

OTDK-DOLGOZAT

2015

**A HALLGATÓI PREFERENCIÁK ELEMZÉSE A GRAVITÁCIÓS ÉS
LOGIT MODELLEK SEGÍTSÉGÉVEL**

**ANALYSIS OF THE STUDENT'S PREFERENCES USING
GRAVITY-, AND LOGIT MODELLS**

Kézirat lezárva: 2014. május. 23.

TARTALOMJEGYZÉK

1.	BEVEZETÉS	1
2.	KUTATÁSI CÉLKITŰZÉS ÉS KUTATÁSI KÉRDÉSEK.....	2
3.	A TÉMÁVAL KAPCSOLATOS SZAKIRODALMI ÁTTEKINTÉS	4
3.1.	A STATISZTIKAI FELTÁRÁS	4
3.1.1.	<i>A statisztikai feltárás fázisai</i>	<i>4</i>
3.1.2.	<i>A statisztikai feltárás során alkalmazott modellek</i>	<i>6</i>
3.2.	A GRAVITÁCIÓS MODELL	8
3.2.1.	<i>A modell elméleti háttere.....</i>	<i>8</i>
3.2.2.	<i>A hallgatói döntéseket befolyásoló tényezők vizsgálatának területei</i>	<i>10</i>
3.3.	A LOGISZTIKUS REGRESSZIÓ	12
3.3.1.	<i>Bináris logit modell.....</i>	<i>12</i>
3.3.2.	<i>Multinomiális és ordinális logit modellek</i>	<i>13</i>
3.3.3.	<i>Logisztikus regresszió előnye és hátránya.....</i>	<i>15</i>
3.4.	A LOGIT MODELL ALKALMAZÁSA SORÁN VIZSGÁLT HALLGATÓI DÖNTÉST BEFOLYÁSOLÓ TÉNYEZŐK	16
3.5.	A DÖNTÉSI FÁK	18
4.	A KUTATÁS BEMUTATÁSA	20
4.1.	A VIZSGÁLT TERÜLET.....	20
4.2.	A VÁLTOZÓK ISMERTETÉSE.....	21
4.2.1.	<i>Független változók.....</i>	<i>21</i>
4.2.2.	<i>Függő változók</i>	<i>23</i>
4.3.	AZ ADATBÁZISOK FELÉPÍTÉSE.....	23
4.3.1.	<i>A logisztikus regresszióhoz felhasznált adatbázis felépítése</i>	<i>24</i>
4.3.2.	<i>Gravitációs modellhez felhasznált adatbázis felépítése.....</i>	<i>25</i>
5.	A KAPOTT EREDMÉNYEK ÉRTELMEZÉSE.....	27
5.1.	A MODELLBE BEKERÜLT ÚJ VÁLTOZÓK.....	27
5.1.1.	<i>Oktatói kiválóság alapján kapott rang</i>	<i>27</i>
5.1.2.	<i>Intézmény megyéjének foglalkoztatási rátája</i>	<i>28</i>
5.2.	A GRAVITÁCIÓS MODELL ALKALMAZÁSA SORÁN KAPOTT EREDMÉNYEK	29
5.2.1.	<i>Az egyes független változók hatásának bemutatása</i>	<i>30</i>
5.2.2.	<i>A gravitációs modellel kapott eredmények összefoglalása:.....</i>	<i>34</i>
5.3.	A BINÁRIS LOGIT MODELL ALKALMAZÁSA SORÁN KAPOTT EREDMÉNY ÉRTELMEZÉSE.....	35
5.3.1.	<i>Az egyes független változók hatásának bemutatása</i>	<i>37</i>
5.3.2.	<i>Tudományterületek és szakok eredményének összehasonlítása</i>	<i>39</i>
5.3.3.	<i>A bináris logit modell alkalmazása során levont főbb következtetések</i>	<i>41</i>
5.4.	AZ ORDINÁLIS (RANK ORDERED) LOGIT MODELL EREDMÉNYÉNEK ISMERTETÉSE	42
5.4.1.	<i>Az ordinális (Rank ordered) logit modell eredménye tudományterületekre vonatkozóan</i>	<i>45</i>

5.4.2.	<i>Az ordinális (Rank ordered) logit modell eredménye szakokra vonatkozóan</i>	46
5.4.3.	<i>A fontosabb megállapítások az ordinális logit modell eredményeinek interpretálása során</i>	48
6.	A GRAVITÁCIÓS ÉS LOGIT MODELLEK ÖSSZEHASONLÍTÁSA	49
6.1.	BUDAPEST KIEMELKEDŐ VONZEREJÉNEK MODELLRE GYAKOROLT HATÁSA:	51
6.2.	AZ ELTÉRŐ TUDOMÁNYTERÜLETEK PREFERENCIÁINAK KÜLÖNBÖZŐSÉGE	52
7.	ÖSSZEFOGLALÁS	55
	IRODALOMJEGYZÉK	58
	MELLÉKLETEK	I
	A LOGIT ÉS GRAVITÁCIÓS MODELLEK EREDMÉNYEINEK ÖSSZEFOGLALÓ ÁBRÁJA	II

TÁBLÁZAT- ÉS ÁBRAJEGYZÉK

1. táblázat: Az ordiális (Rank ordered) logit modell eredményének összefoglaló táblázata tudományterületekre vonatkozóan.....	43
2. táblázat: Az ordinális (Rank ordered) logit modell eredményének összefoglaló táblázata konkrét szakokra vonatkozóan	44
3. táblázat: A gravitációs és bináris logit modellek eredményeinek összefoglaló táblázata	53
1. ábra: A statisztikai feltárás fázisai	4
2. ábra: A statisztikai feltárás modelljeinek típusai	8
3. ábra: A hallgatói döntéseket befolyásoló vizsgált tényezők a gravitációs modell alkalmazása során	10
4. ábra: A logisztikus görbe.....	13
5. ábra: A hallgatói döntéseket befolyásoló vizsgált tényezők a logit modell alkalmazása során.....	16
6. ábra: Az elemzésekhez rendelkezésre álló független változók	21
7. ábra: A gravitációs modell eredménye	29
8. ábra: A bináris logit modell alkalmazása során kapott eredmény tudományterületekre, illetve szakokra vonatkozóan (saját szerkesztés).....	36
9. ábra: A gravitációs és bináris logit modellek eredményeinek összehasonlítása	50

1. BEVEZETÉS

Napjainkban a fogyasztók preferenciáinak megismerése igencsak fontos szerepet játszik a termelő és szolgáltató vállalatok üzleti sikerében. A fogyasztók döntéseire vonatkozó kutatások módszertana folyamatosan fejlődik, a vizsgálatok is egyre szélesebb körben zajlanak.

Módszertani fejlődést azonban nem csak e területen tapasztalhatunk. A felsőoktatási intézményekben jelentkezők preferenciáinak elemzésével kapcsolatban szintén tekintélyes színvonalú kutatómunkák születnek. A hazai kutatások között is találkozunk olyannal, amely a felsőoktatási intézményekben történő jelentkezések alapján állít fel intézményi preferencia sorrendet (Telcs, Kosztyán, Török, 2013).

Munkám során a magyarországi felsőoktatásba jelentkezők preferenciáit elemzem két nagy tudományterületre és ezeken belül egy-egy konkrét szakra vonatkozóan a gravitációs és logit modellek segítségével, melyeket a korábbi kutatásokhoz képest új változókkal egészíték ki.

Dolgozatom első részében bemutatom a témával kapcsolatos hazai és idegen nyelvű szakirodalmat, valamint irodalomkutatást végzek a gravitációs és logit modellek alkalmazására vonatkozóan. Ennek során azt vizsgálom, hogy az egyes modellek esetén melyek azok a független változók, amelyeknek a felsőoktatási intézményekben jelentkezők preferenciáira gyakorolt hatását már vizsgálták a korábbi kutatások alkalmával.

Ezek után bemutatom a vizsgálatok során elemezni kívánt függő és független változókat, illetve a szükséges adatbázisok szerkezetét és felépítésének menetét. A kapott eredményeket külön interpretálom a gravitációs, illetve (bináris) logit modellekre vonatkozóan, továbbá megvizsgálom azt is, hogy a jelentkezők preferenciái hogyan alakulnak attól függően, hogy hányadik helyre tartozó jelentkezésről döntenek. Ehhez ordinális (rank ordered) logit modellt alkalmazok.

Az egyes módszereket összehasonlítom a kapott eredmények alapján, végül pedig összefoglalom a különböző modellek segítségével kapott főbb eredményeket, megállapításokat és további kutatási irányvonalat adok meg.

Az első fejezet a kutatási célkitűzéseket, kutatási kérdéseket és az elemzések során alkalmazandó módszereket mutatja be.

2. KUTATÁSI CÉLKITŰZÉS ÉS KUTATÁSI KÉRDÉSEK

Ahogy azt a korábbiakban is említettem, a felsőoktatási intézményekbe jelentkezők preferenciáinak elemzése igen jelentős terület a statisztikában. A vizsgált területre vonatkozóan tekintélyes színvonalú kutatómunkák széles palettáját olvashatjuk a szakirodalomban.

Jelen kutatás célja a magyarországi felsőoktatási intézményekbe jelentkezők preferenciájának megismerése a korábbi szakirodalmakban található szempontok kiegészítésével. Munkám további célja a preferenciák olyan módszerekkel történő vizsgálata, amely Magyarországon újszerűnek mondható.

