

Bencsik Gábor¹

A NATO-tagországok védelmi kiadásainak vizsgálata klaszterelemzés módszerével

DOI 10.17047/Hadtud.2020.30.E.91

Absztrakt:

A hosszú távú tervezés egyre inkább megköveteli, hogy egy adott ország helyzetét és védelmi kiadásait ne csak a GDP részarányhoz viszonyítottan ítéljük meg, hanem több tényező figyelembe vételével.

Jelen tanulmányban a NATO által évente közzétételre kerülő, esetünkben a 2019-es becsült adatokra épülő elemzésről olvashatunk. A cikkben betekintést nyerhetünk, hogy a védelmi kiadások szerkezeti megoszlását vizsgálva mely országok és milyen tényezők alapján állnak egymáshoz „legközelebb”. Az alkalmazott módszer: a klaszterelemzés.

Kulcsszavak:

NATO, védelmi kiadások, többváltozós adatelemzés, klaszteranalízis, R programozási nyelv

A Cluster Analysis of Defence Expenditures of NATO Countries

Abstract:

Any long-term planning increasingly requires that defense spending should not only be viewed in relation to its share in GDP, as the situation of a given country can be assessed by taking into account several factors. In this study, an analysis, based on the 2019 estimates and published annually by NATO, is presented. This article provides an insight into which countries are the “closest” to each other, on the basis of factors by examining the structural distribution of defense expenditures. The method used: cluster analysis.

Keywords:

NATO, defence expenditure, multivariate data analysis, cluster analysis, R programming language

Bevezetés

A hosszú távú tervezés egyre inkább megköveteli, hogy egy adott ország helyzetét és védelmi kiadásait ne csak a GDP részarányhoz viszonyítottan ítéljük meg, hanem több tényező figyelembe vételével.

A világ legerősebb katonai szövetsége feladatainak finanszírozása érdekében a NATO-tagországok alapvetően két módon, közvetlen és közvetett hozzájárulások

¹ Nemzeti Közsolgálati Egyetem, Katonai Műszaki Doktori Iskola, doktorandusz – National University of Public Service, Doctoral School of Military Engineering, PhD Student; e-mail: bencsik.gabor@hm.gov.hu; ORCID: 0000-0002-1394-6765

formájában teljesítik a védelmi kiadásokhoz történő hozzájárulási kötelezettségeiket.² Az Észak-atlanti Szerződés 5. cikkelye szerinti kollektív védelemhez történő tagországi hozzájárulások – *valamint az ennek érdekében megvalósuló védelmi költségvetés volumenértékei* – tagországokként nagymértékű különbségeket mutathatnak, tekintve, hogy a hozzájárulások a tagországok részéről önkéntesek. A The World Bank, valamint a NATO honlapján elérhető adatbázisok³ alapján az elmúlt közel 60 évben – *a kiadási főösszeg emelkedése mellett* – a világ GDP részarányos védelmi kiadásvolumene egy meredeken csökkenő pályára állt,⁴ mely alól – *habár évente emelkedő védelmi kiadásokkal rendelkezik, és a GDP részarányos mutatót tekintve az utóbbi években fordulatra lehetünk figyelmesek* – a 2019-es becsült adatok alapján 1 039 628 millió USD (~1040 milliárd USD)⁵ védelmi kiadási főösszeggel rendelkező NATO sem jelentett kivételt (1. és 2. ábra).

² Németországban évente – a Global Firepower (GFP) interneten is elérhető 2020-as statisztikája alapján – a népesség 9,8%-a, 7 905 455 fő éri el a (17 éves) sorozási korhatárt, mindezzel a világszinten vizsgált 138 ország közül az előkelő 4. helyet foglalja el. Az említett honlap által közzétett adatok (*Manpower Reaching Military Age Annually*): Németország lakossága 2020-ban: 80 457 737 fő, melyből évente mintegy 7 905 455 fő éri el a sorozási korhatárt. Ezen adatokat vizsgálva, Németországot csupán India (1), Kína (2) és Japán (3) előzi meg. A CIA honlapján elérhető információk szerint a német haderőhöz (Bundeswehr) történő csatlakozás korhatára 17 év. A termékenységi rátát 1960-tól elemezve megállapítható, hogy az 1964-es 2,54-es érték folyamatosan csökkenő pályára állt és a 2000-es évek elején nem volt kiugró érték (2002-ben és 2003-ban a ráta: 1,34). Szintén az adatelírásra enged következtetni a megadott érték utolsó két számjegye, mely azonosan: 5-5. Vélhetően gépelési hiba. https://www.globalfirepower.com/country-military-strength-detail.asp?country_id=germany; <https://www.cia.gov/library/publications/the-world-factbook/fields/333.html>; <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=DE> (Letöltés ideje: 2020. 02. 15.)

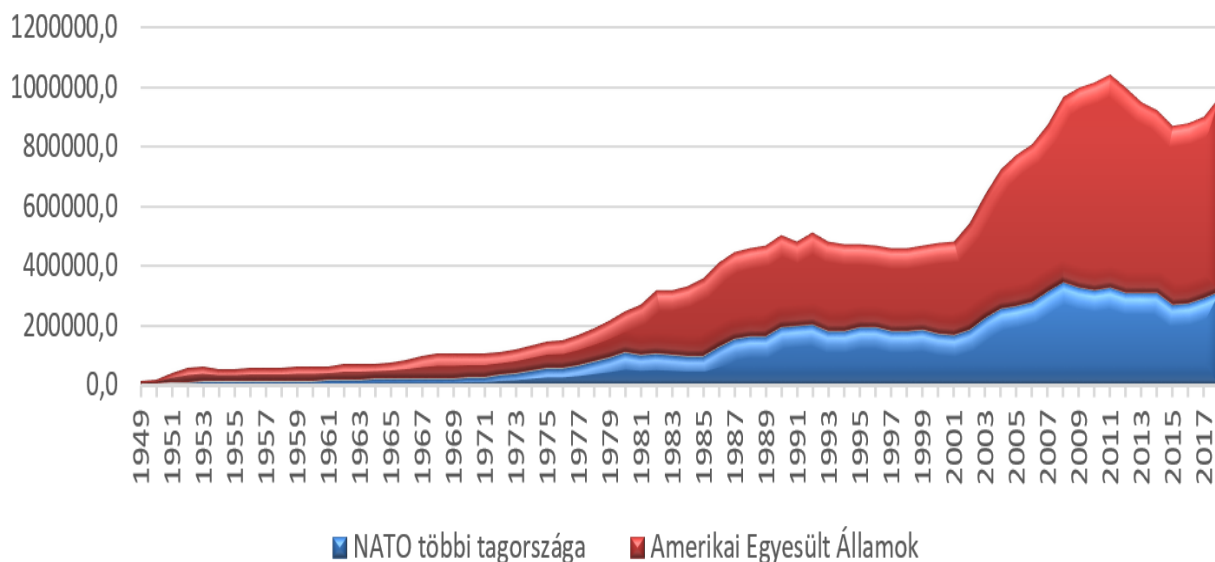
Észtország szintén az említett adatbázis alapján a világ 119. legerősebb hadereje, mely a katonai erővel nem – *így elhanyagolható védelmi kiadással* – rendelkező Izlandot leszámítva, a megmaradó 28 NATO tagországi listán mindössze a 26. pozícióhoz elegendő. Azonban, ha Észtország GDP részarányos védelmi kiadásait vizsgáljuk (2,14%), akkor – holtversenyben az Egyesült Királysággal – a 28-as lista 4. helyét foglalja el. De nem sokkal kedvezőbb a helyzet Spanyolország (20, 7, 27*), Bulgária (62, 19, 2), Olaszország (12, 5, 22) vagy éppen Lettország (102, 24, 8) esetében sem. Egyes országokat leszámítva (pl. Amerikai Egyesült Államok [1, 1, 1], Egyesült Királyság [8, 3, 5], Lengyelország [21, 8, 9] vagy éppen Szlovénia [100, 23, 25]) az előbb felvázolt elemzési megközelítések a helyezések számát tekintve nem sokban térnek el. Akadnak azonban olyan országok is, melyek esetében torzított adatokkal, statisztikai és helyezéssel ellentmondásokkal találkozunk, és az elemzésünk eredménye attól függ, hogy éppen a kétdimenziós kétfaktoros elemzésünk során mely kiválasztott mutatókat viszonyítjuk egymáshoz.

* A számhármassal: (1) a 138 ország közötti GFP általi 2020-as rangsor (20/138), (2) GFP általi rangsor alapján a NATO tagországok** között elfoglalt hely (7/28), (3) NATO tagországokon** belüli GDP részarányú védelmi kiadások mértékét figyelembe vevő hely (27/28). (** Izlandot leszámítva 28 tagország)

³ Military expenditure (% of GDP). <https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS>; Defence expenditures of NATO countries (2013-2019); 2019. november 29; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf (Letöltés ideje: 2020. 02. 15.)

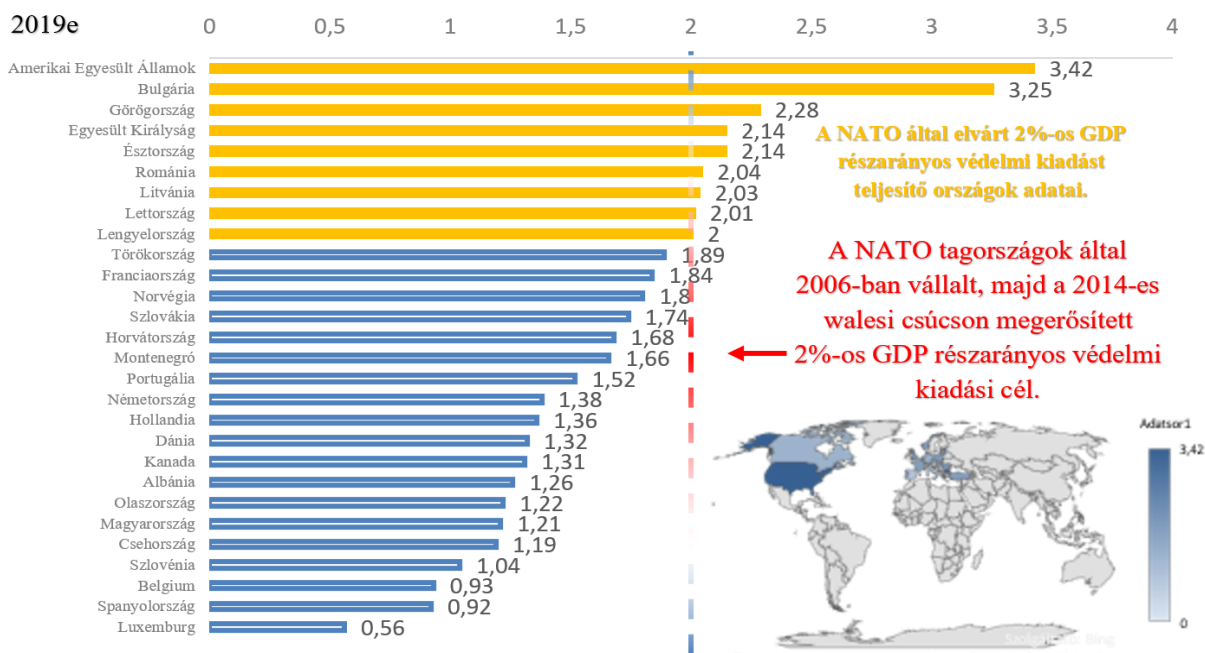
⁴ GDP részarányú védelmi kiadások: 1960: 6,169%; 2018: 2,141%. Lásd bővebben: Military expenditure (% of GDP). <https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS> (Letöltés ideje: 2020. 02. 15.)

⁵ Adatok forrása: Information on defence expenditures (NATO archives). https://www.nato.int/cps/en/natohq/topics_49198.htm (Letöltés ideje: 2020. 02. 15.)



1. ábra

**Az Amerikai Egyesült Államok és a NATO többi tagországnak
védelmi kiadásai (adatok millió USD-ban)⁶**



2. ábra

**A NATO tagországok védelmi kiadásainak alakulása 2019-ben
(GDP részarány; NATO hivatalos becslést adat)⁷**

⁶ Adatok forrása: SIPRI Military Expenditure Database (Data for all countries 1949–2018). <https://www.sipri.org/databases/milex> (Letöltés ideje: 2020. 02. 15.) Az ábra a szerző saját szerkesztése az online adatbázisból – az egyes országok csatlakozási időpontját figyelembe vevő – szűrt adatok alapján.

⁷ Adatok forrása: Defence expenditures of NATO countries (2013-2019); 2019. november 29; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf (Letöltés ideje: 2020. 02. 15.) Az ábra a szerző saját szerkesztése. (Megjegyzés: az elemzésben a

Ezen adatok vizsgálata azonban – *attól függetlenül, hogy a legegyszerűbb összehasonlítási módszereknek tekinthetők* – meglehetősen szubjektív képet mutat.⁸ Egy alapos elemzéshez ugyanis mindig hozzátartozik a szerkezeti vizsgálat, melyben az elemzés folyamatába bevonjuk a rendelkezésre álló adatvolumen szerkezeti összetételét/eloszlását, esetünkben a nemzeti védelmi kiadások mennyiségi és annak GDP részarányos mértékén túlmenően:

- 1) az egy főre jutó védelmi kiadási összegeket;
- 2) a haderők létszámait;
- 3) a védelmi kiadások kategorikus eloszlását, így különösen:
 - a. a főbb eszközök⁹, valamint az azokhoz tartozó K+F(+I)¹⁰ kiadásait;
 - b. a személyi jellegű kifizetéseket;¹¹
 - c. az infrastrukturális kiadásokat;
 - d. valamint a fentiekbe nem tartozó, egyéb (főként működési és karbantartási) kiadásokat.

Az adatmennyiség további bontása azért is szükségszerű, mert az utóbbi időben megfigyelhető és a jelenben is tartó fegyverkezési hajszá¹² a növekvő védelmi kiadások eltérő struktúrában történő felhasználását eredményez(het)i. Az egymással szemben vagy éppen egy oldalon álló felek ugyanis a felhasználás tervezésekor és végrehajtásakor mindig igyekeznek fegyverrendszereik modernizálására, a költség-haszon mutató egyre magasabb szintre történő emelésére, melynek komplexitását nem elegendő csupán egyetlen számadat sorba állításával értékelni. Tudomásul kell venni, hogy a fejlesztések egy komplex rendszerben valósulnak meg, és egyes országok hadi fejlesztései automatikusan indukálják más országok ugyanezen irányú fejlesztési törekvéseit.

