

Havasi Éva

## Burjánzó számok, mutáns adatok<sup>1</sup> – az adat-előállítás és a számhasználat típushibái-

*Zafír Mihályra emlékezve, tanítványaimnak  
ajánlom<sup>2</sup>*

### Előszó

Munkám során, sőt magánemberként is használom, sőt szeretem a számokat. Statisztikusként becsülöm az adat-előállítás mesterségét és művészetét. Meggyőződésem, hogy a számokban rögzített statisztikai adatok, információk hozzásegítenek világunk megértéséhez, hogy érveinket érdemes, sőt esetenként szükséges is számokkal alátámasztani. A háború a „politika folytatása más eszközökkel”<sup>3</sup>. Ma – és ez jellemzi országunk közéletét is – a számháborúk korát éljük, melyben a harcoló felek számokat röpítenek 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, a tipikus hibákról lesz szó a következőkben.<sup>4</sup> A mottóban idézett „Mester” a vele készült interjúban a következőképpen

fogalmazott: „Soha, „első blikkre”, nem szabad egyetlen számot sem elhinni, semmit nem szabad „kapásból” elfogadni. Minden számot a maga összefüggésrendszerében kell vizsgálni és értelmezni.”<sup>5</sup>

### Bevezetés

Nem létezik „neutrális” statisztika. 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 meta-információs rendszer, amely világossá teszi egy-egy szám tartalmát.

A számok használatuk során legtöbbször elveszítük a mögöttük meghúzódó nézőpontokat, módszertani-mérési megfontolásokat és önálló, független életet élnek. Ugyanakkor, eredetük pontos ismerete nélkül is, a számok újabb és újabb számok építőköveivé válnak. Végeredményként egyre bizonytalanabb és követhetlenebb számtenger születik, melyből ki-ki ízlése, érdeklődése, sőt érdeke szerint válogathat. Így lehet végül egy számmal „bebizonyítani” mindent, sőt annak ellenkezőjét még inkább. A számokat is kritikusan kell kezelni. Kritikusan, de nem szkeptikusan, nem cinikusan és nem is naivan.<sup>6</sup> Ehhez azonban számkultúrára lenne szükség. A tapasztalat az, hogy a betűk világában az emberek

<sup>1</sup> A tanulmány rövidebb változata a Statisztikai Szemle 2010. júniusi számában olvasható. Statisztikai Szemle, 88. évf. 6. szám 648-659. oldal.

<sup>2</sup> Zafír Mihály (1925-2007), a „Mester”, évtizedeken keresztül dolgozott a Statisztikai Hivatalban. Nevéhez fűződnek az „Életszínvonal” kiadványok, egy időben a Létfőnyvonal füzetek, illetve a Fogyasztói-árindex füzetek. Egyik legszebb kiadványa a KSH gondozásában megjelent „Múlt és jelen” c. kötet.

<sup>3</sup> Carl von Clausewitz, porosz tábornoktól származó sokak által idézett mondás.

<sup>4</sup> Ezen írásban megfogalmazott mondanivalóm régóta érelődik bennem. „Mániám” bővebb kifejtése túlnyúlik egy cikk keretein, ezért nehéz volt a helyes arányokat megtalálni. Több verzióban készült el az írás, melyet megosztva néhány kollégámmal és barátommal, kaptam hideget és meleget egyaránt. Nem vagyok negatív gondolkodású, és nem is szeretnék annak mutatkozni, ugyanakkor hiszek a kritikus hozzáállás fontosságában, valamint a reflexió-önreflexió gyümölcsöző szerepében. Ezúton is köszönöm az írás során kapott észrevételeket, építő kritikákat.

<sup>5</sup> Beszélgetés Zafír Mihállyal, Statisztikai Szemle, 2007, 85. évfolyam 3. szám.

<sup>6</sup> Joel Best: Damned Lies and Statistics.

jóval otthonosabban mozognak, mint a számokéban. Az „innumeracy”<sup>7</sup>, azaz a „szám-analfabétizmus”, illetve a számok használatában való jártasság hiánya, valamint a „szám-autizmus”<sup>8</sup>, 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 az írás vizsgálati köréből.

### A számmal kifejezett jelenségek még csak jelenségek, de nem maguk a dolgok

*„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 be értesüléseket. Könnyen átsiklunk azonban a felett, hogy az értesülés, a mérés számszerű eredménye nem közvetlen, és ezért csak átiételesen jellemzi a valóságot.”*<sup>9</sup> (Bródy András)

Számtalan szám vesz minket körül. A számadatok egzaktságukkal 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. A jó Arisztotelész helyett Marxot idézve, a Tőke áru 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 nehezét kifejezzük*, *súlyviszonyba* helyezzük a vassal. Ebben a viszonyban a vas olyan test, amely *semmit nem képvisel a nehézségen kívül*. ...Ha mindkettőt a mérlegre tesszük, valóban azt látjuk, hogy *mint nehéz dolog a kettő ugyanaz*.”<sup>10</sup>

**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 egyik 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.”<sup>11</sup> — mondja Arisztotelész. 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

<sup>7</sup> Az angolban az „illiteracy”-re (analfabétizmus) rímelő „numerecy” a magyar nyelvben nem adható vissza ilyen szellemesen.

<sup>8</sup> A szám-autizmus a poszt-autista közgazdaságtani irányzat kiterjesztése a mai számokat, statisztikai adatokat fetisizáló világunkra. A PAE (Post-Autistic Economics) mozgalom tagjai között már szociológusok, jogászok, és számos szakma képviselői megtalálhatóak. A kifejezés innen származik.

<sup>9</sup> Bródy A.: Mennyi az ennyi? Bevezetés a gazdasági mérés elméletébe. Közgazdasági Szemle, XXXVII: évfolyam, 1990. 5. sz., 521-537. oldal.

<sup>10</sup> Marx, Carl: Tőke, I. 63. oldal.

<sup>11</sup> Arisztotelész: Poétika, kategóriák, hermeneutika, Budapest, Kossuth Kiadó, 1997, 66. oldal.

Simmel a pénz filozófiájának 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 értékesítés a modern ember fásultságához vezet,* ami azt jelenti, hogy az illető elvesztette érzékenységét, hogy a dolgok különbözőségeinek finomságait érzékelje.<sup>12</sup>

**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, konkrétabbá téve a fenti gondolatot, először a számglyártás folyamatára, majd pedig annak végtermékére fókuszálunk.

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

*„A statisztikai adatok azonban inkább hasonlítanak a drágakövekhez, amelyeket először kibányásznak, összegyűjtnek, feldarabolnak, csiszolnak, ... majd kiállítják őket az érdeklődők számára.”<sup>13</sup> (Joel Best)*

Van egy elterjedt vélekedés, mely szerint a számok, „az igazság kis aranyrögci”, aranyrögök, amelyeket hasonlóan a markológéphez, mi is csak úgy összekotorhatunk, írja Joel Best. A statisztikai adatok azonban inkább hasonlítanak a drágakövekhez, melyeket egy bonyolult műveletsor eredményeként kapunk. Hasonlóan az értékesebb drágakövekhez a statisztikai adatokhoz is tanúsítványt mellékelnek. A statisztikát valaki(k), valamilyen céllal, valamilyen módon és módszerekkel előállították. A KSH ma oly divatos nyelvezetére fordítva: a statisztika egy „termék”, s az előállítása egy sajátos „termelési folyamat” eredménye. Az értékesebb statisztikai adatokhoz, mintegy tanúsítványként, minőségi riportok is kapcsolódnak.

