

# Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu<sup>1</sup>

Petr Soukup<sup>a</sup>, Ilona Kočvarová<sup>b</sup>

<sup>a</sup>Univerzita Karlova v Praze, Fakulta sociálních věd, Institut sociologických studií

<sup>b</sup>Univerzita Tomáše Bati ve Zlíně, Fakulta humanitních studií, Ústav školní pedagogiky

Redakci zasláno 14. 2. 2016 / upravená verze obdržena 17. 9. 2016 /  
k uveřejnění přijato 21. 9. 2016

**Abstrakt:** Metodologicky orientovaná studie syntetizuje základní požadavky na velikost a reprezentativitu výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu. Prezentuje základní výběrové postupy, jejich parametry a dopad na požadovanou velikost výběru a také dopad na možnosti zobecnění výsledků. Kromě standardních designů náhodného výběru (prostý, vícestupňový, skupinkový a stratifikovaný) je pojednán i dopad velikosti pro nenáhodné výběry. Studie uvádí postupy pro stanovení síly testu a shrnuje základní doporučení statistiků pro velikost výběru s ohledem na sílu testu. Pozornost je věnována také otázce reprezentativity výběru. Součástí studie jsou příklady dobré i špatné praxe z oblasti mezinárodních pedagogických výzkumů.

**Klíčová slova:** kvantitativní výzkum, populace, výběrový soubor, velikost výběru, reprezentativita, zobecnitelnost

Cílem studie je vytvořit integrativní přehled základních požadavků vztahujících se na velikost a reprezentativitu výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu, zasadit je do kontextu parametrů ovlivňujících proces tvorby výběru a jeho kvalitu a ilustrovat tuto problematiku s využitím příkladů dobré i špatné praxe z oblasti mezinárodních pedagogických výzkumů.

Chceme upozornit na problematiku, která dosud v české pedagogické metodologii figurovala spíše v pozadí. Zároveň si však nenárokujeme její komplexní zpracování, protože je široká a nelze ji vtěsnat do rozsahu jednoho časopiseckého článku.<sup>2</sup>

<sup>1</sup> Práce prvního autora byla podpořena z grantu GA ČR P402/12/G130 *Vztahy mezi dovednostmi, vzděláváním a výsledky na trhu práce: longitudinální studie*.

<sup>2</sup> Text zejména neřeší problematiku výběrů z malých populací, která je natolik specifická, že jí bude třeba věnovat samostatný článek.

Přehled požadavků na velikost a reprezentativitu výběrového souboru vychází z české i zahraniční statistické a metodologické literatury věnované oblasti kvantitativního výzkumu.

## 1 Základní pojmy pro výběrová šetření

Před samotným pojednáním tématu velikosti a reprezentativity výběrového souboru stručně definujeme základní pojmy a poukážeme na problematiku malé návratnosti.

Na počátku každého výzkumného šetření stojí stanovený cíl. Ten determinuje celý proces výzkumu, včetně toho, jak široká je oblast našeho bádání a jaké výzkumné jednotky bude třeba do něj zahrnout. Na nejvyšší úrovni si musíme vydefinovat *cílovou populaci (target population)*, což je komplexní soubor elementů – jednotek analýzy (osob, institucí, událostí nebo jiných subjektů), na které chceme zobecnit výsledky výzkumu. Populace může mít jakoukoli velikost, z hlediska statistiky může být teoreticky nekonečná. Populaci bývá těžké jednoznačně vyčíslit a charakterizovat. Ve výzkumu se proto zaměřujeme na určitou část populace, kterou jsme schopni konkrétněji vymezit a je pro naše účely technicky dosažitelná. Označujeme ji jako *dostupnou populaci (survey population)*, příp. *základní soubor* nebo *cílovou skupinu* výzkumu. Dostupnou populaci je nutné definovat co nejpřesněji a také je třeba identifikovat skupiny elementů, které do ní z různých důvodů nemohou být zahrnuty (Kalton, 1983, s. 6–13). Jako cílovou populaci můžeme definovat např. všechny studenty českých vysokých škol, ale při vymezení dostupné populace narazíme na skutečnost, že do ní nebude možné zahrnout studenty, kteří jsou v době šetření na studijních pobytech v zahraničí, jsou dlouhodobě nemocní nebo na čas přerušili studium.

Informace o charakteristikách (ne vždy však o konkrétních jednotkách) populace lze získat z různých databází, např. MŠMT, ČSÚ nebo ČŠI. Potřebujeme znát co nejpřesnější údaje o její velikosti a známých (v případě zkoumání osob většinou demografických) vlastnostech. Zajímá nás také míra její homogenity/heterogenity (např. z hlediska věkového rozpětí, vzdělanosti atp.). Heterogenní populace bývá složitější zahrnout do výzkumného šetření než populace homogenní, protože potřebujeme pro účely šetření získat širší spektrum jednotek splňujících nároky na různé charakteristiky.

Při definování populace musíme brát ohled na limitující faktory výzkumu, primárně na čas a finance, které nás nutí k rychlému a pokud možno finančně nenáročnému sběru dat. Limitující faktory nás často donutí zúžit cíl výzkumu a současně cílovou skupinu výzkumu.

Protože většinou není možné (ani nutné) zahrnout do výzkumu kompletní populaci, vybíráme pro tyto účely její zmenšeninu, tedy *výběrový soubor*, což je množina výzkumných jednotek, které skutečně zkoumáme. V tomto textu dále používáme stručné označení *výběr*. Měl by být dostatečně velký a vykazovat co největší podobnost s populací, na kterou chceme naše závěry zobecnit s využitím inferenční statistiky.

V pedagogice (ale i v jiných vědeckých disciplínách) používáme dva základní typy výběru. *Pravděpodobnostní výběr (probability sample)* je obecně považován za objektivní způsob výběru výzkumných jednotek, protože každá jednotka má předem známou pravděpodobnost, že bude zahrnuta do výběru. Je však možné jej realizovat téměř výhradně v případě, že máme tzv. *oporu výběru (sampling frame)*, tedy seznam jednotek dostupné populace.

V reálných podmínkách často nebývá možné realizovat pravděpodobnostní výběr. Proto se uchylujeme k *výběru nepravděpodobnostnímu (non-probability)*. Ten tvoří jednotky, většinou osoby, které jsou vybrány na základě stanovených kvót, příp. expertního posouzení, jsou dostupné a ochotné se zapojit do našeho šetření (Fink, 2013, s. 88).

I při realizaci pravděpodobnostního výběru dochází ve výzkumech k mnoha zkreslením. Zkreslení – odchylku mezi základním souborem a výběrem – vyjadřuje *výběrová chyba*<sup>3</sup>. Může být způsobena náhodou (*sampling error*) nebo samotným výzkumníkem (*sampling bias*), např. nevhodnou metodou tvorby výběru, při které dojde k tomu, že se z hlediska různých charakteristik více či méně odchýlí od populace. Na základě výběrové chyby lze stanovit *interval spolehlivosti (confidence interval)*, který při námi požadované spolehlivosti udává rozpětí hodnot zahrnujících odhadovaný populační parametr (Hendl, 2015, s. 176).

<sup>3</sup> Bryman (2012, s. 205–206) dělí výběrové chyby na 4 skupiny: *sampling error* (chyba způsobená faktem, že výběrový soubor nikdy neodráží vlastnosti celé populace), *sampling-related error* (chyba související s kvalitou sběru dat a s návratností), *data-collection error* (chyba související s kvalitou použitého výzkumného nástroje) a *data-processing error* (chyba vzniklá při statistickém zpracování dat).

Výběrová chyba, kterou jsme ochotni připustit, souvisí se třemi faktory, jak ukazuje vzorec pro její výpočet v případě prostého náhodného výběru (základní techniky pravděpodobnostního výběru):

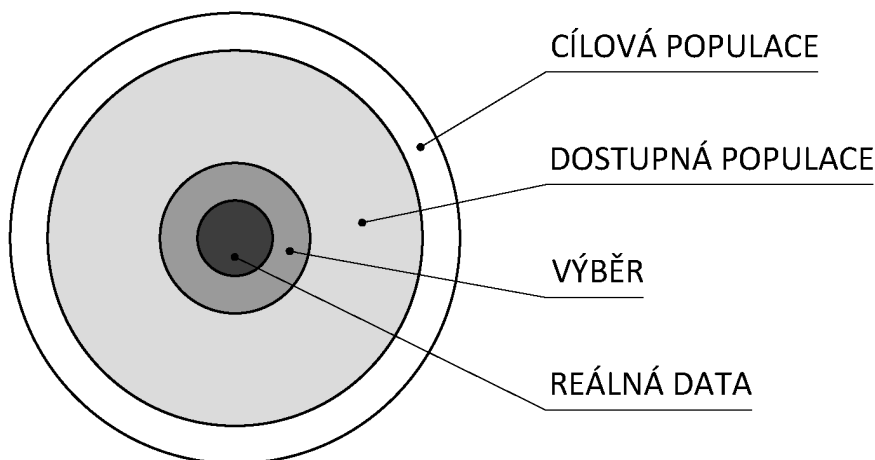
$$se_{(y_0)} = \sqrt{(1-f)s^2/n}$$

v němž  $se_{(y_0)}$  je standardní (výběrová) chyba,  $f=n/N$  a vyjadřuje poměr mezi velikostí výběru a populace (*sampling fraction*),  $s^2$  je směrodatná odchylka a  $n$  je velikost výběru. Pokud je poměr mezi velikostí výběru a populace 1/10 nebo menší, vychází hodnota  $\sqrt{(1-f)}$  blízko k 1, což znamená, že ji lze při výpočtu zanedbat. Při tvorbě výběru není tedy třeba zvažovat velikost populace, pokud je minimálně 10x větší než výběr (podrobně in Kalton, 1983, s. 13–14).

