

# Kapitola 1

## Než začneme

*Jasnost je intelektuální hodnota; ne však přesnost a preciznost. Absolutní preciznost je nedosažitelná; je neúčelné chtít být přesnější, než to vyžaduje naše problémová situace.*

Karl R. Popper

*Aforismus o statistice aneb tři druhy lži: lež prostá, lež sprostá, statistika.*

Benjamin Disraeli

## Memento na začátek

V tomto učebním textu se budeme pohybovat v diskurzu kvantitativního výzkumu, v tzv. kvantitativním paradigmatu. Připomínáme, že sociální vědy jsou vědami multi-paradigmatickými, což znamená, že vedle sebe koexistují různé vzorce a pravidla, jak dělat vědu, jak řešit její hlavolamy. Různost těchto vzorců je v podstatě dána tím, jak si jednotlivá paradigmata odpovídají na tři základní otázky: ontologickou, epistemologickou a metodologickou. 1) Ontologická otázka se ptá, jaká je povaha reality, kterou zkoumáme. 2) Epistemologická otázka řeší, jaká je podstata poznání a jaký je vztah mezi tím, kdo poznává, a tím, co je poznáváno. 3) Metodologická otázka se pídí po tom, jakým způsobem se produkuje vědění, porozumění a pochopení. Na tomto základě se dnes definují tři základní skupiny paradigmat: pozitivistické, interpretativní a emancipativní (Mertens, 1998).<sup>3</sup>

**Kvantitativní paradigma má svůj vzor v přírodních vědách.** Vychází z přesvědčení, že realita je vnější a objektivně poznatelná. Klade velký důraz na měření vlastností, to je na jejich kvantifikaci. Jelikož převážná většina vlastností lidského chování

<sup>3</sup> V literatuře najdeme i další názvy: synonymicky s interpretativním paradigmatem se objevují výrazy etnografické, fenomenologické, hermeneutické nebo naturalistické paradigma. Vedle emancipativního paradigmatu nacházíme také výrazy feministické, participativní nebo kriticky teoretické paradigma.

a lidského světa, jimiž se sociální vědy zabývají, jsou složitými konstrukty a entitami, musíme se ve výzkumu velmi často spokojit s měřením ne přímo těchto vlastností (nejsou totiž přímo pozorovatelné), ale s měřením jejich pozorovatelných indikátorů. Nemůžeme například přímo změřit vzdělanost jedinců, ale na jejich vzdělanost můžeme usuzovat z výše dosaženého vzdělání. Vzdělanost je v tomto případě vlastností, úroveň dosaženého vzdělání jejím indikátorem.

Nemožnost přímého měření vlastností sociálního světa je v sociálněvědním výzkumu zdrojem jistých potíží. Mnozí metodologové proto zdůrazňují – a my s nimi –, že jedním z klíčových momentů kvantifikace a měření v sociálních vědách je **operacionalizace**, tedy převod abstraktních konstruktů do měřitelných znaků. S operacionalizací je spojen důležitý prvek, a to otázka **validity** těchto operací, což je posouzení, zdali námi vytvořený měřitelný znak (indikátor) je dobrým a skutečným reprezentantem vlastnosti, kterou chceme změřit – proto se také validita definuje jako schopnost měřit to, co skutečně měřit chceme.

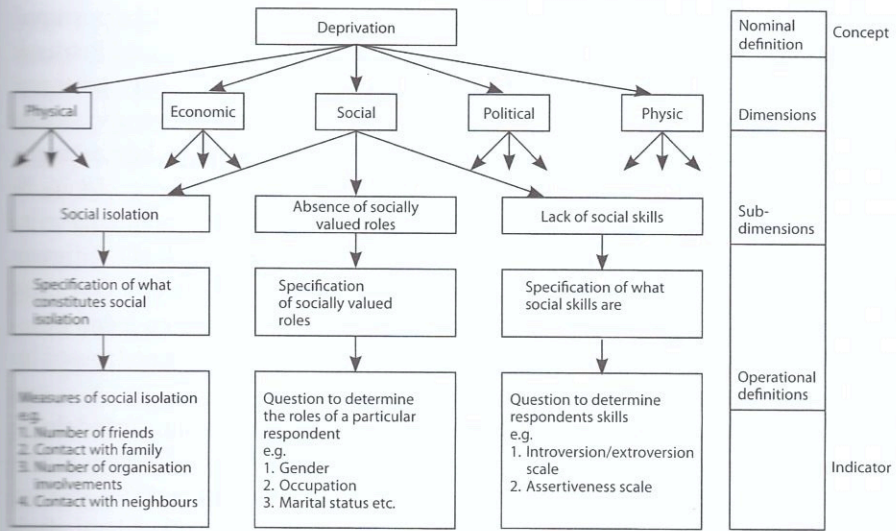
Operacionalizace je náročným tvůrčím procesem, v němž od sociálněvědních konceptů a jejich nominálních definic postupujeme přes odhalování jejich dimenzí a subdimenzí ke konkrétním operacím (operacionálním definicím), které nám říkají, co vlastně máme ve výzkumu zjišťovat a měřit. Názornou ukázkou necht' je příklad schématu operacionalizace, které použil de Vaus (1990) pro koncept (pojem) deprivace – viz obrázek 1.1. Co z něj rozhodně stojí za zapamatování, je, že při zjišťování míry deprivace nevystačíme pouze s jedním indikátorem – de Vaus jich navrhuje zjišťovat devět, a to ještě použití škály introverze/extroverze a škály asertivity obsahuje zjišťování řady dalších údajů.<sup>4</sup>

Operacionalizaci a měření vnímáme jako ústřední metodologické téma kvantitativního výzkumu. Jak konstruovat dobré, tedy validní a samozřejmě i dostatečně spolehlivé (reliabilní) měřicí nástroje a jakým způsobem měřit sociální vlastnosti, to jsou kardinální otázky kvantitativního paradigmatu. Podle zastánců paradigmatu kvalitativního jsou také zásadními – a z jejich pohledu jen obtížně překonatelnými – překážkami vědecké práce.<sup>5</sup>

Problém měření se navíc umocňuje tím, že velkou část našich kvantitativních dat získáváme na základě standardizovaných výpovědí, to je na základě standardizovaných rozhovorů tazatele se subjekty výzkumu. Standardizovanými výpověďmi rozumíme (i v dalším textu) výpovědi, které lze vyjádřit čísly (věk, příjem apod.) nebo číslicemi (přiřazenými jednotlivým variantám možných odpovědí), umožňujícími

<sup>4</sup> Zájemce o detailnější vysvětlení problematiky konceptualizace a operacionalizace odkazujeme např. na pasáže v učebnici E. Babbieho (Babbie, 2001, s. 119–145).

<sup>5</sup> Není cílem tohoto textu tento spor posuzovat, dodejme pouze, že po letech původně nesměřitelných diskuzí v poslední čtvrtině minulého století došlo nyní mezi oběma tábory ke smíru a ke koexistenci, především prostřednictvím tzv. mixed methods research, neboli metod smíšeného výzkumu, který kombinuje (směšuje) relevantní metody kvantitativního a kvalitativního výzkumu. Pro zájemce o metody smíšeného výzkumu může být dobrým úvodem kniha *Advances in Mixed Methods Research: Theories and Applications* editovaná Manfredem Maxem Bergmanem (Sage, 2008).



Podle: de Vaus (1990), str. 52.

Obr. 1.1 Schéma procesu operacionalizace pojmu „deprivace“

jejich statistické zpracování. Klasickým příkladem standardizovaného rozhovoru je použití dotazníku, v němž jsou všem zkoumaným subjektům kladeny stejné otázky ve stejném pořadí, přičemž je na ně možno odpovědět jen volbou (většinou) jedné z předložených variant možných odpovědí, označených číslicemi (např. otázka „jaké je vaše nejvyšší dosažené vzdělání“ s možností zatrhnout jednu z nabízených variant: 1 = základní, 2 = středoškolské; 3 = vysokoškolské).<sup>6</sup> Zde je dobré si připomenout, že při takovém způsobu měření může docházet k celé řadě poruch, neboť zaznamenáváme pouze tzv. verbální (neboli symbolické) chování, z něhož usuzujeme na chování skutečné. Při práci s takto získanými daty je proto nutné si neustále uvědomovat, že mezi skutečným předmětem výpovědi (to je jeho vlastnostmi) a tím, co je obsahem výpovědi (co se sděluje), může být obrovský rozdíl. Platí zde proto základní poučka, kterou si dobře pamatujeme: Výpovědi o realitě zdaleka nemusí být realitou samotnou!

Metodologové sociálních věd se ovšem snaží tyto potíže intenzivně překonávat a snad můžeme konstatovat, že se to daří. Nicméně při formulaci jakýchkoliv závěrů z kvantitativního výzkumu musíme mít neustále na mysli (a musíme to i v našich publikacích sdělovat), že naše závěry mohou být zatíženy nedokonalostmi naší operacionalizace a našeho měření. Permanentně tuto skutečnost reflektujeme!

Permanentní reflexe však není úplně jednoduchá. Ve hře je totiž naše psychika. Po převodu vlastností do znaků (proměnných) a po záznamu jejich hodnot – tedy

<sup>6</sup> Blíže viz oddíl 1.4.2, v němž hovoříme o proměnných.

po vytvoření čísel a jejich nahrání do statistického softwaru počítače – dochází postupně k tomu, že těmto číslům začneme bezmezně věřit. Nastává jev, který označil Petrusek (1993, s. 92) jako tzv. **durifikaci** (ztvrzení) **dat**. Což znamená, že s daty postupně začneme pracovat jako s naprosto přesnými čísly, bezmezně jim věříme a výsledky neproblematizujeme. Může to vést dokonce až k jakési hyper-přesnosti, kdy začneme uvádět (nereflektovaně!) výsledky na několik desetinných míst. Např.: „58,36 % respondentů nesouhlasilo s přijímáním imigrantů do zaměstnání“; „průměrná míra pocitu anomie byla 2,856“; „souvislost mezi úrovní dosaženého vzdělání a příjmem respondenta měřená Spearmanovým pořadovým koeficientem je 0,4681“. Osvícený výzkumník by samozřejmě hovořil o 58 %, průměr anomie by uvedl na jedno desetinné místo (2,9) a hodnotu koeficientu asociace zaokrouhlil na 0,47 – a nedopustil by se přitom žádné redukce informace, spíše naopak.

Z metodologického hlediska je základem pro kvantitativní výzkum metodologie přírodních věd. Organickou součástí sociálněvědního kvantitativního výzkumu je statistická analýza dat. Jsme přesvědčeni o tom, že právě ve statistické analýze spočívá značná síla tohoto přístupu: umožňuje víceméně exaktně (samozřejmě při vědomí všech možných omezení, která kvantitativní paradigma v sobě skrývá) popisovat zkoumané fenomény, navíc často v jejich vývoji, prostřednictvím analýzy časových řad; skýtá možnosti explarovat fenomény nové a umí ověřovat teorie – to vše za pomoci exaktního matematicko-statistického aparátu. Jelikož často pracuje s výběrovými soubory, dokáže s relativně malými náklady zobecňovat (generalizovat) své závěry na velké populace, a přinášet tudíž zobecnitelné poznatky, jež je možné využívat v praxi. Rozvoj sofistikovaných postupů statistické analýzy, jejich přepis do počítačových programů a obecná dostupnost osobních počítačů vede k tomu, že i ty nejmodernější postupy jsou dnes přístupny prakticky každému výzkumníkovi, včetně studentů.

Právě zde leží podle našeho názoru obrovská potence, neboť moderní postupy statistické analýzy pomáhají odkrývat vztahy a souvislosti, které bychom jinak v datech nebyli schopni vidět. Jdou do latentních datových struktur, odhalují efekty působení nezávislé proměnné na proměnnou závislou, očištěnou od efektů dalších proměnných, umí smysluplně pracovat s kategorizovanými daty a tak dále, a to vše ve velmi krátkém čase. Tyto postupy pak vedou k výsledkům, které dříve nebylo možné nalézt a formulovat. Hlavní směr rozvoje postupů moderní statistické analýzy v sociálních vědách byl veden několika základními cíli: popsat a vysvětlit nějaký sociální jev znamená odhalit paralelní působení řady faktorů, které často mají hierarchický charakter.<sup>7</sup> Proto byly zavedeny nejrůznější modelovací techniky. Jelikož sociálněvědní data jsou velmi často data kategorizovaná – a s těmi toho statistika zatím mnoho nesvedla –,

<sup>7</sup> V dalších pasážích této úvodní kapitoly čtenář možná nalezne pojmy, které mu nebudou zcela jasné. Není to jeho chyba, prosíme o strpení, postupně budou všechny vysvětleny v dalších kapitolách. Věříme, že to není ani chyba didaktická, jak by se v takovém případě u učebního textu mohlo zdát. Považovali jsme prostě za nutné na úvod sdělit naše zásadní stanovisko k problematice kvantitativní analýzy dat, byť s vědomím, že mnohé bude jasnější až po přečtení celého textu. Doporučujeme vrátit se k první kapitole po přečtení celé učebnice.

byly vyvinuty nové postupy pro analýzu kategorizovaných dat, které tento hendikep překonávají (např. binární a logitová regrese, loglineární analýza). Většina těchto postupů je součástí běžně dostupných statistických programových balíčků (*statistical packages*). Možnosti statistické analýzy jsou tedy značné. Otázkou ovšem je, zdali je umíme dobře používat. A v tom spočívá, domníváme se, kardinální limit české kvantitativní sociální vědy, sociologii nevyjímaje.

