Mezinárodní vzdělávací studie a práce s jejich daty

Petr Soukup

Plzeň 18. 9. 2015

Základní cíl setkání

- Podat stručný přehled mezinárodních vzdělávacích studií, kterých se ČR účastnila v posledních 15 letech
- Ukázat datové zdroje, které v rámci studií vznikají a jejich složitost

1. ČÁST

- Upozornit na „problémy“, které souvisí s daty z mezinárodních vzdělávacích výzkumů
- Předvést možnosti zpracování dat, které na „problémy“ v datech reagují
- Ukázat reálné analýzy na příkladu dat TALIS 2011

2. ČÁST

Přehled mezinárodních vzdělávacích studií

Přehled studií

- **A) PISA**
- **B) TIMSS**
- **C) PIRLS**
- **D) ICCS resp. CIVED**
- **E) ICILS**
- **E) PIAAC**

Přehled studií

	PISA	TIMSS	PIRLS	ICCS	ICILS	PIAAC
Testovaná oblast	Mat., čtení, přírodní vědy	Mat., přírodní vědy	čtení	Výchova k občanst.	Počítačová a IT gramot.	Mat., čten. gram. a řešení probl.
Testovaný ročník/věk	15 let	4./8. roč.	4. roč.	8. roč.	8. roč.	dospělí
Počet účastníků	2015:60+	2011:60+	2011:40+	2009: 30+	2013: 21	2011: 32

Periodicita a materiály

	PISA	TIMSS	PIRLS	ICCS	ICILS	PIAAC
Perioda	3 roky	4 roky	5 let	7 let	8 let	????
Poslední vlna	2015	2011	2011	2009	2013	2011
Dotazníky	Ředitel, žák (rodič*)	Ředitel, učitel, žák, rodič	Ředitel, učitel, žák, rodič	Ředitel, učitel, žák	Ředitel, učitel, žák	Testovaná osoba

* Není povinné

Mezinárodní koordinace

- **I) PISA + PIAAC - OECD**
- **II) TIMSS+PIRLS+ICCS+ICILS - IEA**

Základní odkazy anglicky

- **PISA: http://www.oecd.org/pisa/home/**
- **TIMSS a PIRLS: http://isc.bc.edu/**
- **ICCS: http://iccs.acer.edu.au/**
- **ICILS: http://www.iea.nl/icils_2013.html**
- **PIAAC:**
- **http://www.oecd.org/site/piaac/surveyofadultskills.htm**

- **Rozcestník pro všechna data s výjimkou PISA a PIAAC je zde: http://rms.iea-dpc.org/#**

- **Poznámka: Stránky obsahují podrobné informace o výzkumech včetně všech publikací,mezinárodní data, výzkumné materiály (dotazníky)**

Další info česky

- **http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PISA**
- **http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/TIMSS**
- **http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PIRLS**
- **http://www.csicr.cz/Prave-menu/Projekty-ESF/Projekt-ESF-Kompetence-I**
- **http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/ICILS**
- **http://www.piaac.cz/**

- **Poznámka: Stránky obsahují informace o české realizaci výzkumu, česká data, výzkumné materiály (dotazníky) a výzkumné zprávy**

Mnohost dat na národní a mezinárodní úrovni

Praktický problém při práci s daty -TIMSS

Typ souboru	Popis
Žáci – 4. ročník	Datový soubor s výsledky testů a odpověďmi žáků z dotazníku
Žáci – 8. ročník	Datový soubor s výsledky testů a odpověďmi žáků z dotazníku
Učitelé - 4. ročník	Datový soubor s odpověďmi učitelů 4. ročníku
Učitelé - 8. ročník - matematika	Datový soubor s odpověďmi učitelů matematiky v 8. ročníku
Učitelé - 8. ročník – přírodní vědy	Datový soubor s odpověďmi učitelů přírodních věd
Školy – 4. ročník	Datový soubor s odpověďmi ředitelů škol, kde se účastnily výzkumu 4. ročníky
Školy – 8. ročník	Datový soubor s odpověďmi ředitelů škol, kde se účastnily výzkumu 8. ročníky

