

Németh Renáta, adjunktus, ELTE TáTK, Statisztika tanszék

nemethr@tatk.elte.hu

nemethr.web.elte.hu

Az alábbi kézirat a socio.hu-nak benyújtott változat. A végleges változat a folyóirat 2014/3-as számában jelenik meg.

kiadó: MTA Társadalomtudományi Kutatóközpont, Szociológiai Intézet.
Minden jog fenntartva @ Németh Renáta.

Módszerek a kvantitatív társadalomkutatási paradigmákban^{1,2}

1. Kivonat

A kvantitatív társadalomkutatás paradigmái nem csak a kutatási problémákat, hanem a módszereket tekintve is behatárolják a szóba jöhető alternatívák körét. Vagyis, bár ritkán reflektálunk erre, a statisztikai eszközök közötti választás nem pusztán technikai kérdés. Hasonlóan ritkán merül fel az a szempont, hogy maga az eszköz is hatást gyakorolhat az általa szolgáltat tudományra, annak szemléletére, fogalmaira, kérdésfeltevéseire. Írásomban ezeket az álláspontokat igyekszem néhány – köztük Bourdieu korrespondenciaelemzését, a mobilitáskutatás paradigmáit, a Big Data társadalomkutatási felhasználását és általában a módszertani innovációk diffúzióját érintő – példával alátámasztani.

kulcsszavak: tudományos paradigmák, kvantitatív módszerek, statisztikatörténet

Mottó:

„Forradalom ment végbe a társadalmi mobilitás kutatásában: az egykoron uralkodó Blau-Duncan paradigmát a loglineáris modellezés döntötte le a trónról. [...] A loglineáris forradalom figyelemre méltó kísérlet volt, és kezdetben úgy tűnt, fényes új jövőt kínál Blau-n és Duncan-en túllépve. [...] Végül, én úgy gondolom, Goldthorpe kísérlete elbukott. Azóta több mint egy évtized telt el. Számptalan modellt illesztettek, design-mátrixok seregeit vették számba, khi-négyzetek ezreit sorakoztatták fel. Mégis, kevés lényegi újdonság derült ki. Teljes részletességgel írhatók már le a szűk keresztátlába szorított apai és fiúi foglalkozás cellagyakorosságai, ám az ehhez kapcsolódó értelmezés ad hoc, az elmélet szegényes. [...] Mindez visszalépést jelent a statisztikailag egyszerűbb, de koncepcionálisan összetettebb Blau-Duncan paradigmához képest.”

(Idézet Jonathan Kelley 1990-es *The failure of a paradigm* című, több kritikát provokált tanulmányából)

2. Motiváció

Kvantitatív szociológiai módszertan a területem, s régóta foglalkoztat a szociológia módszertani divatjainak / módszertani paradigmáinak kérdése. Gyakran szembesültem vele,

¹ A dolgozat a Kulcskérdések a társadalomkutatásban – Kánon és apokrif c. konferencián elhangzott előadásom (Németh, 2013) kibővített változata.

² A tanulmány elkészítése során az MTA Bolyai János kutatási ösztöndíj támogatását élveztem.

hogy különböző korszakokban, különböző tudományterületeken vagy különböző országokban hagyományosan eltérő statisztikai módszereket alkalmaznak, és azzal is, hogy ez jelentősen eltérővé teszi az alkalmazók szemléletét. Írásomban néhány ezzel kapcsolatos megfigyelésemet szeretném rögzíteni.

A mottóbeli idézet jól illusztrálja tanulmányom mondanivalóját: a mobilitáskutatás egymást követő paradigmái nem csak tartalmilag, hanem módszertanukat tekintve is eltérnek egymástól, sőt, sok esetben módszertanuk fémjelzi őket. Jonathan Kelley³ ebben az írásában expliciten használja a paradigma kifejezést, mégpedig kuhniánus megközelítésben, említi a „normál tudomány” és a „tudományos forradalom” kifejezéseket is. Amit kiemelnék: az egyes paradigmákat az általuk használt módszerekről nevezi el, és tulajdonságaikban is technikai-módszertani jegyeket emel ki.

3. Bevezetés

Tanulmányomban azt az álláspontomat szeretném néhány példával alátámasztani, hogy a kvantitatív társadalomkutatás paradigmái nem csak a kutatási problémákat, hanem a módszereket tekintve is behatárolják a szóba jöhető alternatívák körét. Ha a módszerek megválasztása racionálisan történne, akkor kizárólag az adott problémára való érvényesség döntene. Ezzel szemben, bár ritkán reflektálunk erre, számos egyéb, a tudományos közösségből eredő társadalmi szempont is érvényesül. Ez a megfigyelés (1) **a kutatási paradigmának a módszertan megválasztására gyakorolt hatását** hangsúlyozza.

Hasonlóan ritkán merül fel az a szempont, hogy maga a módszer is befolyással lehet az általa szolgált tudományra. Módszereink behatárolhatják a tudományos kérdések körét, formálhatják szemléletünket, alakíthatják fogalmainkat. Ez az állítás a fentihez képest fordított hatásirányt: (2) **a módszertannak a kutatási paradigmára gyakorolt hatását** húzza alá.

Írásomban ezt a két állítást igyekszem néhány – köztük Bourdieu korrespondenciaelemzését, a mobilitáskutatás paradigmáit, a Big Data társadalomkutatási felhasználását és általában a módszertani innovációk diffúzióját érintő – példával alátámasztani. Az argumentáció tehát példákra épül majd. Meggyőződésem, hogy tudományos munkákban a példák felhozása általában, de a jelen tárgykör esetén különösen, nem csupán retorikai fogás, nem csupán illusztrálása a fő mondanivalónak, hanem éppen ellenkezőleg: igazán a példák specifikálják a mondanivalót. A példák (igaz, többnyire implicit módon) leszűkítik a bennfoglalt problémák körét, konkretizálják a területet, a fogalmi kiindulást.

A téma jellegénél fogva tehát nem a klasszikus kutatási kérdés – adatelemzés – konklúzió logikát követem majd, nem is állítok/bizonyítok általánosan, csupán néhány megfigyelésem kiragadásával igyekszem a fent vázolt kétféle hatás plauzibilitását alátámasztani. Fontos azt is megjegyezni, hogy a cikkben hivatkozott argumentumok egy része közismert: például, a tudományos nézetek diffúziójának interperszonális jellegével kapcsolatban számtalan tanulmány ismert. Az újszerűséget inkább ezeknek a tudományos módszerekre való alkalmazása és más argumentumok mellé helyezése, közös keretbe foglalása jelenti.

