



TIM HARFORD

**AKO
SPOČÍTAŤ
SVET**

N

N

TIM HARFORD

AKO SPOČÍTAŤ SVET

TIM HARFORD

**AKO
SPOČÍTAT
SVET**

Tim Harford
How to Make the World Add Up

Copyright © 2020 by Tim Harford
Translation © Filip Pacalaj
Design & Layout © Miro Čevela
Slovak edition © 2021 by N Press, s. r. o.

ISBN 978-80-8230-008-9

Obsah

Úvod	Ako klamať štatistikami	11
Pravidlo č. 1	Skúmaj svoje pocity	29
Pravidlo č. 2	Nezabúdaj na osobnú skúsenosť	59
Pravidlo č. 3	Neponáhľaj sa s výpočtami	79
Pravidlo č. 4	Ustúp o krok a kochaj sa	101
Pravidlo č. 5	Vnímaj príbeh na pozadí	119
Pravidlo č. 6	Pýtaj sa, kto chýba	151
Pravidlo č. 7	Vyžaduj transparentnosť, keď počítač povie „nie“	171
Pravidlo č. 8	Nepovažuj štatistické podložie za samozrejmosť	205
Pravidlo č. 9	Nezabúdaj, že pekne môžu vyzerat' aj nepravdivé informácie	235
Pravidlo č. 10	Zachovaj si otvorenú myseľ	261
Zlaté pravidlo	Buď zvedavý	287
Poznámky		303
Podakovanie		323
Register		327

Úvod

Ako klamať štatistikami

Skutočne podstatný problém... nespočíva v tom, aby sme dokázali, že tvrdenie je nepravdivé, ale v dokázaní, že samotný predmet je daným predmetom.

Umberto Eco (1980)¹

Počuli ste, že bábätká nosí bocian?

Je to pravda.

Štatistiky nám to potvrdia.

Porovnajte si rozšírenie bocianov vo svete a potom si pozrite čísla o pôrodnosti. V európskych krajinách platí priama úmera – viac bocianov, viac bábätiok; menej bocianov, menej bábätiok.

Tento vzorec je dostatočne spoľahlivý na to, aby prešiel tradičným sitom pre publikovanie vo vedeckom časopise. Skutočne sa už stalo, že vyšiel vedecký článok s názvom „Bociany nosia deti ($p = 0,008$)“. Nemusíme byť príliš technicky nadaní, aby nám všetky tieto nuly jasne hovorili, že nejde o náhodu.²

Možno ste už uhádli, v čom je trik. Veľké európske krajiny ako Nemecko, Poľsko a Turecko sú domovom mnohých boci-

nov a rodí sa tu mnoho detí. Malé krajiny ako Albánsko a Dánsko majú málo detí aj bocianov. V údajoch síce badať jasnú súvislosť, táto súvislosť však neznamená, že výskyt bocianov je príčinou pribúdania detí.

Mohlo by sa zdať, že štatistikami sa dá „dokázať“ čokoľvek – ešte aj to, že bociany nosia deti.

Tento dojem by ste určite mali pri čítaní knihy *Ako klamať štatistikami* (*How to Lie with Statistics*). Táto umne cynická knižička, ktorú v roku 1954 napísal pomerne neznámy americký novinár na voľnej nohe menom Darrell Huff, mala okamžite nadšenú recenziu v *New York Times* a s miliónom predaných výtlačkov sa stala azda najpopulárnejšou knihou o štatistikách všetkých čias.

Kniha si popularitu a obdiv zaslúži. Je to skvost štatistickej komunikácie. Z Darrella Huffa sa vďaka nej stala legenda medzi štatistikmi. Ben Goldacre, epidemiológ a autor bestselleru s názvom *Zlá veda* (*Bad Science*), obdivne spomína „toho Huffa“, ktorý napísal „trhák“. Americký spisovateľ Charles Wheelan uvádza svoju knihu *Holá štatistika* (*Naked Statistics*) ako „poctu“ Huffovej „klasike“. Rešpektovaný časopis *Statistics Science* zorganizoval spomienkové stretnutie k päťdesiatročnici od vydania Huffovej knihy.

