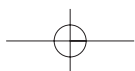
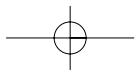
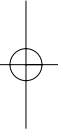
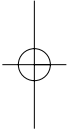




VIII. INTERPRETÁCIÓ





Deirdre N. McCloskey

A szignifikanciasztek retorikája*

Földházi Erzsébet fordítása

A nehézségek forrása a statisztika retorikájának történeti alakulásában keresendő

A szignifikanciasztek retorikájának történeti alakulására visszatekintve látjuk, hogy a közgazdászok (az orvostudományokkal, a szociometriával, a pszichometriával, a politikai tudományokkal, valamint az oktatással foglalkozók) utat tévesztettek. A statisztikai szignifikancia nem megfelelő használata részben a nevében rejlő retorikából ered. Ez azt sugallja, hogy nekünk, komoly tudósoknak mindenekelőtt a „szignifikáns” együtthatókkal kell foglalkoznunk: mi, akik nagyszerűek és kiválóak vagyunk, nem akarhatjuk trivialisokra pazarolni az időnket. A brit statisztikai tradíciónak gyermekkorában Karl Pearson, ifjúkorában pedig R. A. Fisher volt a vezéregyénisége. Mindkettőjüknek, de főleg Fishernek megvolt az a képessége, hogy nevet adjon elképzeléseinek. Ahogyan Kruskal érvelt:

„Tegyük fel, hogy Sir R. A. Fisher – a propaganda mestere – nem vesz át a mindennapi angoltól olyan beszédes szavakat, mint az »elégséges«, »hatásos« és »konzisztens«, és nem teszi ezeket a statisztikai elmélet pontosan definiált szakkifejezéseivé. Ezek után a becslőfüggvények tulajdonságaira vonatkozóan használhatott volna teljesen unalmas kifejezéseket, mondjuk A, B és C jellemzőknek nevezve ezeket. Vajon munkája akkor is ugyanilyen óriási hatást gyakorolt volna? Azt gondolom, hogy nem, vagy legalábbis nem ilyen gyorsan. Vagy nézzük Fisher csodálatos kifejezését: »varianciaanalízis«. Túlságosan cinikusak lennénk, ha azt gondolnánk, hogy ez a remek kifejezés – amely félig misztikus, félig ígéretes – és a szabályos táblázatok hozzájárultak az elfogadásához, az alapját képező elmélettől meglehetősen függetlenül?”

(Kruskal 1978: 98.)

* Eredetileg megjelent: *The Rhetoric of Economics*. (2., javított kiadás) Wisconsin: University Press. /The Rhetoric of Human Sciences./

688 TELEPÜLÉSKUTATÁS – VIII. INTERPRETÁCIÓ

A „szignifikancia” egy még régebbi szóalkotás. A statisztikai szignifikancia régi és jó elgondolás, bár ki van téve annak, hogy nem megfelelően használják. Cicero *De Divinatione* (A jóslásról) című művében Quintus így érvel az égi jelekben való hit mellett: „*Pusztá véletlenek, mondom. Na mármost, valóban így van? Lehet bármi »véletlen«, ami az igazság minden jelét magán viseli? Négy kockát elvetsz, és egy Vénuszt (négy különböző számot) kapsz eredményül – erre van esély; de mit gondolsz, arra lenne esély, hogy ezt százszor megismételve százszor Vénuszt dobj?*” (Div. I, 23.) A 18. század elején John Arbuthnot megfigyelte, hogy Londonban a megelőző nyolcvankét év mindegyikében több lány született, mint fiú, és azt a következtetést vonta le egy ilyen esemény 50% alá csökkenésének alacsony valószínűsége alapján, hogy a Teremtő előnyben részesíti... a férfiakat (Denton 1988: 164.). Az elképzelés első fontos tudományos alkalmazását Laplace 1773-as emlékirataiban találjuk, ahol tizenkét üstökös pályáinak eloszlásáról ír: képes volt elvetni azt a hipotézist, hogy ezek a pályák ugyanolyan síkon helyezkednek el, mint a bolygók, és ezért képes volt bebizonyítani, hogy ezek az üstökösök a naprendszeren kívülről származnak (Scott 1953: 202.).

Lancelot Hogben gondolt először annak a szónak a statisztikai használatára, amiről John Venn beszélt 1888-ban, amikor a különbségeket a valószínűségi hiba egységeiben fejezte ki: „*Ezek arról tájékoztatnak bennünket, hogy a fenti táblázatokban mely különbségek állandóak és szignifikánsak, abban az értelemben, hogy meglehetősen biztosak lehetünk abban, hogy ha egy másik hasonló csoportot veszünk, hasonló különbséget fogunk találni; és melyek pusztán átmenetiek és nem szignifikánsak, abban az értelemben, hogy egy másik hasonló csoport esetében esetleg azt tapasztaljuk, hogy a korábban kapott következtetés megfordul*” (idézi Hogben 1968: 325.). Stephen Stigler kimutatta, hogy nem Venn 1888-ban, hanem F. Z. Edgeworth 1885-ben használta először a megfelelő szót: „*Az ingadozást megállapítva, megbízhatóan állíthatjuk, hogy két statisztikai szám közötti különbség fennáll, még akkor is, ha első ránézésre nem szignifikáns, vagy nem felel meg valódi különbségnek*” (a King’s College előadásaihoz 1885-ben készült összefoglalóból idézi Stigler 1986: 364.). Mindenesre az 1880-as években angolul. Az érvelés ésszerű és elfogadható: amikor megfelelően alkalmazzuk egy mintára, a szignifikancia együtt jár az állandósággal abban a statisztikai értelemben, ahogyan Venn leírta, vagy bizonyos értelemben „*a dolgok közötti valódi különbséggel*”, ahogyan Edgeworth állítja (bár az utóbbi kevésbé megnyugtató, és valószínűleg könnyebben félre lehet érteni, mint a tudományos szignifikanciát). Sőt, egy különbség – például 0,999 és 1,000 közötti különbség az amerikai áraknak a külföldi árakra vonatkozó regressziójában – lehet állandó (ami nem valószínű, hogy a mintából véletlenszerűen adódik) anélkül, hogy bármilyen más értelemben „szignifikáns” lenne.