A tudományos paradigma általam használt értelmét is példákon keresztül világítom meg, itt röviden mégis kitérnék erre az alapvető fogalomra. Mivel nem célom a paradigmák

³ Kelley a magyar olvasónak onnét is ismerős lehet, hogy Kolosi Tamással publikált 1992-ben az American Sociological Review hasábjain egy magyar-ausztrál összehasonlító elemzést.

tudományelméleti elemzése, a fogalomnak nem követem egyik elméleti definícióját sem, a **paradigma fogalmát** hétköznapi értelemben használom. Az általam adott jelentésben a paradigma lehet a tudományos elmélet, iskola vagy kutatási hagyomány többé-kevésbé átfedő megfelelője. Azért használom mégis az utóbbiak helyett, hogy a társadalomkutatás intézményesült jellegét hangsúlyozzam, ezzel kapaszkodót teremtve a statisztikai eszközök kiválasztásának (és ezzel együtt például a kutatások dinamikájának, innováció-terjedésnek, interdiszciplináris hatásoknak) a társadalmi motivációinak megértéséhez. A paradigma általam használt fogalmának lényege, hogy intézményesülve erősen befolyásolja a tudományos gondolkodást, egyfajta elméleti és módszertani keretet teremt, a tudományos szocializációban is fontos szerepet betöltve (egyetemek, alapkönyvek). Nem lesz szükségem rá, ezért nem vállalkozom sem a paradigmák méretbeli lehatárolására („mekkora elmélet a paradigma?”), sem a paradigmaváltás detektálhatóságának kritériumaira. Ugyanakkor nyilván fontos és érdekes kérdések merülnek fel ezekkel a problémákkal kapcsolatban a tanulmány témájával összefüggésben is. Például a mobilitáskutatás mottóban is említett paradigmáival kapcsolatban feltehető a kérdés, hogy vajon összevethetők-e ezek a paradigmák hatékonysági alapon, van-e bennük referenciája az „igaz” fogalmának? Eldönthető-e például empirikus alapon (lásd Saunders, 1997, Savage, Egerton 1997), hogy a Blau-Duncan féle individuális vagy a Goldthorpe-féle strukturális faktorok fontosabb tényezői-e a mobilitásnak? Vagy ugyanazon adatok különböző interpretációival állunk szemben?

Végül a tanulmány felépítéséről. A tanulmány következő részeiben kifejtem és példákkal alátámasztva bemutatom az első fő állítást, azaz, hogy „a módszertan megválasztásának társadalmi motívumai is vannak”: a módszertan megválasztásának társadalmi motívumaira mutatók példát tudományometriai eszközök felhasználásával, majd e motívumok közé a tudományterületek érintkezését, s a történelmi tradíciókat is besorolom, szociológiatörténeti példákon keresztül. Ezt követően a második fő állítást („a módszertan visszahat a kutatási paradigmára”) alátámasztó példákat mutatok. A tanulmányt végül az eredmények összefoglalásával és a hétköznapi kutatói praxisnak szóló mondanivalóval zárom.

4. Szerszám a szerszámosládából?

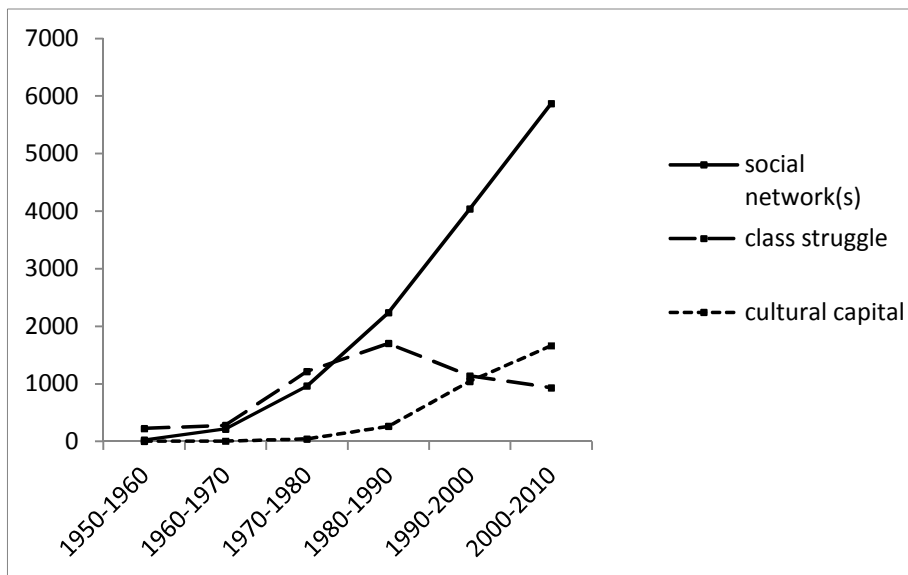
Ha a módszerek kiválasztása tökéletesen racionális módon történne, akkor kizárólag az adott problémára való érvényesség döntene. Valahogy úgy, mint amikor egy szerszámosládából mindig az adott problémára leginkább érvényes eszközt húzzuk elő. Ha ez így lenne, akkor a módszerek elterjedtségét azok (1) használati egyszerűsége nem befolyásolhatná, és (2) tudományterület-specifikusság sem volna jelen. Hasonlóan ellentmondana e szerszámosláda-jellegnek az, ha (3) módszertani divatok lennének kimutathatók. Ebben a fejezetben e három szempont alapján mutatok példákat arra, hogy a módszer megválasztása nem szerszámosláda-jellegű, tehát nem feltétlenül racionális módon történik.

A tudományos paradigmákat legegyszerűbben (és persze kissé leegyszerűsítve) tudományometriai úton közelíthetjük meg. Kulcsszavak vagy hivatkozások időbeli, tudományterületi, földrajzi elterjedtsége alapján empirikusan jól megközelíthető és megérthető a tudományos nézetek diffúziója, interdiszciplináris hatása, közösségképző ereje. Az alábbiakban a JSTOR publikációs adatbázisát felhasználva különböző szakkifejezések előfordulási gyakoriságának időbeli változását vizsgálom, ezzel a mutatóval az adott fogalmaknak a tudományos közösség számára jelentkező relevanciáját, használati értékét mérem.

A kifejezések előfordulását a cikkekben bárhol megengedtem, nem szorítottam meg a kulcsszavakra, az absztraktra vagy a cikk címére, mert az utóbbi megoldással csak azokat a

