

Skálák és statisztikák: a méréselméletről és történetéről*

Kehl Dániel,
a Pécsi Tudományegyetem
egyetemi tanársegédje
E-mail: kehd@ktk.pte.hu

A szerző *Stevens* [1946] nagyhatású cikke nyomán kialakult méréselméleti vitát mutatja be, melynek megoldására nem vállalkozik, de részletesen ismerteti a kifejtett álláspontokat és bőséges irodalomjegyzéket ad az érdeklődő Olvasónak. Hangsúlyozza, hogy a megfelelő módszertanok rendelkezésre állnak, a kutató felelőssége a megfelelő eljárás kiválasztása.

TÁRGYSZÓ:
Méréselmélet.

* A szerző ezúton mond köszönetet a tanulmány jelenlegi és korábbi verzióhoz fűzött megjegyzésekért és tanácsokért *Hunyadi Lászlónak, Szidarovszky Ferencnek, Vita Lászlónak, Vargha Andrásnak* és intézeti kollégáinak, valamint a Rosztoczy Alapítványnak anyagi támogatásáért. Minden fennmaradó hibáért természetesen a szerzőt terheli a felelősség.

Adatok, ezen belül is statisztikai alapadatok jellemzően számlálás, illetve mérés útján keletkeznek. A számlálás útján előállított adatok esetén is találkozhatunk gyakorlati problémákkal, jelen írásunkban azonban a mérés jellemzőivel foglalkozunk. Gondoljunk csak néhány példára: az infláció, a vásárlói attitűd, az életminőség vagy az értelmi képesség számszerűsítésének problémájára. Célunk elsősorban bemutatni a mérési skálák elméletének kialakulását, majd azt a tudományos vitát, amit az irodalom méréselméleti polémiaként ismer. Az Olvasó számára mindehhez bőséges nemzetközi irodalomra vonatkozó forrásanyaggal szolgálunk. Természetesen ma már nem minden bemutatott megállapítással értünk egyet, a tanulmányok főbb tételeinek kiemelését a vita folyamatának bemutatása miatt tartjuk szükségesnek. A történeti áttekintés mellett fontos felhívni a figyelmet a nemparaméteres módszerek tudományos területen történő általánosabb körű alkalmazására és jelentőségére a statisztika oktatásában.

1. A méréselméleti vita története

A mérési skálák napjainkban is alkalmazott típusai *Stanley Smith Stevens* ([1946], [1955]) a Harvard Egyetem pszichológus professzorának klasszikus, sokat hivatkozott tanulmányaihoz kötődnek. A Stevens vezette tudományos bizottság évekig foglalkozott a mérés problematikájával, céljuk annak a kérdésnek a megválaszolása volt, hogy mérhető-e az emberi érzékelés, és ha igen, milyen módon. A legfőbb és talán legfontosabb vita a körül alakult ki, hogy mit is nevezhetünk mérésnek. A bizottság tagjai markánsan eltérő véleményt alakítottak ki. Stevens szerint fontos felismerünk, hogy a mérésnek számos formája létezik, ennek megfelelően különböző mérési skálák definiálhatók, amelyek típusát egyaránt meghatározzák a mérés folyamán alkalmazott konkrét eljárások és a skála matematikai tulajdonságai. A mérési skálától függ, hogy az adott empirikus adatok esetén mely statisztikai módszerek, eljárások alkalmazhatók, és melyek nem. Ez a megállapítás volt Stevens tanulmányának legnagyobb visszhangot keltő kijelentése. Definíciója szerint a mérés nem más, mint számértékek hozzárendelése különböző objektumokhoz vagy eseményekhez, mégpedig meghatározott szabályok szerint. Amennyiben ez a megállapítás helytálló, a skálák problematikája visszavezethető a következőkre.

Meg kell határozunk:

- a számértékek hozzárendelésének szabályait;
- az eredményként előálló skálák matematikai tulajdonságait;
- az egyes mérési skálák esetén alkalmazható statisztikai műveletek, eljárások körét.

A skálák jellemzője, hogy bizonyos hasonlóság van a megfigyelt objektumok tulajdonságai és a számsorok között. Az objektumokat tekintve a következőket vizsgálhatjuk: az egyedek tulajdonságainak egyezőségét, jellemzőinek nagyságrendi sorrendjét, az adott különbségeket és azok egyezőségét, valamint az arányokat és azok egyezőségét. Ezen tulajdonságok leírására a (pozitív) valós számok tökéletesen megfelehetnek, azaz – Stevens szavaival élve – a számok a valós világ jelenségeinek megfelelő modelljét adhatják.

Az elérhető skála minősége természetesen függ a mérni kívánt jelenség jellemzőitől, és a mérés konkrét folyamatától, de a szerző szerint az eredmény a táblázatban szereplő – az alapszintű statisztika tankönyvekből jól ismert – skálák egyike lesz.

Mérési skálák és tulajdonságaik

Skála	Alapvető művelet	Matematikai csoport tulajdonsága	Megengedhető statisztikai műveletek
Nominális	Egyenlőség meghatározása	Permutációs csoport $x' = f(x)$ ahol f tetszőleges, kölcsönösen egyértelmű hozzárendelés	Esetek száma Módusz
Ordinális	Sorrendiség meghatározása	Isotonikus csoport $x' = f(x)$ ahol f tetszőleges, monoton növekvő függvény	Medián Percentilisek
Intervallum	Intervallumok/különbségek egyezőségének vizsgálata	Általános lineáris csoport $x' = ax + b, a > 0$	Számítási átlag Szórás Rangkorreláció Szorzat momentum korreláció
Arány	Hányadosok egyezőségének vizsgálata	Hasonlósági csoport $x' = ax, a > 0$	Mértani átlag Harmonikus átlag Szórás

Forrás: Stevens ([1946] 678. old., [1955] 113. old.).

A skálákon megengedett, azaz elvégezhető műveleteket a táblázat utolsó oszlopában soroltuk fel, mely lista kumulatíván értelmezendő: az alacsonyabb rendű skálák

megengedett műveletei a magasabb rendűeken is elvégezhetők. A „Matematikai csoport tulajdonsága” oszlopban azon matematikai transzformációkat soroltuk fel, melyek nem módosítják a skálátípust. A műveletek megengedhetőségének feltétele az invariancia, melynek jelentését Anderson mutatja be szemléletesen: „Amennyiben adott változóértékekből számítunk mutatót, majd transzformáljuk azt, azonos eredményt kell kapnunk, mintha az egyedi értékeket transzformáltuk volna, és így határoztuk volna meg a mutató értékét.” (Anderson [1961] 309. old.).

A skálák számának növelésére, bővítésére több kísérlet is született. Az intervallumskála mellett (amelyet egyenlő intervallumok skálájának is neveznek) a nem egyenlő intervallumok skáláit is megkülönböztetik esetenként az irodalomban, mint például a logaritmikus skálát, ahol tízes alap esetén minden intervallum pontosan a tízszere az öt megelőzőnek. Az ilyen skálákat azért nem tekintjük külön típusnak, mert megfelelő matematikai művelettel hagyományos intervallumskálává transzformálhatók. A megengedhető statisztikai műveletek ebben az esetben a hatványtranszformációk. A másik kiterjesztés az ún. abszolút skála, melyet Stevens is megemlít, és kardinális skálán mért számnak nevezi. Kizárólag a helybenhagyó művelet megengedett. Az elképzelhető skálátípusok matematikai meghatározásával a modern méréselmélet foglalkozik, alapvető megállapításuk, hogy Stevens besorolása többé-kevésbé teljes, más jelentős struktúrák bizonyíthatóan nem léteznek.

A skálák megkülönböztetése mellett Stevens cikkének legnagyobb jelentősége a megengedhető statisztikai műveletek rögzítésében van. Kategorikusan kijelenti, hogy a kutatók által gyakran alkalmazott ordinális változók esetén „a hagyományos, átlagokon és szórásokon alapuló eljárásokat nem szabadna használni, hisz azok többet tételeznek fel, mint csupán az adatok relatív rangsorának ismeretét”. Következő mondataiban mindenesetre már megengedőbb a szerző: „az eljárások illegális alkalmazása egyetlen dolog miatt bocsátható meg: sok esetben a vizsgálatok gyümölcsöző eredményekre vezetnek” (Stevens [1946] 679. old.). Hangsúlyozza, hogy az ordinális skálákon mért változók esetén számított statisztikákat óvatosan kell kezelni, különösen körültekintően a következtetések levonásakor. Ennek ellenére ordinális skálán mért ismérvekből számított átlagokkal azóta is találkozunk, akár a mindennapi életben, akár a tudományos kutatások területén. Ugyan tudjuk, hogy az iskolai osztályzatok nem mondanak többet az ordinalitásnál, az ösztöndíjak például mégis tanulmányi átlagtól, sőt, kreditpontokkal súlyozott átlagoktól függnak.

