

6.1 ZPŮSOBY VÝBĚRU	126
6.2 PRAVDĚPODOBNOSTNÍ VÝBĚR	127
6.2.1 Prostý náhodný výběr	127
6.2.2 Komplexní náhodné výběry	129
6.3 NEPRAVDĚPODOBNOSTNÍ VÝBĚRY	130
6.3.1 Kvótní výběr	130
6.3.2 Dostupný výběr	131
6.4 VÝBĚR RESPONDENTŮ NA INTERNETU	134
6.5 NÁVRATNOST	134
6.6 SEKUNDÁRNÍ ANALÝZA DAT	136
Shrnutí	138
Doporučená literatura	139

Za doplnění psychologických aspektů děkuji doc. PhDr. Ivě Poláčkové Šolcové, Ph.D., (rámeček 6.4).

Nejčastější motivací výzkumníka při navrhování způsobu **výběru** případů v kvantitativní výzkumné strategii je možnost zobecnění zjištění z vybraného souboru na celou populaci. Obvykle termínem **populace** označujeme určitou jasně vymezenou skupinu lidí. Tedy například všechny občany České republiky, všechny studenty pražských středních škol či každého obyvatele daného města staršího 65 let. Ptát se však každého v dané populaci by bylo příliš drahé až neproveditelné. Techniky kvantitativního šetření spolu se statistickými nástroji nám ale při dodržení určitých předpokladů umožní učinit spolehlivé závěry o populaci i na základě šetření pouze malého zlomku jejich členů.

Problém zobecnění z výběru na celou populaci lze nejzřetelněji ukázat na mediálně často přetřásaném tématu, totiž spolehlivosti a problémů předvolebních výzkumů politických preferencí. Výzkum volebních preferencí vlastně představuje snahu na základě zjištění z určitého relativně malého počtu respondentů usuzovat o budoucím volebním chování celé populace, tedy chování všech oprávněných voličů. Klíčem k úspěšné předpovědi je přitom nejčastěji schopnost výzkumných agentur přiblížit se k předpokladu užití statistických metod zobecnění, tedy k náhodnému výběru respondentů.

Oproti obvyklému intuitivnímu úsudku není ze statistického hlediska pro spolehlivost zobecnění podstatné to, jak velkou část populace do výzkumu zahrneme, ale pouze absolutní velikost výběru. Jinak řečeno, za předpokladu dobře provedeného náhodného výběru můžeme z dotazování 1 000 respondentů získat stejně přesnou informaci o populaci čítající 100 000, milion i sto milionů jedinců (více viz kapitola 10). Proto je nedůvěra k výzkumům veřejného mínění založená na argumentu „mě se nikdo nikdy neptal“ nepřilíš opodstatněná.¹

Rámeček 6.1 → Úplný soubor

V některých případech máme možnost pracovat s **úplným souborem**, tedy s daty zahrnujícími celou populaci, a není tak nutné vybírat pouze část. Může se například jednat o cenové statistiky (sčítání lidu, ale i jiné výstupy statistických úřadů), databáze různých organizací (např. katalog knihovny, seznam prodávaného zboží) či internetových serverů (např. všechna data na Wikipedii, ČSFD.cz). Možnosti práce s tímto typem dat se znatelně rozšiřují v kontextu zvyšování výpočetní efektivity počítačů a s požadavkem na otevřenou státní správu (tzv. otevřená data), ale i s rostoucím množstvím dat, která jsou o lidském chování generována a ukládána v důsledku širšího užívání elektroniky v každodenním životě (mobilní aplikace, platební karty, komunikace na sociálních sítích apod.).

¹ Pro příklad, počítáme-li, že dospělých obyvatel České republiky je přibližně 8 500 000. Pokud je obvyklá velikost výběrového souboru dotazníkových šetření zhruba 1 000 respondentů, znamená to, že šance být jedním výzkumem dotazován je 1 : 8 500. Pokud bychom předpokládali, že se každý rok uskuteční třeba sto reprezentativních dotazníkových šetření, s takovou šancí ani po 50 letech nepřekročí pravděpodobnost, že budete do výzkumu náhodně vybráni, 50 %.

Pokud pracujeme s úplným souborem, není samozřejmě nutné provádět výběr. Zároveň není nutné využívat jinak klíčové statistické procedury zaměřené na zobecňování z výběru na populaci. Pokud jsou data přímo o populaci, tento typ statistické inference je nadbytečný (viz kapitola 10) a zpracování dat se soustředí zejména na užití nástrojů, které umožňují smysluplně zprostředkovat přehled o těchto obvykle extrémně rozsáhlých datových souborech.

Připomeňme si, že populací ve výzkumu zdaleka nemusíme rozumět pouze množinu lidí (respondentů). Zkoumanými případy mohou být třeba i města, státy, novinové články, školy, facebookové statusy, umělecká díla apod., vždy v závislosti na tématu výzkumu a nastavené logice zkoumání. V každém případě je ale pro výzkum potřeba jednoznačně vymezit, které případy do zkoumané populace spadají a které již nikoliv.