Ez a baj. Az 1910-es és 1920-as évekre ez a szóhasználat általánossá vált a kifinomult módszereket alkalmazó kutatók (*sophisticated researchers*) körében (Pearson 1911; Yule és Greenwood 1915; Fisher 1925: 43.). Ahogy terjedt a szignifikancia jelentőségét hangsúlyozó szóhasználat a „kevésbé kifinomult” körökben, a retorikai veszély megszüntetésére irányuló vesztes csata is elkezdődött. A sok munka közül, amelyik ugyanarra a pontra mutatott rá, amiről én is írtam itt, az első már 1919-ben elkészült (Boring 1919). Az érvelés a szignifikancia mechanikus alkalmazása ellen irányult, amely már korán mindennapossá vált a statisztika oktatásában. 1939-re például az intellektuálisan nem túlságosan nagyigényű *Szimbólumok és kifejezések statisztikai szótára* világosan rámutatott: „A szignifikáns különbség nem szükségszerűen nagy, mivel a nagy mintákban még egy kis különbség is lehet a szignifikáns különbség bizonyítéka. Továbbá, a szignifikáns különbség megléte nem feltétlenül jár együtt a gyakorlati értelemben vett szignifikanciával” (Kurtz és Edgerton 1939, s. v. „Szignifikáns különbség”). Kendall és Stuart *A statisztika haladó elmélete* című munkájukban explicit módon elismerik a retorika területén tapasztalt problémát, és olyan színtelen kifejezéseket ajánlanak a „szignifikanciaszint” helyett, mint a „próba mérete” (Kendall és Stuart 1951: 163. vö. Morrison és Henkel 1969, ők a „szignifikanciaszint” helyett a még kevésbé érdekes „mintahiba döntési eljárás”-t ajánlották [1969: 198.]). Elsőként R. A. Fisher fogalmazta meg az 1920-as években azt a véleményt, hogy a szignifikancia fontos tudományos következtetések eszköze. Szembenállása a retorikusan érzékeny eljárásokkal szilárd és elégséges volt. A modern statisztika saját története volt az, amit Fisher legyőzött, szemben a józan ésszel. Az 1930-as években Jerzy Neyman és E. S. Pearson (Karl Pearson fia) és még explicitebb módon Abraham Wald úgy érveltek, hogy a meghozott statisztikai döntéseknek a valódi fontosságtól is függniük kell, nemcsak a statisztikai szignifikanciától. 1933-ban Neyman és Pearson az alábbiakat írta (az első- és másodfajú hibákról):

„Melyik a súlyosabb probléma: egy ártatlan ember elítélése vagy egy bűnös felmentése? Ez függ a hiba következményeitől; attól, hogy a büntetés halál vagy pénzbírság; hogy az elengedett bűnözők milyen veszélyt jelentenek a közösségre nézve; hogy mi az aktuális etikai állásfoglalás a büntetéssel kapcsolatban. A matematikai elmélet szempontjából a legtöbb, amit tehetünk, hogy megmutatjuk, hogy a hibák kockázatát hogyan lehet kontrollálni és minimalizálni. Ezeknek a statisztikai eszközöknek a használatát bármely adott esetben, annak eldöntésében, hogy pontosan hogyan kell megtalálni az egyensúlyt, a kutatóra kell hagyni.”

(Neyman és Pearson 1933: 296.; *kiemelés utólag*)

690 TELEPÜLÉSKUTATÁS – VIII. INTERPRETÁCIÓ

Wald tovább ment: „Az a kérdés, hogy hogyan kell megbatározni a *W* súly- (azaz a veszteség-) függvényt, nem matematikai vagy statisztikai jellegű. A statisztikusnak, amikor tesztelni akar bizonyos hipotéziseket, először meg kell batároznia minden lehetséges hiba relatív fontosságát, ami teljes mértékben kutatásának speciális céljaitól függ” (Wald 1939: 302.; *kiemelés utólag*). Az ilyen megközelítésekbe, vonzóak a közgazdászok számára (és nem csoda, hogy Wald számára is az volt, hiszen Karl Mengerrel tanulmányozta Bécsben az 1920-as években). De javaslatait figyelmen kívül hagyták a közgazdászok a szignifikanciaszintek publikálhatóságára vonatkozó konvenció kedvéért, amelyet még Fisher javasolt jóval korábban.

A statisztikusok jobban tudatában vannak saját elméletük intellektuális alapjainak, de sokan közülük kissé össze vannak zavarodva. Wald elméletének tiszta formában való használata során egy gyakorlati nehézség, ahogyan A. F. Mood és F. A. Graybill panasolja, hogy „ha a veszteségfüggvényt egyáltalán nem ismerjük vagy nem ismerjük pontosan, elegendő a [hipotézis] igazolásához. Ha a veszteségfüggvény nem ismert, úgy tűnik, hogy egy döntésfüggvény az ésszerű eljárás, ami bizonyos értelemben minimalizálja a hibák valószínűségét” (A. F. Mood és F. A. Graybill 1963: 278.). A „bizonyos értelemben” kifejezés egy feltáratlan retorika jelzésének tűnik az intellektuálisan tisztességesek körében. Mindenesetre az általuk ajánlott eljárás ésszerű lehet egy általános statisztikus számára, aki nem támaszt igényt arra, hogy megtudja, melyik az igazság jó vagy rossz megközelítése a statisztikán kívül. Ez azonban nem elfogadható a nemzetközi kereskedelem vagy a makroökonómia specialistája számára. Ha a veszteségfüggvény nem ismert, akkor meg kell találni. És ez együtt fog járni a kérdés retorikájának, a standardok hasznának tanulmányozásával.

