

TUDOMÁNYOS ÉLET

HALABUK LÁSZLÓ

Statisztikai módszerek a megismerés — és a tévedések — szolgálatában*

Varga István egyetemi statisztikai előadásai során ezt a figyelmeztetést intézte hallgatóihoz:

„Ahhoz, hogy Önök statisztikussá váljanak, rá kell jönni arra, hogy a különböző statisztikai eljárásokkal összefüggésben egyrészt mikor mit szabad alkalmazni, másrészt mikor mit nem szabad alkalmazni. Utóbbi majdnem fontosabb, mint az előbbi, mert ha az előbbit nem tudják, legfeljebb nem fognak egyszer helyesen következtetni. A nagyobbik hiba az, amikor valaki azt hiszi, hogy ért a statisztikához és eredményeivel becsapja önmagát és másokat is.”¹

A téves következtetésekre vonatkozóan pedig hadd fűzzünk ide egy konkrét példát:

„Egy valóban bizarr ötlet alapján elektronikus számológép segítségével pontossági vizsgálatnak vetettek alá két karórát. Az egyik törött volt, s nem járt, a másik naponként egy másodpercet késett. A megfelelő adatok betáplálása után az elektronikus számológép azt válaszolta, hogy az az óra rossz, amelyik naponta egy másodpercet késik, mert csak 120 évenként egyszer mutatja a pontos időt, ellentétben a törött órával, amelyik minden 24 órában kétszer mutatja a pontos időt.”²

Ez a tanulmány a címben és az előző idézetekkel jelzett problémával kapcsolatban néhány gondolatot kíván felvetni. Kezdjük egy szembeállítással.

1. A statisztika — hosszú története során — a megismerés robusztus, hatékony eszközének bizonyult. (Ha statisztikáról

beszélünk, ezen a tevékenységet és a módszerek összességét egyaránt értjük.) Kezdvé az első összeírásoktól, az államtudományi szerepen keresztül a modern természettudományi megismerésig a statisztika az emberi tudásnak egyik döntő forrása volt. Ezt az emberi tudás története, az egyes tudományok ismeretanyaga egyértelműen igazolja.

2. „Háromféle hazugság van, az egyszerű hazugság, a rosszindulatú hazugság és a statisztika.” — „A hazugság válfajai az egyszerű hazugság, a szükséghelyzetben való hazugság és a statisztika.” Válogatok egy témára, amely egyesek szerint *Benjamin Disraelitől*, mások szerint *Lord Palmerstontól* származik. A változatoknál és az eredetnél figyelemre méltóbb, hogy a nevezetes negatív értékelést a múlt század óta nem szüntek meg idézni.

Mi rejlik e mögött az ellentmondás mögött? Ha hozzá nem értő, fölényes szellemességre hivatkozunk, ez nem elégséges a teljes indokoláshoz. Az ellentmondás magyarázatához mindenképpen hozzátartozik az is, hogy a statisztika mint módszer és mint hivatkozási információs alap védtelen a rossz felhasználásokkal szemben. Ahogy a nő a módszerek által kínált megismerési kapacitás, oly módon nő a rossz felhasználás lehetősége (és többé-kevésbé ténye) is.³ Ezt a megállapítást próbáljuk ebben a tanulmányban illusztrálni.

A statisztikai kézikönyvek bőven szolgáltatók a vidám, anekdotaszzerű példákat a statisztikai visszaélésre. Ezek a példák

* Megjelent a „Változások, váltások és válságok a gazdaságban. Tanulmányok Varga István emlékezetére” című, Schmidt Ádám és Kemenes Egon szerkesztette kötetben. Közgazdasági és Jogi Kiadó, Budapest 1982. 281—290. old.

¹ Varga István: A statisztika elmélete és módszertana. A Pázmány Péter Tudományegyetem Jog- és Közigazgatástudományi karán az 1948/49. tanév első felében tartott előadásairól készült gyorsírói feljegyzés. Kézirat gyanánt. MEFESZ Jogászkör kiadása. 1949.