Představme si například na první pohled prostý výzkum studentů středních škol v Praze. O jakou populaci se ale vlastně jedná? Měli bychom jasně vymezit ročník studia, který budeme dotazovat. Pokud plánujeme sběr dat v červnu, nemůžeme pak už moc například počítat s tím, že bychom dotazovali maturanty. Jsou součástí výzkumu i studenti víceletých gymnázií? Pokud ano, nezapomeneme dotazované třídy omezit na ročníky odpovídající střední škole? Zaměříme se pouze na veřejné školy (zřizované městem), nebo i na soukromé či církevní? Zajímá nás pouze denní studium, nebo i dálkové? Nezapomeneme na to, že student střední školy často nemusí být zároveň obyvatelem Prahy? Je to pro náš výzkum důležité? A že některé nematuritní obory nemusejí být čtyřleté? Jak je vidět, i zdánlivě jasné vymezení populace může skrývat řadu otázek, které musíme před plánováním sběru dat zodpovědět. Tyto odpovědi nás přitom často mohou přinutit vymezení výzkum posunout.

6.1 ZPŮSOBY VÝBĚRU

V případě, že výzkum nemůže analyzovat data o celé populaci, přistupujeme ve výzkumu ke konstrukci **výběru**. Plánování a volba způsobu výběru jsou často vedeny kompromisem mezi náročností plánovaného výzkumu a ambicí, již na výzkum a kvalitu zobecnění z výběru klademe. Náročnost konkrétního výběrového postupu je přitom často specifická pro různé situace a prostředí. Zatímco provést prostý náhodný výběr ze všech vydání *Lidových novin* v roce 2018 je poměrně snadné, vybrat náhodně respondenty ze všech dospělých obyvatel ČR je extrémně náročné až nemožné. Stejně tak když zkoumáme žáky ve školách, obvykle se vyplatí dotazovat najednou celou třídu a nelosovat jen jednoho či dva žáky. Všechny tyto pragmatické aspekty zasahují do rozhodování o výzkumném plánu a výrazně formují výslednou podobu výzkumu.

Jak již bylo řečeno výše, zobecnění z výběru na populaci lze spolehlivě provést prostřednictvím statistických nástrojů pouze v případě, že výběrový soubor byl z populace vybrán za splnění podmínek náhodného výběru. Základní dělení proto výběrové postupy rozlišuje na pravděpodobnostní a nepravděpodobnostní, tedy na ty, které tento požadavek naplňují a které nikoliv.

6.2 PRAVDĚPODOBNOSTNÍ VÝBĚR

6.2.1 Prostý náhodný výběr

Požadavek náhodného výběru nejuplněji naplňuje postup nazývaný **prostý náhodný výběr**, který je možné velmi snadno teoreticky vymezit, ale v řadě případů je jeho reálné využití značně náročné či nemožné. Při prostém náhodném výběru postupujeme takovým způsobem, že každá jednotka populace má zcela stejnou pravděpodobnost být zařazena do výběru. Pokud bychom například chtěli zkoumat složení městských rad, můžeme ze seznamu měst v ČR náhodně vybrat požadovaný počet případů a ty zařadit do výběru. Podobně ve výše uvedeném případě výběru z novin můžeme sestavit seznam všech dnů v roce, kdy noviny vycházejí, a z nich opět vylosovat vymezený počet.

Podmínkami náhodného výběru ve společenskovědním výzkumu jsou existence a dostupnost **opory výběru**, tedy úplného seznamu všech jednotek v dané populaci. Jinak totiž obvykle nejsme schopni navrhnout takový výběrový postup, který by spolehlivě zaručoval, že každá jednotka má opravdu stejnou šanci být vybrána.

Ve skutečnosti možnost vybírat z populace podle pravidel prostého náhodného výběru, především pokud se jedná o výběr lidí (respondentů), není příliš častá. Často neexistuje spolehlivá opora výběru nebo taková opora není dostupná, a to ani pro účely akademického výzkumu. I pokud ale takovou oporu máme, může být její využití příliš drahé či neefektivní. Jde například o ono dotazování jednoho či dvou žáků ve třídě, jak bylo uvedeno výše. Proto se při výběrových šetřeních používá řada dalších pravděpodobnostních postupů, které se sice od ideálu prostého náhodného výběru vzdalují, ale stále jsou statisticky modelovatelné, a tudíž i dostatečně spolehlivé.

Rámeček 6.2 → Slovníček náhodného výběru

Technika náhodného výběru používá určitou zavedenou specializovanou terminologii. Jelikož provedení reprezentativních náhodných výběrů z populace je obvykle možné pouze prostřednictvím spolupráce se specializovanými výzkumnými agenturami, je znalost jazyka této metody klíčová pro porozumění zadání i výstupů v podobě technických zpráv dotazníkových šetření.

Populace: Množina všech jednotek, které jsou předmětem výzkumu. Jednotky jsou vymezené pro dané šetření, nejčastěji se jedná o jedince (respondenty), ale stejně tak může jít o domácnosti, obce, firmy, novinové články apod. Při dotazníkovém šetření bývá chápání populace často omezeno dostupností jednotlivců. Například předvolební výzkumy stranických preferencí z praktických důvodů obvykle nemohou zahrnout potenciální voliče, kteří dlouhodobě pobývají v zahraničí.

Výběrový soubor: Malá část populace, která je předmětem šetření. Na základě náhodně vybraného výběrového souboru odhadujeme charakteristiky celé populace.

Opora výběru: Seznam všech jednotek dané populace. Na jeho základě můžeme náhodně vybírat ty, které budou předmětem šetření.

Výběrová chyba: Pravděpodobná velikost odchylky, kterou mohou mít charakteristiky náhodně vybraného souboru od skutečných (neznámých) charakteristik celé populace. Jinak řečeno, jak moc se můžeme v našem výběrovém šetření mýlit jen proto, že máme data pouze o náhodně vybrané části (více viz v kapitole 10).

Rámeček 6.3 → Jak se ze seznamu vybírá náhodně?

