

Tim Harford  
ADATDETEKTÍV



**Tim Harford**

# ADATDETEKTÍV

10 egyszerű szabály,  
hogy ne tudjanak manipulálni

A fordítás alapja:  
Tim Harford: *How to Make the World Add up? – Ten Rules  
for Thinking Differently about Numbers*

Copyright © Tim Harford, 2020

Fordította: © Pétersz Tamás, 2021

Szerkesztette: Illényi Balázs

Borítóterv: Lobotworks

HVG Könyvek, Budapest, 2021  
Kiadóvezető: Budaházy Árpád  
Felelős szerkesztő: Dufka Hajnalka  
[www.hvgkonyvek.hu](http://www.hvgkonyvek.hu)

ISBN 978-963-565-030-9

Minden jog fenntartva.

Jelen könyvet vagy annak részleteit tilos reprodukálni, adatrendszerben tárolni, bármely formában vagy eszközzel – elektronikus, fényképes úton vagy más módon – a kiadó engedélye nélkül közölni.

Kiadja a HVG Kiadó Zrt., az 1795-ben alapított  
Magyar Könyvkiadók és Könyvterjesztők Egyesülésének tagja.  
Felelős kiadó: Szauer Péter

Nyomdai előkészítés: Sörfőző Zsuzsa  
Nyomás: Generál Nyomda Kft.  
Felelős vezető: Hunya Ágnes

*Minden tanárnak, különösképpen sajátjaimnak;  
Peter Sinclair kedves emlékére*



# Tartalom

<i>Bevezetés: Hogyan hazudjunk a statisztikával?</i>	9
<i>Első szabály: Vizsgáljuk meg az érzéseinket!</i>	27
<i>Második szabály: Töprengjünk el személyes tapasztalatainkon!</i>	54
<i>Harmadik szabály: Kerüljük az elhamarkodott adatelemzést!</i>	72
<i>Negyedik szabály: Lépünk hátra, és élvezzük a rálátást!</i>	92
<i>Ötödik szabály: Ismerjük meg a háttértörténetet!</i>	108
<i>Hatodik szabály: Kérdezzük meg, ki hiányzik!</i>	138
<i>Hetedik szabály: Követeljük meg az átláthatóságot, ha az algoritmus téved!</i>	156
<i>Nyolcadik szabály: Ne tekintsük maguktól értetődőknek a statisztikai alapokat!</i>	188
<i>Kilencedik szabály: Ne feledjük, hogy a félrevezető információ is lehet mutató!</i>	217
<i>Tizedik szabály: Mindig gondolkodjunk nyitottan!</i>	241
<i>Az aranyszabály: Legyünk kíváncsiak!</i>	266
Jegyzetek	281
Köszönetnyilvánítás	313
Név- és tárgymutató	315
A szerzőről	327





## BEVEZETÉS

---

# Hogyan hazudjunk a statisztikával?

A valódi, tényleges probléma... nem az, hogy valamiről bebizonyítsuk, hogy nem igaz, hanem hogy bebizonyítsuk: a vizsgálat eredeti tárgya valóban eredeti.

*Umberto Eco*<sup>1</sup>

Ismerjük a régi mesét arról, hogy a gólya hozza a kisbabákat. És ez tényleg igaz! Statisztikai adatokkal tudom bizonyítani.

Nézzük meg, hogy a becslések szerint mekkora gólyapopuláció él az egyes országokban, majd azt, hány gyerek születik ott évente. Európa-szerte figyelemreméltóan szoros kapcsolatot találunk e két adat között. Több gólya, több gyerek; kevesebb gólya, kevesebb gyerek.

Az összefüggés elég szoros ahhoz, hogy könnyedén megugorja a lécet egy tudományos folyóiratban való publikációhoz. Sőt „Storks Deliver Babies ( $p = 0,008$ )” (A kisbabákat a gólya hozza) címmel meg is jelent már egy tudományos cikk. Anélkül, hogy nagyon belemenénk a szakmai részletekbe, azok a nullák arról árulkodnak, hogy nem véletlen egybeesésről van szó.<sup>2</sup>

De mostanra alighanem rá is jöttünk a trükkre. A nagy európai országokban, például Németországban, Lengyelországban és Törökországban, sok a gyerek és sok a gólya. A kis országokban, mint például Albánia és Dánia, mindkettőből kevés van. Bár az adatokban könnyen felfedezhető valamiféle összefüggés, ez nem jelenti azt, hogy a csecsemők a gólyáknak köszönhetően érkeznek.

A jelek szerint tehát statisztikával minden „bizonyítható”; még az is, hogy a babákat a gólya hozza. Bizonyára ez lett volna a benyomásunk, ha elolvassuk a *How to Lie with Statistics* (Hogyan hazudjunk

a statisztikával?) című könyvet.<sup>3</sup> Darrell Huff kevésbé ismert, szabadúszó amerikai újságíró 1954-ben kiadott szellemes, cinikus kis műve azonnal lelkes dicséretet kapott a *The New York Times*-ban, és jóval több mint egymillió eladott példányával a statisztikáról valaha megjelent talán legnépszerűbb könyv vált belőle.

Meg is érdemli a népszerűséget és a dicséretet. A statisztika értelmezésének és kommunikációjának valóságos csodája. Ezenkívül Darrell Huffot szinte legendává tette a szakmabeliek körében. A *Rossz tudomány – Hatástalan szerek, gátlástalan kampányok (Bad Science: Quacks, Hacks, and Big Pharma Flacks)* című bestseller<sup>4</sup> epidemiológus szerzője, Ben Goldacre szerint „a Huff” igazi telitalálat. Charles Wheelan amerikai író *Naked Statistics: Stripping the Dread from the Data* (Csupasz statisztika – Hogyan oszlassuk el adatokkal a félelmet?) című könyvét<sup>5</sup> Huff „klasszikusa” előtti „tisztelgésnek” nevezi. A tekintélyes *Statistical Science* folyóirat a *How to Lie with Statistics* megjelenésének ötvenedik évfordulóján megemlékezést szervezett Huff munkásságáról.