V Čechách, na Moravě a ve Slezsku působí zatím jen nemnoho badatelů, kteří umějí možnosti současné statistické analýzy plně využívat. Jednou z příčin je, že statistika bývá při studiu sociálních věd často obávaným předmětem, kterým je potřeba „nějak projít“, bez ambice pochopit její smysl a kouzlo. Studenti se statistiky obávají – čísla jsou pro humanitně orientované osoby strašákem – a mnozí učitelé statistiky studenty tohoto strachu nejenom nezabavují, ale ještě jej zvyšují. A přitom, jak věříme, je možné naučit statistickou analýzu i ty studenty, kteří šli studovat sociální vědy právě z důvodu, že se báli čísel a – jak oni říkají – „matematiky“. Možná právě způsob, jakým se v Česku učí kvantitativní metody výzkumu a analýzy, vede k tomu, že se dnes mezi studenty sociálních věd stal do značné míry módou výzkum kvalitativní (kvali-výzkum). Studenti se domnívají, že je to výzkum lehčí, neboť je založen na analýze slov namísto analýzy čísel. Hluboce se však mýlí – takový výzkum není lehčí, neboť stejně jako kvanti-výzkum klade obrovské nároky na schopnost sociologické imaginace. Jsme pevně přesvědčeni o tom, že student, který nezvládne metodologii výzkumu kvantitativního a analýzu jeho dat, nebude ani dobrým výzkumníkem kvalitativním.

Analýza studentských prací by nám jistě odhalila i to, že v českých sociálních vědách panují některé obsese, jichž je potřeba se urychleně zbavit. Uvádíme je níže – a jelikož jsme je s úpravami převzali od anglického kolegy, který má podobné pocity a zkušenosti (viz Blaikie 2003, s. 6–7), s jistou škodolibou útechou konstatujeme, že v tom asi nejsme v Česku tak úplně sami.

- a) Sociálněvědní výzkum musí vždy začínat s hypotézami.
- b) Testy statistické významnosti (statistické signifikance) jsou esenciálním rysem analýzy dat.
- c) Zjištěná míra asociace nebo korelace mezi dvěma znaky (proměnnými) postačuje k vysvětlení zkoumaného jevu.

Tyto poněkud radikální teze nyní rozvíňme.

### **Ad a) Musí sociálněvědní výzkum začínat s hypotézami?**

Jednoduchá odpověď zní: jak kdy. Argumenty jsou následující:

- Každý sociálněvědní výzkum musí mít na svém počátku nějaký problém, který je přeložen do zkoumatelné podoby formou výzkumné otázky (nebo několika výzkumných otázek).
- Existuje několik typů výzkumných otázek: otázky „co“, „kolik“, „do jaké míry“ popisují věci a jsou tedy typické pro deskripci. Otázka „proč“ je otázkou na příčiny a je

typická pro explanační (vysvětlující) výzkum. Otázka „jak“ je otázkou na sociální mechanismy a je typická pro akční výzkum, pro intervenci.

- Pouze otázka „proč“ vede badatele k výzkumu, který je založen (deduktivně) na teorii, a proto vyžaduje hypotézy.
- Existují dva druhy hypotéz: **teoretické hypotézy**, které jsou odvozovány z teorie a které nabízejí předběžné vysvětlení otázky typu „proč“. Vedle toho jsou zde **hypotézy statistické**, které se používají k **zobecnování výsledků z reprezentativního výběrového souboru na cílovou populaci**,<sup>8</sup> z níž byl výběrový soubor získán.
- Tento rozdíl mezi statistickou a teoretickou (výzkumnou) hypotézou není často akceptován a vede ke zmatku u začínajících výzkumníků.
- Teoretické hypotézy jsou relevantní pouze v případě, kdy hledáme odpovědi na otázku „proč“; statistické hypotézy jsou relevantní, když data pocházejí z pravděpodobnostního (náhodného) či randomizovaného<sup>9</sup> výběru, při kterém mají všechny osoby v dané populaci na začátku výběru stejnou pravděpodobnost, že budou do zkoumaného výběrového souboru vybrány.<sup>10</sup> Některý druh výzkumu může vyžadovat oba druhy hypotéz, některý může vyžadovat pouze jeden z typů; značná část výzkumu ovšem nevyžaduje ani jeden z nich. Co však každý výzkum musí mít, je výzkumná otázka!
- Určitý druh výzkumu hypotézami končí, místo aby jimi začínal (tzv. explorační výzkum).

#### **Ad b) Jsou testy významnosti esenciálním rysem analýzy dat?**

Nejdříve opět jednoduchá odpověď: ne, nejsou. Avšak v českých sociálních vědách (a nejen v nich) bohužel velmi často bývají.

Testy signifikace jsou pravděpodobně nejhůře pochopeným aspektem statistické analýzy dat. Jsou součástí tzv. statistické indukce (nebo také inferenční analýzy či statistického usuzování) a používají se tehdy a jen tehdy, když pracujeme s pravděpodobnostním výběrovým souborem. Slouží k tomu, abychom z charakteristik výběru odhadli charakteristiky populace.

Inferenční analýza se využívá pro dva typy úloh: a) k odhadu charakteristik populace z dat výběrového souboru (odhadujeme např. průměrný příjem); b) ke zjištění, zdali vztah (či rozdíl) nalezený ve výběru je možné také očekávat v populaci, z níž byl

<sup>8</sup> Je zvykem hovořit často spíše o základním souboru než o cílové populaci, ale spolu s řadou dalších autorů rozlišujeme mezi cílovou populací jakožto souborem jednotek, na které chceme zobecnit své závěry, a základním souborem tuto cílovou populaci zastupujícím (ne všechny jednotky této cílové populace jsou totiž ve všech případech v daném okamžiku výzkumu dostupné).

<sup>9</sup> Randomizace jako náhodné přiřazení osob do zkoumaných skupin (experimentální a kontrolní) se často používá v klinickém výzkumu v medicíně či v psychologii, ale i v ekonomii, méně často v sociologii.

<sup>10</sup> Nikoliv tedy nahodilého!

vybírán (toto je důležité). Např. odhadujeme, zdali i v populaci platí, že existuje vztah mezi vzděláním respondenta a jeho intencí účastnit se kurzů celoživotního vzdělávání.

První typ inferenčních úloh není v sociologii příliš častý; výzkumníci málokdy počítají pravděpodobnou hodnotu populační charakteristiky; často se používá tzv. bodového odhadu, kdy se prostě předpokládá, že hodnota výběrové charakteristiky bude stejná i v populaci (což ovšem není úplně v pořádku, měly by se počítat intervaly spolehlivosti). Druhému způsobu inferenční analýzy, testům statistické významnosti (jimiž jsou, jak uvidíme později, např. test chí-kvadrát pro nominální znaky, test významnosti pořadových koeficientů, t-test pro rozdíl dvou průměrů, analýza rozptylu pro rozdíl více průměrů) je naopak věnována velká pozornost a podle našich zkušeností jsou tyto postupy dokonce v české sociální vědě nadužívány, zneužívány a používány špatně (blíže k tomu viz Soukup & Rabušic, 2007). Je to proto, že je špatně pochopen jejich účel a smysl, takže se používají k operacím, pro které nejsou vhodné.

Badatelé a badatelky se zkrátka setrvačně domnívají, že jim testy významnosti řeknou, co je v datech důležitého, a že jim pomohou odhalit těsnost vztahu dvou proměnných. Dále jsou přesvědčeni, že tyto testy musejí být aplikovány na všechny výsledky bez ohledu na to, zdali data pocházejí z vyčerpávajícího zjišťování (z censu), z pravděpodobnostního (náhodného) výběru nebo z výběru nenáhodného (kvótního, záměrného, samovýběru).

**Nic z toho ovšem statistická inference neumí.** Z toho tedy vyplývá, že testy významnosti:

- nemohou v žádném případě sloužit k rozhodnutí, zdali je zjištěný výsledek vědecky nebo prakticky důležitý;
- nejsou míry asociace;
- jsou použitelné pouze tehdy, testujeme-li statistické hypotézy, a ty používáme pouze tehdy, když z dat pravděpodobnostního (náhodného) výběrového souboru odhadujeme charakteristiky populace;
- mohou být aplikovány pouze tehdy, pracujeme-li s výběrovým souborem, který byl vybrán z populace za pomoci postupů pravděpodobnostního (náhodného) výběru a návratnost (např. dotazníků) je relativně vysoká – měla by se, jak nabádá Blaikie (2003, s. 167) pohybovat kolem 85 %;<sup>11</sup>
- je mylné a chybné používat v případech, kdy výběr není pravděpodobnostní. A už žádný smysl nemá jejich použití v případě, kdy nemáme výběr, ale náš soubor je populací (např. když provedeme výzkum na všech žácích gymnázia v městě X);
- nemohou být použity k testování teoretických hypotéz;
- nemohou sloužit ke generalizujícím výpovědím za populaci, z nichž nebyl soubor vybrán. Například pokud provedeme pravděpodobnostní (náhodný) výběr

<sup>11</sup> V současnosti se bohužel tak vysoké návratnosti dosahuje poměrně obtížně – k tomu viz např. článek J. Krejčího „Problém nízké návratnosti výběrových dotazovacích šetření“ (*SDA info 2*, listopad 2006 – dostupný z: <http://archiv.soc.cas.cz/articles/cz/14/SDA-Info.html>) nebo obsáhlejší Krejčí, J. (2007). Non-Response in Probability Sample Surveys in the Czech Republic. *Sociologický časopis / Czech Sociological Review*, 43(3), 561–587.

z populace všech studentů gymnázií v Brně (ze seznamu všech brněnských gymnázií bychom si udělali seznam všech tříd a z tohoto seznamu bychom vylosovali takový počet tříd, aby náš výběrový soubor měl dostatečný počet jednotek), naše závěry nemůžeme zobecňovat na všechny studenty brněnských středních škol (my jsme totiž dělali výzkum pouze na gymnáziích, přičemž víme, že existují ještě další typy středních škol), a už vůbec ne na studenty v jiných městech.

#### **Ad c) Je zjištěná dvourozměrná míra asociace nebo korelace postačující pro vysvětlení?**

Ne, není. Důvody jsou následující:

- Vysvětlením říkáme, proč něco existuje. Nalezení těsnosti vztahu mezi dvěma proměnnými je součástí deskriptivního výzkumu, neboť cílem deskripce je určit charakteristiky nějakého sociálního fenoménu, popsat jeho vývoj v čase a typické vzorce vazby na jiné fenomény.
- Nalezení vztahu mezi proměnnými a jeho změření prostřednictvím měr asociace a korelace je formou vyšší míry deskripce.
- Ačkoliv tato deskripce může posloužit k pochopení fenoménu (a někteří i tvrdí, že může sloužit jako základ k predikci), míry asociace nemohou přinést odpověď na otázku „proč“, neboť otázka „proč“ se ptá na příčiny. Asociace vyjadřuje pouze stochastickou (pravděpodobnostní) souvislost, nikoliv však souvislost příčinnou, kauzální.
- Nicméně, chceme-li vysvětlit nějaký jev, musíme nejdříve najít asociaci – neboť tam, kde není žádná stochastická asociace, nemůže být ani kauzální spojení.

Jak je vidět, dobře pochopená a osvojená statistická analýza dat může být klíčem k badatelsky platným výsledkům. Statistická analýza je mocným nástrojem, ale samozřejmě pouze v takových rukách, které vědí, jak na to. Při provádění kvantitativního výzkumu proto mějme stále na paměti zásady, na nichž musí být každý dobrý kvantitativní výzkum založen (viz Blahuš, 2000):

- 1) Vědecká průkaznost výsledků výzkumu spočívá v logicky správném a metodologicky čistém designu výzkumu.
- 2) Design výzkumu nemohou nahradit žádné dodatečné statistické, byť téměř akrobatické cviky s daty.
- 3) Vědecká průkaznost výsledků výzkumu nespočívá v jejich „statistické významnosti“.



## 1.1 Logika kvantitativního výzkumu

**V kvantitativním výzkumu pracujeme s hromadnými daty.** Tato data jsme získali na základě designu výzkumu, který se odvíjí od naší výzkumné otázky. Ta vzniká většinou na základě naší touhy vysvětlit nějaký problém. Výzkumná otázka určuje, co sledovat, jaké vlastnosti měřit. V sociologii tato hromadná standardizovaná data (viz dále) získáváme většinou dotazníkovým šetřením na výběrových souborech (viz dále), kdy jednotkou dotazování je většinou jedinec, nebo ze statistických výkazů.

