

HALANDÓSÁGI MINTÁZATOK FELTÁRÁSA HELIGMAN-POLLARD HALANDÓSÁGI FÜGGVÉNYEK SEGÍTSÉGÉVEL

Ágoston Kolos Csaba (PhD., habil. egyetemi docens), kolos.agoston@uni-corvinus.hu

ÖSSZEFOGLALÓ

A cikkben bemutatunk egy módszert országok (vagy egyéb kockázatközösségek) halandósági mintázataik alapján történő csoportosítására (klaszterezésére). A bemutatandó módszer a görbék klaszterezésén alapul, ami viszonylag új területnek számít az adatelemzésben. A halandósági elemzésekben többféle görbe is előfordul, gondolhatunk itt a túlélési görbékre, az intenzitási görbékre, de akár a halálózási valószínűséget leíró görbékre is. A cikkben a Gompertz-Makeham törvény esetén a túlélési görbét használtuk, a Heligman-Pollard törvény esetén pedig az esély (odds) görbét. Mindkét megközelítés alkalmas volt arra, hogy az országokat értelmezhető csoportokba sorolja, a Heligman-Pollard törvény esetén még azonos várható élettartam mellett is el tudtuk különíteni a férfi és női adatokat. A cikkben numerikus elemzést is végeztünk európai országok adatait felhasználva. A kapott csoportközépeket az országcsoportokra jellemző halandósági mintáknak is tekinthetjük.

SUMMARY

In this paper we present a method for clustering countries (or other risk communities) based on their mortality data. The suggested method is based on the curve clustering approach, which itself is a fairly new topic in data analysis. We can describe the mortality of a community with multiple curves, like survival curves, hazard curves or mortality rate curves. In the paper we considered the survival curves for the Gompertz-Makeham mortality law and the odds curves for the Heligman-Pollard law. In both cases we were able to sort the countries into reasonable clusters. Furthermore, in case of Heligman-Pollard law even communities with the same life expectancy can be distinguished with respect to gender. A numerical results for European countries is also presented, where the cluster centers can be considered as typical mortality patterns.

Kulcsszavak: halandósági modellezés, klaszterelemzés, Gompertz-Makeham halandósági törvény, Heligman-Pollard halandósági törvény

Key words: mortality analysis, cluster analysis, Gompertz-Makeham mortality law, Heligman-Pollard mortality law

JEL: C38, J11

DOI: 10.18530/BK.2020.1-2.34

<http://dx.doi.org/10.18530/BK.2020.1-2.34>

Bevezetés

A halandósági adatok elemzése régi múltra tekint vissza a demográfiai, közgazdasági elemzésekben. Ilyen jellegű elemzéseknek központi jelentőségük van a gyógyszeriparban, ahol különféle vegyületek hatékonyságát vizsgálják. A közgazdaság területén pedig egyrészt az életbiztosítási díjkalkulációban játszik fontos szerepet, másrészt a nyugdíjrendszer vizsgálatokor, tervezésekor. Innen is látszik, hogy a közgazdasági alkalmazásban jellemzően hosszú távú szerződések elemzésére, modellezésére van szükségünk.

Több kísérlet is történt az emberi élettartamot leíró valószínűségi változó eloszlásának meghatározására: legegyszerűbben kezelhető lenne az exponenciális eloszlás, de ennek örökifjú tulajdonsága ellentmond a legegyszerűbb elvárásoknak. A Gompertz és Makeham nevéhez fűződő halandósági törvénynek már nagyobb relevanciája van. Ők a halálózási intenzitásra tételeztek fel egy függvényformát, ennek felhasználásával meghatározható az eloszlásfüggvény is, és az eloszlás más paraméterei is. A felnőtt népességre a Gompertz-Makeham függvény (eloszlás) jól illeszkedik, a nyers halálózási valószínűségek simítására máig is használják. A teljes életpályára viszont nem illeszkedik jól, a gyermekévek speciális halandósági mintázatait nem tudja lekövetni. Nem ismert olyan eloszlás, amely pár jól magyarázható paraméterrel rendelkezik, és kielégítően illeszkedik a teljes élettartamra.

A gyakorlatban ezért a koréves halálózási valószínűségeket határozzák meg, ami megadja azt a valószínűséget, hogy egy x éves, életben lévő egyén egy éven belül elhalálozik. A koréves halálózási valószínűségek segítségével meghatározható a várható élettartam, amit sok gazdasági elemzésben felhasználnak. A (születéskori) várható élettartam fontos mutatója egy ország (vagy más közösség) halandóságának, ezen értékekre alapozva, akár hasonló csoportokba (klaszterekbe) is sorolhatjuk az országokat (lásd pl. Kovács és Vaskövi, 2019).

Nem nehéz ugyanakkor végiggondolni, hogy két különböző halandósági minta esetén is megegyezhet a várható élettartam. Érdemes ezért olyan eljárásokat is alkalmazni, amelyek nemcsak egy összevont mutató (várható élettartam) értékét veszik figyelembe, hanem a teljes életpálya halandósága alapján határozzák meg a klasztereket. Ágoston et al. (2019a) megad több módszert is, hogyan lehet közösségeket halandósági mintázat alapján csoportosítani. Ezek a módszerek jellemzően a nem-paraméteres eljárások

közé tartoznak. Ágoston, Majstorovic és Vaskövi (2019b) paraméteres eljárást alkalmaz a Gompertz-Makeham halandósági törvény felhasználásával. Ebben a cikkben ezt a módszertant fejlesztettük tovább Heligman-Pollard halandósági törvényekre, ami jól illeszkedik a teljes életpályára.

Gompertz-Makeham halandósági törvényen alapuló klaszterezési eljárás¹

Gompertz-Makeham halandósági törvény esetén a halandósági intenzitásra tételezünk fel egy függvényformát:

$$h(x) = \alpha e^{\beta x} + \gamma$$

A $h(x)$ függvénnyel jelölt halálozási intenzitást 'pillanatnyi' halálozási valószínűségként lehet értelmezni, az $x, x+\epsilon$ intervallumban a halálozási valószínűség jó közelítéssel $h(x)\epsilon$.

A Gompertz-Makeham törvény esetén a β paraméter azt fejezi ki, hogy az életkor előrehaladtával milyen ütemben emelkedik a halálozás valószínűsége. Az α paraméter a halandóság általános nagyságát fejezi ki, a γ paraméter pedig a kortól független (jellemzően baleseti) halandóság nagysága. A Gompertz-Makeham törvényben szereplő α , β és γ paramétereket többféle módon is meg lehet határozni, mi az R statisztikai program MortalityLaws csomagját használtuk. A HMDB (Human Mortality Database) adatbázis adatait felhasználva Magyarországra, 2014-re, a következő értékek adódtak: $\alpha=0,0006610445$, $\beta=0,09182895$ és $\gamma=0,000000115$.