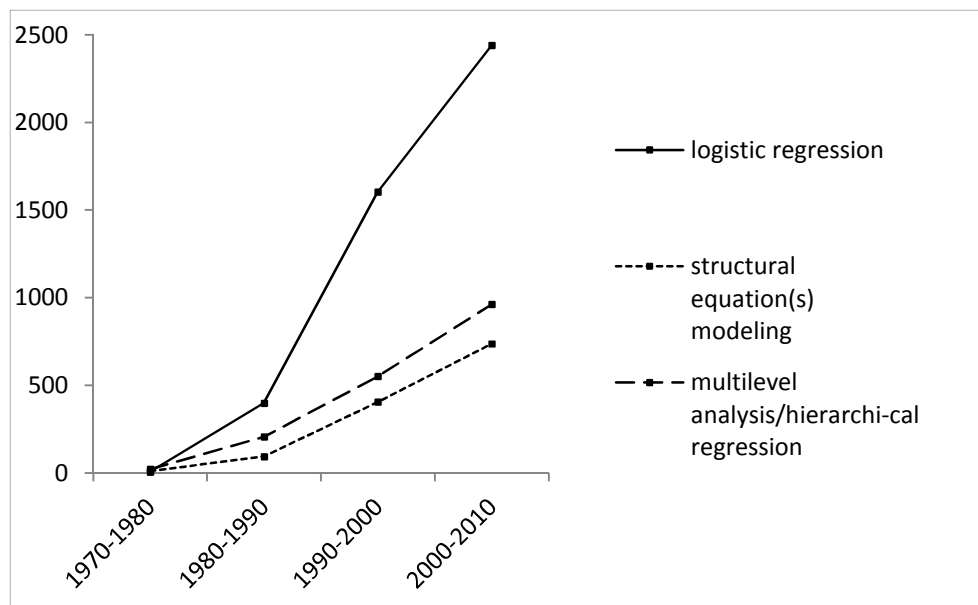
cikkeket értem volna el, melyek az adott szociológiai fogalomra fókuszálnak, azzal kapcsolatban adnak új eredményeket, márpedig egy fogalom kanonizálódása folyamán ezt egyre kevésbé várhatjuk. A témák terjedését/visszaszorulását az adott kifejezést tartalmazó publikációk számának változásával mértem. Alternatívaként felmerülhet az adott típusú cikkeknek a szociológiai cikkekben belüli aránya, de az arányok mindkét alább vizsgált esetben a nyers számokkal teljesen megegyező trendet mutatnak.

Az 1. ábra a *social network(s)*, *class struggle* ill. *cultural capital* kifejezéseket tartalmazó, szociológiai folyóiratokban megjelent publikációk számának változását mutatja az utóbbi hatvan évben, tízéves intervallumokkal. A *class struggle* népszerűsége a '80-as évek óta, a marxi inspirációjú társadalomelméletek visszaszorulásával csökken. A *social network* és a *cultural capital* verseny nem teljesen fair, hiszen az utóbbi fogalmat Bourdieu csak a '70-es években vezette be, de lemaradásának oka nem csak ez lehet. A *social networks* utóbbi évtizedekben tapasztalható hatalmas népszerűség-növekedését nyilván a (digitális kommunikációs csatornákra épülő) hálózati társadalom megjelenése, tehát egy új tárgy, egy újonnan jelentkező igény is generálja.



1. ábra Három téma szociológiai publikációbeli előfordulásának gyakorisága a JSTOR alapján

A társadalmi hálózatok kutatását tehát a kutatás tárgyában, a társadalomban bekövetkező változások is magyarázzák, de a kutatási témák születésének és kihalásának nem csak efféle objektív igények az okai. A tudomány nem vizsgálható függetlenül a tudományt művelő közösségtől, ami esetünkben annak a szempontnak a bevonását jelenti, hogy a vizsgálatra kijelölt problémák körét a közös paradigmát képviselő tudományos közösség határozza meg. Részben ez, a kutatási témák egyezményessé válása magyarázza tehát a fenti ábrán látható növekedő dinamikát.



2. ábra Három módszer tudományos publikációbeli előfordulásának gyakorisága a JSTOR alapján

Ugyanez igaz azonban nem csak a kutatási témákra, hanem a felhasznált módszerekre is: a tudományos közösség által hordozott paradigma jelöli ki a kutatási módszereket. Nézzük most az elterjedési trendet három módszerre vonatkozóan (2. ábra). A *logistic regression*, a *structural equation(s) modeling* és a *multilevel analysis/hierarchical regression*⁴ kifejezések előfordulási gyakoriságát vizsgáltam, s mivel a módszerek szociológiai ismertsége érdekelt, a keresést ismét a szociológiai folyóiratokban megjelent publikációkra szorítottam meg.

Akárcsak a szociológiai fogalmak esetén, a módszereket tekintve is növekvő trendet figyelhetünk meg, de szembeszökő különbségekkel. A logisztikus regresszió módszerét David Cox 1970-ben publikálta elsőként, a strukturális egyenletek modellezése (SEM) is a '70-es években nyerte el Jöreskog és mások által mai tartalmát, a többszintű elemzést is ezekben az években használták először oktatásszociológiai modellekben. Tehát nem indulási időpontjuk eltérése okozza, hogy a logisztikus regresszió a vizsgált időszak egészét tekintve a legnépszerűbb módszer. Ennek oka vélhetően általánosabb problémákra való alkalmazhatósága (a SEM-et elsősorban látens változók vizsgálatára használják, a többszintű elemzést individuális és kontextuális faktorok elkülönítésére). De a logisztikus regresszió elterjedtségében egyszerűbb használatának és könnyebb interpretálhatóságának is szerepe lehet – a módszer néhány óra alatt elsajátítható, míg a másik két technikát önálló féléves kurzusokon oktatják. “Nobody really understands SEM.”- idézi a SEM használhatóságára vonatkozó internetes kutatásuk egyik véleményét Nachtigall et al (2003), és Steiger (2001) is mint legkevésbé sem triviális, sok használati hibalehetőséget rejtő technikát jellemzi a SEM-et. Úgy tűnik tehát, hogy a módszerek elterjedtségét azok **(1) használati egyszerűsége** is befolyásolhatja. Ide, a használati egyszerűséghez tartozik a számítástechnikai kapacitások, a kivitelezhetőség és futásidő problémája is – például a többszintű elemzés maximum likelihood becslései csak az 1970-es év környékén végbement számítástechnikai áttörés révén váltak elérhetővé. Mivel a keresést 1970-től indítottam, ez a szempont nem korlátozza következtetésem érvényességét. Ám talán érdemes megjegyezni, hogy a futásidő sem az adott módszer érvényességének kérdését érinti, hanem a számítástechnikai kapacitás fejlettségi szintjén keresztül szintén a kutatás társadalomba ágyazottságára utal. A tanulmány másik részében (Elmélet és módszertan: hatás mellett visszahatás) a számítástechnika

⁴ A két elnevezés ugyanazt a módszert fedi, előbbi inkább Európában, utóbbi az USA-ban elterjedt.

evolúciójának az elméletre (pontosabban a tudományos igazolás kritériumaira) történő visszahatása kapcsán még visszatérek erre a kérdésre.

A SEM, mint említettük, látens (elsősorban pszichoszociális) faktorok vizsgálatára alkalmas, tehát a módszer logikájában hordoz némi tudományterület-specifikusságot. Máskor ez a specificitás indokolatlannak tűnik. Vegyük például a modellilleszkedés jellemzésére használható BIC (Bayesian Information Criterion) és AIC (Akaike Information Criterion) mutatók közötti választás kérdését. Ezek a mutatók információelméleti alapúak, nem használnak klasszikus hipotézisteszteket, és viszonylag frissek: az AIC-et Hirotugo Akaike, japán statisztikus vezette be 1973-ban, a BIC-et Adrian Raftery amerikai szociológus javasolta alkalmazásra 1986-ban. Nagyon eltérő a két mutató tudományterületenkénti ismertsége, s ez az eltérés nem vezethető le a két mutató matematikai tulajdonságainak egyébként meglévő eltéréseiből (Weakliem, 2004). A JSTOR adatbázisa szerint a BIC inkább a szociológiai, az AIC a közgazdaságtudományi cikkekben használatos (szövegbeli előfordulások száma: szociológia 807 vs. 422, közgazdaságtudomány 977 vs. 2132), biztosan nem függetlenül attól, hogy a BIC-et szociológus vezette be. Ha a módszerek megválasztása csupán érvényességükön alapulna, a BIC-AIC esetén megfigyelt **(2) tudományterület-specifikusság** nem volna jelen.