Annak idején én magam is ugyanígy éreztem. Kamaszkoromban imádtam a *How to Lie with Statistics*-et. A sziporkázó humorral, metsző gúnnyal megírt és játékos rajzokkal illusztrált könyv bevezetett a statisztikai manipuláció kulisszái mögé, és megmutatta, hogyan zajlik az átverés, nehogy még egyszer bedőljek neki.

Huff rengeteg példát sorol fel. Először is azt vizsgálja, mennyi pénz keresnek a Yale Egyetemen végzett egykori diákok. Egy 1950-es felmérés szerint az 1924-ben végzettek éves átlagjövedelme mai árfolyamon számolva közel 500 ezer dollár volt. Ez hihetően hangzik – végül is a Yale-ről beszélünk –, de évi félmillió dollár mégiscsak rengeteg pénz. Tényleg ez volna az átlag?

Dehogy! Huff magyarázata szerint ez a „valószínűtlenül hízelgő” szám önbevallásból származik, vagyis joggal feltételezhetjük, hogy a megkérdezettek hiúságból eltúlozzák a jövedelmüket. A felmérés ráadásul csak azokra terjed ki, akik hajlandók voltak válaszolni, és csak azokra a volt diákokra, akiket a Yale megtalált. És vajon kiket könnyű megtalálni? A híreket és gazdagokat. „Kik azok a kicsiny, elveszett báránycsapatok, akiknek a neve mellett a Yale lajstromában a »lakcím ismeretlen« megjegyzés szerepel?” – teszi fel a kérdést Huff. Az egyetem

nyilván a milliommossá vált hallgatóit követi nyomon, a „futottak még” kategóriába tartozó öregdiákok egy részéről viszont könnyen megféleldekezhet. Mindez pedig azt jelenti, hogy a felmérés durván eltúlzott eredményt ad.

Huff ezután beleveti magát a különféle statisztikai bűncselekmények erdejébe, az alaposan megválogatott kutatásokon alapuló fogkrémreklámoktól az olyan térképekig, amelyekből más és más következtetéseket vonhatunk le attól függően, hogyan színezzük ki őket. Ahogy ő maga fogalmazott: „A csalók már ismerik ezeket a trükköket; az önvédelem érdekében azonban a becsületes embereknek is meg kell tanulniuk őket.”

Ha elolvassuk a *How to Lie with Statisticset*, jóval szkeptikusabbá válunk a számadatokkal kapcsolatban. Okos és nagyon tanulságos könyv. Mivel azonban több mint egy évtizede próbálkozom statisztikai elemzések kommunikálásával és számszerű állítások ellenőrzésével, az évek során egyre nagyobb nyugtalanság töltött el a szóban forgó könyvecske, mindannak kapcsán, amit képvisel. Mit árul el a statisztikáról – és *saját magunkról* – az, hogy a témában írt legsikeresebb mű az elejétől a végéig a félretájékoztatás elleni figyelmeztetés?

Ugyanabban az évben, 1954-ben, amikor Darrell Huff megjelentette a *How to Lie with Statisticset*, azonban történt valami más is. Két brit kutató, Richard Doll és Austin Bradford Hill elvégezte az egyik első meggyőző kutatást annak bizonyítására, hogy a cigarettázás tüdőrákot okoz.<sup>6</sup>

A tudospáros erre a statisztika nélkül nem jött volna rá. A tüdőrákos esetek aránya mindössze tizenöt év alatt hatszorosára emelkedett az Egyesült Királyságban; 1950-re már a legmagasabb volt a világon, és a tüdőrák okozta halálesetek száma először haladta meg a tuberkulózis miatti elhalálozásokét. Már ennek felismeréséhez is statisztikai perspektívára volt szükség. Az egyes orvosoknak csak felületes benyomásuk lehetett az egész jelenségről.

Annak megállapításához pedig, hogy ezért a cigaretta a felelős, megint csak nélkülözhetetlen volt a statisztika. Sokan úgy gondolták,

hogy a tüdőrák elharapózását az autók okozzák, és ez teljesen jogos feltételezés volt. A 20. század első felében vált mindennapossá az autók jelenléte, a kipufogógázzal és az új utakba kevert kátrány furcsán kellemes szagával együtt. A tüdőrákos esetek száma ugyanebben az időben kezdett emelkedni. Az igazság megállapításához – vagyis hogy nem az autók, hanem a cigaretta okozza a betegséget – nem volt elég annyi, hogy valaki egyszerűen körülnézzen. A kutatóknak gondos számolgatásba és összehasonlításba kellett kezdeniük. Röviden szólva, a statisztikához kellett nyúlniuk.

A cigarettaelméletet sokan kételkedve fogadták, pedig nem volt teljesen új keletű. A náci Németországban például nagyarányú kutatást végeztek annak érdekében, hogy bebizonyíthassák a cigaretta veszélyességét; Adolf Hitler ugyanis megvetette a dohányzást. A Führer nyilván nagyon elégedett volt, amikor a német tudósok megállapították, hogy a cigaretta rákot okoz. „A nácik utálják” címke azonban érthető okokból nem csökkentette a dohány népszerűségét a szigetországban.