Praktický problém při práci s daty -TIMSS

- **Za jednu zemi může být až 7 souborů!**
- **Zapojeno je více než 60 zemí**
- **Reálně je cca 250 datových souborů**
- **Co a jak spojit?**
- **Složité, může pomoci IDB Analyzer (viz 2. část)**

Exkurz o identifikaci respondentů

- **ID země**
- **ID školy**
- **ID žáka**
- **Příp. ID učitele a třídy**
- **Ukázka na datech TIMSS**

Problémy v datech z mezinárodních vzdělávacích studií

Problémy v datech

- **Data nepochází z prostého náhodného výběru**
- **nutnost vážení pro korektní odhady populačních parametrů**
- **vážení na více úrovních**
- **nutnost speciálních výpočtů rozptylu výběrových charakteristik**

Problémy v datech

- **Nutnost vážení pro korektní odhady populačních parametrů, plus vážení na více úrovních**
- **Data nepochází z prostého náhodného výběru (nutnost speciálních výpočtů rozptylu výběrových charakteristik)**
- **Některé proměnné neměříme přímo a chceme zohlednit jejich chybu měření (IRT metodologie a plausible values)**

Vážení dat a jeho podstata

Typy vah a vážení

- Designové – víme předem o nerovné pravděpodobnosti vybrání
- Poststratifikační – ex post dovažujeme data, sběr dat, nebyl dle představ
- Boost – připojení dalšího datového souboru, zaměřeného jen určitou podskupinu

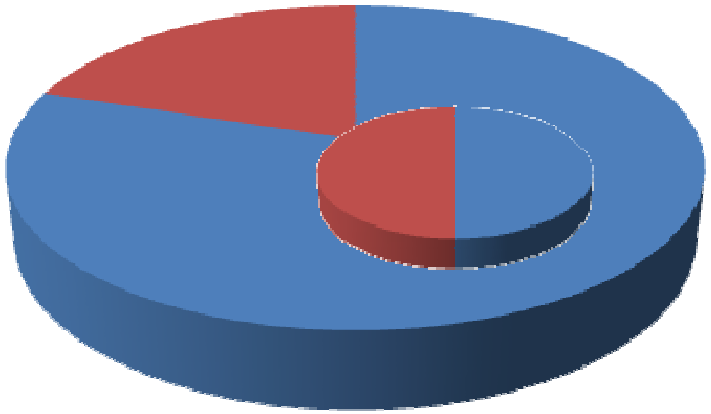
Proč vážíme?

- Díky vážení lze odhadovat výsledky na celou populaci (průměry, procenta atd.)
- Bez vah lze pouze popisovat náš výběrový soubor a usuzovat na rozdíly mezi skupinami, lze též modelovat vztahy mezi proměnnými (ty váhy příliš neovlivní

Základní soubor → výběr

- Příklad: 20% učitelů jsou na soukromých školách a 80% na veřejných.
- Ve výběru ale máme např. proporci 50% a 50%
- Co s tím?

- Řešení: váha, která disproporci napraví



Technický princip vážení I

- Jedna proměnná:
- Příklad: V populaci je poměr veřejné vs. soukromé 80% a 20 %, ale v sebraném výběru je 50% a 50%
- N: 2000
- N_v: 1000
- N_s: 1000
- My chceme ale uměle snížit počet učitelů na soukromých školách a navýšit na veřejných tak, aby výsledný počet vypadal takto:
- N: 2000
- N_v: 1600
- N_s: 400

- Výpočet váhy:
- W_v: 1600/1000 = 1,6, tj. každý učitel z veřejné školy bude v datech obsažen 1,6 krát

- W_s: 400/1000 = 0,25, 4učitelé na soukromé škole budou tvořit jednoho člověka svým vlivem

Co je tedy váha?

