ČŠI | Česká školní inspekce

Metodika sběru a analýzy dat z výsledků šetření České školní inspekce

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

2021/2022

ČŠI | Česká školní inspekce

Metodika sběru a analýzy dat z výsledků šetření České školní inspekce

---

doc. PhDr. Tomáš Lebeda, Ph.D.
Mgr. et Mgr. Jakub Lysek, PhD.
doc. Mgr. Daniel Marek, M.A., Ph.D.
Mgr. Monika Brusenbauch Meislová, Ph.D.
Mgr. Roman Folwarczny
Mgr. Michal Soukop
Mgr. Kateřina Zymová
Mgr. Markéta Zapletalová, PhD.
doc. RNDr. PhDr. Oldřich Hájek, Ph.D., MBA
Bc. Jakub Janega
Bc. Barbora Macková
Bc. Stanislav Daniel

© Česká školní inspekce, Praha 2021
ISBN 978-80-88087-70-0 (online ; pdf)
ISBN 978-80-88087-71-7 (online ; ePub)

OBSAH

1 ÚVOD 6
2 STÁVAJÍCÍ SBĚR DAT A MOŽNOSTI PROPOJOVÁNÍ 8
2.1 INTERNÍ ŠETŘENÍ 8
2.1.1 KOMPLEXNÍ INSPEKČNÍ ČINNOST 8
2.1.2 NÁRODNÍ ZJIŠŤOVÁNÍ VZDĚLÁVACÍCH VÝSLEDKŮ 9
2.2 EXTERNÍ (MEZINÁRODNÍ) ŠETŘENÍ 10
2.3 STRUČNÁ DISKUZE ROZDÍLŮ 12
3 ANALÝZA DAT 16
3.1 KVANTITATIVNÍ VS. KVALITATIVNÍ PŘÍSTUP K VÝZKUMU KAUZALITY 16
3.1.1 KVANTITATIVNÍ PŘÍSTUP 16
3.1.2 KVALITATIVNÍ PŘÍSTUP 16
3.2 ZÁKLADNÍ A VÝBĚROVÝ SOUBOR 20
3.3 PROMĚNNÉ 21
3.3.1 TYPY PROMĚNNÝCH 21
3.3.2 ÚROVEŇ PROMĚNNÝCH 22
3.4 POČÁTEČNÍ KONTROLA A ÚPRAVA DAT 23
3.5 VYBRANÉ ANALYTICKÉ TECHNIKY 24
3.5.1 DESKRIPTIVNÍ ANALÝZA DAT 24
3.5.2 POROVNÁNÍ PRŮMĚRŮ 28
3.5.3 KORELAČNÍ KOEFICIENTY 30
3.5.4 VYBRANÉ TYPY REGRESNÍ ANALÝZY 31
3.5.5 POPISNÁ A KAUZÁLNÍ INTERPRETACE REGRESNÍCH KOEFICIENTŮ 35
3.6 DALŠÍ VYBRANÉ TECHNIKY 36
3.6.1 SHLUKOVÁ ANALÝZA 36
3.6.2 FAKTOROVÁ ANALÝZA 36
3.6.3 STRUCTURAL EQUATION MODELLING (SEM) 37
3.7 POKROČILÉ STATISTICKÉ TECHNIKY PRO VÝZKUM KAUZALITY 39
3.7.1 MATCHING METHODS 39
3.7.2 METODA DIFFERENCE-TO-DIFFERENCE 40
4 VIZUALIZACE VÝSLEDKŮ 42
4.1 VIZUALIZACE POMOCÍ GRAFŮ 42
4.2 VIZUALIZACE POMOCÍ MAP 46
5 CHYBĚJÍCÍ HODNOTY (MISSING VALUES) 50
6 ZÁVĚR 52
LITERATURA 54

1

Úvod

1 ÚVOD

Cílem tohoto výstupu je předložit sadu analytických postupů, které se věnují sběru dat, analýze dat a vizualizaci jejich výstupů. Tento dokument shrnuje metody, postupy a poznatky vycházející z práce na projektu *KSH: Komplexní systém hodnocení,* konkrétně z práce na klíčové aktivitě KA5. Vycházejí tak především z výstupů sekundárních analýz mezinárodních šetření a dílčích výstupů založených na národních datech. Řada z těchto výstupů propojovala data z mezinárodních šetření s domácí datovou základnou ČŠI, MŠMT a ČSÚ. Cílem tohoto dokumentu je shrnout dosavadní zkušenosti výzkumného týmu KA5 do jednoho dokumentu, který standardizuje jednotlivé postupy a usnadní tak v budoucnu ČŠI a dalším uživatelům efektivní analýzu těchto dat. Dokument tedy slouží především pracovníkům ČŠI, ale i odborné veřejnosti, která má zájem analyzovat veřejně dostupná data české vzdělávací soustavy. Naopak snahou dokumentu nebylo navrhovat postupy, jež primárně nesouvisejí s oblastmi řešenými v rámci KA5 projektu KSH.

Součástí snahy o nastavení korektních metodických a analytických postupů je k tomuto dokumentu komplementární výstup *Sada typizovaných analytických nástrojů v prostředí standardních statistických programů vycházejících z metodiky* (dále pak *Sada typizovaných analytických nástrojů)*. Tato *Sada typizovaných analytických nástrojů* je věcně propojena s touto metodikou a společně dávají uživatelům potřebné know how ke sběru, analýze a vizualizaci předmětných dat. Oba dokumenty se snaží na jedné straně respektovat reálná omezení daná specifickými prvky vzdělávací soustavy i dosavadním procesem zjišťování a na straně druhé v maximální možné míře aplikovat moderní statistické metody respektující přísné akademické standardy.

Dokument vychází ze zkušeností získaných během dílčích činností projektu systémového Komplexní systém hodnocení, zejména pak z realizovaných Pilotáží I až III, které se snažily ověřovat navrhované postupy a analytické metody v praxi na dostupných datových zdrojích běžně zpracovávaných Českou školní inspekcí, ať už se jednalo o datové zdroje interní (data získávaná procesem Komplexní inspekční činnosti, elektronickým zjišťováním výsledků či tematickými šetřeními), či externí (zejména mezinárodní šetření gramotností PISA, TIMSS, PIRLS či TALIS, ale i zdroje MŠMT či ČSÚ), nebo zdroje tato šetření propojující. Pilotáže I a II se společně věnovaly problematice sběru dat (s odkazem na nutnost maximálního možného využití pravděpodobnostního, resp. náhodného výběru případů), kontrole a přípravě datových souborů pro využití jak v hlubších analýzách, tak v primárních deskriptivních analýzách a sekundárních analýzách národních datových zdrojů. Pilotáž II kladla důraz zejména na adekvátní prezentaci, vizualizaci a interpretaci analytických výstupů v souladu s posledními akademickými poznatky v této oblasti. Pilotáž III podrobně mapovala a analyzovala dosavadní metody zjišťování a analytické postupy využívané Českou školní inspekcí a na základě těchto poznatků navrhovala řešení respektující reálně existující a těžko odstranitelná omezení dosavadních postupů a navrhovala postupy nové.

V rámci dosavadní inspekční činnosti konané v průběhu trvání projektu KSH byla některá doporučení z Pilotáže I až III úspěšně implementována, zejména doporučení ohledně vizualizace dat a komunikace analýz veřejnosti.

Text metodiky je členěn do čtyř hlavních bloků – stávající sběr dat a možnosti propojování, analýza dat, vizualizace dat a problematika imputace chybějících hodnot.

Nedílnou součástí metodiky je paralelní dokument *Sada typizovaných analytických nástrojů*, který problematiky zde řešené rozšiřuje o praktické hledisko, základní návody a skripty pro realizaci analýz a datových vizualizací. V tomto dokumentu jsou jednotlivé analytické postupy podrobně rozepsány a je uveden návod ve dvou běžně používaných statistických programech, a to IBM SPSS a RStudio. V tomto textu v příslušných sekcích proto odkazujeme na praktickou implementaci v paralelním dokumentu.

2

Stávající sběr dat a možnosti propojování

2 STÁVAJÍCÍ SBĚR DAT A MOŽNOSTI PROPOJOVÁNÍ

2.1 Interní šetření

2.1.1 Komplexní inspekční činnost

V rámci této podkapitoly bude představen typ akce tzv. komplexní inspekční činnost, která v současné praxi ČŠI představuje dominantní a velmi důležitou aktivitu. Ta je zdrojem mnohých zjištění a poznatků o stavu české vzdělávací soustavy, které byly využity při výstupech KA5, jmenovitě sekundárních analýz PISA 2015, TIMSS 2015, PIRLS 2016, TALIS 2018 a PISA 2016, dále pak analýz krajských vzdělávacích soustav v rámci výstupu č. 169. Komplexní inspekční činnost (dále jako „KIČ“) představuje jeden z typů inspekční činnosti. *„Komplexní inspekční činností se rozumí inspekční činnost, jejíž hlavní součástí je inspekční hodnocení podmínek, průběhu a výsledků vzdělávání a poskytovaných služeb. Při komplexní inspekční činnosti se realizuje také kontrola dodržování vybraných ustanovení* [školského] *zákona (…) inspekční tým* [hodnotí] *zejména efektivitu opatření realizovaných vedením školy k dosažení pozitivních kvalitativních změn v podmínkách (včetně finančních), průběhu a výsledcích vzdělávání poskytovaného školou. V případě zjištěných nedostatků a slabých stránek školy identifikuje jejich příčiny. Inspekční tým na základě zjištění učiněných v průběhu inspekční činnosti vyhodnotí silné a slabé stránky školy, specifikuje případná rizika a spolu s vedením školy vymezí příležitosti pro další kvalitativní růst školy.“* (ČŠI 2019b: 3–4).

Komplexní inspekční činnost v sobě kombinuje mnoho různých postupů a metod zjišťování, např. pozorování, dotazníkové šetření, hospitace aj.). Metodika sběru dat vychází z definice KIČ (ČŠI 2019b) a souvisí s § 174 školského zákona.

Sběr dat se v kontextu KIČ váže především na práci krajských inspektorátů, inspekčních týmů a v posledku tedy i na práci jednotlivých školních inspektorů. V tomto směru lze pouze obecně konstatovat, že je nutné, aby práce probíhala co nejkvalitněji ve všech úrovních. Rozhodně je vhodné dbát na standardizaci užívaných metod zjišťování, a hlavně na posuzování, které jako konečný výstup generují inspekční hodnocení s minimalizací chyby měření (ve smyslu rozdílu mezi naměřenou a skutečnou hodnotou). Výsledky inspekční činnosti by měly být maximálně validní i reliabilní, což představuje nutný předpoklad pro následnou datovou analýzu. Z praxe ČŠI je zřejmé, že na těchto bodech kontinuálně pracuje a jednotlivé prvky zdokonaluje. Zde se nabízí propojení dat vycházejících ze subjektivních hodnocení inspektorů s objektivně naměřenými daty či daty z jiných šetření na úrovni dané školy (např. korelace *výsledků Kritéria hodnocení podmínek, průběhu a výsledků vzdělávání na školní rok* s objektivními indikátory školy po propojení například s šetřením TALIS 2018).

V rámci KIČ je administrováno velké množství různých formulářů, které využívají odlišné způsoby zjišťování. Takto získaná data mají relativně složitou víceúrovňovou strukturu. Zpravidla se jedná o pozorování na žákovské, učitelské a školské úrovni. Dále je možné, po případné agregaci, pracovat také s úrovní hospitovaných hodin. Při následné analýze je samozřejmě zapotřebí mít na paměti omezení dat vycházející jednak ze způsobu vytváření jednotlivých výběrových vzorků, jednak z případných omezení při přesnosti měření. Protože se analýzy dat z KIČ zpravidla objevují v podobě každoročně zveřejňované výroční zprávy mající vlastní požadavky a specifika na prezentovaný obsah.

Protože se zjištění (ať už hodnocení, kontrola, nebo jiné) z KIČ zaznamenávají do různých formulářů, je s tímto spojeno také pracnější vytváření výsledných datových souborů, které jsou určeny k následné analýze. Začátek procesu je nutné vnímat již před samotným vygenerováním datových souborů (seskupením jednotlivých vyplněných formulářů) z prostředí InspIS.

Zpravidla je nutná dostatečná znalost o datech a zkoumané problematice. S tím souvisí také znalost veškerých změn, které mohly v čase nastat. Na tomto místě vnímáme změnu v nejširším možném smyslu, kdy se nejedná např. jen o znalost úpravy jednotlivých zajišťovaných položek (změnu formulace otázky a odpovědí či jejich doplnění nebo odstranění), ale také o znalost možných úprav metodiky hodnotitelů (školních inspektorů) a dalších informací, které nejsou v datech přímo pozorovatelné. V této souvislosti by mohly být vytvořeny platforma či systém, v jejichž rámci by případné změny mohly být komunikovány a informace s nimi spojené poskytovány širšímu okruhu aktérů.

Jednotlivě vygenerované datové soubory je třeba náležitě upravit (viz podkapitoly 3.3.2 a 3.4). Pokud byl vytvořen/vyexportován více než jeden datový soubor, je nutné kroky související s přípravou dat replikovat pro každý takový datový soubor. V případech, které to dovolují, je však možné nejprve soubory spojit a pak provádět např. čištění a rekódování dohromady, což může tuto fázi práce urychlit. Pro další práci je klíčová znalost úrovně měření, kdy mohou některé soubory umožňovat samostatnou analýzu (jsou od počátku rozděleny, např. formulář *Kritéria pro hodnocení* je vytvářen odděleně pro GV a SOV, ale např. formuláře *Podmínky, průběh, výsledky SV* či *Hospitační záznam SV* se objevují v mutaci „SV”, tedy za stupeň středního vzdělávání zahrnující jak gymnaziální, tak i střední odborné vzdělávání). Je však potřeba si tyto různé nuance uvědomovat a při spojování náležitě zohlednit (nejčastěji je zde řešeno pomocí výběru relevantních případů či proměnných vztahujících se k naší oblasti zájmu).

Velmi důležitý krok představuje při spojování dat náš výběr způsobu agregace hodnot z nižší jednotky (např. z úrovně navštívených hodin, úrovně dotazovaných žáků či učitelů) na úroveň vyšší (např. na úroveň resortního identifikátoru REDIZO či na úroveň součásti školy, např. základní škola apod.). Nejčastěji volená technika bývá prosté aritmetické zprůměrování hodnot těchto jednotek na nižší úrovni, nabízí se však i jiné způsoby, např. vypočtení mediánu (který není citlivý na extrémní hodnoty) nebo procentuální přepočet dané charakteristiky (např. nás může zajímat počet navštívených hodin, ve kterém hospitující inspektor pozoroval, resp. nepozoroval nějaký jev). Zde je nutné zdůraznit, že neexistuje jediná všeobecně preferovaná cesta a vše je nutné uzpůsobit našemu záměru a další práci s daty, ať už s ohledem na zodpovězení případné výzkumné otázky, ověření statistické hypotézy, nebo „jen“ vytvoření přehledu o daném jevu.

Obecně a zejména při spojování dat je velmi důležitá znalost způsobu vytvoření vzorku dat, se kterým pracujeme (šířeji označován i jako *data generating process*). Což je informace, kterou zpravidla nelze vyčíst ze samotných datových souborů a je nutná její externí znalost. V této souvislosti si klademe řadu otázek typu: Pracujeme s celopopulačními daty? Pokud ne, s jakou množinou dat pracujeme? Byl tento vzorek vybrán za pomoci pravděpodobnostního, nebo náhodného výběru? Jaká byla míra návratnosti? Pokud ne, jak byl vzorek vytvořen? Jsou naše data reprezentativní pro zkoumanou populaci? V jaké charakteristice ano, v jaké charakteristice ne? Do jaké míry tento problém znemožňuje zobecňovat zjištění na širší populaci? V jakém časovém období byla data sbírána? Apod. Tato všechna kritéria musí být vždy zohledněna.

2.1.2 Národní zjišťování vzdělávacích výsledků

Podstatnou část výpovědi o vzdělávání dětí, žáků a studentů, ale také informace o činnosti škol, sbírané pro účel hodnocení efektivnosti vzdělávací soustavy, tvoří sběr a následná analýza dat vzdělávacích výsledků. V kontextu činnosti ČŠI je tento cíl naplňován zejména prostřednictvím národních zjišťování vzdělávacích výsledků (dále jako národní testování) a mezinárodních šetření, přičemž se tato podkapitola soustředí na první z nich.

Národní testování v sobě zahrnuje tři různé podtypy činností: (a) zjišťování výsledků vzdělávání žáků (dále jako testování) 5. a 9. ročníků ZŠ a odpovídajících ročníků víceletých gymnázií, které se opakuje zpravidla každé čtyři roky a soustředí se na zjišťování znalostí žáků v nejrůznějších oblastech, přičemž je nejpočetnější skupinou žáků řešen test z matematiky, českého jazyka a anglického jazyka; (b) testování vybraných gramotností (např. čtenářská, matematická či informační gramotnost), kterého se mohou účastnit žáci základních i různých typů středních škol; a (c) další tematické testování vybrané na základě schválených specifických úkolů ČŠI.

Důležité doplnění samotného zjišťování výsledků představuje zadávání doprovodných dotazníků. Pomocí dotazníků získáváme celou řadu doplňujících informací o žácích a prostředí škol, což umožňuje provádět důkladnější analýzy např. faktorů spjatých s výsledky žáků. V tomto smyslu je vhodné získávat potřebné informace od všech relevantních aktérů. V mezinárodním kontextu zjišťování výsledků vzdělávání (např. šetření PISA, PIRLS či TIMSS) je obvyklá víceúrovňová dotazníková struktura. Typicky na školní úrovni (ředitel), učitelské úrovni (např. učitel v šetření ověřovaného předmětu či vybraní učitelé) a žákovské úrovni (především samostatní žáci, v šetřeních PIRLS i TIMSS byl pak z důvodu nízkého věku testovaných žáků zadáván i dotazník rodičům).

V případě testování ČŠI je již standardní praxí administrace žákovských a učitelských dotazníků, které jsou propojitelné s konkrétní třídou, kterou učitelé vyučují. Nabízí se však také zařadit dotazníky pro ředitele škol, což by případné analýzy mohlo obohatit o hlubší znalost školního kontextu. Ředitelé (a šířeji vedení školy) mohou poskytnout informace, které od žáků či učitelů není možné získat. Data získaná na základě výpovědí ředitelů mohou přidat další informace o školách, ve kterých se žáci učí. Nabízí se samozřejmě využití dat z KIČ, nicméně v důsledku působení inspekčního cyklu a dalších faktorů nemusí být napárování vždy umožněno (viz v podkapitole níže). Důležitou součástí práce s daty z dotazníků je možnost jejich vzájemného propojování (a sice výsledků vzdělávání žáků s daty ze všech zadávaných dotazníků).

Možnost zadávání dotazníkových šetření a možnost vzájemného propojování z nich získaných dat a výsledků žáků otevírá prostor pro získání celé řady informací z různých oblastí a zkoumání souvislostí mezi nimi. Důkladná pozornost věnovaná tvorbě jednotlivých otázek tak může mít značný přínos s ohledem na množství a kvalitu dat, která můžeme analyzovat a na základě nichž lze formulovat doporučení směřovaná ke zkvalitňování výuky českých žáků. V tomto ohledu se tak jako přínosné jeví zvážení zadávání dostatečného množství relevantních otázek z nejrůznějších oblastí. Předně můžeme s ohledem na význam socioekonomického statusu v českém prostředí za stěžejní označit zahrnutí otázek měřících socioekonomický status především u jednotlivých žáků. Je žádoucí, aby byl individuální SES zjišťován, pokud je to možné, v co největším počtu šetření. Mimořádná pozornost by v rámci tvorby dotazníkových položek měla být věnována rovněž vhodné formulaci otázek a nabízených odpovědí.

V souvislosti s tím lze vyzdvihnout výhody případné standardizace dotazníkových položek. Pokud by otázkám a odpovědím byl přiřazen identifikátor a jednotlivé otázky či sady otázek byly zahrnovány do šetření opakovaně, bylo by možné odpovědi porovnávat a sledovat vývoj v čase. Zvážit standardizaci otázek má význam hlavně tehdy, kdy zachycují klíčové jevy, které může být důležité porovnávat napříč šetřeními. Opakující se baterie otázek s jednotnými

identifikátory jsou za účelem porovnatelnosti běžně využívány v mnoha šetřeních včetně mezinárodních šetření výsledků vzdělávání žáků. Souběžně s tím se doporučuje vytvořit katalog otázek (obsahující definici konceptů, otázky měřící koncept a konkrétní položky), který usnadní orientaci a samotnou práci s daty.

Za účelem propojení je obecně nezbytné, aby byly u jednotlivých respondentů přítomny identifikační údaje. Konkrétní postup spojování potom závisí na tom, jaký datový soubor je v dané chvíli žádoucí. V případě, kdy jsou vysvětlovány výsledky žáků, bývá výchozím souborem soubor s výsledky žáků a případně dalšími proměnnými získanými v rámci jejich dotazování. Ve snaze o rozšíření množství informací je na tento soubor následně možné připojovat další data získaná v rámci dotazování učitelů či ředitelů škol, do nichž žáci dochází. V obou případech je postupováno spojováním datových souborů skrze jednotné identifikační údaje o učitelích a školách. Případem výsledného datového souboru je žák, k němuž jsou postupně připojovány údaje o jeho učiteli a/nebo škole, do které dochází. V případě, kdy je více žáků vyučováno jedním učitelem, stejně jako u proměnných na úrovni školy, dochází k připojení těchto údajů ke všem těmto žákům. Jinak řečeno to znamená, že bude mít každý žák jednoho učitele či z jedné školy budou připojeny totožné informace o daném učiteli a škole.

Je však na místě poznamenat, že lze datové soubory spojovat libovolnými způsoby. Výchozím souborem teoreticky může být jakýkoliv soubor, záleží na účelu zamýšlené analýzy. Samozřejmě je vždy potřeba zvážit způsob výběru případů a možné limity zobecňování zjištění získaných z takového datového souboru.

Podobně se nabízí možnost spojovat data ze zjišťování výsledků nejen s daty z doprovodného dotazníkového šetření, ale i s jakýmikoliv dalšími datovými zdroji získanými v rámci odlišných šetření nebo činností (a různě mezi sebou). Podmínkou takového spojování je průnik mezi danými šetřeními na odpovídající úrovni a přítomnost shodného nebo vzájemně převoditelného identifikátoru. Postup je obdobný, jako když jsou příslušné případy datových souborů (na libovolné úrovni) postupně spojovány do jediného souboru skrze identifikátor. Konkrétní postup se odvíjí zejména od zamýšleného účelu takového spojování.

Pokud spojujeme datové soubory na různých úrovních (žák, třída, škola), výsledný datový soubor má hierarchickou strukturu. To má dopady na volbu vhodných postupů analýzy, kdy je zejména potřebné užívat odpovídající techniky, které takovou strukturu zohledňují (např. hierarchické regresní modely využívané v sekundárních analýzách mezinárodních šetření či analýzách národního testování gramotností).

Propojování dat a úprava datasetu, potřebné syntaxe, skripty a návody jsou v kapitole Úprava dat Sady typizovaných analytických nástrojů.

2.2 Externí (mezinárodní) šetření

Česká školní inspekce shromažďuje data sbíraná jak v rámci mezinárodních šetření, tak v rámci národní inspekční činnosti. Další datový zdroj, který nám může poskytnout cenné informace o stavu vzdělávací soustavy, představují právě mezinárodní výzkumy výsledků vzdělávání. Mezinárodní šetření přináší nezastupitelný externí pohled na vzdělávací systém České republiky.

Šetření realizovaná organizacemi OECD a IEA se mezi sebou liší např. z pohledu testované populace žáků (zaměřují se na odlišné kohorty žáků zejména z hlediska věku), metodou sestavení výběrového vzorku (např. výběr celých tříd nebo náhodný výběr žáků napříč třídami), v měřicích nástrojích, resp. používaných dotazníkových položkách a také testovanou oblastí. Všechny tyto rozdíly musíme při sledování výstupu z takových dat zohlednit.

Obecně je mezinárodním testováním společné, že poskytují maximálně objektivní externí hodnocení, a to na úrovni celé vzdělávací soustavy. Jako hlavní nástroj využívají standardizované testy a výběrové vzorky. Tyto jsou sestaveny způsobem, který dovoluje zobecnění nalezených zjištění na zkoumanou populaci (často též dovolují citlivější dělení, např. v šetření PISA dle kraje či dle druhu/typu školy). Protože se šetření realizují v pravidelných cyklech, umožňují nám vytváření časových řad a sledování trendů.

Rozsah problematiky mezinárodního testování byl pro účely této podkapitoly zúžen na dvojici šetření. První z nich představuje šetření PISA (2015) realizované organizací OECD, druhým je šetření PIRLS (2016) realizované IEA. Obě šetření jsou popisována souběžně, pakliže se mezi nimi objeví rozdíl, který by znamenal nutný zásah do způsobu strategie analýzy, je na něj upozorněno. Zároveň text následuje logiku předchozích podkapitol, resp. akcent je kladen především na strukturu datových souborů a jejich možného využití k prezentaci vzdělávacích výsledků.

Jednotlivé datové soubory jsou sesbírány z rozličných zdrojů, proto je nejprve potřeba soubory sloučit. Zpravidla považujeme soubor na úrovni žáka jako výchozí soubor, na který pak podle unikátního identifikátoru napojujeme soubory obsahující data z učitelských, resp. školních (ředitelských) dotazníků. V důsledku sestavení výběrového vzorku v šetření PISA je potřeba s takto připojeným učitelským dotazníkem pracovat na úrovni školy (jedná se o tzv. kontextuální proměnnou). Není totiž zajištěna přímá propojitelnost mezi konkrétním žákem a učitelem. Výsledkem je datový soubor s několika-úrovňovou strukturou.

