



Közzététel: 2026. június 29.

A tanulmány címe:

A standardizálás módszerei és története Magyarországon

Szerzők:

SUGÁR ANDRÁS, a Budapesti Gazdaságtudományi Egyetem egyetemi docense, a Budapesti Corvinus Egyetem másodállású egyetemi docense, a Központi Statisztikai Hivatal Módszertani és innovációs főosztály másodállású elemzője

E-mail: Sugar.Andras@uni-bge.hu; andras.sugar@uni-corvinus.hu; Sugar.Andras@ksh.hu

DOI: <https://doi.org/10.20311/stat2026.06.hu0503>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) *Statisztikai Szemle* c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Szjt.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Szjt. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:

„*Forrás: Statisztikai Szemle* c. folyóirat 104. évfolyam 6. számában megjelent, **Sugár Tamás** által írt, **A standardizálás módszerei és története Magyarországon** című cikk (link csatolása)”

7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Sugár András

A standardizálás módszerei és története Magyarországon*

The methods and history of standardisation in Hungary

Sugár András, a Budapesti Gazdaságtudományi Egyetem egyetemi docense, a Budapesti Corvinus Egyetem másodállású egyetemi docense, a Központi Statisztikai Hivatal Módszertani és innovációs főosztály másodállású elemzője
E-mail: Sugar.Andras@uni-bge.hu; andras.sugar@uni-corvinus.hu; Sugar.Andras@ksh.hu

A standardizálás (két vagy több összetett viszonyszám/főátlag összehasonlítása, az összetételhatás kimutatásával és szükség esetén kiszűrésével) többféle módszerrel hajtható végre. A standardizálás módszere áll a jól ismert standardizált halálhalozási arányszámok, különféle demográfiai, egészségügyi mutatók mellett a konjunktúraindexek, árindexek (egységértékindex-probléma) számítása mögött is, de a standardizálás elve jelenik meg a hedonikus árindexszámításnál vagy a nemek közötti bérkülönbségek (*gender pay gap*) elemzésénél is. Ahhoz, hogy a problémakör megfelelő módszertani eszköztárát helyesen válasszuk ki és interpretáljuk, hasznos adalék, ha megértjük, történetileg hogyan alakultak ki az egyes eljárások, mire reflektáltak a maguk korában, és hogyan fejlődtek. Magyarországon ki kell emelni *Kőrösy József* munkásságát (aki a módszer kidolgozásának és nemzetközi alkalmazásának is az élharcosa volt) az 1940–1950-es évek hosszú ideig ható szovjet örökségén és annak feldolgozásán át egészen a mai sokféle módszertan (ezen belül elsősorban a regressziószámítás standardizálás céljára való felhasználásának) kialakulásáig. A cikk a történeti áttekintés végén azzal a céllal hasonlítja össze a klaszikus standardizálást a regressziószámítással, hogy bemutassa, egyik sem feltétlenül jobb a másiknál, ez a vizsgált probléma jellegétől, a felhasználók felkészültségétől és a felhasználás céljától is függ.

Kulcsszavak: standardizálás, statisztikatörténet, összetett intenzitási viszonyszámok, regressziószámítás, Kőrösy József

Standardisation (comparison of two or more composite ratios/averages, detection of the composition effect and, if necessary, filtering it out) can be performed using several methods. The method/principle of standardisation is behind the calculation of well-known standardized mortality rates, various demographic and health indicators, as well as business cycle indices and price indices (unit-value index problem), but the principle of standardisation also appears in the calculation of hedonic price indices or in the analysis of gender pay gaps. In order to correctly select and interpret the appropriate (relevant) methodological tools for the problem, it is a useful addition to understand how the individual methods developed historically, what they reflected in their time, and how they developed.

* A cikk egy előző változatához *Ferenci Tamás* kollégám sok értékes megjegyzést fűzött, amelyek tanulságait igyekeztem megfogadni, ezúton is köszönet értük. Minden, a cikkben maradt hiba, vitatható állítás az én felelősségem.

In Hungary, the historical development of the methodology is particularly interesting, from the work of József Kőrösy (who was a pioneer in the development and international application of the method) through the long-lasting Soviet legacy of the 1940s and 1950s to the development of today's diverse methodologies (including primarily the use of regression analysis for standardisation purposes). At the end of the historical overview, the article compares classical standardisation and regression analysis, with the aim of demonstrating that neither is necessarily better than the other, as it depends on the nature of the problem under study, the level of preparedness of the users, and the purpose of use.

Keywords: standardisation, history of statistics, composite ratios, regression analysis, József Kőrösy

A magyar statisztikai szakirodalomban (és oktatásban) meghonosodott terminológia szerint a standardizálás módszere két főátlag (vagy általánosabban összetett intenzitási viszonyszám) összehasonlítása, a részhatás (részátlagok, részviszonyszámok különbözőségének hatása) és összetételhatás (a sokaság megoszlásának hatása) megkülönböztetésével, elválasztásával. Az összehasonlítás két vagy annál több csoport (sokaság) párhuzamba állítása, ami egyrészt lehet egy időpontban vagy időszakban különböző csoportok (pl. területek, foglalkozások, ágazatok, intézmények) összevetése, másrészt lehet időbeli változások elemzése is. A keresztmetszeti jellegű összehasonlítások esetében inkább a különbség, az időbelieknél a hányados- (index-) alapú összehasonlítás terjedt el, de a különbségen vagy hányadoson alapuló standardizálás tartalmilag egyenértékű.

Ez a módszertan (legalábbis az egyetemi statisztika tananyagok szerint) a közgazdasági jellegű elemzésekben Magyarországon jóval elterjedtebb, mint pl. angol-szász területen, ahol sok, statisztikával foglalkozó tankönyv egyáltalán nem tárgyalja, vagy csak érdekességként, említés szintjén foglalkozik vele, nemegyszer csak az ún. Simpson-paradoxont kiemelve (ami a standardizálás egy alosa is lehetne megoszlási viszonyszámokra). Példaként hozható erre *Anderson és szerzőtársai (2023)* a magyarországi angol nyelvű oktatásban több helyen használt könyve, amelyben pl. van fejezet az indexszámításról (ez sem igazán jellemző a *business and economic statistics* tankönyvekre), de a standardizálásról nincs, a Simpson-paradoxont a könyv az alapfogalmak között említi¹.

Az egyik bemutató példa a cikkben a férfiak és a nők halandóságának összehasonlítása Budapesten. A 2024-es népességszám (év elején) és a halálozási adatok láthatók az 1. táblázatban. A példa jól szemlélteti az alapproblémát (nem véletlen, hogy a módszer a halandóság vizsgálatára alakult ki, és ma is ez az alkalmazás egyik

¹ Más a helyzet pl. a demográfiai, az orvosi, a biostatisztikai szakirodalommal, ahol a jelenséget részletesen elemzik, de a példaként is bemutatott halálozási arányszámon túl (amit nem lehet megkerülni, tekintettel a módszertan kialakulására és fejlődésében játszott szerepére) a cikkben elsősorban a közgazdasági/biznisz indítatású művekkel, kutatásokkal foglalkozom.

fő terepe). Miközben minden életkori csoportban magasabb (sok csoportban jóval magasabb, akár a duplája) a férfiak halálozási arányszáma a nőkénel, összességében mégis a nők halandósága ér el nagyobb értéket. Az elmúlt 13 évben olyan erős volt az összetételhatás, hogy Budapesten a nők halandósága összességében rendre magasabb volt, mint a férfiaké. A 2012 és 2024 közötti összehasonlító eredmények az 1. ábrán láthatók. A számításhoz minden egyes év férfi-női átlagos népességnagysága adja a standard kormegoszlást, ez az ún. Kitagawa-féle súlyozás.

1. táblázat

**A férfiak, a nők és a halálozások száma,
valamint a halálozási arányszám Budapesten, 2024**
Number of men, women and deaths, and the mortality rate in Budapest, 2024

Életkor, év	Halálozások száma, fő		Népesség száma, fő		Halálozási arányszám, ezrelék	
	férfi	nő	férfi	nő	férfi	nő
0–4	26	19	34 481	32 229	0,8	0,6
5–9	5	0	35 029	33 107	0,1	0,0
10–14	3	0	36 921	34 849	0,1	0,0
15–19	6	5	40 768	39 255	0,1	0,1
20–24	19	5	49 376	48 806	0,4	0,1
25–29	16	17	55 267	54 590	0,3	0,3
30–34	47	15	63 658	62 125	0,7	0,2
35–39	71	34	62 524	60 934	1,1	0,6
40–44	100	62	61 693	62 368	1,6	1,0
45–49	226	98	74 944	77 990	3,0	1,3
50–54	310	170	62 743	66 021	4,9	2,6
55–59	534	264	50 044	55 385	10,7	4,8
60–64	601	361	36 024	43 599	16,7	8,3
65–69	1 005	747	35 448	49 822	28,4	15,0
70–74	1 431	1 263	37 455	59 100	38,2	21,4
75–79	1 432	1 638	26 469	47 438	54,1	34,5
80–84	1 498	2 139	18 823	36 134	79,6	59,2
85–	1 919	4 278	11 333	28 457	169,3	150,3
Összesen	9 249	11 115	793 000	892 209	11,7	12,5

Forrás: saját számítás a KSH adatai alapján.

A szokásos módon számolható a tényleges különbség mellett ($K = 12,5-11,7 = 0,8\%$) a standardizált K' különbség és az összetételhatást mutató K'' különbség is, esetünkben (a nőket viszonyítva a férfiakhoz, mind a két és az átlagos súlyozással is):

$$K'_F = \frac{34481 \cdot (0,6-0,8) + 35029 \cdot (0,0-0,1) + \dots + 11333 \cdot (150,3-169,3)}{793000} = -4,1\%$$

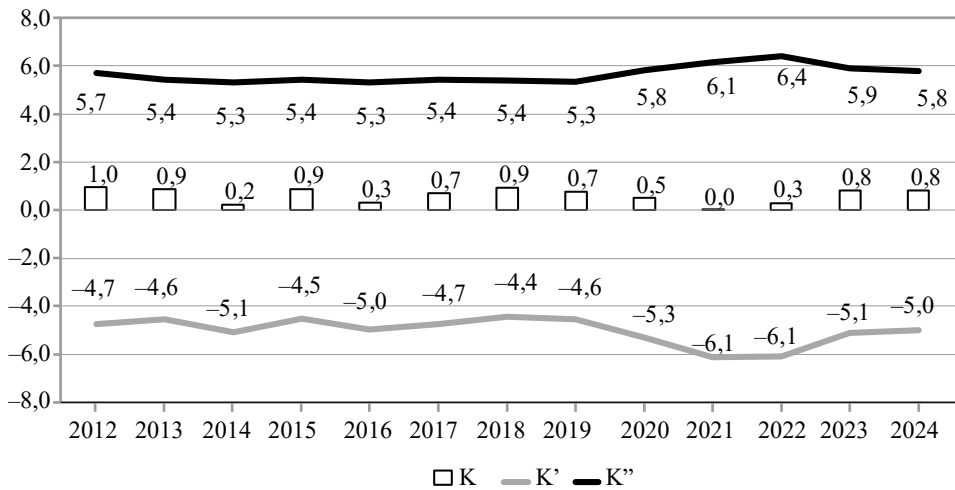
$$K'_N = \frac{32229 \cdot (0,6-0,8) + 33107 \cdot (0,0-0,1) + \dots + 28457 \cdot (150,3-169,3)}{892209} = -5,6\%$$

Az átlagos felbontás: $K = 0,8\%$ $K' = -4,8\%$ $K'' = 5,6\%$.

Azaz a budapesti nők korcsoportonkénti halálozási arányszáma 2024-ben átlagosan 4,8 ezrelékponttal alacsonyabb volt, mint a férfiaké, ugyanakkor a nők esetében magasabb az idősebb (nagyobb halandóságú) népesség aránya, emiatt önmagában (azonosnak feltételezve a korcsoportonkénti férfi-női halálozási arányszámokat) 5,6 ezrelékponttal magasabb lenne a nők átlagos halandósága. Az összetételhatás erősebb, mint a részhatás, ezért bár minden korcsoportban alacsonyabb a nők halandósága, mint a férfiaké, összességében mégis magasabb.