**A mért adat az emberek fejében egyenlővé válik, behelyettesítésre kerül a megmért jelenséggel.** A mérési eredmény esetében így módon elhalványul az a tény, hogy a szám egy adott nézőpontból (egy adott értékrendszer mentén), egy adott mérési eljárás alapján ragadja meg a megfigyelt jelenséget. A „számglyártás” legfontosabb minőségi kritériuma — a statisztikai relevancia, ezen a ponton érhető leginkább tetten. **Akkor releváns a számadat, ha érdeklődésre számot tartó,**

<sup>12</sup> George Simmel: Money in the Modern Culture (Pénz a modern kultúrában). Theory, Culture Society, 1991. 8. szám, 24. oldal.

<sup>13</sup> Joel Best, People Count: The Social Construction of Statistics, 11/2002 Talk at Augsburg College. More on future of Statistical Literacy,

értelemmel bíró nézőpontból „számosítja” az adott jelenséget. S ha a ténylegesen használt (mérhetővé tett) és a felhasználók által elvárt fogalmak közötti megfelelés is megvalósul.<sup>14</sup>

Természetesen 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 nézők, a családszociológusok számára pedig meghatározott családi állású személyek. Így az embereket, mint utasokat, betegeket, nézőket, apákat, anyákat, gyermekeket, stb. számszerűsíthetjük. Egy-egy nézőponton belül is döntő a fogalomhasználat megválasztása. A betegeket osztályba sorolhatjuk, és így számszerűsíthetjük, diagnosztizált betegségük szerint (BEO, azaz betegségek egységes osztályozási rendszerét használva), gyógyulási idejük szerint, táppénzes napjaik száma szerint, betegségük örökletes jellege, maradandó károsodással járó volta, kórházi ellátással járó, vagy szakellátást igénylő volta szerint, stb.

**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 pedig az ezekből kialakuló mintázatok alapján újabb, társadalmunk rétegződését leíró, kategóriákat, osztályokat alkot. Így születnek olyan új osztályok, mint „gazdasági elit”, „vagyonos középosztály”, vagy „elszegényedett alsó középosztály”, és így tovább. Egy-egy

klasszikus statisztikai osztályozási rendszer változása is jól mutatja ezt a folyamatot. Az orvostudomány fejlődésével, a diagnosztikai eljárások finomodásával például a korábban említett BEO is, bizonyos betegségcsoportok szétbomlásával és újabb betegségcsoportok létrejöttével, módosul. Ugyanez elmondható a Foglalkozások Egységes Osztályozási Rendszeréről (FEOR), vagy az utóbbi időben állandóan napirenden levő Tevékenységek Egységes Ágazati Osztályozási Rendszeréről (TEÁOR). Az állami bürokrácia nap mint nap teremt új kategóriákat, új osztályokat, új, korábban nem létező embercsoportokat. Ma statisztikai adatokat közlünk a „rátosokról”, tegnap még a „jövedelempótló támogatásban részesülőkről”, ma a „munkanélküliekről”, tegnap a „közveszélyes munkakerülőkről”.

Még „stabil” kategóriák mellett sem könnyű az osztályba sorolás. Jánossy példájával élve a tehenet és a disznót mérete alapján nem nehéz osztályba sorolni, ugyanakkor a tehén és a ló méretbeli összehasonlítása már nehezebb. Bródy, Jánossynak<sup>15</sup> a golyóscsapágyra és kavicsra vonatkozó példáját továbbfűzve azt fejtegeti, hogy a kavicsok osztályozása esetén nem csak a formagazdagság nehezíti a méretbeli összehasonlítást, hanem a gondok már ott elkezdődnek, hogy mettől meddig kavics a kődarab? Miközben a hétköznapokban gondolkodás nélkül használjuk a „gyerek” fogalmát, a társadalomstatisztikusok számára a gyerekes családok vizsgálata esetén komoly nézetkülönbségek vannak a fogalom meghatározása, „operacionalizálása” körül. A szüleivel élő 30 éves

<sup>14</sup> Szilágyi, 198. oldal

<sup>15</sup> Jánossy Ferenc: A gazdasági fejlettség mérhetősége és új mérési módszere, 27-32. oldal.

agglegény bizonyos szempontból gyerek, de a gyerekes családok vizsgálatának ez a családtípus semmiképpen nem része. Ha 20 évnél húzzuk meg a határt, akkor kérdéses, hogy a már keresettel rendelkező „gyerek családi állású” személy gyermeknek tekinthető-e? És így tovább.

**A relevancia esetében az érdeklődésre számot tartó nézőpont mellett, a „fit for use” részét képezi a „kategóriába sorolás” mellett 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 bontásban, hanem a település népességszáma alapján képzett nagyság-kategóriák 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 „3 napi járóföld” a távolság idővel való meghatározásának szép példája. A kifejezés utal a korra, amelyben helye és relevanciája volt az effajta „mérési módnak”. 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 érdemes inkább használni.<sup>16</sup>

### **A számok önállósodása, és ennek következményei**

A számok, elveszítve a háttérükben meghúzódó tartalmi és módszertani-mérési megfontolásokat, önálló életet kezdenek élni, s relevanciájuk sokszor túlnő saját határaikon, „érvényességi körükön”. Egy példát említve, a mai társadalmunkban a szegények aránya lehet 5-8-10-25, de akár 100% is. Attól függően, hogy mivel mérjük a szegénységet, és hol húzzuk meg a szegénységi küszöbszintet. Ha „csak” jövedelemmel mérjük, akkoris fennáll az a kérdés, hogy mekkora összegnél húzzuk meg a határt. 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 számot, nem téveszthetjük szem elől annak mögöttes mérési történetét.**

Persze vannak már sztemderdek bizonyos jelenségek mérésére. Ezek azonban egyfajta 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 meghúzódó mérési nézőpont, vagy a mérési metódus. Ennek egyik eklatáns példája a GDP széleskörű — gazdasági fejlettségre, jólétre vonatkozó — fetiszizált használata. Bár a GDP kiszámításának többszáz oldalas dokumentációja van (hosszú lábjegyzet lenne belőle), és igény szerint szinte bárki számára hozzáférhető, érthető módon, kevesen böngészik a számhasználók közül.

<sup>16</sup> Hüttl Antónia: A gazdasági mérés történetéről: adatok, elmélet, gazdaságpolitika c. Közgazdasági Szemlében megjelent írása.

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, felróva a környezetvédelem szempontjainak hiányát (zöld számlák), az információs társadalom teljesítményének figyelmen kívül hagyását, a magánháztartás keretei között végzett munkák felértékelésének elmaradását, stb.<sup>17</sup>

Mások szerint mi végre a magas GDP, 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. Egyre többen, merőben közgazdasági szempontból is, a (jövedelmi egyenlőtlenséggel, bizonyos termelési eredmények elhagyásával) korrigált GDP-t ajánlják<sup>18</sup>

„Ha a GDP felmegy, Amerika miért megy lefelé?”<sup>19</sup> című sokat idézett már 1995-ben megjelent írás is ezt a kérdést feszegeti. Hasonló gondolat jelenik meg Gyurcsány Ferencnél, amikor egy interjúban a következőket mondja: „nem csak a számoknak kell jóknak lenniük, hanem az életminőségnek is”.<sup>20</sup> A számok interpretációs kerete nem feszíthető túl a ténylegesen megmérten. Valószínű, hogy a GDP emelkedésével az emberek jóléte is nőni fog, de ez számos egyéb tényezőtől, például a jövedelmek

elosztásának egyenlőtlenségétől, is függ. Lehet, hogy a GDP nagy része az államadósság törlesztésére megy, a környezet lerabolásának árán, vagy bizonyos társadalmi csoportok aránytalanul nagy tehervállalásán keresztül valósul meg.

