

---

**Havasi Éva,**  
a KSH főtanácsadója  
E-mail: Eva.Havasi@ksh.hu

**Burjánzó számok,  
mutáns adatok\***

*Havasi Éva* írásának megjelentetésével vitára invitáljuk a statisztikatudomány művelőit. A témával kapcsolatban Olvasóink figyelmébe ajánljuk a szerző korábban megjelent *David Boyle*: „A számok zsarnoksága, avagy miért nem tesz minket boldoggá a számolás” című könyvének ismertetését.<sup>1</sup>

Semleges statisztika nem létezik. A számokban tömören megnyilvánuló jelenségek nem maguk a vizsgált dolgok, hanem azok meghatározott nézőpontból számosított tulajdonságai. Egy-egy jelenség egyenértékű módon többféle számmal is leírható. Fel kell hagyni azzal a gyakran hangoztatott nézettel, hogy „a számok önmagukért beszélnek”. Minden számhoz tartozik egy értelmezési keret, mérési mód, azaz egy módszertani metainformációs rendszer, amely világossá teszi a számok tartalmát.

Használatuk során a számok legtöbbször elveszítik a mögöttük meghúzódó nézőpontokat, módszertani-mérési megfontolásokat és önálló, független életet élnek. Egyre bizonytalanabb és követhetlenebb számtenger születik, melyből – látszólag – ki-ki ízlése, érdeklődése, sőt érdeke szerint válogathat. Így lehet végül egy számmal mindent „bebizonyítani”, sőt annak ellenkezőjét is. Ahogy *Bródy András* írja: „A mérés és a megfigyelés mindig valamilyen elméletet tételez fel a háttérben. A megfigyelés eredménye, illetve a mért érték csak ezen elmélet alapján értelmezhető. Mit is teszünk tulajdonképpen, amikor „mérünk”? Bizonyára a körülöttünk levő valóságról szerzünk értesüléseket. Könnyen átsiklunk azonban a felett, hogy az értesülés, a mérés számszerű eredménye nem közvetlen, és ezért csak áttételesen jellemzi a valóságot” (*Bródy* [1990]).

A számokat kritikusan kell kezelni. Kritikusan, de nem szkeptikusan, nem cinikusan és nem is naivan (*Best* [2001]). Ehhez azonban „számkultúrára” van szükség. A

---

\* A „Műhely” rovatban megjelenő tanulmány az eredeti írás rövidített és a rovat műfajához igazított változata. Ezúton köszönöm *Lakatos Mária* közreműködését.

<sup>1</sup> *Havasi É.* [2010]: Davide Boyle: „A számok zsarnoksága, avagy miért nem tesz minket boldoggá a számolás” c. könyvének margójára. *Statisztikai Szemle*. 88. évf. 2. sz. 197–201. old. [http://www.ksh.hu/statszemle\\_archive/2010/2010\\_02/2010\\_02\\_197.pdf](http://www.ksh.hu/statszemle_archive/2010/2010_02/2010_02_197.pdf)

tapasztalat az, hogy a betűk világában az emberek jóval otthonosabban mozognak, mint a számokéban. A „számanalfabétizmus”, illetve a számok használatában való jártasság hiánya, valamint a „számautizmus”, a számok vég nélküli, már-már beteges használata, a mindent egzakt módon számokkal, indikátorokkal „bizonyítani akaró” számelőállítói-számhasználói magatartás egyaránt jelen van világunkban.

A következőkben a statisztikai számhasználat buktatóiról lesz szó, főként az ún. leíró statisztikák területéről. A magasabb matematikai eszköztárral dolgozó statisztikai modellezés, szimuláció stb. kimaradnak írásom vizsgálati köréből.

## 1. A számok ismeretelmélete

Számtalan szám vesz bennünket körül. A számmal kifejezett jelenségek még csak jelenségek, de nem maguk a dolgok

A számadatok egzaktágukkal azt a látszatot keltik mintha a minket körülvevő jelenségek eredendően rendelkeznének a számokban kifejezett tulajdonságokkal, hasonlóan a számháborúban a játékosok fejére tűzött számjegyekhez, csak le kéne olvasnunk a számadatokat. *Marx* „A Tőke” „Az áru” című fejezetében, érzékletesen fogalmazta meg a gondolatot a cukorsüveg példáján. „Egy cukorsüveg, minthogy test, nehéz, és ezért súlya van, de semmiféle cukorsüvegen sem látható vagy tapintható a súlya... A vas testi formája, önmagában véve, éppoly kevésbé megjelenési formája a nehézségnek, mint a cukorsüveg testi formája... Mindazonáltal, hogy a cukorsüveget, mint nehezett kifejezzük, súlyviszonyba helyezük a vassal. Ebben a viszonyban a vas olyan test, amely semmit nem képvisel a nehézségen kívül. ...Ha mindkettőt a mérlegre teszszük, valóban azt látjuk, hogy mint nehéz dolog a kettő ugyanaz” (*Marx* [1955]).

