (pro SPSS, SAS a volném textovém formátu). Jednotlivé datové soubory obsahují kromě meritorních proměnných též identifikační proměnné (identifikátory země, školy, učitele, žáka), příslušné výběrové váhy a proměnnou obsahující replikační zónu a podskupinu žáků sloužící k replikaci. Detailní popis jednotlivých proměnných a jejich logiku může čtenář vyhledat v publikaci IEA (Foy et al., 2013, s. 82–98).

## 2.2 Spojování dat

Po stručném úvodu do datové struktury šetření TIMSS ukažme, jak prakticky spojit různé datové soubory, které chceme analyzovat.<sup>25</sup> Pokud se rozhodneme provádět spojování bez využití IDB Analyzery (viz dále), musíme minimálně vědět:

1. jaká je konvence pojmenovávání jednotlivých souborů, tj. která část názvu charakterizuje zemi, která charakterizuje cílový ročník, žáka, učitele či ředitele apod.;
2. jaká je konvence pojmenovávání klíčových proměnných pro spojování, tj. identifikátoru školy, třídy, učitele, žáka;
3. jak se jmenují jednotlivé proměnné s výběrovými vahami a které z nich máme do spojeného datového souboru uložit;
4. jak se jmenují jednotlivé proměnné s replikačními vahami, které máme do spojeného datového souboru uložit.

<sup>22</sup> V některých zemích proběhlo šetření též v 8. a 12. ročníku školní docházky.

<sup>23</sup> Tři volné pozice v názvu souboru jsou vyhrazeny pro ISO kód země, data od ředitelů v ČR mají tedy název ACGCZEB1), druhé a třetí písmeno vždy označuje příslušný typ dat, tj. pro ředitele, učitele atd.

<sup>24</sup> Soubor nadto obsahuje i vícenásobné odhady latentních proměnných pro oblast matematiky a přírodních věd.

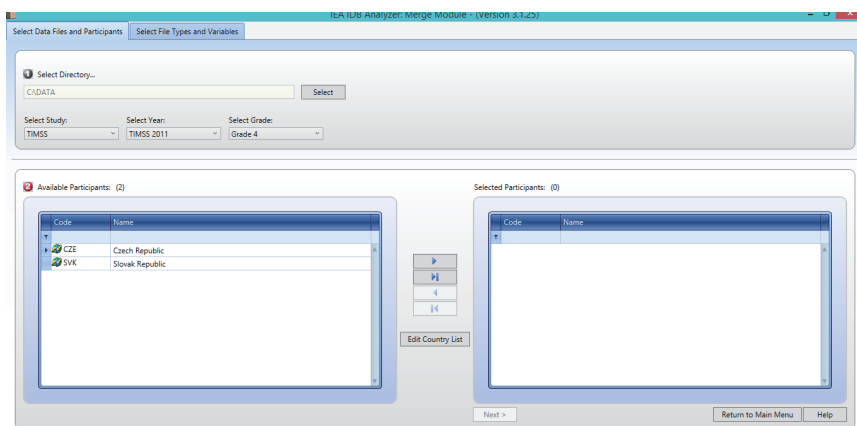
<sup>25</sup> Běžně si při získávání dat z mezinárodních vzdělávacích studií z webových stránek stáhneme mnoho jednotlivých souborů, které poté spojujeme.

Všechny výše uvedené informace lze nalézt v technické dokumentaci k jednotlivým šetřením. Pokud je výzkumník nechce nastudovat, musí využít některé připravené pomůcky, z nichž uživatelsky nejpřívětivější je IDB Analyzer, který spolupracuje s SPSS.<sup>26</sup>

Poté co výzkumník nainstaluje IDB Analyzer<sup>27</sup> a má ve svém počítači též SPSS (verze 15 a vyšší) a data z některé mezinárodní vzdělávací studie, může začít se spojováním těchto dat. Po spuštění IDB Analyzera je nutno vybrat první volbu „Merge Module“. Pak na základní obrazovce (obr. 1) musíme vybrat:

1. adresář, kde jsou data určena ke spojování;
2. studii, s níž chceme pracovat (v našem případě TIMSS 2011);
3. rok, kdy byla studie provedena;
4. ročník, na který se chceme v analýze zaměřit.

IDB Analyzer nám dle našeho zadání automaticky určí seznam zemí, pro které máme v námi vybraném adresáři dostupná data (levá strana dolní části obrazovky na obr. 1). Pokud chceme spojovat jen některé země, učiníme tak přesunem doprava za použití symbolu jednoduché šipky. Chceme-li spojovat data pro všechny nabízené země, přesuneme je doprava jedním stiskem dvojité šipky. Ve spojeném souboru pak budou jen ty země, které vidíme na pravé straně v dolní části obrazovky.



**Obrázek 1** Spojování dat – výběr studie, ročníku a spojovaných zemí

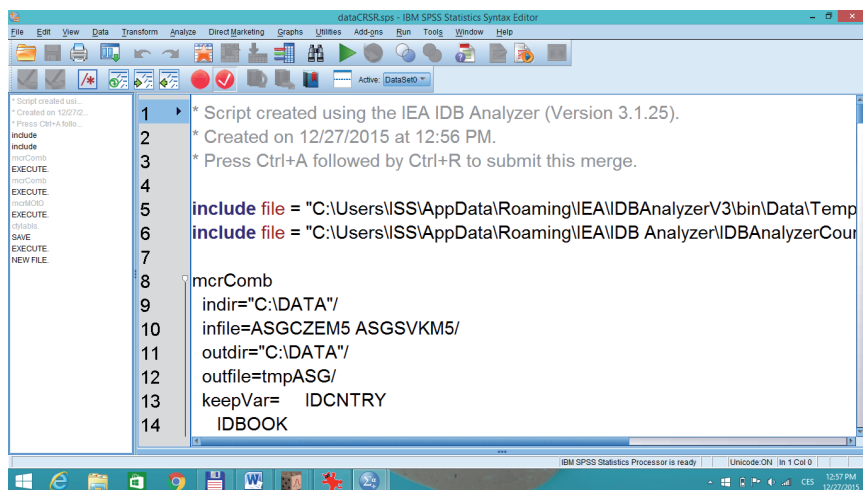
Tento obrázek i všechny i následující obrázky v článku jsou screenshoty z IDB Analyzer. resp. z SPSS a Excelu.