1. ábra

A férfiak és a nők halálozási arányszáma közötti különbség (K), a standardizált különbség (K') és az összetételhatás (K'') Budapesten, 2012–2024
Difference between male and female mortality rates (K), the standardised difference (K') and the composition effect (K'') in Budapest



Forrás: saját számítás a KSH adatai alapján.

1. Történet a tankönyvek, publikációk tükrében

Több magyar tankönyv (pl. *Hunyadi-Vita, 2003*) és a hazai közvélekedés szerint is a standardizálás módszerét Kőrösy József (1844–1906) magyar statisztikus, a Magyar Tudományos Akadémia rendes tagja dolgozta ki. A demográfiával kapcsolatos területeken (halálozási arányszámok, egyéb egészségügygel, orvosi gyakorlatokkal is összefüggő területek) érthető a standardizálás elterjedtsége Magyarországon, hiszen ez a nemzetközi gyakorlat, amire még ráerősít Kőrösy kétségtelenül

jelentős hatása. A gazdaságstatisztikai gyakorlatban való hangsúlyos megjelenést egyrészt az 1940-es évek végétől jelentkező szovjet módszertan, másrészt *Köves Pál* (1925–2021) és *Párniczky Gábor* (1925–1992) munkássága magyarázhatja, akik nemcsak leírták a módszert, hanem új alkalmazási területeket mutattak, és kitartóan igyekeztek beilleszteni mindezt a statisztika általános elméletébe is. A módszer tételes leírása is az ő nevükhöz fűződik (*Köves, 1956; Párniczky, 1956*). Az árindexek számításánál természetesen a nemzetközi gyakorlat is foglalja az összetételhatással, de általában direkt módon nem kapcsolják össze a standardizálással, *unit-value* (egységérték) index problémakör néven jelenik meg. A standardizálás aztán bevonul a számvetési elemzések területére is, ennek tárgyalása is megér egy részletesebb kifejtést. Cikkemben röviden kitérek arra is, hogyan jelenik meg a standardizálás, mint a zavaró/összemosódó tényezők (*confounder*) kiszűrésének egy módszere, és milyen alternatív lehetőségek merülnek fel, illetve ezek a (hivatalos/gazdaságstatisztikai/elemező/modellező) gyakorlatban mennyire jelennek meg.

1.1. Az első korszak: a 19. század vége és a 20. század első fele

A módszertan kialakulásának története – a halálzási arányszámok elemzése területén – meglehetősen kalandos. Itt érdemes arra is kitérni, mennyire állja meg a helyét az a vélekedés, hogy *Körösy József* volt a kidolgozója.

A standardizálás módszere jóval *Körösy* előtt megjelent (*Neison, 1844; Farr, 1859*). *Neison* a foglalkozási halálzási arányszámok elemzése kapcsán – rámutatva az eltérő korszerkezet torzító hatására – kritizálta az átlagos halálzási arányszám használatát, továbbá ő vezette be a direkt és indirekt standardizálás fogalmát². *Farr* nevéhez fűződik a várható halálzási számok, azaz az indirekt standardizálás rutinszerű alkalmazásának meghonosítása a hivatalos statisztikákban. A statisztikai standardizálás alapjait – *Keiding (1987)*, valamint *Keiding és*

² A direkt standardizálás (*CMF*) feltételezi, hogy ismertek a részsokaságok, pl. a korcsoportok arányszámai, amit egy standardizált megoszlással súlyoz. Így készül pl. az egyes megyék standardizált halálzási arányszáma, vagy az egyes országoké is egy standard (pl. országos vagy európai átlagos) kormegoszlással súlyozva. Az indirekt standardizálás (*SMR*) abból indul ki, hogy esetleg nem ismertek vagy megbízhatatlanok a részsokaságok arányszámai, csak a teljes arányszám, valamint a részsokaságok megoszlása ismert. Ebben az esetben a standard sokaság csoportonkénti részviszonyait kell súlyozni a részsokaságban jellemző megoszlással, és ezt a fiktív főátlagot kell hasonlítani a standardnak tekintett sokaság főátlagához. Az indirekt standardizálás egy jellemző csoport (pl. egy foglalkozás) halandóságát hasonlította össze a teljes halandósággal. *Körösy József* a direkt standardizálás módszerét javasolta a nemzetközi alkalmazásokban is, a mai magyar standardizálás leírása – ami a tankönyvekben is megjelenik – ez a megkülönböztetést nem használja. Maga *Körösy* sem alkalmazta a direkt elnevezést, ő a standard kormegoszlással „helyesbített vagy tisztított” halálzási ráták kifejezéseket preferálta.

Clayton (2014) cikkei alapján – korábban fektették le, már a 18. században is léteztek erre vonatkozó alkalmazások.

Körösy József a Nemzetközi Statisztikai Intézet 1891. évi bécsi kongresszusán terjesztette elő javaslatát, amelyben egy általa korindexnek nevezett mutatószámot ajánlott a halandóság nemzetközi összehasonlítására. Ezzel a mutatóval – a népességet³ négy korcsoportra bontva – a korcsoportonkénti (nyers) halálozási arányszámokat standard népességmegoszlással súlyozták. Annak tiszteletére, hogy Svédországban készült az első népességregiszter, a svédet javasolta standard népességmegoszlásnak, de azt is fontos megjegyezni, hogy egy összeurópai kormegoszlást – ezt használják jelenleg is az EU-ban – megbízhatatlannak tartott a sokféle rendszerből való átszámítási nehézségek miatt. Körösy csak később értesült arról, hogy *Ogle* már 1883-ban hasonló módszertant (*Ogle, 1883*) publikált. Az 1893-as chicagói kongresszuson erre is reagálva az *Ogle* módszerét javító témaként tárgyalta a sajtóját. Ugyanebben az évben jelentkezett *Koch* is (*Koch, 1883*), aki némileg bírálta *Ogle* számításait, és erről 1891 után értesítette Körösyt. A 19. században Körösy munkásságának lett a legnagyobb nemzetközi hatása (*Sailé, 1927*), mivel ő védte meg a standardizálás módszertanát, és cáfolta a bírálók érveit, valamint általa érte el a széles körű nemzetközi elfogadottságot, ami ahhoz vezetett, hogy az 1895-ös, bérni Nemzetközi Statisztikai Kongresszuson hivatalossá tették a halálozási arányszámok nemzetközi összehasonlításában.

Összefoglalva:

1. A módszer Körösy József munkássága előtt is már megjelent. „Demológiai” módszertani alapművében (*Körösy, 1892*) hivatkozott Farr és Neison tevékenységére is: Farr indirekt standardizálásával szemben a standard kormegoszlás használatát javasolta, ebben támaszkodott Neison munkásságára, sőt az ő angol kormegoszlási számait használta a saját, Budapestre vonatkozó standard halálozási arányszámok számításához.

2. Körösy munkásságának eredményként a standard kormegoszlás megjelent és elfogadottá vált a nemzetközi összehasonlításokban. A zavaró tényezők hatása elkülönítésének is szellemi előfutára volt, mivel minden elemzésében hangsúlyozta, hogy a halandóság és az olyan változók között, mint pl. a lakásviszonyok, a vagyoni helyzet vagy a vallás kapcsolat áll fenn, amit a kormegoszlás különbözősége elfedhet.

A Körösy által formált gyakorlati, demográfiai és társadalomstatisztikai irányzat a 20. század első felének hazai oktatására is mély benyomást gyakorolt, ezt

³ Körösy által javasolt korcsoportok: 0–1 évesek, 1–20 évesek, 20–50 évesek és 50 évnél idősebbek. Ebben az időben, miután még nagyon magas volt a születés utáni halandóság szintje, a nyers halandóságra leginkább a legfiatalabbak eltérő számaránya gyakorolt hatást. Ezt az 1895-ös nemzetközi ajánlás annyiban módosította, hogy a javasolt korcsoportokat 1 éven aluli, 1–19 évesek, 20–39 évesek, 40–59 évesek, valamint 60 éven felüliekre bontotta, hogy az idősebb népesség súlyát jobban figyelembe vehessék.

igazolja a II. világháború előtti statisztika könyvek tartalma is. *Kenéz Béla* elméleti statisztikáról írott könyvében (*Kenéz, 1903*) még nincs standardizálás fejezet, de a szerző standard számítás néven – szintén Svédország kormegoszlását alapul véve – ismertette a standardizált halálozási arányszám számítását, értelmezését, és példát is mutatott rá. A két háború között a standardizálás módszere a halálozási arányszámok elemzése során bekerült a statisztikai módszertanba, majd más (demográfiai) területeken is megjelent. A *Schweng Lóránd* által jegyzett statisztikai tankönyv (*Schweng, 1944*) egyik fejezete már foglalkozott a viszonyszámokkal, ezen belül az „arányszámokkal” (a későbbi terminológiához hasonlóan megkülönböztette a viszonyszámokat az „igazi” indexektől, de az egyes viszonyszámfajtáknak még némileg más volt a megnevezése: megoszlási, fejlődési viszony- és arányszámok, ez utóbbiak a ma intenzitási viszonyszámoknak nevezett mutatók). Schweng külön alfejezetben vezette le a standardizáláson alapuló különbségek számítását olyan arányszámokon, ahol a nevezőben a népesség nagysága állt (a kormegoszlás a kategóriaképző ismérv), szemléltette a halálozási arányszámok példáját is: alapul a születési arányszám időbeli változását vette Magyarországon, az itt megkülönböztethető tiszta hatással (ez a részhatás), valamint a kormegoszlás változásának hatásával. (Az 1900-tól 1944-ig terjedő időszakban a tiszta születési arányszám is folyamatosan csökkent, de az összetételhatás erre még ráerősített, hiszen ebben a korszakban folyamatosan csökkent a fiatalabb nők aránya, különösen az időszak végén, amikor szülőképes korba kerültek az I. világháború alatt világra jött – jóval alacsonyabb számú – korosztályok.) A fejezet végén Schweng minden hasonló esetre – ahol érdekes az összetételhatás – a standard átlagok használatát javasolta, egy termelékenységi mutató elemzésén keresztül bemutatva a „munkahatályt”. (A munkahatály a szénbányák esetében egy adott időszak alatt egy vājárra jutó kitermelt szénmennyiség volt, ami bányánként, területenként elég különböző lehetett.)

1.2. Standardizálás az 1940-es és az 1950-es években

1948 után hazánkban szovjet hatásra vezették be a változó, változatlan állományú, valamint arányeltolódási indexeket (az elnevezések az orosz eredetű fordításai). Itt ugyan a standardizálás módszere jelent meg (a változatlan állományú azonban a standardizált főátlag/összetett intenzitási viszonyszámok változását mutatja), de más a módszer elhelyezése, és főleg a gyakorlati felhasználása, mint a klasszikus standardizálás esetében. Ahhoz, hogy meg lehessen érteni az 1945 utáni magyar fejleményeket az indexszámítás területén, vissza kell menni ezen indexek társadalmi-gazdasági indíttatásáig. Magyarul is megjelent a gazdasági indexek filozófiáját tag-

laló alapcikk (*Malij, 1949*), ez alapján rekonstruálható, miért került előtérbe a változó és változatlan állományú index: *Malij* ezeket nem a standardizáláshoz kapcsolta, hanem az indexszámításhoz, valamint az ár- és a konjunktúraindexszel állította szembe. *Irving Fisher* egyik művéből indult ki, amelyben Fisher úgy vélte – miután az egyedi árváltozások nagyon különbözőek lehetnek –, hogy az árszínvonal változását az átlagos árváltozás mértéke, az index mutatja. (Az átlag az arany középpút.) *Malij* szerint az index átlagformája (akár ár-, akár konjunktúraindexről van szó) nem a gazdasági jelenségek természetéhez igazodik, hanem a lényegét elfedő mesterséges konstrukció. Fisher életművének jelentős részét tette ki az ideális (minél több szempont szerint jó tulajdonságú) formula keresése, hogy az index a lehető legjobb tulajdonságú legyen, ami *Malij* felfogása szerint „nemcsak elszakad a valóságtól, de azt meg is hamisítja”.