Náhodný výběr by měl být ekvivalentem hodů kostkou. Obvykle ale není možné ani praktické vyrábět kostku, která by počtem stěn odpovídala počtu jednotek ve zkoumané populaci. V dnešní době však již díky počítačovým programům není problém náhodně vygenerovat libovolně velký soubor případů ze zadaného seznamu (přesněji pseudonáhodný, ale to pro naše účely bohatě stačí).

Tento mechanický, neosobní prvek při generování náhodného výběru je velmi podstatný. Náhodný totiž není nahodilý. To, co považujeme za náhodnou volbu, s sebou obvykle nese řadu nežádoucích systematických zkreslení. Pokud bychom chtěli do našeho výzkumu náhodně vybírat kolemjdoucí, nejspíš se budeme vědomě či nevědomě častěji zaměřovat na osoby, které působí otevřeně, přístupně či jinak sympaticky.

Ale i při eliminaci těchto vnějších charakteristik lze ukázat, že lidské myšlení obsahuje řadu zavádějících intuicí, které se od matematicky přísné náhodnosti značně odchyľují. Představme si, že z fronty sta lidí máme pro výzkum vybrat náhodných deset. Bude nám jako náhodný připadat takový výběr, kde se vyskytne dvojice lidí, kteří ve frontě stáli hned za sebou? Nebo dokonce ten, kde takové dvojice budou dvě? Řada z nás by nejspíš intuitivně považovala takový výběr za nějakým způsobem zkreslený a chybně zkonstruovaný. Přitom při čisté náhodném losování by taková dvojice nebyla nějak neobvyklá. Dokonce by se měla objevit téměř ve dvou třetinách případů. A téměř pětina potenciálně provedených výběrů by takové dvojice měla obsahovat dokonce dvě.

6.2.2 Komplexní náhodné výběry

Problémy a náklady spojené s organizací sběru dat často v praxi vedou k tomu, že při plánování způsobu výběru jsou užívány postupy, které se sice odchyľují od výše zmíněného ideálu prostého náhodného výběru, zároveň ale zachovávají možnost tuto odchylku statisticky popsat a vyčíslit. Jedná se zejména o užití shlukování a stratifikace.

Shlukování (někdy též skupinkový výběr) se využívá v momentě, kdy jsou respondenti přirozeně shromážděni v určitých skupinách. Může se například jednat o obce (pokud zkoumáme celou populaci), školy (při zaměření na žáky), firmy (na zaměstnance) a jiné. V takovém případě můžeme nejdříve náhodně vybrat určitý počet těchto skupin a v dalších krocích výběru pokračovat pouze v rámci této množiny. Nejdříve tedy můžeme z rejstříku všech škol v zemi náhodně vybrat určitý počet škol, kde budeme provádět výzkum, a žáky dotazovat pouze v rámci těchto vybraných škol. Tento postup často výrazně zefektivňuje sběr dat. Tazatelé nemusejí cestovat do malých obcí jen kvůli jednomu výzkumnému rozhovoru, ale rovnou jich zde realizují třeba pět. Stejně tak není nutné domlouvat vyplnění dotazníku ve škole kvůli jednomu náhodně vybranému žákovi, ale rovnou dotazujeme žáky v jedné nebo více třídách.

Využití shlukování ale také přináší riziko nárůstu výběrové chyby. Čím méně skupin výběr tvoří a čím jsou skupiny uvnitř více homogenní, tím větší chyba může být. Pokud například existují velké rozdíly mezi výkonem studentů jednotlivých škol, může se při náhodném výběru malého počtu škol lehce stát, že náhodou budou vybrány školy se studenty s velmi dobrými či naopak velmi slabými výsledky. Aby byl tento postup smysluplný, je potřeba uvažovat spíše o desítkách či stovkách skupinek, proto se obvykle nejedná o řešení použitelné v malých studentských projektech.

Stratifikování výběru představuje postup, který naopak míru výběrové chyby nezvyšuje. Při tomto postupu je celá populace rozdělena do určitých kategorií (strat), které poskytují rámec pro další kroky výběru. Aby mělo smysl takto postupovat, musíme znát podíl jednotlivých kategorií v celé populaci. Příkladem může být již zmiňovaný výzkum žáků škol, kde při výběru můžeme postupovat stratifikovaně podle typu škol. Zjistíme podíl studentů navštěvujících jednotlivé typy škol (gymnázia, obchodní akademie, střední odborné školy atd.) a v rámci těchto typů je pak v dalším kroku vybírán odpovídající počet respondentů k oslovení. Na rozdíl od kvótního výběru (viz dále) se ale jedná stále o pravděpodobnostní postup. V dalším kroku musíme jednotlivé respondenty stále vybírat náhodně.

Výše uvedené postupy lze v některých situacích kombinovat do tzv. **komplexních designů výběru**, které umožní maximalizovat efektivnost sběru dat s ohledem na časové a finanční náklady. Můžeme se tak setkat s **vícetupňovými výběry**, které na různých úrovních kombinují shlukování, stratifikaci i náhodný výběr. Konkrétní návrh je ale vždy kombinací našich prostředků, požadované velikosti výběrového souboru a povahy zkoumané populace.

6.3 NEPRAVDĚPODOBNOSTNÍ VÝBĚRY

V určitých situacích není možné navrhnout postup, který by splňoval požadavky náhodného výběru, případně by takový postup značně přesahoval zdroje, které jsou pro výzkum vyhrazené. V takovém případě lze zvažovat několik možných cest nepravděpodobnostního výběru, byť možnost zobecnění z dat získaných touto cestou bude vždy značně limitovaná.