Amikor a második világháború után a különböző x-ometriák – a filozófiában a pozitívizmussal, a politikában a társadalmi tervezéssel összekapcsolódva – a társadalomtudományokban megszokottá váltak, a szignifikanciaszint egyetemes reflexszé vált. Az 1950-es évek végén és az 1960-as években néhány szociológus és pszichológus tiltakozott. Tiltakozásuk a szignifikanciaszint-vita néven ismert, s a Denton E. Morrison és Ramon E. Henkel (1970) szerkesztette kiadványban jelent meg (lásd még Lieberman 1971; és Frank Schmidt, Iowai Egyetem, Gerd Gigerenzer, Münchener Egyetem és Louis Guttman, Héber Egyetem ugyanerre a pontra vonatkozó munkáit). A közgazdászok hozzászórtak ahhoz a feltételezéshez, hogy a statisztikai igényesség terén más társadalomtudósok előtt járnak. Ebben az ügyben, néhány kivételtől eltekintve (Arrow 1959; Griliches és Ohta 1976), egyáltalán nem ez a helyzet.

A régi klasszikus W. Allen Wallistól és Harry V. Robertstól, a *Statisztika: Új megközelítés*, amelyet 1956-ban publikáltak először, világossá teszi ezt a pontot:

„Lényeges, hogy ne keverjük össze a »szignifikáns« szó statisztikai használatát a mindennapi használattal. A mindennapi használatban a »szignifikáns« azt jelenti, hogy »gyakorlati fontossággal bír« vagy egyszerűen »fontos«. A statisztikai használatban a »szignifikáns« annak a populációnak egy jellemzőjét jelenti, amelyből a mintát vettük, tekintet nélkül arra, hogy ez a jellemző fontos-e vagy sem” (W. Allen Wallis és Harry V. Roberts 1956: 385.). Ez a nézőpont újraéledt a bevezető statisztikai könyvekben, bár csak néhányan fektettek rá hangsúlyt. Kiváló bevezető könyvükben David Freedman, Robert Pisani és Roger Purves (1978) nem is fogalmazhattak volna világosabban. A számos helyek egyikén, ahol ezt a pontot vizsgálták, azt írják: „Ez a fejezet (...) a szignifikanciatesztek korlátait magyarázza meg. Az első az, hogy a »szignifikancia« egy technikai kifejezés. Egy teszt vagy próba csak azzal a kérdéssel foglalkozhat, hogy vajon egy különbség valódi-e (állandó-e Venn értelmezése szerint), vagy csak egy variációs lebetősség. Nem arra lett kitalálva, hogy megmutassa, hogy a különbség fontos-e.” (David Freedman, Robert Pisani és Roger Purves 1978: 487.; *kiemelés utólag*) a különbségtétel szintén ilyen élesen jelenik meg David S. Moore és George P. McCabe írásában (1993: 474.)

Morris DeGroot, egy statisztikus némi közgazdasági kifinomultsággal, határozott nézeteket vall ebben a kérdésben:

„Különösen fontos (...) hogy különbséget tegyünk aközött, hogy az U egy megfigyelt értéke statisztikailag szignifikáns, és a paraméter aktuális értéke között (...) Egy adott probléma esetén az U megfigyelt értékéhez tartozó szélső terület lehet nagyon kicsi; sőt az aktuális érték lehet olyan közel [nullához], hogy gyakorlati szempontokat figyelembe véve, a kutató nem tekinteti azt [lényegében] különbözőnek [nullától] (...) Igen valószínű, hogy egy 20 000-es mintán alkalmazott t -próba U statisztikailag szignifikáns értékéhez vezet (...) [A kutató] előre tudja, hogy nagy a valószínűsége a [nulla] elutasításának még akkor is, ha a valódi érték (...) [aritmetikailag] csak alig különbözik [a nullától].”

(DeGroot 1975: 469–497.)

De nem sok más ökonométer különbözteti meg a közgazdasági szignifikanciát a statisztikai szignifikanciától: Frank Denton például megkülönbözteti (1988); Goldberger, ahogy korábban már megjegyeztük, és néhányan mások is. És még kevesebben hangsúlyozzák a közgazdasági szignifikanciát. Az 1970-es és 1980-as években, amikor ez a gyakorlat standarddá vált, a széles körben használt ökonometriai szövegekben, mint amilyen Jan Kmentától *Az ökonometria elemei* (1971) vagy Johnstontól az *Ökonometriai módszerek* (1984, 1963-ban adták ki

692 TELEPÜLÉSKUTATÁS – VIII. INTERPRETÁCIÓ

először), nem említik meg közgazdasági szignifikanciát a statisztikai szignifikanciával szemben. Peter Kennedy *Útmutató az ökonometriához* (1985) c. könyvében röviden megemlíti, hogy elég nagy minta esetén mindig statisztikailag szignifikáns különbségeket kapunk. Ez csak egy része az érvelésnek, nem az egész, és Kennedy mindenesetre ezt a részleges érvelést is a jegyzetekhez teszi (62.).

Bevezető könyvükben, *A statisztika erejének felfedezésében* Ronald J. Wonnacott (közgazdász) és Thomas H. Wonnacott dióhéjban így foglalják össze a problémát:

„A »statisztikai szignifikancia« kifejezéssel van egy kis probléma. Ez egy technikai kifejezés, ami egyszerűen azt jelenti, hogy elegendő adatot gyűjtöttek össze annak megállapításához, hogy egy különbség létezik. Ez nem jelenti azt, hogy a különbség szükségképpen fontos. Például, ha egy óriási mintát vettünk egy közel azonos populációból, a különbség olyan apró lehet, hogy figyelmen kívül hagyhatjuk, nem lévén valódi szignifikanciája, még akkor sem, ha statisztikailag szignifikáns. Más szavakkal, a statisztikai szignifikancia egy technikai kifejezés, amelynek jelentése nagymértékben eltér a közönséges szignifikanciától és fontosságtól (...) Sajnálatos, de értetető módon sok ember bajlamos arra, hogy összekeverje a statisztikai szignifikanciát a közönséges szignifikanciával.”