Doll és Hill tehát úgy döntöttek, hogy saját statisztikai nyomozásba kezdenek. Richard Doll jóképű, csendes és minden körülmények között udvarias fiatalember volt. Amikor hazatért a II. világháborúból, rengeteg ötlete volt, hogyan forradalmasíthatná a statisztika az orvostudományt. Mentora, Austin Bradford Hill pilótaként szolgált az I. világháborúban, mielőtt kis híján elvitte a tuberkulózis.<sup>7</sup> Karizmatikus, éles eszű férfiú volt, és a 20. század legnagyobb orvosi statisztikusának tartották.<sup>8</sup> Közös adatnyomozói munkájuk életmentőnek bizonyult.

A páros első, dohányzással és rákkal kapcsolatos kutatása 1948. újév napján vette kezdetét. Húsz északnyugat-londoni kórházra összpontosítottak, és a munkát Richard Doll vezette. Valahányszor rákos beteg érkezett a kórházba, az ápolók véletlenül kiválasztottak az intézményben egy másik, ugyanolyan nemű és hasonló korú beteget. Ezután a rákbeteget és a másik páciens is alaposan kifaggatták arról, hol él és dolgozik, milyen az életmódja és az étrendje, valamint milyenek a dohányzási szokásai. Heteken, hónapokon át érkeztek az eredmények.

Kevesebb mint két évvel a felmérés kezdete után, 1949 októberében Doll abbahagyta a dohányzást. Harminchét éves volt ekkor, és

egész addigi felnőtt életében cigarettázott. Hill-lel ugyanis felfedezték, hogy az erős dohányzás nem csupán megduplázza, megháromszorozza vagy negyedszeresíti a tüdőrák kockázatát. A dohányosnál tizenhatszor akkora eséllyel alakul ki a betegség.<sup>9</sup>

Hill és Doll 1950 szeptemberében tették közzé az eredményeiket, és rögtön egy nagyobb, hosszabb távú és nagyratörőbb kísérletbe fogtak. Hill írt az Egyesült Királyság összes orvosának – mind az 59 600-nak –, és megkérte őket, hogy töltsenek ki egy „kérdőívet” az egészségi állapotukról és dohányzási szokásaikról. Doll és Hill úgy gondolták, az orvosok pontosan tudják követni, mit és mennyit szívnak. Ezenkívül mindig szerepelni fognak az országos nyilvántartásban, így bármikor könnyen elérhetik őket. Ha pedig egy orvos meghal, jó diagnózist várhatnak a halál okáról. Dollnak és Hillnek ezután nem kellett egyebet tennie, mint várni.

Több mint 40 ezer orvos válaszolt Hill megkeresésére, de nem mindegyikük volt elragadtatva a dologtól. Ehhez tudnunk kell, hogy a dohányzás akkoriban nagyon elterjedt volt, és senkit sem lepett meg, hogy az eredeti mintában szereplő férfi orvosok 85 százaléka hódolt ennek a szenvedélynek. Márpedig senki sem veszi jó néven, ha közlik vele: lehet, hogy lassan megöli magát; különösen akkor, ha az öngyilkosság módja ennyire addiktív.

Az egyik orvos egy londoni társasági eseményen számon is kérte ezt Hillen.

– Maga az a fickó, aki azt akarja, hogy hagyjuk abba a dohányzást?! – szegezte neki a kérdést.

– Egyáltalán nem – felelte Hill, aki akkor még maga is pipázott. – Csupán érdekel, hogyan hal meg, ha folytatja a dohányzást. De akkor is érdekel, hogyan hal meg, ha abbahagyja. Szóval ön dönti el, abbahagyja-e vagy folytatja. Ez teljességgel közömbös a számomra. A halálát így is, úgy is fel fogom jegyezni.<sup>10</sup>

Említettem már, hogy Hill eredetileg közgazdász volt? Ott tanulta ezt az elbűvölő stílust.

Az orvosokkal végzett kísérlet még évtizedeken át tartott, de Dollnak és Hillnek hamarosan elegendő adat állt a rendelkezésére, hogy egyértelmű konklúziót vonhassanak le: a dohányzás tüdőrákot okoz,

és minél többet dohányzik valaki, annál magasabb a kockázata, hogy a betegség kialakul. Sőt kiderült – és ez újdonság volt –, hogy a dohányzás szívrohamhoz is vezethet.

Az orvosok nem ostobák. Amikor a kutatás eredménye 1954-ben megjelent saját szakmai folyóiratukban, a *The British Medical Journal*-ben, ki-ki levonhatta magának a tanulságot. Hill még abban az évben abbahagyta a dohányzást, és sok orvoskollégája is így tett. Ők lettek az első, jól behatárolható társadalmi csoport az Egyesült Királyságban, amelynek tagjai nagy számban szoktak le erről a szenvedélyről.

A statisztikának tehát egy időben két különböző aspektusa is megmutatkozott 1954-ben. Darrell Huff *How to Lie with Statistics*ének számos olvasója szemében a statisztika csalókkal és csalásokkal teli játszma volt, a trükköző gazemberek leleplezése pedig szórakoztató időtöltés. Austin Bradford Hill és Richard Doll azonban egyáltalán nem tekintette nevetségesnek ezt a tudományágot. Az ő játszmájuk ugyanis az elképzelhető legnagyobb térről szólt, és ha jól és tisztességesen játszották, életeket menthetett.

2020 tavaszára – amikor az utolsó simításokat végeztem ezen a könyvön – hirtelen minden eddiginél világosabbá vált, mekkora jelentősége van a pontos, időben érkező és tisztességes statisztikának. A világon a koronavírus egy új változata söpört végig. A politikusoknak az elmúlt évtizedek legsúlyosabb döntéseit kellett meghozniuk, mégpedig gyorsan. E döntések közül több is adatnyomozói munkát igényelt, amelyet epidemiológusok, orvosi statisztikusok és közgazdászok igyekeztek elvégezni. Emberek tízmillióinak élete és milliárdjainak megélhetése foggott kockán.