Zdánlivě triviální technický úkon má však podstatný důsledek pro následné datové analýzy a interpretaci dat. V případě analýz takového souboru totiž nalezená zjištění z vyšší úrovně interpretujeme stále jako charakteristiku náležící k žákovi, resp. ke kohortě testovaných žáků. Pro příklad tohoto jevu použijeme dotazníkovou položku ze školního dotazníku ze šetření PISA 2015 pro ČR: *Do jaké míry ohrožuje záškoláctví žáků výuku?* Možnými odpověďmi byly: *značně*, *do určité míry*, *velmi málo* a *vůbec ne*. Podíváme-li se na zastoupení odpovědi *velmi málo* (51 %), nemůžeme tento výsledek vysvětlovat tak, že 51 % ředitelů na otázku takto odpovědělo. Protože pracujeme s datovým souborem na úrovni žáků, čteme zjištění následovně: *„Pro 51 % 15letých žáků v České republice platí, že navštěvují školu, jejíž ředitel deklaruje, že záškoláctví ohrožuje výuku žáků pouze velmi málo.“* Tento poznatek samozřejmě nevylučuje možnost, že školu, pro kterou ředitel takovou odpověď uvedl (a je tedy pro všechny žáky školy stejná), mohou navštěvovat žáci, pro které je záškoláctví značným problémem.

Stejně opatrně je potřeba interpretovat agregované testové výsledky za celorepublikovou či krajskou úroveň. V případě České republiky u mnohých mezinárodních šetření pozorujeme, že se testové výsledky značně odlišují, pokud na ně pohlížíme z krajské úrovně, či podle druhu školy. Obecně identifikujeme jednotky, které stabilně dosahují vysokých testových výsledků (např. hl. m. Praha a víceletá gymnázia) a nízkých testových výsledků (typicky Ústecký či Karlovarský kraj, resp. střední odborná učiliště). Stejný rozklad bychom však mohli vést z krajské úrovně na nižší úroveň. Neznamená to však, že každému žákovi náleží charakteristika z vyšší úrovně, resp. každý testovaný žák v Praze nedosahuje ve srovnání s jakýmkoliv žákem jiného kraje vyššího výsledku, totéž platí i při porovnání druhů škol. V kontextu prezentování a vizualizace dosahovaných výsledků je proto vhodné tuto skutečnost zohlednit. Nabízí se například mapová vizualizace či sloupcové grafy, které zobrazují výsledky pro každý kraj.

Datová základna mezinárodních šetření je velmi bohatá a vedle dat o vzdělávacích výsledcích nabízí také informace z dotazníků určených žákům, rodičům, učitelům či ředitelům. V tomto textu se však omezíme na hlavní závisle proměnnou, kterou představují testové výsledky. Ty jsou v mezinárodních šetřeních vyjádřeny dvojím způsobem. Za prvé s pomocí průměrného skóre. Druhým způsobem je pohled na podíly žáků dosahujících jednotlivých gramotnostních/dovednostních úrovní. V obou případech jsou základem výpočtů skupiny proměnných, které v datových souborech vyjadřují výsledek žáků. V mezinárodních šetřeních se používá metoda tzv. plausibilních hodnot (metoda vícenásobné imputace). Téma výpočtu plausibilních hodnot je mnohem obsáhlejší a pro detailnější informace lze odkázat na technické manuály k příslušným šetřením (např. Technical report, PISA 2015).

Pro první způsob prezentace výsledků, průměrné skóre, je výsledná stupnice stanovena zpravidla na průměrné hodnotě 500 se směrodatnou odchylkou 100. Podle typu měřítka se jedná o spojitou proměnnou, která bývá nejčastěji prezentována pro úroveň celé testované populace. Můžeme ji reportovat i na nižší úrovni, např. krajské, či podle druhu školy (vždy záleží na parametrech výběrového vzorku a na tom, pro jakou nižší úroveň je vzorek ještě reprezentativní a dovoluje zjištění zobecňovat). Protože se jedná o standardizovanou hodnotu, její interpretace je odvislá od hodnot ostatních případů v souboru. Z toho vyplývá, že interpretace izolované hodnoty nám mnoho neřekne, pokud ji neuvedeme v kontextu ostatních, ať už se jedná o výsledky ostatních zemí, nebo o časové srovnání s výsledky v minulých cyklech.

Testové výsledky můžeme vyjádřit i druhým způsobem, na gramotnostních/dovednostních úrovních, tedy na ordinální škále. V obou šetřeních je ordinální škála reprezentována procentuálním podílem žáků, kteří dosáhli určité úrovně. Namísto plausibilních hodnot tak pracujeme s tzv. gramotnostní/dovednostní úrovní (v originále „proficiency level“). Pro čtenáře může být tento způsob prezentace vhodnější v tom ohledu, že každá jednotlivá úroveň obsahuje vlastní kvalitativní definici. Popisuje, jaké dovednosti a vědomosti mají žáci, kteří dosáhli určitých výsledků a byli do těchto úrovní zařazeni (šetření PISA i PIRLS se liší v počtu úrovní u jednotlivých gramotností, jejich princip je však totožný). Pro ilustraci uvádíme pouze dvě gramotnostní úrovně (nejvyšší 6 a nejnižší 1b) z šetření PISA 2015, které byly definovány pro testovanou přírodovědnou gramotnost:

TABULKA 1 | Příklady gramotnostních úrovní v mezinárodních šetřeních a jejich definice – PISA 2015, přírodovědná gramotnost

Úroveň	Popis
6	Žáci používají obsahovou, procedurální a epistemickou znalost k důslednému vysvětlování, vyhodnocování a navrhování vědeckých výzkumů. Interpretují údaje rozmanitých složitých životních situací vyžadujících vysokou úroveň poznání. Umí vyvozovat odpovídající závěry z řady různých složitých zdrojů dat v rozmanitých souvislostech a podat vysvětlení vícenásobných vzájemných vztahů. Umí důsledně rozlišovat vědecké a nevědecké otázky, vysvětlovat účely výzkumu a ovlivňovat významné proměnné veličiny v každém vědeckém pokusu nebo v návrhu pokusu. Umí převádět všechna datová zobrazování, vysvětlovat složitá data a prokazují schopnost správně posoudit spolehlivost a přesnost veškerých vědeckých tvrzení. Žáci důsledně prokazují pokročilé vědecké myšlení a uvažování vyžadující použití modelů a abstraktních myšlenek a mají schopnost používat takový způsob uvažování v neznámých a složitých situacích. Umí hledat důkazy k posouzení a vyhodnocení výkladů, modelů a vysvětlování dat a navrhovat pokusy na osobní, místní/národní a globální úrovni.
1b	Žáci mají jenom minimální obsahové, procedurální a epistemické znalosti k vysvětlování, vyhodnocování a navrhování vědeckých výzkumů. Interpretují údaje pouze několika známých životních situací vyžadujících nízkou úroveň poznání. Jsou schopni určit přímé vzory v jednoduchých zdrojích dat v několika známých souvislostech a umí nabídnout pokusy o popis jednoduchých vzájemných vztahů. Umí určit nezávislou proměnnou veličinu v některých vědeckých pokusech nebo v jednoduchých návrzích. Pokouší se převádět a popisovat jednoduchá data a použít je přímo v několika známých situacích.

2.3 Stručná diskuze rozdílů

Jak může být už z prvního pohledu patrné, všechny typy zmíněných šetření se mezi sebou více či méně odlišují. V principu, pohledem z úrovní sběru a propojování dat, jsou si mezinárodní šetření zjišťující vzdělávací výsledky žáků (2.2) podobnější s národním zjišťováním vzdělávacích výsledků (dále pro snadnější odlišení jako *národní testování*, 2.1.2) spíše než s KIČ (2.1.1). Nejprve budou diskutovány vybrané možnosti a limity ve sběru i propojování mezinárodních a národních šetření, následně budou obě diskutována v souvislosti s možnostmi KIČ.

V rámci mezinárodních šetření jsou shodně jako v národním testování administrovány testy, které žáci vyplňují a jejichž výsledkem je zpravidla nějaký ukazatel úspěšnosti. Na individuální úrovni žáků si jej můžeme představit např. jako ukazatel procentuální úspěšnosti v řešení vybraných úloh (národní testování) či jako sadu imputovaných plausibilních hodnot (mezinárodní šetření).

Vedle testů jsou taktéž administrovány dotazníky, které sbírají kontextuální informace o žácích. Jejich rozsah i kvalita zpracování se však mezi oběma typy značně liší. V mezinárodních šetřeních lze vypozorovat často řadu dotazníků (pro žáky, jejich učitele nebo vybrané učitele v dané škole, dotazníky pro ředitele školy aj.). Cílem dotazníků v rámci mezinárodních šetření je zjistit nejrůznější kontextuální informace (podmínky ke vzdělávání, rodinné zázemí, nálady, motivace, postoje a další) o různých aktérech ve vzdělávání. Výsledky těchto dotazníků pak mohou být velice efektivně propojeny právě např. s výsledky žáků a v rámci následných analýz dovolují popisný i exploratorní přístup, spočívající např. v reportování četností, procentuálního zastoupení či asociací vybraných znaků, resp. vztahů mezi nimi, či dovolují konfirmovat v úvodu analýzy stanovené hypotézy. Národní testování pak také nabízí dotazníky (zpravidla testovaným žákům a jejich učitelům). Ty jsou však ve srovnání s dotazníky mezinárodních šetření obsahově velmi redukované. Zde je patrný vliv nízké časové dotace, který je v rámci národního testování věnován vyplňování těchto dotazníků. V důsledku tohoto omezení (časového, resp. obsahového) nemůže být plně naplněn potenciál pro výzkum a detailnější pohled na četné relevantní aspekty z těchto testování.

V otázce propojování dat jsme schopni standardně spojit data na úrovni škol (prostřednictvím identifikátoru REDIZO). Nižší jednotky agregace, tedy např. učitelé a testovaní žáci, jsou anonymizovány, a proto je propojování na individuální úrovni znemožněno. Data však vzájemně dovolují propojení na individuální úrovni dříve agregovanou proměnnou. Zpravidla pak výsledky interpretujeme na úrovni žáků sdílejících určitou charakteristiku[1]. Zde je však velmi důležité si uvědomit, že šetření (jak mezinárodní, tak i národní) probíhají v různých časových obdobích, ale také jsou do něj zahrnovány odlišné školy. Pokud se tedy nejedná o plošné národní šetření, je možnost vzájemného propojování velmi omezena. Je tomu tak především z důvodu velmi malého průniku mezi množinou škol, které byly zahrnuty do národního testování, a množinou škol zahrnutých do mezinárodního šetření. Na druhou stranu pokud se škola účastní vícera testování či šetření, hrozí riziko nižší reliability šetření a vyšší míra chybějících odpovědí na různé sledované položky. Snaha o průniky domácích šetření a mezinárodních šetření tak musí často nabývat kompromisu zvažujícího všechny faktory a omezení.

KIČ se z podstaty své realizace dvěma předchozím typům šetření zcela vymyká. Při sběru dat je hlavní výhodou možnost skutečného nahlédnutí do navštívených škol. Nemusíme tedy „pouze“ spoléhat na data, která jsme získali z vyplněných a odevzdaných dotazníků, přičemž výsledky některých mohou být a pravděpodobně jsou zatíženy

[1] Formulace může být typově následující: Žáci, jejichž učitelé v průměru odpovídají určitým způsobem, dosahují nižších/vyšších výsledků apod.

náhodnou i systematickou chybou, ale školní a kontrolní pracovníci provádějící prezenční inspekční činnost mohou nahlédnout i za tato data. Cenné poznatky těchto aktivit je samozřejmě nutné vnímat i v kontextu náročnosti této činnosti obecně a odpovídajícím způsobem do našich úvah promítnout i možnost odchýlení zaznamenaných/pozorovaných a následně analyzovaných dat ve formulářích od skutečného stavu v dané škole či školském zařízení. Jinak řečeno, navzdory profesionální práci budou do výsledku vstupovat různé vlivy mající náhodný i systematický charakter včetně např. samotného pozorování (tzv. efekt pozorovatele hovoří o tom, že pozorované subjekty mění své chování právě v důsledku tohoto pozorování), které mohou více či méně pohled na školu zkreslovat.

S ohledem na spojování dat z KIČ pak musíme zohledňovat období sběru, které je spojeno zejména s posouzením toho, nakolik se vychyluje návštěva konkrétní školy v rámci KIČ od realizace testu. Zde však neexistuje univerzálně platné měřítko. Je potřeba zohledňovat kontext i cíle, kterých chceme v rámci následné analýzy dosáhnout. Obecně je však vhodné zpozornit v případech, kdy byla KIČ realizována několik let od proběhlého šetření (před i po). Zde je možné se domnívat, že poznatky získané z té doby mohly projít určitou reflexí. Například při velmi špatném kriteriálním hodnocení mohlo vedení školy změnit dle doporučení školních inspektorů způsob vedení školy, ředitel mohl být za tu dobu vyměněn a nové vedení započít řadu změn atd. KIČ navíc neprobíhá v úzce vymezeném časovém období, ale v rámci celého školního roku, jistá variabilita může být podmíněna i návštěvou v konkrétním období. V kontextu vzdělávacího výzkumu se nicméně setkáváme i se situací dlouhé „trvanlivosti“ určitých aspektů, tradičně se hovoří např. o postojích či hodnotovém ukotvení pedagogického sboru. Je tedy potřeba postupovat velmi individuálně a všechny tyto parametry při propojování vyhodnocovat a zvážit.

Kromě období sběru je také nutné vyhodnotit velikost průniku vzorku navštívených a vzorku testovaných škol. I v tomto případě bude výsledné posouzení velmi individuální, avšak obecně platí, že při malém průniku ztrácíme možnost poznatky zobecňovat a vypovídáme v podstatě jen o dostupných datech, která ovšem mohou být velmi zkreslena vůči základní populaci ve smyslu nadreprezentace nebo podreprezentace určité charakteristiky, případně též nízkým počtem případů (viz výše). Spíše než možnost propojování dat směrem z KIČ na data z mezinárodního či národního šetření se nabízí propojení opačné, kdy namísto snahy o odhad populačních parametrů, testování hypotéz či kvantifikaci vybraných aspektů doplňujeme bohatou kvalitativní informační základnu (KIČ) o nové poznatky.

3

Analýza dat

3 ANALÝZA DAT

3.1 Kvantitativní vs. kvalitativní přístup k výzkumu kauzality

3.1.1 Kvantitativní přístup

Kvantitativní výzkum se v oblasti kauzální inference zpravidla zaměřuje na identifikace a odhadování důsledků příčin (*effects-of-causes*) nezávisle proměnné na závisle proměnnou, ve srovnání s kvalitativním výzkumem, který vychází z přístupu odhadování příčin důsledků (*causes-of-effects*) mnohdy v jednotlivých případech. Cílem kvantitativního přístupu je odhad tzv. průměrného kauzálního efektu (případně nějaké z jeho variant) sledované nezávisle proměnné (která mnohdy představuje zaváděnou intervenci) na hodnoty závisle proměnné. Pro tyto účely v kvantitativním výzkumu často využíváme a analyzujeme velké množství případů, které jsou v optimální situaci z populace vybírány náhodně, resp. jsou hodnoty intervence náhodně přiřazovány (Goertz, Mahoney 2012).

V mnoha situacích v rámci kvantitativního výzkumu využíváme statistické procedury, s jejichž pomocí ověřujeme dopředu stanovené hypotézy předpokládající vzájemný vztah (či jeho absenci) mezi dvěma či více proměnnými a které mj. předpokládají dostatečně velké množství případů a kvantifikaci proměnných ještě před začátkem analýzy.

Typicky disponujeme buď daty za celou zkoumanou populaci, nebo výběrem z této populace. Pokud jsou při tvorbě výběru (vzorku) splněna kritéria náhodného výběru případů, kdy má každý případ (jedinec) dané populace předem známou pravděpodobnost, že bude do výběru zahrnut, a výběr je proveden metodou vycházející z těchto pravděpodobností, můžeme při potřebném množství případů soudit, že je tento výběrový vzorek pro danou populaci reprezentativní. Zjištění z takového vzorku lze pak, zpravidla při použití odpovídajících statistických technik, zobecňovat na celou populaci, kterou takový vzorek reprezentuje. V tomto kontextu pak hovoříme o tzv. externí validitě odhadu (o možnosti zobecňovat zjištění).

Kvantitativní techniky jsou využívány jak u experimentů, tak u observačních studií. Zatímco randomizované experimenty v důsledku náhodného rozřazení případů do skupin (jinak řečeno manipulací s hodnotami intervenující proměnné či přiřazováním odlišných tzv. *causal states*) v podstatě izolují vliv třetích proměnných, které by jinak mohly odhad ovlivnit a zkreslit, umožňují s použitím technik statistické inference odhadovat, při splnění dalších podmínek, nezkreslený (*unbiased*) průměrný kauzální efekt intervenující proměnné, který je platný v rámci skupiny zkoumaných případů (Cartwright, Deaton 2018), u observačních studií toto až na vzácné výjimky možné není.

Jestliže v rámci observační studie, a šířeji pak ve výsledných observačních datech, nad kterými neměl výzkumník přímou kontrolu (viz výše), pozorujeme souvislost mezi dvěma proměnnými (např. A a B), ještě to neimplikuje nalezení příčinného vztahu, tedy že A způsobuje B. Ač je souběžná změna (kovariace) v hodnotách obou proměnných jednou z nutných podmínek pro určení kauzality, za pozorovanou souvislostí mezi A a B může stát např. jiný faktor (jakákoliv třetí proměnná), který jsme buď nezahrnuli do analýzy, nebo ho ani neumíme měřit. Pro formulaci závěrů o příčinné souvislosti je tedy nutné, abychom vliv těchto „třetích proměnných“, které mohou do vztahu mezi A a B vstupovat, kontrolovali. S tím zároveň souvisí potřeba prokázat, že A způsobuje B, nikoliv naopak. Nutné je také prokázat logický časový soulad, tedy že A časově předchází B.

Jednu z nejpoužívanějších statistických technik, která umožňuje sledovat působení celé řady nezávisle/třetích proměnných na vysvětlovanou proměnnou, představuje regresní analýza (viz kapitola 3.5.4). Efekty jednotlivých proměnných v regresním modelu jsou odhadovány tak, že je zároveň kontrolováno působení dalších proměnných zahrnutých do modelu. Možnost kontroly celé řady proměnných najednou tak alespoň částečně snižuje riziko falešné korelace, přestože ho nedokáže eliminovat úplně.

3.1.2 Kvalitativní přístup

Přestože není kvalitativní přístup v činnostech ČŠI stěžejní a v mnoha případech by byl činností suplující spíše akademický výzkum, můžeme nalézt případy, kdy se jeho použití jeví jako vhodné. Při práci s kvantitativními šetřeními ČŠI pravidelně odhaluje různé specifické, těžko vysvětlitelné nebo přímo odchylné případy, pro které nejsou dostupná další data, která by pomohla identifikovat možné příčiny. Přijímání opatření k nápravě se v takovém případě děje do značné míry na základě zkušeností pracovníků ČŠI, MŠMT a škol jako takových. V případě identifikování takových případů a snaze o hlubší vysvětlení, v zájmu přijetí přesně cíleného opatření, se využití některých kvalitativních technik přímo nabízí.

Pro účely této metodiky chápeme kvalitativní přístupy jako doplňkové k analytické a evaluační činnosti ČŠI, které mohou přitom pomoci zkvalitnit proces přijímání některých opatření.

3.1.2.1 Kvalitativní výzkum zaměřený na variaci

Komparativní případová studie

Komparativní případová studie je nástrojem, který má za cíl do hloubky analyzovat několik případů. Cílem je určit klíčové nutné nebo dostačující podmínky (deterministické vztahy). Tím se liší oproti statistickým technikám, ve kterých se analyzují pravděpodobnostní vztahy a cílem je určit sílu efektu a míru zobecnitelnosti na základní soubor. Komparativní studie jsou založeny na logice Millových metod. Praktické využití může být v případě inspekční činnosti, kdy se cíleně vyberou školy na základě kvalitativních kritérií. Zejména ty, které si jsou v některých klíčových faktorech podobné (např. SES), ale výsledky ve vzdělávání se v těchto školách liší. Tímto způsobem se kvalitativní analýza může soustředit na jiné klíčové proměnné, které vysvětlí lepší či horší výsledky žáků v těchto školách, popřípadě vysvětlí sledované jevy v hospitační činnosti. Zde je třeba ale mít na paměti, že charakter studie neumožňuje výsledky zobecnit na populaci všech škol, ani na subpopulaci daných škol, které se svými charakteristikami blíží školám vybraným do kvalitativní analýzy.

QCA

Kapitola 10 v Sadě typizovaných analytických nástrojů

Kvalitativní komparativní analýza (*Qualitative comparative analysis* – QCA) (Ragin 1987) je založena na principech kvalitativní komparativní případové studie s tím rozdílem, že jejím cílem je analyzovat kombinace nutných a dostatečných příčin, které vedou k důsledku. QCA dokáže pracovat s vyšším počtem případů. V případě, že máme střední počet případů, vhodnou alternativou ke klasické regresi či korelaci je právě QCA. Obě metody mohou být kombinovány a jsou běžně využívány v analýze veřejných politik (Vis 2012). QCA ale oproti klasické regresi má výhodu v tom, že dokáže modelovat ekvifinalitu, tedy rozdílné kombinace odlišných příčin, a asymetrickou kauzalitu, což znamená, že přítomnost důsledku vyžaduje jiná vysvětlení než absence důsledku. Například pomocí této logiky můžeme zkoumat příčiny, které vedou k horšímu klimatu třídního sboru, ty mohou být zcela jiné než ty, které vysvětlují lepší klima třídy. Zatímco korelační metoda předpokládá symetrický vztah, QCA dokáže modelovat i asymetrie.

Asymetrické vztahy se dají modelovat i pomocí regresních modelů vložením interakčního efektu do rovnice regresního modelu. Tyto vztahy popisují například nedávno vydané Sekundární analýzy PISA 2018 a TALIS 2018. Problém je ale ten, že je problematické analyzovat vztah interakce mezi více než dvěma proměnnými. Rovněž je celá řada problémů při vložení více interakčních efektů do jednoho modelu (multikolinearita, pokles stupňů volnosti). QCA by tak mohla být vhodnou alternativou pro celou řadu analytických zpráv, které ale ne nutně vycházejí z kvantitativního paradigmatu sběru dat. QCA analýza by tak mohla být vhodným nástrojem pro analýzu těch šetření, která neprobíhají tak, že je zajištěna reprezentativita vzorku. I s ohledem na to, že celá řada proměnných je při inspekční činnosti měřena/zaznamenávána na nominální úrovni, jeví se jako vhodné do budoucna tento typ analytické metody zvážit.

SCHÉMA 1 | Rozdíl mezi QCA a regresní analýzou

Kvalitativní model	Kvantitativní model
$Y = (A*B*c)^{p1} + (A*C*D*E)^{p2}$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_{12} X_1 * X_2 + \varepsilon$
• Ekvifinalita je explicitně modelována. • * = AND • + = OR • K důsledku vede buď příčina **p1** nebo příčina **p2.** • Empiricky můžeme pozorovat případy, kdy jsou obě příčiny přítomné, nebo jen jedna z nich.	• Ekvifinalita je implicitní. • Na důsledek Y má vliv/efekt několik příčin X_1, X_2 či X_3. • Velikost efektu udává koeficient. • Ve statistickém modelování jde o změření čistého efektu proměnné X_1 očištěného o vliv dalších proměnných X_n. • Lze modelovat interakční efekt, ale nejedná se o vztah A*B jako v případě kvalitativní logiky.

Zdroj: Autoři

Například v Sekundární analýze PISA 2018 bylo zjištěno, že používání ICT ve výuce nemá jednoznačný přímý efekt na výsledky žáků v testu čtenářské gramotnosti. Efekt byl podmíněn klimatem třídy, způsobem využívání učitelem, přizpůsobením výuky učitelem atd. Všechny tyto faktory působí ve vzájemné kombinaci. Kombinace těchto příčin pak lépe vysvětlí efekty používání ICT ve výuce. QCA analýza by tak teoreticky mohla odhalit i ty vztahy, které v rámci standardních asymetrických vztahů zůstávají skryty (viz Vis 2012).

QCA může pracovat jak s kategorickými proměnnými (csQCA), tak se škálovými proměnnými, které se ale musí transformovat (fsQCA). Mluvíme tak o QCA s ostrými množinami v prvním případě a o QCA s mlhavými množinami v případě druhém. Využití QCA na úrovni žáka není běžně používáno, ale nabízí se možnost analyzovat úroveň školy, a to i po agregaci jak dat z národních šetření, tak i z mezinárodních šetření. QCA analýza byla například využita ve výzkumu soutěživosti školních vzdělávacích soustav (Choi, Lee 2012) či při studiu nerovností ve vzdělávání (Cooper, Glaesser 2011).

