

5.1 Úvod

Žijeme v informačním věku. Informace na nás útočí ze všech stran, topíme se v nich. Ze všech stran se dozvídáme, co si máme myslet, co si máme koupit, koho máme volit. Tvrzení nám předkládají „odborníci“, jsou „výsledkem vědeckých výzkumů“, jsou „statisticky dokázána“. Obecně rozšířená představa o podvodné roli statistiky je tak paradoxně provázena i svým opakem: magie čísel s mnoha desetinnými místy působí, že argument „je statisticky dokázáno“ je považován div ne za vrchol nevyvratitelnosti.

Dvě protichůdné představy vytvářejí v mysli normálního člověka obraz statistiky coby podezřelé, obtížné a tajemné vědy. To by ani nevadilo, pokud by s ní přicházel do styku jen zřídka. Je tomu ale právě naopak, ze všech masových médií je občan neustále zaplavován statistickými údaji, které se snaží člověka přesvědčit a získat pro jisté účely a cíle. Čísla však prokazují argumenty, které jsou na nich zkonstruované, jen někdy. Častěji vytvářejí pouze jakousi vidinu pravdy jako důsledek nevědomosti nebo nepozornosti veřejnosti. Někdy je i přímým úkolem číselných údajů úmyslně zmást.

Následující kapitola se pokusí nabídnout návod, jak se ve statistických informacích úplně neutopit, nenechat se jimi ovládnout, vyznat se v nich, mít nad nimi pokud možno „navrch“.

5.2 Hierarchie vědění a statistika

Začneme od hierarchie vědění a jednotlivých používaných termínů. Podle systémového teoretika Russella Ackoffa (Veselý, Kalous, Marková, 2004) můžeme obsah lidské mysli klasifikovat pomocí těchto pěti kategorií:

1. *Data*: Symboly.
2. *Informace*: Data, která jsou zpracována tak, aby poskytovala odpovědi na zjišťovací otázky typu „kdo“, „co“, „kdy“, „kde“.
3. *Vědění*: Aplikace dat a informací, odpověď na otázku „jak“.
4. *Porozumění*: Pochopení otázky „proč“.
5. *Moudrost*: Zhodnocení porozumění.

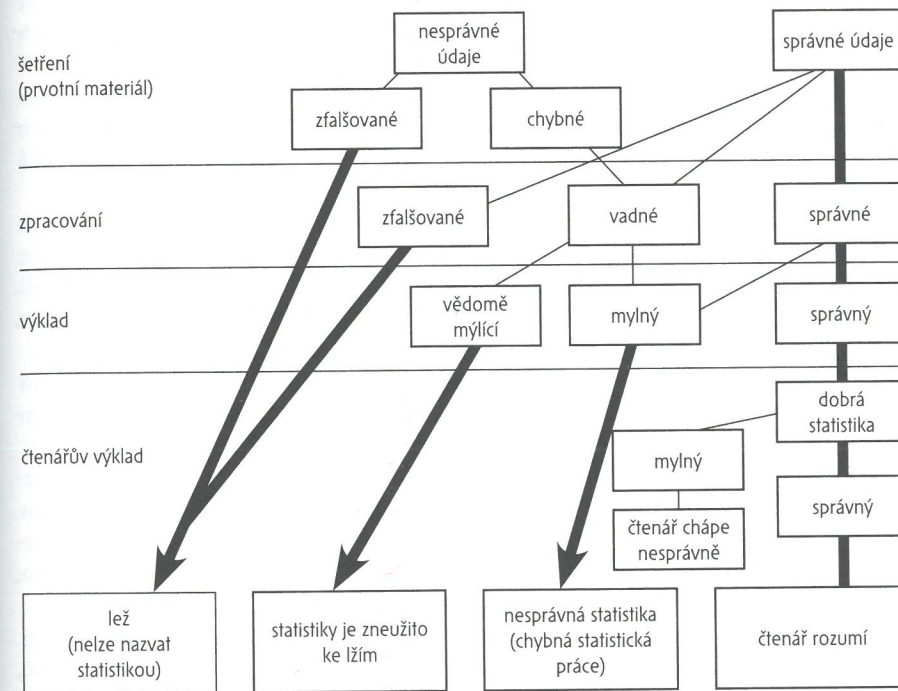
První čtyři kategorie se týkají minulosti; zabývají se tím, co bylo či co je známo. Pouze pátá kategorie, moudrost, se zabývá budoucností, obsahuje vizi, záměr, plán. S moudrostí mohou lidé vytvářet budoucnost spíše než se jen snažit pochopit současnost a minulost. Ale dosažení moudrosti není snadné, a lidé proto musí projít postupně nižšími kategoriemi.