A jelentkezők döntését befolyásoló tényezők vizsgálata több érdekes kérdést is felvet, melyek közül dolgozatomban elsősorban a következőkre keresem a választ:

1. Melyek a jelentkezők döntését legnagyobb mértékben befolyásoló szempontok?
2. Budapest kiemelkedő gazdasági vonzereje gyakorol-e hatást az egyes intézmények megítélésére a jelentkezők döntésekor?
3. Mutatkoznak-e különbségek az eltérő tudományterületeken továbbtanulni kívánók preferenciájában a felsőoktatási intézmény(ek) megjelölésekor?
4. Befolyásolja-e döntésének szempontjait, hogy a jelentkező hányadik helyre vonatkozó megjelölésről dönt a jelentkezéskor (abban az esetben, ha több intézményt jelöl meg)?

A kutatás során alkalmazott módszerek:

A kutatási kérdések megválaszolásához több módszert is alkalmazni fogok. Először is szükségem lesz az adatok rendelkezésre állítására, adatbázisok felépítésére, melyeken a szükséges vizsgálatokat elvégezhetem.

Munkám során a következő módszereket fogom alkalmazni:

- Gravitációs modell
- Bináris logisztikus regresszió
- Ordinális logisztikus regresszió (rank ordered logit)
- Döntési fa módszer

A gravitációs modell alkalmazása a szakirodalomban sok esetben fellelhető térbeli pontok közötti áramlások modellezésére. Dolgozatomban én ezt a módszert a magyaror-

szági kistérségek és felsőoktatási intézmények közötti hallgatói áramlás vizsgálatára fogom alkalmazni, új szempontokkal kiegészítve.

A bináris és ordinális logisztikus regresszió esetén azt fogom vizsgálni, hogy mely tényezők és hogyan befolyásolják azt, hogy a jelentkezők hányadik helyen jelölnek meg egy adott intézményt. Az ordinális logisztikus regresszióra azért lesz szükségem, hogy a negyedik kutatási kérdést megválaszolhassam. Segítségével megvizsgálom, hogy az első helyen történő megjelöléstől a sorban magasabb helyek felé haladva milyen megállapításokat tehetek a független változókra kapott eredmények tekintetében. A magyarországi elemzésekben ezen a területen a szóban forgó kérdés ordinális regresszióval történő vizsgálata újszerűnek mondható.

A logisztikus regresszió alkalmazása mellett a független változók erőssorrendjét döntési fa módszerrel fogom meghatározni. A döntési fák alkalmazására találunk példát a szakirodalomban e területre vonatkozóan, de magyar viszonylatban e konkrét területen történő alkalmazása újszerű.

Kutatási célkitűzés teljesítéséhez, valamint kutatási kérdéseim megválaszolásához a fentiekben felsorolt módszerek alkalmazására mind szükségem van munkám során. A következő nagy fejezet a témával kapcsolatos szakirodalom és módszertan bemutatásával foglalkozik, melyben sorra ismertetem az említett módszereket.

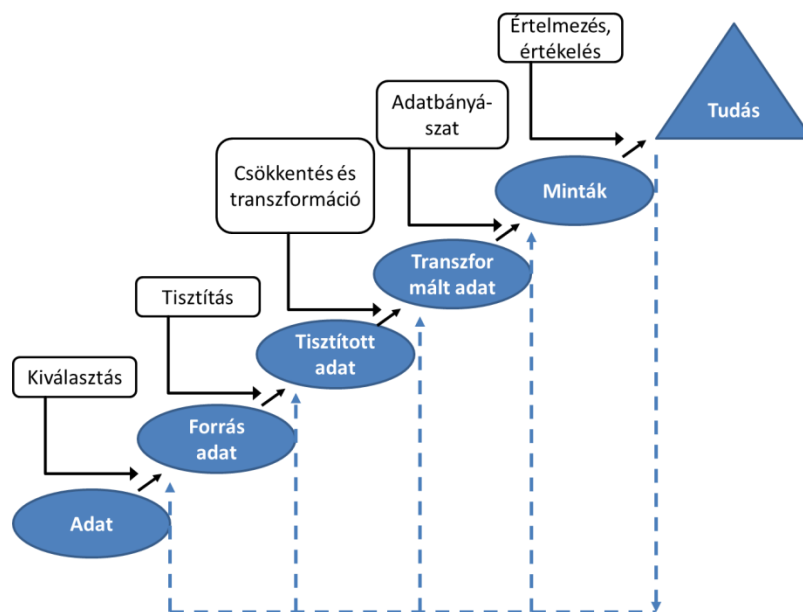
3. A TÉMÁVAL KAPCSOLATOS SZAKIRODALMI ÁTTEKINTÉS

3.1. A STATISZTIKAI FELTÁRÁS

A gravitációs modell felírásához megfelelő nagyságú adathalmazra van szükség. Az információk rendelkezésre állítása gyakran nem kis feladat, hiszen meg kell terveznünk az információk jellegét, hogy milyen változókra lesz szükségünk, esetlegesen azt is, hogy az egyes adatokból, hogyan számíthatjuk, származtathatjuk a számunkra fontos információkat. Ehhez olyan eszközre van szükségünk, amely képes arra, hogy a nyers adatokból információt tudjunk kinyerni.

3.1.1. A statisztikai feltárás fázisai

Téves lenne azt feltételezni, hogy a statisztikai feltárás az egyes adatok, adathalmazok rendelkezésre állását jelenti. A folyamat során a látens, számunkra hasznos tudást, illetve összefüggéseket nyerjük ki egy adathalmazból. A tudás feltárásának folyamatában több szakaszt különböztethetünk meg, ezeket az 1. ábra mutatja be.



1. ábra: A statisztikai feltárás fázisai

Forrás: Fayyad, U M., Shapiro, P. G., Smyth, P, 1996, Han, J., Kamber, M., 2006

Vegyük sorra az ábrán szereplő szakaszokat!

Kiválasztás:

Az első fázis az alkalmazás területének illetve a későbbi céloknak meghatározása. A továbbiakban egy kiválasztott adatbázist használunk fel, melyet az eredetileg meghatározott halmazból választunk. Így egy céladatbázishoz jutunk.

Tisztítás, csökkentés és transzformáció:

Következő lépésként a céladatbázist fogjuk tovább finomítani. Ennek során kijavítjuk a benne szereplő esetleges hibákat. Ilyen lehet például, hogy egy számérték helyett szöveges érték szerepel az adatok között. Ez többek közt azért is fontos lehet, mert egyes programok nem tudják kezelni a szöveg adattípusú cellákat. Ilyenkor ezeket a mezőket kódszámmal szokás ellátni.

A feldolgozás szempontjából fontos lépés az adatok integrációja, az esetlegesen felosztott adathalmazok egyesítése. Előfordulhat, hogy több adathalmaz áll egy időben rendelkezésünkre, ebben az esetben szükséges az adatok egy összesített halmazba történő integrálása, melyből a továbbiakban meghatározhatjuk a szükséges attribútumokat, tovább finomítva ezáltal a felhasználni kívánt információk halmazát.

Adatbányászat:

Ekkor már rendelkezésünkre áll egy olyan adatbázis, amelyből kinyerhetjük a szükséges információkat. Ehhez azonban először meg kell határoznunk, hogy milyen módszerek (Klaszterezés, Diszkriminancia elemzés, Regresszió analízis, stb...) alkalmazhatók erre a célra. Figyelembe kell vennünk a módszerek alkalmazhatóságát korlátozó tényezőket, (pl.: Regresszió elemzésnél heteroszkedaszticitás megléte, normalitás feltétele, stb...), illetve az egyes módszerek előnyeit, hátrányait.

Értelmezés, értékelés:

A kiválasztott algoritmus futtatása után a kinyert információkat értelmezni kell. Ezeket célszerű összevetni a korábbi elvárásokkal, továbbá magyarázatot keresni a várakozásokhoz viszonyított esetleges eltérésekre. Itt azonban meg kell jegyeznünk, hogy előfordulhat az is, hogy a kapott információk nehezen értelmezhetőek. Ez abból is következhet, hogy torzítás jelentkezik a felhasznált adatokat illetően. Ez esetben szükséges lehet visszanyúlnunk korábbi lépésekhez és az adatokat tovább kell tisztítanunk, finomítanunk. (Bodon F., 2010)

A következőkben vizsgáljuk meg a statisztikai feltárás során alkalmazott modelleket részletesen is!

3.1.2. A statisztikai feltárás során alkalmazott modellek

Az statisztikai feltárás során alapvetően két modellt különböztethetünk meg, az előrejelző, illetve leíró modelleket. Az előrejelző modellek esetében bemeneti adatokat alkalmaznak egy olyan modell létrehozására, amelyben a következmények ismertek. Ezután ezt a modellt alkalmazzák olyan bemeneti adatokra vonatkozóan, ahol a következmények ismeretlenek.

A leíró modellek mintákat, illetve kapcsolatokat írnak le a vizsgálatnak alávetett adatokban. Az így kapott eredményeket különböző döntéseknél használhatjuk fel.

E két alapvető modelltípuson belül is különíthetünk el kategóriákat a megoldandó feladat típusa alapján. Ez utóbbi szempontjából hat modell típust különböztethetünk meg:

- Osztályozás
- Regresszió számítás
- Idősor elemzés
- Klaszterezés
- Kapcsolatelemzés
- Sorozatkeresés

Nézzük e modellek részletesebb ismertetését is!

Osztályozás:

Osztályozás során csoportosítjuk az esetleírások azon jellemzőit, amelyek azt határozzák meg, hogy az eset melyik előre definiált csoportba tartozik. A modellben a csoportok származtatására több lehetőség is van. Egyrészt alkalmazhatók történeti adatbázisok a csoportok származtatására, másrészt szakértői segítség vonható be. Ennek során a szakértő osztályozza az adatbázis egy részét, majd az így meghatározott elven tovább haladva az osztályozást kiterjesztjük az adatbázis egészére. Harmadik lehetőségként pedig kísérlet alapján is felépíthetjük a modellt.