Jedním ze základních parametrů kvality výběrových dat je kromě jejich velikosti měřené počtem jednotek (typicky respondentů) i *návratnost* (*response rate*), které jsme při získávání výběru dosáhli. Kromě velikosti námi získaných dat bychom měli při publikaci výsledků uvádět i návratnost, aby měl čtenář možnost posoudit kvalitu námi používaných dat. Pokud má náš výběr vypovídat o situaci v populaci, pak je třeba logicky dosáhnout co nejvyšší návratnosti. Doporučené míry návratnosti (měřené poměrem jednotek v našem datovém souboru ku počtu původně vybraných jednotek) se v literatuře liší, nicméně běžně se pohybují okolo 60 %. Reálné míry návratnosti mohou být však výrazně nižší (srov. Krejčí, 2011), nebo také vyšší (srov. například šetření PISA či TIMSS). Důvodem obecně nízké návratnosti je paradoxně prosperita výzkumu, která vede k tomu, že počet šetření se zvyšuje a „populace je označována za tzv. přezkoumanou“ (Krejčí, 2008, s. 76). Specifika při výběrech v pedagogice diskutuje Mertensová (2015, s. 340–341), podle které je v současnosti největším problémem zajistit si povolení k výzkumu a také získat zájem a podporu zainteresovaných stran.

S návratností souvisí také to, jaké informace se podaří získat od jednotlivých jednotek v datovém souboru. Například v případě dotazníkového šetření je na místě kontrolovat množství chybějících informací, které respondenti z různých důvodů neuvedou, což může vést k další redukci konečného souboru reálně získaných dat.

Od cílové populace přes dostupnou populaci a požadovaný výběr se konečný datový soubor zredukuje na reálně získaná data, jak ukazuje obrázek 1.



Obrázek 1. Vizualizace tvorby výběru z populace a jeho redukce na soubor reálně získaných dat.

Namísto je položit si otázku, co dělat v případě, že návratnost našeho šetření je poměrně malá (nedosahuje ani zmíněných 60 %). Nejdříve bychom měli řešit, zda respondenti chybí v našem výběru náhodně, tj. nechybí systematicky nějaká skupina. Toto lze zjistit v podstatě jediným způsobem, a to využít dále zmíněné testování reprezentativity (viz další část článku), kterým se snažíme zjistit co nejvíce o struktuře populace a porovnat náš výsledný datový soubor s populací. Pokud zjistíme výrazná (rozuměj statisticky významná) vychýlení, máme v zásadě jen dvě možnosti práce:

(1) na soubor dále aplikujeme jen nástroje popisné statistiky a rezignujeme na zobecňování na populaci (tento způsob uplatňujeme i u nepravděpodobnostních výběrů, srov. jejich popis v závěru článku);

(2) data zkusíme za pomoci vah upravit tak, aby byla reprezentativní pro populaci a způsobilá ke zobecňování výsledků na populaci.

První uvedený režim práce nepotřebuje dalšího komentáře, proto soustředíme krátce pozornost ke druhému. Váhy (zde používáme tzv. *výběrové váhy* – srov. Soukup, 2016) můžeme využít k navýšení vlivu podreprezentovaných skupin v našich datech a naopak ke snížení vlivu skupin, které máme v datech zastoupeny více než v populaci. Technický popis postupu vážení v SPSS lze nalézt například v učebnici Mareše, Rabušice a Soukupa (2015, s. 197–201). Je namísto upozornit, že pokud víme, že v našem datovém souboru nějaká

skupina z populace zcela chybí, pak nám vážení nepomůže a nelze z dat zobecnění provádět. Obdobně, pokud bude zastoupení skupiny velice malé (například jediným respondentem), hrozí při použití vah výrazná zkreslení výsledků. Lze doporučit, aby výsledné váhy nepřesahovaly hodnoty 5 (resp. 1/5), tj. jeden jedinec z našeho datového souboru je brán pětinásobně (resp. jednou pětinou). Dodejme, že při používání statistických testů s váženým datovým souborem je vhodné respektovat doporučení, která podali Soukup a Rabušic (2007, s. 391–393).

## 2 Reprezentativita výběru

První charakteristikou výběru, které věnujeme v textu pozornost, je jeho *reprezentativita*, která „znamená takovou kvalitu výběru a dalších postupů metodiky výběrového šetření, díky níž dotázaný soubor jednotek získá rozložení charakteristik, jež můžeme považovat za shodné s populací, ze které vybraný soubor pochází“ (Krejčí, 2011, s. 3). Jedná se tedy o to, jak věrně výběr reprezentuje známé (a tudíž kontrolovatelné) znaky cílové populace (např. věkovou, vzdělanostní či genderovou strukturu), které jsou podstatné vzhledem k cíli našeho šetření.

Reprezentativita závisí především na způsobu tvorby výběru, kdy ideální je pravděpodobnostní výběr, a také souvisí s velikostí výběru, která zvyšuje přesnost odhadu populačních parametrů (Řehák, 1978, s. 492, 496).

V kontextu reprezentativity považujeme za podstatné položit si dvě otázky:

- 1) Musí všechny kvantitativně orientované výzkumy splňovat nároky na reprezentativitu výběru? Většina šetření realizovaných na území ČR nesplňuje tyto nároky (Sedláková, 2014, s. 92). Důvody jsou následující: neumožnost přesně vydefinovat populaci, nedostatek financí a času a také neumožnění přístupu do terénu a nezájem či neochota respondentů. To, že šetření nejsou postavena na náhodném výběru a dalších důležitých faktorech, však neznamená, že nemají pro společnost význam.

Každé šetření, i to, které má charakter případové studie a věnuje se pouze jednomu výskytu daného jevu, pokud je provedeno korektně, přináší relevantní zjištění, obohacuje stávající vědění a rozšiřuje daný obor. Je proto třeba zbavit se neopodstatněného dojmu, že pouze reprezentativní šetření jsou přínosná a žádaná. (Sedláková, 2014, s. 93)

- 2) Je reálné zajistit reprezentativitu výběru? Tento problém diskutuje Babbie (2013, s. 133–134), který uvádí, že reprezentativita nemá precizní vědecký význam. Ačkoli bychom ji měli brát velmi vážně, musíme si uvědomit, že výsledky jakéhokoli sociálního výzkumu jsou vždy relativní a kontextuálně podmíněné, jejich reprezentativita je proto diskutabilní i za dodržení požadavků na náhodný a dostatečně velký výběr respondentů.

Reprezentativní výběr z hlediska známých demografických znaků automaticky nezaručuje důvěryhodnost sebraných dat, a to především v případě zkoumání osob. Jedná se o to, jak se sebraná data odlišují od reality z důvodu záměrné či nevědomé modifikace reality v chování a výpovědích respondentů. Ukazuje se, že respondenti mají tendence prezentovat sebe a své bezprostřední okolí sociálně přijatelně (více viz Urbánek, Denglerová, & Širůček, 2011, s. 134) a vyhýbají se odpovědím na nepříjemné otázky, čímž dochází k porušení reprezentativity (Řehák, 1978, s. 490). V pedagogických výzkumech se stává, že respondenti jsou ochotni odpovědět i na věci, které neznají nebo se jich netýkají, pokud se jim přirozeně nenabízí odpověď typu „nevím“ nebo „netýká se mě“. Je také známo, že se respondenti nechávají ovlivnit sugestivními či ofenzivními podněty, které by odborně sestavené techniky sběru dat vůbec neměly obsahovat. Důvěryhodnost výpovědí respondentů můžeme pozitivně ovlivnit pečlivou tvorbou a aplikací výzkumných nástrojů, není však možné na ni absolutně spoléhat.

Výběr a získaná data lze podrobit kontrole reprezentativity z hlediska vybraných charakteristik. Jako nejjednodušší možnost se nabízí aplikace testu dobré shody chí-kvadrát, ve kterém porovnáváme, zda se námi získané četnosti v různých kategoriích zkoumaných jednotek shodují s očekávanými četnostmi, které stanovíme na základě znalosti populace. Například tak můžeme porovnat vzdělanostní strukturu našeho výběru se známou vzdělaností strukturou populace. Vzhledem k tomu, že chí-kvadrát test je možné spočítat i v běžně užívaném SPSS, lze čtenáře odkázat na jeho využití. Postup, jak test vypočítat, lze včetně ukázky nalézt např. v učebnici Mareše, Rabušice a Soukupa (2015, s. 259–263). Samozřejmě, že nejobtížnějším momentem při testování reprezentativity bude získání charakteristik námi zkoumané populace.

