

# A TÖBBSZÖRÖS KORRESPONDENCIA-ELEMZÉS (MCA) ELMÉLETE ÉS GYAKORLATA

*Molnár László, laszlo.molnar@uni-miskolc.hu*  
*Miskolci Egyetem, Marketing Intézet*

## Abstract

Due to the evolution of market-research, which is one of the disciplines of science, better and better methods for analysis of data have become everyday practice. Among these, are such method is MCA, the theoretical basis of which was founded in 1960s.

This method is well known in Hungary and widely used by researchers is market-research and public opinion. However, it is rarely come across in economic studies in higher education or even at doctoral level. Also, the method is only briefly discussed in economic text books, written in Hungarian.

This paper has two aims: firstly, I will describe the theoretical details of MCA, paying particular attention to the different steps in the application. Secondly, I will demonstrate the efficiency of the method with an example, for which the input data was generated by using a special questionnaire.

## 1. Bevezetés

Ahogy fejlődik a marketingkutatás, mint tudományos diszciplína, újabb és újabb (matematikai-statisztikai) adatelemzési módszerek épülnek be a mindennapi gyakorlatba. Ezek közé a módszerek közé sorolható a többszörös korrespondencia-elemzés (Multiple Correspondence Analysis, MCA), amely elméleti kereteit az 1960-as években már rögzítették.

Természetesen a hazai gyakorlat számára sem ismeretlen a módszer, sőt a piac- és közvélemény-kutató cégek előszeretettel alkalmazzák is. Ugyanakkor ritkán találkozhatunk vele a felsőoktatásban, de még a doktori iskolákban sem gyakori az oktatása. Nem is beszélve arról a néhány magyar nyelvű szakirodalomról, amelyek csupán rövid említést tesznek erről a módszerről.

Jelen tanulmány célja kettős. Egyrészt ismertetni kívánom a többszörös korrespondencia-elemzés legfontosabb elméleti tudnivalóit, különös tekintettel a módszer egyes lépéseire. Másrészt egy konkrét gyakorlati példán keresztül kívánom bemutatni az alkalmazását, amely input adatait külön ebből a célból készített kérdőíves megkérdezés nyújtotta.

## 2. A többszörös korrespondencia-elemzés elmélete

A többszörös korrespondencia-elemzés lényegét Hajdu [5] definíciója alapján határozom meg, vagyis egy olyan exploratív többváltozós technikáról van szó, amely az asszociáció kapcsolat vizuális elemzése érdekében egy három- vagy többváltozós kontingencia tábla adatait grafikus ábrává konvertálja. Egyszerűbben fogalmazva a többszörös korrespondencia-elemzés a három- vagy többváltozós keresztábra sorait az oszlopok, míg az oszlopait a sorok tengelyeinek tekintetében egy „pontfelhő” pontjaiként

értelmezi. A módszer eredményeként egy redukált, alacsony dimenziójú térben grafikusán ábrázoljuk a „pontfelhő” pontjait. Mindezek után vizuális elemzéssel következtetni tudunk arra, hogy a vizsgált változók mely kategóriái vonzzák és melyek taszítják egymást.

### 2.1. A többszörös korrespondencia-elemzés menete

**A probléma meghatározása.** A probléma megfogalmazása során a kutatónak meg kell határoznia azt a célt, amelyre a többszörös korrespondencia-elemzés eredményeit szeretné felhasználni, és ki kell választani azokat a változókat, amelyek bekerülnek az elemzésbe. Minden változó esetében definiálni kell a kategóriákat, hogy elkészülhessen az észlelési térkép, azonban 24 kategória értékelése már nehézkes, az eredmények értelmezése túlságosan bonyolulttá válik.

**A kiinduló adatok összegyűjtése.** A többszörös korrespondencia-elemzéshez csak nominális skálán mért változókra van szükség, vagyis a kiinduló adatok összegyűjtése során arra kell megkérni a válaszadókat, hogy ítéljék meg, hogy egy változó mely kategóriái vonatkoznak a többi változó kategóriáira.

**A kiinduló többdimenziós keresztábra előállítás.** A kiinduló többdimenziós keresztábra csupán a vizsgálatba vont változók adott kategóriáira vonatkozó pozitív válaszokat tartalmazza.