Dotazujeme-li se jedince, chceme:

- zjistit jeho stav, identifikovat jeho měřitelné vlastnosti (například: je to muž, nebo žena?; jakou stranu volí?... ) a míru těchto vlastností (jaké má vzdělání?; jakou mírou anomie se vyznačuje?; do jaké míry souhlasí s určitým názorem?...);
- klasifikovat jedince podle těchto zjištění do obecnější kategorie jednotek (volí levice, nebo pravici?; je liberál, nebo etatista?...);<sup>12</sup>
- usuzovat z výskytu nějaké vlastnosti (intenzity vlastnosti) na jinou vlastnost (respektive její intenzitu);
- sledovat vývoj kvantifikovaných vlastností jedinců v čase.

Vlastnostmi neboli charakteristikami, kterými je jedinec popisován, nejsou jen jeho psychické atributy a stavy (rasy osobnosti, inteligence, neurotické stavy, emoční vyladění, frustrační tolerance), demografické charakteristiky (pohlaví, věk, počet dětí) a jeho zařazení do sociální struktury (sociální třída, profese, velikost místa bydliště), či výrazy tohoto zařazení (sociální status, prestiž), ale i jeho postoje a preference (volební preference, hodnotové orientace, míra xenofobie či rasismu, distance od jiných osob či sociálních skupin, obliba určitých TV programů, religiozita či náboženská konfese, etatismus či liberalismus, způsob trávení dovolené), aktivity či jednání (náplň volného času, účast v sociálních hnutích, návštěvy kulturních akcí, účast na stávkách), vědomosti či míra informovanosti a stavy či podmínky, jimž je vystaven (nezaměstnanost, anomie, nemoc, dlouhodobý stres, deprivace) atd.

### Rámeček 1.1 Vlastnosti jedince

V sociologii jsou ovšem údaje o jedincích jen přechodnou informací a slouží k tomu, abychom získali informace o sociálních útvech (skupinách, kategoriích, institucích), jejichž členy/členkami tyto jedinci jsou. Nezajímají nás tedy ani tak naměřené

<sup>12</sup> Na to se lze v prvním případě zeptat přímo (předložením škály se stupnicí a póly levice/pravice se žádostí, aby se na ni respondenti zařadili podle svých politických názorů), ale zařazení lze také opřít o klasifikaci jednotlivých politických stran jako levicových, pravicových či středových a jedince přiřadit k levici, středu či pravici podle jím volené strany. V druhém výše uvedeném případě se ptát přímo nemůžeme nejen proto, že mnozí z dotazovaných by neměli představu, na co se jich vlastně ptáme. Problémem by bylo i to, že ti znalější by mohli používat vlastní definice liberalismu a etatismu – a my bychom nevěděli, co si vlastně pod těmito pojmy jednotliví respondenti představovali. Výzkumník musí mít vlastní definici obou postojů a z ní odvozené otázky indikující jejich přítomnost u daného jedince. Na tuto skutečnost ovšem nesmí zapomínat ani při interpretaci svých výsledků. Nejde o liberalismus/etatismus obecně, abstraktně, ale o liberalismus/etatismus jím konkrétně definovaný.

hodnoty u jednotlivých subjektů, ale tendence, která se v naměřených hodnotách projevuje u – z jistého hlediska – homogenních skupin či kategorií těchto objektů (u žen, u osob s vysokoškolským vzděláním, u osob s vysokou mírou anomie apod.). Míry závislosti či souvislosti mezi vlastnostmi jsou měřitelné jen v souborech objektů.

### Ilustrace

Měření souvislosti mezi pohlavím a vzděláním není na úrovni jednoho subjektu smysluplné. Smysl dostává až na úrovni „z jistých hledisek homogenních souborů“ subjektů, kde můžeme konstatovat například: *Mezi nezaměstnanými* (soubor osob homogenní z hlediska jejich postavení na trhu práce) *je více žen než mužů* (soubory osob homogenní z hlediska jejich pohlaví). Nebo z jiného úhlu: *Mezi ženami je více nezaměstnaných než mezi muži*. Popřípadě: *Příslušníci jednoho pohlaví (ženy/muži) mají vyšší/nížší míru nezaměstnanosti než příslušníci druhého pohlaví*. A tak dále.

V případě souboru, skupiny chceme:

- zjistit stav vlastností jednotek, z nichž je soubor složen, identifikovat jejich vlastnosti (míru těchto vlastností), popsat je pomocí vlastností souboru a zjistit rozložení těchto vlastností v souboru: u kategorizovaných proměnných počty a podíly jednotek s určitou vlastností, intenzitou vlastností v souboru (například: *kolik bylo ve sledovaném souboru katolíků a jaký podíl v něm tvoří? ...*), u spojitých proměnných stanovit střední hodnoty (například: *jaký je průměrný věk, příjem, ... v daném souboru?*);
- hledat příčiny variability neboli hledat vztahy či souvislosti mezi vlastnostmi (proměnnými): pohlaví, vzdělání nebo věk, příjem, respektive vztah mezi pohlavím a vzděláním či věkem a příjmem (*jak se mění výše příjmu v závislosti na vzdělání?*);<sup>13</sup>
- sledovat vývoj kvantifikovaných vlastností v souboru. Z hlediska sociologické metodologie zde lze použít opakované výzkumy, kohortní analýzu či panelová šetření (což jsou formy longitudinálních výzkumů). Ke zpracování dat shromážděných těmito technikami má statistika vlastní nástroje, s nimiž se ovšem v rozsahu našeho textu seznamovat nebudeme.

## 1.2 Hromadná data

Kvantitativní výzkum soustředící se na to, **jak mnoho** (jaký podíl) něčeho ve společnosti existuje a **jak něco s něčím souvisí**, pracuje s hromadnými daty. V hromadnosti se projevují pravidelnosti, které u jednotlivých případů nemůžeme identifikovat (setkáváme-li se s jednotlivými případy, působí na nás dojemem obrovské proměnlivosti a individuálnosti), zatímco v hromadných datech existují jisté tendence. Například individuálně se můžeme setkat s velkou variabilitou příjmů a můžeme nalézt případy, kdy osoby se středoškolským vzděláním mají vyšší příjmy než některé osoby se vzděláním

<sup>13</sup> Problém závislosti statistické a kauzální jsme již otevřeli v úvodu tohoto textu. Blíže bude popsán v kapitole o měření statistické souvislosti. Dopředu pouze upozorňujeme, že mluvíme-li v tomto textu o vztahu, nemáme ve většině případů na mysli vztah kauzální (příčinný).

vysokoškolským,<sup>14</sup> přesto v hromadných datech nepochybně zjistíme již zmíněnou tendenci: čím vyšší stupeň vzdělání, tím je mezi osobami, které ho dosáhly, vyšší průměrný příjem.

### Ilustrace

Lidé uzavírají sňatky v nejrůznějších životních situacích a s partnery s nejrůznějšími osobními i sociálními charakteristikami – v reálném životě se setkáme se všemi kombinacemi podle výše vzdělání obou partnerů. Jakkmile ale přestaneme posuzovat jednotlivé případy a začneme brát v úvahu větší počty, začnou se projevovat určité pravidelnosti, respektive nepravidelnosti, například tendence k homogamii (to je tendence brát si partnera s analogickým sociálním statutem, vzděláním apod.). A zjišťovat tyto tendence je jedním z důležitých úkolů statistické analýzy hromadných dat.

Hromadná data se zpracovávají statistickou procedurou, jejíž obsah tvoří:

- **výběr jednotek pro sledování** (zjišťování jevů či charakteristik). Zde se rozhodujeme, jaký druh výběru použijeme, zdali pravděpodobnostní, kvótní, záměrný apod.;
- **zjišťování údajů** (prostřednictvím pozorování a měření), většinou u velkého počtu jednotek (pozor: sama statistika nemůže nahradit konceptualizaci pojmů a konstrukci nástroje sběru dat). Zde měříme vlastnosti jednotek, které nás na základě naší výzkumné otázky zajímají;
- **kontrola údajů**, jak formální, to je kontrola návratnosti a úplnosti údajů, tak i logická, kdy kontrolujeme nerozpornost údajů. Zde tzv. „čistíme“ data;
- získání informace o struktuře souboru z hlediska jednotlivých sledovaných vlastností jeho jednotek. **Třídění údajů** ústící do informace o rozložení jevů (třídění prvního stupně), popřípadě o rozložení kombinace jevů či charakteristik objektů (neboli třídění vyšších stupňů). Zde se již dostáváme do oblasti statistické analýzy;
- **výpočet statistik** (středních hodnot charakterizujících rozložení hodnot proměnných, koeficientů asociace charakterizujících sílu, respektive směr vztahů mezi proměnnými) a sledování časové (časové řady) i věčné (testování významnosti rozdílů a shody) proměnlivosti jevů;
- **vyjádření rozložení jevů** (charakteristik) v tabulkách a grafech, časových řadách; **relační neboli vztahová analýza** (hledající souvislosti), která může být jednoduchá a vícefaktorová nebo také kauzální (hledající příčinné závislosti);
- **statistická verifikace hypotéz**. Pozor, jde o verifikaci operacionalizovaných hypotéz, které lze dosáhnout jen s určitou přesností – určitou pravděpodobností chyby. Není verifikací věcnou!
- **inferenční statistika** (statistická indukce) neboli zobecnění výsledků našeho výběrového souboru na cílovou populaci, z níž byl vybrán (jde o pravděpodobnost platnosti našich výsledků i v cílové populaci, samozřejmě za předpokladu, že pracujeme s reprezentativním výběrovým souborem).

<sup>14</sup> Nezapomeňme ovšem na možnost intervence dalších faktorů do vztahu dvou vlastností (proměnných), to je mezi vyšší vzdělání a vyšší příjmu. K tomuto problému se ještě dostaneme, zde jen připomeňme, že je třeba vzít v úvahu působení přinejmenším třetího faktoru, jímž je věk jedince. To je příklad uváděný Dismanem (1993, s. 24) jako dvojí příčina nebo také nepravá závislost.

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 4 100 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 vymezit 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.

## 1.4 Měření

Chceme-li analyzovat výzkumná data, musíme je nejdříve získat. Ačkoliv se kolem nás pohybuje obrovské množství dat, která jsou již někde uložena a dychtivě čekají, abychom je dále zpracovali,<sup>17</sup> ne vždy jsou po ruce data taková, abychom mohli vyřešit naši výzkumnou otázku. Proto musíme jít často do terénu a vlastní výzkumná data získat, tj. musíme fenomény, s nimiž budeme operovat, změřit.

Již Galileo Galilei formuloval na přelomu 16. a 17. století pro vědu požadavek **měřit všechno, co je měřitelné, a snažit se učinit měřitelným vše, co dosud měřitelné není.**

Tento přístup vedl v přírodních vědách v 18. století, a zejména v 19. století k revoluci poznání. Není divu, že učaroval i mnoha sociálním vědcům. V pozitivistické empirické sociologii bylo měření od počátku chápáno jako jediná záruka vědeckosti jejich výsledků. Tento radikální požadavek je dnes již minulostí, model poznávání sociální reality skrze měření ovšem (a to je dobře) mrtvý není. Jen se musel tvářit v tvář kvalitativním přístupům zřici monopolu, který byl nejzřetelněji zformulován a uplatňován v 30. až 50. letech dvacátého století.<sup>18</sup>

Základním předpokladem měření v sociologii je, že sociální objekt či jev, respektive jeho atributy (vlastnosti) jsou popsateľné pomocí čísel nebo číslic, které jsou jim připsány v procesu měření. Tyto číselné vlastnosti ovšem objekt nemá de facto – jsou mu připsávány teprve v tomto procesu.

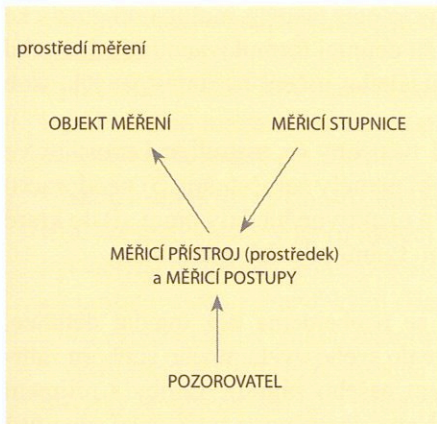
Měření lze jednoduše definovat slovy klasika měření Stanleyho Stevense – který vlastně jen parafrázoval definici jiného klasika měření, Normana Campbella – jako přiřazování čísel objektům a jevům podle pravidel. Český statistik Jan Řehák tuto definici rozvinul a formuloval ji následovně: „Měření je proces realizace homeomorfních zobrazení empirického systému s relacemi do jednoznačně určeného abstraktního systému s relacemi“ (Řehák, 1971, s. 646).<sup>19</sup>

Celý proces měření lze znázornit jednoduchým schématem (viz obr. 1.2). Popíše-li jej slovně (a zde si vypůjčujeme formulace Řeháka), pak říká: Pozorovatel měří vlastnosti **objektu měření** tak, že mu přiřazuje hodnoty **měřicí stupnice** (škály) pomocí **měřicího přístroje** a měřících postupů v určitém **prostředí měření**. Z toho vyplývá,

<sup>17</sup> Data např. leží (v agregované podobě) na statistických úřadech (v Českém statistickém úřadě, v evropské databázi Eurostat, datové databázi OSN atd.) nebo jsou – jako primární data – uložena v datových archivech sociálněvědních výzkumů a čekají na sekundární analýzu, to je na nové zpracování. Blíže k tomu viz přílohu, která obsahuje internetové linky aktuální v době vydání této publikace.