S tématem reprezentativity výběrového souboru souvisí *reprezentativita obsahu*, která vyjadřuje, „do jaké míry je v projektu zachycen obsah cílových

témat“ (Řehák, 1978, s. 490).<sup>4</sup> Kromě toho, že požadujeme, aby soubor osob v našem výzkumu byl reprezentativní pro zkoumanou populaci (chceme-li pak na populaci zobecňovat výsledky našeho výzkumu), je potřeba, abychom reprezentativně vystihli zkoumané téma (obsah). Ostatně již například klasický postup Thurstonova škálování (Schenk & Hrabovská, 2010) jasně vystihoval tuto potřebu a úkol svěřoval do ruky expertům. Tento požadavek platí jak pro dotazník (tj. naše otázky musí reprezentativně pokrývat zkoumané téma), tak pro test (testové úlohy musí reprezentativně pokrývat testované kurikulum). Již při započetí plánování výzkumu je tedy třeba výzkumný nástroj (dotazník či test) plánovat tak, aby byl obsahově reprezentativní vzhledem ke zkoumanému problému (ideální je vytvoření tzv. specifikační tabulky, kde se v řádku zachytí jednotlivá zkoumaná témata a v sloupci jejich dimenze a navrhne se počet otázek, resp. testových úloh, pro jednotlivé buňky tabulky). Pro vyhodnocení obsahové reprezentativity je vhodné provést pilotní šetření, které umožňuje poté skrze statistické nástroje vyhodnotit „kvalitu“ jednotlivých otázek (testových úloh) a posoudit i celkovou obsahovou reprezentativitu. S ohledem na omezený rozsah článku jen zmiňme, že pro jednotlivé testové úlohy je vhodné minimálně vypočítat jejich obtížnosti a citlivosti (analogické charakteristiky je vhodné počítat i pro otázky tvořící baterie v dotazníkových šetřeních). Pro jednotlivé zkoumané oblasti je pak vhodné vyhodnotit reliabilitu (nejčastěji skrze Cronbachovo alfa<sup>5</sup>) a validitu našeho výzkumného nástroje (Štuka et al., 2013, s. 62–93; Řehák, 1998a). Složitější přístupy pak využívají strukturního modelování (srov. zejména Řehák, 1998b). Výše uvedený popis je spojen s klasickou teorií testování (případně s postupy likertovského škálování, srov. Schenk & Hrabovská, 2010). V dnešní době se namísto klasické teorie testování (CTT) používá často teorie odpovědi na položku (IRT), nicméně výše popsané principy platí i zde, jen postupy IRT umožňují další možnosti vyhodnocení reprezentativity a zajišťují tak lépe optimální výběr otázek, resp. testových úloh (Jelínek, Květoň, & Vybořil, 2011).

S reprezentativitou i velikostí výběru úzce souvisí již zmiňovaná *zobecnitelnost (generalizability)*<sup>6</sup>, která vyjadřuje, zda a s jakou jistotou platí výsledky

<sup>4</sup> Děkujeme anonymnímu recenzentovi za upozornění na tento aspekt výzkumné činnosti.

<sup>5</sup> Nutno ovšem upozornit na problematičnost tohoto oblíbeného koeficientu, viz např. článek Zváry (2002) nebo novější text Sijtsma (2009).

<sup>6</sup> V kontextu se zobecnitelností se také mluví o *externí validitě* a *transferabilitě*. Všechny tři pojmy vyjadřují, jak dalece je možné přenést výsledky z výběru na populaci (Mertens, 2015, s. 320). Někteří autoři též rozlišují zobecnitelnost vzhledem k populaci a vzhledem k teorii (Saunders, Lewis, & Thornhill, 2007, s. 227).

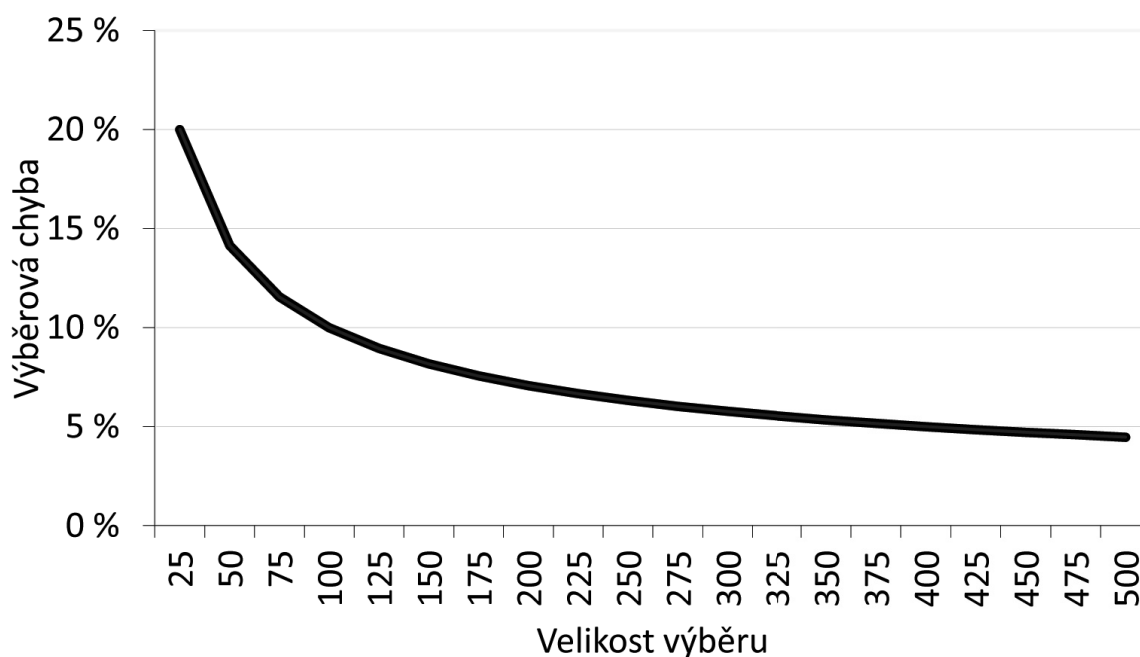


získané z výběru na populaci. Ačkoli se běžně spojuje s kvantitativně orientovaným výzkumem, je vysoce spekulativní (Vogt & Johnson, 2011, s. 156). Lidé jakožto nejčastěji zkoumané subjekty pedagogického, potažmo sociálního výzkumu jsou příliš rozdílní na to, abychom je mohli považovat za soubor relativně homogenních jednotek. Jejich charakteristiky se navíc vyvíjejí v čase, tudíž míra zobecnitelnosti každého sociálního výzkumu teoreticky klesá od chvíle, kdy byl realizován. Proto je zobecňování vždy riskantní.

### 3 Velikost náhodného výběru

V této a následující části článku se věnujeme problematice velikosti výběru.

Pro pravděpodobnostní výběry obecně platí, že čím menší je výběr, tím vyšší je pravděpodobnost statistické chyby 2. druhu, tedy chyby, kdy s využitím statistických testů neprokážeme statisticky významnou souvislost mezi jevy, přestože reálně existuje. Proto se obecně doporučuje realizovat kvantitativně orientovaný výzkum na co největším výběru, protože platí, že s rostoucí velikostí výběru klesá výběrová chyba (srov. též obrázek 2).<sup>7</sup>



Obrázek 2. Velikost výběrové chyby pro relativní četnost (předpokládaná relativní četnost 50 %) s ohledem na velikost výběru pro prostý náhodný výběr.

<sup>7</sup> I velké výběry však mají svá negativa, která se projevují především při realizaci inferenční statistiky, jak uvádí např. Field (2013, s. 72–73).

V praxi se však podřizujeme limitujícím faktorům jako čas a finance, které snižují naše ambice na velikost výběru. Snažíme se proto odhadnout jeho velikost tak, aby byl co nejmenší, ale zároveň vyhovoval základním metodologickým standardům. Poznamenejme již zde, že velikost výběru nelze zjednodušovat na pravidlo „čím větší, tím lepší“ (viz varovný příklad na konci článku). Velikost výběrového souboru souvisí především s designem výběru (srov. další část) i s plánovanými statistickými operacemi, které chceme na výběr aplikovat (srov. praktická doporučení na konci další části).

### 3.1 Základní stanovení velikosti výběru pro prostý náhodný výběr

Vzorec pro odhad velikosti prostého náhodného výběru (*simple random sampling*) uvádí Kalton (1983, s. 82):

$$n' = 1,96^2 \frac{PQ}{SE^2}$$

kde  $n'$  je základní odhad velikosti výběru (bez ohledu na velikost populace),  $P$  je procentuální proporce, ve které se sledovaný znak v populaci vyskytuje,  $Q$  je procentuální proporce, ve které se sledovaný znak nevyskytuje ( $Q = 100 - P$ ), a  $SE$  je tolerovaná výběrová chyba (uvádí se v %, standardně se počítá s hodnotou 5 nebo nižší).