Statistika je věda o získání informací z dat. Nejstarší statistikou byl „popis státu“ spočívající v zobrazení daného zeměpisného, hospodářského a politického stavu. Takové informace byly často numerického charakteru, a jak narůstalo jejich množství, přecházely postupně do tabulkové formy prezentace. Přirozeným posunutím významu začalo slovo „statistika“ koncem minulého století znamenat jakýkoli numerický materiál

vzniklý pozorováním vnějšího světa. Dnes se toto slovo užívá v několika významech: od statistiky jako vědního oboru přes statistiku jako metodu až po statistiku jako jistým způsobem prezentovanou informaci.

5.3 Statistické zpracování dat a statistikulace

Všekerou statistickou činností lze rozdělit do tří základních etap: získávání prvotních dat, jejich zpracování a interpretaci. Pouze správný postup ve všech třech etapách může vést k pravdivé výpovědi. Nesprávný postup v kterékoli etapě vede k nepravdivému tvrzení, a to až uť jde o chybu neúmyslnou – omyl, nebo o vědomé falšování skutečnosti – podvod.



Obr. 5.1 Schéma zdrojů chyb a možností klamání při zpracování a interpretaci dat

Zdroj: Swoboda, 1977

Jestliže manipulace znamená v hanlivém smyslu „nepoctivý způsob jednání“ a dezinformace je „úmyslné uvádění nepravdivých nebo zkrácených skutečností a údajů, a tím uvedení v omyl“, pak vedle statistiky jako vědecké disciplíny si svůj název zaslouží i velice

rozšířená statistikulace, kterou definujeme jako „dezinformaci a manipulaci lidí pomocí statistiky“.

„Jsou tři druhy lži: lež z nouze, sprostá lež a statistika.“ Tento výrok se vyskytuje v různých obměnách, všechny se však shodují v tom, že statistika je zvláště rafinovanou formou lži. Proti nařčení statistiky ze lži se ale statistici brání: „Statistika nelže, ale lháři užívají statistiku.“ Profesionální demagogové už dávno přišli na to, že statistikulace je mnohem užitečnější než prostá lež. Plní totiž stejný úkol, ale nikdo ji nemůže přímo označit jako nepravdu. Statistikulace je umění dokázat správnými čísly něco nesprávného.

5.4 Metody statistikulace

Hlavní otázkou při statistikulaci není „Jaký je výsledek?“, ale „K čemu je výsledku zapotřebí?“. Cíl, ke kterému mají být data užita, rozhoduje o výběru dat, nejnvýhodnější metodě jejich zpracování i o nejšikovnější interpretaci.

Zatímco statistika je podložena náročným matematickým aparátem, statistikulace ve směr vystačí s procenty, průměry a grafy – látkou známou už ze základní školy. Ale jaká kouzla s nimi dokáže provádět!

Statistikulace užívá většinou velmi přesná čísla s mnoha desetinnými místy. Jestliže například prohlásíte, že lidé denně spí v průměru 7,83 hodiny, bude to znít, jako že víte naprosto přesně, o čem mluvíte, a nikdo se nebude ptát, jak jste k tomu číslu přišli. Kdybyste však byli tak pošetilí a prohlásili, že lidé spí obvykle něco méně než osm hodin, nebylo by na tom už nic ohromujícího.

