

TÖBBDIMENZIÓS SKÁLÁZÁS

Takács Szabolcs'

'Károli Gáspár Református Egyetem
Levelező szerző: Takács Szabolcs
E-mail: takacs.szabolcs@kre.hu

Beérkezett: 2013.03.01 – Elfogadva: 2013.04.12.

A cikkben egy olyan eljárást kívánunk bemutatni, mely nem mindenhol része a pszichológiai képzések törzsanyagának, azonban több olyan kutatási kérdés felvethető, melynek megoldásában jelentős szerepet tölthet be e módszer. Ezért e módszer rövid, áttekintő bemutatására vállalkozunk cikkünkben, melyben 2 gyakorlati példán keresztül ismertetjük a módszer legfontosabb bemeneti és kimeneti adatait, mutatóit.

A MÓDSZER BEMUTATÁSA

A többdimenziós skálázás (angolul: multidimensional scaling, a továbbiakban MDS) szakirodalma széleskörű, ismert és gyakran használt statisztikai módszer. Segítségével sokdimenziós objektumok olyan 2 vagy 3 dimenziós ábrázolása válik lehetségessé (nagyobb dimenziós alkalmazás is lehetséges, azonban ilyenkor az ábrázolás értelemeszerűen nehézkessé válik), melyben az eredeti pontthalmaz pontjai között meglévő távolságok nagyságrendi viszonyai megőrződnek. A távolságokat olykor kényelmesebb hasonlóságként/különbségként értelmezni – hiszen a vizsgált objektumok nem feltétlenül vannak közel vagy távol egymáshoz/egymástól, hanem azt vizsgáljuk, hogy mennyire hasonlítanak/különböznek egymásra/egymástól.

A módszert két, egymástól jelentős mértékben eltérő helyzetben fogjuk bemutatni:

- Először egy olyan változatot láthatunk, ahol előre megadott távolságok alapján ábrázoljuk az objektumokat 2 dimenzióban.
- Másodszer a program számítja ki az objektumok (változók) közötti távolságot és ad számunkra egy kétdimenziós ábrát.

Az ábrázolást természetesen nem szükséges kizárólag 2 dimenzióban elképzelnünk, azonban sok szempontból ez tűnik leginkább kényelmesnek. Részint a 3-dimenziós felületek megjelenítése nehézkes lehet, másik oldalról pedig nem is mindig egyértelmű, hogy mely irányokból látható vagy láttatható legjobban az eredmény.

A többdimenziós módszerek egyik sarokkövét jelentik, hogy milyen típusú változókon alkalmazhatók, illetve milyen módon tudjuk ellenőrizni a sokdimenziós eljárások jóságát, megbízhatóságát – továbbá, hogy az eredmények milyen módon interpretálhatók szakmai, és nem statisztikus szemmel.

Az MDS alkalmazása során lényegében egyetlen elvárásunk van: amely objektumok eredetileg hasonlítottak egymásra, azok az ábrázolás után is hasonlítsanak – amelyek pedig különböztek, azok az alacsonyabb dimenziós térben is legyenek különbözőek.

AZ MDS MATEMATIKAI TULAJDONSÁGAI, MEGBÍZHATÓSÁGÁNAK MÉRÉSE

Az MDS egyik matematikai sarokköve a távolságok definiálása: az eljárásban – bár némiképpen rejtettem – három ponton is mérünk eltéréseket, ha úgy tetszik távolságokat.

- Egyik oldalról adott az eredeti pontok, objektumok közötti távolság (itt nem feltétlenül a geometriából jól ismert euklideszi távolságra kell gondolni, azaz nem feltétlenül a két pontot összekötő szakasz hossza lesz a két pont közötti távolság).

Használhatunk euklideszi távolságot, mely adekvát lehet például normális eloszlású változók esetén.

Példaként érdemes megemlíteni a Manhattan távolságot is, mely úgy is értelmezhető, mint a négyzetrácson való, egyik rácspontból a másik rácspontba való séta hossza – a négyzetek élhosszúságainak összegét értjük alatta. Jól alkalmazható például akkor, ha egész értékű, intervallum skálájú változókkal dolgozunk.

A módszer néhány általánosítása és azok tesztjei olvashatók Níroomand és társainak (Níroomand, 2011) cikkében, amelyben a szerzők kidolgozzák a módszer legáltalánosabb alakját, majd két fontosabb, de nem euklideszi távolságra specifikálják eljárásukat.