A Gompertz-Makeham törvény esetén az eloszlás egyéb jellemzői is meghatározhatóak, az eloszlásfüggvény például a

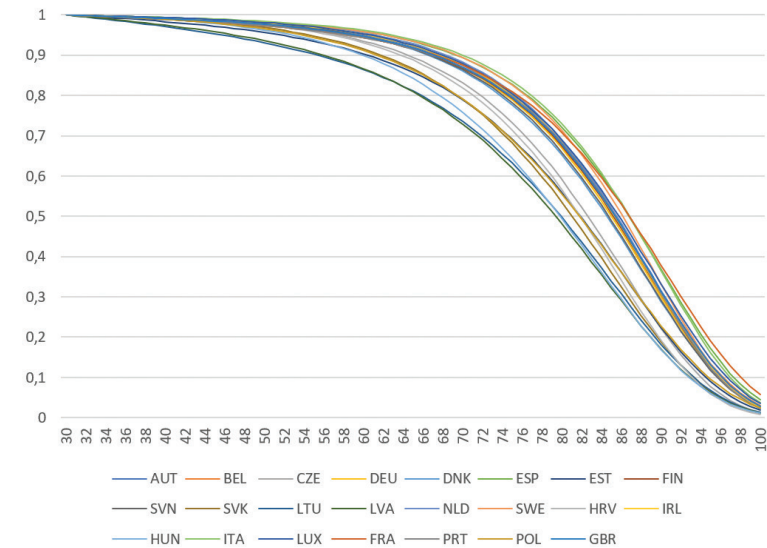
$$F(x) = 1 - \exp\left(-\gamma x - \frac{\alpha}{\beta}(\exp(\beta x) - 1)\right)$$

alakot ölti, de halandósági elemzések esetén inkább a túlélési ($S(x)$) függvény használatosabb, ami $1-F(x)$ módon számítható:

$$S(x) = \exp\left(-\gamma x - \frac{\alpha}{\beta}(\exp(\beta x) - 1)\right)$$

Kiszámítottuk a HMDB adatbázisban elérhető országok esetén a 2014-es évre a túlélési függvényeket, ezek láthatóak az 1. ábrán.

1. ábra: Túlélési görbék a Gompertz-Makeham törvény alapján



Forrás: saját szerkesztés

Az 1. ábrán szereplő túlélési görbék alapján szeretnénk hasonló csoportokat létrehozni. Görbék klaszterezése nem általánosan elterjedt, a Sangalli et al. (2010) és Majstorovic et al. (2018) munkákban megtalálható a módszer szabatos leírása. Mindkét munkában ugyanaz a hasonlósági mérték szerepel, ezt használtuk mi is. Adott függvény f és g . Az $[a, b]$ intervallum esetén a köztük lévő hasonlóságot a

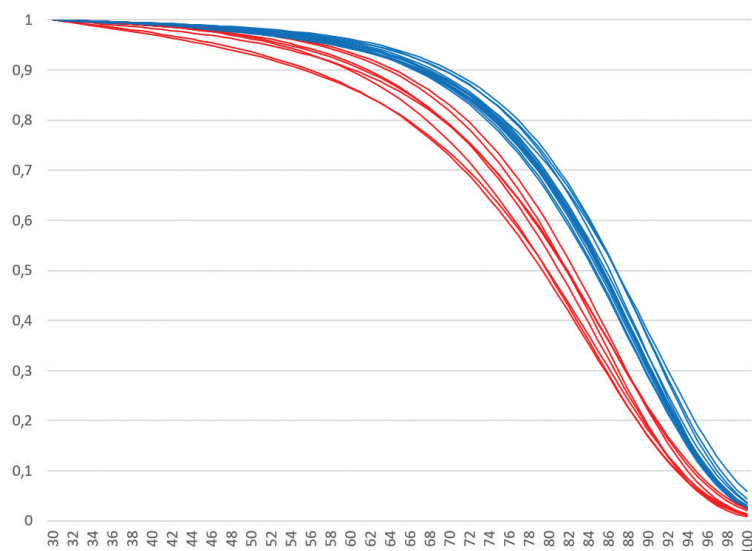
$$d(f, g) = \int_a^b \left(\frac{f'(t)g'(t)}{\sqrt{\int_a^b (f'(s))^2 ds} \sqrt{\int_a^b (g'(s))^2 ds}} \right)^2 dt \quad (1)$$

képlet adja meg. Az (1) képletet mi a túlélési függvényekre használtuk. Ha páronként kiszámítjuk a túlélési függvények hasonlóságát, akkor adva van egy 'távolság' mátrix. A leggyakoribb klaszterelemzési módszer a k-középpontú algoritmus, ami egyrészt elméleti oldalról kellemetlen tulajdonságokkal rendelkezik (heurisztikus algoritmus, ami érzékeny az induló klaszterek kiválasztására), másrésztől a mi esetünkben jelentős a számítási igénye is. Ehelyett a Kmedian-eljárást alkalmaztuk, amelyhez elegendőek a távolságmátrixban szereplő értékek, és gyorsan lefut (lásd pl. Vinod, 1969).

A 2. ábrán az első ábrán szereplő görbék vannak, színükkel jelezve, hogy melyik görbe melyik klaszterbe tartozik. A kialakult klaszterek teljesen hihetőek: a nyugat-európai országok

kerültek egy klaszterbe (kék színű görbék), a másikba a kelet-közép-európai, illetve balti országok (piros színű görbék).

2. ábra: Túlélési görbék esetén kialakult csoportok (piros és kék színnel van színezve két kialakult csoport, a kékekkel jelölt klaszterbe tartoznak a nyugat-európai országok, a piros színnel jelöltbe pedig a kelet-közép-európai országok, illetve a balti országok)



Forrás: saját szerkesztés

A 2. ábrán a görbék jól elkülönülnek, és a csoportok is jól értelmezhetők, de a várható élettartamban is jelentősek az eltérések. Ha egyszerűen a várható élettartam alapján klasztereznénk az országokat, hasonló eredményt kapnánk. Megvizsgáltuk, hogy akkor is hasonlóan szép eredményt kapunk-e, ha olyan túlélési függvényeket klaszterezünk, ahol a várható életkorok azonosak. Ezért minden egyes ország esetén külön tekintettük a férfi és női halandóságot, és a naptári évet úgy választottuk meg, hogy 30 évesen a várható élettartam 50 év legyen (legalábbis amennyire lehet, közelítse meg). Kérdés, hogy a férfi és női adatokat el tudjuk-e választani. A korábban bemutatott módszer alapján kapott csoportokat az 1. táblázat mutatja. Az 1. táblázat harmadik oszlopában megadjuk, hogy a klasztereken belül hogy oszlanak meg a férfi és nő adatok. Bár az első klaszterben arányaiban többen vannak a nők, a másodikban pedig a férfiak, azért az átfedés jelentős, nem túlzás kijelenteni, hogy a Gompertz-Makeham törvény esetén a férfi és női halandóság nem különíthető el.