A szocialista statisztikának tehát olyan gazdasági indexeket kellett definiálnia, amelyek:

- i) Nem szakadnak el a tényleges gazdasági folyamatoktól, azokat részterületenként mérik.
- ii) Ezek egységes rendszere jön létre, de nem egy mechanikus átlagolás eredményeként. Azt a ma trivialitásként kezelt kiindulópontot kérdőjelezték meg, hogy az index egy olyan összetett összehasonlító viszonyszám, ahol az aggregálás a különböző mértékegységű termékeket beárazva, értékben összegzi.

Ezek után már meg lehet fogalmazni a tervutasításos rendszerre jellemző indexek legfontosabb sajátosságait.

- a) Az indexanalízis tárgya csakis a tényleges gazdasági jelenség lehet, az indexek mögött ott kell lenni a meghatározott abszolút mennyiségeknek/összegeknek is.
- b) Ezt a lehető legrészletesebb bontásban, részterületenként kell számolni, ami nem csupán a termék/szolgáltatásfajták szerinti bontásokat jelenti, hanem pl. azt is, hogy a folyó áron számolt, a mennyiségi és az árváltozásokat meg kell különböztetni egymástól, és külön kell őket jellemezni.
- c) Az egyes területek összekapcsolásának terepe a változó és a változatlan állományú index. Itt is az aggregát forma az alapvető: bár levezeti a súlyozott (mérlegelt) számtani és harmonikus átlagformákat, ezek szerepe másodlagos, legfeljebb a gyakorlati számítások meghatározására szolgálhat.

A szovjet hatásra megjelent standardizált index ideológiailag és tartalmilag a tervutasításos rendszer logikájához kötődött, amennyiben a tervutasítás nem tényleges tervezést jelent, hanem a hatalomgyakorlás formája a gazdaságban, aminek része a minél részletesebb lebontás, a naturális mértékegységek preferálása. Ez a sztálini párt és a tömegek közötti hajtószíjak modell leképezésének tekinthető a statisztikában. A statisztikai mutatószámok rendszere szempontjából a változatlan/változó

állományú index ezzel kikerült a standardizálás logikai egységéből, és az indexszámítás része lett, mintegy ellenpontozva a valódi indexeket. Ez a módszertani „újítás” nagyon erős, évtizedekig tartó hatást gyakorolt a magyar statisztikusokra, annak ellenére, hogy rögtön a változó/változatlan állományú indexek megjelenése után megjelent a kritikája is (pl. *Drechsler–Köves, 1952*), és az a törekvés, hogy a változó/változatlan állományú indexeket visszailleszték a standardizálás logikájába, illetve megteremtették az indexekkel való kapcsolatot is. Ez azonban eltartott egy ideig: a különbségen és indexeken alapuló standardizálás – mint azonos tartalmú, de különböző technikájú módszerek –, valamint a standardizálás helye az indexszámításban pontos kifejtése csak *Köves és Pármiczky 1981-es, Általános statisztika I–II.* című alaplívében jelent meg teljesen. Köves és Pármiczky az 1950-es évek végétől elvetette a változatlan állományú index elnevezést is, helyette a részhatásindexet használta. Ugyanakkor más tananyagokban, alkalmazásokban a szovjet hatás nagyon erős volt. A szakközépiskolai egy- vagy kétéves, az adott korban színvonalas oktatásnak számító statisztika tananyagaiban még az 1980-as évek végén is az indexek között, változó és változatlan állományú indexnek nevezve, a standardizálástól elszakítva tanították ezt a részt (lásd ennek legnagyobb tanulói létszámhoz eljutó könyvét {*Móroc, 1987*}).

Az 1945 és 1956 közötti statisztikaelmélet és gyakorlat egyik összefoglaló műve a *Péter György* (a KSH elnöke 1948 és 1969 között) nevével fémjelzett statisztikakönyv (*Péter, 1955*). A könyv szerzői között ott van *Rédei Jenő* és *Zala Júlia* is, de közreműködőként megjelent Köves Pál és Pármiczky Gábor is. Az átlagokról szóló fejezet egyik alpontja tárgyalja a standardizálást. Abból indultak ki, hogy a részátlagok megoszlással súlyozott átlaga a főátlag, azaz a részátlagok különbözősége mellett a két sokaság összetétele is hat a főátlag értékére. A mintapélda (amelyet a könyv viszonylag részletesen, alaposan tárgyalt) a halálozási arányszámok összehasonlítása volt. Bár megemlítették, hogy a módszer kidolgozása *Kőrösy József* nevéhez fűződik, a példa nem különböző országok lakóinak halandóságát, hanem – az 1910-es angol adatok alapján – a papok és a nyomdászok halandóságát vetette össze, bemutatva az összetételhatást, ami olyan erős (a papok jóval magasabb arányban található az idősebb, magasabb halandósággal jellemezhető életkori csoportokban, mint a nyomdászok), hogy összességében a papok halandósága magasabb volt. A standardizált átlag (itt halálozási arányszám, ahol a standard kormegoszlás nem valamelyik foglalkozásé, hanem a teljes népességé) „valóban tükrözi a társadalmi osztály helyzetének befolyását a halálozási arányszám alakulására és csak azt tükrözi”.

A papok, a bányászok és a nyomdászok halálozási arányszámainak standardizált összehasonlítása eredetileg *Yule*, majd *Yule* és *Kendall* együtt írt könyveiben is szerepelt (természetesen ott a standardizálás módszerének osztálytartalma ilyen plasztikusan nem jelent meg), de ebben a tankönyvben nem erre hivatkozott a szerző, hanem egy szovjet demográfiai könyvből vette át azt (*Bojarszkij–Suserin, 1952*).

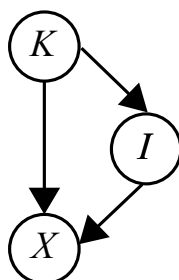
Bojarszkij és *Suserin* érzékletesen mutatta be, hogy a „polgári apologetika” ugyan eltüntette a halandósági statisztikából az osztálytartalmú elemzéseket, de a foglalkozási csoportok halálzási arányszámainak elemzése/közlése „oázisként a sivatagban” megmaradt. Ugyanakkor rámutattak arra, hogy hiába a gazdag (kb. 600 foglalkozást és betegcsoportokat is megkülönböztető) halandósági kimutatás, ha ez sok szempontból a valós osztályellentétek elkendőzésére szolgál. Szerencsére maga a standardizálás módszere (amelyet ismertetnek is) osztálytartalmát tekintve pozitív minősítést kapott, hiszen éppen abban segített, hogy „átlássunk a burzsoá megtévesztések sűrű hálóján”.

A Péter György nevével fémjelzett könyv egy terjedelmes indexszámítási fejezetet is tartalmaz, ennek több alfejezete is tárgyalta a változó/változatlan állományú és arányeltolódási indexet. A változó állományú index viszonyszámok „együttes/átlagos” változását mutatja. Ugyan az alfejezet nem nyilvánította ki, hogy ez a standardizálásnak megfelelő problémakör, a bemutató példában jól látszott, hogy egy vállalatnál a szak- és a segéd munkások átlagbére egyaránt nőtt, ugyanakkor összességében az átlagbér csökkent, és a „látszólagos ellentmondás” feloldása éppen az, hogy a foglalkoztatottak aránya az alacsonyabb átlagbérű segéd munkások felé tolódott el. A változatlan állományú index esetében azt közölte, hogy ez a standardizáláshoz hasonló megoldás, hiszen standard súlymegoszlást használtak. Az alfejezet csak egyféle súlyozást mutatott be (tárgyidőszaki létszámok), de annak nemcsak az aggregát, de az átlagformáit is. Az arányeltolódási index esetében a bázisidőszaki átlagokat tekintette változatlanoknak, és szemléltette azt is, hogy ebben az esetben a változatlan állományú és az arányeltolódási index szorzata kiadja a változó állományú indexet, de a különböző súlyozásokról nem esett szó. Az utolsó alfejezet jó pár területet felsorolt, ahol ezek az indexek eredményesen használhatóak (termelékenység-, önköltségelemzések).

1.3. A standardizálás módszerének kiteljesedése 1956-ot követően

A standardizálás magyarországi történetének egyik alapműve Párniczky Gábor *Statisztikai Szemlé*ben megjelent cikke (*Párniczky, 1956*), amely történetileg, az akkori legújabb eredményeket becsatlakoztatva, gyakorlati vonatkozásokkal együtt komplexen mutatta be a standardizálás módszerét. Ismertette Kőrösy eredeti adatain az alapproblémát, az 1890-es években lezajlott (standardizált halálzási arányszámokra vonatkozó) vitákra is kitérve. Majd általánosította a problémát, itt érdemes a gondolatmenetet szó szerint idézni: „Tekintsünk két (A és B) statisztikai sokaságot, melyeknek egységei egymástól valamely K ismerv szempontjából különböznek. (Ha tehát a két sokaság két adott foglalkozási ágban működő kereső népesség, akkor K a foglalkozást jelenti; ha több város lakosságát tekintjük, akkor K területi ismerv

stb.) A sokaságok időben is különbözhetnek, amely esetben K az időt jelenti. Mindkét sokaság egyedeire vonatkozólag egy bizonyos mennyiségi ismerv átlagos színvonalának tanulmányozását tűzzük ki. Ez lehet például a halálozási arány, kereset, főiskolai képzettséggel rendelkezők aránya stb. Nevezzük ezt az ismérvet X -nek. Bennünket elsősorban az érdekel, hogy az X ismerv általános színvonala szempontjából milyen különbség mutatkozik a két sokaság között (amelyek K -ban különböznek egymástól). Világos azonban, hogy az A és B sokaságra vonatkozó egyes színvonalakat – és így ezek különbözetét is – a két sokaság egyedeinek még több tulajdonsága, ismérve befolyásolja. Számunkra elsősorban olyan ismerv bír fontossággal, amely a) lényegesen befolyásolja X -et, b) nem független K -tól. Az első követelmény kézenfekvő. Ha egy ismerv nincs hatással X -re (vagy csak jelentéktelen hatással van rá), akkor a vizsgálat szempontjából elhanyagolható. Érthető a második követelmény is, mert ha egy ismerv független K -tól, akkor nyilván A és B sokaságok eloszlása ezen ismerv szempontjából többé-kevésbé megegyezik és így ez az ismerv szintén nem zavarhatja vizsgálatunkat. (Legyen I olyan ismerv, amely mindkét követelménynek megfelel. (Például halandósági vizsgálatnál lehet A egy város lakossága, B egy másik város lakossága, K tehát a területi ismerv, X a halálozási arány és I az életkor.) Az ismérvek egymásra hatásának sémáját ez esetben a következőképpen szemléltethetjük:



vagyis K hatása X -re egyrészt közvetlenül, másrészt I ismérven keresztül érvényesül.” (Párniczky, 1956, 435–436. oldal).

A leírás – kimondatlanul, de logikáját és az ábrát tekintve – a *confounding* kimutatásának elvét is tartalmazza. A cikk részletesen ismertette (nagyjából a ma is tanított módon) a standardizálás végrehajtásának menetét, kitért arra, hogy nemcsak a standardizált átlag a fontos, hanem az összetételhatást számszerűsítő (az elemzés másik felét képező) különbség is. Bemutatta az indexösszefüggéseket is (változó és változatlan állományú index), valamint kitért *Evelin Kitagawa* 1955-ben publikált eredményeire is, a Kitagawa-féle súlyozásra (Kitagawa, 1955), valamint a többszörös standardizálásra (több ismerv szerinti egyidejű standardizálás) is.

Köves Pál 1956-ban jelentette meg az első indexszámításról szóló könyvét, amelyben a standardizáláson alapuló indexekre egy 60 oldalas fejezetet szánt.

