
KAPITOLA 1

Vzorek s genetickou vadou

Časopis *Time* jednou komentoval něco z deníku *New York Sun* a poznamenal, že „průměrný student univerzity Yale, který absolvoval v roce 1924, vydělává ročně 25 111 dolarů“.

Ten se tedy má!

Ale moment. Co znamená tohle působivé číslo? Vypadá to, že když pošlete svého synka na univerzitu Yale do New Havenu, nebudete muset k stáru pracovat ani vy, ani on. Ale je to opravdu tak?

Už při prvním pozornějším pohledu jsou na tom čísle podezřelé dvě věci. Jednak je překvapivě přesné a zároveň je až nepravděpodobně příznivé.

Jen těžko se dá věřit tomu, že by průměrný příjem jakékoliv tak rozsáhlé skupiny lidí byl známý s přesností na dolar. Ani vy sami nejspíš neznáte svůj příjem za loňský rok tak přesně, leda by celý pozůstal jen ze mzdy. Ale příjmy kolem 25 000 dolarů ročně obvykle

nepocházejí jen ze mzdy; lidé v této příjmové kategorii mají většinou náležitě diverzifikované investiční portfolio.

Navíc byl tento roztomilý průměr nepochybně vypočítán z obnosů, o kterých absolventi Yale *řekli*, že je vydělávají. I kdyby na téhle univerzitě v New Havenu v roce 1924 všichni ctili tuto povinnost a sdělovali pravdivé údaje, není moc jisté, že to tak funguje i po čtvrt století a že tedy všechna hlášení jsou poctivá. Někteří lidé, když se jich zeptáte na jejich příjmy, tak je nadsadí. Z ješitnosti, nebo z optimismu. Jiní naopak uvedou příjmy nižší – nejspíš s ohledem na svá daňová přiznání; už je odevzdali a nechťejí teď na žádném dalším papíře uvádět něco jiného. Kdo ví, k čemu se daňoví úředníci mohou dostat? Tyto dvě odchylky, nadsazení a snížení, se jistě mohou navzájem vynulovat, ale to je dost nepravděpodobné. Naopak – jedna z těch odchylek může být podstatně větší a my nevíme, která to je.

Začali jsme tedy pracovat s údajem, o němž nám zdravý rozum říká, že jen těžko může být pravdivý. Zaměřme se teď na pravděpodobný zdroj největší chyby, totiž na to, z jakých podkladů lze odvodit závěr, že 25 111 dolarů je „průměrný příjem“ nějaké skupiny lidí, jejichž skutečný průměrný příjem může být klidně i poloviční.

Zdrojem největší chyby je způsob výběru vzorku, což je bolestí většiny statistik, se kterými se setkáte v nejrůznějších oblastech. Princip je celkem jednoduchý, ale praktické provádění si prošlapalo různé cestičky, přičemž některé jsou dost nekorektní. Když máte hrnec fazolí, z nichž některé jsou červené a některé bílé, je jen jediný způsob, jak lze přesně zjistit, kolik jich je od té které barvy: spočítat je. Podstatně jednodušším způsobem můžete však zjistit, kolik přibližně je v hrnci červených fazolí, a to tak, že si naberete hrst těch fazolí, spočítáte červené a budete předpokládat, že poměr červených a bílých je stejný v celém hrnci jako ve vaší hrsti. Pokud váš vzorek bude dostatečně velký a bude správně vybraný, bude pro většinu účelů docela dobře reprezentovat celý hrnec. Pokud nebude dost velký a nebude dobře vybraný, může být daleko méně přesný než rozumný odhad. Pak tato metoda nemá žádnou jinou výhodu, než pochybnou auru vědecké přesnosti. Je to mrzutá pravda, ale závěry, učiněné na základě podobných vzorků, příliš malých, špatně vybraných, nebo závadných oběma těmito způsoby, stojí za většinou toho, co čteme nebo o čem si myslíme, že to víme.