V dalším kroku je nezbytné zvolit, jaké instrumenty (typicky dotazníky či testy) a jejich jednotlivé proměnné zahrne spojený datový soubor. To určíme poté, co nahore přejdeme na záložku „Select File Types and Variables“. Zcela vlevo (obr. 2) máme seznam dostupných instrumentů a můžeme vybrat jeden nebo více. IDB Ana-

<sup>26</sup> IDB Analyzer je zdarma dostupný doplněk, který spolupracuje s SPSS (plným názvem IBM SPSS Statistics). SPSS je placený software, který umožňuje širokou škálu statistických operací.

<sup>27</sup> Aktuální verzi najde on-line například na webové adrese <http://www.iea.nl/eula0.html>.





Obrázek 3 Ukázka skriptu v SPSS pro spojování dat

zúčastněné země. Pokud výzkumník chce využívat jen data pro některé země, musí z této (obrovské) databáze udělat podvýběh.

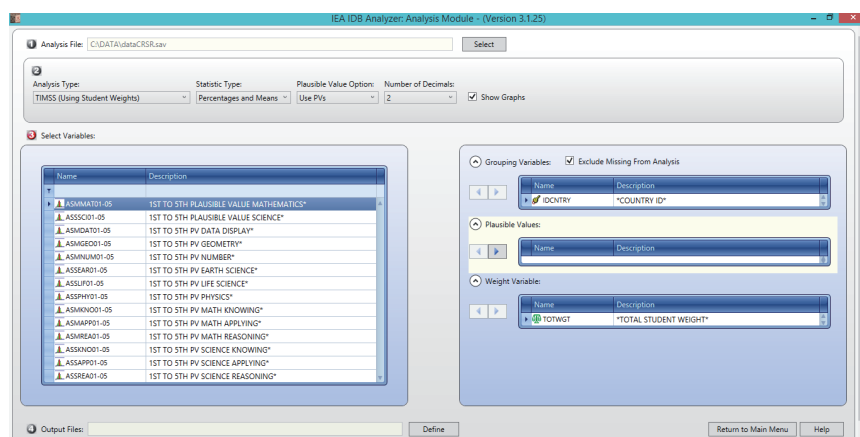
### 2.3 Výpočty průměrů a intervalů spolehlivosti

Analogicky ke spojování dat lze IDB užít, opět ve spojení s SPSS, i pro výpočty, tj. prostřednictvím IDB Analyzeru definujeme proměnné a parametry výpočtu a poté IDB Analyzer otevře SPSS a v něm spustíme připravený skript. Při generování analytických výsledků máme nadto ještě na výběr, zda chceme výsledky zobrazit přímo ve výstupním okně SPSS nebo zda kromě něj chceme výsledky zobrazovat v souboru typu MS Excel.<sup>30</sup>

Pro jednoduchou ilustraci vypočteme průměr znalostí z matematiky českých a slovenských žáků. V IDB Analyzeru se vrátíme do hlavní nabídky (dole vpravo volba „Return to Main Menu“) a vybereme „Analysis Module“. Poté najdeme námi vytvořený soubor dataCRSR.sav a v rámci nabídky analýz zvolíme „TIMSS (Using Student Weights)“, protože chceme provést analýzy na úrovni jednotlivých žáků (viz obr. 4). Dále zvolíme v nabídce „Statistic Type“ volbu „Percentages and Means“. S ohledem

<sup>30</sup> Bohužel IDB Analyzer pracuje tak, že pro každou jednotlivou analytickou proceduru (tj. jeden výpočet) vytvoří samostatný soubor pro Excel, případně výstup v SPSS (konkrétně uloží výstup pro Excel – \*.xls, data ve formátu SPSS – \*.sav, výstup z SPSS – \*.spv a skript pro SPSS – \*.sps). Toto je pro analytickou praxi mírně nepraktické, analytik pak pracuje s mnoha soubory a musí si výsledky přenést do jednoho souboru (typicky do textového dokumentu nebo do prezentace). Z vlastní analytické zkušenosti lze doporučit užívání výstupu pro MS Excel, protože tento formát je na rozdíl od výstupu v SPSS zobrazitelný téměř ve všech počítačích bez ohledu na verzi Excelu či obdobného freeware (např. Open Office Calc). Výhodou práce s IDB Analyzerem je možnost připravit si několik analýz, poté IDB Analyzer vypnout a postupně analýzy spouštět v SPSS. Zkušený výzkumník navíc může modifikovat skripty pro SPSS připravené v IDB Analyzeru přímo v SPSS (typicky tím, že změní proměnné pro příslušnou analýzu).