A mérés műveletében látszólag eltűnik az elméleti elvonatkoztatás, ami megelőzi magát a mérést. Pedig a mérés eredménye mindig csak valamilyen elmélet alapján értelmezhető. Azért nem szembetűnő az elmélet, azért feledjük a háttérben maradó elvonatkoztatást, mert vannak esetek – éppen ezeken kezdjük a mérést – amikor a valóság könnyen és közvetlenül követhető – olvashatjuk Bródy András már idézett, nagy vitát kiváltó cikkében.

Már *Arisztotelész* is figyelmet szentelt a problémának a cserére vonatkozó fejtegetései kapcsán. *George Simmel* is a cserére és a pénzre, mint általános egyenértékesre vonatkozó elméletével kapcsolatban foglalkozik a mérés ismeretelméleti kérdéseivel.

„A mennyiség legjellegzetesebb sajátossága, hogy egyenlőség, vagy nem-egyenlőség állítható róla.” – mondja *Arisztotelész* [1997]. Kifejti, hogy a csere éppen a minőségi különbözőség, az eleve „össze nem mérhetőség” talaján jön létre.

A méréssel, a számszerűsítéssel – hasonlóan a cseréhez – a minőséget összemérhetővé, mennyiséggé alakítjuk. Ezt a gondolatot viszi tovább *Simmel* a pénz filozófi-

ájának kifejtése során. A „mennyire értékes” kérdés helyett egyre inkább a „mennyibe kerül” kérdése kerül a középpontba. Ez azt jelenti, hogy a tárgyak kvalitatív jellegét fokozatosan háttérbe szorítja a kvantitatív vonás. A dolgok minőségi jellege elveszti pszichológiai jelentőségét. Simmel az „egyenlősítés tragédiájának” nevezi, hogy amikor a dolgok pénzbeli értéke egyforma, akkor nem számít, hogy más szempontok alapján hogyan viszonyulnak egymáshoz. Ez az értékvesztés a modern ember fásultságához vezet, vagyis az illető elvesztette érzékenységét arra, hogy a dolgok különbözőségeinek finomságait érzékeldje (Simmel [1991]).

Mint ahogy a pénz is eszközből egyre inkább céllá válik, úgy a statisztikusok számára is a számszerűsítés, amely egy eszköz arra, hogy a dolgok lényeges jegyeit megragadják és összehasonlítsák egyre inkább abszolút céllá válik és magát a megismerés értelmét jelenti. A számszerűsítés, a dolgok széles skálája fölött egy olyan absztrakt közös nevezőt alkot, amelyben a legellentétebb, legidegenebb és legtávolabbi dolgok is kifejezhetővé válnak. Amikor a sokoldalú minőségnek ez az „egyenlősítése”, számmá alakítása megtörténik, akkor az így nyert szám önálló életet kezd élni és abszolutizálódik.

A következőkben a számgyártás folyamatát, majd pedig annak végtermékét tekintjük át.

## **2. Az adat, mint a „termelési folyamat” terméke, avagy az adatokat emberek csinálják**

Van egy elterjedt vélekedés, mely szerint a számok, „az igazság kis aranyrögei”, aranyrögök, amelyeket hasonlóan a markológéphez, mi is csak úgy összekotorhatunk – írja *Joel Best*. „A statisztikai adatok azonban inkább hasonlítanak a drágakövekhez, amelyeket először kibányásznak, összegyűjtenek, feldarabolnak, csiszolnak, ... majd kiállítják őket az érdeklődők számára.” A statisztikai adatok azonban inkább hasonlítanak a drágakövekhez, melyeket egy bonyolult művelet sor eredményeként kapunk. Hasonlóan az értékesebb drágakövekhez a statisztikai adatokhoz is tanúsítványt mellékelnek.

A statisztikát valaki(k), valamilyen céllal, valamilyen módon és módszerekkel előállítják. Divatosan szólva a statisztika „termék”, mégpedig egy sajátos „termelési folyamat” eredménye. Az értékesebb statisztikai adatokhoz mintegy tanúsítványként, minőségjelentések, riportok is kapcsolódnak. A mért adat az emberek fejében egyenlővé válik, behelyettesítésre kerül a megmért jelenséggel. Ily módon, a mérési eredmény esetében, elhalványul az a tény, hogy a szám egy adott nézőpontból, egy adott mérési eljárás alapján mutatja fel a megfigyelt jelenséget.