<sup>18</sup> Upozornění na limity tohoto přístupu naleznete již v klasické práci Charlese Wrighta Millse *Sociologická imaginace* z roku 1959 (česká vydání v letech 1968 a 2002).

<sup>19</sup> Ostatně jeho stať o měření nazvaná „Definice měření ve společenských vědách“, kterou publikoval v *Sociologickém časopise* v r. 1971, stojí rozhodně za přečtení.



Obr. 1.2 Obecné schéma procesu měření

Podle: Řehák (1971), str. 639.

že abychom mohli úspěšně měřit, musíme mít 1) jednoznačně určenou stupnici, jejímž prostřednictvím měříme; 2) musí být zkonstruován reliabilní a validní měřicí přístroj a musíme vědět, jaké jsou adekvátní měřicí postupy; 3) musí být maximálně eliminován vliv prostředí a 4) musí být maximálně eliminován vliv subjektu pozorovatele na výsledek měření. Žádná z těchto podmínek není triviální a v sociálních vědách může být vážnou překážkou měření. Pro ilustraci: Jak bychom asi v tomto kontextu změřili míru individuálního blaha (tzv. well-being), která, jak naznačují někteří ekonomové, by mohla sloužit jako obecný ukazatel celkové ekonomické úspěšnosti a ekonomického pokroku společnosti? Na měření blaha přece nemáme ani žádnou reliabilní stupnici, ani validní měřicí prostředek a rovněž by se poněkud obtížně eliminoval vliv prostředí (pokud by v době měření ve společnosti převažovala „blbá nálada“, byly by výsledky zřejmě jiné, než kdyby panoval duch veselí a optimismu).

Znovu jsme se tak obloukem dostali k problematice operacionalizace. Teoretický rozbor našeho konceptu a určení jeho dimenzí a subdimenzí po vzoru de Vausova schématu (viz obr. 1.1) by nás jistě přivedl k formulaci adekvátních a validních indikátorů a ke způsobu jejich měření – vždyť přece ve vědě je vše měřitelné, a co není měřitelné, se měřitelným musí učinit!<sup>20</sup>

### 1.4.1 Koncepty a jejich operacionalizace – indikátory

Chceme-li nějaký jev změřit, musíme vědět, jak je daný jev vymezen, ohraničen, definován. Musíme tedy mít jeho koncept a jeho definici. Když budeme zjišťovat např. podíl nezaměstnaných v ekonomicky aktivní populaci (míra nezaměstnanosti), musíme mít definici nezaměstnanosti – kdo je za nezaměstnaného považován. Definice

<sup>20</sup> A skutečně, základní kroky již byly v tomto ohledu učiněny – viz zprávu: Stiglitz, J., Sen, A., & Fitoussi J.-P. (2008). *Issues paper – Commission on the Measurement of Economic Performance and Social Progress*, která je dostupná na: <http://www.stiglitz-sen-fitoussi.fr/en/index.htm>.

konceptu může být ovšem více – ostatně v sociologii jakožto multiparadigmatické vědě je to jev docela častý. Můžeme např. použít definici formulovanou Mezinárodní organizací práce (ILO), která je základem oficiálního určení nezaměstnaných osob v zemích EU, a tedy i v České republice:

Definice ILO považuje za nezaměstnané ty osoby (v metodice Eurostatu ve věku 15–74 let), které: a) v referenčním období neměly zaměstnání, b) neodpracovaly ani jednu hodinu za mzdu nebo odměnu a c) aktivně hledaly práci, d) do které by byly schopny nastoupit nejpozději do dvou týdnů ode dne zjišťování.

Ještě ilustrativnějším příkladem toho, že se neobejdeme bez úvodní definice, může být zkoumání chudoby. Jaké osoby máme do svého výzkumného souboru zahrnout (kdo je vlastně chudý, a tudíž předmětem našeho zájmu)? Osoby s příjmem pod hranici stanoveného životního minima? Nebo domácnosti vydávající více než 30 % svých příjmů na potraviny? Či osoby s příjmem nedosahujícím 60 % mediánu příjmového rozložení v dané zemi (což je definice Eurostatu)? Anebo osoby či domácnosti dosahující určité hodnoty na indexu deprivace? Popřípadě osoby, respektive domácnosti, které se deklarují jako chudé (což je tzv. subjektivní chudoba)? To vše jsou příklady definic chudoby a nejde vůbec o jejich vyčerpávající výčet.<sup>21</sup> Chudobu tedy nikdy nezkoumáme samu o sobě, vždy půjde o její určitý koncept. Proto budeme-li analyzovat data o chudobě, nebudeme vypovídat o chudobě jako takové, ale o chudobě, jak jsme ji definovali (nebo podle definice chudoby, kterou jsme přijali). Tento aspekt sociálních věd mějme neustále na paměti, když vypovídáme o sociálních jevech, které zkoumáme – to znamená, řečeno jinými slovy – při interpretaci výsledků našich analýz.

V sociálních vědách se celá věc ještě komplikuje tím, jak jsme již uvedli na začátku této kapitoly, že vlastnosti jednotek často nejsme schopni měřit přímo, takže musíme měřit pouze indikátory (ukazatele) těchto vlastností.<sup>22</sup> Například politickou orientaci jedince z hlediska levice či pravice jsme schopni určit na základě toho, jakou volil stranu v parlamentních volbách. A to ještě pouze na základě jeho výpovědi.

Při měření jevů (konceptů, indikátorů) měříme různé aspekty:

**1. Intenzitu vlastností** zkoumaných jednotek výzkumu a také vlastností objektů vnějšího světa, který je obklopuje (kontextuální vlastnosti – podmínky). V zásadě ale zjišťujeme:

<sup>21</sup> Čtenáři by se velmi podivili, kdyby porovnali výsledky získané různým měřením chudoby a předpokládali, že všechna měří totéž – chudobu jakožto jakýsi esenciální stav. Podle definice Eurostatu (tzv. příjmová chudoba) bylo v roce 2001 v České republice asi 8 % chudých, měřeno ale dotazem na pocit chudoby (tzv. subjektivní chudoba) jich bylo 36 % (Sirovátka & Mareš, 2006). Podobně bychom ale dopadli při měření v jiných evropských zemích.

<sup>22</sup> Hledáme něco, co indikuje existenci nějaké vlastnosti, která sama o sobě není pozorovatelná. Čtenáři si to mohou sami zkusit, chtějí-li „změřit“ lásku svého partnera, respektive své partnerky. Říkájí-li si „miluje mě“, z čeho tak usuzují – z jakých jeho/jejich projevů?



- jaká je intenzita nějaké vlastnosti nějakého zkoumaného jevu (objektu) v určitém okamžiku;
  - k jaké změně intenzity vlastnosti došlo v nezměněných podmínkách v určitém čase;
  - k jaké změně intenzity určité vlastnosti došlo v podmínkách, které se v určitém čase známým způsobem změnily.
2. **Distance objektů** (vlastností), což je ve svém primárním významu geometrický pojem, konkretizovaný v teorii měření metrickou veličinou délky. Tato distance může být měřena i v určitém konstruovaném prostoru, jak je tomu v případě prostoru znaků postulovaném Lazarsfeldem (Barton, 1955), v sémantickém prostoru, či v prostoru vytvářeném technikou faktorové analýzy. Ve svém sekundárním významu jde o vztah mezi dvěma bezprostředně sousedními škálovými hodnotami, který je numericky reprezentován jako jejich rozdíl, vyjádřený v absolutních hodnotách.
  3. **Závislosti** (asymetrický vztah) či **souvislosti** (symetrický vztah) mezi dvěma vlastnostmi zkoumaných jednotek (to je subjektů či objektů výzkumu) či mezi dvěma jevy, respektive mezi proměnnými, jež je reprezentují.
  4. **Globální vlastnosti souborů** (například průměrný věk okresů ČR, průměrný příjem vzdělanostních skupin, porodnost v zemích EU apod.).

## 1.4.2 Proměnná

Kvantitativní sociálněvědní výzkum může nalézt řešení jen pro problémy, které je možno popsat v termínech vztahu mezi pozorovatelnými proměnnými. Proměnné představují logicky uspořádané charakteristiky/vlastnosti zkoumaných jednotek (hodnoty proměnných). Dovolují zkoumané jednotky podle jejich vlastností pouze zařadit do kategorií (nominální proměnné – např. pohlaví, umožňující jednotky zařadit mezi muže nebo ženy), nebo je zařadit do kategorií uspořádaných podle nějaké míry (ordinální proměnné – např. vzdělání, umožňující jednotky zařadit do kategorií seřazených podle stupňující se míry dané vlastnosti, mezi absolventy pouze základního vzdělání, absolventy středních škol a absolventy vysokých škol) nebo určit číselně intenzitu, jakou daná vlastnost nabývá (spojité neboli kardinální proměnné – např. věk, příjem apod.).<sup>23</sup>

### Ilustrace

- Muž a žena jsou vlastnosti (*attributes*) a současně hodnoty (*values*) proměnné (*variable*) nazývané „pohlaví“. Tyto dvě vlastnosti představují obor možných hodnot této proměnné (její varianty).
- Zaměstnaný a nezaměstnaný představují dvě vlastnosti a současně dvě hodnoty proměnné „postavení na trhu práce“.
- Vzdělání základní, středoškolské a vysokoškolské představují tři vlastnosti a současně tři hodnoty proměnné nazývané „nejvyšší dosažené vzdělání“. Jak proměnnou konstruujeme, závisí

<sup>23</sup> Někdy se namísto výrazu „proměnná“ používá výraz „znak“. Pojem znaku se často používá tehdy, jestliže je kladen důraz na to, že něco značí, zastupuje (nějakou vlastnost, stav apod.). Pojem proměnné používáme tehdy, když se naše úvahy koncentrují na to, jak se tento znak mění, jakých hodnot sledovaný objekt nabývá.

na našich výzkumných otázkách. Proměnnou „vzdělání“ si jistě dovedeme představit i s jiným počtem hodnot (variant). Například ve variantě základní vzdělání můžeme odlišit ty, kdo základní vzdělání nedokončili, od osob s dokončeným základním vzděláním; nebo v rámci vysokoškolského vzdělání můžeme rozlišit stupeň bakalářský a magisterský. V případě, že budeme chtít srovnávat vzdělání v různých zemích, musí být mezi nimi dosaženo shody o obsahu jednotlivých kategorií (jejich standardizace). I v rámci pouhé Evropy se totiž školské soustavy jednotlivých zemí vyvíjely různým způsobem, takže je nutná společná metodika, jak absolventy různých typů škol na škále nejvyššího dosaženého vzdělání umisťovat. K tomu slouží Mezinárodní standardní klasifikace vzdělávání neboli ISCED (International Standard Classification of Education). Vydalo ji UNESCO v roce 1976 a od té doby je pravidelně aktualizována. V současnosti se používá verze z roku 1997, která má 7 úrovní vzdělávání, v každé z nich je navíc ještě vnitřní členění A až C.<sup>24</sup> V České republice ji používá například Výběrové šetření pracovních sil, prováděné Českým statistickým úřadem, běžně se používá ve velkých komparativních výzkumech (například European Social Survey nebo European Values Study zahrnujícím většinu evropských zemí).

- Silný souhlas, souhlas, nesouhlas a silný nesouhlas s privatizací představují čtyři vlastnosti a současně hodnoty proměnné nazývané „postoj k privatizaci“ apod.