110 na skutečnost, že znalosti z matematiky jsou v datech zaznamenány ve formě pěti odhadů, zaškrtneme dále v „Plausible Value Option“ variantu „Use PVs“. SPSS skrze skript připravený v IDB Analyzeru bude automaticky počítat výsledky odděleně pro jednotlivé země<sup>31</sup> (srov. vpravo proměnná IDCNTY zařazená jako „Grouping Variable“ dále níže vpravo určí správné výběrové váhy (proměnnou TOTWGT). Na nás je již jen definice proměnné, pro niž chceme počítat průměry, tj. nalezení matematických znalostí. Poté co vpravo klikneme na okénko u „Plausible Values“, dojde vlevo k aktualizaci seznamu proměnných jen na ty, které mají vícenásobný odhad pro latentní proměnné. Hned první položka seznamu je pro matematické znalosti (srov. označení „1<sup>ST</sup> to 5<sup>TH</sup> PLAUSIBLE VALUE MATEMATICS“). Právě tuto proměnnou přeneseme šipkou u „Plausible Values“ doprava.



Obrázek 4 Ukázka zadání výpočtu průměru znalostí z matematiky žáků 4. ročníku (ČR a SR)

Posledním krokem je poté zadání názvu a formátu výstupu (dole „Output Files“). V souladu s výše uvedenou poznámkou vybereme formát v Excelu a název bude například PRUMER. Pak stiskneme „Start SPSS“ a analogicky k operaci v rámci spojování dat vybereme celý skript a spustíme jej. Po chvilce výpočtů<sup>32</sup> se zobrazí výsledek, my jej ukážeme v Excelu, tj. v našem adresáři najdeme soubor PRUMER.XLS (obr. 5).

Průměr matematických znalostí najdeme ve sloupci „mnpv“, tj. průměr (mean) z pěti vícenásobných odhadů hodnot příslušné latentní proměnné. V našem případě činí průměr pro ČR 510,85 bodu, pro Slovensko pak 506,77 bodu. Pokud bychom

<sup>31</sup> SPSS skrze skripty z IDB Analyzeru počítá analýzy vždy automaticky pro jednotlivé země, které jsou v našem datovém souboru, a tuto volbu nelze nijak změnit. Přes možnost „Grouping Variables“ lze zadat další třídící proměnné a počítat v rámci jednotlivých zemí pro námi zvolené podskupiny, například pro chlapce a dívky nebo pro děti z různých typů škol.

<sup>32</sup> Připomeňme, že SPSS musí počítat 75krát s různými replikačními vahami a poté ještě pětkrát pro 5 různých odhadů latentní proměnné, tj. výpočet se provádí 375krát.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	IDCOUNTRY		n	TOTWGT	sumw_se	pct	pct_se	mnpv	mnpv_se	sdpv	isdpv_se	weight	reps	method	shortcut	date
2	Czech Republic	ASMMATO	4.578.00	87.193.80	2.517.23	65.10	0.82	510.85	2.42	70.30	1.69	TOTWGT	75.00	JRR	Y	27-DEC-15
3	Slovak Republic	ASMMATO	5.616.00	46.706.13	1.147.96	34.90	0.82	506.77	3.75	79.62	2.56	TOTWGT	75.00	JRR	Y	27-DEC-15
4	x International Average	ASMMATO	#NULL!	#NULL!	#NULL!	50.00	0.58	508.81	2.23	75.00	1.53	TOTWGT	75.00	JRR	Y	27-DEC-15

Obrázek 5 Ukázka výsledku výpočtu průměru matematické znalosti žáků

chtěli konstruovat 95% interval spolehlivosti,<sup>33</sup> musíme využít standardní chyby odhadu v dalším sloupci „mnpv\_se“ a k průměru přičíst, respektive od něj odečíst, 1,96násobek této standardní chyby, tj. pro ČR, respektive SR, získáme tyto výsledky: <506,1; 515,6> a <499,4; 514,2>.

Kromě výsledků pro ČR a SR získáme ještě výsledek označený „x.International Average“. Název je ale matoucí, nejde totiž obecně o mezinárodní průměr ze šetření TIMSS 2011, ale jen o průměr z výsledků zemí, které jsou v našem datovém souboru, v našem případě tedy průměr českých a slovenských žáků dohromady. Z dalších údajů ve výstupu je již zajímavý jen počet žáků, na nichž je výsledek založen (srov. sloupec „n“), ostatní údaje jsou pro běžného uživatele zbytečné.<sup>34</sup> Pro úplnost dodejme, že pokud porovnáme naše výsledky s oficiálními výsledky z mezinárodní zprávy z TIMSS 2011 (např. pro ČR činí průměr 511 bodů a standardní chyba odhadu 2,4 bodu), zjistíme, že náš výpočet je korektní.

Pokud bychom nerespektovali výše uvedené postupy a počítali průměr v SPSS (typicky přes proceduru MEANS bez vážení) z našich dat, získali bychom tyto výsledky:

<sup>33</sup> Připomeňme, že standardní chybu odhadu lze využít i například pro jednovýběrový  $t$ -test. Pokud bychom například chtěli otestovat nulovou hypotézu, že průměr českých žáků ve 4. ročníku dosahuje hodnoty 500 bodů ( $\mu_0 = 500$ ), dosadili bychom do vzorce  $t = (\bar{x} - \mu_0) / se$ , kde  $\bar{x}$  je námi vypočtený průměr a  $se$  námi vypočtená standardní chyba odhadu. V našem případě by hodnota  $t$  činila 4,48 a na 5% hladině statistické významnosti bychom zamítli nulovou hypotézu, tj. prokázali, že průměr matematických znalostí měřených TIMSS testem českých čtvrtáků je odlišný od hodnoty 500.

<sup>34</sup> Nelze ale nijak nastavit, aby se nezobrazovaly.