Hasonlóan ellentmond a szerszámoszláda-jellegnek az, hogy **(3) vannak „módszertani divatok”**, melyek tudományos interszociális kapcsolatokon és intézményesült csatornákon (folyóiratok stb.) keresztül terjednek. Ezt, a tudományos nézetek és módszerek diffúzióját magyarázó jól ismert tény illusztrálja az a megfigyelés, hogy a fenti kereső-kifejezések idősorainak kezdeti szakaszán a módszerek hálózatokon keresztül történő terjedésére utaló exponenciális növekedést, vagyis járványszerű terjedést láttunk. A divatoknak a folyóiratok bírálati mechanizmusán keresztül érvényesülő hatását emeli ki Nachtigall et al. (2003), akik szubjektív tapasztalataik alapján úgy látják, hogy sok esetben a SEM pszichológiában megfigyelhető divatossága miatt alkalmazzák a szerzők a technikát, csupán azzal a céllal, hogy növeljék tanulmányaik közlési valószínűségét.

5. Tudományterületek érintkezése

Az előző fejezetben, tudományometriai eszközökkel, három szempontra koncentrálnak mutatam a módszer megválasztásának racionális, szerszámoszláda-jellegét cáfoló példákat. A szerszámoszláda ellen szól az is, hogy **(4) tudományterületek érintkezésénél módszertani változások is beállnak**. Például Bernert (1983) a kauzalitás fogalmának az amerikai szociológia története során megfigyelhető változásait vizsgálva felhívja a figyelmet arra, hogy mindig a szociológia perifériáiról jövő aktorok és más tudományokból jövő diffúziók okozták az oksággal kapcsolatos szemléletváltozást, s ezek a változások lényegileg változtatták meg a módszereket is. Két nagy fordulatot azonosít. Az első, a XX. század elején bekövetkezett fordulat a Karl Pearson körül intézményesedett brit statisztikai iskola hatásaként fogható meg. A fordulat az okság fogalmának teljes elutasításával járt, melyet újonnan definiált kapcsolati mutatókkal helyettesítettek, survey módszereket használva. A másik fordulat a II. világháború után következett be, a biometrika, statisztika, pszichometria, ökonometria, oktatástatisztika felől érkező hatásokra. E generáció képviselői ismét életre keltették az okokat és okozatokat, statisztikai modellek (faktorelemzés, útelemzés) segítségével.

Mindezzel összecseng a statisztikus fizika hatása az utóbbi évtizedekben létrejött sajátos szociológiai kutatási módszertanokra. A legjellemzőbb változás a statisztika kihátrálása a szociológia korábban általa uralt módszertani mezejéről. Ennek oka egyrészt az, hogy sok esetben teljeskörűek az adatok, ezért nincsen szükség valószínűségi modellre. Egy másik ok a szimulációk használatában keresendő. A statisztikus fizikában ez a módszer biztosítja a

makroszinten megfigyelt törvényszerűségek és a mikroszinten történő interakciós mechanizmusok közötti összekötő kapcsot. A szociológiai, elsősorban analitikus szociológiai alkalmazásokban ágens alapú szimulációként jelent meg a módszer. Logikája lényegileg tér el a statisztikai modellek megszokott tesztelésétől. A mikrocelekvés paramétereinek valamely beállítása mellett szimulálják a populáció viselkedését, megfigyelik, ez milyen makromintázatot eredményezne, majd ezt vetik össze a valós mintázattal. Ám ez az eljárás nem teszteli a mikrocelekvés modellben alkalmazott módját. Barabási-Albert László ismert tételét véve példaként: ha egy hálózat önszerveződését a Máté-effektussal, mint mechanizmussal irányítjuk a szimulációban, akkor a létrejövő hálózat fokszámeloszlása – sok valós hálózathoz hasonlóan – skálafüggetlen lesz. Ez nem bizonyítéka ugyanakkor a Máté-effektusnak, viszont megerősíti azt, hogy egy egyszerű és ráadásul plauzibilis feltétellel skálafüggetlen hálózathoz juthatunk.⁵ Az eset jól példázza a tudományterületek érintkezési pontjain beálló alapvető jelentőségű módszertani változásokat.

6. Történelmi tradíciók

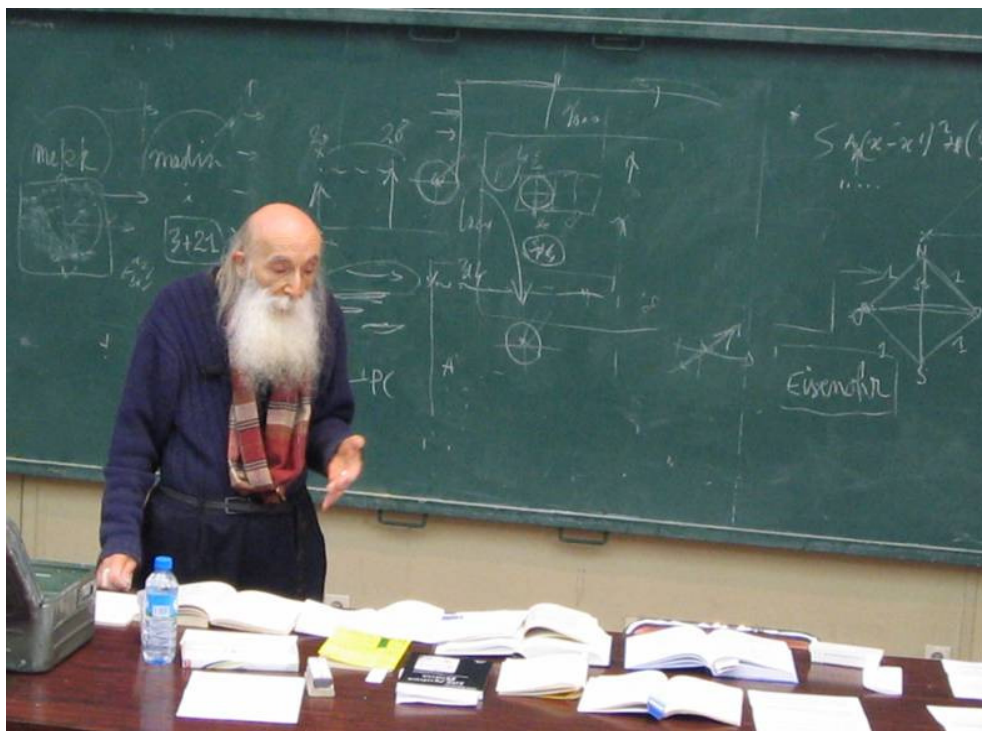
Ugyancsak a szerszámosládával szembeni érv az, hogy **(5) az eszközválasztásra nyilvánvaló módon hatnak történelmi tradíciók is.** Holmes (2007) például a többváltozós elemzések „francia módjáról” beszél. A francia statisztikusok, köztük Jean-Paul Benzécri a '60-as '70-es években a valószínűségi absztrakciókat praktikus szempontból haszontalannak ítélték, helyette adataik vizuálisan jól interpretálható geometriai reprezentációján alapuló elemzést ajánlották. Ez az angolszász statisztikus iskolától nagyon eltérő szemlélet. Maga a geometriai/vizuális gondolkodás Descartes óta kimondottan hangsúlyos a francia matematikai szemléletben, ami kihat a kutatásra (analízis, topológia⁶) és a matematikaoktatásra. Úgy tűnik tehát, hogy a módszerek megválasztásában tágabb gondolkodási hagyományok is közrejátszhatnak.