Praktická implementace QCA analýzy v programu R je v kapitole *Kvalitativní komparativní analýza (QCA)* v dokumentu *Sada typizovaných analytických nástrojů.*

3.1.2.2 Zaměření na kauzální mechanismy a explanační techniky

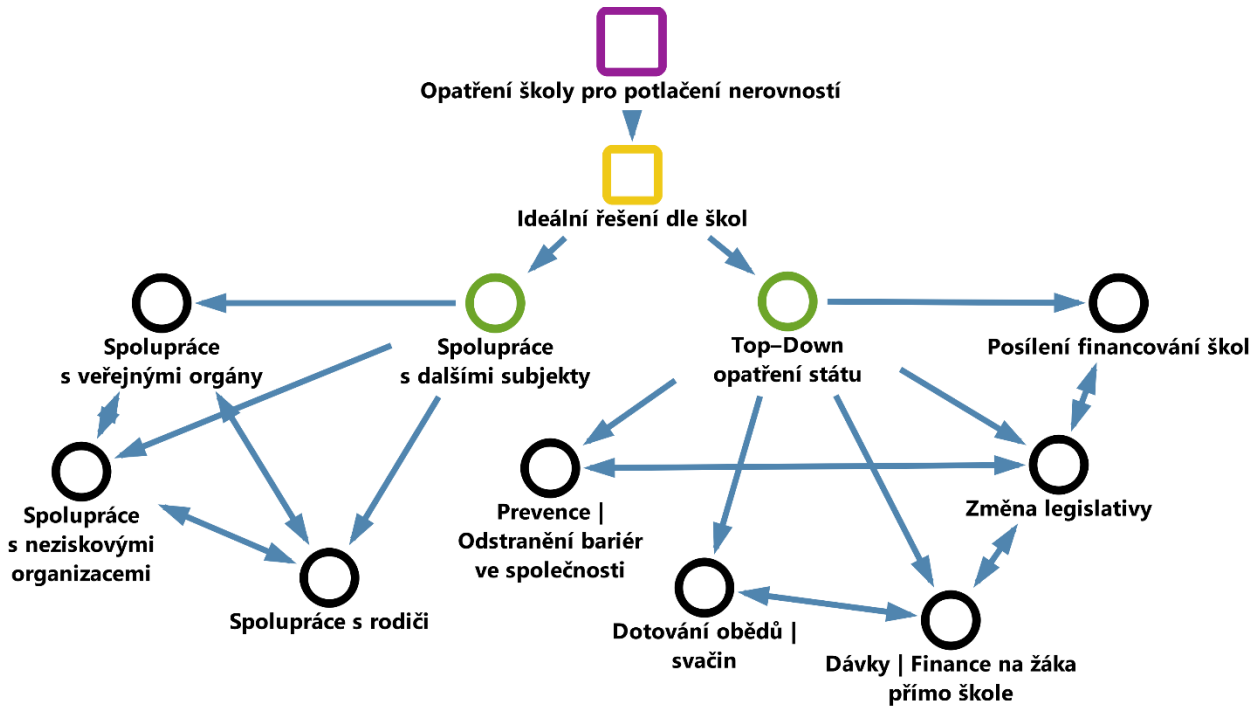
Rozhovory

Rozhovory v kvalitativním výzkumu obecně pomáhají odhalovat a pochopit svět pohledem respondenta. Pomáhají odhalovat faktory nezachytitelné v kvantitativních šetřeních, které ovšem mohou mít klíčový vliv na jednání daného respondenta, čímž mohou významně formovat výsledná doporučení ČŠI. Ve vztahu k inspekční činnosti se jako přínosné jeví rozhovory zejména s řediteli škol či dalšími aktéry v systému vzdělávání vybíranými pomocí kvantitativních technik. Tato forma tzv. vnořené analýzy (kdy jednou z variant takové analýzy je výběr případů z kvantitativních šetření pro kvalitativní rozhovory) pomáhá propojovat zobecňující zjištění z dotazníkových šetření (za předpokladu, že jsme si vědomi limitů daných vybraným vzorkem) s hlubším pochopením motivace aktérů k některým činnostem či rozhodnutím a nepřímo tak pomáhá pochopit očekávané kauzální procesy, které samy o sobě nemohou být zobecňovány. Pokud nejsme schopni zjištění z kvantitativních šetření, jež se pohybují v rovině korelací, podložit dostatečně relevantní literaturou, tvoří kvalitativní rozhovory alternativní nebo doplňující vrstvu pro posilování validity výsledných zjištění z inspekční činnosti. Rozhovory přitom nemusí evokovat akademický výzkum, ale mohou být prováděny v rámci standardně zavedených hospitačních aktivit.

Varianta zvolených rozhovorů bude vycházet z předmětu inspekční činnosti, ve většině případů ovšem postačí prosté hloubkové rozhovory s přesně vytipovanými aktéry, které se pokusí rozvést některá z témat pokrytých v rámci mezinárodních šetření gramotností či tematických šetření ČŠI, s cílem zvýšit validitu případných zjištění a dodat další vrstvu argumentace pro podporu reálného zavedení těchto doporučení.

Užití rozhovorů v inspekční činnosti již bylo v rámci projektu KSH pilotováno v problematice škol vyučujících znevýhodněné nebo naopak nadané žáky (viz interní dokumenty ČŠI *Znevýhodnění žáci na českých školách: Analýza rozhovorů* a *Systém práce s nadanými žáky na českých školách: Analýza rozhovorů*). Pracováno bylo s velmi omezeným vzorkem škol, který sám o sobě neumožňuje zobecňování, avšak jednotlivé případy byly vybírány pomocí regresního modelování, kdy byly přesně vytipovány školy, které dosahovaly výrazně horších nebo naopak výrazně lepších výsledků v testech PISA, TIMSS aj., než jakých by měly dosáhnout na základě predikce z tohoto modelu kontrolujícího proměnné spojené s nízkým SES, vyšším podílem žáků se sociálním znevýhodněním apod. Tento design umožnil zaměření rozhovorů na hlubší pochopení předpokládaných příčin, které stojí za úspěchem či naopak neúspěchem škol při práci se sociálně znevýhodněnými skupinami žáků.

SCHÉMA 2 | Jedna z možností prezentace otevřeného a částečného axiálního kódování konceptů z kvalitativních rozhovorů (výstup programu MaxQDA)



Zdroj: Znevýhodnění žáci na českých školách: Analýza rozhovorů (nepublikováno)

Musíme zdůraznit, že v kontextu inspekční činnosti ČŠI je naprosto nutné navázat podobné kvalitativní techniky na běžně probíhající kvantitativní šetření. Zatímco kvantitativní šetření pomůže odhalit šíři potenciálních problémů (v rovině korelací), kvalitativní techniky se zaměří na hlubší pochopení a eventuálně i komplexní kauzální mechanismy těchto problémů, pomohou přidat další rovinu argumentace a zvýší validitu výsledných zjištění. Pouhé izolované použití kvalitativních technik by pro reálný přínos k výsledným zjištěním potřebovalo přísné dodržování metodologických postupů, které je v praxi inspekční činnosti ČŠI mnohdy nerealizovatelné.

Focus groups

Jinou variantu rozhovorů představují již v práci zmíněné focus groups, které kromě hlubšího pochopení sledovaného problému poskytují také možnost sledování interakce mezi respondenty. Výhodou je též možnost cílení skupinových rozhovorů na různé úrovně vzdělávací soustavy a také zahrnutí úrovně žáků. Pokud bude cílem zjišťování např. práce konkrétního učitele, je možné zvolit homogenní skupinu žáků pro dotazování. Pokud bude naopak cílem zjišťování situace na dané škole a nepostačí individuální rozhovor s ředitelem školy, je možné zvolit heterogenní skupinu žáků napříč jednotlivými třídami a tím získat validnější obrázek o sledované problematice.

I v případě užití varianty focus groups je vhodné její zařazení do běžných aktivit v rámci hospitací a samozřejmě provázání s kvantitativními šetřeními, které by sloužily pro výběr případů a témat k hlubšímu zkoumání. Izolované použití techniky focus groups nepřinese v kontextu inspekční činnosti ČŠI požadovanou přidanou hodnotu kvalitativního přístupu k problematice.

Process-tracing

Třetí a nejkomplexnější techniku kvalitativního přístupu ke zjišťováním v rámci hodnocení výsledků vzdělávání, jež vychází z mechanistického přístupu, představuje technika process-tracing (dále jen PT) (Mazák 2017; Beach, Pedersen 2019). PT je technika vycházející přímo ze základních principů mechanistického přístupu, kdy příčina a důsledek sledovaného problému jsou známy (ev. předpokládány) a pozornost směřuje pouze ke kauzálnímu mechanismu (odtud mechanistický přístup), který příčinu a důsledek spojuje. Kauzální efekty (1) jsou standardně odhalovány v rámci kvantitativních šetření (jedná se např. o regresní koeficienty), pokud u těchto předpokládáme kauzalitu, podloženou např. literaturou. Zpravidla ovšem nedokážeme s jistotou říci, že zde kauzalita je skutečně přítomna, protože nevíme, zda mezi příčinou a následkem existuje kauzální mechanismus (2). Záleží pak na zvoleném přístupu, jakým chceme existenci kauzality podpořit.

SCHÉMA 3 | Zjednodušené ilustrativní schéma principu metody process-tracing

Zdroj: Autoři

Přestože se metoda PT může pro použití v rámci inspekční činnosti zdát jako příliš složitá a příliš evokující akademický výzkum, což je jistě pravda, v některých typech inspekční činnosti je velice vhodné její použití ne jako standardizované metody, ale flexibilní využití jednotlivých principů při dokládání některých problémů a tím zvýšení validity zjištění a doporučení ČŠI. Metoda PT, při použití pouze minimalistického designu kauzálního mechanismu (tj. jednotlivé kroky kauzálního mechanismu mohou být obecnějšího charakteru nebo nemusí být zcela doložitelné), umožňuje využití izolovaných dat sbíraných ČŠI v rámci různých typů inspekční činnosti, které jsou samy o sobě prakticky nepoužitelné pro odhalování složitých asociací. Taková data jsou ovšem dostačující pro dokládání jednotlivých kroků kauzálního mechanismu. Ve výsledku tak metoda PT může za využití stávajících datových zdrojů z běžné inspekční činnosti ČŠI a dostupné literatury dostatečně doložit předpokládanou kauzalitu a obejít tak problémy s reprezentativitou způsobené již při výběru vzorku do kvantitativních šetření ČŠI. Potenciál zvýšení validity zjištění a doporučení ČŠI z inspekční činnosti se významně zvýší jen při pouhé systematické práci s dosavadními datovými zdroji. Nezbytnými předpoklady pro použití metody PT při zpracování inspekčních dat je, obdobně jako v případě rozhovorů, propojení s kvantitativními šetřeními a také rezignace na metodologickou přesnost, a to ve prospěch efektivity zjištění.

3.2 Základní a výběrový soubor

Při realizaci kteréhokoliv šetření musíme vedle otázky *Co chceme zkoumat?* položit také otázku *Koho chceme zkoumat?* Právě druhou otázkou se bude zabývat stručný přehled nabídnutý v této podkapitole. Klíčové termíny zde představují základní a výběrový soubor. Základním souborem (též populace) obecně myslíme množinu prvků (někdy též jako jednotky či případy), výběrový soubor (též vzorek, subpopulace) pak tvoří jeho podmnožinu. Problematiku můžeme přiblížit na příkladu navštěvovaných škol, ve kterých školní inspektoři provádějí tzv. komplexní inspekční činnost (KIČ), časově období vymezíme na jeden školní rok.

Protože není možné během jednoho školního roku navštívit a provést KIČ ve všech školách a školských zařízeních (dále v textu budou školská zařízení zahrnována pod „školami“), jinak řečeno, neprovádí se cenzus neboli plošné šetření celé populace škol, je zřejmé, že je nutné zvolit nějakou z forem či typů výběrového šetření. Výběrová šetření z podstaty definice pracují pouze s podmnožinou základního souboru, tedy s výběrovým souborem či vzorkem. Tento výběrový soubor pak zahrnuje jednotky, které skutečně zkoumáme (například právě školy navštívené při KIČ). Mezi hlavní vlastnosti vzorku řadíme dostatečnou velikost a podobnost s populací, na kterou chceme naše zjištění zobecnit (tedy tu, kterou má vzorek reprezentovat).

Způsoby výběru můžeme odlišit na dva základní, a to: (1.) *pravděpodobnostní výběr* (též jako náhodný výběr) a (2.) *nepravděpodobnostní výběr* (též nenáhodný výběr) (Soukup, Kočvarová 2016: 514). Pro pravděpodobnostní výběr je charakteristické, že (a) každá jednotka ze základního souboru má známou pravděpodobnost pro zařazení do vzorku a (b) výběr je proveden metodou vycházející z těchto pravděpodobností (Hendl 2009: 57–58). V mnoha reálných situacích může být provedení tohoto způsobu výběru značnou komplikací, protože realizace předpokládá existenci a dostupnost tzv. opory výběru (seznam všech případů, které se nachází v populaci, ze které chceme vybírat). Například pro výběr škol je taková opora dostupná v podobě školského rejstříku, resp. v příslušných výkonových výkazech MŠMT. U jiných populací, se kterými ČŠI pracuje, (např. v rámci administrace dotazníků to jsou učitelé, žáci nebo i hospitované hodiny) je situace složitější a jedinou takovou oporu bude zřejmě představovat seznam příslušných jednotek vygenerovaný v dané škole.

Druhým typem je tzv. nepravděpodobnostní výběr. U něj nejsou oproti pravděpodobnostnímu výběru vybírány jednotky z populace náhodně. To znamená, že jednotky nemají známou (resp. stejnou) pravděpodobnost, že budou do vytvořeného výběrového vzorku zahrnuty. Pod nepravděpodobnostní výběr můžeme zahrnout výběr na základě dobrovolnosti, dostupnosti či kvótní výběr. Získaná data prostřednictvím těchto technik jsou však ve srovnání s daty získanými pravděpodobnostním způsobem výběru ve značné „nevýhodě“. Problémem je zde riziko tzv. systematické

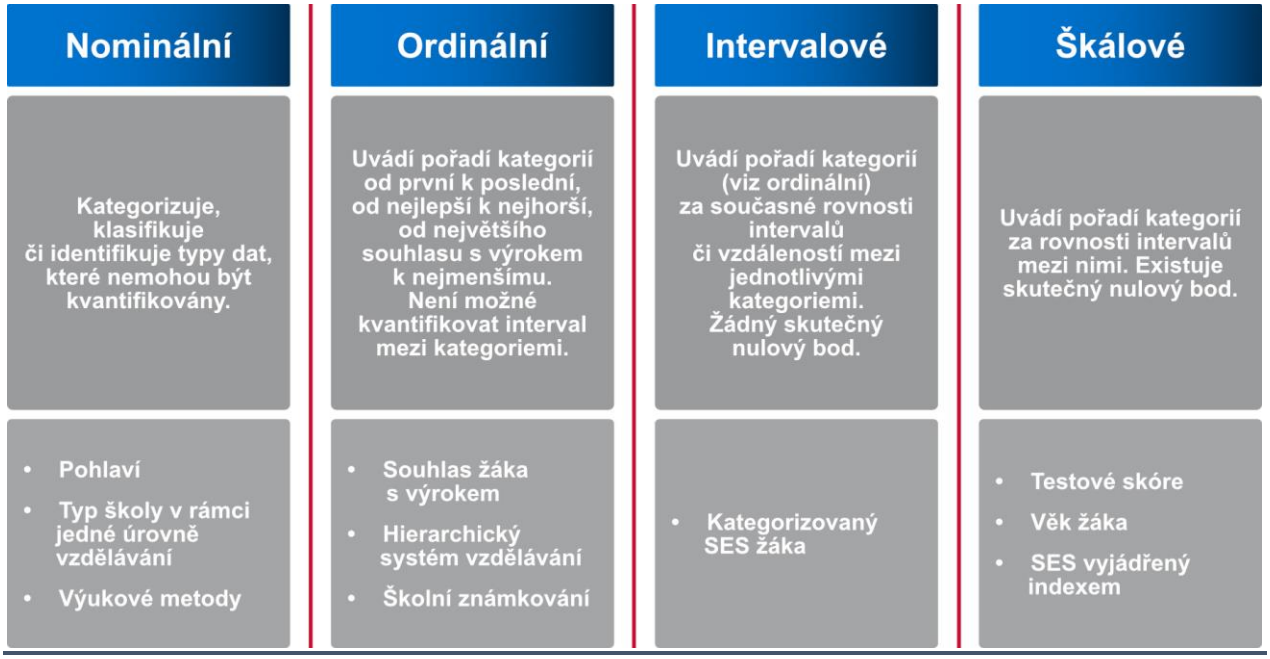
chyby či zkreslení (sampling bias), které se projeví jako nadreprezentace či podreprezentace určité subpopulace, resp. charakteristiky ve srovnání se základní populací. Jinými slovy lze říci, že pravděpodobnost pro vybrání konkrétního případu (např. jedince či školy) byla nižší, resp. vyšší ve srovnání s pravděpodobností jeho zvolení při realizaci náhodného výběru. Pravděpodobnostní výběr toto výběrové zkreslení eliminuje (Hendl 2009: 57).

3.3 Proměnné

3.3.1 Typy proměnných

Typ proměnné přímo ovlivňuje aplikovatelnost analytických metod a následnou vizualizaci dat. Pro každý typ proměnné je nutné používat k tomu vhodné metody analýzy a vhodný vizualizační nástroj. Statistická literatura rozlišuje čtyři základní typy proměnných – nominální, ordinální, intervalové a škálové – jak je znázorňuje schéma 4. Nejčastější metody analýzy a vizualizace dat (avšak ne všechny možné) znázorňují tabulky 2 až 4. Jednotlivé příklady analýzy a vizualizace dat jsou podrobně popsány v dalších kapitolách.

SCHÉMA 4 | Shrnutí typů využívaných proměnných



TABULKA 2 | Možnosti analýzy a vizualizace nominálních proměnných (shrnutí)

Nominální proměnné	
Analytické nástroje	Vizualizace dat
Deskriptivní statistika (počty a podíly) Chí-kvadrátový test Millovy metody Cramerovo V Různé varianty logistické (či logitové) regrese	Sloupcové a pruhové grafy (kategorizace při třídění prvního i druhého stupně) Paprskové grafy (kategorizace) Mapová vizualizace (bodová) Infografika

TABULKA 3 | Možnosti analýzy a vizualizace ordinálních proměnných (shrnutí)

Ordinální proměnné	
Analytické nástroje	Vizualizace dat
Deskriptivní statistika (počty a podíly) Korelační analýzy (Spearmanovo Rhó, Kendallovo Tau) Ordinální logistická regrese Jednoduchá i multinominální lineární regrese	Sloupcové a pruhové grafy (kategorizace při třídění prvního i druhého stupně) Koláčové grafy Mapová vizualizace Infografika

TABULKA 4 | Možnosti analýzy a vizualizace intervalových a škálových proměnných (shrnutí)

Intervalové a škálové proměnné	
Analytické nástroje	Vizualizace dat
Deskriptivní statistika (počty a podíly) Korelační analýzy (Pearsonovo R) Jednoduchá i vícenásobná (ne)lineární regrese	Sloupcové a pruhové grafy (vyjma kategorizace při třídění) Boxplot grafy (s analýzou odchylných případů) Bodové grafy (s uvedením regresní přímky) Mapová vizualizace (plošná) Infografika

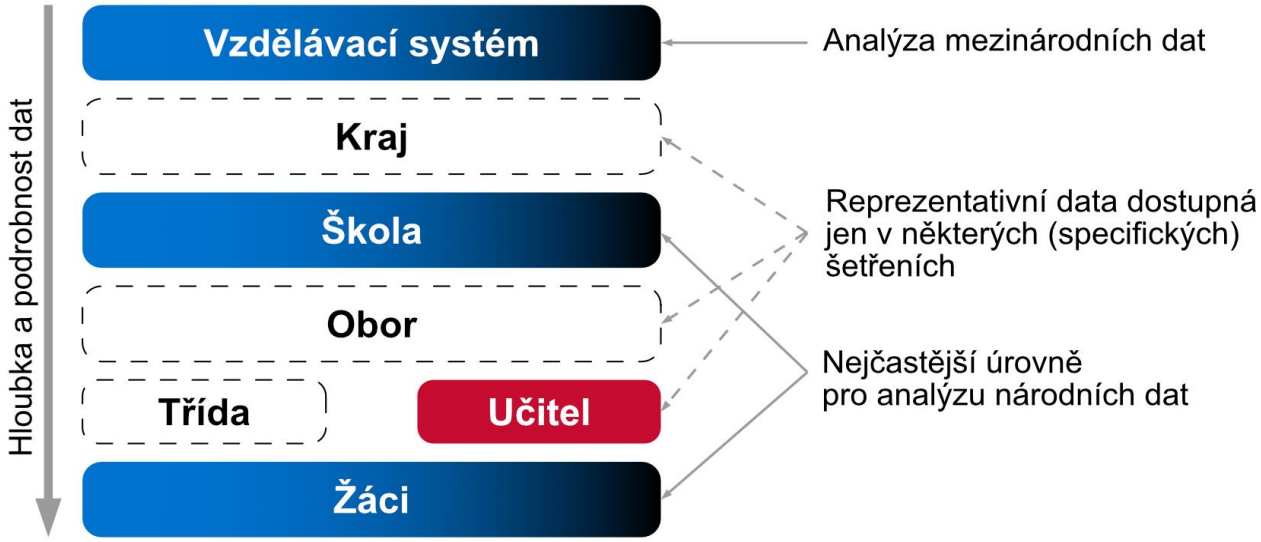
3.3.2 Úroveň proměnných

Kromě typu proměnných ovlivňuje analýzu a následnou vizualizaci výsledků samotná úroveň proměnných, resp. úroveň, na které jsou měřeny jednotky, k nimž se proměnné vztahují (pro účely tohoto textu toto stručně označujeme jako „úroveň proměnných“). V oboru edukačních věd se nejčastěji setkáváme s nejnižší úrovní – žáků škol. Na této úrovni probíhá řada analýz, vážou se k ní skóre dosažená v mezinárodních testech gramotností (PISA, TIMSS, PIRLS apod.), analýzy využívají i názorové postoje žáků a jejich charakteristiky (zejména SES). Všechny vyšší úrovně (kromě úrovně učitele) se zpravidla tvoří pouhou agregací dat na vyšší úroveň spolu s doplněním specifických kontextuálních proměnných pro danou úroveň. Nejen v České republice je nejčastější druhou úrovní škola, pro kterou se agregují data za žáky. Úroveň třídy je sice možné agregovat, sběr dat však zpravidla nereflektuje potřebu reprezentativnosti dat. Výběrová testování škol a žáků v rámci těchto škol oproti např. Cenzu (testování celé populace) probíhají mj. i kvůli efektivitě nákladů. Vzhledem k velmi rozdílným velikostem škol dochází často ke sběru dat pouze za jednu třídu (v případě mezinárodních testů gramotností ovšem u šetření TIMSS a PIRLS, kde to velikost škol dovolí, jsou sbírány třídy dvě, u šetření PISA se pak jedná o výběr žáků dle roku narození bez ohledu na třídu), v takovém případě se úroveň třídy rovná úrovni celé školy, ovšem s problematickou zobecnitelností závěrů interpretovaných za jednu třídu. Úroveň žáků a úroveň školy jsou v ČR nejčastěji využívané úrovně pro dvouúrovňovou analýzu národních dat (třetí úroveň analýzy [třídy] běžná jen ve větších zemích typu USA). Pro mezinárodní analýzu se pak využívá i úroveň vzdělávacího systému jako takového.

V šetřeních se objevuje několik dalších úrovní dat, které jsou v datových analýzách ČŠI využitelné, ale ne běžné – úroveň učitele, oboru a regionální úroveň kraje. Úroveň učitele je typická pro mezinárodní šetření TALIS, které se vzorkem učitelů přímo pracuje. Design šetření naopak znemožňuje spolehlivé usuzování na úroveň konkrétní školy, usuzování na úroveň žáků je pro jejich absenci v tomto šetření vyloučena. Úroveň oboru je specifická pro některá zvláštní šetření prováděná ČŠI, např. výsledky maturit. Opět je zde ale velice problematické usuzování na úroveň školy.

Úroveň, jejíž reprezentativnost není v šetřeních ČŠI standardně zajištěna, ale pro provádění komplexních analýz v kontextu českého vzdělávacího systému a jeho problémů je klíčová, představuje úroveň kraje. Zatímco méně často využívané úrovně třídy a oboru není nezbytně nutné designovat reprezentativně, protože u nich dochází k pravidelné obměně a rychlému zastarávání problémů (což znemožňuje jakékoli přijímání opatření pro aktuální výsledky šetření, případná opatření jsou přijímána zpětně do již obměněného kontextu), úroveň kraje a výsledky šetření z ní vzešlé jsou v čase relativně stabilní. Vzhledem k problémům ve vzdělávání založeným na strukturální postiženosti některých regionů (zejména Ústecký, Karlovarský a Moravskoslezský kraj) by se snahy ČŠI v oblasti redesignování budoucích šetření měly soustředit na zajištění reprezentativnosti na úrovni krajů (v případě SŠ jen tam, kde to umožňuje oborová struktura škol v jednotlivých krajích, spádovost aj.), jako doplňující úrovně reprezentativnosti na úrovni žáků a škol. Implementace úrovně krajů do standardně vybíraného reprezentativního vzorku zajistí do budoucna kvalitnější a přesnější zhodnocení výsledků vzdělávání souvisejících s jednotlivými kraji a umožní krajským orgánům přijímání efektivních opatření pro nápravu zjištěných problémů a další zvýšení kvality vzdělávání. Návaznost jednotlivých úrovní proměnných shrnuje schéma 5.

SCHÉMA 5 | Úrovně proměnných využívané v analýzách vzdělávacích systémů



3.4 Počáteční kontrola a úprava dat

Kapitola 3 v Sadě typizovaných analytických nástrojů

Kontrola dat představuje počáteční fázi datové analýzy. V rámci ní identifikujeme možné problémy a snažíme se předejít nepřesnostem v následných analýzách. Zároveň se snažíme o nápravu chyb. Za optimální považujeme takovou korekci, při které nedojde ke ztrátě požadované informace. Pokud taková náprava již není možná, pak je v některých případech potřeba záznam odstranit, přičemž se může jednat pouze o odebrání několika buněk, ale také celých případů či proměnných. Kontrole dat musí být věnován náležitý čas, protože platí, že nelze očekávat kvalitní analytický výstup s nekvalitními vstupními daty.