Protože parametry  $P$  a  $Q$  většinou neznáme (nebo se zaměřujeme na více sledovaných parametrů), dosadíme za obě veličiny nejkonzervativnější hodnotu 50 (podle úvahy, že sledovaný znak se může a nemusí vyskytovat u každého jedince s pravděpodobností „50 na 50“). Po dosazení do vzorce získáme při nejkonzervativnějším nastavení hodnot  $P$  a  $Q$  a za tolerance 5% chyby základní odhad velikosti výběru  $n' = 1,96^2 \frac{50 \times 50}{5^2} = 384$ .

Povšimněme si, že ve vzorci nefiguruje velikost cílové populace, tj. obecně platí (výjimku srov. dále), že velikost cílové populace neovlivňuje velikost výběru a pro stejnou přesnost výsledků potřebujeme stejně velký výběr bez ohledu na velikost cílové populace (tedy například výběr 400 jednotek z 10 tisíc zajišťuje srovnatelnou přesnost jako výběr z 10 milionů, za předpokladu srovnatelné variability dat v desetitisícovém i desetimiliónovém souboru).

Velice snadno lze pro stanovení velikosti výběru použít uvedený obrázek 1. Dle požadované velikosti výběrové chyby najdeme minimální velikost výběru. Požadujeme-li například 10 %, pak vedeme vodorovnou linii ve výši 10 % a na ose  $x$  odhadneme, že potřebujeme cca 110 jednotek ve výběru.

Podobně pro 5 % získáme námi výše vypočtenou hodnotu ve výši cca 380 výběrových jednotek.

Speciální situace nastává, pokud námi zkoumaná populace je malá a my chceme přesto namísto vyčerpávajícího zjišťování provést výběr a zobecnit poté výsledky na celou tuto malou populaci za pomoci statistických testů. Je-li náš výběr více než desetinou zkoumané populace, pak neplatí běžné vzorce pro statistické testy a je třeba je modifikovat. Praktický příklad výpočtu lze nalézt v článku Soukupa a Rabušice (2007, s. 388–389). Obecně platí, že pokud vybíráme velkou proporcí z malé populace, výběr může být při stejné požadované přesnosti menší. Dodejme zároveň, že v případě, kdy provádíme výběr z malých populací, je ještě více třeba dbát na to, aby výběr byl striktně náhodný, sebemenší vychýlení či systematickosti zcela zničí možnost zobecňovat výsledky na populaci.

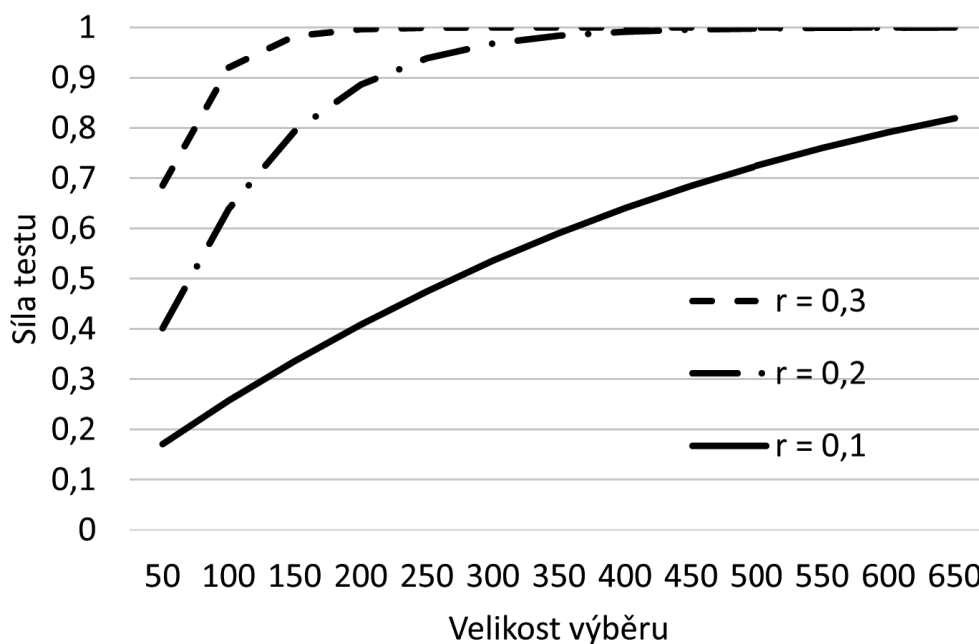
### 3.2 Stanovení velikosti prostého náhodného výběru s ohledem na sílu testu

V posledních letech se prosazuje zejména v pedagogickém a psychologickém kvantitativním výzkumu doporučení, aby byla uváděna *síla testu* u výsledků statistických testů (tj. stanovení síly testu *ex post*). Obdobně lze nalézt doporučení pro stanovení velikosti výběrového souboru s ohledem na předpokládanou sílu plánovaných statistických testů (tj. stanovení síly testu *ex ante*). S ohledem na téma článku, stanovení velikosti výběrového souboru před jeho provedením, bude dále diskutováno jen stanovení síly *ex ante*, resp. určení velikosti výběrového souboru, aby bylo dosaženo dostatečné síly testu. Nejhlásitějším propagátorem síly testu byl Jacob Cohen, který tématu věnoval celou knihu (Cohen, 1988) a tato publikace je pomyslným zlatým standardem. Pro úplnost uveďme, že síla testu je pravděpodobnost, že zamítneme nulovou hypotézu za předpokladu, že tato opravdu v populaci neplatí, tj. síla testu měří schopnost statistického testu detekovat rozdíly či souvislosti, o které se zajímáme (Hendl, 2015, s. 417–422). Běžně se doporučuje minimální síla testu o velikosti 0,8. Stanovení síly testu je pro jednodušší statistické testy možné provést složitějšími vzorci, pro složitější statistické techniky se pak nejčastěji užívá simulačních postupů Monte Carlo (Lafaye de Micheaux & Tran, 2016). S ohledem na velice časté používání korelací a t-testů v českém pedagogickém výzkumu se zde omezíme na ukázky stanovení velikosti výběrového souboru jen u těchto dvou situací, pro ostatní odkážeme na užitečné pomůcky na internetu a relevantní programy.

Začneme se stanovením minimální velikosti výběrového souboru s ohledem na sílu testu pro korelační studii. Pro stanovení velikosti výběrového souboru při respektování síly testu musíme znát:

- požadovanou míru statistické významnosti (typicky volíme 0,05);
- předpokládanou velikost korelačních koeficientů;
- požadovanou sílu testu (typicky minimálně 0,8).

S ohledem na běžné výsledky pedagogických studií jsme připravili graf na obrázku 3, který ukazuje vztah mezi velikostí výběrového souboru (osa x) a silou testu pro korelační koeficienty o velikosti 0,1, 0,2, resp. 0,3 (míra statistické významnosti je konvenčně 0,05).

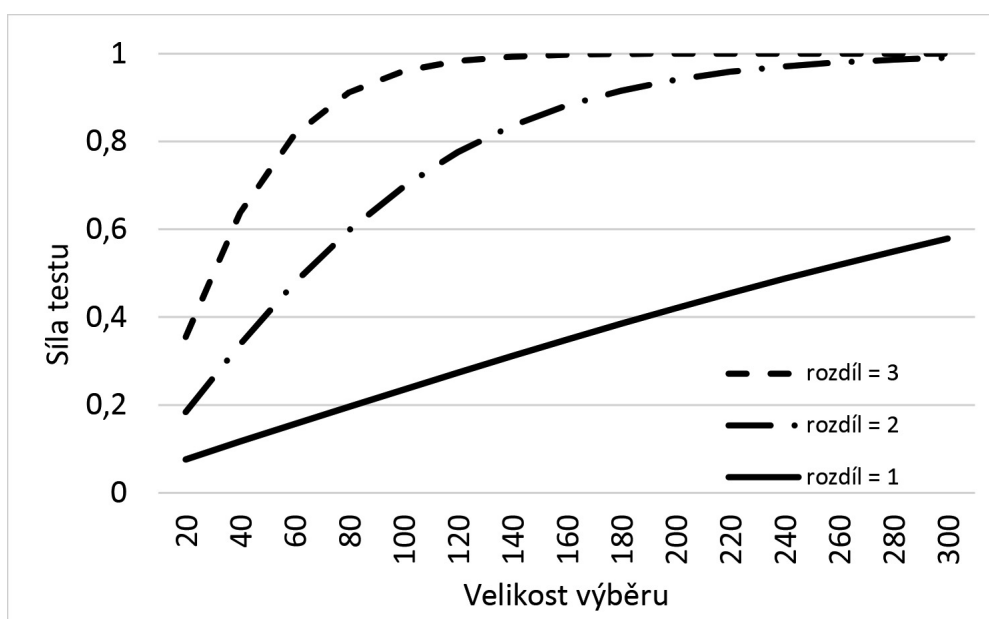


Obrázek 3. Velikost síly testu pro korelační koeficient s ohledem na velikost výběru pro prostý náhodný výběr ( $\alpha = 0,05$ ).