(R. J. Wonnacott és T. H. Wonnacott 1982: 160.; a szerzők kiemelése)

Egyetlen kifogásom az, hogy Wonnacotték nem mennek elég messzire. Az nem igaz, hogy a statisztikai szignifikancia azt mutatja, hogy „elegendő adatot gyűjtöttek össze annak megállapításához, hogy egy különbség létezik”. A kivitelezés módja következtében tűnik úgy, hogy a statisztikai szignifikancia az elemzés megfelelő eszköze. De nem az.

A statisztikai szignifikancia használata sok bajt okozhat

Nem könnyű a valószínűségi modellek használatát megindokolni akkor, amikor nem valószínűségi kérdésekre keresünk választ. Visszavághatsz azzal, hogy a közgazdászok, legalábbis akik jók a szakmájukban, nem követnek el ilyen hibákat. De igen, megteszik, ahogyan az a legjobb munkáikból is látható, az *Amerikai Közgazdasági Szemlében*. Stephen Ziliak és én megnéztünk az 1980-as években a *Szemlében* megjelent minden teljes terjedelemben közölt empirikus

tanulmányt, és megvizsgáltuk egy tizenkilenc kérdésből álló kérdőív alapján (McCloskey és Ziliak 1996). A kutatás kérdései a következők voltak:

1. *A tanulmány kisszámú megfigyelést használ, amelynek esetében nem találhatók statisztikailag szignifikáns különbségek a konvencionálisan elfogadott szinteken pusztán nagyszámú megfigyelés kiválasztásával?* A próba ereje magas, ha 30 vagy 300 elemű minta helyett $N=30000$ esetén vizsgáljuk a szignifikanciaszintet. Például Glen C. Blomquist, Mark C. Berger és John P. Hoehn-nél, ahol $N=34414$ lakóegység és 46004 egyén szerepel a mintában (March 1988: 93.). Ilyen nagyméretű minták esetén a szerzőnek oda kell figyelnie a próba erejének és méretének felcserélődésére és a próba erejének közgazdasági szignifikanciájára az alternatívákkal szemben.
2. *Az egységek és a leíró statisztikák minden regressziós változóra vonatkozóan szerepelnek?* A közgazdaságtanban az empirikus munka mérést jelent. Alapvető követelmény a változók egységeinek figyelembevétele és az átlagok megadása.
3. *Az együttthatókról valamilyen rugalmas formában számolnak-e be, vagy valamilyen olyan értelmezhető formában, amely a tárgyalt probléma szempontjából releváns, és összhangban van a közgazdaságtan elméletével, úgy, hogy az olvasók meg tudják ítélni a regressziós változók közgazdasági hatását?* Wallis és Roberts régen arról panaszkodott, hogy „időnként a szerzők olyan fondorlatosan bánnak a szignifikanciatesztekkel, hogy még a hatás aktuális mértékét (amount), sem sikerül megállapítani, és még kevésbé sikerül a gyakorlati fontosságát értékelni” (Wallis és Roberts 1956: 409.). Bizonyos területeken (nem annyira a közgazdaságtanban, bár egy példát itt is találtunk) a kutató olyan táblákat publikál, amelyek csak a szignifikanciaszinteket jelentő csillagokból állnak.
4. *A megfelelő nullhipotézist specifikálták-e?* A legáltalánosabb probléma a zéró érvénytelenségének tesztelése, amikor valamilyen más feltevés érvénytelenségének kimutatása a cél. Egy ilyen hiba lehet annak az eredménye, ha megengedjük, hogy egy előre megírt számítógépes program hozza meg a tudományos döntéseket. Ha a nullhipotézis az, hogy $\beta_1 + \beta_2 = 1$, nem sokat nyerünk annak a hipotézisnek a teszteléséből, hogy mindegyik együtttható statisztikailag szignifikánsan különbözik nullától. A Neyman–Pearson-teszt leggyümölcsözőbb alkalmazása úgy határozza meg a nullhipotézist, mint valami olyasmit, amiről a kutató azt gondolja, hogy igaznak kell lennie. Az egyetlen olyan eredmény, amely határozott következtetéshez vezet, a nullhipotézis elvetése. Az elvetés megghiúsulásából természetesen nem következik, hogy a nullhipotézis igaz. És a nullhipotézis elvetéséből nem következik az, hogy az alternatív hipotézis igaz: lehetnek más alternatívák (olyan

694 TELEPÜLÉSKUTATÁS – VIII. INTERPRETÁCIÓ

lehetőségek köre, amelyben a kutatók egyetértenek, például) amelyek a nullhipotézis elvetéséhez vezetnek. A hipotézis elvetésének retorikája elősegíti azt a könyv ízű eljárást, amely szerint „vedd a magasság, jövedelem, ország, életkor regresszióját”; vizsgálj meg a t -értékeket; ne tekintsd fontosnak, ha $t < 2$; tekintsd fontosnak, ha $t > 2$.