Zpráva o absolventech Yale pochází ze vzorku. Můžeme si tím být naprosto jisti, neboť zdravý rozum

nám říká, že se těžko někdo může dostat ke všem žijícím absolventům ročníku 1924. Musí být hodně těch, jejichž adresy jsou po pětadvaceti letech neznámé.

A z těch zbývajících, jejichž adresy známé jsou, mnozí nevyplnili dotazníky. Zejména, když se jednalo o tak intimní otázky. U některých typů rozesílaných dotazníků je dobrým výsledkem, když se jich vyplněných vrátí 5 až 10 procent. Tento by třeba mohl být i úspěšnější, ale rozhodně se to neblížilo stu procent.

Je tedy jasné, že údaj o příjmech je založen na vzorku, sestávajícího z těch absolventů v daném ročníku, jejichž adresy byly dosažitelné a kteří zároveň na dotazník odpověděli. Je to reprezentativní vzorek? Jinými slovy: můžeme předpokládat, že v této skupině jsou příjmy rozděleny stejně jako mezi zbývajících absolventy, kteří do tohoto vzorku nebyli zahrnuti, tedy mezi těmi, kteří nemohli být zastiženi, nebo kteří neodpověděli?

Kdo jsou ty malé ovečky, zatoulané na lukách v Yale, s označením „adresa neznámá“? Jsou to lidé s vysokými příjmy? – Chlapíci z Wall Street, ředitelé korporací, vedoucí pracovníci továren a institucí? Ne, ti to nejsou. K adresám bohatých lidí se dá dostat snadno. Ti nejlépe prosperující absolventi daného ročníku se většinou dají najít v knize „Kdo je kdo

v Americe“ a v dalších pramenech, a to i v případě, že zanedbali svou povinnost a neohlásili se v kanceláři pro absolventy. Celkem spolehlivě se dá odhadnout, že ztracená jména patří lidem, kterým se za těch pětadvacet let od promoce na Yale nepodařilo naplnit žádnou ze skvělých vyhlídek. Jsou z nich prodavači, mechanici, tuláci, nezaměstnaní alkoholicí, sotva přežívající spisovatelé a umělci..., lidé, kterých by se muselo složit nejméně půl tuctu, aby dali dohromady příjem 25 111 dolarů. Ti se většinou neregistrují na srazy absolventů. Už jen proto, že si takový výlet nemohou dovolit.

Pak jsou tu lidé, kteří zahodili dotazníky do nejbližšího odpadkového koše. Jací lidé to jsou? Jistotu samozřejmě mít nemůžeme, ale přinejmenším se dá poctivě odhadnout, že většinou prostě nevydělávají tolik, aby se s tím mohli chlubit. Jsou trochu jako ten chlapík, co na své první výplatní pásce našel přilepený lísteček s poznámkou, aby výši své mzdy považoval za důvěrnou informaci a nesvěřoval se s ní dál. Šel za svým šéfem a povídá: „Nedělejte si starosti, já se za tu almužnu stydím stejně jako vy.“

Je nám tedy stále jasnější, že vzorek pominul dvě skupiny, které by průměr nejspíš posunuly směrem dolů. To číslo 25 111 dolarů se začíná samo vysvět-

lovat. Pokud o něčem vypovídá pravdivě, pak jedině o vybrané skupině absolventů z ročníku 1924, jejichž adresy jsou známé a kteří jsou ochotní povstat a říci, kolik vydělávají. Ale i potom je ve hře pouhý předpoklad, nikoli jistota, že tito pánové mluví pravdu.