Regresszió számítás:

A regresszió számítás egy régóta alkalmazott statisztikai eljárás. Gyakori problémát jelenthet azonban, hogy a feltárni kívánt összefüggések nem lineárisak. Ez a probléma jelenik meg (amint azt a dolgozat későbbi fejezeteiben látni fogjuk) a közgazdasági elemzések során felállított gravitációs modell esetében is. Ilyen esetekben szükséges transzformációt végrehajtanunk és linearizálnunk kell az összefüggéseket.

Idősor elemzés:

Az idősor elemzés szintén a statisztika egyik meghatározó területe. Alapvetően beszélhetünk determinisztikus, illetve sztochasztikus idősor elemzésről. Az idősorok fő komponensei az alapirányzat (trend), szezonális hatások, véletlenszerű hatások és a ciklikusság.

Jelen kutatásban ugyan az idősor elemzéssel nem foglalkozok, de a későbbiekben érdekes lehet megvizsgálni a hallgatói preferenciák alakulásának összevetését akár több időszak összevetésével is.

Klaszterezés:

Az eljárás az adatbázisban szereplő adatok csoportosítását végzi, mégpedig úgy, hogy a meghatározott csoportok különbözzenek egymástól, az egyes csoportok elemei között pedig minél nagyobb mértékű hasonlóság álljon fenn. A klaszterezés során fellépő probléma lehet (a szubjektivitásból kifolyólag), hogy különböző távolságfüggvények alkalmazásával különböző eredményeket kaphatunk. Ennek kiküszöbölésére egy lehetséges megoldás lehet egy, a vizsgált terület szakértőjének bevonása. Segítségével meghatározható, hogy melyik eredmény áll legközelebb a gyakorlati megfigyelésekhez.

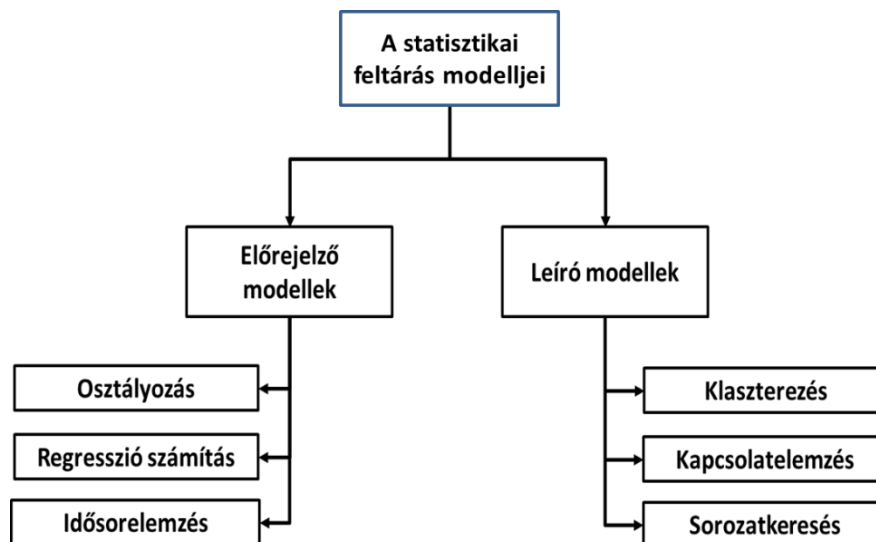
Kapcsolatelemzés:

A kapcsolatelemzés során olyan adattételeket keresünk, amelyek egy időben vannak jelen egy eseményben vagy az adatbázis egy rekordjában.

Sorozatkeresés:

Ez az eljárás a kapcsolatelemzéshez hasonló, azzal a különbséggel, hogy az egyes adatok nem egy időben vannak jelen, hanem különböző időpontokból származnak (Sramó A., 1999).

Míg az osztályozás, regresszió elemzés és idősor elemzés az előrejelző modellek esetében, a klaszterezés, kapcsolatelemzés és sorozatkeresés főként a leíró modellek esetében használatos. A modellek típusait a 2. ábra szemlélteti.



2. ábra: A statisztikai feltárás modelljeinek típusai

Forrás: Saját szerkesztés Sramó A., 1999, pp. 350-359 alapján

Ebben az alfejezetben bemutatam a statisztikai feltárás során alkalmazott módszereket. A következő részben ismertetem a gravitációs modellt.

3.2. A GRAVITÁCIÓS MODELL

A matematikai statisztikában széles körben alkalmazott módszer a gravitációs modell. Segítségével modellezhetők az emberi viselkedés által indukált áramlások, melyek vonatkozhatnak több tényezőre is, mint például tőke, információ, vagy akár a migráció.

3.2.1. A modell elméleti háttere

A gravitációs modell a húszas-harmincas évek fordulóján jelent meg az Amerikai Egyesült Államokban. Ekkor még nagyvárosok vonzáskörzetének valamilyen szempont alapján történő lehatárolására használták. Az ötvenes-hatvanas évek fordulópontján a gravitációs modellt már nagyvárosokon belül történő mozgások elemzésére alkalmazzák. Az igazi áttörésekre a hetvenes évektől kerül sor, ekkor születtek meg a ma is nagy jelentőséggel bíró összefoglaló munkák (Nagy G., 1996).

A gravitációs modell a fizikából megismert általános tömegvonzás törvényén alapszik, mely kimondja, hogy két test között fellépő vonzóerő egyenesen arányos a két test tömegével (m_1 , m_2) és fordítottan arányos a köztük lévő távolság négyzetével (d):

$$F = G \frac{m_1 m_2}{d^2} \quad (1)$$

Az ehhez hasonló közgazdaságtani gravitációs modell legegyszerűbb formáját pedig a következőképpen írhatjuk fel:

$$D_{ij} = g \left(\frac{P_i P_j}{d_{ij}^2} \right) \quad (2)$$

Ahol D_{ij} az i és j pontok közötti „népességi erő”, P_i és P_j az i és j pontokhoz tartozó népesség, d_{ij} pedig az i és j pontok távolsága (Stewart, 1948).

A gravitációs modell ennél bonyolultabb, Bramhall által megfogalmazott leírásával a következőképpen módosul az összefüggés:

$$N_{ij} = g \frac{(w_i P_i^\alpha w_j P_j^\beta)}{d_{ij}^\gamma} \quad (3)$$

Ahol N_{ij} a várt áramlás i pontból j pontba adott időegység alatt, w_i és w_j súlysza-
mok, d az i és j pontok közötti távolság, α , β , γ pedig paraméterek, melyek meghatáro-
zása más módszerekkel történhet (Dusek T., 2003).

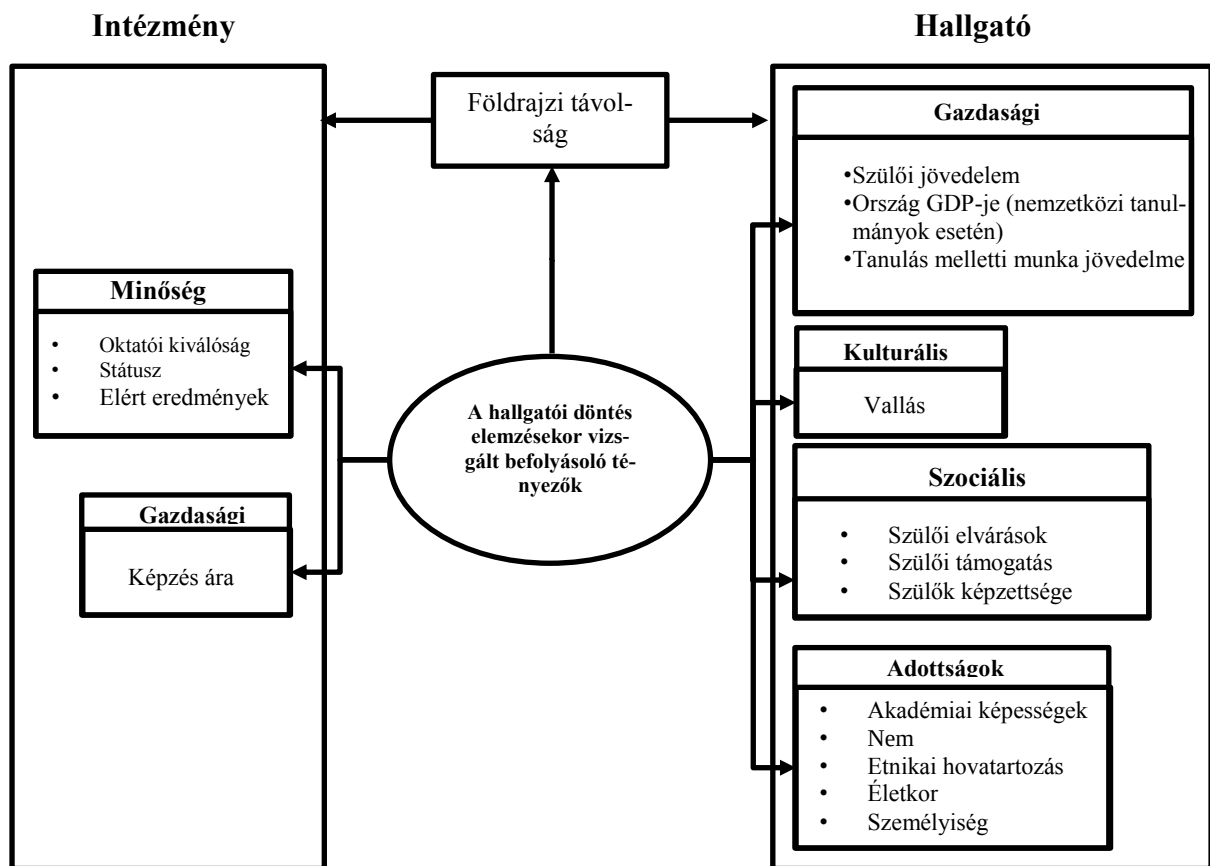
A modell alkalmazásának területét alapvetően két nagy csoportra oszthatjuk. Egy-
részt alkalmazzák a modellt tömegpontok közötti áramlások intenzitásának becslésére,
másrészt tömegpontok vonzáskörzetének meghatározására. Számos kutatás foglalkozik
a felsőoktatásban megjelenő hallgatói döntések vizsgálatával. A gravitációs modell ezen
a területen is alkalmazható (Neumanné V. I., 2010).