² Az idézet Kovács István: Illúziók és borulátás nélkül című cikkéből való. Népszabadság, 1968. január 12.

³ O. W. Haseloff — H. J. Hoffmann: Kleines Lehrbuch der Statistik. Walter de Gruyter Co., Berlin, 1968.

főleg a korrelációs számítás területéről meríteneek.⁴ A korrelációs számítás eredményeinek hibás felhasználásait még rendszerezni is szokták. Egyik csoportba sorolják a helytálló, de semmitmondó statisztikai összefüggést, amikor például egymás komplementumainak minősíthető változók korrelációs koefficiense -1 . Ilyenek az egymást 100-ra kiegészítő százalékszámok. Például ha a dohányzók százalékos aránya változik, a nem dohányzók százalékos aránya -1 értékű korrelációs koefficienssel kapcsolódik az előzőhöz. Egy másik csoportba sorolják a „közös harmadik tényező” fennállásának esetét. („Ha Magyarországon fagy, Svájcban több tüzelőanyagot vásárolnak.”) Végül mit szóljunk a véletlen kapcsolatok vagy értelmetlen következtetések humorba torkolló, vidám példáihoz:

„Egy zenerajongó floridai fiatalember lakása három helyiségében szobai növényeinek különféle zenedarabokat szokott játszani, és úgy találta, hogy abban a szobában fejlődnek legjobban a növények, ahol beatzenét játszik.”

„Aránytalanul több ember hal meg az ágyban, mint az utcán, a tengerben vagy a lebujsokban. Az ágy tehát bizonyíthatóan a legveszélyesebb tartózkodási hely.”

„Az Egyesült Államokban, ahol az egyetemek is hirdetik, reklámozzák magukat, az egyik egyetem buzdításul közölte, hogy az utolsó évben végzett nőhallgatók 50%-át az egyetem professzora vette nőül.” (Két nőhallgatója volt.)

Az előző példák statisztikai tankönyvekből, kézikönyvekből származnak. Az a helyzet azonban, hogy az ilyen riasztó példák inkább szórakoztatásra alkalmasak; a kézikönyvek, tankönyvek sajnos nem tekintik feladatuknak, hogy a statisztikai módszerek rendszeres bemutatása során a módszerek képességeinek korlátaival, a téves felhasználás, a hibás interpretáció körvonalazásával is foglalkozzanak.

A statisztikai módszertan eszköztára egymás mellett élő, egymással párhuzamosan és egymással összefüggésben (részben azonban egymástól függetlenül) fejlődő témakörökből áll. Ezeket tartalmazza a statisztikai kézikönyvek,⁵ tankönyvek, így a középértékek és szóródás vizsgálatát, a korreláció- és trendszámítás módszereit, az indexszámítás és időselemlés, a minta-

vétel, a becslésmélet, a hipotézisellenőrzés kérdéseit stb. A modern vizsgáldásokban azonban egyre nagyobb szerephez jutnak az olyan komplex vizsgálatok, elemzések, amelyek a statisztikai módszertan mind több részét kombinálják és egyesítik a valóság valamely részének sokoldalú felderítése, megismerése céljából. Ezek a komplex jellegű vizsgálatok vezetnek el a matematikai megfogalmazású, különféle (de főleg statisztikai) információs bázisból kiinduló, a (matematikai) statisztikai inferencia eszközeivel operáló diszciplínákhoz, mint az ökonometria, biometria, szociometria stb. Ezeknek a komplex módszertanoknak, diszciplínáknak gyakorlati megjelenési formái a többváltozós, többrelációs modellek.⁶

Kézenfekvő, hogy ezek a komplex módszerek és modellek a megismerésnek nagy lehetőséget nyitó, robusztus eszközei. Természetes az is, hogy kumuláltan produkálják a tévedés, félreismerés lehetőségeit is. A következőkben főleg e területek problémái foglalkoztatnak.⁷