- Hodnota přiřazená každému respondentovi a (v případě vzdělávacích studií je to ještě komplikovanější váha je přiřazena též škole – důvod = vícestupňový výběr)
- Váha udává kolik jednotek v základním souboru reprezentuje vybraný jedinec (škola, učitel, apod.)
- Jde o hodnotu v intervalu <0;∞>, hodnoty nad 1 značí, že daná jednotka je v datech podreprezentována, hodnota pod 1 naopak značí nadreprezentaci

Vážení ve vzdělávacích výzkumech

Úrovně vážení ve vzdělávacích výzkumech (complex design)

- Váhy jsou kombinací designových vah a poststratifikačních
- Designové váhy se uplatňují na jednotlivých úrovních, kde se vybírá a vyrovnávají nestejné pravděpodobnosti vybrání
  - Vysoká pravděpodobnost vybrání → malá váha
  - Nízká P vybrání → velká váha.
- Rozlišují se tedy váhy:
  - školní,
  - učitelské atd.

Váhy poststratifikační – úpravy neúčasti žáků

- Váhy se korigují díky neúčasti žáků na výzkumu
- Neúčast opět může nastat na všech úrovních (vypadne škola, žák atd.)
- Poststratifikační korekce na neúčast řeší tento problém (vyšší váha tam, kde se neúčastnilo více žáků a vice versa)

Shrnutí k vahám

- Váhy napravují designová vychýlení a neúplnou návratnost dotazníků
- Váhy je nutno užívat, aby byly správné odhady charakteristik v populaci (průměrů, procent atd.)
- Váhy mohou být různého typu (školní, žákovská, národní), vždy musíme zvolit tu správnou případně jejich kombinaci

Jiné než náhodné výběry: Standardní chyby odhadu a jejich korektní odhady

Co je to standardní chyba odhadu (SE)?

- Pro každou odhadovanou (výběrovou) charakteristiku (průměr, procento atd.) lze stanovit standardní chybu jejího rozdělení
- Výběrové rozdělení je rozdělení sledované charakteristiky pro všechny potenciálně možné výběry získané stejnou technikou o stejné velikosti
- Samozřejmě toto je teorie, v praxi nevybíráme všechny možné výběry, ale provádíme odhad SE za pomoci chytrých vzorců či složitějších postupů

K čemu jsou dobré standardní chyby odhadů?

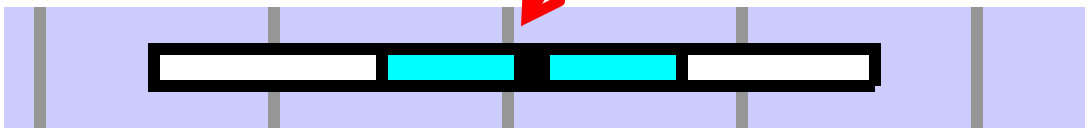
- Výzkumy (vč. mezin. vzděl.) vycházejí z výběrových dat
- Naším cílem je zobecnit výsledky na populaci (cílovou skupinu)
- Počítáme intervaly spolehlivosti (průměrů, procent), provádíme statistické testy (odlišností, závislostí – např. t-test, regrese, korelace atd.)
- Pro tyto účely je třeba znát standardní chyby odhadovaných parametrů (určují přesnost našich závěrů, které vztáhneme na populaci)

Standardní chyba odhadu a interval spolehlivosti

- ε – sledovaná charakteristika (průměr, procento...)
- 95% interval spolehlivosti lze vypočítat:

$$\varepsilon \pm 1.96 \times SE_{\varepsilon}$$

- V tabulkách z výzkumů bývá značen jako černý pruh:



569 (2.4)

- 95%,interval spolehlivosti pro průměr v populaci je mezi 554.3 a 563.7...

Jak spočítat st. chybu?

- Prostý náhodný výběr:
  - Směrodatná odchylka sledované proměnné (*s*) se dělí odmocninou počtu pozorování (*n*)

$$\hat{SE}_{\varepsilon} = \frac{s}{\sqrt{n}}$$

- V reálných výběrech (*complex samples),* které se realizují to ale nejde tak snadno

Základní efekty

- Vybírání ve školách (tzv. clustered sample):
  - Žáci v jedné škole jsou si navzájem podobnější než z různých škol (vliv učitele, prostředí, sousedství atd.)
  - Tento proces vede k navýšení standardní chyby (pokles přesnosti odhadů)
- Stratifikovaný výběr
  - Vybíráme separátně v různých skupinách škol (např. v ČR soukromé vs. veřejné atd.)
  - Tento proces vede ke snížení standardní chyby (nárůst přesnosti odhadů)

I váhy mají dopad

- Platí (prokázáno zejména simulacemi, zde už vzorce nestačí):
  - Při užití vah roste standardní chyba (klesá přesnost odhadů)
  - Samozřejmě se komplikují výpočty

- Závěr: Je nutno vše zohlednit a počítat s tím, že běžně jsou standardní chyby větší než v prostém náhodném výběru

Jak lze řešit?