Benzécri és köre adatelemzésnek (*l'analyse des données*) nevezte az új megközelítést, és olyannyira önálló paradigmának tartották, hogy Benzécri tudománytörténeti-filozófiai kontextusát is fontosnak tartotta önálló monográfiaként megírni (Benzécri, 1982). Ebből a kép-központú nézőpontból fejlesztette tovább és tette népszerűvé a '60-as években Franciaországban Jean-Paul Benzécri és iskolája a korrespondencia-analízist. Ez a módszer angolszász nyelvterületen sokkal kevésbé népszerű: a google találati listája a *logistic regression* vs. *réggression logistique* keresőszavakra az angol kifejezés javára 30-szoros, míg a *correspondence analysis* vs. *L'Analyse des Correspondances* keresőszavakra csak háromszoros különbséget mutat. A magyar korrespondencia(-)analízis/elemzés keresőszó csak 230 találatot ad.⁷

⁵ Az ágens alapú szimuláció evolúciós biológiában alkalmazott változatának hasonló értelmezési problémáira hívja fel a figyelmet Huneman (2011). Ahogy írja: ha a madarak idegrendszerének három speciális sajátosságát feltéve a szimuláció a valóshoz hasonló repülési mintázatot ad, az nem e három sajátosságra nyújt bizonyítékot, hanem csak arra, hogy nem szükséges központi irányítást feltételezni az összehangolt repülés létrejöttéhez.

⁶ „Van a francia és van az orosz topológia. Lehet úgy is művelni a topológiát, mint az oroszok, de nem érdemes” – vezette be a matematikai szépséget mindig fontosnak tartó analízistanárom, Czách László az ELTE TTK-n előadását.

⁷ A módszer Magyarországon meglepő ismertségnek örvend véleménykutatói körökben, annak ellenére, hogy általában nem része a társadalomtudományi képzésnek. A módszer diffúzióját talán az Ipsos francia cég segítette elő magyar leányvállalatán keresztül.



Jean-Paul Benzécri⁸

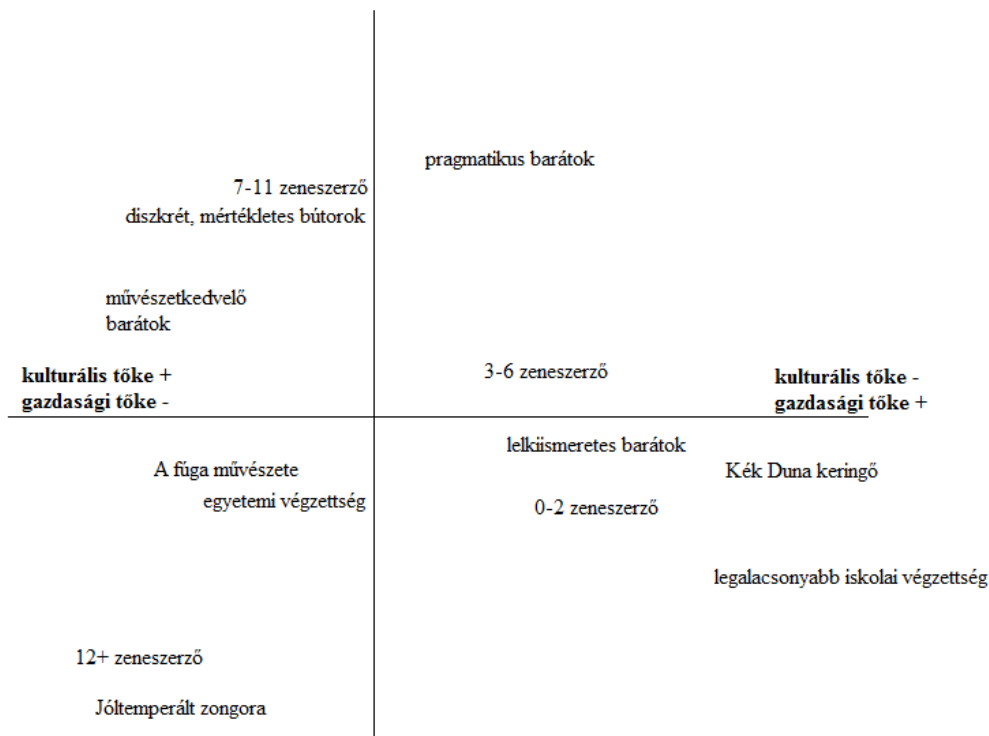
A korrespondenciaanalízis a legtöbb statisztikai módszerrel szemben nem numerikus, hanem kvalitatív változók közötti kapcsolat elemzésére alkalmas. Kreatív és gondos használatának legismertebb szociológiai példája Bourdieu *La Distinction* (1979) c. könyve.⁹ Ez az empirikus vizsgálatokra épülő ízlésszociológiai munka annyira tipikus alkalmazása a korrespondencia-analízisnek, hogy azóta gyakran „*La Distinction*-paradigma”-ként is utalnak az eszköz ilyen használatára. Nem csak tudománytörténetileg, hanem témánk, a módszertani paradigmák szempontjából is fontos megemlíteni, hogy Bourdieu Benzécrihez és köréhez kapcsolódva ismerte meg a módszert, és azt is, hogy Bourdieu és Benzécri már az École Normale Supérieure-n ismerték egymást, sőt Benzécri szerint Bourdieu haláláig leveleztek, tudományos és baráti kapcsolatban voltak.

