

nost připravených otázek, o vhodnost pozorovacích kategorií, o zkoumání, zda informace, na kterou se hodláme ptát, je vůbec ve zkoumané populaci známa atp. Na základě takovéto pilotáže pak bývá konečná podoba konkrétní techniky (maketa dotazníku, záznamový arch rozhovoru) přílohou definitivní verze projektu. (Může však jít i o zkoušku určitého postupu při zpracování dat, při výběru objektů atp.)

Pokud se prověřují úplně všechny aspekty připravovaného výzkumného postupu (což nebývá příliš běžné), jedná se o **zmenšené, modelové provedení plánované akce**.

Za **další variantu předvýzkumu** lze považovat tzv. **sondáž** (což není právě šťastně zvolený pojem, neboť v cizojazyčné literatuře může představovat celé, dokonce dosti rozsáhlé šetření). V tomto případě se ale sondáží myslí **hlubší analýza některého ze zkoumaných aspektů**, hlubší průnik do nějakého z prvků či fenoménů zkoumaného problému. Nedá se vyloučit ani **kvalitativní charakter sondáže**, jež má podobu například přímého zkoumání prostředí v určité instituci (v našem případě např. osobní účast na některé vzdělávací akci) atp. Podle mého názoru lze pak **některé druhy výzkumu orientačního a také deskriptivního**, čili sociografického (o nich viz v 2.3), chápat jako **kvantitativní typy sondáže**.

## 4. VÝBĚR ZKOUMANÝCH OBJEKTŮ

Přípravná fáze výzkumného úkolu tedy končí vypracováním projektu (viz 3.6). Jak jsem uvedl, v jeho metodologické části je záhodno představit objekt či objekty zkoumání, jejich soubor, množinu, populaci apod. Při tvorbě konceptuálního rámce úkolu (3.2) totiž postupně přichází rozhodnutí, jakým způsobem budou získávány údaje a jak budou zpracovány. V těsné souvislosti s tím si však též projasňuje, o jaké se bude jednat objekty a případně o kolik z nich. Jinými slovy, uvažuje se o výběru zkoumaných jednotek, tj. také o možných redukcích jejich počtu, je-li tento příliš vysoký (viz dále, o redukcích též v 3.2). **V přípravě zkoumání patří výběr objektů k jedné z klíčových otázek**. Proto mu věnuji samostatnou kapitolu

*Problematika výběru se objevuje, byť v nestejném rozsahu, ve většině metodologických publikací. Zájemcům o podrobnější poznatky mohu doporučit například Hendla (2004, s. 51–59), Janouška aj. (1986, s. 70–71), Jeřábka (1992, s. 37–51), Chráska (2007, s. 20–27), Dismana (1993, s. 92–116), Lamsera (1966, s. 234–257), Pelikána (1998, s. 47–54), Kerlingera (1972, s. 61–73) ad., jakož i Maříkovou, Petruska, Vodákovou aj. (1996, s. 1396–1401 a další navazující hesla), též neobvykle stručný pohled Punchův (2008, s. 80–81). K otázkám výběru v kvalitativním zkoumání Hendla (2005, s. 150–154) a Miovského (2006, s. 127–140) aj. To vše jsou prameny, ze kterých jsem čerpal mnohé informace i pro následující výklad.*

### 4.1 ZÁKLADNÍ A VÝBĚROVÝ SOUBOR

Přes některé názorové a terminologické nuance v literatuře je v podstatě možné shrnout, že stěžejními pojmy zde jsou základní a výběrový soubor, někdy se lze setkat ještě s označením cílový soubor. Postupně je objasním.



**Cílový, též výzkumný soubor** (popř. populace) je množina objektů, kterých se zkoumání týká, na které se mají jeho výsledky vztáhnout. Doporučuje se hovořit o něm v případech, kdy se dá předpokládat, že ne všechny objekty zkoumání budou/Jsou pro účely výzkumné dostupné. V případě populace občanů naší republiky to mohou být ti, co jsou dlouhodobě v cizině, vážně nemocní atd., jedná-li se o studenty univerzity, jsou to ti, co momentálně studují v zahraničí, přerušili studium atp. Toto rozlišení se užívá i v případech, kdy z takovýchto nebo podobných důvodů není možné pořídit kompletní seznam jednotek či zcela přesně vymezit populaci, jíž se výzkum týká.

**Základní soubor** je množina jednotek, která ve výzkumné situaci figuruje právě v roli cílového souboru. Jde tedy o **souhrn objektů, který lze vymezit, podchytit, identifikovat a z něhož se pak vybírá**. Ve skutečnosti se tedy může jednat o podobně rozsáhlou populaci objektů, nicméně nepatrně zmenšenou o nedosažené či nedosažitelné jednotky. Třeba říci, že cílový a základní soubor se v mnoha případech ani nerozlišují, uvádím je tu pouze pro úplnost. Většinou bývá celá záležitost chápána tak, že výsledky výzkumu se prostě vztahují, zobecňují na základní soubor.