A felvázolt szerkezetben megbontott adatmennyiség azonban meglehetősen bonyolulttá és első látásra inkább átláthatatlanná teszi az elemezni kívánt adatstruktúrát, hiszen el kell vonatkoztatnunk a jól megszokott kétdimenziós,

NATO 29 tagországa közül 28 tagország szerepel, tekintettel arra a tényre, hogy Izland esetében értékelhető adatok nem állnak rendelkezésre.)

⁸ Vö. a 2020. 02. 17-én megjelent NATO Spending by Country 2020 online elérhető elemzés, amely nem a NATO honlapján elérhető 2019. 11. 29-i, legfrissebb elemzésére, hanem a fél évvel korábbi, 2019. 06. 25-i dokumentumra támaszkodik. Természetesen az öt hónappal később megjelent elemzésre történő támaszkodás a 2019-es becsült adatokat érintően nagyobb magabiztossággal tehető meg, így már ennél a résznél szükséges alaposan fontolóra vennünk a következtetések helytállóságát. (Csupán a kvantitatív jellegű védelmi kiadások egymásután állításával láthatunk egy rangsort és nem veszi figyelembe az egyéb tényezők jellegét és befolyásoló szerepét [pl. népesség, haderő létszáma, védelmi kiadási főösszeg, stb.]) <http://worldpopulationreview.com/countries/nato-spending-by-country/> (Letöltés ideje: 2020. 02. 20.)

⁹ Főbb fegyverrendszerek (és azok beszerzései).

¹⁰ K+F(+I): Kutatás + Fejlesztés (+ Innováció).

¹¹ A személyi jellegű kifizetések egyes országok esetében magukban foglalják a nyugállományúak nyugdíjait, a teljes és volt személyzet, valamint családjaik szociális ellátásait is. (Bővebben: <https://www.sipri.org/databases/milex/definitions> Hozzáférhető: 2020. 02.15.)

¹² „Tehát ismét fegyverkezési hajszá zajlik [...]”. Dr. Tömösváry Zsigmond nyá. dandártábornok: Új fegyverkezési hajszá? Mindenki Hadtudománya Szabadegyetem; 2019. http://bhd.honvedseg.hu/cikk/mindenki_hadtudomanya_amerikai-orosz_viszony (Letöltés ideje: 2020. 02. 15.)

kétfaktoros [pl. évek ~ védelmi kiadások (USD) vagy tagországok ~ védelmi kiadások (GDP részarány)]; lásd 1–2. ábrák] elemzési metodikáinktól, és az elméletben már kellően megalapozott matematikai, statisztikai és informatikai¹³ eszközökhöz szükséges fordulnunk. A következő fejezetben ezen elemzés végrehajtásakor az R programozási nyelv¹⁴ használatán keresztül nyerünk betekintést a klaszterelemzés említett területet érintő gyakorlati megközelítésébe.

Klaszteranalízis

A kvantitatív és kvalitatív jellegű adathalmazok elemzése során egyre gyakrabban találjuk magunkat szemben a kérdéssel, hogy a többdimenzióba tartozó adatok vajon milyen tulajdonság, jellemző alapján sorolhatók azonos csoportokba. Erre nyújt kiváló segítséget a klaszteranalízis, a többdimenziós csoportok megtalálásának egyik kedvelt módszere. Petrovics fogalmi meghatározása szerint a klaszteranalízis „*olyan dimenziócsökkentő eljárás, amellyel adattömböket – megfigyelési egységeket – tudunk viszonylag homogén csoportokba sorolni, klasszifikálni*”.¹⁵ Kiemeli ugyanakkor, hogy az eljárás célja „*megmutatni, hogy léteznek olyan csoportok, amelyek jobban hasonlítanak egymáshoz, mint más csoportok tagjai*.” Szüle írásában rávilágít, hogy habár a klaszteranalízis népszerűségét matematikai előfeltevéseinek rugalmassága adja, az eredmények értékelésekor a „*jóságvizsgálat*” végrehajtásakor számos nehézségbe ütközhetünk.¹⁶

És hogy miért klaszteranalízis? Mert:

- 1) a megszokott egy-két faktor alapján történő elemzés adatszerkezetét újabb dimenziókkal bővíthetem;
- 2) biztos lehetek abban, hogy (esetlegesen) egyes dimenziók elhagyásával (dimenziócsökkentés) végeredményként ahhoz az osztályozási eredményhez jutok, mely az adott országok (pl. humán és pénzügyi) erőforrás korlátait is

¹³ A számítógép elmúlt évszázadbéli megjelenése hamar lehetővé tette az adatok, adatmennyiségek tömeges, akár több dimenzióra kiterjesztett, többfaktoros elemzését. Elég, ha csak az 1930-as, 1940-es években ismertté vált – Alan Turing (1912-1954) brit matematikus, a modern számítógép-tudomány atyja által kidolgozott – Turing gépre (Colossus), vagy az IBM által évekig fejlesztett Deep Blue szuperszámítógép '90-es évekbéli eredményeire gondolunk, melyek kivétel nélkül rámutattak: az ember igenis képes alulmaradni a gépi gondolkodással és teljesítménnyel szemben. (Gondoljunk csupán az 1996-1997-es összecsapássorozatra, melyben az akkori világbajnok sakkozó, Garry Kasparov ellen vívott csatában végül a Deep Blue győzedelmeskedett. (1996: Kasparov – Deep Blue: 4-2; 1997: Deep Blue – Kasparov: 3½ – 2½). Forrás: <https://www.chess.com/article/view/deep-blue-kasparov-chess> (Letöltés ideje: 2020. 02. 15.)

¹⁴ Az R programozási nyelv két új-zélandi egyetemi oktató, Ross Ihaka és Robert Gentleman által, neveik kezdőbetűi után (R, mint Ross és Robert) elnevezett, 1993-ban megalkotott, a Bell laboratórium által 1976-ban fejlesztett S nyelv alapjain nyugvó, elsődlegesen matematikai és statisztikai elemzésekre használt magas szintű programozási nyelv.

R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing., Vienna, Austria 2019. URL: <https://www.R-project.org/>

¹⁵ Petrovics Petra: Klaszterelemzés az SPSS-ben; előadásjegyzet; Miskolci Egyetem, Gazdálkodástudományi Kar; online: https://nemzetkozi-gazdalkodas.hu/files/950/7_MM_klaszter.pdf (Letöltés ideje: 2020. 02. 28.)

¹⁶ Szüle Borbála: Klaszterszám-meghatározási módszerek összehasonlítása. *Statisztikai Szemle*, 97. évfolyam (2019), 5. szám, p. 422. http://real.mtak.hu/93712/1/2019_05_421.pdf (Letöltés ideje: 2020. 02. 15.)

figyelembe véve az elemzésbe bevont változók alapján történő optimálisabb csoportosítást eredményezi;

- 3) a szintén népszerű diszkriminanciaanalízissel ellentétben itt nem követelem meg az előzetes csoportba tartozást és a kialakított osztályokhoz keresem a magyarázó tulajdonságokat. A klaszteranalízis során – mint említettük – az előfeltevések rugalmasságát használjuk ki.

Látható tehát, hogy a klaszterelemzési eljárás önmagában feltételezi, hogy „*az egyes klasztereken belüli adatok valamilyen dimenzió szerint hasonlítanak egymáshoz, és e dimenzió mentén különböznek a többi klaszter elemeitől*”.¹⁷ A diszkriminanciaanalízistől eltérően azonban szabad kezet kapunk a klaszterek kialakításában, és a kialakult osztályok és azok magyarázó tényezői a dimenziószám növelésével folyamatosan változhat.

Alapadatok lehatárolása, elemzési célok megfogalmazása

Mint korábban láthattuk, az alapsokaságot alkotó NATO-tagországok védelmi kiadásainak mennyiségi és azok GDP részarányos mutatói az elemzés során számos egyéb, az említett mutatókhoz valamilyen formában kapcsolódó tényezőkkel (faktorokkal, dimenziókkal) egészülnek ki. Az elemzésbe bevont adatváltozók homogenitását a NATO hivatalos honlapján 2019. november 29-én közzétett *Defence expenditures of NATO countries (2013–2019)* elnevezésű, Nr.: PR/CP(2019)123 számmal ellátott dokumentumban¹⁸ elérhető 2019-es évre vonatkozó becült adatok biztosítják. A becült adatok választásának indoklása:

- 1) a védelmi kiadások mértékét és megoszlását bemutató, NATO hivatalos dokumentum megjelenési idejét (2019. november 29.) kellő mértékben alkalmasnak tartom a becült adatokban esetlegesen előforduló hibahatár elfogadható szignifikanciaszint alatt tartására;
- 2) a korábbi adatok nem egyértelmű garanciák arra vonatkozóan, hogy az adott adat nem fog változni az idő előre haladásával [vö. a nemzeti valutákban megadott 2014. évi védelmi kiadási mértékeket Horvátországot érintően (egyik sem becült adat):
 - a. 2016. január 28-as publikációs adat alapján:¹⁹ 4.625 millió HRK
 - b. 2019. november 29-ei publikációs adat alapján: 6.113 millió HRK]

¹⁷ Baranyai Dávid, Dr. Mándoki Péter, Dr. Kővári Botond, Dr. Török Ádám: Magyarországi gyalogos és kerékpáros balesetek elemzéseinek módszerfejlődése. *IFFK* (Innováció és fenntartható felszíni közlekedés), 2016., Budapest, 2016. augusztus 29–31. p. 257.; http://kitt.uni-obuda.hu/mmaws/2016/pages/program/papers/Paper_44_Baranyai_et_al_IFFK%202016..pdf (Letöltés ideje: 2020. 02. 20.)

¹⁸ Defence expenditures of NATO countries (2013-2019); 2019. november 29; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf (Letöltés ideje: 2020. 02. 15.)

¹⁹ Defence expenditures of NATO countries (2008-2015); 2016. január 28; Nr.: PR/CP(2016)011; NATO Public Diplomacy Division. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2016_01/20160129_160128-pr-2016-11-eng.pdf (Letöltés ideje: 2020. 02. 15.)

3) meglátásom szerint a legfrissebb adatokra épített elemzések következtetései nem összehasonlíthatók az öt vagy tíz évvel ezelőtti adatokra épített megállapításokkal.

Fontos azonban megjegyezni, hogy a rendelkezésre álló információk közül Izland esetében értékelhető adatok nem állnak rendelkezésre, így a NATO jelenleg 29 tagországa közül 28 tagország képezi a vizsgálat alapját, valamint Olaszország vonatkozásában a 8 faktort érintően 3 faktor [védelmi kiadások részaránya; 1. táblázatban: (g), (h) és (i) oszlopok] esetében nem rendelkezünk adatokkal. Az elemzésbe bevont, rendelkezésre álló nyers adatokat az alábbi, 1. táblázat tartalmazza:

1. táblázat
Az elemzésbe bevonásra kerülő függő és független változók halmaza
(NATO tagországok; 2019. évi becsült adatok)²⁰

NATO-tagország neve (ABC)	Védelmi kiadás (millió USD)	Védelmi kiadás (GDP %)	Védelmi kiadás lakosságra vetítve (USD/fő)	Haderő létszám (ezer fő)	Védelmi kiadás részaránya (%; főbb eszközök, K+F(+I))	Védelmi kiadás részaránya (%; személyi jell.)	Védelmi kiadás részaránya (%; infrastr.)	Védelmi kiadás részaránya (%; egyéb)
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
<i>Albánia</i>	198	1,26	58	6,8	14,42	64,18	1,98	19,41
<i>Amerikai Egyesült Államok</i>	730 149	3,42	2072	1338,1	27,51	38,62	1,27	32,59
<i>Belgium</i>	4 921	0,93	392	25,9	10,78	68,65	1,3	19,27
<i>Bulgária</i>	2 179	3,25	268	24,8	58,97	28,73	3	9,3
<i>Csehország</i>	2 967	1,19	236	25	14,42	53,16	7,04	25,39
<i>Dánia</i>	4 651	1,32	742	18	18,06	48,19	1,74	32,01
<i>Egyesült Királyság</i>	60 761	2,14	985	153,3	22,35	34,52	2,17	40,95
<i>Észtország</i>	670	2,14	430	6,3	17,54	35,47	5,5	41,49
<i>Franciaország</i>	50 729	1,84	710	207,8	24,45	45,78	3,12	26,65
<i>Görögország</i>	4 940	2,28	441	107,6	12,27	76,82	1,25	9,66
<i>Hollandia</i>	12 478	1,36	658	41,7	22,98	45,25	3,19	28,58
<i>Horvátország</i>	1 032	1,68	229	15,1	6,73	72,96	1,64	18,67

²⁰ Adatok forrása: Defence expenditures of NATO countries (2013-2019); 2019. november 29; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division; online: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf; (Letöltés ideje: 2020. 02. 15.) A táblázat a szerző saját összeállítása, szerkesztése.