- Másik oldalról hasonló dilemmával állhatunk szemben akkor, amikor a reprezentált 2-dimenziós objektumok egymástól való távolságát szeretnénk beállítani – ez az ábrázolás miatt hagyományosan az euklideszi távolság, de nyilván elfogadható a módszer olyan irányú általánosítása is, ahol ezt a távolságot is tetszőlegesen definiálhatjuk.
- A harmadik távolság rejtettebb: gondoljunk arra, hogy valamilyen módon mérnünk kell azt, hogy az eredeti objektumok által definiált alakzat és a 2 vagy 3 dimenzióban definiált, reprezentáló alakzat milyen módon, mennyire van „közel” egymáshoz. Ennek a mértéknek külön mutatója

lesz a módszerben – és az eljárásunk „jóságát” e mutató mértéke fogja számunkra meghatározni.

Itt is felhívjuk a figyelmet arra, hogy ez a távolság (ha úgy tetszik, eljárás pontosságát mérő mérték) szintén általánosítható, újradefiniálható. Azonban a módszertan általánosságban ezt a mértéket többé-kevésbé költöztetnek gondolja, mint azt a későbbiekben majd szemléltetjük is.

Cikkünkben – teljes egészében – csak a klasszikus megoldást ismertetjük. Az MDS eljárásnak azonban több alternatív (robosztusnak tekinthető) változata is ismert (skálatípustól és hasonlósági definícióktól függően).

- Az egyik első alkalmazását az alternatív eljárásoknak Kruskal (Kruskal, 1964) ismerteti cikkében, melyben ordinális (kvalitatív) változók és az alapvetően kvantitatív eljárásként működő MDS eljárás között teremt kapcsolatot, lehetőséget adva ezzel a kvalitatív változókra való alkalmazásra. Ezzel az eljárással utat nyitott az MDS eljárás robusztus változatainak kifejlesztésére, tanulmányozására.
- Fasham (Fasham, 1977) cikkében több eljárással is összehasonlítja az MDS ordinális változatát: különböző átlagokat összehasonlító és elemző eljárásokkal, illetve a főkomponens-analízissel. Ebből a cikkből kiderül, hogy számos esetben a robusztus (nem metrikus) MDS jobb, megbízhatóbb eredményt ad, mint más eljárások abban az esetben, ha a változók például ordinális skálájúak.

A módszerre vonatkozó matematikai statisztikai tételek egyértelműek arra az esetre, amikor mindhárom távolság az általánosan megszokott, euklideszi távolságot jelenti.

A klasszikus MDS eljárás optimalitási tulajdonságát két matematikai tételből ismerhetjük. Ezek megtalálhatók több tankönyvben is, így nem matematikai stilisztika alapján megfogalmazva az alábbi formában foglalhatjuk őket össze (a tételek megtalálhatók: Móri (1986), XIII / 4.4 és 4.6):

Amennyiben adott az eredeti pontthalmazunk dimenzió száma (p) és a reprezentáns pontthalmaz dimenzió száma (valamely fix), továbbá a reprezentánsok közötti távolságot euklideszi távolság alapján mérjük, úgy ha négyzetes eltérésben keressük a legközelebbi megoldást – függetlenül attól, hogy az eredeti p dimenziós pontthalmaz távolságait euklideszi vagy nem euklideszi módon definiáljuk – a klasszikus MDS által szolgáltatott megoldás adja az optimális megoldást.

A matematikai minőségét az MDS eljárásnak többek között az úgynevezett s -stressz mutató segítségével mérhetjük. E mutató nem más, mint az ábrázolt és az eredeti pontok koordinátáinak eltéréséből számított mutató. Az ábrázolás minőségét az alábbi szabályt mondhatjuk:

Az adott dimenziószámánál csak nagy információ-vesztéssel ábrázolható. Érdeemes nagyobb dimenziószámmal próbálkozni.

S-STRESS	MINŐSÉG	MEGJEGYZÉS
0-0.05	Kiváló	Valószínűleg minden releváns információt tartalmaz
0.05-0.1	Jó	Megfelelő, az eredmények értelmezhetők.
0.1-0.2	Elfogadható	Érdeemes foglalkozni vele. Az eredmény még többnyire értelmezhető
0.2 fölött	Nem megfelelő	Az adott dimenziószámmal csak nagy információ-vesztéssel ábrázolható. Érdeemes nagyobb dimenziószámmal próbálkozni.

1. táblázat S-stress mutató értéke és minősége

Megjegyzendő, hogy néha az s-stress mutató helyett az úgynevezett stress mutató értékét használják, mely az s-stress mutató négyzetgyöke.