1. táblázat: Klaszterelemzés eredménye a túlélési görbék alapján, 30 évesen azonos várható életkor esetén

Klaszter	Országok	Esetszám
1	AUT_F_1991, BEL_F_1989, CZE_F_2006, DEU_F_1992, ESP_F_1980, FIN_F_1993, SVN_F_2000, SVK_F_2011, SWE_F_1981, HRV_F_2007, IRL_F_2000, ITA_F_1984, LUX_F_1991, FRA_F_1983, PRT_F_1992, NLD_M_2013, SWE_M_2009, ITA_M_2009	Férfi: 3 Női: 15
2	DNK_F_2002, EST_F_2008, LTU_F_2012, LVA_F_2015, NLD_F_1979, HUN_F_2016, POL_F_2004, GBR_F_1995, AUT_M_2017, BEL_M_2018, ESP_M_2013, IRL_M_2014, LUX_M_2013, FRA_M_2014, GBR_M_2016	Férfi: 7 Nő: 8

Forrás: saját szerkesztés

Heligman-Pollard halandósági törvényen alapuló klaszterezési eljárás

Az előző alfejezet tapasztalatait figyelembe véve érdemes más halandósági törvénnyel is próbálkozni. Bár meglehetősen sok halandósági törvény ismert (egy részletes összefoglalót közöl Tabeau (2001)), közülük sok nem a teljes életpályát modellezi, vagy akár nem is emberi sokaságokra lettek kigondolva. A lehetőségek közül a Heligman-Pollard törvényt láttuk a legígéretesebbnek (Heligman és Pollard, 1980). A törvény több alakja is ismert, mi az első változatot választottuk. A Heligman-Pollard törvény nem a halálozási intenzitásra fogalmaz meg függvényformát, hanem a halálozási odds-ra:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D(\exp(-E(\ln(x) - \ln(F))^2)) + GH^x \quad (2)$$

A (2) képletben megadott törvénynek 8 paramétere (A, B, C, D, E, F, G és H) és 3 tagja van. Az első tag a gyermekkori halandóságot írja le, a második az ún. baleseti púpot, a harmadik pedig az időskori halandóságot. Talán az időskori halandóságot leíró 3. tagot a legegyszerűbb értelmezni. Az életkor növekedésével a halandóság exponenciálisan növekedik (hasonlóan a Gompertz-Makeham törvényhez), a H paraméter írja le a növekedés ütemét, amely jellemzően 1,1 körüli érték. G paraméter egy skála paraméter, értéke közel van a 0-hoz.

Az első tag a gyermekkori halandóságot írja le. C paraméter értéke 1-nél kisebb, tehát ez a tag a kor növekedésével exponenciálisan csökken, a C paraméter értéke jellemzően 0,1 és 0,2 közötti érték. A B paraméter nullához közeli érték, ebből az is látszik, hogy egyéves korban a tag értéke megközelítőleg q_1 . Ha a B paraméter 0 lenne, akkor újszülött korban a tag értéke 1, függetlenül A és C értékétől. Újszülöttek halandóságát a B paraméter igazítja a tényleges értékhez.

A középső tag az ún. baleseti púp. Baleseti púp alatt azt értjük, hogy fiatal emberek (jellemzően férfiak) halandósága megnő a felnőtt kor elején (18-20 éves kor között), a nevéből jól láthatóan a balesetek miatt. A nyugati országokban a húszas évek első harmadában (20-23 év) kisebbek a halálzási valószínűségek, mint 1-2 évvel korábban (18-20). Hazánkban ez a csökkenés nem figyelhető meg, de a felnőttkor elején intenzívebben növekszik a halálzási valószínűség, a növekedés üteme mérséklődik a húszas években. A baleseti púp a modern társadalom terméke, de ez a mintázat korábban is megfigyelhető volt (bár nem pont ezekben az életkorokban), férfiak esetén a háborúk, nők esetében a szülések következményeként. A (2) képletben a középső tag akkor maximális, ha $x=F$, tehát az F paraméter a baleseti púp közepét adja meg. Ha x értéke kisebb vagy nagyobb, mint F , akkor a tag exponenciálisan csökken (a becsült odds nem biztos, hogy csökken, mert a második tag csökkenését a harmadik tag növekedése ellensúlyozhatja). Az E paraméter mutatja a baleseti púp kiterjedtségét, a D paraméter pedig a nagyságát.

A Heligman-Pollard törvényt szintén lehetne az R programcsomag MortalityLaws csomagjával illeszteni, de sajnos az eredmények nem stabilak, nem robusztusak, értelmetlenül próbálkozva sokszor jobb eredményt kapunk.

Heligman és Pollard (1980), illetve Kostaki (1992) leírja (vázlatosan) a paraméterbecslési eljárást, ami lényegileg numerikus nemlineáris optimalizálás; a

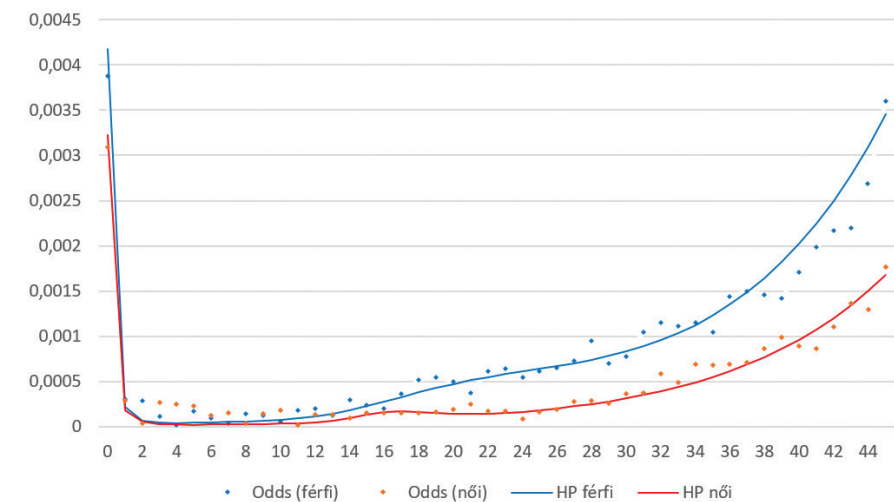
$$\sum_0^N \left(\frac{\hat{q}_x}{q_x} - 1 \right)^2 \quad (3)$$