5. *Az együtthatókat megfelelő körültekintéssel értelmezték-e?* Goldberger erre mutat példát, amely hasonló ahhoz, mint amilyenek sokszor előfordulnak a közgazdaságtanban (Goldberger 1991). Tegyük fel, hogy a függő változó „súly fontokban”, a nagyobb együttható a „magasság”-nál szerepel, a kisebb együttható a „testgyakorlás”-nál, és a becsült együtthatóknak ugyanaz a standard hibájuk. Sem az orvosnak, sem pedig a páciensnek nem sok haszna lenne egy olyan elemzésből, amely azt mondja, hogy a magasság a „fontosabb” (együtthatója nagyobb standard eltérést mutat a nullától ebben a mintában), és valójában azt a magyarázatot adja a túlsúlyos páciensnek, hogy ő nem túl kövér, pusztán túlságosan alacsony ehhez a súlyhoz. *„Ennek a példának a tanulsága az, hogy a »fontosság« statisztikai mérése elterelés a kutatás valódi céljától – a fontos paraméterek becslésétől – a függő változó »szórásának magyarázata« irányába”* (Goldberger 1991: 241.).
6. *A tanulmány elkerüli-e azt, hogy az összes t - vagy F -statisztikáról vagy standard hibáról beszámoljon, tekintet nélkül arra, hogy a szignifikanciateszt megfelelő-e vagy sem?* A statisztikai számításokat végző szoftverek rutinszerűen megadják a t -statisztikákat minden becsült együtthatóra. De az a tény, hogy a program megadja ezeket, nem jelenti azt, hogy az információ lényeges a tudomány számára. Azt gyanítjuk, hogy a beszámolókat készítőik erőltetik a jelentés nélküli t - és F -statisztikák elburjánzását, abból a hiedelemből kiindulva, hogy a statisztikai és a valóságos szignifikancia ugyanaz.
7. *A statisztikai szignifikancia az első alkalmazáskor, általában a tanulmány tudományos csúcspontjánál, a „fontosság” egyetlen kritériuma?* „Csúcsponton” azt a helyet értjük a tanulmányban, ahol a szerző megállapítja, hogy milyen nyilvánvaló következtetést vont le a kritikus próba alapján.
8. *Említi-e a tanulmány a próba erejét?* Például Frederic S. Mishkin ezt teszi, szokatlan módon két lábjegyzetben (Mishkin 1981: 298., 11. lj.; 305., 27. lj.) a próba erejének hiánya állandó nehézség a tőkepiaci tanulmányokban, de ritkán szembesülünk vele). Ahogyan DeGroot kimutatta, egy próba ereje lehet alacsony egy szomszédos és valóságosan szignifikáns alternatívával szemben. Másrészt, az erő lehet magas egy szomszédos és triviális alternatívával szemben.
9. *Ha a tanulmány megemlíti a próba erejét, tesz valamit ezzel kapcsolatban?* Igaz, hogy a próba erejét csak egy explicit ellenhipotézishez viszonyítva lehet meghatározni, és a próba erejének elemzése bizonyos ellenhipotézisek ese-

tén nehéz. Egy példa erre az, amikor Durbin-Wu-Hausman azt vizsgálja, hogy vajon két becslőfüggvény konzisztens-e. (A kutatás beszámol a nehézségről, az érintett tanulmányoknál jelezve, hogy „nem alkalmazható”.)

10. *Elkerüli-e a tanulmány a „csillag-ökonometriát”, azaz az együttthatók ragso-rolását a t-statisztikák abszolút nagyságának megfelelően?*
11. *Elkerüli-e a tanulmány a „jel-ökonometriát”, azaz, csak jelzi a szignifikanciát, de az együttthatók nagyságát nem adja meg?* Az ökonometriai könyvekben van egy kevés statisztikai elmélet, amely emögött a szokásos gyakorlat mögött áll (Goldberger 1991, 22. fejezet), bár legnagyobb részben a szokásos alkalmazás megelőzi az elméletet. De a jel nem *közgazdaságilag* szignifikáns, hacsak a terjedelme nem elég nagy, hogy számítson. A statisztikai szignifikancia nem mondja meg, hogy vajon a méret elég nagy-e, hogy számítson. Nem igaz az, ahogyan a szokásos érvelés hangzik, hogy a jel statisztikailag független a terjedelemtől.
12. *Megadja-e a tanulmány az együttthatók nagyságát?* Azaz ha egyszer a regressziós eredményeket bemutatták, rámutat-e a tanulmány arra, hogy bizonyos együttthatók és a hozzájuk tartozó változók *közgazdaságilag* befolyásosak, miközben más változók nem azok? Blomquist, Berger és Hoehn részben ezt tették, amikor a lakást és a lakóköznyezetet kellemesebbé tévő dolgokra vonatkozó együttthatókat dollár formájában adták meg. Ők azonban nem fejtették ki azt, hogy ezek nagysága tudományosan indokolható-e, vagy valamilyen más módon fontos. Ezzel szemben Christina Romer egy tizenkilenc oldalas, kizárólag empirikus tanulmányában azt írja: „*Valóban, a leltári mozgások korrigálása csökkenti a diszkrépanciát (...) körülbelül a felével. Ez azt mutatja, hogy a leltári mozgások [közgazdaságilag] fontosak*” (Romer 1986: 327.). M. Boissiere, J. B. Knight és R. H. Sabot tükrözi a tipikusabb gyakorlatot: „*A kognitív teljesítmények mindkét országban nagymértékben szignifikáns kapcsolatot mutatnak az oktatás szintjével (...) Kenyában a középfokú oktatás H-t 11.75 ponttal emeli, vagy az átlag 35 százalékkal*” (M. Boissiere, J. B. Knight, és R. H. Sabot 1985: 1026.). Ők a „szignifikancia” kifejezést nem egyértelműen használják, amikor visszamennek a közgazdasági szignifikancia releváns kérdéséhez. Később ebben a részben ismét visszatérnek a statisztikai szignifikancia kizárólagos használatához: a „szignifikánsan pozitív” és a „csaknem szignifikánsan pozitív” ismét a fontosság egyedüli kritériumává válik.

