

MAGYAR GAZDASÁGTUDOMÁNYI KUTATÓK TELJESÍTMÉNYMUTATÓINAK ÉS CSOPORTJAINAK ELEMZÉSE A SCOPUS/SCIVAL TÜKRÉBEN

THE PERFORMANCE OF HUNGARIAN ECONOMICS RESEARCHERS BASED ON SCOPUS AND SCIVAL

A tanulmány a magyar gazdálkodás- és közgazdaságtudományi kutatók tudományos teljesítményét elemzi a Scopus és SciVal adatbázisokból származó és szabadon elérhető adatok alapján. A munka súlypontja három irányba mutat. Az első irány azt próbálja meghatározni, hogy a rendelkezésre álló és az előbbi adatbázisokból vett kritériumok között milyen lineáris statisztikai kapcsolat van. Ezzel együtt hogyan lehet a szempontok számát csökkenteni? Ez a kérdés annak segítségével dönthető el, hogy a kutató publikációi milyen mértékben kerülnek be a gazdasági folyóiratokba. Erre egy hüvelykujjszabály állítható fel. Ezt a multikollinearitás meghatározásával, korrelációs számítással, főkomponens-elemzéssel lehet elérni. A kutatás másik iránya azt deríti fel, hogy a két adatbázisból származó teljesítményváltozók mentén milyen ok-okozati kapcsolatok állnak fenn. Ezt parciális korrelációs számítással végezték el. Végül arra próbáltak választ kapni, hogy milyen jellemző kutatói csoportokat lehet meghatározni, amelyek jellemeznék a magyar gazdasági kutatókat.

Kulcsszavak: tudománymetria, gazdaságtudomány, többváltozós statisztika

The present study analysed the scientific performance of Hungarian researchers in economics and management, based on freely available data from the Scopus and SciVal databases. The study had three foci. First, it aimed to determine the linear statistical relationship between the available criteria taken from the databases and reduce the number of criteria. The issue was decided by examining the extent to which researchers' publications are included in economic journals through multicollinearity, correlation calculation, and principal component analysis. Second, it aimed to establish the causal relationship between performance variables from the two databases through partial correlation calculation. Third, it revealed the typical groups of researchers so that Hungarian economics researchers could be characterized.

Keywords: economics and management, multivariate statistics, scientometrics

Finanszírozás/Funding:

A szerző a tanulmány elkészítésével összefüggésben nem részesült pályázati vagy intézményi támogatásban. The author did not receive any grant or institutional support in relation with the preparation of the study.

Köszönetnyilvánítás/Acknowledgements:

A szerző köszöni Sasvári Péternek, a Nemzeti Közszolgálati és Miskolci Egyetemek egyetemi docensének a cikk írása során adott hasznos ötleteit és jó tanácsait.

The author thanks to Péter Sasvári, the associate professor at the University of Public Service and University of Miskolc, for his useful ideas and good advices during the writing of this article.

Szerző/Author:

Dr. Dobos Imre^a (dobos.imre@gtk.bme.hu) egyetemi tanár

^aBudapesti Műszaki és Gazdaságtudományi Egyetem (Budapest University of Technology and Economics), Magyarország (Hungary)

A cikk beérkezett: 2021. 08. 09-én, javítva: 2021. 11. 02-án és 2021. 11. 10-én, elfogadva: 2021. 11. 10-én.

This article was received: 09. 08. 2021, revised: 02. 11. 2021 and 10. 11. 2021, accepted: 10. 11. 2021.

A felsőoktatás és azon belül a gazdasági képzések tudományos teljesítményének mérése időről időre az érdeklődés homlokterébe kerül. Hogy néhány újabban megjelent tanulmányt említsünk, Csóka et al. (2019) egy új súlyozásos módszert javasolnak a magyar intézmények rangsorolására. Cikkünkben röviden kitérnék a nagy nem-

zetközi egyetemi rangsorokra, amit később érintünk is röviden. Más szerzők a magyar gazdasági doktori iskolákat hasonlítják össze a Nyugat-Európában meglévő hasonló intézményekkel (Sebrek, 2020). A szerző eredményeit abban összegezhethetnénk, hogy hazánkban még mindig hiányoznak a nívónak tekinthető publikációk, és a doktori iskolák tagjainak a nemzetközi beágyazottsága is kívánni-

valót hagy maga után. Húsz év távlatából azt kell megállapítanunk, hogy a magyar gazdasági kutatók teljesítménye kevésbé fejlődött úgy, ahogy azt az Európai Unióba lépésünk után elvárhattuk volna. Ezt a 2000 és 2004 között lezajlott vita akkori eredményei is visszaigazolták (Valentinyi, 2000; Simonovits, 2004). Az oktatási teljesítmény mérése is ugyanígy felmerül időnként (Hetesi & Kürtösi, 2009), a továbbiakban azonban csak a kutatási, publikációs teljesítményre összpontosítunk.

Az egyetemi rangsorok alapvetően két nagyobb, fontosabb mutatócsoporton alapulnak. Az egyik az oktatáshoz kapcsolódó indikátorok, a mási csoport a kutatási mutatószámok rendszere, ami főként a publikációs és hivatkozási indikátorok nyers és relatív viszonyszámokon alapuló rendszere. A nemzetközi irodalomban fellelhető dolgozatok nagyobb része a „nagy hármas” (big three) egyetemi rangsort emelik ki, illetve tanulmányozzák. Ezek az Academic Ranking of World Universities (ARWU), a QS World Rankings (QS) és Times Higher Education (THE) lista (Angelis et al., 2019; Doğan & Al, 2018; Fauzi et al., 2020; Mammadli, 2021; Sheeja et al., 2018).

Az ARWU rangsorában 20-20 százalékot ad a Nature és Science folyóiratokban megjelent cikkek száma és a Web of Science (WoS) adatbázisban indexált dolgozatok száma, ami összesen 40 százalékot jelent. A QS rangsora sokkal inkább kérdőíves lekérdezéseken, semmint kvantitatívan ellenőrizhető adatokon alapszik. Az adatbázisokból kinyerhető információk csak a hivatkozásokon alapszanak, ami 20 százalékos súllyal szerepel. Az információt a Scopus adatbázisból nyeri a minősítő. A THE listán a legmagasabb a számszerűsíthető kutatási eredmények aránya a maga 60 százalékaival. A 60 százalékból 30 százalék a kutatásból, míg a másik 30 százalék a hivatkozásból áll össze (Mammadli, 2021).

A dolgozat célja, annak vizsgálata, hogy a magyar kutatók milyen mértékben járulnak hozzá intézményük globális listákon elfoglalt helyezéshöz a Scopus és SciVal adatbázisokból szabadon letölthető adatok feldolgozása alapján. Ezt figyelembe véve a kinyert információk leginkább a QS és THE listákon elért magyar teljesítményeket támasztják alá, annak is a gazdasági kutatók eredményességében. A cél eléréséhez a két adatbázisból tizenkét kritérium választható ki összesen. Az első kérdés az volt, hogy lehet-e csökkenteni a kritériumok, statisztikai értelemben vett változók számát. Erre azért van szükség, mert a rendelkezésre álló tizenkét szempont kicsit soknak tűnik. Ehhez a változók közötti korrelációra alapuló eljárások alkalmazása tűnik megfelelőnek, mint a főkomponens-elemzés és a kollinearitás kiszűrésére használt varianciainflációs tényező módszere (variance inflation factor, VIF). A második kérdés azt célozta, hogy a megmaradt változók között milyen kauzalitás áll fenn. Ehhez a parciális korrelációelemzést lehet alkalmazni. Végül a harmadik, utolsó kérdés arra keresett választ, hogy a kutatók milyen csoportokba oszthatók. Ehhez a hierarchikus klaszterelemzés módszerét, semmint a K-means klaszterelemzést használtuk.

A dolgozat következő fejezetében az adatállomány összeállítását tárgyaljuk, arra is felhívva a figyelmet, hogy

milyen problémákkal járhat az adatállomány kinyerése a Scopus és SciVal adatbázisokból. Majd a rendelkezésre álló adatok, változók közötti kapcsolatokat elemezzük. Ezt a kollinearitás elemzésével kezdjük, majd a korreláció és főkomponens-elemzéssel folytatjuk, végül a változók közötti parciális korreláció vizsgálatával, azaz az ok-okozati elemzéssel zárjuk. A negyedik részben négy adatállomány segítségével kísérletet teszünk a kutatók csoportba sorolására, ami nem sok eredményt hozott. Végül összegezzük a dolgozat eredményeit.