- Statistici vyvinuli v zásadě dva základní postupy:
  - Replikace
  - Užití Taylorových polynomů
- Běžný software často toto neumí (platilo donedávna i o SPSS), případně užívá jednodušší postup (Taylor), který je zpravidla horší

- Řešení: Lze užít speciální software (např. VESVAR), případně doplňkové programy k běžnému software (tak pracuje IDB Analyzer)

Jackknife Repeated Replication

- Jedna z možností replikace: Jackknife Repeated Replication (JRR), uplatněna v IDB Analyzer
- Lze užít pro odhad standardních chyb v tzv. complex designs
- Základní postup: provést výpočet sledované charakteristiky (průměru, procenta) na mnoha výběrech (z výběru):
  - Tak, že nastavíme váhy pro některé školy na nulu,
  - A naopak pro jiné na dvojnásobek
- Odhad standardní chyby vychází z rozptylu sledované charakteristiky z těchto replikovaných výběrů

JRR-ukázka

School	Student	JKZone	JKRep	Score	Weight	RWgt1	RWgt2	RWgt3	RWgt4	RWgt5
01	0101	1	1	96	6	**12**	6	6	6	6
01	0102	1	1	74	7	**14**	7	7	7	7
02	0201	1	0	56	9	**0**	9	9	9	9
02	0202	1	0	62	2	**0**	2	2	2	2
03	0301	2	0	34	2	2	**0**	2	2	2
03	0302	2	0	53	7	7	**0**	7	7	7
04	0401	2	1	96	5	5	**10**	5	5	5
04	0402	2	1	57	2	2	**4**	2	2	2
05	0501	3	0	78	2	2	2	**0**	2	2
05	0502	3	0	87	2	2	2	**0**	2	2
06	0601	3	1	71	8	8	8	**16**	8	8
06	0602	3	1	47	7	7	7	**14**	7	7
07	0701	4	0	56	1	1	1	1	**0**	1
07	0702	4	0	25	2	2	2	2	**0**	2
08	0801	4	1	82	2	2	2	2	**4**	2
08	0802	4	1	68	2	2	2	2	**4**	2
09	0901	5	1	16	10	10	10	10	10	**20**
09	0902	5	1	80	8	8	8	8	8	**16**
10	1001	5	0	20	5	5	5	5	5	**0**
10	1002	5	0	35	9	9	9	9	9	**0**
Unweighted Mean:		**59.7**		**Sum of Weights:**		**100**	**96**	**109**	**99**	**102**
Weighted Mean:		**57.2**		**Weighted Means:**		**60.7**	**60.0**	**56.6**	**58.6**	**58.8**

*2

*0

Shrnutí – důvody užívání speciálních sw k odhadu SE

- Běžně užívané výběrové designy jsou výrazně odlišné od prostého náhodného výběru, standardní chyba je díky tomu větší než udávají běžné výpočty
- Běžný software nemá postupy pro korektní odhad standardních chyb, případně neužívá úplně nejvhodnější postupy