Ezeket a sorokat 2020 áprilisának elején írom: a világ országaiban néhány hete karantént vezettek be, a halottak száma nemrég lépte át a 60 ezret, és még távolról sem világos, hogyan alakul majd a helyzet. Előfordulhat, hogy mire ez a könyv az olvasók kezébe kerül, az 1930-as évek óta nem tapasztalt gazdasági válságban vergődünk, az áldozatok száma pedig megsokszorozódik. Az is lehet, hogy az emberi zsenialitásnak vagy a jó szerencsének köszönhetően az ilyen apokalip-

tikus félelmek már csak rossz emlékként élnek majd bennünk. Sokféle forgatókönyv elképzelhető. És éppen ez a probléma.

John Ioannidis epidemiológus március közepén azt írta, hogy a Covid-19 „évszázadonként egyszer előforduló bizonyítékfiaskó lehet”.<sup>11</sup> Az adatnyomozók ugyan minden tőlük telhetőt megtesznek, de hiányos és egymásnak ellentmondó adatokkal kénytelenek dolgozni, amelyek sajnos elégtelenek ahhoz, hogy a kívánt magabiztossággal hozzassunk döntéseket élet-halál kérdésében.

A kudarcot minden bizonnyal évekig tanulmányozzák majd. De bizonyos jelenségek már most egyértelműnek tűnnek. A válság kezdetén például a politikusok mintha akadályozták volna a hiteles statisztikai adatok szabad áramlását; ehhez a problémához a nyolcadik fejezetben még visszatérünk. Tajvan arról panaszkodott, hogy 2019 decemberének végén fontos információkat szolgáltatottak a betegség emberről emberre való terjedéséről az Egészségügyi Világszervezetnek (World Health Organization, WHO), a szervezet azonban még január közepén is azzal nyugtatta a közvéleményt a Twitteren, hogy Kína nem talált bizonyítékot az emberről emberre való terjedésre. (Tajvan nem tagja a WHO-nak, mert Kína jogot formál a szigetre, és rossz néven veszi, ha független államként kezelik. Lehetséges, hogy ez a geopolitikai konfliktus vezetett az állítólagos késlekedéshez.)<sup>12</sup>

Vajon számított ez? Majdnem biztosan; mivel az esetek száma kétféle naponként megduplázódott, sohasem fogjuk megtudni, mi történt volna másképpen, ha két héttel hamarabb figyelmeztetik a világot a veszélyre. Sok vezetőnek időbe telt felmérni a fenyegetés súlyosságát. Donald Trump amerikai elnök például február végén a következőképpen fogalmazott: „El fog tűnni. Egy szép napon, mint egy csoda, el fog tűnni.” Négy héttel később, amikor már ezerháromszáz amerikai halt meg a vírus miatt, és az Egyesült Államokban minden más országnál több igazolt esetet tartottak nyilván, Trump elnök még mindig reménykedve arról beszélt, hogy húsvétkor már mindenki mehet templomba.<sup>13</sup>

Miközben ezeket a sorokat írom, világszerte heves viták dúlnak. Vajon a gyors teszteléssel, az elkülönítéssel és a kontaktkutatással háttározatlan időre kordában tartható-e a járvány, vagy csak lassíthatjuk

a terjedését? A kis, zárt térben tartott vagy a nagy, szabadtéri összejövetelek miatt kell-e jobban aggódnunk? Megakadályozza-e a vírus terjedését az iskolák bezárása, vagy nagyobb kárt okoz, ha a gyerekeket a veszélyeztetett nagyszülőkhöz küldik? Mennyit segít a maszk viselése? Ezekre és sok más kérdésre csak akkor válaszolhatunk, ha megbízható adatok állnak rendelkezésünkre arról, ki és mikor fertőződött meg.

A fertőzések nagy részét azonban a tesztelés hiánya miatt semmilyen hivatalos statisztikában nem regisztrálták. Az elvégzett tesztek pedig torz képet adtak, mivel az egészségügyi dolgozókra, a súlyos betegekre és – valljuk be – a híres és gazdag emberekre korlátozódtak. E sorok írása idején az adatokból egész egyszerűen nem állapítható meg, hány enyhe vagy tünetmentes eset van, így az sem, valójában mennyire halálos a vírus. A halálozási ráta márciusban exponenciális növekedésnek indult, kétnaponta duplázódott, ezért nem volt idő rá, hogy várjunk, míg tisztábban látunk. A vezetők mesterséges kómába küldték országaik gazdaságát; március végén egyetlen hét leforgása alatt több mint hárommillió amerikai igényelt munkanélküli-segélyt, ami ötszörösen múlta felül az addigi rekordot. A következő hét még ennél is rosszabbul alakult: további hat és fél millió igénylést nyújtottak be. Vajon valóban akkora egészségügyi katasztrófa fenyegetett, ami indokoltta tette ennyi ember jövedelmének lenullázását? Úgy tűnt, hogy igen – ámbár az epidemiológusok a nagyon korlátozott információk miatt csak találgathattak.

Nehéz volna elképzelni olyan helyzetet, amely ennél drámaibb módon illusztrálja, általában mennyire maguktól értetődőnek tekintjük a pontos, szisztematikusan összegyűjtött számadatok meglétét. A koronavírus megelőzően a szorgalmas statisztikusok éveken át aprólékos gondnal gyűjtötték rengeteg fontos jelenség statisztikai adatait, és ezeket gyakran szabadon hozzáférhetővé is tették bárhol a világon. Ez a luxus azonban elkényeztetett bennünket, és hajlamosak vagyunk olyasféle jópofaságokkal dobálózni, mint hogy a hazugságnak három fajtája van: a hazugság, a szemenszedett hazugság és a statisztika. A Covid-19 esete azonban rávilágít, hogy milyen kiszolgáltatottá válunk, ha ezek a statisztikák hiányoznak.