Stevens nagy hatású cikkei, és az ezekből levonható tanulságok követőkre találtak, melyek a társadalomtudományi módszertanokkal foglalkozó tankönyvekben hamarosan meg is jelentek: *Sidney Siegel* [1956] határozottan elutasította a paraméteres eljárások alkalmazását ordinális skálák esetén, azokat csak legalább intervallum erősségű skáláknál tartotta elfogadhatónak. A táblázatban bemutatottak alapján egy intervallumskála lineáris transzformációja megengedett, hasonlóan az ordinális skála nemlineáris (de monoton) transzformációjához. Az intervallumskála nemlineáris transzformációja

ellenben nem, hiszen például az átlag nem viselkedne az elvárások szerint, azaz nem invariáns. Ez vezet el a statisztikai eszköztár két részre bontásához: a paraméteres és nemparaméteres eljárások megkülönböztetéséhez. A paraméteres eljárások legalább egy sokasági érték, azaz paraméter becsléséből indulnak ki, méghozzá leggyakrabban normális eloszlású sokaságból származó minták esetére. Mindez legalább intervallum erősségű skálát követel meg a Stevens tanait követő konzervatívok szerint. A nemparaméteres eljárások nem követelik meg sokasági paraméterek becslését, nem teszik fel az egység állandóságát a skála teljes értelmezési tartományán, nincsenek olyan erős előzetes feltevések a sokasági eloszlását illetően sem. A két megközelítés hipotézisrendszerei nem mindenkor feleltethetők meg egymásnak teljes mértékben.

A Stevens elméletét támogató kutatók mellett mások határozottan támadták álláspontját, ellenezték az abból levonható következtetéseket. Szélsőséges véleményét fejtette ki *Lord*, aki rövid, nem egészen két oldalas, szarkasztikus cikkében futballmezek számozásáról írt, azt bizonygatva, hogy akár nominális változókon is eredményesen alkalmazhatók paraméteres statisztikai eljárások. A tanulmány egy képzeletbeli futballcsapat mekszámairól és egy professzorról szól, akinek azt kellett kiderítenie, hogy a másodévesek tényleg magasabb számú mezekben szerepelnek-e, mint az elsőévesek. Véleménye szerint az elvégezhető statisztikai műveleteket egyáltalán nem befolyásolják a mérési skálák, „a számok nem emlékeznek rá, honnan jöttek, származásuktól függően mindig ugyanúgy viselkednek” (*Lord* [1953] 751. old.).

Lord cikkére reagálva rövid választ adott *Behan* és *Behan*, melyben azt hangsúlyozták, hogy „amennyiben egy empirikus folyamat nem tartalmaz olyan műveletet, mely matematikai műveletnek megfeleltethető, akkor az adott művelet nem használható az eredményül kapott számokon” (*Behan–Behan* [1954] 263. old.). Természetesen, ha „valaki ismeri a képleteket, tud írni, vagy van számológépe, az bármilyen számok átlagát, szórását stb. kiszámíthatja. De miután végzett, csupán a jelekről tud valamit megállapítani, nem a futballistákról vagy azok tulajdonságaikról” (*Behan–Behan* [1954] 262. old.). Az eredeti, *Lord* által írt cikk jelentőségét mutatja, hogy az még napjainkban is foglalkoztatja a tudományos közösséget. Például *Scholten* és *Borsboom* [2009] szerint *Lord* példája helyesen értelmezve épp alátámasztja, tökéletesen illusztrálja Stevens elméletét, nem pedig ellentmond annak.

A Stevens tételeit egyértelműen tagadó ultraliberális szemlélet mellett számos tanulmány született a paraméteres próbák robusztusságával kapcsolatban is. Ezek a tanulmányok elsősorban azt a vitát voltak hivatottak eldönteni, mely a paraméteres (például *Anderson* [1961], *Baker–Hardyck–Petrinovich* [1966], *Labovitz* [1967]) és a nemparaméteres próbákat előnyben részesítő kutatók (például *Siegel* [1956], *Townsend–Ashby* [1984]) között alakult ki. A központi kérdés emellett az ordinális és intervallumskálák megkülönböztetése volt. A paraméteres és nemparaméteres eljárások alkalmazhatósága, illetve a mérési skálák hatása a kérdéssel kapcsolatban a kutatókat azóta is megosztja. Megtalálhatók szélsőséges vélemények is, miszerint a nem-

paraméteres eljárások szinte teljesen feleslegesek, mert majdnem minden esetben alkalmazhatók a jól ismert paraméteres megfelelőik, azok robusztussága miatt, ráadásul – érvelnek egyes kutatók – a paraméteres próbák statisztikai ereje jóval nagyobb.

A két tábor közötti vita alapvetően a következő témákban zajlott (*Gardner* [1975] 45. old.):

1. Adott mérési eszköztár mellett hogyan határozható meg a skála erőssége, típusa?
2. A skála típusának meghatározása után az mennyiben határozza meg a vizsgálat során alkalmazható statisztikai eljárásokat?

Az első kérdés megválaszolása alapvetően méréselméleti feladat (mivel a legtöbb pszichológiai kutatás a szóban forgó ordinális/intervallumskálán mért változókkal dolgozik, ezért a pszichometria rendelkezik komoly eredményekkel, kutatásokkal a témával kapcsolatban). Néhány kutató véleménye szerint (például *Gaito* [1980]) az intervallumskála elégséges bizonyítéka az is, ha a mérés eredményei normális eloszlást követnek. Ez a vélekedés valószínűleg a következő gondolatmenet – hibás – megfordításán alapul: sok arányskálán (azaz az intervallumskála feltételeit is kielégítő) mérhető ismérv esetén a különböző statisztikai mutatók normális eloszlást követnek. A következtetés visszafelé történő alkalmazása azonban csupán egy gyakran elkövetett logikai fallácia. *Thomas* [1982] szintén a következtetés hibás voltára hívja fel a figyelmet. Annak bizonyítása, hogy a pszichológiában (és egyéb társadalomtudományokban) alkalmazott skálák, mérési módszerek intervallum erősségű eredményeket szolgáltatnak, még napjainkban is várat magára.

A második kérdés megválaszolása, azaz a statisztikai eljárások alkalmazhatósága tisztán statisztikai probléma, amelyhez ismerni kell:

- az alkalmazandó statisztikai eljárás feltételrendszerét;
- a feltételrendszer teljesülésének jelentőségét, azaz a teszt robusztusságát;
- a feltételek teljesülésének mértékét az adott empirikus vizsgálat esetében.

A robusztusság és az előfeltételek vizsgálatához a már említett, a szakirodalomban ultrakonzervatívnak tartott *Siegel* munkássága szolgáltatott alapot, aki négy pontban gyűjtötte össze a paraméteres eljárások (főként a *t*-próba, ANOVA) alkalmazhatóságának feltételeit (*Siegel* [1956]-t idézi *Gardner* [1975]):

1. A megfigyelések függetlenek, egy adott egyed mintába kerülése nincs hatással sem a többi elem mintába kerülésének valószínűségére, sem a rájuk vonatkozó mérés eredményére.

2. Normális eloszlású sokaságokból kell származniuk a megfigyeléseknek.
3. A sokaságoknak azonos varianciával kell rendelkezniük.
4. A szóban forgó változóknak legalább intervallumskálán mértnek kell lenniük.

Az első pont tartalma egyaránt érvényes feltétel a paraméteres és a nemparaméteres eljárások esetében is. A második és harmadik pont a robusztusság kérdését veti fel a paraméteres próbákkal kapcsolatban, ami alatt azt értjük, hogy mennyiben maradnak érvényesek az adott statisztikai próba következtetései, ha egy vagy több kiinduló feltétel nem teljesül.

A t -próba tekintetében már korán születtek tanulmányok (például *Gayen* [1949], *Boneau* [1960]), melyek szerint jelentős torzítást sem a sokaság normálistól eltérő eloszlása, sem a varianciák különbözősége nem okozott. Azonban a két feltétel együttes megsértése esetenként erős hatást mutatott. Később a t -próba és a Wilcoxon-féle teszt összehasonlítását *Blair* és *Higgins* [1980, 1985] több tanulmányukban végezték el, és azt állapították meg, hogy a statisztikai próba ereje szempontjából nem bizonyítható a t -próba előnye a Wilcoxon-próbával szemben. A *Statisztikai Szemle* hasábjain *Vargha* [2003] foglalkozott az egymintás t -próbával, szimulációk segítségével bizonyítva, hogy az alapeloszlás csúcsosságától és ferdeségétől is függ a robusztusság.

Az 1940-es évek elejétől kezdődően az F -eloszlással kapcsolatban folytatott vizsgálatok (*Godard-Lindquist* [1940], *Box* [1953], *Glass* [1972]) hasonló eredményekre vezettek. A témakör jelentőségét mutatja, hogy neves szerzők által írt tanulmányokból tematikus lapszámok is jelentek meg (*Eisenhart* [1947], *Cochran* [1947], *Bartlett* [1947]). *Gaito* [1972] véleménye szerint összességében elmondható, hogy az eltérő mintaelemszám és variancia negatív hatása akkor a legnagyobb, amikor két csoportot hasonlítunk össze. A csoportok számának növekedésével ez a hatás csökken. A vizsgálatok tehát azt mutatják, hogy a paraméteres eljárások mégis csak robusztusak, azonban ezt a tulajdonságot nem szabad minden határon túl kihasználni. Különösen fel kell hívni a figyelmet arra, hogy egy feltétel sérülése kis hatást válthat ki, de több kiinduló feltevés egyidejű megsértése komoly következményekkel járhat. Szintén a paraméteres tesztek robusztusságát járja körbe *Wiley*, *Bunderson* és *Olsen* [2000] tanulmánya. *Vargha* [2004] mutatja be magyar nyelven a kétszemponos sztochasztikus összehasonlítás modelljét, amely ordinális változókra alkalmazható.