Některé chyby mohou vzniknout například chybným zaznamenáním původních dat (typografické chyby, posunutí desetinné čárky, při telefonickém zjišťování přeslechnutí respondenta), jiné chyby mohou vzniknout při importování dat v důsledku špatné kompatibility jednotlivých aplikací. V textu představíme vybrané dílčí aspekty, které se s tématem kontroly dat pojí.

Různé statistické výpočty, resp. programy vyžadují, aby byla proměnná nastavena na určité úrovni měření (viz 2.2 Typy proměnných). Jednou z prvních věcí, kterou bychom měli po importu dat prověřit, je **typ měřítka proměnných**. V praxi se můžeme setkat s případy, kdy software přiřadí proměnné chybné měřítko, což může znamenat problém pro následnou analýzu dat (v lepším případě pouze časové zdržení). Některé programy navíc neumožňují vykonat požadovaný výpočet, aniž je proměnná nastavena na požadované úrovni.

Další věc, kterou je nutné zjistit, je **způsob kódování** proměnných. Důležité to je zejména v případech, kdy nepracujeme s vlastními datovými soubory, ale naopak s daty z externích zdrojů. Např. k datasetům z mezinárodních šetření bývá připojen číselník, ve kterém je přehledně uvedeno kódování jednotlivých odpovědí. Znalost kódování je také důležitá pro interpretaci výsledků. Programy, resp. vzorce samy o sobě nedokážou odlišit kvalitativní rozměr jednotlivých odpovědí. Pracují pouze s čísly, proto je na samotném uživateli, aby věděl, jaké kvantitativní hodnoty byly k jednotlivým položkám přiřazeny.

Jako příklad uveďme kódování odpovědi vyjadřující míru souhlasu respondenta s určitým tvrzením. Jednotlivé možnosti byly kódovány následovně: *1 = rozhodně ano*, *2 = spíše ano*, 3 = *spíše ne* a 4 = *rozhodně ne*. V tomto případě bychom mohli na základě předchozí zkušenosti usuzovat, že vyšší hodnota značí silnější míru souhlasu, a tím se ve výsledku dopustit dezinterpretace zjištění. Problém to může představovat zejména v korelační analýze (resp. i dalších statistických technikách, které jsou na jejím principu založeny), kde v konečném výstupu není znění konkrétních odpovědí známo. Další potenciální chybou je sdružení dvou odlišných odpovědí pod stejnou číselnou hodnotu, např. *1 = ano, 1 = nevím*, 3 = *ne*. Takovéto sloučení lze podle potřeby provést až při práci s daty, a to způsobem, aby původní proměnná a její kódování byly zachovány. Obecně je nevhodné slučovat dvě či více vzájemně se vylučující kategorie. Nelze říci, že je nějaký způsob kódování apriorně špatný, je však potřeba mít znalost toho, jak jsou jednotlivé proměnné a odpovědi kódovány v datovém souboru, se kterým pracujeme.

Chybějící hodnoty (missing values) označují chybějící informace v datovém souboru, kdy případu či několika případům nebyla přiřazena hodnota v některé z proměnných. Existují různé příčiny vzniku chybějících hodnot. Ta může např. nastat pochybením na straně tazatele, který nezapsal odpověď respondenta, nebo také tehdy, když respondent odmítl odpovědět. V tomto případě je však nutné takovou skutečnost také zaznačit (kódovat).

V některých datových souborech zůstávají chybějící hodnoty neoznačeny a nelze již zpětně identifikovat příčinu. Toto odlišení ovšem může být pro vyhodnocování dat klíčové (je podstatný rozdíl, pokud dotázaný odmítl odpověď, nebo pokud tazatel odpověď nezaznamenal). Chybějící hodnoty mohou vznikat také logicky, například větvením formuláře, kdy je odpověď na určitou otázku podmíněna odpovědí na předchozí otázku.

Uveďme si příklad s řetězcem dvou otázek: (1) *Vyučujete cizí jazyk? – ano/ne* a (2) *Jaký cizí jazyk vyučujete? – výběr možností*. Z uvedeného příkladu je zjevné, že pokud respondent v prvním případě zvolil odpověď *ne*, již nemůže zodpovědět druhou otázku, a proto mu bude přiřazena chybějící hodnota. Tato skutečnost musí být v datovém souboru zjistitelná.

Vhodné je také prozkoumat, zda jsou chybějící hodnoty rozmístěny náhodně, nebo jsou naopak systematicky nadreprezentovány u některé subpopulace. Analýza chybějících hodnot může sama o sobě představovat plnohodnotnou náplň výzkumu. Pokud například zjistíme, že se chybějící hodnoty systematicky vyskytují u nějaké podskupiny respondentů, nejspíše nás začne zajímat i důvod. Podle předpokládaného dopadu, který mohou mít chybějící hodnoty na výsledky analýzy, je potřeba rozhodnout, jak s případy dále pracovat. V některých situacích postačí takové případy odstranit (resp. do analýzy ani nevstoupí), v jiných situacích je možné chybějící hodnoty imputovat.

Poslední vybraný aspekt související s problematikou kontroly dat představuje identifikace **odlehlých hodnot** (outliers). Odlehlé hodnoty jsou takové, které se výrazně odlišují od ostatních hodnot dané proměnné. Vhodným způsobem k jejich identifikaci je provedení průzkumu dat ještě před analýzou, kdy zkoumáme vlastnosti jednotlivých proměnných. Typicky se může jednat o použití základních ukazatelů míry centrální tendence a rozptýlenosti dat (viz 4.3 Univariační analýza dat). Velmi efektivní je grafická vizualizace hodnot (např. pomocí boxplotu, histogramu nebo bodového grafu).

Odlehlá hodnota může vzniknout nepřirozeně jako výsledek chybného zápisu dat (například posunutím desetinné čárky, zaznačení výsledku v odlišných jednotkách apod.). Stejně tak můžeme pozorovat netypickou hodnotu zcela přirozeně. Použití některých statistických technik předpokládá neexistenci těchto hodnot (v kontextu bivariační a multivariační analýzy pak můžeme mluvit o odlehlých případech), je proto na zvážení, zda bychom měli nechat takové hodnoty, resp. případy vstoupit do analýzy a riskovat zkreslení výsledků, či proměnnou jinak transformovat. Neexistuje standardizovaný způsob, jak se s odlehlými hodnotami vypořádat. Řešení je potřeba promýšlet vždy v kontextu námi zamýšleného analytického záměru a také v souvislosti s příčinou vzniku takové hodnoty.

3.5 Vybrané analytické techniky

3.5.1 Deskriptivní analýza dat

Kapitola 4 v Sadě typizovaných analytických nástrojů

Základním a často prvotním krokem, kterým práci s připravenými daty zahajujeme, je popis dat. K tomu slouží deskriptivní statistika, jejímž cílem je sumarizace charakteristik určité proměnné obvykle skrze popis míry centrální tendence a rozptýlenosti dat.

Asi nejtypičtější střední hodnotou, vhodnou zejména pro škálové proměnné, je **aritmetický průměr**. Průměr vypočítáme tak, že sečteme hodnoty všech případů v souboru a součet vydělíme celkovým počtem případů. Jedná se o běžně používaný ukazatel, který je však velmi náchylný na odlehlé hodnoty.

Další možný ukazatel centrální tendence představuje **medián**. Medián získáme jednoduše tak, že seřadíme hodnoty a vezmeme tu, která se nachází uprostřed. Přesně polovina případů pak bude mít hodnotu nižší a polovina případů bude mít hodnotu vyšší, než je medián. Medián tedy není citlivý k odlehlým hodnotám jako průměr.

Za účelem prozkoumání variability proměnné můžeme pracovat s kvartily, kterými podobným způsobem rozdělíme soubor na čtvrtiny. Hodnota prvního kvartilu je hodnotou, pod kterou se nachází přesně 25 % případů, druhý kvartil je medián a pod hodnotou třetího kvartilu se nachází 75 % případů. Jakmile známe hodnoty kvartilů, můžeme vypočítat **mezikvartilové rozpětí**, které je rovno rozdílu mezi hodnotou třetího a prvního kvartilu.

Další běžně uváděnou mírou variability je **rozptyl**. Abychom vypočítali rozptyl, musíme nejdřív od hodnoty každého případu odečíst celkový průměr. Takto získané odchylky od průměru následně umocníme na druhou a sečteme. Součet umocněných odchylek nakonec vydělíme počtem případů.

Odmocněním rozptylu získáme **směrodatnou odchylku**. Tím v podstatě vracíme hodnoty proměnné do původní jednotky, což usnadní interpretaci.

Pro popisnou statistiku rozdělení proměnné stačí několik základních údajů, a to nezávisle na úrovni měření. Základem je průměr, směrodatná odchylka, minimum a maximum. Pokud se medián odchyluje výrazně od průměru, je vhodné jej rovněž doplnit do tabulky popisné statistiky.

TABULKA 5 | Příklad popisné statistiky

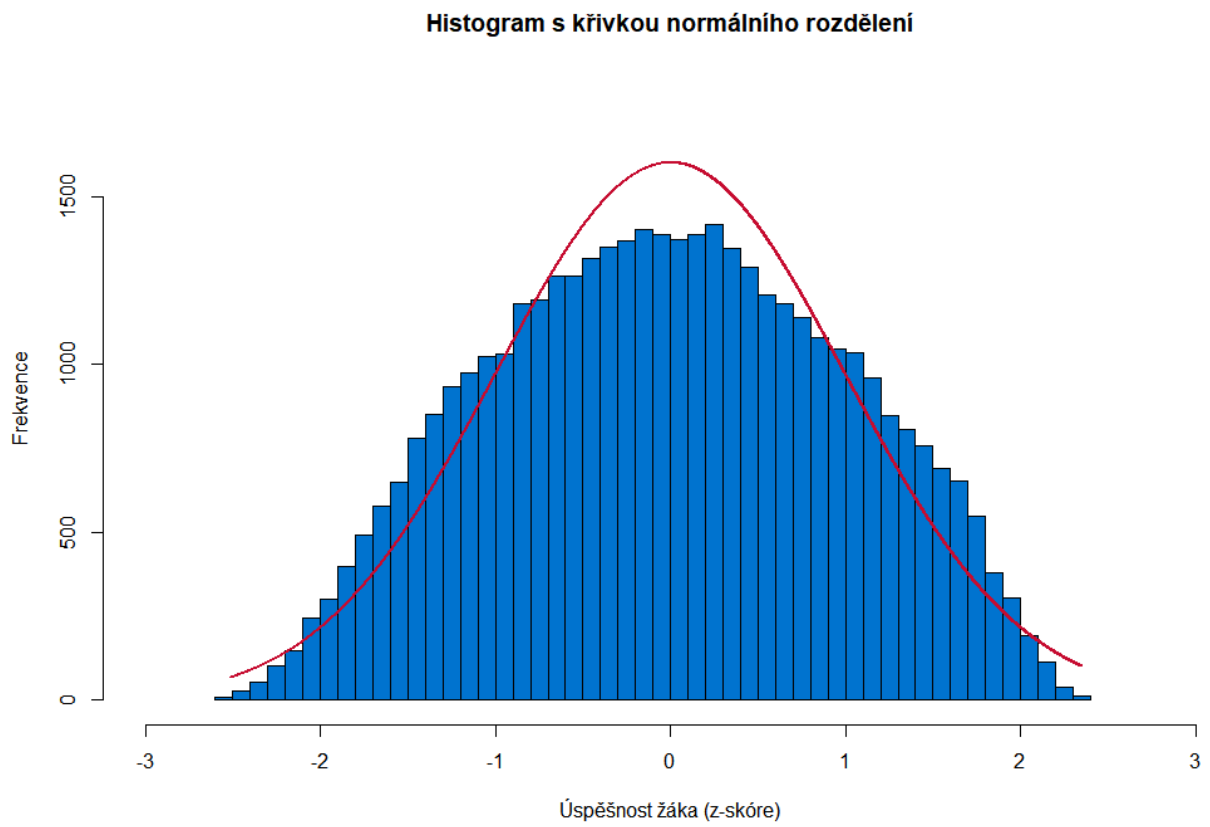
Název proměnné	Počet pozorování	Průměr	Směrodatná odchylka	MIN	MAX
Délka praxe v letech	5 537	20,26	10,61	1	45
Obliba čtení	5 518	9,47	1,73	2,55	14,58
Předškolní aktivity žáka	5 269	10,52	1,70	2,14	14,70
Předškolní dovednosti žáka	5 230	9,91	2,01	4,87	13,88
Sebejistota ve čtení (žák)	5 496	9,96	1,82	2,96	13,47
Socioekonomický status školy	5 518	10,58	0,61	7,75	12,05
Socioekonomický status žáka	5 202	10,58	1,43	4,16	14,82
Spokojenost učitele	5 537	8,87	2,17	4,39	12,00
Vztah rodičů k četbě	5 257	9,80	1,88	2,72	13,88

Poznámka: Za účelem ukázky uvedeny jen vybrané proměnné z šetření PIRLS 2016.

Tvar rozdělení dat je důležitý pro rozhodnutí dalšího analytického postupu. Pokud je proměnná normálně rozdělena, můžeme použít běžně používané analytické techniky. Zejména parametrické korelační koeficienty, parametrické testy rozdílů v průměrech (t-test, ANOVA) a regresní analýzu. Rozložení dat zjistíme pomocí grafu typu histogram, který ukazuje četnosti výskytu hodnot proměnné v daném intervalu (software většinou určí automaticky).

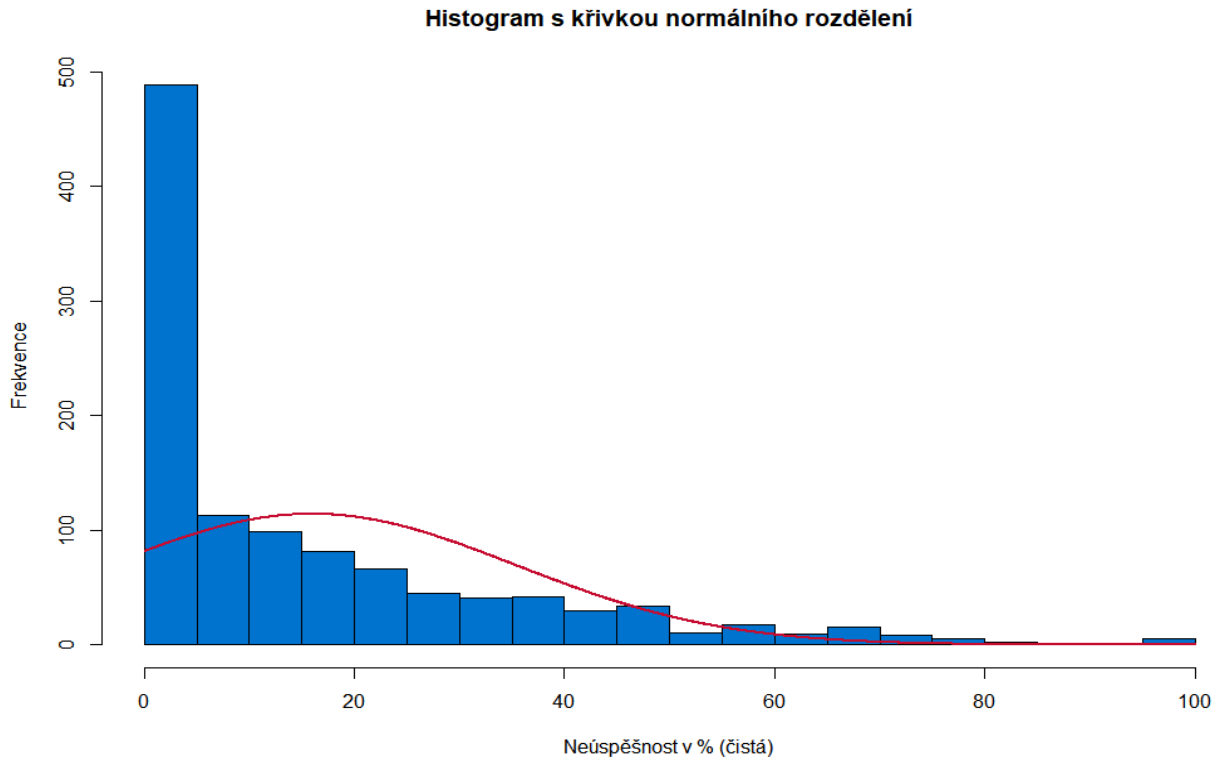
Pro první graf zobrazuje výsledky žáků 9. tříd z *Výběrového zjišťování výsledků žáků 9. ročníku*, a to v matematice. Na ose x jsou hodnoty proměnné úspěšnost žáka (z-skóre) a na ose y jsou četnosti daných intervalů, které jsou znázorněny sloupcovým grafem. Spolu s křivkou normálního rozdělení tvoří námi sledovaný histogram. Při porovnání rozložení této proměnné můžeme vizuálně říci, že proměnná je zhruba normálně rozdělena, byť v případě hodnot okolo průměru je rozdělení tzv. platykurtické, tedy nižší, než je očekávaná frekvence hodnot v případě normálního rozdělení (opakem by bylo rozdělení špičaté). Obecně ale proměnná splňuje předpoklady normálnosti a může vstoupit do celé řady parametrických statistických testů. Navíc počet případů je více než dostačující (N = 40 115).

GRAF 1 | Příklad rozložení proměnné



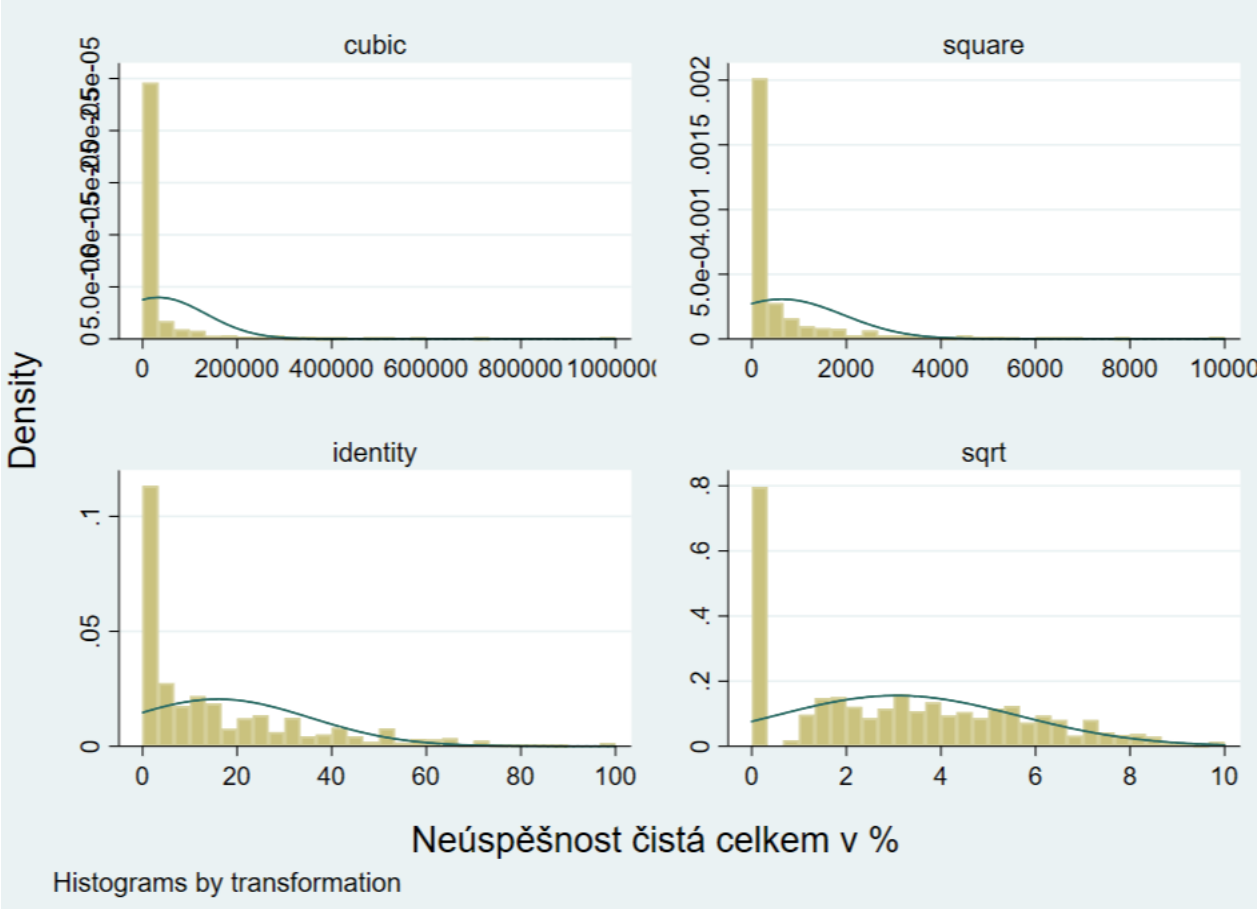
Druhým grafem je ilustrace proměnné hrubá neúspěšnost žáků u maturitní zkoušky 2018 (po podzimním zkušebním období; v %). Data pocházejí z datového souboru Centra pro zjišťování výsledků vzdělávání (CVVZ) a případy tvoří jednotlivé středoškolské obory, na nichž se v daném školním roce konala maturitní zkouška. Histogram zde jasně ukazuje, že tato proměnná normálně rozdělena není a vůbec se normálnímu rozdělení ani neblíží. Data jsou výrazně vychýlena vpravo, jedná se tedy o pozitivní zešikmení. Současně je proměnná rozdělena špičatě. Obecně se jedná o příliš rozptýlená data (overdispersed data), navíc s omezeným intervalem 0–100 %. Z tohoto důvodu už není možné použít klasickou lineární regresi, ale specifický model pro tento typ dat (ukázáno v části regresní modely). Nabízí se Poissonova regrese nebo negativní binomická regrese (v případě, že je proměnná příliš rozptýlená a parametr disperse je statisticky významný, je vhodnější použít negativní binomickou regresi). Tento typ modelů zvládá specifický software jako Stata a R.

GRAF 2 | Příklad rozložení hodnot proměnné



Jakým způsobem postupovat, pokud naše data nesplňují charakteristiky normálního rozložení, ale zároveň se nejedná o typ dat, který by byl vhodný pro modelování za pomoci specifických technik? Vhodným řešením je transformace proměnné a použití běžných technik jako korelace a regrese. Nejběžnější transformací je převedení do logaritmické škály (zlogaritmování proměnné zpravidla o základu 10), převedení na z-skóre, umocnění na druhou, druhá odmocnina či inverze proměnné (1/x). Všechny tyto transformace neznamenají manipulaci s daty! Data jsou jen transformována pro potřeby lineární funkce. Jakým způsobem zjistit, která transformace je nejvhodnější? Ideální je funkce „gladder" v programu Stata. Ta provede automaticky celou řadu transformací a zobrazí v grafu ty, které nejlépe adresují původní rozložení proměnné. Příkladem výstupu je obrázek 14. Z histogramů transformací pak můžeme vidět, že nejlepší řešení transformace je druhá odmocnina proměnné. Ostatní software (SPSS či R) tuto funkci nenabízí, a tudíž je nutné transformace vyzkoušet – manuálně provést výpočet nové proměnné a zobrazit v histogramu.

GRAF 3 | Transformace proměnných pomocí funkce „gladder“ v programu Stata



K transformacím obecně viz manuál: https://www.stata.com/manuals13/rladder.pdf

Doporučení

- Pro všechny proměnné vytvořit deskriptivní tabulku obsahující průměr, směrodatnou odchylku, minimální a maximální hodnotu.
- Vytvořit graf typu histogram a porovnání s normálním rozdělením.
- Případně provést testy normality dat (Smirnov-Kolmogorův test).
- Případné zvolení způsobu transformace pro účely lineární funkce.

3.5.2 Porovnání průměrů

Kapitola 5 v Sadě typizovaných analytických nástrojů

Základní statistickou technikou pro bivariační analýzu je porovnání průměrů u zvolených skupin. Například můžeme porovnat průměrné skóre z výsledků testování, jak se liší mezi chlapci a dívkami, popřípadě jak se liší průměrné skóre mezi vícero kategoriemi, například kraji. V prvním případě použijeme klasický t-test, v druhém případě, kdy porovnáváme více než dvě kategorie, analýzu rozptylu (ANOVA).