A fejezet első része a standardizálás problémáját tárgyalta, a Párniczky-cikk kapcsán ismertetett ábrák segítségével megvilágítva a confounding témakört. Mintapéldának az átlagbérek (Y) különbözőségének elemzését választotta különböző iparágakban (X), ahol a „zavaró” (összetételhatást okozó) tényező a dolgozók végzettségi szintje (Z). Az elemzés során a szerző pontosan megfogalmazta, mitől tekinthető zavaró tényezőnek az iskolai végzettség: „Az X ismerv szerinti csoportosítás önmagában nem oldja meg a problémát. Közelebb jutunk a megoldáshoz, ha az egész sokaságot Z ismerv szerint csoportosítjuk. Az így kapott csoportokon belül ugyanis a Z ismerv szempontjából minden egység többé-kevésbé egyforma, tehát a Z ismerv, mint zavaró körülmény nem szerepel. Ezeken a csoportokon belül hajtva végre az X ismerv szerinti csoportosítást, és így számítva ki az Y ismerv átlagos értékét, zavartalanul ki tudjuk mutatni az Y ismerv és az X ismerv közvetlen kapcsolatát.” (Köves, 1956, 60.) A standardizálás elvének általános ismertetése után az időbeli változások felbontási módszerét elemezte az előbb ismertetett logika mentén egy termelékenységi példán (ahol az összetételt két különböző vállalat jelenti {munkahely szerinti összetétel}). A példa végén, az eredmények összefoglalása során – az értelmezést olvasva és Köves Pál egyéniségét figyelembe véve biztos, hogy – enyhe ironia is van a szövegben: a termelékenység változását két tényezőre bontotta, az I' számszerűsítette a tényleges termelékenységváltozás mértékét, ami a szocializmus vívmányainak tulajdonítható, I'' pedig az egyéb tényezők hatását, ami az emberi (vezetői) gyarlóság számlájára írható. Köves Pál a standardizálást és az indexszámítást úgy próbálta összeilleszteni, hogy az eddig szigorúan átlagokon végzett műveleteket intenzitási viszonyszámokra terjesztette ki, bemutatva a kettő közötti tartalmi és technikai különbségeket. A fejezet nagyrészt az egyenes és fordított jellegű mutatókon alapuló standardizálás eredményeinek összefüggéseivel foglalkozott. Ez a későbbi művekből teljesen kimaradt, illetve a Párniczky-cikkhez hasonlóan megjelent a többszörös standardizálás, amire részletes példát is mutatott.

Az első Köves–Párniczky-féle könyvben (Köves–Párniczky, 1960) a standardizálás az átlag és a viszonyszám csoportosítással való kapcsolatáról szóló fejezetében található, ahol a csoportosítás sokaságon belüli szerepét tárgyalták, bemutatva a főátlagok gyakorisággal vagy relatív gyakorisággal súlyozott számtani, vagy értékösszeggel súlyozott harmonikus átlagát, majd ehhez kapcsolódva a szórásnégyzet-felbontást is, bebizonyítva, hogy a teljes szórásnégyzet a belső és külső szórásnégyzet összege. Ez alapján azonban a későbbi művekben (ahol a szórásnégyzet-felbontás a kapcsolatvizsgálat része) explicit nem definiált kapcsolatszorosság-mérő mutatót.

Ezek után következett a részviszonyszámokból történő összetett viszonyszám számolása (nevező megoszlással súlyozott számtani, vagy számlálómegoszlással súlyozott harmonikus átlagforma). Az összes Köves–Párniczky szerzőpáros által

jegyzett tankönyvben (és Köves Pál más írásaiban is) szerepel az átlag és intenzitási viszonyszám összevetése, Köves Pál felfogása szerint ugyanis az átlag csak akkor számít intenzitási viszonyzámmak, ha az értékösszegnek van tárgyi jelentése. Pl. az átlagbér felfogható intenzitási viszonyzámmak, az átlagos életkor (ami egy rendes átlag) ugyanakkor nem.

Ezekhez a fogalmakhoz kapcsolódott a standardizálás módszerének ismertetése. Ugyan a bevezetőben megemlítették, hogy sok olyan mutató van, amit érdemes időben és/vagy térben összehasonlítani, köztük pl. a halálozási arányszámot is, de a mintapélda (különböző fajta szőlők) átlagára vonatkozott. Ez alapján két város piacán 4 szőlőfajtát vizsgáltak didaktikusan, és bár az egyik városban mind a 4 szőlőfajta átlagára alacsonyabb, összességében mégis magasabb (mert a drágább/jobb minőségű szőlőfajták aránya nagyobb a felvásárlásban). Megemlítették, hogy a két város piaca helyett lehetne egy város piaca bázis- és tárgyidőszakban is az összehasonlítás, de a tárgyalás itt végig területi alapon történt, részletesen kifejtve a tartalmat, de emellett a lehetséges súlyozási lehetőségeket is bemutatva (a standard súlyok lehetnek mind a két sokaságé vagy akár a két sokaság összességéé, átlagos súly). A szerzők a „vizsgált jelenség színvonalának különbözősége, valamint az összetétel különbözősége” elnevezést adták a két hatásnak. Ezek után általánosították a standardizálás módszerét, megadva az általános képleteket, és csak ezután említették Kőrösyt – az ő nevéhez fűződik a halálozási arányszámok nemzetközi összehasonlítása –, majd példákkal is illusztrálva elemezték a halálozási arányszám változását Magyarországon. Noha a fejezetet azzal zárták, hogy a standardizálást a halálozási arányszámokon „találták fel”, ugyanakkor számos egyéb területen alkalmazható, pl. az egyik legfontosabbnál, az indexszámításnál.

Az indexek tárgyalását ebben a cikkben csak a standardizálás szempontjából veszem szemügyre. Az első Köves és Párniczky által írt könyv a különbségen alapuló standardizálás után részletesen tárgyalta az indexeket (érték-, ár-, volumen-indexkör, aggregát, átlagformák, összefüggések, indexsorok, alkalmazások), majd az utolsó alfejezetben tért ki „indexszámítás standardizálás alapján” címmel az akkoriban változó és változatlan állományú indexeknek nevezett, manapság fő-átlag- és részátlagindexnek nevezett mutatókra, gyakorlatilag megismételve a különbségeknél kielemezteket. Ezek után ismertette az összetétel- (vagy arányeltolódási) indexet, és a 3 index közötti összefüggésrendszert.

Az utolsó alfejezetben tértek ki arra, hogy az indexszámítás standardizálás alapján miért inkább az indexszámításhoz és nem a standardizáláshoz kapcsolódik. A fő indok, amit egy átlagolás példán keresztül mutattak be, az az árindex és a változatlan állományú index azonos tartalma, de a szerzők elsősorban a különbségeket hangsúlyozták. Ennek az alfejezetnek az lett az igazi erőssége, hogy a stan-

dardizáláson alapuló indexek többféle gyakorlati felhasználását is bemutatták (átlagárváltozás, termelékenységelemzés, önköltségszámítás, átlagkeresetek elemzése), és két fontos dologra is felhívták a figyelmet:

- i) Az elemzett területtől függetlenül azonos módszertani tartalmat, számításokat kell használni.
- ii) Az összetétel változása, ennek kimutatása is lehet fontos elemzendő terület, nemcsak a részek átlagos változásának kimutatása (lehet pl. átlagos önköltséget csökkenteni úgy, hogy az egyes munkahelyeken nem mérséklődik az átlagos önköltség, csak eltolódik a termelt mennyiség a kisebb fajlagos önköltségű munkahelyek felé, vagy emelkedhet úgy az átlagbér, hogy minden beosztásban marad ugyanaz az átlagbér, csak elmozdul a foglalkoztatottak szerkezete a magasabb bérszínvonalú beosztási csoportok felé).

Az alfejezet végén bemutatták, hogy nemcsak a nevező, hanem a számláló változatlanosságát feltételezve is lehet változatlan állományú indexeket számítani (pl. a termelékenység változása esetében nem a munkaidő vagy foglalkoztatottak száma, hanem az érték az állandó), ez aztán több ágazati gazdaságstatisztikai részterületen el is terjedt.

Érdeemes megnézni, hogyan alakult a standardizálás kielemezése az utolsó Köves–Párniczky-féle tankönyvben, milyen változásokat jelentett kb. 20 év állandó tananyagfejlesztés (*Köves–Párniczky, 1981*). Az egyes témakörök tárgyalásának sorrendje módosult. (Az asszociációs kapcsolatoknál megjelent a több ismérv szerinti kapcsolatszorosság mérése, Yule és Kendall könyvére hivatkozva a közvetett kapcsolat léte, annak kimutatásának lehetősége.)

A középértékeknél megmutatkozott a csoportosítás és átlag viszonya, de a sztochasztikus kapcsolatot itt még csak a részátlagok különbözősége segítségével érzékeltették. A következő, szóródásszámítás fejezet utolsó alfejezete elemezte a szórásnégyzet-felbontást, a vegyes kapcsolat vizsgálatát. A teljes indexszámítási fejezetet követve a leíró statisztika legvégén jelent csak meg a standardizálás témaköre (a főátlagok összehasonlítása standardizálás segítségével). A problémakör vázolója után a különbségen alapuló elemzés következett, a mintapéldában a férfiak és nők átlagkeresetének különbségét vizsgálták, ahol a szakképzettség szerinti összetétel hatását szűrték ki. Az első kiadáshoz hasonlóan a K' (részátlagkülönbség) tárgyalása végén hivatkoztak Kőrösyre, és egy teljes, kidolgozott példát is hoztak, Magyarország és Jugoszlávia halandóságát összevetve. Részletesen elemezték a K'' összetételhatás-különbség számítását, értelmezését, a súlyozási kérdésekre is kitérve. Ezt követték a főátlag-, részátlag-, összetételhatás-indexek, ahol tulajdonképpen megismételték az értelmezéseket. A fejezetet a két indexkör közötti összefüggésekről szóló alfejezet zárta, ami jóval pontosabb volt a 20 évvel azelőtti változatnál. Az árváltozás példáján a p, q, v adatokat a V, B, A adatoknak feleltették meg, bemutatva az átlagárváltozás felbontását az árak átlagos változása

és az összetételhatás eredőjére, valamint az értékváltozást az ár- és mennyiségi változások kiindulópontjára, rámutatva, hogy az ár- és a részátlagindex ugyanaz. Azt is prezentálták, hogy az átlagárváltozás az ár- és az összetételhatás-index szorzata összefüggés mellett a volumenindex is felbontható a természetes volumenváltozás és az összetételhatás-index szorzatára, azaz az összetételhatás is mennyiségi változás, a mennyiségekben bekövetkező szerkezeti változás (a nagyobb használati értékű termékek felé tolódik a kereslet).

Az alfejezet végén egy további érdekes példán keresztül rámutattak, hogy a két indexkör nem csupán az olyan árváltozások elemzésénél használható, ahol a termékek/szolgáltatások mennyisége természetes mértékegységben is összegezhető. Az értékindexkör tagjai minden esetben számolhatók, ha a standardizálás értelmes (míg ez fordítva nem igaz, csak az előbb említett természetes összegzés értelme esetén). Egy termelékenységi példán (egy főre jutó termelés) vázolták fel azt az összefüggést, hogy az A aggregátum változása 3 részre bontható: a B értékek összegének változására, a viszonyszámok átlagos változására és az összetételhatásra. (Tehát a termelés 3 ok miatt változhat $\{I_A\}$, ha változik a foglalkoztatott létszám $\{I_B\}$, ha változik átlagos mértékben a részek megkülönböztetésre használt csoportokban a termelékenység $\{I\}$ és ha változik az összetétel $\{I^*\}$: annak hatása, hogy eltolódhat a létszám a termelékenyebb/kevésbé termelékeny csoportok felé).

Köves és Párniczky könyve – ismereteim szerint a magyar statisztikai irodalomban először – összekötötte a standardizálás problémakörét a regressziószámítással. A többváltozós regressziószámítás minőségi ismérvek alkalmazásáról szóló részében volt egy példa, ahol a foglalkoztatottak keresetét magyarázták az életkorral, szakképzettséggel, nemmel, és ennek kapcsán szerepelt, hogy a minőségi ismérvekhez tartozó paraméterbecsléseket hasonló módon lehet értelmezni, mint a standardizálásban a részhatást számszerűsítő különbséget, hiszen a többi tényező változatlanóságát feltételezzük, azaz kiszűrjük a hatásukat ebből az összehasonlításból. Azt is említették, hogy a regressziószámítás az általánosabb módszer, hiszen több változó hatását is elkülöníthetjük, valamint hibaszámításra is ad lehetőséget, ugyanakkor 1981-ben a hátrányának rótták fel, hogy számításigényes, kézzel nem lehet realizálni, mint a standardizálást.