<i>Izland</i>	NA	NA	NA	NA	NA	NA	NA	NA
<i>Kanada</i>	22 485	1,31	585	69,4	13,34	49,4	3,45	33,81
<i>Lengyelország</i>	11 902	2	295	118,5	24,01	48,47	5,35	22,17
<i>Lettország</i>	724	2,01	325	6,4	24,18	34,91	9,27	31,64
<i>Litvánia</i>	1 107	2,03	346	15,7	30,04	42,36	5,02	22,59
<i>Luxemburg</i>	395	0,56	558	0,9	45,06	30,25	6,71	17,98
<i>Magyarország</i>	2 080	1,21	178	19,7	23,48	41,72	3,52	31,29
<i>Montenegró</i>	92	1,66	127	1,5	14,19	64,17	3,97	17,67
<i>Németország</i>	54 751	1,38	597	182	16,55	44,02	3,82	35,61
<i>Norvégia</i>	7 708	1,8	1384	20,7	29,27	34,23	6,44	30,06
<i>Olaszország</i>	24 482	1,22	385	178,1	20,57	NA	NA	NA
<i>Portugália</i>	3 613	1,52	323	28,6	14,78	69,77	0,32	15,13
<i>Románia</i>	5 050	2,04	225	73,1	25,74	54,79	4,42	15,06
<i>Spanyolország</i>	13156	0,92	264	121	20,72	60,8	1,1	17,39
<i>Szlovákia</i>	1905	1,74	322	11,7	41,68	40,61	1,9	15,8
<i>Szlovénia</i>	581	1,04	253	6,3	7,63	68,65	0,65	23,07
<i>Törökország</i>	13919	1,89	222	435,5	38,6	46,75	2,04	12,61

A többváltozós statisztikai adatelemzésben a hiányzó adatok kezelésére több alternatíva közül választhatunk.²¹ Tekintettel azonban arra a tényre, hogy a GFP 2020-as statisztikája alapján Olaszország a világszinten vizsgált 138 ország közül a 12. helyet foglalja el (mely a vizsgálatban maradt 28 tagország²² közül az 5. pozíció megszerzését jelenti), így Olaszország elemzési folyamatban tartása érdekében a hiányzó adatok kiegészítése szükséges, amelyeket a 2000–2018. közötti tényadatokra (2. táblázat) épített lineáris regresszió alapuló előrejelző függvény használatával hajtunk végre.

²¹ Vö. Kovács Erzsébet: Többváltozós adatelemzés. Budapest, Budapesti Corvinus Egyetem, 2014. http://etananyag.ttk.elte.hu/FileS/downloads/14_KOVACS_E_Tobbvalt_adatelemzes.pdf (Letöltés ideje: 2020. 02. 16.)

²² Izland értékelhető adatokkal nem rendelkezik, így a vizsgált alapadathalmazból az Izlandhoz tartozó sort eltávolítottuk.

2. táblázat
Olaszország védelmi kiadási részarányainak megoszlása (2000–2019)²³
 (* Védelmi kiadások részaránya)

Dátum	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
VKR* (% főbb eszközök, K+F(+I))	14,3	10,3	12,4	12,9	11,7	9,1	7,2	14	12,7	11,3	10,9	11,7	8,9	12,51	10,92	9,72	19,09	20,68	21,12	20,57
VKR* (% személyi jell.)	71,4	72,3	74	72,7	75,3	77,1	81,9	72,8	70,8	73,9	75,1	74,8	77,1	75	76,41	77,55	70,79	67,58	65,66	NA
VKR* (% infrastr.)	1,2	0,9	0,8	1,1	0,6	0,8	0,6	1	1,6	1,4	1,4	1,3	1	1,57	1,4	1,3	0,7	0,94	1,92	NA
VKR* (% egyéb)	13,1	16,4	12,8	13,3	12,4	13	10,3	12,2	14,9	13,3	12,6	12,2	13,1	10,93	11,27	11,42	9,42	10,8	11,3	NA

²³ Adatok forrása:

2000–2002: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2005_06/20090902_p050609.pdf;

2003–2005: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2007_12/20090327_p07-141.pdf;

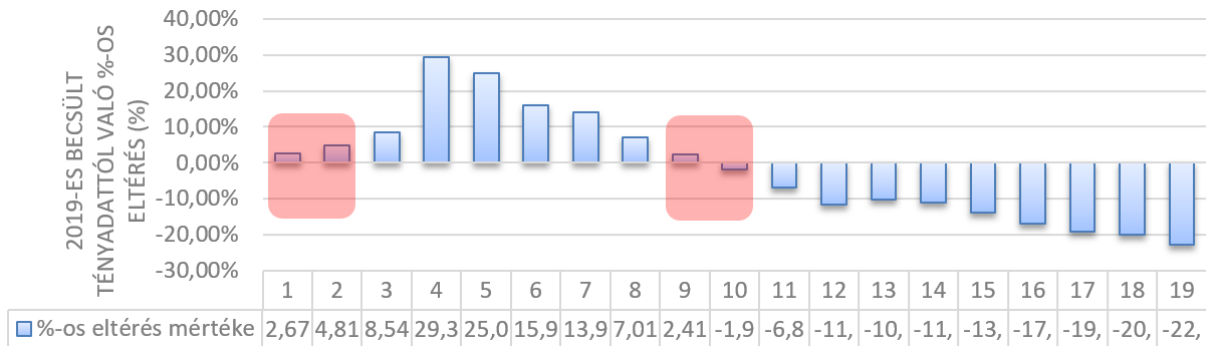
2006–2007: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2011_03/20110309_PR_CP_2011_027.pdf;

2008–2012: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2016_01/20160129_160128-pr-2016-11-eng.pdf;

2013–2018: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf

(Letöltés ideje: 2020. 02. 15.) A táblázat a szerző saját összeállítására, szerkesztésére.

A 2. táblázat adatai már elegendő információt nyújtanak a helyes becslőfüggvény felállítására, melynek megtalálásában segítségünkre lesz egy ismert 2019-es adat [Védelmi kiadások részaránya (% , főbb eszközök, K+F(+I)): 20,57%]. A becslőfüggvény adathalmazra történő alkalmazása esetén a 2018-as adatoktól visszafelé történő adatbevonás eredményét az alábbi, 3. ábra szemlélteti:



3. ábra

Olaszország védelmi kiadásai közül a főbb eszközökre, valamint K+F(+I)-re fordított 2019-es tényadat eltérése a historikus adatokra épített regressziós modelltől²⁴

Látható, hogy összesen 4 esetben (1, 2, 9, valamint 10 historikus adat bevonásakor) tekinthető az abszolút eltérés a megszokott 5%-os hibahatár alattinak, így elfogadhatónak. A lehetőségek közül célszerű a 9-10 historikus adat bevonásával felállított alternatívákat választani, hiszen ebben az esetben a modellben szereplő zajsztint nagyobb mértékben csökkenthető. A kiválasztott modelleket alkalmazva az alábbi választási lehetőségek állnak rendelkezésünkre:

A 3. táblázatból látható, hogy az ismert 2019-es becsült adatot behelyettesítve a legkedvezőbb eredményt a 10 tényező bevonására épített modell biztosítja, ugyanis ebben az esetben az összérték eltérése a 100%-tól csupán 0,4143% (3. táblázatban szürke háttérrel jelölve). Ennek megfelelően a 28 vizsgált elemet tartalmazó, 1 független és 8 függő változóból összeállított adathalmaz hiányzó adataink²⁵ [a 10 tényezőt figyelembe vevő (2018–2009) lineáris regressziós modellre építetten] az alábbi értékekkel rendelkeznek:

²⁴ Az ábra a szerző saját szerkesztése. Magyarázat az ábrában található táblázathoz:

(1): 2,67% → egy historikus adat (2018-as) bevonásakor a regressziós modell eltérése a 2019-es adattól;

(2): 4,81% → kettő historikus adat (2018-2017) bevonásakor a regressziós modell eltérése a 2019-es adattól;

(9): 2,67% → kilenc historikus adat (2018-2010) bevonásakor a regressziós modell eltérése a 2019-es adattól;

Cél: minél nagyobb historikus adatmennyiségre építetten az eltérés minimalizálása → ez biztosítja a modell elfogadhatóságát.

²⁵ 1. táblázat: Olaszország sor (g), (h) és (i) oszlopok.

3. táblázat

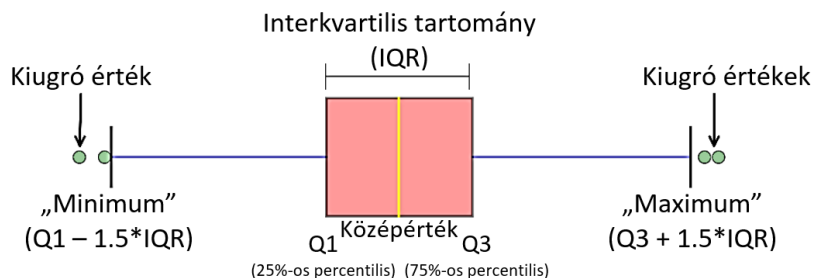
A 9 és 10 historikus adatra épített regressziós függvénnyel kapott eredmények a 2019-es tényadatra épített összehasonlításban²⁶

	Dátum	VKR (% főbb eszközök, K+F(+I))	VKR (% személyi jell.)	VKR (% infrastr.)	VKR (%, egyéb)	Σ	100%-tól való teljes eltérés
reg. fgv. értékeivel	2019(9)	21,0664	67,5414	1,2919	10,0931	99,9928	0,0072%
	2019 (10)	20,1727	68,5860	1,2660	9,9940	100,0187	0,0187%
2019-es tényadat behelyettesítésével	2019 (9)	20,5700	67,5414	1,2919	10,0931	99,4964	0,5062%
	2019(10)	20,5700	68,5860	1,2660	9,9940	100,416	0,4143%

- a) VKR (%
főbb eszközök, K+F(+I)): 20,57%
 b) VKR (%
személyi jell.): 68,59%
 c) VKR (%
infrastr.): 1,27%
 d) VKR (%
egyéb): 9,99%

Kiugró értékek kezelése

Miután az adatbázisban kezeltük a hiányzó értékeket, a következő lépés az adattisztítás, a kiugró értékek, vagyis az outlierok kezelése. Outliereknek nevezzük azokat az értékeket, melyek az adatkészlet többségétől nagyfokú eltérést mutatnak. Statisztikailag:



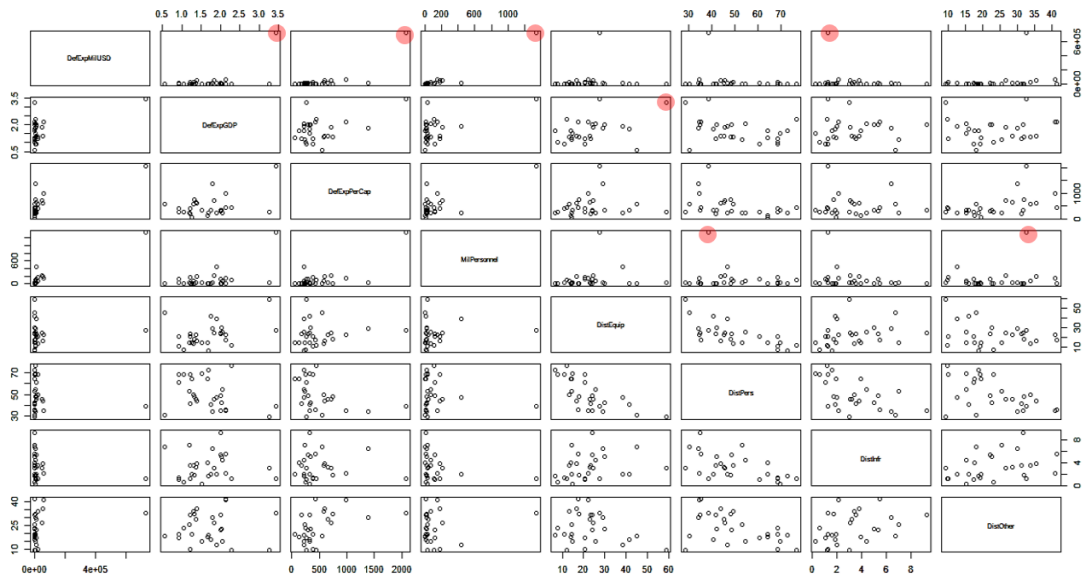
4. ábra

Outlierek szemléltetése boxplot diagrammal²⁷

Látható tehát, hogy minden olyan adat, amely jelentősen ($<Q1-1,5*IQR / Q3+1,5*IQR<$) eltér az adathalmaz többi elemének átlagától, statisztikailag outliernek tekintendő és egyes esetekben kezelésük szükséges. Adatbázisunkat tekintve (Izland értékeit eltávolítva és a becsült értékeket behelyettesítve) az alábbi plot diagramot kapjuk:

²⁶ A táblázat a szerző saját szerkesztése.

²⁷ Forrás: Michael Galarnyk: Understanding boxplots alapján; 2018. szeptember 12. <https://towardsdatascience.com/understanding-boxplots-5e2df7cbcd51> (Letöltés ideje: 2020. 02. 29.)