T-test testuje statistickou významnost rozdílu dvou středních hodnot (dvou aritmetických průměrů). Pomocí t-testu de facto odpovídáme na otázku, zda jsou dva populační průměry stejné, nebo zda je mezi nimi rozdíl. Celkem můžeme rozlišit tři formy t-testu. Prvním je *t-test pro jediný výběr (One-Sample T-Test)*, který zjišťuje, zda se střední hodnota v populaci liší od námi předpokládané hodnoty, nebo zda výběr pochází z populace, jejíž průměr známe. Druhým je *t-test pro dva nezávislé výběry (Independent-Samples T-Test)*, který srovnává průměry dvou skupin případů, a tedy ukazuje, zda se statisticky liší dva populační průměry. Typickým příkladem je srovnání průměrného skóre z výsledků testování mezi chlapci a dívkami. Třetí formou je *t-test pro párová data (Paired-Samples T-Test)*. Párová data získáme závislým výběrem, technicky se pak jedná o porovnání průměrů dvou vzájemně provázaných proměnných. Typickým příkladem

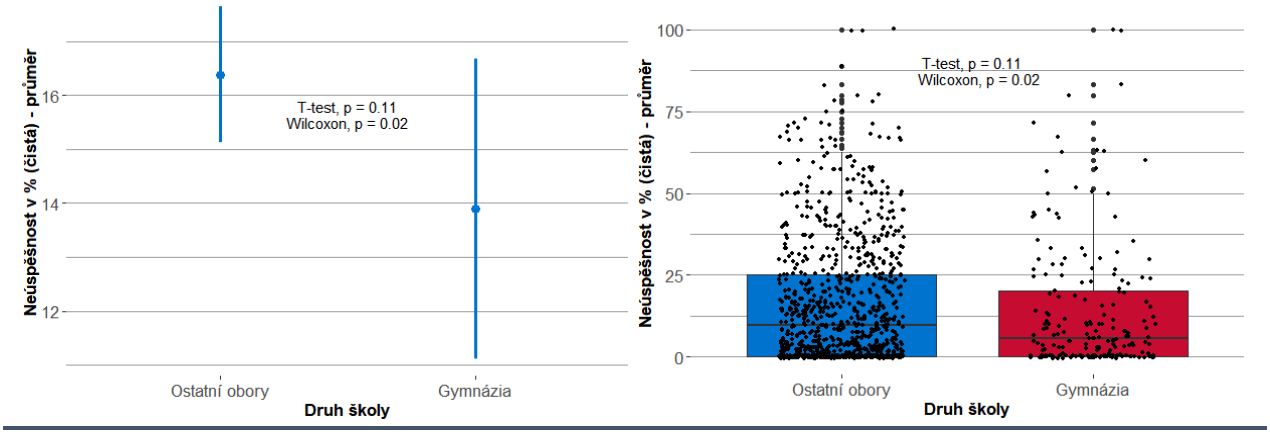
jsou panelová šetření, kdy se jednoho vzorku dotazujeme opakovaně. Můžeme poté změřit rozdíl v průměrech u tohoto jednoho vzorku např. před a po nějaké změně.

ANOVA neboli analýza rozptylu (Analysis of variance) má proti t-testu pro sociálněvědní výzkum jednu zásadní výhodu. Tou je skutečnost, že na rozdíl od t-testu zjišťuje, zda se v populaci liší průměry více skupin najednou. ANOVA rozkládá rozptyl v datech výběrového souboru na dvě složky a zkoumá jednak, jak se liší hodnoty uvnitř jednotlivých skupin (vnitroskupinový rozptyl) a jednak, jak se liší průměry mezi skupinami (meziskupinový rozptyl). Při hledání statistické významnosti v rozdílech průměrů můžeme použít jednoduchou *jednofaktorovou analýzu (one-way analysis of variance),* nebo dvoufaktorovou *(two-way analysis of variance).* Jednofaktorová analýza zjišťuje, jak se liší hodnoty závisle proměnné u kategorií jedné nezávisle proměnné. Dvoufaktorová analýza pak zjišťuje, jak se liší hodnoty závisle proměnné u kategorií dvou nezávisle proměnných. Příkladem může být sledování průměrného skóre v mezinárodním testování žáků mezi jednotlivými kraji v ČR (jednofaktorová analýza), nebo sledování skóre mezi jednotlivými kraji a současně ještě např. na základě pohlaví (dvoufaktorová analýza).

Ideálním způsobem prezentace rozdílů v průměru je graf chybových úseček (error bar chart). Tento graf by měl být upřednostněn před výstupy v podobě tabulek. Umožní přehledněji ukázat rozdíly v průměrech a zároveň ukazuje intervaly spolehlivosti. Pokud se intervaly mezi kategoriemi neprotínají, jedná se o statisticky významný rozdíl. Jestliže porovnáváme dvě skupiny případů, je vhodnějším grafem tradiční krabicový graf doplněný o hodnoty testovací statistiky. Pokud ale v analýze nesplňujeme předpoklady jako homogenitu rozptylu, máme malý počet případů či je naše závislá proměnná na delší ordinální stupnici, je třeba provést příslušné specifické statistické testy vhodné pro tyto situace. Rovněž je třeba nastavit v případě mezinárodních datových souborů odpovídající vážení.

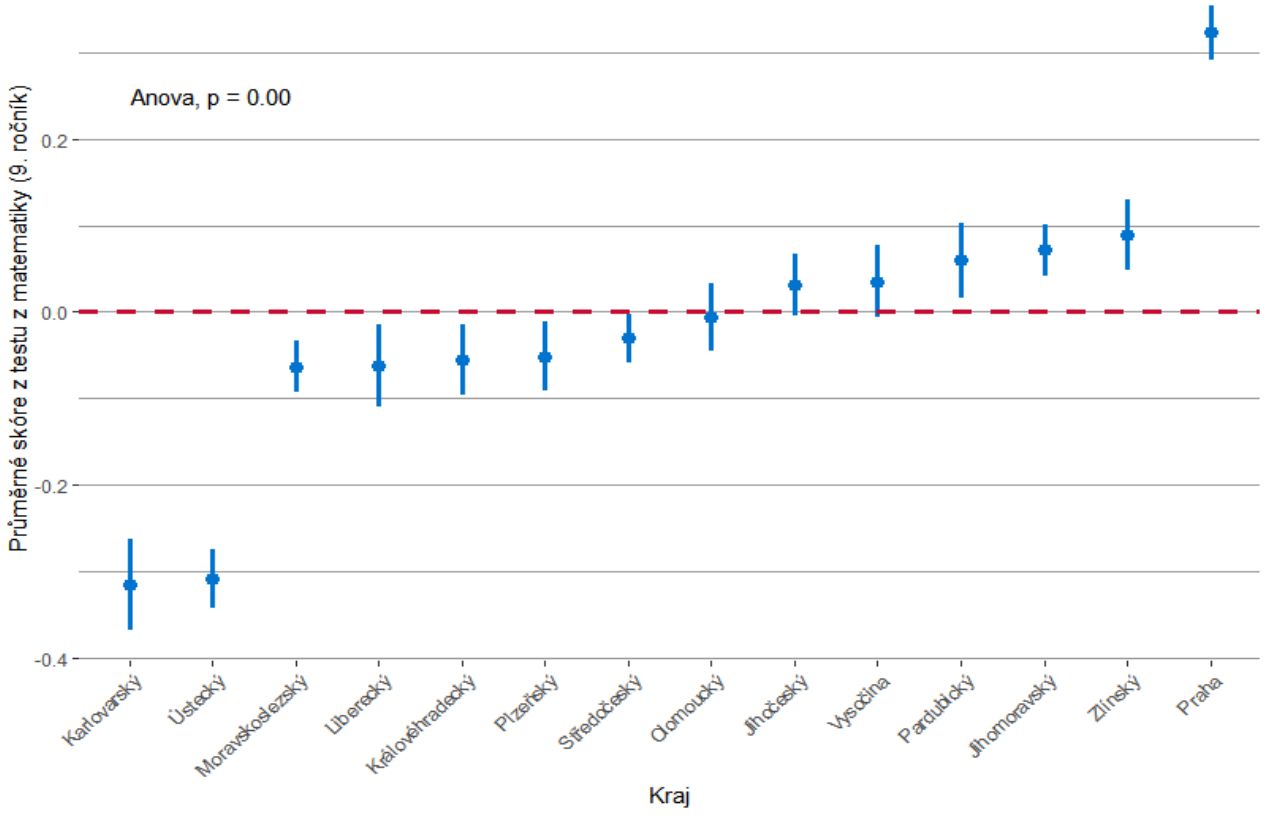
Prvním příkladem je porovnání průměrů opět na datech z maturit u proměnné hrubá neúspěšnost žáků u maturitní zkoušky 2018. Pro porovnání je ukázán jak graf chybových úseček (vlevo), tak i krabicový graf (vpravo). Oba grafy jsou doplněny testovací statistikou, tradičně t-testem, ale protože v předchozí kapitole jsme zjistili, že je tato proměnná velmi zešikmena a neblíží se normálnímu rozdělení, je použit i neparametrický Wilcoxonův test průměrů.

GRAF 4 A 5 | Průměrná neúspěšnost maturantů z matematiky – porovnání grafu chybových úseček a krabicového grafu



Příkladem pro porovnání průměrů u více kategorií jsou výsledky žáků 9. tříd z matematiky. Jednotkou je žák a jeho standardizované skóre z testu z matematiky. Zde je do grafu nanesen bodový odhad spolu s 95% intervalem spolehlivosti (úsečka). Ten je pro zjednodušení konstruován tak, že předpokládá náhodný výběr z neznámé populace. Nicméně šetření nebylo konstruované jako plošné testování celé populace, ani jako kvótní výběr na základě populačních parametrů. Interval spolehlivosti by se tak měl konstruovat dle metody sběru dat. Graf nám ukazuje, jak se liší výsledné průměry žáků v krajích. Pokud se interval spolehlivosti neprotne s daným krajem, se kterým srovnáváme, průměry se statisticky významně liší. Například tak můžeme říci, že Praha se odlišuje od všech krajů a má statisticky nadprůměrné výsledky a průměr je statisticky významně vyšší než ve srovnání se všemi zbylými kraji. Naopak celá řada krajů dosahuje podobných výsledků. Průměr za všechny kraje je přesně -0,02, což je hodnota téměř blízká nule, která je v grafu zobrazena červenou přerušovanou linií. Pokud interval spolehlivosti daného kraje protíná hodnotu nula, není jeho výsledek statisticky významně lepší od průměru. Dle osy y také můžeme říci, že se kraje odlišují od dané hodnoty.

GRAF 6 | Průměrné skóre z testu z matematiky (9. ročník)



V případě porovnání průměrů máme několik možností, jak data prezentovat a interpretovat. Nejvhodnější je graf typu boxplot nebo graf chybových úseček. Pokud se porovnávají průměry dvou kategorií, je vhodné zvolit graf typu boxplot doplněný o výsledky testovací statistiky. Pokud porovnáváme více kategorií, nejvhodnějším typem grafu je graf chybových úseček.

V případě splnění podmínek pro parametrický test zobrazujeme p hodnotu t-testu, v případě nesplnění podmínek je vhodnější použít neparametrický Wilcoxonův test. Pro vícekategorické proměnné pak výsledky testu ANOVA a její neparametrická alternativa Kruskalův-Wallisův test.

3.5.3 Korelační koeficienty

Kapitola 5.3 v Sadě typizovaných analytických nástrojů

Jednou ze základních analýz, ze kterých již můžeme odhadovat míru asociace mezi dvěma proměnnými (nikoli odhalit, která z proměnných je skutečná příčina a která skutečný důsledek uvažovaného vztahu), je analýza asociací či korelační analýza.[2] Korelační analýza je souhrnné označení pro analýzy založené na různých typech korelačních koeficientů s odlišnými předpoklady a odlišnými výpočty, které se volí dle typů proměnných, jež vstupují do výpočtu. Uvedené příklady nejsou úplným výčtem všech korelačních koeficientů, přibližují nejčastěji využívané koeficienty. V metodologické literatuře je možné dohledat další příklady koeficientů jak pro souhlasné proměnné, tak pro proměnné na různých úrovních měření (např. Koeficient eta pro nominální*škálovou [intervalovou] proměnnou). Možnost výpočtu všech prezentovaných koeficientů je běžnou součástí všech statistických programů.

Při zjišťování míry asociace mezi dvěma nominálními proměnnými se nejčastěji využívá **Koeficient phi** (dvě binární proměnné se dvěma kategoriemi, např. pohlaví*dosažení vysokoškolského vzdělání) či **Cramerovo V** (pokud alespoň jedna z proměnných má více než dvě kategorie – např. typ střední školy – SOU, SOŠ, Gymnázium). Výstupem obou koeficientů, založených na hodnotě chí-kvadrát, je číselná hodnota v intervalu <0;1>, přičemž čím vyšší hodnota, tím silnější asociace mezi dvěma proměnnými.

V případě zjišťování míry asociace mezi dvěma ordinálními proměnnými se často využívají koeficienty založené na srovnávání dvojic hodnot a jejich míry souhlasu či nesouhlasu – (dis)konkordance – jako např. **Kendalova tau b**

[2] Ke korelačním koeficientům viz Mareš, Rabušic, Soukup. 2015. Analýza sociálněvědních dat (nejen) v SPSS.

a **tau c**. Nejčastěji se však setkáváme s použitím **Spearmanova rhó** jako korelačního koeficientu pořadí, který nesrovnává souhlas či nesouhlas párů hodnot dvou proměnných, ale jejich vzájemné pořadí. Popularita Spearmanova rhó spočívá v nižší citlivosti na nenormální rozložení dat a odchylné případy (koeficient je neparametrický). I při existenci odchylných případů, které u jiných koeficientů vedou k nezanedbatelnému zkreslení, dokáže relativně spolehlivě odhalit míru asociace mezi dvěma ordinálními proměnnými.

Zdaleka nejpoužívanějším koeficientem pro zjišťování míry asociace mezi intervalovými a škálovými proměnnými je **Pearsonův korelační koeficient** (či Pearsonovo r), který pracuje s hodnotou kovariance, tj. rozptylu mezi dvěma proměnnými děleného součinem jejich směrodatných odchylek. Nevýhodou tohoto koeficientu je poměrně vysoká citlivost na odchylné případy (Pearsonův korelační koeficient vyžaduje předpoklad lineárních vztahů).

Jak pro ordinální, tak pro intervalové a škálové proměnné platí, že výstupem korelačních koeficientů je číselná hodnota v intervalu <-1;1>, přičemž čím více se hodnota blíží hodnotě +-, tím silnější asociace mezi dvěma testovanými proměnnými. Hodnota 0 přitom znamená buď neexistenci asociace, nebo nelineární průběh dat. V takovém případě je nutné využít jinou metodu zjišťování korelací.

Interpretace korelačních koeficientů vychází z jednoduchého předpokladu. Kladná hodnota značí pozitivní vztah (čím více X, tím více Y), záporná naopak vztah negativní (čím více X, tím méně Y). V případě Cramerova V a jiných koeficientů pro nominální proměnné (včetně Koeficientu eta) pracujeme pouze s kladnými hodnotami, nejsme schopni určit směr vztahu. Hodnota koeficientu pak určuje samotnou sílu asociace. Škála určující sílu asociace ovšem není přesně definována, záleží spíše na interpretaci autora analýzy. Orientačně lze využít hodnoty v tabulce 6.

TABULKA 6 | Orientační hodnoty síly asociace v korelační analýze

Pozitivní asociace	Míra asociace	Negativní asociace
0,91–1,00	Velmi silná	-0,91 – -1,00
0,71–0,90	Silná	-0,71 – -0,90
0,51–0,70	Střední	-0,51 – -0,70
0,31–0,50	Slabá	-0,31 – -0,50
0,10–0,30	Velmi slabá	-0,10 – -0,30
0–0,10	Žádná	0 – -0,10

3.5.4 Vybrané typy regresní analýzy

Kapitola 6 v Sadě typizovaných analytických nástrojů

Předchozí techniky braly v potaz pouze vztahy mezi dvěma proměnnými, popřípadě mezi kategorií a průměry za tyto kategorie. Přestože máme možnosti, jak kontrolovat vliv dalších proměnných (parciální korelační koeficient, přidání dalšího faktoru v rámci porovnání průměrů ANOVA), předchozí techniky nejsou vhodné pro modely, které obsahují velký počet nezávisle proměnných. K tomu slouží především regresní modely.

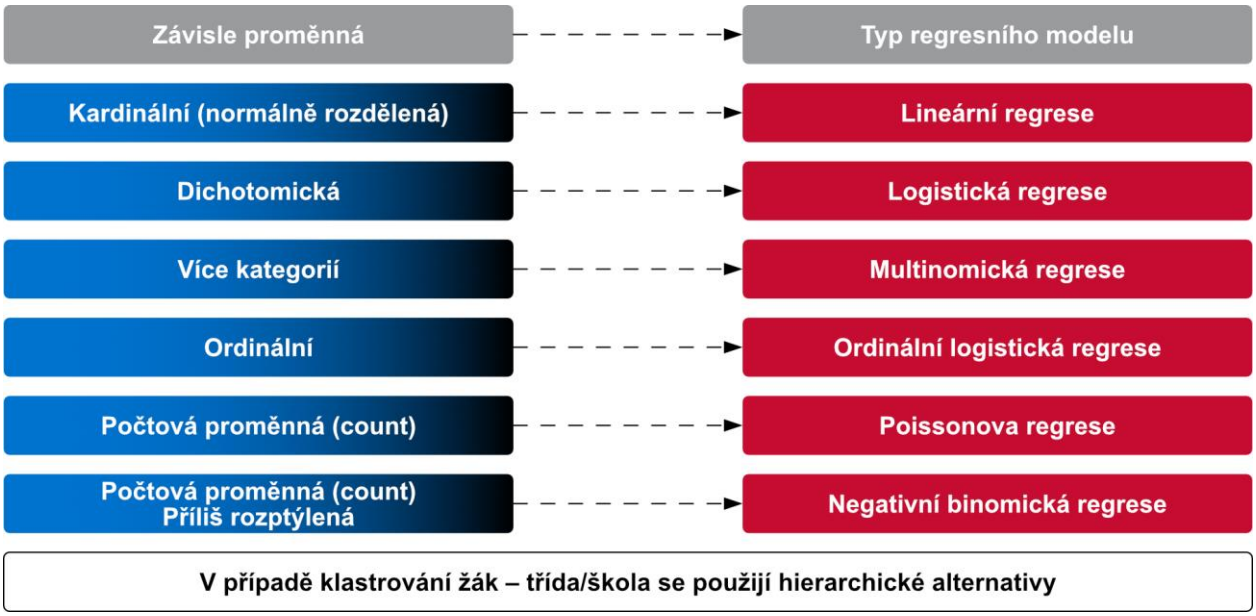
Jejich cílem je určit asociační vztah mezi ZP a NP při kontrole dalších (třetích) proměnných, které mohou být zkorelované jak s ZP, tak NP. Základní technikou je mnohonásobná lineární regrese, která předpokládá normální rozdělení reziduí (chyb), linearitu vztahu mezi ZP a NP, nezávislost případů a homoskedasticitu.

Při analýze školních dat jsou některé tyto předpoklady narušeny. V případě, kdy jednotkou analýzy je žák, jsou pozorování na sobě závislá, protože žák chodí zpravidla do stejné školy či třídy jako jeho spolužáci, kteří byli také vybráni do výzkumu. Škola či třída pak ovlivňuje hodnoty případů (žáků), respektive hodnoty ZP budou zkorelované. Z tohoto důvodu je správnou technikou hierarchické modelování, které bere v potaz klastrování případů (žáků) za třídy či školy. Výpočet směrodatných chyb bere tuto datovou strukturu v potaz, proto jsou směrodatné chyby vyšší a konzervativnější než v případě použití standardní mnohonásobné regrese. To znamená, že hranice pro překročení statistické významnosti je vyšší, tudíž je menší pravděpodobnost chyby prvního řádu. Hierarchický model navíc umožňuje modelovat efekty proměnných na vyšší úrovni, zpravidla na úrovni školy a třídy. Navíc umožňují meziúrovňové interakce (cross-level interaction), kdy efekt individuální proměnné na první úrovni je podmíněn hodnotou proměnné na vyšší úrovni, například individuální SES žáka v závislosti na celkovém SES školy.

Hierarchické modely (*random intercept* nebo *random intercept and slopes*) předpokládají, že chybová složka v regresi na obou úrovních není zkorelovaná s prediktory. To znamená, že přímo umožňují modelovat efekt proměnných na druhé a vyšší úrovni. Alternativou těchto modelů pro hierarchický typ dat jsou modely s fixními efekty, které se v edukačních vědách také používají. Jejich výhoda je, že postihnou veškeré efekty měřitelných i neměřitelných proměnných na druhé úrovni, nicméně neumožňují z tohoto důvodu modelovat vliv proměnných na druhé úrovni, protože jsou perfektně korelované s druhou úrovní, která nevariuje. Tyto modely jsou vhodné jen v případě, že nás zajímá efekt proměnné na první (individuální) úrovni.

Obvykle se používají hierarchické lineární modely, nicméně pokud je proměnná dichotomická, používá se hierarchická logistická regrese (v případě více kategorií multinomická regrese), v případě číselné proměnné, která je příliš rozptýlená, se používá hierarchická Poissonova nebo negativní binomická regrese.

SCHÉMA 6 | Přehled typů regresních modelů pro různý typ a charakter dat



Ve schématu výše jsou popsány jednotlivé modelové varianty regresní analýzy v závislosti na typu a charakteru dat. Klasické hierarchické modely jsou vysvětleny v příslušných sekundárních analýzách mezinárodního šetření (nejnověji PIRLS 2016).

V rámci tohoto textu se ale podíváme na možnosti analýzy domácích šetření, jejichž datová podoba je specifická a od mezinárodních dat se liší. Za prvé data nejsou náhodným výběrem z populace, popřípadě z jednotlivých strat (velikost školy, geografie školy atd.). Za druhé k šetřením zpravidla nemáme individuální dotazník a chybí tak celá řada dat na individuální úrovni, ať už sociodemografická (SES), či postojová. Možnosti analýzy těchto dat jsou zatím omezené. V datasetu je k dispozici identifikátor kraje a anonymizovaný identifikátor školy. Proto můžeme spočítat z výsledků jejich rozptyl nebo průměr za školu. Náš model výsledků žáků přidává rozptyl výsledků, SES kraje (hodnota ze šetření PIRLS 2016) a rozvodovost jako proxy proměnnou pro sociální charakteristiky kraje (ČSÚ). Model je pouze ilustrační, protože bychom mohli modelovat více proměnných na úrovni krajů (podíl exekucí, HDP na obyvatele atd.), nicméně průměrné SES kraje vysvětlí velkou porci variace v průměrných výsledcích škol napříč kraji.

Kdybychom měli individuální proměnné, mohli bychom přidat i je, takto nám model vysvětluje pouze rozptyl na vyšší úrovni (školy) a říká nám, jak se žáci v dané škole v daném kraji v průměru liší na základě rozptylu výsledků v dané škole, rozvodovosti v kraji a SES kraje.

TABULKA 7 | Příklad regresní analýzy výsledků testů z matematiky, češtiny a anglického jazyka – 9. ročník

	Matematika	Čeština	Anglický jazyk
Rozptyl výsledků ve škole	-0.188***	-0.311***	-0.322***
	(0.015)	(0.012)	(0.013)
SES kraje	0.082***	0.069***	0.104***
	(0.015)	(0.012)	(0.014)
Rozvodovost v kraji	-0.082***	-0.060***	-0.048***
	(0.015)	(0.012)	(0.013)
(Konstanta)	-0.063***	-0.074***	-0.097***
	(0.014)	(0.012)	(0.013)
AIC	102942.891	103556.941	94604.102
BIC	102994.488	103608.563	94655.258
Log likelihood	-51465.446	-51772.471	-47296.051
Počet pozorování	40115	40280	37272
Počet skupin 2 úrovně - škola	1444	1443	1334
Rozptyl - škola	0.251	0.164	0.187
Rozptyl - reziduály	0.702	0.715	0.689

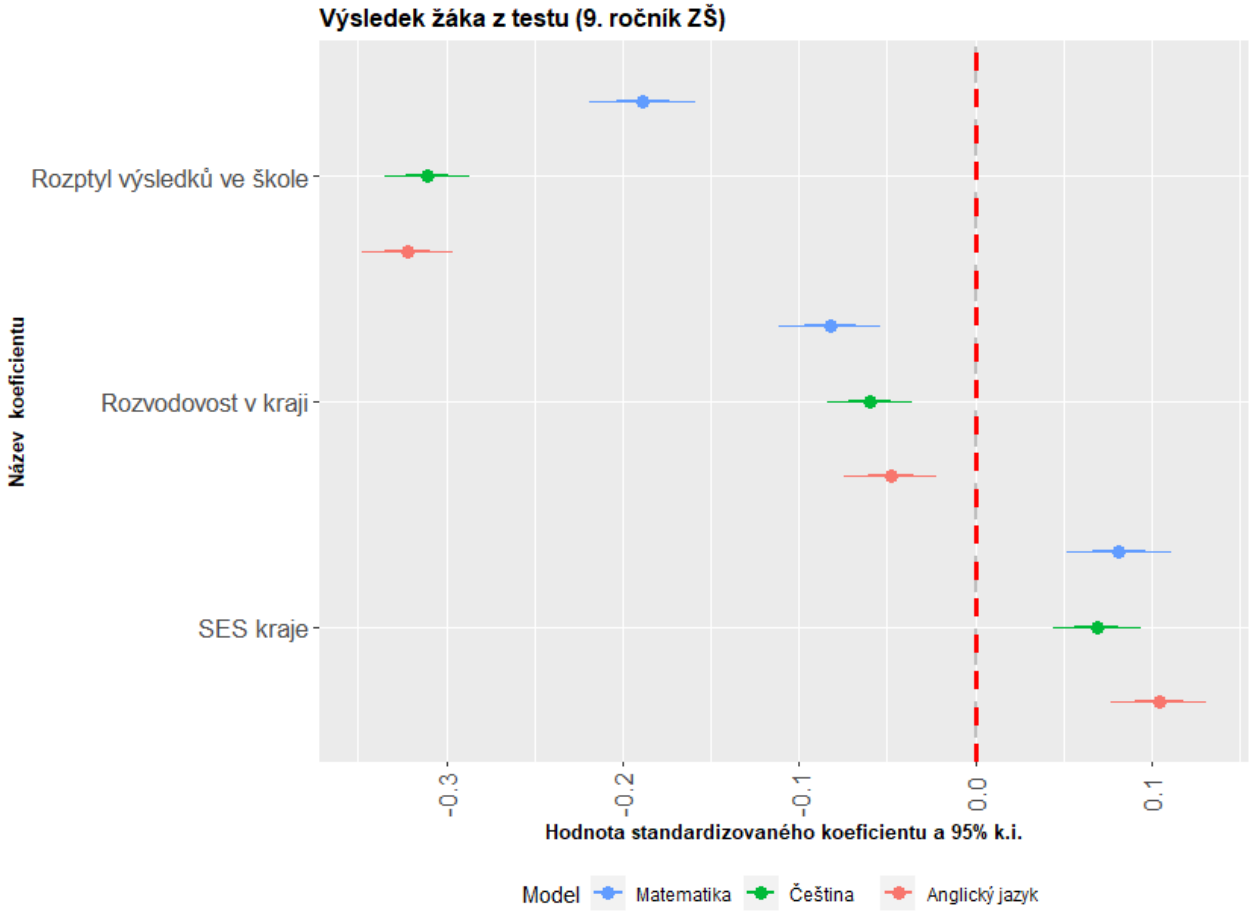
***p < 0.001, **p < 0.01, *p < 0.05

Poznámka: Balíček lmer a texreg v R. Viz Sada typizovaných analytických nástrojů.