Darrell Huffnak köszönhetően a statisztika színpadi bűvésztükknek tűnik: jó móka, de sohasem szabad komolyan vennünk. Már jóval a koronavírus előtt aggódni kezdtem, hogy ez a hozzáállás nem igazán válik a javunkra. Elvesztettük a hitünket abban, hogy a statisztika érthetővé teszi a világot. Nem arról van szó, hogy minden statisztikát hazugságnak tekintünk, hanem arról, hogy úgy érezzük, nem tudjuk kiszűrni az igazságot. Így hát elhisszük, amit el akarunk hinni (erről bővebben a következő fejezetben), minden egyéb esetben pedig Huff reakcióját alkalmazzuk: gúnyos kacaj, vállrándítás, vagy mindkettő.

Ez a statisztikával szembeni cinizmus nemcsak hiba – egyenesen tragédia. Ha beletörődünk, hogy nem áll hatalmunkban megállapítani, mi az igazság, akkor nagyon fontos eszközt veszítünk el. Ez az eszköz mutatta meg nekünk, hogy a cigarettázás halálos. Ez az eszköz lehet az egyetlen igazi esélyünk rá, hogy kiutat találjunk a koronavírus-válság útvesztőjéből, vagy – tágabb értelemben – arra, hogy megérthessük a bonyolult világot, amelyben élünk. Ez az eszköz azonban használhatatlanná válik, ha rászokunk, hogy minden nemkívánatos statisztikai állítást csípőből elutasítsunk. Persze az sem jó, ha naivak és hiszékenyek vagyunk. A hiszékenység ellentéte azonban nem az, amikor nem hiszünk el semmit, hanem ha megvan bennünk a szükséges magabiztosság, hogy kíváncsian és egészséges kétkedéssel értékeljük a kapott információkat.

A jó statisztika nem trükk, bár bizonyos értelemben valóban varázslat. A jó statisztika nem csak ködösítés; sőt inkább segít, hogy jobban lássunk. A jó statisztika olyan, mint a csillagász távcsöve, a bakteriológus mikroszkópja vagy a radiológus röntgenkészüléke. Ha hajlandók vagyunk megengedni neki, akkor a jó statisztika segít olyan – kicsi és nagy – dolgokat meglátnunk a minket körülvevő világgal és saját magunkkal kapcsolatban, amelyeket másképp nem vennénk észre.

Ezzel a könyvvel mindenekelőtt azt szeretném elérni, hogy Huff cinizmusa helyett Doll és Hill felfogását tegyük magunkévá. Szeretnék meggyőzni mindenkit, hogy a statisztika igenis képes tisztán és őszintén rávilágítani a valóságra. Ehhez meg kell mutatnom, hogy magunk is alkalmazhatjuk a statisztikai érvelést, és mérlegre tehetjük a hozzáunk a médián, a közösségi oldalakon és a hétköznapi beszélgetéseken

keresztül eljutó információkat. Segíteni szeretnék abban, hogy előképzettség nélkül is mérlegre tehesük a különböző állításokat, és – ami ugyanilyen fontos – megtaláljuk ehhez a megbízható segítséget.

A jó hír, hogy mindez kifejezetten szórakoztató lesz. A statisztikai sztori mélyére nézni igazi örömet rejt: közben magabiztosabbá válunk, kielégítjük a kíváncsiságunkat, a végén pedig úgy érezzük majd, hogy magas szinten elsajátítottunk valamit. Megértjük a dolgokat, ahelyett hogy a pálya széléről gúnyolódnánk. Darrell Huff megközelítése olyan, mint a gyorskaja: elsősre felszínes módon megragad bennünket, egy idő után azonban unalmassá válik. Ez pedig árt nekünk. A statisztikai gyorskaja ellentéte azonban nem a nyers zab és petrezselyem, hanem a finom és ínycsiklandozóan változatos étrend.

Ebben a kötetben azt fogom leírni, amit 2007 óta magam tanultam, amikor a BBC felkért a hírekben és az életben felbukkanó számadatokról szóló „More or Less” című rádióműsor vezetésére, miután a műsor megalkotói, Michael Blastland újságíró és Sir Andrew Dilnot közgazdász távoztak. Nem voltam annyira alkalmas a szerepre, mint azt a BBC alighanem gondolta, én ugyanis nem statisztikával, hanem gazdaságelmélettel foglalkoztam. És bár ez a tudás adott némi magabiztosságot a számok világában, ez inkább védekező jellegű volt: megtanultam, hogyan fedezzem fel a hibákat és a trükköket, de ennél sokkal többre nem voltam képes.

Ezzel kezdődött a Darrell Huff nézőpontjától való eltávolodásom.

A munkatársaimmal hétről hétre górcső alá vettük a politikusok szájából elhangzó vagy az újságok címlapján virító statisztikai állításokat. Ezek az állítások gyakran feszegették az igazság határait, de egy egyszerű tényellenőrzés önmagában valahogy sohasem tűnt kielégítő válasznak. Arra jöttünk viszont rá, hogy minden állítás mögött – legyen az igaz, hamis vagy a kettő határán egyensúlyozó – lenyűgöző világ rejlik, arra várva, hogy felfedezzük és megmagyarázzuk. Akár a stroke gyakoriságát elemeztük, akár azt az állítást, hogy az adósság árt a gazdasági növekedésnek, vagy éppen azt, hányszor szerepel *A hobbit* vagy *Oda-vissza (The Hobbit: There and back Again)*<sup>14</sup> című könyvben

nőnemű személyes névmás, a számok éppúgy érthetőbbé teheték a világot, ahogy ködösítésre is alkalmasak voltak.