A következőkben ismertetem irodalomkutatásom eredményét, mely azt vizsgálja,
hogyan az egyes kutatómunkákban melyek azok a vizsgált befolyásoló tényezők, melyek
hatását összefüggésbe hozzák az adott felsőoktatási intézmény megítélésével és kivá-
lasztásával.

3.2.2. A hallgatói döntéseket befolyásoló tényezők vizsgálatának területei

Az egyes felsőoktatási intézmények hallgatók általi megítélését és kiválasztását több tényező is együttesen befolyásolhatja. Az irodalomkutatás során azt vizsgáltam, hogy az egyes kutatások során mely befolyásoló tényezőket veszik figyelembe az összefüggések vizsgálatakor.

Irodalomkutatásom eredményét az 3.ábra mutatja be:



3. ábra: A hallgatói döntéseket befolyásoló vizsgált tényezők a gravitációs modell alkalmazása során

Forrás: Saját szerkesztés Joachim Z, Prathibha J, 2005, Ahmet G., Büsra K., et. al., 2011, Saisana, M., d'Hombres, B., et. al., 2011, Leung, A., Hou, Z., et al., 2011, Larose, S. Cyrenne, D., et. al., 2009, Horstschräer, J., 2012, Asari, F. F. A. H, Idris, et al., 2011, Germeijs, V., Verschueren, K., 2007, Vrontis, D., Alkis Thrassou, A., et al., 2007, Gibbons, S., Vignoles, A., 2012, Toutkoushian, R. K., 2001, Bruno, G., Improta, G., 2008, Alm, J., John V. et al., 2009 alapján.

A már említett területtel foglalkozó munkák alapvetően két megközelítésből, a hallgató, valamint az intézmény megközelítéséből végzik elemzéseiket. Elsőként vegyük sorra a hallgatói oldalon fellépő, a döntést befolyásoló tényezőket!

3.2.2.1. A vizsgált befolyásoló tényezők a jelentkező oldaláról

E vizsgált tényezők is több csoportba oszthatók. Számos kutatás irányul a jelentkezők gazdasági hátterére, hiszen a döntéshozatalkor fontos szempont lehet egy intézmény kiválasztásakor, hogy a döntéshozó milyen anyagi háttérrel rendelkezik, mivel ez lehetőségeit is korlátozhatja. A gazdasági háttér vizsgálatára irányuló munkák a szülői keresetet, és a hallgató jövedelmét veszik figyelembe. Utóbbit értelemszerűen abban az esetben, ha a tanulmányok mellett végez munkát az illető. Ezen kívül nemzetközi tanulmányokra is irányulnak kutatómunkák, melyek az adott ország GDP-jének és a hallgatók döntéseinek összefüggéseit elemzik.

További döntést befolyásoló tényezőként jelenik meg a kulturális szempont, ezen belül pedig a vallás, ami főként a távol-keleti országokban játszik fontos szerepet.

A szociális háttér esetében a szülők képzettsége, a szülők elvárásai és az általuk nyújtott támogatás került a figyelem középpontjába. Ezek komoly hatást fejthetnek ki egy leendő hallgató esetében a felsőoktatási intézmény kiválasztásakor. A negyedik csoportba a jelentkező adottságai, illetve tulajdonságai tartoznak. Ilyenek az életkor, az etnikai hovatartozás, az akadémiai képességek, a nem, illetve a személyiségi jellemzők. A következőkben bemutatom, melyek azok a tényezők, amelyek a döntést intézményi oldalról befolyásolhatják.

3.2.2.2. A hallgatói döntést intézményi oldalról befolyásoló tényezők

A meglévő kutatásokban megjelenik az intézmények kiválasztásakor azok megítélésének elemzése is. A munkákban főként a minőség és a gazdasági szempontok kerülnek a vizsgálódások középpontjába. Minőség szempontjából az oktatói kiválósággal, az intézmény által képviselt státusszal, valamint az elért eredményekkel (pl.: minőségi díjak), míg a gazdasági szempontból az intézmény által indukált árakkal, mint például a tandíj foglalkoztak.

Végül pedig essen szó egy olyan szempontról, amely nem sorolható egyoldalúan a hallgatói vagy az intézményi oldalra, hiszen egyaránt kapcsolható hozható mindkettő megközelítéssel! Ez a földrajzi távolság. Számos kutatás foglalkozik azzal a kérdéssel, hogy az egyes jelentkezők lakóhelye és a potenciális felsőoktatási intézmények között fennálló földrajzi távolság hogyan befolyásolja a döntésüket az intézmény kiválasztásakor.

Ahogy azt az 3. ábra is mutatja, a gravitációs modellel végzett kutatások terén a felsőoktatásban nagyobb hangsúllyal szerepelnek olyan munkák, melyek jelentkezői oldalról végzik főként a vizsgálatokat. Az ábrán azt is észrevehetjük, hogy nem szerepelnek olyan változók, amelyek az intézmény környezetét jellemeznék. A döntéskor ezek is szerepet játszhatnak egy intézmény megítélésében. Érdeemes lehet megvizsgálni, hogy az ilyen változók milyen hatással lehetnek a jelentkezők döntésére vonatkozóan.

A gravitációs modell mellett számos kutatás elemzi a hallgatói preferenciákat a logisztikus regresszió segítségével. A következő fejezetben bemutatom a logisztikus regressziót, illetve szintén bemutatom a módszer alkalmazása során eddig vizsgált változókat a jelentkezők preferenciájára vonatkozóan.

3.3. A LOGISZTIKUS REGRESSZIÓ

A lineáris regresszió esetén magas mérési szintű változók magyarázatát végezzük. Ez az elemzési módszer azonban csak a lineáris kapcsolatok erősségének és irányának leírására alkalmas. Előfordulhat olyan eset, hogy vizsgált függő változó dichotóm vagy kategorikus. Ebben az esetben logisztikus regressziót (más néven logit) alkalmazunk a kapcsolatok irányának és erősségének meghatározására.

Attól függően, hogy a magyarázott változónk dichotóm vagy kategorikus, megkülönböztetünk bináris, illetve multinomiális logisztikus regressziót.

3.3.1. Bináris logit modell

A lineáris regresszió esetén a függő változónk magas mérési szintű. Előfordulhat azonban olyan eset is, mikor a függő változó dichotóm jellegű (egy adott esemény bekövetkezik-e vagy sem, például egy jelentkező megjelöl-e egy adott felsőoktatási intézményt vagy sem). Legyen adott k független változó: $X_1, X_2, \dots, X_k$, az ezekhez tartozó $b_0, b_1, \dots, b_k$ együtthatók, valamint Y célváltozó, ami esetünkben két értéket vesz fel, 1-et, ha az esemény bekövetkezik és 0-át, ha nem következik be.

Ekkor a független folytonos változók segítségével akarjuk megbecsülni Y -t vagyis az esemény bekövetkezését. Ehhez pedig a független változóinknak megkeressük azt a

$$Z = B_0 + B_1 \cdot X_1 + B_2 \cdot X_2 + \dots + B_k \cdot X_k \quad (4)$$

lineáris összefüggését, amely exponenciális függvényével a legjobban tudjuk közelíteni a vizsgált esemény „esélyét”. Vagyis

$$\frac{P(Y=1)}{P(Y=0)} = \frac{P(A)}{1-P(A)} \approx e^Z \quad (5)$$

Az egyes eseményekre kifejezve a következő egyenletek írhatók fel:

$$P(Y = 1) = P(A) \approx \frac{1}{1+e^{-Z}} \quad (6)$$

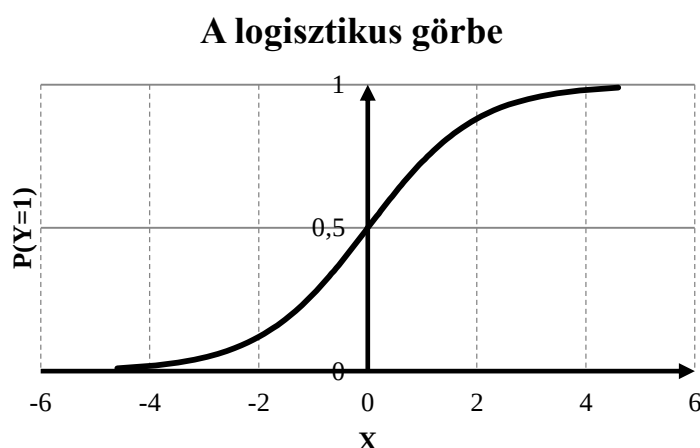
$$P(Y = 0) = 1 - P(A) \approx \frac{1}{1+e^Z} \quad (7)$$

Ennek megoldásához a legnagyobb valószínűség elvét alkalmazzuk. Az $Y, X_1, X_2, \dots, X_n$ változókra n elemű $(\alpha_1, X_{11}, X_{21}, \dots, X_{k1}), \dots (\alpha_n, X_{1n}, X_{2n}, \dots, X_{kn})$ minta alapján a következő ún. loglikelihood függvény értékét maximalizáljuk.

$$L = (\alpha_1, \alpha_2, \dots, \alpha_n) = \ln(P(\alpha_1, Y_2 = \alpha_2, \dots, Y_n = \alpha_n)) \approx \sum_{i=1}^n \ln\left(\frac{1}{1+e^{(-1)^{\varepsilon_i} B_0 + B_1 \cdot X_{1i} + B_2 \cdot X_{2i} + \dots + B_k \cdot X_{ki}}}\right) \quad (8)$$

A megoldásban szereplő $B_0^*, B_1^*, \dots, B_k^*$ együtthatók maximalizálják a fenti összeget, és ezeket alkalmazzuk az Y -ra vonatkozó előrejelzésnél.