Proměnné vstupují do modelu standardizované, standardizované jsou i výsledky žáků. V tabulce jsou uvedeny rovnou tři modely – pro každou oblast testování zvlášť. Protože jsou proměnné standardizovány, můžeme přímo koeficienty porovnat napříč modely. Z tabulky lze vyčíst, že rozptyl výsledků ve škole je silněji asociován (negativně) s výsledky z anglického a českého jazyka než z matematiky. U těchto oblastí má heterogenita výsledků větší vliv (ve smyslu asociačního vztahu, ne příčinného) než v případě matematiky. Interpretace oproti klasické lineární regresi se příliš neliší. Rozdíl je pak v hodnocení modelu, kdy nás zajímá zlepšení oproti tzv. nulovému modelu v případě kritérií AIC a BIC. Důležitější je ale hodnocení samotného efektu proměnných.

Regresní modely lze prezentovat buď formou tabulek, nebo formou grafů regresních koeficientů. Pokud je hlavním příjemcem zprávy či textu veřejnost, je výsledky statistických modelů vhodné prezentovat formou grafu regresních koeficientů. Tato forma je uživatelsky přívětivá pro čtenáře těchto zpráv, kteří nedisponují znalostmi složitých statistických analýz, zejména hierarchického modelování. Tato forma je vhodnější než souhrnný popis analýzy ve formě tabulky či textu. V nejnovějších tematických zprávách se již takto postupuje (viz *Podpora rozvoje osobnostních a komunikačních dovedností žáků středních škol ve školním roce 2019/2020*). Za zvážení stojí další úprava, kdy lze koeficienty pro srovnatelnost škálových a indikátorových proměnných (typ/druh školy) použít standardizaci dle Gelmanovy metody. Standardizace je provedena do dvou směrodatných odchylek (Gelman 2008). To má za následek, že hodnota koeficientu zhruba odpovídá změně z minimální hodnoty na maximální, podobně jako u indikátorové proměnné. Indikátorové proměnné vstupují do modelu nestandardizovány, zatímco škálové proměnné (a případně ordinální) vstupují do modelu standardizované na základě dvou směrodatných odchylek. Jedná se pouze o vizualizaci koeficientů.

GRAF 7 | Příklad regresní analýzy výsledků testů z matematiky, češtiny a anglického jazyka – 9. ročník



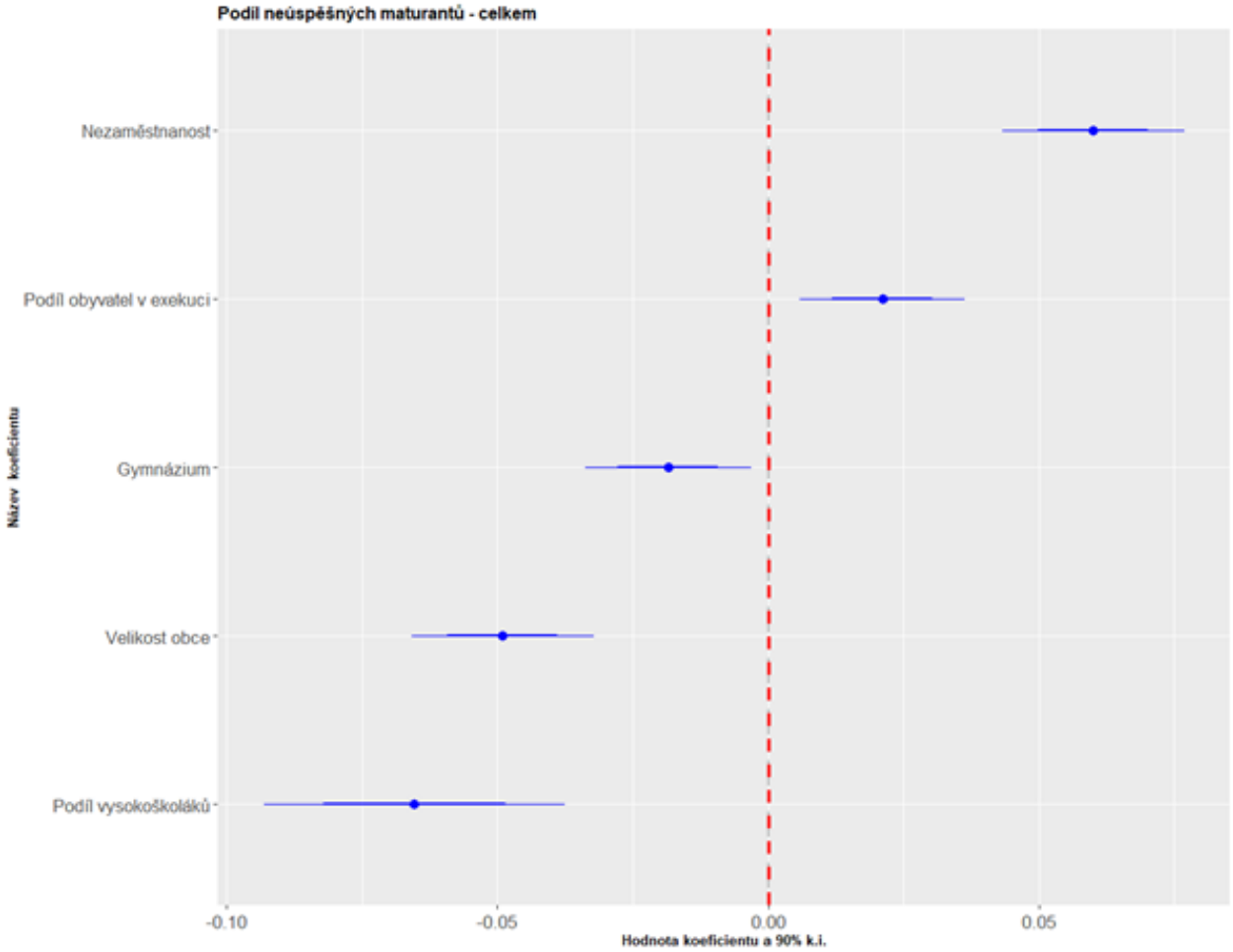
Poznámka: Balíček coefplot v R. Viz Sada typizovaných analytických nástrojů.

Doporučení

Datový soubor napárovat s dalšími daty na úrovni obcí a krajů. Možnost propojování dat jako podmínka pro další statistické modelování.

Pro ilustraci použití nelineárních modelů je příklad závisle proměnné podíl neúspěšných maturantů. Jak bylo ukázáno v předchozích kapitolách, proměnná je velice zešikmená. Její distribuce se vůbec neblíží normálnímu rozdělení. Vhodným modelem pro tento typ je Poissonova regrese. Rozdíl oproti lineární regresi je v interpretaci koeficientů, které jsou v logaritmické škále, nicméně se interpretují jako procentuální navýšení počtu. Zvýší-li se nezaměstnanost v kraji o jednu jednotku (v našem případě o 1 %), zvýší se podíl neúspěšných maturantů v krajích v průměru o 6 % (procentních bodů, hodnota koeficientu 0,06).

GRAF 8 | Model vysvětlující podíl neúspěšných maturantů z roku 2018



Pozn. Jedná se o Poissonovu regresi pro číselná data s velkou disperzí, vytvořeno v jazyku R. Podíl osob v exekuci, zdroj www.mapaexekuci.cz. CERMAT, ČŠI, ČSÚ.

Regresní modely jsou patrně nejvyužívanější metodou. Pro jejich plnohodnotné využití je nutné mít vytvořeny datové soubory, které by obsahovaly celou řadu proměnných na různých úrovních. Zpravidla žák, třída, škola, okres, kraj. To, jakou úroveň budeme modelovat, záleží na účelu výzkumu.

3.5.5 Popisná a kauzální interpretace regresních koeficientů

Přestože se v mnohých odborných článcích a (bohužel) i učebnicových textech často setkáváme s interpretací regresních koeficientů, která obecně říká *„…pokud zvýšíme hodnotu nezávisle proměnné o jednu jednotku, zvýší se tím hodnota závisle proměnné o příslušnou hodnotu koeficientu, a to při kontrole ostatních proměnných“*, je potřeba mít na paměti, že taková interpretace má tzv. kontrafaktuální podobu související s *potential outcome* rámcem (co by stalo, kdyby…) a odráží příčinnou (kauzální) souvislost mezi nezávisle a závisle proměnnou. V podstatě nám říká, jak velký efekt očekáváme, že nastane s hodnotou ZP při manipulaci s hodnotou NP.

Taková interpretace regresního koeficientu je však možná pouze při splnění velmi striktních podmínek a klade velmi vysoké nároky na kvalitu získaných dat a zejména na proces, který stál za generováním těchto dat. V praxi může být její použití odůvodnitelné pouze vzácně, a to zejména ve vazbě na předchozí realizaci randomizovaného experimentu (RCT), kdy provádíme skutečnou intervenci (tedy manipulujeme s hodnotou nezávisle proměnné) a máme k dispozici dvě (či více) optimálně vybalancované skupiny případů. Vybalancování v tomto smyslu odkazuje na to, že jsou hodnoty všech relevantních proměnných, které by vztah mezi ZP a NP mohly ovlivnit, mezi dvěma porovnávanými skupinami v průměru identické. Navíc musíme vyloučit, že v průběhu takového experimentu nedošlo k dalším potenciálním problémům (viz Gelman et al. 2020: 363–379).

Za jistých podmínek lze kauzální interpretaci přisoudit i regresním koeficientům, které vycházejí z dat, v jejichž rámci nejsou dvě intervenované skupiny dostatečně vybalancovány, a tedy je velmi pravděpodobné, že vliv intervence (změněná hodnota NP) je silně ovlivněna i dalšími třetími proměnnými (uvažuje se tzv. samovýběr, *selfselection bias*). V takových případech je nutné do regresního modelu zahrnout všechny relevantní proměnné, které by tento vztah

mohly ovlivnit (snažíme se o dosažení tzv. *no ommited variable bias*). Pro kauzální interpretace je samozřejmě nutná i správná specifikace modelu (viz Gelman et al. 2020: kap. 20).

Pointou je, že mnohdy taková data v potřebné kvalitě nemáme k dispozici. Často pracujeme s výběrovým šetřením a analyzované NP nepředstavují randomizovanou intervenci. Například při analýze dat z mezinárodních šetření (PISA, TIMSS apod.) není ani možné splnit základní předpoklad temporality, tedy že intervence v čase předchází zvýšení/snížení hodnot ZP. V těchto případech je sice stále možné výsledky regresního modelu smysluplně interpretovat, je však nutné zjištěním nepřisuzovat kauzální charakter.

Gelman et al. (2020: 134) doporučuje vycházet z popisné, resp. *predictive interpretation* regresních koeficientů. Ta nám umožní vyjádřit, *„…nakolik se hodnota ZP liší v průměru při komparaci dvou skupin případů, jež se mezi sebou liší o jednu jednotku v námi zkoumané NP, zatímco jsou identické s ohledem na všechny ostatní v modelu zahrnuté třetí proměnné"*. Přičemž musíme dbát na to, aby interpretace vždy odkazovala na změnu hodnoty ZP mezi případy, nikoliv v rámci případu, a tak by měla být vždy doprovázena slovem „v průměru".[3]

3.6 Další vybrané techniky

3.6.1 Shluková analýza

Kapitola 7.1.2 v Sadě typizovaných analytických nástrojů

Shluková analýza je vícerozměrnou explorační metodou – podobně jako například dále řešená explorační faktorová analýza. Faktorovou analýzu ale obvykle používáme k hledání podobných proměnných, kdežto shlukovou analýzu využíváme spíše při hledání podobných případů (ačkoliv využitelná je pro obě situace). Obecně je smyslem shlukové analýzy klasifikace objektů či jinak řečeno seskupování či shlukování podobných objektů do skupin. S jejím využitím se proto mnohdy můžeme setkat při utváření typologií.

Prakticky shluková analýza postupuje tak, že prozkoumává datový soubor a zjišťuje, zda je možné vytvářet v něm určité nižší počty shluků či skupin. Případy jsou sdružovány tak, aby si v rámci dané skupiny byly co nejpodobnější a zároveň se co nejvíce lišily od případů v jiných skupinách (Everitt et al. 2011). Způsobů, jakými lze utvářet shluky, nicméně existuje velmi mnoho. Při rozhodování obvykle postupujeme podle dat, s nimiž pracujeme.

Poměrně často využívaným přístupem je *hierarchické shlukování*. Hierarchická shlukovací analýza konkrétně může postupovat dvěma způsoby, na základě měření vzdáleností mezi objekty (i tady existuje více způsobů) jsou objekty buď postupně spojovány, až vytvoří jeden shluk, nebo jsou postupně rozpojovány, až každý představuje shluk samostatný. V tomto procesu a jeho posuzováním dochází k identifikaci situace, která je ideální, tedy k vytvoření optimálního počtu skupin. U *K-Means shlukování* zase dopředu určujeme požadovaný počet skupin a výpočet postupuje tak, že jsou objekty náhodně rozděleny do skupin a následně přeřazovány podle blízkosti k průměru skupin, což se děje, dokud se průměry mění. Existuje nicméně celá řada různých postupů a jejich variant.

Při shlukování tedy dochází k seskupování objektů či případů do skupin na základě sady zvolených proměnných. Jistou nevýhodu zde však představuje, že není možné určit, jestli byly proměnné, na základě nichž ke shlukování dochází, zvoleny vhodně. Rozhodnutí o optimálním řešení navíc může být poměrně volné a do značné míry v rukou samotného výzkumníka, přestože existují ukazatele vhodnosti provedení, které v obecné rovině zachycují, zda shluky opravdu odrážejí struktury v datech, či jsou spíše uměle vytvořeny. Stojí za zmínku, že kromě shlukování existují i další podobné techniky. Jednou z nich je například *mnohorozměrné škálování*, jehož výstupem je grafické zobrazení vzdálenosti mezi případy.

3.6.2 Faktorová analýza

Kapitola 7.7.1 v Sadě typizovaných analytických nástrojů

Faktorová analýza se rovněž řadí mezi vícerozměrné metody. Jejím smyslem je nahradit sadu vzájemně spjatých proměnných malým počtem ne přímo pozorovaných znaků, kterým se říká latentní proměnné či jinak faktory. Existuje explorační a konfirmační verze faktorové analýzy. *Explorační faktorovou analýzu* využíváme v situaci, kdy máme soubor proměnných a snažíme se zjistit, zda za nimi stojí jeden, či více faktorů. Jak je tedy patrné rovněž z jejího názvu, jedná se o metodu průzkumnou, využitelnou k formulování hypotéz. Oproti tomu *konfirmační faktorová analýza* neslouží k prozkoumání dat, ale k ověření již existujících předpokladů ohledně struktur v datech. Nepoužíváme ji tak na začátku analýzy k tomu, abychom data prozkoumali a zjistili, zda se v nich nacházejí nějaké faktory, ale

[3] „When comparing two children whose mothers have the same level of education, the child whose mother is x IQ points higher is predicted to have a test score that is 6x higher, on average." Or, „Comparing two items i and j that differ by an amount x on predictor k but are identical on all other predictors, the predicted difference yi − yj is βk x, on average." (Gelman et al. 2020: 134).

naopak tehdy, kdy už existenci faktorů předpokládáme a naše předpoklady chceme ověřit. Jinak řečeno konfirmační faktorovou analýzu využíváme k ověření hypotéz.

Provedení faktorové analýzy je založeno na výpočtu korelací mezi proměnnými. Běžně je za tímto účelem nastaveno využití Pearsonova korelačního koeficientu. Stále častěji však bývá upozorňováno na to, že využití Pearsonova korelačního koeficientu nemusí být vždy vhodné, zejména s ohledem na situace, kdy nepracujeme s kardinálními proměnnými. Taková situace přitom může být v sociálněvědním výzkumu poměrně častá, neboť celá řada proměnných je ordinálních či dichotomických. V takovém případě proto bývá doporučováno pracovat s jinými korelačními koeficienty – u ordinálních proměnných se konkrétně doporučuje využití polychorických korelací, u dichotomických proměnných poté využití korelací tetrachorických (Holgado-Tello et al 2010, Soukup 2021). Využití těchto korelací při výpočtu faktorové analýzy obvykle řešení vyjasní a zesílí hodnoty faktorových zátěží. Prakticky ovšem není vždy snadné změnit před provedením faktorové analýzy korelační koeficient. Mnohdy je potřeba nejdřív vypočítat korelační matici a následně provést faktorovou analýzu postavenou na této matici. Podobné nastavení ovšem např. základní verze SPSS neumožňuje, musíme proto pracovat v kombinaci v R, případně v některém z dalších softwarů.

Další otázkou spjatou s faktorovu analýzou je metoda extrakce faktorů. Dvěma základními postupy jsou metoda faktorové analýzy a metoda hlavních komponent. Rozdíl mezi nimi spočívá v tom, že se v prvním případě snažíme vysvětlit vztahy mezi proměnnými, kdežto ve druhém vysvětlujeme rozptyl proměnných. Například SPSS má v základu nastavenou metodu hlavních komponent. Pokud bychom chtěli využít první z možností, museli bychom tak učinit v nastavení. Neplatí to ale pro všechny softwary, neboť kupříkladu Stata automaticky pracuje s první zmíněnou možností, a proto je v tomto případě naopak nezbytné určit, pokud chceme pracovat s metodou hlavních komponent. Volba metody by přitom měla vycházet z konkrétní řešené situace.

Často diskutované téma představuje rovněž provedení rotace faktorů. Tuto otázku na rozdíl od předchozích zmíněných neřešíme při samotném provádění faktorové analýzy, ale tehdy, kdy už máme výsledek v podobě faktorových zátěží (tedy korelací mezi vstupními položkami a faktorem) a snažíme se jej interpretovat. Na řadu rotování přichází konkrétně tehdy, kdy jsou u některých položek faktorové zátěže silné ve dvou či více faktorech (jinak řečeno položka sytí více faktorů). Cílem provedení rotace potom je přimknout položky pouze k jednomu z faktorů. Způsob rotace závisí na konkrétní situaci. Pokud předpokládáme, že spolu faktory souvisejí, používá se rotace šikmá (např. oblimin), pokud naopak předpokládáme, že spolu nesouvisejí, používáme rotaci kolmou (např. varimax). Pokud se řešení provedením rotace vyjasní, tedy žádná z proměnných není silná ve více faktorech, a zároveň jsou faktorové zátěže dostatečně silné (u větších souborů alespoň 0,3), můžeme uložit faktorová skóre jako novou proměnnou do datového souboru a v analýzách s nimi dále pracovat.

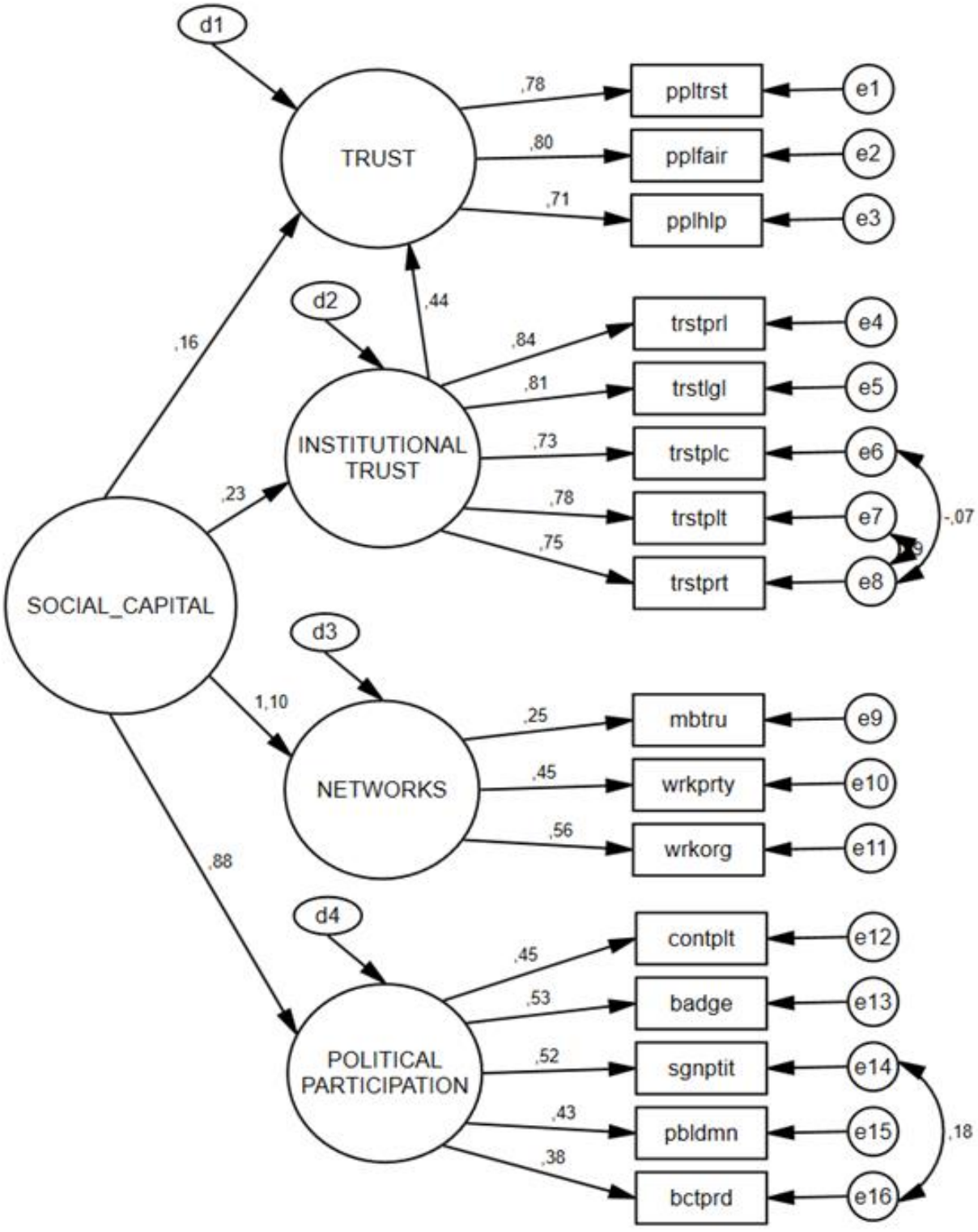
3.6.3 Structural equation modelling (SEM)

Kapitola 8 v Sadě typizovaných analytických nástrojů

Strukturní modelování je technika, která kombinuje faktorovou analýzu, regresi a úsekovou analýzu. Základním modelem je tzv. measurement model, kdy pomocí více proměnných měříme to, jestli dané proměnné měří danou latentní proměnnou („faktor“). SEM je využívána tymem PISA a OECD pro konstrukci indexů a pro testování tzv. measurement equivalence, tedy konceptu ekvivalence měření, kdy se snažíme zjistit to, že latentní koncept je napříč skupinami respondentů chápán stejně. Pokud by tomu tak nebylo, znamenalo by to, že položky dotazníků vhodně neměří daný koncept.

SEM analýza se provádí v programu AMOS, který běží na enginu IBM SPSS. Přestože se jedná o samostatný program, je součástí softwaru IBM. SEM analýzu lze provést i v programu Stata a v R pomocí balíčku „lavaan“. V prvním kroku vždy určíme, jaké proměnné (či položky dotazníku) měří daný latentní koncept, který v programech nakreslíme pomocí kruhu. Jako obdélník se standardně označují přímo měřené proměnné, které jsou v datasetu. K těmto proměnným program automaticky vytvoří chyby měření *e*. Schéma 7 níže pak ukazuje příklad konfirmační faktorové analýzy druhého řádu, kdy nadřazený latentní koncept, sociální kapitál, je měřen pomocí dalších čtyř latentních proměnných. Model tak testuje, jestli teorie sociálního kapitálu a jeho dimenze empiricky existují ve společnosti a jsou tak měřitelné.

SCHÉMA 7 | Konfirmační faktorová analýza druhého řádu (Second-order CFA model); Data ESS čtvrtá vlna 2008



Poznámka: Indikátory z datasetu jsou v obdélnících s původními názvy z datasetu ESS (European Social Survey), ty jsou seřazeny dle tabulky č. 5. Vytvořeno v programu AMOS. CFI=0,952, RMSEA=0,053.

SEM analýzy jsou velmi komplexní, zejména pokud kombinujeme faktorové analýzy s regresními modely. Můžeme tak testovat moderační a modifikující efekty proměnných. Pomocí úsekové analýzy můžeme testovat, prostřednictvím jakých cest spolu proměnné souvisí. Můžeme také analyzovat, jestli některé proměnné nepůsobí skrze další proměnné v teoretickém kauzálním řetězci. Zde je nutné upozornit, že vzhledem k povaze dat dotazníkových šetření ani SEM analýza nedokáže potvrdit či vyvrátit kauzální vztahy. Blíže k problematice SEM kniha autorky Barbary Byrne (2016)

a manuál k balíčku R lavaan. Kód pro tvorbu SEM v programu R a interpretace výstupů analýzy je pak v příslušné kapitole Sady typizovaných analytických nástrojů.