Daniel Hamermesh, ezzel ellentétben, a *K* kritikus paramétert becsüli, és az első említésnél azt mondja: „*A K becslései elég nagyok, ebből arra lehet következtetni, hogy a cég csak az igazán nagy megrázkódtatásokra válaszol foglalkoztatáspolitikájának változtatásával (...) Vegyük figyelembe, amit egy ilyen becslésnél a nagy értékek jelentenek*” (Hamermesh 1989: 683.). Itt a forma közel van az ideálshoz: tudományos kérdéssé válik, hogy mit jelent a

696 TELEPÜLÉSKUTATÁS – VIII. INTERPRETÁCIÓ

fontosság mértéke. Két bekezdéssel később arról beszél, hogy „elég nagy”, „igazán fontos”, „kicsi” és „fontos”, anélkül azonban, hogy utalna arra, hogy ezek mit jelentenek a statisztikai szignifikancia szempontjából. Goldberger kifejezéseivel, „*a kutatás sajátos céljára koncentrálni – a fontos paraméterek becslésével*”. (Később azonban Hamermesh visszaesik a szokásos gyakorlatba: „*A K-kalap az aggregált adatokra a 2. táblázatban nem szignifikáns*”, bár bölcsen hozzáteszi, „*és igazán kicsi; és a p-kalap átlagos értékei sokkal magasabbak, mint az egyesített adatoknál*” [Hamermesh 1989: 685.].)

13. *Tárgyalja-e a tanulmány tudományos megközelítésben, hogy mely együttbátókat kell „nagyinak” vagy „kicsinek” tekinteni?* Romer például megjegyzi, hogy „*annak a határozott ténynek (azaz a tudományos konszenzusnak) a létezése, hogy a közgazdaságtan megállapodott, egy általános konszenzust von maga után*” (Romer 1986: 322.).
14. *Elkerüli-e a tanulmány azt, hogy kizárólag a statisztikai szignifikancia alapján válassza ki a változókat?* A szokásos érvelés az, hogy ha bizonyos változók szignifikánsan lépnek be a modellbe, az információt nem kell elutasítani. De az ilyen érvelés egybeolvasztja a statisztikai és a tényleges szignifikanciát.
15. *A csúcspontot követően, elkerüli-e a tanulmány a statisztikai szignifikancia fontossági kritériumként való használatát?* A szakértők gondolkodás nélkül ragaszkodnak a szignifikanciateszthez, a körültekintő szerző beleegyezik a ragaszkodásukba, de miután erről beszámolt, vissza kell térnie a fontosság más tudományosan releváns kritériumaihoz.
16. *A statisztikai szignifikancia döntő-e, az érvelést megáll-e ennél a pontnál, a befejezettség érzetét kelti-e?* Romer és Jeffrey Sachs (1980) mindketten használják a statisztikai szignifikanciát, és mindketten rosszul használják – mindkét esetben úgy tekintenek a statisztikai szignifikanciára, mint annak a kritériumára, hogy mennyire nagy a nagy. De egyik tanulmányban sem terjed ki a statisztikai szignifikancia az empirikus munkára. A nem megfelelő használat Michael Darbynál (1984) még leplezetlenebb: egyetlen érve egy együttbátó mellett, amikor regressziót futtat, a statisztikai szignifikancia (Darby 1984: 311, 315.), másrészt azonban, megállapításai nem utalnak vissza a regresszió eredményeire.
17. *A tanulmány mindig szimulációt alkalmaz-e annak meghatározására, hogy az együttbátók elfogadhatóak-e (szemben a regresszió inputként való használatával a további érvelés során)?* Bizonyos mértékig Blomquist, Berger és Hoehn ezt teszik. Szimulálják a városok vonzás szerinti sorrendjét, és ha az együttbátók nem megfelelőek, akkor magát a rangsort nem tekintik elfogadhatónak. Santa Barbara magas sorszámot kap, bár a kellemesség legrosszabbtól a legjobbig tartó skálán való megítélése, 5,146 \$ alacsonynak tűnik, ha jártál már Santa Monicában és East St. Louisban (Blomquist, Berger,

Hoehn 1988: 96.). A regressziós együtthatókat használó szimulációk informatívak lehetnek, de természetesen nem kell a statisztikai szignifikanciát használni az input meghatározásához.

18. A „következtetések” és „implikációk” részekben a statisztikai szignifikanciát külön kezelik-e a közgazdasági, politikai és tudományos szignifikanciától? Boissiere, Knight és Sabot (1985) tanulmányukban a képesség hatását jól elkülönítik, de a közgazdasági szignifikanciával nem foglalkoznak.
19. Elkerüli-e a tanulmány a „szignifikancia” szó kétértelmű használatát, vagy előfordul, hogy az egyik mondatban a „statisztikailag szignifikáns” kifejezést használja, a másokban pedig azt, hogy „elég nagy abhoz, hogy a politika és a tudomány számításba vegye”? Például Darby (1984) – „Először azt kívánjuk vizsgálni, hogy az olajárak, az árak kontrollja vagy mindkettő szignifikáns hatást gyakorol-e a termelékenység növekedésére” (Darby 1984: 310.). – a különböző jelentéseket egybemossa.

A sok-sok elrontott dologról

Az *Amerikai Közgazdasági Szemlé*ben az 1980-as években számos hiba fordult elő a statisztikai szignifikancia alkalmazása során.