A továbbiakban a tananyagok változásának illusztrálására csak a „Közgáz” (Marx Károly Közgazdaságtudományi Egyetem, Budapesti Közgazdaságtudományi Egyetem, Budapesti Közgazdaságtudományi és Államigazgatási Egyetem, majd Corvinus Egyetem) tananyagait használom, ahol a következő tankönyvíró generáció (Hunyadi László, Mundruczó György, Kerékgyártó Györgyné, Vita László) az indexszámítás és a regresszió területén is kiemelkedő felkészültségűek voltak Magyarországon. Bár az ő tollukból is több könyv, jegyzet jelent meg 25 év alatt, csak egyet (Hunyadi-Vita, 2003) hasonlítok össze az eddigiekkel. A könyvben a standardizálás és indexszámítás egy fejezetben (*Összehasonlítás*

standardizálással és indexszámítással) jelent meg, először a standardizálás, aztán az indexszámítás, és ezt követi a két indexkör összefüggéseiről szóló alfejezet. Az elemzés itt is az intenzitási viszonyszámokra vonatkozott (a szerzők az átlagot tényleg ennek speciális esetének fogják fel). A tárgyalásmód tartalmát tekintve nagyjából megegyezik a Köves–Párniczky-féle könyvvel, a közlés tömörebb, kiforrottabb, például úgy, hogy a különbség és hányadoselemzésnél plasztikusabban jelent meg, hogy ugyanarról van szó, csak nem eredeti mértékegységben, hanem relatív módon történt a felbontás. A halálozási arányszám elemzése is felbukkant egy példán keresztül tipikus alkalmazási területként, annak megemlítésével, hogy ez alapozta meg a módszert. A standardizáláson alapuló indexek esetében, új vagy újszerű az eddigiekhez képest:

- i) Az I'' index árnyalt értelmezése, amennyiben ez „nem egyszerűen az összetétel-változás tényét, hanem annak valamely összetett intenzitási viszonyszám változására gyakorolt hatását fejezi ki”. Ez azt jelenti, hogy nem elég a különböző összetétel, ennek érvényesülnie is kell, és ehhez kell, hogy a pozitív és negatív hatások ne semlegesítsék egymást,
- ii) Ezzel összefüggően precízen felsorolták, milyen esetekben lehet $I = I'$.
- iii) Egyértelmű, hogy a standardizálás módszere a „tisztá” és a „közvetett” hatást különíti el, és az $I'' = 1$ eseteinek végiggondolásával elemezni lehet, milyen esetekben nem jelentkezik közvetett hatás a vizsgált közgazdasági-társadalmi problémánál.

Az indexszámítás fejezet utolsó (nagyon tömör) alfejezete foglalkozott a két indexkör összefüggésével, kifejezetten az árváltozások olyan termékkörön való számszerűsítésével, ahol a mennyiségek naturáliában is összegezhetők. Érdemes megemlíteni, hogy az árindexek számítása esetén az összetételhatás-problémakör komplex elemzése is megjelenik (pl. két indexkör összefüggése mellett a unit-value) az egyik szerző egy másik művében (*Hüttl-Vita, 2004*) ami – véleményem szerint – ma is az indexek/standardizálás témakörben a legjobb, elméleti igényességű, de gyakorlati célra készült gazdaságstatisztikai jellegű mű Magyarországon.

Az indexek elnevezésének 50 éves története során eleinte volt a változatlan és változó állományú, valamint az arányeltolódási index, majd következett a főátlag, részátlag-, összetételindex, végül az összhatás-, a részhatás- és az összetételhatás-index terminológia vált elfogadottá. Az elnevezések aszerint változtak, ahogy a szerzők az egyes mutatószámok tartalmát és nem a számolási módját szerették volna minél pontosabban megragadni.

Érdemes felidézni *Yule és Kendall* Bevezetés a statisztika elméletébe című könyvének tárgyalási módját e témakörben. Ez a tankönyv nagyon sok kiadást megért, többször átdolgozott alapkönyv volt évtizedeken keresztül, és miután magyar fordításban is megjelent (*Yule–Kendall, 1964*), ezért a magyar elméletre és

gyakorlatra is hatott. Két, egymástól távoli helyen jelent meg benne a standardizálás gondolata, mindkettő a kapcsolatvizsgálat látszólagos jellege témakörben. Az asszociációs kapcsolat esetében feltűnt a látszólagos asszociáció fogalma. Ez tulajdonképpen a Simpson-paradoxonnak is keresztelt jelenséget fedí – a szerzők nem így hívták, mivel *Yule (1911)* kb. negyven évvel korábban leírta már, amit *Simpson (1951)* –, azaz egy összességében számolt megoszlási viszonyszám lehet kisebb/nagyobb, mint bármelyik részenkénti az eltérő összetétel miatt. Itt hozták be a híres példájukat, miszerint a nyers halálozási arányszám a bányászok esetében alacsonyabb, mint a papoknál, de nem az egészségesebb munkakörülmények, hanem a kormegoszlás különbsége miatt, és csak a 65 éves kor alatti és feletti bontás szerepelt (hogy a Yule-féle 2×2 -es keresztábra problémájára lehessen ezt is visszavezetni), meg sem említve az akkor már sok évtizede bevett nemzetközi összehasonlításokra használt több korcsoportos standardizálási gyakorlatot.

A problémakör még egyszer megjelent a korrelációs együtthatóval kapcsolatos különböző tételek között, ezen belül is a súlyozott számtani közép vonatkozó gondolatmenetben⁴. Azért a korrelációs kapcsolatnál bukkant fel, mert Yule és Kendall levezetései szerint a súlyozás csak akkor téríti el a súlyozott átlagot a súlyozatlanól, ha korreláció áll fenn a súlyok és a változó értékei között. (Ha pozitív a korreláció, akkor a súlyozott átlag nagyobb, ha negatív, akkor kisebb lesz, mint a súlyozatlan átlag.) A mintapélda ebben az esetben is a halálozási, illetve a születési, házassági arányszám volt, részletes kormegoszlás alapján. Bemutatták a nyers és standardizált halálozási arányszám számítását, és megemlítették, hogy a módszer általánosítható más problémákra is (pl. iskolai tanulók iskolánkénti átlagos testmagasságának összehasonlítása standard nem és kormegoszlás feltételezésével).

2. A standardizálás alternatívái

A standardizálás problémakörének több módszer szerinti vizsgálata a magyar szakirodalomban talán *Ferenci Tamás* biostatistikus munkáiban jelenik meg a leginkább. Az egyes módszerek közös alapját más szerzők is, pl. *Pintér és Rappai*

⁴ A súlyozott számtani átlag a középértékek tárgyalásánál annyiban bukkant fel, hogy ha az adatok számossága nagy, akkor az adatok összegzése nehézkes, ezért egy gyorsító, közelítő eljárás lehet, ha osztályközös gyakorisági sort készítünk, és az osztályközöket annyiszor adjuk össze, ahány eleme van az osztályköznek, azaz csak a gyakorisággal súlyozott osztályközép jelenik meg korábban.

(2007) használta, abból kiindulva, hogy a *ceteris paribus* (a többi tényező változatlanágát feltételező) elv több módszer közös filozófiája, idetartozik a regressziószámítás, a standardizálás (és az indexszámítás is).

Ferenci Tamás az összetételhatást a leggyakrabban *confoundinghatásnak* (összszemosódás) nevezi, de az összetételhatás kifejezés is megjelenik a leírásaiban. A *confounding* Ferenci megfogalmazásában „a vizsgált tényezőbeli eltérés (a becsülni kívánt hatás) összemosódott más eltéréssel vagy eltérésekkel” (Ferenci, 2021a). A problémakör megoldásaira (a *confoundinghatások* kiszűrésére) Ferenci 5 módszert sorolt fel (Ferenci, 2021b):

1. Kísérleti jellegű vizsgálatnál a véletlenség (randomizáció) biztosítása. Ennek az a lényege, hogy az összehasonlítandó két csoportba (pl. kap-e gyógyszert vagy sem) véletlenszerűen kerülnek be a vizsgálandó egyedek. Az összes további módszertől eltérően ennek az a legnagyobb előnye, hogy nem kell feltevés arra, mi okozhat összetételhatást. A gazdasági vizsgálatok esetében a kísérleti megfigyeléseknek elég körülhatárolt a szerepe (bár a viselkedési közgazdaságtan terjedésével jelentősen nőtt), de azért a gazdasági-társadalmi vizsgálatok döntő része megfigyeléses jellegű.
2. Rétegzés. Az összetételhatást okozó szempontok szerint részekre bontjuk a sokaságot/mintát, és a rétegekben külön-külön végezzük el az elemzést (vizsgáljuk a részhatást). A módszer hátránya, hogy több összetételhatást okozó réteggépző ismérv kombinációi szerint vizsgálva a sokaságot/mintát, az egyes ismérvváltozatok száma összeszorozódik (ellentétben az interakciók nélküli regresszióval, ahol dummyváltozókat használva „csak” összeadódik), azaz a rétegek számossága áttekinthetlenné teheti az elemzést, amihez az egyre kisebb elemszámok is hozzájárulnak.
3. Standardizálás. Ezt a módszertant részletesen szemügyre vettük tartalmában és történetiségében is. Ferenci Tamás leírása is a halálozási arányszámot választotta alapul, ami az összetételhatás-kezelés egyik általános módszere.
4. Propensity score (kezelési valószínűség) módszere (feltörekvő, de a regressziószámításnál ritkábban használt). Megfigyeléses problémáknál szokott felmerülni, hogy a kezelést kapó és a kontrollcsoport egyedei nem egyforma valószínűséggel szerepelnek az egyik vagy másik csoportban, ami az összehasonlítást torzítható összetételhatás. A *propensity score* tulajdonképpen a randomizáció szimulálása, logisztikus regresszióval becsülik a kezelési valószínűségeket az egyedi megfigyelésekre, és a hasonló score-ral rendelkező kezeltlen és kezelt megfigyeléseket párosítják, majd az így kiegyensúlyozott mintán végzik az összehasonlítást.
5. Regressziószámítás. Ferenci a megfigyeléseken alapuló módszerek esetén egyértelműen ezt tekinti a legjobb megoldásnak. Cikkemben is kifejtem, ösz-

szeheonlítva az eredményeit más eljárásokkal. A standardizálás és a regressziószámítás előnyeire és hátrányaira a cikk végén mintegy összegzőként visszatérek még.

A fenti csoportosításhoz gyakorlati alkalmazási kiegészítést tennék, amelyek gazdagíthatják az összetételhatás kiszűrése kapcsán felvázolt képet.

A rétegzés és a standardizálás összekapcsolódik egymással (ez Ferencinél is megjelenik természetesen, a rétegzéssel „szétszedett” eredeti sokaságot rakja össze a standard súlyok segítségével). A halálózási arányszámok használata mellett talán ez a statisztikai gyakorlatban leggyakrabban megjelenő standardizáláson alapuló eljárás. Sok konjunktúraindex – pl. a BUX-index vagy az Eurostat által harmonizált üzleti és fogyasztói bizalmi indexek is – I' (részhatásindex) típusúak, mert az összetétel megváltozása nem számít konjunkturális változásnak.