Cítil som to rovnako. Ako tínedžer som jeho knihu čítal s nadšením. Jasne a ostro písaný text sprevádzaný hravými karikatúrami mi dával možnosť nahliadnuť za oponu štatistickej manipulácie a vysvetľovaním rôznych podvodov budoval moju imunitu voči podobným lžiam.

Huff nešetří príkladmi. Začína úvahou o platových pomeroch absolventov Yale. Podľa prieskumu z roku 1950 mal ročník 1924 priemerný ročný príjem na úrovni dnešných 500 000 dolárov. Nie je to nepredstaviteľná suma – koniec koncov je to Yale. No pol milióna dolárov ročne je naozaj veľa peňazí. Je to reálny priemer?

Vôbec. Huff vysvetľuje, že toto „neuveriteľne priaznivé“ číslo vychádza z údajov, ktoré nahlasovali sami absolventi, z čoho

môžeme očakávať, že títo márnomyseľní boháči budú svoj príjem skôr nadhodnocovať. Prieskum navyše zachytil iba tých, ktorí sa obťažovali odpovedať – a iba tých, ktorých univerzita dokázala kontaktovať. A ktorí sú ľahko dohľadateľní? Bohatí a slávni. „Kde sú všetky tie zatúlané ovečky z grúňov Yale s poznámkou adresa neznáma?“ pýta sa Huff. Yale si starostlivo odsleduje milionárov spomedzi svojich absolventov, ale ich nenápadnejší kolegovia ľahko prepadnú okami týchto sietí. To všetko prispieva k premršteným výsledkom prieskumu.

Huff svižne prechádza širokou škálou štatistických zločinov, od reklám na zubné pasty založených na výskume, ktorý vyberal hrozienka z koláča, až po mapy meniace svoj význam podľa toho, ako ich vyfarbujete. Uzatvára slovami: „Podvodníci už tieto triky poznajú; čestní ľudia sa ich musia naučiť v sebaobrane.“

Pri čítaní knihy *Ako klamať štatistikami* bude rásť vaša skepsa k číslam a povedomie o možnostiach, ako s vami dokážu vypieť. Je to šikovná a poučná kniha.

Ja som však strávil vyše dekády pokusmi o komunikáciu štatistických poznatkov a o preverovanie numerických tvrdení – a v priebehu tých rokov mi tá kniha spolu so svojimi dôsledkami prekážala čoraz viac. Čo to vlastne hovorí o štatistike – a vlastne aj o nás –, ak je najúspešnejšia kniha v tomto odbore od začiatku do konca jedným veľkým varovaním pred dezinformáciami?

Darrell Huff napísal svoju knihu v roku 1954. V tom istom roku sa udiala ďalšia významná vec: dvaja britskí vedci, Richard Doll a Austin Bradford Hill, prišli s prvou presvedčivou štúdiou, ktorá dokazovala, že fajčenie spôsobuje rakovinu pľúc.³

Doll s Hillom by to nedokázali, nebyť štatistik. Výskyt rakoviny pľúc vo Veľkej Británii vzrástol v priebehu piatich rokov šesťnásobne; do roku 1950 zaujala Británia v týchto číslach neľichotivé svetové prvenstvo, pričom úmrtia na rakovinu pľúc

prvýkrát prevýšili počty úmrtí na tuberkulózu. Už tieto zistenia si vyžadovali štatistickú perspektívu. Keby bol dáta zberal každý doktor sám za seba, vyšiel by z toho azda len úsmevný obrázok.

No dokázať, že za to treba viniť práve cigarety, je bez štatistiky nemožné. Veľa ľudí si totiž myslelo, že za nárastom výskytu rakoviny pľúc je rozmach automobilizmu. Dokonale to sedelo. V prvej polovici dvadsiateho storočia sa automobil s výfukovými plynmi a čudnými dechtovými výparmi na nových asfaltových cestách stal bežnou súčasťou života. Zároveň s tým pribúdali prípady rakoviny pľúc. Dokázanie pravdy – teda toho, že za rakovinu môžu práve cigarety, a nie autá – si vyžadovalo viac než len letný pohľad. Potrebné bolo vedecké počítanie a starostlivé porovnávanie. Inými slovami, potrebná bola štatistika.