eltérésfüggvény minimalizálása, ahol \hat{q}_x a Heligman-Pollard törvényből számolható becsült koréves halálzási valószínűség, a nevezőben szereplő q_x pedig a nyers halálzási valószínűség. Ezzel kapcsolatban érdemes megjegyezni, hogy ha q_x értéke 0, akkor a hányados értelmezhetetlen. Kisebbségi országok esetén (lásd például Litvánia esetét a későbbiekben) könnyen előfordulhat, hogy bizonyos korévek (jellemzően gyermekkor) esetén nem történt haláleset, ezért a nyers halálzási valószínűség 0. Ezeket a koréveket nem vettük figyelembe. Másik megjegyzés, hogy ha a nevező nagyon kicsi (ami könnyen előfordul megint csak gyermekkorban), akkor kismértékű eltérés esetén is nagy lesz a hányados értéke. Ezért azt fogjuk tapasztalni, hogy a görbe felett sokkal nagyobb eltérések adódnak, mint alatta.

Megpróbáltuk reprodukálni a módszert, az illesztésnél az R programcsomag optim függvényét használtuk. Még így sem kaptunk minden esetben kifogástalan eredményt. Közismert, hogy nemlineáris optimalizálásnál könnyen egy lokális optimumba ragad bele a megoldó, a mi esetünkben is sokszor ez történik. A probléma orvoslására egy többlépcsős módszert alkalmaztunk. A B paraméter lényegében az újszülött-halandóságnál játszik szerepet, első lépésként ezért ezt a paramétert 0-nak tekintettük, és az újszülöttek odds-át nem is vettük figyelembe. Az A paraméter értéke, ha $B=0$, akkor az egyévesek halálzási valószínűségének felel meg, ezt az értéket rögzítettük. Eltekintettünk a baleseti púp illesztésétől is, így lényegileg maradtak a C , G és H paraméterek. Több kísérlet során arra jutottunk, hogy nem érdemes a nagyon idős kor adatait figyelembe venni, mert az időskori halandóság így a kettőnél nagyobb hangsúlyt kap. Az illesztésnél csak az 1-60 éveket vettük figyelembe. Második lépésként az optimalizálásba

bevtünk mind a 8 paramétert. A C , G és H paraméterek kiinduló értéke a korábbi optimalizálás során kapott értékek voltak, az A paraméter kiinduló értéke q_1 volt, a B paraméter kezdő értékét pedig úgy állítottuk be, hogy a (korábban meghatározott) A és C paraméter kiinduló értéke mellett az újszülöttkorban tökéletes legyen az illeszkedés. A Heligman-Pollard törvény esetén a legnagyobb kihívás a baleseti púp jó illesztése (Heligman és Pollard, 1980., Kostaki, 1992.), ezért a D , E és F paramétereknek többféle lehetséges kezdőértéket is adtunk. Elvégeztük a nemlineáris optimalizálást az optim megoldóval, és a legjobb illeszkedést tekintettük a végső megoldásnak. A 3. ábrán a magyar halandósági adatok láthatóak, illetve a magyar adatokra illesztett Heligman-Pollard törvény.

3. ábra: Férfi, illetve női halálzási odds-ok, és az illesztett Heligman-Pollard halandósági törvény (Magyarország, 2017)

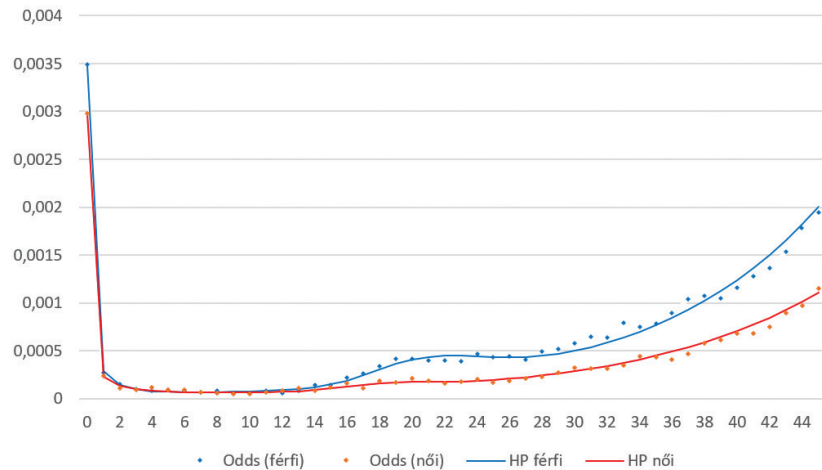


Forrás: saját szerkesztés

A 3. ábra adatait tanulmányozva azt mondhatjuk, hogy a férfiak esetén a Heligman-Pollard görbe jobban egybevág az előzetes várakozásokkal: a baleseti púp közepe (az F paraméter értéke) 22,6 éves korban van. Nők esetén a baleseti púp közepe 16,6 éves kor, amit nehezebb szakmai alapon magyarázni. De nők esetén nagyon kicsi a baleseti púp, még azt is meg lehetne kockáztatni, hogy nincs is. Az átfogóbb kép érdekében a 4. ábrán a német, az 5. ábrán pedig a litván adatokat mutatjuk be. Németország jelentősen nagyobb lélekszámú ország, ezért a nyers halálzási valószínűségek is kevésbé ingadoznak, a baleseti púp jobban kirajzolódik nyers halandósági adatokból is, így becslése is könnyebb. Litvánia pedig még Magyarországnál is kisebb ország, a nyers halandósági adatok itt még nálunk is jobban szóródnak, így még nehezebb a baleseti púp helyének becslése. Férfiak esetén nem is 'púp', hanem 'gödör' adódott. Ez egyrészt a nyers halálzási valószínűségek relatív nagy ingadozásának, másrészt annak a következménye,