A „számgyártás” legfontosabb minőségi kritériuma, a statisztikai relevancia ezen a ponton érhető leginkább tetten. Akkor releváns a számadat, ha érdeklődésre számot

tartó, értelemmel bíró nézőpontból „számosítja” az adott jelenséget. S ha a ténylegesen használt (mérhetővé tett) és a felhasználók által elvárt fogalmak közötti megfelelés is megvalósul (*Szilágyi* [1997]). Egy-egy jelenségről több releváns és kevésbé releváns számadat is születhet. Az emberek a közlekedési szakemberek számára utasok, az orvosok számára betegek, a sportesemények szervezői számára szurkolók, a családszociológusok számára például meghatározott családi állású személyek. Tehát az embereket mint utasokat, betegeket, szurkolókat, apákat, anyákat, gyermekeket stb. számszerűsíthetjük. Egy-egy nézőponton belül is döntő a fogalomhasználat megválasztása. A betegeket osztályba sorolhatjuk, és így számszerűsíthetjük, diagnosztizált betegségük (BNO – betegségek nemzetközi osztályozási rendszerét használva), gyógyulási idejük, táppénzes napjaik száma, betegségük örökletes jellege, maradandó károsodással, kórházi ellátással járó vagy szakellátást igénylő volta szerint stb.

Ezekben az esetekben a statisztikai számbavételi tevékenység alapja a dolgoknak egy adott fajtához sorolása. Ez az egyszerűnek tűnő első lépés gyakran a mérés műveletének utolsó lépése is. A rétegződéskutató iskolai végzettség, gazdasági aktivitás, foglalkozás, beosztás, jövedelem, vagyoni helyzet, fogyasztási sajátosságok stb. alapján kategóriákba sorolja az embereket, majd a belőlük kialakuló „mintázatok” alapján újabb (társadalmunk rétegződését leíró) kategóriákat, osztályokat alkot. Így születnek az olyan új osztályok, mint „gazdasági elit”, „vagyonos középosztály” vagy „elszegényedett alsó középosztály” és így tovább. Egy-egy klasszikus statisztikai osztályozási rendszer változása jól mutatja ezt a folyamatot. Az orvostudomány fejlődésével, a diagnosztikai eljárások finomodásával például a korábban említett BNO is módosul, bizonyos betegségcsoportok szétbomlásával és újabbak létrejöttével. Ugyanez elmondható a Foglalkozások Egységes Osztályozási Rendszeréről (FEOR) vagy az utóbbi időben állandóan napirenden levő Tevékenységek Egységes Osztályozási Rendszeréről (TEAOR). Az állami bürokrácia nap mint nap létrehoz új kategóriákat, új osztályokat, korábban nem létező embercsoportokat. Tegnap még „közveszélyes munkakerülőkről”, ma „jövedelepótló támogatásban részesülőkről”, „munkanélküliekről” beszélünk, illetve beszélünk.

Még „stabil” kategóriák mellett sem könnyű az osztályba sorolás. *Jánossy Ferenc* példájával élve a tehenet és a disznót mérete alapján nem nehéz osztályba sorolni, ugyanakkor a tehen és a ló méretbeli összehasonlítása már nehezebb (*Jánossy* [1963]). *Bródy András*nak *Jánossynak* a golyóscsapágyra és kavicsra vonatkozó példáját továbbfűzve azt fejtegeti, hogy a kavicsok osztályozása esetén nemcsak a formagazdagság nehezíti a méretbeli összehasonlítást, hanem a gondok már ott kezdődnek, hogy mettől meddig kavics a kődarab (*Bródy* [1990]). Miközben a hétköznapi életben gondolkodás nélkül használjuk a „gyermek” fogalmát, a társadalomstatistikusok számára a gyermekes családok vizsgálata esetén komoly nézetkülönbségek vannak a fogalom meghatározása, „operacionalizálása” körül. A szüleivel élő 30

éves aggregény bizonyos szempontból gyermek, de a gyermekes családok vizsgálatának ez a családtípus semmiképpen nem része. Ha 20, 25 évnél húzzuk meg a határt, akkor kérdéses, hogy a „gyerek családi állású” személy gyermeknek tekinthető-e, még akkor is, ha esetleg nem aktív kereső, és így tovább.

A relevancia esetében az érdeklődésre számot tartó nézőpont mellett a felhasználó számára való alkalmasság („fit for use”) részét képezi a megfelelő (operacionális) fogalmakon túl a mértékegység megválasztása is. Lehet, hogy az adathasználót nem érdekli, hogy kik tartoznak a „vagyonos középosztályba”, de érdekli az emberek vagyoni helyzet szerinti tagolódása. Lehet, hogy érdekli az emberek településtípus szerinti megoszlása, de nem falu-város szerint, hanem a település népességszáma alapján képzett nagyságkategóriák szerint. A mértékegység adekvát megválasztása kardinális kérdés. A munkahelyem otthonomtól való távolságát meghatározhatom méterben, de mint notórius késő számára relevánsabb az időmennyiségben, percben való távolság megadása. A „három napi járőföld” a távolság idővel való meghatározásának szép példája. A kifejezés utal a korra, amelyben helye és relevanciája volt az effajta „mérési módoknak”. Az adózóknak „kéményekkel”, az „egy füst alatt” élőkkel való számbavétele (füstadó) a maga korában releváns volt, mára azonban a keresettel, piaci jövedelemmel rendelkezők (személyi jövedelemadó) számát használjuk (Hüttl [2003]).

