Preferenciaadatok kiértékelése

Szakdolgozat

Írta: Simonka Fruzsina

Matematika BSc

Matematikai elemző szakirány

Témavezető:

Pröhle Tamás

egyetemi tanársegéd

Valószínűségelméleti és Statisztika Tanszék

Eötvös Loránd Tudományegyetem

Természettudományi Kar

2013

Tartalomjegyzék

1. fejezet

Bevezetés

Szakdolgozatom abból az alapvető kérdésből indul ki, hogy ki lehet-e fejezni igazságosan az egyéni véleményekből a közösség akaratát. A mindennapi élteben gyakran találkozhatunk azzal, hogy bizonyos preferenciasorrendet kell felállítanunk. Néha a saját preferenciánkat össze kell egyeztetnünk mások döntéseivel. Ilyen lehet például egy család lakóhelyváltoztatása, egy baráti kör szabadidős programjának kiválasztása. Sokszor ez a döntéshozatal a lakosság vagy a társadalom egy nagyobb csoportját érinti. Ilyen például az a kérdés, hogy hol épüljön szemétégető vagy éppen ki nyerjen meg szavazást.

Ez a kérdéskör olyan tág és komplex, hogy csak néhány részét érintem. Szakdolgozatomban bemutatom Kenneth Arrow közgazdász lehetetlenségi tételét, miszerint nem létezik olyan demokratikus szavazás, amely nem torzít, azaz minden egyén preferenciáját tükrözi. Áttekintem a többdimenziós skálázás módszerét is, melynek segítségével adatokat tudunk térben elhelyezni, ez alapján esetleg rejtett kapcsolatokat fedezhetünk fel az adatok között. Bemutatom az **R** program egyik skálázó függvényét is, aminek segítségével el is tudjuk végezni a többdimenziós skálázát. Végül kitérek egy valós esetre is, az Eurovíziós Dalfesztiválra. Ebben a témakörben az a kérdés foglalkoztat, hogy hogyan alakul a szavazás, fel tudunk-e fedezni a szavazásban valamilyen összefüggést.

2. fejezet

Preferencia

A preferencia különböző lehetőségek közötti képzeletbeli, illetve tényleges választást jelent. Ha egy személy preferál egy általa választható alternatívát egy másikkal szemben, akkor azt előnyben részesíti, tehát szívesebben választja, mint a másikat. A preferenciáknak két fajtája van: szigorú és megengedő preferencia. Dolgozatomban szigorú – a rendezésben egyenlőséget nem megengedő – preferenciákat használok. Egy szavazó preferencialistája az, amikor a jelölteket sorba állítja asszerint, hogy melyiket szeretné leginkább, és melyiket legkevésbé. Például, ha a, b és c lehetőségek közül választhatunk, akkor egy választó preferencialistája így nézhet ki: $b > c > a$. Ez azt jelenti, hogy b-t szeretné választani, de ha őt nem lehet, akkor c-nek jobban örülne, mint a-nak.

3. fejezet

Kiértékelési módszerek

Ebben a fejezetben a preferenciaadatok kiértékelésének néhány módszerét szeretném bemutatni, a teljesség igénye nélkül.

3.1. Arrow tétele

A társadalmi választás elméletében Arrow lehetetlenségi tétele – az általános lehetőség tétel, vagy Arrow paradoxona – azt állítja, hogy egy választási rendszer képes átalakítani az egyének rangsorolt preferenciát egy közösségi szintű csoportos döntéssé bizonyos kritériumok alapján. Ezek a kritériumok: az úgynevezett univerzális értelmezési tartomány, a diktátor -mentesség, a Pareto -elv és a lényegtelen alternatíváktól való függetlenség. Röviden a tétel azt állítja, hogy olyan választási rendszert kell kialakítani, amely kielégíti a következő követelményeket:

- Ha minden szavazó az X alternatívát preferálja Y -al szemben, akkor a csoport is X -et helyezi előtérbe Y -al szemben.

– Ha minden szavazó úgy változtatja meg preferenciáját, hogy az X és Y között változatlan marad, akkor a csoport preferenciája X és Y között is változatlan. Azaz a csoport preferenciája két lehetőség között kizárólag a csoport tagjainak két lehetőség közötti preferenciájától függ.

– Nincs olyan vezérszavazó, akinek a szavazata adná a végeredményt. Azaz nincs olyan szavazó, akinek preferenciájával a közös preferencia mindig megegyezik.