Az adatállomány összeállítása

A szerzők a korábbi munkáikban többször kerestek arra választ, hogy ki tekinthető gazdasági kutatónak, vagyis ki jöhet számba az adatállomány összeállításakor (Dobos et al., 2020; 2021a; 2021b). Mivel az adatállomány összeállításakor az adatok forrása célzottan a Scopus és SciVal adatbázisok voltak, ezért a kérdés eldöntéséhez is a két adatbázist, valamint az ezekhez szorosan kapcsolódó SCImago adatállományt választottuk. A SCImago, mivel alapvetően a Scopus-ra alapozva folyóiratokat indexál, megfelelő alaphoz bizonyult a gazdaságtudományi folyóiratok meghatározására. A SCImago-ban ugyanis szakterület (subject area) szinten a gazdaságtudományok mindkét ága, vagyis a gazdálkodás- és a közgazdaságtudomány ezen a szinten szerepel. Megjegyezzük, hogy vannak tudományterületek, amelyek folyóiratai csak mélyebb szinten, azaz szakágzat (subject category) szinten gyűjthetők össze. Dobos et al. (2020) és (2021a) tanulmányaikban négy szakterületen publikáló kutatót neveztek meg, mint gazdasági kutatót egy kétlépéses algoritmus alapján. Ennek a kiválasztási folyamatnak a leegyszerűsítése az, amikor azt tekintjük gazdasági szakembernek, akinek a Scopus adatbázisban a publikációinak 30%-a két gazdasági szakterületre, vagyis Business, Management, and Accounting (BMA, Üzleti tudományok, menedzsment és számvitel) és Economics, Econometrics, and Finance (EEF, Közgazdaságtan, ökonometria és pénzügyek) esik a SCImago besorolása alapján (Dobos et al., 2021b). Ebben a dolgozatban ez utóbbi algoritmust alkalmazzuk a gazdasági kutató kategóriájának meghatározására azzal az élesítéssel, hogy azt tekintjük gazdasági szakembernek/kutatónak, akinek a cikkeinek 40 százaléka a fentebb említett két szakterületen került publikálásra. Ez azt jelenti, hogy ezzel szűkítettük a kutatóknak azt a halmazát, akiket egyáltalán figyelembe vehetünk.

Jegyezzük itt meg, hogy 2021 júliusáig a SCImago adatbázisban 2150 gazdálkodás- és közgazdaságtudományi folyóirat került be. Ebből 1427 darab tartozik a gazdálkodástudományhoz, 1111 a közgazdaságtudományhoz és 388 folyóiratot mindkét területen jegyeznek.

A kutatásra felhasznált adatállományt kétlépcsőben, 2020 augusztusában és 2021 márciusában vettük fel. 2021 márciusában még nem álltak rendelkezésre a SciVal adatbázisban a 2016 és 2020 közötti publikációs és hivatkozási adatok, ezért a 2015 és 2019 közötti teljes évek információit használtuk. Az adatállományba bekerülő kutatókat úgy vettük fel, hogy a gazdasági területre eső folyóiratcikk-

kek szerint csökkenő sorrendbe állítottuk a Scopus-ban szereplő magyarországi affiliációjú szakembereket. Mivel egy lekérdezésnél a Scopus csak 160 gazdasági kutatót tudott rendelkezésre bocsátani, ezért volt szükség további lekérdezésre, mert az adatainkat tisztítani kellett. A Scopus ugyanis olyan kutatókat is listázott, akiknek valaha volt magyar kötődésük, alkottak hazánkban. Ezért egyesével végig kellett néznünk, hogy a szakember affiliációja magyar-e. Itt egy újabb nehézségbe ütköztünk. A Közép-Európai Egyetem (CEU) kutatóit csak abban az esetben vettük fel az adatállományunkba, ha rendelkeztek másodállásban magyar affiliációjú munkahellyel, és ezt a publikációikban fel is tüntették. Az is a lekérdezésünk érdekessége volt, hogy azt is megvizsgáltuk, hogy mennyi olyan folyóiratcikke volt a kutatóknak, amelyet magyar oktatási/kutatási intézmény munkatársaként publikált. Ez olyan kutatóknál volt fontos, aki ugyan már Magyarországon dolgozik, de életútja során külföldön is eltöltött valamennyi időt, és ott publikált is. Ezzel a gazdasági tematikájú folyóiratcikkek számát is korrigáltuk. Felmerül a kérdés, hogy miért volt szükség kétlépcsős lekérdezésre?

1. táblázat

Az adatállományba bekerült kutatók affiliáció szerinti megoszlása

Intézmény	Gyakoriság	
	147 elem	116 elem
Budapesti Corvinus Egyetem	51	45
MTA Közgazdaság- és Regionális Tudományi Kutatóközpont	23	20
Budapesti Műszaki és Gazdaságtudományi Egyetem	18	13
Szent István Egyetem	10	4
Debreceni Egyetem	7	5
Miskolci Egyetem	5	4
Nemzeti Közszolgálati Egyetem	5	5
Pécsi Tudományegyetem	5	3
Budapesti Gazdasági Egyetem	4	4
Eötvös Loránd Tudományegyetem	4	2
Pannon Egyetem	4	2
Agrárközgazdasági Intézet	1	1
Kodolányi János Egyetem	1	1
Magyar Nemzeti Bank	1	1
Metropolitan Egyetem	1	1
Óbudai Egyetem	1	1
Pénzügykutató Intézet	1	1
Széchenyi Egyetem	1	1
Szegedi Tudományegyetem	1	1
International Business School	1	0
Semmelweis Egyetem	1	0
Morgan Stanley Capital Int Inc	1	1
Total	147	116

Forrás: saját számítások a Scopus/SciVal adatbázisokból

A feltett kérdésre félig már választ kaptunk: az első 160 kutató közé nem magyar affiliációjú kutatók is bekerültek. Azonban a lista ismétlődést is tartalmazott, több magyar kutató többször szerepelt a Scopus listájában, mert több azonosítóval is rendelkezett. Így az első lépésben csak 104 kutatót tudtunk, mint magyar illetőségű gazdasági szakembert azonosítani. Ezen kívül az adatállományba bekerült még 51 nem magyar illetőségű kutató és 5 magyarországi kutató kétszer szerepelt a Scopus adatbázisban. A 104 magyar kutató legalább 7 munkája szerepelt a lekérdezés pillanatában a Scopus-ban. Hogy az adatállományban szereplő kutatók számát növelhessük, második lépcsőben öt és hat gazdasági munkával a Scopus-ban szereplő magyar kutatót is kikerestünk, ami további 41 kutatóval növelte az adatállományt. Ezzel alakult ki a 147-es induló adatállományunk; ami azt jelenti, hogy ennyi magyar kutató publikált legalább öt BMA-s és/vagy EEF-es gazdasági folyóiratban egyáltalán.

A 147 kutató munkahelyi megoszlását az 1. táblázat mutatja. A táblázat utolsó két sorában szereplő intézmény hovatartozását nem sikerült kideríteni. Ilyen magyarországi illetőségi felsőoktatási intézmények nem léteznek, azok valószínűleg vállalatok. A felsőoktatási intézmények nevei úgy szerepelnek, amint az a Scopus-ban szerepelt, a folyó intézményi átalakulást nem vettük figyelembe. Az 1. táblázat utolsó oszlopában szereplő adatokat az adatállomány szűkítése után értelmezzük.

A legtöbb kutatót a Budapesti Corvinus Egyetem adja, a kutatók több mint harmadával. Ezt követi a volt MTA Gazdasági Kutatóközpont és a Műszaki Egyetem, szinte fej-fej mellett. Térjünk rá ezután az egyes kutatókhoz rendelt változók ismertetésére.

A Scopus-ból a kutatók adatlapjain szabadon hozzáférhető hét változón keresztül mértük a teljesítményüket, és a SciVal adatbázisból a 2015 és 2019 közötti teljesítménymutatókat gyűjtöttük össze. A változók tartalmazták a publikációs, hivatkozási és társszerzői mutatókat is. Ezek a változók a következők (zárójelben a rövidítésekkel):

Scopus-ból

- az összes magyarországi affiliációval publikált folyóiratcikkek száma (PA_{BE}),
- az összes Magyarországon született folyóiratcikkek száma (PA_{ALL}),
- az összes publikáció darabszáma (DOC_{ALL}),
- az összes hivatkozás (CIT_{ALL}),
- egy publikációra eső hivatkozásszám (CIT/DOC_{ALL}),
- a Hirsch-index (H-I),
- a társszerzők száma (C-A).