Při striktním výkladu nelze u dat získaných nepravděpodobnostním výběrem používat nástroje statistické inference, tedy zobecnění výsledků na populaci. Základním předpokladem těchto nástrojů je totiž právě náhodný výběr případů z populace, který u nepravděpodobnostních výběrů chybí. V praxi se však můžeme setkat s užitím takových statistických metod i u datových souborů sebraných například kvótním výběrem, a i když se nejedná o korektní postup a měli bychom při něm vždy postupovat opatrně, ve výsledku se často nemusí jednat o zásadní problém.

6.3.1 Kvótní výběr

Většina levnějších reprezentativních dotazníkových šetření zaměřených na celou dospělou populaci je obvykle realizována právě prostřednictvím kvótního výběru. **Kvótní výběr** vychází z předpokladu, že pokud náš výběr bude ve svých klíčových charakteristikách kopírovat strukturu cílové populace, lze na jeho základech provádět poměrně spolehlivé zobecnění. Pokud se tedy například zastoupení mužů a žen, jednotlivých věkových či vzdělanostních skupin ve výběru bude shodovat s odpovídajícím zastoupením v celé populaci, je soubor respondentů sestaven vhodně a můžeme jej považovat za poměrně spolehlivý. Podmínkou kvótního výběru je, že musíme znát strukturu celé populace pro použité kvóty (v příkladu výše tedy pohlaví, věk a vzdělání). Kvótou tedy může být pouze to, co je již zmapováno na základě úplných statistik, tedy např. ze sčítání lidu.

V praxi se při plánování stanoví, jaké kvóty budou při sběru dat sledovány, a podle toho mohou jednotliví tazatelé dostat předpis, kolik výzkumných rozhovorů má být uskutečněno s muži či ženami, kolik s lidmi jednotlivých vzdělanostních stupňů apod. Tazatel pak hledá respondenty jakkoli, pouze s ohledem na naplnění předepsaných kvót.

Realizace výzkumu založeného na kvótním výběru je podstatně levnější než u náhodných výběrů. Na druhou stranu kvalita dat a možnosti zobecnění jsou značně nespolehlivé. Problematický je i předpoklad, že zvolené kvóty jsou tím klíčovým, co strukturuje populaci v otázce, kterou ve výzkumu řešíme, ať již je jakákoliv. Akademické projekty proto většinou upřednostňují použití některé z forem pravděpodobnostních výběrů.

6.3.2 Dostupný výběr

Dotazovat se můžeme zcela prostě i kolemjdoucích na ulici, studentů na chodbě či v učebně, lidí, kteří náhodou otevřeli naše webové stránky. V takovém případě hovoříme o **dostupném výběru**.² Tento postup má samozřejmě jednu výhodu – je nenáročný. Stojí nás relativně malé úsilí, nemusí respondenty složitě hledat, cestovat za nimi. Zároveň má ale jednu zcela zásadní nevýhodu – nemáme žádné vodítko pro zobecňování dat získaných tímto způsobem na jakoukoliv populaci.

Pokud se postavíme na ulici, nejspíš budeme často potkávat lidi, kteří v okolí pracují, bydlí či chodí nakupovat. Obvykle přitom nevíme, jaká všechna specifika dané místo má. Je nedaleko velký kancelářský objekt nebo velká škola? Je zde spíše drahé či levné bydlení? Objeví se zde i lidé z jiných částí města, nebo ne? A co ti, kteří jezdí výhradně autem? O čem jiném než jen o skupině daných dotazovaných lidí budou vlastně vypovídat takto získaná data? Z těchto důvodů se jedná o postup, který nemá pro většinu výzkumů usilujících o dosažení zobecnění na populaci příliš valný význam.

Dostupný výběr je odůvodnitelný ve specifických případech, například se jedná o častý postup při realizaci psychologických experimentů (viz rámeček 6.4). V určitém slova smyslu se pak tento postup blíží účelovému výběru (viz kapitola 12.4), jelikož se výběr zaměřuje spíše na naplnění základních kritérií důležitých pro danou studii než na reprezentativitu z hlediska celé populace. Přestože je tato praxe v posledních letech předmětem kritické diskuze, je spíše neobvyklé, že by účastníci experimentů byli vybíráni náhodným výběrem. Zároveň je ale patrná snaha hledat účastníky experimentů i mezi méně pohodlně dostupnými skupinami, než jsou studenti. Kromě experimentů je dostupný výběr odůvodnitelný spíše jen u přípravných fází výzkumu, tedy například pro pilotáž či testování jednotlivých otázek (viz rámeček 6.4, viz kapitola 7).

Rámeček 6.4 → Dostupný výběr v psychologickém výzkumu

V psychologii (ale nejen tam) je dostupný výběr využíván poměrně často například pro tzv. sondy, předběžné a pilotní studie v kvantitativní výzkumné strategii (srov. Howitt, Cramer 2011). Autory takových výzkumů jsou často experti na vysokých školách, kteří jako svoje výzkumné subjekty využívají studenty – osoby, které jsou pro ně dobře dostupné. Výběr je tedy v mnoha ohledech především praktický: účastníci studie jsou dobře dostupní z hlediska času, místa a ochoty se účastnit studie třeba v rámci plnění studijních povinností. Vždy ale záleží na vytyčeném výzkumném problému. Pokud psychologové například uvažují o nějakém univerzálním mechanismu psychiky, je dostupný výběr z hlediska praktičnosti ideální k prvotnímu zkoumání. Pokud se projeví

² V české literatuře se také někdy hovoří o příležitostném výběru, tj. výběru, který je založen na využití určité příležitosti. Anglický ekvivalent je v obou případech convenience sample.