Ferge Zsuzsa egy vele készített interjúban ezt szellemesen úgy fogalmazta, hogy „mielőtt egy számot használunk érdemes kinézni az ablakon”.<sup>21</sup> Ez segít abban, hogy a számok interpretációjánál, ne veszítsük el a valósággal való kapcsolatunkat. A számok soha nem helyettesíthetik a valóságot a maga komplexitásában, ugyanakkor értő használat mellett hozzájárulnak világunk megértéséhez.

### A mérhetetlen mérhetővé tételének csapdája és a szám-autizmus

*A lélek ereje mérhetetlen. Mégis méri a bennünk lévő végtelent. "Mérni magunkban mérhetetlent" (József Attila)*

A címben jelzett szám-autizmus<sup>22</sup> a számokba való megszállott belefeledkezésre utal. Arra az érzésre, hogy a számok és az indikátorok özőnében elveszítjük a valósággal való kapcsolatunkat, kiszakítjuk magunkat a reális világból. A mérhetetlen mérésére tett kísérletek, a valóságtól való eltávolodásunk, a gyors és „egzakt” válaszok iránti vágyunk — internetről kapásból kiragadott — példái a „szerelem-kalkulátor”<sup>23</sup>, vagy a „szerelmi hőmérő”. Az egyikből megtudom, hogy kit kell választanom, a másik kiszámítja, hogy mennyire

<sup>17</sup> Számos tanulmány mellett lásd az egyik legfrissebbet, Sarkozy francia elnök megrendelésére készült bizottsági jelentést. J. E. Stiglitz, A Sen, J. P. Fitoussi: Report by the Commission on the Measurement of Economic Performance and Social Progress.

<sup>18</sup> Az írás kereteit szétfeszítené a GDP mérőszámának és korlátainak részletes bemutatása. Itt csak utalunk néhány e tárgykörben született meghatározó jelentőségű írásra.

<sup>19</sup> Lásd többek között: Beyond GDP: Measures of Economic, Social and Environmental Progress, 2. oldal. Az „If the GDP is Up, Why is America Down?” című cikk az Atlantic Monthly hasábjain jelent meg 1995 októberében.

<sup>20</sup> Gyurcsány Ferenc: Reformok megszorítás nélkül, Metro, 2008. május 9. 2. oldal.

<sup>21</sup> Interjú Ferge Zsuzsával, Statisztikai Szemle, 2003. augusztusa, 702-704. oldal.

<sup>22</sup> A szám-autizmus a poszt-autista közgazdaságtani irányzat kiterjesztése a mai számokat, statisztikai adatokat fetiszizáló világunkra.

<sup>23</sup> Szerelmekalkulátor, lásd <http://szeretnemszeret.info/?page=80844>

vagyok szeretve. Igaz némi pénzbe kerülnek a válaszok, de nincs bonyolult dialektika, „egzakt”, „egyértelmű” válaszok vannak.

Túl gyakran próbáljuk meg kvantifikálni azt, amit valójában nem tudunk, vagy csak nagyon korlátozottan tudunk megmérni. Legtöbbször a kvantifikálás folyamán számos csúsztatást is végrehajtottunk, 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? (1. válasz: Mert nem is akar!, 2. válasz: Mert az állam nem gondoskodik arról, hogy mindenkinek legyen munkája!, 3. válasz: Mert az emberekben nincs kezdeményezőkézség, rugalmasság,...) 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 egyenlőségelet teszünk bizonyos jövedelem nagyság és a szegénység közé. Alacsony jövedelem egyenlő 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.

A mérési igyekezet szép példáit eleveníthetjük fel a szépségversenyek, művészeti vetélkedők pontozására gondolva. Például a vadászgörények szépségversenyének pontozási rendszerét<sup>24</sup> tanulmányozva, nyilvánvalóvá válik, hogy a görények estében a hétköznapi ember számára oly fontos „kisugárzás” (mivel számunkra éppen ettől olyan nevezetesekek) nem játszik. Ugyanakkor 15 pontot ér a bunda és ugyanennyit a vérmérséklet és kezelhetőség. A tökéletes fülekért és farkakért 5-5 pontot lehet kapni. Szerencsére az összesen 100 pontból 10 pont erejéig lehetőség van az „esztétikai összkép” értékelésére. Amit, a magam részéről egyébként a szépség szinonimájának gondolnék. Felgyorsult világban élünk, ahol „az idő pénz”. Nem szánunk időt Arany János költeményeinek komplex megközelítésére — hovatovább élvezésére sem —, de pár perc alatt meg tudjuk mondani kompjúterünk segítségével, hogy hány szóból áll a Toldi, vagy

<sup>24</sup> Lásd: Vadászgörények küllembírálati kritériumai, a szépségverseny alapvető szabályai és menete.

Arany János életműve. **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. Meg hát — valljuk be — igény se nagyon. Mit is kezdhethetnénk vele?

És mindezek után jönnek a különböző mérési-módszertani hibák, értelmezési nehézségek, csúsztatások, torzítások. Akkor mire való mégis ez a számosítási igyekezet? Egy olyan világban élünk, amikor minden megmértetik, s „ami meg van mérve, az el is van végezve”<sup>25</sup>. 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.** Ennek egyik látványos kifejeződése a ma oly divatos indikátorhasználat. Indikátorokkal mérjük a dolgozók, a vállalat, az ország teljesítményét, hazai és nemzetközi projektjeink előrehaladását és eredményességét, a „fenntartható fejlődést”, a „haladást” és mindent. **Amit nem mérünk valamilyen mutatószámmal az a teljesítmény praktikusán nem létezik. Az lesz az „eredmény”, ami eredményességi mutatónk**

**részét képezi.** Ha gyermekünket megdicséri a tanító néni, annak örülünk, de mégis inkább arra vagyunk kíváncsiak, hogy kapott-e jegyet az iskolában? Ebből lesz a gyerek bizonyítványa, s majd ezt cipeli magával tudása bizonyítékaként egy életen át.

### Mutáns statisztika, számokat szülő számok

A gyakorlatban a „háttérüket” elvesztő számok önálló életet élnek, továbbszámolnak velük, összehasonlítják őket más számokkal. A GDP, a GPI (Genuine Progress Indicator, azaz a Valódi Fejlődés Indikátora), az ISEW (Index of Sustainable Economic Welfare, a MEW (Measure of Economic Welfare, azaz a Gazdasági Jólét Mérőszáma), a HDI-index (Human Development Index) és egyéb vég nélkül sorolható mutatószámok építőkockájává válik. Az inflációs rátát a különböző időpontra vonatkozó adatok esetében automatikusan használjuk, anélkül, hogy foglalkoznánk kiszámításának menetével, korlátaival. A nyugdíjak értékállóságának vizsgálatánál, idős mesterem 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 fejükbe, használják, és bekerül a köztudatba. Ugyanakkor a számok egyre több újabb számot szülnék. Így születik a mutáns statisztikák legtöbbször.