Megállapítható, hogy euklideszi távolságok esetén a klasszikus MDS adja az optimális megoldást, tehát amennyiben minden távolság euklideszi, úgy 2 vagy 3 dimenzióban a klasszikus MDS-nél jobb ábrázolást nem tudunk készíteni. A matematikai tételek azt is biztosítják számunkra, hogy a klasszikus megoldás – amennyiben az ábrázolás euklideszi távolságok szerint történik és az ábrázolás jóságát mérő mutató is a klasszikus négyzetes különbséggel számolt, úgy szintén a klasszikus MDS adja a legjobban illeszkedő megoldást.

A KLASSZIKUS MDS ELJÁRÁSBÓL SZÁRMAZÓ INFORMÁCIÓK INTERPRETÁLÁSA

Fontos kiemelniünk, hogy az ábrázoláshoz használt koordináta-rendszer tengelyei nem feltétlenül jelentenek háttérváltozókat. Azonban előfordulhatnak olyan esetek, amikor tudunk jelentést adni az egyes tengelyeknek – esetlegesen többletjelentést adva így magának a modellnek, keletkező konfigurációnak. Erre láthatunk majd példát az első alkalmazásban.

EGY GYAKORLATI PÉLDA ELŐRE MEGADOTT TÁVOLSÁGOK ALAPJÁN

A módszer technikai megvalósítását több helyen is megtalálhatjuk, például Ketskeméty (2006) és Székelyi (2008), SPSS programrendszer használatáról szóló könyveiben.

Az első példában egy olyan adatállománnyal dolgozunk, melyben az objektumok távolságai előre adottak és e távolságok alapján szeretnénk egy olyan 2-dimenziós ábrázolást készíteni, melyben az objektumok távolságai a lehető legjobban közelítik az eredeti objektumok által definiáltakat.

A feladatot magyarországi városokkal mutatjuk be. Az objektumok közötti távolságok alapján történő ábrázolás láttathatósága nagymértékben függ az objektumok sorrendjétől, felsorolásától. Azonban mindez megfogalmazható abban a formában is, hogy az ábrázolás nem egyértelmű: különböző forgatások, eltolások,

	SOPRON	K.VÁR	TATAB.	PÉCS	B.P.	B.CSABA	DEBR.	MISK.
SOPRON	0	220	154	287	210	414	436	390
KAPOSVÁR	-	0	184	67	189	314	381	368
TATABÁNYA	-	-	0	211	56	259	282	235
PÉCS	-	-	-	0	198	283	367	377
BUDAPEST	-	-	-	-	0	203	226	179
BÉKÉSCSABA	-	-	-	-	-	0	130	228
DEBRECEN	-	-	-	-	-	-	0	98
MISKOLC	-	-	-	-	-	-	-	0

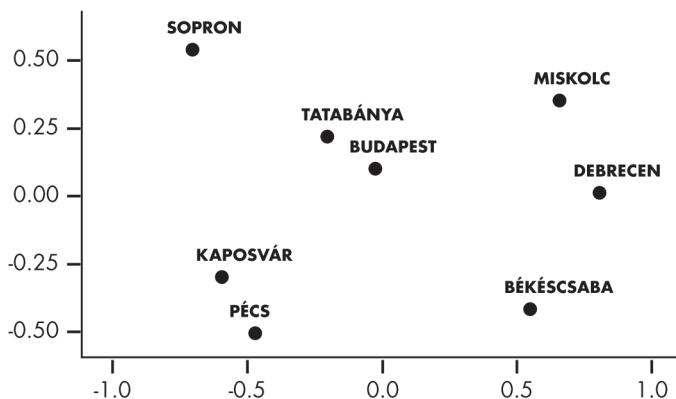
2. táblázat Magyarországi városok távolsága km-ben

tükrözések (klasszikus értelemben vett egybevágósági transzformációk) nem befolyásolják az ábrázolás jóságát, minőségét – azonban nehezíthetik vagy könnyíthetik a végeredmény interpretálhatóságát.¹

8 magyarországi várost választottunk ki, a közöttük lévő távolságot pedig az autóval megtehető legrövidebb (nem feltétlenül leggyorsabb) távolságokban adtuk meg:

A városok közötti távolságok szimmetrikusak, de természetesen ez sem kell, hogy igaz legyen – gondolhatunk itt a repülőutakra, ahol ismert tény, hogy az oda-vissza utak rendszerint nem azonosak. Hasonlóan: nem feltétlenül szimmetrikus az adatállomány akkor, ha emberek közötti szimpátiát kell pontosítani, értékelni: a szimpátia ugyanis nem feltétlenül kölcsönös.