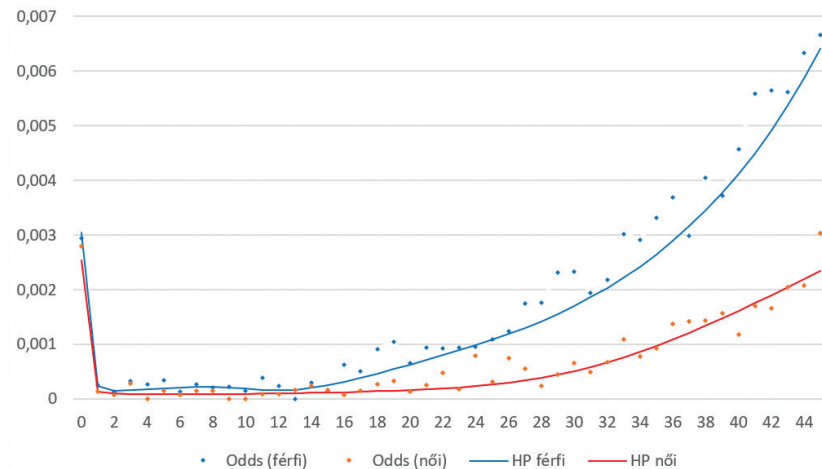
hogyan a görbe jobban „akar” illeszkedni a kisebb (nyers) valószínűségek esetén, mint a kicsit nagyobbakon. Ezen okok miatt a kisebb létszámú országokat ki is hagyhattuk volna, de egyrészt így jelentősen csökkent volna a vizsgálatba bevont országok száma, másrészt arra is kíváncsiak voltunk, hogy ilyen mértékű pontatlanságra mennyire érzékeny a modell.

4. ábra: Férfi, illetve női halálodds-ok és az illesztett Heligman-Pollard halandósági törvény (Németország, 2017)



Forrás: saját szerkesztés

5. ábra: Férfi, illetve női halálodds-ok és az illesztett Heligman-Pollard halandósági törvény (Litvánia, 2017)



Forrás: saját szerkesztés

Nem ismerünk olyan eredményt/levezetést, ami a Heligman-Pollard törvény esetén az eloszlásfüggvényt vagy a túlélésfüggvényt adná meg. Bár elméleti akadályja nincs, a képletek nagyon elbonyolodnak. Ezért a klaszterelemzésnél közvetlenül az odds görbékre számítottuk ki a hasonlósági mértékeket.

Az előző alfejezetben bemutatott vizsgálatot megismételtük, azzal a módosítással, hogy minden országhoz és mindkét nemhez olyan évszámot választottunk, hogy a születés kori várható életkor legyen 70 év. A kiválasztott országok és a Heligman-Pollard függvény paraméterei a 3. táblázatban megtalálhatók. A bemutatott klaszterelemzési eljárást elvégeztük két csoportot kérve. A klaszterek elemei a 4. táblázatban találhatók, a 6. ábrán pedig a két klaszterhez tartozó görbék.

2. táblázat: 70 éves várható életkorhoz tartozó naptári évek és országok, illetve a Heligman-Pollard függvény paramétereinek becsült értékei

Halandósági tábla	A *1000	B *10	C *10	D *10000	E	F	G *100000	H
AUT_M_1984	1,291	1,179	1,726	14,510	19,08	20,48	7,426	1,098
AUT_F_1953	3,237	0,422	1,946	2,934	148,75	23,45	15,105	1,074
BEL_M_1979	0,880	0,327	1,436	11,733	25,50	19,86	11,836	1,083
BEL_F_1952	3,073	0,386	1,839	1,702	10,53	23,89	10,699	1,080
BGR_M_2008	0,865	0,575	1,444	6,588	16,50	21,76	9,496	1,093
BGR_F_1960	5,816	1,147	2,333	2,610	84,91	19,18	14,696	1,070
CZE_M_1996	0,394	0,327	1,250	5,919	25,88	19,95	6,560	1,100
CZE_F_1953	5,034	1,936	2,591	2,784	8,55	20,30	12,703	1,076
DNK_M_1953	2,378	0,300	1,613	7,086	13,02	21,52	9,983	1,083
DNK_F_1948	3,687	1,042	2,039	3,419	13,87	25,54	15,388	1,072
GBR_M_1977	0,996	0,068	1,028	6,753	17,73	19,70	4,415	1,104
GBR_F_1948	2,877	0,457	1,741	8,500	12,13	20,22	21,023	1,067
EST_M_2009	0,591	2,067	1,962	18,818	18,21	27,62	7,531	1,099
FIN_M_1982	0,541	0,689	1,459	4,878	13,16	23,21	10,205	1,090
FIN_F_1953	3,813	1,183	1,799	-2,955	21,86	9,62	15,242	1,074
FRA_M_1978	1,221	0,702	1,542	13,345	25,50	19,77	12,756	1,085
FRA_F_1952	6,071	1,657	2,744	3,377	10,17	22,32	14,329	1,077
NLD_M_1950	4,101	0,724	1,673	4,760	11,79	20,62	11,728	1,079
NLD_F_1947	4,569	1,427	1,996	3,291	13,49	21,81	18,818	1,068
IRL_M_1980	0,960	0,101	1,001	5,830	18,44	19,99	6,276	1,096
IRL_F_1956	21,556	10,323	5,010	-4,282	19,99	18,27	16,121	1,073
POL_M_2001	0,505	0,658	1,711	5,669	23,31	20,91	8,901	1,096
POL_F_1960	3,351	0,423	2,092	1,970	16,93	22,73	13,260	1,074

LVA_M_2017	3,675	9,599	5,941	-2,344	35,25	14,96	8,770	1,100
LTU_M_2017	3,480	10,446	6,838	-2,176	7,29	13,68	12,013	1,092
HUN_M_2008	0,554	1,399	2,026	3,526	20,77	18,30	4,012	1,115
HUN_F_1960	3,215	0,680	2,284	3,231	19,26	19,64	9,637	1,082
DEU_M_1990	0,798	0,919	1,627	6,838	26,10	20,31	9,317	1,087
ITA_M_1974	0,977	0,021	1,041	6,316	17,81	18,66	9,209	1,088
ITA_F_1957	4,948	0,795	2,105	-5,469	19,45	41,35	16,245	1,071
PRT_M_1986	1,990	0,913	1,888	10,542	19,75	19,73	19,392	1,073
PRT_F_1970	6,529	0,658	2,133	2,026	82,44	23,54	10,644	1,075
ESP_M_1972	1,523	0,152	1,368	4,358	16,99	20,51	13,853	1,078
ESP_F_1958	4,876	0,726	2,042	1,806	19,02	25,10	16,273	1,068
SWE_M_1950	2,322	0,239	1,283	6,870	15,24	21,90	13,500	1,077
SWE_F_1946	1,866	0,486	1,705	6,945	10,22	21,20	22,569	1,066
SVK_M_2002	0,808	3,063	3,162	5,655	15,62	21,10	5,380	1,107
SVK_F_1958	4,126	1,420	2,220	2,224	57,74	15,50	10,919	1,077
SVN_M_1994	16,280	26,791	4,654	-3,639	9,15	11,21	20,049	1,076