A preferenciák összesítésének, csoportpreferenciák kialakításának igénye sok különböző tudományágban felbukkan: a jóléti közgazdaságtanban, a marketingben, döntéselméletben, és természetesen a szavazási rendszerekben. Tegyük fel, hogy minden szavazó megadja a saját egyéni preferencia-sorrendjét. Keressük azt az úgynevezett társadalmi jóléti függvényt, amely átalakítja a szavazók egyéni preferenciáit egyetlen átfogó társadalmi preferencia-sorrenddé. Ezt a függvényt angolul *social welfare function* néven ismerjük. Dolgozatomban gyakran használom a szó szerinti fordítását is, de véleményem szerint ezt a kifejezést nem lehet igazán jól magyarosítani, így néha a szociális függvény kifejezést használom. A tétel a következő tulajdonságokat – egy korrekt szavazási rendszer feltételezetten ésszerű követelményeit – veszi tekintetbe:

– ***univerzális értelmezési tartomány***: az egyéni választói preferenciák bármely halmazára a társadalmi jólétű függvénynek meg kell adnia a társadalmi alternatívák egyértelműen meghatározott és az összes alternatívára kiterjedő sorrendjét. Ezt oly módon kell megtennie, hogy a társadalom számára egy teljes preferenciasorrendet eredményezzen.

Minden alkalommal, amikor a szavazói preferenciákat azonos módon határozzuk meg, akkor determinisztikusan megegyező sorrendet kell adnia.

- ***irreleváns alternatíváktól való függetlenség***: az X és Y közötti szociális preferenciának csak az X és Y közötti egyéni preferenciától kellene függenie. Általánosabban: az irreleváns alternatívák egyéni sorrendjében történő változások nem lehetnek hatással a társadalmi sorrendre.

- ***Pareto-elv***: ha minden egyén jobban preferál egy bizonyos választási lehetőséget egy másikkal szemben, akkor a társadalmi preferenciasorrendben is így kell lennie.

- ***diktatúra-mentesség***: a szociális függvény nem tükrözheti mindig egy és ugyanazon szavazó preferenciáját. A közös eredményt bármely szavazó kívánsága befolyásolhatja.

Ezeket a feltételeket sokféleképpen meg lehet fogalmazni. Bizonyos esetekben nagymértékben enyhíthetőek azonos kimenetel mellett. Ugyanakkor szigorúbb feltételek mellett a tétel áttekinthetőbb és könnyebben bizonyítható lehet.

Legyen $A = \{a, b, c, \dots\}$ az alternatívák véges, n elemű halmaza, N pedig a szavazók, illetve a döntési szempontok száma. Jelölje $L(A)$ az A halmaz összes teljes lineáris rendezéseit. Minden ilyen $>$ reláció az A elemeinek egy permutációját adja meg úgy, hogy ha valamely $a, b \in A$ elemekre $a > b$, akkor az elrendezésben a megelőzi b -t. Minden $i = 1, \dots, N$ -re az i. választó megad egy R_i lineáris teljes rendezést A -n.

A szociális jóléti függvény egy $F : L(A)^N \to L(A)$ függvény, ami aggregálja a szavazók preferenciáit egyetlen preferenciarendezéssé az A halmazon, tehát $R = F(R_1, R_2, \ldots, R_N)$. A szavazók preferenciáinak egy $(R_1, \ldots, R_N)$ N -esetét preferencia profilnak nevezzük.

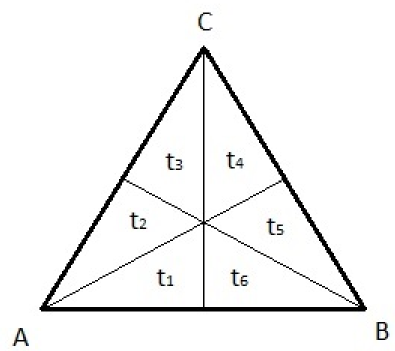
3.1. Tétel. (Arrow) *Tegyük fel, hogy a szavazók több mint kettő alternatíva közül választhatnak, azaz $|A| > 2$. Ekkor nem létezik olyan társadalmi jóléti függvény, amely a következő feltételeket együttesen teljesítené:*

- ***Pareto-elv:*** *ha az a alternatíva előrébb található, mint a b minden $R_1, \ldots, R_N$ rendezésben, akkor a magasabban van rangsorolva, mint b $F(R_1, R_2, \ldots, R_N) = R_i$ által is.*
- ***diktátor-mentesség:*** *nincs olyan $i \in 1, 2, \ldots, N$, amelyre minden $(R_1, \ldots, R_N) \in L(A)^N$ esetén $F(R_1, R_2, \ldots, R_N) = R_i$.*
- ***irreleváns alternatíváktól való függetlenség:*** *ha $(R_1, \ldots, R_N)$ és $(S_1, \ldots, S_N)$ két preferencia profil úgy, hogy minden i -re az a és b alternatívák sorendje ugyanaz R_i -ben és S_i -ben, akkor az a és b alternatívák sorrendje ugyanaz $F(R_1, \ldots, R_N)$ -ben és $F(S_1, \ldots, S_N)$ -ben is.*

Az utóbbi feltételt a későbbiekben a következő formában fogom használni: bináris függetlenség az, amikor az a, b alternatívapáros rendezése a kollektív rendezés szerint csak az egyének a -ra és b -re vonatkozó preferenciájától függ. Arrow lehetetlenségi tételének számos különböző bizonyítása van. Számomra a geometria szemléletű bizonyítás a legszimpatikusabb, talán ez tűnik ki a legjobban a többi közül, így ezt a bizonyítási módszert szeretném a továbbiakban bemutatni.