112 Tabulka 2 Průměry proměnných charakterizujících matematické znalosti v ČR a SR

Země		N	Průměr	
			hodnota	standardní chyba
Česká republika	*1ST PLAUSIBLE VALUE MATHEMATICS*	4578	514,8005	1,04103
	*2ND PLAUSIBLE VALUE MATHEMATICS*	4578	515,3556	1,02885
	*3RD PLAUSIBLE VALUE MATHEMATICS*	4578	515,4251	1,02867
	*4TH PLAUSIBLE VALUE MATHEMATICS*	4578	514,9547	1,03550
	*5TH PLAUSIBLE VALUE MATHEMATICS*	4578	514,7762	1,04049
	Valid N (listwise)	4578		
Slovenská republika	*1ST PLAUSIBLE VALUE MATHEMATICS*	5616	505,9169	1,06897
	*2ND PLAUSIBLE VALUE MATHEMATICS*	5616	505,0494	1,06877
	*3RD PLAUSIBLE VALUE MATHEMATICS*	5616	505,9885	1,07430
	*4TH PLAUSIBLE VALUE MATHEMATICS*	5616	505,7910	1,07570
	*5TH PLAUSIBLE VALUE MATHEMATICS*	5616	504,8844	1,07369
	Valid N (listwise)	5616		

Zdroj: Vlastní výpočty

Pro Českou i Slovenskou republiku máme 5 hodnot průměrů (pro jednotlivé vícenásobné odhady). Pro ČR oscilují okolo hodnoty 515 (tj. cca o 4 body vyšší, než je správná hodnota), pro SR oscilují okolo hodnoty 505 (tj. o 2 body nižší, než je správná hodnota). Vidíme tedy, že dopad nesprávného výpočtu (nepoužití vah) posunuje nepředvídatelně výsledky (jednou hodnotu průměru navyšuje, v druhém případě snižuje). Dramatičtější jsou ovšem dopady na standardní chybu odhadu. Pro Česko osciluje hodnota kolem 1,0, z předchozích správných výpočtů ale víme, že hodnota je cca 2,4 (tj. téměř 2,5násobně vyšší<sup>35</sup>). Díky tomu bychom konstruovali výrazně užší interval spolehlivosti (tj. předstírali větší přesnost měření), případně bychom častěji (nesprávně) zamítali nulové hypotézy o určité velikosti námi měřeného fenoménu. V dalších ukázkách už tato srovnání neprovádíme, ale závěry by byly obdobné.

<sup>35</sup> Pro úplnost doplníme, že pro Slovensko je poměr ještě větší, nesprávně spočtená standardní chyba je cca 1,1 a správně spočtená 3,75 (viz tab. 2, resp. obr. 5).



## 2.4 Výpočty procent (podílů) a intervalů spolehlivosti

Po detailnějším rozboru výpočtu průměrů již stručně představíme další analytické možnosti práce prostřednictvím IDB Analyzery. Pro výpočet procent (výskytu jednotlivých odpovědí) a jejich intervalu spolehlivosti užíváme druhou analytickou proceduru („Statistic Type“) nazvanou „Percentages Only“. Pro jednoduchost ukážeme výsledky pro první proměnnou v našem spojeném souboru, která zjišťuje, zda rodičovský dotazník vyplňovala matka (proměnná ASBH01A). Proměnné, pro něž chceme počítat podíly (procenta), přesouváme doprava do okna „Grouping Variable“.<sup>36</sup>

Výsledek, tj. procenta odpovědí, najdeme ve sloupci „pct“ (obr. 6), jejich standardní chyby pak ve sloupci označeném „pct\_se“. V ČR byla tedy odpověď *ano* (dotazník vyplňovala matka) v 89,4% případů, odpověď *ne* ve zbývajících 10,6% případů. Pro případný výpočet intervalu spolehlivosti či testování hypotéz je opět možné využít vypočtené standardní chyby odhadu.<sup>37</sup>

IDCOUNTRY	ASBH01A	n	TOTWGT	sumw_se	pct	pct_se	weight	method	date	time		
2	Czech Republic	YES	3 861	74 041 42	2 275 78	89 40	0 62	TOTWGT	75 00	JRR	27-DEC-15	14h 0m
3	Czech Republic	NO	494	8 780 09	603 06	10 60	0 62	TOTWGT	75 00	JRR	27-DEC-15	14h 0m
4	Slovak Republic	YES	4 752	38 841 26	1 089 58	89 14	0 55	TOTWGT	75 00	JRR	27-DEC-15	14h 0m
5	Slovak Republic	NO	615	4 852 07	237 62	10 86	0 55	TOTWGT	75 00	JRR	27-DEC-15	14h 0m
6	x International Average	YES	#NULL!	#NULL!	#NULL!	89 27	0 42	TOTWGT	75 00	JRR	27-DEC-15	14h 0m
7	x International Average	NO	#NULL!	#NULL!	#NULL!	10 73	0 42	TOTWGT	75 00	JRR	27-DEC-15	14h 0m