Mint arra a koronavírus-járvány olyan kíméletlenül felnyitotta a szemünket, a döntéseink meghozatalához egyénként, szervezetként és társadalomként is megbízható számokra van szükségünk. És ahogy az a vírus esetében is történt, a statisztikák összegyűjtése gyakran csak akkor történik meg, amikor már szembesültünk a válsággal. Vegyük például a munkanélküliségi mutatót, amely azt jelzi, hányan vannak, akik szeretnének dolgozni, de nincs hol. Ez ma már alapvető információ minden kormány számára ahhoz, hogy átláthassa a gazdaság helyzetét, 1920-ban azonban senki sem tudta volna megmondani, hányan keresnek állást.<sup>15</sup> A kormányok csak akkor kezdtek erre vonatkozó adatokat gyűjteni, amikor a súlyos válságok miatt a kérdés politikai jelentőséget kapott.

Kitágult és ijesztően bonyolult világunk tele van olyan kérdésekkel, amelyekre csak a számok gondos vizsgálata után válaszolhatunk. Vajon boldoggá vagy szomorúvá tesz bennünket a Facebook, és megállapítható-e, miért reagálunk rá különbözőképpen? Hány fajt fenyeget a kihalás, ez hányad része az összesnek, és az oka az éghajlatváltozás, a mezőgazdaság terjeszkedése, vagy valami teljesen más tényező? Gyorsul vagy lassul-e az emberi innováció? Mennyire jelentős hatást gyakorol az opioidválság az amerikaiak egészségére? Visszaszorulóban van-e az alkoholfogyasztás a tizenévesek körében, és ha igen, miért?

Egyre kényelmetlenebbül éreztem magam, amikor a „More or Less” rajongói rendszeresen gratuláltak hozzá, hogy „leleplezzük a hamis statisztikákat”. Persze, tényleg ezt csináltuk, és remekül szórakoztunk. Közben azonban szép lassan megtanultam, hogy az igazi örömet nem az jelenti, hogy megcáfoljuk a hazugságokat, hanem az, hogy megértjük, mi az, ami igaz.

A „More or Less” készítése közben arra is rájöttem, hogy a józan ész meglepően nagy segítséget jelent az adatdetektív munkájában. Ebben a könyvben ezeket az elveket foglalom össze. A kutatók és producerek nagy része, hozzám hasonlóan, nem rendelkezett semmilyen komolyabb képzettséggel az adatok kezelésére vonatkozóan. Ám egy-két egyszerű kérdés – kiegészítve némi internetes keresgéssel – még

a nagyon szakmai területeken is sokszor hasznos válaszokat hozott. Bár néha jól jött volna a valódi statisztikai tudás, a helyes kérdések feltevéséhez sohasem volt szükségünk rá. Ahogyan az olvasóimnak sem lesz.

Nem sokkal 1953 karácsonya előtt a nagy dohányipari cégek vezetői összegyűltek a New York-i Plaza Hotelben. Doll és Hill nagyszabású tanulmánya csak egy évvel később jelent meg, a cigarettagyártók azonban már akkor érezték, hogy tudományos szempontból nem áll jól a szénájuk. Azért ültek le tanácskozni, hogy kitalálják, miként reagáljanak a fenyegető válságra.

A válaszuk – sajnos – zseniális volt, és a propagandisták számára azóta is iránymutató.

Elkezdtek összezavarni a közvéleményt. Megkérdőjelezték az elvégzett kutatás hitelességét; további vizsgálatokat követeltek; kutatásokat finanszíroztak más területeken, mint a beteg épület szindróma vagy a kergemarhakór, hátha rá tudják venni a médiát, hogy inkább ezekkel foglalkozzon. Mesterségesen szították a kételkedést.<sup>16</sup> Egy titkos belső körlevél később arra emlékeztette a bennfenteseket, hogy „a mi termékünk a kétely”.<sup>17</sup>

Ha a meggyőzésre gondolunk, érthető módon az jut eszünkbe, hogy az emberekkel olyasmit hitetnek el, amit nem kellene elhinniük, és erről a jelenségről lesz is szó a következő fejezetben. A probléma azonban néha nem az, hogy túl lelkesen hiszünk valamiben, hanem az, hogy okunk lesz nem hinni semmiben. A dohányosok szerettek dohányozni, testi szinten függtek a nikotintól, és ha egy mód volt rá, ezt folytatni is akarták. A dohányiparban érdekelték számára az volt a legkedvezőbb helyzet, ha a dohányos vállat von, és azt mondja magában: „Nem tudom, melyik az igaz a sok ellentmondó érv közül”. Nem az volt tehát a feladatuk, hogy meggyőzzék a dohányosokat a cigaretta biztonságosságáról, hanem az, hogy kételyt ébresszenek a veszélyességét igazoló statisztikai bizonyítékokkal szemben.