### 3. A számok önállósodása és következményei

A számok, elveszítve a háttérükben húzódó tartalmi és módszertani-mérési megfontolásokat, önálló életet kezdenek élni, s relevanciájuk sokszor túlnő saját határainkon, „érvényességi körükön”. Például mai társadalmunkban a szegények aránya lehet 5-8-10-25, de akár 100 százalék is, attól függően, hogy mivel mérjük a szegénységet, és hol húzzuk meg a szegénységi küszöbszintet. Ha „csak” jövedelemmel mérjük, akkor is fennáll az a kérdés, hogy mekkora összegnél húzzuk meg a szegénység határát. Természetesen minden határválasztásnak megvan, meg lehet a „relevanciája”, attól függően, hogy mire vagyunk kíváncsiak. Csak úgy, minden „lábjegyzet” nélkül, nem mondhatunk egy számot, arányt a szegényekre. Amikor használatba vesszük a szegénységi mutatót, nem téveszthetjük szem elől tartalmának mögöttes mérési történetét.

Vannak már standardok bizonyos jelenségek mérésére. Ezek azonban csak egy bizonyos megközelítései a jelenség vizsgálatának. Sokszor ezek a bevált és használt mérési módok, mérőszámok, mutatók nem elégítik ki az adathasználók igényeit, számukra ugyanis nem vagy csak korlátozottan releváns a mögöttük húzódó mérési nézőpont vagy metódus. Ennek egyik jó példája a GDP széles körű fetiszizált, gazdasági fejlettségre, jólétre vonatkozó használata.

Bár a GDP kiszámításának többszáz oldalas dokumentációja van és szinte bárki számára hozzáférhető, érthető módon, kevesen böngészik a számhasználók közül. Ugyanakkor tartalmának pontos ismerete nélkül, mintegy automatikusan, használják a gazdaság eredményességi mutatója mellett, a gazdasági fejlettség vagy a jólét, az ország életszínvonalának mérőszámaként. Egyre többen a gazdasági mérőszám funkcióját is megkérdőjelezzik, felröví a környezetvédelem szempontjainak hiányát (zöld számlák), az információs társadalom teljesítményének figyelmen kívül hagyását, a magánháztartás keretei között végzett munkák felértékelésének elmaradását stb. (*Stiglitz–Sen–Fitoussi* [2009]).

Mások szerint mi végre a magas GDP-érték, ha nem arra, hogy hosszú boldog életet éljünk. Ezért az ország teljesítményét vagy a jólétét inkább mérnék, sőt esetenként mérik is a várható élettartammal és a lakosság „boldogságszintjével”. Merőben közgazdasági szempontból is egyre többen a jövedelmi egyenlőtlenséggel, bizonyos termelési eredmények elhagyásával korrigált GDP-t ajánlják mérőszámnak.

#### 4. A mérhetőség csapdája és a számautizmus

A számautizmus a posztautista közgazdaságtani irányzat kiterjesztése a mai számokat, statisztikai adatokat fetiszizáló világunkra. A kifejezés a számokba való megszállott belefeledkezésre utal. Arra az érzésre, hogy a számok és az indikátorok özönében elveszítjük a valósággal való kapcsolatunkat, kiszakítjuk magunkat a reális világból.

Túl gyakran próbáljuk meg kvantifikálni azt, amit valójában nem, vagy csak nagyon korlátozottan tudunk megmérni. A kvantifikálás folyamán (igen gyakran) számos csúsztatást is végrehajtunk, hogy mindenképpen mérni tudjunk.

Amikor a munkanélküliséget, a jólétet, a szegénységet kvantifikáljuk, akkor egészen más a helyzet. Mivel a hétköznapiakban is használt fogalmakról van szó, mindenkinek van egy képzelete arról, hogy mit jelentenek ezek a fogalmak. Ráadásul a fogalmakhoz értéktartalmak is kapcsolódnak. Munkanélküli az, aki nem dolgozik. S miért nem dolgozik? Mert nem is akar. Mert az állam nem gondoskodik arról, hogy mindenkinek legyen munkája. Mert az emberekben nincs kezdeményezőkézség, rugalmasság... stb. A munkanélküliségi ráta az ország egyik kiemelten kezelt gazdasági-társadalmi mutatószáma is. Ezért számos érdek fűződik annak bizonyításához, hogy ez a szám alacsony vagy magas, nő vagy csökken. Ugyanez vonatkozik a jólétre vagy annak hiányára, a szegénység alakulására. Munkanélküli az, akinek nincs munkája. Munkanélküli az, aki annak tartja magát. Munkanélküli az, akit a Munkaügyi Központ munkanélküliként tart nyilván. Munkanélküli az, akinek az elmúlt héten nem volt egyetlen fillér keresetet biztosító munkája sem, ugyanakkor aktívan keres munkát és 30 napon belül tudja vállalni a munkába állást. Ezek a definíciók mind

a munkanélküliség értelmes, releváns és használt definíciói. Az utolsóként említett ún. ILO-definíció az, ami a munkanélküliségi ráta kiszámításának alapját képezi. Ugyanakkor kevés ember gondol éppen erre, amikor a rátát használja.