Složitě vlastnosti lidí jsou takto převáděny na relativně jednoduchý soubor informací. Člověk se tak jeví jako statistická jednotka (nositel statistické informace), jako nositel těchto proměnných, lépe řečeno jako nositel určitých vlastností uvnitř daných proměnných. V kvantitativním výzkumu je člověk:

- mužem či ženou (proměnná pohlaví);
- osobou s určitým počtem let (proměnná věk);
- příslušníkem vyšší, střední či nižší třídy (proměnná sociální třída);
- osobou s určitou výší příjmu (proměnná příjem);
- vlastníkem určitého statku (proměnná vlastnictví tohoto statku);
- osobou s určitým dosaženým vzděláním (proměnná vzdělání);
- příslušníkem nějaké profese (proměnná profese);
- residentem v určité komunitě (proměnná bydliště);
- potenciálním nebo i reálným voličem jedné z politických stran (proměnná volební preference);
- jedincem, který nějak tráví svůj volný čas (proměnná struktura volného času);
- osobou, která má určité postavení na trhu práce – má či nemá placené zaměstnání (proměnná ekonomická aktivita);
- nositelem určitého postoje či názoru (proměnnou je postoj: například nesouhlas se snahou vlády zvyšovat daně);
- sociálním aktérem (proměnná akce: například účast na antiglobalistické demonstraci, volební účast – účastní se/nezúčastní se voleb; apod.

<sup>24</sup> **ISCED** (International Standard Classification of Education) je nástrojem, který vyvinulo UNESCO pro porovnávání indikátorů a statistik vzdělávání ve svých členských zemích, který bere v úvahu rozdíly jejich vzdělávacích systémů. Více informací najdete v knize *Cesta k datům* (Soukup, 2012) nebo přímo na <http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>. Uživatelskou příručku pro zavádění ISCED 97 v zemích OECD vydal v roce 1999 Ústav pro informace ve vzdělávání, Praha.

### 1.4.3 Typy škál – proč jsou důležité

Při měření používáme měřicí stupnice (měřicí škály). Tím se rozumí obvykle číselná stupnice, předem daný jednoznačný systém, jehož hodnoty jednoznačně odpovídají úrovním měřené vlastnosti. Měřicí stupnice bývá určena svým měřítkem nebo jednotkou a způsobem uspořádání svých hodnot. Pod vlivem Stanleyho Stevense rozlišujeme různé úrovně měření, což závisí na tom, jaký je vztah mezi tím, co se měří, a číslem, které reprezentuje výsledek měření. Tyto úrovně (škály) měření jsou: nominální, ordinální a kardinální (a ty se ještě dělí na intervalové a podílové). Podle těchto úrovní potom rozeznáváme proměnné nominální, ordinální a kardinální.

**Nominální proměnná** je taková, jejíž hodnoty jsou kategorie označené číselnými kódy, které jim výzkumník arbitrárně přiřadil. Příkladem nominální proměnné jsou např. barva vlasů (1. černá, 2. kaštanová, 3. rezavá, 4. blond), okresy ČR, druhy zaměstnání. Jednotlivé kategorie nominální proměnné, byť mají číselná označení, není možné uspořádat do pořadí, neboť přiřazené číslo je pouhým symbolem a v žádném případě neoznačuje množství měřené vlastnosti – objekty zde můžeme pouze přiřazovat do číselných kategorií podle předem stanovených pravidel.

**Ordinální proměnná** je taková, jejíž kategorie lze uspořádat do pořadí. Takže zatímco u kategorií nominální proměnné pouze zjišťujeme, zdali se jednotlivé kategorie vyskytly, nebo ne, a pokud se vyskytly, pak jak často, o kategoriích ordinální proměnné jsme schopni říci, která je v pořadí výše než jiná. Příkladem ordinální proměnné je míra spokojenosti (stupnice může mít podobu: 1. velmi spokojen, 2. spokojen, 3. nespokojen, 4. velmi nespokojen), stupeň dosaženého vzdělání, výsledky v soutěži krásy. Pozor tedy: ordinální stupnice zobrazují pouze pořadí, a nikoliv stupeň odlišnosti – nedokážeme zde totiž určit, o jaké množství spokojenosti se liší „velmi spokojeni“ od „spokojených“, byť číselná řada by nezkušenému výzkumníkovi sugerovala, že jde o rozdíl/množství jednoho stupně.

**Kardinální proměnná** je taková, jejíž číselné kategorie již vyjadřují skutečné množství sledované vlastnosti (číselné kódy tedy nejsou arbitrární). Jednotlivé kategorie této proměnné jsme proto nejenom schopni seřadit do pořadí, ale umíme i říci, o kolik (o jaké množství) se liší. Kardinální proměnné rozdělujeme na **proměnné intervalové** (jejich stupnice nemá přirozenou smysluplnou nulu) a **proměnné poměrové** (s existencí přirozené nuly) – poměrové proměnné jsou snem každého sociologického kvantitativního výzkumníka. Je to tím, že u nich jsme schopni říci nejenom, o kolik se kategorie liší, ale také kolikrát je nějaká kategorie vyšší než jiná (je to dáno právě tím, že poměrové škály mají přirozenou nulu). Příkladem kardinální proměnné je věk, příjem respondenta, počet dětí, které žena porodila, apod. V jemnějším pohledu lze kardinální proměnné ještě členit na **diskrétní**, tedy takové, které mohou nabývat pouze určitých hodnot – většinou to jsou celá čísla, příkladem je počet dětí –, a na jejich opak, tedy proměnné **spojité** (*continuous*), které mohou, jak naznačuje název, nabývat jakýchkoliv hodnot: např. věk člověka může být měřen – podle potřeby přesnosti – nejenom na roky, ale i na dny, sekundy, popřípadě mikrosekundy a nanosekundy...

Zvláštním případem je proměnná, která nabývá pouze dvou kategorií. Označuje se jako proměnná **dichotomická** (nebo také binární). Příkladem je muž x žena, zaměstnaný x nezaměstnaný, živý x mrtvý apod. Chování dichotomických proměnných je trochu zvláštní. Z pohledu metodologie jde o nominální proměnnou, ale pokud užíváme zavedených kódovacích schémat (často 0 vs. 1 či 0 vs. 100), lze s ní ve statistice často operovat jako s proměnnou kardinální.

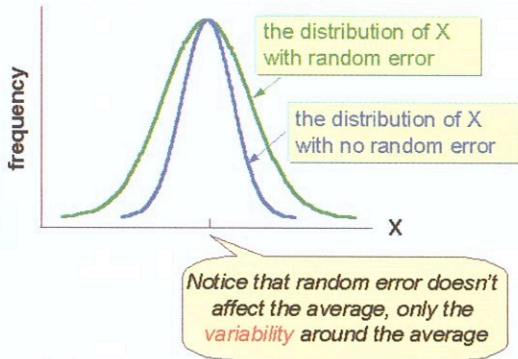
Rozlišovat úrovně měření a druhy proměnných je velmi důležité, neboť na druhu proměnných jsou závislé statistické operace, které můžeme s daty provádět. Ideálem pro statistiku jsou pochopitelně proměnné kardinální, neboť umožňují nejvyšší množství statistických operací. V sociálních vědách jich ovšem není příliš mnoho, velmi často se pohybuje pouze v oboru proměnných ordinálních a nominálních.

Typy proměnných je třeba doplnit ještě o jednu důležitou klasifikaci, a to o **proměnné nezávislé a závislé**. Základem pro toto dělení již není množství vlastností a způsob měření, ale úvaha o tom, co je „příčinou“ a co „následkem“. Když hledáme souvislosti mezi jevy, obvykle začínáme úvahou o vztahu dvou jevů, dvou proměnných. Přitom předpokládáme, že jeden jev je příčinou, druhý následkem. Proměnná reprezentující v našich úvahách (v naší hypotéze) příčinu je proměnná nezávislá (*independent variable*), proměnná reprezentující uvažovaný následek je proměnná závislá (*dependent variable*). Např. hledáme-li vztah mezi pocitem štěstí a věkem, můžeme formulovat předpoklad, že mladší lidé budou v průměru šťastnější než senioři, to je že s rostoucím věkem se bude snižovat průměrný pocit štěstí. Rozlišit nezávisle a závisle proměnnou není obvykle problém; pokud si nejsme jisti, vezměme v úvahu časový průběh: v našem případě pocit štěstí prostě nemůže být nezávisle proměnná, pocit štěstí nemůže být příčinou určitého věku.<sup>25</sup> V analýze dat se bez rozlišování nezávisle a závisle proměnných neobejdeme, Hendl dokonce říká, že „výzkum začíná určením nezávisle a závisle proměnných...“ (Hendl, 2004, s. 40).

Při měření se pochopitelně můžeme dopustit řady chyb. Ostatně **základní teorie měření** (anglicky *true score theory*) s chybou počítá, neboť říká, že každé měření sestává ze dvou (aditivních) složek: ze skutečné správné hodnoty a z chyb měření. Vyjádřeno jednoduchým vzorcem:

$$X = T + e,$$

<sup>25</sup> Někteří autoři namítají, že tato klasifikace je patřičná pouze pro experimentální výzkum, v němž jsme schopni manipulovat s příčinami: experimentátor vystavuje subjekty experimentu různým podnětům – nezávisle proměnným – a zkoumá jejich reakce (závisle proměnné). Pro analýzu neexperimentálních dat, např. pro data ze sociologických výzkumů založených na dotazníku, navrhuji nazývat nezávisle proměnné **prediktorem** (*predictor variable*), závisle proměnné pak **výsledkem** (*outcome variable*). Česká sociologická praxe je taková, že používáme termíny nezávisle a závisle proměnná.



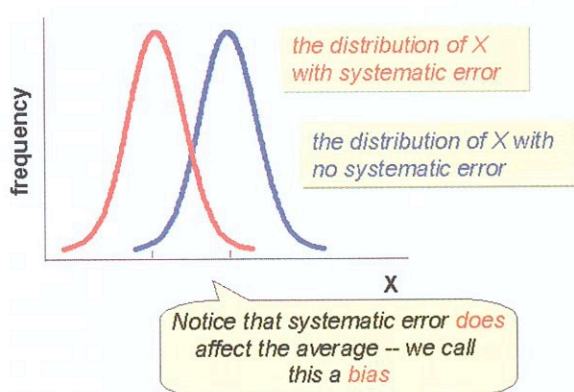
Obr. 1.3 Graf rozložení dat při náhodné chybě měření a při měření bez chyby

Zdroj: True Score Theory, dostupné z: <http://www.socialresearchmethods.net/kb/truescor.php>

kde  $X$  je naměřená hodnota vlastnosti,  $T$  je její skutečná, správná (*true*) hodnota, kterou ale neznáme (snažíme se ji totiž zjistit právě našim měřením), a  $e$  je celková chyba měření (*error*). Celková chyba měření může být systematická nebo náhodná. **Náhodná chyba** (*random error*) většinou nezpůsobuje větší problém, neboť odchylky od měření sice způsobují větší variabilitu v datech, ale jelikož mají tendenci se náhodně vychylovat v kladném i záporném směru (někdy naměříme více, než je skutečná vlastnost, jindy zase méně), vzájemně se vyruší, takže průměr měřené vlastnosti zůstává nedotčen. Proto se někdy o náhodné chybě hovoří jako o šumu (*noise*). Hezky to vyjadřuje obrázek 1.3, v němž modrá křivka ukazuje, jaké bychom měli hodnoty bez náhodné chyby měření, zatímco zelená křivka zobrazuje hodnoty s náhodnou chybou. Vidíme, že střední hodnota (průměr) je u obou křivek shodná, rozdíl mezi křivkami je pouze v šířce rozložení hodnot – zelená (s chybou) má širší rozložení, hodnoty zde mají větší rozptyl (variabilitu), modrá je sevřenější, rozptyl je menší.

**Systematická chyba** (*systematic error*) – viz obr. 1.4 – představuje trvalou, nikoliv náhodnou chybu, a pokud se při měření vyskytne, je to skutečný problém. Jelikož jde o systematické zkreslení, jsou hodnoty při měření opakovaně vychýleny buď kladně, nebo záporně, takže se navzájem nemohou vyrušit. Výsledné měření je pak systematicky vychýlené (na obr. 1.4 je to ta červená křivka), a tudíž zkreslující – průměr hodnot se systematickou chybou (u červené křivky) je odlišný od průměru hodnot, které chybu neobsahují (modrá křivka). Systematickou chybu nazýváme **zkreslením** (*bias*).

Zdroj systematické chyby je různý. Může být na straně pozorovatele, pokud je jeho pozorování selektivní (viz schéma měření na obr. 1.2), ale chyba může být způsobena i měřicím nástrojem (pokud je například nevalidní). Jejím zdrojem může být ale i měřený objekt, například když u výběrových souborů jde o chybný výběr jejich jednotek. Zobecnujeme-li z výběrového souboru na cílovou populaci, může jít o chybu spočívající v neoprávněném zobecnění – hovoříme zde o chybě neoprávněného rozšíření domény



Obr. 1.4 Graf rozložení dat při systematické chybě měření a při měření bez chyby

Zdroj: True Score Theory, dostupné z: <http://www.socialresearchmethods.net/kb/truescor.php>

měření, kdy zobecňujeme své výsledky na širší soubor, než z něhož jsme provedli náš výběr. Může jít například o snahu zobecnit volební preference z výběrového souboru vybraného z pražské populace na celou populaci ČR.