Na cigaretovú hypotézu sa mnohí pozerali skepticky, hoci nebola úplne nová. Napríklad veľké výskumné úsilie na dokázanie škodlivosti cigariet vynaložilo nacistické Nemecko; Adolf Hitler fajčením opovrhoval. Führer sa nepochybne potešil, keď mu nemeckí doktori priniesli dôkazy o tom, že cigarety spôsobujú rakovinu. Z pochopiteľných dôvodov však „nenávisť Nemcov“ nebola v Británii tým pravým argumentom proti tabaku.

Doll a Hill sa teda rozhodli spustiť vlastné štatistické pátranie. Richard Doll bol pohľadný, mierny a neskonale jemnocitný mladík. Z druhej svetovej vojny sa vrátil domov s hlavou plnou nápadov, akú revolúciu by mohla štatistika vyvolať v medicíne. Jeho mentor Austin Bradford Hill slúžil v prvej svetovej vojne ako pilot a neskôr nechýbalo veľa, a bol by zomrel na tuberkulózu. Hill bol charizmatiký ostrovtipný muž s povestou najprecíznejšieho medicínskeho štatistika dvadsiateho storočia.⁴ Ich spolupráca ako dátových detektívov mala viesť k záchrane ľudských životov.

* Chorobe sa Hill neskôr pomstil výskumom o liečení tuberkulózy, ktorý sa zapísal do dejín ako prvé rigorózne randomizované klinické testovanie.

Dvojica spustila štúdiu o fajčení a rakovine na samom začiatku roku 1948. Zamerali sa na dvadsať nemocníc v severozápadnom Londýne, ktoré mal na starosti Richard Doll. Za každým, keď nemocnica prijala pacienta s rakovinou, sestrička k nemu – náhodne – priradila iného pacienta rovnakého pohlavia a veku. Pacienta s rakovinou aj pacienta bez rakoviny sa potom detailne pýtali na pomery, v akých žijú, ich životný štýl, stravu a fajčiarsku históriu. Dotazníky postupne, týždeň po týždni, mesiac po mesiaci, pribúdali.

Doll prestal fajčiť v októbri 1949, čiže menej než dva roky po spustení štúdie. Mal tridsaťsedem a fajčil od ranej dospelosti. Spolu s Hillom totiž zistili, že tuhý fajčiar nemá dvojnásobné riziko, že dostane rakovinu pľúc; ani trojnásobné, dokonca ani štvornásobné. Mal šestnásobne vyššie riziko, že skončí s rakovinou.⁵

Hill a Doll výsledky publikovali v septembri 1950 a hneď sa pustili do ambicióznejšieho a dlhodobejšieho výskumu. Hill napísal každému doktorovi v Británii – bolo ich 59 000 – a požiadal ich, aby vyplnili dotazník o vlastnom zdravotnom stave a fajčiarskych návykoch. Doll s Hillom predpokladali, že práve doktori budú vedieť najlepšie odsledovať, koľko fajčia. Navyše ostanú v zdravotníckych registroch, vďaka čomu ich bude možné dohľadať aj v budúcnosti. A v momente doktorovej smrti je vysoká šanca, že mu bude správne stanovená diagnóza aj príčina úmrtia. Napokon im ostávalo už len čakať.

Na dotazník odpovedalo vyše 40 000 doktorov, hoci nie všetci boli nadšení. Musíte chápať, že fajčenie bolo v tom čase extrémne bežný zlozvyk a nikoho neprekvapilo zistenie, že medzi mužmi z prvej vzorky bolo až 85 % doktorov fajčiarov. Nikto nepočúva rád, že sa pomaly pripravuje o roky života, obzvlášť v prípade, ak je táto samovražedná metóda návyková.