A korrespondencia-analízis lényege, hogy tulajdonságok együttjárását az őket reprezentáló síkpontok geometriai közelségébe fordítja. Alább a könyv egyik ábrájának, egy korrespondencia-elemzési outputnak illusztratív céllal rekonstruált változatát láthatjuk. Az ábra segítségével a fogyasztási és életmód-preferenciák jellemző halmazai különíthetők el. Leolvasható például, hogy a diszkrét, mértékletes bútorok preferálása a pragmatikus barátok kereséséhez közel helyezkedik el, míg az egyetemi végzettség a (Bourdieu szavaival) legtisztább esztétikai ízlésre valló Bach-darab, *A fűga művésze* kedvelése mellett van. Az ábra jobb alsó negyedében a legalacsonyabb iskolai végzettséggel rendelkező, lelkiismeretes barátokat preferáló és „másodvonalbeli” zeneműveket (*Kék Duna*) kedvelők csoportja található. De általánosabb strukturális következtetések is levonhatók a tengelyek mögött álló látens faktoroknak történő jelentéstulajdonítással: például az iskolai végzettség a síkon jobbról balra haladva növekszik, tehát a pontfelhő vízszintes tengelyének egyik fontos meghatározója a kulturális tőke lehet. A gazdasági helyzet, jövedelem indikátorainak az ábrán

⁸ Benzécri 2006 októberében, helyszín: Institut National Agronomique Paris-Grignon (fotó: Guiseppe Giordano-Univ. Salerne) Forrás: Modulad, N 35, 2006, INRIA.

⁹ Bourdieu már korábban, 1976-ban Az ízlés anatómiájában (*Anatomie du goût*) is használta ezt a módszert.

történő elhelyezése pedig megmutatná, hogy a gazdasági tőke éppen ellenkező irányban, balról jobbra nő, igaz, kevésbé lineárisan.



A La Distinction korrespondenciaelemzési outputjának illusztrációs céllal rekonstruált változata

Ez a térbeli reprezentációt kínáló statisztikai eszköz párhuzamba állítható Bourdieunek a társadalomra vonatkozó térbeli víziójával, ahogyan azt már mások is észrevételezték (pl. Lebaron 2009, Rouanet és társai 2000). Bourdieu számára a társadalom térben létezik, számára a társadalmi kapcsolatok és távolságok elsősorban térbeli kapcsolatok és távolságok. (Térben, szó szerint: térképen. Lásd például a *Les Règles de l'art* elemzését az *Érzelmek iskolájá*-ról: egy korabeli Párizs-térkép segítségével világítja meg, hogy Flaubert hőseinek lakóhelyei, költözései a szimbolikus társadalmi rendszerben történő elhelyezkedésnek, az abban való mozgásnak feleltethetők meg). Bourdieu komplex struktúrában gondolkodik, nem egyszerűen vertikálisan, egy-egy dimenzió mentén tagolja a társadalmat, amint azt például a lineáris regresszió igényelné/sugallná. Bourdieu maga is többször utal rá, hogy számára a regresszió nem megfelelő eszköz, mert túlságosan leegyszerűsít, és a különbségeket hangsúlyozza a hasonló csoportok keresése helyett. A bevett statisztikai eljárások (amik ebben az időben főként folytonos változókat használó többváltozós technikák, mint a lineáris regresszió) helyett ezért döntött a korrespondencia-analízis mellett. Így ír erről *A szociológus mestersége (Le Métier de Sociologue)* 1991-es német kiadásának előszavában:

„Gyakran használok korrespondenciaanalízist, mivel azt gondolom, hogy ez alapvetően egy olyan, relációkra épülő módszer, aminek a filozófiája tökéletesen kifejezi mindazt, ami véleményem szerint a társadalmi valóságot felépíti. Olyan eljárás, ami relációkban „gondolkodik”, ahogyan arra én is kísérletet teszek a mező fogalmának használatakor.”

7. Elmélet és módszertan: hatás mellett visszahatás

Eddig a statisztikai módszer megválasztásának paradigmaticus jellegére hoztam példákat. Mindegyik példa amellett szól, hogy a módszerek elterjedését, ezt az innovációs diffúziót az

egyes technikák használhatóságán túl társadalmi tényezők is befolyásolják: a terjedési csatornák minősége, a terjesztő ágensek és a tudományos közönség fogékonysága stb.

Az utolsó példa a történelmi tradíciókról, a francia statisztika vizuális jellegéről szól, azonban egy lépéssel tovább is vezethet. Láttuk, hogy maga Bourdieu erős párhuzamot érzett szociológiai rendszere és statisztikai módszere között. Elemzői is tételről tételre azonosítják a korrespondenciaanalízis fogalmait Bourdieu szociológiai fogalmaival (például a disztinkív jegy nem más, mint az a kategória egy látens fogalom, azaz tengely szempontjából, ami az origótól messze van stb.). Talán érdemes felvetni azt a kérdést, hogy maga a módszer nem gyakorol-e hatást nem csak az eredményekre, hanem a társadalomszemléletre, a használt fogalmakra, kérdésfeltevésekre is. Gondoljunk csak arra például, hogy a korabeli technika számára a háromdimenziós ábrázolás még nehézkes volt, csak a kétdimenziós síkábrázolás (mint amilyent a fenti ábrán láttunk) volt elérhető, még akkor is, ha a pontok struktúrája kettőnél több fontos dimenziót mutatott. Ez szükségszerű leegyszerűsítése a modellnek, és óhatatlanul egyszerűsíti az interpretációt, a társadalomképet is. Vagy gondoljunk arra, hogy a látens dimenziók fogalmát (ami a kulturális, gazdasági, társadalmi tőke fogalmához vezet el) a módszer a koordinátatengelyekkel maga implikálja, s ezek a tengelyek mint látens dimenziók fel sem merültek volna, ha Bourdieu például hálózatelemzési módszereket használt volna.

Sokan negatívumként értékelik, hogy a modellezés logikája szabja meg az elméleti keretet, s ezáltal kimondatlanul is a módszertan irányítja az elméletalkotást. Az amerikai szociológiában a II. világháború után elterjedt, survey-adatokra épülő, többváltozós statisztikai modelleket használó empirikus kutatási irányzatot az '50-es évek óta érik ilyen jellegű kritikák. A változó-szociológia (*variable-sociology*) kritikusai (például. Sorensen, 1998) szerint a legtöbb statisztikai elemzés a változók lineáris hatását tételezi fel, továbbá azt, hogy ezek a hatások additív módon összegződnek, anélkül, hogy elméletileg igazolnák ezeket a feltevéseket. Továbbá: a paradigma művelői eleve azokat a kutatási problémákat részesítik előnyben, melyek az elfogadott módszerekkel vizsgálhatók.