A 4. ábra mutatja be annak valószínűségét, ha az független változóhoz tartozó függő érték éppen 1.



4. ábra: A logisztikus görbe

Forrás: Kecskeméty L., Izsó L., 2009

3.3.2. Multinomiális és ordinális logit modellek

Az előzőekben bemutattam a bináris logisztikus regressziót, amely esetén a függő változó dichotóm jellegű volt. A multinomiális logisztikus regresszió ennek kiterjesztése, mikor a függő változó kategoriális és kettőnél több értéket is felvehet.

A multinomiális logisztikus regresszió visszavezethető $L-1$ számú bináris logisztikus modellre ($L \geq 2$ a függő változó lehetséges értékeinek a száma). Az alapmodellt $L=3$ esetén keresztül mutatom be. Például tegyük fel, hogy azt vizsgáljuk, hogy egy adott jelentkező egy felsőoktatási intézménybe első, második vagy harmadik helyen jelentkezik-e.

Ekkor a modell a következő három egyenletből áll:

$$P(Y = 1) = \frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3}} \quad (9)$$

$$P(Y = 2) = \frac{e^{z_2}}{e^{z_1} + e^{z_2} + e^{z_3}} \quad (10)$$

$$P(Y = 3) = \frac{e^{z_3}}{e^{z_1} + e^{z_2} + e^{z_3}} \quad (11)$$

Ezzel az egyenletrendszerrel a probléma az, hogy ha minden z értékhez hozzáadunk egy C konstanst, akkor a valószínűségek változatlanok maradnak, ami azt jelenti, hogy különböző z értékekhez ugyanaz a valószínűség tartozna. Ezt úgy hidalhatjuk át, hogy az egyik kiválasztott z értéket 0 szinten rögzítjük. legyen esetünkben ez $z_3=0$. tehát az $Y=3$ értéket ún. referencia kategóriának jelöltük ki.

Ezzel pedig két esély értéket is definiáltunk:

$$esély_{13} = \frac{p_1}{1-p_3} \quad (12)$$

$$esély_{23} = \frac{p_2}{1-p_3} \quad (13)$$

Ha ezt az $L=2$ esetben végezzük el, akkor visszakapjuk a bináris logisztikus regresszió modelljét:

$$P(Y = 1) = \frac{e^{z_1}}{e^{z_1} + e^{z_2}} \quad (14)$$

$$P(Y = 2) = \frac{e^{z_2}}{e^{z_1} + e^{z_2}} \quad (15)$$

Legyen $Y=1$ a referencia kategória, és $Y=2$ pedig feleljen meg az A következmény bekövetkezésének. Így z_2 indexét elhagyva a következő egyenlethez jutunk:

$$P(Y = 2) = P(Y = A) = p_A = \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}} \quad (16)$$

A bináris esetben a referencia a 0 értékű kategória és ehhez képest becsüljük az 1 értékű eset (esemény bekövetkezése) előfordulásának valószínűségét. A multinomiális logisztikus regressziónál a referencia kategória a függő változó bármely értékéhez tartozó kategória lehet. Ekkor a függő változó összes kategóriáját összehasonlítjuk a kiválasztott referencia kategóriával minden egyes független változóra (Kecskeméty, Izsó, et al. 2009).

Ahogy azt a későbbiekben majd láthatjuk, ezt fogom alkalmazni a jelentkezők preferenciáinak elemzésénél, ahol azt vizsgálom majd, hogy mely tényezők befolyásolják, hogy egy adott jelentkező egy adott intézmény hányadik helyen jelöl meg.

Olyan eset is előfordulhat, hogy sorrendet kívánunk felállítani a függő változó lehetséges felvehető értékei között. Ekkor alkalmazható az úgynevezett ordinális logit modell. Tegyük fel, hogy a véletlen hatások $(\varepsilon_{i,j})$ függetlenek, vagyis:

$$U_{i,j} = V_{i,j} + \varepsilon_{i,j} \quad (17)$$

Ha egy adott válaszadó a legpreferáltabb lehetőséget választja J alternatíva közül, akkor $Y_{ij}=1$ jelöli azt, hogy az i -edik válaszadó a j -edik alternatívát preferálja legjobban, tehát az i -edik válaszadó szempontjából a j -edik alternatíva utilitása nagyobb, mint a többi alternatíváé:

$$U_{ij} \geq \max\{U_{i1}, \dots, U_{ij}\} \quad (18)$$

A valószínűséget pedig a következő egyenlettel adhatjuk meg:

$$P(Y_{ij} = 1; B) = P(U_{ij} \geq \max\{U_{i1}, \dots, U_{ij}\}) = \frac{e^{V_{ij}}}{\sum_{l=1}^J e^{V_{il}}} \quad (19)$$

(Dijk, B., et al., 2012)

A logisztikus regresszió sok tulajdonságában különbözik a lineáris regressziótól. A következőkben bemutatom milyen előnyei, illetve hátrányai vannak az alkalmazásának.

3.3.3. Logisztikus regresszió előnye és hátránya

A logisztikus regresszió egyik előnye, hogy a diszkriminancia-analízisnél szélesebb körben alkalmazható. Ennek oka, hogy úgy végzi el a függő változó magyarázatát, hogy nem kívánja meg a független változók struktúrába rendeződését. További előnyt jelent, hogy nem kívánja meg a heteroszkedaszticitás teljesülését, illetve a változók eloszlása sem kötött. Ezzel szemben hátrányként jelenhet meg, hogy nehézkes a modell magyarázóerejének értelmezése. A lineáris regresszió esetén az R^2 érték mutatja számunkra, hogy a független változók hányad részét magyarázzák a függő változó varianciájának. Logisztikus regressziónál a gond az, hogy a függő változó varianciája (ezzel együtt a logisztikus regresszió R^2 mutatója) függ magának a változónak az eloszlásától.

Adódnak módszerek, melyek segítségével értéket kaphatunk a modell magyarázóerejére vonatkozóan. Ilyen például a Cox-Snell,¹ valamint a Nagelkerke-féle R^2 mutató. Ezeken kívül használatos az R^2 értékének egy módosított változata is a magyarázóerő mérésére. Ezt R_L^2 -el jelöljük és a következőképpen számítjuk ki:

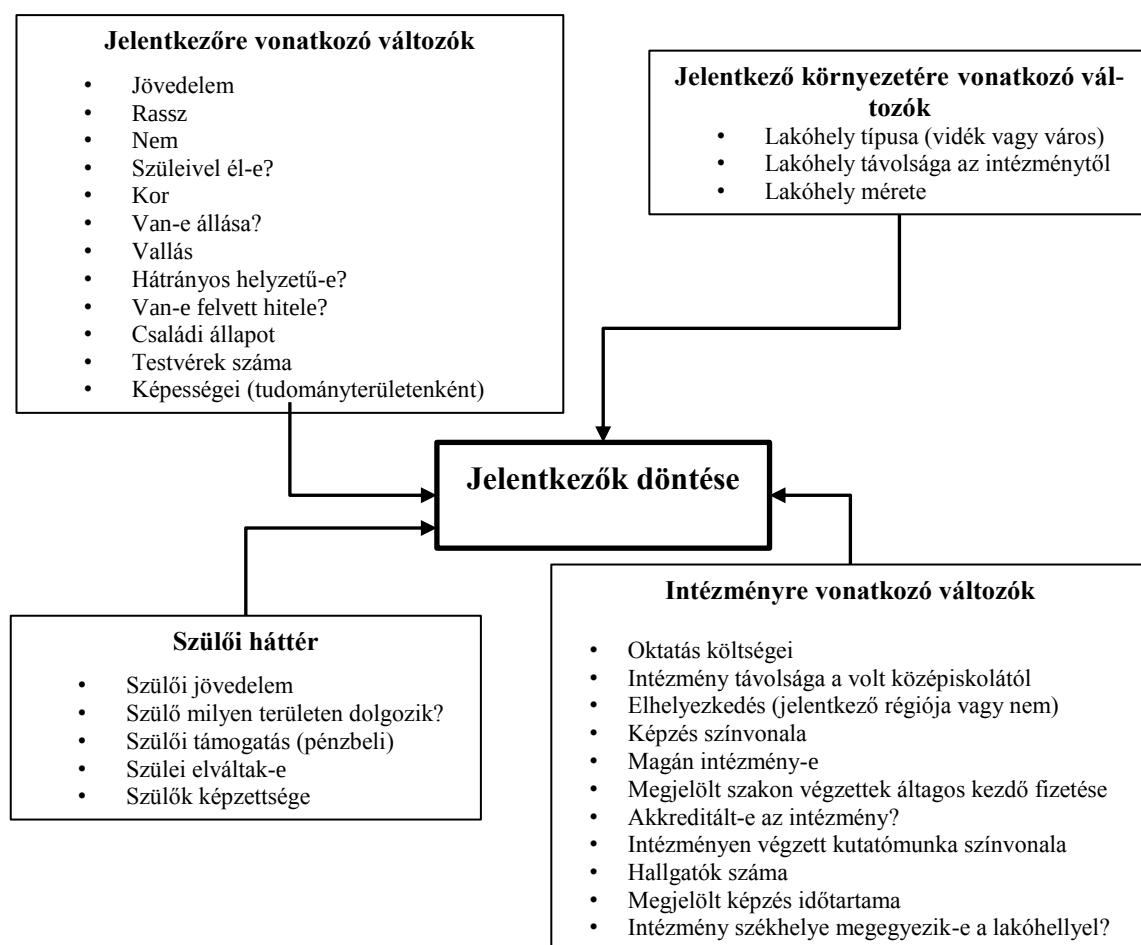
$$R_L^2 = \frac{D_0 - D_M}{D_0} \quad (20)$$