Latentní proměnné, IRT a chyba měření

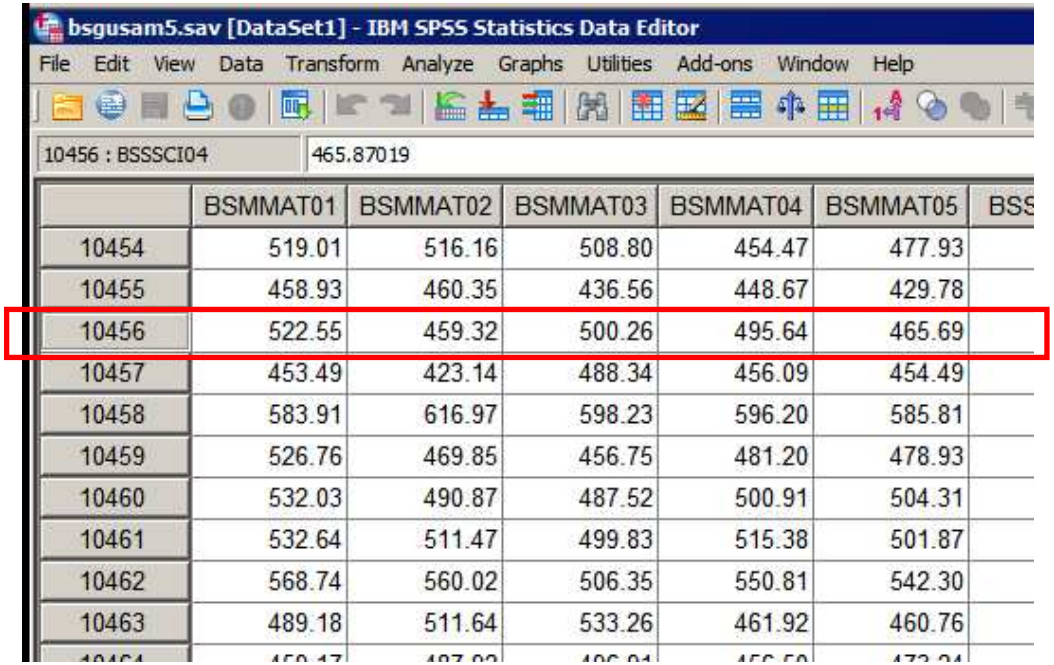
Latentní proměnné, IRT

- Mnoho věcí neměříme přímo (příklady)
- Běžné testování a počítání výsledků (skrze CTT) nezohledňuje obtížnost jednotlivých úloh, jejich diskriminační schopnost, možnost hádání atd.
- Běžné testování neumí podchytit chybu měření
- Řešení: IRT metodologie, která tyto problémy bere v potaz při odhadu výsledného skóre (více viz Urbánek, Šimeček)

Reálné využití IRT ve vzdělávacích studiích

- Běžně pro odhad výsledků testované gramostnosti – čtení, matika atd.
- V datech pak máme zpravidla 5 proměnných, které charakterizují jedincovu schopnost a chybu měření (tzv. plausible values)
- Korektní práce s daty: 5x spočti analýzu a výsledky poté zprůměruj (to je ale trochu moc práce a tak se to automatizuje – viz dále IDB Analyzer)
- Poznámka: Pro proxy výsledky lze pracovat i s jedinou hodnotou PV, když náš SW více neumí

Reálná ukázka IRT v TIMSS



	BSMMAT01	BSMMAT02	BSMMAT03	BSMMAT04	BSMMAT05	BSS
10454	519.01	516.16	508.80	454.47	477.93	
10455	458.93	460.35	436.56	448.67	429.78	
10456	522.55	459.32	500.26	495.64	465.69	
10457	453.49	423.14	488.34	456.09	454.49	
10458	583.91	616.97	598.23	596.20	585.81	
10459	526.76	469.85	456.75	481.20	478.93	
10460	532.03	490.87	487.52	500.91	504.31	
10461	532.64	511.47	499.83	515.38	501.87	
10462	568.74	560.02	506.35	550.81	542.30	
10463	489.18	511.64	533.26	461.92	460.76	

Rozdíl 523-459=64 bodů, tj. velká chyba měření u konkrétního žáka

Poznámka: Mezinárodní výzkumy většinou testové výsledky standardizují, mezinárodní průmer činí 500 jednotek a mezinárodní směrodatná odchylka činí 100 jednotek (toto pomáhá interpretaci výsledků).