Obecně ve statistice **chyba (error) znamená rozdíl mezi pozorováním a predikcí** (či odhadem) a je spíše nepřesností než skutečnou chybou. Jde většinou o náhodnou chybu. V tomto ohledu se v kapitole o inferenční statistice (viz kapitolu 5) seznámíme se **standardní chybou (standard error)** a **výběrovou chybou (sampling error)**. Zde jde o to, jak blízko (s určitou zvolenou pravděpodobností) jsou výsledky získané výzkumem ve výběrovém souboru příslušným charakteristikám populace.<sup>26</sup>

#### 1.4.4 Aspekty měření

Při měření si musíme všimnout jeho několika aspektů. Zásadně platí, že použitý způsob měření by měl být relevantní, validní, reliabilní, senzitivní a přesný. To jsou hlavní aspekty měření (viz též Disman, 1993, s. 62 nebo Babbie, 2001, s. 140–145). **Relevance** znamená vhodnost použité procedury měření ve vztahu k problému. **Validita** se týká rozsahu, v němž měření korespondují se skutečnou vlastností, která má být měřena (měříme skutečně to, co měřit chceme?).

Základní otázkou, kterou si můžeme u každého měřicího nástroje položit, je: „Co měří?“ Validitou se zhruba chápe platnost měřících procedur, získaných údajů, měřících nástrojů, prostě všech složek měření a škálování. **Měření je validní, jestliže měří to, co myslíme (očekáváme), že měří.** Ve skutečnosti to ovšem není měření, které je validní nebo nevalidní, ale jeho použití. Validita tak závisí na tom, jak je

<sup>26</sup> V jazyce analýzy označujeme výsledky získané výpočty (ve výběrovém souboru) výrazem „statistiky“ a číselným charakteristikám populace říkáme „parametry“.

definován měřený pojem.<sup>27</sup> Babbie (2001) konstatuje, že konvenční používání termínu validita se vztahuje k rozsahu, ve kterém empirická míra adekvátně reflektuje skutečný význam uvažovaného pojmu.

### Ilustrace

Použijeme-li proměnnou „úroveň dosaženého vzdělání“ k měření sociálního statusu, problémem nebude, zda měříme s úspěchem tuto úroveň vzdělání, ale zda tímto způsobem (prostřednictvím tohoto ukazatele) skutečně měříme to, co měřit chceme: tedy sociální status.

**Reliabilita** označuje rozsah, ve kterém způsob měření dává konzistentní výsledky. U reliability se tedy ptáme, do jaké míry jsou výsledky opakovaného měření shodné s původním měřením. Ilustrativní pro reliabilitu je Segalův zákon, který praví: Když má člověk jen jedny hodinky, ví vždycky, kolik je hodin. Když má člověk dvoje hodinky, nemůže si být nikdy jist.

Reliabilitou se zhruba chápe spolehlivost výsledků měření v závislosti na a) objektivní spolehlivosti měřicích procedur, technik či nástrojů, b) subjektivní spolehlivosti respondentů a experimentátora.

Reliabilita je také považována za míru stability měřicích nástrojů, s jakou lze při opakovaných měřeních či testech získávat přibližně stejné výsledky. Jde o záruku, že proměnlivé výsledky měření nejsou způsobeny špatným měřicím prostředkem, ale skutečnou variabilitou měřené vlastnosti. Jestliže lidé odpovídají na otázky při opakovaném dotazování stejným způsobem, pak jsou otázky reliabilní. Synonyma reliability jsou: spolehlivost, stabilita, konzistence, prediktabilita neboli předpověditelnost, přesnost (Kerlinger, 1972, s. 421).<sup>28</sup> Zdrojem nereliability mohou být špatné formulace otázek, rozdílné kulturní významy vkládané do použitých termínů v různých sociálně kulturních prostředích, vliv tazatelů (a to nejen záměrný). Pohlaví tazatele, jeho etnický původ, oblečení, sociální zařazení, to vše může při dotazování (to je sběru dat) hrát značnou roli. Jinak odpovídá žena ženě, jinak muži a podobně, a v tomto odlišném odpovídání se skrývá zdroj chyb.

Jiným zdrojem chyb může být kódování, neboť různí tazatelé mohou kódovat stejné odpovědi rozdílně, uplatňuje se selektivní slyšení a podobně. Reliabilita je nízká i u otázek, na které lidé nemají názor (mínění, postoj), a otázka tento názor uměle vytváří (může, ale spíše nemusí zůstat stabilní). Reliabilitu se proto snažíme zvyšovat

<sup>27</sup> Měříme-li příjmovou chudobu definovanou například poměrem příjmu jednotky s příjmovým rozložením (například za chudé považujeme všechny jednotky s příjmem nižším než 60 % příjmového mediánu), bude dotaz typu „cítíte se být chudou rodinou“ nepochybně velmi zajímavý, ale neměří její (definovanou) chudobu – je z tohoto hlediska nevalidní. I když ve vztahu k subjektivní chudobě jistě validní je.

<sup>28</sup> Koncept reliability je pro nás důležitý obecně, ale i specificky – například v případě sumačních indexů (viz kapitolu 6 o transformaci proměnných), kde je nutno změřit reliabilitu jednotlivých položek (jejich vnitřní konzistenci), z nichž je sumační index vytvořen.

použitím více indikátorů (u jednoduchých položek je vždy obtížnější zajistit reliabilitu než u použití více indikátorů), pečlivou formulací otázek, ale i určitým výcvikem (instruktážemi) tazatelů a standardizací způsobu kódování.

**Senzitivitou** se rozumí schopnost testu dávat pozitivní odpověď, jestliže daná osoba má příslušnou vlastnost, **specificitou** pak schopnost testu dávat negativní odpověď, jestliže daná osoba nemá příslušnou vlastnost. Toto rozlišení zná velmi dobře například medicína a medicínský výzkum.

## 1.5 Hypotézy a modely

### 1.5.1 Od tématu přes problém k výzkumné hypotéze

Hypotéza je určité očekávání o povaze věci odvozené většinou z teorie (je to tvrzení o tom, jaká tato povaha má být, má-li být teorie, z které je odvozena, pravdivá). **Výzkumná hypotéza** je předběžný předpoklad, domněnka o 1) existenci a 2) příčinách jevů, o 3) vztahu mezi jevy, o 4) průběhu nějakého procesu, o 5) změně apod. Svou povahou leží na jakémisi myšlenkovém rozhraní mezi teoretickou a empirickou fází výzkumu. Je návodem k výzkumu, který je na jedné straně determinován dosavadním poznáním, na druhé straně je orientován na další poznání. Může být odvozena z kontextu vědy (formulována z teorie nebo z jiných hypotéz), ale i ze zkušenosti. Má charakter výroku, tvrzení, modelu, či dokonce teorie, ale jsou to výroky, tvrzení, modely a teorie, jež nebyly dosud přijaty jako obecně platné. Její podstatnou charakteristikou je, že ji můžeme různým způsobem empiricky ověřovat. Podle etologa Konráda Lorenze (1903–1989) si nejdříve něco myslíme, pak to srovnáváme se zkušeností a s dalšími smyslovými daty, až nakonec – podle toho, shoduje-li se to s nimi, nebo ne – rozhodneme o správnosti či chybnosti toho, co jsme vymysleli. Hypotéza je něco, co dosud nebylo ověřeno, a má proto v rámci vědy dočasný status (Lorenz, 1990, s. 72–77).<sup>29</sup>

Velmi často se setkáváme s nepochopením při rozlišování mezi tématem výzkumu a výzkumnou otázkou, respektive výzkumnou hypotézou. Rozlišuje se:

- **téma**, respektive předmět výzkumu. Jestliže řekneme, že chceme zkoumat sociální nerovnost, neboli vidíme v nerovnosti problém (sociální či výzkumný) a chceme výzkumem získat odpovědi na některé otázky týkající se „sociální nerovnosti“, vymezili jsme si prozatím jen předmět výzkumu;
- **výzkumný problém**, na který hledáme prostřednictvím výzkumu odpověď a který je zúžením tématu na některý z jeho (pro nás) významných aspektů;
- **výzkumná otázka/hypotéza**, která problém dále specifikuje do takové podoby, abychom z odpovědi na ni tento problém mohli pochopit – teprve výzkumná otázka určuje otázky, s nimiž se obrácíme ke zkoumaným jednotkám: například otázky dotazníku.

<sup>29</sup> Připomeňme si Popperovo stanovisko, že všechna vědecká tvrzení jsou vlastně jen hypotézy, jež mají dočasný status (Popper, 1974, s. 87–88).



### Ilustrace

Téma (předmět) výzkumu je sociální nerovnost.

Jaké odpovědi a na jaké otázky chceme získat (co je výzkumný problém)?

- Jaký typ nerovnosti budeme zkoumat (ekonomickou, mocenskou, politickou)? Jak ji budeme měřit (rozdíly ve statusech, v příjmech)?
- Jaký výzkumný prostor zvolíme? Celou společnost, nebo pouze její určitý výsek (organizaci, komunitu, sociální skupinu, rodinu)?
- Jaké období bude časovým rámcem výzkumu? Soustředíme se na jednu generaci, nebo budeme zkoumat i mezigenerační vztahy a procesy (například přenos nerovností z generace na generaci)?
- Jaký aspekt nerovnosti budeme studovat? Zajímá nás rozsah, jakého nerovnost ve společnosti nabývá, její rozložení ve společnosti, její příčiny, její důsledky, funkce ve společnosti nebo ještě něco jiného (například extrémní forma nerovnosti)?
- Bude nás zajímat, které instituce, ideologie, hodnoty a normy nerovnost podporují a jaké mechanismy ji zajišťují, mechanismy, jimiž se nerovnost ve společnosti reprodukuje?
- Jde nám o to, jaké zájmy stojí za udržováním nerovnosti či jak je ideologizována, respektive legitimizována, jak ji lidé hodnotí či jak vysvětlují její příčiny?
- Budeme zkoumat prožitek nerovnosti a deprivaci, anomii či marginalizaci, které ji doprovázejí?

Všechny tyto otázky ovlivňují povahu hypotéz, které v návaznosti na volbu výzkumného problému formulujeme. Například si zvolíme jako prostor rodinu (nerovnosti v rodině) a budeme se zajímat o genderový aspekt nerovnosti, konkrétně v dispozici rodinnými zdroji. Můžeme formulovat hypotézu: *Pohlaví bude mít vliv na osobní spotřebu finančních zdrojů rodiny*. Tu bychom mohli ještě zpřesnit (zaměřit) tím způsobem, že budeme předpovídat, že: *Muži pro svou osobní spotřebu využívají více finančních zdrojů rodiny než ženy*. (Vyšší výdaje na vlastní spotřebu indikují větší moc nad finančními prostředky rodiny a tím i nerovnost v rodině.)

## 1.5.2 Typy hypotéz

Hypotézy lze klasifikovat podle různých hledisek. Jaké druhy tedy rozlišujeme?

Hypotézy **teoretické** a **empirické**. Teoretické hypotézy jsou výroky formulované jazykem teorie. Empirické hypotézy jsou výroky, které jsou empiricky testovatelné.

Hypotézy **výchozí** a **pracovní**. Výchozí hypotézy bývají obvykle hypotézami teoretickými nebo empirickými na vyšší úrovni obecnosti a hypotézami komplexními. Statistická analýza slouží k ověřování pracovních hypotéz, které Disman definuje následujícím způsobem: 1) Pracovní hypotéza je tvrzení předpovídající souvislosti mezi dvěma nebo více proměnnými; 2) všechny proměnné zmíněné v hypotéze musejí mít validní operační definici; 3) soubor pracovních hypotéz musí zahrnovat nejen proměnné reprezentující zkoumané koncepty, ale i ty proměnné, které mohou významně zkreslit interpretaci testovaných vztahů (Disman, 1993, s. 79).

Hypotézy **kauzální** a **vztahové**.<sup>30</sup> Hypotézy mohou být hypotézami o stavu či struktuře (zde je užitečná jednorozměrná analýza, neboť technicky jde o otázky týkající se hodnot

<sup>30</sup> Blíže k otázkám kauzality v sociálním výzkumu Babbie (2001, s. 68–87) nebo také Field (2009, s. 13–14).

proměnných) – někdy o nich lze hovořit jako o hypotézách popisných. Většinou však ve statistické analýze testujeme hypotézy o vztahu dvou proměnných (dvojezměrná analýza) nebo několika proměnných (vícezměrná, multivariační analýza), které mohou být označeny také jako hypotézy vysvětlující (explanační). Zajímá nás, zda mezi určitými sledovanými proměnnými existují nějaké vztahy a jakou povahu tyto vztahy mají. Primárně nás zajímá, zda jde o vztah statistický, nebo kauzální. Považujeme-li vztah za kauzální, uvažujeme o kauzální hypotéze. Babbie (2001, s. 86) uvádí tři základní kritéria kauzality: 1) proměnné musejí být empiricky asociovány nebo korelovány, musejí existovat jejich souběžné změny; 2) kauzální proměnná, rozumějme příčina, musí v čase předcházet proměnné, kterou ovlivňuje, rozumějme důsledek; 3) pozorovaný důsledek nemůže být vysvětlen působením jiných proměnných. Pouhá (ani perfektní) asociace nebo korelace nepředstavuje kauzalitu. Popravdě řečeno, empirická sociologie ve většině případů pracuje se statistickými vztahy (které nesplňují třetí ze zmíněných kritérií), a její závěry proto mají vždy pravděpodobnostní charakter: např. čím má člověk vyšší příjem, tím je pravděpodobnější, že bude volit pravicové strany (proměnná  $X$  má vliv na proměnnou  $Y$ ).