A tévedések néhány típusa

E tanulmány keretei nem elégségesek ahhoz, hogy a teljesség igényével tekintsük át a hibákat, amelyek a statisztikai eszközök segítségével történő valóságmegismerést tévútra vezetik. Ha mégis — a teljes rendszerre való törekvés nélkül — pragmatikusan tipizálni kívánjuk a legfontosabb hibákat, a következő csoportosulásokat fogalmazhatjuk meg:

- a fétések felelőssége,
- a rossz specifikáció hatása,
- a kovariancia—kauzalitás fogalmainak összekeverése.

a) A „fétések felelősségét” abban látjuk, hogy egy-egy kiragadott szempont, cél, módszer túlzott jelentőséget kap, az elengedhetetlen mellérendelt tényezők elhanyagolásával.

— Az „illeszkedés” fétise. Amikor egy-egy (egy- vagy többtényezős) összefüggést vagy összefüggésrendszert megfogalmazzunk, gyakran a legjobb illeszkedés a cél. Lehet törekedni a priori ismeretek, a plauzibilitás alapján reálisnak tűnő specifikációra, de törekedni lehet az összefüggések szorosságának jelzésére rutinszerűen

⁴ H. Swoboda: Knaurs Buch der modernen Statistik, 1971.

⁵ K. A. Yeomans: Statistics for the social scientist. Bristol, 1968.

⁶ Halabuk László: Ökonometriai modellek és módszerek kutatása és alkalmazása Magyarországon. Statisztikai Szemle, 1975/8—9.

⁷ Halabuk László: Some experiences with econometric model building. European Meeting of the Econometric Society, Budapest, 1972.

használt mutatók legmagasabb értékének biztosítására is. Ennek az utóbbi célnak érdekében, a becslési eredmények feljavítására gyakran folyamodnak megengedhetetlen eszközökhöz is, mint amilyen a változók számának, a kisleltetések számának gátlástalan növelése, a mesterséges változók (dummy variables) könnyed alkalmazása. Ezek az eljárások esetenként a sztochasztikus kapcsolatok determinisztikussá való degradálásához közelítenek. Ehhez a kérdéskomplexumhoz tartozik annak a megállapítása is, hogy a gyakorlati szempontból nagy jelentőségű előrejelzésnél a mintaidőszakhoz való tökéletes közelítésnél nagyobb jelentősége van a mintaidőszak és az előrejelzési időszak paraméterrendszere közötti eltérés helyes felmérésének.

— A pontbecslés uralkodó szerepe. Elsősorban az előrejelzésnél észlelhető, hogy a pontbecslés mellett az intervallumbecslés szerepe háttérbe szorul. Ez pedig egyenértékű annak a lehetőségnek az elhanyagolásával, amit a sztochasztika azáltal nyújt, hogy minden megállapítás mellé valószínűségi eloszlást rendel.

— A mennyiségi közelítés prioritása a minőségivel szemben. Az előrebecslés szokásos sémája szerint az időskálához egy mennyiségi alakulást feltüntető függvényt rendelnek. Az előrejelzések számos kudarca sem volt elég ahhoz, hogy e kvantitatív közelítés mellett megfelelő elismerésben részesüljön a kvalitatív közelítés. Arról van szó, hogy az aprólékos számszerű determináció helyett ígéretesebb és tartalmasabb lehetne a tendenciák, a fordulópontok, a struktúraváltozások előrejelzése. Az irodalmi eredmények mellett az ilyen irányú gyakorlati alkalmazás alig kapott még szerepet.