mechanismus na dostupném souboru studentů (např. frustrace vždy vyvolává agresii), mohl by platit i jinde a může být ověřován i v dalších „nedostupných“ vzorcích. Pokud je ale záměrem výzkumu porozumět například generativitě čtyřicátníků, studentský vzorek není ideální a ani trochu praktický. Psychologové na dostupném výběru ale ověřují také metodologii, metodologické postupy, techniky sběru/tvorby dat, porozumění otázkám, položkám, podnětovému materiálu ad. S výsledky z výzkumů, v nichž byla data získávána prostřednictvím dostupného výběru, je pracováno jako s „předběžnými zjištěními“, která je nutno ověřit i na jiných (větších nebo naopak menších a specifitějších) výzkumných souborech. Výzkumník svoje zjištění vždy prezentuje s akcentací limitů šetření a uvažuje možná zkreslení, která jsou právě dostupným výběrem dána: např. v případě studentů vysoké školy se jedná o zkreslení vyplývající z věku, vzdělání, socioekonomického statusu, pohlaví atd. Je velkým nedorozuměním (nejenom) v psychologii, pokud jsou tyto sondy, pilotáže a předběžná zjištění chápány jako např. potvrzení nějakého psychického mechanismu či (ne)platnosti nějaké teorie.

Rámeček 6.5 → Šetření malého rozsahu

Jak je zřetelné z výkladu v této kapitole, pokud plánujeme dotazníkové šetření, je realizace náhodného či alespoň kvótního výběru v rozsahu, který je obvykle nutný pro relevantní kvantitativní analýzu, finančně náročná a značně přesahuje možnosti studentských, ale i malých badatelských projektů. Zdálo by se tedy, že navrhovat vlastní dotazník a sběr dat je zbytečné.

Punch (2008: 37, 52–57) ale právě s ohledem na četné studentské projekty – třeba v rovině bakalářských, diplomových či disertačních prací – přináší návrh výzkumného designu, který je pro takovou situaci vhodný, a přitom má stále určitý potenciální přínos pro daný obor. Navrhuje používat teoreticky podložený účelový výběr, který se zaměřuje na maximalizaci variability u klíčové proměnné, a naopak vědomé oslabení reprezentativity výběru a s tím i vědomé omezení možnosti zobecnění.

Šetření malého rozsahu tedy například tvoří dvě – nebo více – účelově vybrané skupiny, které jsou do určité míry podobné, ale zároveň u nich předpokládáme výraznou odlišnost v některé klíčové proměnné. Pokud bychom například chtěli studovat souvislost mezi trávením volného času dětí a sociálním statusem rodiny, mohli bychom se po rozvaze zaměřit na dvě skupiny: studenty základní školy a víceletého gymnázia ze stejného ročníku a města. Jelikož víceletá gymnázia jsou velmi selektivní, co se týče rodinného zázemí, lze očekávat, že se studenti z obou typů škol budou právě v této proměnné výrazně lišit (maximalizace variability).

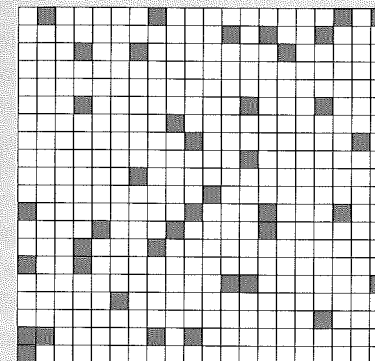
Tento design umožňuje i s relativně omezenými prostředky dobře prozkoumávat teoreticky podložené vztahy mezi proměnnými. Tedy například výše zmíněnou souvislost

rodinného zázemí a trávení volného času. Zároveň je ale nutné mít na paměti výrazné omezení, které je v něm obsaženo. Platnost zjištění je limitována provedeným výběrem a výsledky mohou být velmi citlivé na to, jaké konkrétní skupiny byly do dat zařazeny. Je dost možné, že pokud bychom vybrali jiné gymnázium či jinou základní školu, byly by i rozdíly jiné – například proto, že by se jednalo o sportovní školu či gymnázium se zaměřením na jazyky. Toto riziko je potřeba zohlednit při diskuzi výsledků a odpovídajícím způsobem promyslet.

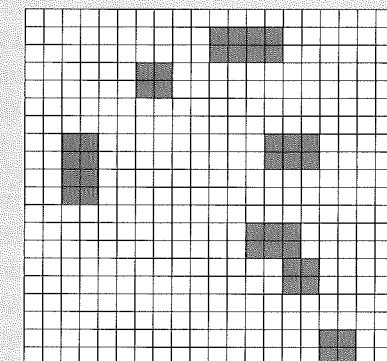
Rámeček 6.6 → Grafické znázornění jednotlivých způsobů výběru

Následující náčrtky srovnávají čtyři způsoby výběru v kvantitativní výzkumu. Představte si, že vybíráme 40 čtverců z prostoru (populace) 20×20 . Vhodný výběrový postup by měl být takový, že na jeho základě můžeme získat poměrně věrnou představu o podobě celého prostoru.