Pro čtení grafu je možné volit následující způsob: povedu pomyslnou vodorovnou čáru na úrovni síly testu 0,8 a zjistím, jaká je minimální požadovaná velikost výběru pro příslušnou velikost korelačního koeficientu. Závěr plynoucí z grafu si lze snadno zapamatovat. Pro prokázání korelačních koeficientů o velikosti 0,3 (nejvyšší křivka v grafu) se silou testu 0,8 postačuje velikost výběru cca 65 jednotek. V případě korelace na úrovni 0,2 je to již minimálně

150 jednotek a pro korelace o velikosti 0,1 pak téměř 620 jednotek! Je tedy opět patrné, že čím menší je předpokládaný zjištěný efekt, tím větší výběr musíme volit.<sup>8</sup> Elegantní pomůcku k obecnému výpočtu velikosti korelační studie poskytuje např. web StatsToDo.

V případě studií, v nichž se porovnávají průměry (či mediány ve skupinách), lze obdobně jako u korelačních studií stanovit vztah mezi velikostí výběrového souboru a silou testu. Namísto hodnoty korelace je zde potřebné stanovit velikost předpokládaného rozdílu (typicky mezi dvěma skupinami). Pro jednoduchou ilustraci uvádíme graf na obrázku 4. Graf opět vykresluje pro hladinu statistické významnosti 0,05. Namísto různých úrovní korelace zobrazujeme různé velikosti předpokládaného rozdílu (o velikosti  $\frac{1}{4}$ ,  $\frac{1}{2}$  a  $\frac{3}{4}$  směrodatné odchylky<sup>9</sup>), konkrétně v našem případě máme data se směrodatnou odchylkou 4 jednotky a předpokládaný rozdíl je tedy 1, 2, resp. 3 jednotky (viz legenda v grafu).



Obrázek 4. Velikost síly testu pro dvouvýběrový t-test s ohledem na velikost výběru pro prostý náhodný výběr ( $\alpha = 0,05$ ).

<sup>8</sup> Poznamenejme, že korelace o hodnotě 0,1 je ale věcně nezajímavá, a tak lze vycházet z požadavku pro korelační studie, a to prokazovat korelace o hodnotě 0,2 a vyšší (tj. stačí nám výběr o velikosti 150 jednotek).

<sup>9</sup> Tyto hodnoty zhruba odpovídají velikosti *Cohenova d*, přičemž rozdíl se označuje jako malý (0,2), střední (0,5) a velký (0,8) – srov. Soukup (2013, s. 131).

Opět můžeme graf snadno přečíst. *Budeme-li předpokládat rozdíl o velikosti  $\frac{3}{4}$  směrodatné odchylky (v grafu nejvyšší křivka s označením „rozdíl = 3“), pak pro sílu testu 0,8 postačuje výběr o velikosti cca 60 jednotek. V případě, že předpokládaný rozdíl bude maximálně  $\frac{1}{2}$  směrodatné odchylky, výběr musí být minimálně 130 jednotek a pro detekci statisticky významného rozdílu o velikosti pouze  $\frac{1}{4}$  směrodatné odchylky potřebujeme zhruba 600 jednotek ve výběru (není uvedeno v grafu)!<sup>10</sup>*

Pro složitější statistické techniky lze buď využít online kalkulačky, nebo speciální software. IBM distribuuje společně s SPSS software Sample Power (nyní verze 3.0, jde o placený software), který umožňuje stanovení síly testu. Ve volně dostupném prostředí R existuje mnoho balíčků pro stanovení síly testu (za všechny zmiňme balíček Power; Lafaye de Micheaux & Tran, 2016). Dále lze doporučit zdarma dostupný software G\*Power (Heinrich-Heine-Universität Düsseldorf, ©2016). Ex post sílu testu pro běžné postupy (t-test, analýza rozptylu a regrese) lze vypočítat i v SPSS, konkrétně v procedurách General linear model, samozřejmě lze využít i mnohé online kalkulátory.

S ohledem na skutečnost, že v pedagogickém výzkumu se často užívá namísto t-testů neparametrických obdob (Mann-Whitney či Wilcoxonův test), platí doporučení ohledně velikosti výběrového souboru s úpravou. Obecně platí, že síla neparametrických testů je oproti parametrickým zhruba 90 %, tj. velikost výběrového souboru by se měla oproti dříve uvedeným hodnotám cca o pětinu navýšit.

### 3.3 Některá praktická doporučení pro velikost výběru (zejména pro vícerozměrné statistické techniky)

V této pasáži zmíníme některá doporučení, která se uvádějí ve statistických či metodologických textech. Ve valné většině případů jde o tzv. *rules of thumbs*, tj. doporučení plynoucí ze simulačních studií, která jsou zjednodušena pro maximální praktičnost při používání. První doporučení říká, že pokud populaci tvoří méně než 100 jednotek, není důvod z ní tvořit výběr, ale je praktičtější podrobit výzkumu celou populaci (Gay, Mills, & Airasian, 2014, s. 139).<sup>11</sup> Za minimální velikost výběru vhodného pro kvantitativně orientovaný výzkum bývá považováno 30 jednotek, avšak pouze v případě, že tento

<sup>10</sup> Opět poznamenejme, že rozdíl o velikosti čtvrtiny směrodatné odchylky je již věcně poměrně nezajímavý.

<sup>11</sup> Stovka je v kvantitativním výzkumu téměř magická, protože platí, že ve výběru menším než sto jednotek není jedno procento reprezentováno ani jedním jedincem, a je tak obtížně představitelnou jednotkou.

soubor neplánujeme dělit pro účely vztahové analýzy na podskupiny (Gay et al., 2014, s. 139). Soukup a Rabušic (2007, s. 380, 385) uvádějí jako minimum 30–50 jednotek, pro větší kvalitu však 80–100.<sup>12</sup> Samozřejmě platí, že pokud chceme pracovat na úrovni podskupin, musíme minimální velikost vztáhnout na jednotlivé podskupiny.

Pro faktorovou či regresní analýzu se užívá pravidlo, které vychází z počtu použitých proměnných, resp. počtu odhadovaných koeficientů. Běžně je doporučeno, aby počet jednotek výzkumu (respondentů) byl 10násobkem počtu odhadovaných parametrů. Demonstrujme toto na dvou příkladech:

- 1) Regresní analýza s devíti prediktory – odhadujeme devět regresních koeficientů a jednu konstantu, výběr by tedy měl mít minimálně  $(9 + 1) * 10 = 100$  jednotek.
- 2) Explorační faktorová analýza s deseti indikátory a dvěma faktory – odhadujeme  $2 * 10$  faktorových zátěží, výběr by tedy měl mít  $20 * 10 = 200$  jednotek.

Z uvedeného pravidla plyne i běžně doporučené pravidlo pro stabilitu faktorového řešení zajistit minimálně 200 výzkumných jednotek (Tinsley & Tinsley, 1987; Comrey, 1988, cit. podle DeVellis, 2012, s. 157–158).

Obdobně se v případě realizace konfirmační faktorové analýzy doporučuje, aby počet zkoumaných jednotek byl minimálně desetinásobkem počtu parametrů, které musí být v daném modelu odhadnuty (srov. Schreiber et al., 2006, s. 326), stejná pravidla bývají formulována pro obecné strukturní modely. Doplňme, že tato pravidla vychází ze skutečnosti, že by měla být zajištěna dostatečná síla testů pro jednotlivé parametry. Zájemcům lze doporučit například Soperův webový kalkulátor (©2016).

Veškerá uvedená doporučení však platí pouze rámcově, vždy je nutné přihlížet ke konkrétním podmínkám výzkumu (více in Bertrand & Hughes, 2005, s. 65; Babbie, 2013, s. 125–126). Doplňme, že pokud náš výzkumný soubor nedosahuje požadované velikosti (stanovené dle výše uvedených pravidel), není to automaticky překážkou publikování výsledků (ideálně bez používání statistických testů) vypočtených z tohoto souboru. Autor by měl na tuto skutečnost upozornit a je pak na poučeném čtenáři, jak s tímto upozorněním naloží.<sup>13</sup>

<sup>12</sup> Samozřejmě je možné aplikovat kvantitativně orientované metody sběru dat i na menší skupiny nebo jednotlivce, avšak v takovém případě lze pracovat pouze na popisné rovině (např. sledovat výkon jednotlivce v kontextu s populačními normami).

<sup>13</sup> Děkujeme anonymnímu recenzentovi za tento podnět.

### 3.4 Dopad složitějších designů pravděpodobnostních výběrů na velikost výběru

Tabulka 1 shrnuje základní informace o nejpoužívanějších typech pravděpodobnostních výběrů (Kalton, 1983) a jejich obecné srovnání s prostým náhodným výběrem z hlediska požadavku na velikost. Detailnější pojednání o problematice včetně výpočetních postupů nalezne čtenář zejména v klasické monografii Kishé (2014) či modernější monografii Lévyho a Lemeshowa (2008).