Kutatási kérdés	A tanulmányok száma, amelyekre a kérdés vonatkozott	Az „igen” válaszok aránya százalékban
A tanulmány		
8. Figyelembe veszi-e a próba erejét?	182	4,4
6. Elkerüli-e azt a hibát, hogy minden standard hibáról, t- és F-statisztikáról beszámoljon, akkor is, amikor ezek az információk irrelevánsak?	181	8,3
17. Használ-e szimulációt annak eldöntésére, hogy az együtthatók elfogadhatóak-e?	179	13,2
9. Vizsgálja-e az erőfüggvényt?	12	16,7
13. Tudományos megközelítésben tárgyalja-e, hogy mely együtthatók tekinthetők nagyoknak és melyek kicsinek?	181	28,0
16. A gyakorlati érvelés során figyelembe vesz-e a statisztikai szignifikancián kívül más bizonyító erejű tényezőket?	182	29,7
18. A következtetések levonása során különbséget tesz-e a statisztikai és a tényleges szignifikancia között?	181	30,1

698 TELEPÜLÉSKUTATÁS – VIII. INTERPRETÁCIÓ

2.	Megadja-e a regressziós változók leíró statisztikáit?	178	32,4
15.	A csúcspontot követően a statisztikai szignifikancián kívül használ-e más fontossági kritériumot?	182	40,7
19.	Elkerüli-e a „szignifikancia” szó nem egyértelmű használatát?	180	41,2
5.	Megfelelő körültekintéssel értelmezi-e az együttthatókat? Például, figyelmet fordít-e a mérési egységre és az adatok korlátaira?	181	44,5
11.	Elkerüli-e a „jel-ökonometriát”, vagyis azt, hogy az együttthatókat ne a nagyságukkal, hanem csak jelekkel jelölje?	181	46,7
7.	Első alkalmazásakor a statisztikai szignifikanciát csak a fontossági kritériumok egyikének tekinti-e?	182	47,3
3.	Az együttthatókról rugalmasan számol-e be, vagy valamilyen más alkalmas formában, amely felveti azt a kérdést, hogy „milyen nagy a nagy”?	173	66,5
14.	Elkerüli-e azt, hogy a változókat kizárólag a statisztikai szignifikancia alapján válassza ki?	180	68,1
10.	Elkerüli-e a „csillag-ökonometriát”, vagyis az együttthatóknak a teszt statisztika abszolút nagysága szerinti ragsorolását?	182	74,7
12.	Tárgyalja-e az együttthatók nagyságát, rámutat-e a tényleges szignifikanciára?	182	80,2
1.	Olyan kevés számú megfigyelést használ-e, hogy nem találhatók statisztikailag szignifikáns különbségek – csak nagyon nagy minta választásával?	182	85,7
4.	Azokat a nullhipotéziseket teszteli-e, amelyekről a szerzők azt állították, hogy fontosak?	180	97,3

A 3–7. táblázatokhoz felhasznált források: Az 1980–1989 között az *Amerikai Közgazdasági Szemlében* teljes terjedelemben megjelent összes olyan tanulmány, amely regresszióelemzést alkalmazott, a konferenciaanyagok kivételével.

Megjegyzések: az „igen válaszok aránya százalékban” nem más, mint az „igen” válaszok száma osztva az érintett tanulmányok számával (amely legfeljebb 182 lehet). Bizonyos kérdések nem alkalmazhatók általánosságban egyes tanulmányokra, és bizonyos kérdések azért nem alkalmazhatók, mert csak a meghatározott jellemzőkkel rendelkező tanulmányokra vonatkoznak. A 3. kérdés esetén például azokat a tanulmányokat kódoltuk a „nem vonatkozik rá” kóddal, amelyek kizárólag nem paraméteres statisztikákat alkalmaztak. A 19. kérdés azon tanulmányok esetén kapta a „nem vonatkozik rá” kódot, amelyekben nem használták a „szignifikancia” szót.

A vizsgálat során a következő fontosabb eredményekre jutottunk:

- Az *Amerikai Közgazdasági Szemlében* az empirikus írások hetven százaléka nem különböztette meg a statisztikai szignifikanciát a közgazdasági, politikai vagy tudományos szignifikanciától. (...)
- (...) egy együttható becslési eljárását követően 72 százaléuk nem vette figyelembe (...) hogy más szerzők milyen standardokat alkalmaztak a „fontosság” meghatározására (...)
- Ötvenkilenc százaléuk nem egyértelműen használta a „szignifikancia” szót (...)
- A több szerző által jegyzett tanulmányok esetén (...) gyakrabban fordult elő a „szignifikancia” nem egyértelmű használata, a jel-ökonometria, a becslült együtthatók nagyságának figyelmen kívül hagyása.
- (...)

Reménykeltő jel, hogy a statisztikai szignifikancia helytelen használata – úgy tűnik – részben „évjárat”-hatás... Azok a tanulmányok, amelyeknek szerzői PhD-fokozatukat 1975 és 1979 között szerezték – amikor az olcsón előállítható t-próbák először jutottak el szélesebb körhöz –, lényegesen rosszabbak más tanulmányoknál, ha a közgazdasági és a statisztikai szignifikancia közötti különbségtételt nézzük. (...)

Ha a közgazdászok nem akarják kockára tenni a tudomány hitelességét, újra kell gondolniuk a statisztikai szignifikancia retorikáját. Valamit tenni kell. (...) Segítene, ha a standard programcsomagok nem állítanák elő a t-statisztikákat ilyen pazarlóan. (...) A programok ilyeneket kérdezhetnének, jó nagy betűkkel: „VALÓBAN VÉLETLEN MINTÁD VAN?” „FIGYELEMEBE VETTED A PRÓBA EREJÉT?”, és mindenekelőtt, „MILYEN TUDOMÁNYOS STANDARD ALAPJÁN ÍTÉLED MEG, HOGY EGY EGYÜTTHATÓ KICSI VAGY NAGY?”

IRODALOM:

- ARROW, KENNETH 1959: Decision Theory and the Choice of a Level of Significance for the t-Test. In Ingram Olkin és mások (szerk.): *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford: Stanford University Press.
- BLOMQUIST, GLENN C., MARK C. BERGER és JOHN P. HOEHN 1988: New Estimates of Quality of Life in Urban Areas. *American Economic Association*, vol. 78 (1): 89–107.
- BOISSIERE, M., J. B. KNIGHT és R. H. SABOT 1985: Earnings, schooling, ability, and cognitive skills. *American Economic Association*, vol. Review 75 (5): 1016–1030.