— A méretek féltése.⁸ Ennek a bűvöletnek talán a legprimerebb típusa a teljes körű statisztikai felvétel iránti tisztelet. A teljes körű felvétel különösen fontos eseténél, a népszámlálásnál nem szabad azonban figyelmen kívül hagyni, hogy ennek pontossága sem abszolút, sőt az elért pontosság is csupán egy eszmei pillanatra érvényes. Amint ezen a pillanaton túllépünk, a legteljesebb körű felvétel is már csak extrapolációs alapnak tekinthető, és — különösen erősen növekvő népességszám esetén — a vizsgált sokaságnak már csak egy részét — éspedig eszerélődő részét — képviseli.

Az ökonometriai modellek esetében nagy jelentőségű a közgazdasági változó-

kat reprezentáló idősorok hosszúsága. A becslésmélet annál megbízhatóbbnak tekint az információk bázist, minél nagyobb a minta, idősorok esetében minél hosszabbak a megfigyelt idősorok. Nem szokták azonban hangsúlyozni a paraméterbecsléshez felhasznált idősorok homogenitásának fontosságát. Ha viszont a megfigyelési minta nagysága — az idősorok hossza — a heterogenitás növekedésével jár együtt, akkor nagyon körültekintően meghatározott kompromisszum indokolt. (Ez a kérdés elvezet a *Bortkiewicz* által megfogalmazott ún. kis számok törvényéhez.)

Hasonló túlértékelésben részesülhet egy összefüggérendszer specifikálása esetén a változók és a relációk számának nagysága. Igazolható, hogy adott esetekben a változók vagy relációk számának csökkentése nem jár információvesztéssel, legfeljebb munkamegtakarítással.

— Megemlíthetjük a pontosság féltését is. Nem lehetne klasszikusabban fogalmazni, mint *Oskar Anderson*: „1. Nines gyakorlati értelme egy szekér szénát kémiai precíziósmérleggel mérni. 2. Nem ér semmit, ha két város távolságát kb. ezer lépésekben lemérjük, és ehhez hozzáadjuk a városkapu milliméterben mért vastagságát.”

b) Az előzőekben módszertani megközelítési hibákat említettem. A következő néhány hibapélda a szakmai adottságok, ismeretek hiányára, elhanyagolására, valamely releváns körülmény figyelmen kívül hagyására vezethető vissza. Nevezhetjük e példákat a *rossz specifikáció* eseteinek. Ilyenek például:

— Az összehasonlítási bázis problémája. Magyarországon az ásványolaj termelése 1938-ban, a rotációsipari termelése 1929-ben még kísérleti stádiumban volt. Ha tehát 1929-et vagy 1938-at tekintjük bázisnak, a következő évek termelése robbanásszerű fejlődést tükröz.⁹ Ez a helyzet minden új termék vagy gyártási ág megindulásakor. Ha ezzel kapcsolatban utalunk arra, hogy Magyarországon 1929 konjunkturális csúcs volt, és a következő években általános volt a hanyatlás, illetőleg pangás, akkor az a látszat születik, mintha a rotációsipari-termelés konjunktúramentes ágazat lett volna.

— Egy ország népsűrűsége úgy adható meg, mint a népesség és a terület hányadosa. De ha például a terület jelentős része (50—70—90%-a) emberi lakhatásra alkalmatlan, akkor nyilvánvalóvá válik az ún. nyers arányszámokkal szemben a tiszta arányszámok jelentősége, például a nép-

⁸ *Halabuk László—Hulyák Katalin*: Az időjárás és a mezőgazdasági termelési eredmények. Ökonometriai Füzetek, 10. sz. KSH, 1968.

⁹ Ezek egyébként a gazdaságtörténet, illetve a gazdaságstatistika szokásos bázisai.

sűrűség esetében az ország össznépszerűségének (nem a teljes, hanem) a lakhatásra, művelésre alkalmas területtel képezett hányadosa.