**A kiinduló többdimenziós keresztábra konvertálása kétdimenziós táblázattá.** A kiinduló többdimenziós keresztábrát ún. indikátor-változó bevezetésével kétdimenziós táblázattá konvertálhatjuk.

**Sor- és oszlopprofilok előállítás.** A sor- és oszlopprofilok előállítása egyszerűen történik: A sorváltozó kategóriáinak értékeit a megfelelő sorösszegekkel, míg az oszlopváltozó kategóriáinak értékeit a megfelelő oszlopösszegekkel kell elosztani.

**A kategóriák geometriai súlypontjainak előállítás.** A kategóriák geometriai súlypontjának előállítása azért szükséges, hogy egyszerűen kiszámítható legyen az adatok saját centroidjaik, azaz peremprofiljaik körüli szóródási mutatószámok.

**A kategóriák közötti távolság kiszámítása.** A kategóriák közötti és geometriai súlyponttól mért távolság kiszámításával kapunk információt a hasonlóságokról és a különbözőségekről, amelyet normál esetben Euklédészi-távolsággal határozunk meg. A korrespondencia-elemzés esetében azonban  $\chi^2$ -távolsággal célszerű számolni, mivel eliminálja az eltérő peremgyakoriságokból fakadó torzító hatást.

**Az adatok súlyozott varianciájának kiszámítása.** Az adatok súlyozott varianciájának (inercia) kiszámítása azért szükséges, hogy információt kapjunk az egyes kategóriák geometria súlypontjaik körüli szóródásának mértékéről.

**A dimenziók számának csökkentése.** A dimenziószám csökkentésekor azt az alacsonyszámú többdimenziós teret keressük, amely a legkisebb információ-vesztéssel jár. Fontos megjegyezni, hogy a vizuális elemzése csak két-, legfeljebb három dimenzióban lehetséges. A dimenziószám csökkentésének lineáris algebrai módszere a szinguláris érték felbontás (Singular Value Decomposition, SVD).

**A konfiguráció interpretálása.** A konfiguráció vagy a többszörös korrespondencia-térkép a koordináták és a vizsgálatba vont változók kategóriái pozíciójának elemzésével értelmezhető. Az egymáshoz közel elhelyezkedő kategóriák hasonlóknak érzékelik a válaszadók, míg a távol elhelyezkedőket különbözőeknek. Azok a kategóriák, amelyek közel találhatóak az origóhoz, kevésbé markáns, míg azok, amelyek távol találhatóak attól, markáns tulajdonságokkal, egyedi imázssal rendelkeznek.

**A megbízhatóság és az érvényesség értékelése.** A többszörös korrespondencia-elemzés megbízhatóságát és érvényességét többféle technika segítségével ellenőrizhetjük:

- A kiinduló adatokat két vagy több részre osztjuk, majd a többszörös korrespondencia-elemzést mindegyik részre külön elvégezzük, és az eredményeket összehasonlítjuk.
- A változók kategóriáit felváltva eltávolítjuk az eredeti adatokból és elvégezzük az elemzést a megmaradt kategóriákra.
- Véletlen hibát adunk az eredeti adatokhoz, majd az így kapott adatokra elvégezzük az elemzést, és az eredményeket összehasonlítjuk.
- Az eredeti adatok gyűjtését két különböző időpontban tesszük meg, majd tesztadatok-kontrolladatok módszerrel meghatározzuk az elemzés megbízhatóságát.

### 3. A többszörös korrespondencia-elemzés gyakorlata

A többszörös korrespondencia-elemzés legfontosabb elméleti tudnivalóinak áttekintése után az életből vett gyakorlati példán keresztül mutatom be a használatát, különös tekintettel az input adatokra, az elemzés menetére és az eredmények értelmezésére.

#### 3.1. A kutatás módszertana

A többszörös korrespondencia-elemzés demonstrálása érdekében önkéntes kérdőívvel támogatott megkérdezést<sup>1</sup> végeztem, melynek célja, annak feltárása, hogyan észlelik a fiatalok a különböző sörmárkákat és ezekhez az észlelésekhez a fiatalok milyen demográfiai jellemzői kapcsolódnak. A kutatás alapsokaságát a Miskolci Egyetem azon hallgatói képezték, akik havonta legalább 1 alkalommal fogyasztanak sört. A mintanagyság – a szűkös anyagi erőforrások következtében – 100 fő.