Takový předpoklad by se ovšem neměl dělat automaticky. Na základě zkušeností s jedním typem výběrových šetření, s tzv. „průzkumem trhu“, se dá říct, že podobný předpoklad hraje velkou roli a je otázkou, zda si jej vůbec lze dovolit. Jednou byl proveden průzkum od domu k domu s cílem analyzovat čtenost časopisů. Základní otázka zněla: Jaké časopisy čtete ve vaší domácnosti? Když se výsledky sečetly a zpracovaly, ukázalo se, že velké množství lidí má rádo *Harper's* a naopak velmi málo lidí čte *True Story*. Za stejné období tu ale zároveň byla čísla od vydavatelů a ta velmi jasně ukazovala, že *True Story* má několik milionů prodaných výtisků, zatímco *Harper's* se prodaly jen stovky tisíc. Autoři výzkumu si říkali, že asi oslovili nesprávné lidi. V tom ale problém nebyl. Otázka byla pokládána ve všech typech domácností po celé zemi. Jediný rozumný závěr tehdy byl, že významné množství respondentů (jak se říká lidem v těchto šetřeních), nemluvilo pravdu. Všechno, co ten průzkum odkryl, bylo snobství.

Nakonec se zjistilo, že když chcete vědět, co určití lidé čtou, nemá cenu se jich na to ptát. Pochodíte lépe, když k nim zajdete domů a řeknete, že chcete koupit staré časopisy a jestli by nějaké neměli. Vezmete je a spočítáte, kolik je *Recenzí z Yale* a kolik *Milostných romancí*. Ani tento pochybný postup vám samozřejmě neřekne, co lidé čtou, ale jen – čemu byli vystaveni.

Podobně až se zase jednou z nějakého článku dozvíte, že průměrný Američan (v poslední době o něm slyšíte často a většinou nic, co by bylo alespoň trochu pravděpodobné) si čistí zuby 1,02 krát denně – to číslo jsem si právě teď vymyslel, ale je stejně dobré jako kterékoliv jiné – položte si otázku: Jak může někdo zjistit něco takového? Copak existuje žena, která by byla ochotna cizímu člověku přiznat, že si nečistí pravidelně zuby, když si v tolika reklamách přečetla, že kdo si nečistí zuby, je sociální vyvrhel? Taková statistika má smysl pro toho, kdo chce zjistit jen to, co lidé o čištění zubů říkají. Nejistí ale nic podstatného o tom, s jakou frekvencí se štětiny kartáčků dotýkají našich stoliček a řezáků.

Stará moudrost říká, že řeka nevyteče výš, než je její pramen. Když tam ale někde zapojíte čerpadlo, tak se to stát může. Stejně pravdivé je i tvrzení, že výsledek výběrové studie nemůže být spolehlivější, než

nakolik spolehlivá je kvalita výběru vzorku, na kterém ta studie byla provedena. Jestliže se ale data proženou filtrem statistických manipulací a průměr se vyčíslí na několik desetinných míst, vzbuzuje výsledek dojem přesvědčivosti. Ten se ovšem rozplyne, když se podíváte na to, jak kvalitně byl proveden výběr vzorku.

Když se rakovina včas odhalí, může to zachránit život? Snad ano. Ale o číslech, kterými se toto tvrzení běžně dokazuje, se dá v nejlepším případě říct, že to prostě nedokazují. Mám na mysli záznamy z Registru nádorů v Connecticutu. Vracejí se do roku 1935, kdy byla tzv. doba dožití po zjištění nádoru v průměru pět let. Během následujících šesti let do roku 1941 se tato doba výrazně prodloužila. Ve skutečnosti bylo ale celé sledování zahájeno v roce 1941 a všechno, co se dělo předtím, bylo zjišťováno zpětně. Mnoho pacientů za tu dobu Connecticut opustilo, a tak se nedalo zjistit, jak dlouho tito lidé žili a kdy zemřeli. Podle Leonarda Engela, komentátora, který se specializoval na zdravotnictví, bylo takto zabudované zkreslení vzorku „samo o sobě schopno způsobit prakticky celé deklarované prodloužení doby dožití“.

Aby mělo výběrové šetření skutečnou hodnotu, musí být vybrán reprezentativní vzorek, tedy takový, který není v žádném ohledu vychýlený a svým slože-

ním odpovídá struktuře celku. Proto je ten údaj z Yale bezcenný. A když se z tohoto hlediska zamyslíte nad většinou věcí, které čtete v novinách a časopisech, zjistíte, že postrádají velmi podobně pevný základ.