Obrázek 6 Ukázka výsledku výpočtu procentního podílu

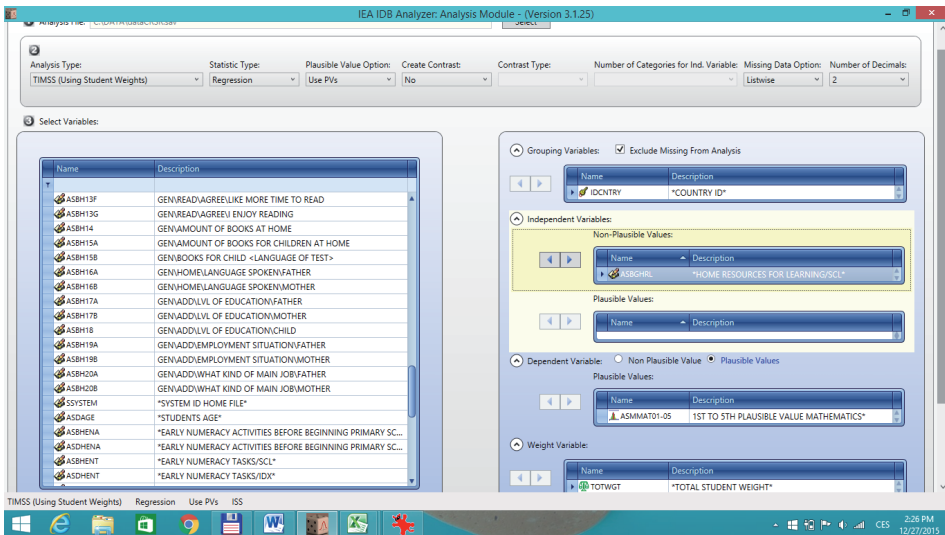
## 2.5 Regresní analýza

Další procedurou nabízenou v prostředí IDB Analyzery je regrese, konkrétně jde o běžnou lineární regresní analýzu počítanou metodou nejmenších čtverců (srov. Hendl,

<sup>36</sup> S ohledem na analogičnost dříve popsaného postupu u obrázku 4 zde již tento neopakujeme a obrázek neuvádíme.

<sup>37</sup> Meritorně to však pro námi vybranou proměnnou nemá valného smyslu, proto od tohoto kroku upustíme. Vzorce pro výpočet intervalu spolehlivosti či pro testování jsou totožné s výše uvedenými vzorci pro průměr, není je tedy nutné opakovat. Výsledky budou pouze vypovídat o procentním podílu, nikoli o průměru.

114 2012). V této proceduře je nutné odlišit závisle („Dependent Variable“) a nezávisle proměnné („Independent Variables“), tj. zvláště musíme přesouvat tyto typy proměnných. V případě užití latentních proměnných s vícenásobnými odhady je pak třeba ještě odlišit tyto proměnné a zadávací dialog má pak více nabídek (srov. pravou stranu obr. 7). Nadto je ještě možné použít jako jednu nezávisle proměnnou nominální či ordinální proměnnou (volby „Create Contrast“, „Contrast Type“ a „Number of Categories for Ind. Variable“ v obr. 7 nahoře).<sup>38</sup>



Obrázek 7 Ukázka zadávacího dialogu regresní analýzy (včetně vícenásobných odhadů latentních proměnných)

Výstupy z regresní procedury jsou poměrně nepřehledné, proto opět uvedeme jednoduchou ukázkou. Konkrétně zvolíme matematické znalosti jako závisle proměnnou a pro vysvětlení jejich úrovně vybereme stupnici měřící vybavenost domácnosti pro učení (proměnná ASBGHRL, viz opět obr. 7). Výstupy se nyní nabízí v jednotlivých tabulkách, které najdeme v jednotlivých souborech v Excelu, případně přímo ve výstupním okně SPSS. Ve výstupu s názvem „Model“ najdeme hodnotu indexu determinace ( $R^2$ ), která v našem případě pro ČR i SR shodně činí 0,17. Skrze vybavenost domácnosti se nám tedy daří vysvětlit cca 17% rozdílů naměřených žákovských znalostí z matematiky. Další výstup (viz obr. 8) nese název „Coefficients“. V jednotlivých řádcích jsou hodnoty jednotlivých regresních koeficientů („b“) včetně konstanty, společně s hodnotami standardních chyb odhadu („b.se“) a hodnotami testových kri-

<sup>38</sup> S ohledem na nutnost obšírnějšího výkladu v této problematice odkazujeme zde čtenáře na detailní návod k programu IDB Analyzer. Zároveň upozorňujeme na radikální omezení této možnosti práce s ordinálními či nominálními nezávisle proměnnými, které spočívají v tom, že lze využít toliko jedinou nezávisle proměnnou, která nesmí mít více než devět kategorií. Běžnější tedy bude postup, kdy si analytik připraví data (typicky skrze tzv. dummy coding) přímo v datovém souboru v SPSS a poté použije IDB Analyzer pro takto připravená data.

térií pro dílčí  $t$ -testy (b.t). Dále nalezneme hodnoty  $\beta$ -koeficientů („beta“) a jejich standardních chyb („beta.se“).<sup>39</sup>

S ohledem na to, že nabízených výsledků je mnoho a orientace v nich není zcela snadná, ukažme na našem příkladu (obr. 8), jak výsledky korektně vyhodnotit. Zaměříme se jen na Českou republiku. Vliv vybavení domácnosti na matematické znalosti najdeme ve třetím řádku tabulky, hodnota regresního koeficientu činí 19,61 bodu, tj. s nárůstem vybavenosti o jednotku se zvyšují znalosti z matematiky žáka průměrně cca o 20 jednotek.<sup>40</sup> Otázka řešitelná skrze naše výsledky může znít následovně: Lze v ČR prokázat (tj. u žáků 4. ročníku) vliv vybavení domácnosti na matematické znalosti? Lze vyjít z výsledku dílčího  $t$ -testu (srov. hodnotu „b.t“ ve třetím řádku tabulky na obr. 8, tj. hodnotu 16,46). Platí, že pokud absolutní hodnota testového kritéria přesáhne cca hodnotu 2<sup>41</sup>, můžeme vliv nezávisle proměnné na závisle proměnnou brát za statisticky průkazný na 5% hladině statistické významnosti (formálně statisticky tedy zamítáme nulovou hypotézu o neexistenci vlivu, respektive o nulové hodnotě příslušného regresního koeficientu v populaci). Pokud tomu tak není, vliv nebyl statisticky prokázán.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	IDCNTRY	EqVar	beta	b.se	beta.se	b.t	beta.t	xvar	dvar	weight	method	shortcut	pairwise	date	time	reps	
2	Czech Republic	(CONSTANT)	306.68	#NULL!	12.97	#NULL!	23.64	#NULL!	ASBGHRL	ASMMATO	TOTWGT	JRR	Y	N	27-DEC-15	14h27m	75
3	Czech Republic	ASBGHRL	19.61	0.41	1.19	0.02	16.46	17.30	ASBGHRL	ASMMATO	TOTWGT	JRR	Y	N	27-DEC-15	14h27m	75
4	Slovak Republic	(CONSTANT)	326.38	#NULL!	20.57	#NULL!	15.87	#NULL!	ASBGHRL	ASMMATO	TOTWGT	JRR	Y	N	27-DEC-15	14h27m	75
5	Slovak Republic	ASBGHRL	18.45	0.41	1.94	0.03	9.52	13.19	ASBGHRL	ASMMATO	TOTWGT	JRR	Y	N	27-DEC-15	14h27m	75