- BORING, EDWIN G. 1919: Mathematical versus Scientific Significance. *Psychological Bulletin*, 16 (October): 335–338.
- CICERO, MARCUS TULLIUS 1942: *De Oratore*. Trans. E. W. Sutton. Vol. 1. Cambridge MA: Harvard University Press.
- COASE, RONALD 1982: *How Should Economists Choose?*. The G. Warren Nutter Lectures in Political Economy, Washington, D.C.: American Enterprise Institute.
- DARBY, MICHAEL R. 1984: The US Productivity Slowdown: A Case of Statistical Myopia. *American Economic Association*, vol. 74 (3): 301–322.
- DEGROOT, MORRIS H. 1975: *Probability and Statistics*. Menlo Park, CA: Addison-Wesley.
- DENTON, FRANK 1983: Econometric Data Mining as an Industry. Unpublished MS, Department of Economics, McMaster University, October (forthcoming in the *Review of Economics and Statistics*).
- FISHER, RONALD A. 1925: *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- FREEDMAN DAVID, ROBERT PISANI és ROGER PURVES 1978: *Statistics*. New York: Norton.
- GOLDBERGER, ARTHUR 1991: *A Course in Econometrics*. Harvard MA: University Press.
- GREENWOOD, Major és G. UDNEY YULE 1915: *The Statistics of Antityphoid and Anticholera Inoculations, and the Interpretation of such Statistics in General*. *Pro. Roy. Soc. Med.* 1914–15; 8: 113–194.
- GRILICHES, ZVI és MAKOTO OHTA 1976: Automobile Prices Revisited: Extensions of the Hedonic Hypothesis. In N. E. TERLECKYJ (szerk.): *Household Production and Consumption*. Studies in Income and Wealth, vol. 40. New York: National Bureau of Economics Research. 325–390.
- HAMERMESH, DANIEL S. 1989: Labor Demand and the Structure of Adjustment Costs. *American Economic Association*, vol. 79 (4): 674–689.
- HOGBEN, LANCELOT [1968]: *Statistical Theory: The Relationship of Probability, Credibility, and Error*. New York: Norton.
- JOHNSTON, JOHN 1984, 1963: *Econometric Methods*. New York: McGraw Hill.
- KENDALL, M. G. és STUART, A. 1951: *The Advanced Theory of Statistics*. 3rd ed. Vol. 2. London: Griffin.
- KENNEDY, PETER 1985: *A Guide to Econometrics*. Cambridge, MA: MIT Press.
- KMENTA, JAN 1971: *Elements of Econometrics*. New York: MacMillan.
- KRUSKAL, WILLIAM 1978: Formulas, Numbers, Words: Statistics in Prose. *The American Scholar*, 47 (Spring): 223–229. Reprinted in D. Fiske (szerk.): *New Directions for Methodology in Social and Behavioral Sciences*. San Francisco: Jossey-Bass, 1981.

- KURTZ, A. K. és EDGERTON, H. A. (szerk.) 1939: *Statistical Dictionary of Terms and Symbols*. New York: Wiley.
- LIEBERMAN, BERNHARDT 1971: *Contemporary Problems in Statistics: A Book of Readings for the Behavioral Sciences*. New York: Oxford University Press.
- MCCLOSKEY, DEIRDRE N. és STEPHEN T. ZILIAK 1996: The Standard Error of Regressions. *Journal of Economic Literature*, vol. 34 (1): 97–114.
- MISHKIN, FREDERIC S. 1981: Are Market Forecast Rational, *American Economic Association*, vol. 71 (3): 295–306.
- MOOD, A. F. és GRAYBILL, F. A. 1963: *Introduction to the Theory of Statistics*. 2nd ed. New York: McGraw-Hill.
- MOORE, DAVID S. és GEORGE P. McCABE 1993: *Introduction to the Practice of Statistics*. Second ed., New York: WH Freeman.
- MORRISON, DENTON E. és HENKEL, RAMON E. 1970: *The Significance Test Controversy: A Reader*. Chicago: Aldine.
- MORRISON, DENTON E. és HENKEL, RAMON E. 1969: Significance Tests Reconsidered. *American Sociologist*, 4 (May): 131–140. Reprinted in Morrison és Henkel 1970.
- MOSTELLER, FREDERICK és TUKEY, JOHN W. 1977: *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley.
- NEYMAN, JERZY és PEARSON, E. S. 1933: On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society*, ser. A, 231: 289–337.
- PEARSON, KARL 1911: Probability That Two Independent Distributions of Frequency Are Really Samples from the Same Population. *Biometrika*, vol. 8.
- ROMER, CHRISTINA 1986: Is the Stabilization of the Postwar Economy a Figment of the Data? *American Economic Association* vol. 76 (3): 314–334.
- SACHS, JEFFREY 1980: The changing Cyclical Behavior of Wages and Prices 1890–1976. *American Economic Review*, American Economic Association, vol. 70 (1): 78–90.
- SCOTT, ELIZABETH 1953: Testing Hypotheses. In R. J. TRUMPLER és H. F. WEAVER (szerk.): *Statistical Astronomy*. New York: Dover.
- STIGLER, STEPHEN M. 1986: *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press.
- WALD, ABRAHAM 1939: Contributions to the Theory of Statistical Estimation and Testing Hypotheses. *Annals of Mathematical Statistics*, 10 (December): 299–326.
- WALLIS, W. ALLEN és HARRY V. ROBERTS 1956: *Statistics: A New Approach*. Chicago: Free Press.
- WONNACOTT, RONALD J. és THOMAS H. WONNACOTT 1982: *Statistics: Discovering its Power*. New York: Wiley.