SciVal-ból

- a publikációk darabszáma 2015 és 2019 között (DOC_{15-19}),
- a hivatkozások száma 2015 és 2019 között (CIT_{15-19}),
- egy publikációra eső hivatkozásszám 2015 és 2019 között (CIT/DOC_{15-19}),
- a 2015 és 2019 közötti publikációk Hirsch-indexe (H5-I),
- a Field-Weighted Citation Impact ($FWCI_{15-19}$).

Ez azt jelenti, hogy összesen 12 változóval számoltunk, ami soknak mondható, de ugyanakkor magas információ-tartalmat is jelent. A folyóiratcikkek számát szerepeltettük, a Q1-Q4 minőségi besorolástól eltekintettünk.

A változók közül az FWCI₁₅₋₁₉ minden bizonnyal bővebb magyarázatra szorul, a többi, a Hirsch-indexet is beleértve, jól ismert. Az FWCI₁₅₋₁₉ alapvetően azt mutatja, hogy a szerző publikációi milyen hivatkozásvonzó hatással bírnak. Ha az FWCI egynél nagyobb, akkor több hivatkozás várható a publikációtól. Az FWCI₁₅₋₁₉ mutató számítási algoritmus sajnos nem érhető el, így csak a konkrét eredmények használhatóak szabadon a Scopus-ból (Purkaystha et al., 2019).

eloszlást követ, hanem exponenciális eloszlást. Ez különösen igaz a 2015-2019 közötti teljesítményadatokra. Ez arra utal, hogy az egyes mutatók növekedésével egyre csökken az azt teljesítők száma. A normális eloszláshoz leginkább a cikkek eloszlásai és a Hirsch-index emlékeztet a 2019-ig terjedő adatok megtekintésekor. Azonban mivel próbával nem teszteltük az eloszlások illeszkedését, ezért a leírtak csak vizuális sejtésnek tekinthetők. A sejtéseinket az SPSS által generált hisztogramokra alapoztuk. Mivel nem a változók eloszlásának tesztelése volt a célunk, attól eltekintettünk.

Azonban mielőtt elkezdenénk az adataink elemzését, arról kell döntenünk, hogy kit tekintünk a 147 ku-

2. táblázat

A változók átlaga és szórása a 147 és 116 elemű adatbázisokra

	Átlag		Szórás		Relatív szórás	
	147 elem	116 elem	147 elem	116 elem	147 elem	116 elem
PA _{BE}	11,136	11,009	8,576	8,830	0,770	0,802
PA _{ALL}	24,197	15,250	28,161	12,676	1,164	0,831
DOC _{ALL}	35,605	21,578	45,649	17,428	1,282	0,808
CIT _{ALL}	427,986	230,509	955,168	443,083	2,232	1,922
CIT/DOC _{ALL}	9,294	8,859	10,603	11,224	1,141	1,267
H-I	7,837	5,871	6,771	3,673	0,864	0,626
C-A	50,537	22,397	92,599	36,169	1,832	1,615
DOC ₁₅₋₁₉	12,721	8,284	14,455	7,924	1,136	0,957
CIT ₁₅₋₁₉	109,449	59,517	224,370	162,633	2,050	2,733
CIT/DOC ₁₅₋₁₉	6,522	6,132	10,916	11,993	1,674	1,956
H5-I	3,163	2,293	2,910	1,764	0,920	0,769
FWCI ₁₅₋₁₉	0,980	0,929	1,140	1,234	1,163	1,328

Forrás: saját számítások a Scopus/SciVal adatbázisokból

Első elemzésként a tizenkét változó statisztikai átlagát és szórását vizsgáljuk meg, amihez a relatív szórás is meghatározzuk. Az eredményeket a 2. táblázatban mutatjuk be. A táblázat azt mutatja, hogy az átlagos magyar gazdasági kutató a Scopus-ban 11 olyan cikkel rendelkezik, amelyet gazdasági folyóiratban publikált, az összes cikkeinek száma mintegy 24 és 35 dokumentuma van ugyanitt. Az átlagos hivatkozásainak száma 428, és egy Scopus-os dolgozatára 9 hivatkozást kap átlagosan. A H-indexe pedig körülbelül 8, míg az átlagos társszerzőinek száma 51. Ugyanakkor 2015 és 2019 között 13 olyan dokumentumot publikált, amit a SciVal is szemlél, amire átlagosan 109 hivatkozást kapott, amivel az egy dolgozatára eső ötéves átlagos citációja mintegy 7. Az ötéves H-indexünk átlagosan három, és az FWCI mutatónk 0,980. Azonban ehhez hozzá kell tennünk, hogy a legkisebb relatív szórásunk mintegy 77%, ami viszonylag magasnak tekinthető, ami arra utal, hogy mind a tizenkettő változónk nagy szórással rendelkezik. Ráadásul a legnagyobb relatív szórással éppen az összes hivatkozás rendelkezik, ami 223%-ot tesz ki.

Az egyes változók eloszlásának statisztikai próbáit ugyan nem végeztük el a változók többsége nem normális

tató közül gazdasági szakembernek. Ehhez a 3. táblázat nyújthat segítséget. A rövidítéseinket használva a PA_{BE} / PA_{ALL} értéket százalékban kifejezve keressük. A táblázatban felülről kumuláltuk az egyes százaléktizedekbe eső kutatók számát. Pl. a gazdasági cikkek az összes dolgozatokhoz aránya 70 és 80 százalék közé eső kutatók száma 17. Ugyanakkor a 70 százalékos aránynál több gazdasági cikket publikáló kutatók összes száma 92, ami a kutatók közel 63%-át teszi ki. Mivel a Dobos et al. (2021b) a 30 százalékos arány mellett döntött, most a 40 százalékos arány mellett döntöttünk, amivel a kutatók 78,91%-át, azaz majdnem négyötödét vetettük vizsgálat alá.

A 2. táblázat ugyanakkor tartalmazza a tizenkét változónk átlagát, szórását és relatív szórását a 116 kiválasztott gazdasági kutatóra is. A 116 kutató intézményi hovatartozását az 1. táblázatban tüntettük fel. Az intézményi arányok nagyon hasonlatosak az alapadat-állományban lévő adatokkal. Azt állapíthatjuk meg, hogy a relatív szórásunk többségében alatta marad a 147 elemű adatállományának, vagyis ezzel homogénebbé tettük az adatállományt. Ez azt is jelenti, hogy több olyan kutató volt az induló adatbázisunkban, akik magas publikációs számmal és hivatkozással rendelkeztek. Ugyanakkor mindegyik változó átlaga is

A gazdasági folyóiratokban megjelent cikkek aránya az összes dolgozaton belül

Százalék (PA_{BE} / PA_{ALL} %)	Kutató	Kumulált kutató	Százalék	Kumulált százalék
90-100	36	36	24,49	24,49
80-90	18	54	12,24	36,73
70-80	17	71	11,56	48,30
60-70	21	92	14,29	62,59
50-60	9	101	6,12	68,71
40-50	15	116	10,20	78,91
30-40	9	125	6,12	85,03
20-30	12	137	8,16	93,20
10-20	8	145	5,44	98,64
0-10	2	147	1,36	100,00
	147			

Forrás: saját számítások a Scopus/SciVal adatbázisokból

csökkent. Az adatállomány ismertetése után rátérünk az elemzések bemutatására.

Az adatállomány statisztikai vizsgálatai

A 116 elemű adatállományunk mentén tizenkét változó vonatkozásában öt elemzést végzünk el a változók közötti kapcsolatot vizsgálva. Először kísérletet teszünk a változók számának csökkentésére varianciainflációs tényező, VIF segítségével a változók közötti multikollinearitást elemezve. Ezután a változók közötti sztochasztikus kapcsolatot térképeztük fel a korrelációs mátrix elemzésével. Majd főkomponens-elemzéssel a változók számát csökkentettük. A negyedik elemzés a változók közötti ok-okozati kapcsolatot térképezte fel a parciális korreláció segítségével. Legvégül klaszteranalízis segítségével azt vizsgáltuk, hogy milyen jellegzetes csoportba oszthatók a magyar kutatók. Ez utóbbi elemzést egy külön fejezetben mutatjuk be.