Jistý psychiatr jednou prohlásil, že prakticky každý je neurotik. Pomiňme skutečnost, že když se pojem „neurotik“ použije tímto způsobem, ztrácí to slovo jakýkoliv smysl a podívejme se na vzorek, se kterým ten dobrý muž pracoval. Tedy, koho ten psychiatr pozoroval? Ukázalo se, že k tomuto objektivnímu závěru došel studiem svých pacientů. Ti ovšem zdaleka nemohou být vzorkem celé populace. Pokud je někdo normální, psychiatr se s ním v životě nesetká.

Nechcete-li vstřebávat spoustu věcí, které jsou ve skutečnosti jinak, je třeba věnovat trochu kritické pozornosti všemu, co čtete.

Je také třeba mít na paměti, že spolehlivost, reprezentativnost vzorku může být velmi snadno narušena nejen viditelnými, ale i neviditelnými zdroji vychýlení. Zachovejte si tedy určitou míru pochybností i v případech, kdy nevidíte zjevný zdroj prokazatelného vychýlení a to tak dlouho, dokud nějaká možnost vychýlení existuje. Ona totiž existuje vždycky. Pokud o tom pochybujete, prezidentské volby v letech 1948 a 1952 to jasně potvrdily.

Pro další důkaz se můžeme vrátit do roku 1936 ke slavnému flasku časopisu *Literary Digest*. Redaktoři tohoto nebohého plátku oslovili deset milionů lidí podle telefonního seznamu a seznamu předplatitelů *Digestu*. Ti všichni je ujistili, že výsledek bude pro Landonu 370 hlasů a pro Roosevelta 161 hlas. Byli to stejní lidé, kteří velmi přesně předpověděli výsledek předcházejících voleb v roce 1932. Kde by se tedy vzalo vychýlení vzorku, když už se jednou tak dobře osvědčil? Samozřejmě, vychýlení existovalo. Ukázaly to akademické a další analýzy, provedené post mortem: Lidé, kteří si mohli v roce 1936 dovolit telefon a předplatné na časopis, nebyli průřezem všech voličů. Ekonomicky to byla zvláštní skupina lidí, vychýlený vzorek, protože v něm byla převaha republikánských voličů. Vzorek zvolil Landonu, ale celý soubor voličů rozhodl jinak.

Nejjednodušší vzorek je založen na náhodném výběru. Ten se provádí čistě namátkou z „vesmíru“, pod kterýmžto pojmem rozumí statistik onen celek, jehož částí má vybraný vzorek být. Z kartotéky se vytáhne každé desáté jméno. Z klobouku se vytáhne padesát papírků. Každý dvacátý člověk, kterého potkáte na Market Street, je dotázán. (Pamatujte však, že tento poslední příklad není vzorkem světové popu-

lace nebo USA nebo San Franciska, ale pouze vzorkem lidí na Market Street v určitém čase. Jedna tazatelka ve výzkumu veřejného mínění řekla, že své lidi oslovila na železniční stanici, protože „na nádraží lze najít všechny typy lidí“. Bylo jí pak vysvětleno, že zde nemusejí být dostatečně zastoupeny například matky s malými dětmi.)

Test náhodného vzorku provedete snadno. Položte si otázku: Má každé jméno nebo věc z celého souboru stejnou šanci dostat se do vzorku?

Čistě náhodný vzorek je jediný, který může být zkoumán statistickou teorií s naprostou důvěrou. Je tu ale jeden problém. Získat náhodný vzorek je v mnoha případech tak složité a drahé, že to už sama nákladnost vylučuje. Levnější náhražkou, užívanou téměř univerzálně v oblastech, jako jsou průzkumy názorů a výzkumy trhu, je tzv. stratifikovaný náhodný výběr.