Joel Best szellemes könyvében azt fejtegeti, hogy vannak olyan statisztikák, számok, amelyek torzszülöttként jönnek a világra. Míg mások később,

<sup>25</sup> „What gets measured gets done”.

a használatuk során torzulnak el. A torzításhoz komoly érdekek is fűződhetnek. Egy „jól megválasztott”, vagy „átértelmezett” szám alkalmas lehet a közvélemény hangulatának felkorbácsolására, vagy akár lehűtésére, öngazolásra és mások vádolására, félelemkeltésre és sok minden egyébre. A téves információ téves következtetésekhez vezet, s ez elhibázott döntéshozatalt, hibás cselekvést eredményez(het). Ezért a kritikátlan adathasználat komoly veszélyeket rejt magában. A hibás, vagy hibásan értelmezett szám továbbél és mutáns statisztikák sokaságát hozza világra, melyek nem csak szakmai, hanem etikai kérdéseket is felvetnek. Éles szakmai viták folynak arról, hogy lehet-e korrigálni a lakosság bevallott jövedelmi adatait, természetesen megfelelő külső, megbízhatóbb adatok segítségével? Ha a vállalkozók és a nyugdíjasok bevallott jövedelmét összehasonlítom, akkor — minden lakossági adatgyűjtés szerint — a nyugdíjasok sokkal több jövedelemmel rendelkeznek. Kérdés, hogy a két társadalmi csoport jövedelemadatainak megbízhatósága azonos-e? Eltérő megbízhatóságú adatok esetében az egyszerű, mechanikus összehasonlítás szakmailag és etikailag is megkérdőjelezhető. Ettől még igaz lehet, hogy a nyugdíjasok több jövedelemből élnek, de az összehasonlítás nagyobb körültekintést igényel.

Best az adathasználókat három nagy csoportra osztja: vannak a naivak, a cinikusak és a kritikusak. 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 komoly 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.

### **Ne szólj „szám”, avagy kiáltó számok — a statisztikai adatok relevanciájának kiemelt fontossága**

*„Legtöbbször nem hazugsággal,  
hanem hallgatással kendőzzük el az igazságot”  
(Szent Ágoston)*

*„A politikai küzdőtér harcosai nem a tudományos igazságot keresik, még kevésbé a pontosságot.  
Lőfegyverre van szükségük az információs háborúban.” (Alvin Toffler)<sup>26</sup>*

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

<sup>26</sup> Az idézet a „Data, Information, Knowledge, and Wisdom cikk egyik mottójának laza fordítása.

hazudjunk a statisztikával?” c. munkája 1954-ben jelent meg először, s mindmáig a statisztikai adatok interpretációjának egyik legfontosabb „alaptankönyve”. 2003-ban kínai nyelvre is lefordították. A siker titka a könyv szöveges tartalma mellett képanyaga. Mindjárt a címdalán egy joviális 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, témákra vonatkozóan 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.

Legtöbbször úgy hazudunk, hogy „mellébeszélünk”. Mellébeszélhetünk szavakkal és számokkal egyaránt. Lehet, hogy hiányoznak a valóság adekvát kifejezéséhez szükséges szavaink, kategóriáink, számaink, máskor lennének, de inkább nem azokat használjuk. Mai társadalmunkban valós problémát jelent, hogy a rendszerváltás utáni világunk leírására még nincs elméletünk, nincsenek releváns kategóriáink és így releváns számaink sem. Sokszor azonban csak „mellébeszélünk” és „mellészámolunk”. „Ne szólj szám...!” Amikor forráshiányos településen élő, aktív korú alacsony

iskolai végzettségű inaktív munkaeő-piaci helyzetét számszerűsítjük, akkor érdemes, sőt releváns lenne az etnikai hovatartozás figyelembe vétele is. Így egzakt módon, számszerű formában megadható lenne a roma emberek munkanélküliségi-, inaktivitási rátája. Ferge a vele készült interjúban ezt így fogalmazta: „Van az indirekt hazugság, amely sokkal bonyolultabb. Amikor csak olyan kérdéseket tesz fel, csak olyan adatokat gyűjtök, amelyek a hatalmat nem zavarják.”<sup>27</sup> A relevancia tartalom és forma egyszerre. Arra vonatkozóan gyűjtünk adatokat, amelyekre valós igények vannak, és olyan kategóriákat alkalmazunk, amelyek közelebb visznek világunk megértéséhez, a megfelelő mérőeszköz, mértékegység megválasztásával. Ez utóbbit a „3 napi járóföld” példájával már korábban illusztráltuk. Hiába pontos, érthető, összehasonlítható, naprakész és hozzáférhető egy számadat, akár erős módszertani háttérrel is, ha teljesen érdektelen.

**A statisztikai adatokat gyakran fegyverként használják a politikai küzdőtéren. Minél nagyobbat szól egy szám, annál jobb.** Az igazság ez esetben másodlagos. „Fogy a magyar”, halljuk, majd egy konzervatív folyóiratban a következő adatot olvassuk: 2002-2006 között a népesség száma 10 millió 175 ezerről, 2006-ra 10 millió 077 ezerre csökkent. Vagyis a szocialista kormányzás négy éve alatt a lakosság száma 98 ezerrel csökkent. *A szám igaz, az interpretáció hamis!* A baloldali sajtóban ezt a számadatot a következőképpen fogalmazhatnánk meg: 2002-2006 között a magyar lakosság száma

<sup>27</sup> Ferge Zsuzsával készült interjú, 705. oldal.



kisebbs mértékben csökkent, mint 1998 és 2002 között (a jobboldali kormányzás idején). Az akkori 105 ezres népességszámcsökkenés helyett kormányzásunk négy éve alatt mindössze 98 ezerrel lettünk kevesebben. Ez a szám is igaz, de az interpretáció továbbra is hamis. A népesség csökkenése ugyanis rendszereket és kormányokat átívelően 1980 óta folyamatosan, kisebb-nagyobb ütemben, de csökken. Ez nem hazai sajátosság, hanem szerves része a fejlett országokra jellemző nemzetközi trendnek.<sup>28</sup> A számokat interpretációval kell ellátni. Az interpretációs tévedésekkel, hamisításokkal szemben ugyanolyan szigorúan kell fellépni, mint a számok előállításánál során vétett hibákkal szemben. Az adathasználat típushibáiról később lesz szó. Itt az adat-előállításban, illetve használatban rejlő eltérő érdekek felvillantása volt a cél.

### Mutánsok és torzszülöttek - a hamis adathasználat tipikus példái

*„Az emberek nyolcvan százaléka átlag fölöttinek tartja magát.”*

„Sokkoló írás: Statisztikai adatok Magyarország állapotáról 47 tételben”, olvasom az egyik internetes portálon. A következőkben a Hítel c. folyóiratban megjelenő cikk „sokkoló” számait idézik. Az eredeti írás, mely 47 területről emel ki egy-egy statisztikai adatot, mutatószámot, döntően két időpontra vonatkozóan, mindenhol nélkülözi a forrásmegjelölést. A számokat egy-egy jól megválasztott irodalmi idézet zárja, itt szintén

irodalmi műveltségünkre hagyatkozhatunk csak. Példaként álljon itt közülük két tétel:

„6. Minimálbér

1998: 19 ezer Ft, 2002: 50 ezer Ft, 2008: 69 ezer Ft

„...*Nem kérdez senki semmit...*”, vagy egy másik:

„23. Táppénzes napok száma

2001: 43 millió nap, 2006: 36 millió nap.