Novinový titulek hrdě hlásá: „Každý osmý delegát mladší 35 let!“ Jakou informační hodnotu nám však poskytujete? Cílem výroku je zřejmě vzbudit dojem o mládí delegátů a můžeme i věřit tomu, že je pravdivý. Ale zrovna tak pravdivá mohou být tvrzení „Většina delegátů je v důchodovém věku!“ anebo vědecky vypadající „Průměrný věk delegátů je 56,31 roku“. Tutéž skutečnost je obvykle možné vyjádřit nesčetným množstvím různých způsobů, z nichž každý si všímá pouze jednoho aspektu reality. Základní metoda statistikulace pak spočívá v tom, že se ze všech možných údajů uveřejní pouze ten, který zní nejvíce ve prospěch sledovaného záměru.

„V Meritánii mají 273 traktorů.“ Co si myslet o této informaci? Takový číselný údaj sám o sobě nemá valný význam, jenom z něj nemůžeme nikdy usoudit, je-li to hodně, či málo. K tomu míváme obvykle nápovědu ve formě doplňujících slůvek „pouze“, „již“ apod. Minimum objektivity však vyžaduje alespoň jedno srovnání (v našem případě s předchozími lety, s jinými zeměmi, v přepočtu na počet obyvatel, rozsah orné půdy apod.). Pak už je jen třeba dát si pozor na to, co a s čím se srovnává. Podle výběru srovnávaných charakteristik usuzujeme na míru objektivity. Existuje mnoho způsobů, jak něco spočítat a pak to interpretovat jako něco jiného.

Statistický údaj může mít mnoho různých výkladů. Statistikulant (tj. člověk provozující statistikulaci) vybere ten, který nejvíce vyhovuje jeho záměrům, a ostatní ignoruje. Často se třeba užívá postup „post hoc, ergo propter hoc“, známý už od starověku. Jde

o tvořivou aplikaci chybné úvahy „Jestliže B následuje po A, pak A je příčinou B“, kterou statistikulanti často a úspěšně užívají. Například všechna tvrzení, kdy nějaký dnešní stav B je předkládán jako následek jisté události A před x lety, jsou tohoto typu. Z časové následnosti jeví A a B není však možné jednoznačně usuzovat na jejich příčinnou souvislost, tu je třeba vždy řádně prokázat.

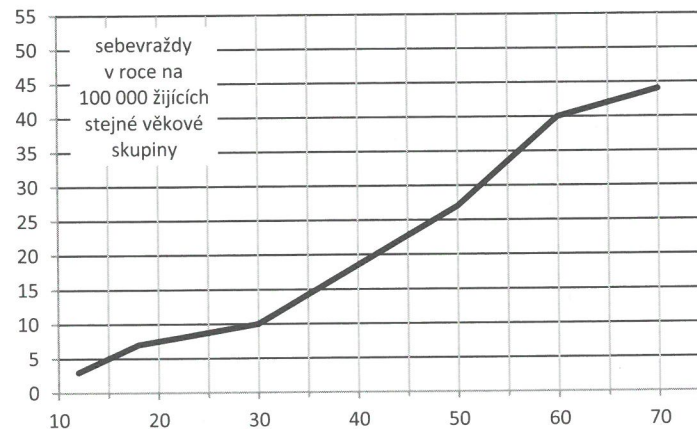
5.5 Tabulky a grafy

Statistické zpracování znamená především uspořádání a shrnutí informací, což se dělá nejdříve pomocí tabulek a grafů. Tabulky a grafy jsou vlastně přehlednou formou prezentování výsledků. Jejich výhodou je, že jsou úsporné a vyjádří více než dlouhé slovní pasáže. Z tohoto důvodu není třeba tabulky podrobně a zdlouhavě slovně popisovat.