¹ A Cox-Snell-féle mutató a nulladik és az aktuális modell log likelihoodjait hasonlítja össze. Ezzel a probléma az, hogy maximális értéke soha nem éri el az egyet, ezért pedig nem lehet pontosan értelmezni. A Nagelkerke-féle mutató maximális esetben felveszi az egyet, ezt pedig úgy éri el, hogy a Cox-Snell-féle R^2 értékét osztja el a mutatónak az adott modellhez tartozó lehetséges maximum értékével.

ahol D_0 jelöli a log likelihood függvény értékét a logisztikus regresszió nulladik lépésében, D_M pedig a függvény értéke az adott lépésben (Barna I., Székelyi M., 2002, pp. 378-385, Shtatland, E. S., Kleinman, K., Cain, E. M., 2008).

3.4. A LOGIT MODELL ALKALMAZÁSA SORÁN VIZSGÁLT HALLGATÓI DÖNTÉST BEFOLYÁSOLÓ TÉNYEZŐK

A hallgatói döntések elemzésével kapcsolatban számos munka született. Ezek közül is sok olyannal találkozhatunk, amely logisztikus regressziót alkalmaz ez elemzések során. A 4. ábra ismerteti azt, hogy melyek azok a tényezők, amelyek megjelennek független változóként a hallgatói döntések elemzésében.



5. ábra: A hallgatói döntéseket befolyásoló vizsgált tényezők a logit modell alkalmazása során

Forrás: Saját szerkesztés Schwartz, B. 1985, Ordovensky J. F., 1995, Weiler, W. C., 1989, Weiler, W. C., 1986, Niu, S. X., Tienda, M., 2008, Cohen-Zada D., Justman M., 2012, Montgomery, M., 2000, Montmarquette, C., Cannings, K., et al.: 2002, Reynolds, L. C.: 2012, Horstschräer, J.: 2012, Jung, J., Diane M. et al., 2012, DesJardins, S. L., Dundar, H., et al.: 1999, Toutkoushian, R. K.: 2001, Long, B. T.: 2004, Coelli, M. B.: 2011, Flannery, D., Kennelly, B., et al.: 2013 alapján.)

Munkám során megvizsgáltam, hogy a szakirodalomban fellelhető logit modell alkalmazások során melyek azok a független változók, melyeket megvizsgáltak, mint hallgatói döntést befolyásoló tényezőket. E kutatásokban döntésként megjelenik az intézmény kiválasztása, illetve számos cikk foglalkozik azzal a kérdéssel, hogy melyek azok a tényezők, amelyek befolyásolják, hogy a vizsgált egyén milyen típusú felsőoktatási intézménybe adja be jelentkezését.

Ahogy az ábrán is látjuk számos tényező hatását vizsgálták a hallgatói döntésekre vonatkozóan. Ezeket a következő négy csoportba soroltam:

- Jelentkezésre vonatkozó változók
- Jelentkező környezetére vonatkozó változók
- Szülői háttér
- Intézményre vonatkozó változók

E kutatásokban, a legnagyobb hangsúlyt magára a jelentkezőre, illetve az intézményre vonatkoztatott megközelítések kapják. Ezen kívül szerepet kapnak a szülői háttérrel jellemező változók, mint például a szülő végzettsége, valamint a hallgató környezetére vonatkozó változók, de ezek előfordulásának gyakorisága elmarad a korábban említett két kategória mellett.

Az ábrán azt is láthatjuk, hogy nem szerepelnek olyan változók az elemzésekben, melyek az intézményben oktatók színvonalával, illetve az intézmény környezetével foglalkoznának. Az intézmény környezetére vonatkozó változók a gravitációs modell esetében is hiányoztak. A továbbiakban érdemes lehet e területeket is mélyrehatóbban elemezni.

A következő részben egy olyan módszert mutatok be, amely szintén alkalmazható a döntések elemzésére. Ez pedig a döntési fa módszer.

3.5. A DÖNTÉSI FÁK

A döntési fák elsősorban csoportosításra és előrejelzésre alkalmazott eszközök. A döntési fák élekből, csúcsokból, épülnek fel, ahol az élek jelentik az egyes tulajdonságokhoz tartozó értékeket, míg a csúcsok az egyes osztályokhoz tartozó értékeket (Duchessi, P., Lauria, E.J.M, 2013).

Egy út a döntési fában egy lehetséges osztályozási, vagy előrejelzési megoldást jelent. A döntési fákat a tanító adatbázisból rekurzív vágással tudjuk előállítani. Ez azt jelenti, hogy a teljes tanító adatbázisból indulunk ki, és olyan kérdést keresünk, ami alapján úgy tudjuk szétválasztani az adatbázist, hogy a független változó eloszlása a keletkezett részekben kevésbé legyen szórta, mint a kiinduló esetben. Vannak olyan algoritmusok is, melyek arra is törekednek, hogy a keletkező részek nagysága közel azonos legyen. Ezt az eljárást a keletkezett részekre kell rekurzívan alkalmazni. Ezt a rekurziót a következő esetekben kell megszakítani:

- Nincs több attribútum, ami alapján tovább tudnánk osztályozni.
- Az adott mélység elért egy előre megadott korlátot.
- Nincs már olyan vágás, ami javítani tudna a jelenlegi osztályozáson.

Azt, hogy hogyan is válasszuk szét a vizsgált elemeket, a vágási függvény adja meg. Ez alapján többféle módszert különböztetünk meg, melyek különféle vágási függvényt alkalmaznak (Hátori G., 2001).

Egy ilyen alapvető módszer az úgynevezett CRT (Classification and Regression Tree) módszer, melynek lényege a következő:

Legyen X egy diszkrét valószínűségi változó, amely k értéket vehet fel. Az eloszlásfüggvény értékei pedig: $P=(p_1, p_2, \dots, p_k)$. Az úgynevezett Gini-index a következőképpen adható meg:

$$Gini(P) = 1 - \sum_{j=1}^k p_j^2 \quad (21)$$

A CRT algoritmus úgy működik, hogy a Gini-index értékét minimalizálja. A CRT eljárás mindig bináris döntéseket használ a csomópontokban. Alkalmazása során olyan attribútumot keresünk, amellyel a legnagyobb homogén osztályt tudjuk leválasztani az adatbázisból.

Egy másik eljárás a CHAID (Chi-squared Automatic Interaction Detection) módszer, amely, ahogy a nevéből is adódik, a χ^2 -tesztet használja. A leglényegesebb különbség a két eljárás között, hogy mit tekintenek jó vágásnak. Ha az Y függő változó inter-

vallumskálán mért, akkor a CRT eljárás a varianciájának csökkentésére törekszik, míg a CHAID módszer F-tesztet használ. Ezért a CRT eljárás nagyméretű fát hoz létre, míg a CHAID módszer egy konzervatív eljárás, vagyis csak addig növeli a fát, amíg a vágás szignifikancia szintje meg nem halad egy előre adott értéket. Mindkét módszer esetén elmondhatjuk, hogy egyes vágási módszerekkel ellentétben mindkettő alkalmazható klasszifikációra és előrejelzésre egyaránt. (Bodon, 2010)

Eddig áttekintettünk a témával kapcsolatos hazai és idegen nyelvű szakirodalmat. A következő nagy fejezet a kutatómunkám bemutatásáról szól.

4. A KUTATÁS BEMUTATÁSA

E fejezetben azt a tevékenységet fogom bemutatni, amely során létrehoztam az elemzésekhez szükséges adatbázisokat. Ehhez elsőként ismertetem, hogy munkám során mely területre koncentrálok.

4.1. A VIZSGÁLT TERÜLET

Dolgozatomban a hallgatói preferenciák elemzését két nagy tudományterületre, és ezen belül egy-egy szakra vonatkozóan végeztem el. A két tudományterületre való koncentrálnak okai a következők: Az összes tudományterület vizsgálata oly nagy mennyiségű adathalmazt jelent, hogy azt e dolgozat keretein belül, a szakmai kíváncsiságnak megfelelő részletességgel és pontossággal megvalósítani a rendelkezésre álló informatikai eszközök teljesítményéből kifolyólag igencsak nehézkes lenne. A tudományterületekre való szétbontás további előnye a nagyobb mértékű homogenitás a vizsgált sokaságra vonatkozóan. Így külön vizsgálhatom a különböző érdeklődésű hallgatók preferenciáit, ezzel tovább növelve az elemzések részletességét és pontosságát.

Az elemzésekhez a 2011-es felsőoktatási jelentkezések adatait tartalmazó teljes adatbázist használtam fel. Az adatbázist megvizsgálva azt tapasztaltam, hogy a vizsgált évben a legtöbb jelentkező a gazdaságtudományi, illetve a bölcsész tudományi területhez tartozó szakokra adta be jelentkezését. Munkám során e két tudományterületre vonatkoztatva végzem elemzéseimet logit, illetve gravitációs modell segítségével.