— Igen fontos, hogy egy ország nemzeti jövedelméből mekkora részt fordít bizonyos célokra, például katonai, illetőleg jóléti vagy egészségügyi kiadásokra. Ez a kérdés a politikai publicisztikának is kedvelt, gyakori témája. Amikor azonban az említett célokra fordított kiadásokat nem a nemzeti jövedelem (vagy valamely rokon kategória) szintjén nézik, hanem az állami költségvetés síkján, erősen torzított, félrevezető kép adódik, hiszen az egészségügyi, jóléti kiadások országonként eltérően oszlanak meg az állami („föderális”) költségvetés és a helyi kormányzati egységek költségvetése között.

Az előző három példa az elemi statisztikai tankönyvek, kézikönyvek szférájába tartozik. A hibás specifikáció igazi „érvényesülési területe” azonban az a vizsgálati szféra, amelyet az előzőekben a komplex, alkalmazott módszerekhez kapcsoltam, amely végső fokon a „modellek” fogalmkörébe torkollik. Vegyünk e hatalmas komplexumból példákat a rossz specifikációra.¹⁰

— Az ökonometriai modellek logikai sémáját tekintve interdependens és rekurzív modelleket szoktak megkülönböztetni.¹¹ Az interdependens modellek olyan rendszert tükröznek, amelyben a változóknak — legalábbis a rendszer endogén változóinak — egymásra való hatása (elvből) kölcsönös, és a modell számszerűsítése során határozzák meg ezen egyedi összefüggések erősségét, fontosságát. (Ebben a kvantifikálásban a 0 értékű koefficiens is helyet kaphat.) Amellett, hogy az ökonometriai modellek túlnyomó többsége ismeri és elismeri ennek az összefüggérendszernek az interdependenciáját, vannak, akik a rekurzív specifikáció hívei. Ez utóbbi rendszerben az összefüggések, hatások egyirányúak. Ebben a specifikációban az összefüggések paramétereit meghatározásával kapcsolatos módszertani nehézségek áthidalása érdekében a közgazdasági valóság megfogalma-

zása szenved súlyos, alapvető sérelmet.¹² Ilyen hiányosság jelentkezik például akkor, amikor a termelés (kínálat, kapacitás) hatását a fogyasztásra (keresletre) megfogalmazzák ugyan, de a fordított irányú hatást (visszacsatolás, feedback) nem dolgozzák ki.¹³

— Némileg rokon ezzel a problémával a következő. Szokásos a termelés színvonalát kapacitásbeli (kínálati) vagy pedig igény szerinti (keresleti) tényezőkkel meghatározni. Előfordul a termelés szintjének meghatározására keresleti és kínálati tényezők vegyes alkalmazása is. A (leegyszerűsített) valóság azonban az, hogy olyan helyzetben, amikor az igény (kereslet) meghaladja a kapacitást, a termelés színvonalát a kapacitás határozza meg, és a kapacitást meghaladó többletigényeknek az alakulása a folyó termelés szempontjából irreleváns. Amikor az igények nem érik el a termelési kapacitást, a termelés színvonalát az igények határozzák meg, és a termelési kapacitásoknak a termelés meghatározásában játszott szerepe legalábbis csökken. A termelési színvonal meghatározásában tehát a szűk keresztmetszetet képviselő tényezőrendszer szerepe a döntő, vagy legalábbis elsődleges. A világszerte nagy számban készülő ökonometriai modellek specifikációja nem tükrözi ennek a ténynek a felismerését.¹⁴

c) A tévedések egy jelentős része a kovariancia — korreláció — kauzalitás helytelen felfogásából, a mutatószámok téves értelmezéséből és hibás következtetések-ből adódik. Anélkül, hogy a korreláció- és regressziószámítás nélkülözhetetlen és kitűnő módszertanának szakmai tárgyalásába bocsátkoznánk, a következőket emelhetjük ki:

A kovariancia két vagy több jelenség („változó”) alakulása közötti statisztikai összefüggésre, a kauzalitás viszont a változók közötti okozati kapcsolatra utal. Nehéz, szakmai tájékozottságot és józan gondolkodást igénylő feladat annak elbírálnia, hogy statisztikai összefüggés mennyiben utal funkcionális, kauzális kapcsolatra, és ha ezt létezőnek tekintik, a kauza-

¹⁰ Ph. J. Dhrymes—Ph. Howrey—S. H. Hymans—J. Kmenta—E. E. Leamer—R. E. Quandt—J. B. Ramsey—H. T. Shapiro—V. Zarnowitz: Criteria for Evaluation of Econometric Models. *Annals of Economic and Social Measurement*, 1/3. 1972.