A szegénység esetében hasonló a helyzet. Mérésére legtöbbször az emberek jövedelmét használjuk. Majd egyenlőségjelet teszünk bizonyos jövedelem nagyság és a szegénység közé. Alacsony jövedelem = szegénység. Majd elfelejtjük, hogy mi volt a mérőeszközünk és újból szegényekről/szegénységről beszélünk, mintha a fogalom sokszínűségét figyelembe véve mértük volna meg a szegénység nagyságát, kiterjedtségét.

Mérjük, amit gyorsan és látványosan mérni tudunk. A szegények számát jövedelemmel mérve gyors és „egzakt” eredményt kapunk. Ezt a számot össze tudjuk hasonlítani az előző évvel és a következő évvel egyaránt. A jövedelemhiány mellett, a mobilitási képtelenség, a megalázottság, a lelki gyötrellem, az egészségtelen lakáskörülmények, a rossz munkafeltételek, a perspektívátlan fásultság stb. számbavétele a szegénység mérésénél már olyan „kihívást” jelentenek, amelyekre gyors és látványos eredményeket követelő menedzseri világunkban csak korlátozottan van igény és lehetőség.

Mindezek után jönnek a különböző mérési-módszertani hibák, értelmezési nehézségek, csúsztatások, torzítások. Akkor mire való mégis ez a számosítási igyekezet? Olyan világban élünk, amikor minden megmértetik, s „ami meg van mérve, az el is van végezve” – ahogy mondják. És ezt megtoldhatjuk azzal, hogy amit nem mérünk az praktikusán nincs is. Többek között éppen ezért van az, hogy a mérési folyamatnak és eredményének egyre nagyobb jelentőséget tulajdonítunk.

## 5. Mutáns statisztika – számokat szülő számok

A gyakorlatban a „háttérüket” elvesztő számok önálló életet élnek, továbbszámolnak velük, összehasonlítják őket más számokkal. A Bruttó Hazai Termék (Gross Domestic Product – GDP), a Valódi Fejlődés Indikátora (Genuine Progress Indicator – GPI), a gazdasági jólét mérőszámai (Index of Sustainable Economic Welfare – ISEW, Measure of Economic Welfare – MEW) a Humán Fejlettségi Mutató (Human Development Index – HDI) és egyéb vég nélkül sorolható mutatószámok építőkövéjé válnak. Az inflációs rátát a különböző időpontra vonatkozó adatok esetében automatikusan használjuk, anélkül, hogy foglalkoznánk kiszámításának menetével, korlátaival. A nyugdíjak értékállóságának vizsgálatánál Zafir Mihály halála óta senki nem figyelmeztet minket, hogy az általános fogyasztói árindex helyett nyugdíjas árindexet használtunk-e. A szegénységi ráta adatgyűjtések sokaságából számolódik, anélkül, hogy a felhasználók emlékeznének a forrásra, ismernék a mérési módszert. A szám megragad a fejükben, használják és bekerül a köztudatba. Ugyanakkor a számok egyre több újabb számot szülnék. Így születik a mutáns statisztikák legtöbbször.

Joel Best „Damned Lies and Statistics” (Átkozott hazugság és statisztika) c. könyvében [2001] azt fejtegeti, hogy vannak olyan statisztikák, számok, amelyek torzszülöttként jönnek a világra, míg mások később, a használatuk során torzulnak. A torzuláshoz, torzításhoz erős érdekek is fűződhetnek.

Best az adathasználókat három nagy csoportra osztja: a naivakra, a cinikusakra és a kritikusokra. A naiv felhasználó a számok bűvöletében él és kritikátlanul elfogadja azokat. Ha egy vizsgálatból „ez jött ki”, akkor ez az igaz. A cinikus felhasználókat jól illusztrálhatjuk *Churchill* elhíresült mondásával: „csak annak a statisztikának hiszek, amit magam hamisítottam”. A kritikus adathasználó pedig „a számok önmagukért beszélnek” gondolkodásmód helyett megpróbálja értelmezni, megérteni egy-egy szám valós tartalmát. Ehhez azonban számkultúrára és esetenként kitartó elszántságra van szükség.