Forrás: saját szerkesztés

3. táblázat: Klaszterelemzés eredménye a Heligman-Pollard odds görbék alapján, születéskor azonos várható élettartam esetén

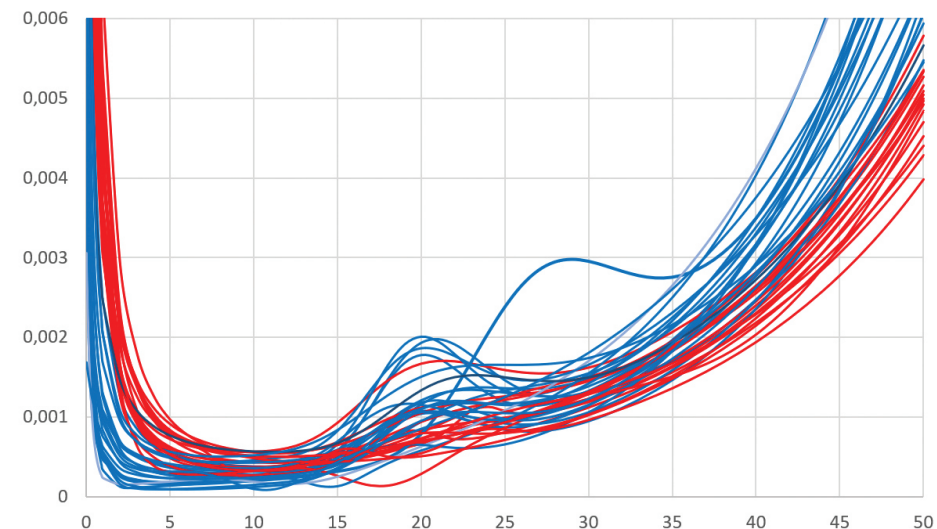
Klaszter	Országok
1	AUT_M_1984 BEL_M_1979 BGR_M_2008 CZE_M_1996 DNK_M_1953 GBR_M_1977 EST_M_2009 FIN_M_1982 FRA_M_1978 IRL_M_1980 POL_M_2001 LVA_M_2017 LTU_M_2017 HUN_M_2008 DEU_M_1990 ITA_M_1974 PRT_M_1986 ESP_M_1972 SWE_M_1950 SWE_F_1946 SVK_M_2002 SVN_M_1994
2	AUT_F_1953 BEL_F_1952 BGR_F_1960 CZE_F_1953 DNK_F_1948 GBR_F_1948 FIN_F_1953 FRA_F_1952 NLD_M_1950 NLD_F_1947 IRL_F_1956 POL_F_1960 HUN_F_1960 ITA_F_1957 PRT_F_1970 ESP_F_1958 SVK_F_1958

Forrás: saját szerkesztés

A 2008-as magyar férfi halandósági esélyek az első klaszterbe kerültek, ahol a fejlettebb országok férfi népességének jóval korábbi, 1950-1986 közötti értékei találhatóak, jelezve egy 40-50 évnyi lemaradást, míg a térségbeli országok férfi lakossága 2008-2017 közötti halandósági adatokkal látható itt.

A magyar nők 1960-as halandósági esélyei hasonlítanak leginkább a többi európai ország női népességének halandóságához, de itt nem válik élesen ketté a magyar adatokhoz képest a fejlett és feltörekvő országok köre.

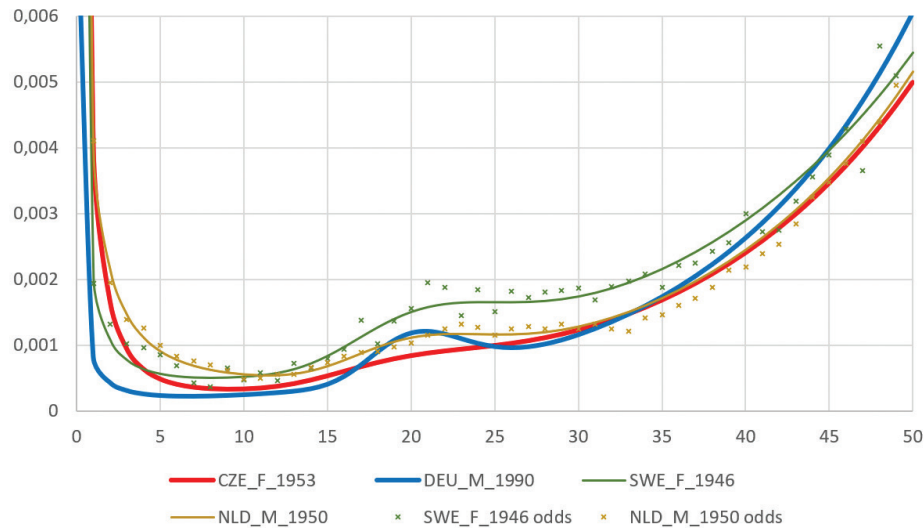
6. ábra: Heligman-Pollard odds görbék esetén kialakult csoportok (a 4. táblázatban az első klaszter elemei kék színnel, a második klaszter elemei piros színnel vannak jelölve)



Forrás: saját szerkesztés

A 4. táblázatból jól látszik, hogy ebben az esetben sikerült a férfi és a női halandóságokat elkülöníteni, két eset kivételével: az 1950-es holland férfi halandósági tábla a női csoportban tűnik fel, míg az 1946-os svéd női pedig a férfi csoportban. Ezen két halandósági tábla nyers adatait, az illesztett Heligman-Pollard odds függvényeket, illetve a klaszterközepek értékeit a 7. ábra tartalmazza. Látható, hogy bár rossz helyre lettek besorolva, nem arról van szó, hogy a módszer látványosan téved, hanem ezek a halandósági táblák alig térnek el a másik nemet jellemző értékektől. A 4. táblázat másik érdekessége, hogy kilógó adatok nem zavarták meg a klaszterezést, ez részben a Kmedian algoritmus robusztusságának, részben a választott hasonlósági mértéknek a következménye.