5. ábra

Az elemzésbe bevont adatbázis plot diagramja²⁸

A 5. ábrán jól látható, hogy számos esetben rendelkezünk kiugró értékekkel [néhányat (változó-páronként egyet) az ábrán piros körlappal jelöltem]. Ezen adatok értékelése azonban a statisztikai modell figyelembevételével egyedi elbírálás tárgyát képezi. A kiugró értékek keresésének ugyanis elsődleges célja az adatbázis esetleges hibáinak feltárása [pl. elgépelésből adódóan (lásd a bevezetőben hivatkozott németországi adat²⁹)]. Ennek megfelelően mérlegelnünk szükséges, ugyanis egy esetlegesen feltárt hiba esetén az elemzésből történő kizárás melletti döntés, vagy a becslésen alapuló adatkorrekció elengedhetetlen. Akadnak továbbá olyan esetek is, melyek nem téves adatgyűjtésből származnak, azonban az alapsokasághoz viszonyított értékeik az aránytalanság következtében mérlegelendő. A védelmi kiadások (millió USD) függő változót tekintve az alábbi eredményeket kapjuk:

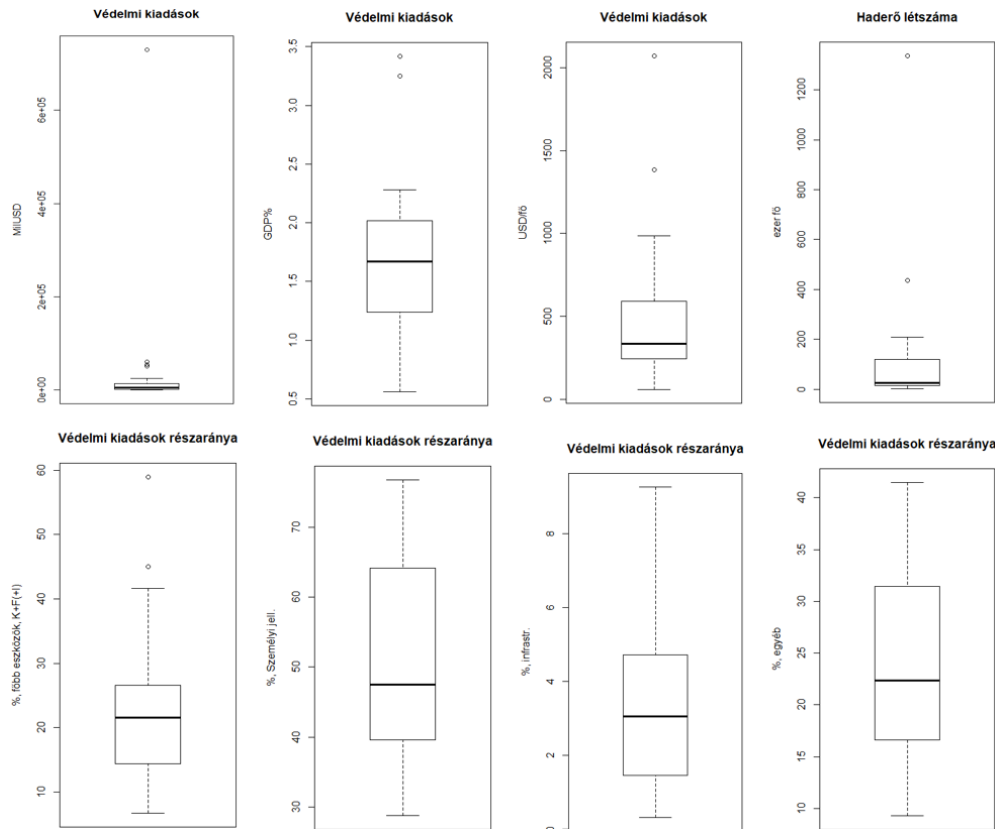
\$stats	[1] 92.0 1069.5 4786.0 13537.5 24482.0
\$n	[1] 28
\$conf	[1] 1063.156 8508.844
\$out	[1] 730149 60761 50729 54751

Látható, hogy az elemzésbe bevont 28 tagország védelmi kiadásait (millió USD) érintő outlier meghatározás 4 kiugró értéket jelez, melyek rendre: 730.149 (Amerikai Egyesült Államok), 60.761 (Egyesült Királyság), 50.729 (Franciaország), 54.751

²⁸ A szerző saját szerkesztése az R programnyelv használatával (Package 'graphics' version 3.6.2.). A korábban bevezetett 8 függő változót az ékezetek és a megnevezés-hosszúság elkerülése érdekében az ábrán azok angol nyelvű megnevezésének rövidítésével helyettesítem.

²⁹ „Németországban évente – a Global Firepower (GFP) interneten is elérhető 2020-as statisztikája alapján – a népesség 9,8%-a, 7 905 455 fő éri el a (17 éves) sorozási korhatárt, mindezzel a világszinten vizsgált 138 ország közül az előkelő 4. helyet foglalja el.”

(Németország). Itt bármilyen megengedők is vagyunk, be kell látnunk, hogy az Amerikai Egyesült Államok ezen adata rendkívül extrém értéket képvisel. A 6. ábra jól rámutat, hogy elemzésünket érintően is számos hasonló esettel állunk szemben.



6. ábra

Az adatbázis boxplotjai³⁰

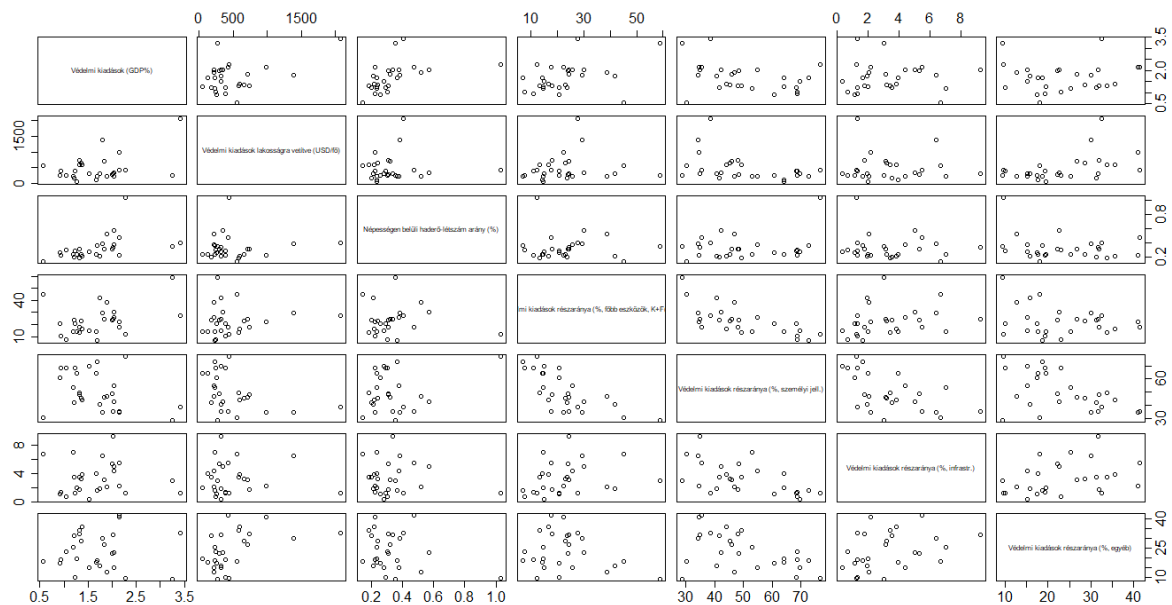
A 6. ábrán feltüntetett boxplotok rávilágítanak arra a tényre, hogy egyes országok esetében extrém értékekkel is számolnunk kell [pl. 1. ábrán bemutatott aránytalanság miatt a védelmi kiadások (millió USD) ábra extrém (legfelső) értékét az Amerikai Egyesült Államok képviseli]. Tekintettel azonban arra, hogy az elemzésből nem célozom kizárni egyik országot sem, így olyan megoldást szükséges alkalmazni, mely megfelelő szinten biztosítja az adatok összehasonlíthatóságát, valamint „*tompítja*” a kiugró értékek adatbázisra gyakorolt hatását.

Ennek megfelelően az eredeti, nyers adatbázison az alábbi módosításokat hajtom végre:

- 1) a védelmi kiadások (millió USD) függő változót kiveszem az elemzésből. Ennek visszatükröződését a védelmi kiadások lakosságra vetítve (USD/fő) fogja képviselni;
- 2) a haderő létszámát arányosítom a teljes lakossághoz.³¹

³⁰ A szerző saját szerkesztése az R programnyelv használatával (Package 'graphics' version 3.6.2.).

A módosítások eredményeképp a 28 vizsgált elemet tartalmazó adathalmaz 1 független és 7 függő változó általi összetételre módosul. A plot diagram az alábbiaképpen alakul:



7. ábra

A módosított adatbázis plot diagramja³²

Az összehasonlíthatóság biztosítása – Skálatranszformáció és távolságmeghatározás

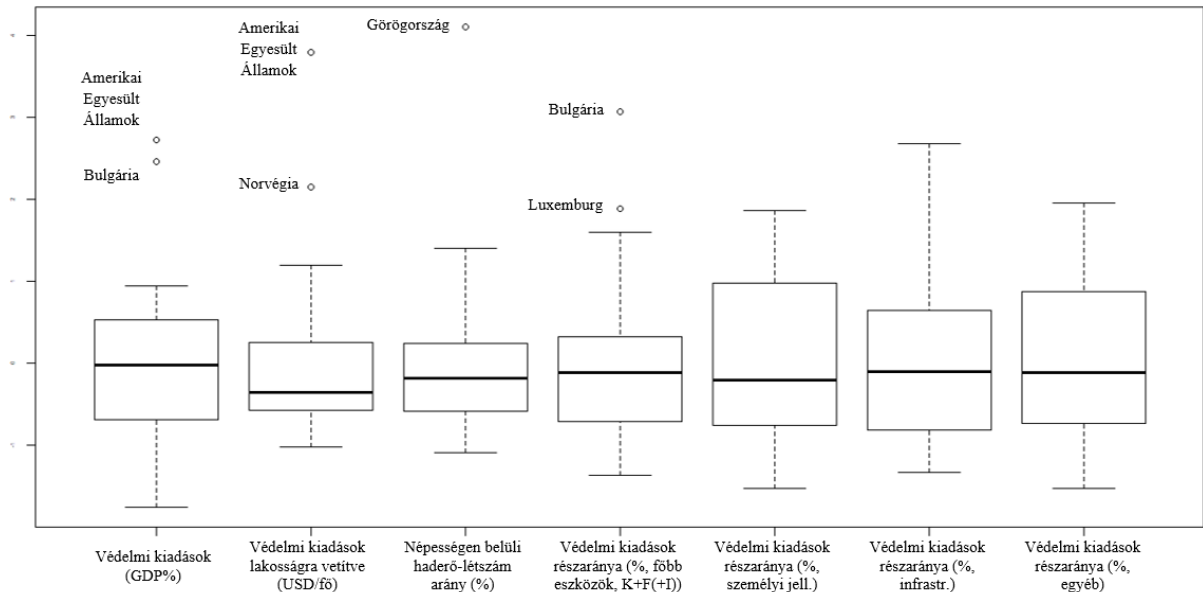
A korábbi fejezetekben ismertetett adatbázist elemezve láthattuk, hogy az azokban foglalt 8 függő változó nemcsak, hogy rendre eltért egymástól, de az eltérő skáláknak és az extrém kiugró értékeknek köszönhetően az eloszlások átlaga és szórása nagyfokú eltérést feltételez. Kijelenthetjük, hogy ebben a formában a nyers adataink heterogén adatbázisnak tekinthetők. A homogenitás biztosításához összehasonlíthatóvá kell tennünk az adatokat, melynek folyamán egy lineáris transzformáció végrehajtásával biztosítjuk az eltérő dimenziók közös nevezőre hozását. Az alkalmazott módszer a magyar származású statisztikus, szántói Kőrösy József (1844–1906) által kidolgozott standardizálás.

Standardizálás alkalmazása esetén (hasonlóan a minimum-maximum normalizáláshoz) a meglévő elemeink függő változónkénti skáláit, az eloszlás megtartása mellett összehangoljuk.³³ Az eljárással biztosítjuk, hogy a függő változók átlaga 0, míg standard hibájuk (szórásuk) várhatóan 1 legyen. A normalizálási eljárás végrehajtásával a keletkező, elemezni kívánt táblánk az alábbi:

³¹ A teljes lakosság adatainak forrása: World Population Review, Total Population by Country 2020. <http://worldpopulationreview.com/countries/> (Letöltés ideje: 2020. 02. 29.)

³² A szerző saját szerkesztése az R programnyelv használatával (Package 'graphics' version 3.6.2.).

³³ Az eljárás során az átlagot kivonjuk az egyes értékekből és a különbséget elosztjuk a szórással.



8. ábra
A normalizált adatbázis boxplotjai³⁴

4. táblázat
Az elemzésbe bevont független és módosított, normalizált függő változók halmaza
(Nyers adatok: 1. táblázat)³⁵

NATO-tagország neve (ABC)	Védelmi kiadás (GDP%)	Védelmi kiadás lakosságra vetítve (USD/fő)	Népesége n belüli haderő-létszám arány (%)	Védelmi kiadás részaránya (% főbb eszközök, K+F(+I))	Védelmi kiadás részaránya (% személyi jell.)	Védelmi kiadás részaránya (% infrastr.)	Védelmi kiadás rész-aránya (% egyéb)
Albánia	-0.664684047	-1.0230202	-0.56052720	-0.716669254	0.96996625	-0.59279280	-0.43453165
Amerikai Egyesült Államok	2.724868328	3.7901080	0.44571065	0.393719717	-0.83208309	-0.91121270	0.99266272
Belgium	-1.182532327	-0.2248152	-0.62834884	-1.025440519	1.28511338	-0.89775834	-0.44969153
Bulgária	2.458098002	-0.5211548	0.13704497	3.062385648	-1.52935492	-0.13534448	-1.52929151
Csehország	-0.774530652	-0.5976295	-0.57275019	-0.716669254	0.19302636	1.67650965	0.21301178
Dánia	-0.570529814	0.6116271	-0.11305198	-0.407897989	-0.15737213	-0.70042769	0.92985750
Egyesült Királyság	0.716244698	1.1923571	-0.61325774	-0.043988999	-1.12114421	-0.50758183	1.89792409
Észtország	0.716244698	-0.1340015	0.85014522	-0.452008170	-1.05416663	0.98585238	1.95639791

³⁴ A szerző saját szerkesztése az R programnyelv használatával (Package 'graphics' version 3.6.2.).

³⁵ A szerző saját szerkesztése.