És mint kiderült, a kétely igen könnyedén előállítható termék. Kari Edwards és Edward Smith pszichológusok két évtizeddel ez-

előtt elvégeztek egy kísérletet, amelyben arra kérték amerikai alanyaikat, hogy hozzanak fel pró és kontra érveket olyan, politikailag érzékeny témákban, mint az abortusz, a gyerekek verése, a homoszexuális párok örökbefogadáshoz való joga, a kisebbségekhez tartozó munkavállalóként történő alkalmazásának kvótái és a tizenhat év alattiakra kiszabható halálbüntetés.<sup>18</sup> Nem meglepő, hogy arra jutottak, az emberek részrehajlók: a megkérdezetteknek nehézséget okozott megfogalmazni azokat az érveket, amelyekkel ellenfeleik védének saját nézetüket. Érdekes módon azonban Edwards és Smith megfigyelése szerint ez a részrehajlás sokkal egyértelműbben mutatkozott meg a negatív érvek terén, vagyis a hitetlenség könnyebben jött, mint a meggyőződés. A kísérlet résztvevőinek könnyebb volt olyan álláspontok ellen érvelni, amelyeket nem kedveltek, mint olyanok mellett, amelyekkel egyetértettek. A kételynek tehát különleges ereje volt.

A kételyt azért is könnyű eladni, mert szervesen hozzátartozik a tudományos felfedezés és vita folyamatához. A legtöbbszörnek azt tanítják az iskolában – legalábbis így kellene, hogy legyen –, hogy kérdőjelezzük meg a bizonyítékokat. Az egyik legrégebbi tudományos társaság, a brit Királyi Természettudományos Társaság (The Royal Society) mottója is erre utal: *nullius in verba*, vagyis nagyjából „a szó semmi”. A statisztikai bizonyítékokat tagadni igyekvő érdekcsoport mindig meg fogja találni a tudománynak azokat a területeit, amelyekben még nincs egyetértés, mindig előállhat azzal, hogy a kérdéskör rettentően bonyolult, és mindig követelhet további kutatásokat. És ezek a kijelentései tudományosan, sőt bölcsen hangzanak majd. Mégis hamis és veszélyes benyomást keltenek: azt üzenik, hogy valójában senki nem tud semmit.

A dohányipar módszereit azóta széles körben átvették.<sup>19</sup> Manapság a legnyilvánvalóbban a klímaváltozás tagadói használják őket, de már átlépték a tudomány határait, és begyűrűztek a politikába is. Robert Proctor történész, aki évtizedek óta tanulmányozza a dohányipart, a modern politikát „a tudatlanság aranykorának” nevezi. A sok dohányoshoz hasonlóan, aki szeretne tovább füstölni, közülünk is sokan ragaszkodnak ösztönös megérzéseikhez a politikai kérdésekben.

A politikusoknak így nem kell egyebet tenniük, mint meggyőzni bennünket, hogy kételkedjünk az ezeket az ösztönöket esetleg megingató bizonyítékokban.

Ahogy Donald Trump egykori főstratégája, Steve Bannon elhíresült gondolatában megfogalmazta Michael Lewis írónak: „A demokraták nem számítanak. Az igazi ellenzék a média. Velük pedig úgy lehet elbánni, hogy elárasztjuk az egészet szarral.”<sup>20</sup>

Egy másik, szintén Donald Trumphoz köthető kifejezés – a *fake news*, vagyis álhír – története is tanulságos. Ez eredetileg nagyon is konkrét jelenségre utalt: egyes weboldalak hamis cikkeket tesznek közzé abban a reményben, hogy kattintásokat szerezzenek a közösségi médiában, és így reklámbevételekhez jussanak. Ennek szembeötlő példája volt az az állítás, mely szerint a pápa támogatta Trump elnökjelöltségét. Amikor aztán Trump győzött, egy időre eluralkodott az erkölcsi pánik; mértékadó elemzők amiatt aggódtak, hogy a hiszékeny választók azért szavaztak Trumpra, mert elhitték ezeket a felháborító hazugságokat.

A pánik felesleges volt. Tudományos vizsgálatok szerint az álhír sohasem volt elterjedt, és nagy befolyással sem bírt; javarészt kis számú, szélsőségesen konzervatív, idős szavazó fogyasztotta, akik valószínűleg amúgy is mindvégig Trumpot támogatták. Ezek az álsztorik rövidesen jóval kisebb problémát jelentettek, a közösségimédia-oldalak ugyanis felléptek ellenük.<sup>21</sup>

Az *álhír* mint fogalom azonban igen erőteljes eszközzé, kényelmes kifogássá vált, amivel bármilyen forrásból származó állítást vissza lehetett utasítani; a cinikus „hazugság, szemenszedett hazugság és statisztika” aforizma modern változata lett belőle. Trump, akinek már-már beteges tehetsége van ahhoz, hogy bonyolult problémákat politikai bunkósbotná egyszerűsítsen, a kifejezést az újságírók demonizálására használta. Ahogy tette azt sok más politikus is, köztük Theresa May, az Egyesült Királyság akkori miniszterelnöke és politikai ellenlábasa, Jeremy Corbyn munkáspárti vezető is.

Az *álhír* azért csengett jól, mert sajnálatos igazságra világított rá: mint ugyanis látni fogjuk, még a fősodorhoz tartozó hírszolgáltatóknál is bőven akad példa gyenge minőségű, összezsapott újságírói mun-

kára. Hiába vannak komolyan vehető és felelősségteljes újságírók, akik gondosan feltüntetik állításaik forrását, ők is ugyanabban a mentális szemétkosárban találták magukat, mint a Trump pápai támogatásáról cikkező álhírterjesztők.

Aggasztó lenne egy olyan világban élni, amelyben sokan készek bármit elhinni, de még ennél is aggasztóbb lenne, ha az emberek semmiben sem hinnének a saját előítéleteiken kívül.