IDB Analyzer (dále jen „IA“)

IA – základní informace

- Software vyvinutý IEA DPC (Hamburg) pro **SPOJOVÁNÍ** a **ANALÝZU** dat z velkých mezinárodních výzkumů (TALIS, TIMMS, ICCS, nově i PISA)
- Spolupracuje se statistickými pakety (konkrétně s SPSS od verze 15) a dále umožňuje export výsledků též do Excelu
- Ke stažení na http://www.iea.nl/data.html (pro PISA existuje modifikovaná verze)
- V současnosti verze 3.1.25 (září 2015)

IA – technické informace

- Nutno mít instalován MS Excel a .NET4
- Nutno mít administrátorská práva
- Poměrně malý program, který „jen“ generuje programovací kódy (syntax) pro SPSS

Spojování dat (Merge Module)

IA – Merge Module I

- Lze spojovat data za různé země
- Lze spojovat data z různých instrumentů
- Lze vybrat jen některé proměnné ke spojení
- Automaticky se do spojených dat kopírují váhy, replikační proměnné a identifikační proměnné (IDSCHOOL, IDSTUD, aj.)
- Výstupem je datový soubor pro SPSS (*.sav) nebo syntaxe pro spojení (*.sps)

IA – Merge Module II

- Jak funguje?
  - Připraví syntax pro SPSS, krerý se odkazuje na makra předem napsaná v IEA DPC (uloží se při instalaci IDB Analyzer na C:\Program Files\IEA\IDBAnalyzer\data\templates
  - Výstupem jsou syntaxe pro SPSS a po jejich spuštění data, která se ukládají do adresáře WORK (ten se automaticky vytváří při instalaci IDB Analyzer)

IA – technické finty spojování

- Jedna šipka – přesun proměnné, dvojitá šipka – přesun všech proměnných
- Nutno vždy stisknout tlačítko Start SPSS (pak se otevře SPSS)
- V proměnných se dá vyhledávat (buď v technických názvech – NAME nebo v popiscích – DESCRIPTION)
- S připraveným spojeným souborem lze provádět analýzy
- **! Zrada – nejdříve nutno naklikat typy souborů ke spojení (učit a žákovské) a pak teprve přesouvat proměnné**
- Spouštění syntaxu SPSS: Ctrl+A a poté Ctrl+R

Analýza dat (Analysis Module)

IA – Analysis Module I

- Proč jej užívat?
  - Zohledňuje design (pracuje se správnými vahami), odhady charakteristik pro populaci jsou správně (samozřejmě jen pokud váhy jsou správně☺)
  - Počítá korektně standardní chyby odhadů (JRR replikací) – díky tomu jsou korektní testy a intervaly spolehlivosti
  - Umožňuje pracovat se škálami vzniklými z IRT modelů

IA – Analysis Module II

- Co umí?
  - Výpočet procent
  - Výpočet průměrů
  - Korelace (dle Pearsona)
  - Lineární regresi (tj. i ANOVA)
  - Výpočet skupin znalostí dle testů (tzv. benchmark)
  - Průměry, regrese, korelace jsou možné i s tzv. plausible values z IRT