<i>Franciaország</i>	0.245473535	0.5351524	-0.07061411	0.134148269	-0.32728366	-0.08152703	0.34945069
<i>Görögország</i>	0.935937908	-0.1077133	4.10509042	-0.899047886	1.86112054	-0.92018228	-1.49030896
<i>Hollandia</i>	-0.507760326	0.4108810	-0.51367954	0.009452182	-0.36465010	-0.05013352	0.55844046
<i>Horvátország</i>	-0.005604419	-0.6143584	0.20372639	-1.368990964	1.58898007	-0.74527557	-0.51466244
<i>Kanada</i>	-0.586222187	0.2364230	-0.85793246	-0.808282706	-0.07206384	0.06647096	1.12477024
<i>Lengyelország</i>	0.496551489	-0.4566292	-0.10769404	0.096824270	-0.13763137	0.91858057	-0.13566545
<i>Lettország</i>	0.512243861	-0.3849342	0.02722153	0.111244906	-1.09364815	2.67661724	0.88979211
<i>Litvánia</i>	0.543628605	-0.3347476	1.40241180	0.608332711	-0.56840294	0.77058258	-0.09018581
<i>Luxemburg</i>	-1.763150095	0.1718974	-1.08984708	1.882438315	-1.42219080	1.52851166	-0.58937899
<i>Magyarország</i>	-0.743145908	-0.7362400	-0.75236208	0.051865817	-0.61352468	0.09786447	0.85189241
<i>Montenegró</i>	-0.036989163	-0.8581215	-0.54340475	-0.736179526	0.96926122	0.29967990	-0.62294729
<i>Németország</i>	-0.476375582	0.2651010	-0.66684686	-0.535987168	-0.45136844	0.23240809	1.31968297
<i>Norvégia</i>	0.182704047	2.1459013	0.31720971	0.543015713	-1.14159000	1.40742240	0.71870204
<i>Olaszország</i>	-0.727453536	-0.2415441	-0.21751951	-0.194981540	1.28088321	-0.91121270	-1.45457496
<i>Portugália</i>	-0.256682372	-0.3897138	-0.30278161	-0.686131437	1.36407642	-1.33726751	-0.89799082
<i>Románia</i>	0.559320977	-0.6239177	0.27390042	0.243575448	0.30794578	0.50149534	-0.90557076
<i>Spanyolország</i>	-1.198224699	-0.5307141	-0.42527168	-0.182257450	0.73166708	-0.98745409	-0.65326705
<i>Szlovákia</i>	0.088549814	-0.3921037	-0.68757836	1.595722140	-0.69178269	-0.62867110	-0.82543997
<i>Szlovénia</i>	-1.009916233	-0.5570023	-0.16479123	-1.292646421	1.28511338	-1.18926952	-0.03820908
<i>Törökország</i>	0.323935396	-0.6310872	1.12579817	1.334454147	-0.25889603	-0.56588407	-1.17086865

A 4. táblázat és a 8. ábra adatait áttekintve megállapítható, hogy az adatbázisunk készen áll a klaszterelemzés alkalmazására. Vannak ugyan kiugró értékek, de azok nem adatelírásból származnak, és megszüntetésük adatvesztéssel [sor(ok) vagy oszlop(ok) eltávolításával, így az elemzésbe bevont függő vagy független változók mennyiségének csökkenésével] járna. Láthatjuk továbbá, hogy a kiugró értékek sem konkrétan egy tagországhoz köthetők, így bízunk benne, hogy sikerült kezelnünk az Amerikai Egyesült Államok jelenlévő és kimutatható dominanciáját. Ez azért is fontos, mert mint a tanulmány elején említettem, az elemzés elsődleges célja a szerkezetvizsgálatra alapozott elemzés végrehajtása.³⁶ A valódi különbséget azonban, vagyis hogy az adatok homogenitását biztosító standardizálás végrehajtásával hova jutottunk, az 5. és 7. ábrák összehasonlításával kapjuk. Az adattisztítás és normalizálás eredményeképp összehasonlítható adatokat kaptunk, egyik adat sem dominálja le a másikat, a kiugró értékek elfogadhatók. A

³⁶ Az outlierok kezelésének lehetőségéről lásd bővebben: Karen Grace – Martin: Outliers: To Drop or Not to Drop; The Analysis Factor (Making Statistics, Make Sense). <https://www.theanalysisfactor.com/outliers-to-drop-or-not-to-drop/> (Letöltés ideje: 2020. 03. 04.)

tovább lépéshez tekintsük át, hogy az egyes változók egymással milyen szoros kapcsolatban állnak, mennyire korrelálnak³⁷ (ez alapvetően determinálja a távolságmódszer kiválasztását):

5. táblázat
A függő változók korrelációs mátrixa³⁸

	Védelmi kiadás (GDP%)	Védelmi kiadás lakosságra vetítve (USD/fő)	Népessége n belüli haderőlétszám arány (%)	Védelmi kiadás részaránya (% főbb eszközök, K+F(+I))	Védelmi kiadás részaránya (% személyi jell.)	Védelmi kiadás részaránya (% infrastr.)	Védelmi kiadás részaránya (% egyéb)
Védelmi kiadás (GDP%)	1.000000000	0.42121520	0.47779196	0.38525118	-0.3467259	0.008841989	0.03720918
Védelmi kiadás lakosságra vetítve (USD/fő)	0.421215200	1.00000000	0.05643904	0.11223735	-0.3994857	-0.026218514	0.47596113
Népességen belüli haderőlétszám arány (%)	0.477791961	0.05643904	1.00000000	-0.03157562	0.2099479	-0.097014666	-0.25903936
Védelmi kiadás részaránya (% főbb eszközök, K+F(+I))	0.385251178	0.11223735	-0.03157562	1.00000000	-0.7054040	0.235140219	-0.25025802
Védelmi kiadás részaránya (% személyi jell.)	-0.346725918	-0.39948569	0.20994792	-0.70540404	1.00000000	-0.567017810	-0.49629157
Védelmi kiadás részaránya (% infrastr.)	0.008841989	-0.02621851	-0.09701467	0.23514022	-0.5670178	1.000000000	0.32792052
Védelmi kiadás részaránya (% egyéb)	0.037209176	0.47596113	-0.25903936	-0.25025802	-0.4962916	0.327920515	1.00000000

Az 5. táblázatban látható, hogy az elemzett függő változók között akadnak a közepesnél kicsit magasabb korrelációval rendelkező változópárok (pl.: -0.7054040),

³⁷ Atáblázatban szereplő értékek (x) magyarázata:

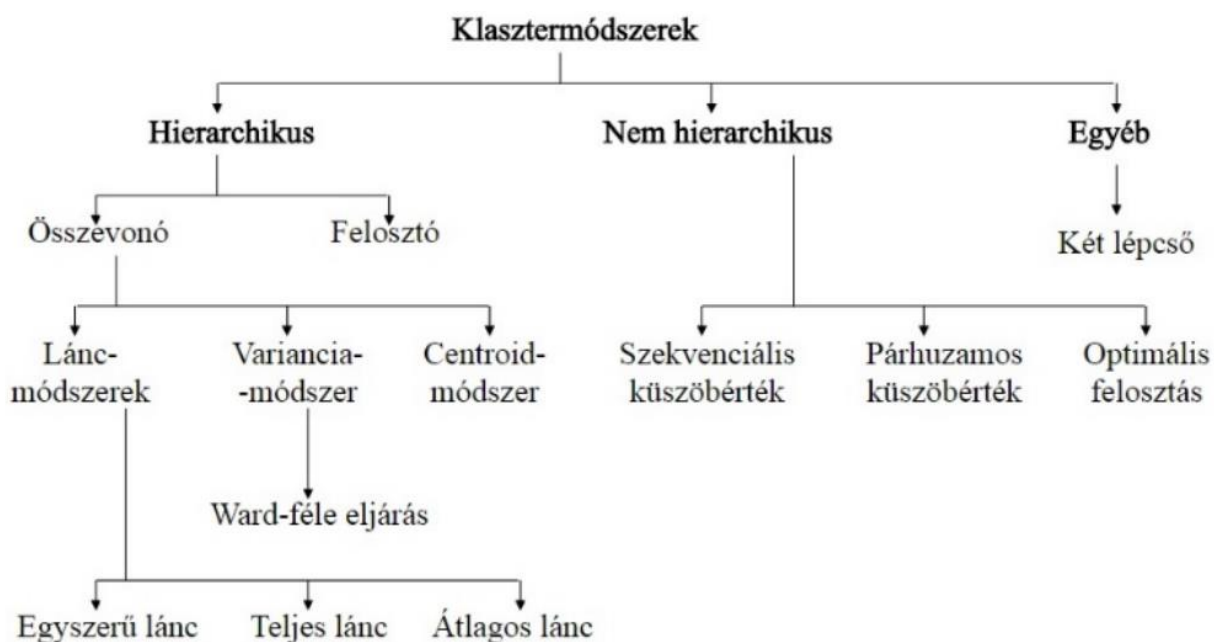
x = 0	 x =]0,0,2]	 x =]0,2-0,4]	 x =]0,4-0,7]	 x =]0,7-0,9]	 x =]0,9-1]
nincs lineáris kapcsolat	gyenge korreláció, szinte elhanyagolható kapcsolat	biztosnak tekinthető, de gyenge kapcsolat	közepes erősségű, jelentős kapcsolat	magas korreláció, erős kapcsolat	kiemelkedően magas korreláció, erős függő kapcsolat

³⁸ A szerző saját szerkesztése az R programnyelv használatával (Package 'stats' version 3.6.2.).

melyek az adott eseteket érintően a jelentős függőségi kapcsolatot feltételezik. Az 5. táblázatban szereplő adatok nagy része azonban ezen erősségű függőségi kapcsolattól elmarad (jellemzően gyenge, vagy közepes erősségű kapcsolatok figyelhetők meg), így a metrika használatát érintően a további fejezetek és számítások alapját az Euklideszi távolságmérték³⁹ képezi.

Klasztermódszer kiválasztása, klaszterszám meghatározása

A klaszteranalízis további lépése a klasztermódszer kiválasztása, valamint szükség esetén a klaszterszám meghatározása. A klasztermódszer kiválasztására számos lehetőségünk kínálkozik, melyek áttekintése érdekében vessünk egy pillantást az alábbi, 9. ábrára:



9. ábra

Klasztermódszerek csoportosítása⁴⁰

A szakirodalomban és a gyakorlatban azonban leggyakrabban három klasztermódszert hivatkoznak és alkalmaznak:

- a hierarchikus teljes láncmódszerrel előállított klaszterezést,
- a hierarchikus átlagos láncmódszerrel előállított klaszterezést, valamint
- a nem hierarchikus k-közép klaszterezést.⁴¹

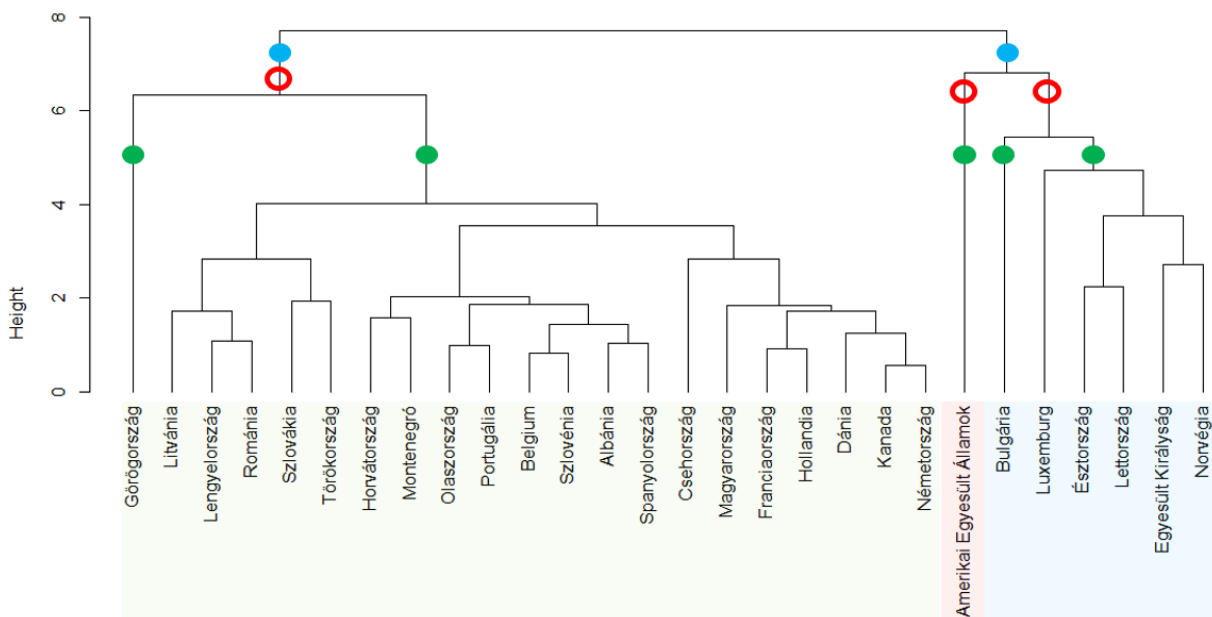
³⁹ Euklideszi metrika: két pont távolsága a két pontot összekötő szakasz hossza.

Koordináta geometriában ('a' és 'b' pontok távolsága): $d(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$

⁴⁰ Forrás: Kovács Erzsébet: Többváltozós adatelemzés. Egyetemi jegyzet. Budapest, Budapesti Corvinus Egyetem, 2014. Idézi: Baranyai Dávid, Dr. Mándoki Péter, Dr. Kővári Botond, Dr. Török Ádám: Magyarországi gyalogos és kerékpáros balesetek elemzéseinek módszerfejlődése. *IFFK* (Innováció és fenntartható felszíni közlekedés), 2016. augusztus 29–31. p. 258.; http://kitt.uni-obuda.hu/mmaws/2016/pages/program/papers/Paper_44_Baranyai_et_al_IFFK%202016..pdf; (Letöltés ideje: 2020. 02. 20.)

Lehetőség van azonban arra is, hogy az elemzés során ún. vegyes megoldást, vagyis különböző algoritmusra épített klaszterezési csoport kombinációját alkalmazzuk. Esetünkben ezen lehetőséget kihasználva a hierarchikus teljes-, valamint átlagos láncmódszerrel előállított klaszterezést alkalmazzuk.⁴²

A teljes-, valamint az átlagos láncmódszerek a klaszterezési eljárások hierarchikus csoportján belül az összevonó, vagyis az alulról felfelé építő módszerek közé tartoznak. A módszer lényege, hogy a függő változókat egytől-egyig különböző klaszterként határozza meg, majd a kiválasztott eljárási metodikának megfelelően (egymáshoz mért távolságuk alapján)⁴³ összevonja az egyes klasztereket. Az összevonási folyamatok ábrázolását ún. dendrogramon (összevonási fa) láthatjuk. Esetünkben az elemzésbe bevont 28 NATO tagországot tartalmazó dendrogramok a klaszterek kialakítására teljes láncmódszerrel készített dendrogramok esetében 2 (10. ábra kék színnel jelölve), valamint 5 csoport kialakítását (10. ábra zöld színű jelölés), míg átlagos láncmódszerrel készített dendrogramok esetén 4 klaszter kialakítását szemléltetik optimálisnak (11. ábra piros színnel jelölve).