„...*bordott sérvig tönköt, gabonát...*”<sup>29</sup>

Az egyik hetilap és annak internetes portálja átveszi, és újabb interpretációval bővíti az idézett folyóiratban megjelent számsorokat. Valahogy így születik a mutáns statisztika. Egyébként a KSH kereseti adatai szerint a minimálbér 1998-ban 19.500 Ft volt.<sup>30</sup> A minimálbér az akkori bruttó átlagkereset 28,8%-a volt, 2002-ben 40,8% (se korábban, se később nem volt ilyen magas) és 2008-ban 34,7%-a. Ez utóbbi nagyon közel van az 1992-es, 1993-as arányokhoz. Egyébként 1992-től napjainkig a minimálbér bruttó bérben kifejezett aránya erősen ingadozó. 1999-ben és 2000-ben volt a legalacsonyabb szinten (29,1%), 2001-ben ugrásszerűen emelkedett (38,6%), majd 2002-ben tovább nőtt, s attól kezdve kis mértékben, de folyamatosan csökken. Ha a minimálbér összegét nemzetközi összehasonlításban vizsgáljuk, akkor, euróban kifejezve, 2009 elején Magyarország a 16. helyen áll a 270 eurós havi minimálbérrel. Ugyanez vásárlóerő-paritással számolva 408 PPS, ami a 15. helyre volt elegendő.<sup>31</sup>

<sup>29</sup> Heti Válasz 8. évfolyam 46. száma, illetve Hítel, 2008. október, Csendes Csaba írása.

<sup>30</sup> Forrás: KSH portál, kereseti adatok, minimálbér 1992-től, [http://portal.ksh.hu/pls/ksh/docs/hun/xstadat/xstadat\\_eves/tabl2\\_01\\_40i.html](http://portal.ksh.hu/pls/ksh/docs/hun/xstadat/xstadat_eves/tabl2_01_40i.html)

<sup>31</sup> Hol kaphatjuk minimálbérként a hazai ötszörösét! Nézze meg!, 2009-07-31, VGO.

<sup>28</sup> Forrás Magyar Statisztikai Évkönyv, 2007, KSH, Budapest, 2008. 1. oldal.



A táppénzes napok számánál a KSH adatai közül<sup>32</sup> két dolgot mindenképpen figyelembe kell venni. A táppénzes napok száma függ a táppénzre jogosultak számától és a táppénzen töltött időtől. 1990-ben még 4,5 millió táppénzre jogosult személy volt. Számuk a kilencvenes évek első felében erőteljesebben, majd ezt követően kisebb mértékben, de egyenletesen csökkent. 1990-ben az egy jogosultra jutó táppénzes napok átlagos száma még 22 volt, ami 2006 után 10 nap alá csökkent. Az interpretáció kicsit bonyolultabb a használt szép idézetnél. Mindkét példánál figyelembe kell venni a jogszabályi változásokat és a mögöttük meghúzódó érdekeket. A hosszú táppénz napjainkban már nem menekülési útvonal a munkahely, illetve az álláshely megszűnte elől. A magánszektor erősödésével egyre nagyobb a távolság a munka ideje alatt realizált kereset és a táppénz között. Sokan kisebb magáncégeknél vannak alkalmazásban, vagy önmagukat alkalmazzák, ezért sokkal nehezebb táppénzre menni. Az időbeli összehasonlítást nehezíti a táppénzszabályok módosítása is. A lakosság egészségügyi állapotának változása, bizonyos szempontból javulása, az egyes betegségek gyógyítási-gyógyulási idejének változása is tényező lehet.

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, még inkább,

elmúlt félevesi jövedelmi adatokhoz viszonyítom a lakosság mai jövedelmi helyzetét. Részben ebbe a kategóriába sorolható a magyarok fogyasztást bizonyító korábbi példánk.

A mutáns statisztikák eklatáns példái az egyre inkább elharapódzó *telefonos, betelefonálás felmérések*. Az ilyen-olyan megrendelések alapján végrehajtott telefonos adatgyűjtések sokasága megkérdőjelezhető eredményeket produkál. Ha a Tv műsor egy meghatározott irányultságú célközönségnek szól, akkor a csatorna által megrendelt adatgyűjtés eredménye is nagy valószínűséggel elégedetté teszi a nézőket, hiszen saját véleményükkel megegyező végeredményt láthatnak a képernyőn. „Igaz” az adat, mert egyezik a néző véleményével is. Ugyanezeket a kérdéseket szó szerint feltéve egy másik csatorna másik műsorában nagy valószínűséggel más eredményeket kapnánk, mellyel annak a műsornak a nézőközönsége legalább annyira elégedett lesz. A csapda elkerülését szolgálják az ún. reprezentatív adatgyűjtések.

A *hibás mintakiválasztásból adódó hamis eredmény* legismertebb példáit a választási eredményekre vonatkozó előrejelzések tévedései adják. A Literary Digest Egyesült Államok elnökeire vonatkozó 1936-os választási veresége annak volt köszönhető, hogy a megkérdezettek kiválasztása telefonkönyvek és gépkocsi-tulajdonosok névjegyzékei alapján történt. Az így kiválasztott emberek „felülreprezentálták a jómódú, városi, képzetesebb 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

<sup>32</sup> Forrás: KSH portál, STADAT, Egészségbiztosítás, táppénz, 1990-, [http://portal.ksh.hu/pls/ksh/docs/hun/xstadat/xstadat\\_eves/tabl2\\_04\\_03i.html](http://portal.ksh.hu/pls/ksh/docs/hun/xstadat/xstadat_eves/tabl2_04_03i.html)

csoportja a teljes választói népességhez viszonyítva. Az akkor felbukkanó G. Gallup újfajta mintakiválasztási módszerével, az ún. kvótás kiválasztással, helyesen jelezte előre Roosevelt sikerét Landonnal szemben. Itt jegyeznénk meg, hogy a valószínűségszámítás 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 cm, akkor az adott mintanagyság mellett azt is meg tudjuk válaszolni, hogy a definiált 95%-os, 99%-os, valószínűségi szinten +/- hány centiméter eltérés mellett „jó” a mintából nyert adatunk. Vagyis például 99%-os valószínűséggel állíthatjuk, hogy a tanulók átlagmagassága 163 és 167 cm 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.