**Výběrový soubor, též vzorek**, je množina objektů/jednotek, která ve výzkumu zastupuje základní soubor. **Výsledky zkoumání zjištěné na výběrovém souboru by měly být, pokud má jít o výzkum reprezentativní, zobecnitelné na soubor základní, tzn. i na cílový.** (O reprezentativitě viz zmínku v souvislosti s kritérii kvality zkoumání v 3.5.) **Určité způsoby výběru, jak dále vyplyne, mohou lépe či hůře zajistit, aby vzorek (tj. výběrový soubor) reprezentoval soubor základní, tzn. aby údaje, které byly na vzorku zjištěny, bylo pak možné vztáhnout na celý soubor, z něhož tento vzorek vznikl.**

*V tomto kontextu se někdy hovoří o tzv. **opoře výběru**, kterou se rozumí jednak seznam jednotek základní populace, z něhož vzorek vybíráme, případně konkrétní předpis jak výběrový soubor koncipovat. V praxi jsou výběrové opory v podobě seznamů jednotek (např. soupisy voličů, seznamy klientů, kartotéky zákazníků apod.) dosti cenným materiálem a nezřídka bývají předmětem obchodních transakcí mezi jejich majiteli a výzkumnými či reklamními agenturami.*

*Pokud takováto podoba výběrové opory není k dispozici, výběr ze základní populace je nutné nějakým způsobem vytvořit. Jedním z nich je tzv. **náhodná procházka**. V rámci ní se obyvatelé či domácnosti určitého sídelního celku (vesnice, městské čtvrti, ulice apod.) vybírají podle instrukce podrobně ur-*

*čující, který dům (popř. který byt v něm a které jeho uživatele) do šetření zahrnout. To může potom představovat tzv. náhodný výběr (viz 4.2), nebo být ještě zpřesněno požadavkem, jaké parametry, tj. charakteristiky (věk, pohlaví, vzdělání atp.), má takto vybíraná osoba mít (viz kvótní výběr v 4.3).*

Při úvahách o možnostech a potřebách výběru objektů zkoumání lze též dospět ke dvěma specifickým způsobům výběru. Jedním je případ, kdy **předmětem zkoumání je pouze jeden jediný objekt** (firma, agentura, škola apod.). Tehdy se jedná o tzv. **monografickou studii** (srv. v 2.2 poznámku o výzkumných procedurách, k nimž patří právě též monografická). Tato jednotka může být vybrána náhodně (viz dále), ale obvykle bývá volena jako představitelka standardu, extrému atp., pokud máme dostatek informací k tomu, abychom rozpoznali, co lze za standard, deviaci apod. považovat.

Druhou variantou je opačná situace, tj. rovná-li se základní soubor výběrovému, jinými slovy, **jsou-li zkoumány všechny objekty, jichž se šetření týká. Tento výběr se nazývá úplný, totální, vyčerpávající**, také tzv. **cenzus**. (O jeho reprezentativitě viz tab. 3 v 3.5.) Důvodem k jeho realizaci může být to, že celkový počet jednotek, které jsou předmětem zkoumání, je organizačně, časově aj. zvládnutelný.

*Například někdy koncem 70. let jsem se podílel na výzkumu otců v kraji Praha, jimž soudy svěřily děti do péče. Tehdy se jednalo o něco přes dvě stě případů, takže bylo zbytečné provádět jakýkoli výběr a naopak se jevílo jako účelné zahrnout do zkoumání všechny. Oporou našeho výběru potom byly evidence odborů péče o dítě jednotlivých městských částí, tehdy obvodních národních výborů.*

Nezáviděníhodnou situací je, když je počet jednotek těžko zvládnutelný a přesto je třeba šetření uskutečnit. Příkladem jsou tzv. **cenzovní výzkumy**, které provádí Český statistický úřad (pro ně se toto označení standardně užívá), tj. šetření všech domácností České republiky (též tzv. Sčítání lidu, domů, bytů...). Je to záležitost nepředstavitelně nákladná, nicméně pro potřeby statistických analýz vývojových trendů v mnoha sférách týkajících se celé společnosti naprosto nezbytná.

Asi nejčastěji však dochází k tomu, že z **důvodů časových, logistických a též pochopitelně finančních, popř. jiných, nelze zkoumat všechny jednotky základního souboru a je potřeba z něj vytvořit výběrový soubor**. Výběrových postupů (předpisů, způsobů, technik) je poměrně hodně, takže se omezím jen na některé z nich, především ty běžněji užívané. Obecně



je lze rozdělit na dva základní typy – na výběry náhodné, pravděpodobnostní (4.2) a záměrné, nepravděpodobnostní (4.3). Posléze si představíme některé specifické výběrové postupy (4.4).

Nyní si ještě vzpomeňme na přehlídku různých typů výzkumů (2.3). Bylo mj. řečeno, že výběrový soubor (vzorek) lze zkoumat jednorázově či opakovaně, přičemž tomu **opakovaně šetřenému, je-li tvořen živými osobami, se říká panel**. Při práci s ním je nezbytné pamatovat na **některé jeho zvláštnosti**. Během zkoumání totiž u něj dochází k jakési „únavě materiálu“ – opakovaným šetření (rozhovorem, dotazníkem, ale i experimentem, viz 5.4) jsou zkoumané osoby svým způsobem již poučeny, neboť vědí, co je „čeká“, stručně řečeno získávají určitou výzkumnou rutinu.