Hypotézy **věcné a statistické**. Z hlediska statistické analýzy dat má toto rozlišení zásadní význam. Věcnou hypotézou je domněnka o existenci vztahu mezi dvěma nebo více proměnnými. Statistická hypotéza je ve statistických termínech vyjádřené hypotetické tvrzení o relacích vyvozených ze vztahu ve věcné hypotéze. Věcné hypotézy se mohou týkat: 1) existence, výskytu a stavů předmětů, jevů, událostí, lidí, skupin (existují či proběhly, jak byly početné?); 2) vlastností předmětů, jevů, událostí, lidí, skupin (jak byly početné jejich určité kategorie, jakou měly intenzitu?); 3) vztahů mezi předměty, jevy, událostmi, lidmi, skupinami, respektive zda mezi jejich vlastnostmi existovaly, jaké povahy byly (příčinné, statistické), kdy nastaly atd.; 4) vývoje předmětů, jevů, událostí, lidí, skupin, jeho etap a stadií a jejich charakteristik (v jakých etapách se vyvíjely, co bylo pro tyto etapy charakteristické?); 5) procesů, jichž se lidé či skupiny účastní nebo jež v nich probíhají; respektive se mohou týkat vztahů mezi těmito skutečnostmi (existencí, výskytem, stavy, vlastnostmi, procesy apod.).

Věcnou hypotézu není možno podle Kerlingera (1972) samu o sobě testovat, proto testujeme tzv. statistickou hypotézu (představuje tedy způsob testování věcné hypotézy) – což je matematický model sociální reality o chování proměnných, který se snažíme zamítnout (viz pasáž o nulové hypotéze níže).<sup>31</sup> Jde de facto o formulaci věcné hypotézy ve statistických termínech. Protože statistická teorie dovoluje testovat jen jednoduché hypotézy, rozkládá se často komplexní sociologická (věcná) hypotéza na řadu dílčích statistických hypotéz. To, že má hypotéza statistický charakter, znamená podle Nowaka (1975), že jednoznačně nehovoří o vlastnostech nebo chování každého člověka, nýbrž vypovídá o relativních četnostech, statistických závislostech

<sup>31</sup> Pozor! Mnohdy se statistické hypotézy neodlišují od věcných a testování statistických hypotéz je neoprávněně chápáno jako testování hypotéz věcných.

mezi vlastnostmi a chováním lidí. Tyto závislosti pak platí buď pro určitá seskupení lidí, nebo jsou formulovány jako univerzální teze.

Hypotéza v empirickém výzkumu je výrokem o vztahu alespoň dvou jevů a v konečné fázi (v kvantitativním výzkumu) výrokem o vztahu (alespoň) dvou nebo více proměnných. Čistě teoreticky může být hypotézou i výrok: „staří lidé jsou konzervativní“, protože obsahuje implicitně dvě proměnné – věk a politickou ideologii. Lépe je ovšem formulovat hypotézu „věk je pozitivně vztažen ke konzervativismu“ nebo „s růstem věku roste příklon osob ke konzervativismu“ či „s růstem věku roste počet osob volících strany s konzervativním programem“ nebo také „čím jsou lidé starší, tím vyšší dosahují skóre na škále konzervativismu“ (všimněte si, že tyto hypotézy nevyjadřují vždy totéž).

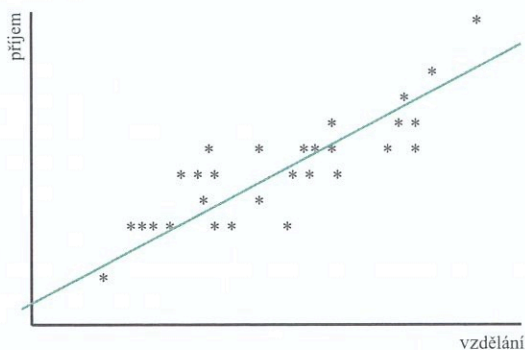
V posledních dvou případech máme již pracovní hypotézu, neboť obě proměnné jsou operacionalizovány (věk vyjadřujeme v letech a na základě definice konzervativismu jsme schopni určit, které strany můžeme označit jako konzervativní, popřípadě máme k dispozici konstruovanou škálu konzervativismu). Ve zmíněné podobě jde o hypotézu vztahovou a pravděpodobnostní. Netvrdíme, že mezi oběma proměnnými existuje kauzální vztah a že všichni staří lidé jsou konzervativní. Jen předpokládáme, že pravděpodobnost výskytu konzervativismu (podíl osob zastávající konzervativní postoje) mezi starými lidmi je vyšší než mezi lidmi mladými. Je to hypotéza věčná, testovat bychom ji mohli jako statistickou (nulovou) hypotézu „neexistuje vztah mezi věkem jedince a skórem, kterého tento jedinec dosahuje na škále konzervativismu“ nebo „koeficient korelace vztahu proměnné věk a proměnné hodnota dosažená jedincem na škále měřící míru konzervativismu je roven nule“.<sup>32</sup>

### 1.5.3 Složitější modely

Modely jsou zvláštním způsobem výkladu reality, který reprodukuje aspekty jevu a umožňuje deduktivní odvození a výroky, jež mohou být přezkoušeny ve zkušenosti. Nejjednoduššími modely jsou **abstraktní popisy systémů** jako modely jejich jistých aspektů. Někdy jsou modely jen synonymem pro teorie – obvykle velmi jednoduché – nebo pro část teorie. Identifikujeme i tzv. **konceptuální modely**. Ty jsou pokusem prezentovat sociální svět v pojmech řady vzájemně vztažených pojmů (blahobyt, moc, prestiž jako odměny za výkon určité role). **Teoretické modely** představují zabudování určité teorie, která vysvětluje výběr a uspořádání prvků modelu (např. teorie racionální volby umožňuje uspořádat jednotlivé kroky jednajícího aktéra). Mohou mít i charakter metaforických analogií. Modely mohou mít verbální prezentaci či matematickou prezentaci, mohou být také prezentovány jako diagramy.

<sup>32</sup> Měli bychom být ale opatrní, protože jak si ukážeme dále, nulová hodnota korelačního koeficientu znamená pouze neexistenci asociace u vztahů lineárních (mohlo by se stát, že vztah existuje, ale není lineární).

Model na obr. 1.5 je matematické vyjádření vztahu mezi výší příjmu a výší vzdělání. Modelem tohoto vztahu je přímka, která říká, že vztah mezi oběma proměnnými je lineární. Model navíc vyjadřuje, že čím vyšší je vzdělání, tím vyšší je také příjem.<sup>33</sup> Pokud bychom tento vztah vyjádřili matematicky ve formě rovnice, mohli bychom přesně predikovat, o kolik jednotek vzroste příjem (závisle proměnná), zvýší-li se výše vzdělání o jednotku.<sup>34</sup>



Obr. 1.5 Vztah mezi vzděláním a příjmem (smyšlená data)

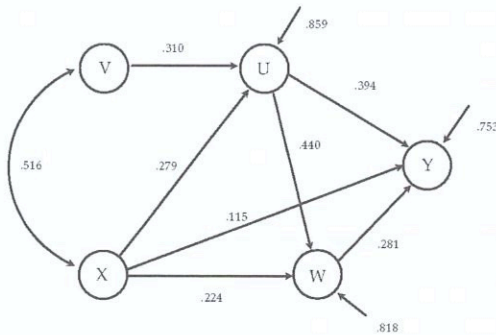
Na dalším obrázku (viz obr. 1.6) je uveden příklad složitějšího modelu vyjadřujícího vztah mezi jednotlivými proměnnými, které byly předmětem zkoumání (čísla zde představují sílu vztahu mezi jednotlivými prvky, vyjádřenou specifickými koeficienty asociace). V tomto případě jde o model dosahování sociálního statusu respondenta/respondentky v závislosti na vzdělání a zaměstnaneckém statusu otce.

Sociální systémy, které jsou předmětem zkoumání v sociologii, jsou ovšem systémy a) komplexními a b) otevřenými. Jejich **komplexnost** má za následek, že když je popíšeme jednoduchým modelem, dochází ke značnému zjednodušení, k simplifikaci. Ta je ovšem nezbytná (čím složitější modely konstruujeme, tím více se vymykají možnosti testovat je) a většinou postačující, protože se nesnažíme o úplné vysvětlení jevů, ale spíše o odkrytí míry vlivu určitých konkrétních stimulů na ně.

<sup>33</sup> Takový vztah je nejen lineární, ale i přímo úměrný, neboť s růstem hodnoty jedné proměnné roste současně i hodnota druhé proměnné. U nepřímo úměrného vztahu s růstem hodnot jedné proměnné hodnoty druhé proměnné klesají: například bylo empiricky ověřeno, že s růstem příjmu klesá míra anomie měřená na Sroleho škále (viz Rabušic & Mareš, 1996) – tento vztah je nejen nepřímo úměrný, ale také asymetrický, což znamená, že jedna proměnná je jasně nezávislá a druhá závislá. Hypoteticky bychom sice mohli předpokládat, že anomie jako nezávisle proměnná bude ovlivňovat závisle proměnnou, to je výši příjmu, což bychom spekulativně formulovali tak, že anomičtí lidé jsou natolik ve svých životech frustrováni, že nepodávají v práci řádný výkon, a tím jsou jejich příjmy nižší, ale logičtější je očekávat příčinu a důsledek v opačném gardu: příjem ovlivňuje míru anomie. K otázce aspektů vztahů mezi proměnnými se ještě dostaneme.

<sup>34</sup> V tomto konkrétním případě bychom ale museli vzdělání vyjádřit ve formě kardinální proměnné, například počtem let strávených ve škole (studiem), a nikoliv, jak je obvyklé, stupněm dosaženého vzdělání (základní, středoškolské, vysokoškolské), což je proměnná ordinální. S ordinálními proměnnými totiž, striktně řečeno, nemá regresní analýza pracovat.

**Otevřenost** sociálních systémů znamená, že žádný z námi konstruovaných modelů nevystačí jen s proměnnými, které jsou do něho zahrnuty. Vždy musíme počítat s vlivem proměnných, které v našem modelu nejsou obsaženy, neboť nebyly např. dostupné našemu měření nebo o jejich vlivu, či dokonce o jejich existenci nemáme ani ponětí. Všimněme si, že Blau a Duncan v demonstrovaném modelu s těmito vlivy počítají (viz šipky směřující zvenčí do bodů U, Y a W v obr. 1.6).



Obr. 1.6 Model proměnných ovlivňujících dosažení sociálního statusu

V = vzdělání otce

X = zaměstnanecký status otce

U = dosažené vzdělání

W = zaměstnanecký status v prvním povolání

Y = zaměstnanecký status v době šetření

Podle: Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley. Převzato z Matějů, P. (2005). Ke kořenům sociálně psychologického modelu sociální stratifikace. *Sociologický časopis*, 41(1), 7–30.

I když se svými modely pracujeme jako s uzavřenými systémy, statistika nám v některých svých procedurách umožňuje alespoň odhadnout, jaký podíl variance závisle proměnné/proměnných z našeho modelu je vysvětlen nezávislými proměnnými zahrnutými do našeho modelu a jaký je podíl vlivu proměnných do modelu nezahrnutých.<sup>35</sup> Podíl vysvětlené variance závisle proměnné vlivem námi použitých nezávislých proměnných může být někdy i překvapivě vysoký, většinou jsme však nadměru spokojeni i s 50–60 %. Zbylé procento variance v závisle proměnné (dopočítáváme do 100 %) jde pak na vrub proměnných, s kterými v modelu nepočítáme, a to většinou proto, že ani nevíme, které by to mohly být. Můžeme sice experimentovat a propočítávat modely s různými proměnnými (měnit jejich počet) i různými předpoklady vztahů mezi nimi, jsme však přitom omezeni jen na proměnné, které máme ve svých datech.

### Ilustrace

Zjistili jsme, že výše příjmu silně koreluje s věkem, ale koeficient determinace nám říká, že věk jako nezávisle proměnná vysvětluje jen 36 % variance v příjmu, zbylých 64 % jde na vrub vlivu jiných proměnných (tušíme, že ve hře je přinejmenším i vzdělání).

<sup>35</sup> U procedur, se kterými se zde seznamujeme, se s tím setkáme například u lineární regrese (koeficient determinace) nebo u faktorové analýzy.