Na jednej párty v Londýne sa na Hilla oboril istý doktor: „Ty si ten chlapík, ktorý chce, aby sme prestali fajčiť!“

„Ale vôbec nie,“ odpovedal Hill, v tom čase ešte fajčiar klasickej fajky. „Mňa len zaujíma, ako zomriete, keď budete fajčiť ďalej. A tiež ma zaujíma, ako zomriete, ak prestanete. Či budete pokračovať, alebo prestanete, si spokojne vyberte sám. Na tom mi vôbec nezáleží. Vašu smrť rovnako rád vyhodnotím v jednom aj druhom prípade.“⁶

Spomenul som, že Hill pôvodne študoval ekonómiu? Odtiaľ sa asi zobral jeho povestný jemnocit.

Doktorská štúdia trvala desaťročia, už po krátkom čase však nabrala dostatok dát, aby Doll s Hillom publikovali jasný poznatok: fajčenie spôsobuje rakovinu pľúc a čím viac fajčíte, tým je riziko vyššie. Ba čo viac – a toto bola novinka – fajčenie spôsobuje aj infarkt.

Doktori nezvyknú byť hlúpi. Keď v roku 1954 vyšiel článok zo štúdie v ich vlastnom odbornom časopise *British Medical Journal*, vedeli z toho sami pre seba vyvodit' dôsledky. Hill prestal fajčiť ešte toho roku a pridalo sa k nemu veľa priateľov a kolegov. Stali sa tak prvou veľkou identifikovateľnou sociálnou skupinou v Británii, ktorá sa hromadne vzdávala tohto zlozvyku.

V rovnakom roku sa tiež zrodili dva pohľady na štatistiku. Pre zástupy čitateľov Huffovej knihy *Ako klamať štatistikami* bola štatistika hrou plnou švindľovania a podvodov – a hlavnou zápravou pri nej bolo odhaľovanie nečestných ťahov. Austin Bradford Hill a Richard Doll ju nepovažovali až za takú roztopašnú. V ich štatistickej hre figurovali tie najvyššie stávky a ak sa hrala čestne a dobre, bola schopná zachraňovať životy.

Na jar 2020 – práve som dopĺňal posledné riadky tejto knihy – sa zrazu opäť celkom jasne ukázalo, že veľmi záleží na presných, včasných a čestných štatistikách. Nový koronavírus vyčíňal všade po svete. Politici boli nútení prijímať konsenzuálne rozhodnutia tak rýchlo ako už dávno nie. Mnoho týchto rozhodnutí záviselo od nadľudského úsilia pri detektívnej práci

epidemiológov, lekárskejších štatistikov a ekonómov. Hrozba vi-sela nad desiatkami miliónov životov. A tiež nad miliardovými sumami na ich živobytie.

Tieto slová píšem začiatkom apríla 2020: v krajinách po celom svete sa začínajú týždne obmedzení, počet globálnych úmrtí práve prekročil 60 000 a nie je jasné, ako sa bude príbeh vyvíjať*. Kým sa táto kniha dostane do vašich rúk, možno sa budeme topiť v najhlbšej hospodárskej kríze od 30. rokov dvad-siateho storočia a obeť budú pribúdať neskutočným tempom. Tiež je možné, že zafunguje ľudská vynaliezavosť či šťastie a ta-kéto apokalyptické obavy rýchlo vymiznú z našich spomienok. Otvára sa pred nami mnoho možných scenárov. A to je problém.

Epidemiológ John Ioannidis v polovici marca 2020 napísal, že Covid-19 „môže byť dôkazovým fiaskom storočia“⁷. Dátoví detek-tívi sa snažia zo všetkých síl – pracujú však s nepravidelnými, nekonzistentnými a žalostne nedostatočnými údajmi, ktoré nestačia na to, aby sme dokázali nadobudnúť istotu pri rozho-dovaní o živote a smrti.