Ennek fényében a következő kutatási kérdéseket állíthatjuk fel.

1. Kutatási kérdés:

Milyen összefüggés, kapcsolat van a vizsgálatra kiválasztott 12 teljesítménymutató között? Milyen redundancia van a változók között? Hogyan csökkenthető a teljesítménymutatók/változók száma a további vizsgálatok elvégzéséhez?

2. Kutatási kérdés:

Milyen lineáris kapcsolat, korreláció van a 12 teljesítménymutató között? Mely változóknál erős a lineáris kapcsolat?

3. Kutatási kérdés:

Hogyan tömöríthető a teljesítménymutatók információtartalma? Hány látens komponensbe/változóba nyomható össze a változók varianciája?

4. Kutatási kérdés:

Milyen ok-okozati kapcsolat van a változók között? Mely teljesítménymutatók azok, amelyek leginkább hatnak a többire?

5. Kutatási kérdés:

Milyen kutatói csoportok határozhatók meg a teljesítménymutatók alapján?

A következőkben a kérdések vizsgálatára térünk rá.

A multikollinearitás vizsgálata varianciainflációs tényezővel és többdimenziós skálázással

A szakirodalomban nincs egységes szabály arra nézve, hogy mely érték felett és milyen mutató esetén tekinthetők a változók kollinearitásnak. Ugyan vannak bizonyos empirikusan tesztelt VIF-küszöbértékek, amelyek 2,5 és 10 között szóródnak. A redundancia kiszűrése esetén nem létezik olyan elméleti/logikai szabályrendszer, amely alapján ezeket megbízhatóan meg lehetne határozni. Ezért több tanulmány (Lafi & Kaneene, 1992; Liao & Valliant, 2012; O'Brien, 2007) ajánlását elfogadva hoztunk döntést e tekintetben. Küszöbértéknek az 5-öt választottuk. Hasonló elemzést végzett Vörösmarty & Dobos (2020) is.

A 4. táblázatban mutatjuk be a változók szekvenciális kiszűrését. Előre bocsájtjuk, hogy nincs determinisztikus algoritmus a kollinearitás változók kiszűrésére. Első lépésként a legnagyobb VIF-értékkel rendelkező változó kiszűrését ajánlják, de bármelyik küszöbérték feletti változó is megfelelő az első lépés megtételéhez. A következő lépésben újra két lehetőség áll rendelkezésre: vagy újra a legnagyobb VIF-értékű elemet választjuk, vagy azt a változót, amelynél a legnagyobb mértékben csökken a VIF-értéke. Esetünkben az első lehetőséget választottuk, vagyis a megmaradt változók közül a legnagyobb VIF-értékűt kiszűröltük ki. Az második algoritmus is öt változó kiszűrését eredményezte.

Az induló VIF-értékek vizsgálatánál azonnal kiderült, hogy mindegyik változó induló VIF-értéke 5-nél nagyobb, vagyis elvileg mindegyik változó kollinearitásnak tekinthető.

A VIF-értékek alakulása az algoritmus során

	0. lépés	1. lépés	2. lépés	3. lépés	4. lépés	5. lépés
PA _{BE}	10,302	9,818	2,226	2,218	2,208	2,099
CIT/DOC _{ALL}	7,944	6,856	6,851	6,850	6,801	2,887
H-I	9,989	5,320	5,243	5,083	4,853	4,282
C-A	5,016	4,838	4,699	4,699	2,278	2,234
DOC ₁₅₋₁₉	7,264	3,976	3,753	3,752	3,747	3,609
CIT/DOC ₁₅₋₁₉	11,919	11,154	11,072	3,648	2,609	2,397
H5-I	7,394	5,610	5,442	5,429	4,861	4,729
CIT _{ALL}	7,633	7,056	7,046	7,044	6,826	
CIT ₁₅₋₁₉	8,945	8,816	8,799	8,706		
FWCI ₁₅₋₁₉	12,104	11,612	11,158			
PA _{ALL}	13,156	12,972				
DOC _{ALL}	13,791					

Forrás: saját számítások a Scopus/SciVal adatbázisokból

Ez azt jelenti, hogy az összes dokumentum száma, a hazánkban született folyóiratcikkek száma, az FWCI-mutató, valamint a 2015 és 2019 közötti hivatkozások száma lineárisan függ a többi változótól.

A varianciainflációs tényezőkön kívül még két utunk lehet a változók számának csökkentésére. Az egyik az a megfontolás lehet, hogy az összes publikáció és hivatkozás, valamint a 2015-2019 közötti adataink halmazódást tartalmaznak. Ezt azzal szűrhetjük ki, hogy a két változóhalmazt szétbontjuk. Az egyik változóhalmazba kerülhet az a hét változó, amelyet a Scopus nyílt hozzáférésű felületéről gyűjthetünk ki; míg a másik megoldásban a SciVal-ból letölthető adatok lehetnek, amelyekhez az intézmények külön előfizetés útján juthatnak hozzá. Ezzel már három csökkentett változójú adatállományunk lehet, amelyek öt, hat és hét változóból állnak. Ezek után elemezzük a változókat többdimenziós eljárással.

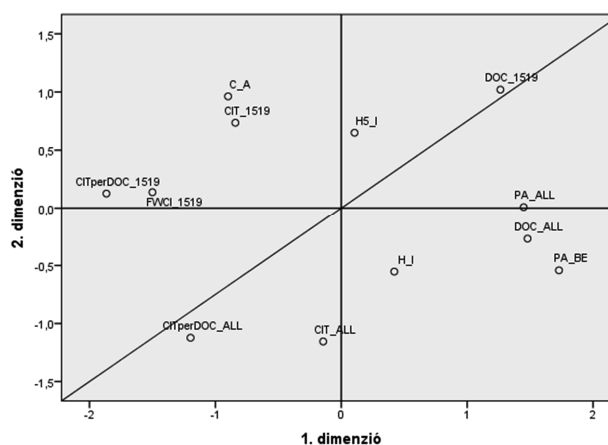
A többdimenziós eljárással, mint az MDS-Alscal, a változókat mint pontokat képezhetjük le egy alacsonyabb dimenziós térbe. A tér esetünkben 116 dimenziós. Ezzel a vizsgálattal ellenőrizhetjük, hogy a varianciainflációs elemzésünk megfelelő eredményt adott-e. Ezt a vizsgálatot hierarchikus klaszterelemzéssel is elvégezhetnénk, hiszen abba a Pearson-korreláció, mint (fél-)távolság be van építve. Az Alscal-ban távolságként az euklideszi távolságot választottuk, és a változók adatait a z-score segítségével normáltuk, vagyis nulla várhatóértékű és egy szórású adatokká transzformáltuk azokat. Erre azért volt szükségünk, mert a hivatkozási adatok többszöröse voltak a többi adatnak, ezért hasonló skálára hoztuk a változókat. Az eredményt az 1. ábrán szemléltetjük.

Az euklideszi távolsággal és átírással nyert modell stress értéke 0,049, míg az R² értéke 0,985, ami nagyon jó modellt mutat. Az y=x egyenest is berajzoltuk az ábrába, ami azt mutatja, hogy az összes publikációs teljesítmény és a 2015 és 2019 közötti változók szépen két részre bonthatók, a társszerzőszám kivételével. Az ábra azt is jól mutatja, hogy három változócsoport van nagyon közel

egymáshoz. Az összes gazdasági folyóiratcikk száma, az összes folyóiratcikk száma és az összes publikált dokumentum közel vannak. A VIF-elemzés is mutatta, hogy ebből a három változóból csak a gazdasági folyóiratcikkek száma maradt kollinearitás nélkül a maradékváltozók között. A másik csoport a 2015 és 2019 közötti hivatkozások száma és a társszerzők száma is közel vannak. Ez talán azzal magyarázható, hogy az utóbbi években a társszerzőszám emelkedése a hivatkozások számának növekedését is maga után vonta. Ebből a két változóból a társszerzők száma maradt meg. Végül, a 2015 és 2019 közötti egy publikációra eső átlagos hivatkozásai és a 2015-2019 közötti FWCI₁₅₋₁₉ mutató közül ez utóbbi került ki. Mivel mindkét mutató a hivatkozásra alapuló fajlagos mutató, ezért valószínűleg ez halmazódást jelenthet, ezért az egyik változó kiesett, amint a VIF-elemzés is mutatta. Azzal összegezhettük a többdimenziós skálázás elemzését, hogy az vizsgálta a varianciainflációs tényezőkön alapuló vizsgálatot.