Az oktatási rendszer elhanyagolja a számok használatának, „olvasásának” tanítását, így az adat-előállítókra fokozott szakmai és etikai felelősség hárul. Az adatgyűjtés eredményeihez érdekek is kapcsolódnak. Ezek lehetnek egyszerűen szakmai, de szakmán túlnyúló érdekek is, amelyek hozzájárulhatnak ahhoz, hogy egy-egy adat csak látens módon legyen jelen vagy éppen fordítva, a lehető legnagyobb ricsajt csapja. Ahogy Best állítja, a „kritikus”, de nem naiv és nem cinikus közvélemény szerepe napjainkban különösen fontos.

## 6. A statisztikai adatok relevanciájának fontossága

Vannak számok, melyek hozzásegítenek világunk megértéséhez, míg mások segítenek „elkendőzni” azt. A világ számokban kifejezett feltárásához, illetve „számlepel” alatt tartásához is fűződik érdek. Vannak helyzetek, amikor egyszerűen nem találjuk azokat a szempontokat, kategóriákat, mérési eszközöket, amelyek közelebb vinnének minket egy-egy jelenség megértéséhez.

*Darell Huff*: „Hogyan hazudjunk a statisztikával?” c. munkája 1954-ben jelent meg először, s mindmáig a statisztikai adatok interpretációjának egyik legfontosabb „alaptankönyve” (*Huff* [1976]). 2003-ban kínai nyelvre is lefordították. A siker titka a könyv szöveges tartalma mellett képanyaga. A címlapon egy jovialis arcú embert látunk partvissal a kezében, amint számokat söpör a szőnyeg alá. Régi jól bevált gyakorlat, hogy nem hazudunk, csak „eltüntetjük” az eredményeket, mintha nem is léteznének. A hivatalos statisztikákból is számos példát hozhatunk erre a megoldásra. A számozott, kis példányszámban megjelenő, csak egy szűk kör számára hozzáférhető statisztikai kötetek árulkodnak az ilyen múlt század ötvenes-hatvanas éveire oly jellemző hivatali gyakorlatról.

Mai megfelelője ennek a „Ne szólj szám, nem fáj fejem” megoldás. Amikor politikailag vagy társadalmilag érzékeny témákról inkább nem gyűjtünk adatot vagy nem



olyan metszetben, nézőpontból, amely a közvélemény számára releváns lenne. Nem vizsgáljuk, hogy a lakosság hány százaléka roma, vagy nem mondunk számot a hazai szegénységre stb.

A relevancia tartalom és forma egyszerre. Arról gyűjtsünk adatokat, amelyekre valós igények vannak, és olyan kategóriákat alkalmazzunk, amelyek közelebb visznek világunk megértéséhez, a megfelelő mérőeszköz, mértékegység megválasztásával. Ez utóbbit a „3 napi járó föld” példájával már korábban illusztráltuk. Hiába pontos, érthető, összehasonlítható, naprakész és hozzáférhető egy számadat, akár erős módszertani háttérrel is, ha teljesen érdektelen.

## 7. A hamis adathasználat következménye – mutánsok és torzszülöttek

A számokat interpretációval kell ellátni. Az interpretációs tévedésekkel, hamisításokkal szemben ugyanolyan szigorúan kell fellépni, mint a számok előállításánál során vétett hibákkal szemben.

Az időbeli összehasonlítások típushibája a bázis téves, esetenként hamis megválasztása. Ha a magyar lakosság jövedelmi helyzetét 1995-höz hasonlítom, akkor a gazdasági recesszió nehézségei ellenére is, ma jobban élünk, mint „korábban”. Ha azt akarom hangsúlyozni, hogy mekkorát zuhant az életszínvonal, akkor az elmúlt évi vagy a két évvel ezelőtti jövedelmi adatokhoz viszonyítom a lakosság mai jövedelmi helyzetét.

A hibás mintakiválasztásból adódó hamis eredmény legismertebb példái a választási eredményekre vonatkozó előrejelzések tévedései adják. A Literary Digest Egyesült Államok elnökére vonatkozó 1936-os választási veresége annak volt köszönhető, hogy a megkérdezettek kiválasztása telefonkönyvek és gépkocsitulajdonosok névjegyzékei alapján történt. Az így kiválasztott emberek „felülreprezentáltak” a jómódú, városi, képzetlenebb lakosságot. Ezt a hibát erősítette az önkéntes kérdőív használata, melynek visszaküldési aránya, mint mindig, ez esetben is alacsony volt, de politikai hovatartozástól nem teljesen független. Így duplán torzult a válaszadók csoportja a teljes választói népességhez viszonyítva. Az akkor felbukkanó G. Gallup újfajta minta-kiválasztási módszerével, az ún. kvótás kiválasztással, helyesen jelezte előre Roosevelttel szemben Landonnal szemben (Babbie [1999]). Itt jegeznék meg, hogy a valószínűség-számítás fejlődésével már a megfelelő nagyságú ún. „véletlen mintákat” preferáljuk, mert megbízhatóságuk mellett az eredmények konfidencia-intervalluma is meghatározható, vagyis meg tudjuk mondani, hogy a mintából nyert adataink mennyire pontosak, milyen határok között érvényesek. Ha egy ilyen mintából tudjuk, hogy a tanulók magassága 165 centiméter, akkor az adott mintanagyság mellett azt is meg tudjuk válaszolni, hogy a definiált 95, 99 százalékos, valószínűségi szinten  $\pm$  hány centiméter eltérés mellett „jó” a mintából nyert