Obrázek 8 Ukázka výsledku regresní analýzy žakovské znalosti z matematiky v závislosti na vybavení domácnosti

## 2.6 Korelační analýza

Poslední ukázkou je výpočet korelačního koeficientu. Jde o Pearsonův korelační koeficient, který je vhodný jen pro kardinální proměnné, uživatel tedy na to musí při výpočtech pamatovat (software to nijak automaticky nekontroluje). Podobně jako v regresní analýze či proceduře pro výpočet průměrů (srov. výše) je i zde možné pracovat s proměnnými, které jsou vícenásobnými odhady latentních proměnných (opět je nutno v zadání nastavit). Pro jednoduchost spočteme korelační koeficient mezi výsledkem matematických znalostí a znalostmi v oblasti přírodních věd (ASCI01-05).

<sup>39</sup> Připomeňme, že  $\beta$ -koeficienty či standardizované koeficienty slouží pro srovnání vlivu jednotlivých nezávisle proměnných, protože všechny proměnné převedou na stejné měřítko skrze standardizaci (tj. odečtou od hodnot jednotlivých proměnných jejich průměr a poté podělí jejich směrodatnou odchylkou).

<sup>40</sup> Obě škály jsou umělé a nemají žádnou přirozenou interpretaci, proto je hodnota koeficientu v našem případě spíše iluzorní. Nicméně například srovnání se Slovenskem (či jinou zemí) již může přinést meritorně zajímavé výsledky.

<sup>41</sup> Přesněji pro 5% hladinu významnosti již uvedenou hodnotu 1,96.

116 Výsledkem bude tabulka s korelačními koeficienty a jejich standardními chybami (obr. 9). Příslušný korelační koeficient najdeme tam, kde se kříží sloupec a řádek s označením příslušných proměnných (v obr. 9 je to buňka v Excelu s označením D2 pro ČR a D4 pro SR). Korelace tedy činí 0,82 pro Česko a 0,88 pro Slovensko. Věcně lze konstatovat, že tyto dvě dimenze (matematická a přírodovědná) souvisí tedy velice úzce, na Slovensku je souvislost ještě těsnější než u nás. Pro konstrukci intervalu spolehlivosti korelačního koeficientu by opět šlo užít standardních chyb (buňky F2, resp. F4 v obr. 9) a postupů uvedených výše u výpočtu průměrů.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	IDCNTRY	variable	ASMMAT0	ASSSCIO	ASMMAT0	ASSSCIO_s_xvar	weight	method	shortcut	pairwise	date	time	reps	
2	Czech Republic	ASMMAT0	1.00	0.82	0.00	0.01	ASMMAT0_ASSSCIO_TOTWGT	JRR	Y	N	27-DEC-15	14h51m	75.00	
3	Czech Republic	ASSSCIO	0.82	1.00	0.01	0.00	ASMMAT0_ASSSCIO_TOTWGT	JRR	Y	N	27-DEC-15	14h51m	75.00	
4	Slovak Republic	ASMMAT0	1.00	0.88	0.00	0.01	ASMMAT0_ASSSCIO_TOTWGT	JRR	Y	N	27-DEC-15	14h51m	75.00	
5	Slovak Republic	ASSSCIO	0.88	1.00	0.01	0.00	ASMMAT0_ASSSCIO_TOTWGT	JRR	Y	N	27-DEC-15	14h51m	75.00	

Obrázek 9 Ukázka výsledku korelační analýzy znalostí z matematiky přírodních věd

## 2.7 Slabiny IDB Analyzery

Po několika ukázkách v prostředí IDB Analyzer je namístě se zmínit o slabinách tohoto prostředku. Jako slabiny lze jmenovat následující:

1. vazba na jediný komerční software, tj. na SPSS;
2. neobratnost některých zadání (srov. zejména výpočty v regresní analýze);
3. neobratnost výstupů (srov. všechny předchozí ukázky);
4. omezenost na nejjednodušší analytické procedury (tj. umí jen lineární a binární logistickou regresi, neumí jiné než Pearsonovy korelace, nesvede dvouvýběrové  $t$ -testy, analýzu rozptylu<sup>42</sup> a nedokáže nic ze složitějších postupů, např. víceúrovňové či strukturní modely).

Na druhou stranu nutno konstatovat, že IDB Analyzer se ovládá velice jednoduše a pro základní analytické postupy jej lze použít bez větších problémů.