¹¹ H. T. Shapiro: Is verification possible? The evaluation of large econometric models. *American Journal of Agricultural Economics*, 15/2. 1973. május.

¹² Halabuk László: Kérdésfeltevés és módszerek viszonya az ökonometriai modellekben. A magyar statisztikai felsőoktatás kétszáz éve c. KSH, 1979. kiadványban.

¹³ H. T. Shapiro—Halabuk L.: Macroeconometric model building in socialist and non-socialist countries: A comparative study. *International Economic Review* 17/3. 1976.

¹⁴ Halabuk László: A népgazdaság ökonometriai modellezésének néhány időszerű kérdése. *Sigma*, 1977. 4.

litás milyen jellegű.¹⁵ Itt elkerülhetetlenül némi ismétlésbe kell bocsátkoznunk, ha a problémakört vázolni akarjuk. Olyankor, amikor két változó között statisztikai összefüggés észlelhető, a korreláció típusai — a kauzalitás szempontjából — a következők lehetnek:

— egyirányú kapcsolat, például a csapadék és a mezőgazdasági termésátlagok között:

— interdependens, kölcsönös kapcsolat, például a bérek és termelékenység között. Konkrét ismeretek szükségesegek annak elbírálásához, hogy az elvben kölcsönös kapcsolaton belül adott esetben melyik irány a lényegesebb;

— közös meghatározó tényező, például a kukorica és a cukorrépa átlagtermésének alakulása nagyfokú kapcsolatot mutat; a két változó azonban nem egymás befolyása alatt áll, hanem mindkettőre közös (meteorológiai és egyéb) tényezők hatnak. Ide tartozik az a fontos eset is, amikor egy gazdaság erős általános növekedés fázisában van. Ilyenkor úgyiszlólván bármely gazdasági változók között megállapítható az erős statisztikai összefüggés („a változók kollinearitás”), ezt azonban „közös harmadik” tényező idézi elő. Az említett esetben a változók kapcsolatának vizsgálatánál a trendjellegű fejlődési tényezőt ki is szokták szűrni;

— már utaltunk arra a tautológia jellegű kapcsolatra, amely valamely egész komplementáris részei között létezik;

— vég nélkül lehetne sorolni a véletlenszerű statisztikai kapcsolatokat.

Nyilvánvaló, hogy a különféle típusú kauzalitást (vagy függetlenséget) takaró statisztikai összefüggések gyakran ténvútra esábítják az óvatlan elemzőt. Utalunk ezenkívül arra is, hogy a döntést, következtetést további tényezők bonyolíthatják. Mindenekelőtt az, ha nem két-, hanem

többváltozós kapcsolatról van szó. Megfontolásra int az is, hogy a változók kapcsolata döntően megváltozhat, ha közöttük időbeli késleltetést alkalmazunk. Végül szinte csüggesztő lehet, hogy egy („szerencsésen specifikált”) összefüggésben a magyarázó változók kicserélése hasonlóan szoros összefüggéshez vezethet.

Két szempont azonban nagyon fontos. Először az elemi összefüggések mellett¹⁶ óriási a szerepük a sokaságok közötti összefüggéseknek. Egy közeledő esőfelhő esetében minden egyes vízmolekula helyzetének és pályájának tanulmányozása helyett megelégszünk annak megállapításával, hogy az okok halmaza (esőfelhő stb.) erős valószínűséggel esőt jósol.