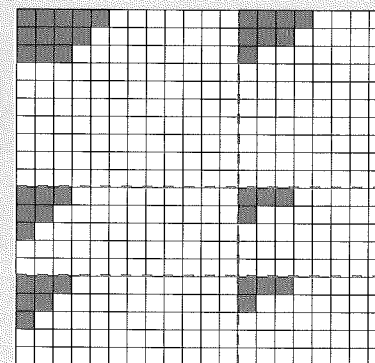
Prostý náhodný výběr



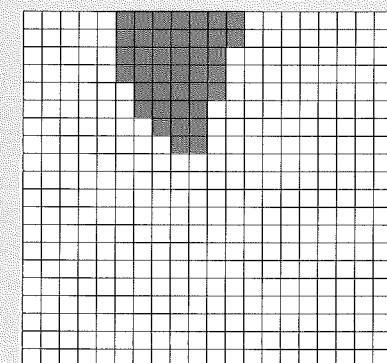
Skupinkový výběr



Kvótní výběr



Dostupný výběr



6.4 VÝBĚR RESPONDENTŮ NA INTERNETU

Dotazování přes internet patří mezi populární způsoby výzkumu, obzvláště u studentských projektů či bakalářských prací (viz také kapitola 7.3.2). Jedná se o potenciálně velmi efektivní způsob sběru dat, avšak pokud mají být jeho výsledky důvěryhodné, je potřeba postupovat velmi obezřetně. Prostřednictvím internetu lze snadno získat poměrně velký počet vyplněných dotazníků. Pokud však nevíme, jakým způsobem takový soubor dat vznikl, provedený výzkum se může stát prakticky bezcenným.

Příkladem velmi špatné praxe, která je pouze velmi obtížně obhajitelná z hlediska případné zobecnitelnosti, je distribuce dotazníků prostřednictvím odkazů sdílených na sociálních sítích či přes různé online dotazovací služby, které hovoří o „zajištění“ získání respondentů. V obou případech bude výsledný výběrový postup do určité míry netransparentní variantou dostupného výběru. Na základě takových dat můžeme stěží vyvozovat závěry o populaci, protože lidé se podstatně liší v tom, zda mají přístup k internetu, i jakým způsobem a jak často ho využívají. Stejně tak je podstatný problém tzv. samovýběru, tedy zkresení výzkumu na základě toho, že se ho zúčastní lidé, které téma zaujme.

Provést reprezentativní výzkum na internetu je náročné. Výzkumné agentury obvykle vytvářejí prostřednictvím tradiční tazatelské sítě pravidelně obměňovaný panel registrovaných respondentů, kteří jsou následně vyzýváni k účasti na několika málo výzkumech. Tím mohou prověřit, že respondenti skutečně existují a že spadají do deklarovaných sociodemografických kategorií, tj. že skutečně mají určitý věk, pohlaví, místo bydliště apod.

Pro výzkum realizovaný vlastními silami může být elektronické dotazování velice efektivní a užitečné, pokud jsme v situaci, kdy chceme prostřednictvím e-mailu systematicky oslovit všechny členy určité organizace, skupiny či komunity (např. studenty školy, účastníky určitého komunitního serveru). Téma výzkumu ale musí samozřejmě odpovídat tomuto zúžení cílové populace.

6.5 NÁVRATNOST

Problém získání kvalitních dat o populaci na základě dotazování bohužel nekončí pouze tím, že zajistím adekvátní výběr respondentů. Na rozdíl od zkoumání neživých předmětů nemůžeme předpokládat, že každý vybraný potenciální respondent bude ochotný se výzkumu zúčastnit. Pokud by tato ochota či neochota byly zcela náhodné, nijak by to předpoklady náhodného výběru neohrozilo. Často ale máme důvod domnívat se, že účast ve výzkumu systematicky souvisí s některými charakteristikami oslovených osob, a tím může vést ke zkresení celkových výsledků. Například se může jednat o pečlivost, respekt k autoritám, otevřenost, aktivitu, ale též o nezamýšlené bariéry, jako třeba nedostatek času, časté cestování mimo domov apod.

V souvislosti s mírou neúspěchu při oslovení hovoříme u dotazníkových šetření o **návratnosti**, tedy o poměru mezi počtem oslovených respondentů a počtem uskutečněných rozhovorů či sebraných dotazníků. Samozřejmě v ideálním případě bychom chtěli dosáhnout toho, aby se návratnost co nejvíce blížila 100 %. V dnešní době může být nicméně u obecné populace považován za velmi úspěšný výzkum, ve kterém bylo dosaženo návratnosti kolem 60–70 %. Nezřídka pak dotazníková šetření dosahují návratnosti jen kolem 30–40 %.

Návratnost výzkumu lze ovlivňovat různými aspekty spojenými se sběrem dat. U výzkumů s tazatelem bývají klíčové náklady vynaložené na výzkum. Pokud tazatelé umožníme (tj. proplatíme) více opakovaných pokusů o kontaktování vybraných respondentů, dokážeme tak snížit pokles návratnosti způsobený nezastižením respondentů v jejich domovech. Podobně u výzkumů vyplňovaných respondenty samostatně bývá užitečné rozesílat s odstupem připomínku (viz rámeček 6.7).

Míra návratnosti také může být spojena se způsobem prezentace výzkumu potenciálním respondentům, tedy způsobem, jak výzkum tazatelé představují nebo jak je formulován průvodní dopis či e-mail. Nepochybně také souvisí s mírou zainteresovanosti populace na tématu. Studenti středních škol budou mít spíše blíže k výzkumu zabývajícím se tím, jak tráví volný čas, než k detailní studii jejich rodinného socioekonomického zázemí. Při plánování výzkumu lze také uvažovat o finanční či symbolické odměně za participaci na šetření, byť je vzhledem k velikosti výběrového souboru vždy potřeba zvažovat její únosnou míru.