*Ta se mnohdy projevuje tendencí odpovídat/reagovat stereotypně, jednoduše, nebo naopak fabulovat, snažit se o netypický, soutěživý výkon aj. Je to zřetelné také v případě, má-li se něco hodnotit, což se často stává například v mediálních výzkumech (hodnocení pořadů). To pak šetřené osoby časem přestávají být obyčejnými laiky a mění se v laiky velmi zasvěcené, ne-li přímo v experty. Není výjimkou, že některý jedinec využívá svých dosavadních zkušeností hlavně k tomu, aby vším rychle a nekomplikovaně nějak „propnul“, veškeré zkoumání mohl absolvovat bez přílišného zájmu a nasazení atd.*

To výpovědní hodnotu získaných údajů a tím tedy i efektivitu celého výzkumu podstatným způsobem nabourává. Není-li zcela nezbytné zachovat vzorek v nezměněné podobě (např. kvůli komparacím), **jako účelná obrana se jeví postupné a průběžné obměňování panelu**. Například kvartálně jeho třetinu aj. (samozřejmě záleží na frekvenci užívání panelu), čímž do šetření přicházejí stále čerstvé a „dosud nezkažené“ posily.

## 4.2 NÁHODNÝ, PRAVDĚPODOBNOSTNÍ VÝBĚR

Podstatou náhodných, pravděpodobnostních výběrů je, že **každý prvek, každá jednotka základního souboru má stejnou šanci, stejnou pravděpodobnost stát se prvkem výběrového souboru**. Reprezentativita výběru je tu tedy založena na principu náhodnosti. (Tj. bude-li základní soubor tvořen např. muži a ženami v poměru 80:20, náhodnost výběru s velkou pravděpodobností zajistí jejich podobný poměr i v souboru výbě-

rovém.) Jednotlivé typy náhodných výběrů se pokoušejí různě **řešit praktickou stránkou výběrového postupu a případně i zvyšovat přesnost prováděného výběru**. V literatuře je asi nejčastěji zmiňován náhodný výběr prostý, stratifikovaný, skupinkový a vícestupňový.

**Prostý náhodný výběr** se provádí tak, že z celkového seznamu prvků, případně také vloženého do osudí či očíslovaného, se jednotky výběrového souboru určují buď losem, nebo na základě vygenerovaných náhodných čísel. Tento postup zcela eliminuje možnost ovlivnění výběru výzkumníkem.

**Stratifikovaný náhodný výběr** probíhá ve dvou krocích. V prvním **nahlížíme základní soubor jako množinu tvořenou určitými podsoubory, kterým se říká strata (čili vrstvy)**. Vyznačují se tím, že jejich prvky mají jednu nebo více společných vlastností (např. věk, pohlaví, vzdělání), takže se v tomto ohledu od sebe navzájem neliší tolik, jako jednotlivá strata mezi sebou. **Ve druhém kroku se v každém ze strat provede náhodný výběr**. Volba strat vychází z výzkumného záměru (např. s ohledem na hypotézy, viz 3.3.3). To znamená, že předpokládáme-li třeba, že vztah k dalšímu vzdělávání se diferencuje podle věku, vzdělání atd., volíme stratum věkové, vzdělanostní apod.

Pokud jsou u základního souboru známy **proporce sledovaných proměnných** (věku, pohlaví atp.), lze je použít též pro výběrový soubor, tzn. vybírat z jednotlivých strat podle nich. Touto cestou sice snižujeme náhodnost, ale současně tím zpřesňujeme podíly relevantních reprezentantů určitých charakteristik ve výběrovém souboru. Tento postup se nazývá **proporcionální**. Oproti tomu při výběru **neproporcionálním** bereme úmyslně z jistých důvodů, například pro přesnější komparaci, u některého ze strat neúměrný počet jednotek, obvykle vyšší, než odpovídá proporcím základního souboru. (Znalost parametrů základního souboru je též nutná i při jiných výběrových postupech, jak ještě poznáme.)

**Skupinkový náhodný výběr** (též clusterový či shlukový, skupinový) v podstatě vyžaduje tři kroky. V prvním opět nahlížíme základní soubor rozdělený do podsouborů, ale na rozdíl od strat teď vidíme **skupinky větší-nou přirozeného původu, tzv. clustery, shluky**. Od sebe se výrazně neliší, v zásadě jsou navzájem zastupitelné. **Ale jejich rozdílnost tkví v tom, že jsou vnitřně poměrně heterogenní, tj. tvoří je prvky vzájemně dosti odlišné**. (Typickým příkladem je tady např. rodina.) Ve druhém kroku vybereme **náhodným způsobem některé z těchto shluků** a ve třetím potom **všechny jejich prvky zařadíme jako samostatné jednotky do výběrového souboru**.