A negyedik ponttal kapcsolatosan Siegel maga is elismerte, hogy ez a követelmény nem a paraméteres modellekkel, azok matematikai formájával, hanem a modellek alkalmazhatóságával kapcsolatos. Az ellenérvek (miszerint a paraméteres próbák nyugodtan elvégezhetők) két alapvető indoklásra alapulnak. Az első, logikai okfejtés szerint a statisztikai módszereknek semmi közük ahhoz, hogy milyen

jellemzőkkel bíró adatokon alkalmazzák őket. A statisztika csupán a számokról mond valamit, az a mérést végző feladata, hogy a számok helyesen írják le a valóságot. A legsarkosabban valamivel később talán *Baker, Hardyck* és *Petrinovich* [1966] fogalmaznak, miszerint a statisztikai eljárások számokra vonatkoznak. Tanulmányuk címe a gyenge mérés elméletének híveiként (Stevens és tanai), valamint az erős statisztika követőinek aposztrofálja a két, eddigiekben bemutatott tábor (weak measurement vs. strong statistics) Ezt az első érvet ma már igen kevés kutató véli tarthatónak. A második érvet a paraméteres eljárások alkalmazhatósága mellett empirikus tanulmányok adják a liberálisok szerint, melyekben a vizsgált adatokat transzformációknak vetik alá, majd megvizsgálják a számított statisztikákhoz tartozó valószínűségi szinteket. Baker, Hardyck és Petrinovich idézett tanulmányukban például azt találták, hogy csak a lineáristól erősen eltérő transzformációk voltak jelentős hatással a t -értékekre, így az elemzések alapján levonható következtetésekre. A tanulmány arra az ötletre épül, hogy különböző skála-transzformációkat definiáltak, majd szimuláció segítségével vizsgálták azok hatását a t -próba eredményére (1 és 5 százalékos szignifikanciaszintek esetére). Összességében azt a következtetést vonták le, hogy „az erős statisztikai eljárások, mint például a t -próba, több mint megfelelőek a gyenge mérések vizsgálatához, néhány apróbb feltétellel a t -eloszlásból becsült valószínűségeket csak kis mértékben befolyásolja az alkalmazott mérési skála” (*Baker–Hardyck–Petrinovich* [1966] 308. old.). Hasonló tanulmányok a későbbiekben is születtek, melyek közül *Zimmerman* és *Zumbo* ([1989], [1990], [1993]) munkáit említhetjük.

Anderson [1961] objektív tanulmányt írt a paraméteres és nemparaméteres tesztekkel kapcsolatban, két – a gyakorlati statisztikai és a méréseméleti – szempont figyelembevételével. A tanulmány alapvetően a hipotézisellenőrzés problematikájával foglalkozik, a leíró statisztikai és becslési aspektusoktól eltekint.

A gyakorlati statisztikai megfontolások esetén a szerző legnagyobb, nemparaméteres próbákkal kapcsolatos problémája, hogy a kidolgozott eljárások nem elég változatosak, több kérdésre, hipotézisre csak paraméteres próba ad adekvát választ. Összefoglalóan az a véleménye, hogy „amíg a nemparaméteres próbák nem fejlődnek olyan szintre, hogy a kutatók rutin szükségleteit kielégítsék, addig nem tekinthetők a paraméteres próbák igazi versenytársainak. Addig a napig a nemparaméteres tesztek a numerikus adatok elemzésének hasznos, de mellékes eszközeinek kell tekinteni” (*Anderson* [1961] 307. old.). Tanulmánya második felében *Anderson* a következő két fontos kérdésen keresztül mutatja be véleményét a mérési skálák és a statisztikai eljárások kapcsolatával összefüggésben:

1. Ordinális skálán mért adatokra is alkalmazható az F -próba?
2. A skála megváltozása esetén a statisztikai eredmények invariánsak lesznek?

Az első kérdéssel kapcsolatban (ahol a szerző *F*-próba alatt általánosan a paraméteres tesztekért) véleménye szerint az irodalomban nagy egyetértés uralkodik a tekintetben, hogy Siegelnek nem volt igaza, amikor azt állította: az alkalmazáshoz a változóknak legalább intervallumskálán mérteknek kell lenniük. Ami megkérdőjelezheti a módszerek alkalmazását, az a normalitás és/vagy egyenlő variancia feltételének sérülése. Ebben az esetben azonban a nemparaméteres próbák melletti döntés nem a skála tulajdonságai, hanem pusztán statisztikai jellemzők alapján történik.

A második kérdéssel összefüggésben a szerző elismeri, hogy könnyen található olyan példa, melyben ordinális skálákon mért változóval kapcsolatosan levont hipotézisellenőrzési következtetést megváltoztat a skála megengedhető (szigorúan monoton) transzformációja. Azaz például az eredeti skálán nem mutatkozik két minta alapján különbség a sokaságok között, azonban a transzformációkat valamennyi értéken végrehajtva már szignifikáns a különbség. A szerző elismeri a rangstatisztikák „logikai előnyét” a paraméteres próbákhoz képest, de a gyakorlatban ezt elhanyagolhatónak tartja, főként a már említett változatosság hiánya miatt.

Egy másik, az adott időszakra jellemző tanulmányban *Labovitz* [1967] egy négyfokozatú ordinális skálát mutat be, melyen terápiákat kell értékelniük a betegeknek. A lehetséges válaszok:

1. ártott a kezelés (-);
2. nem volt hatása a kezelésnek (0);
3. valamelyest segített a beavatkozás (+);
4. sokat segített a terápia (++)

Egy hipotetikus példán keresztül azt mutatja be a szerző, hogy a számértékek ordinális skálához történő hozzárendelésének módja csak kevéssé érinti a levonható következtetéseket. Az alkalmazott pontozási rendszerek mindössze annyi hasonlóságot mutatnak, hogy az első (-) válaszhoz 0 értéket, míg az utolsó (++) válaszhoz 10 értéket rendelt. A nemparaméteres, csupán a válaszlehetőségek sorrendiségét kihasználó Wilcoxon-próba szignifikáns különbséget jelez a kétfajta terápia között a hipotetikus válaszok alapján. A tanulmány további vizsgálatai arra vonatkoznak, hogy a különböző pontozási rendszerrel nyert értékek paraméteres tesztjei (kétmintás *t*-próba) milyen szignifikanciaértékeket eredményeznek, ezek mennyiben különböznek egymástól, illetve a nemparaméteres próbáétól. Az eredmények alapján *Labovitz* azt a következtetést vonja le, hogy „jogos a rangskálán mért adatokhoz tetszőlegesen számokat rendelni, amennyiben a hozzárendelés a monotonitás követelményének megfelel” (*Labovitz* [1967] 154. old.). A különböző skálák közötti választás – amennyiben pótlólagos „távolságinformáció” nem áll rendelkezésünkre – leghatékonyabb módja a lineáris pontrendszer alkalmazása. Ez a választás egyrészt közepes, biztosítja, hogy nem tévedünk „túl sokat” rossz irányban, másrészt a Wilcoxon-

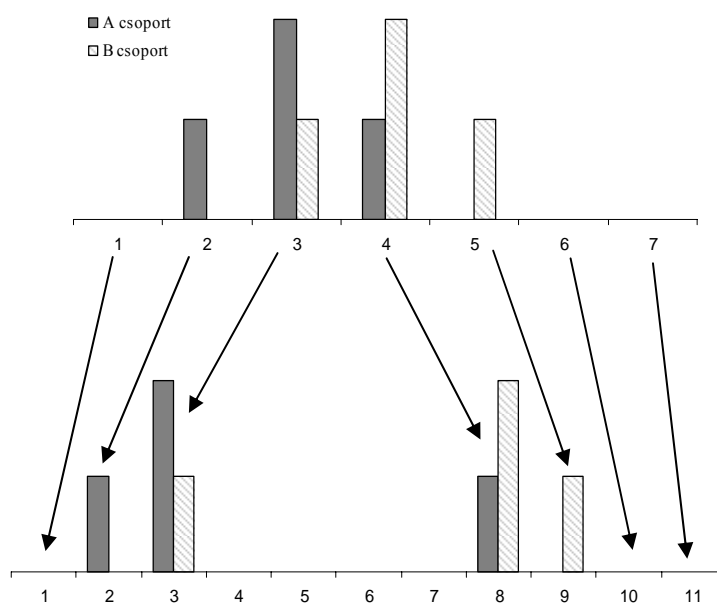
teszthez nagyon hasonló eredményt szolgáltat. Labovitz – helyesen és tudományos korrektséggel – maga is elismeri: a bemutatott példa természetesen nem jelenti azt, hogy minden esetben megfelelő eljárás a paraméteres próba, ezen eljárás automatikus alkalmazásával kapcsolatban – bizonyítékok hiányában – óvatosságra inti a kutatókat. Javasolja, hogy egy konkrét kutatási probléma megoldásakor ne az igazi pontozási rendszer kialakításán fáradozzanak a tudósok, hanem teszteljék az adataikat több elképzelhető pontozási rendszerrel is. A paraméteres eljárások robusztusságával és a próbák erejével összefüggésben, az erős statisztikák vonal képviselőit idézve, az ismert következtetéseket vonja le: a paraméteres próbák nagy robusztusságot mutatnak, több kiinduló feltevés egyidejű megsértése azonban komoly következményekkel járhat. Összességében a szerző óvatosan fogalmaz a nemparaméteres tesztek szerepével kapcsolatban, mégis úgy gondolja, hogy az elkövetett hiba nagyságát és az elméleti követelmények megsértését bőven ellensúlyozzák a paraméteres próbák előnyei. Azokban az esetekben, ha a kategóriák száma nem túl kicsi, valamint a számok kategóriákhoz rendelése nem „extrém” módon (például dichotomizálás) történik, úgy az viszonylag szabadon elvégezhető, azaz kicsi a hatása a levonható következtetésre, véli a szerző. A korábbi tanulmányokkal ellentétben Labovitz az alkalmazható leíró és következtetési statisztikai eszköztár mellett érintőlegesen megemlíti a kapcsolatszorossági mérőszámokat.