Másodszor egy közgazdasági összefüggésből (modellből) nem zárható ki a direkt, közvetlen kauzalitások léte már csak azért sem, mert ez a gazdaságirányítás, szabályozás funkcióját zárja ki a modellből.

Ismeretelméleti megfontolások

Az előzőekben vázoltuk és némileg tipizáltuk a statisztikai módszerek felhasználásával végrehajtott téves eljárásokat, következtetéseket, törekvéseket. Most megkíséreljük annak észlatos megmutatását, hogy eme tévedések mögé sokszor egy ismeretelméleti háttér rendelhető.

A tudományos ismereteket sokféleképpen osztályozzák. Egy bizonyos szempont alapján a tudományos ismereteket nomografikus és idiografikus ismeretekre oszthatjuk.¹⁷ A nomografikus kutatás során a közvetlenül tapasztalt kölcsönösen összefüggő jelenségek sorozatos felbontása révén a környező világban uralkodó „elemi okok” és „elemi következmények” összefüggésének megállapítására törekszünk. Az idiografikus típusú megismerés folyamán a

¹⁵ S. H. Hymans: On the use of leading indicators to predict cyclical turning points. Brookings Papers on Economic Activity, 2. 1972.

¹⁶ Az emberi tudásnak az elemi összefüggések teljes birtoklására alapozott ideálját fogalmazta meg Laplace: „Une intelligence qui pour un instant donné connaitrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si le d'ailleurs elle était assez vaste pour soumettre ces données à l'anayse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome: rien ne serait incertain pour elle, et l'avenir comme le passé serait présent à ses yeux.” (P. S. Laplace: Essay philosophique sur les Probabilités. 1814.)

Ez a gondolat — hogy ti. egy olyan elme, amely adott pillanatban a természetet mozgató összes erőket és a természetet alkotó lények helyzetét is ismerné, s ha emellett ez az elme elég hatalmas lenne, hogy mindezeket az adatokat elemzés alá vegye és ugyanabba a formulába foglalná a világmindenség legnagyobb testeinek és legkisebb atomjainak mozgását, akkor számára semmi sem lenne bizonytalan, mert mind a jövő, mind a múlt jelen lenne előtte — valamennyi tudomány számára érdekes lehet.

¹⁷ Oskar N. Anderson: Einführung in die mathematische Statistik. Verlag Springer, Wien, 1935.

kutató elme elemeknek valóságos térbeli és időbeli megoszlásának, megjelenésének ismeretére, leírására törekszik.¹⁸

Mivel a statisztika a megfigyelt jelenségek leírásában nem tud téről és időtől elvonatkoztatni, alapvető jellege idiografikus. Másrészt módszertani oldalról, a valószínűségszámítás, a matematikai statisztikai inferencia alkalmazásán keresztül fokozatosan nőtt a szerepe egy nomografikus elemnek is. A konfliktusnak — amely az ökonometriai modellezésben a legélesebb — gyökere ez: lehet-e a térben és időben meghatározott, egyszeri, megismételhetetlen jelenségek elemzésére téről és időtől elvonatkoztatott, igen nagy vagy végtelen sokaságok vizsgálatára alkalmas matematikai statisztikai módszereket alkalmazni? Tekinthető-e egy időben és térben izolált eseményhalmaz egy magasabb rendű sokaság, halmaz mintájának? Első megközelítésben úgy tűnik, hogy összeegyeztethetetlen dolgokról van szó. A probléma a sztochasztikus ökonometriai modellek esetében azért különösen éles, mert kezelésükhöz, elsősorban a paraméterbecsléshez elengedhetetlen a megfigyelt adathalmaznak mintaként való kezelése.