#### 3.2. Az input adatok

Az elemzéshez nominális skálán mért input adatokra van szükség, vagyis elő kell állítani egy olyan többváltozós keresztábrát, amely dimenziói a termékjellemzőket<sup>2</sup>, a sörmárkákat<sup>3</sup>, és a demográfiai ismérveket (Nem: Férfi, Nő; Évfolyam: I., II., III., IV., V.; Kar: ÁJK, BTK, GÉK, GTK, MAK, MFK; Lakhely: Albérlet, Kollégium, Szülőknél) tartalmazza. Az egyes cellákban pedig azon válaszadók száma szerepel, akik azonos demográfiai profillal rendelkeznek és egy adott termékjellemzőt igaznak vélnék egy adott sörmárkára. Fontos megjegyezni, hogy a keresztábra celláiban nem szerepelhet nulla, ellenkező esetben a többszörös korrespondencia-elemzés megbízhatósága gyengül.

#### 3.3. Az elemzés menete

Az elemzéshez az SPSS 15.0 statisztikai szoftvercsomag Categories modulját használtam, azon belül pedig a Leideni Egyetem adatok skálázási rendszereivel foglalkozó elméleti csoportja által kifejlesztett Multiple Correspondence 1.0 programot (Analyze/Data Reduction/Optimal Scaling). Az elemzés változóinak (Analysis Variables) a termékjellemzőket, a sörmárkákat és a demográfiai ismérveket választottam. Ezt követően meghatároztam a végső dimenziószámot (Dimensions in solution), amelynek érdemes 2-t vagy 3-at választani, ugyanis nagyobb dimenziószámokban gyakorlatilag lehetetlen az ábrázolás. A program további beállítási lehetőségeket nyújt, úgymint:

<sup>1</sup> A gyakorlati példa kiválasztásában nagy szerepet játszott a sörpiacon szerzett tapasztalatom, amelyet a Dreher Sörgyárak Rt. [6], valamint a Borsodi Sörgyár Rt. [1] részére készített marketingkutatások elemzőjeként szereztem az elmúlt években.

<sup>2</sup> A jellemzők a Dreher Sörgyárak Rt. részére készített kvantitatív kutatásból [6] származnak.

<sup>3</sup> A sörmárkák a Borsodi Sörgyár Rt. részére készített szakértői jelentésből [1] származnak.

- a változók súlyozása (Define Variable Weight...),
- újrakódolása (Discretize...),
- a hiányzó értékekkel rendelkező változók imputálása (Missing...),
- a kezdő konfiguráció kiválasztása (Options.../Configuration),
- az iteráció és a konvergencia kritériumok specifikációja (Options.../Criteria).

Mindezen felül lehetőségünk van az elemzés outputjainak testre szabására (Output...), azok elmentésére (Save...), valamint különböző ábrák, diagramok és grafikonok generálására (Object..., Variable...).

Annak érdekében, hogy a program előállítsa a változók közötti sztochasztikus kapcsolatot vizuális elemzéséhez szükséges többszörös korrespondencia térképet, meg kell jelölni a következő opciót: Joint Category Plot (Variable...) és ki kell választani valamennyi vizsgálatba vont változót. A program futtatásához és a térkép elkészítéséhez több beállítás nem szükséges.

### 3.4. Az eredmények értelmezése

Az eredmények értelmezését először a többváltozós keresztábra vizsgálatával kezdem, ezt követően az elemzés legfontosabb eredményeként előállt ún. többszörös korrespondencia-térkép (Lásd. 1. ábra) alapján levonható általános következtetéseket fogalmazom meg.