1. ábra

A többdimenziós skálázás outputja



Forrás: saját számítások a Scopus/SciVal adatbázisokból

Az 1. ábrán azt is megfigyelhetjük, hogy az egyenesünk két oldalán struktúrájukban hasonló változók maradtak, amiből csak a társszerzők száma lóg ki. Ugyanis a H-indeket a publikációs szám és az egy dolgozatra eső átlagos hivatkozási szám határozza meg, amihez 2015-2019 közötti öt évben még a társszerzők száma adódik hozzá.

viszonylag alacsony. Viszont a mátrix jobb felső, illetve bal alsó sarka arra utal, hogy mindegyik kollinearitás tekintett változó erős korrelációt mutat a nem korrelált változók csoportjával. Ez segít megérteni, hogy milyen lineáris összefüggések vezetnek a kollinearitás változók elhagyásához.

5. táblázat

A korrelációs mátrix bal alsó részmatrixa

	PA _{BE}	CIT/DOC _{ALL}	H-I	C-A	DOC ₁₅₋₁₉	CIT/DOC ₁₅₋₁₉	H5-I
CIT _{ALL}	,428	,832	,790	,415	,345	,396	,513
CIT ₁₅₋₁₉	,214	,511	,481	,880	,443	,750	,726
FWCI ₁₅₋₁₉	,135	,629	,461	,649	,198	,942	,555
PA _{ALL}	,923	,164	,683	,318	,689	,115	,606
DOC _{ALL}	,799	,203	,794	,303	,760	,074	,559

Forrás: saját számítások a Scopus/SciVal adatbázisokból

Korrelációs számítás

A két adatbázisból nyert tizenkét változó közötti korrelációt az online mellékletben mutatjuk be. A varianciainflációs tényezővel kapott, és az 5. táblázatban bemutatott sorrendben tüntettük fel a változókat. A korrelációs mátrixban szürkével jelöltük a két változó egymáson belüli korrelációs együtthatókat. Ez a két részmatrix arra utal, hogy a két csoporton belüli korreláció

6. táblázat

A 12 változó rotált komponensei

A változók rotált komponens mátrixa

	Komponens				
	1	2	3	4	5
PA _{BE}	,970	,059	,108	,083	-,007
PA _{ALL}	,916	,059	,107	,273	,152
DOC _{ALL}	,801	-,055	,277	,411	,096
CIT/DOC ₁₅₋₁₉	-,001	,927	,228	,028	,233
FWCI ₁₅₋₁₉	,053	,901	,272	,126	,241
CIT _{ALL}	,316	,160	,881	,133	,167
CIT/DOC _{ALL}	,001	,420	,866	,004	,130
H-I	,560	,200	,635	,359	,097
DOC ₁₅₋₁₉	,475	,002	,038	,823	,200
H5-I	,314	,368	,243	,718	,313
C-A	,101	,386	,186	,231	,851
CIT ₁₅₋₁₉	,135	,591	,202	,286	,673
Megjegyzés. alkalmazott módszerek: főkomponens-elemzés és varimax rotáció Kaiser-normalizálással.					

Forrás: saját számítások a Scopus/SciVal adatbázisokból

Az 5. táblázat azt szemlélteti, hogy az öt kimaradt változó közül kettő, vagyis a PA_{ALL} és a FWCI₁₅₋₁₉ két maradt változóval nagyon erős, 0,9 feletti lineáris kapcsolatot mutat, amit az 1. ábra is visszaigazol. A két hivatkozási szám mindegyike a megfelelő átlagos citációs értékkel mutat magas korrelációt, ugyanakkor az összes dokumentummal kiegészítve, ez a három változó két, vagy több változóval mutat erősebb korrelációt, amit vastagon szedtünk. Ez is mutatja, hogy magas, tehát 0,9 feletti korrelációnál biztosak lehetünk a változók közötti kollinearitásba. Viszont ennél alacsonyabb korrelációs hányadosnál több változóval kell lineáris kapcsolatban lennie a kollinearitás változónak, mint a DOC_{ALL}, a CIT₁₅₋₁₉ és a CIT_{ALL} változókkal. Erre az utóbbi összefüggésre a korrelációs együtthatók alaposabb vizsgálata nélkül nehezen jöhetünk rá. A korrelációs együtthatóink többsége, de különösen a magasabb értékek mindegyike kétoldalú szignifikanciát mutat.

Főkomponens-elemzés

Mielőtt a hét, már nem kollinearitás változó főkomponens modelljét bemutatnánk, röviden visszautalunk a varianciainflációs tényezővel végzett elemzéseinkre és a többdimenziós skálázással nyert eredményekre egy olyan főkomponens modellel, ahol mind a tizenkét változót be vesszük a vizsgálatba. Egy olyan modellt építünk, ahol a legtöbb komponens akarjuk elérni, de úgy, hogy a változóink közül legalább egy a komponensekkel legalább 0,4-es korrelációt mutasson. Esetünkben öt ilyen komponens van, mert a hatodik komponensnek már nincs meg ez a tulajdonsága. A modell megfelelése a Kaiser–Meyer–Olkin-teszt alapján 0,761, ami egy közepes modellt jelent az elfogadott kategorizálás szerint. A rotált komponensek mátrixát a tizenkét változóra a 6. táblázatban mutatjuk.

A táblázatból világos, hogy az első komponens a publikációs blokk, a második a relatív mutatókat tartalmazza, a harmadik hivatkozási alapú mutatókat tartalmaz, a negyedik publikációs mutatókat, míg az utolsó, ötödik azt a két mutatót tartalmazza, amelyik az 1. ábrán is közel

volt egymáshoz. Az öt komponens közül csak a negyedikből nem maradt ki multikollinearitás miatt változó. Az eredmények tehát alátámasztják az MDS-moddal kapott eredményeket. Még jegyezzük meg, hogy az öt komponens a variancia szinte egészét, azaz 94,137 százalékot adja vissza.

A maradt hét változó főkomponens-elemzésénél három olyan komponenst kaptunk, amelyek a variancia 85,734 százalékát adják (7. táblázat), ami magasnak mondható. A modell megfelelése a Kaiser–Meyer–Olkin-teszt alapján 0,721, ami egy közepes modellt jelent.

7. táblázat

A változók komponensei
A változók rotált komponens mátrixa

	Komponens		
	1	2	3
CIT/DOC _{ALL}	,907	,255	,025
CIT/DOC ₁₅₋₁₉	,805	-,089	,378
PA _{BE}	-,013	,893	,125
H-I	,469	,772	,275
C-A	,478	-,021	,778
H5-I	,294	,450	,767
DOC ₁₅₋₁₉	-,164	,575	,735

Megjegyzés. alkalmazott módszerek: főkomponens-elemzés és varimax rotáció Kaiser-normalizálással.

Forrás: saját számítások a Scopus/SciVal adatbázisokból

Amint azt már a korrelációs elemzés alapján várhattuk, az átlagos hivatkozási szám mindegyike egy főkomponensbe került. A másik két komponensben a publikált dolgoza-

tok száma és a Hirsch-index került egy komponensbe. A társ szerzők száma, amint korábban is, a 2015-2019 évekre jellemző mutatók mellé került. Az első komponens a szórásnégyzet 29,039%-át, a második komponens a szórásnégyzet 28,522%-át, míg a harmadik komponens 28,143%-ot magyarázta. Ez azt is jelenti, hogy a három komponens szinte azonos varianciát magyaráz a rotáció után. A komponens modell érdekessége, hogy a 15-19 között publikált dokumentumok száma két komponensbe is bekerült.

A parciális korrelációk elemzése: ok-okozat

A parciális korreláció alkalmas arra, hogy egy lineáris modellben két változó közötti korreláció meghatározásánál kiszűrjük a többi változó hatását. Ezt úgy is interpretálhatjuk, hogy a két változó közötti kauzális kapcsolatot térképezzük fel. A 8. táblázatban szerepeltetjük a parciális korrelációkat, amelyek segítségével az ok-okozati kapcsolatokat írjuk le.