A mérési skálák és a statisztika kapcsolatát taglaló viták újbóli kirobbanását néhány, 1979-ben megjelent könyv okozta, melyekről *Gaito* [1980] írt kemény hangú kritikát. Ezekben a – főként szociológusoknak írt – tankönyvekben a liberálisok által támadott Stevens-féle megközelítés szerepelt, ezt a nézetet képviselte a németnyelvű, *Bünning* és *Trenkler* [1978] által jegyzett kötet is. A *Gaito* által felhozott ellenérveket (*Anderson* [1961]; *Boneau* [1960], [1961]; *Burke* [1953]; *Gaito* [1960]) szellemiségében már korábban bemutattuk. A szerző szerint élesen szét kell választani a méréselméletet és a statisztikai elméletet. Annak eldöntése és ellenőrzése kizárólag a méréselmélet feladata, hogy a jelenségekhez rendelt számok megfelelően visszaadják-e a szóban forgó tulajdonságokat, a statisztikának csupán a statisztikai próbák előfeltételeire szabad koncentrálniuk, de a gyakran alkalmazott eljárások robusztusságát újfent hangsúlyozza.

Gaito cikkére válaszul jelent meg némileg később *Townsend* és *Ashby* [1984] írása, ami egyértelműen azt állítja, hogy a mérési skálát figyelembe kell vennünk az elemzésünk készítése során. A szerzők véleménye szerint a mérés „számok objektumokhoz rendelését jelenti, méghozzá olyan módon, hogy a köztük empirikusan megfigyelhető, kvalitatív kapcsolatokat maguk a számok, illetve a számok rendszere jól leírja” (*Townsend–Ashby* [1984] 394. old.). A liberálisok talán legismertebb alakjainak, *Lord*nak és a *Baker–Hardyck–Petrinovich* szerzőhármasnak munkáit veszi a továbbiakban górcső alá a tanulmány. *Lord* futballmezék számozásáról írt gondolatait egyértelmű kritika éri. Egyrésztől azért, mert a számok eredetileg csupán a játékosok

azonosítását szolgálták, így azok nagysága semmilyen összefüggésben nem állt a valósággal: a statisztikai elemzés nem választható el a vizsgált jelenségtől. A Baker és társai által írt – a korábbiakban bemutatott, a robusztusságot hangoztató, alátámasztó – cikkkel kapcsolatban Townsend és Ashby megmutatja, hogy könnyedén konstruálható olyan transzformáció, mely segítségével a szignifikáns különbség „eltüntethető”, illetve fordítva: a jelentéktelen különbség felnagyítható. Az ábrán egy egyszerű példát mutatunk be az utóbbi esetre. Mivel ordinális skáláról beszélünk, az eredeti (az ábra felső része) helyett új (az ábra alsó része) kódolással, vagy az ábrán látható (monoton) transzformációval elérhető, hogy a paraméteres próba az egyik esetben szignifikáns különbséget mutasson az értékek között, míg a másik esetben a két csoport különbsége nem jelentős.

Nem szignifikáns különbség szignifikánssá tétele



Forrás: Townsend–Ashby ([1984] 399. old.).

A *Psychological Bulletin*-ben zajló Gaito–Ashby-vita mellett a tudomány más területein is hasonló cikkpárbajok bontakoztak ki (lásd például Armstrong [1981], [1984]; Knapp [1984], [1990]), melyek részletes bemutatásától eltekintünk, esetükben az érvek és ellenérvek hasonló logikai utakat jártak be. Ez a fajta párhuzamosság azonban önmagában is érdekes.

A későbbiekben a kutatók egyre inkább úgy vélték, hogy a skálák közötti választás, a skála erősségének meghatározása nem olyan egyértelmű, mint ahogy azt Stevens gondolta. Később *Knapp* [1990] már „ordinális”, „kevesebb mint ordinális” és „több mint ordinális” skálákról is beszél. A Soha, Ritkán, Gyakran, Mindig ismérvváltozatokból álló skálát a legtöbb kutató úgy elemezné, hogy számokat rendel a négy kategóriához (lineárisan vagy akár nemlineárisan). Amennyiben azonban a lehetséges válaszok például: Soha, Esetenként, Néha, Mindig lennének, úgy a két középső lehetőség sorrendjének megállapítása komoly problémát jelentene. Az ordinális és intervallumskálák között létezik bizonyos átmenet, mely a használatban levő mérési skálák jó részének sajátja. Az ilyen, „átmeneti” skálák esetén alkalmazható módszerek köre természetesen ugyancsak kérdéses.

Joel Michell [1986] cikkében mintegy válaszul az eddigiekben bemutatott tudományos vitára a méréselmélet tanait három nagy iskolára osztotta, és a kutatók közötti ellentéteket erre vezette vissza. Tanulmányának célja csupán a különböző irányzatok követőinek azonosítása, és az általuk képviselt tanok bemutatása volt, ahogy ez esetünkben is igaz. Michell a három különálló – reprezentációs, operacionalista és klasszikus – elmélet szemléletét (néhol az általa is hivatkozott, alapvető munkák mélyebb ismertetésével) mutatjuk be a következőkben. A tanulmányban a pszichológia területén tetten érhető iskolákat tárja fel, amelyek jelenléte valamennyi, kvantitatív, statisztikai módszereket alkalmazó tudományágban kimutatható. A reprezentációs elmélettel bővebben foglalkozunk, egyszerűen azért, mert képviselői jóval nagyobb irodalmat tudhatnak magukénak, az elmélet összetettsége, matematikai alapjai miatt.

1.1. Reprezentációs elmélet

Michell a reprezentációs elmélet korai megjelenésének tartja *Helmholtz* [1887], *Hölder* [1901], *Russel* [1903] és *Campbell* [1920] műveit, melyek alapul szolgáltak *Suppes* [1951], [1959], valamint *Suppes* és *Zinnes* [1963] munkásságához, akik az elmélet matematikai alapjait fektették le. Hölder eredeti, német nyelvű írását Michell és munkatársa fordították angolra (*Michell–Ernst* [1996], [1997]). A reprezentációs elmélet fejlődése főként néhány kutató nevéhez köthető, a megjelent tanulmányok, cikkek jó részét ők jegyzik, melyek közül néhányat az irodalomjegyzékben felsoroltunk. Ennél részletesebb jegyzék található például *Khurshid* és *Sahai* [1993] és *Luce* [1996] munkáiban. A következőkben a már megismert, Stevens-féle skálákkal kapcsolatban mutatjuk be a reprezentációs elmélet definícióinak jellegét, eltekintve a pontos matematikai leírástól.

Tegyük fel, hogy a vizsgálandó jelenség a hajszín. A reláció ebben az esetben, hogy két személy hajszíne megegyezik, vagy sem. Úgy kell számokat személyekhez

rendelnünk, hogy bármely két személy esetén, ha x hajszíne megegyezik y -ével, akkor és csak akkor $M_x = M_y$, ahol M_x az x -hez, M_y az y -hoz rendelt szám. Nevezzük ezt, a Stevens által nominálisnak nevezett skálát X hajszínskálának. A nominális skálák esetén bármely egy-egy értelmű hozzárendelés megengedhető (admissible) transzformáció.

Tekintsünk következőként egy gyenge sorrend relációt (weak order relation). Az alábbi megállapítást tehetjük: x dolgozata legalább olyan jó, mint y -é. Ha ez a reláció tranzitív és kapcsolt (connected), akkor azt leírhatjuk a matematikai \geq jellel. Ekkor a hozzárendelést úgy kell elvégezni, hogy x minősége akkor és csak akkor legalább olyan jó, mint y -é, ha $M_x \geq M_y$ állítás igaz. Az eredmény egy ordinális skála. Az ordinális skálák esetén csak a (szigorúan) monoton növekvő transzformációk megengedettek.

Nézzük valamely attribútumra vonatkozóan a különbségek sorrendjét. Ekkor – néhány könnyen tesztelhető feltételezés fennállásakor – a számok hozzárendelése megtörténhet. Amennyiben w és x közötti különbség legalább akkora, mint y és z között, akkor és csak akkor $M_w - M_x \geq M_y - M_z$. Az eredmény egy intervallumskála, ahol a megengedhető transzformációk halmaza valamennyi pozitív lineáris transzformációból áll.