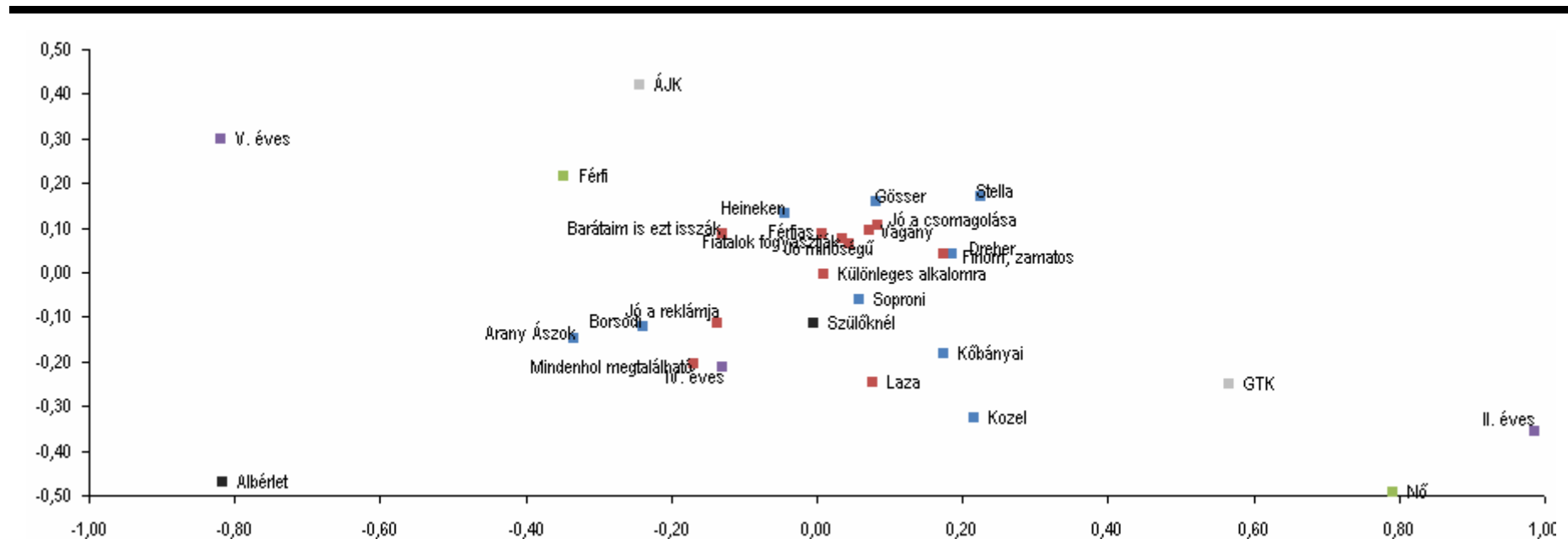
A keresztábra elemzésekor – mint bármely más sztochasztikus kapcsolat vizsgálatakor – arra keressük a választ, hogy van-e összefüggés a két – ez esetben minőségi – ismerv között. Abban az esetben, ha találunk szignifikáns kapcsolatot a két változó között, megvizsgáljuk, hogy milyen erős ez a kapcsolat. Az első kérdésünk megválaszolására nem paraméteres hipotézisvizsgálatot, ún. Pearson-féle  $\chi^2$ -próbát<sup>4</sup> kell végeznünk. A második kérdés megválaszolására több jó megoldás is kínálkozik, azonban a marketingkutatók gyakorlatában Cramer-féle V-mutató<sup>5</sup> terjedt el leginkább.

Jelen esetben egyértelmű a szoros összefüggés a termékjellemzők és sörmárkák, sörmárkák és évfolyam, nem és az évfolyam, nem és a kar, nem és a lakhely, évfolyam és a kar, évfolyam és a lakhely, valamint akar és a lakhely változópárok között, hiszen a  $\chi^2$ -próba kétoldali aszimptotikus szignifikancia értéke 0,05-nél kisebb. A kapcsolat erőssége azonban nem mondható erősnek, hiszen a Cramer-féle V-mutató értéke kisebb, mint 0,5.

---

<sup>4</sup>  $\chi_p^2 = \sum_{ij} \frac{(f_{ij} - E_{ij})^2}{E_{ij}}$ , ahol  $f_{ij}$  a megfigyelt,  $E_{ij}$  pedig a függetlenség esetén elvárt gyakoriság.