Abyste získali takový stratifikovaný vzorek, rozdělíte si svůj „vesmír“ na několik skupin, jejichž podíl na celku znáte. A právě tady můžete narazit na problém: vaše informace o jejich podílu na celku nemusí být správná. Dáte svým tazatelům instrukci, aby zajistili, že budou hovořit s určitým počtem lidí podle barvy pleti, s určitým procentem lidí v jednotlivých příjmových kategoriích, s určitým počtem farmářů a tak dále.

Při tom všem musí být skupina přiměřeně rozdělena mezi osoby nad čtyřicet a pod čtyřicet let věku.

Zní to pěkně – ale co se stane? V otázce barvy pleti se váš tazatel bude většinou rozhodovat správně. Pokud jde o příjmy, velmi často se netrefí. Nu a co se týče farmářů – jak budete klasifikovat člověka, který dělá na farmě na půl úvazku a zároveň pracuje ve městě? Dokonce i otázka věku může být poněkud problematická. Tazatel to nejspíš vyřeší tím, že si vybere pouze respondenty, kteří budou opravdu zřetelně pod nebo zřetelně nad čtyřicet let. V tom případě bude vzorek opět vychýlený, protože v něm budou chybět skupiny starších třicátníků a čerstvých čtyřicátníků. Nemůžete vyhrát.

Navíc je otázkou, jak získat náhodný vzorek v rámci určitého rozvrstvení? Nejjednodušší je začít se seznamem všech, vzít jméno za jménem a náhodně vybírat: to je ale příliš drahé. Tak vyrazíte do ulic – a vyřazujete ze svého vzorku lidi, kteří jsou doma. Jdete od domu k domu během dne – a pomínete tak většinu zaměstnaných lidí. Přejdete na večerní dotazování – a zanedbáte návštěvníky kin a nočních klubů.

Provádění průzkumů se tak stává nepřetržitým bojem se zdroji vychýlení a tento boj vedou soustavně všechny uznávané agentury pro výzkum veřej-

ného mínění. Když čteme jejich zprávy, musíme si uvědomit, že tento boj není nikdy vítězně dobojován. Když čteme nějaké tvrzení, že „67% Američanů je proti něčemu“ a podobně, vždycky bychom si měli položit stejnou otázku: 67% JAKÝCH Američanů?

Stejně je to i v případě „ženské sexuality“, jak ji nastolil dr. Alfréd C. Kinsey. Jako u všeho, co je založeno na výzkumu vzorku, je otázkou, jak tuto knihu (či populárnější výtah z ní) číst, aniž bychom podleli informacím, které nemusejí být úplně pravdivé. Jsou zde uplatněny přinejmenším tři úrovně výběru vzorků. Vzorky populace, jak je vybíral dr. Kinsey (jedna úroveň), jsou na hony vzdálené náhodným vzorkům a nebudou ani příliš reprezentativní. Jsou to ale obrovské vzorky v porovnání s čímkoliv, co bylo v jeho oboru provedeno dříve. Proto musejí být jeho čísla přijímána jako objektivní a důležitá, i když ne nezbytně úplně přesná. Důležitější možná je, že i každý dotazník je jen vzorkem (další úroveň) možných otázek a že odpověď, kterou ona dáma dává, není nic jiného než opět vzorek (třetí úroveň) jejich postojů a zkušeností ke každé z položených otázek.

Lidé, kteří sestavují týmy tazatelů, mohou ovlivnit výsledek zajímavým způsobem. Před několika lety, během války, vyslalo Národní centrum pro výzkum

názorů dva týmy tazatelů, aby položili tři otázky pěti stovkám barevných lidí v jednom jižanském městě. Bílí tazatelé tvořili jeden tým, barevní druhý.

Jedna otázka zněla: „Zacházelo by se tu s barevnými lépe nebo hůř, kdyby Japonci dobili Spojené státy?“ Černí tazatelé zjistili, že 9% dotázaných řeklo „lépe“. Bílí tazatelé získali takových odpovědí jen 2%. A zatímco černí tazatelé zaznamenali jen 25% těch, kteří si mysleli, že by se s černými nakládalo hůř, bílí tazatelé jich objevili 45%.