Végezetül tekintsünk egy sorrendi relációt az objektumok valamely jellemzőjének összegére (összekapcsolására) vonatkozóan. Példaként a fizikai hosszúságot véve, legyen A szilárd rudak halmaza. Bármely, A -ban levő x és y rúdra vonatkozóan legyen $x \cdot y$ a két rúd (végeiknél való) összeillesztéséből származó rúd (\cdot az összekapcsolás jele). Amennyiben a feltételek megfelelnek az extenzív struktúra (extensive structure) követelményeinek, úgy a hosszúsági sorrend leképezhető numerikusan. Amennyiben $w \cdot x$ legalább olyan hosszú, mint $y \cdot z$, akkor és csak akkor $M_w + M_x \geq M_y + M_z$. Az eredmény egy arányskála, és a megengedhető transzformációk a hasonlósági transzformációk.

A reprezentációs elmélet hívei úgy gondolják, hogy a szóban forgó relációk jellegetől függ, hogy milyen módszerek alkalmazhatók velük kapcsolatban. A fő probléma annak meghatározása, hogy mely skálák esetén milyen eljárások ezek. Stevens szerint az egyes skálákat tekintve az objektumokhoz rendelt számokkal kapcsolatban vannak nem megengedhető műveletek, melyek eredményei nem invariánsak az elvégezhető, megengedett skála-transzformációk tekintetében. Stevens invariancia definíciója azonban távolról sem volt pontos, ráadásul a későbbi vélemények szerint „a tudományban minden tény megengedhető” (Michell [1986] 399. old.). A megfelelő (appropriate) statisztikákról többek között Suppes, valamint Adams, Fagot és Robinson [1965] pontosították Stevens elképzeléseit. A logikai, többnyire intuitív vagy néhány példán alapuló érvelést felváltotta a tételek matematikai bizonyítása.

A megengedhetőség, megfelelőség fogalmához szorosan kapcsolódik a statisztikai értelmesség (meaningfulness) koncepciója, melyet Suppes fektetett le, miszerint „egy empirikus hipotézis vagy bármilyen állítás, mely numerikus mennyiségeket tartalmaz, csak abban az esetben értelmes (meaningful), ha igazságtartalma változatlan a numerikus mennyiség megfelelő transzformációi esetén is” (Suppes [1959] 131. old.). Az egyik klasszikus, sokat idézett példamondat a következő: „Kétszer olyan magas vagyok, mint a Sears Tower” (Marcus-Roberts–Roberts [1987] 384. old.). A kijelentés jól láthatóan hamis, azonban értelmes, hisz igazságtartalma változatlan centiméterben, méterben, lábban vagy hüvelykben mért magasság esetén is.

Michell az értelmesség két változatát, megközelítését különbözteti meg: a skála-specifikus megállapításokra vonatkozó és a skálafüggetlen megállapításokra vonatkozó értelmességet. Míg a skálaspecifikus megállapítások tartalmazzak egy bizonyos mérési skálára vonatkozó hivatkozást, addig a skálafüggetlen megállapítások esetén ez nem igaz. Suppes előzőekben bemutatott definíciója jól láthatóan skálaspecifikus megállapításoknál alkalmazható. A megközelítés egyik problémája: attól függetlenül, hogy egy állítás értelmetlen, még lehet tudományos értelemben hasznos, segítségével valós következtetéseket vonhatunk le a vizsgált egyedekről. Így az értelmetlen megállapítások nem „számúzhetők” automatikusan. Például, ha azt állítjuk, hogy a mai hőmérséklet a tegnapi kétszerese, állításunk könnyen beláthatóan értelmetlen lehet (meg kell adnunk a hőmérsékleti skálát is, hiszen, ha mi Celsius fokban értjük, annak számára, aki ezt Fahrenheitben gondolja, a viszony nem kétszeres). Ennek ellenére hasznos, hisz tudjuk, hogy melegebb van, mint tegnap volt (0 fok feletti hőmérsékletet feltételezve). Egy másik példa szerint a hajszínt számokkal jelölve az X hajszínskálán értelmetlen megállapítást tehetünk, amennyiben azt mondjuk, hogy a mintánkban a hajszínek összege 10. Ha tudjuk azonban, hogy a vörös hajúakat 3-as számmal jelöltük, akkor az értelmetlen megállapítás haszna az, hogy tudjuk, nem minden egyed vörös hajú.

Adams, Fagot és Robinson [1965] az értelmesség skálafüggetlen definícióját javasolták, hiszen a mérés során célunk általában nem az, hogy skálafüggő megállapításokat tegyünk, hanem a jelenségekről szeretnénk skálafüggetlen információkhoz jutni. A szerzők által javasolt definíció lényege, hogy egy skálafüggetlen kijelentés akkor és csak akkor értelmes, ha az igazságtartalma valamennyi skálaspecifikus változatának azonos. Egy skálafüggetlen kijelentés skálaspecifikus változatát úgy nyerjük, hogy minden mérendő változót valamely skálára vonatkozóan írunk le, természetesen minden értékhez azonos skálát rendelve. A skálafüggetlen kijelentések értelmessége különös jelentőséggel bír, de ebben az esetben is szembesülhetünk nehézségekkel. A következő két állítás minden kétséget kizárólag értelmetlen skálafüggetlen megállapítás: A magassága 6,4; B magassága 3,2 (Michell [1986]). A két kijelentés alapján levonható következtetés: A magassága kétszerese B -ének, ami kétségkívül értelmes skálafüggetlen megállapítás, és akár igaz is lehet.

A reprezentációs elmélet követői rengeteg tételt dolgoztak ki és bizonyítottak az utóbbi évtizedekben. Az értelmesség definícióján túl, ehhez kapcsolódóan az invariancia (invariance), homogenitás (homogeneity) és a megfelelő statisztikák (appropriate statistics), valamint ezek kapcsolódási pontjai álltak a kutatások közép-pontjában. A Stevensi tanokhoz kapcsolódóan sikerült nagyrészt tisztázni azt a kérdést, hogy milyen mérési skálák lehetségesek. A kutatások alapján jól látszik, hogy Stevens nem tévedett nagyot akkor, amikor igen kevés skálát vezetett be elméletébe (Narens [1981a], [1981b]). A hasznosságelméletek területén értek el további jelentős eredményeket a reprezentációs elmélet követői. Az extenzív struktúrákon kívül egyéb struktúrák feltárása is sikerrel járt, melyek közül a conjoint struktúra (Luce–Tukey [1964], Tversky [1967]) a legjelentősebb, melynek gyakorlati alkalmazása mára szélesebb körben elterjedt, igen jelentős a már említett hasznosság mellett az attitűdök, képességek stb. mérésében is. Bizonyítható, hogy amennyiben a matematikai axiómák teljesülnek, úgy az eredmény intervallumskála erősségű lesz (Krantz et al. [1971]).

A reprezentációs elmélet nézeteinek szélesebb körben történő elterjedését leginkább az hátráltatja, hogy a gyakorlatban alkalmazott módszerek döntő többsége az elmélet szerint nem minősül mérésnek, így az azokat felhasználó kutatók nem mutatnak kellő érdeklődést. A matematikai-logikai tételek ráadásul nehezen érthetőek, sok szempontból túlságosan elméletiek (Velleman–Wilkinson [1993]). A mérési hiba jelensége nem került beépítésre a reprezentációs elmélet logikai keretrendszerébe, ami kritikákat váltott ki, melyre azonban született reakció (Luce–Narens [1994]). A témakörrel foglalkozó, legfrissebb kézikönyveket Narens ([2002], [2007]) jegyzi. Az érdeklődőknek ajánljuk továbbá a Krantz et al. [1971], Suppes et al. [1989] és Luce et al. [1990] által jegyzett háromkötetes sorozatot.

1.2. Operacionalista elmélet

A társadalomtudományok területén rengeteg olyan információforrás van, amely kvantitatív eredményeket szolgáltat, reprezentációs értelemben mégsem tekinthetjük azokat mérésnek. Gondoljunk egy szellemi képességeket mérő tesztre, amely több kérdésből áll, és miután az egyed kitöltötte, az eredmény a kérdéseknek megfelelő számú helyes/helytelen válaszokat tartalmazó sor lesz. Az adatokkal kapcsolatos fontos empirikus reláció az, hogy *A* személy legalább azokat a kérdéseket jól megválaszolta-e, mint *B*. Ebben az esetben *A* teljesítménye legalább olyan jó, mint *B*-é. Amennyiben ez a reláció tranzitív és kapcsolt valamennyi válaszadóra, úgy az eredmény ordinális skálán mért (a reprezentációs elmélet alapján). Ez az eset azonban a legkritikában fordul elő, így ez a teszt még az ordinális erőt sem éri el. Másik lehetőség, hogy az empirikus kapcsolat vizsgálata helyett a tesztet végzők összeadják a he-

lyes válaszok darabszámát, és ezt az értéket tekintik az adott személy tesztértékének, azonban ekkor nem világos, hogy mi az a reláció, amit vizsgálunk.

Az operacionalista definíció szerint a mérés nem más, mint egy művelet, ami számot eredményez, ami hasonló Stevens értelmezéséhez. A számok tehát műveletek eredményei: „a szigorúan vett operacionalista számára a tudomány egyszerűen a műveletek tanulmányozása, nem pedig a valóságé” (Michell [1986] 404. old.). Ebben a tekintetben pedig a számok valóban nem tudják, honnan jöttek, azaz szabadon végezhetők velük műveletek; a skáláknak és a statisztikai módszereknek egymáshoz nincs közük. Mindez azonban a mérés és a tudomány kapcsolatát egészen más fényben mutatja be, mint a reprezentációs elmélet esetén.