V dnešní době se v kontextu velkých dotazníkových šetření hovoří o krizi návratnosti, tedy o celkově nízké ochotě lidí účastnit se těchto výzkumů. Metodologické studie zabývající se tématem návratnosti ale také často upozorňují na to, že problémem ani tak nemusí být nízká míra návratnosti jako spíše potenciální zkresení, nad- či podreprezentování určitých skupin respondentů. Při posuzování kvality výběru tak lze u výzkumu sledovat odchylku výběrového souboru z hlediska známých charakteristik populace (věk, vzdělání, ekonomický status) a na základě toho usuzovat o možném riziku zkresení výsledků z důvodu nestejné návratnosti.

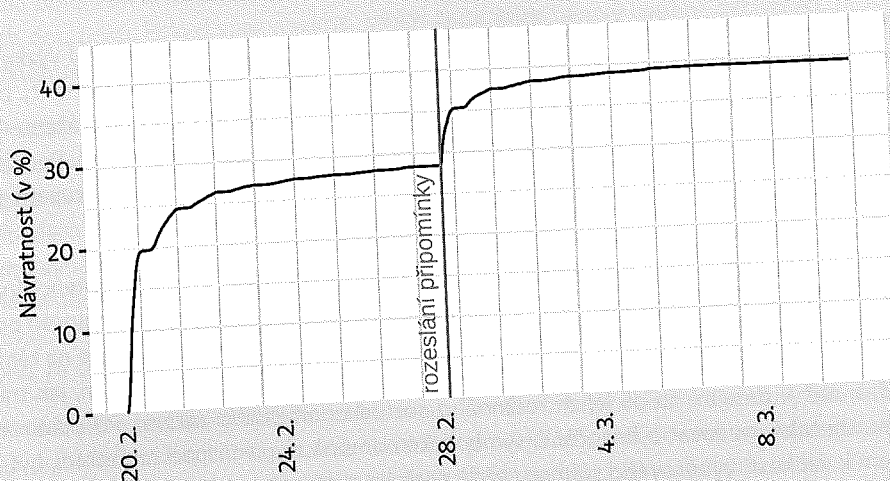
Rámeček 6.7 → Návratnost v čase: příklad online výzkumu studentů

V únoru a březnu 2017 proběhlo na Univerzitě Karlově rozsáhlé online dotazníkové šetření, při kterém byli osloveni všichni studenti v 1. a 2. ročníku. Celkově bylo e-mailovou výzvou osloveno 13 022 potenciálních respondentů. Studenti byli k účasti v dotazníkovém šetření motivováni možností být zařazeni do slosování o finanční odměnu.

Následující graf ukazuje vývoj návratnosti tohoto šetření po rozeslání 20. února. Přibližně čtvrtina studentů reagovala bezprostředně v den oslovení nebo den následující. Během následujícího týdne již podíl vyplněných dotazníků příliš nerostl a ustálil se

přibližně na 28%. Týden od původního oslovení byla studentům rozeslána připomínka možnosti účastnit se výzkumu. To během dne vedlo k dalšímu skokovému zvýšení účasti. Výzkum tak dosáhl celkové návratnosti 39%.

Vývoj návratnosti výzkumu Kulturní spotřeba studentů Univerzity Karlovy (2017)



Další připomínka již studentům rozesílána nebyla. Na základě zkušeností z jiných výzkumů podobného typu lze předpokládat, že by sice ještě mohla vést k zvýšení návratnosti o jednotky procent, nicméně větší efekt by nejspíš již neměla.

6.6 SEKUNDÁRNÍ ANALÝZA DAT

Kvantitativní výzkumné projekty nezdávka využívají již dostupné existující soubory dat. Jedná se zejména o rozsáhlé a neustále se rozšiřující katalogy volně dostupných datových souborů, které byly vytvořeny pro účely nejrůznějších výzkumných projektů a následně byly dány k dispozici širší odborné veřejnosti (viz rámeček 6.8). Velká dotazníková šetření obvykle zahrnují desítky až stovky proměnných a nelze říci, že by jejich potenciál bylo možné vyčerpat. Vždy lze najít relevantní a nezodpovězené výzkumné otázky, problémy či analytické přístupy, které dosud na daná data nebyly aplikovány.

Tento postup se označuje jako **sekundární analýza dat**, ale v žádném případě ji nelze chápat jako něco druhotného či vedlejšího (viz též kapitola 15.4.1). Naopak, celá řada významných studií v nejrůznějších oblastech bádání vznikla právě na základě inovativní analýzy již existujících dat. Z praktického hlediska jsou výhodami sekundární analýzy zejména její dostupnost

a úspornost. Využití existujících datových souborů umožňuje ušetřit zdroje, které si jinak náročný sběr dat vyžaduje. V případě, že výzkum má ambici vypovídat o celé společnosti, jedná se pro výzkumníka, který nemá dostatečné grantové zdroje financování, o jedinou možnost.

Kromě praktických důvodů ale v současné době existující datová základna umožňuje provádět i dříve nemyslitelné srovnávací výzkumy. Především se otevírá možnost analýzy unikátních časových řad. Spolehlivá archivovaná data z dotazníkových šetření lze dohledat z celé druhé poloviny 20. století, ve výjimečných případech i dříve. Kromě toho se nabízí i možnost mezinárodního srovnání různých společností, a to nejen díky širokým mezinárodním projektům, zahrnující pravidelný sběr dat stejným výzkumným nástrojem (viz rámeček 6.8).

Sekundární analýza dat má ale samozřejmě i svá úskalí. V první řadě je výzkumník silně limitován podobou dotazníku, který byl pro sběr dat použit. Při operacionalizaci konceptů, se kterými chceme ve výzkumu pracovat, tak často musíme přistupovat na řadu kompromisů – a v některých případech ani operacionalizace zamýšlených konceptů není možná. Přestože pak víme, že by se v analýze měly objevit určité proměnné, v datech prostě nejsou.