Podrobnosti o fiasku sa nepochybne budú v nadvádzajú-cich rokoch vedecky skúmať. Sú však veci, ktoré sa už teraz zdajú jasné. Napríklad pri začiatku krízy to vyzeralo tak, že voľnému toku poctivých štatistik bránia politici – je to prob-lém, ktorému sa venuje ôsma kapitola. Taiwan sa sťažoval, že koncom decembra 2019 poskytol Svetovej zdravotníckej orga-nizácii (WHO) dôležité informácie o prenose z človeka na člo-veka. WHO však ešte v polovici januára upokojujúco tvrdovala, že Čína o tom nenašla nijaké dôkazy. (Taiwan nie je členom WHO, pretože Čína sa domáha zvrchovanosti nad týmto ostro-vom a požaduje, aby sa s ním nezaobchádzalo ako s nezávislým štátom. Je možné, že k spomínanému oneskoreniu viedla táto geopolitická prekážka.)⁸

* V čase ukončovania prekladu sú globálne úmrtia na čísle 2 600 000 a naďalej nie je jasné, aký vývoj máme očakávať. (pozn. prekl.)

Malo to význam? Takmer určite. Ak sa počet prípadov každé dva alebo tri dni zdvojnásobí, sotva sa niekedy dozvieme, v čom by bol vývoj iný, keby varovanie bolo prišlo už po pár týždňoch. Je zrejmé, že mnohým vedúcim funkcionárom trvalo nejaký čas, kým pripustili potenciálnu závažnosť hrozby. Napríklad prezident Trump koncom februára oznámil: „Zmizne to. Jedného dňa to z ničoho nič zmizne.“ O štyri týždne neskôr, keď zomrelo už 1 300 Američanov a USA sa v počte potvrdených prípadov dostali na čelo tabuľky, Trump naďalej dúfal, že na Veľkú noc otvoria kostoly v celej krajine.⁹

Kým píšem, všade sa živo debatuje. Môže rýchle testovanie, izolácia a sledovanie kontaktov eliminovať ohniská nákazy alebo iba spomalí jej šírenie? Mali by sme sa viac vyhýbať malým stretnutiam v interiéri alebo veľkým v exteriéri? Pomáha zatvorenie školy zabrániť šíreniu vírusu alebo škodí viac, keď deti posielame k ohrozenejším starým rodičom? Ako veľmi pomáha nosenie masiek? Na tieto a mnohé ďalšie otázky možno odpovedať iba na základe dobrých údajov o tom, kto a kedy bol infikovaný.

No obrovské množstvo infekcií oficiálne štatistiky nezaznamenali kvôli nedostatku testov. A vykonané testy poskytovali skreslený obraz, keďže zachytili hlavne zdravotnícky personál, kriticky chorých pacientov a – priznajme si to – bohatých a slávnych ľudí. V čase písania týchto riadkov nám údaje ešte nedokážu povedať, koľko z prípadov má len mierne alebo žiadne príznaky – nevieme teda, nakoľko je vírus v skutočnosti smrtiaci. Kvôli tomu, že počet obetí v marci exponenciálne stúpil a zdvojnásoboval sa každé dva dni, nám neostal čas na premýšľanie. Štátna moc uviedla ekonomiky do umelého spánku – koncom marca sa viac ako tri milióny Američanov uchádzali o príspevok v nezamestnanosti, čím päťnásobne prekonal posledný rekord. Nasledujúci týždeň prišiel s novým rekordom: bolo podaných ďalších šesť a pol milióna žiadostí. Boli potenciálne zdravotné následky skutočne dosť apokalyptické na to,

aby ospravedlnili, že toľko ľudí prišlo o príjem? Všetko tomu nasvedčovalo – ale epidemiológovia aj tie najlepšie odhady stali na výrazne limitovaných informáciách.

Fakt, že v bežnom živote považujeme presné a systematicky zhromažďované čísla za samozrejmosť, by sotva lepšie podčiarkla akákoľvek iná situácia. Masívne štatistiky o dôležitých problémoch, ktoré tu boli dávno pred koronavírusom, usilovní štatistici dlhé roky dôsledne zhromažďovali a často ich aj zadarmo sprístupňovali na stiahnutie odkiaľkoľvek na svete. Napriek tomu sme takým luxusom natoľko rozmaznaní, že šmahom ruky odmietame „klamstvá, hnusné lži a štatistiky“. Covid-19 nám pripomína, do akej zúfalej situácie sa môžeme dostať, keď štatistiky jednoducho nebudú k dispozícii.