Érdeemes ennek a témának a kapcsán felidézni a lakásépítési költségindex számításának történetét a Központi Statisztikai Hivatalban (KSH). A hivatal 1975-től kezdve gyűjtött adatot az épített új lakások költségeiről a kivitelező szervezetektől, elsősorban az állami építőipari vállalatoktól. (Az 1975 előtti és utáni index számításáról lásd *Drechsler–Kerekes, 1974.*) Ebben az esetben – mint a lakásárindex típusú adatoknál – a minőségváltozás kiszűrése a feladat, mert az jelenti az összetételhatást. A KSH megfigyelési rendszerében 1975-től a ténylegesen felmerülő költségeket vizsgálták meg, és a megfigyelt lakások az összes új épített lakás mintegy felét fedték le. A használt építőipari árindex az ún. komponensmódszerrel készült, ami a minőségi változásokat olyan módon igyekezett kiszűrni, hogy nem a heterogén épületeket/lakások árát figyelte, hanem a jóval homogénebb komponensekét. Ezen az adatbázison mutatta be egyébként később Párniczky a hedonikus árindex számítását is (lásd a cikkben később), de ezt akkor még nem vezették be a gyakorlatban. A rendszerváltás után, 2000-től – addigra az addigi adatgyűjtés a lakások egyre kisebb körét fedte le – alakult ki az új gyakorlat. Ez már a rétegzés–standardizálás következetes alkalmazásának tekinthető (*KSH, 2000*), amikor gondos elemzőmunka után 28 rétegre bontották a Magyarországon épített lakásokat (területi hovatartozás, településtípus, lakástípus, lakások minősége szempontok kombinálásával), és a rétegenként számolt költségek alapterülettel súlyozott/standardizált átlagának változása alapvetően az index. Ma is ezt a megoldást használják.

A minták átsúlyozása a rétegzés–standardizálás egy speciális, de gyakori alkalmazási területe. Ennek kiindulópontja a nem arányosan rétegzett minta esete. Ebben az esetben a minta reprezentativitását direkt „elrontjuk” (általában azért, hogy a nagyobb szóródású rétegeket az arányoshoz képest túlreprezentáljuk), de ezt a súlyozásnál rendbe kell hozni. Technikailag ez éppen a standardizálás logikáját jelenti, nem a mintabeli, hanem a sokasági arányokkal súlyozzák a rétegenkénti becsléseket. Ennek általánosítása egy nem reprezentatív minta átsúlyozása. Rétegenként az egyszerű véletlen minta szabályai szerint történik a becslés, és ezt a N_j/n_j súlyokkal kell

súlyozni, ahol N a sokasági, n a mintaelemszám, és $j = 1, 2, \dots, M$ a rétegek. A súlyozás egész bonyolult is lehet, mert több szempont szerint, azokat kombinálva a rétegek száma nagyon nagy, ahol az N_j -k nem is feltétlenül ismertek, és az egyes cellákban is sok lehet a nulla, vagy nagyon kicsi az elemszám. Ezért különböző iterációs eljárások alakultak ki a konkrétan felmerülő súlyozások megoldására. A statisztikai hivatalokban, közvélemény- és piackutató cégeknél ez egy gyakori probléma, még ha nem is nevezik meg, hogy ez is a standardizálás alkalmazásának egy esete.

Az elmúlt években még egy alternatív lehetőség merült fel: a gépi tanulós módszerek standardizálásra való alkalmazása. Ezek általában a regressziószámítással állíthatók párhuzamba, de általánosabb eljárások annál, sokféle (nemcsak lineáris) kapcsolatrendszeret képesek számszerűsíteni, amelyeket előre nem is kell feltételezni.

A később még részletesebben is bemutatott nemi bérkülönbség elemzése területén prezentált egy gépi tanulós alkalmazást és a regressziós eredményekkel való összehasonlítást *Takács (2021)*.

2.1. A regressziós standardizálás

Az összetételhatás regressziószámítással való kiszűrésének (vagy éppen kimutatásának) alap gondolata, hogy az összehasonlítandó két csoportban külön, vagy egy modellben (de akkor a kétértékű megkülönböztetést dummyváltozóval szintén beépítjük a modellbe) kell szerepeltetni az összetételhatást okozható ismérveket (ez lehet mennyiségi ismérv, mint az életkor, vagy minőségi, mint a nem, területi elhelyezkedés), és ezek hatását becsüljük a két csoportban összehasonlítandó mennyiségi ismérvre. Az összetételhatás kiszűrésének ilyenkor az az alapja, hogy egy többváltozós parciális regressziós együttható a ceteris paribus hatást számszerűsíti.

A többváltozós lineáris regresszióban a tiszta és zavaró hatások megkülönböztetését és számszerűsítését a paraméterek szokásos értelmezése mellett többféleképpen lehet interpretálni. Megmutatható pl. (ez a parciális korrelációs együttható számításának az alapja is), hogy kiszűrve Y -ból és X_i -ből is a többi tényező hatását (a regresszió maradéktagjának segítségével) a két új változóval számolt lineáris regresszióban az X_i -hez tartozó maradék tényező regressziós együtthatója megegyezik az eredeti többváltozós modell X_i -hez tartozó parciális regressziós együtthatójával (bizonyítását 3 változós esetre pl. *Maddala, 2004, 182–184.*, illetve általánosítása a könyv függelékében). Az útelemzésnek nevezett összefüggésrendszer is idekapcsolódik, ennek egyszerű bemutatása megtalálható *Hajdunál (2001)* a paraméterek értelmezése címszó alatt.

Az összetételhatás regressziószámítással való kiszűrésének Magyarországon talán két legismertebb területe a lakásárindex (tisztá árváltozás indexe) és a férfi-női átlagkeresetek közötti (összetételhatástól megtisztított) különbség (*gender gap*).

A KSH-ban is használt lakásárindexet 2011 óta publikálják (visszavezetve az előző évekre is, lásd *Horváth–Székely, 2009; Székely, 2020*). Ez az árindexszámítás az ún. hedonikus árindexszámítás elvén, azaz regressziószámítással történik. A hedonikus árindex nem csak a lakáspiacon használható, a lényege, hogy a termék árát a minőségi jellemzőivel egy regressziós összefüggéssel magyarázzuk, ami tulajdonképpen a hasznossági függvény becslése. Az első erre vonatkozó leírás Magyarországon szintén Párniczky nevéhez fűződik (*Párniczky, 1982*), ezért az ő szóhasználatát, jelöléseit használom.

Feltételezzük, hogy egy termék árát különböző jellemzői magyarázzák (a jellemzők lehetnek mennyiségi ismérvek vagy minőségek, ebben az esetben a szokásos módon dummyváltozókat alkalmazunk). Feltételezzük, hogy a termék ára (Y) a jellemzők (X_i -k) lineáris függvénye:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_i \cdot X_i + \dots + \beta_k \cdot X_k + \varepsilon$$

Jelöljük a becsült paramétereket b_i -vel és az egyszerűség kedvéért a becslést írjuk továbbiakban tömör formában az alábbiak alapján (vastagítva jelezzük a vektorokat, a T betű a transzponálást jelöli):

$$\mathbf{X}^T = [1, X_1, X_2, \dots, X_k] \quad \mathbf{b}^T = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]$$

A becsült regressziókra a gyakorlati alkalmazás során elvárják (illetve saját gyakorlati modellezési tapasztalatom alapján törekszem ezekre), hogy a magyarázó változók legyenek szignifikánsak, a modell magyarázóereje legyen viszonylag magas, a véletlenre tett feltételek teljesüljenek, de olyan kompromisszummal, hogy ha több modellt becsülünk (pl. külön modellek vannak a két időbeli vagy térbeli sokaságra), akkor ugyanazok legyenek az egyes modellek magyarázóváltozói.

Az árváltozások megfigyelése esetében két időszak van, a 0-val jelölt bázis, és az 1-gyel jelölt tárgyidőszak. Legyenek X_0 a bázisidőszaki, X_1 a tárgyidőszaki minőségi jellemzői a terméknek.

A) Külön-külön becsülve a bázis- és tárgyévre a regressziós összefüggést, felírható a szokásos módon a regressziós összefüggés realizálódása a bázis- és tárgyévre:

$$\hat{Y}_0 = \mathbf{b}_0^T \cdot \mathbf{X}_0 \quad \text{és} \quad \hat{Y}_1 = \mathbf{b}_1^T \cdot \mathbf{X}_1$$

Legyenek az X_i változók átlagos értékét tartalmazó vektorok a bázis- és a tárgyidőszakra $\bar{\mathbf{X}}_0, \bar{\mathbf{X}}_1$.

Ebben az esetben az árindexnek (vagy részhatásindexnek) megfeleltethető formálisan Laspeyres–Paasche-típusú indexek is értelmezhetők.

$$I_P^L = \frac{\mathbf{b}_1^T \cdot \bar{\mathbf{X}}_0}{\mathbf{b}_0^T \cdot \bar{\mathbf{X}}_0} \quad I_P^P = \frac{\mathbf{b}_1^T \cdot \bar{\mathbf{X}}_1}{\mathbf{b}_0^T \cdot \bar{\mathbf{X}}_1}$$

Miután az átlagár- (összhatás-) index is számolható, az összetételhatás-index is közvetve megadható.

- B) Egy másik megoldás, ami kevésbé a standardizálás–indexszámítás, hanem inkább a regressziószámítás szellemiségét követi, ha egy modellbe helyezzük a két időszak egyesített adatait, és egy plusz dummyváltozó ($X_{k+1} = 0$, ha bázis-, 1, ha tárgyidőszaki a megfigyelés) az időtényező. Ebben az esetben X_{k+1} változó becsült b_{k+1} paramétere éppen a többi tényező változatlanóságát feltételező árszintváltozás mértékeként értelmezhető eredeti mértékegységben, de ebből (egyfajta) árindex is egyszerűen képezhető.

Az ún. *gender pay gap* (nemek közötti bérkülönbség) problémakör szintén egy standardizálási probléma. A férfiak és a nők közötti átlagbér különbségét ebben az esetben is az összetételhatásra (a férfiak és nők aránya más lehet a jobban/rosszabbul fizetett területeken) és a részhatásra (azonos feltételek mellett is kereshetnek a nők átlagosan kevesebbet, ez a bérdiszkrimináció). A regresszióalapú standardizálás talán legismertebb formája az ún. Blinder–Oaxaca-felbontás (Blinder, 1973; Oaxaca, 1973, illetve magyar adatokon Galasi, 2001). Ez a regressziós módszer (az eddig jelöléseket használva) a kereseteket függő változóként, illetve annak logaritmusát, legyen ez az Y , magyarázza releváns tényezőkkel (mint az életkor, munkatapasztalat, iskolai végzettség, beosztás, ágazat stb., legyenek ezek X_i -k.) Külön regressziós becslés készül a férfiakra és a nőkre, amelyekbe beírjuk X átlagos értékeit, akkor visszakapjuk Y átlagos értékét:

$$\bar{Y}_F = \mathbf{b}_F^T \cdot \bar{\mathbf{X}}_F \quad \text{illetve} \quad \bar{Y}_N = \mathbf{b}_N^T \cdot \bar{\mathbf{X}}_N$$

Átrendezve és bővítve a fentieket a következő összefüggést kapjuk:

$$\bar{Y}_F - \bar{Y}_N = (\bar{\mathbf{X}}_F - \bar{\mathbf{X}}_N) \cdot \mathbf{b}_F + \bar{\mathbf{X}}_N \cdot (\mathbf{b}_F - \mathbf{b}_N)$$

Az elv ugyanaz, mint a hedonikus árindexnél, csak ebben az esetben nem hányadosokat, hanem különbségeket írunk fel (amennyiben nem a béreket, hanem azok logaritmusát magyarázzuk X -ekkel, akkor tulajdonképpen az is a hányados felbontásához vezet, amikor visszatranszformálunk eredeti mértékegységre), és a triviális bővítés után a standardizálásnak megfelelő felbontást kapjuk.

Az átlagos keresetkülönbség két részre bontható, az első rész az összetételhatás, a második a részhatás, és az összetételhatás kimutatásánál változatlanak tekintjük a férfiakra jellemző regressziós paramétereket, míg részhatás-kimutatásnál a nőkre jellemző összetételt vesszük standardnak. Itt is lehetne fordítva súlyozni, vagy akár a Kitagawa-átlagsúlyozásnak megfelelő megoldást választani. A hedonikus árindex esetén nem szokták ezt a bővített formát használni, noha lehetne, mert az árindex (a tiszta árváltozás hatása) az érdekes, az összetétel megváltozása kevésbé, bár közvetve ott is megadható az I^* index.