Bizonyítás:

Könnyen látható, hogy tetszőleges számú alternatíva mellett elégséges három alternatívára belátni a tételt. Így a bizonyítást három alternatívára végezzük el. Vegyünk egy szabályos háromszöget, melynek csúcsai (A, B, C) legyenek a jelöltek. Húzzuk be az összes felezőmerőlegest. Ezzel hat egyenlő részre osztottuk a háromszöget. A kis háromszögeket az alábbi ábrán látható módon elnevezzük:



A hat kis háromszög egyértelműen megfeleltethető a jelöltek közötti összes lehetséges permutációnak. Ez úgy lehetséges, hogy ha vesszük a háromszög egy tetszőleges p pontját, akkor az meghatároz egy sorrendet a csúcsok között az alapján, hogy p melyik kis háromszögben van. A preferencia rendezést a csúcsoktól való távolság határozza meg:

$p \in t_1 : A > B > C$

$p \in t_2 : B > A > C$

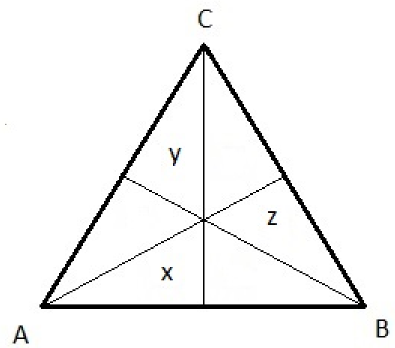
$p \in t_3 : B > C > A$

$p \in t_4 : C > B > A$

$p \in t_5 : C > A > B$

$p \in t_6 : A > C > B$

Legyen a szavazók száma $n \in \mathbb{N}$, $n \geq 3$ és tegyük fel, hogy az összes szavazó a t_1, t_3, vagy a t_5 kis háromszögeknek megfeleltetett preferencia sorrendekből választott. Jelöljék rendre az $x, y, z \in \mathbb{N}$ számok, hogy mennyi szavazat esett a t_1, t_3 és t_5 kis háromszögekbe:



Ekkor nyilvánvalóan teljesülnie kell az $x + y + z = n$ egyenletnek. Legyen G azon profilok halmaza, amelyekre $x + y + z = n$ teljesül, és legyen $H \subseteq G$ azon elemek összessége, amelyekre a szociális függvény az $A > B > C$ rangsort generálja. Ha $x = n$ lenne, akkor a Pareto- elv miatt a rendezés $A > B > C$, tehát ez a profil benne van a H halmazban. Ha viszont az $x + y + z = n$ egyenlet teljesül, akkor az a profil biztosan nem lesz benne H-ban, mert akkor a Pareto-elv miatt $C > A$, ez pedig ellentmond a H halmazba kerülés feltételének. Tekintsük H elemei közül azokat, amelyekben a legkevesebb t_1 -es típusú szavazó van, azaz ahol x minimális. Ez a profil legyen p. Tehát $F(p) = A > B > C$. Vegyünk egy t_1 -es típusú szavazót p-ben és változtassuk meg a preferenciáját úgy, hogy a többi szavazó preferenciái változatlanok maradjanak. Belátjuk, hogy ez a szavazó egy diktátor. Kezdjük azzal az esettel, amikor a szavazónk t_3 -as típusú lesz. Így keletkezik egy új profil, amelyet p_3 -al jelölünk. Ha $F(p_3) = A > B > C$ lenne, akkor ellentmondásra jutnánk a kezdeti feltevéssel. A bináris függetlenség miatt $F(p_3) = C > A > B$ vagy $F(p_3) = A > C > B$. Vegyük az utóbbit.