Když se do otázky místo „Japonců“ dosadili „nacisté“, výsledky byly obdobné. Třetí otázka testovala postoje, které by mohly vycházet z pocitů, vyvolaných prvními dvěma otázkami. „Myslíte si, že je důležitější soustředit se na porážku německo-italsko-japonské Osy nebo zlepšit fungování demokracie u nás doma?“ Podle barevných tazatelů odpovědělo 39% dotázaných, že „porazit Osu“; podle bílých tazatelů dalo tuto odpověď 62% dotázaných.

Tady se projevuje zkreslení (vychýlení), způsobené netušeným faktorem. Největší vliv tu zřejmě měla tendence, která musí být vždy brána v úvahu při čtení průzkumů názorů. Tou je snaha dávat tazateli odpovědi, které ho potěší. Mohli bychom se divit jižanskému černochovi, který odpovídá na otázku

související s loajalitou v době války, že řekne bílému člověku raději to, co zní dobře, než to, čemu ve skutečnosti věří? Je také možné, že různé skupiny tazatelů vyhledávají různé typy lidí, které oslovují.

V každém případě byly výsledky tohoto průzkumu evidentně tak nereprezentativní, že byly prakticky bezcenné. Můžete sami pro sebe posoudit, kolik dalších tvrzení, založených na průzkumech veřejného mínění, je tak vychýlených, až jsou bezcenné – ovšem bez možnosti to nějak zkontrolovat a prokázat.

Máte-li pocit, že průzkumy jsou obecně postavené na nereprezentativním vzorku, pak v jednom ohledu máte nejspíš pravdu a existuje k tomu dostatek dobrých důkazů – je to vychýlení, které se projevilo i u časopisu *Literary Digest*. Do vybraných vzorků se dostávají v převaze lidé, kteří mají víc peněz, vyšší vzdělání, jsou lépe informovaní a bdělejší, mají lepší vzhled, lepší způsoby chování a ustálenější návyky než průměr populace, kterou by měl vzorek reprezentovat.

Snadno se přesvědčíte o tom, jak to vzniká. Představte si, že jste tazatel a vaším úkolem je udělat na rohu ulice jeden rozhovor. Všimnete si dvou mužů, kteří vypadají, že patří do kategorie, kterou potřebujete doplnit: přes čtyřicet, barevní a žijící ve městě.

Jeden má čisté montérky, na nich úhledné záplaty a je čistý. Druhý je špinavý a tváří se mrzutě. Abyste tu práci měl už za sebou, přistoupíte k tomu chlapíkovi, který vypadá lépe. Vaši kolegové po celé zemi se rozhodnou úplně stejně.

Největší odpor k průzkumům veřejného mínění mají lidé v liberálních nebo levicových kruzích. Ti v drtivé většině věří tomu, že průzkumy jsou celkově zmanipulované. Je to způsobeno tím, že výsledky průzkumů jsou velmi často v rozporu s názory a touhami lidí, kteří nepřemýšlejí konzervativním směrem. Zdá se, říkají, že průzkumy veřejného mínění volí Republikány, i když voliči krátce nato rozhodnou jinak.

Ve skutečnosti, jak jsme si ukázali, nemusí být průzkumy nezbytně zmanipulované v tom smyslu, že by výsledky byly úmyslně překrouceny, aby vytvořily falešný dojem. Vzorek sám může mít sklon být nereprezentativní (určitým směrem vychýlený) a může výsledek zkreslit automaticky.