„A lelkiismeretes embereknél ritkább az Alzheimer-kór”, olvasom a főcímet az InforMed c. hírújságban. „Chicagói kutatók összefüggést véltek felfedezni a lelkiismeretesség és az Alzheimer-kór előfordulása között. Ezer idősebb katolikus apácát, papot és szerzetest vizsgáltak meg, egyikük sem szenvedett demenciában. A résztvevők kérdőíveket töltöttek ki személyiségükre, kórtörténetükre, fizikai és szellemi aktivitásukra vonatkozóan. A személyiségtesztek alapján a kutatók lelkiismeretességi pontszámokat

határoztak meg. A 12 éves vizsgálat alatt a résztvevőket évente orvosi vizsgálatnak vetették alá, és többek közt memóriateszteket is ki kellett tölteniük. Kiderült, hogy a magasabb lelkiismeretességi pontszámmal rendelkezőknél ritkábban fordult elő Alzheimer-kór, mint azoknál, akiknek a pontszáma alacsony volt. Ez az összefüggés érvényes maradt a többi tényezővel való egyeztetés után is”. Bár a vizsgálat hosszú ideig tartott, a mintába került emberek nem képviselik az amerikai társadalom széles spektrumát. Az Alzheimer-kór egyre komolyabb betegség, de előfordulási gyakorisága még mindig alacsony ahhoz, hogy egy ezer fős önkéntesekből álló minta elégséges legyen. A 65 éves lakosság körében becslések szerint kb. 1%-os az előfordulása. A nők gyakrabban betegszenek meg és egyes koponyasérülések, pajzsmirigybetegségek, depresszió és a keringés rendellenességei (érelmeszesedés, magas vérnyomás) fokozhatják a kockázatot.<sup>33</sup> Ez azt jelenti, hogy ezer emberből 10 körüli betegségszám a reális. Ehhez nem kell túl sokat hozzátenni. Nyilvánvalóan az évek előrehaladtával, a betegség megjelenésével a lelkiismeret mérésére szolgáló tesztek kitöltése is egyre nehezebb, s ez megmutatkozik a teszteredményekben. És akkor még arról az apróságról ne is beszéljünk, hogy mi is a „lelkiismeretesség” kritériuma?

Az *óvatlan, hamis adatértelmezés* területéről végtelen példát idézhetünk. Az egyik iskolapélda az Amerikai Gyermekvédelmi Alap Évkönyvének szövege, melyet egy amerikai folyóirat főcímben idézett: „A pisztoly általi gyermekgyilkosságok száma az

<sup>33</sup> Lásd <http://www.hazipatika.com/services/betegseglexikon/view?id=46>.



Egyesült Államokban évente megduplázódik.” A pistollyal 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 gyilkosságok száma – amennyiben évi egy gyilkossággal számolunk - az éves duplázódás eredményeként 1965-re 32768-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 végbement. Amit tudunk az két adat két időpontra, 44 év elteltével. Néhány napja olvastam az egyetemistáknak, főiskolásoknak készült magazinban ”A kutatók szerint 2020-tól kezdődően minden évben egy évvel meg tudják növelni az átlagéletkort...”<sup>34</sup> Először is az átlagéletkor nem azonos a születéskor várható élettartammal. A mai magyar lakosság egészét tekintve az átlagéletkor 40 év, ebben benne vannak az újszülöttek és az „ezüstkorúak” is. Ha több gyerek születik, akkor az átlagéletkor csökken, ha több idős van, akkor növekszik, akkor is ha életkilátásaik változatlanok maradnak. Ezt hívják összetételhatásnak. A születéskor várható élettartam 2007-ben a férfiaknál 69,19 év volt, egy évvel később 69,79 évre emelkedett. Ez a közel 1 éves javulás számos társadalmi-gazdasági, egészségügyi- tényező együttesének eredménye, de semmiképpen sem Michele Rose kutató csodaszerének hatása. Bízva a kutatási eredményekben és elfogadva, hogy az évi

egy éves emelkedés a várható élettartamra vonatkozik, akkor 50 év múlva egy magyar férfi, minden egyéb életkort befolyásoló hatást kiszűrve, 119 életévre számíthat születésekor. A nők születők pedig 132 évet remélhetnek. Ha az országok közötti várható életévekben meglévő mai különbségek fennmaradnak, akkor egy svéd nő másfél évszázadnyi életre számíthat születésekor és még többre, ha elérte a 20-30-40 éves életkort. Gondoljuk el mi lesz 100 év múlva! Jó hinni egy-egy hírnek, hiszen bearanyozza a napot, kivéve, ha statisztikusként dolgozunk.

*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 40km/óra sebességgel tesszük meg, akkor könnyen azt gondolhatjuk, hogy átlagsebességünk 50 km/óra volt. Ugyanezt a hibát követjük el akkor, amikor a fővárosban élő 5%-os szegénységi rátát és a vidéki 15%-os szegénységi rátát átlagolva, a hazai szegénységi rátát 10%-nak gondoljuk. A szegénységi rátákat nem adhatjuk össze (ha Budapesten 5% és vidéken 15%, akkor együtt, országosan 20% — 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%-os szegénységi ráta a lakosság alig egyötödére vonatkozik, míg a 15%-os szegénységi arány a lakosság négyötödére. A szegénységi ráta értelemszerűen nagyobb lesz, mint 10% (valahogy így:  $4/5 \cdot 15\% + 1/5 \cdot 5\%$ , ami 13%).*

<sup>34</sup> Moha, 2010. január, 4. oldal. Az írás a hplusmagazine.com-ra hivatkozik.

A hétköznapi emberek szeretik az átlagokat. Amikor statisztikai adatokat használnak, legtöbbször, tudatosan, vagy látens módon, de átlagokról beszélnek.

*„Egyik lábunkkal egy vödör jeges vízben, másikkal egy vödör forrásban levő vízzel üldögélve, átlagban komfortosan érezhetjük magunkat.”*

Egy átlagos napon egy átlagember átlagosan „x” óra tévét néz, „y” Ft-ot költ makarónira és „z” mértékben elégedett a családi életével. D. Boyle, szellemes könyvében, rengetek példát hoz erre a gyakorlatra. „Az emberek életük során (átlagosan) 8 pókot esznek meg álmukban.” „Az Egyesült Királyságban az emberek szexre fordított ideje (átlagosan) 21 perc, Olaszországban (átlagosan) 14 és Brazíliában (átlagosan) 30 perc.” „Az angol iskolás gyerekek 4%-a szerint Adolf Hitler volt a brit miniszterelnök a II. világháború idején.”<sup>35</sup>

*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) átlagiövedelemnek és a medián alapú átlagiövedelemnek is van értelme. A kettő azonban nem ugyanaz. Ha egy társadalomban nagyok a jövedelmi különbségek, akkor a számtani átlagiövedelem 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 átlag változásából nem lehet elhamarkodott következtetéseket levonni. Amikor arról beszélünk, hogy az elmúlt 20 évben a tévézésre fordított idő lényegesen nőtt a hazai lakosság körében, akkor azt is figyelembe kell venni, hogy a népesség öregszik, nőtt az inaktívok száma, így a tévézésre fordított átlagidő mögött a lakosság kormegoszlásának és gazdasági aktivitásának a hatása is megmutatkozik. Bizonyos esetekben, a tévézési szokások vizsgálatánál ezt az ún. összetételhatást érdemes kiszűrni. Az átlagot interpretációval kell ellátni, s ezzel mintegy megerősítettük a korábbi állításunkat is.