Amennyiben e sorrendben szerepeltetjük a városokat, úgy az alábbi ábrázolást nyerjük:

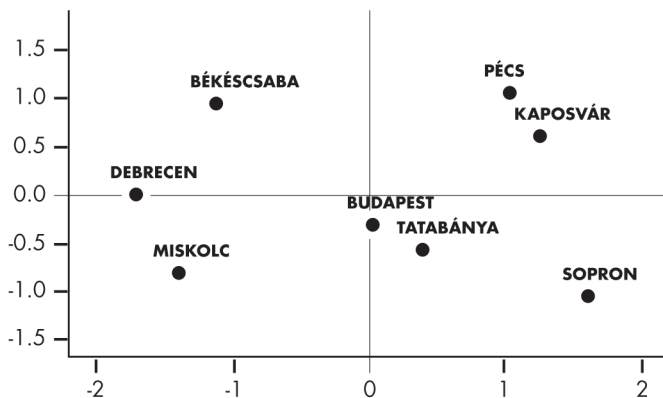


1. ábra Magyarországi városok távolsága km-ben

Észrevehető, hogy az eredeti ábrázolás mértékegységei nem köszönnek vissza az ábráról (egy majdnem origó középpontú téglalap lett az ábrázolásunk alapja, melyben elhelyeztük a városokat). A városok azonban megközelítően felismerhető helyen vannak – nyilvánvalóan nem illeszthető rá teljes egészében egy jól ismert Magyarország térkép, de „elég közelinek” tekinthető a megoldás a valósághoz.

Fontos azonban kiemelni, hogy semmivel sem jobb vagy rosszabb ábrázolást nyernénk, ha a tengelyek skálázása nem -1 és 1 (illetve -0,5 és 0,5) közötti lenne, hanem például -10 és 10 közötti, illetve ha az egész ábrát néhány fokkal jobbra vagy balra elforgatnánk, netán az egészet az origóra tükröznénk.

Néhány ilyen transzformációt végrehajtva például az alábbi ábrázolás nyerhető:



2. ábra Magyarországi városok ábrázolása transzformált alakban, átskálázás után

A városok távolságai egymáshoz képest nem változtak meg, azonban a tengelyek más skálán szerepelnek és az Észak-Dél, illetve Kelet-Nyugat tengelyek is megfordultak. De az ábrázolás matematikai minősége semmit sem javult vagy romlott.

A MODELLEK ÉRTÉKELÉSE

A modelleket az s-stress mutató segítségével értékeljük tehát, mely mindegyik ábrázolás esetén 0,04331 volt, ami azt jelenti, hogy 0,05 alatti, tehát kiváló ábrázolást találtunk mindkét esetben – matematikailag nyilvánvalóan semmi különbség nem lehetett a különböző ábrákban, ábrázolásokban.

A távolságok alapján számított ábrázolás esetén fontos lehet az objektumok felsorolásának sorrendje – és általánosságban nem számít, hogy milyen beosztású koordináta-rendszerben, vagy milyen elhelyezkedéssel

ábrázolunk. Minden egybevágósági transzformáció megengedett (forogatások, eltolások), ha segítségével könnyebben interpretálható, jobban értelmezhető eredményeket nyerünk.

EGY GYAKORLATI PÉLDA VÁLTOZÓKBÓL, PROGRAM ÁLTAL SZÁMÍTOTT TÁVOLSÁGOK ALAPJÁN

Második példánkban változókat szeretnénk elhelyezni egy 2-dimenziós ábrán – így keresve kapcsolatokat, összefüggéseket. Az ábrázoláshoz egy politikai felmérés adatait használjuk, melyben parlamenti pártokról különböző, a választók számára fontos értékeket kellett pontozni³. Az alacsony pontszámok azt jelentették, hogy az adott párt nem jellemezhető az adott értékrend hatékony képviselőjével – míg a magas értékek azt jelentették, hogy kellő képviselőt, hangsúlyt kap az adott értékrend a párt célkitűzéseiben.

Ezen túlmenően a megkérdezettek azt is elmondták, hogy számukra mennyire fontos az adott értékrend a személyes életükben, életvitelükben.