10. ábra

Teljes láncmódszerrel készített dendrogram⁴⁴

⁴¹ A módszerekről érthetően és bővebben Valentina Alto: Unsupervised Learning: K-means vs. Hierarchical Clustering. <https://towardsdatascience.com/unsupervised-learning-k-means-vs-hierarchical-clustering-5fe2da7c9554>

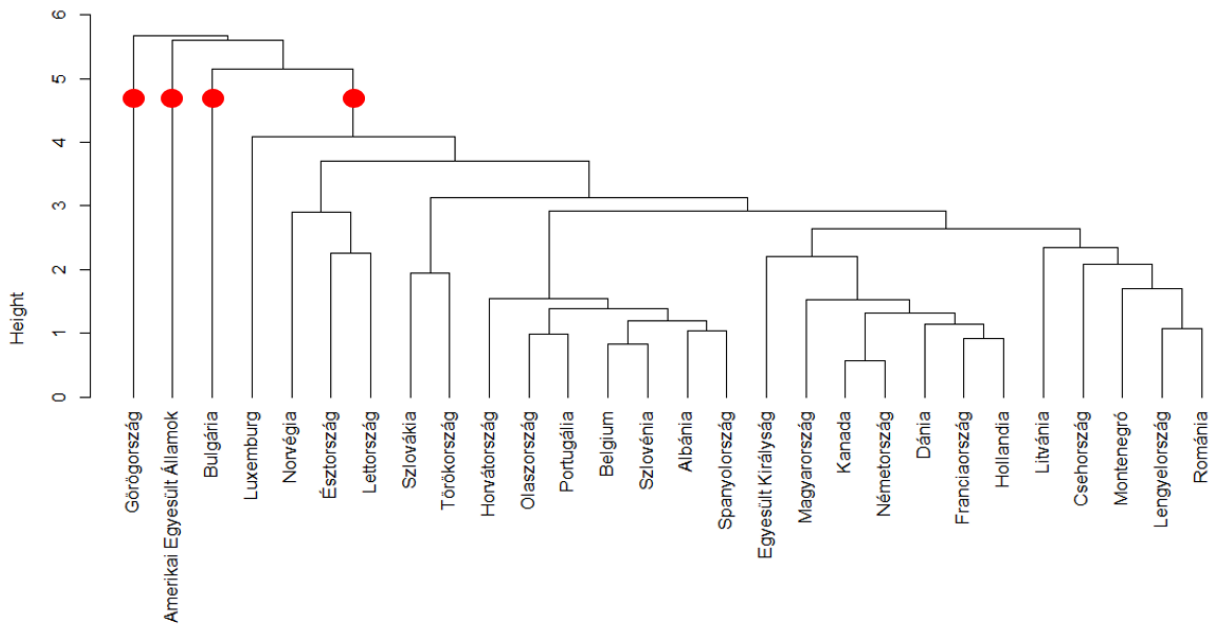
⁴² A kiválasztott (hierarchikus klaszterezési) módszerek alapvető tulajdonsága, hogy (szemben a nem hierarchikus eljárásokkal) nem követelik meg előfeltételként a klaszterszámok előzetes ismeretét.

⁴³ Távolság meghatározása: (Euklideszi metrika alapján)

Teljes láncmódszer esetén: két klaszter távolsága = klasztereken belüli két legtávolabbi pont távolsága.

Átlagos láncmódszer esetén: két klaszter távolsága = a megfigyelési egységek páronkénti távolságának átlaga.

⁴⁴ A szerző saját szerkesztése az R programnyelv használatával (Package 'stats' version 3.6.2.).



11. ábra

Átlagos láncmódszerrel készített dendrogram⁴⁵

Mint korábban említettem, a hierarchikus klaszterezés egyik legnagyobb előnye, hogy a módszer nem kívánja meg a csoportszám előzetes ismeretét, ugyanis az összevonó láncmódszerek alulról építik fel a klasztereket, az egymáshoz mért távolságuk (több féle!) alapján. A másik oldalról ugyanakkor meg kell említenünk, hogy ezen előny hátránya a dendrogramok leolvasásának rugalmas kezelésénél, a klaszterszám meghatározásánál jelentkezik. A klaszterszám meghatározása azon eljárások összessége, mely alapján meghatározzuk, hogy a függő változók értékeinek (esetünkben 7 db) „*egymáshoz állása*”⁴⁶ alapján hány csoportba (klaszterbe) soroljuk a vizsgálatba bevont független változókat. Ezen számadat meghatározásához (támogatásához) alapvetően két közkedvelt módszer áll rendelkezésünkre:

- a) a klaszterkönyök eljárás és
- b) a sziluettmódszer.

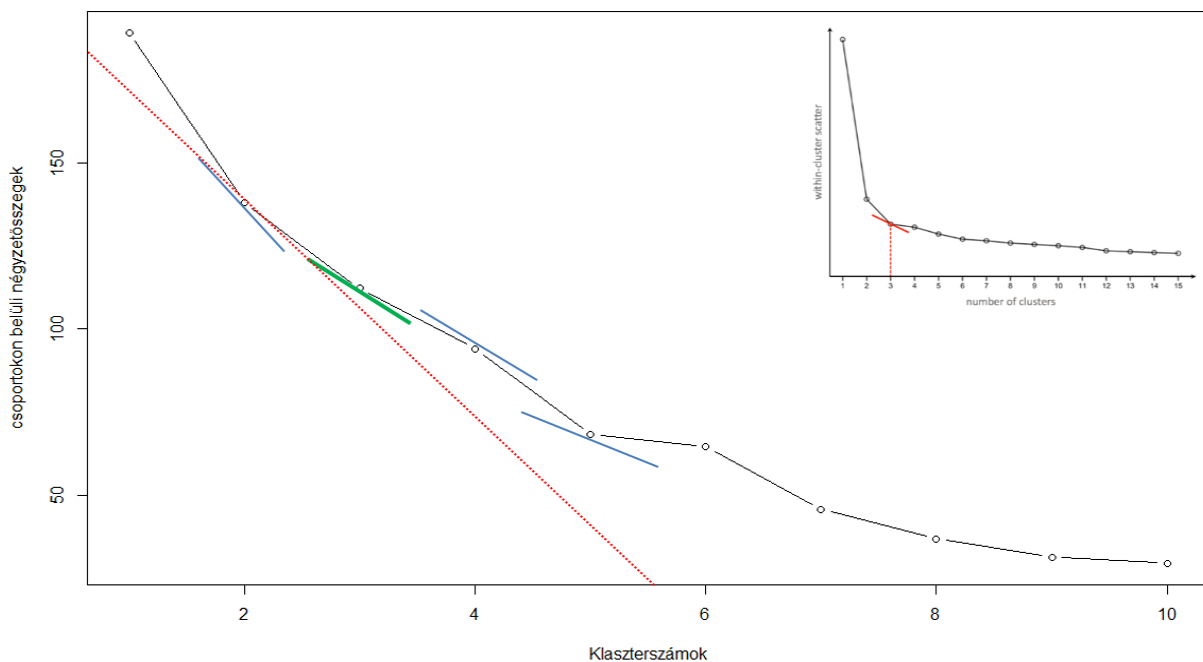
Szüle írásában kiemeli, hogy „*a könyök- és sziluettmódszer fontosságát jelzi, hogy néhány elemzés kiemelten foglalkozik ezek eredményeivel*”,⁴⁷ megjegyzi továbbá, hogy „*az optimális klaszterszám a könyökmódszernél az ábrázolt értékek*

⁴⁵ A szerző saját szerkesztése az R programnyelv használatával (Package 'stats' version 3.6.2.).

⁴⁶ Elsődleges cél a csoportokon belüli négyzetösszegek minimalizálása, mely a klaszterszám elemszámmal megegyezősége esetén nulla értéket vesz fel. A klaszteranalízis célja azonban ezzel egyidejűleg a dimenziócsökkentéssel történő csoportokba sorolás, melynek megfelelően a négyzetösszegek kárára áldozatot kell hoznunk. Meg kell tehát határozni azt a pontot (klaszterszámot), melynél a négyzetösszegek csökkenés mértéke a klaszterszám növelését nem ellensúlyozza.

⁴⁷ Forrás: Szüle Borbála: Klaszterszám-meghatározási módszerek összehasonlítása. *Statisztikai Szemle*, 97. évfolyam (2019), 5. szám, p. 423. http://real.mtak.hu/93712/1/2019_05_421.pdf (Letöltés ideje: 2020. 02. 15.)

nagymértékű meredekségváltozása [...], a sziluettmódszernél pedig az átlagos sziluettérték maximuma alapján azonosítható”,⁴⁸ azonban hozzáteszi: „a klaszterkönyök meghatározása kapcsán [...] két fontos probléma adódhat: egyrészt az eljárás „számításintenzív” (ami miatt nagy adatbázisoknál viszonylag lassan számíthatók az eredmények), másrészt a könyökpont megtalálása olykor nehézségekbe ütközik, hiszen az ábrán nem mindig rajzolódik ki a könyök alakzat.”⁴⁹ Esetünkben az utóbbival állunk szemben, vagyis a meredekségváltozás intenzitása nem rajzolja ki a keresett „könyököt”. (Összehasonlításként a 12. ábra jobb felső sarkában egy elfogadhatónak tekinthető és az ábráról egyértelműen leolvasható klaszterkönyök ábrát helyeztem el⁵⁰):



12. ábra

Optimális klaszterszám meghatározása klaszterkönyök módszerrel⁵¹

A 12. ábrán megfigyelhetjük, hogy az egy klaszterről kettő klaszterre való áttérés egy masszív csökkenést eredményez a klasztercsoportokon belüli négyzetösszegek értékét érintően. Látható ugyanakkor, hogy a klasszikus „könyök” ábra nem vehető ki egyértelműen. Belátható, hogy a módszer lényege a klaszterszám folyamatos növelésével minél nagyobb „haszon” (minél nagyobb mértékű négyzetösszegcsökkenés) elérése. Kompromisszumos megoldásnak tekinthetjük, ha az adott ponthoz húzott érintő meredeksége alapján hozzuk meg a döntésünket. Látható, hogy a 12. ábra jobb felső sarkában feltüntetett klasszikus „könyök” ábra 3-

⁴⁸ Uo. p. 423.

⁴⁹ Uo. p. 428.

⁵⁰ Forrás: <https://algobeans.com/2015/11/30/k-means-clustering-laymans-tutorial/> (Letöltés ideje: 2020. 03. 04.)

⁵¹ A szerző saját szerkesztése az R programnyelv használatával (Package 'graphics' version 3.6.2.).

as klaszterszámánál az adott ponthoz húzott érintő meredeksége a $-1/2$ értéket átlépve, a $-1/2$ -ed értéktől a 0 érték felé tart. Ha jobban megfigyeljük a 12. ábrát, akkor látható, hogy esetünkben ez a folyamat a 3-as klaszterszámnál rajzolódik ki. Mivel célunk a dimenziócsökkentés, így a „*nehezen kirajzolódó könyvkalakzat*”, valamint az ismertetett könyvmegtalálást-támogató módszer alapján a 3-as klaszterszám választása célszerű. A vizsgált elemeket érintően a kapott csoportokat a 10. ábrán az országnevek különböző színű bekeretezésével, valamint a klaszterképzéshez szükséges vágás helyét piros körrel jelöltem.

Klaszterek értelmezése

A végrehajtott elemzéssel a hierarchikus, alulról építő, teljes láncmódszerrel előállított klaszterezési eljárás során az elemzésbe bevont 28 NATO tagországot az elemzés alapját képező hét változó⁵² alapján három különálló klaszterbe sorolhatjuk, melyek az alábbiak:

1. klaszter:

Elemei (21 tagállam): Albánia, Belgium, Csehország, Dánia, Franciaország, Görögország, Hollandia, Horvátország, Kanada, Lengyelország, Litvánia, Magyarország, Montenegró, Németország, Olaszország, Portugália, Románia, Spanyolország, Szlovákia, Szlovénia, Törökország.

Jellemzői: A csoport kialakításának elsődleges szempontja a személyi jellegű kiadásoknak a védelmi kiadásokon belüli túlsúlya. Érdekes tényként figyelhető meg ugyanakkor, hogy ezen tényező dominanciája a teljes, 28 független változót magába foglaló csoporton belül kiemelkedő, ugyanis ha a 7 függő változóból épített adatmodellt csak ezen tényező alapján rangsoroljuk, hasonló eredményre jutunk. A 10. ábrán feltüntetett dendogramra tekintve azonban láthatjuk, hogy a klaszter 3 fő, nagyobb klasztercsoport és egy kiugró értékkel rendelkező ország (Görögország; lásd 11. ábra) összevonásából keletkezett. Láthatjuk továbbá, hogy:

- a klaszteren belüli két egymáshoz legközelebb álló ország Németország és Kanada (majd Belgium-Szlovénia, Franciaország-Hollandia, Olaszország-Portugália, Albánia-Spanyolország, Lengyelország-Románia és így tovább);
- az elemezett változókat érintően, két egymástól legtávolabb eső országpáros a Németország-Kanada országpáros és Görögország;
- Görögország kiemelkedően magas népességarányos haderejével (1,03%; 107,6 ezer katona a megközelítőleg 10,5 milliós lakosságon belül) relatív távol került a többi tagországtól, így a klaszterbe csak utolsóként került bevonásra [érdekességképpen: Görögországot a népességarányos haderő adatait érintően Litvánia (0,57%), Törökország (0,52%), Észtország (0,48%), majd az Amerikai Egyesült Államok (0,41%) követi];

⁵² Védelmi kiadások: (1) GDP%; (2) lakosságra vetítve (USD/fő); (3) népességen belüli haderőlétszám arány (%); (4-7) Védelmi kiadások részaránya (%): (4) főbb eszközök kiadásai (K+F(+I)) – (5) személyi jellegű kiadások – (6) infrastrukturális kiadások – (7) egyéb kiadások.