7. ábra: Az illesztett Heligman-Pollard odds függvények a klaszterközepek, illetve a kilógó értékek esetén



Forrás: saját szerkesztés

Szerettük volna a világ összes országára kiterjeszteni a klaszterelemzési eljárást, jellemző halandósági mintázatokat meghatározva. Halandósági adatok kellő részletezettséggel Európán kívül csak a fejlett országokra érhetőek el, az elemzés így majdhogynem értelmét vesztené. Ezért inkább a 3. táblázatban szereplő országok esetén növeltük a klaszterszámot, először háromra, majd négyre. A kapott klasztereket az 5. és 6. táblázat adja meg, a Heligman-Pollard görbéket pedig a 8. és 9. ábra. (A jobb áttekinthetőség kedvéért csak a klaszterközepek esetén rajzoltuk ki a görbéket, és a táblázatban a magyar adatok mellett ezeket is vastag betűvel jelöljük.)

4. táblázat: Klaszterelemzés eredménye a Heligman-Pollard odds görbék alapján, születéskor azonos várható élettartam esetén, 3 klaszter

Klaszter	Országok
1	AUT_F_1953 BEL_F_1952 CZE_F_1953 DNK_F_1948 GBR_F_1948 FIN_F_1953 NLD_M_1950 POL_F_1960 HUN_F_1960 SWE_F_1946 SVK_F_1958
2	AUT_M_1984 BEL_M_1979 BGR_M_2008 CZE_M_1996 DNK_M_1953 GBR_M_1977 EST_M_2009 FIN_M_1982 FRA_M_1978 IRL_M_1980 POL_M_2001 LVA_M_2017 LTU_M_2017 HUN_M_2008 DEU_M_1990 ITA_M_1974 PRT_M_1986 ESP_M_1972 SWE_M_1950 SVK_M_2002 SVN_M_1994
3	BGR_F_1960 FRA_F_1952 NLD_F_1947 IRL_F_1956 ITA_F_1957 PRT_F_1970 ESP_F_1958

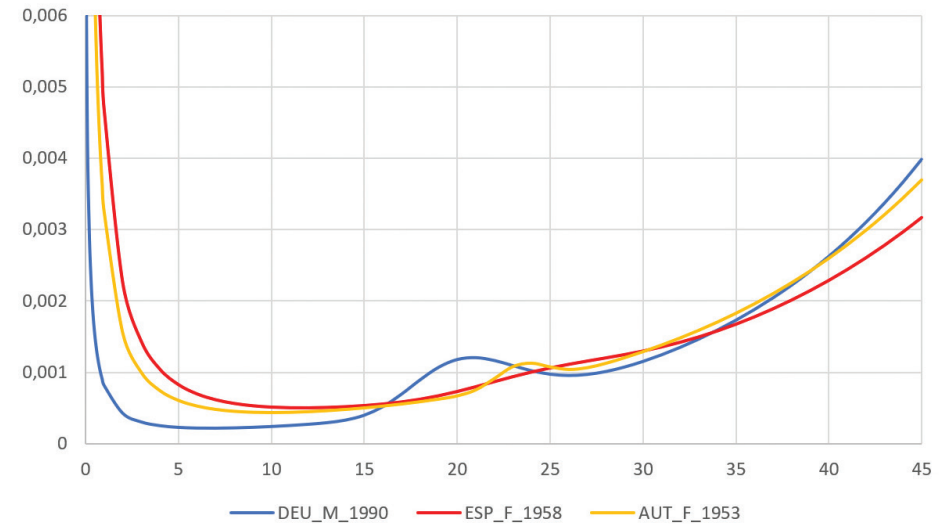
Forrás: saját szerkesztés

5. táblázat: Klaszterelemzés eredménye a Heligman-Pollard odds görbék alapján, születéskor azonos várható élettartam esetén, 4 klaszter

Klaszter	Országok
1	AUT_M_1984 BEL_M_1979 BGR_M_2008 CZE_M_1996 GBR_M_1977 EST_M_2009 FIN_M_1982 FRA_M_1978 IRL_M_1980 POL_M_2001 LVA_M_2017 LTU_M_2017 HUN_M_2008 DEU_M_1990 ITA_M_1974 SVK_M_2002 SVN_M_1994
2	AUT_F_1953 BEL_F_1952 CZE_F_1953 DNK_F_1948 GBR_F_1948 FIN_F_1953 FRA_F_1952 NLD_M_1950 NLD_F_1947 POL_F_1960 HUN_F_1960 SVK_F_1958
3	BGR_F_1960 IRL_F_1956 ITA_F_1957 ESP_F_1958
4	DNK_M_1953 PRT_M_1986 ESP_M_1972 SWE_F_1946 SWE_M_1950

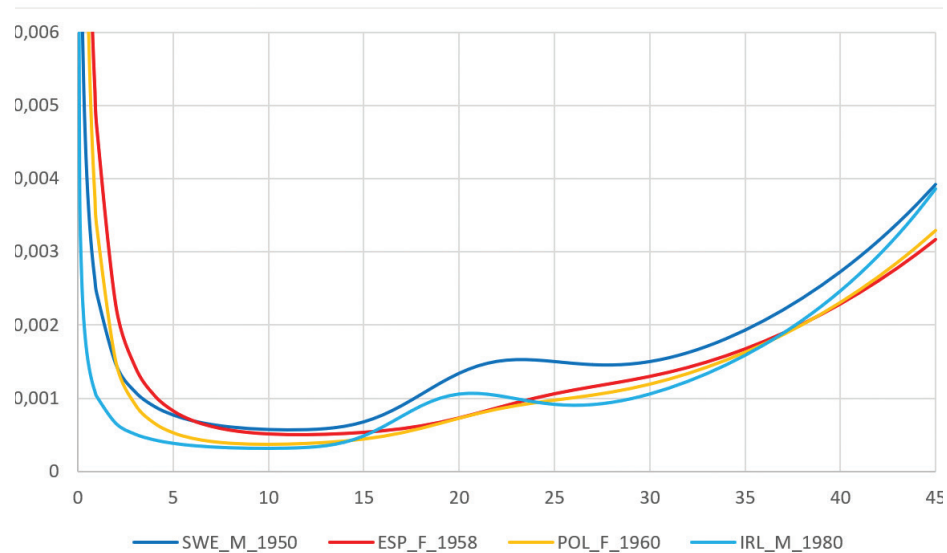
Forrás: saját szerkesztés

8. ábra: Klaszterközeppek esetén a Heligman-Pollard odds görbék, 3 klaszter



Forrás: saját szerkesztés

9. ábra: Klaszterközpontok esetén a Heligman-Pollard odds görbék, 4 klaszter



Forrás: saját szerkesztés

Az 5. táblázat és a 8. ábra alapján elmondhatjuk, hogy három klaszter esetén sem kaptunk vegyes klasztert (a korábban is tapasztalt kivételtől eltekintve), a nőket osztotta két részre az eljárás. Ez megint a módszer stabilitását jelzi. A 6. táblázat és a 9. ábra alapján négy klaszter esetén is tiszta klasztereket kaptunk, a várakozásnak megfelelően két férfi és két női csoportot. A 9. ábrán kirajzolt odds függvényeket tekinthetjük akár a háború utáni Európa esetén jellemző, tipikus halandósági mintázatoknak is.