Tabulka 1

Základní informace o nejpoužívanějších typech náhodných výběrů

Označení výběru	Stručný popis	Požadovaná velikost ve srovnání s prostým náhodným výběrem
stratifikovaný ( <i>stratified sampling</i> )	populace se rozdělí na co nejmenší vnitřně homogenní skupiny (subpopulace, strata) a z těch poté náhodně vybíráme jednotky	srovnatelná, pokud je výběr ze skupin proporční vzhledem k jejich velikosti
systematický ( <i>systematic sampling</i> )	do výběru je zahrnuta každá n-tá jednotka dostupné populace na základě požadované proporce mezi populací a výběrem	srovnatelná, pokud jsou jednotky v opoře výběru náhodně seřazeny
skupinkový ( <i>cluster sampling</i> )	populace se rozdělí na co nejmenší vnitřně heterogenní skupiny (svazky), z nich náhodně některé vybereme a v těchto skupinách pak využijeme všechny jednotky	výrazně horší (je třeba asi 2–3násobně velký výběr)
vícestupňový ( <i>multistage sampling</i> )	populaci postupně rozdělíme na co nejmenší administrativní skupiny (např. města – školy – třídy) a z nich náhodně vybereme jednotky	

V případě, že náš výběr není proveden jako prostý náhodný výběr, ale je užíván některý z výše popsaných složitějších designů pravděpodobnostních výběrů, musíme korigovat doporučení pro velikost výběrového souboru. Obecně bude v pedagogickém výzkumu platit, že budeme používat typicky vícestupňového, často v kombinaci se skupinkovým výběrem nebo stratifikací. Je tedy zapotřebí vzít v potaz doporučení uvedená v tabulce 1 a velikost výběru oproti doporučením pro prostý náhodný výběr z předchozí části textu navýšit. Nadto pokud plánujeme pracovat s jednotlivými skupinami vymezenými straty, musíme dbát na dostatečnou velikost jednotlivých skupin,



což může celkovou velikost výběru výrazně navýšit (zejména pokud chceme prezentovat výsledky samostatně za jednotlivé skupiny).

Na příkladu výzkumu PISA 2012 si ukažme, jak se provádí výběr žáků v tomto mezinárodním šetření. Detailní popis najde čtenář v publikaci OECD (2014, s. 65–88). Cílovou populací šetření PISA jsou patnáctiletí žáci, kteří jsou minimálně v 7. ročníku školní docházky. S ohledem na dostupnost i snadnější proces testování jsou z cílové populace pro výzkum zpravidla vyloučeny školy hůře dostupné, případně školy, které vzdělávají jen žáky se speciálními vzdělávacími potřebami; maximální míra tohoto vyloučení je stanovena na 5 %. Pro výběr konkrétních žáků je v šetření PISA užito *dvoustupňového stratifikovaného pravděpodobnostního výběru*. Stratifikace záleží na rozhodnutí každé země (typicky v ČR jsou separátně vybírány v jednotlivých státech základní školy a různé typy středních škol, výběr probíhá separátně v jednotlivých krajích a ve třech velikostních kategoriích škol). V prvním stupni se vybírají náhodně školy ze seznamu všech škol, kde mohou být 15letí žáci, s tím, že seznam škol se náhodně seřadí a poté je vybíráno systematickým výběrem. Pravděpodobnost vybrání škol navíc odráží velikost školy (tj. větší školy mají větší pravděpodobnost být vybrány než malé). V takto náhodně vybraných školách je ve druhém stupni ze všech patnáctiletých žáků náhodně vybráno 35 žáků, pokud je v příslušné škole méně než 35 žáků, pak jsou za výběr považováni všichni patnáctiletí žáci. Celkově je vybíráno v každé zemi minimálně 150 škol, minimální velikost výběru je stanovena na 4500 žáků a požadovaná návratnost (tj. reálná spolupráce škol) je 85 % (tj. minimálně 85 % vybraných škol se musí účastnit výzkumu, aby byla data z výzkumu PISA příslušné země zahrnuta do mezinárodní databáze a výsledků mezinárodní zprávy). Obdobně je stanovena i míra návratnosti pro druhý stupeň výběru, tj. pro žáky. Pro úplnost dodejme několik údajů o českém sběru dat PISA 2012. Šetření se účastnilo 297 (ze 300 náhodně vybraných škol) ze všech 14 krajů a k testování se dostavilo celkem 6 413 žáků. Návratnost na úrovni škol činila 98,15 % a návratnost na úrovni žáků činila 90 %. ČR tedy splnila všechny požadavky.

Praktická informace, která se váže k pravděpodobnostním výběrům, se týká jejich nákladnosti. Pokud budeme chtít pořizovat pravděpodobnostní výběr například z dospělé populace, pak zpravidla realizace tohoto úkolu nebude v silách výzkumného týmu, ale bude muset využít služeb výzkumných agentur. Ceny se samozřejmě budou lišit, ale lze odhadnout, že pouhý sběr dat u výzkumného souboru od tisícovky respondentů se bude pohybovat v řádu

mnoha set tisíc korun. Pro úplnost ještě dodejme, že například náklady na šetření PISA (příprava, pilotáž a sběr dat včetně přípravy národní zprávy, vše se realizuje během čtyř let) budou okolo deseti milionů. Proto pro získání dostatečně velkého a kvalitního pravděpodobnostního výběru musí získat výzkumník externí zdroje (typicky grantové prostředky či prostředky z Evropských fondů). Nákladnost je jedním z důvodů, proč se v praxi často používají výběry nepravděpodobnostní.

Doplňme zároveň, že v případě, kdy používáme složitější designy pravděpodobnostního výběru, musíme při zpracování dat tuto skutečnost zohlednit. Typicky pak nevystačíme s běžně užívaným softwarem (zejména SPSS či Excel), ale musíme používat speciální doplňky či zcela jiný software. S ohledem na komplexnost problému lze odkázat opět na technického průvodce PISA (OECD, 2014), případně na detailní diskusi této problematiky na datech TIMSS za ČR v článku Soukupa (2016). Mnoho problémů tedy neohrožuje výzkumníka jen na počátku výzkumu (při jeho přípravě), ale i na konci (při zpracování získaných dat). To by však nemělo výzkumníka odradit, ale obojí by měl přijmout jako, moderně řečeno, výzvu.

## 4 Velikost výběru v případě nepravděpodobnostních výběrů

V pedagogice oblíbené a nejčastěji aplikované nepravděpodobnostní výběry existují v několika typech, z nichž základní shrnuje tabulka 2, která obsahuje také jejich obecné srovnání s prostým náhodným výběrem z hlediska kvality (Kalton, 1983).

Tabulka 2

*Základní informace o nejpoužívanějších typech nepravděpodobnostních výběrů*

Označení výběru	Stručný popis	Kvalita ve srovnání s prostým náhodným výběrem
kvótní ( <i>quota sampling</i> )	záměrný výběr jednotek na základě známých vlastností populace podle daných kvót	horší, za předpokladu reprezentativity lze uvažovat o zobecnění na populaci
expertní ( <i>expert choice</i> )	výběr jednotek na základě úsudku výzkumníka	nesrovnatelně horší, ani za předpokladu reprezentativity nelze uvažovat o zobecnění na populaci
dostupný ( <i>convenience sampling</i> )	založený na dostupnosti/dobrovolnosti zkoumaných jednotek	nesrovnatelně horší, ani za předpokladu reprezentativity nelze uvažovat o zobecnění na populaci

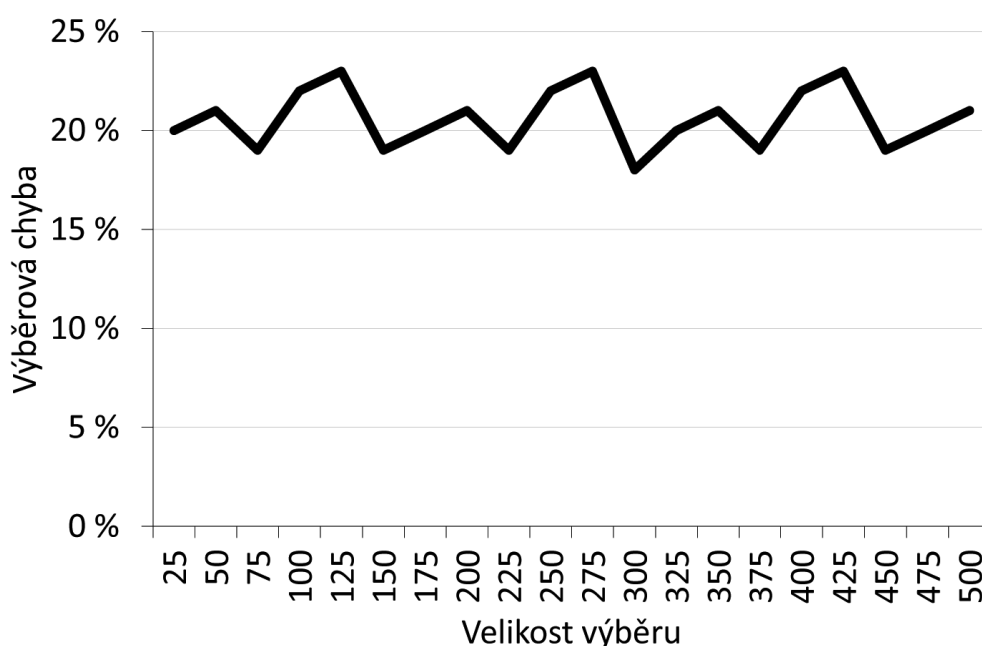
Z hlediska reprezentativity (viz první část článku) lze konstatovat, že reprezentativní bude z nepravděpodobnostních výběrů maximálně kvótní výběr. *Pro všechny nepravděpodobnostní výběry ovšem společně platí, že při práci s daty nelze automaticky používat statistické testy, protože nelze vyčíslit výběrovou chybu (srov. např. Kalton, 1983, s. 90). To ovšem neznamená, že pro tato data nelze používat vůbec statistické nástroje. Bez problémů lze využívat veškeré nástroje popisné statistiky, včetně počítání korelačních či kontingenčních koeficientů, případně regresních koeficientů. Obdobně lze využít nástrojů explorační statistiky, typicky explorační faktorové analýzy, seskupovací (shlukové) analýzy, mnohorozměrného škálování, explorační analýzy latentních tříd apod. Jediné, co není dovoleno, je používat statistické testy či intervaly spolehlivosti pro zobecnění výsledků z nepravděpodobnostních výběrů na cílovou populaci. Obecně platí, že je vždy nutné popsat, jak byl výběr proveden, jeho velikost a omezení. Platí, že s velikostí nepravděpodobnostních výběrů se nemění přesnost odhadů, tj. výběrová chyba zůstává zhruba konstantní (viz obrázek 5). Na základě uvedených skutečností lze podat některá obecná doporučení pro nepravděpodobnostní výběry:*