IA – Analysis Module III

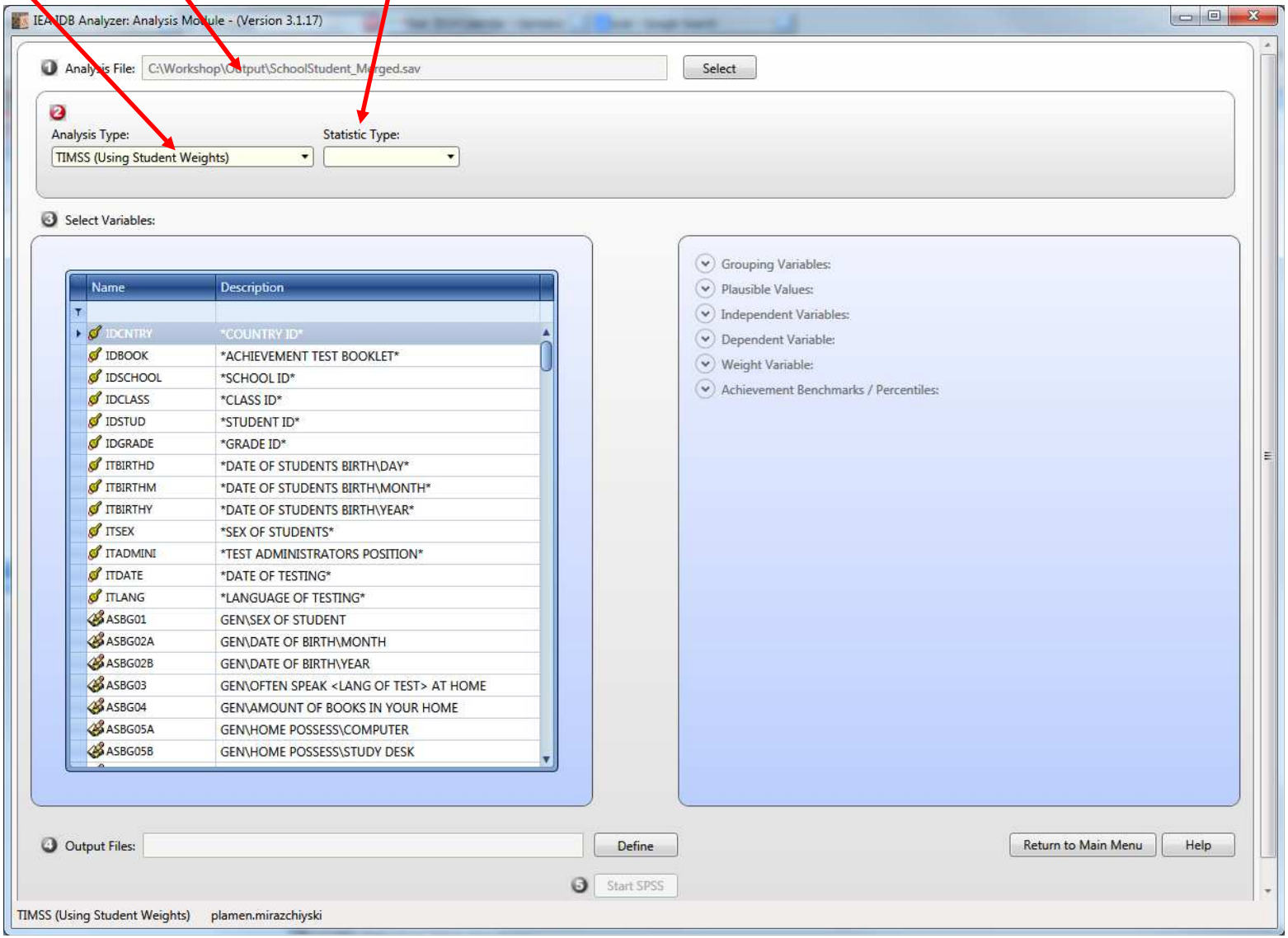
- Jak pracuje?
  - Připraví syntax pro SPSS či SAS
  - Tento syntax za pomoci předdefinovaných maker provádí JRR replikace případně jiné potřebné výpočty
  - Výsledky se vytvoří jako textový výstup do SPSS nebo se uloží do Excelu (dle požadavku, který zadáme v okénku Output Files)
  - Poznámka: Výstup v Excelu je pro další editace (zejména automatické) vhodnější)
- Spouštění syntaxu SPSS: Ctrl+A a poté Ctrl+R

IA – Analysis Module IV

- Jaké má nedostatky?
  - Umí ze statistiky poměrně málo (nutno tak užívat další software či doplňkové pomůcky)
  - Při zadání jakékoli volby smaže vše ostatní
  - Spolupracuje jen s jedním komerčním balíkem (SPSS)
  - Výstupy jsou v mírně neobvyklé podobě