adatunk. Vagyis például 99 százalékos valószínűséggel állíthatjuk, hogy a tanulók átlagmagassága 163 és 167 centiméter közötti. Adataink jelentős része ún. reprezentatív (véletlen) mintákból származik, s ezekben az esetekben, az adatok valójában nem „pontok”, hanem „tartományok”. „Minden szám egy kis „felhő”, tanította *Szalai Sándor*, az első magyar szociológia tanszék megalapítója.

Az *óvatlan, hamis adatértelmezés* területéről végtelen példát idézhetünk. Az egyik iskolapélda az Amerikai Gyermekvédelmi Alap Évkönyvének szövege, melyet egy amerikai folyóirat főcímében idézett: „A pisztoly általi gyermekgyilkosságok száma az Egyesült Államokban évente megduplázódik”. A pisztollyal elkövetett gyermekgyilkosságok száma ugyanis 1994-ben kétszer akkora volt, mint 1950-ben. Ha a címben jelzett állítást elhinnénk, vagyis az éves duplázódást, akkor az 1950-ben elkövetett 1 gyilkosság az éves duplázódás eredményeként 1965-re 32 768-ra nőtt volna, s 1987-re elérte volna a 137 milliárdot. Az 1950 és 1994 között eltelt 44 év alatt megduplázódott a gyermekgyilkosságok száma értelmezés is korlátok között igaz. Nem tudjuk, hogy 1951 és 1993 között mi történt, lehet, hogy a gyilkosságok számában semmilyen törvényszerűség nem érvényesült, lehet, hogy közben ugrásszerű emelkedés is végbe ment. Amit tudunk az két adat két időpontra, 44 év elteltével.

Az *adatokkal végzett átgondolatlan, hibás számtani műveletekre* végtelen példát hozhatnánk. Ha Budapestről Hatvanba 60 kilométer/óra sebességgel 1 óra alatt érünk le, s visszafelé ugyanezt az utat 40 kilométer/óra sebességgel tesszük meg, akkor könnyen azt gondolhatjuk, hogy átlagsebességünk 50 kilométer/óra volt. Ugyanezt a hibát követjük el akkor, amikor a fővárosban élő 5 százalékos és a vidéki 15 százalékos szegénységi rátát átlagolva, a hazai szegénységi rátát 10 százaléknak gondoljuk. A szegénységi rátákat nem adhatjuk össze (ha Budapesten 5 és vidéken 15 százalék, akkor együtt, országosan 20 százalék – bár ez is elő szokott fordulni), de nem is átlagolhatjuk automatikusan ( $5\%+15\%=20\%$  és az osztva kettővel 10%). Figyelembe kell venni, hogy az 5 százalékos szegénységi ráta a lakosság alig egyötödére vonatkozik, míg a 15 százalékos arány a lakosság négyötödére. A szegénységi ráta értelemszerűen nagyobb lesz, mint 10 százalék (valahogy így:  $4/5 \times 15\% + 1/5 \times 5\%$ , ami 13 százalék).

*Többféle átlagot használhatunk a statisztikában.* A két legismertebb: a számtani átlag és a medián (helyzeti középérték). Hogy melyiket választjuk, az attól függ, hogy mire akarjuk használni a számot. A (számtani) átlagjövedelemnek és a medián alapú átlagjövedelemnek is van értelme. A kettő azonban nem ugyanaz. Ha egy társadalomban nagyok a jövedelmi különbségek, akkor a számtani átlagjövedelem inkább matematikai jelentéssel bír, és kevésbé alkalmas az emberek valós jövedelmi helyzetének bemutatására. Ez esetben inkább a mediánt használjuk, ami nem érzékeny a szélsőségesen nagy, illetve alacsony jövedelmekre.

Az egyik leggyakoribb hiba a *túláltalánosítás, a szelektív/előítéletes észlelés*. Ez utóbbira történt már utalás, amikor a célközönség véleményét tükröző adatot a célkö-

zönség „igaznak”, „valósnak” éli meg. A jelenség lényege, hogy túl hamar megálunk a megismerésben. A barátnőm sokat beszél, jön a következtetés: a nők szeretnek beszélni. A statisztikusokra, a statisztikára vonatkozó idézeteket tartalmazó honlapon erre vonatkozóan számos vicces példát találunk, többek között a beteg és a sebész közötti párbeszédet: „Túl fogom élni ezt a nagyon kockázatos operációt?” Válasz: „Minden bizonnyal.” „Miért olyan biztos ebben doktor úr?” „Tízből kilenc beteg meghal az operáció során, és tegnap hunyt el a kilencedik.”