1.3. Klasszikus elmélet

A klasszikus elméletet Michell egészen *Arisztotelész*ig és *Euklidesz*ig vezeti vissza. Az elmélet szerint a mérés nem más, mint annak a megállapítása, hogy az egység hányszor szerepel egy adott mennyiségben. Mindez a mérés, a mérhető jellemzők körét erősen leszűkíti, bár a reprezentációs elmélettel szemben nem követeli meg a vizsgált objektumok közötti empirikus kapcsolatrendszer létét. A klasszikus elmélet szerint a mérés nem számok hozzárendelését jelenti az objektumokhoz, ahogyan azt a reprezentációs és operacionalista tábor véli. A klasszikusok szerint a mérés nem más, mint számszerű kapcsolatok felfedezése, feltárása a kvantitatív jellemzők között.

Az eltérő definícióból adódik, hogy a klasszikusok esetén nem beszélhetünk mérési skálákról, hiszen a számok mindig ugyanabból a folyamatból, a mennyiségi kapcsolat (arányosság) feltárásából származnak. Ez a felfogás leginkább a Stevensi arányskála tulajdonságait hordozza, az elvégezhető műveletek köre a lehető legszélesebb, mérési skálák nincsenek, így ezek korlátozást sem jelentenek az alkalmazható statisztikák tekintetében.

2. Napjaink gondolatai

A statisztikai szoftverek elterjedésével párhuzamosan ismét felvetődött a kérdés, hogy a mérési skálák befolyásolják-e, és ha igen, mennyiben az alkalmazható módszereket. Szükséges-e, hogy a szoftverek beépítetten korlátozzák a felhasználót az elérhető módszerek tekintetében, a mérési szintet figyelembe véve. *Velleman* és *Wilkinson* [1993] azt állítják, hogy a korlátozás néhol felesleges, sőt rossz lehet. Legfontosabb megállapításuk, hogy a skálátípus nem tisztán az adatok jellemzője, hanem

attól is függ, mi a kérdésfeltevésünk. Hasonló gondolattal találkozhatunk *Surányi-Vita* [1972] tanulmányában is, akik a keresetek példáján keresztül tárgyalják a jelenséget. Pusztán pénzügyi szempontból a kereset arányskálán (vagy abszolút skálán) mért jellemző, közgazdasági vagy szociológia szempontú kérdésfeltevés esetén azonban nem egyértelmű, hogy ugyanez igaz-e. *Hand* ([1996], [2004]) véleménye szerint minden gyakorlati életben történő mérés egyfajta keveréke a reprezentációs és a pragmatikus nézőpontoknak. A skálák abban különböznek, hogy a két szemlélet különböző „súlyokkal” szerepel bennük.

Michell napjainkban is a mérés elméletével, a társadalomtudományok területén betöltött szerepével foglalkozik. Munkái ([1986], [1994], [1999], [2005], [2008]) szélsőségesen kritikus áttekintést adnak a területről, melyek szerint súlyos hiba bizonyíték hiányában elfogadni, azt az állítást, hogy egyes jellemzők kvantitatívak. Természetesen Michell saját tudományterületének, a pszichológiának a méréseiről mond véleményt, de gondolatai más területek mérési módszereire is vonatkoznak. Szélsőséges álláspontja szerint komoly erőfeszítések soha nem történtek a különböző jellemzők kvantitatív voltának bizonyítására, a méréseket végzők azt vizsgálatok nélkül elfogadják. Michell szerint nem is ez a legnagyobb probléma, hanem az, hogy az empirikus adatokkal dolgozó kutatóknak eszükbe sem jut: a vizsgált jelenség esetleg nem is kvantitatív, azaz a pszichometria (és minden méréseken alapuló társadalomtudomány) ilyen értelemben „kóros tudomány” (pathological science). Michell szerint túl nagy hatással volt a tudományra Stevens tág mérésdefiníciója, ami minden olyan folyamatot, ami számértékeket eredményez, mérésnek tart. *Michell* [2008] két olyan körülményt azonosít tanulmányában, melyek megmagyarázhatják ezt az állapotot. Az első ok ideológiai: a tudományosság (scientism). Sokan a mai napig is úgy gondolják, hogy tudományos megismerés csak mérés útján érhető el, ami egyben a tudományosság mérőfoka. A mérés és a statisztikai módszertan bevezetése a (kvantitatív) tudományok körébe emelte az pszichológiát. Ehhez kapcsolódik a gyakorlatiasság (practicalism) szükségességének elterjedése, ami a tudomány sikerét gyakorlati problémák megoldásában méri, szemben azzal a klasszikus tudományfelfogással, ami a vizsgált rendszer megértését helyezi előtérbe. A másik, Michell által említett ok gazdasági jellegű: komoly, méréseken alapuló tudományok könnyebben kaptak a világháború után állami vagy ipari megbízásokat, így a pszichológiai jellemzőknek egyszerűen kvantitatívnak kellett lenniük. Ez a vélemény természetesen messzemenőig szélsőséges, a tudományos közvélemény által erősen vitatott, azonban megemlítését fontosnak tartottuk. Michell természetesen ezzel nemcsak a pszichológiai mérést véleményezi, hanem gyakorlatilag a társadalomtudományok, döntő többségét és némely természettudományt is. A túlságosan radikális vélemény – mellyel jelen tanulmány szerzője nem ért egyet – azonban felhívhatja a figyelmet arra, hogy nem csupán attól válhat valami tudománnyá, tudományossá, ha mérések társulnak hozzá; ha egyes jelenségek nem (megfelelően) mérhetők, merjük ezt kijelenteni.

Hasonló folyamatoknak más területeken is tanúi lehetünk: a marketingkutatás, a szociológia és sok társadalomtudomány attitűdök, képességek, belső értékek vizsgálatát végzi. Rengeteg mérési módszer került kidolgozásra, melyek a mérni kívánt jellemzők széles skáláját fedik le, például: szükségletek, preferenciák, önképek, értékek, érzelmek, reakciók stb. mérése. *Bearden és Netemeyer* [1999] például több száz mérési módszert felsorakoztató kötetet állított össze, azok eredeti megjelenésével, rövid leírásával, az addigi kutatási tapasztalatokkal. Így ezek a mérési módszerek „legjobb gyakorlatként” terjednek egy-egy nevesebb kutató vagy kutatócsoport publikációi nyomán. Mindez azt okozza, hogy a hasonló témakörben tevékenykedő tudósok hasonló módon készítik el kérdőíveiket. Kérdés természetesen, hogy adott jelenség mérhetőségét mi módon állapíthatjuk meg.

Meg kell ugyanakkor jegyeznünk, hogy napjainkra egyszerűen nem tartható Anderson „kifogása” az ordinális és nominális ismérvekkel kapcsolatos módszerek szűkösségével kapcsolatban. A nemparaméteres eljárások és kategóriás adatok elemzésére alkalmas módszerek sokasága került kifejlesztésre az utóbbi évtizedekben, nem kis mértékben a bemutatott méréselméleti vita folyamányaként. Napjainkban nemzetközi standard kézikönyvként *Agresti* [2002] művét alkalmazzák, oktatják a klasszikus statisztika területén, a kategóriás adatok bayesi modelljeit leíró munkát pedig *Congdon* [2005] jegyzi. Magyar nyelven a legátfogóbb összefoglalót *Vargha* [2007] műve adja, mely részletesen mutat be ordinális skálákon végezhető műveleteket és tesztek is, különös tekintettel a sztochasztikus egyenlőségek, különbségek vizsgálatára. A szükséges módszerek tehát rendelkezésre állnak, a kutató felelőssége azok megismerése és alkalmazása, amennyiben az adatok jellemzői ezt megkívánják.

3. Összegzés

Tanulmányunkban a méréselmélettel és skálákkal kapcsolatos tudományos vitát mutattuk be az adott időszakokra jellemző publikációkon keresztül, ezzel bővítve az irodalomjegyzéket adva az érdeklődő Olvasónak. Természetesen ma már minden megállapítással értünk egyet, de úgy gondoljuk, hogy a tudományos világban ma sincs konszenzus a témakörrel kapcsolatban. Véleményünk szerint a méréselmélet fogalmainak és történetének megismerése az alkalmazott kutatásokat végző tudósok számára alapvető fontossággal bír. A kategóriás adatokra kidolgozott modellek, a nemparaméteres próbák – főként a népszerű statisztikai programcsomagoknak köszönhetően – várhatóan egyre elterjedtebbekké fognak válni. Az oktatásban, ennek megfelelően, nagyobb szerep vár az ilyen irányú ismeretek továbbadására.