Protože sebrané údaje můžeme přeskupovat a kombinovat nejrůznějšími způsoby, je možné vyrobit velké množství tabulek a grafů. Otázkou je, jaké má být jejich rozumné množství. Ve zprávě o výzkumu je třeba uvést podstatné výsledky vzhledem k cíli výzkumu, proto tabulek má být jen tolik, kolik se takových výsledků zjistilo.

Příliš velké množství tabulek má obvykle za úkol jen omráčit čtenáře. Někdy je též svědectvím o tom, že výzkumník se v údajích „ztratil“, neumí v nich najít správnou hierarchii, a proto uvedl všechno, co měl k dispozici. „Když jsou škálovací procedury používány ve spojení se systematickým empiricismem, zdá se, že neexistují žádné hranice počtu hodin, které mohou být intenzivně stráveny při hledání ničeho.“ (Willer, Willer, 1973)

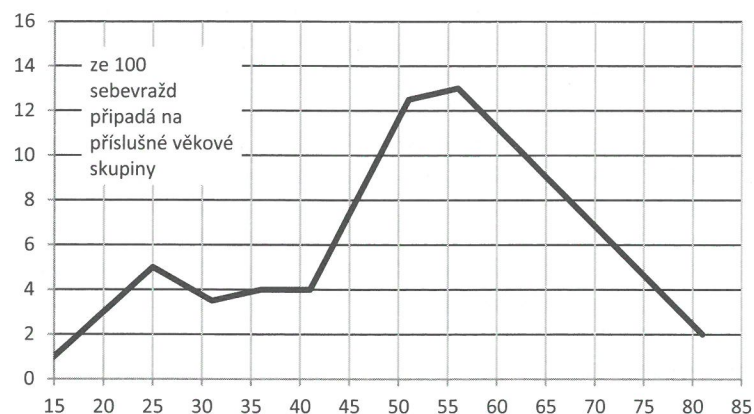
Dále uvedeme příklad, jak zavádějící mohou být grafy a že „obraz může lhat víc než tisíc čísel“ (Swoboda, 1977). Všechny tři následující grafy se týkají závislosti sebevraždění na věku.



Obr. 5.2 Závislost sebevraždění na věku I.

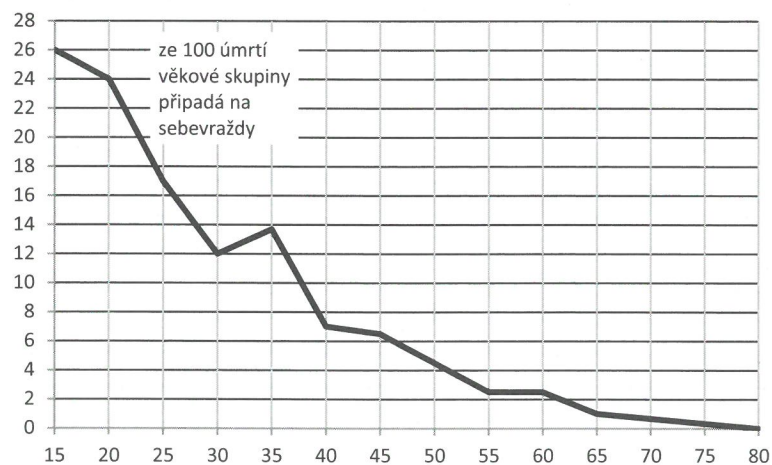
Zdroj: Swoboda, 1977

Obrázek 5.2 demonstruje „statisticky dokázané zjištění“, že sebevraždnost se zvyšuje s rostoucím věkem. Průběh křivky grafu zřetelně ukazuje, že počet sebevrahů se s věkem plynule zvyšuje a ve věkové skupině „70 a starší“ činí více než desetinásobek skupiny mladistvých.



Obr. 5.3 Závislost sebevraždnosti na věku II.