- 1) Pokud chceme zobecňovat výsledky na cílovou populaci, nepravděpodobnostní výběry neužívejme.
- 2) Nepravděpodobnostní výběry jsou vhodné pro indikaci možných výsledků a generování teorií<sup>14</sup>, nikoli k jejich ověřování.
- 3) Absolutní hodnoty (ve formě průměrů, mediánů, četností) z nepravděpodobnostních výběrů nelze zobecňovat na populaci a nelze je ani považovat za hodnoty, které budou populačním hodnotám blízké (s výjimkou kvótních výběrů).
- 4) Vztahy mezi veličinami (typicky zjišťované kontingenčními či korelačními koeficienty) budou i v případě nepravděpodobnostních výběrů zhruba pro populaci platné (samozřejmě za předpokladu, že náš výběr netvoří jedna zcela atypická podskupina cílové populace).
- 5) Z hlediska velikosti lze pro získávání indikativních výsledků či generování hypotéz doporučit provádět nepravděpodobnostní výběry spíše menší (v řádu do cca 150 jednotek) a tyto výsledky či hypotézy poté ověřovat na reprezentativních pravděpodobnostních výběrech.

---

<sup>14</sup> Tímto vyjádřením nezpochybňujeme úlohu kvalitativního výzkumu při generování teorií.

Připomeňme, že v případě nepravděpodobnostních výběrů neplatí výše uvedené pravidlo snižování výběrové chyby (srov. obrázek 2), že s růstem velikosti klesá (nelineárně) *výběrová chyba*. Tato chyba *se s velikostí v zásadě nemění a osciluje kolem neznámé hodnoty* (srov. ilustraci v obrázku 5). Hodnota výběrové chyby je v případě nepravděpodobnostních výběrů neznámá, protože výše uvedené vzorce zde neplatí a nelze ji tedy vyčíslit.



Obrázek 5. Velikost výběrové chyby pro proporci s ohledem na velikost výběru pro nepravděpodobnostní výběr.

Pro ilustraci skutečnosti, že velikost výběrového souboru nehraje u nepravděpodobnostních výběrů roli, připomeňme jeden příklad z české praxe, konkrétně šetření *Eurostudent* provedené v roce 2007 (Pabian & Provázková, 2008, s. 15).<sup>15</sup> Výzkumný soubor byl nejdříve získáván skrze pop-up okno na portálu centrum.cz a několika dalších webech, kde byl očekáván výskyt osob mladšího věku. Poté, co bylo zjištěno výrazné vychýlení takto získaného samovýběru (měl cca 1700 jednotek), došlo k dodatečnému výběru dalších studentů z databáze firmy iVýzkumy a z dalších zdrojů. Celkově tak bylo

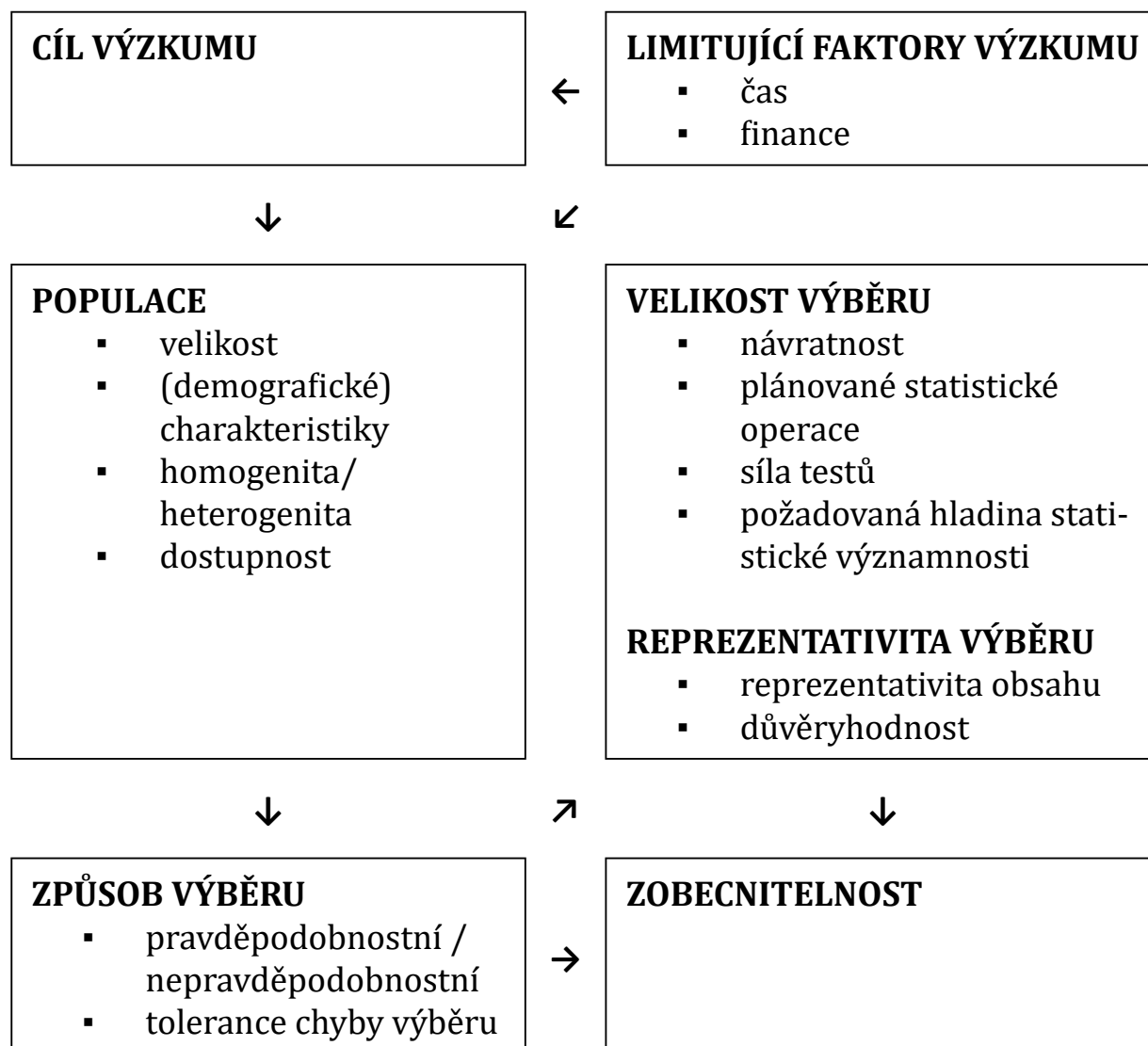
<sup>15</sup> Šetření *Eurostudent* je evropským srovnávacím šetřením podmínek studia vysokoškolských studentů. Jedním ze základních požadavků kladených na toto šetření je reprezentativita výběrového souboru, tj. soubor by měl odrážet strukturu vysokoškolské populace v jednotlivé zemi, ideálně díky náhodnému vybírání studentů.

získáno 3355 vyplněných dotazníků od vysokoškolských studentů. Na první pohled impresivně velký výběrový soubor (spíše samovýběrový) však nebylo možné považovat za reprezentativní, což bylo ostatně konstatováno i v recenzním řízení a poprvé v historii tak nebyla česká data odeslána do mezinárodní databáze. Bylo tak promarněno nemálo finančních prostředků i času jednotlivých aktérů včetně studentů, kteří v dobré víře a zdarma (nicméně s příslibem možného vítězství v loterii pro účastníky) vyplňovali dotazníky. Z tohoto příkladu plyne minimálně dvojí poučení:

- 1) velikost výběru negarantuje jeho kvalitu;
- 2) samovýběry a jiné nepravděpodobnostní výběry nelze využívat pro zobecnění na populaci.