A hedonikus árindex értelmezésének illusztrálására egy saját számítást mutatok. Egy használt autót hirdet, internetes portál adatai alapján az öt leggyakoribb

elektromosautó-típus (BMW, Nissan, KIA, Tesla, Volkswagen) autóinak átlagos ár-változása a kérdés 2025 januárjától 2026 januárjáig. (2025-ben $n = 2470$, 2026-ban $n = 2904$ a megfigyelések száma.) Az egyszerű értelmezhetőség kedvéért lineáris összefüggéseket feltételezünk. Az autók árát az életkorukkal (év), a megtett kilométerrel (ezer kilométer), az autó teljesítményével (kW), a típussal (4 dummyváltozó, a referenciakategória a Nissan), a típus és az autó korának az interakciójával, valamint az autó állapotával (az átlagos vagy normál állapothoz, mint referenciához képest a sérült/hibás, valamint a kitűnő/újszerű állapotnak megfelelő két dummyváltozó) magyarázzuk. Az alábbi táblázat mutatja a becsült paramétereket 2025. és 2026. januárra, valamint együtt a két hónapra. A modellek magyarázóereje egyaránt 75–76%, minden becsült paraméter legalább 1%-os szinten szignifikáns, ezért a t-próbákat, p-értékeket nem tüntetjük fel.

2. táblázat

**Becsült regressziós paraméterek a három modellben,
illetve az átlagvektorok**

*Estimated regression parameters in the three models
and the corresponding mean vectors*

Megnevezés	Becsült paraméterek			Átlagok	
	2025	2026	együtt	2025	2026
Konstans	2 228	3 871	3 503	1	1
Teljesítmény, kW	53	41	46	201,900	210,430
Futott táv, ezer km	-2	-2	-2	80,349	85,951
Életkor, év	-132	-213	-170	4,737	5,040
BMW dummy	9 371	8 837	9 270	0,240	0,211
Hyundai dummy	3 198	2 214	2 707	0,136	0,188
Tesla dummy	-4 594	-2 589	-3 318	0,337	0,328
Volkswagen dummy	4 047	4 188	4 020	0,143	0,156
BMW *kor	-1 284	-1 059	-1 194	1,033	0,964
Hyundai *kor	-376	-271	-322	0,592	0,834
Tesla *kor	112	-2	21	1,611	1,693
Volkswagen *kor	-520	-543	-508	0,612	0,769
Sérült, hibás dummy	-2 760	-2 476	-2 707	0,010	0,008
Kitűnő/újszerű dummy	807	515	686	0,204	0,173
2026 dummy			-794		

Forrás: hasznaltauto.hu és KSH alapján saját számítás.

A paramétereket természetesen önmagukban is lehet értelmezni (X egységnyi változása átlagosan milyen változást okoz Y -ban a többi tényező változatlanságát feltételezve), de esetünkben az árindexek számolása a cél. A két külön modelltől számszerűsítve az árindexeket:

$$I_p^L = 0,934 \quad I_p^P = 0,919 \quad I_p^F = 0,927$$

Azaz a tiszta árváltozás mértéke a Fisher-indexet használva: $-7,3\%$. Az összevont mintára 2026 adatait jelölve dummyval a 794 ezer forintos – a többi tényező

változatlanságát feltételező – árcsökkenés pontosan 0,927-es indexnek felel meg, azaz itt a kétféle megoldás (ha az első esetben a Fisher-típusú árindexet használjuk) gyakorlatilag ugyanarra az eredményre vezet. Számolható az átlagárindex (0,936) segítségével közvetve egy összetételhatás-index is, ennek értéke 0,99. Azaz a tiszta árváltozás nagyobb mértékű, mint az átlagárváltozás, mert volt némi eltolódás az olcsóbb árfekvésű autók felé (pl. 2026-ra nőtt a meghirdetett autók átlagos életkora, futásteljesítménye, illetve csökkent némileg a kitűnő/újszerű állapotú autók aránya).

A regressziós standardizáláshoz fűznék még egy megjegyzést. A halandóság elemzésekor inkább a Poisson-regresszió alkalmazása a célszerű: ebben az esetben a függő változó a meghaltak száma (fő), azaz gyakoriság, ami ráadásul kis értékeket is felvehet (lásd erről egy hasonló alkalmazással együtt: *Moksony, 2006*). A Poisson-regresszió esetében a regressziós modell a következő formában írható:

$$\ln\left(\frac{A}{B}\right) = b_0 + b_1 \cdot x$$

ahol ebben a konkrét esetben az 'A' a meghaltak száma (az egyes életkorokban), 'B' a népesség száma, 'x' pedig az életkor. A Poisson-regresszió gyakran készül olyan jelenségekre, ahol amellet, hogy a függő változó gyakoriság, értelmes az egy-egy-egyre jutó érték, mint esetünkben az ezer főre jutó halálozás. A programokban ez a vetítési alap nem egy másik függő változóként szerepel (bár az összefüggés ilyen formára is átírható, ebben az esetben B logaritmusával 1 együtthatóval magyarázóváltozóként szerepel), hanem mint ún. rizikónépesség (*offset*).

A hagyományos és regressziós standardizálás illusztrálására álljon itt egy példa a gender gap témaköréből (szintén saját számítás). Az egyéni bérek és keresetek adatfelvétele alapján a főállású egészségügyi dolgozók esetén a férfi-női bruttó keresetekben megmutatókozó különbséget, és ennek okait kerestem. (Az adatok 2021. végére vonatkoznak, forrásuk a KSH.) Az adatbázisban n = 47468 fő adatai szerepelnek, a nemek mellett ismert a foglalkoztatott területi elhelyezkedése, életkora, iskolai végzettsége, munkaköre, annak ténye, hogy az állami vagy versenyszférában dolgozik-e. Az egészségügyben dolgozók között összességében női túlsúly van, a foglalkoztatottak 29%-a férfi, 71%-a nő.

A férfiak átlagkeresete 972 ezer, a nőké 645 ezer forint volt, azaz jelentős, K = 327 ezer forintos, I = 1,507, azaz 50,7%-os különbség jelentkezik összességében a férfiak javára. Az alábbiakban látható, milyen különbségek mutatkoznak a keresetekben az esetleges összetételhatást okozó háttérismérvek alapján, és mekkora a K (teljes különbség) mellett a K' (részhatastkülönbség) és K'' (összetételhatás-különbség) értéke, ha csak egy-egy háttérváltozót veszünk figyelembe. A standardizálás a kétféle súlyozás átlaga alapján történik (Kitagawa-féle súlyozás).

3. táblázat

A férfi-női kereseti különbség felbontása az öt háttérváltozó alapján
Decomposition of the male-female earnings gap by five background variables

a) Terület szerint
By region

Terület	Foglalkoztatottak megoszlása, %			Átlagkereset, ezer forint/fő	
	férfi	nő	férfi	nő	k
Vidék	51	73	1 139	601	537
Budapest	49	27	798	765	32
Összesen	100	100	972	645	327

K = 327 ezer forint, K' = 346 ezer forint, K'' = -20 ezer forint

b) Szektor szerint
By sector

Terület	Foglalkoztatottak megoszlása, %			Átlagkereset, ezer forint/fő	
	férfi	nő	férfi	nő	k
Versenyszféra	5	8	1 042	542	499
Költségvetés	95	92	968	654	315
Összesen	100	100	972	645	327

K = 327 ezer forint, K' = 326 ezer forint, K'' = 1 forint

c) Életkor szerint
By age

Terület	Foglalkoztatottak megoszlása, %			Átlagkereset, ezer forint/fő	
	férfi	nő	férfi	nő	k
-29	12	10	415	415	0
30-39	21	12	719	613	106
40-49	28	35	765	553	212
50-59	23	34	1 098	632	466
60-	16	9	1 887	1 331	556
Összesen	100	100	972	645	327

K = 327 ezer forint, K' = 288 ezer forint, K'' = 39 ezer forint

d) Iskolai végzettség szerint
By educational attainment

Terület	Foglalkoztatottak megoszlása, %			Átlagkereset, ezer forint/fő	
	férfi	nő	férfi	nő	k
Legfeljebb 8 általános	11	5	368	325	44
Szakmunkás, szakiskola	5	8	408	371	37
Érettségi szakma nélkül	25	16	409	424	-15
Érettségi és szakma	13	40	450	444	6
Felsőfokú végzettség	45	30	1 648	1 164	484
Összesen	100	100	972	645	327

K = 327 ezer forint, K' = 187 ezer forint, K'' = 140 ezer forint

(A táblázat folytatása a következő oldalon)

(folytatás)

e) Munkakör (FEOR) szerint
By occupation (FEOR)

FEOR	Foglalkoztatottak megoszlása, %			Átlagkereset, ezer forint/fő	
	férfi	nő	férfi	nő	k
Egészségügyi intézmény vezetője	9	6	1 846	1 111	735
Általános orvos	5	3	1 030	936	95
Szakorvos	30	13	1 774	1 571	203
Felsőfokú képzettségű ápoló	1	8	563	566	-3
Ápoló, szakápoló	35	60	420	438	-18
Segédápoló, mütős	20	9	385	309	76
Összesen	100	100	972	645	327

K = 327 ezer forint, K' = 106 ezer forint, K'' = 221 ezer forint

A terület és a szektor esetében nem értelmezhető a klasszikus összetételhatás, mert ugyan a megoszlások különböznek egymástól, de ez nem tud érvényesülni, mert a részhatás ettől nagyjából független. (A terület szerint a nők között magasabb a vidékiek aránya, de a férfiaknál a vidékiek, a nőknél a budapestiek keresnek átlagosan többet. a szektor szerinti megoszlások között sincs nagy különbség, valamivel magasabb a nőknél a versenyszférában dolgozók aránya, de itt is ellentmondásos a részátlagok viszonya, a nők esetében a költségvetési fenntartású helyeken, a férfiaknál a versenyszférában dolgozók keresnek átlagosan többet.)

Életkor szerint is némi ellentmondásosságot lehet tapasztalni, de miután a legidősebb, legjobban kereső csoportok aránya a férfiaknál magasabb, összességében enyhén pozitív (a részhatással egyező irányú) összetételhatás tapasztalható. Az utolsó két ismérv esetén egyértelműbb a helyzet, a férfiak között jóval magasabb a legjobban kereső felsőfokú végzettségűek, vezetők és orvosok aránya.

Az öt háttérismérvet össze lehetne kombinálni, de ez mind a férfiak, mind a nők esetén $6 \times 6 \times 5 \times 2 \times 2 = 720$ kategóriát hoz létre. Illusztrációként összevariáltuk őket, és összesen 123 olyan társítást kaptunk, ahol mind a férfiak, mind a nők esetén volt legalább 1 megfigyelés a kategóriában. Természetesen ez alapján is lehet hagyományosan standardizálni, amiatt, hogy össze lehessen hasonlítani a regressziós elemzéssel, közöljük a standardizálás eredményét 5 kombinált szempont szerint is. K = 327 ezer forint, K' = 78 ezer forint, K'' = 249 ezer forint, azaz összességében az erős pozitív összetételhatás dominál.

Vizsgáljuk meg a regressziós eredményeket! A bruttó havi kereset a felsorolt ismérvekkel magyarázható külön a férfiak és a nők esetén. Az életkor (év) mennyiségi ismérv, a többi kategóriaképző, mind a négy esetben a legalacsonyabb átlagú referenciakategóriához képest.

4. táblázat

A regressziós becslés eredményei
Results of the regression estimation

Ismérvek	Férfiak			Nők		
	b	exp(b)	t-érték	b	exp(b)	t-érték
Konstans	12,034	168 384	652,758	11,931	151 903	1 062,987
Életkor, év	0,013	1,013	53,075	0,011	1,011	70,351
Budapest (vidékhez képest)	-0,002	0,998	-0,327	0,107	1,113	30,31
Egészségügyi intézmény vezetője	1,023	2,782	47,144	0,729	2,073	71,334
Általános orvos	0,718	2,05	30,67	0,802	2,23	64,685
Szakorvos	1,052	2,863	50,49	1,145	3,142	113,061
Felsőfokú képzettségű ápoló	0,179	1,196	6,639	0,303	1,354	31,117
Ápoló, szakápoló	0,07	1,073	6,427	0,279	1,322	43,258
Szaktunok, szakiskola	0,07	1,073	4,985	-0,045	0,956	-4,872
Érettségi szakma nélkül	0,138	1,148	10,106	0,106	1,112	12,104
Érettségi és szakma	0,183	1,201	13,995	0,112	1,119	13,376
Felsőfokú	0,413	1,511	19,207	0,399	1,49	37,91
Költségvetési (versenyszférához képest)	0,146	1,157	11,434	0	1	0,138
R ²	83,6%			75,6%		

A regressziós becslések önmagukban is értelmezhetők, érdekesek, mutatják a hatásmechanizmusokat. A regresszió alapuló standardizálás eredménye is az összetételhatás dominanciáját mutatja, de nem olyan mértékben, mint a klasszikus standardizálás esetében, $K = 327$ ezer forint, $K' = 108$ ezer forint, $K'' = 215$ ezer forint.