Vicestupňový náhodný výběr má první dva kroky totožné s předcházejícím, tj. **identifikaci skupinek/clusterů a náhodný výběr z nich**. Třetí krok spočívá v tom, že ze souboru vybraných skupinek **dále vybereme náhodným postupem pouze některé z jejich jednotek, nebo provedeme ještě užší náhodný výběr z původně vybraných skupinek a teprve z nich pak náhodně určíme jejich konkrétní členy**, tj. definitivní podobu výběru. (Například náhodně určíme některé z regionů, ze vzdělávacích agentur, jež v nich působí, několik náhodně vybereme, v každé náhodně zvolíme některé vzdělávací kurzy a z každého náhodně nominujeme jistý počet frekventantů.)

### 4.3 ZÁMĚRNÝ, NEPRAVDĚPODOBNOSTNÍ VÝBĚR

Pro **záměrný, nepravděpodobnostní výběr** neplatí princip předchozích výběrů, tedy v tomto případě **nemá každý z prvků základního souboru stejnou možnost, tj. pravděpodobnost stát se součástí výběrového souboru**.

Mezi typické nepravděpodobnostní výběry patří způsob hojně využívaný například v sociologických výzkumech rozsáhlejších populací, tzv. **kvótní výběr**. Jeho princip, a de facto i záruka reprezentativity, spočívá v tom, že **výběrový soubor je tvořen podle poměrů četnosti výskytu jednotlivých variant vlastností/znaků v základní populaci**. Těmito znaky bývají většinou některé tzv. **sociodemografické ukazatele**, jako pohlaví, věk, vzdělání, příjem, velikost místa bydliště, rodinný stav, charakter povolání atp., ale i **další charakteristiky** (počet dětí, vlastnictví auta či nemovitosti, připojení k internetu, národnost aj.). Volba znaků pochopitelně plyne ze zaměření výzkumného úkolu. Mimo jiné též z hypotéz (a z celkové konceptualizace výzkumného problému), kdy rozhodujeme, **jaké znaky lze vzhledem k šetřenému problému považovat za určující** (popř. za nezávisle proměnné, viz 3.3.2). Často se vychází i z dosavadních zkušeností s řešením takového či obdobného problému. (Například se ukazuje, že na volební preference nemá nějaký zásadní vliv rodinný stav, zatímco vzdělání, věk, příjem, velikost místa bydliště ad. zde naopak hrají dosti významnou roli.)

Je evidentní, že **kvótní výběr lze uskutečnit pouze v případě, známe-li podíly jednotlivých variant zvolených znaků v základní populaci**. Určení jejich kvót potom nejčastěji vychází z dostupných statistických ukazatelů. (Například pro populaci občanů ČR jsou to ročenky Českého statistického

úřadu – jedním z nenahraditelných zdrojů údajů jsou pro ně mj. cenzovní výzkumy domácností, viz 4.1). Zjistíme-li tedy konkrétně procentuální poměr mužů a žen v české populaci například 46:54 a v armádě 78:22, ve výběru „českém“ bude z 1000 osob 460 mužů a 540 žen a v „armádním“ 780 mužů a 220 žen. Obdobně rozložení podle výše vzdělání v procentech 40:55:15 (základní, středoškolské a vysokoškolské) u základního souboru znamená pak ve výběru 400 lidí se vzděláním základním, 550 středoškolačků a 150 vysokoškolačků atd. Takový je postup, pokud jsou **kvóty vybírány nezávisle**.

Dost často se ale **dimenzují ve vzájemných vazbách**, tj. nalezený podíl mužů → u nich ve zjištěných proporcích výše vzdělání → současně v těchto podskupinách podíly podle věkových kategorií apod. **Tento způsob daleko přesněji vytváří výběrový soubor jako jakýsi model základní populace**. (Je ale hodně složitý a musím přiznat, že jsem se jej nikdy nemusel naučit konstruovat, neboť v týmu byl vždy statistik znalý demografie.) V praxi se často používá kombinace tak 4–7 znaků (byť jsou i výjimky).

*Ale i čtveřice „zřetězených“ kvót může být pro tazatele (viz 5.2.4) velkým problémem. Stačí si představit, že má za úkol najít a „vyzpovídat“ – v místě třeba s počtem 200–499 obyvatel – vdovu s vysokoškolským vzděláním ve věku 30–39 let nebo svobodného muže se základním vzděláním ve věkové kategorii 40–49 atp.*

Ať již jde o kvóty navzájem nezávislé či nikoli, zřetelně se ukazuje, proč **nelze tento druh výběru považovat za náhodný. Pravděpodobnost jednotlivých prvků, že se stanou součástí výběrového souboru, není rozhodně stejná**, naopak je od samého začátku jasně vymezena právě těmito kvótami. Ty na druhé straně, tím, že jejich různé kombinace lehce přibližují výběrový soubor určitému miniaturnímu modelu základního souboru, **zajišťují jeho reprezentativnost**.

S jiným postupem, tzv. **systematickým** (mechanickým, systémovým) **výběrem**, je to v literatuře poněkud nejednoznačné. Bývá považován za variantu náhodného (např. Hendl, 2004, s. 54; Lamser, 1966, s. 246; aj.), nebo není z tohoto hlediska explicitně zařazován (srv. Chráska, 2007, s. 23; Maříková, Petrušek, Vodáková aj., 1996, s. 1400; ad.), případně je považován za nepravděpodobnostní, nenáhodný (viz Jeřábek, 1992, s. 49–50; aj.). Přikláním se k této třetí variantě a jak bude patrné, ne zcela bezdůvodně.