*Minden negyedik ember szenved valamilyen mentális betegségben. Nézd meg három barátodat, és ha ők rendben vannak, akkor te vagy az.*

*Beteg: Túl fogom élni ezt a nagyon kockázatos operációt?*

*Sebész: Minden bizonnyal.*

*Beteg: Miért olyan biztos ebben doktor úr?*

*Sebész: Tízből kilenc beteg meghal az operáció során és tegnap hunyt el a kilencedik.*

A statisztika a tömegesen előforduló jelenségekre fókuszál, az adatok pontosságának, megbízhatóságának meghatározása a nagy számok törvényén alapul. Vannak teljeskörű megfigyelések, amikor minden érintettet megmérnek, de a gyakorlatban legtöbbször a vizsgált sokaság egy részének megfigyeléséből következtetnek az egészre. A teljeskörű megfigyeléseknek anyagi- és technikai akadályai, illetve időbeli korlátai egyaránt vannak. Mint ahogy erről korábban már szót ejtettünk, a valószínűségszámítás alapján az egészből vett „jó” minta segítségével a megfigyelt részsokaság tulajdonságai alapján az egész sokaságra vonhatunk le következtetéseket. Így igaz lehet, bizonyos

<sup>35</sup> Idézett mű 231. oldal.

megbízhatósági szinten, bizonyos eltérést elfogadva, hogy egy adott betegséget operálva tízből kilenc beteg meghal. Ez az összefüggés azonban nem ad választ az összefüggés érvényesülésének sorrendjére. Lehet, hogy először 10 beteg meghal, s ezt követően 90 egymás után életben marad, de történhet bármilyen más sorrendben is. Indiába sikerült a legolcsóbb jeggyel utaznunk, mivel a jegyvásárlást megelőzően két gépe is lezuhant az adott légitársaságnak. Statisztikusként nem utaztam nagyobb biztonságban, hogy a „neheze már mögöttünk van”, ugyanakkor a drágább jeggyel utazók sem választottak jól, mivel a légitársaság repülési mutatói nem voltak rosszabbak, mint a konkurens légitársaságoké.

Az egyik leggyakoribb hiba a *túláltalánosítás*, a *szekletív/előítéletes észlelés*. Ez utóbbira történt már utalás, amikor a célközönség véleményét tükröző adatot a célközönség „igaznak”, „valósnak” éli meg. A jelenség lényege, hogy túl hamar megállunk a megismerésben. A barátnőm sokat beszél, jön a következtetés: a nők szeretnek beszélni. A statisztikusok honlapján erre vonatkozóan számos vicces példát találunk, többek között a mottóként választott fenti kettőt is.

*„Bizonyított tény, hogy azok élnek legtovább, akik a saját születésnapjuk megünneplésében is vezetnek.”*

Nagyon 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.

A nélkül, hogy az összes adatértelmezési hibaforrást lajstromba vennénk, hozzunk egy példát az *összetételhatás figyelmen kívül hagyására*. A napokban megjelent egy újsághír: „Létszámstopot rendelt el a államigazgatásban a miniszterelnök – jelentette be a kormány szóvivő. Év elejétől amúgy bő 50 ezerrel hízott a költségvetési szektorban dolgozók létszáma.”<sup>36</sup> A növekedett helyett a „hízott” szó használata már sejtet valamit. Az 50 ezer fővel történő növekedés pedig némi szakmai felvezetést igényel. Tapasztalati tény, melyet hosszú idősorok támasztanak alá, hogy az év elejéhez képest az év közepére a munkavállalók száma növekszik. A tendencia, létszámokban megnyilvánuló szabályszerűség, 2009-ben is igaznak bizonyult. Ez a közsférában döntően annak a „pozitív” ténynek volt köszönhető, hogy az év elejéhez képest az ágazatban erőteljesen nőtt, már-már „hízott” a közmunkában foglalkoztatottak száma. A

<sup>36</sup> Lásd az Index honlapján, a gazdaság hírei között. [http://index.hu/gazdasag/magyar/2009/07/28/letszamstopot\\_vezetnek\\_a\\_kozszferaban/](http://index.hu/gazdasag/magyar/2009/07/28/letszamstopot_vezetnek_a_kozszferaban/)

költségvetési szervek és intézményeinél a közmunkások januári 20 ezer alatti száma, májusra 68 385 főre emelkedett. Az ágazaton belül egyébként a köztisztviselők és közalkalmazottak száma valamelyest még csökkent is. Így az adat interpretációjának pontosításával a radikális intézkedés indoklása gyengébb lábakon áll.

Ez a példa összefügg a statisztikában *ökológiai tévkövetkeztetésnek* nevezett jelenséggel, az *aggregációs torzítással*. A csoportosított, összesített adatok alapján fennálló összefüggéseket nem vetíthetjük ki automatikusan a csoport tagjaira. *Aggregált adatokból ugyanis elvi, logikai okok miatt csak korlátozottan lehet következtetni az egyéni viselkedésekre*, az adatokból nehezen lehet következtetni arra, hogy az aggregálás után fennálló összefüggés az aggregálás előtt is létezett-e. Egy életből vett példával illusztrálva: „aggregált adatokból könnyen kimutatható, hogy minél többet locsoljuk a kertünket, annál valószínűbb, hogy kiszárad.” Nyilvánvalóan olyan térségekben locsolnak sokat az emberek, ahol nagy a szárazság, és ott a kiszáradás veszélye is nagyobb.<sup>37</sup> Aggregált adatokból ennek ellenkezője is bebizonyítható. Egyéni szinten ismerve az adatokat nyilvánvaló, hogy azoknak, akik locsolnak, kevésbé kell félniük kertjük kiszáradásától, mint azoknak, akik nem.

Robinson klasszikus példája a feketék és az írástudatlanság között tagállami szinten kimutatható szoros kapcsolat, amely egyedi adatokon már sokkal kisebb arányban érvényesül. Nagy valószínűséggel igaz, hogy azokban a térségekben, megyékben, régiókban, ahol a roma népesség aránya magas, ott a

romák iránti ellenszenv is magas. Ebből azonban semmiképpen nem következik, hogy a romák ellenszenvvel viseltetnek a romák iránt.

A „statisztikai nyelvhasználat” típushibáinak teljes körét nehéz lenne számba venni. A felsorolt példák azonban arra mindenképpen elégségesek, hogy felhívják a figyelmet a körültekintő adathasználat fontosságára, a számokban rejlő üzenetek megértésének finomságaira. Főként a „naivitásból”, illetve a statisztikai műveltség, képzettség hiányából adódó hibákkal foglalkoztunk. A tudatos torzításoknak, és ide sorolom a *nekünk tetsző statisztikákból történő szemérmesítést* is, kevesebb figyelmet szenteltünk, ezek ugyanis sokkal inkább az etika, mint statisztikai műveltség területére tartoznak.

### Következtetések és ajánlások

*„Könnyű a statisztikával hazudni, de még könnyebb a nélkül.”*  
(F. Mosteller)

A számokat vég nélkül használó korunkban, a „számháborúk” világában nem tehetjük meg, hogy ne szerezzünk jártasságot az adathasználatban, a számok értelmezésében. Jó ha tudjuk mikor, és mire használhatunk egy adatot. Hogy mire jó és mire nem jó a rendelkezésünkre álló szám. Ez az előfeltétele annak, hogy elkerüljük az eredetileg helyes adat téves interpretálását és ezzel lavinát indítva, mutáns statisztikák sorozatához járuljunk hozzá. Az adatok előállításának és használatának etikai vetületei is vannak, melyet mindig szem előtt kell tartanunk. Marc Isaacson a „statisztikai műveltség”

<sup>37</sup> Mérő Csabától vett példa.