Most nézzük meg azt az esetet, amikor a szavazónk t_5 -ös típusú lesz. Az így keletkezett profil legyen p_5. Ez a mozgatás a szavazót a B, C közömbösségi vonaltól jobbra tartja, ezért B, C szociális rangsora egyezik p -ével, vagyis $B > C$. $F(p_3)$ és a bináris függetlenség miatt A, C szociális rangsora p_3 -nál és p_5 -nél azonos. Azt tudjuk, hogy $F(p_5) \neq A > B > C$. Ezért $F(p_5) = B > A > C$ lehet. Ugyanígy, ha a p_4 profilt tekintjük, akkor $F(p_4) = B > A > C$ lehet. A bináris függetlenség miatt B, C szociális rangsora egyezik p_3 -nál és p_5 -nél. Ez ellentmondás. Tehát tudjuk, hogy $F(p) = A > B > C$ és $F(p_3) = C > A > B$. Innen egyszerűen adódik, hogy a többi $F(p_i)$ azonos a szavazónk rangsorával, azaz azokkal, amit a t_1 -es kis háromszög meghatároz. Meg kell jegyezni, hogy csak a t_1 -es típusú szavazó preferenciáját változtattuk meg. Figyeljük meg, hogy mi történik akkor, amikor egy szavazó az óramutató járásával megegyező irányban lépked az egyes területekre. Az történik, hogy a páros preferenciák közül csak egy változik meg, a többi marad. Például, ha a t_1 -es típusú személy t_6 -ra mozdul, akkor csak A, B sorrendje változik, a B, C és A, C sorrendje marad. A választott t_1 -es típusú szavazónkat ezentúl kvázi-diktátornak fogjuk hívni. Jelöljük $p_j \{1 \to 6\}$ -al azt, hogy a p profilban a t_1 -es típusú szavazóból és néhány másik t_1 -es típusúból t_6 -os lesz. Ekkor így keletkezik a p_j profil. Ismét a bináris függetlenséget alkalmazva azt kapjuk, hogy $F(p_j \{1 \to 6\}) = A > C > B$. Hasonlóan okoskodva innen megkapjuk az összes $F(p_j)$ -t. Azt kapjuk, hogy $j = \{1, 3, 4, 6\}$ -ra a szociális rangsor megegyezik a kvázi-diktátoréval. Látható, hogy az új szavazó megint helyet cserélhet, mondjuk t_5 -re. Ahhoz, hogy a kvázi-diktátor egyben diktátor is legyen, csak az kell, hogy hagyjuk a többi szavazót is változtatni. Így p -ről bármely más profilra mozdulhatunk. $\square$

3.2. Többdimenziós skálázás

A többdimenziós skálázás egy többváltozós statisztikai módszer, melynek segítségével egy adathalmazt térben elhelyezkedő ponthalmazként tudunk ábrázolni. Általánosan arra törekedünk, hogy úgy ábrázoljunk adatokat, hogy az egymáshoz valamilyen szempontból közelebbinek vélt objektumok az ábrázolásban is közel kerüljenek egymáshoz, a távolabbinak vélt adatok pedig az ábrázolásban is távolibbak legyenek egymástól. Ezek az ábrázolások olyan geometriai reprezentációk, amelyek az ábrázolt adatok viszonyát valamilyen szempontból helyesen, de legalább közelítőleg helyesen adják vissza. A többdimenziós skálázás módszerének segítségével adott objektumokra vonatkozó észlelt hasonlósági, vagy éppen különbözőségi adatokból szisztematikusan létrehozhatunk olyan geometria reprezentációkat, amelyek ezen objektumok észlelt viszonyát egy megfelelő dimenziószámú geometriai térben a lehető legkisebb torzítással tükrözik vissza.

Az eljárás eredménye mindig egy ponthalmaz "térképe", amelyen az egyes pontok úgy helyezkednek el, hogy az egymás közötti távolságaik megfelelnek az adott pontokhoz tartozó objektumok közötti különbözőségeknek. Talán a többdimenziós skálázás legfőbb ereje az, hogy a pszichológiai eszközökkel nyert különbözőség-érzékelési adatok alapján lehetővé teszi addig nem ismert, de sokszor jelentős szerepű dimenziók felismerését.

3.2.1. Matematikai modellje

A többdimenziós skálázást sokféleképpen és különböző feltételek mellett lehet végrehajtani. Szakdolgozatomban a teljesség igénye nélkül, általánosan szeretném bemutatni ezt a módszert.

Első lépésben szükségünk lesz egy $k \times n$ -es X adatmátrixra, amellyel az $\mathbb{R}^k$ -beli, n pontból álló ponthalmazt azonosítjuk. Ennek a mátrixnak az oszlopai azok a vektorok, amelyek az origóból a ponthalmaz egyes pontjaiba mutatnak. Szükségünk lesz egy távolságmátrixra is, melynek i. sorának j. eleme egyenlő az (i, j) pontpár euklideszi távolságának négyzetével.

3.1. Definíció. *A $T \in \mathbb{R}^{n \times n}$ mátrixot **távolságmátrix**nak nevezzük, ha a következő tulajdonságokkal rendelkezik:*

(i) elemei nemnegatívak, azaz minden $i, j \in \{1, \dots, n\}$ -re $t_{ij} \geq 0$,

(ii) szimmetrikus, tehát $t_{ij} = t_{ji}$, ahol $1 \leq i < j \leq n$ és

(iii) diagonálisa nulla, azaz minden $i \in \{1, \dots, n\}$ -re $t_{ii} = 0$.