Jeho podstatou je výběr jednotek ze seznamu základního souboru, kdy **nejprve je náhodně (losem, náhodným číslem) stanoven první prvek**



a pak každý  $x$ -tý podle kroku, jehož rozsah se určí jako podíl velikosti základního a výběrového souboru. Například máme základní soubor čítající 120 000 jednotek a potřebujeme výběrový soubor o velikosti 1000 prvků, tedy jeho jednu stodvacetinu. Začátek výběrového kroku stanovíme losem z čísel 1–120, například je to 10. Odtud pak vybíráme každou stodvacátou jednotku (mechanicky, systematicky), takže do výběru přicházejí prvky, které vystupují v seznamu základního souboru pod pořadovými čísly 10, 130 (10+120), 250 (130+120), 370 (250+120) atd. V tomto ohledu je tak systematický výběr velmi blízký náhodnému. Celý problém ovšem spočívá v onom seznamu jednotek základního souboru, ze kterého vybíráme a na němž je celý výběr naprosto závislý. Podmínkou realizace samozřejmě je, aby takový seznam vůbec existoval, jenže i pak zůstává otázkou, jakým způsobem vlastně vznikl, což zásadně ovlivňuje princip náhodnosti celého výběru a tedy též míru jeho reprezentativity.

Odhlédněme od skutečnosti, že seznamů čítajících statisíce jednotek asi příliš nenajdeme. U desetitisíců a tisíců bude situace patrně lepší, takže tento výběr se hodí spíše pro soubory takovýchto řádů. Současně ovšem nebývá pravidlem, že by jejich seznamy byly tvořeny dle úplně náhodného pořadí prvků, spíše v sobě častěji obsahují jistý organizační klíč. A ten je právě „kamenem úrazu“ náhodnosti těchto výběrů, neboť zvolený výběrový krok může kopírovat nějakou (nám skrytou) pravidelnost uspořádání seznamu. Tím se do výběru mohou dostat prvky, jejichž vlastnosti jsou pro základní soubor nejen netypické, ale i naprosto unikátní. A tato **apriorní a vcelku opodstatněná nedůvěra k náhodnosti uspořádání základních seznamů** je důvodem, proč mnohdy nebývá systematický výběr řazen mezi náhodné (pravděpodobnostní).

*Jako příklad může sloužit v literatuře nezřídka uváděné hypotetické šetření, kdy máme k dispozici seznam mnoha např. stočlenných armádních útvarů uspořádaných tak, že prvních pět míst vždy patří důstojníkům, dalších pět poddůstojníkům a zbylých devadesát prostým vojínům. Při velikosti výběrového kroku 100 a počátku 1–5 budou ve výběrovém souboru sami důstojníci, při počátku 6–10 pouze poddůstojníci, přestože 90 % souboru představují vojini. Další ilustrací je typicky americký sídlištní satelit tvořený stovkami rodinných domků, kde každá první a desátá parcela (začátek a konec bloku) je rohová – o hodně větší a též s větší stavbou. Tudiž zřetelně dražší a tak obývaná příslušníky jiné příjmové a sociální kategorie, kteří by se při pro ně příznivém výběrovém kroku stali představiteli ostatních 80 % zdejších obyvatel.*

Typickým nepravděpodobnostním, záměrným výběrovým předpisem je postup nazývaný **výběr úsudkem, účelový** (též ad hoc či kvalifikovaný, zaměřený). Používá se hlavně pro nevelké výběry (nepřesahující desítky jednotek) a jeho podstata není složitá. **Výzkumník znalý kontextů zkoumané problematiky vybere prvky sám.** Vychází při tom z výzkumných záměrů (obecně z konceptualizace výzkumného problému) a také z uvažované podoby zkoumání. (Například v případě experimentu výzkumník jako odborník dokáže určit, které osoby jsou k tomuto účelu vhodné.) Tento postup se často používá v kvalitativních zkoumáních (viz o teoretickém výběru v 4.5), též při konstrukci panelu expertů (viz 6.1). Jeho reprezentativita je diskutabilní, spíše minimální, přesněji řečeno se s ní většinou vůbec nepočítá.