A módszereknek az elméletre való visszahatásának egy másik példája a 4. fejezetben már említett számítástechnikai kapacitásfejlődésnek, a futási idő lerövidülésének következményével kapcsolatos. Annak, hogy a korábban több órán át futó alkalmazások ma milliós számú végrehajthatók egymás után, komoly visszahatása van a tudományos elméletre, pontosabban annak igazolási kritériumaira nézve is. Például míg korábban egyetlen, mély szakmai háttérismeretet igénylő tudományos állítás felállítása és statisztikai tesztelése volt a klasszikus eljárás, addig ma a legjobb modellt kereső adatbányászati algoritmusok az összes lehetséges modellt képesek megvizsgálni és közülük (akár minden háttértudás nélkül) kiválasztani a statisztikailag legjobban illeszkedőt. Ezeket a *data-driven* eljárásokat több kritika is éri, éppen a valóságtól való távolságuk miatt. A matematikai jellegű kritikák jó példája Freedman (2010), a szociológiaiaké Sorensen (1998), utóbbi szerző a számítástechnikai kapacitások fejlődését egyenesen a kvantitatív kutatás minőségi hanyatlásával hozza összefüggésbe. Hasonlóan negatív a konklúziója Steigernek (2001), aki kitűnő cikkében a SEM felhasználóbarát szoftveres elérhetőségével hozza összefüggésbe, hogy a módszer használói sokszor alapvető hibákat vétének. A cikk a probléma tudomány-szociológiai elemzését is tartalmazza, említve például azt az érdekes megfigyelést, hogy a kiadók üzleti érdekei miatt a szoftver-központú, de matematikailag felületes könyvek kiadása a jellemző, vagy hogy a társadalomtudományi tanszékek a felhasználóbarát szoftverek elérhetősége miatt megváltak a statisztikus-kollégától.

Az utóbbi példában a számítástechnikai kapacitások bővülése, a következő példánkban az empirikus bázis megváltozása hat vissza az elméletre, sőt, teszi indokolhatóvá a teljes tudományos paradigma lecserélését, amint arra a „Big Data” egyes kutatói utalnak. A

hedonometer.org tudományos publikációi például egy olyan boldogságmérő rendszer működtetéséről számolnak be, ami angol nyelvű szöveges inputok (Twitter-szövegek, könnyűzenei számok, blogok) alapján méri a boldogság (időszakokra vagy földrajzi területekre átlagolt) szintjét. Egy szöveg boldogsága szavai pozitivitásának súlyozott átlagából jön. A szavakat előzőleg vizsgálati alanyokkal pontoztatták 1 és 9 között, ezért pozitívítás-szintjük kontextustól függetlenül adott. Pl. a *love* és *gold* magas átlagpontot kap, a *war*, *don't* és *no* alacsony. Zárójelben megjegyzem, hogy a publikációk csak elnagyoltan érintik az operacionalizálás problémáit. Emiatt nehéz az olyan eredmények értelmezése, mint hogy a könnyűzenei szövegek a '60-as évektől egyre boldogtalanabbak, vagy hogy a Twitter-használók hálózatában adott csúcs fokszáma és boldogsága között pozitív korreláció van.

A hedonométer példája is jól mutatja az új empirikus bázis jellegzetességeit: a Big Data a hagyományos survey-módszerek helyett digitálisan és automatikusan rögzített hatalmas adatbázisokkal dolgozik. Párezres keresztmetszeti minták helyett teljeskörű, real-time adatokkal. Szubjektív önbevalláson alapuló kérdésekre („Mennyire elégedett ön az életével általában?”) adott válaszok helyett bizonyos tevékenységeink digitális rögzítésének melléktermékéből (itt: tweetek) definiált mutatókkal (itt: szövegek átlagos pozitivitása). A hedonometer kutatói (a Vermont Egyetem matematikusai) expliciten is kifejtik (Dodds et al, 2011), hogy ezeknek a hatalmas adattömegeknek véleményük szerint első szinten csak a leírása, mintázatfeltárása történhet meg, a mélyebb magyarázat csak ezután következik. Az új adatforrást olyan fontos megkülönböztető jegynek tartják, ami új társadalomtudományi paradigmát nyit, szükségszerűen új módszerekkel:

„A big-data-társadalomtudomány kora kétségtelenül megkezdődött. Egy ilyen adatgazdag tudomány esetén először a leírás és mintázatkeresés jelenti a kihívást, s csak azután jön a magyarázat és kísérletezés ideje. Ahelyett, hogy rögtön hipotézisekkel állnánk elő, kénytelenek vagyunk sok időt és energiát fordítani a leírásra. Az adatszegény tudomány megközelítéseinek továbbra is megmarad az értéke, de most hatalmas új ablakok nyíltak a társadalmi és pszichológiai magatartásra, az ezeken való betekintéshez új eszközök szükségesek, melyek részben még kidolgozásra várnak.”

Utolsó példában ismét visszatérnék tanulmányom mottójára, mely a mobilitáskutatás eszközeinek paradigmatisztikus jellegére utalt. Kelley az idézetben az általuk használt statisztikai módszerekkel azonosítja a mobilitáskutatási paradigmákat, és összevetésükkor is a statisztikai módszerek előnyeire/hátrányaira hivatkozik. Bár nyilván a kutatási problémák változása is maga után vont a paradigmák (és a vizsgálatokra alkalmas eszközök) változását, Kelley, úgy tűnik, a fordított hatásirányt tartja elsődlegesnek: a változó módszerek határozzák meg a vizsgálatra alkalmas problémákat.

Ezeket a mobilitáskutatási paradigmákat alapvetően a társadalmi státusz mérésére használt változók típusa (folytonos vagy kategoriális) különbözteti meg. A Blau-Duncan paradigma a foglalkozási státuszt mérő folytonos Duncan-féle társadalmi-gazdasági indexet (SEI) használta. Az index eredetijét az amerikai National Office of Vital Statistics munkatársai hozták létre az '50-es években; azért, mert „olyan módszerre volt szükségük, mellyel a foglalkozási kategóriák apró részletei néhány, lehetőség szerint egyetlen kvantitássá redukálhatók, mely kvantitás azután korreláció mérésére használható”. (Hout, 2007, kiemelés tőlem). Az indexnek ezek a tulajdonságai tették lehetővé később a foglalkozás útelemzésben szerepeltetését, a Blau-Duncan modell megszületését, a mobilitási folyamat hatásainak elkülönítését és számszerűsítését, a direkt és indirekt hatás fogalmának megjelenését. Ettől kezdve az Egyesült Államokban a társadalmi helyzet operacionalizálása inkább a (folytonos) foglalkozási presztízzsel történik, míg Európában elsősorban (kategoriális) osztályokat használnak (a legismertebb Erikson, Goldthorpe és Portocarero után a tizenegy kategóriás

EGP osztályséma). Ez persze struktúraelméleti különbséget is jelez: az osztályok kevés, nem feltétlenül rangsorolható, több dimenzió mentén tagolódó, belül homogén csoportból állnak; míg a foglalkozási presztízs sok-sok, egyetlen dimenzió mentén rendezhető csoportot hoz létre. De a kiemelés Hout szövegéből - miszerint azért hozták létre az indexet, hogy korrelációt tudjanak számolni (!) - és talán Kelly írása is elképzelhetővé teszi az operacionalizálás módjának pusztán módszertani indíttatását.