## 3 Přehled speciálního softwaru pro práci s daty z mezinárodních vzdělávacích studií

S ohledem na výše uvedené výhody a pro větší obecnost textu se na závěr zaměříme na přehled dalších softwarových produktů, které umožňují korektní práci s daty z mezinárodních vzdělávacích studií. Cílem této části je opravdu podat přehled,

<sup>42</sup> Tato výtku není technicky zcela korektní, skrze regresní analýzu s nominální či ordinální nezávisle proměnnou lze i skrze IDB Analyzer spočítat dvouvýběrový  $t$ -test (máme-li proměnnou s dvěma kategoriemi), respektive analýzu rozptylu (máme-li nezávisle proměnnou s více kategoriemi).

konkrétní postup pak musí čtenář vyhledat v návodech k těmto produktům. Přehled nadto zcela jistě není úplný, ale uvádí software, který je v současnosti nejčastěji pro tyto účely používán.

### 3.1 Mplus

V současnosti zřejmě nejsnazším komerčním softwarem, který je vhodný pro analýzu dat z mezinárodních vzdělávacích studií, je Mplus (Muthen & Muthen, 2010)<sup>43</sup>. Tento software sice není na první pohled uživatelsky zcela přátelský (ovládá se skrze speciální příkazový jazyk), ale umožňuje snadnou práci s vahami (výběrovými i replikačními) a umí zcela snadno počítat s latentními proměnnými s vícenásobnými odhady. Nadto tento software dovede téměř všechny postupy pokročilé statistické analýzy, mj. víceúrovňové modely, strukturní modely, nelineární regresní modely, které jsou pro data z mezinárodních vzdělávacích studií vhodné. Na stránkách softwaru je dostupné velice bohaté fórum, kde může uživatel načerpat mnoho zkušeností, případně dotazem získat řešení svého problému od tvůrců softwaru. Slabinou Mplus je nemožnost přípravy dat (tj. spojování dat a případné úpravy proměnných je nutno provést jinde). Nadto v českém prostředí je nevýhodou, že Mplus využívá jen několik málo výzkumníků a neexistuje žádná oficiální podpora.

### 3.2 R

V rámci programovacího prostředí R (Venables, Smith, & R Development Core Team, 2013) lze mj. provádět statistickou analýzu dat. Výhodou R je skutečnost, že je zdarma jak základní prostředí, tak i všechny doplňkové balíčky, protože je vyvíjejí jednotliví statistici a matematici z celého světa a předávají je bezplatně ostatním. Pro práci s daty z mezinárodních vzdělávacích studií existuje speciální balíček s názvem „intsvy“<sup>44</sup>. Autorem je Daniel Caro, působící v Oxfordu v týmu zaměřeném právě na zpracování dat z mezinárodních vzdělávacích studií. Obdobně jako IDB Analyzer umí tento balíček spojovat a analyzovat data. Obdobně jako IDB Analyzer je i balíček intsvy omezen na základní procedury, tj. počítá průměry, procenta, regrese a korelace. Při výpočtech přitom korektně zohledňuje váhy (výběrové i replikační) a umí správně pracovat s případnými vícenásobnými odhady hodnot latentních proměnných.

### 3.3 Další software

Kromě výše uvedených produktů je možné využívat i některé další statistické programy. Pro akademiky se konkrétně nabízí hojně na západních univerzitách užívaná Stata či SAS, který je díky štědrému univerzitnímu programu dostupný také na mno-

<sup>43</sup> Software (demo zdarma) i veškeré informace jsou dostupné on-line na adrese <http://www.statmodel.com>.

<sup>44</sup> Jde o zkratku International Assessment Data Manager, základní popis je dostupný on-line na adrese <http://CRAN.R-project.org/package=intsvy>.

118 ha vysokých školách (včetně ČR). Oba produkty jsou obecné statistické programy analogické například k SPSS, nicméně na rozdíl od SPSS umí snáze pracovat s replikačními vahami či latentními proměnnými s vícenásobnými odhady jejich hodnot. Nadto oba produkty dokážou pracovat s pokročilejšími statistickými technikami (mj. víceúrovňovými<sup>45</sup> či strukturními modely). Výhodou SAS oproti Stata je skutečnost, že data z mezinárodních vzdělávacích studií jsou kromě formátu SPSS nabízena běžně též ve formátu SAS. Stata i SAS mají velice bohatou podporu ve formě literatury i uživatelských fór, v případě Stata je nadto možné stávající procedury dotvářet, respektive přidávat vlastní nové procedury. I při práci se Stata či SAS lze doporučit nejdříve spojit data skrze IDB Analyzer, a teprve poté provádět vlastní analýzy, aby nedošlo k chybám již při přípravě dat.

#### 4 Doporučení pro práci s daty z mezinárodních vzdělávacích studií

Cílem tohoto článku je upozornit na úskalí při práci s daty z mezinárodních vzdělávacích studií a podat přehled možností korektní práce s těmito daty. Z detailních ukázek prostředí IDB Analyzer i předchozího přehledu plyne ne zcela lichotivé vyznění: *korektní práce s daty z mezinárodních vzdělávacích studií není zcela snadná*. Vždy je nutné vybrat některý softwarový prostředek, a ten se naučit a pochopit jeho případné slabiny. Nadto je nutné mít poměrně slušný přehled o designu příslušné mezinárodní vzdělávací studie. Menší přehled je potřebný při užívání IDB Analyzera nebo balíčku intsvy v prostředí R. Pro složitější analýzy je pak nutná detailnější znalost designu i znalost z oblasti pokročilých statistických metod, aby nedošlo ke zneužití statistiky, ale jejímu korektnímu užívání. Obecně lze doporučit, aby si uživatel vybral jedno prostředí, které mu vyhovuje nebo které už zná, a to užíval, práce s několika programy najednou spíše analytickou činnost komplikuje. V nesnázích lze využít rad statistiků, zejména těch, kteří data z mezinárodních vzdělávacích šetření znají.