Az X ponthalmaz középpontja vagy centruma az a pont, amelynek minden koordinátája a pontok megfelelő koordinátáinak átlaga. Egy ponthalmazt centrált ponthalmaznak nevezünk, ha centruma az origó.

Szükségünk lesz még egy S mátrixra, amely úgynevezett skalárszorzat mátrix. Ez a mátrix $n \times n$ méretű és i. sorának j. eleme annak a két vektornak a skalárszorzata, amelyek a ponthalmaz középpontjából az i és a j pontba mutatnak. Jelöljük a csupa egyesekből álló n hosszú vektort u -val, tehát $u = (1, \dots, 1)^T$, és U -val a csupa egyesekből álló $n \times n$ -es mátrixot, azaz $uu^T = U$.

3.2. Definíció. *Az $S \in \mathbb{R}^{n \times n}$ mátrixot **skalárszorzat mátrix**nak nevezzük, ha a következő tulajdonságokkal rendelkezik:*

(i) *szimmetrikus, tehát $s_{ij} = s_{ji}$, ahol $1 \leq i < j \leq n$,*

(ii) *egy nullához tartozó sajátvektora az u vektor és*

(iii) *pozitív szemidefinit, azaz minden sajátértéke ≥ 0.*

Szükségünk lenne arra, hogy minden ponthalmaz középpontja az origóba essen, azaz centrált legyen. Ezt úgy tudjuk elérni, hogy bevezetünk egy úgynevezett centráló mátrixot. Ha ezzel a mátrixszal jobbról megszorozzuk a ponthalmazt, akkor ez egy olyan eltolást eredményez, hogy a kapott új ponthalmaz már centrált lesz.

3.3. Definíció. *A $C \in \mathbb{R}^n$ mátrixot **centráló mátrix**nak nevezzük, ha $C = I - \frac{1}{n}U = I - \frac{1}{n}uu^T$.*

A centráló mátrix a következő tulajdonságokkal rendelkezik:

(i) *szimmetrikus,*

(ii) $u^T C = Cu = 0$,

(iii) $CU = UC = 0$ *és*

(iv) *idempotens, azaz $CC = C$. Ezt értelmezhetjük úgy is, hogy mivel a C -t C -vel jobbról szorozva C -t kapunk, (a baloldali) C egy centrált ponthalmaz.*

3.1. Állítás. *Egy X általános helyzetű ponthalmaz skalárszorzat mátrixa $S_X = CX^T XC$, egy Y centrált ponthalmazé pedig $S_Y = Y^T Y$.*

Legyenek az S sajátvektorai a hozzájuk tartozó $\lambda_1, \dots, \lambda_n$ sajátértékek csökkenő sorrendjében $v_1, \dots, v_k, v_{k+1} = u, v_{k+2}, \dots, v_n$. Ekkor $\lambda_{k+1} = \dots = \lambda_n = 0$. Az X ponthalmaz $k \times n$ -es mátrixa $X = (\sqrt{\lambda_1}v_1, \dots, \sqrt{\lambda_k}v_k)^T$. Ez a ponthalmaz centrált, hiszen a benne szereplő sajátvektorok ortogonálisak a $v_{k+1} = u$ sajátvektorra, ezért $Xu = 0$.

Jelöljük $Diag(Z)$ -vel azt a mátrixot, melynek diagonális elemei megegyeznek Z-vel, de a többi eleme nulla. Ennek segítségével felírhatjuk zárt alakban is a T_X -et. Egy-egy háromszög és a koszinusz-tétel segítségével kiszámíthatjuk a pontpárok távolságait a skalárszorzatokból. Ehhez a pontpár mellé szükségünk van egy tetszőleges harmadik pontra. Írjuk fel két különböző harmadik pont esetén az X ponthalmazhoz tartozó távolságmátrixot. Az egyik ilyen pont legyen az origó, a másik pedig a ponthalmaz középpontja. Nyilván a két távolságmátrix egyenlő: $T_X = Diag(X^TX)U + UDiag(X^TX) - 2X^TX = Diag(CX^TXC)U + UDiag(CX^TXC) - 2CX^TXC$.

Előzőleg láttuk, hogy $S_X = X^TX$, valamint $S_X = CX^TXC$. Ezeket felhasználva az előző képletben két leképezést értelmezünk. Az első a τ leképezés, amelyet a ponthalmazhoz tartozó skalárszorzat mátrixon értelmezünk: $\tau(S_X) = Diag(S_X)U + UDiag(S_X) - 2S_X$. A másik leképezés a σ, ezt pedig a távolságmátrixon érlemezzük: $\sigma(T_X) = -\frac{1}{2}CT_XC$. Ezeket felhasználva közvetlenül adódik, hogy $T_X = \tau(S_X)$ és $S_X = \sigma(T_X)$.