A társadalomkutatási praxisban gyakran magunk is az aktuális módszer által megkívánt mérési szint szerint operacionalizáljuk változóinkat. Azt is meg lehet itt jegyezni, hogy (leginkább történeti okokból, a hagyományt követve) a társadalomtudományi képzés statisztika kurzusain a világ minden táján jelentősen nagyobb szerep jut a folytonos változókat használó többváltozós elemzéseknek, annak ellenére, hogy a kutatói gyakorlatban előforduló változók nagyobb része kategoriális. Ezért, a jobbára kategoriális típusú változók folytonos változókat igénylő módszerekkel történő elemzése miatt van szükség indexképzésre vagy a változók „felskálázására”. Márpedig a folytonos vagy kategoriális változók melletti döntésnek lényegi következményei vannak az adatokon végezhető elemzésre nézve. A folytonos megközelítés a majdnem mindig automatikus normalitás-feltétellel egyszerűsége és függvény-jellegű kapcsolatokra törekszik, míg a kategoriális megközelítés finomabb modellezést, nagyobb szabadságot tesz lehetővé (nem-függvény jellegű kapcsolatok, komplex interakciók). Az alkalmazást tekintve (lásd pl. Sorensen, 1998, Goldthorpe, 2005) pedig e matematikai szabadságpéldául a mobilitás tipikus csatornáinak és akadályainak megragadását teszi lehetővé. A folytonos megközelítés ugyanis kiátlagolja a különböző osztályokban megjelenő eltérő rekrutációs mintákat, a kategoriális megközelítés lehetőséget ad ezek osztályspecifikus megragadására. A kategoriális megközelítés ugyanakkor viszonylag kevés változót és kevés kategóriát képes kezelni (egyébként hamar üres cellákhoz jutunk). A döntés tehát a szociológiai koncepciót (például rangsorolható-e a társadalmi osztályok) és a szóba jöhető módszereken keresztül az elemzési keretet és az eredményeket is befolyásolja.

8. Összefoglalás

Tanulmányom első részében a kvantitatív kutatási paradigmáknak a módszertan megválasztására gyakorolt hatásáról írtam. Példákkal támasztottam alá, hogy a módszer kiválasztásánál több, a tudományos közösségből eredő társadalmi szempont is érvényesül, úgymint

- használati egyszerűség,
- tudományterület-specifikusság,
- tudományos kapcsolatokon keresztül terjedő „divatok”,
- tudományterületek érintkezése, vagy
- történelmi tradíciók.

Írásom utolsó fejezetében azt igyekeztem alátámasztani, hogy bizonyos esetekben a fordított hatásirány is elképzelhető: a módszerek is hatást gyakorolhatnak a kutatási paradigmára. Az itt felsorolt példákban az alkalmazott módszer hatást gyakorolhat

- a társadalomképre (egyetlen dimenzió mentén hierarchikusan rendezhető / több látens dimenzió strukturálta többdimenziós alakzat / néhány homogén osztály alkotja / adott relációk szervezte hálózat stb.),
- a használt fogalmakra (olyannyira, hogy néha e fogalmak a statisztikai modell elemeiként azonosíthatók),

- a kérdésfeltevésre (melyek az elfogadott módszerrel vizsgálható kérdések?),
- sőt magára az igazolási kritériumokra (*data-driven* eljárások) vagy az elméletre (Big Data) is.

A módszer megválasztásánál tehát az adott problémára való érvényesség mellett számos irracionális szempont is érvényesül. Másfelől, aktuálisan használt módszereink óhatatlanul behatárolják a tudományos kérdések és társadalomreprezentációk körét, formálják szemléletünket. Az intézményesült társadalomkutatás paradigmatis keretei között a kvantitatív módszerek adta objektivitás is korlátozott. Írásom az ezzel kapcsolatos önreflexiókhoz próbált hozzájárulni.

9. Hivatkozások

Benzécri, J. P. (1982): *Histoire et préhistoire de l'analyse des données*. Paris: Dunod.

Bernert, C. (1983): The career of causal analysis in American sociology. *British Journal of Sociology*, 24(2): 230-254.

Dodds, P. S., K. D. Harris, I. M. Kloumann, C. A. Bliss, C. M. Danforth (2011): Temporal Patterns of Happiness and Information in a Global-Scale Social Network: Hedonometrics and Twitter. *PLoS ONE*, 6, e26752.

Freedman, D. (2010): *Statistical Models and Causal Inference*. (Collier, D., Sekhon, J., Stark, J., szerk.). New York: Cambridge University Press.

Goldthorpe, J. H. (2005): Progress in Sociology: The Case of Social Mobility Research. In: S. Svallfors (ed.): *Analyzing Inequality: Life Chances and Social Mobility in Comparative Perspective*, pp. 56–82. Stanford: Stanford University Press.

Holmes, S.P. (2007): Multivariate Analysis: The French Way. *Probability and Statistics*, Volume 2, pp 219-233.

Hout, M. (2007): Otis Dudley Duncan's major contributions to the study of social stratification. *Research in social stratification and mobility*, 26:109-118.

Huneman, Philippe (2011): Computer sciences meet evolutionary biology: issues in gradualism. In: Torres J. L., Pombo O., Symons J., Rahman S. (szerk.): *Special sciences and the Unity of Science, Volume 24: Logics, epistemology and the unity of science*, pp.200-225. Dordrecht: Springer.

Kelley, J. (1990): The failure of a paradigm: Log-linear models of social mobility. In Clarke, Modgil and Modgil (szerk.): *John Goldthorpe: Consensus and Controversy*, London: Falmer Press. Pp. 319-346.

Lebaron, F. (2009): How Bourdieu „quantified” Bourdieu: the geometric modelling of data. In: Robson and Sanders (eds): *Quantifying theory: Pierre Bourdieu*. Springer.

Nachtigall, Ch., Kroehne, U., Funke, F., Steyer, R. (2003): (Why) Should We Use SEM? Pros and Cons of Structural Equation Modeling. *Methods of Psychological Research Online*, (8)2, pp. 1-22.

Németh Renáta (2013): „I just ran two million regressions”, avagy módszertani paradigmák a kvantitatív társadalomkutatásban. Konferencia-előadás. In: Bacsák, D., Krámer, L., Szabó, M. (szerk.): *Kulcskérdések a társadalomtudományban 2011-2012*, pp. 189-200. Budapest, ELTE.

Rouanet, H., Ackermann, W., Le Roux, B. (2000): The geometric analysis of questionnaires: The Lesson of Bourdieu's La Distinction. *Bulletin de Méthodologie Sociologique*, 65, 5-15.

Saunders, P. (1997) Social Mobility in Britain: an empirical evaluation of two competing explanations. *Sociology*, 31(2): 261-88.

Savage, M., Egerton, M. (1997) Social mobility, individual ability and the inheritance of class inequality, *Sociology*, 31(4):645-672.

Sorensen, A. B. (1998): Theoretical Mechanisms and the Empirical Study of Social Processes. In: Hedstrom, P. – Swedberg, R. (szerk.): *Social Mechanisms. An Analytical*

Steiger, J.H. (2001): Driving Fast In Reverse. The Relationship Between Software Development, Theory, and Education in Structural Equation Modeling. *Journal of the American Statistical Association*, (96) 453: 331–338.

Weakliem, D. L. (2004). Introduction to the special issue on model selection. *Sociological Methods & Research*, 33(2), 167-186.