Darrell Huff predstavil štatistiky ako kúzelnícky trik na scéne: jednoducho dobrá zábava, ktorú nikdy netreba brať vážne. Dávno pred koronavírusom ma začala mátať obava, že tento prístup nám dnes veľmi nepomôže. Stratili sme cit pre to, že práve štatistiky nám môžu pomôcť dobre spočítať svet. Problém nepredstavuje až tak veľmi to, že by sme každú štatistiku považovali za lož, ako skôr to, že strácame nádej na poznanie pravdy. Veríme teda čomukoľvek, čomu veriť chceme (viac hneď v nasledujúcej kapitole), a na zvyšok odpovedáme ako Huff: výsmechom, pokrčením pliec či obidvoma.

Takýto štatistický cynizmus nie je len hanbou – je to tragédia. Ak zoberieme do úvahy, že sme sa pripravili o schopnosť rozoznať pravdu, pripravili sme sa o základnú životnú funkciu. Je to nástroj, ktorý nám ukáže, že cigarety zabíjajú. Je to naša jediná skutočná šanca pri hľadaní cesty z koronavírusovej krízy – alebo, v širšom zmysle, pochopenie komplexného sveta, v ktorom žijeme. Nástroj je však nepoužiteľný, ak sa zamotáme v reflexívnom odmietaní akéhokoľvek štatistického tvrdenia, ktoré nám práve nevyhovuje. Samozrejme, nemali by sme byť zbytočne dôverčiví, ale protilátkou na prehnanú dôverčivosť nie je

to, že prestaneme veriť všetkému, ale to, že budeme seabavedome vyhodnocovať informácie a zapojíme pri tom jednak vlastnú zvedavosť a jednak zdravý skepticizmus.

Dobré štatistiky nie sú podvod, hoci istá mágia za nimi stojí. Dobré štatistiky nie sú zahmlená sála plná krivých zrkadiel; naopak, pomáhajú nám vidieť jasnejšie. Dobré štatistiky sú tým, čím je astronómovi ďalekohľad, virológovi mikroskop či rádiológovi röntgen. Ak im to dovolíme, pomôžu nám vnímať skutočnosti okolitého sveta aj poznávať seba samých – vo veľkom i v malom – tak, ako to nedokážeme nijakým iným spôsobom.

Mojím hlavným cieľom v tejto knihe je presvedčiť vás, aby ste si osvojili pohľad Dolla a Hilla, a nie Huffov cynizmus. Chcem vás presvedčiť, že štatistiky sa hodia na jasné a čestné osvetlenie reality. Aby sme to dosiahli, musím vám ukázať, že pre vlastnú potrebu dokážete použiť štatistické uvažovanie, keď budete pochybovať o tvrdeniach, ktoré počúvate v médiách, na sociálnych sieťach a v každodenných rozhovoroch. Chcem vás od piky naučiť vyhodnocovať tvrdenia, a čo je nemenej dôležité, naviesť, kde prípadne hľadať pomoc, ktorej sa dá dôverovať.

Dobrá správa je, že to bude zábava. Najväčšie uspokojenie pocítite, keď sa dostanete až na dreň štatistického príbehu: získate istotu a popri tom zapojíte svoju zvedavosť, až vás napokon naplní pocit, že ste zvládli niečo nové. Skôr budete veci *rozumieť*, než sa kdesi zboku na vec uškŕňať. Prístup Darrella Huffa je ako nezdravé jedlo: napohľad lákavé, ale potom dlho leží v žalúdku. A tomu sa chcete vyhnúť. Opakom štatistických bôčikov však nie je surový ovos a repa, ale výživné a príjemne rozmanité menu.

V tejto knihe som zhrnul to, čo som sa začal učiť od roku 2007, keď ma BBC požiadala, aby som uvádzal rozhlasový program s názvom *Viac či menej* (*More or Less*), reláciu o číslach v spravodajstve aj v každodennom živote. Tvorcovia relácie, novinár Michael Blastland a ekonóm sir Andrew Dilnot, mali