A kauzális kapcsolatok feltárásánál az abszolút értékben 0,25 feletti parciális korrelációs értékeket vesszük figyelembe. 0,50 és 0,65 között három érték fekszik, míg 0,30 és 0,50 között további két érték van, és még 0,25 és 0,30 között fekszik öt parciális korreláció A 8. táblázatban kiszíneztük szürkével a vizsgált parciális korrelációkat.

A 2. ábra mutatja a változók közötti ok-okozati összefüggéseket. Fekete színnel a 0,386 és 0,643 közötti kapcsolatokat, míg szürkével a 0,25 és 0,30 közötti korrelációkat jeleztük. Az ábrán azonnal látható, hogy két csoportra esnek szét a változók, mégpedig aszerint, hogy a teljes teljesítményt, vagy csak az utolsó öt év teljesítményét vizsgáljuk. Ebből ebben az elemzésben is elüt a

8. táblázat

Parciális korrelációk

		CIT/DOC _{ALL}	H-I	C-A	DOC ₁₅₋₁₉	CIT/DOC ₁₅₋₁₉	H5-I
PA _{BE}	Szign. (2-oldalú)	-0,264	0,550	-0,150	0,205	0,096	-0,084
	N	,005	,000	,115	,031	,315	,382
CIT/DOC _{ALL}	Szign. (2-oldalú)		0,620	0,008	-0,277	0,386	0,051
	N		,000	,933	,003	,000	,596
H-I	Szign. (2-oldalú)			0,023	0,133	-0,165	0,249
	N			,808	,164	,083	,009
C-A	Szign. (2-oldalú)				0,124	0,403	0,251
	N				,194	,000	,008
DOC ₁₅₋₁₉	Szign. (2-oldalú)					-0,272	0,643
	N					,004	,000
CIT/DOC ₁₅₋₁₉	Szign. (2-oldalú)						0,289
	N						,002

Forrás: saját számítások a Scopus/SciVal adatbázisokból

2. ábra

A változók közötti ok-okozati összefüggések



Forrás: saját számítások a Scopus/SciVal adatbázisokból

társszerzőszám, ami az utolsó öt év hivatkozási mutatóval, vagyis a H5-I-vel és a CIT/DOC₁₅₋₁₉-cel mutat kapcsolatot.

Megállapítható, hogy az az ok-okozati összefüggésrendszer írható le, miszerint a társszerzők számának növekedése növeli az utolsó öt év publikációs számát, viszont a publikációk száma növelheti a hivatkozások számát, majd ezzel együtt a Hirsch-indexet. Megjegyezzük azt is, hogy az említett logikai láncot úgy is interpretálhatjuk, hogy több publikáció több társszerzőt is jelenthet, tehát az összefüggésrendszer nem írható le egy irányított gráffal.

Az előidejűség feltételezne egy dinamikus vizsgálatot, vagy a változók közötti ok-okozati kapcsolatokat logikai úton kellene bizonyítani, pl. a megfordítás logikai kizárásával. Azonban ezeket az elemzéseket egy későbbi dolgozatra hagyjuk.

A kutatók csoportosítása klaszterelemzéssel

A klaszteranalízis elvégzésekor arra voltunk kíváncsiak, hogy a kutatókat milyen csoportba lehet beosztani, felismerhetők-e jellegzetes csoportok. A kérdés megválaszolásához négy elemzést végeztünk, a változók már korábban említett csoportosítása szerint. Ezek

- a 12 Scopus-ból és SciVal-ból vett változókkal,
- a kollineáris változók kiszűrése után maradt 7 változóval,
- a 7 darab 2019-ig elért teljesítményt mutató változóval és
- az 5 darab 2015 és 2019 közötti teljesítménymutatóval.

Várakozásunk az lehet, hogy az első kettő változócsoporthal kapott eredmény hasonló, de az első három változócsoporthal meghatározott klaszterek nagyon eltérhetnek a legutóbb megfigyelt öt év eredményétől, ha a kutató ebben az időintervallumban nem folytatott aktív kutatást, viszont azelőtt igen.

A négy klaszterezésnél minden esetben 13 klaszter meghatározását kértük az SPSS 26-os programtól. Azért

kértünk ennyit, mert ekkora elemszámnál váltak szét a csoportok erősebben. Mind a négy elemzésnél euklideszi távolsággal számoltunk, és a változókat z-score-ra normaltáltuk. Amint azt a többdimenziós skálásnál említettük, ezzel a változók skálaterjedelme közötti eltéréseket akaruk elkerülni. Mielőtt azonban a négy klaszterezés együttes eredményével foglalkoznánk, a 12 változóval kapott csoportokat tekintjük át.

A 12 változóval nyert csoportosítást azért mutatjuk be, mert a másik három elemzés is nagyon hasonló eredményt mutat. Az eredmények a 9. táblázatban láthatók. Azonnal feltűnik, hogy van egy klaszter, a harmadik, amelyben 101 kutató található. Ez a vizsgálatba bevont kutatók 87,1 százaléka, ami nagyon magas. Ezen kívül van még egy három kutatót tartalmazó, egy két kutatót, valamint tíz egy kutatót tartalmazó klaszter.

A klaszteranalízis szempontjából ez azt jelenti, hogy a tizenkét klaszterbe került kutatók valamilyen változó, vagyis kritérium mentén kiemelkednek, vagy rosszabbul teljesítenek. Azonban a magasabban vagy alacsonyabban teljesítő kutatók csak egy-egy szempontban térnek el lényegesen a többi kutatótól. Mivel nem a rangsorolás, hanem a csoportképzés a klaszteranalízis célja, ezért ez az eltérés nem ítéhető meg a szerint, hogy ez pozitív vagy negatív irányban tér-e el. A másik három, más-más változóval elvégzett klaszterezés is hasonló eredményt mutatott, ezért az elemzést nem mutatjuk be. Összegezve azt jelenthető ki, hogy a 116 vizsgálatba bevont kutató közül csak tizenöt tér el lényegesen a többitől. Végig nézve, hogy milyen kritériumok alapján jelenik meg, és milyen irányba, azt mondható ki, hogy az eltérések pozitív irányban jelennek meg, pl. a publikációk, a hivatkozások számában stb. Tehát az eltérő kutatók esetünkben a valamilyen szempontból jobban teljesítők közül kerülnek ki.

A 10. táblázat tartalmazza az összesített eredményeket. A viszonylag magasan választott klaszterszámnál sem vált szét az adatállományunk értelmezhető klaszterekre, azaz csoportokra. A táblázatunknak az az érdekessége, hogy mind a négy klaszterezés eredményét tartalmazza. A táblázat azt mutatja, hogy a klaszterelemzésben mennyi kutató került azonos klaszterbe mind a négy esetben. A táblázatban tehát az adott változóhalmaz klaszterei szerepelnek. A négy elemzéssel 26 csoport jött ki, ötféle elemszámmal. Van egy csoport, amely 78 elemű, vagyis a vizsgálatainkba bevont kutatók alig több, mint kétharmada került bele. Ez arra utal, hogy nagyon homogén az adatállományunk. Ezen kívül egy-egy csoport hét, illetve hat kutatóból tevődött össze, valamint még két csoportba két-két kutató került. Ezzel eddig a kutatók mintegy 70 százalékát csoportosítottuk. A többi 21 csoportba csak egy-egy kutató került. Mindez azt mutatja, hogy nem képezhetünk homogén kutatói csoporto-

9. táblázat

A 12 változós klaszterelemzés csoportjai

Klaszterszám	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	Összesen
Kutatók száma a klaszterben	1	1	101	1	3	1	2	1	1	1	1	1	1	116

Forrás: saját számítások a Scopus/SciVal adatbázisokból

kat. A 10. táblázat alapján leolvasható, hogy az egyes klaszterezési modellek egyes klasztereibe hány kutatót rendelt az algoritmus. A csoportszámokat a 12 változóval képzett klaszterelemzés adja ki.

Kérdezzük meg azt is, hogy az egyes klasztercsoportosítások között milyen asszociációs mérőszámokat tudunk felállítani. Ehhez a Cramér-féle V-t használtuk, amit az egyes elemzések közötti kereszttáblák segítségével határoztunk meg. Az eredményeket a 11. táblázat szemlélteti.