## 1.6 Jak získat data pro analýzu

Hromadná data pro analýzu lze získat mnoha způsoby. Primární podmínkou je, aby byla získávána jako data standardizovaná. Například používáme-li pro jejich sběr dotazník, musíme zajistit vedle takových podmínek, aby všem byly kladeny stejné otázky ve stejném znění a stejném pořadí, také podmínky plynoucí ze statistických požadavků na proměnné:

- Proměnná musí variovat, musí tedy nabývat alespoň dvou hodnot (diskriminabilita čili **rozlišitelnost** mezi vlastnostmi objektu uvnitř proměnné: například u proměnné pohlaví lze rozlišit mezi muži a ženami).
- Ke každému stavu vlastnosti existuje příslušná hodnota znaku (**zařaditelnost**). Všechny pozorovatelné vlastnosti objektu musejí být zařaditelné do některé z hodnot proměnné (např. proměnná „volební preference“ má mít tolik hodnot, kolik politických stran postavilo své kandidátky do parlamentních voleb).
- Dvě různé hodnoty znaku nemohou odpovídat jednomu stavu vlastnosti (**jednoznačnost**). U žádné z pozorovatelných vlastností objektu nemůžeme být na rozpacích, jakou hodnotu proměnné jí přiřadit neboli do které kategorie ji zařadit. Například proměnná věk nemůže mít vedle hodnoty 20–30 let hodnotu 30–40 let, neboť by nebylo možné jednoznačně rozhodnout, do které kategorie by měly být zařazeny třicetileté osoby.
- Naše data by měla být reprezentativní (viz předchozí text), tak aby nám umožnila zobecnění výsledků našich výpočtů z výběrového souboru na soubor základní (statistickou inferencí – viz dále) a do výzkumu musí být zahrnut dostatečný počet výzkumných jednotek.<sup>36</sup>

Poslední podmínku je ovšem třeba mírně korigovat: ne vždy musejí být naše data reprezentativní. Máme-li například malou populaci (studenty jednoho gymnázia), uděláme vyčerpávající šetření (census), při kterém do výzkumu jednoduše zahrneme všechny její členy. V takovém případě ovšem ztrácí smysl tzv. inferenční statistika (viz kapitolu 5), kterou lze aplikovat pouze v souborech, jehož jednotky byly vybrány náhodně.

Zdrojem hromadných dat pro statistickou analýzu může být především vlastní sběr dat, dále data posbíraná jinými výzkumníky nebo statistické výkaznictví a speciální šetření, jako je například sčítání lidu nebo mikrocensus.

<sup>36</sup> Mlhavý výraz „dostatečný“ si blíže specifikujeme, až se dostaneme k inferenční statistice a k otázce výběrové chyby. Co tímto výrazem míníme, závisí na velikosti výběrové chyby, kterou jsme ochotni připustit, a také na hloubce zamýšlené analýzy (do kolika podsouborů budeme soubor, s nímž pracujeme, členit). Může nás uspokojit třeba i 300–400 výzkumných jednotek, ale nemusí nám stačit ani 5 000 výzkumných jednotek. Vše závisí na řadě okolností.

### 1.6.1 Sekundární analýza dat

Zdrojem dat pro nás mohou být i datové archivy (blíže o nich píšeme v dodatku IV). V nich jsou umístěny soubory, jež lze získat buď za úplatu, nebo se pro jejich získání požaduje jen registrace, popřípadě jsou volně k dispozici bez dalších podmínek (včetně dotazníků – bez nich jsou data často jen obtížně luštitelná nebo i zcela nečitelná a jejich analýza může pro neznalost přesného znění otázek a hodnot proměnných<sup>37</sup> být až zavádějící). Většina datových archivů je spravována různými výzkumnými pracovišti (viz například GESIS – Archive for Empirical Social Research – University Köln am Rhein apod.), nadacemi (viz European Social Survey apod.), mezinárodními organizacemi (UNESCO či Světová banka apod.) nebo státní správou (v České republice jde zejména o Český statistický úřad, v EU o Eurostat; může jít o dohodu sdružení států jako v případě The European Community Household Panel). Většina datových archivů je také na webu a data si lze přímo stáhnout. Nejvýznamnějším datovým archivem kvantitativních dat v České republice je datový archiv Sociologického ústavu Akademie věd České republiky (viz <http://archiv.soc.cas.cz/>). Doporučujeme webové stránky všech zmíněných institucí navštívit a s jejich archivy se seznámit. Více o těchto službách najdete v dodatku této učebnice.

**Naléhavá připomínka: Mějme zdravou skepsi k datům, s nimiž pracujeme!**

Pro vědeckou práci platí základní dictum: „buďme zdravě skeptičtí“. Každý výzkumník je pochopitelně přesvědčen, že jeho nástroje jsou validní a reliabilní a že jeho data jsou určitě reprezentativní. Do jaké míry je však třeba uchovávat si v tomto ohledu určitou skepsi, můžeme ilustrovat na několika příkladech tzv. rutinálních statistik. Nejsou primárně shromažďovány pro výzkumné účely, ale mohou být pro ně využity stejně jako například jiné údaje shromažďované v systému statistického výkaznictví. Příklady databází vznikajících takovým způsobem můžete nalézt na webových stránkách Českého statistického úřadu, Ministerstva práce a sociálních věcí (například data o struktuře nezaměstnanosti) i jinde.

Při jejich studiu si ale musíme klást otázky. Například můžeme věřit policejním nebo zdravotnickým statistikám? Policie se může soustředit na určité typy trestných činů a to se projeví v policejních statistikách. Nikoliv tedy změna ve struktuře, povaze a množství deviantního jednání, ale změna v přístupu této sociální instituce může mít za následek pohyb ve statistikách o trestných činech.

#### Ilustrace

- Poklesne-li velká kriminalita, má policie více času na malou kriminalitu a vzroste její objasněnost. Výskyt nemocí ve statistikách může být ovlivněn schopností diagnostikovat je. Pokud nebyla nemoc identifikována (nebyly popsány její příznaky, indikátory), respektive pojmenována, nemůže

<sup>37</sup> V datových souborech jsou proměnné a jejich hodnoty opatřeny jen velmi hrubými a zkratkovitými orientačními popisky (*variable label* a *value labels*).

být diagnostikována. Podobně též policejní statistiky zaznamenávají jen odhalené trestné činy (existuje například skrytá kriminalita, jako jsou zločiny bílých límečků nebo nehlášená trestná činnost uvnitř rodin, jako je třeba domácí násilí).

- V Brně vypadá jinak mapa kriminality (počet určitých trestných činů nebo trestných činů obecně) podle odhalených trestných činů, jinak by vypadala podle skutečně spáchaných trestných činů (včetně neodhalených a neohlášených). Zřejmě by se také ukázalo, že mnohem více, než by se čekalo, je trestných činů uvnitř rodin.

Mohou se také měnit definice. Například ve zdravotnických statistikách lze jen s obtížemi provádět historické komparace údajů. Vezměme třeba rozdíl mezi úmrtími na srdeční choroby v 21. století a v 19. století, kdy (právě s ohledem na zmíněnou schopnost diagnostikovat a rozlišovat mezi diagnostikami) byla do této kategorie zahrnována úmrtí, která se dnes zahrnují do kategorií zcela jiných. Analogicky se mohou měnit klasifikace určitých osob či domácností mezi chudé či nechudé, aniž by se změnila jejich situace (příjmy či vydání), a sice prostým administrativním úkonem – změnou hranice chudoby.

Vše komplikuje i to, že tvůrci statistik, respektive ti, kdo pro ně poskytují údaje, mohou přitom přihlížet ke svým zájmům. Policie je například hodnocena podle míry objasněnosti svých případů a policisté mají zájem, aby některé činy nebyly ve statistikách zachyceny a jiné ano (nejde jen o přímé falšování údajů, ale i o takové aktivity, jako je neochota přijímat trestní oznámení drobných činů, které mají tradičně nízkou míru objasnění). Podobně se někdy licituje v nemocnicích, zda dítě zachytit jako mrtvě narozené nebo jako novorozeneckou úmrtnost apod. (takto manipulovaly a manipulují se statistikami i některé státy). Policisté mohou sympatizovat s hnutími vyjadřujícími xenofobii a rasismus, jako jsou například skinheads, a snažit se – když už musejí jejich činy vyšetřovat – neklasifikovat je jako motivované rasovou nesnášenlivostí. Naopak může docházet k růstu podílu takto motivovaných činů ve statistikách pod vlivem vnějších tlaků, ať již mocenských ze strany různých příslušných státních institucí (ministerstva vnitra, státních zastupitelství, soudů), nebo pod vlivem změn legislativy – např. změnou definice, jejím jednoznačnějším výkladem apod., popřípadě pod tlakem veřejného mínění či kontrolou masmédií, zejména jsou-li vyburcovány díky zvláště brutálnímu činu.

Existují i jiné způsoby „práce“ se statistikami. Jen jeden z možných příkladů: v 19. století bylo velmi obtížné dosáhnout rozvodu. Důvody k rozvodu, které soudy uznávaly, byly velmi omezené a možnost rozvodu po vzájemné dohodě tehdy v právních normách neexistovala. Přesto se však tento způsob rozvodu v některých zemích fakticky vyskytoval. Manželé si totiž například účelově po vzájemné dohodě vyměňovali urážlivé dopisy a mohli pak být rozvedeni pro oboustranné těžké urážky (viz Bologne, 1996).

Absolutní čísla jsou často interpretačně málo zajímavá, zajímavější jsou čísla poměrná (přepočtená na určitý základ – často se o nich hovoří jako o specifických



ukazatelích). Policejní statistika je jen hrubou informací o struktuře trestných činů (o podílu dané činnosti v celku trestné činnosti), zajímavější údaj však dostaneme, přepočteme-li čísla například na celkový počet obyvatel určitého území nebo na cílové skupiny (např. počet trestných činů mladistvých nikoliv v celku trestných činů, ale v přepočtu na počet osob této věkové skupiny v populaci a porovnání tohoto přepočtu s ekvivalentními přepočty ostatních věkových skupin). Velmi cenné jsou komparace s jinými prostředními, například s jinými městy, státy – ovšem za předpokladu, že je používána ve všech prostředích stejná metodika a tato metodika je stejným způsobem vykládána.

### Literatura

- Babbie, E. (2001). *The Practice of Social Research* (9th ed.). Belmont: Wadsworth Publishing Company.
- Barton, A. H. (1955). The Concept of Property-Space in Social Research. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The Language of Social Research* (pp. 40–53). New York: The Free Press.
- Bell, D. (1999). *Kulturní rozpory kapitalismu*. Praha: Slon.
- Blahušík, P. (2000). Statistická významnost proti vědecké průkaznosti výsledků výzkumu. *Česká kinantropologie*, 4(2), 53–72.
- Blaikie, N. (2003). *Analyzing Quantitative Data*. London: Sage.
- Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.
- Bologne, J. C. (1996). *Svatby. Dějiny svatebních obřadů na Západě*. Praha: Volvox Globator.
- Disman, M. (1993). *Jak se vyrábí sociologická znalost*. Praha: Slon.
- Field, A. (2009). *Discovering Statistics Using SPSS for Windows. Advanced Techniques for the Beginner*. London: Sage.
- Hendl, J. (2004). *Přehled statistických metod zpracování dat*. Praha: Portál.
- Kerlinger, F. N. (1972). *Základy výzkumu chování*. Praha: Academia.
- Lorenz, K. (1991). *8 smrtelných hříchů*. Praha: Panorama.
- Mertens, D. M. (1998). *Research Methods in Education and Psychology*. London: Sage.
- Mills, Ch. W. (2002). *Sociologická imaginace*. Praha: Slon.
- Nowak, S. (1975). *Metodologie sociologických výzkumů*. Praha: Svoboda.
- Petrusek, M. (1993). *Teorie a metoda v moderní sociologii*. Praha: Karolinum.
- Popper, K. R. (1974). Thesis in Defence of Science and Rationality. In Notturmo, M. A. (Ed.), *The Myth of the Framework*. London: Routledge.
- Rabušic, L., & Mareš, P. (1996). Je česká společnost anomická? *Sociologický časopis*, 32(1), 175–188.
- Řehák, J. (1971). Definice měření ve společenských vědách. *Sociologický časopis*, 7(6), 639.
- Sirovátka, T., & Mareš, P. (2006). Poverty, Social Exclusion and Social Policy in the Czech Republic. *Social Policy and Administration*, 40(3), 288–303.
- Soukup, P., & Rabušic, L. (2007). Několik poznámek k jedné obsesi českých sociálních věd – statistické významnosti. *Sociologický časopis / Czech Sociological Review*, 43(2), 379–395.
- Vaus, D. de. (1990). *Survey in Social Research*. London: Unwin Hyman.

**Petr Mareš / Ladislav Rabušic / Petr Soukup**

# **Analýza sociálněvědních dat (nejen) v SPSS**

Masarykova univerzita  
Brno 2015