3.7 Pokročilé statistické techniky pro výzkum kauzality

Přestože pracujeme s daty, která neumožňují kauzální či jinými slovy příčinnou interpretaci nalezených vztahů, ale pouze interpretaci ve formě asociací a korelací, existují výzkumné designy a statistická modelování, které se snaží přiblížit experimentálnímu výzkumnému designu.

3.7.1 Matching methods

Kapitola 9 v Sadě typizovaných analytických nástrojů

První metodou jsou „matching methods" (dále MM) algoritmy. Obvykle máme nezávisle proměnnou (v případě těchto metod mluvíme o „treatment variable"), která nabývá kategorických hodnot, ideálně se jedná o dichotomickou proměnnou. Modelovým příkladem je zjištění, jestli používání alternativních učebnic matematiky má kauzální efekt na výsledky žáků. Při běžné regresi z průřezových dat i při kontrole dalších proměnných nejsme schopni kontrolovat všechny třetí proměnné, které mohou způsobit falešnou korelaci. Navíc je zde závažný problém endogenity. Tento metodologicky složitý jev lze vysvětlit tak, že nemůžeme vyloučit, že školy, které si vybraly alternativní učebnice matematiky, nemají nějaké vlastnosti, které predikují zvolení těchto metod. Tyto vlastnosti mohou ale současně působit na naši závisle proměnnou, kterou je úspěšnost žáků v testování. Cílem MM je porovnat skupinu škol používajících alternativní učebnice se skupinou škol, které nepoužívají alternativní učebnice (kontrolní skupina), ale které jsou si ideálně ve všem podobné s první testovací skupinou škol. Důležité je proto vybrat co nejvíce vlastností, dle kterých se do kontrolní skupiny škol vyberou pouze školy, které jsou co nejpodobnější školám používajícím alternativní učebnice. Algoritmus pak vybírá ke každé škole používající alternativní učebnice kontrolní školu, která je nepoužívá, ale která je co nejvíce podobná dané škole.

Na výběr máme několik algoritmů pro výběr kontrolní skupiny, které tak simulují experimentální výzkumný design (pozor, nejedná se ale o experiment). „Exact matching", který vybere identické školy, je základní metodou, jež je ale v praxi nereálná, zejména pokud máme několik prediktorů. Nejpoužívanějším algoritmem je „propensity score matching", který na základě hodnot z probitového modelu vybere ty školy, které jsou si nejpodobnější. Algoritmů je více a odkazujeme na odbornou literaturu a na autora softwaru Matchit Garyho Kinga. Dalším metodologickým krokem je zvolení, jestli daná kontrolní škola bude přiřazena jen té nejpodobnější škole, nebo bude přiřazena vícero školám v „experimentální" skupině. Obecně se doporučuje druhá možnost, abychom zajistili, že ke každé škole se napáruje co nejpodobnější.

Příkladem použití metody jsou data TIMSS 2015 a data ze šetření ČŠI (2014 INEZ_Alternativní metody výuky), která byla napárována pro potřeby Sekundární analýzy TIMSS 2015. Prezentován je výstup z programu Stata, který tento typ metod nově obsahuje. Jako proměnné, dle kterých se počítala podobnost škol, byly vybrány: SES školy, třídní klima školy, oblíbenost výuky matematiky (agregace za školu), velikost obce, ve které se škola nachází, velikost školy, velikost třídy. V tabulce X je ukázáno, kolik případů (pozor: úroveň je žák, ale proměnné jsou na úrovni školy, tudíž pro každý případ žáka dané školy je zde konstantní hodnota), kde školy používají alternativní učebnice (testovací skupina), je napárováno s kontrolními případy, kde školy nepoužívají učebnice, ale jsou si co nejvíce podobné. Jak je vidět z tabulky, některé školy (*Treated,* cca 5 škol) nebyly napárovány s kontrolní skupinou.

TABULKA 8 | Výstup z programu R | spárované případy

Matching statistics

	Matched			Controls			Band-
	Yes	No	Total	Used	Unused	Total	width
Treated	2765	152	2917	3101	0	3101	.02276
Untreated	3021	80	3101	2568	349	2917	.00997
Combined	5786	232	6018	5669	349	6018	.

Výstupem analýzy je pak tabulka, která ukazuje průměrný efekt dané proměnné na výsledky žáků z testů z matematické gramotnosti TIMSS 2015. V tabulce je kromě alternativních učebnic více klíčových proměnných, jejichž vliv chceme měřit. Hodnota „kauzálního efektu" dané proměnné nám říká, o kolik se v průměru zvýší dosažené skóre z testu matematické gramotnosti (TIMSS 2015, průměr 530 bodů, min. 244 bodů, max. 753 bodů, směrodatná odchylka 70 bodů). V případě učebnice se jedná o 9,29 bodu. Tabulka obsahuje i konfidenční interval. V populaci tak skutečný efekt bude variovat zhruba v rozmezí 4,55 až 14 bodů. Tento efekt je očištěn od vlivu dalších třetích proměnných, protože obě skupiny případů (testovací a kontrolní) jsou shodné v průměrech u proměnných, podle kterých jsme školy

srovnávali. Samozřejmě protože nebyl proveden zcela náhodný výběr škol do testovací a kontrolní skupiny, stále nejsou vyloučeny vlivy dalších neměřitelných a endogenních faktorů. Přesto jsou výsledky těchto metod pro zhodnocení teoretických kauzálních předpokladů vhodnější než standardní statistické metody na průřezových datech.

TABULKA 9 | Výstup z programu R – tabulka výsledků (upraveno)

Propensity score matching

Proměnná	Kauzální efekt*	Konfidenční interval 95 %	
Alternativní učebnice	9.29	4.55	14.03
Budování schémat	4.25	-1.72	10.23
Činnostní učení	4.68	0.41	8.96
Využívá metody	8.58	4.15	13.01

**ATT (Average Treatment Effect for the Treated)*

MM analýzy mají za cíl simulovat experiment z průřezových dat, nicméně jedná se pořád o výzkumný design, který není schopný experiment nahradit. Přesto výsledky analýz umožňují interpretovat závěry silněji ve vztahu k teoretickému kauzálnímu efektu než jen čistě na základě korelační či regresní analýzy.

Doporučení

Při analýze kategorických nezávisle proměnných zkusit použít MM a vybrat kontrolní vzorek škol na základě co nejvíce charakteristik na úrovni školy.

Vytvoříme tak datový soubor, který bude obsahovat „experimentální" a „kontrolní" skupinu případů. Na takovém datasetu provedeme jednoduchý t-test, popřípadě regresní modelování. Statistický software automaticky zobrazí hodnotu „kauzálního efektu", buď jako rozdíl v průměru, nebo jako regresní koeficient.

3.7.2 Metoda difference-to-difference

Dalším pokročilým výzkumným designem a statistickou metodou je *difference-to-difference* (DiD) metoda. DiD měří kauzální efekt pomocí rozdílu v čase. Pro tento typ analýzy je potřeba mít data za dva časové úseky (t a t_{+1}). Naše klíčová nezávislá proměnná je ideálně dichotomická, popřípadě kategorická. Potřebuje tak mít například skupinu škol nebo tříd, které zavedly novou výukovou metodu, a skupinu tříd a škol, které ji nezavedly. V čase se srovnají výsledky škol první skupiny s kontrolní skupinou. Výsledný rozdíl je pak kauzální efekt zavedení nové výukové metody. Testovat tímto způsobem lze celá řada opatření, nutnou podmínkou jsou data před zavedením daného opatření a po zavedení daného opatření. A také to, že dané opatření nebude pro všechny školy či třídy plošné, tudíž bude mít konstantní hodnotu. Jedná se o výzkumný design ideální pro případy, kdy stát uvolní nějakou politiku a nechá její zavedení v diskreci škol. Některé školy danou politiku zavedou, jiné nikoliv. Z tohoto důvodu by měly být nachystány sběry dat před a po intervenci tak, aby tato metoda mohla být ČŠI využita. Výsledky těchto metod mají silné kauzální závěry.

Doporučení

Identifikace předmětu a tématu šetření, kdy je možnost sběru dat ve dvou časových okamžicích pro dvě a více skupin škol. Závěry při použití této metody budou vhodné pro ověřování kauzálních vztahů.

4

Vizualizace výsledků

4 VIZUALIZACE VÝSLEDKŮ

Pro techniky tvorby vizualizací viz příslušné kapitoly a metody Sady typizovaných analytických nástrojů

Podkapitola se bude zabývat problematikou prezentování a vizualizace dat. Budou představeny vybrané způsoby vizualizace dat. Konkrétně se bude jednat jak o některé typy grafů, které jsou pro prezentaci výsledků vzdělávání vhodné, tak o vizualizaci pomocí kartogramu. Některé pokročilé vizualizační nástroje, např. Gaussův model či regresní coefplot, jsou součástí paralelního dokumentu *Sada typizovaných analytických nástrojů*.

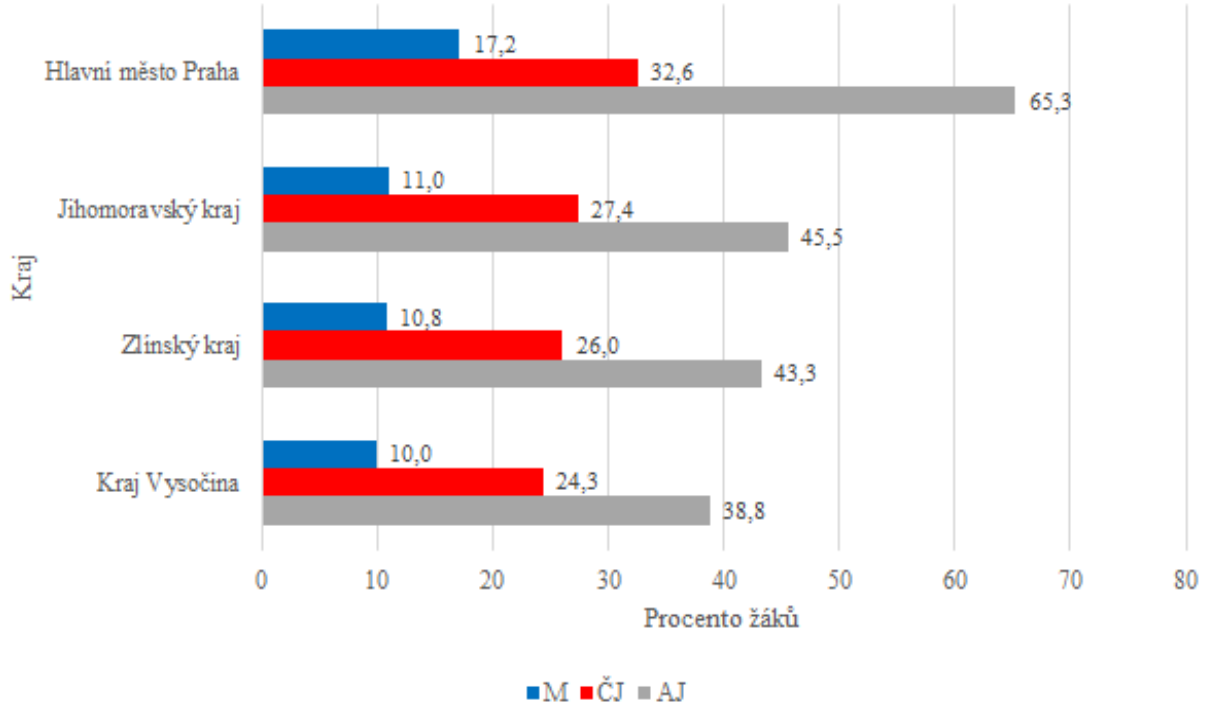
4.1 Vizualizace pomocí grafů

Grafy představují hlavní způsob vizualizace výsledků vzdělávání žáků. Pomocí grafů můžeme prezentovat popisné statistiky proměnných nebo vztahy mezi nimi formou, která usnadňuje orientaci a pochopení. Grafy proto tvoří důležitou součást všech výstupů.

Obecně představuje první krok tvorby grafů příprava dat. Podobně jako u všech analýz bychom měli nejdřív provést kontrolu dat, případně vybrat či připravit ta data, s kterými chceme pracovat. Při samotném výběru typu grafu je potřeba vycházet z informace, kterou se snažíme ukázat. Pro různé typy proměnných volíme odpovídající typy grafů (viz 1.2 Typy proměnných). Poslední fázi představuje úprava vzhledu grafu a doplnění náležitostí, jakými jsou například názvy os nebo popisky dat. Některým konkrétním náležitostem je dále věnována větší pozornost v rámci představení příkladů typů grafů vhodných pro vizualizaci vzdělávacích výsledků žáků.

Za účelem vizualizace výsledků vzdělávání žáků jsou běžně používány **sloupcové** nebo **pruhové grafy**. Tyto grafy jsou vhodné zejména pro srovnání hodnot určitých skupin. V případě analýzy výsledků vzdělávání se jedná například o porovnání průměrného skóre žáků nebo porovnání podílů žáků v různých gramotnostních kategoriích či podle jiného kritéria (neúspěšnost u maturit). Jako příklad porovnání hodnot různých kategorií může posloužit následující pruhový graf zobrazující podíly žáků 9. ročníku, kteří v národním testování dosáhli výborného výsledku (úspěšnost nad 80 %) v matematice, českém jazyce a anglickém jazyce v jednotlivých krajích České republiky.

GRAF 9 | Podíly žáků s výborným výsledkem (Výběrové zjišťování výsledků žáků 9. ročníku)



Poznámka: Za účelem ukázky uvedena data jen pro vybrané kraje.

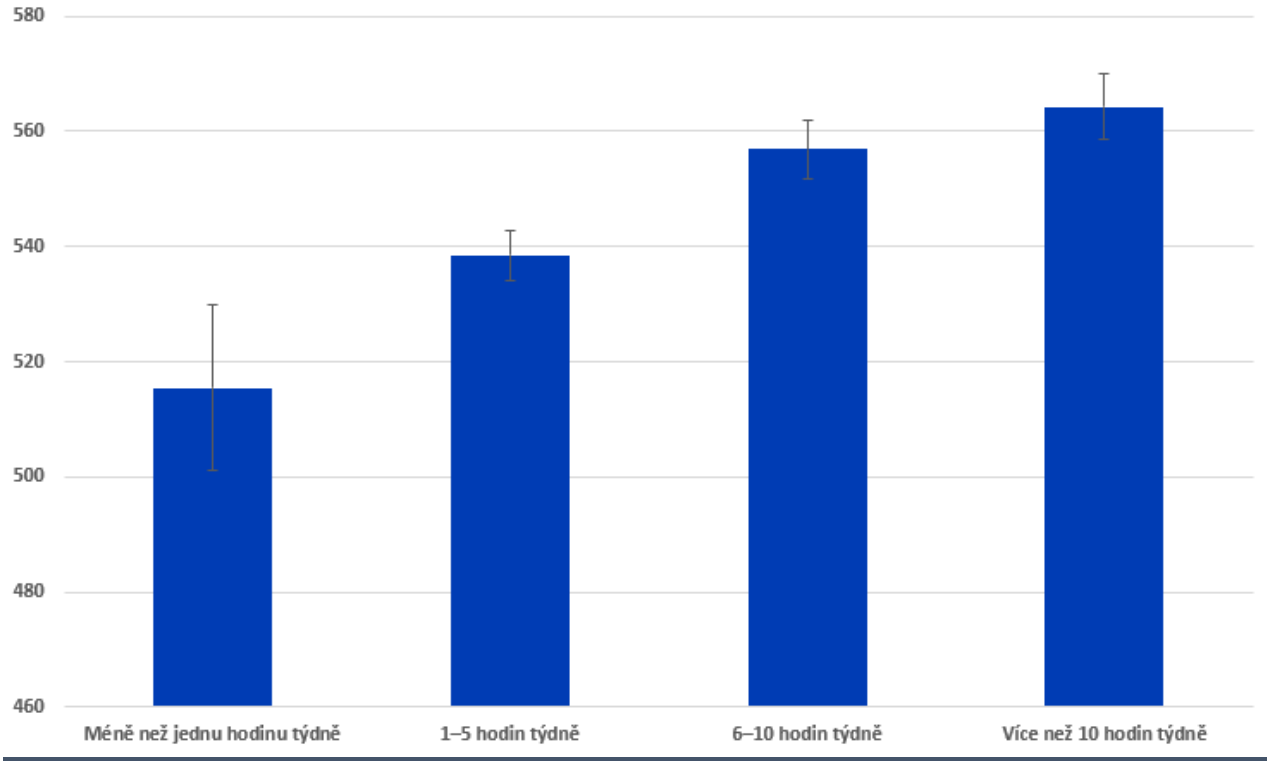
Velikost sloupce či délka pruhu v tomto typu grafu odpovídá množství dané jednotky. Na první pohled je tak z grafu patrné, která kategorie se vyznačuje vyšší či nižší hodnotou. Jelikož byly do grafu naneseny tři proměnné (za každý testovaný předmět), bylo potřeba určit, podle které hodnoty budou řazeny. V tomto případě bylo vybráno procentuální

zastoupení žáků dosahujících výborné úrovně v testu z matematiky, což však nemusí znamenat, že byli testovaní žáci daného kraje stejně úspěšní také ve dvou zbývajících předmětech. Na tuto skutečnost je potřeba při interpretaci grafu upozornit.

Abychom podobné porovnání na základě grafu nečinili zavádějící, je vhodné příliš nemanipulovat s osami. Rozdíl, který se mnohdy může zdát velký, ve skutečnosti může být jen minimální. S tím souvisí rovněž nutnost uvádět ke každé ose hodnoty (v tomto případě na ose x 0 až 80). Ke snazší orientaci v grafu celkově přispívá důkladný popis dat. Neměli bychom proto zapomínat ani na názvy kategorií a os. Z uvedených názvů by vždy mělo být hned patrné, co přesně zobrazujeme.

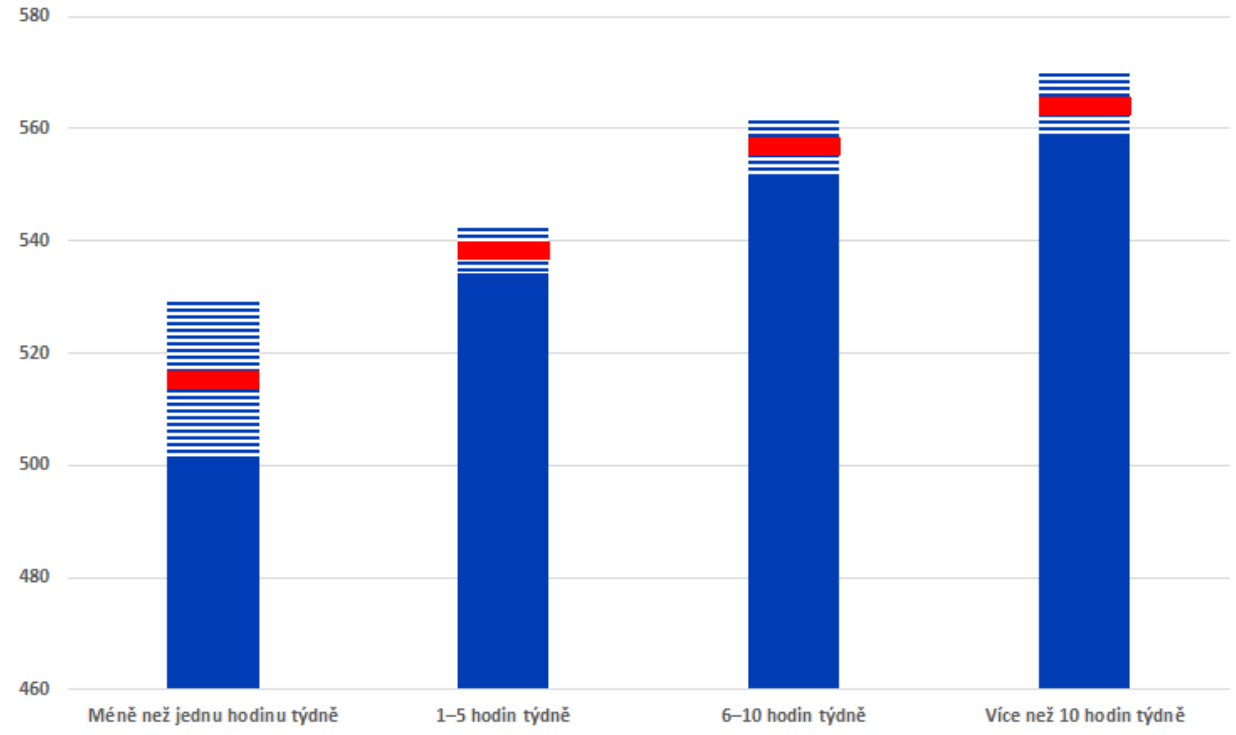
Pokud nedisponujeme daty za celou populaci, ale máme náhodně sesbíraný vzorek, který umožňuje zobecnění (například data z mezinárodních šetření), je vhodné do grafu vynést nejen hodnotu samou, ale také interval spolehlivosti. To lze učinit jednoduše tak, že ke sloupcům přidáme chybové úsečky, jak je provedeno v dalším grafu porovnávajícím průměrné skóre žáků podle toho, jak často jejich rodiče uvedli, že si čtou pro radost. Jedná se konkrétně o data z šetření PIRLS 2016.

GRAF 10 | Průměrné skóre žáků ze čtení podle četnosti četby rodičů pro radost (PIRLS 2016)



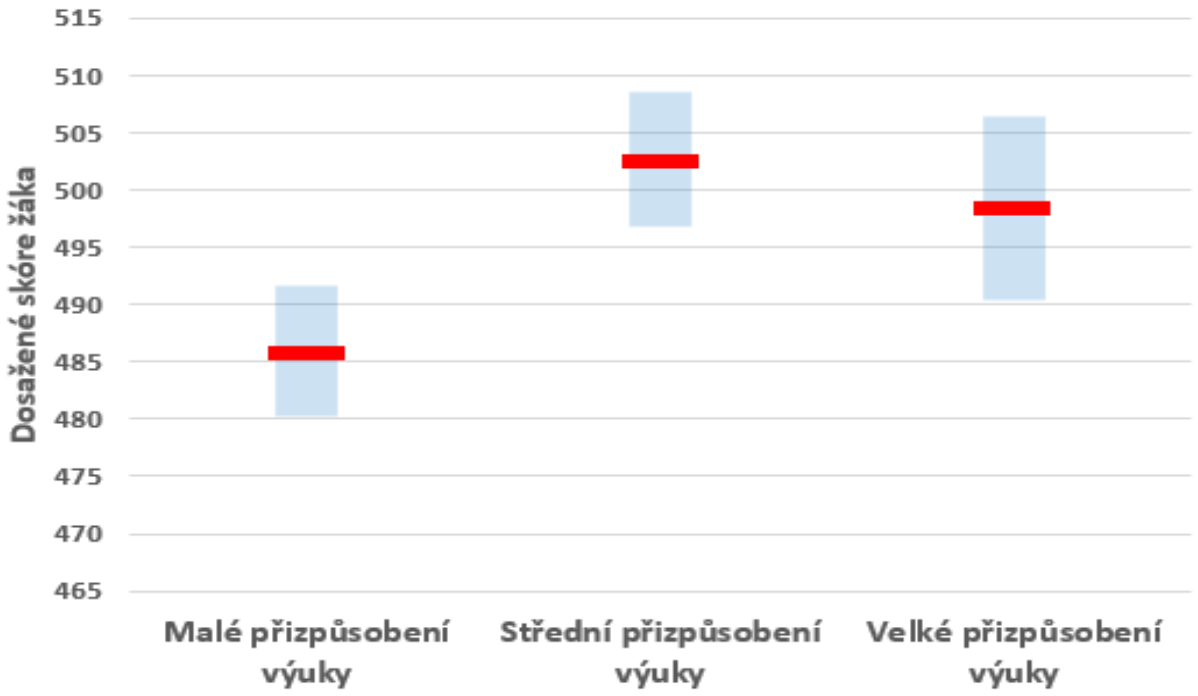
Následující graf demonstruje možnou inovaci vizuální úpravy grafu. Jedná se o totožný sloupcový graf, jaký byl vytvořen v předchozím případě, liší se ale právě způsobem zobrazení konfidenčního intervalu. Průměrná hodnota je v tomto případě značena červenou linkou a konfidenční interval zobrazuje pruhovaná oblast okolo.

GRAF 11 | Průměrné skóre žáků ze čtení podle četnosti četby rodičů pro radost (PIRLS 2016) – sloupcový graf s konfidenčními intervaly



Jiné alternativní zobrazení pracuje pouze s průměrnými hodnotami (červená linie) a samotnými konfidenčními intervaly. Toto zobrazení je přehlednější zejména v těch případech, kdy do grafu vykreslujeme více kategorií.