Végezetül a hagyományos standardizálás és a regressziószámítás összehasonlításának általános szempontjait (előnyöket és hátrányokat) érdemes külön is megfogalmazni:

- Regresszióval több változó is kezelhető, bonyolultabb összetételhatás-mechanizmusok is bemutatathatók, hatásuk kiszűrhető.
- A regresszió lehetővé teszi folytonos változók egyszerűbb és jobb kezelését is.
- A regresszió biztosítja a feltevések, összefüggérendszer explicit, számszerűsített vizsgálatát (pl. interakciót tartalmazó és a nélküli modell összevetését F-próbával).
- A regresszió ad hibajellegű, modell-jóságot jellemző eredményeket.

Ugyanakkor:

- Több olyan alkalmazás van, ahol a standardizálás hagyományos módszerét szokták használni, mint pl. a minta átsúlyozása, itt rutinszerűség esetén nem is merül fel a regresszió alkalmazása.

- A standardizálás módszertanilag egyszerűbb, jobban áttekinthető a kevésbé képzett felhasználók számára. A hivatalos statisztikában is gyakran alkalmazzák, különösen a konjunktúraindexek esetén.
- Kevés (és főleg diszkrét) háttérismérv esetén az eredmények egyszerűen áttekinthetők, interpretálhatók.
- Nagyon sok területen használták a múltban és használják jelenleg is, a szakemberek kiterjedt tapasztalatokkal rendelkeznek az alkalmazásáról (elég csak a halálzási arányszámokra gondolni).

Mindkét módszer (és esetleges más eljárások) esetében is szerettem volna felhívni a figyelmet, hogy bár a gazdasági–társadalmi alkalmazások esetében a leggyakrabban a figyelem központjában a részhatás elemzése (a standardizált eltérés/index) áll, sokszor az összetételhatás legalább olyan érdekes lehet nem csupán nagyságrendileg, de tartalmilag se sikkadjon el annak elemzése.

A cikk – szándékom szerinti – legfőbb tanulsága az lehet, hogy egy problémát (esetünkben a főátlagok összehasonlítását, az összetételhatás kimutatását és esetleg kiszűrését) többféle módszerrel is meg lehet közelíteni. Ahhoz, hogy a problémakör megfelelő módszertani eszköztárát helyesen válasszuk ki és interpretáljuk, hasznos adalék, ha megértjük, történetileg hogyan alakultak ki az egyes módszerek, mire reflektáltak a maguk korában, hogyan fejlődtek. A klasszikus standardizálás és a regressziószámítás összehasonlításával az is célom volt, hogy bemutassam, egyik sem feltétlenül jobb a másiknál, ez a vizsgált probléma jellegétől, a felhasználók felkészültségétől, a felhasználás céljától is függ. A történeti bemutatással szerettem volna azt is megvilágosítani, hogy bár egyes felhasználók idegenkednek attól, hogy a rögzült módszerektől eltérjenek, pedig egy másik módszer célravezetőbb lehet (pl. jobb a hedonikus árindexszámítás akkor, ha a reprezentáns alapú árindexszámítás nyilvánvalóan torzító hatású.)

Irodalom

- Anderson, D. R. – Sweeney, D. J. – Williams, T. A. – Camm, J. D. – Cochran, J. J. – Fry, M. J. – Ohlmann, J. W. (2023): *Statistics for Business & Economics* (14. kiadás). p. 1120.
<https://www.cengage.com/c/statistics-for-business-economics-14e-anderson-sweeney-williams-camm-cochran-fry-ohlmann/9781337901062/> \t " blank
- Blinder, A. S. (1973): Wage discrimination: reduced form and structural variables. *Journal of Human Resources, Fall*, 1113–1124. <https://www.jstor.org/stable/144855>
- Bojarszkij, A. Ja. – Suserin, P. P. (1952): *Népességi statisztika*. Statisztikai Kiadóvállalat, A szocialista statisztika könyvtára, Budapest.
- Drechsler L. – Köves P. (1952): Hozzászólás Kornis P. Andorné „Változó és változatlan állományú indexek” c. cikkéhez. *Statisztikai Szemle*, 30(8), 680–686.
https://www.ksh.hu/statszemle_archive/all/1952/1952_08/1952_08_0680_0686.pdf

- Drechsler L. – Kerekes O. (1974): Az építőipari indexszámítás egyes kérdései. *Statisztikai Szemle*, 52(10), 886–906.
https://www.ksh.hu/statszemle_archive/all/1974/1974_10/1974_10_0886_0906.pdf
- Farr, W. (1859): *Letter to the Registrar General. Appendix to the 20th Annual Report of the Registrar General for England and Wales*. London: General Registrar Office.
- Ferenci T. (2021a): *A klinikai gyógyszervizsgálatok alapjai* kézirat.
https://tamas-ferenci.github.io/FerenciTamas_AKlinikaiGyogyszervizsgalatokAlapjai/FerenciTamas_AKlinikaiGyogyszervizsgalatokAlapjai.pdf
- Ferenci T. (2021b): *Az orvosi megismerés módszertana (és az orvosi kutatások kritikus értékelése)* kézirat.
<https://ferenci-tamas.github.io/orvosi-megismeres-modszertan/ferenci-tamas-orvosi-megismeres-modszertan.pdf>
- Galasi P. (2001): A nő–férfi kereseti függvények Magyarországon. *Statisztikai Szemle*, 79(1), 18–34.
https://www.ksh.hu/statszemle_archive/all/2001/2001_01/2001_01_018.pdf
- Hajdu O. (2001): Összefüggések a lineáris regressziós modellben. *Statisztikai Szemle*, 79(10–11), 885–898. https://www.ksh.hu/statszemle_archive/all/2001/2001_10-11/2001_10-11_885.pdf
- Horváth Á. – Székely G. (2009): Hedonikus módszer alkalmazása a használt lakások áralakulásának megfigyelésében. *Statisztikai Szemle*, 87(6), 594–607.
https://www.ksh.hu/statszemle_archive/all/2009/2009_06/2009_06_594.pdf
- Hunyadi L. – Vita L. (2003): *Statisztika közgazdászoknak*. KSH, Budapest, p. 770.
- Hüttl A. – Vita L. (2004): *Gazdaságstatisztika egyetemi jegyzet*. BKÁE Statisztika tanszék, Budapest, p. 260.
- Keiding, N. (1987): The method of expected number of deaths, 1786–1886–1986. *Correspondent Paper International Statistical Review / Revue Internationale de Statistique*, 55(1), 1–20.
<https://www.jstor.org/stable/1403267>, <https://doi.org/10.2307/1403267>
- Keiding, N. – Clayton, D. (2014): Standardization and control for confounding in observational studies: a historical perspective. *Statistical Science*, November 2014, Vol. 29, No. 4, Special Issue on Semiparametrics and Causal Inference (November 2014), pp. 529–558.
<https://www.jstor.org/stable/43288498>, <https://doi.org/10.1214/13-STS453>
- Kenéz B. (1903): *A Statisztika elmélete*. Stampfel Károly kiadása, Budapest–Pozsony, p. 123.
- Kitagawa, E. M. (1955): Components of a difference between two rates. *Journal of the American Statistical Association*, 50, 1168–1194. <https://www.jstor.org/stable/2281213>, <https://doi.org/10.2307/2281213>, <https://doi.org/10.1080/01621459.1955.10501299>
- Körösy J. (1892): *Demologische Beiträge*. Berlin.
- Köves P. (1956): *Statisztikai indexek*. Közgazdasági és jogi könyvkiadó, Budapest, p. 203.
- Köves P. – Párniczky G. (1960): *Általános statisztika*. Közgazdasági és Jogi Könyvkiadó, Budapest, p. 410.
- Köves P. – Párniczky G. (1981): *Általános statisztika I–II*. Közgazdasági és Jogi Könyvkiadó, Budapest, p. 749.
- Központi Statisztikai Hivatal [KSH] (2000): *Lakásépítési költségindex 2000. I. negyedév*. KSH, Budapest, p.10.
- Maddala, G. S. (2004): *Bevezetés az ökonometriába*. Nemzeti Tankönyvkiadó, Budapest, p. 704.
- Malij, I (1949): A gazdasági indexek néhány módszertani kérdése. *Statisztikai Szemle*, 27(9), 211–221. (eredeti: *Voproszii ekonomiki*. „Pravda” kiadás, Moszkva, 1949(5), 21–35.)
https://www.ksh.hu/statszemle_archive/all/1949/1949_09/1949_09_0211_0221.pdf

- Moksony F. (2006): A Poisson-regresszió alkalmazása a szociológiai és demográfiai kutatásban. *Demográfia*, 49(4), 366–382.
<https://www.demografia.hu/demografia/index.php/demografia/article/view/576/479>
- Mórocz B. (1987): *Általános statisztika (a közgazdasági szakközépiskola számára)*. Tankönyvkiadó, Budapest, p. 301.
- Neison, F.G.P. (1844): On a method recently proposed for conducting inquiries into the comparative sanitary condition of various districts, with illustrations, derived from numerous places in Great Britain at the period of the last census. (London). *Journal of the Statistical Society of London*, 7, 40–68. <https://doi.org/10.2307/2337745>
- Oaxaca, R. L. (1973): Male-female wage differentials in urban labor markets. *International Economic Review*, 22(4), 724–732. <https://doi.org/10.2307/2525981>
- Pármiczky G. (1956): A standardizálás néhány újabb módszere. *Statisztikai Szemle*, 34(5), 433–449. https://www.ksh.hu/statszemle_archive/all/1956/1956_05/1956_05_0433_0449.pdf
- Pármiczky G. (1982): Az árszínvonal-változás mérése hedonikus módszerrel. *Statisztikai Szemle*, 60(5), 475–485. https://www.ksh.hu/statszemle_archive/all/1982/1982_05/1982_05_0475_0485.pdf
- Péter Gy. (1955): *Általános statisztika*. Tankönyvkiadó, Budapest, p. 395.
- Pintér J. – Rappai G. (2007): *Statisztika egyetemi jegyzet*. PTE Közgazdaságtudományi Kar, Pécs, p. 508.
- dr. Saïle T. A. (1927): *Kőrösy József hatása a statisztika fejlődésére*. Magyar Tudományos Akadémia Kiadása, Budapest, p. 134–148.
- Schweng L. (1944): *Statisztika*. Stephaneum nyomda, Budapest, p. 366.
- Simpson, E. H. (1951): The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(2), 238–241.
<https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>
- Székely G. (2020): Száz év a magyarországi lakásstatisztikában. *Statisztikai Szemle*, 98(6), 642–664.
<https://doi.org/10.20311/stat2020.6.hu0642>
- Takács O. (2021): A nemek közötti bérkülönbségek Magyarországon: a véletlenerdő- és az OLS-becslésen alapuló Blinder–Oaxaca-dekompozíció eredményeinek összehasonlítása. *Statisztikai Szemle*, 99(1), 5–45. <https://doi.org/10.20311/stat2021.1.hu0005>
- Yule, G. U. (1911): *An introduction to the theory of statistics*. London, C. Griffin and Company, p. 379.
<http://hdl.handle.net/2027/mdp.39015033708259>
- Yule, G. U. – Kendall, M. G. (1964): *Bevezetés a statisztika elméletébe*. Közgazdasági és Jogi Könyvkiadó, Budapest, p. 699.