Poznámka Evy Zamrazilové:

Nejprve je třeba vysvětlit, že pojem „statistika“ se běžně používá v mnoha souvislostech. Zcela základním rozlišením je statistika popisná a induktivní. Popisná statistika se v zásadě zabývá soupisem a je schopna prošet-

řit kompletně celý základní soubor jednotek, který nás zajímá. Typickým příkladem úplného šetření je „Sčítání lidí, domů a bytů“, které provádí ČSÚ zhruba jednou za deset let a které tedy patří do domény popisné statistiky. Většinou však z technických, finančních, časových a dalších důvodů nelze prozkoumat celý základní soubor (někdy říkáme populaci). Základní soubor nemusí vždy tvořit lidé (a už vůbec ne všichni občané ČR) – může být definován jakkoli, podle účelu šetření. Může se jednat o veškeré obyvatelstvo státu, ale také např. o muže ve věku 25 – 50 let, studenty středních škol v Brně, vlastníky rodinných domů v Praze apod. V případech, kdy nelze prozkoumat celý základní soubor, je využívána statistika induktivní: ze základního souboru je vybrán vzorek – tzv. „výběrový soubor“ a z výsledků na něm zjištěných pak vyvozujeme závěry pro celý „základní soubor“. Odtud pojem indukce (na rozdíl od deduktivních metod Sherlocka Holmese).

Celému procesu, který začíná výběrem vhodného vzorku a končí interpretací výsledků, říkáme „výběrové šetření“. Prezentace výsledků výběrových šetření v médiích je jednou z nejčastějších situací, kdy se běžný občan setkává se statistikou. Výsledky výběrových šetření jsou vždy zatíženy chybou, protože závěry z průzkumu, provedeného na části základ-

ního souboru, na tzv. výběrovém souboru (vzorku) se zobecňují pro celý základní soubor. Výsledky výběrových šetření jsou proto vždy jen odhadem, platícím s určitou pravděpodobností. Výběrový soubor je určitá výseč základního souboru, která by měla být jeho reprezentativním vzorkem. Velmi často se provádí tzv. kvótní výběr, kdy výběrový soubor zachovává z hlediska zásadních důležitých charakteristik strukturu souboru základního. V zásadě by výběrový soubor měl být co nejdokonalejší miniaturou základního souboru. Vybranými charakteristikami (ve statistice říkáme statistickými znaky) bývají tzv. socio-demografické charakteristiky – věk, vzdělání, pohlaví, ekonomická aktivita atd. Ty by měly být dobře zvoleny s ohledem na účel výběrového šetření.

Není-li výběrové šetření provedeno metodicky správným způsobem, bude zatíženo tzv. systematickou chybou. Tento druh chyby nevzniká náhodou a při opakování průzkumu či šetření se bude stále opakovat. Zdrojem chyby může být nedostatečně velký nebo nereprezentativní vzorek. Malé a/nebo nereprezentativní vzorky častěji poskytují extrémní výsledky. Dalším zdrojem systematických chyb jsou nesprávně formulované otázky nebo nesprávný způsob dotazo-

vání – příkladem je kladení takzvaných návodných otázek.

V textu D. Huffa jsou nekorektní postupy velmi dobře popsány. Pouliční dotazování je typický případ, jak získat nereprezentativní vzorek – nezkoumáme-li ovšem zrovna způsob dopravy obyvatel po městě. Další ukázkou, jak vytvořit nereprezentativní vzorek ukazuje pasáž o volebním průzkumu. Volební modely nebo modely stranických preferencí jsou vlastně speciálním případem průzkumu veřejného mínění a setkáváme se s nimi v médiích velmi často. I když se výběrové šetření provede bezchybně a korektně, výsledky jsou vždy zatíženy určitou chybou, tzv. statistickou chybou. Té se vyhnout nelze, protože vzniká tím, že nebyl prošetřen celý základní soubor, ale jen jeho vzorek. Solidní zpráva o průzkumu veřejného mínění by proto měla uvést, jak vysoká je pravděpodobnost přibližné správnosti předkládaného údaje. Statistika vymezí tuto pravděpodobnost, za což zaplatí určitým zpochybněním přesnosti. Jinak to ale nejde. Výběrové šetření je často jedinou možností, jak zjistit názor obyvatel na určitou otázku nebo prošetřit určitý jev v rozumném čase a za rozumné náklady.