## 5 Shrnutí požadavků

S oporou o schéma na obrázku 6 integrujeme základní požadavky vztahující se na velikost a reprezentativitu výběru v kvantitativně orientovaném pedagogickém výzkumu, a to v kontextu dalších parametrů tvorby výběru a jeho kvality. Velikost a reprezentativitu výběru udává v první řadě cíl výzkumu. Na jeho základě je nutné definovat populaci a co nejpřesněji specifikovat její velikost, (demografické) charakteristiky, zvážit míru homogenity/heterogenity, ale také její dostupnost. Je nutné vzít v úvahu limitující faktory jako čas a finance, případně zúžit cíl výzkumu i cílovou populaci. Posléze je třeba určit způsob tvorby výběru (pravděpodobnostní či nepravděpodobnostní a jejich různé podoby) a míru tolerance chyby výběru. V případě pravděpodobnostního výběru, ze kterého lze zobecňovat na populaci, je třeba uvažovat plánované statistické operace a sílu testů, které plánujeme použít v rámci analýzy. S využitím výše uvedených informací je možné odhadnout minimální potřebnou velikost výběru, avšak vždy s ohledem na předpokládanou návratnost. Získaný výběr lze posuzovat s využitím statistických metod z hlediska reprezentativity a sledovat, jak věrnou kopii populace představuje. Reprezentativitu ovlivňuje také důvěryhodnost dat. Zobecnitelnost výsledků na populaci pak lze zvažovat s ohledem na získané parametry výběru.



Obrázek 6. Kontext požadavků vztahujících se na velikost a reprezentativitu výběru.

## Závěr

Pozornost této studie je věnována základním požadavkům na velikost a reprezentativitu výběru v kvantitativně orientovaném pedagogickém výzkumu. Tyto dva parametry kvality výběru považujeme za zásadní, avšak závisí na celé řadě dalších faktorů, které prolínají celým výzkumným designem. Pokusili jsme se je shrnout, schematizovat a doplnit o odkazy na relevantní literaturu včetně uvedení příkladů z reálných pedagogických studií.

Samozřejmě že s ohledem na komplexnost tématu a omezený rozsah textu je předložený článek jen úvodem do tématu a složitější postupy vyžadují od čtenáře další studium a naučení se novým postupům, případně novým softwarům.

V současnosti, kdy se výzkum stále více internacionalizuje, považujeme za důležité, aby se česká pedagogika více orientovala na výzkum, jehož výsledky lze s určitými limity zobecnit na širší populaci. Zároveň však chceme zdůraznit, že zobecnitelnost je velmi subtilní konstrukt, jehož funkčnost v pedagogickém výzkumu nelze jednoznačně zaručit, a není jediným kritériem kvality kvantitativně orientovaného výzkumu. Ať už je či není záměrem výzkumníků zobecnit výsledky výzkumu „za hranice“ jejich výběru, v každém případě je třeba věnovat v každé studii dostatek prostoru informacím o cílové populaci, procesu tvorby výběru a jeho kvalitě nejen z hlediska velikosti a reprezentativity (APA, 2010, s. 247–248).

## Literatura

- APA. (2010). *Publication manual of the American Psychological Association*. Washington: American Psychological Association.
- Babbie, E. R. (2013). *The practice of social research*. Belmont: Wadsworth Cengage Learning.
- Bertrand, I., & Hughes, P. (2005). *Media research methods: Audiences, institutions, texts*. New York: Palgrave Macmillan.
- Bryman, A. (2012). *Social research methods*. New York: Oxford University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science*. Hillsdale: Erlbaum.
- DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks: Sage.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.
- Fink, A. (2013). *How to conduct surveys: A step-by-step guide*. Los Angeles: Sage.
- Gay, L., Mills, G. E., & Airasian, P. (2014). *Educational research: Competencies for analysis and applications*. Harlow: Pearson Education.
- Heinrich-Heine-Universität Düsseldorf. (2016). *G\*Power: Statistical power analyses for Windows and Mac* [online]. Dostupné z <http://www.gpower.hhu.de/>
- Hendl, J. (2015). *Přehled statistických metod: analýza a metaanalýza dat*. Praha: Portál.
- Jelínek, M., Květoň, P., & Vybořil, D. (2011). *Testování v psychologii: teorie odpovědi na položku a počítačové adaptivní testování*. Praha: Grada.
- Kalton, G. (1983). *Introduction to survey sampling*. Thousand Oaks: Sage.
- Kish, L. (2014). *Survey sampling*. New York: Wiley.
- Krejčí, J. (2008). *Kvalita sociálněvědních výběrových šetření v České republice*. Praha: Sociologické nakladatelství (SLON).
- Krejčí, J. (2011). *Kvalita výběrových šetření a standardy v sociálním výzkumu*. Praha: Sociologický ústav AV ČR.

- Lafaye de Micheaux, P., & Tran, V. A. (2016). PoweR: A reproducible research tool to ease Monte Carlo power simulation studies for goodness-of-fit tests in R. *Journal of Statistical Software*, 69(3), 1–42.
- Levy, P. S., & Lemeshow, S. (2008). *Sampling of populations: Methods and applications*. New York: Wiley.
- Mareš, P., Rabušic, L., & Soukup, P. (2015). *Analýza sociálněvědních dat (nejen) v SPSS*. Brno: Masarykova univerzita.
- Mertens, D. M. (2015). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Los Angeles: Sage.
- OECD. (2014). *PISA 2012 technical report*. Paris: OECD.
- Pabian, P., & Provázková, K. (2008). Přehled sociologických šetření CSVŠ 1991–2008. *AULA* 16(3), 1–24.
- Řehák, J. (1978). K pojmu „reprezentativita“ v sociologických výzkumech. *Sociologický časopis/Czech Sociological Review*, 14(5), 489–507.
- Řehák, J. (1998a). Kvalita dat I. Klasický model měření reliability a jeho praktický aplikační význam. *Sociologický časopis/Czech Sociological Review*, 34(1), 51–60.
- Řehák, J. (1998b). Kvalita dat II. Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur. *Sociologický časopis/Czech Sociological Review*, 34(2), 195–204.
- Saunders, M., Lewis, P., & Thornhill, A. (2007). *Research methods for business students*. Harlow: Prentice Hall/Financial Times.
- Sedláková, R. (2014). *Výzkum médií: nejužívanější metody a techniky*. Praha: Grada.
- Schenk, J., & Hrabovská, A. (2010). *Škálovanie. Základné jednodimenzionálne metódy*. Bratislava: Univerzita Komenského.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120. Dostupné z <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2792363/>
- Soper, D. (2016). *A-priori sample size calculator for structural equation models* [online]. Dostupné z <http://www.danielsoper.com/statcalc/calculator.aspx?id=89>
- Soukup, P. (2013). Věcná významnost výsledků a její možnosti měření. *Data a výzkum-SDA Info (Data and Research-SDA Info)*, 7(2), 125–148.
- Soukup, P. (2016). Možnosti praktické práce s daty z mezinárodních vzdělávacích studií: problémy a jejich praktická řešení. *Orbis Scholae*, 10(1), 97–120.
- Soukup, P., & Rabušic, L. (2007). Několik poznámek k jedné obsesi českých sociálních věd – statistické významnosti. *Sociologický časopis/Czech Sociological Review*, 43(2), 379–395.
- StatsToDo Trading Pty Ltd. *Computer program to calculate sample size requirement for estimating the correlation coefficient*. StatsToDo [online]. Dostupné z [http://www.statstodo.com/SSizCorr\\_Pgm.php#Single calculation : sample size estimation](http://www.statstodo.com/SSizCorr_Pgm.php#Single%20calculation%20-%20sample%20size%20estimation)
- Štuka, Č., Martinková, P., Vejražka, M., Trnka, J., & Komenda, M. (2013). *Testování při výuce medicíny. Konstrukce a analýza testů na lékařských fakultách*. Praha: Karolinum.



- Urbánek, T., Denglerová, D., & Širůček, J. (2011). *Psychometrika: měření v psychologii*. Praha: Portál.
- Vogt, W., & Johnson, B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences*. Thousand Oaks: Sage.
- Zvára, K. (2002). Měření reliability aneb Bacha na Cronbacha. *Informační bulletin České statistické společnosti*, 13(2), 13–20. Dostupné z <http://www.statspol.cz/bulletiny/ib-02-2.pdf>

## Autoři

PhDr. Ing. Petr Soukup, Univerzita Karlova v Praze, Fakulta sociálních věd, Institut sociologických studií, U Kříže, 158 00 Praha 5, e-mail: soukup@fsv.cuni.cz

Mgr. Ilona Kočvarová, Ph.D., Univerzita Tomáše Bati ve Zlíně, Fakulta humanitních studií, Ústav školní pedagogiky, Mostní 5139, 760 01 Zlín, e-mail: kocvarova@fhs.utb.cz

## Size and representativeness of research sample in quantitatively oriented educational research

**Abstract:** This methodologically oriented study summarizes the basic requirements for the size and representativeness of research sample in quantitatively oriented educational research. The study presents basic sampling techniques, their characteristics and impact on the required sample size and the possibility to generalize results. Not only standard designs of probability sampling (simple random, multistage, cluster and stratified) but also non-probability sampling designs are discussed in this manner. The study provides approaches for determining the statistical power and summarizes some basic statistical recommendations for the sample size with respect to the statistical power. Attention is also paid to the issue of sample representativeness. Examples of good and bad practice in the field of international educational research are included in the study.

**Keywords:** research, population, research sample, size of research sample, representativeness, generalizability