Přitom je nutné si uvědomit, že i když při sekundární analýze výzkumník nezodpovídá za způsob sběru používaných dat, zodpovídá za jejich užití a způsoby interpretace. Pokud tedy otázky v původním dotazníku nebyly formulovány adekvátně našemu výzkumnému záměru, nemůžeme omlouvat nedostatky naší analýzy nevhodností dat. Je vždy úkolem výzkumníka prokázat, že zvolená cesta byla správná či alespoň přijatelná.

Při použití sekundární analýzy dat ve výzkumu je vždy nutné na základě dostupné dokumentace nastudovat způsob, jakým byla data získána. Konkrétně tedy zjistit cílovou populaci šetření, způsob výběru, sběru dat i další charakteristiky kvality, jako je například návratnost.

Rámeček 6.8 → Datové archivy

Současné sociální vědy v čím dál větší šíři rozvíjejí infrastrukturu datových archivů, která na jednom místě v přenositelné a přehledné podobě shromažďuje existující datové soubory. Většina datových archivů umožňuje procházet či prohledávat dostupné datové soubory. V případě výzkumného zájmu pak lze při dodržení podmínek poskytovatele kromě základních výstupů získat i úplné datové soubory. Datové archivy shromažďují jednotlivé výzkumy stejně tak jako rozsáhlá a dlouhodobá mezinárodní šetření, která často zahrnují řadu zemí a poměrně dlouhé časové řady.

Vybrané datové archivy

- Český sociálněvědní datový archiv (archiv.soc.cas.cz)
Národní centrum poskytující primární data z českých a československých výzkumů. Zahrnuje převážně sociologická dotazníková šetření, a to jak ze současnosti, tak historická (70. a 80. léta, ale i starší). Kromě jiného jsou zde archivována i data z výzkumů veřejného mínění CVVM.

- ZACAT – GESIS (zacat.gesis.org)
Katalog datových souborů z dotazníkových šetření spravovaných hlavní německou institucí poskytující služby pro sociální vědy – GESIS. Zahrnuje mimo jiné data z velkých mezinárodních dotazníkových šetření (např. ISSP, EVS či Eurobarometr).

Významná mezinárodní sociologická šetření

- International Social Survey Programme (ISSP) – mezinárodní dotazníkové šetření, které se každý rok zaměřuje na jiné téma (jako např. náboženství, sociální nerovnosti, pracovní orientace, národní identita apod.). Sběr dat probíhá ve více než 30 zemích, Česká republika se výzkumů účastní nepřetržitě od roku 1992.
- European Social Survey (ESS) – mezinárodní dotazníkové šetření zaměřené na evropské země se specifickým důrazem na vysokou kvalitu metodologie. Sběr dat probíhá každé dva roky, zahrnuje pravidelnou část (např. politika, hodnoty, důvěra) a nepravidelné moduly (např. zdraví a péče, spravedlnost, demokracie). Česká republika se výzkumů účastní od počátku (2002).
- Eurobarometr – výzkumy veřejného mínění pro Evropskou komisi, které probíhají několikrát do roka ve všech zemích Evropské unie. Výzkumy se zaměřují na nejrůznější otázky se specifickým důrazem na témata a problémy spojené s fungováním Evropské unie.

Shrnutí

- V kvantitativní výzkumné strategii se nejčastěji snažíme navrhnout výběr případů tak, aby na jeho základě bylo možné spolehlivě popsat celou populaci.
- Požadavek zobecnění na populaci nejlépe splňuje náhodný výběr, který je vymezen tím, že každý případ v populaci má známou (nejčastěji stejnou) pravděpodobnost, že bude zařazen do výběru.
- Realizace náhodného výběru je často velmi náročná a v ideální podobě u řady populací nemožná. Postupy komplexních náhodných výběrů mohou náklady na realizaci snížit. Zároveň musíme vždy počítat s tím, že nedosáhneme abstraktního ideálu.
- Ve výzkumech se často pracuje s nepravděpodobnostními výběry, např. kvótním výběrem či dostupným výběrem, které však nesplňují předpoklad náhodného výběru. I v tomto případě výzkumná praxe používá statistické metody analýzy, přestože zde nejsou striktně vzato smysluplné.
- Klíčovou charakteristikou výběru v případě dotazníkových šetření je návratnost, tedy poměr mezi počtem oslovených respondentů a počtem těch, kteří se výzkumu skutečně zúčastní.
- Data z mnoha dotazníkových šetření jsou shromažďována v archivech sociálněvědních dat. Zde jsou zájemcům k dispozici pro sekundární analýzu.

Doporučená literatura

- Babbie, E. 2014. *The Practice of Social Research*. Boston: Cengage Learning. Kapitola „The Logic of Sampling“.
- Bryman, A. 2012. *Social Research Methods*. Oxford: Oxford University Press. Kapitola „Sampling“.
- Disman, M. 2018. *Jak se vyrábí sociologická znalost*. Praha: Karolinum. Kapitola „Kolik vran musíme pozorovat?“.
- Hibberts, M., Johnson, R. B., Hudson, K. 2012. *Common Survey Sampling Techniques*. In: Gideon, L. (ed.). *Handbook of Survey Methodology for the Social Sciences*. New York: Springer, s. 53–74.
- Fowler, F. J. 2014. *Survey Research Methods*. London: SAGE. Kapitola „Sampling“.