11. táblázat

A Cramér-féle V értékei

	Független változók	7 változó	5 változó
12 változó	0,873	0,857	0,790
Független változók		0,854	0,788
7 változó			0,659

Forrás: saját számítások a Scopus/SciVal adatbázisokból

A 11. táblázat arra utal, hogy az első három klasztercsoport között erős asszociáció mutatható ki, ami azt jelenti,

hogy a változók csoportjaira elég stabil a csoportosítás. Ugyanakkor az öt 2015-2019 közötti változóval képzett csoportosítás kicsit gyengébb kapcsolatot mutat a többi klaszterezés eredményével.

Ennek az lehet az oka, hogy nagyon sok olyan kutatóknak volt, akik a tizenkét (valamint hét és öt) változó mentén valamiben kiemelkedtek a többiekhez képest. Szinte kizárólag a „jobban” teljesítő kutatókat, és szinte egyesével szakította le a klaszteranalízis megoldása. Az sem adott jobb megoldást, ha más csoportosítási technikát vagy távolságdefiníciót választottunk.

Az eredmények diszkussziója

Az elemzések elvégzése után az eredményeket foglaljuk össze.

Az első kutatási kérdésre adott válasz, hogy a kollinearitás vizsgálatával a 12 változót 7-re csökkenthetjük. Az összes hivatkozás (CIT_{ALL}) az egy dolgozatra eső hivatkozással és a Hirsch-indexszel mutat erős korrelációt. A 2015-2019 között kapott hivatkozások (CIT_{15-19}) a társszerzők számával, a 2015-2019 közötti dolgozatokra eső hivatkozással és az ötéves Hirsch-indexszel áll erős lineáris kapcsolatban; az $FWCI_{15-19}$ mutató a 2015-2019 közötti publikációkra eső citációval, az összes folyó-

10. táblázat

A klaszterelemzés csoportjai

Csoport-szám	Kutatók száma a csoportban	12 változó klaszterei	Független változók klaszterei	7 változó klaszterei (összes, 2019-ig)	5 változó klaszterei (2015-2019)
6.	78	3	3	8	3
10.	7	3	3	8	5
5.	6	3	3	3	3
4.	2	3	10	9	1
7.	2	3	3	9	3
1.	1	1	1	1	1
2.	1	2	2	2	2
3.	1	3	3	8	1
9.	1	3	3	8	5
8.	1	3	7	11	3
12.	1	3	7	3	10
11.	1	3	10	9	5
13.	1	3	10	8	13
14.	1	4	4	4	4
15.	1	5	3	5	3
16.	1	5	3	5	5
17.	1	5	5	6	6
18.	1	6	6	7	7
19.	1	7	7	3	8
20.	1	7	7	8	10
21.	1	8	8	9	9
22.	1	9	9	10	11
23.	1	10	10	9	12
24.	1	11	11	12	2
25.	1	12	12	11	5
26.	1	13	13	13	5
Összesen	116				

Forrás: saját számítások a Scopus/SciVal adatbázisokból

iraticikk száma (PA_{ALL}) az összes gazdasági folyóiratban megjelent cikkek számával. Végül az összes dolgozatok száma (DOC_{ALL}) a gazdasági folyóiratcikkek számával, a Hirsch-indexszel és a 2015-2019 közötti dolgozatok számával. Ez annyit is jelent, hogy a nyers publikációs és citációs teljesítménymutatók a fajlagos változókkal való erős lineáris kapcsolat miatt kollineárisnak mutatkoznak, így a nyers mutatók elhagyhatók. Az $FWCI_{15-19}$ mutatót azért zárhatjuk ki a változók közül, mert 2015-2019 közötti fajlagos hivatkozási mutatóval erős a korrelációja. Ezt a multikollinearitási összefüggést a 12 változóra elvégzett többdimenziós skálázási elemzés is alátámasztja annyiban, hogy az egymással kollineáris változók a térképen egymáshoz közel esnek.

A második kutatási kérdés, azaz a korreláció elemzése arra hívja fel a figyelmet, hogy a 12 teljesítménymutató között a közepes és az erős korrelációs hányadosok vannak többségben. Amint azt a VIF-vizsgálat is alátámasztotta, az öt kollineáris változó erős korrelációt mutatott a hét, már kevésbé kollineáris változóval.

A harmadik kutatási kérdésre a válaszunk az lehet, hogy ez az elemzés is alátámasztja a változók és csoportjai közötti erősebb lineáris kapcsolatot. A hét, már kevésbé redundáns változót három komponensbe, azaz látens változóba tette a főkomponens-elemzés. Az első komponens az egy dolgozatra eső hivatkozási mutatókkal mutat erősebb korrelációt, míg a másik két komponens a Hirsch-indexekkel és a dolgozatok számával, amit a harmadik komponensben kiegészít a társszerzők száma. Ez annyit is jelent, hogy minden komponens erős kapcsolatban van valamilyen hivatkozási teljesítménymutatóval.

A negyedik kutatási kérdéssel arra kerestünk választ, hogy milyen ok-okozati kapcsolatot lehet feltárni a hét, nem kollineáris változó között. Azt látjuk, hogy három-három teljesítménymutató szorosan együtt mozog. Ezek a változók publikációk száma, az átlagos hivatkozásszám és a Hirsch-index az életpálya mentén és a 2015-2019 közötti időintervallumban. A társszerzők száma pedig a 2015 és 2019 közötti mutatókhoz kapcsolódik.

Az ötödik kutatási kérdés a kutatók csoportját vizsgálta. A vizsgálatokat négy adatbázis segítségével végeztük arra keresve a választ, hogy a kollinearitás módosítja-e lényegesen a kutatói csoportokat. Arra az eredményre jutottunk, hogy mind a négy hierarchikus klaszterelemzéssel elvégzett vizsgálatban a kutatók döntő többsége, vagyis mintegy kétharmada azonos csoportba került. Ez arra hívja fel a figyelmet, hogy a kutatói többség közel azonos teljesítménymutatóval bír. A maradék egyharmad egy-egy mutató mentén jobban teljesít. Ha csak egy adatbázison végezzük az eredményt, és a választott klaszterszám 13, akkor nagyon feltűnő, hogy egy nagy elemszámú klasztert kapunk, és 12 olyan klasztert, amiben az elemszám nem haladja meg a hármat. Ez a magyar kutatók teljesítménymutatóinak homogenitására utal. A 12 klaszterben olyan kutatók szerepelnek, akik valamilyen mutató mentén kiemelkednek.

Összegzés

A dolgozat célja a magyar gazdasági, vagyis gazdálkodás- és közgazdaságtudományi kutatók teljesítménymutatóinak feltérképezése volt a Scopus és SciVal adatbázisokból nyilvánosan elérhető adatok alapján. A felméréshez először 147 olyan kutatót azonosítottunk, akik legalább öt cikket publikáltak gazdasági folyóiratban. Ezután azt elemeztük, hogy az összes publikált folyóiratcikk hány százaléka volt gazdasági folyóiratbeli cikk, százalékban kifejezve. Azt a kutatót tekintettük gazdasági szakembernek, akinek a folyóiratcikkeinek 40 százaléka gazdasági folyóiratban jelent meg. 116 ilyen kutatót találtunk. Elemzéseinket e kutatók mutatóiból képzett adatállománnyal végeztük el.

Az adatállományban először a változók közötti lineáris kapcsolatot vizsgáltuk. Itt nem csak a lineáris kapcsolat erőssége volt érdekes, de a változók közötti kapcsolat nagysága is, ami a kollinearitás vizsgálatát is feltételezte. Mivel 12 változónk volt, azért a változók számának csökkentésére két lehetőség állt rendelkezésre: varianciainflációs és főkomponens-elemzés. Mindkét elemzés a korrelációs mátrixra alapozódik. Azt az eredményt kaptuk, hogy a 12 változóból elegendő azt a hét változót kiválasztani, amelyek között a lineáris kapcsolat viszonylag alacsony. Ez arra is felhívja a figyelmet, hogy kevesebb változóval is elvégezhető a vizsgálatok. A megmaradt mutatók között a gazdasági folyóiratokban megjelent dolgozatok száma, a társszerzők száma és további öt hivatkozási mutató volt, mint a Hirsch-indexek, publikációkra eső átlagos hivatkozás stb.