A tudományterületek elemzése után mindegyiken belül egy-egy konkrét szakra, nevezetesen a gazdálkodás menedzsment, illetve szabad bölcsész szakokra is elvégzem a vizsgálatokat. A vizsgálatot azért végzem el így is, mert a konkrét szakokra történő szűréskor feltételezhetjük, hogy az újonnan kapott sokaság homogénebb jellemzőkkel bír, mivel így már a vizsgált adatbázisban csak egy-egy szak lesz jelen, és ezzel kiszűrhetők a kiugró értékkel rendelkező elemek.

Kutatásom során e modellek alkalmazásával arra keresem a választ, hogy melyek azok a tényezők, amelyek a jelentkezők döntését befolyásolják az egyes felsőoktatási intézmény megjelölése során.

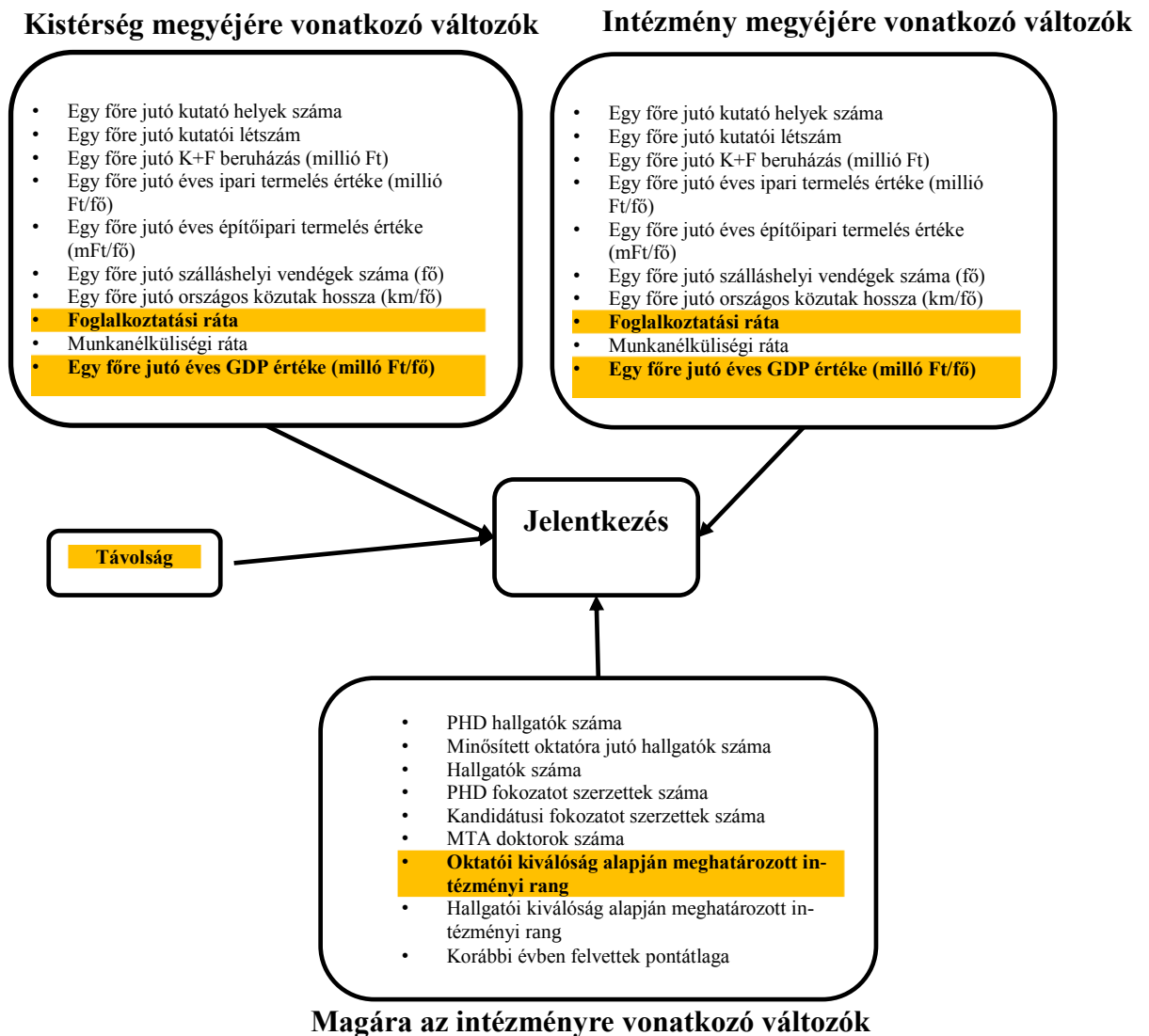
A következő részben bemutatom a modellek függő és független változóit.

4.2. A VÁLTOZÓK ISMERTETÉSE

A változók közül először a független (magyarázó), majd a függő (magyarázott) változókat mutatom be.

4.2.1. Független változók

Dolgozatomban azt vizsgálom, hogy melyek azok a változók, melyek hatást gyakorolnak a jelentkezők döntésére. Az összegyűjtött független változókat mutatja be a 6. ábra.



6. ábra: Az elemzésekhez rendelkezésre álló független változók

Forrás: Saját szerkesztés

Az ábrán szerepel a kistérség megnevezés. A kistérség itt a jelentkező lakóhelyét jelöli, vagyis innen jelentkezik egy (vagy több) intézménybe. Ez a társadalmi mobilitás vektorának kezdőpontja, ahová a vektor mutat, az pedig az intézmény székhelye.

Ahogy az ábrán is láthatjuk, a rendelkezésre álló független változókat négy megközelítésben csoportosítottam, ezek a következők:

- magára az intézményre vonatkozó változók
- az intézmény környezetére (megye) vonatkozó jellemzők
- az jelentkező környezetére (kistérség) vonatkozó jellemzők
- a távolság (itt tulajdonképpen a változó maga egy külön csoport, mert egyértelműen nem sorolható be a másik három csoportba)

Az ábrán látható számos változó közül célszerű kiválasztani azokat, amelyek talán legjobban jellemzik csoportjukat. Ez azért lényeges, mert figyelembe kell venni azt is, hogy a független változók között ne álljon fenn multikollinearitás. A számos változó közül e dolgozat keretein belül csak az egyes csoportokból csak a *feltételezhetően* legszignifikánsabb változókat vizsgálom (a változók összevont elemzésére egy későbbi kutatásban kerül sor), melyek a következők:

- Oktatói kiválóság alapján meghatározott rang
- Távolság
- Intézmény megyéjének egy főre jutó GDP-je
- Intézmény megyéjének foglalkoztatási rátája
- Kistérség megyéjének egy főre jutó GDP-je
- Kistérség megyéjének foglalkoztatási rátája

Az itt felsorolt független változók narancssárga kiemeléssel szerepelnek a fenti ábrán. Az oktatói kiválóság alapján meghatározott rangszám az Eduline.hu adatbázisából származó változó. Segítségével a 2011-ben felállított felsőoktatási rangsort tudjuk megadni az oktatók kiválósága alapján. Az intézmény és a jelentkező környezetére vonatkozó változók megyékre vonatkoznak. Ezek a foglalkoztatási ráta, illetve az egy főre jutó GDP értéke. Ezek azért kerültek be a független változók közé, mert feltételezésem szerint az egyének döntését az intézmény kiválasztásakor egyrészt befolyásolja az is, hogy milyenek a munkalehetőségek, illetve a fejlettség az adott helyen, ahova jelentkezni szándékoznak. Másfelől a saját környezetére vonatkozó jellemzők is befolyásolhatják döntését. Ha például alacsony a foglalkoztatási ráta az egyén környezetében, az motiválhatja, hogy olyan helyen tanuljon tovább, ahol jobbak a lehetőségek, így a tanulmányok végeztével könnyebben tudna elhelyezkedni.

Végül pedig a távolság is szerepet kap, mivel a döntéskor fontos szempont lehet az, hogy milyen messzire kell a jelentkezőnek utaznia, hiszen ez anyagi vonzattal is jár.

4.2.2. Függő változók

A két modellben a függő változók nem ugyanazok, ugyanis a logit modell esetén azt vizsgálom, hogy az egyes független változók hogyan befolyásolják a jelentkezők preferenciáit arra vonatkozóan, hogy az adott intézményeket hányadik helyen jelölik meg, a gravitációs modell estében pedig azt, hogy e tényezők milyen hatással vannak arra, hogy az egyes kistérségekből az egyes intézményekbe hányan jelentkeznek. Így a logit modellnél a függő változó a jelölés sorszáma, míg a gravitációs modell esetén a jelentkezések száma.

4.3. AZ ADATBÁZISOK FELÉPÍTÉSE

A következőkben bemutatom az elemzés alapjául szolgáló adatbázisok felépítését, majd a vizsgálatok menetét.

Elsőként bemutatom a logit modell alapjául szolgáló adatbázist.

Felhasznált adatbázisok:

Az adathalmaz két kisebb egységből épül fel, egy intézményi és egy jelentkezőre vonatkozó részből.