A magyar férfi és női adatok több klasztert készítve is külön csoportban találhatóak. A női halandóság a fejlettebb országok pár évvel korábbi halandósági mintájához hasonlít. Három klaszter esetében az osztrák 1953-as adat a klaszter középpontja, négy csoportnál pedig a magyarral azonos, 1960-as évi lengyel adat a klaszter átlagos értéke. A férfi 2008-as halandóság azonban három és négy klasztert vizsgálva is több évtizednyi lemaradást mutat.

Összefoglaló

A tanulmányban a halandóságot leíró görbéket klasztereztünk. A bemutatott klaszterelemzési eljárással működőképes, hihető, reális klasztereket kaptunk. Először a Gompertz-Makeham halandósági törvény alapján soroltuk csoportokba az európai országokat. Az eredmény megfelel a várakozásoknak, de ha úgy ismétljük meg a módszert,

hogyan különböző naptári éveket tekintünk azért, hogy a mintában szereplő országok esetén a 30 évesen várt élettartam megegyezzen, akkor a férfi és női halandósági mintázatot nem sikerült megkülönböztetni.

Ezért megismételtük a vizsgálatainkat a Heligman-Pollard halandósági törvénnyel is. A Heligman-Pollard törvény illesztése sem triviális, az implementált függvények eredményét nem ítéltük kielégítőnek. A tanulmányban bemutattunk egy módszert, amely kielégítő eredményre vezet, bár kis országok és ebből kifolyólag nagy ingadozású nyers halálozási valószínűségek esetén itt is kaphatunk furcsa eredményeket.

A kapott függvények alapján elvégeztük olyan közösségek vizsgálatát, ahol a születés kori várható élettartam 70 év volt. Ebben az esetben a férfi és női halandósági mintázatot szét tudjuk választani, a jelenlévő outlier-ek megléte mellett is. Az elkülönülés megmarad nagyobb klaszterszám esetén is. A bemutatott görbék alapján képet kaphatunk a háború utáni Európára jellemző halandósági mintázatokról is.

A bemutatott görbék alapján képet kaphatunk a háború utáni Európára jellemző halandósági mintázatokról is. Az elmúlt évtizedekben többek között a jobb egészségügyi ellátás és a magasabb iskolázottság miatt a halandóság fejlődése és a várható élettartam - újabban lassuló - növekedése (lásd: Horváth 2019, Vékás 2019) miatt ez a terület az élet- és nyugdíjbiztosítók, a járadékszolgáltatók számára fontos változásokat hozott. A folyamat azonban nem állt meg, ezért a halandósági mintázat változása további kutatásokat igényel.

Köszönetnyilvánítás

A szerző ezúton szeretné megköszönni két lektorának a tanulmány gondos átnézését, jobbító szándékú javaslataikat.

Jelen publikáció az Európai Unió, Magyarország és az Európai Szociális Alap társfinanszírozása által biztosított forrásból az EFOP-3.6.2-16-2017-00017 azonosítójú „Fenntartható, intelligens és befogadó regionális és városi modellek” című projekt keretében jött létre.

HIVATKOZÁSOK

¹Az alfejezet nagyban épít Ágoston, Majstorović és Vaskövi (2019) munkájára, de bizonyos helyeken eltér attól (nem spektrál klaszterelemzést használunk, hanem Kmedian-módszert), illetve ott nem szereplő elemzéseket is közlünk most.

IRODALOMJEGYZÉK

- Ágoston, K. Cs. – Burka, D. – Kovács, E. – Vaskövi, Á. – Vékás, P. (2019a): Klaszterelemzési eljárások halandósági adatokra. Statisztikai szemle, 97(7), pp. 629–655.
<https://doi.org/10.20311/stat2019.7.hu0629>
- Ágoston, K. Cs. – Majstorovic, S. – Vaskövi, Á. (2019b): Spectral Clustering of Survival Curves. Proceedings of The 15th International Symposium on Operations Research in Slovenia (ISBN 978-961-6165-55-6), pp. 81–86.
- Heligman, L. – Pollard, J. H. (1980): The age pattern of mortality. Journal of the Institute of Actuaries, 107, pp. 49–80.
<https://doi.org/10.1017/s0020268100040257>
- Horváth, Gy. (2019): Fordulat a halandóságban? Biztosítás és kockázat, 6(3), pp. 18–33.
<https://doi.org/10.18530/bk.2019.3.18>
- Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on [22/03/2019]).
- Majstorović, S. – Sabo, K. – Jung, J. and Klarić, M. (2018): Spectral methods for growth curve clustering. Central European Journal of Operations Research, 26(3), pp. 715–737.
<https://doi.org/10.1007/s10100-017-0515-6>
- Kovács, Erzsébet – Vaskövi, Ágnes (2019): Living Longer, Working Longer? Life Expectancy and Retirement Age Trends in OECD Countries. Proceedings of 33rd Conference on Modelling and Simulation (ISBN 978-3-937436-65-4) pp. 103–108.
<https://doi.org/10.7148/2019-0103>
- Kostaki, A. (1992): A nine-parameter version of the Heligman–Pollard formula. Mathematical Population Studies, 3(4), pp. 277–288.
<https://doi.org/10.1080/08898489209525346>
- Sangalli, L. M. – Secchi, P. – Vantini S. and Vitelli V. (2010): K-mean alignment for curve clustering. Computational Statistics and Data Analysis, 54, pp. 1219–1233.
<https://doi.org/10.1016/j.csda.2009.12.008>
- Tabeau, E. (2001): A Review of Demographic Forecasting Models for Mortality. In: Tabeau E., van den Berg Jeths A., Heathcote C. (szerk.) Forecasting Mortality in Developed Countries. European Studies of Population, vol 9, Springer, Dordrecht.
- Vékás, P. (2019): Rotation of the age pattern of mortality improvements in the European Union. Central European Journal of Operations Research.
<https://doi.org/10.1007/s10100-019-00617-0>
- Vinod, H. D. (1969): Integer Programming and the Theory of Grouping. Journal of the American Statistical Association, 64, pp. 506–519.
<https://doi.org/10.1080/01621459.1969.10500990>