– a klaszterbe vont országok személyi jellegű kiadásai a teljes védelmi kiadásokon belül kivétel nélkül 40% fölött vannak (az első helyen Görögország (76,82%, mely haderejének aránytalan méretéből adódik); míg az utolsó helyen Szlovákia (40,61%))

– a klaszterbe a NATO által 2006-ban megfogalmazott, majd a 2014-es newporti (walesi) NATO csúcstalálkozón megerősített GDP részarányos 2% mértékű védelmi kiadási szintet teljesíteni képes 9 tagország közül 4 ország került bevonásra [Görögország (2,28%), Románia (2,04%), Litvánia (2,03%) és Lengyelország (2,00%)];

– a GFP interneten is elérhető 2020-as 138 ország védelmi összehasonlítását végrehajtó statisztikája alapján az alábbi (zöld színnel jelölt) „legerősebb” indexszámok (és országok) kerültek az 1. klaszterbe.⁵³

1	7	8	1 1	1 2	1 3	2 0	2 1	2 4	3 1	33	3 4	3 6	3 9	4 9	5 4	5 8	6 1	6 2	6 9	7 0	8 3	10 0	10 2	11 5	1 1 9	12 3
X	<input type="checkbox"/>	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X	<input type="checkbox"/>	X	<input type="checkbox"/>
Amerikai Egyesült Államok	Franciaország	Egyesült Királyság	Törökország	Olaszország	Németország	Spanyolország	Lengyelország	Kanada	Norvégia	Görögország	Csehország	Hollandia	Románia	Dánia	Magyarország	Szlovákia	Portugália	Bulgária	Horvátország	Belgium	Litvánia	Szlovénia	Lettország	Albánia	Észtország	Montenegró

– a népességen belüli haderőlétszámot érintően az első három helyezett a klaszter elemeit képezi (Görögország (1,03%), Litvánia (0,57%) és Törökország (0,52%)).

2. klaszter:

Elem (1 tagállam): Amerikai Egyesült Államok

Jellemzői: Az egy elemű klaszter létrehozásának indoka, hogy még a normalizálással sem sikerült ellensúlyozni az Amerikai Egyesült Államok dominanciáját az adatszerkezeten belül. A klaszter kialakításának elsődleges szempontjaként az algoritmus:

– a védelmi kiadások GDP %-os részarányát (értéke: 3,42 % (730 149 millió USD); 2. helyezett: Bulgária (3,25%, 2.179 millió USD)⁵⁴);

– az egy főre jutó védelmi kiadási összeget (2.072 USD/fő; 2. helyezett: Norvégia 1.384 USD/fő értékkel);

– a népességen belüli haderő-létszám arányt [értéke: 0,41% – 1,34 millió fő; függő változóbeli helyezés értéke: 5. hely; megelőzi: Görögország (1,03% – 107,6

⁵³ Az elemzésbe bevont 28 NATO tagország közül Luxemburgot érintően nem rendelkezünk értékelhető helyezéssel, így a táblázatban nem került feltüntetésre.

⁵⁴ Ezzel az értékkel a 7 millió főt számláló Bulgária (3,25%, 2.179 millió USD) megelőzte a 9,7 millió lélekszámmal rendelkező Magyarországot (1,21%, 2.080 millió USD).

ezer fő), Litvánia (0,57% – 15,7 ezer fő), Törökország (0,52% – 435,5 ezer fő) és Észtország (0,48% – 6,3 ezer fő)], valamint

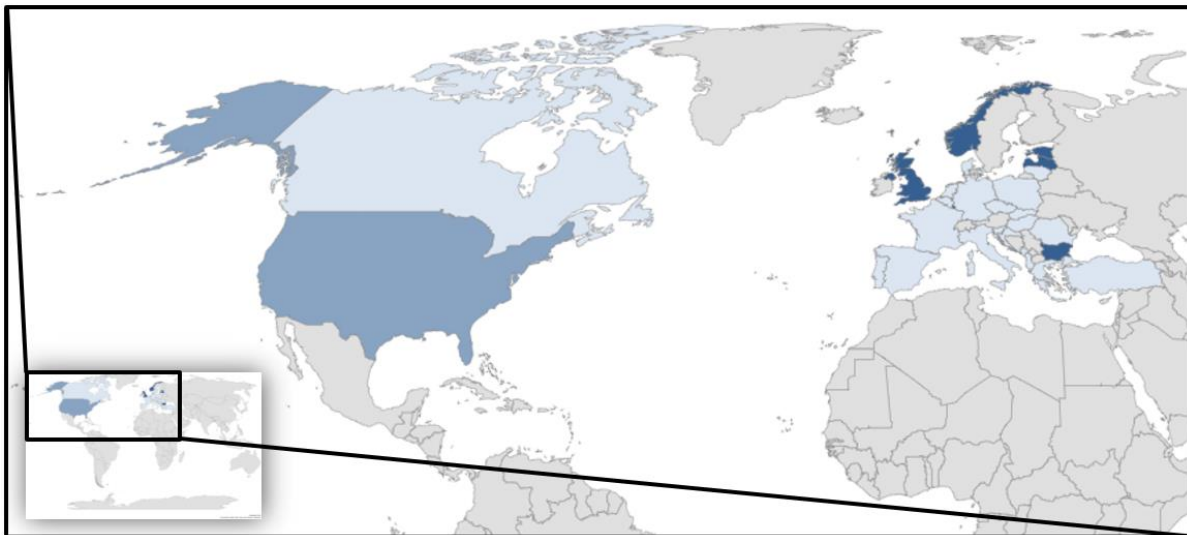
– az egyéb védelmi kiadási részarányt (teljes adathalmaz 5. helyezette 32,59%-os értékkel)
azonosította.

3. klaszter:

Elemi (6 tagállam): Bulgária, Egyesült Királyság, Észtország, Lettország, Luxemburg, Norvégia.

Jellemzői: A 3. klaszter legfőbb magyarázó jellemzője (habár meglehetősen nagy szórás figyelhető meg az adatokat érintően) a főbb eszközök (+ K+F(+I)), valamint az infrastrukturális védelmi kiadások részarányai. Ennek szintén magyarázat, hogy az alulról építő modell, az érintett klaszterben történő első tagországi összevonását megelőzően az 1. klaszter 21 eleme már 6 kisebb klaszterbe került besorolásra (a besorolás természetesen, mint azt már korábban, többször is hangsúlyoztam, a tagországok egymáshoz közelsége alapján valósult meg). Látható tehát, hogy a 3. klaszterbe sorolt tagországok egy klaszterbe kerültek, azonban a klaszteren belüli elemtávolságok nem tekinthetők a legszorosabbnak [különösen igaz ez a 3. klaszterbe sorolt kettő, valamint egy kiugró értékkel rendelkező Bulgária és Luxemburg esetében (lásd 10. ábra)].

A 3 klaszter földrajzi eloszlását, így a klaszteralapú besorolást az alábbi ábra szemlélteti:



13. ábra

Az elemzésbe bevont 28 NATO tagország 3 klaszterbe sorolása⁵⁵
(1. klaszter: világos kék; 2. klaszter: kék; 3. klaszter: sötét kék)

⁵⁵ A szerző saját szerkesztése a Microsoft Excel program alkalmazásával.

Megállapítások

A 29 NATO tagország közül a 2019-es becsült adatok alapján a 2006-ban vállalt, majd a 2014-es walesi csúcson megerősített 2%-os GDP részarányos védelmi kiadási mértéknek az akkori 3 tagország⁵⁶ közül immár 9 tagország képes eleget tenni, így a vállalt kötelezettséget:

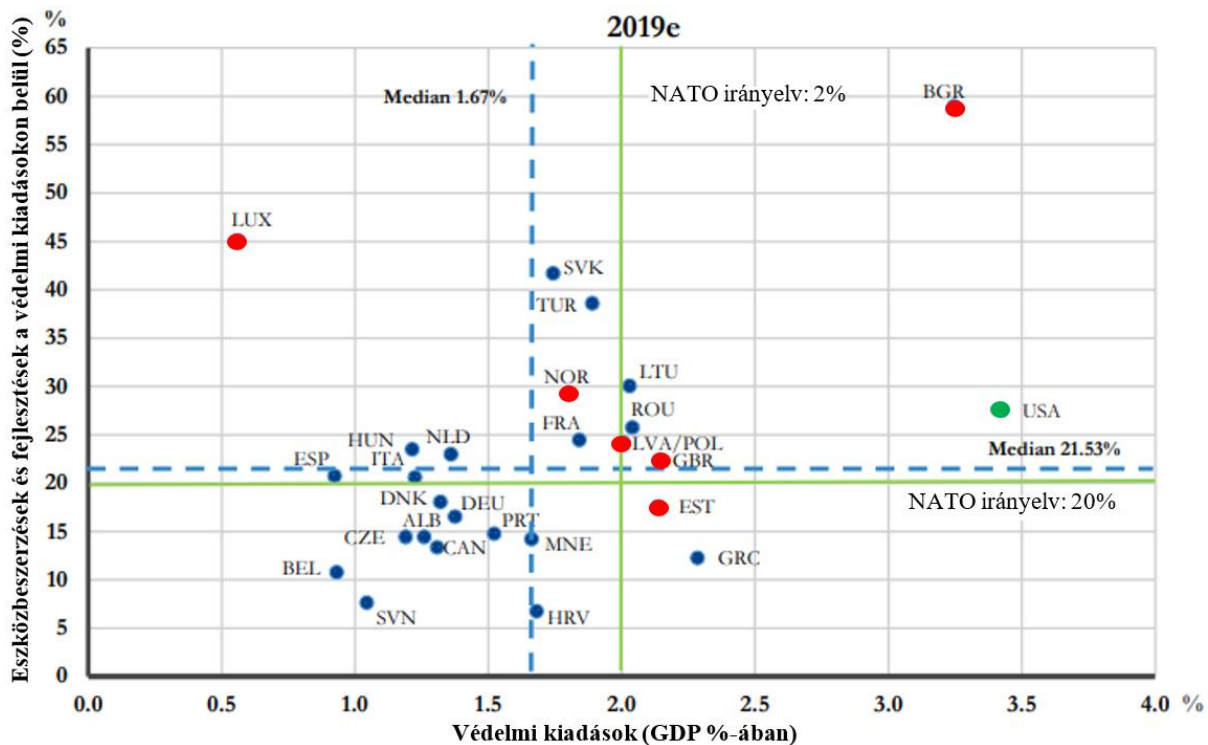
- a) az Amerikai Egyesült Államok 3,42%-os;
- b) Bulgária 3,25%-os;
- c) Görögország 2,28%-os;
- d) az Egyesült Királyság és Észtország 2,14-2,14%-os;
- e) Románia 2,04%-os;
- f) Litvánia 2,03%-os;
- g) Lettország 2,01%-os;
- h) Lengyelország 2%-os mértékben teljesíti.⁵⁷

Látható, hogy a Szövetségben meghatározott és elfogadott 2%-os GDP részarányú védelmi kiadási mértéket teljesíteni akaró és tudó tagországok köre folyamatosan bővül. Lassan el kell fogadnunk, hogy a megszokott kétdimenziós elemzési módszereink nem elegendők az értékelésekhez és a megfelelő következtetések levonásához.⁵⁸ Mindettől függetlenül a NATO 2019. november 29-én megjelent elemzésében – *mint a 14. ábrán láthatjuk* – még jelenleg is kiemelt figyelmet fordít a tagországok GDP részarányos védelmi kiadásainak (láthattuk, hogy ez a változó a 2. klaszter kialakításánál játszott jelentős szerepet), továbbá a védelmi kiadásokon belüli főbb eszközök, valamint az azokhoz tartozó K+F(+I) kiadásainak (mely a 3. klaszter kialakításánál volt meghatározó) összehasonlító elemzésére. (Megjegyzendő és értékelendő ugyanakkor, hogy a dokumentumban ez az egyetlen olyan ábra, mely az általános elemzési módszerekkel szakítva túlmerészkedik az egyfaktoros elemzésen).

⁵⁶ 2014: Amerikai Egyesült Államok, Görögország és az Egyesült Királyság (https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf; p. 3.; Letöltés ideje: 2020. 02. 15.)

⁵⁷ Uo.

⁵⁸ Már csak azért sem, mert míg Bulgária 2018-ban 961 millió USD védelmi kiadással rendelkezett (GDP részarány 1,48 %), addig 2019-ben ez az érték 2.179 millió USD (GDP részarány: 3,25 %). A változás 1218 millió USD, ami 1,77 %pontos GDP részarány növekedést eredményezett. Ugyanezt a %pont mértékű elmozdulást az Amerikai Egyesült Államok több mint 701 %-os (!), 853 939 millió USD, míg Montenegró 8,6 %-os, 105 millió USD összeggel lenne képes teljesíteni.



14. ábra

A NATO tagországok GDP részarányos védelmi kiadásainak, valamint a védelmi kiadásokon belüli főbb eszközök és az azokhoz tartozó K+F(+I) kiadásainak bemutatása⁵⁹

(Összehasonlításképp a kapott klasztercsoportok szemléltetése végett 2. klaszter eleme zöld színnel, a 3. klaszter elemei piros színnel kerültek átszínezésre)

Az 14. ábrán látható, hogy a kétfaktoros elemzés alapján történő csoportosítást alapvetően a NATO irányelvekben megfogalmazott 2% (GDP részarány), valamint 20%-os (K+F(+I) irányszám) mértékekhez igazítja. Problémát okozhat ugyanakkor az egyes tagországok ábra alapján történő valódi hovatartozásának megállapítása. Anélkül azonban, hogy a hivatkozott ábrát érintően mélyebbreható elemzésbe temetkeznénk, megállapíthatjuk, hogy – *csak egy példát említve a sok közül* – Bulgária és az Amerikai Egyesült Államok egy csoportba történő sorolása (átszínezés nélkül ugyanis kihívás lenne bármilyen csoport azonosítása is) alapos magyarázatot igényel.⁶⁰

Jelen elemzésben az általánosan megszokott 1 faktoros, majd a 14. ábrán bemutatott 2 faktoros adatelemzést további dimenziókkal bővítettem. Az összesen 7 függő változóval rendelkező adathalmaz elemeit egymáshoz való „közelségük” alapján 3 különböző klaszterbe sorolhattuk. A 3 klaszter különböző „fő tulajdonság” alapján került kialakításra, majd ennek megfelelően a 12. ábrán ábrázolásra.

⁵⁹ Forrás: Defence expenditures of NATO countries (2013–2019); 2019. november 29; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division. p. 2. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf (Letöltés ideje: 2020. 02. 15.)

⁶⁰ Mindentől függetlenül érdemes kiemelni Bulgária elmúlt időszakban tett intézkedéseit, melynek eredményeképp elérte és jelentősen meg is haladta a Szövetség tagjaként vállalt egyik fő kötelezettségvállalás (GDP részarányos védelmi kiadások) teljesítését.