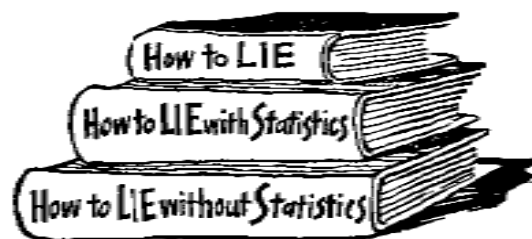
szempontjából hat területet emel ki:<sup>38</sup> 1. A statisztika mögött meghúzódó történet ismeretét, 2. CARE, azaz a gondosságot (de egyben egy mozaikszó, ahol C=Confounding, azaz zűrzavar, összekeverés, A=Assembly, „összeszerelés”, R=Randomness, véletlenszerűség és E=Error, hiba)<sup>39</sup>, 3. A mérési művelet megértését, 4. Az arányok és százalékok leírását, 5. Az arányok és százalékok összehasonlítását és 6. Az arányok és százalékok interpretálásának tudását.

Ez az írás a „leíró statisztika” területére korlátozódott. Főként a mindennapi adathasználat buktatóira hívta fel a figyelmet. Abból a meggyőződésből született, hogy a számokban kifejezett információk rendkívül hasznosak, sőt nélkülözhetetlenek. A számokban megragadott tények hozzásegítenek a világ megértéséhez. Részben azért, hogy az egyébként nehezen belátható, vagy beláthatatlan összefüggések a számok közvetítésével világossá válnak, de oly módon is, hogy a számösszefüggések újabb és újabb kérdések megfogalmazásához vezetnek.

Huff már idézett könyvében egy szellemes illusztrációt találunk: három könyvet egymás tetején<sup>40</sup>. Legfölül egy kisebbet „Hogyan hazudjunk?” címmel. Alatta egy nagyobbat „Hogyan hazudjunk a statisztikával?” címmel és legalul egy még nagyobbat „Hogyan hazudjunk statisztika nélkül?”. A statisztikát művelő szerző a „közbülső” könyvből szemezgetett.

## 2. ábra

### A hazugság irodalma



Forrás: D. Huff, id. műve, 131. oldal

### Irodalomjegyzék

A Brief History of the Post-Autistic Economics Movement,  
[http://www.paecon.net/#\\_A\\_Brief\\_History](http://www.paecon.net/#_A_Brief_History)

Babbie, Earl: A társadalomtudományi kutatás gyakorlat, Balassi Kiadó, Budapest, 1999

Bellingren, Gene, Durval, Castro and Mills, Anthony: Data, Information, Knowledge, and Wisdom,

Best, Joel: Damned Lies and Statistics, University of California Press, Berkeley and Los Angeles, California, 2001

Best, Joel: People Count: The Social Construction of Statistics, 11/2002 Talk at Augsburg College. More on future of Statistical Literacy,  
<http://www.statlit.org/pdf/2002BestASA.pdf>

Boyle, David: The Tyranny of Numbers, HarperCollinsPublishers, London, 2001

Bródy A.: Mennyi az ennyi? Bevezetés a gazdasági mérés elméletébe. Közgazdasági Szemle, XXXVII: évfolyam, 1990. 5. sz., 521-537. oldal

Coffey, Richard D.: Assessing Sow Body Condition,  
[www.Pigsite.com](http://www.Pigsite.com)

Dusek Tamás: Az ökológiai tévkövetkeztetés, ELTE, Regionális Földrajzi Tanszék, 2005,  
[http://geogr.elte.hu/ref/REF\\_Kiadvanyok/REF\\_RT10/RTT-10-6resz.pdf](http://geogr.elte.hu/ref/REF_Kiadvanyok/REF_RT10/RTT-10-6resz.pdf)

<sup>38</sup> Isaacson, Marc: Statistical Literacy — An Online Course at Capella University.

<sup>39</sup> A terminust Schieldtől kölcsönzi. Lásd Schild, Milo: Statistical Literacy.

<sup>40</sup> Huff idézett műve, 121. oldal.



Havasi Éva: Az indikátorok, indikátorrendszerek jellemzői és statisztikai követelményei, Statisztikai Szemle 2007, 85. évfolyam, 8. szám, 677-689. oldal.

Havasi Éva: Beszélgetés Zafir Mihállyal, Statisztikai Szemle, 2007, 85. évfolyam, 3. szám.

Havasi Éva: Interjú Ferge Zsuzsával, Statisztikai Szemle, 2003, 81. évfolyam, 8. szám.

Havasi Éva-Marton Ádám: Vita a statisztika minőségéről, Statisztikai Szemle, 2002.

Huff, Darell: How to Lie with Statistics, W. W. Norton & Company Inc, New York, , 1976

Hunyadi László - Rappai Gábor: Gondolatok a statisztikáról, Statisztikai Szemle,

Hüttl Antónia: A gazdasági mérés történetéről. Adatok, elméletek, gazdaságpolitika, Közgazdasági Szemle, 2003. február.

Interjú Ferge Zsuzsa szociológussal, Statisztikai Szemle, 2003. augusztusa.

Jánossy Ferenc: A gazdasági fejlettség mérhetősége és új mérési módszere. KJK, 1963, 27–32. old.

Isaacson, Marc: Statistical Literacy — An Online Course at Capella University, Augsburg College, Minneapolis, MN, 2005.

Pukli Péter: A statisztika minősége, valóság-hűsége, In.: A statisztika és a közigazgatás elkötelezettje, ELTE Állam-és Jogtudományi Kar, KSH, Budapest, 2008, 399-414. oldal

Gyurcsány Ferenc: Reformok megszorítás nélkül, Metro, 2008. május 9

Mérő Csaba: A részben megfigyelt keresztátlak problémáinak bemutatása, [mero.csaba.googlepages.com/mero\\_csaba\\_tdk.doc](http://mero.csaba.googlepages.com/mero_csaba_tdk.doc)

Schild, Milo Statistical Literacy: Thinking Critically about Statistics, Augsburg College, Minneapolis, <http://web.augsburg.edu/~schild/MiloPapers/984StatisticalLiteracy6.pdf>

Steele, Michael, J.: Darell Huff and Fifty Years of How to Lie with Statistics, Statistical Science, 2005, Volume 20, No. 3, 205-209

Szép Katalin – Vigh Judit: Minőség a hivatalos statisztikában, Statisztikai Szemle, 2004, 82. évfolyam 8. szám

Szilágyi György: Indikátorok, statisztikák, avagy mire van szükségük a felhasználóknak?, In.: A statisztika és a közigazgatás elkötelezettje, ELTE Állam-és Jogtudományi Kar, KSH, Budapest, 2008, 463-472. oldal

Szilágyi György: Érteni a számok nyelvén, Statisztikai Szemle, 199

Szilágyi György: A jelenkori statisztika alapkérdései, Statisztikai Szemle, 1997

Marx, Carl: Tőke, I. Szikra Kiadás, Budapest, 1955., <http://mek.oszk.hu/04700/04724/04724.doc>

Tammet, Daniel: The Boy With The Incredible Brain, I-V. <http://www.youtube.com/watch?v=AbASOcqc1Ss>

Vadászgörények küllembírálati kritériumai, a szépségverseny alapvető szabályai és menete, MaVOE Elnökség kiadványa, [www.vadaszgoreny.net/letolthetok/szepsegverseny.doc](http://www.vadaszgoreny.net/letolthetok/szepsegverseny.doc)