3.2. Állítás. *Tetszőleges T távolság- és S skalárszorzat mátrixra $\tau(S) = Diag(S)U + UDiag(S) - 2S$ és $\sigma(T) = -\frac{C^TC}{2}$.*

3.3. Állítás. *Tetszőleges S skalárszorzat mátrix és T távolságmátrix esetén*

(i) $\sigma(\tau(S)) = S$, *valamint*

(ii) $\tau(\sigma(T)) = T$.

Bizonyítás:

(i) Tetszőleges S skalárszorzat mátrixra a $\tau(S)$ nemnegatív, szimmetrikus és diagonálisa 0. Ezért $\sigma(\tau(S))$ értelmezhető és $\sigma(\tau(S)) = -C(Diag(S)$
$U + UDiag(S) - 2S)\frac{C}{2} = -\frac{(CDiag(S)UC+CUDiag(S)C-2CSC)}{2} = S$. Az utolsó egyenlőség tényleg fennáll, mert C -t kifejtjük a definíciója szerint, és felhasználjuk az S szimmetriája miatt érvényes $US = SU = USU = 0$ egyenlőséget.

(ii) Tetszőleges T távolságmátrixra $\sigma(T) = -\frac{CTC}{2}$ szimmetrikus. $Cu = 0$ miatt $\sigma(T)u = 0$. Tehát $\sigma(T)$ skalárszorzat mátrix, és így $\tau(\sigma(T)) = Diag(\sigma(T))U + UDiag(\sigma(T)) - 2(\sigma(T))$ értelmezhető. Ha felhasználjuk a következőket: $\sigma(T)$ és C deffiníció szerinti értékeit, $Diag()$ additív tulajdonságát, a $Diag(T) = 0$, $Diag(UT) = Diag(TU)$, $Diag(UT)U = TU$, $UDiag(TU) = UT$ és a $Diag(UTU)U = U$ $Diag(UTU) = UTU$ egyenlőségeket, akkor tényleg teljesül, hogy $\tau(\sigma(T)) = T$.

□

3.2. Tétel. *Egy k dimenziós X ponthalmazhoz tartozó T távolságmátrix akkor reprezentálható, ha a $\sigma(T) = -\frac{1}{2}CTC$ – vagyis a T-hez tartozó skalárszorzat mátrix – k rangú pozitív szemidefinit.*

3.4. Állítás. *Legfeljebb $n-2$ dimenzióban minden távolságmátrix ábrázolható konstans hibával.*

4. fejezet

Az R program egy skálázó függvénye

Az **R** szoftverben találhatunk egy `cmdscale` nevű beépített függvényt. Ez a legegyszerűbb **R** -beli skálázó függvény és a következőképpen néz ki: `cmdscale(d, k = 2, eig = FALSE, add = FALSE, x.ret = FALSE)`. A `d` argumentum jelöli a távolságmátrixot, a `k`$\in \{1, \ldots, n-1\}$ pedig annak a térnek a maximális dimenzióját, amelyben az adatokat ábrázolni szeretnénk. Az `eig` argumentummal megadhatjuk, hogy a sajátértékek szerepeljenek-e (`TRUE`), vagy sem (`FALSE`) az eredményben. Ha használjuk az `add=TRUE` argumentumot, akkor keres egy olyan `ac` konstanst, amivel megnövelve az ábrázolandó távolságokat, a pontpárok távolságait jól közelítő ábrát ad vissza. Az `x.ret=TRUE` input paraméterrel egy kitüntetett szerepű szimmetrikus mátrixot kaphatunk vissza.

A `d` távolságmátrixot a `dist` fügvénnyel készíthetjük el: `dist(x, method = "euclidean", diag = FALSE, upper = FALSE, p = 2)`. Az `x` argumentum az adatmátrix. A `method` input paraméterrel adhatjuk meg, hogy melyik távolság-formulát alkalmazza a távolságmátrix kiszámításában. A következő távolságok közül választhatunk:

- `euclidean`: $\sqrt{\sum |x_i - y_i|^2}$
- `maximum`: maximális koordináta-különbség
- `manhattan`: a koordináta-különbségek összege
- `canberra`: $\sum \frac{|x_i - y_i|}{|x_i + y_i|}$
- `binary`: ha az objektumok vektorai 0/1 vektorok, akkor azon pozíciók hányada, amelyekben a két bit különböző
- `minkowski`: $\sqrt[p]{\sum |x_i - y_i|^p}$.

A `diag=TRUE` argumentum segítségével a távolságmátrix főátlóbeli, az `upper=TRUE` pedig a felsőháromszög-mátrix elemeit kérhetjük vissza. A `p` paraméter a Minkowski-féle együttható.

