

Pro sběr hromadných dat je v sociologii typický zejména *survey* – nejčastěji používané dotazníkové šetření (vedle například rutinně shromažďovaných statistik nebo panelových šetření) na výběrovém souboru. Blíže k logice *survey* a postupu při něm viz například de Vaus (1990).

### 1.3 Soubory a způsoby výběru jednotek

Již bylo řečeno, že data pro statistickou analýzu většinou shromažďujeme standardizovaným způsobem. Máme-li je statisticky zpracovat, musejí mít standardní podobu a musejí být zaznamenána jako čísla (neboli kardinální proměnné) nebo číslice (proměnné ordinální a nominální). Navíc bychom si měli uvědomit, že jen málokdy pracujeme s vyčerpávajícími šetřeními zahrnujícími všechny dostupné členy dané cílové populace neboli se základními soubory. Každý výzkum vyžaduje určení okruhu zkoumaných jednotek (osob), tedy určení „zkoumané populace“. Jestliže říkáme, že se naše „zkoumání zaměří“ na určitý soubor jednotek, máme na mysli většinou širší množinu jednotek, než postihne naše výzkumná činnost. Je to tzv. cílová populace (*general universum/population*). **Cílová populace** představuje okruh osob (soubor jednotek) vymezených nějakými sociálními či demografickými vlastnostmi, pro který chceme vyslovit závěry. Toto vymezení cílové populace může být různé a určuje i její velikost (zahrnuje všechny jednotky s vymezenými charakteristikami).

#### Ilustrace

Cílovou populaci může tvořit 4100 studentů, kteří navštěvují Fakultu sociálních studií v Brně; nebo to mohou být všechny ženy, které porodily v roce 2008 v porodnici města X své dítě; příjemci sociálních dávek s trvalým bydlištěm v Ostravě; obyvatelé celé ČR ve věku 15–65 let, ekonomicky aktivní obyvatelé ČR, obyvatelé ČR s volebním právem apod.<sup>15</sup>

Určení zkoumané populace je dáno především naším výzkumným tématem. Většinou přímo nezkoumáme, nemusíme zkoumat a mnohdy ani fyzicky zkoumat nemůžeme celou cílovou populaci – kdybychom tak učinili, provedli bychom tzv. **vyčerpávající zjišťování** neboli **census**. Proto převážně pracujeme s **výběrovými soubory** (*samples*), které jsou tvořené jednotkami vybranými podle určitých kritérií z cílového souboru, provádíme tedy **výběrová šetření**.<sup>16</sup> Jednotky pro náš výzkum vybíráme obvykle tak, aby byl výběrový soubor pro základní soubor **reprezentativní**.

<sup>15</sup> Všimněme si, že vymezení cílové populace vyžaduje často definici (co rozumíme ekonomickou aktivitou, jak je vymezeno volební právo, jaké sociální dávky máme na mysli apod.).

<sup>16</sup> Způsoby, jak tyto jednotky vybíráme, aby byl náš soubor reprezentativní, zde nepopisujeme, čtenář se s nimi může seznámit v publikacích věnovaných sociologické metodologii, jako jsou například Babbie, 2001, de Vause, 1990 či Disman, 1993.

Reprezentativnost znamená dvojí věc:

- Výběrový soubor má strukturu analogickou struktuře cílové populace – z hlediska známých i neznámých charakteristik jeho prvků –, což ovšem zajistí jen jejich **pravděpodobnostní (náhodný) výběr** (viz například Babbie, 2001 nebo Disman, 1993).
- Výsledky v něm zjištěné proto mohou být zobecněny na cílovou populaci. Co platí ve výběrovém souboru, platí i v cílové populaci. Takové zobecnění je možné ovšem jen za dodržení určitých podmínek při výběru jednotek – při dodržení pravidel pravděpodobnostního (náhodného) výběru.

Znovu připomínáme, že zobecnění je možné jedině pro cílovou populaci, z níž byl výběrový soubor vybrán, a pro žádnou jinou. Na to často autoři zapomínají, a proto se lze setkat i se statěmi, které mají například tendenci vypovídat o povaze sexuálního života české populace na základě dat získaných mezi klienty sexuologických ordinací a poraden. Potíž u takovýchto závěrů spočívá jednoduše v tom, že klienti sexuologických ordinací, byť by byli vybráni na základě všech pravidel pro reprezentativní výběr (tj. i kdyby všechny jednotky cílové populace – zde je možná na místě hovořit o základním souboru cílovou populaci zastupujícím – měly stejnou pravděpodobnost být vybrány do souboru výběrového), prostě nejsou reprezentanty pro výpovědi o sexuálním chování české populace, nýbrž pouze pro výpovědi o sexuálním chování souboru klientů poraden a ordinací (se všemi jejich zvláštními charakteristikami).

**Proto pozor: před získáním výběrového souboru musíte vždy vymežit cílovou populaci, a to není vždy tak jednoduché, jak by se na první pohled mohlo zdát.**

#### Ilustrace

Nelze zkoumat hodnotu mateřství jen v souboru žen s dětmi, ani jen v souboru žen vdaných. Můžeme se samozřejmě na tuto cílovou populaci omezit, ale pak si musíme být vědomi limitů svých výsledků. Jaký by asi byl příspěvek k poznání toho, „jakou hodnotu má dítě pro ženu“, bez dotazování žen, které dítě chtějí, ale mít nemohou (neplodnost, zdravotní problémy), či žen, které mít dítě programově odmítají?

Upozorňujeme také, že z jedné cílové populace, respektive z jednoho základního souboru lze učinit celou řadu výběrů (k této otázce se dostaneme podrobněji v pasážích věnovaných testování hypotéz a inferenční statistice). Závisí to na velikosti obou, nebo lépe na poměru jejich velikostí.