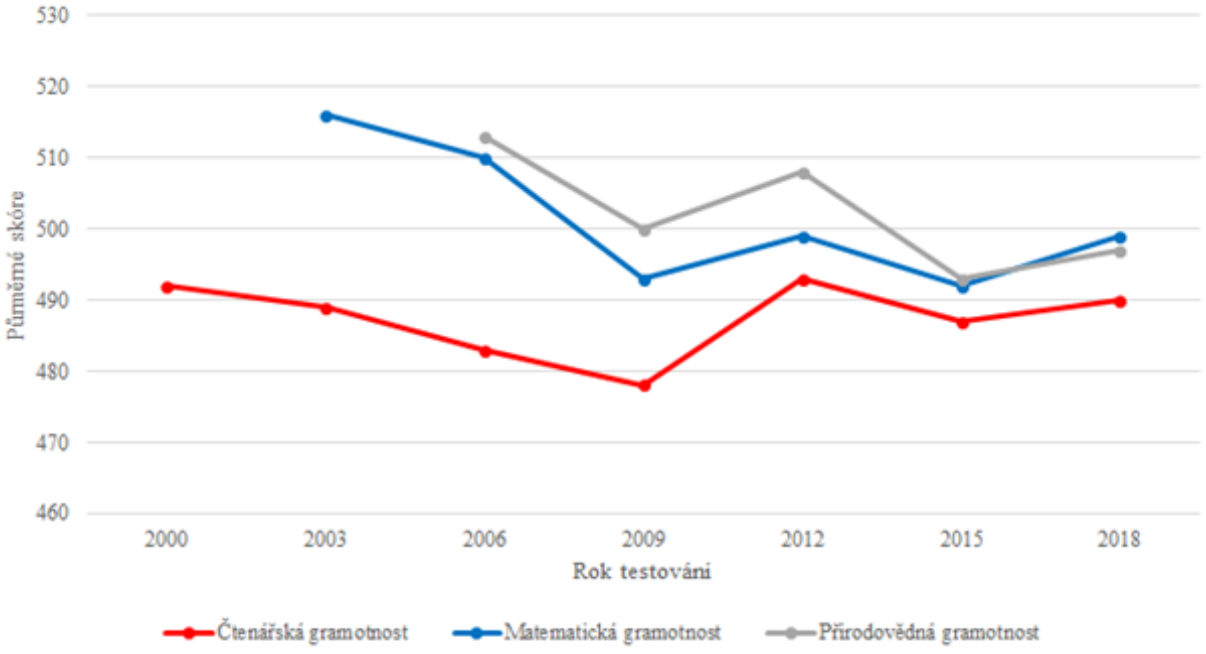
GRAF 12 | Průměrné skóre žáka dle míry přizpůsobování výuky ze strany učitele – zobrazení průměrných hodnot s konfidenčními intervaly



Pro zobrazení změn v průběhu času se hodí **spojnicový graf**. Využíváme ho například tehdy, kdy máme údaje o výsledcích vzdělávání za různé časové jednotky (cykly mezinárodního testování, výsledky maturit v letech apod.) a chceme zachytit jejich vývoj. Konkrétně jsou do grafu za každý příslušný čas naneseny hodnoty (průměrné skóre z testování, podíl neúspěšných maturantů), které se spojí, čímž se výrazně usnadní posouzení toho, jak se v čase mění.

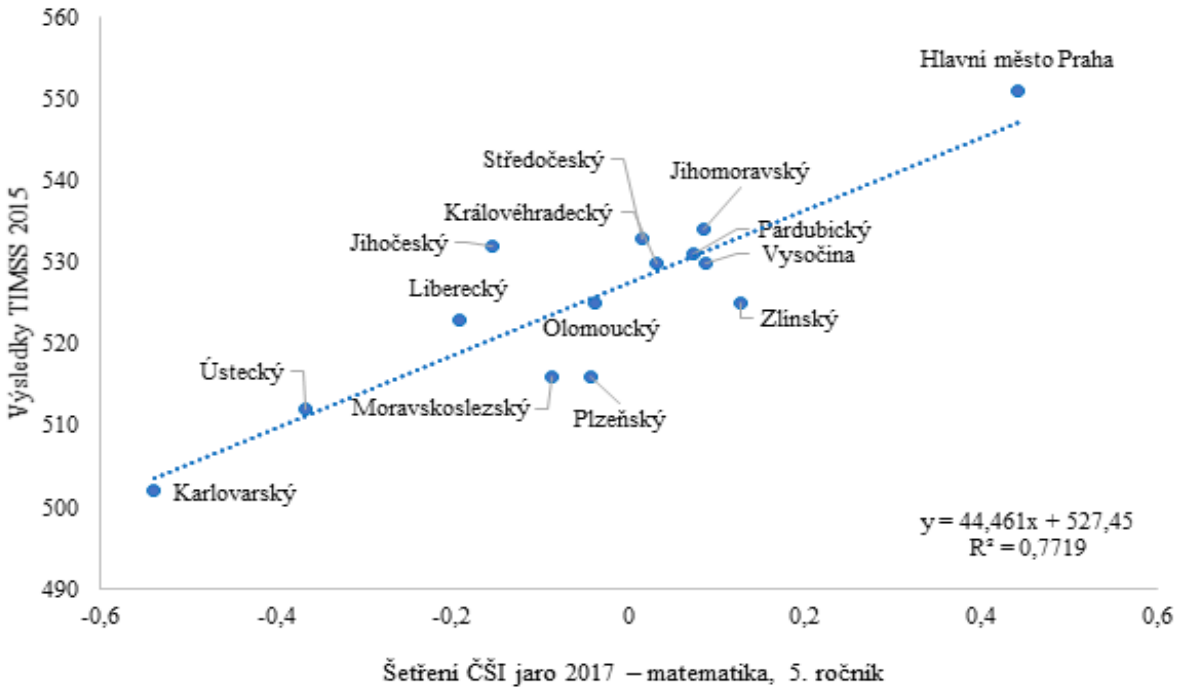
Příkladem může být následující spojnicový graf, který takto mapuje průměrné výsledky českých žáků v oblasti čtenářské, matematické a přírodovědné gramotnosti z mezinárodního šetření PISA za jednotlivé roky, v nichž dané šetření probíhalo.

GRAF 13 | Průměrné skóre žáků z testování čtenářské, matematické a přírodovědné gramotnosti v čase (PISA)



Grafem využívaným pro vizualizaci vztahu mezi dvěma škálovými proměnnými je **bodový graf**. Při analýze výsledku vzdělávání pomocí něj typicky zachycujeme vztah mezi dosaženým výsledkem a nejrůznějšími proměnnými. To samozřejmě můžeme činit na jakékoliv úrovni. Například následující bodový graf zobrazuje vztah mezi průměrnými výsledky ve dvou různých testováních, konkrétně se jedná o testování TIMSS a šetření ČŠI, v krajích.

GRAF 14 | Vztah mezi výsledky mezinárodního šetření TIMSS 2015 (matematická gramotnost) a Výběrového zjišťování výsledků žáků 5. a 9. tříd ZŠ (matematika) zobrazený bodovým grafem



Na horizontální osu je v daném grafu vyneseno průměrné skóre z matematiky z národního šetření žáků 5. ročníků a na vertikální průměrný výsledek z matematiky z šetření TIMSS. Body v grafu odrážejí pozici každého kraje s ohledem na jeho hodnoty. Do grafu byla navíc vložena lineární spojnice trendu, která vystihuje průběh dat. Z grafu je tak ihned patrný určitý pozitivní trend, tedy že kraje, které se vyznačují vyšší průměrnou hodnotou v národním šetření, mají vyšší průměrné skóre rovněž v mezinárodním testování. Do grafu je navíc vložena hodnota R^2, která značí, jak dobře se data přimykají k přímce, a tedy i to, jak silný vztah je.

Jak bylo zmíněno v textu dřív, vztahy mezi proměnnými jsou poměrně složité a za účelem vysvětlení výsledků vzdělávání je nutné pracovat s celou řadou nezávisle proměnných najednou, což už tímto způsobem vizualizovat nelze.

4.2 Vizualizace pomocí map

Některé výsledky šetření je vhodné vizualizovat pomocí GIS softwaru. Jedná se například o data z mezinárodních šetření, ale i domácích šetření. Mapy lze vytvořit v softwaru Esri ArcGis nebo v softwaru QGIS. Cílem je postihnout prostorovou dimenzi dat.

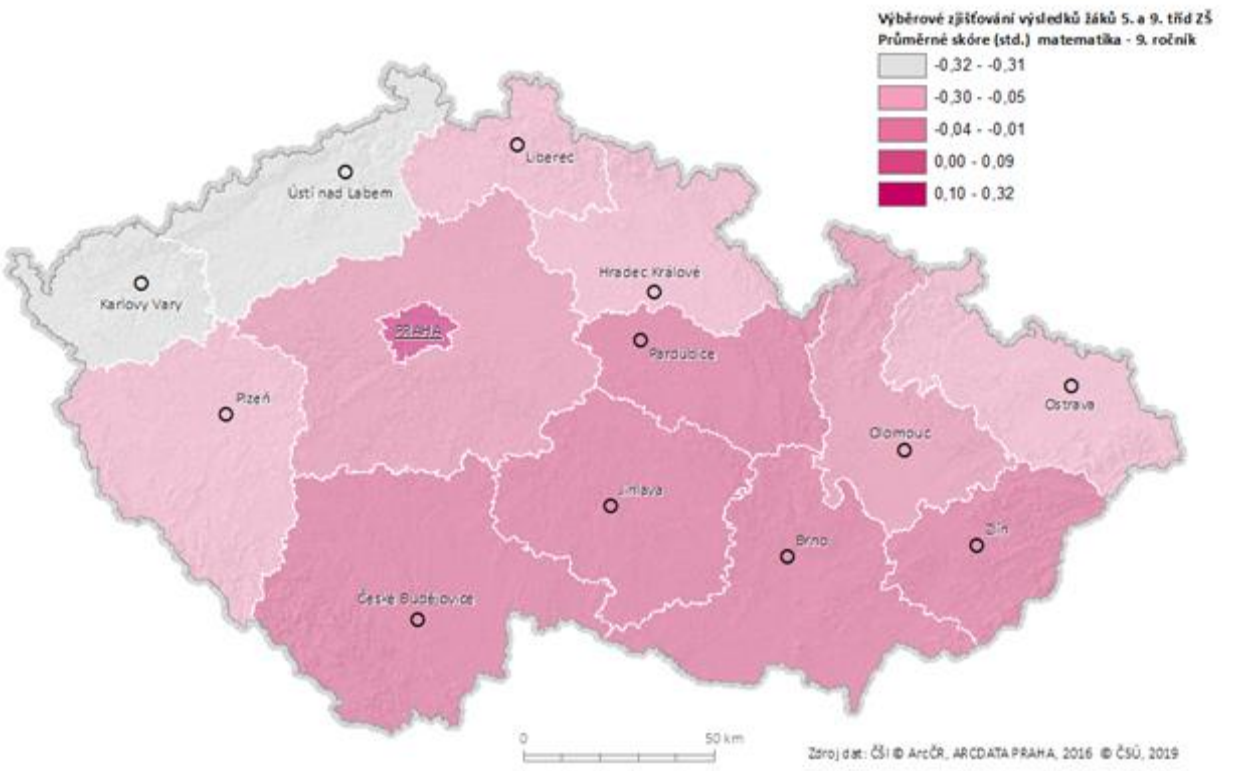
Mapu lze vytvořit v několika krocích:

1) přidání vrstvy krajů a vrstvy ČR, zvolení koordinačního systému,
2) přidání excelového souboru jako zdroje dat (ideálně formát .csv),
3) spárování mapové vrstvy s daty – dle číselníku krajů (identifikátor),
4) vizualizace dat.

U kroku 4 máme několik analytických možností, jak data vizualizovat.

Defaultní nastavení je „Natural breaks" – česky přirozené zlomy (Jenks). To je založeno na výpočtu přirozených zlomů v datech. Metoda nebere v potaz rozložení dat, ale přirozené zlomy v hodnotách, proto v každé z kategorií je zpravidla jiný počet případů. Přirozené zlomy jsou prvotní možností vizualizace dat v mapách, často dostačující. Příkladem je mapa zobrazující data z Výběrového zjišťování výsledků žáků 9. ročníku v matematice. Protože se jedná o škálovou proměnnou, je vhodné zobrazit výsledky pomocí přirozených zlomů a barevně odlišit 5 kategorií, což je defaultní nastavení většiny GIS softwarů.

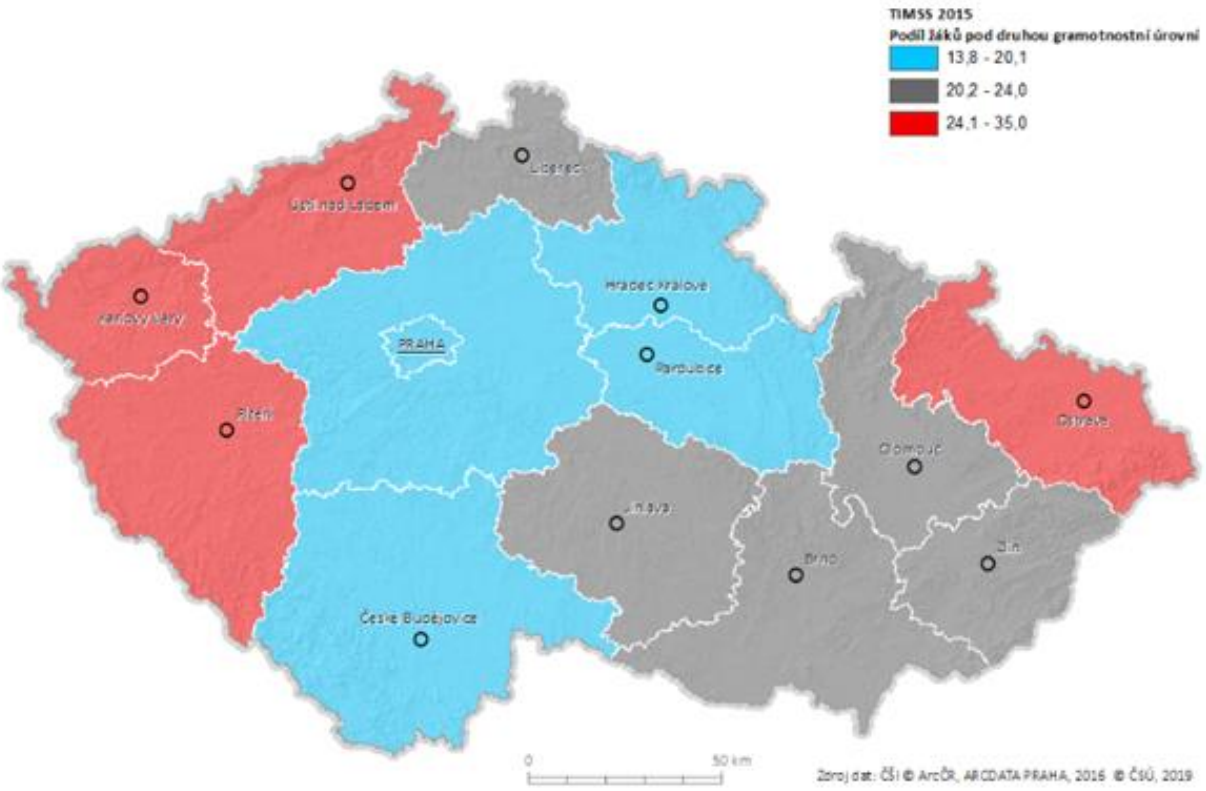
MAPA 1 | Výběrové zjišťování výsledků žáků 9. ročníku – matematika



Druhá možnost vizualizace je pomocí definovaných kvantilů. Určíme si, kolik barevných polí chceme zobrazit, v každé kategorii je pak stejné množství případů. Obecně tato metoda vizualizace může být použita v případě malého počtu kategorií, případně pokud škálovou či intervalovou proměnnou chceme vizualizovat pomocí několika kategorií a udělat

vizualizaci přehlednější. Tato metoda vizualizace (klastrování) je použita například v mapě zobrazující podíl žáků pod druhou gramotnostní úrovní v krajích ze šetření TIMSS. Cílem bylo zpřehlednit hodnoty krajů tak, aby mapa jasně odlišila ty kraje, které mají v této oblasti velmi špatné výsledky. Na mapě jsou proto vidět tři skupiny krajů. Kraje s vysokým podílem žáků pod druhou gramotnostní úrovní jsou zobrazeny červenou barvou, kraje uprostřed jsou zobrazeny šedou barvou, kraje s nejnižšími hodnotami pak modrou barvou. Ve dvou posledních kategoriích je 5 krajů, v první kategorii pak 4 kraje (soubor není dělitelný třemi, proto nejde rozdělit na tři zcela stejné skupiny).

MAPA 2 | Podíl žáků pod druhou gramotnostní úrovní v krajích – TIMSS



Další metodou rozdělení jsou stejné intervaly (equal interval), které ale neberou v potaz rozdělení dat. Tato metoda rozdělí případy dle definovaného počtu skupin do kategorií, které nabývají stejného intervalu hodnot mezi minimem a maximem. Poslední vhodnou metodou pro rozdělení hodnot do několika skupin je metoda na základě směrodatných odchylek (Standard Deviation). Jedná se o využití vlastností normálního rozložení, metoda disproporčně rozdělí případy tak, že největší počet se shlukuje kolem průměru, menší počet krajních hodnot pak na oba konce normálního rozdělení. Metoda je vhodná pro vizualizaci odlehlých případů.

Při tvorbě map je nutné dodržovat správný koordinační systém. V Česku se pro civilní data používá systém S-JTSK (EPSG:5514). Jeho použití v katastrálních mapách a státním mapovém díle (tedy např. ZABAGED, RÚIAN a většina tematických databází) je stanoveno zákonem. Běžně používaná varianta v GIS softwarech se označuje jako S-JTSK Krovak East North. Tento koordinační systém ale není vhodný pro kartografické účely a pro vizualizace dat. Důvodem je stočení severního směru o několik stupňů k východu. Lepším zobrazením mapy ČR je pomocí koordinačního systému WGS84 / UTM Zone 33 North (EPSG:32633), který používá například Armáda ČR. V tomto případě je mapa České republiky natočena správně. Mapy by měly být zobrazeny a natočeny tak, jako na výše uvedených obrázcích. Každá mapa by měla obsahovat zdroj dat a měřítko, popřípadě Severku.

5

Chybějící hodnoty (Missing Values)

5 CHYBĚJÍCÍ HODNOTY (MISSING VALUES)

Kapitola 11 v Sadě typizovaných analytických nástrojů

Chybějící hodnoty je možné nalézt téměř ve všech typech datových souborů. Zpravidla se jedná o chybějící případy proměnných, přičemž důvody absence daných hodnot se mohou různit. V tomto ohledu jsou vydělovány celkem tři druhy chybějících proměnných:

- hodnoty chybějící zcela náhodně (missing completely at random; MCAR),
- hodnoty chybějící náhodně (missing at random; MAR),
- hodnoty chybějící nenáhodně (missing not at random; MNAR).

V případě první skupiny dochází k absenci dat na základě zcela nahodilé příčiny. Neexistuje zde tedy žádný vztah mezi danou chybějící hodnotou a dalšími chybějícími, ale i zaznamenanými daty. Příkladem zde může být neúmyslná ztráta vyplněného dotazníkového formuláře, ze kterého měla být data čerpána.

U druhé skupiny chybějících hodnot lze již pozorovat systematický vztah mezi chybějícími a pozorovanými hodnotami, nikoliv však mezi chybějící hodnotou a dalšími chybějícími daty. Jinými slovy to, zda nějaké pozorování chybí, nemá nic společného s chybějícími hodnotami, ale může zde existovat souvislost s dalšími měřenými parametry. Příkladem mohou být chybějící data o tělesné váze. Muži zpravidla nemají takový problém se sdělováním svých údajů o váze jako ženy. Právě podmíněnost chybějících hodnot pohlavím v tomto případě poukazuje na hodnotu chybějící náhodně.

V poslední skupině chybějících hodnot následně dochází ke vzniku podmíněného vztahu mezi chybějícími daty a jejich hodnotou. Příkladem v tomto případě může být student, který se nedostaví k vykonání zkoušky, protože nemá dostatečné vědomosti k jejímu složení.

Chybějící hodnoty do značné míry ztěžují práci s daty. V řadě případů tak může docházet ke zkreslení a snižování hodnoty výsledků analýz, do kterých tato neúplná data vstupují. Obdobným problémem je pak také neschopnost některých algoritmů pracovat s datasety, které obsahují chybějící hodnoty. Příkladem mohou být mnohonásobné regresní modely, které chybějící hodnoty, potažmo celé případy, nejsou schopny v rámci analýzy zpracovat. Menší počet pozorování následně snižuje přesnost odhadů a sílu statistických testů. Z těchto důvodů je vhodné před započetím samotné analýzy provést imputaci chybějících dat. Díky doplnění chybějících případů dojde zpravidla ke zkvalitnění využívaných dat a zpřesnění výsledků prováděné analýzy.

Praktický návod, jak implementovat analýzu chybějících hodnot a imputovat tyto hodnoty, je v dokumentu *Sada typizovaných analytických nástrojů* v sekci Chybějící hodnoty a jejich imputace.

6

Závěr

6 ZÁVĚR

Předkládaný dokument přiblížil stávající sběr dat a jejich propojování napříč různými datovými zdroji přes příslušné identifikátory, metodologické přístupy k analýze a následné vizualizaci dat. Stěžejní částí metodiky je metodologická debata nad interpretací jednotlivých analýz a přístupů. Pokud výzkumný design sběru dat není experiment, kde máme zpravidla dvě náhodně vybrané skupiny, ale jedná se o tzv. observační sběr dat, kauzální interpretace výsledků jsou velmi limitující. Z tohoto důvodu je na místě opatrnost s ohledem na tvorbu striktních kauzálně vyznívajících doporučení pro vzdělávací soustavu, které vycházejí z pouhých korelací a asociací. Důraz je třeba dávat na interpretační omezení.

Existují však výzkumné designy, které umožňují doporučení a výsledné interpretace vyplývající z analýz opatrně interpretovat kauzální terminologií. Jedná se o tzv. kvaziexperimenty či případně metody, které simulují výběr dat jako v případě experimentu (tzv. matching methods). Doplňujícím typem analýzy ke klasickým kvantitativním technikám mohou být některé dílčí kvalitativní techniky. V tomto ohledu se jedná např. o metodu process-tracing, případně focus groups či různé formy rozhovorů s respondenty, jež je možné implementovat do stávajících technik hospitací. Tyto kvalitativní informace pak mohou posloužit k lepšímu zacílení a tvorbě doporučení ze strany ČŠI. Jejich roli nicméně nelze přeceňovat a stále je nutné dbát na specifická interpretační omezení.

Dokument vychází ze zkušeností z již provedených sekundárních analýz, ale i dílčích zpráv analyzujících prostorová data, popřípadě komplexních analýz krajských vzdělávacích soustav. V těchto analýzách byla propojena data na rozdílných úrovních, a to nejen z datových zdrojů ČŠI, ale i MŠMT, ČSÚ a dalších organizací. Metodika představila nejběžnější typizované analytické techniky pro data v oblasti vzdělávání, které byly úspěšně testovány v praxi v běžném procesu Komplexní inspekční činnosti a tvorby analytických výstupů ČŠI. Cílem bylo představit jednotlivé analytické postupy, jejich praktická implementace ve statistickém softwaru je pak tématem navazujícího dokumentu *Sada typizovaných analytických nástrojů v prostředí standardních statistických programů vycházejících z metodiky*.

Literatura

LITERATURA

Beach, D. & Pedersen, R. B. 2019. *Process-Tracing Methods: Foundations and Guidelines*. Michigan: The University of Michigan Press.

Brinkmann, S. & Kvale, S. 2015. *InterViews: Learning the Craft of Qualitative Research Interviewing*. London: Sage Publications.

Byrne, B. M. 2016. *Structural Equation Modeling with AMOS. Basic Concepts, Applications and Programming (3rd edition)*. New York/London: Routledge.

Cooper, B. & Glaesser, J. 2011. „Paradoxes and Pitfalls in Using Fuzzy Set QCA: Illustrations from a Critical Review of a Study of Educational Inequality.“ *Sociological Research Online* 16 (3): 106–119.

ČŠI. 2019b. *Metodika inspekční činnosti ve školách a školských zařízeních*. Praha: ČŠI.

Dunning, T. 2008. „Improving Causal Inference: Strengths and Limitations of Natural Experiments.“ *Political Research Quarterly* 61 (2): 282–293.

Everitt, B. S., Landau, S., Leese, M. & Stahl, D. 2011. *Cluster Analysis*. London: Wiley.

Gelman, A., Hill, J. & Vehtari, A. 2020. Regression and Other Stories. Cambridge University Press.

Goertz, G. & Mahoney, J. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press.

Hendl, J. 2009. *Přehled statistických metod zpracování dat: Analýza a metaanalýza dat*. Praha: Portál.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I. & Vila-Abad, E. 2010. „Polychoric versus Pearson Correlations in Exploratory and Confirmatory Factor Analysis of Ordinal Variables.“ *Quality & Quantity* 44 (1): 153–166.

Lichtman, M. 2013. *Qualitative Research in Education. A user's guide*. London: Sage Publications.

Mazák, J. 2017. „Process tracing: zkoumání kauzality v případových studiích.“ *Sociológia* 49 (1): 55–80.

Miles, M. B., Hubermann, A. M. & Saldana J. 2014. *Qualitative Data Analysis: A Methods Sourcebook*. London: Sage Publications.

PISA. 2015. *PISA 2015 Technical Report.* Dostupné z: https://www.oecd.org/pisa/data/2015-technical-report/.

Ragin, C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies.* Berkeley: University of California Press.

Soukup, P. & Kočvarová, I. 2016. „Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu.“ *Pedagogická orientace* 26 (3): 512–536.

Soukup, P. & Kočvarová, I. 2016. „Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu.“ *Pedagogická orientace* 26 (3): 512–536.

Soukup, P. 2021. „Faktorová analýza jako známá neznámá (aneb metoda hlavních komponent a varimax není vždy ideální postup).“ *Sociologický časopis* 57 (4): 1–30.

Vis, B. 2012. „The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses.“ *Sociological Methods & Research* 41 (1): 168–198.

Young-Chool, Ch. & Ji-Hye, L. 2015. „What most Matters in Strengthening Educational Competitiveness?: An Application of FS/QCA Method.“ *Procedia – Social and Behavioral Sciences* 197: 2182–2190.

Česká školní inspekce

Fráni Šrámka 37 | 150 21 Praha 5 | www.csicr.cz