Poměrně efektivním (a svým způsobem i efektním) je **výběr nabalováním** (tzv. **snowball** čili sněhová koule). Je velmi vhodný k výběrům ve specifických souborech, které jsou výzkumníkům neznámé, pro které neexistují seznamy, ani jiné vhodné opory výběru (například při zkoumání sběratelů známek, lidových léčitelů, struktur moci apod.). Užívá se též k výběru expertů. Základní princip je poměrně jednoduchý. Na počátku je několik málo osob (ze zkoumaného souboru, obeznámených s ním, případně s důvody výběru) požádáno, aby **každá určila cca 3–5 jedinců, které považuje za vhodné do výběru zařadit. Tito jmenovaní jsou potom dotázáni na to-též, čímž množina tipů strmě narůstá.** Počet kroků záleží na rozsahu monitorovaných struktur, každopádně se po jisté době začínají některé typy opakovat a nakonec žádné nové nepřibývají. V této situaci můžeme výběr ukončit a začít se šetřením na základě takto vzniklého seznamu, nebo jej podle okolností omezit pouze na typy s relativně největší frekvencí. V určitém pohledu (např. právě vzhledem ke zkoumané specifické skupině) je možné tyto výběry považovat za **svým způsobem reprezentativní**, nicméně záleží především na tom, co je předmětem zkoumání. Někdy nemusí jít o reprezentativitu ve smyslu čistě statistickém, ale o jakousi reprezentativitu (popř. dominanci) názorovou, nebo při šetřeních expertů o tzv. reprezentativitu problémového pole (viz 6.1).



#### 4.4 VÝBĚRY ZALOŽENÉ NA DOSTUPNOSTI A DOBROVOLNOSTI

Kromě předchozích čtyř výběrových postupů pravděpodobnostních a stejného počtu nepravděpodobnostních se v praxi lze setkat i s dalšími výběrovými mechanismy. Mimo jiné se jedná o **výběry založené na dobrovolnosti a dostupnosti**. Jsou spojeny například se specifickým způsobem sběru dat (anketa, experiment, blíže viz 5), případně s nedostatkem informací o základním souboru (to ostatně nebývá tak výjimečný případ). Odborné publikace se o nich občas zmiňují, takže upozorním alespoň na některé jejich podoby. (O určité jejich období, tzv. příležitostném výběru, srv. např. Hendl, 2004, s. 52; popř. Miovský, 2006, s. 134.)

**Výběr na základě dobrovolnosti** se často objevuje například v experimentech (viz 5.4), kdy námi apriori vybrané a podle našeho názoru vhodné osoby však pochopitelně nelze nutit, aby do tohoto druhu zkoumání vstoupily. Je tedy na jejich rozhodnutí, zda se stanou členy zkoumaného vzorku, či nikoli. Tento aspekt (nehledě na to, že už jsme u nich stejně nejprve provedli výběr záměrný, úsudkem) je evidentní překážkou tomu, aby získané údaje mohly být považovány za reprezentativní a jakkoli zobecňovány. Ambicí většiny experimentů ale není reprezentativnost, takže jejich realizátoři takový výběrový postup oprávněně považují za vhodný.

Určitou podobou výběru založeného na dobrovolnosti je tzv. **samovýběr**, s nímž se lze setkat především u různých druhů anket (viz 5.2.6). Mechanismus samovýběru je dán způsobem distribuce podnětu, zde tedy konkrétně nějakého anketního lístku, archu apod. Záleží na čtenářích novin a časopisů, zákaznících nákupního centra, klientech jisté ordinace či nemocnice atp., zda se ankety zúčastní. I přes stovky a tisíce či desetitisíce odpovědí však **výsledky tohoto šetření nejsou zobecnitelné, resp. jsou zobecnitelné pouze na populaci těch, kteří odpověděli**. Takže je nelze jakkoli vztahovat na ostatní populaci. (Obdobně jako Zlatý či Český slavík nevypovídá nic reprezentativního o hudebních preferencích obyvatel ČR.) Jak patrně, reprezentativita **není určována jen počtem** zkoumaných jednotek, ale **i mechanismem jejich výběru** ze základního souboru.

**Výběr na základě dostupnosti** také není tak vzácný, jak by se na první pohled zdálo. V případech, kdy se má zkoumat nějaká blíže neznámá populace, o které nejsou žádné údaje, neexistuje její soupis atp. (návštěvníci divadel, kin, hypermarketů, hradů a zámek atp.), nezbyvá nic jiného než se pokusit o jakýsi druh náhodného výběru osob přímo na místě. (Jisté dimenze náhodnosti má. Jestliže obchod obvykle navštěvuje 70 % rodičů s dětmi,

dá se takováto proporce očekávat i při náhodném výběru návštěvníků.) Někdy se tomuto postupu říká **výběr namátkou**. **Reprezentativnost tohoto výběru se zvyšuje**, dokážeme-li, například na základě předvýzkumu, přesněji určit případné rozdíly v množství a ve složení (třeba z hlediska pohlaví a věku) návštěvníků podle denní a týdenní doby apod., a takovéto rámcové kvóty potom promítnout do struktury a počtu výběrů. (Toto se někdy označuje jako **strukturovaný namátkový výběr**.)

Na každý pád je nezbytné stanovit pro všechny výzkumníky v terénu **závazný výběrový krok** (např. pauza 5 minut po ukončení posledního rozhovoru, potom oslovit desátého návštěvníka, který vejde do obchodu / vyjde ze zámecké brány / odejde od pokladny atd.). Tím se posiluje náhodnost výběru a současně se též částečně minimalizují možné (i třeba podvědomé) preference výzkumníků (např. ve věku, pohlaví aj.). Při rozhovorech se mnohdy všechna odmítnutí zaznamenávají včetně dostupných údajů (stanoviště, den v týdnu, hodina, odhadovaný věk, pohlaví aj.). Obdobných výběrových kroků je samozřejmě možné užívat i v případě pozorování.