Az Egyesült Államokban egy szenátusi bizottság 1965 tavaszán azt az élet-halál kérdést mérlegelte, hogy tegyenek-e egészségkockázati figyelmeztetést a cigarettásdobozokra. Az egyik beidézett szakértőt azonban nem győzték meg teljesen a tudományos bizonyítékok, ezért elővette a gólyák és a kisbabák témáját. Mint elmagyarázta, határozott kapcsolat áll fenn a születő csecsemők és a vidéken élő gólyák száma között.<sup>22</sup> A régi hiedelem, amely szerint a kisbabákat a gólya hozza, nem igaz, folytatta a szakértő; persze, hogy nem az. A korreláció nem jelent ok-okozati viszonyt. Nem a gólyák hozzák a kisbabákat. A nagyobb helyeken azonban több gyerek és több gólya fér el. Ennek megfelelően az, hogy a dohányzás előfordulása korrelál, vagyis együtt mozog a tüdőrák előfordulásával, még nem jelenti azt – egy pillanatra sem –, hogy a dohányzás *okozza* a tüdőrákot.

– Ön tényleg úgy gondolja, hogy a dohányzást a betegséggel összekapcsoló statisztikák éppen olyan érintőleges viszonyt fejeznek ki, mint a gólyákra vonatkozó statisztikák? – tette fel a kérdést a bizottság elnöke. A szakértő erre azt felelte:

– Nekem a kettő ugyanolyannak tűnik.<sup>23</sup>

Az illetőt Darrell Huffnak hívták.

A dohányipari lobbizás fizetett neki, hogy tegye azt, amihez a legjobban ért: szellemes példák, némi statisztikai tudás és megfelelő mértékű cinizmus elegyével vonja kétségbe az állítást, hogy a dohányzás veszélyes. Még mesterműve folytatásán is dolgozni kezdett, bár a könyv végül nem jelent meg. A második rész tervezett címe a következő volt: *How to Lie with Smoking Statistics* (Hogyan hazudjunk a dohányzási statisztikával?).<sup>24</sup>

A kétely erős fegyver, a statisztika pedig sebezhető célpont. Ez a célpont védelemre szorul. Igen, valóban könnyű hazudni a statisztika felhasználásával, nélküle azonban még könnyebb.<sup>25</sup>

Ami pedig ennél is fontosabb, statisztika nélkül lehetetlen igazat mondani; nélküle nem érthetjük meg a világot, és nem próbálhatunk meg javítani rajta, ahogy azt Richard Doll és Austin Bradford Hill tették. Ahhoz, amit véghez vittek, kellett némi éleslátás és eltökéltség, de nem volt szükség hozzá sem zsenialitásra, sem érthetetlen matematikai számításokra. Megszámolták, ami fontos volt: a dohányosokat, a nemdohányzókat, a tüdőrákos eseteket, a szívbetegségeket. Türelmesen és módszeresen számlálták meg őket, és az összegyűjtött bizonyítékok alapján körültekintő következtetést vontak le. Ezek a következtetések az elkövetkező évek alatt emberek tízmillióinak az életét mentették meg, köztük talán a sajátjukat is: Hill letette a pipát és követte a nem dohányzó Doll példáját, és mindkét kutató a kilencvenes éveiben hunyt el.

Ha biztonsággal és okosan használjuk a statisztikát, olyan tendenciákat is felfedezünk, amelyeket egyébként nem vennénk észre. Modern világunk óriási, nagyon összetett és nagyon-nagyon érdekes. Már majdnem nyolcmilliárd ember él itt. Nap mint nap több ezer-milliárd dollár cserél gazdát. Az átlagos emberi agy nyolcvanhat milliárd idegsejtet tartalmaz.<sup>26</sup> Az interneten mintegy kétmilliárd weboldal található. Egy új vírus pedig egyetlen emberről több ezer, több millió vagy akár több milliárd másakra is átterjedhet. Bármit akarjunk is megérteni a világgal, egymással és saját magunkkal kapcsolatban, statisztika nélkül bizony nem jutunk messzire. Mintha röntgen nélkül próbálnánk vizsgálni a csontokat, mikroszkóp nélkül a baktériumokat vagy távcső nélkül a csillagos égboltot.

Kering egy népszerű történet Galilei teleszkópjáról: a római katolikus egyház eretnekséggel vádolta a csillagászat atyját, de a bíborosok a tárgyaláson nem voltak hajlandók belenézni az általa készített eszközbe, mert azt varázstrükknek tartották. Galilei azt állítja, hegyeket látott a Holdon? Bizonyára koszos volt a teleszkóp üvege. Látta a Jupiter holdjait? Ugyan már! A holdak magában a teleszkópban vannak. Nem néztek bele.



Négy évszázaddal később könnyen nevetünk ezen a történeten, amelyet egyébként némileg eltúloztak az idők során.<sup>27</sup> Nem kellene ennyire önelégültnek lennünk. Sokan nem vagyunk hajlandók áttekinteni a statisztikai bizonyítékokat, mert félünk, hogy átvernek bennünket. Dörzsölnék és tapasztaltnak tartjuk magunkat, csak mert a Huff-féle megközelítést alkalmazva cinikus módon legyintünk minden statisztikára. Pedig nem vagyunk azok. Ezáltal ugyanis elismerjük a vereséget a populistákkal és propagandistákkal szemben, akik éppen azt akarják elérni, hogy legyintsünk, vessük el a logikát és a bizonyítékokat, húzódjunk vissza a csigaházunkba, és higgyük azt, amitől jól érezzük magunkat.

Én azt szeretném, ha az emberek valami mást tennének. Szeretnék segíteni abban, hogy magabiztosan vegyék kézbe a statisztika teleszkópját, és bátran vizsgálják meg vele a világot. Segíteni akarok, hogy megértsék a statisztikai igazságok mögött húzódó logikát, és megmeneküljenek a hazugságokat szülő hibás logika, érzelmek és kognitív előítéletek karmaiból.

Emeljük hát a szemünkhöz a statisztika távcsövét, és nézzünk körül! Meg fogunk lepődni, milyen tisztán látunk mindent.