Ne feledjük azonban – Lord szavaival élve –, hogy a számok nem tudják honnan jöttek, de tegyük hozzá, a szoftverek sem, így a kutatónak kell azt észben tartaniuk. Az adatainkból bármilyen mutatót kiszámíthatunk, modellt illeszthetünk, de vigyázzunk a belőlük levonható következtetésekkel! Amennyiben hipotéziseket tesztelünk, csak értelmes (meaningful) kérdéseket tegyünk fel, nehogy levont következtetésünk csak a skála sajátja legyen. Ha szükséges, tegyünk skálafüggő kijelentéseket skálafüggetlenek helyett, ezzel jelezve, hogy más mérési módszerrel akár más eredményeket is kaphattunk volna. A és B csoport vevői elégedettségének mérése esetén megállapításunkat fogalmazzuk meg úgy, hogy A csoport alacsonyabb elégedettségi pontszámmal rendelkezik, mint B csoport az adott mérési módszerrel, ne pedig úgy, hogy A csoport kevésbé elégedett a termékkel/szolgáltatással. Ha nem vagyunk biztosak abban, hogy az elégedettség mérése ordinálisnál erősebb skálát eredményez, alkalmazzunk ennek megfelelő teszteket. Ezzel szemben tegyünk nyugodtan skálafüggetlen kijelentéseket, ha ezt a körülmények megengedik. Az elemzési módszertan kiválasztásakor legyünk egészségesen szkeptikusak azok mérési szintjével kapcsolatban, de vegyük figyelembe a felhasználási területet is. Nem várhatjuk, hogy a már említett iskolai osztályzatok alapján két tanulócsoporthoz teljesítményét ezentúl valamilyen nemparaméteres próbával hasonlítsa össze a tanulmányi osztály. Ennek ellenére egy tudományos munkában a megfelelő eljárások és számítógépes háttér birtokában mindez már nem okozhat problémát.

Ne felejtjük el, hogy egy-egy változó több információt tartalmazhat, mint az első ránézésre látszik. A magyar gépkocsik rendszáma például nominális skálán mért. Ez nem jelenti azt, hogy egy autó rendszámából (és egyéb kiegészítő információkból) ne tudnánk igen fontos következtetéseket levonni! Az autó kora és rendszáma között igen szoros a kapcsolat: egy autókereskedő a rendszámából és az autó típusából igen pontosan meg tudja mondani, hogy Magyarországon eladott, vagy külföldről behozott autóról van-e szó? Mivel a rendszámokat az okmányirodák „csomagokban” kapják, az azonos betűkből álló rendszámok egy területen csoportosulnak, így azok a forgalomba helyezés helyére is utalhatnak. Ugyancsak többletinformációt szolgáltat az, hogy kért rendszámmal rendelkezik a gépkocsi: nagy valószínűséggel céges gépjárműről van szó vagy tehetős tulajdonosról. A rövid példa annak érzékeltetésére szolgál, hogy a skálatípus nem feltétlenül keverendő össze az információtartalommal. Vegyük észre, hogy a gyakorlati esetekben a változók mérési skálához rendeltése közel sem triviális, valamint a mérési skála egyazon adatsornál függhet a felhasználási céltól. A rendszám nem vált ugyan ordinális skálán mért ismérvvé, a változó által hordozott információt azonban kár lenne elveszíteni. Ilyen értelemben a reprezentációs elmélet képviselője számára a szigorú skálafeltételek fontosak, az operacionalista számára pedig az információtartalom.

Irodalom

- ADAMS, E. W. – FAGOT, R. F. – ROBINSON, R. E. [1965]: A Theory of Appropriate Statistics. *Psychometrika*. Vol. 30. No. 2. pp. 99–127.
- AGRESTI, A. [2002]: *Categorical Data Analysis*. John Wiley & Sons, Inc. New York.
- ANDERSON, N. H. [1961]: Scales and Statistics: Parametric and Non-Parametric. *Psychological Bulletin*. Vol. 58. No. 4. pp. 305–316.
- ARMSTRONG, G. D. [1981]: Parametric Statistics and Ordinal Data: A Pervasive Misconception. *Nursing Research*. Vol. 30. No. 1. pp. 60–62.
- ARMSTRONG, G. D. [1984]: Letter to the editor. *Nursing Research*. Vol. 33. No. 1. p. 54.
- BAKER, B. O. – HARDYCK, C. D. – PETRINOVICH, L. F. [1966]: Weak Measurements vs. Strong Statistics: An Empirical Critique of S. S. Stevens' Proscriptions on Statistics. *Educational and Psychological Measurement*. Vol. 26. No. 2. pp. 291–309.
- BARTLETT, M. S. [1947]: The Use of Transformations. *Biometrics*. Vol. 3. No. 1. pp. 39–52.
- BEARDEN, W. O. – NETEMEYER, R. G. [1999]: *Handbook of Marketing Scales*. Sage Publications, Inc. Thousand Oaks.
- BEHAN, F. L. – BEHAN, R. A. [1954]: Football Numbers (continued). *American Psychologist*. Vol. 9. No. 6. pp. 262–263.
- BLAIR, R. C. – HIGGINS, J. J. [1980]: A Comparison of the Power of Wilcoxon's Rank-Sum Statistic to that of Student's t Statistic under Various Nonnormal Distributions. *Journal of Educational Statistics*. Vol. 5. No. 4. pp. 309–335.
- BLAIR, R. C. – HIGGINS, J. J. [1985]: Comparison of the Power of the Paired Samples t Test to that of Wilcoxon's Signed-Rank Test under Various Population Shapes. *Psychological Bulletin*. Vol. 97. No. 1. pp. 119–128.
- BLALOCK, H. M. [1968]: The Measurement Problem: A Gap between the Languages of Theory and Research. In: *Blalock, H. M. – Blalock A. B. (eds): Methodology in Social Research*. McGraw-Hill. New York. pp. 5–28.
- BONEAU, C. A. [1960]: The Effects of Violations of Assumptions Underlying the t Test. *Psychological Bulletin*. Vol. 57. No. 1. pp. 49–64.
- BONEAU, C. A. [1961]: A Note on Measurement Scales and Statistical Tests. *American Psychologist*. Vol. 16. No. 5. pp. 260–261.
- BOX, G. E. P. [1953]: Non-Normality and Tests on Variances. *Biometrika*. Vol. 40. No. 3–4. pp. 318–335.
- BÜNING, H. – TRENKLER, G. [1978]: *Nichtparametrische statistische methoden*. Walter de Gruyter GmbH & Co. KG. Berlin, New York.
- BURKE, C. J. [1953]: Additive Scales and Statistics. *Psychological Review*. Vol. 60. No. 1. pp. 73–75.
- CAMPBELL, N. R. [1920]: *Physics, the Elements*. Cambridge University Press. London.
- COCHRAN, W. G. [1947]: Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied. *Biometrics*. Vol. 3. No. 1. pp. 22–38.
- CONGDON, P. [2005]: *Bayesian Models for Categorical Data*. John Wiley & Sons, Ltd. Chichester.
- COOMBS, C. H. – RAIFFA, H. – THRALL, R. M. [1954]: Some Views on Mathematical Models and and Measurement Theory. *Psychological Review*. Vol. 61. No. 2. pp. 132–144.

- EISENHART, C. [1947]: The Assumptions Underlying the Analysis of Variance. *Biometrics*. Vol. 3. pp. 1–21.
- FALMAGNE, J. C. – NARENS, L. [1983]: Scales and Meaningfulness of Quantitative Laws. *Synthese*. Vol. 55. No. 3. pp. 287–325.
- GAITO, J. [1960]: Scale Classification and Statistics. *Psychological Review*. Vol. 67. No. 4. pp. 277–278.
- GAITO, J. [1972]: An Index of Estimation to Ascertain the Effect of Unequal n on ANOVA F Tests. *American Psychologist*. Vol. 27. No. 11. pp. 1081–1082.
- GAITO, J. [1980]: Measurement Scales and Statistics: Resurgence of an Old Misconception. *Psychological Bulletin*. Vol. 87. No. 3. pp. 564–567.
- GARDNER, P. L. [1975]: Scales and Statistics. *Review of Educational Research*. Vol. 45. No. 1. pp. 43–57.
- GAYEN, A. K. [1949]: The Distribution of Students's t in Random Samples of Any Size Drawn from Non-Normal Universes. *Biometrika*. Vol. 36. No. 3–4. pp. 353–369.
- GLASS, G. V. [1972]: Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*. Vol. 42. No. 3. 237–288.
- GODARD, R. H. – LINDQUIST, E. F. [1940]: An Empirical Study of the Effect of Heterogeneous Within-Groups Variance upon Certain F-tests of Significance in Analysis of Variance. *Psychometrika*. Vol. 5. No. 4. pp. 263–274
- HAND, D. J. [1996]: Statistics and the Theory of Measurement. *Journal of the Royal Statistical Society*. Series A. Vol. 159. No. 3. pp. 445–492.
- HAND, D. J. [2004]: *Measurement Theory and Practice: The World Through Quantification*. Arnold. London.
- HELMHOLTZ, H. VON [1887]: Numbering and Measuring from an Epistemological Viewpoint. In: *Hertz, P. – Schlick, M.* (eds): *Hermann von Helmholtz: Epistemological Writings*. Reidel. Dordrecht. pp. 77–114.
- HÖLDER, O. [1901]: Die Axiome der Quantität und die Lehre vom Mass, Berichte über die Verhandlungen der königlich sächsischen Gesellschaft der Wissenschaften zu Leipzig. *Mathematisch-Physische Klasse*. Vol. 53. pp. 1–64.
- HUNYADI L. [2001]: *Statisztikai következtésemélet közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- HUNYADI L. – VITA L. [2008]: *Statisztika I–II*. Aula Kiadó. Budapest.
- KHURSHID, A. – SAHAI, K. [1993]: Scales of Measurements: An Introduction and a Selected Bibliography. *Quality and Quantity*. Vol. 27. No. 3. pp. 303–324.
- KNAPP, T. R. [1984]: Letter to the editor. *Nursing Research*. Vol. 33. No. 1. p. 54.
- KNAPP, T. R. [1990]: Treating Ordinal Scales as Interval Scales: An Attempt to Resolve the Controversy. *Nursing Research*. Vol. 39. No. 2. pp. 121–124.
- KRANTZ, D. H. – LUCE, R. D. – SUPPES, P. – TVERSKY, A. [1971]: *Foundations of Measurement*. Vol. 1. *Additive and polynomial representations*. Academic Press. New York.
- LABOVITZ, S. [1967]: Some Observations on Measurement and Statistics. *Social Forces*. Vol. 46. No. 2. pp. 151–160.
- LABOVITZ, S. [1970]: The Assignment of Numbers to Rank Order Categories. *American Sociologist Review*. Vol. 35. No. 3. pp. 515–524.