5. fejezet

Eurovíziós dalfesztivál

Az Eurovíziós Dalfesztivál egy évente megrendezett zenei verseny az Európai Műsorsugárzók Uniójának (EBU) aktív tagállamai között. A fesztivál keretében minden résztvevő ország benevez egy zeneszámot. Az élőben közvetített döntőben minden versenyző előadja dalát. A döntőt mindig abban az országban rendezik meg, akinek versenyzője az előző évben nyertes volt. A versenyt 1956-os kezdése óta minden évben megtartották, így egyike a világ leghosszabb ideig tartó televíziós műsorainak.

5.1. Szavazás szabályai

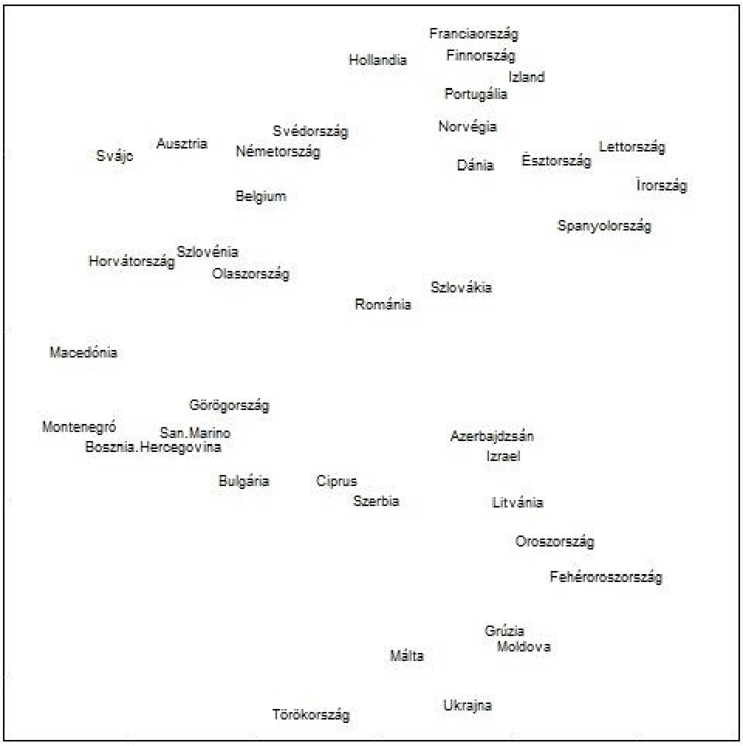
A dalfesztivál több részből áll: országonként 2 elődöntőből és a végső közös döntőből. Mivel dolgozatom szempontjából a döntő lényeges, így csak ennek a szavazási rendszerére térek ki. A szavazási rendszer az évek folyamán sokat változott, talán még nem is fejeződött be ez a folyamat. A következőekben a jelenlegi (2013) szavazási rendszert mutatom be.

A résztvevő országokban a nézők telefonos szavazás keretein belül szavazhatnak kedvenc dalaikra. Természetesen a saját országuk által benevezett dalra nem szavazhatnak.

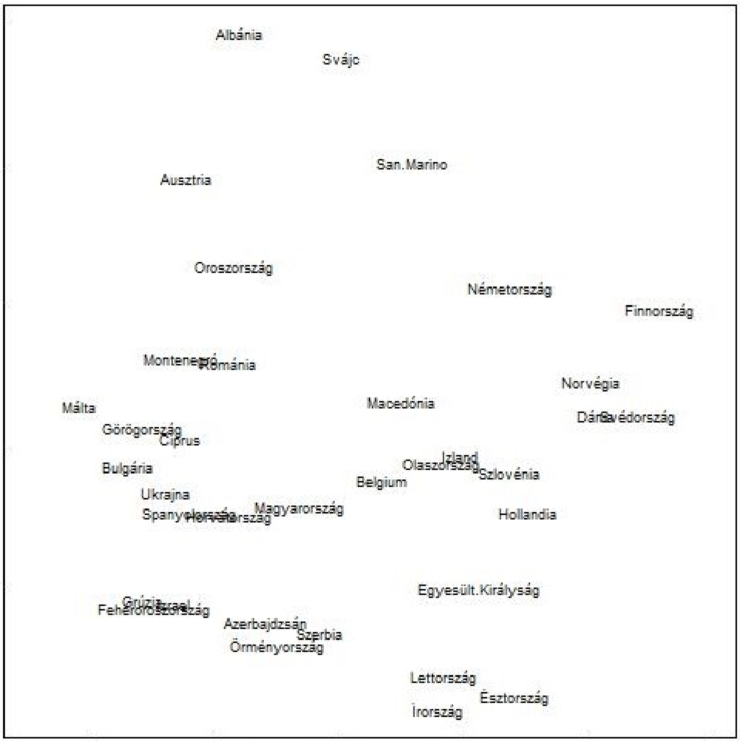
Továbbá minden országban a résztvevő műsorszolgáltató kijelöl egy öt tagú nemzeti zsűrit. A zsűri és a nézői szavazatok fele-fele arányban befolyásolják az adott ország döntését. A nézői szavazatokat és a zsűri döntését összefésülve meghatározzák a 10 legtöbb szavazatot elért versenyzők sorrendjét. Az első helyezett 12 pontot, a második 10 pontot, a harmadik 8 pontot, minden további helyezett pedig egyel kevesebb pontot kap. Így a 10. helyen végzett versenyzőnek 1 pontot ítélnek meg. Ha az együttes rangsorban döntetlen alakulna ki, akkor az a dal nyer, amelyik több nézői szavazatot kapott. Végül az élőben közvetített döntőn minden résztvevő ország szóvivője közli a saját országa által megítélt pontokat. Minden ország kapott pontszámait összeadják. Az nyer, aki a legtöbb összpontszámot kapja. Amennyiben a döntő végén az első helyen döntetlen alakulna ki, akkor megszámolják, hány országtól kaptak pontokat a döntetlen országok, és amelyik a legtöbb másik országtól kapott pontot, az lesz a győztes. Ha az eredmény még így is döntetlen, a maximális pontszámok (12 pont) elnyerésének a számát hasonlítják össze. Ha még mindig döntetlen lenne, akkor a 10 pontok számát, majd a 8 -ét, stb. hasonlítják össze. Abban a nagyon valószínűtlen esetben, ha még mindig döntetlen lenne az állás, akkor ezeket az országokat közösen kiáltják ki győztesnek.