4 kiválasztott értéket fogunk vizsgálni:

7. szabadság
8. hagyományőrzés
9. hit
10. türelem

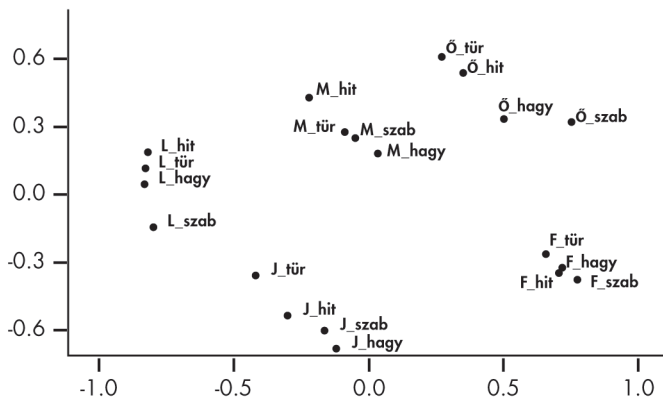
4 parlamenti pártra vonatkoztatjuk a kérdéseket (azaz: a megkérdezett szerint az adott párt számára mennyire fontos az adott értékrend képviselője). A 4 parlamenti párt a FIDESZ, az MSZP, az LMP és a Jobbik. Az ábra átláthatósága miatt minden esetben az alábbi kódokat használtuk:

ELNEVEZÉS	Megkérdezett	FIDESZ	MSZP	Jobbik	LMP	szabadság	hagyomány- őrzés	hit	türelem	
RÖVIDÍTÉS		Ő	F	M	J	L	szab	hagy	hit	tűr

3. táblázat Az ábrán alkalmazott rövidítések

Ezek után például a megkérdezett számára mennyire fontos a hagyományőrzés az „Ő_hagy” felirattal szerepel, míg például az a változó, mely azt méri, hogy az LMP számára mennyire fontos a megkérdezett megítélése szerint a szabadság, az „L_szab” címkével jelenik meg az ábrán.

Az MDS segítségével rajzolt „politikai térkép” így az alábbi formát ölti:



3. ábra A parlamenti pártok politikai térképe

A jobb felső kvadránsban helyezkednek el a megkérdozettek és az ő általuk értékelt értékek. Ennek alapján elmondható, hogy a felsorolt értékek közel azonos mértékben voltak fontosak a megkérdozettek számára – továbbá az is elmondható, hogy megítélésük szerint az MSZP és a FIDESZ értékei mind egymástól, mind a megkérdozettek értékeitől közel azonos mértékben voltak távol. Ugyanakkor az LMP és a Jobbik által képviselt értékek távolra kerültek a megkérdozettektől.

Jól kirajzolódni látszik az is, hogy a 4 parlamenti párt értékei egymástól igen távolinak mondhatók – és hogy mindez a két kisebb párt esetén fokozottan távoli a választók értékeitől.

A MODELL ÉRTÉKELÉSE

Az s-stressz mutató értéke 0,04482 volt, tehát 0,05 alatti, azaz ismét kiváló ábrázolást adhattunk a fenti változók alkotta objektumok elhelyezkedéséről. Világos, hogy itt a távolságok nem annyira egzakt módon megfogható fogalmak, mint az első esetben: egy-egy változó pár távolságát az definiálja, hogy az egyes megkérdozettek szerint milyen mértékben köthető 1-1 párthoz 1-1 érték. Az így kialakuló skálán való értékelésbeli különbségek adják a hasonlóságokat/különbségeket az egyes pártok egyes értékei kapcsán.

A változók szempontjából tehát irreleváns a tájolás – hiszen nincsen jelentősége északnak vagy délnek, ellentétben például a térképek ábrázolásánál. Számít viszont az objektumok/változók egymástól való távolsága, egymáshoz való viszonya. Ennek segítségével dönthetünk például arról, hogy

egy-egy kérdés egy-egy indexben, skálában összevonhatók-e (bár ezek csak kiinduló pontjai lehetnek skálák megalkotásának). Ugyanakkor ne feledjük el ilyen esetben azt sem, hogy egyes, egymástól távol kerülő változók⁴ esetén könnyen elképzelhető, hogy azonos jelenséget mérnek, csak a kérdések fordított tételeket tartalmaznak és így egymástól „távolra” kerültek.

ÖSSZEFOGLALÁS

Az MDS módszer széles körben alkalmazható olyan esetekben, amikor valamilyen módon csoportosításokat, kategorizálásokat szeretnénk készíteni. Ilyen szempontból rokonságban áll akár a klaszterezéssel (egyedek közötti kapcsolatok, hasonlatosságok keresésével), másik oldalról például a faktorelemzéssel (vagy főkomponens-elemzéssel), ahol a változók között keressük kapcsolatokat.