Gyakran találkozunk az *együtt járásnak, a látszólagos kapcsolatnak ok-okozati összefüggésként történő kezelésével*. Ezt esetenként nem könnyű észrevenni, de legtöbbször nem igényel különösebb szellemi erőfeszítést a felismerése. Az átlaghőmérséklet, illetve a napsütéses órák számának alakulása és a házasságkötések száma közötti kapcsolat fennállásából – jelesül, ha nő a napi átlaghőmérséklet, vagy a napsütéses órák száma, akkor nő a házasságkötések száma – nem következik az, hogy ha növekszik az átlaghőmérséklet, globális felmelegedés van, akkor nő a házasságkötések száma is. Az összefüggés látszólagos jellege esetenként kevésbé nyilvánvaló. Például kimutatták, hogy az egyetemeken a dohányzó diákok tanulmányi eredménye rosszabb. Ebből nem vonható le az a következtetés, hogyha leszokik egy diák a cigarettáról, akkor jobb jegyei lesznek.

Szeretem a számokat. Statisztikusként becsülöm az adat-előállítás mesterségét és művészetét. Meggyőződésem, hogy a számokban rögzített statisztikai adatok, információk hozzásegítenek világunk megértéséhez, ahhoz, hogy érveinket érdemes, sőt esetenként szükséges is számokkal alátámasztani. A háború a „politika folytatása más eszközökkel” – mondta *Carl von Clausewitz*, a porosz tábornok, katonai teoretikus. Ma – és ez jellemzi országunk közéletét is – a számháborúk korát éljük, melyben a harcoló felek számokat röpipitenek egymás fejébe. Esetenként légből kapott, összefüggésrendszeréből kiragadott számokat. A számokra akkor figyelünk oda, ha robbannak, vagy legalábbis robbanásveszélyesek. Ezért különös figyelmet igényel az adat-előállítás és az adathasználat korrektsége és átláthatósága. A veszélyforrásokról és a tipikus hibákról igyekeztem néhány gondolatot felébreszteni.

## Irodalom

- ARISZTOTELÉSZ [1997]: *Poétika, kategóriák, hermeneutika*. Kossuth Kiadó. Budapest.
- BABBIE, E. [1999]: *A társadalomtudományi kutatás gyakorlata*. Balassi Kiadó. Budapest.
- BEST, J. [2001]: *Damned Lies and Statistics*. University of California Press. Berkeley és Los Angeles.
- BRÓDY A. [1990]: Mennyi az ennyi? Bevezetés a gazdasági mérés elméletébe. *Közgazdasági Szemle*. XXXVII. évf. 5. sz. 521–537. old.
- HUFF, D. [1976]: *How to Lie with Statistics*. WW. Norton and Company Inc. New York.

- HÜTTL A. [2003]: A gazdasági mérés történetéről. Adatok, elméletek, gazdaságpolitika. *Közgazdasági Szemle*. L. évf. 2. sz. 164–182. old.
- JÁNOSSY F. [1963]: *A gazdasági fejlettség mérhetősége és új mérési módszere*. Közgazdasági és Jogi Könyvkiadó. 27–32. old.
- MARX, C. [1955]: *A Tőke*. Szikra Kiadás. Budapest. <http://mek.oszk.hu/04700/04724/04724.doc> (Elérés dátuma: 2010. május 20.)
- SIMMEL, G. [1991]: Money in the Modern Culture. *Theory, Culture Society*. 8. sz. 24. old.
- STIGLITZ, J. E. – SEN, A. – FITOUSSI, J. P. [2009]: *Report by the Commission on the Measurement of Economic Performance and Social Progress*. [http://www.stiglitz-sen-fitoussi.fr/documents/rapport\\_anglais.pdf](http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf)
- SZILÁGYI GY. [1997]: A jelenkori statisztika alapkérdései. *Statisztikai Szemle*. 75. évf. 3. sz. 197–209. old. [http://www.ksh.hu/statszemle\\_archive/1997/1997\\_03/1997\\_03\\_197.pdf](http://www.ksh.hu/statszemle_archive/1997/1997_03/1997_03_197.pdf)
- SZILÁGYI GY. [2000]: Érteni a számok nyelvén. *Statisztikai Szemle*. 78. évf. 1. sz. 5–12. old. [http://www.ksh.hu/statszemle\\_archive/2000/2000\\_01/2000\\_01\\_005.pdf](http://www.ksh.hu/statszemle_archive/2000/2000_01/2000_01_005.pdf)
- SZILÁGYI GY. [2008]: Indikátorok, statisztikák, avagy mire van szükségük a felhasználóknak? In.: Pukli P. (szerk.): *A statisztika és a közigazgatás elkötelezettje*. ELTE Állam-és Jogtudományi Kar – KSH. Budapest. 463–472. old.
- Quotations about Statistics*. <http://www.quotegarden.com/statistics.html>