Zdroj: Swoboda, 1977



Obr. 5.4 Závislost sebevraždnosti na věku III.

Zdroj: Swoboda, 1977

Druhé „statisticky dokázané zjištění“: ve středním věku dosahuje sebevraždnost vrcholu. Vzácné jsou naproti tomu sebevraždy mezi mladými a nejstaršími. Průběh křivky grafu zřetelně ukazuje, že z každých 100 sebevražd připadá asi čtvrtina na věkovou skupinu 51 až 60, na sedmdesátileté (71–80) jen asi 10 % a na skupinu 21 až 30 let pouze 8 %.

Třetí „statisticky dokázané zjištění“: sebevraždnost je nemocí mládeže a její význam klesá s rostoucím věkem. Průběh křivky grafu zřetelně ukazuje, že počet sebevrahů na celkovém počtu úmrtí jedné věkové skupiny klesá z 23 % mezi mladými lidmi (15–25 let) na méně než 1 %, jakmile se věk blíží sedmdesátce.

Všechny tři grafy se týkaly sebevraždnosti, ale každý vztahoval počet sebevražd k jinému základu: 1. k žijícím stejné věkové skupiny, 2. k celkovému počtu sebevražd, 3. k počtu všech úmrtí v dané věkové skupině. Uvedené grafy se týkají tří různých zemí a různých období, nejsou tedy v jednotlivostech srovnatelné. Ale tím není dotčena podstata sdělení: v závislosti na volbě základu můžeme dostat absolutně rozdílné průběhy křivek.

5.6 Interpretace

Export výrobku A se loni zvýšil o 7 %, výrobku B ale o 700 %! Má vůbec A vůči raketovému vzestupu B nějakou šanci? Máme-li k dispozici kompletní údaje, vidíme, že ano (viz tab. 5.1).

Tab. 5.1 Srovnání exportu výrobku A a B

Výrobek	Předloni	Loni	Přírůstek	Přírůstek v %
A	400 000	430 000	30 000	7
B	2 500	20 000	17 500	700

Zdroj: modelová data

Ve světle tohoto příkladu je patrné, že i výroky typu „roční přírůstek se zvýšil o třetinu“ jasně patří k použitelnému statistickému arzenálu. Může totiž znamenat jak vzrůst z 0,3 % na 0,4 %, tak i z 60 % na 80 %, ale stále není jasné, zda je to hodně, nebo málo.

Vysokých procent přírůstku lze dosáhnout tím snadněji, čím nižší je výchozí základ. Dítě například během prvního roku svého života v průměru ztrojnásobí svou hmotnost a vyroste o 50 %. Kdyby tento proces takto pokračoval, vážil by člověk v patnácti letech 50 000 tun a byl by vysoký 200 metrů. Na začátku růstu jsou relativní přírůstky vysoké téměř vždy, ale pak se růst musí zpomalit.

Každý má rád procenta. Symbol „%“ jako by měl magickou moc. Reálnou představu o situaci můžeme však získat pouze tehdy, jestliže známe nejen procenta, ale i absolutní čísla. Kdo zamlčuje absolutní čísla a ohání se jen procenty, chce zřejmě něco skrýt.

Skrývání nemusí nutně znamenat zastírání neúspěchů, může jít pouze o snahu utajit jisté skutečnosti a k tomu se procenta skutečně výborně hodí.

Někdy se používá i postup opačný: místo procent se imponuje velkými absolutními čísly. Kdyby kupř. letošní sklizeň pšenice byla pouze o 0,1 % vyšší než loňská, určitě by lépe znělo prohlášení, že letos bylo sklizeno o 3 tisíce tun pšenice více než loni (při roční sklizni 3 milionů tun může být pravda obojí).

Často je užitečné uvažovat různé indexy: podělit uvažovanou charakteristiku podle okolností počtem obyvatel, dní, km² apod. Obraz je pak mnohem názornější. Jasnější představu o reálné situaci umožňují i různá mezinárodní srovnání. Viz Internetové zdroje.