#### 5 Omezení při práci s daty z mezinárodních vzdělávacích studií a praktická doporučení

Kromě výše uvedených problémů na nás číhají ještě další v článku nezmíněné komplikace, které přináší analytická práce s daty z mezinárodních vzdělávacích studií. Tyto problémy by jistě vydaly na samostatné texty (snad v budoucnu vzniknou), nicméně zde zmiňme dva jako ukázkou:

1. problematičnost mezinárodních srovnání, odlišnost kurikula v jednotlivých systémech, případně jejich neznalost;
2. víceúrovňovost dat a jejich korektní analýza.

<sup>45</sup> Ve Stata jde o sadu programů s názvem GLAMM, v SAS o proceduru MIXED nebo NLMIXED, respektive o proceduru CALIS.

Ad 1. Při práci s daty z mezinárodních vzdělávacích studií je velikým pokušením pro analytika, aby činil mezinárodní srovnání, protože má zpravidla k dispozici data z mnoha desítek zemí (a často ještě více vzdělávacích systémů<sup>46</sup>). Musíme si uvědomit, že kurikula v jednotlivých vzdělávacích systémech mohou být velice odlišná, a v extrémní situaci tak může dojít k tomu, že znalost testovaná v některém mezinárodním vzdělávacím šetření není v testovaném ročníku (ani dříve) ještě rozvíjena. Bez zohlednění této skutečnosti může být interpretace mezinárodního srovnání zcela nesprávná. Problémem ale je, že zřejmě žádný výzkumník nezná do detailu kurikula všech vzdělávacích systémů a tato skutečnost znamená, že mezinárodní srovnání je téměř nemožné. Není tedy zcela od věci vzít v potaz doporučení traktovaná na schůzkách mezinárodních vzdělávacích studií, tj. že *se má přednostně provádět národní analýza těchto dat a srovnání mají být prováděna jen s obdobnými vzdělávacími systémy, které nadto výzkumník zná.*

Ad 2. Kromě všech komplikací při práci s daty z mezinárodních vzdělávacích studií popsaných na počátku článku je nutno upozornit na víceúrovňový charakter dat ze vzdělávacích studií (srov. Soukup, 2006, a literatura tam uvedená). V případě, že je studie zaměřená na žáky z jedné školy, respektive třídy, platí, že tito žáci jsou si vzájemně podobnější než žáci z různých škol či tříd. Korektní práce s těmito daty pak vyžaduje používání víceúrovňových modelů, které jsou rozšířením běžné regresní analýzy či dalších statistických postupů. Kromě výše popsaných modifikací je tak nutné ještě využít víceúrovňové výpočty, a tím se situace ještě komplikuje (z popsaných produktů jsou na tyto výpočty připravené Mplus, SAS a Stata<sup>47</sup>). Do budoucna lze očekávat další rozšiřování všech popsaných produktů a díky tomu bude korektní práce s daty z mezinárodních vzdělávacích studií snazší a dostupnější.

### Poděkování

Autor touto cestou děkuje oběma anonymním recenzentům a dále pak Dominiku Dvořákovi, který autora upozornil na mnohé nedostatky v textu. Samozřejmě, že za případné chyby textu je plně odpovědný autor textu.

### Literatura

- Basl, J. (2014). Statistika ve školství: Eurostat, OECD, PISA, IEA. In J. Hendl et al., *Statistika v aplikacích* (s. 287–304). Praha: Portál.
- Foy, P., Arora, A., & Stanco, G. M. (Eds.). (2013). *TIMSS 2011 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Hendl, J. (2012). *Přehled statistických metod: analýza a metaanalýza dat*. Praha: Portál.
- Mareš, P., Rabušic, L., & Soukup, P. (2015). *Analýza sociálněvědních dat (nejen) v SPSS*. Brno: Masarykova universita.

<sup>46</sup> Vycházíme zde ze skutečnosti, že v rámci některých zemí existuje paralelně více vzdělávacích systémů.

<sup>47</sup> Více se lze o těchto problémech dočíst například v Rutkowski et al. (2010).

- 120 Muthen, L. K., & Muthen, B. (2010; 6th Ed.). *Mplus user's guide*. Los Angeles, CA: Muthen & Muthen.
- Petrůšek, I. (2015). *Analýza chybějících hodnot*. Praha: Sociologický ústav AV ČR.
- Rutkowski, L., Gonzalez, E., Joncas, M., & Davier, M. von. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151.
- Soukup, P. (2006). Proč užívat hierarchické lineární modely? *Sociologický časopis / Czech Sociological Review*, 42(5), 987–1012.
- Soukup, P. (2012). Mezinárodní výzkumy v oblasti vzdělávání. In J. Krejčí & J. Leontiyeva (Eds.), *Cesty k datům: zdroje a management sociálněvědních dat v České republice* (s. 287–304). Praha: Sociologické nakladatelství (SLON) a Sociologický ústav AV ČR.
- Urbánek, T., & Šimeček, M. (2001). Teorie odpovědí na položku. *Československá psychologie* 45(5), 428–440.
- Venables, W. N., Smith, D. M., & R Development Core Team (2013). *An introduction to R*. Vienna: R Foundation for Statistical Computing.

PhDr. Ing. Petr Soukup  
Fakulta sociálních věd UK  
U Kříže 8, 158 00 Praha 5 – Jinonice  
soukup@fsv.cuni.cz