V praxi se lze setkat ještě s dalšími podobami výběru na základě dostupnosti (či s výběry postavenými na obdobných nebo jiných principech), nicméně se domnívám, že předchozí ilustrace jsou pro naše potřeby dostačující.

#### 4.5 VÝBĚR V KVANTITATIVNÍM A KVALITATIVNÍM ZKOUMÁNÍ

Až dosud jsme si o rozdílech mezi zkoumáním kvantitativním a kvalitativním řekli již dost, takže jistě nepřekvapí, že určité diference lze najít též v jejich přístupech k výběrům. Předně jde o otázku zajištění reprezentativity, o níž jsem se v předchozím textu rámcově zmiňoval a k níž ještě učiním několik poznámek. Dále se jedná o rozsahy výběrů a za třetí o konkrétní rozdíly při jejich přípravě a realizaci, při tzv. vzorkování.

V **kvantitativním zkoumání** tkví zásadní požadavek na výběry v tom, aby námi vytvořený **výběrový soubor byl pro základní soubor reprezentativní**, jinými slovy **aby jej vhodně a spolehlivě zastupoval**. Z toho důvodu se v těchto případech pracuje s takovými výběrovými technikami, které to dokážou zajistit. Jak vyplývá z předchozího, jsou to postupy pravděpodobnostní (4.2) a z těch nepravděpodobnostních ty, u kterých známe stěžejní charakteristiky základního souboru, tedy výběr systematický a kvótní (viz 4.3). U těchto druhů výběru lze totiž **charakteristiky základního**



a vytvořeného výběrového souboru komparovat a zjišťovat, zda se případně odlišují, tj. identifikovat tzv. **výběrovou chybu**. Při jejím objevení je možné výběr ještě určitým způsobem poupravit, či získané výsledky s vědomím její existence interpretovat. (Výběrové chyby ovšem nelze určit u **namátkových výběrů**, viz 4.4, které se užívají dosti často. V těchto situacích totiž nemáme o základním souboru potřebné informace a nezbyvá než mít za to, že **reprezentativita je zajišťována spolehlivostí výběrového postupu**.)

*V praxi bývá analyzováno několik druhů výběrových chyb, nepovažují však za nezbytné se o nich rozepisovat a zájemcům o bližší informace nabízím literaturu v úvodu kapitoly.*

**Výběry v rámci kvantitativních zkoumání** se v některých pramenech souhrnně nazývají **statistické** (např. Hendl, 2005, s. 150–152). Je pro ně typické, že jejich rozsah je předem dán (viz dále), je u nich známo rozložení hlavních zkoumaných znaků nebo je lze alespoň částečně odhadnout, provádějí se jednorázově předepsaným způsobem a končí naplněním určeného rozsahu (tab. 4 na konci kapitoly). Třeba zdůraznit, že jejich **reprezentativitu podporuje přesné a jasné definování zkoumané populace, tj. základního souboru**.

Další **klíčovou podmínkou k dosažení reprezentativity je adekvátní rozsah výběru, tj. velikost vzorku**. Jeho určení v zásadě vychází z následujících úvah: Předně je třeba mít na paměti, že dle zákona velkých čísel platí, že **čím větší výběr, tím menší chyba**. (Až k extrému totálního výběru, viz 4.1, kdy je reprezentativita absolutní, neboť základní soubor se rovná výběrovému, což jak víme, nelze většinou uskutečnit.) **Výběr je tedy vždy jakýmsi kompromisem** mezi snahou o co nejpřesnější údaje a efektivitou beroucí v potaz jak nezbytnost zvoleného postupu, tak prostředky, které jsou pro jeho realizaci k dispozici. Obecně však platí, že rozhodování o rozsahu vzorku musí známost či odhadovanou **velikost zkoumané populace určitým způsobem reflektovat** (nikoli ale nějak mechanicky, např. v podobě přímé úměry).

Též je nutné – v souvislosti s konceptualizací problému, zvolenými hypotézami a znaky – zvážit **míru homogenity základního souboru**, neboť větší heterogenita ve sledovaných znacích vyžaduje rozsáhlejší výběr. (Například populace všech žáků pátých tříd základních škol bude v řadě charakteristik jako věk, sociální zařazení, příjmová skupina, ale také postoj

k určitým hodnotám apod. patrně homogennější než soubor televizních diváků.)

Konečně je nezbytné v rámci přípravy akce (viz 3.3, a posléze projektu, viz 3.6) promyslet všechny operace, kterým mají být získaná data podrobena. **Velikost vzorku** totiž určuje též **plánovaná podoba statistických výpočtů a jejich očekávaná přesnost** (rámcový výklad této problematiky v 7.1.4). Významnou roli tu hrají **tři aspekty** úzce s plánovanými operacemi související: Předně **členitost zkoumaných znaků, tj. počty jejich variant**. (Čím má znak více kategorií, na tím více podskupinek se pak při analýze souboru rozpadne. Tím by měl být rozsáhlejší, aby se v jednotlivých kategoriích tohoto znaku ukázaly nějaké zřetelnější četnosti.) Současně jde o **užití dalších stupňů třídění**, tj. o vzájemné porovnávání dvou a více znaků současně (blíže opět v 7.1.4). Přitom též, a to za třetí, záleží, **jak vysokou míru pravděpodobnosti mají statistické výpovědi mít** (viz v 7.1.4 zmínku o hladině významnosti).