- LORD, F. M. [1953]: On the Statistical Treatment of Football Numbers. *The American Psychologist*. Vol. 8. No. 12. pp. 750–751.
- LUCE, R. D. [1996]: The Ongoing Dialog between Empirical Science and Measurement Theory. *Journal of Mathematical Psychology*. Vol. 40. No. 1. pp. 78–98.
- LUCE, R. D. [2005]: Measurement Analogies: Comparisons of Behavioral and Physical Measures. *Psychometrika*. Vol. 70. No. 2. pp. 227–251.
- LUCE, R. D. – KRANTZ, D. H. – SUPPES, P. – TVERSKY, A. [1990]: *Foundations of Measurement. Vol. III. Representation, axiomatization and invariance*. Academic Press. New York.
- LUCE, R. D. – NARENS, L. [1994]: Fifteen Problems Concerning the Representational Theory of Measurement. In: *Humphreys, P.* (ed.): *Patrick Suppes: Scientific Philosopher. Vol. 2. Philosophy of Physics, Theory Structure and Measurement Theory*. Vol. 234. pp. 219–249.
- LUCE, R. D. – TUKEY, J. W. [1964]: Simultaneous Conjoint Measurement: A New Scale Type of Fundamental Measurement. *Journal of Mathematical Psychology*. Vol. 1. No. 1. pp. 1–27.
- MARCUS-ROBERTS, H. M. – ROBERTS, F. S. [1987]: Meaningless Statistics. *Journal of Educational and Behavioral Statistics*. Vol. 12. No. 4. pp. 383–394.
- MICHELL, J. [1986]: Measurement Scales and Statistics: A Clash of Paradigms. *Psychological Bulletin*. Vol. 100. No. 3. pp. 398–407.
- MICHELL, J. [1994]: Numbers as Quantitative Relations and the Traditional Theory of Measurement. *British Journal for the Philosophy of Science*. Vol. 45. No. 2. pp. 389–406.
- MICHELL, J. [1999]: *Measurement in Psychology: Critical History of a Methodological Concept*. Cambridge University Press. Cambridge.
- MICHELL, J. [2005]: The Logic of Measurement: A Realist Overview. *Measurement*. Vol. 38. No. 4. pp. 285–294.
- MICHELL, J. [2008]: Is Psychometrics Pathological Science? *Measurement: Interdisciplinary Research and Perspective*. Vol. 6. No. 1–2. pp. 7–24.
- MICHELL, J. – ERNST, C. [1996]: The Axioms of Quantity and the Theory of Measurement. *Journal of Mathematical Psychology*. Vol. 40. No. 3. pp. 235–252.
- MICHELL, J. – ERNST, C. [1997]: The Axioms of Quantity and the Theory of Measurement. *Journal of Mathematical Psychology*. Vol. 41. No. 4. pp. 345–356.
- NARENS, L. [1981a]: A General Theory of Ratio Scalability with Remarks About the Measurement-theoretic Concept of Meaningfulness. *Theory and Decision*. Vol. 13. No. 1. pp. 1–70.
- NARENS, L. [1981b]: On the Scales of Measurement. *Journal of Mathematical Psychology*. Vol. 24. No. 3. pp. 249–275.
- NARENS, L. [2002]: *Theories of Meaningfulness. Scientific Psychology Series*. Lawrence Erlbaum Associates. Mahwah.
- NARENS, L. [2007]: *Introduction to the Theories of Measurement and Meaningfulness and the Use of Invariance in Science*. Lawrence Erlbaum Associates. Mahwah.
- RAPPAI G. – PINTÉR J. (szerk.) [2007]: *Statisztika*. Pécsi Tudományegyetem Közgazdaságtudományi Kar. Pécs.
- RUSSEL, B. [1903]: *The Principles of Mathematics*. Cambridge University Press. London.
- SCHOLTEN, A. Z. – BORSBOOM, D. [2009]: A Reanalysis of Lord's Statistical Treatment of Football numbers. *Journal of Mathematical Psychology*. Vol. 53. No. 2. pp. 69–75.
- SIEGEL, S. [1956]: *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Co. New York.

- STEVENS, S. S. [1946]: On the Theory of Scales of Measurement. *American Association for the Advancement of Science*. Vol. 103. No. 2684. pp. 677–680.
- STEVENS, S. S. [1955]: On the Averaging of Data. *American Association for the Advancement of Science*. Vol. 121. No. 3135. pp. 113–116.
- SUPPES, P. [1951]: A Set of Independent Axioms for Extensive Quantities. *Portugaliae Mathematica*. Vol. 10. No. 4. pp. 163–172.
- SUPPES, P. [1959]: Measurement, Empirical Meaningfulness and Three-Valued Logic. In: Churchman, C. W. – Ratoosh, P. (eds.): *Measurement: Definitions and Theories*. John Wiley & Sons, Inc. New York. pp. 129–143.
- SUPPES, P. – KRANTZ, D. M. – LUCE, R. D. – TVERSKY, A. [1989]: *Foundations of Measurement. Volume II: Geometrical, Threshold and Probabilistic Representations*. Academic Press. New York.
- SUPPES, P. – ZINNES, J. L. [1963]: Basic Measurement Theory. In: Luce, R. D. – Bush, R. R. – Galanter, E. (eds.): *Handbook of Mathematical Psychology. Vol. I*. John Wiley & Sons, Inc. New York. pp. 3–76.
- SURÁNYI B. – VITA L. [1972]: A mérési szintek elmélete és értéke a társadalomstatisztikában. *Statisztikai Szemle*. 50. évf. 7. sz. 731–743. old.
- THOMAS, H. [1982]: IQ Interval Scales, and Normal Distributions. *Psychological Bulletin*. Vol. 91. No. 1. pp. 198–202.
- TOWNSEND, J. T. – ASHBY, F. G. [1984]: Measurement Scales and Statistics: The Misconception Misconceived. *Psychological Bulletin*. Vol. 96. No. 2. pp. 394–401.
- TVERSKY, A. [1967]: A General Theory of Polynomial Conjoint Measurement. *Journal of Mathematical Psychology*. Vol. 4. No. 1. pp. 1–20.
- VARGHA A. [2003]: Robusztussági vizsgálatok az egymintás *t*-próbával. *Statisztikai Szemle*. 81. évf. 10. sz. 872–890. old.
- VARGHA A. [2004]: A kétszemponos sztochasztikus összehasonlítás modellje. *Statisztikai Szemle*. 82. évf. 1. sz. 67–82. old.
- VARGHA A. [2007]: *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya Kiadó. Budapest.
- VARGO, L. G. [1971]: Comment on „The Assignment of Numbers to Rank Order Categories”. *American Sociological Review*. Vol. 36. No. 3. pp. 517–518.
- VELLEMAN, P. F. – WILKINSON, L. [1993]: Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading. *The American Statistician*. Vol. 47. No. 1. pp. 65–72.
- WILEY, D. A. – BUNDERSON, C. V. – OLSEN, J. A. [2000]: *An Exploratory Study of the Statistical and Educational Implications of Violations of the Assumptions of Parametric Analysis Techniques*. <http://opencontent.org/docs/parametric.pdf>
- ZIMMERMAN, D. W. – ZUMBO, B. D. [1989]: A Note on Rank Transformations and Comparative Power of the Student *t*-test and Wilcoxon–Mann–Whitney Test. *Perceptual and Motor Skills*. Vol. 68. No. 3. pp. 1139–1146.
- ZIMMERMAN, D. W. – ZUMBO, B. D. [1990]: The Relative Power of the Wilcoxon–Mann–Whitney Test and Student *t*-test Under Simple Bounded Transformations. *Journal of General Psychology*. Vol. 117. No. 4. pp. 425–436.
- ZUMBO, B. D. – ZIMMERMAN, D. W. [1993]: Is the Selection of Statistical Methods Governed by Level of Measurement? *Canadian Psychology*. Vol. 34. No. 4. pp. 390–400.

Summary

The main objective of this review is to introduce the so-called measurement controversy arising from the influential paper of *Stevens* [1946]. This article does not attempt to solve the long standing debate, we only give a detailed bibliography as a guideline to show the reader the different opinions. The main conclusion of the paper is that the appropriate methods are currently available, the researcher is responsible for knowing and choosing the correct one.