Mégis létrehozható a *történeti jellegű minta* („idiográfia”) és a *matematikai statisztikai inferencia* („nomográfia”) közötti *kompromisszum*. Ennek döntően fontos elméleti alapja és ugyanilyen fontos előfeltétele van. Az elsőt, az áthidaló eszközt a kis minták elmélete nyújtja, amely lehetővé teszi a statisztikai inferencia alkalmazását a tömegjelenségek fogalmköréhez képest elenyészően kis mintákra, 10—20—30 éves idősorokra is. Ennek az áthidalásnak a megteremtésében elévülhetetlen érdeme van *R. A. Fishernek*, a 20. század kiemelkedő statisztikusának.¹⁹ A statisztikai inferenciának a kis mintákra való alkalmazása feltételként nyomatékosan igényli azonban az inferencia korlátozottságának minden rendelkezésre álló külső, szakmai információval való ellensúlyozását. A történeti jellegű kis minták és a

sztochasztika szintézise századunk egyik nagy vívmánya, ez az alkalmazókra kötelezettségeket is ró.

Ezzel eljutunk a módszer és a vizsgálat tárgyának kettősségéhez. A módszer neutrális, közömbös, a tárgynak viszont saját arculata van.

— A módszer passzív. Hagyja magát felhasználni okosan és oktalanul. Rövid eszmefuttatásunkhoz elsősorban közgazdasági összefüggésrendszerek adják a hátteret. Ezeknek az összefüggésrendszereknek a paraméterbecslése engedelmesen túri a matematikai optimalizálást akkor is, ha a külső, a priori, nem is vitatott összefüggések szerepét figyelmen kívül hagyják, és ezáltal szabadon engedik az esetleges, véletlenszerű, abszurd összefüggések megállapításának lehetőségeit.

— A vizsgálat tárgyának saját karakterisztikumai, strukturális adottságai, okozati összefüggései, funkcionális törvényei vannak. Ezekről a vállalt, konkrét vizsgálat megkezdése előtt is vannak a priori ismereteink. Az a priori ismeretanyagoknak a mellőzése gyengíti a megállapítások erejét, megbízhatóságát. A tárgy sajátosságainak mellőzhetetlen voltára utal az ökonometria, szociometria, biometria diszciplínáinak önállósulása is, ugyanakkor, amikor a statisztikai inferencia és a matematikai kezelés mindnyájuk közös vonása.

*

Korábban utaltunk arra, hogy az újabb, potenciálisan erőteljes módszerek a téves alkalmazások lehetőségét is megteremtik. Ennek a tanulmánynak azonban nem az a célja, hogy — egy korlátozott területen — a megismerés és a tévedés küzdelmének reménytelenségét konstatálja. Az ismeretek szédítő kibővülésének keretében a józan ítélet fontosságát kívántam hangsúlyozni. Mert a legragyogóbb módszerek birtoklása sem menti fel az embert — a statisztikust és a statisztika alkalmazóját sem — a gondolkodás szükségességétől.

¹⁸ Ennek az osztályozásnak kialakításában kiemelkedő szerepet játszott *Csuprov*, *Couturat*, *Windelband*, *Kries*, *Rickert*, *Cournot*. Az itt használt terminusok mellett a „nomotetikus”, „nomologikus”, „ontológiai”, „törvénytudományok”, illetőleg „eseménytudományok”, „történeti tudományok” kifejezés is szerepel. Kiemelkedő hozzájárulás e problémákörhöz például: *L. Bortkiewicz*: Die Iterationen, ein Beitrag zur Wahrscheinlichkeitstheorie. Berlin, 1917.; *L. Bortkiewicz*: Homogenität und Stabilität in der Statistik. Skandinavisk Aktuarietidskrift, H. 1/2. Uppsala, 1918.; *A. A. Csuprov*: Statistik als Wissenschaft. Archiv für Sozialwissenschaft und Sozialpolitik, Bd. 23., 1906.; *A. A. Csuprov*: Das Gesetz der grossen Zahlen und der stochastisch-statistische Standpunkt in der modernen Wissenschaft. Nordisk Statistik Tidskrift, I/1., 1922.

¹⁹ *R. A. Fisher*: Statistical Methods for Research Workers. London, 1932.