*Budeme-li pro ilustraci vzájemně porovnávat dva znaky, například region bydliště a příjem domácnosti, z nichž každý bude mít deset variant/kategorií (což není vůbec nereálné), vytvoří nám tabulku 10 × 10, tj. o 100 políčkách. Jak budou asi jednotlivá pole obsazena, kdybychom pracovali s výběrem třeba 100 osob, přičemž námi plánované statistické výpočty (viz výklad v 7.1.4) přitom vyžadují, aby četnost v každém políčku nebyla nižší než 5? A co teprve v případě, kdyby nás rozložení příjmových kategorií v regionech zajímalo (na základě formulovaných hypotéz) ještě i z hlediska pohlaví, zastoupeného ve výběru v poměru 50:50 – to by potom tuto tabulku o 100 políčkách v jednom případě zaplnilo pouze 50 mužů a ve druhém 50 žen! Veškeré výpočty by tedy byly statisticky nekorektní a pro naše další interpretace v podstatě nepoužitelné.*

*Rozsah výběrového souboru je tedy zásadním způsobem ovlivněn mj. členitostí znaků a plánovanými stupni vzájemného třídění. Jeho konkrétní výpočet je ale trochu složitější a je třeba pro něj znát některé statistické hodnoty rozložení zkoumaných znaků v základním souboru či jejich odhady. (Izde jsem vždy spoléhal na týmového statistika. Takže ty, co by se o tom chtěli dozvědět něco více, musím pouze odkázat na uváděnou literaturu.) Některé publikace formulují na základě dlouhodobé praxe orientační rozsahy výběrového souboru např. takto: při velikosti základního souboru do 100 jednotek je nutné zkoumat 80 % z nich, do 1000 jednotek 40 %, do 10 000 7,5 % z nich, do 100 000 1,5 %, do 1 000 000 0,25 % a do 10 000 000 stačí 0,045 %.*



V **kvalitativním zkoumání**, které ohledně reprezentativity nemá zásadně žádné ambice, je proto otázka výběru řešena zcela odlišně. Vždy se jedná o **šetření nepřilíš velkých souborů** (několika málo desítek prvků) a vzhledem k základním rysům tohoto výzkumného přístupu většinou nejde o takové podoby výběrů, které jsme v předchozím výkladu označili za statistické. (I když pochopitelně též můžeme vybírat třeba náhodně, systematicky apod., nejsou to však výběry v tom smyslu, jak už je známe, hlavně co se jejich rozsahu a snahy o reprezentativitu týče.)

Pomiňme případy, kdy předmětem kvalitativního zkoumání je soubor, jehož rozsah a podoba jsou předem dány (např. všichni pracovníci malé firmy, všichni frekventanti určitého vzdělávacího kursu), tedy **výběry totální, cenzoční**. Patrně častější a typickou podobou pro tento druh zkoumání je nějaký nepravděpodobnostní výběr, jako **úsudkem nebo nabalováním** (viz 4.3), případně **na základě dobrovolnosti** (4.4).

Takovéto i další jim podobné výběry **pro potřeby kvalitativního šetření** bývají v literatuře někdy označovány jako **teoretické** (srv. Hendl, 2005, s. 151–152). Je pro ně charakteristické, že základní soubor často není definován, stejně jako klíčové zkoumané znaky, či rozsah takového výběru. Postup kvalitativního zkoumání podle teoretického výběru potom vypadá tak, že výzkumník (na základě úsudku, snowballu atp.) **postupně šetří** další a další jednotky. **Zkoumání – a tedy i výběr – ukončí** ve fázi, kdy se již ke zkoumanému problému přestává dozvídat cokoli nového, rozšiřujícího, obohacujícího, vysvětlujícího atp., **kdy dochází k tzv. teoretickému nasycení, teoretické saturaci**.

**Zásadní rozdíly v pojetí výběru u kvantitativního a kvalitativního zkoumání** je možné demonstrovat právě na odlišnostech mezi výše zmiňovaným statistickým a tímto teoretickým výběrem, které jsou pro každý z těchto přístupů typické. Jejich základní charakteristiky si na závěr shrneme do následující tabulky:

Tab. 4 Základní rysy statistického a teoretického výběru

	statistický výběr (kvantitativní zkoumání)	teoretický výběr (kvalitativní zkoumání)
základní soubor	musí být definován	není definován
rozsah výběru	je předem znám	není předem znám
znaky základního souboru	jejich rozložení je známé nebo je lze odhadnout	nejsou předem známé
provedení výběru	pouze jednou, podle předem zvoleného plánu	opakovaně, podle dosahovaných výsledků
ukončení výběru	po dosažení zvoleného rozsahu	po dosažení teoretického nasycení

Poznámka: srv. Hendl, 1997, s. 69 (též Hendl, 2005, s. 152).