V příkladech by bylo možno pokračovat, ale provést úplný výčet metod statistikulační nelze. Čtenář statistických údajů by si však měl zapamatovat, že bitva není nikdy vyhrána!

5.7 Závěr

Statistikulaci se pochopitelně hůře daří tam, kde může proudit volný tok informací, kde hlavním cílem je skutečně informovat, a ne pouze manipulovat. Když se v takovém prostředí někdo přece jen pokusí zneužít statistiky ve svůj prospěch, vystoupí okamžitě opozice či konkurence a uvede svůj pohled na věc. Z těchto různých pohledů má přemýšlející člověk vždy možnost si utvořit reálnější obraz skutečnosti.

Naopak statistikulační přímo kvete, když se s pravdivými informacemi zachází jako s privilegiem, jehož se dostává pouze těm, kteří skutečně nutně musí být informováni. Prostý občan – adresát statistických informací – nemá většinou skoro žádnou možnost, jak prověřit věrohodnost prezentovaných údajů. Je mu předložen výsledný produkt, do kuchyně ho nikdo nepustí, a tak nikdy nevidí, z jakých surovin a jakým způsobem byl připraven. Zcela bezmocný však také není. Může se připravovat tak, aby různé formy statistikulační snáze prohlédl a dokázal se jim lépe bránit. Existují i jiné návody, jak prohlédnout manipulaci a bránit se jí – viz Argumentační fauly (2012).

Nad předkládanými statistickými údaji je vždy užitečné položit si tyto základní otázky:

1. Kdo to říká?
2. Jak to ví?
3. Co chybí (absolutní čísla, nebo procenta)?
4. Co s čím bylo zaměněno?
5. Dává to smysl?
6. Cui bono — komu to slouží?

A to je i shrnutí a hlavní poučení celého našeho výkladu.

5.8 Literatura

- Argumentační fauly* (2012). Dostupné na <http://www.obcanskevzdelavani.cz/argumentacni-fauly>.
HENDL, J. (2012). *Přehled statistických metod. Analýza a metaanalýza dat*. 4. vyd. Praha: Portál.
HUFF, D. (2013). *Jak lhát se statistikou*. Praha: Brána.
KALOUS, J. (1989). Statistikulace. *Vesmír*, roč. 68, č. 7, s. 404–406.
MAGNELLO, E., VAN LOON, B. (2010). *Seznamte se ... Statistika*. Praha: Portál.
SWOBODA, H. (1977). *Moderní statistika*. Praha: Svoboda.
VESELÝ, A., KALOUS, J., MARKOVÁ, J. (2004). *Kultivace vědění v klíčový faktor produkce*. Praha: UK FSV CESES, Studie CESES.
WALKER, I. (2013). *Výzkumné metody a statistika*. Praha: Grada.
WILLER, D., WILLER, J. (1973). *Systematic Empiricism: Critique of a Pseudoscience*. Englewood Cliffs, New Jersey: Prentice-Hall.

Internetové odkazy

- ✓ NATION MASTER – World Statistics, Country Comparisons: <http://www.nationmaster.com/index.php>
- ✓ CIA – The World Fact book: <https://www.cia.gov/library/publications/the-world-factbook/index.html>
- ✓ WB – World Bank Data: <http://data.worldbank.org/>
- ✓ WB – Knowledge Assessment Methodology: <http://www.worldbank.org/kam>
- ✓ UNESCO Institute for Statistics: <http://www.uis.unesco.org>
- ✓ UNDP – Human Development Reports: <http://hdr.undp.org/en/data>
- ✓ ILO – Labour Office database on labour statistics: <http://laborsta.ilo.org/>
- ✓ OECD – Statistics: <http://www.oecd.org/statistics/>
- ✓ EUROSTAT – Statistical Office of the European Communities: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>