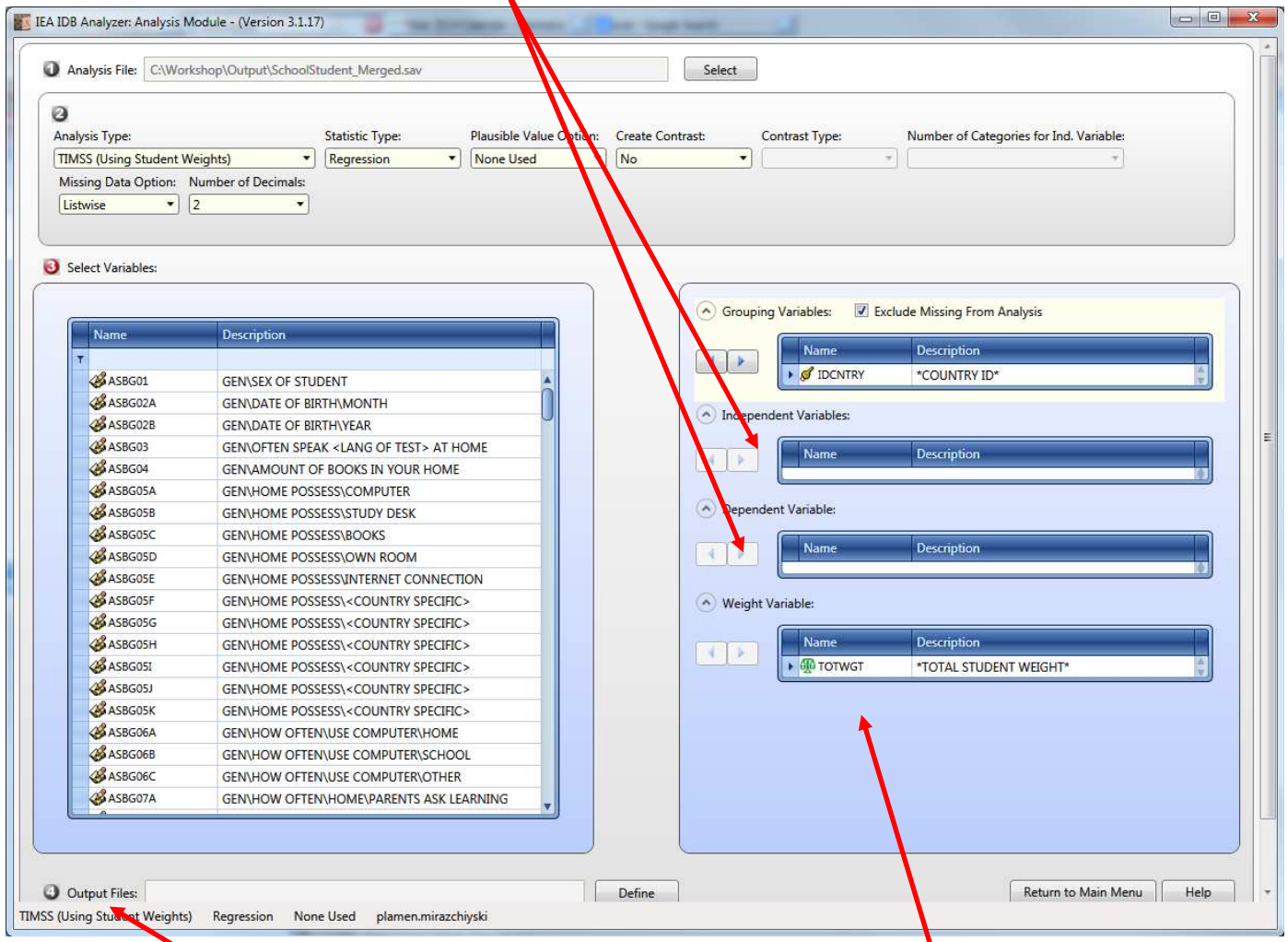
IA – ukázka

1. Vybrat soubor
2. Typ analýzy
3. Co chci počítat



IA – ukázka

4. Poté vybírám proměnné, které chci analyzovat (různé procedury různé počty)



5. Na konci musím určit, kam chci uložit výsledky

Poznámka: Váha se mi vybere automaticky a správně sama

Užitečné informace

IDB Analyzer – další info

- Obsahuje Help (viz 3. nabídka při spouštění programu)
- Verze lze aktualizovat opět přes spouštění programu přes web IEA (při instalaci novější verze se automaticky odinstaluje verze předchozí)

Ukázky práce

Ukázky v IDB Analyzer

- Spojení českých dat TIMSS
- Průměry po skupinách
- Korelace s 5 PV hodnotami mat. dovedností (Listwise/pairwise pro MV)
- Regrese s 5 PV hodnotami mat. dovedností

Díky za pozornost

soukup@fsv.cuni.cz