5.2. Skálázás eredménye

Zárásként nézzük meg a 2012-es és 2013-as évi Eurovíziós Dalfesztivál adataira elvégzett skálázás eredményét. Az előző fejezetben bemutatott **R** függvény felhasználásával készített eredmények a következők:



2012-es adatok alapján



2013-as adatok alapján

Mindkét "térkép" azt ábrázolja, hogy az egyes országok kire is szavaznak. Talán a legtöbben arra számítanak, hogy a szomszédos országok egymásra szavaznak. Ez a feltevés mindkét évben több esetben is látszik. Mindkét esetben alakultak ki klikkek. Például az északi országok (Dánia, Finnország és Norvégia) mindkét vizsgált évben egymásra szavaztak. Viszont néhány esetben megdőlni látszik ez az elmélet. Például 2012-ben Portugália, 2013-ban pedig Izland esetében. Tehát biztosan nem jelenthetjük ki, hogy az országok szomszédsága befolyásolná az Eurovíziós Dalfesztivál szavazását.

Köszönetnyilvánítás

Ezúton szeretném megköszönni témavezetőmnek, Pröhle Tamásnak, hogy ötleteivel, tanácsaival és rengeteg segítségével hozzájárult szakdolgozatom elkészítéséhez. Köszönettel tartozom családomnak és barátaimnak támogatásukért és bátorításukért.

NYILATKOZAT

Név: Simonka Fruzsina

ELTE Természettudományi Kar, szak: Matematika BSc

ETR azonosító: SIFQAAT.ELTE

NEPTUN kód: ANVPEB

Szakdolgozat címe: Preferenciaadatok kiértékelése

A szakdolgozat szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló munkám eredménye, saját szellemi termékem, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2013. 05. 31.

Irodalomjegyzék

[1] Pröhle Tamás: *Elektronikus statisztika jegyzet II. kötet, 3. fejezet* (2013)

[2] https:en.wikipedia.orgwikiArrow_Impossibility_theorem

[3] http:hu.wikipedia.orgwikiEurov%C3%ADzi%C3%B3s_Dalfesztiv%C3%A1l

[4] http:kgk.uni-obuda.husitesdefaultfilesdtr.pdf

[5] http:stat.ethz.chR-manualR-patchedlibrarystatshtmlcmdscale.html

[6] http:stat.ethz.chR-manualR-patchedlibrarystatshtmldist.html

[7] http:www.cs.elte.hublobsdiplomamunkakalkmat2008/kiss_csaba.pdf

[8] http:www.cs.elte.hublobsdiplomamunkakalkmat2011lorincz_geza.pdf

[9] http:www.cs.elte.hublobsdiplomamunkakbsc_alkmat2012karpati_laszlo.pdf

[10] http:www.eurovision.tvpagetimeline

[11] http:www.erg.bme.huszakkepzes2felevmds.pdf

[12] http:www.escholarship.orgucitem96n108tsh

[13] http:www.kommunista.netblogegy-kis-matematika-a-demokratikus-szavazas-lehetetlensegerol

[14] http:www.math.upenn.edu˜kazdan210votinggeanakoplos.pdf