<sup>5</sup>  $V = \left( \frac{\chi_p^2}{W(q-1)} \right)^{1/2}$ , ahol W a főösszeg és  $q = \min\{R, C\}$ , ahol R a sorok, C pedig az oszlopok



1. ábra  
Többszörös korrespondencia-térkép<sup>6</sup>

A többszörös korrespondencia-térkép elemzéséből levonható legfontosabb tanulság, hogy a férfiakhoz a Heineken, a nőkhöz pedig a Kozel áll a legközelebb, annak ellenére, hogy a sörmárkák és a nem között nem mutatható ki szignifikáns összefüggés. Hasonló kijelentést tehetünk a sörmárkák és az évfolyam esetében, hiszen a II. évesekhez a Stella, míg az V. évesekhez az Arany Ászok áll közel és ebben az esetben a statisztikai összefüggés is fennáll. A többszörös korrespondencia térképről egyértelműen leolvashatók azok az összefüggések, amelyeket a többszörös keresztábra-elemzés még csak nem is érzékeltet, ezért a minőségi ismérvek közötti kapcsolat mélyebb vizsgálatára sokkal alkalmasabb módszer, mint a többszörös keresztábra-elemzés.

<sup>6</sup> Fontos megjegyezni, hogy több demográfiai kategória lekerült a térképről a jobb átláthatóság kedvéért.

#### 4. Következtetések

A többszörös keresztábla-elemzés korlátozottan alkalmas az asszociációs kapcsolat vizsgálatára, vagyis két minőségi ismerv közötti összefüggés feltárására, leírására. A módszer nagy hátránya, hogy mélyebb elemzések elvégzésére nem alkalmas. Példánkban szignifikáns, de gyenge kapcsolat van az a termékjellemzők és sörmárkák, valamint a sörmárkák és évfolyam változópaárok között.

A mélyebb elemzések elvégzésére sokkal alkalmasabb módszer a többszörös korrespondencia-elemzés, amely úgy alakítja át az elemzésbe vont többváltozós keresztábla sorait és oszlopait, hogy grafikusán ábrázolható legyen az alacsonyszámú többdimenziós térben. A gyakorlati példában több demográfiai kategória és sörmárka között érzékelhető összefüggés, amely a többszörös-korrespondencia térképről egyértelműen leolvasható.

#### Irodalomjegyzék

- [1] BERECKINÉ F. E. – HAVRILÓ A. – MOLNÁR L. – PISKÓTI I. – SCHUPLER H.: **A fiatalok életstílusának, alkohol- és sörfogyasztási szokásának kvalitatív vizsgálata – Szakértői jelentés a Borsodi Sörgyár Rt. részére**, Miskolc, 2007
- [2] BOLLA M. – KRÁMLI A.: **Statisztikai következtetések elmélete**, Budapest, Typotex Kiadó, 2005
- [3] FÜSTÖS L. – KOVÁCS E. – MESZÉNA GY. – SIMONNÉ M. N.: **Alakfelismerés – Sokváltozós statisztikai módszerek**, Budapest, Új Mandátum Könyvkiadó, 2004
- [4] GREENACRE, M. J.: **Theory and applications of correspondence analysis**, Academic Press, London, 1984
- [5] HAJDU O.: **Többváltozós statisztikai számítások**, Budapest, Központi Statisztikai Hivatal, 2003
- [6] HÓDOSI L. – KÓSA Á. – KULCSÁR I. – MOLNÁR L. – SZIVÓS J.: **Kvantitatív kutatás a Dreher Sörgyárak Rt. részére**, Miskolc, 2004
- [7] KETSKEMÉTY L. – IZSÓ L.: **Bevezetés az SPSS programrendszerbe**, Budapest, ELTE Eötvös Kiadó, 2005
- [8] MALHOTRA, N. K.: **Marketingkutatás**, Budapest, KJK-KERSZÖV Jogi és Üzleti Kiadó, 2002
- [9] MOLNÁR L.: **A korrespondencia-elemzés (CA) elmélete és gyakorlata**, Miskolc, microCAD 2008 International Scientific Conference, Miskolci Egyetem, Q szekció: Kihívások a gazdaságban, Konferenciakiadvány, 2008.
- [10] SPSS Inc.: **SPSS Categories**, Chicago, 2005
- [11] VERES Z. – HOFFMANN M. – KOZÁK Á.: **Bevezetés a piackutatásba**, Budapest, Akadémiai Kiadó, Budapest, 2006