A faktorelemzés előnye az MDS eljárással szemben az, hogy a bevont változók egészen változatos helyekről, felmérésekből is származhatnak. Az MDS eljárásban a „távolság”, illetve „hasonlóság” feltételezése okán a változók bekerülése előre eldöntött, célirányos kutatások végeredményeként állnak elő.

Azonban a faktoranalízis során a változók közötti kapcsolat igen szigorú szabályszerűséggel kötött (lineáris viszonyok leírására alkalmas a módszer). Az MDS eljárás ezt nem követeli meg, így ebben az értelemben a „kapcsolatok” feltárása lényegesen kötetlenebbnek, általánosabbnak tekinthető, mint a faktoranalízis esetén.

IRODALOMJEGYZÉK

- Fasham, M. J. R. (1977). A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines, and coenoplanes. *Ecology*, 58(3), 551-561. Ketskeméty L. & Izsó L. (2006). *Bevezetés az SPSS programrendszerbe: Módszertani útmutató és feladatgyűjtemény statisztikai elemzésekhez*. Budapest: ELTE Eötvös Kiadó Kft.
- Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1-27.
- Móri, F. T. & Székely, J. G. (Szerk.). (1986). *Többváltozós statisztikai analízis*. Budapest: Műszaki Tankönyvkiadó.
- Niroomand, S., Takács, Sz., Vizvári, B. (2011). To lay out or not to lay out. *Annals of Operation Research*, 191(1), 183-192.
- Székely, M., Gyene, Gy., Pörzse, K. & Takács, Sz. (2008) Values and lifestyle: Do consumers declaring sustainable values demonstrate sustainable consumption patterns? Paper presented at the Sustainable Consumption

2008 Conference, Budapest, Corvinus University, October 8, 2008. In Vadovics E. & Gulyás E. (Eds.), *Sustainable consumption 2008: Academic conference proceedings* (pp. 134-149). Retrieved from: http://www.greendependent.org/konferencia/Sustainable%20Consumption%202008_Proceedings_final.pdf

Barna, I. & Székelyi, M. (2008). *Túlélőkészlet az SPSS-hez: Többváltozós lelemzési technikákról társadalomkutatók számára*. Budapest: Typotex.

-
- 1 Gondolhatunk arra, hogy a faktoranalízis során bevett eljárás az úgynevezett rotáció: ebben az esetben sem tudunk semmit javítani a modell teljes illeszkedésén, a modell matematikai jósága semmit sem változik a forgatástól – azonban a modell által magyarázható információk, a modellben rejlő összefüggések felfedezése, azok interpretálhatósága jelentős mértékben javítható, megkönnyítve így a modellel való további munkánkat.
 - 2 Említettük korábban, hogy a transzformációk hatására a tengelyek módosulhatnak. Míg a korábbi ábrán a két tengely jól láthatóan É-D és K-Ny tengelyek voltak, addig a transzformáció hatására a tengelyeink megfordultak. Nyilvánvaló, hogy jelen esetben a korábbi értelmezéshez képest rosszabb helyzetbe kerültünk – de az is nyilvánvaló, hogy adott esetben ilyen transzformációkkal találhatunk olyan tengelyeket, melyek segítségével a modellnek többletjelentést is adhatunk.
 - 3 Az adatok egy Dr. Székely Mózes által végzett kutatás adataiból származnak, melyről konferencia-kiadvány formájában (Székely, 2008) jelent meg e cikk leadásáig publikáció.
 - 4 Gondolhatunk itt arra, hogy normális eloszlású változók esetén negatívan korreláló változók (főként, ha erős a közöttük lévő korreláció) távolság vagy hasonlóság szempontjából igen távolinak látszódnak – míg az egymástól független változók közötti korreláció 0, ami távolság szempontjából megtévesztő lehet. Ezt azt jelenti, hogy amikor változók között keresünk kapcsolatokat, távolságokat, akkor e jelenséggel mindig számolnunk kell. Például egy $1 - |r|$ függvény segítségével nagyobb lesz az érték két olyan változó között, amelyek függetlenek. Így ha a két változó közötti korreláció 1, akkor ott lineáris függvénnyel leírható a közöttük lévő kapcsolat, így jogos, hogy távolságuk 0 legyen. Független esetben viszont $r=0$, tehát a közöttük lévő távolság 1 lesz.