Láthatjuk, hogy az egyes klaszterek (Bulgáriát leszámítva) földrajzilag is jól elkülöníthetők. Az elemzés eredményeképp az alábbi következtetéseket vonhatjuk le:

- 1) a NATO tagországok védelmi szempontú, eddig megszokott egyfaktoros elemzése sokszor téves csoportosításokhoz, ezáltal téves következtetések levonásához vezethet⁶¹;
- 2) az Amerikai Egyesült Államok NATO-n belüli súlya több tényező alapján is kiemelendő;
- 3) a 2019-es adatokat elemezve (Bulgáriát leszámítva)⁶² az európai NATO tagországok horizontális kettészakadását figyelhetjük meg (látható, hogy a közép és déli tagországoknál a személyi jellegű kiadások magas részaránya erősen dominál);
- 4) míg Kanada, valamint a közép- és dél-európai NATO tagországoknál a személyi jellegű kiadások védelmi kiadásokon belüli magas részarányára lehetünk figyelmesek, addig az a többi tagország védelmi kiadásának meghatározóbb részét haditechnikai eszközeinek fejlesztésére, modernizálására fordítja.

Megállapíthatjuk, hogy a következtetések helytállóságát az adatelemzésbe bevont függő változók sokszínűsége, valamint az alkalmazott elemzési módszerek elméleti megbízhatósága biztosítja. Láthatjuk ugyanakkor, hogy minél nagyobb adatbázis áll rendelkezésünkre, annál pontosabb következtetéseket tudunk levonni, mivel annál nagyobb az alkalmazott technika által nyújtott mozgásterünk (kiugró értékek szűrése, függő változók kizárása stb.).

A hasonló módszerrel végrehajtott elemzések különösen hasznos információkat nyújthatnak számos területen, így a védelmi tervezés területén egyaránt. Németh írásában⁶³ kiemeli, hogy a védelmi tervezés elmúlt időszakban koncepcionális újragondolás küszöbén állt. Hangsúlyozza: *„Szenes Zoltán korunk környezeti tényezőit vizsgálva tett javaslatot egy új stratégiai koncepcióra, [2005-ben megjelent cikkében] helyesen felismerve azt a tényt, hogy a tervezésre kívülről ható tényezők dinamikája új koncepció szükségessége felé mutat.”*⁶⁴ Ez később a NATO által meg is valósult, 2010-ben új stratégiai koncepciót hagytak jóvá. Németh kiemeli ugyanakkor, hogy a tervezési alapvetés kiemelten fontos tényezőként azonosítja „a

⁶¹ Nem elegendő a csupán GDP részarányos, vagy védelmi kiadási mérték alapján történő sorba rendezés.

⁶² Bulgária 2018-ról 2019-re jelentősen, mintegy 126,7%-kal, 961 millió USD-ről 2.179 millió USD-re növelte védelmi kiadásainak összességét, melynek eredményeképp GDP részarányos védelmi kiadásai 1,48%-ról 3,25%-ra emelkedtek. Szintén fontos megjegyeznünk, hogy a jelentős növekedés nagymértékű eszközmodernizációnak tulajdonítható, mely eredményeképp a személyi jellegű védelmi kiadási részarány 2018-ról 2019-re 62,99%-ról 28,73%-ra, míg az eszközökre költött védelmi kiadási részarány 2018-ról 2019-re 9,65%-ról 58,97%-ra változott. Forrás: Defence expenditures of NATO countries (2013–2019); 2019. november 29; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division. p. 2.: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf; (Letöltés ideje: 2020. 02. 15.)

⁶³ Németh Gergely: A NATO védelmi-tervezés fejlődéstörténete (1949–1967). *Hadtudomány*, 2012. 3–4. szám, pp. 89–103. http://mht.eu/hadtudomany/2012/3_4/HT_2012_3-4_Nemeth_Gergely.pdf (Letöltés ideje: 2020. 03. 10.)

⁶⁴ Uo. p. 90.

változást érzékelő, feldolgozó és a szükséges lépéseket tervezni képes szakemberek, törzs meglétét”, akik „képes[ek] becsatornázni a védelmi szféra politikai, társadalmi, gazdasági és katonai környezetében bekövetkező jelentős változásokat, valamint az azokból eredő következtetéseket.”⁶⁵ Jelen tanulmányban bemutatott statisztikai módszer csupán adatbázissal operál. Amennyiben a Németh által felvetett változók számszerűsíthetők, úgy a bemutatott klaszteranalízis eljárás képes és alkalmas a függő változók további „becsatornázására”, legyen az politikai, társadalmi, gazdasági vagy éppen katonai eredetű.

Összefoglalás

Az elmúlt időszakban különösen nagy mennyiségű adatnövekedés szemtanúi lehettünk, és jelenleg is lehetünk. Míg 2005-ben 130 exabájtnyi (10^{16} bájtnyi) adatmennyiségről, addig 2020-ra a becslések szerint ennek 315-szöröséről, 40 900 exabájt méretű adattömegről beszélünk.⁶⁶ Ez óriási mennyiségű adat, kiemelkedő növekedési ütemmel. Nincs ez máshogy a védelmi szférát érintően sem. Az egyes országok védelmi kiadásainak mértéke, annak összetétele, az alkalmazott (hadi)technikai eszközök sokszínűsége, a védelmi tervezés területén megfogalmazott politikai, társadalmi, gazdasági és katonai tényezők bevonásának igénye egytől egyig megköveteli egy többdimenziós adatszerkezet létrehozását és az elemzés megfelelő eszközzel történő végrehajtását.

Ahogy Szenes cikkében⁶⁷ hivatkozik a hadtudomány ún. „*hagymaszerkezetére*”, úgy a fent bemutatott elemzésre alapozva láthatjuk és megállapíthatjuk, hogy a védelmi kiadásoknak is van egyfajta rétegződése, „*hagymaszerkezete*”. Az elemzésben megfigyelhettük, hogy a védelmi kiadásokat nem elegendő csupán a GDP részarányhoz viszonyítottan, vagy egzakt védelmi kiadási mértékre alapozottan elemezni, mert ezen elemzések sokszor téves következtetések leszűréséhez vezetnek. Az elemzés teljeskörűségéhez elengedhetetlen a „*hagyma leveleinek*” leválasztása (vagy éppen megfelelő felépítése), valamint az így kapott adathalmaz alaposabb vizsgálata. A megfelelő vizsgálatból megállapíthatjuk és következtetéseket vonhatunk le. Láthatjuk, hogy az Amerikai Egyesült Államok nélkül, a 2019-es becsült adatok alapján 309 479 millió USD költségvetéssel (GDP részarány: 1,57%), míg az Amerikai Egyesült Államokkal együtt mindösszesen 1 039 628 millió USD védelmi kiadási összeggel (GDP részarány: 2,52%), megközelítőleg 3,26 milliós katonai létszámmal számon tartott Szövetség védelmi kiadási adatait tekintve milyen stacionárius helyzetek figyelhetők meg, mely

⁶⁵ Uo. p. 102.

⁶⁶ Kirill Eremenko, Hadelin de Ponteves, SuperDataScience Team, SuperDataScience Support : Machine Learning A-Z: Hands-On Python & R In Data Science; Udemey course. <https://www.udemy.com/course/machinelearning/> (A szerző által előfizetett tartalom.)

⁶⁷ Szenes Zoltán: Akadémiai viták a hadtudomány struktúrájáról. *Hadtudomány*, 2013. 3-4. szám, p. 64. http://mhtt.eu/hadtudomany/2013/3_4/Hadtudomany_2013_3-4_6.pdf (Letöltés ideje: 2020. 03. 10.)

tagországok milyen erőfeszítéseket tesznek a vállalt kötelezettségek teljesítése érdekében.

A tanulmányban láthattuk, hogy az elemzésbe bevont 28 NATO tagország mely tényezők mentén sorolható a kialakított 3 klaszterbe. Bemutattam a kiugró értékek (outlierek) kezelését, a hiányzó adatok becslésének egyik módszerét (lineáris regresszió), majd az R programozási nyelv által kínált technikák egy meghatározott részének gyakorlatban történő alkalmazhatóságát. Megállapítottam, hogy az Amerikai Egyesült Államok szövetségén belüli dominanciája még standard normalizálási eljárás alkalmazásával is nehezen kezelhető, a külön klaszterben biztosított hely szinte kézenfekvő. Rávilágítottam az európai NATO-tagországok horizontális kettészakadására, melyet a klaszterképzés eredményeként, a védelmi kiadásokon belüli személyi jellegű kiadási részarány magas mértékének tulajdoníthatunk. Láthattuk és láthatjuk, hogy míg ezen országok kiadásaik jelentős részét a humán erőforrás pénzügyileg történő elismerésre fordítják, addig más országokban (észak-európai NATO-tagországok esetén) a főbb fegyverrendszerek és azok fejlesztései kerülnek előtérbe.

A bemutatott eljárást érintően további érdeklődésre számot tartó elemzésnek tartom egy kibővített adattömb dinamikus alapokra helyezett vizsgálatát, melybe a NATO-tagországokon túlmenően számos más tagország (nem csak védelmi kiadási, de haditechnikai, politikai, társadalmi, gazdasági és katonai) adatait, több évre visszamenőleg bevonjuk. Az elemzés során ugyanakkor tisztában kell lennünk azzal a ténnyel, hogy az adatok között számos esetben találkozhatunk kiugró értékekkel, melyek vagy emberi figyelmetlenségből (pl. elgépelés, lásd Németországban évente a lakosság közel 10%-a éri el a sorozási korhatárt), vagy valami érdekes folyamatból eredeztethetők (pl. Bulgária megemelt 2019. évi becsült védelmi kiadása). Mint láthattuk, az adatok előkészítése a teljes elemzés oroszán részét képezi, azonban a kapott eredmények és az azokból levonható következtetések ezen erőfeszítéseinket alaposan kompenzálják. Zárásként pedig a 43 éves cseh sztárközgazdász, Tomáš Sedláček szavait szeretném idézni, aki szerint: *„Még a legkifinomultabb matematikai modell is, de facto, egy történet, egy példázat, kísérlet arra, hogy – racionális módon – megértsük a minket körülvevő világot.”*⁶⁸

⁶⁸ Tomáš Sedláček: A jó és a rossz közgazdaságtana (A Gilgames-eposztól a Wall Streetig). Budapest, HVG Kiadó Zrt., 2012. pp. 19–20.

FELHASZNÁLT IRODALOM

- Alto, Valentina: Unsupervised Learning: K-means vs. Hierarchical Clustering. <https://towardsdatascience.com/unsupervised-learning-k-means-vs-hierarchical-clustering-5fe2da7c9554>
- Baranyai Dávid, Dr. Mándoki Péter, Dr. Kővári Botond, Dr. Török Ádám: Magyarországi gyalogos és kerékpáros balesetek elemzéseinek módszerfejlődése. *IFFK* (Innováció és fenntartható felszíni közlekedés), 2016. augusztus 29–31. http://kitt.uni-obuda.hu/mmaws/2016/pages/program/papers/Paper_44_Baranyai_et_al_IFFK_%202016.pdf
- Defence expenditures of NATO countries (2013–2019). 2019. november 29.; Nr.: PR/CP(2019)123; NATO Public Diplomacy Division. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf
- Galarnyk, Michael: Understanding boxplots; 2018. szeptember 12.
- Grace, Karen – Outliers, Martin: To Drop or Not to Drop; The Analysis Factor (Making Statistics, Make Sense). <https://www.theanalysisfactor.com/outliers-to-drop-or-not-to-drop/>
- Information on defence expenditures (NATO archives). https://www.nato.int/cps/en/natohq/topics_49198.htm
- Kovács Erzsébet: Többváltozós adatelemzés. Egyetemi jegyzet. Budapest, Budapesti Corvinus Egyetem, 2014. http://etananyag.ttk.elte.hu/FiLeS/downloads/14_KOVACS_E_Tobbvalt_adatelemzes.pdf
- Németh Gergely: A NATO védelmi-tervezés fejlődéstörténete (1949–1967). *Hadtudomány*, 2012. 3–4. szám, pp. 89–103.
- Petrovics Petra: Klaszterelemzés az SPSS-ben. Előadásjegyzet; Miskolci Egyetem, Gazdálkodástudományi Kar. https://nemzetkozi-gazdalkodas.hu/files/950/7_MM_klaszter.pdf
- R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing., Vienna, Austria 2019. URL: <https://www.R-project.org/>
- Sedláček, Tomáš: A jó és a rossz közgazdaságtana (A Gilgames-eposztól a Wall Streetig). Budapest, HVG Kiadó Zrt., 2012.
- SIPRI Military Expenditure Database (Data for all countries 1949–2018). <https://www.sipri.org/databases/milex>
- Szenes Zoltán: Akadémiai viták a hadtudomány struktúrájáról. *Hadtudomány*, 23. évfolyam (2013), 3–4. szám, pp. 59–66. http://mhtt.eu/hadtudomany/2013/3_4/Hadtudomany_2013_3-4_6.pdf

- Szüle Borbála: Klaszterszám-meghatározási módszerek összehasonlítása. *Statisztikai Szemle*, 97. évfolyam (2019), 5. szám, pp. 421–438.
http://real.mtak.hu/93712/1/2019_05_421.pdf
- Dr. Tömösváry Zsigmond nyá. dandártábornok: Új fegyverkezési hajsza? Mindenki Hadtudománya Szabadegyetem; 2019.
http://bhd.honvedseg.hu/cikk/mindenki_hadtudomanya_amerikai-orosz_viszony
- World Population Review. Total Population by Country 2020.
<http://worldpopulationreview.com/countries/>

INTERNETES FORRÁSOK

- <https://algobeans.com/2015/11/30/k-means-clustering-laymans-tutorial/>
- <https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS>
- <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>
- <http://worldpopulationreview.com/countries/nato-spending-by-country/>
- https://www.globalfirepower.com/country-military-strength-detail.asp?country_id=germany
- https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2005_06/20090902_p050609.pdf
- https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2007_12/20090327_p07-141.pdf
- https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2011_03/20110309_PR_CP_2011_027.pdf
- https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2016_01/20160129_160128-pr-2016-11-eng.pdf
- https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_11/20191129_pr-2019-123-en.pdf
- <https://www.cia.gov/library/publications/the-world-factbook/fields/333.html>
- <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=DE>