

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.²¹ 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.²² 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:

²¹ 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.

²² 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. Říkají-li si „miluje mě“, z čeho tak usuzují – z jakých jeho/jejích 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 délkou. 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.).²³

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

²³ 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.²⁴ 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.

²⁴ **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.

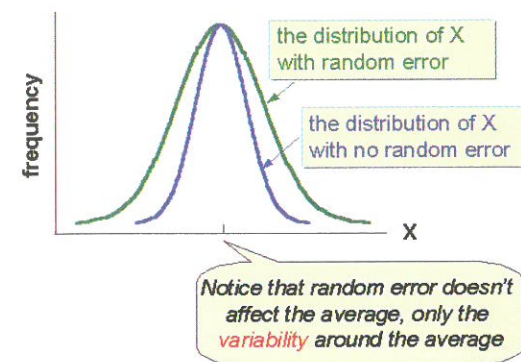
Rozlišovat úroveň měření a druhy proměnných je velmi důležité, neboť na druhu proměnných jsou závislé statistické operace, které můžeme s daty provádět. Ideálem pro statistiku jsou pochopitelně proměnné kardinální, neboť umožňují nejvyšší množství statistických operací. V sociálních vědách jich ovšem není příliš mnoho, velmi často se pohybujeme 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 vlastnosti 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.²⁵ 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:

$$X = T + e,$$

²⁵ 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