A következő, második elemzés a változók közötti kauzális kapcsolatot tekintette, természetesen irány nélkül, a parciális korrelációelemzés segítségével. Azt kaptuk, hogy a dolgozatok száma, az egy dolgozatra eső hivatkozás és a Hirsch-indexek befolyásolják az ok-okozati kapcsolatokat, az időtávától függetlenül.

Végül csoportokat próbáltunk képezni a kutatók között. Ez a vizsgálat négy adatállomány esetén is hasonló eredményt mutatott. A 116 kutató közül 78 mind a négy elemzés esetén azonos csoportba került, vagyis a kutatók mintegy négyötödének a teljesítménye nem mutat lényeges különbséget. A többi kutató egyes változók mentén kiemelkedő eredményt mutatott, ami azt jelentette, hogy nagyon sok egyelemű klaszter állt rendelkezésünkre. Ebből arra következtethetünk, hogy csak néhány kutató teljesítménye tér el lényegesen az átlagtól. A magyar kutatók tehát homogén módon oszthatók csoportokba.

Egy további kutatásban a Scopus-ban megjelent folyóiratcikkek minőségét vehetnénk figyelembe aszerint, hogy a mennyi a Q1-Q4 minőségű folyóiratcikkek száma kutatonként. Ezzel a publikációk minőségét is figyelembe vennénk az elemzéseink során. Azt sem vizsgáltuk meg, hogy az egyes kutatók milyen mértékben járultak hozzá saját egyetemük listában elfoglalt helyéhez. Ehhez definiálnunk kellene a hozzájárulás mértékét az egyes egyetemi rangsorokhoz.

Felhasznált irodalom

Angelis, L., Bassiliades, N., & Manolopoulos, Y. (2019). On the necessity of multiple university rankings. *COLLNET Journal of Scientometrics and Information Management*, 13(1), 11-36. <https://doi.org/10.1080/09737766.2018.1550043>

Csóka, I., Neszveda, G., & Sebestyén, G. (2019). Tudományos teljesítmény mérése a magyar felsőoktatás gazdasági képzéseiben. *Közgazdasági Szemle*, 66(7-8), 751-770. <http://dx.doi.org/10.18414/Ksz.2019.7-8.75>

Dobos, I., Michalkó, G., & Sasvári, P. (2020). Messze még a híd? Kelet-Közép-Európa gazdaságtudományi kutatóinak összehasonlítása=A Bridge Too Far? Comparison of Economics and Management Researchers' Publication Performance in Central and Eastern Europe. *Statisztikai Szemle*, 98(8), 981-1000. <https://doi.org/10.20311/stat2020.8.hu0981>

Dobos, I., Michalkó, G., & Sasvári, P. L. (2021a). The publication performance of Hungarian economics and management researchers: A comparison with the Visegrád 4 countries and Romania. *Regional Statistics*, 11(2), 165-182. <https://doi.org/10.15196/RS110207>

Dobos, I., Urbanovics, A., & Sasvári, P. L. (2021b). A visegrádi négyek, Ausztria és Románia gazdaságtudományi publikációs teljesítményének összehasonlító elemzése. Területi Statisztika, 61(6), 739-768. <https://doi.org/10.15196/TS610603>

Doğan, G., & Al, U. (2018). Is it possible to rank universities using fewer indicators? A study on five international university rankings. *Aslib Journal of Information Management*, 71(1), 18-37. <https://doi.org/10.1108/AJIM-05-2018-0118>

Fauzi, M. A., Tan, C. N. L., Daud, M., & Awalludin, M. M. N. (2020). University rankings: A review of methodological flaws. *Issues in Educational Research*, 30(1), 79-96. [HTTP://WWW.IIER.ORG.AU/IIER30/FAUZI.PDF](http://www.iier.org.au/iier30/FAUZI.PDF)

Hetesi, E., & Kürtösi, Z. (2009). A felsőoktatás képzési teljesítményének mérési problémái. In: Hetesi E., Majó Z. & Lukovics M. (Eds.), *A szolgáltatások világa* (pp. 168-185). Szeged: JATEPress.

Lafi, S. Q., & Kaneene, J. B. (1992). An explanation of the use of principal-components analysis to detect and correct for multicollinearity. *Preventive Veterinary Medicine*, 13(4), 261-275. [https://doi.org/10.1016/0167-5877\(92\)90041-D](https://doi.org/10.1016/0167-5877(92)90041-D)

Liao, D., & Valliant, R. (2012). Variance inflation factors in the analysis of complex survey data. *Survey Methodology*, 38(1), 53-62. <https://www.rti.org/publication/variance-inflation-factors-analysis-complex-survey-data/fulltext.pdf>

Mammadli, A. (2021). Global University Rating Indicators and Suggestion For Establishment of Entrepreneur Universities in Azerbaijan. *InterConf*, (42) 192-210. <https://doi.org/10.51582/interconf.19-20.02.2021.016>

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673-690. <https://doi.org/10.1007/s11135-006-9018-6>

Purkayastha, A., Palmaro, E., Falk-Krzesinski, H. J., & Baas, J. (2019). Comparison of two article-level, field-independent citation metrics: Field-Weighted Citation Impact (FWCI) and Relative Citation Ratio (RCR). *Journal of Informetrics*, 13(2), 635-642. <https://doi.org/10.1016/j.joi.2019.03.012>

Sebrek, S. S. (2020). A magyar gazdasági felsőoktatás egy nyugati doktori fokozattal rendelkező itthon oktató nézőpontjából – kitörési pontok beazonosítása és egy lehetséges cselekvési terv. *Vezetéstudomány*, 51(4), 51-62. <https://doi.org/10.14267/VEZTUD.2020.04.05>

Sheeja, N. K., Mathew, S., & Cherukodan, S. (2018). Impact of scholarly output on university ranking. *Global Knowledge, Memory and Communication*, 67(3), 154-165. <https://doi.org/10.1108/GKMC-11-2017-0087>

Simonovits, A. (2004). Publikációs szelekció a közgazdaságtanban. *Budapesti Könyvszemle-BUKSZ*, 16(02), 158-164. <https://efolyoirat.oszk.hu/00000/00015/00034/pdf/07simon.pdf>

Valentinyi, Á. (2000). A tudomány piaca és a hazai közgazdaságtan. *BUKSZ*, 12, 144-150.

Vörösmarty, G., & Dobos, I. (2020). A vállalatméret hatása a zöldbeszerzési gyakorlatra. *Statisztikai Szemle*, 98(4), 301-323. <http://doi.org/10.20311/stat2020.4.hu0301>

Melléklet

A 12 változó korrelációs mátrixa

	PA _{BE}	CIT/DOC _{ALL}	H-I	C-A	DOC ₁₅₋₁₉	CIT/DOC ₁₅₋₁₉	H5-I	CIT _{ALL}	CIT ₁₅₋₁₉	FWCI ₁₅₋₁₉	PA _{ALL}	DOC _{ALL}
PA _{BE}	1,000	,137	,629	,147	,535	,065	,437	,428	,214	,135	,923	,799
CIT/DOC _{ALL}	,137	1,000	,614	,432	,080	,613	,423	,832	,511	,629	,164	,203
H-I	,629	,614	1,000	,437	,562	,357	,690	,790	,481	,461	,683	,794
C-A	,147	,432	,437	1,000	,415	,611	,643	,415	,880	,649	,318	,303
DOC ₁₅₋₁₉	,535	,080	,562	,415	1,000	,093	,758	,345	,443	,198	,689	,760
CIT/DOC ₁₅₋₁₉	,065	,613	,357	,611	,093	1,000	,468	,396	,750	,942	,115	,074
H5-I	,437	,423	,690	,643	,758	,468	1,000	,513	,726	,555	,606	,559
CIT _{ALL}	,428	,832	,790	,415	,345	,396	,513	1,000	,477	,464	,454	,554
CIT ₁₅₋₁₉	,214	,511	,481	,880	,443	,750	,726	,477	1,000	,781	,361	,303
FWCI ₁₅₋₁₉	,135	,629	,461	,649	,198	,942	,555	,464	,781	1,000	,188	,162
PA _{ALL}	,923	,164	,683	,318	,689	,115	,606	,454	,361	,188	1,000	,844
DOC _{ALL}	,799	,203	,794	,303	,760	,074	,559	,554	,303	,162	,844	1,000

Forrás: saját számítások a Scopus/SciVal adatbázisokból