Az intézményi adatbázisban a következő változók kapnak helyet:

- az intézmény oktatói kiválóság alapján meghatározott rangja
- az intézmény megyéjére vonatkozó foglalkoztatási ráta
- az intézmény megyéjére vonatkozó GDP érték egy főre számítva
- az egyes intézmények központjának x és y GPS koordinátái

Ezek rendelkezésre állításához az oktatói kiválóság alapján meghatározott rangsorhoz az Eduline.hu adatbázisából származó értékeket használtam fel. Ahhoz, hogy a foglalkoztatási rátát illetve az egy főre jutó GDP értékeit intézményekhez társíthassam, meg kellett határozni mindegyik intézmény központjának megyéjét. Itt úgy jártam el, hogy azon intézmények, melyek központja Budapesten található, külön halmaznak számítsanak, mert a későbbi elemzések során ezeket kiszűrve azt is megvizsgálhatom, hogy hogyan alakulnak az eredmények, ha a budapesti felsőoktatási intézményeket kiveszem a modellből. Továbbá szükséges volt minden egyes intézmény központi GPS koordinátáinak meghatározása, mert ezek felhasználásával tudom kiszámítani a hallgató kistérségének és a megjelölt intézmény központjának a légvonalbeli távolságát. Fontos meg-

jegyezni, hogy a GPS koordináták meghatározásánál az összes 2011-ben megjelölt 2874 szakra vonatkozóan határoztam meg a GPS koordinátákat, hiszen előfordulhatnak kihelyezett szakok is. Ebben az esetben, a szakokra vonatkozó elemzésnél, torzítás lépne fel abban az esetben, ha az illető egy kihelyezett szakot jelöl meg, az adatbázisban pedig a szakhoz eredetileg tartozó intézmény GPS koordinátájával számolnék.

A logit modell felépítéséhez ezek után egy, a jelentkezőkre vonatkozó adatbázisra is szükségem volt, melyben a következő változók kaptak helyet:

- A jelentkező által megjelölt intézmény (ez tulajdonképpen 5 változó itt még, mert az első 5 öt helyen megjelölt intézményt vettem figyelembe.)
- A jelentkező állandó lakcímének kistérsége
- Az adott jelentkező vonatkozó kistérség megyéje
- A kistérség megyéjének foglalkoztatási rátája
- A kistérség megyéjének egy főre jutó GDP értéke
- A kistérség központjának x és y GPS koordinátái

A jelentkező által megjelölt intézmény esetében összesen öt oszlopra volt szükség. Itt, ha az adott helyen nem jelölt meg intézményt, vagy ha igen, de a jelentkezést nem a vizsgált tudományterületre vonatkozóan (gazdaságtudományi vagy bölcsész tudományi) adta be, akkor nulla szerepel, egyébként pedig a megjelölt intézmény száma. A logit modellre vonatkozó adatbázisban ezen oszlopokból képezem a függő változót, vagyis azt hogy az adott intézménybe jelentkezett-e az illető, és ha igen, akkor hányadik helyen jelölte meg azt.

Természetesen az x és y GPS koordinátákra itt is szükség van, hiszen ezek a távolságszámításhoz szükségesek. Most pedig ismertetem, hogy a rendelkezésre álló adatbázisokból hogyan hoztam létre a logit modell adatbázisát.

4.3.1. A logisztikus regresszióhoz felhasznált adatbázis felépítése

A logit modellhez szükséges adatbázis felépítése a fentiekben bemutatott két részből tevődik össze, összeállítását Matlab program segítségével oldottam meg. Felépítését tekintve úgy került kialakításra, hogy minden egyes jelentkező esetében felsorolásra került az összes intézmény (de csak azok ahol az adott tudományterületre, vagy a szakok esetében a vizsgált szakra jelentkezni lehet). Tehát egy jelentkező annyi sort jelent az adatbázisban, ahány intézmény szerepel a vizsgálódásainkban. Az egyes változók megfelelő értékei a program segítségével lettek társítva a megfelelő helyre. Így már rendel-

kezésre álltak egymáshoz társítva a megfelelő GPS koordináta párok is, melyből légvonalbeli távolságot számítottam a Föld lapultságának figyelembevételével.

A függő változó (jelentkezés) esetében azonban kétféle módon jártam el, mert a kétféle logisztikus regresszió alkalmazásához különféle megoldás szükséges.

A bináris logisztikus regresszióhoz úgy kellett meghatároznom a függő változót, hogy értéke nulla legyen abban az esetben, ha az illető nem jelentkezett az adott intézménybe (első öt helyen), illetve egy legyen akkor, ha igen.

A multinomiális logisztikus regresszió esetében összetettebb a feladat. Ott ugyanis nem elegendő az, hogy jelentkezett-e a hallgató az intézménybe vagy nem. Itt arra is szükség van, hogy ha jelentkezett, akkor az adott intézményt hányadik helyen jelölte meg.

Ezek összeállításával a logisztikus regresszióhoz rendelkezésemre állt az elemzések elvégzéséhez szükséges adatbázis. Ez azonban munkámnak csak egyik része, ahogy a korábbiakban írtam, vizsgálataimat nem csak logit, hanem gravitációs modell segítségével is elvégzem. A következőkben a gravitációs modell adatbázisának felépítését ismertetem.

4.3.2. Gravitációs modellhez felhasznált adatbázis felépítése

A gravitációs modell esetében az adatbázis soraiban már nem a jelentkező egyének szerepelnek. Ebben a modellben térbeli pont-párok kerültek kialakításra. Magyarországon összesen 175 kistérség van, ezek közül minden egyeshez társítani kell az összes intézményt (természetesen az összes alatt azon intézményeket értem, melyek az adott vizsgálódás szempontjából relevánsak, tehát például gazdaságtudományi terület esetén csak azok, ahol gazdasági irányultságú szakra lehet jelentkezni).

A függő változók beépítése itt is Matlab program segítségével történt. A különbség ez esetben az, hogy itt az intézményi adatbázison kívül nem jelentkezői, hanem egy kistérségi adatbázisra volt szükségem, mely az egyes kistérségekre vonatkozóan tartalmazza a független változókat.

A gravitációs adatbázisban a függő változó pedig az egyes térbeli pont-párokhoz tartozó jelentkezések száma. Ez azt mutatja meg, hogy az egyes kistérségekből az egyes intézményekbe hányan jelentkeztek. Az adatbázison azonban még egy fontos transzformációt el kell végezni. A gravitációs modell alkalmazása során ugyanis lineáris regresszió elemzést végzünk, és meghatározzuk az egyes változókra számított β értékeket. Ezt ebben az esetben nem tehetjük meg így, ugyanis esetünkben Cobb-Douglas típusú

függvényről beszélünk. Ez esetben pedig a β értékek a változók kitevőjében helyezkednek el. Ennek megfelelően a társadalmi interakciókat leíró függvényt a következő alakban írhatjuk fel:

$$F_{ij} = G \frac{M_{ij}^{\alpha} \cdot M_{ij}^{\beta}}{D_{ij}^{\theta}} \quad (22)$$

Esetünkben a felírt képletben i jelöli a jelentkező kistérségét, j pedig a megjelölt intézményt. Ez a multiplikatív formájú modell pedig logaritmizálva lineárisá alakítható. Ezért szükséges a függő változó, valamint a független változók elemeinek a természetes alapú logaritmusát kiszámítani és a modell összeállítása során az így kapott értékekkel dolgozni tovább (Neumanné V. I., 2010).

Ebben a fejezetben ismertettem vizsgálódásom területét, illetve bemutattam a vizsgált független és függő változókat, valamint az elemzésekhez szükséges adatbázisokat. Munkám során megvizsgálom a térbeli pont-párok közötti áramlást a gravitációs modell segítségével, majd ezután a jelentkezői preferenciákat bináris, illetve ordinális (rank ordered) logit modell segítségével is elemzem. A következő nagy fejezet a számítások során nyert eredményeket mutatja be. Elsőként tekintsük a gravitációs modell eredményeit!

5. A KAPOTT EREDMÉNYEK ÉRTELMEZÉSE

E fejezet tartalmazza a korábban is említett elemzések eredményeit, illetve a modellekből levont következtetéseket. A korábbiakban már láthattuk, hogy eddig milyen szempontokat vizsgáltak az egyes kutatómunkák során a hallgatói preferenciák elemzésében. Dolgozatomban új független változók hatását is vizsgálni fogom. Mielőtt rátérnék az eredmények értelmezésére, essen szó e változókról részletesebben is!

5.1. A MODELLBE BEKERÜLT ÚJ VÁLTOZÓK

A korábbi kutatásokban, ahogy azt a korábbiakban is már ismertettem, számos tényező hatását vizsgálták a jelentkezők döntésére vonatkozóan. Ott arra is utaltam, hogy nem jelennek meg a gravitációs modell esetén az intézmény környezetére vonatkozó változók, míg logit modell esetén ezen kívül az oktatók kiválóságát jellemző független változók. Ugyan vannak olyan kutatások, melyek független változóként a kereseti lehetőséget veszik figyelembe az adott szak elvégzése után, mégsem ekvivalens ez az intézmény környezetében tapasztalható elhelyezkedési lehetőségekről, melyre a foglalkoztatási ráta is utal. A következőkben bemutatom a modellekbe a korábbi kutatásokhoz képest újonnan beépített változókat!

5.1.1. Oktatói kiválóság alapján kapott rang

Ez a változó intézményi oldalról került beépítésre. Az oktatók színvonala nagymértékben kihat az adott intézményben magára az oktatási folyamat színvonalára is. Egy jelentkező döntésekor fontos tényező, hogy milyennek ítéli meg az adott egyetemet. Ez a megítélés természetesen nem csak egy szempont eredménye. Az elemzéshez az intézmény reputációját talán legjobban jellemző változót, az oktatói kiválóság alapján meghatározott rangot választottam ki. Így a vizsgálatok során azt is kideríthetem, hogy az oktatói kiválóság alapján meghatározott rang százalékos változása a térbeli pont-párok között fennálló hallgatói áramlásban mekkora változást eredményez.

Itt meg kell jegyezni, hogy ez a független változó csak a logit modell esetében számít új változónak a korábbi kutatásokhoz viszonyítva. Ezzel szemben az intézmény megyéjének foglalkoztatási rátája mindkét modell esetében új változónak számít.

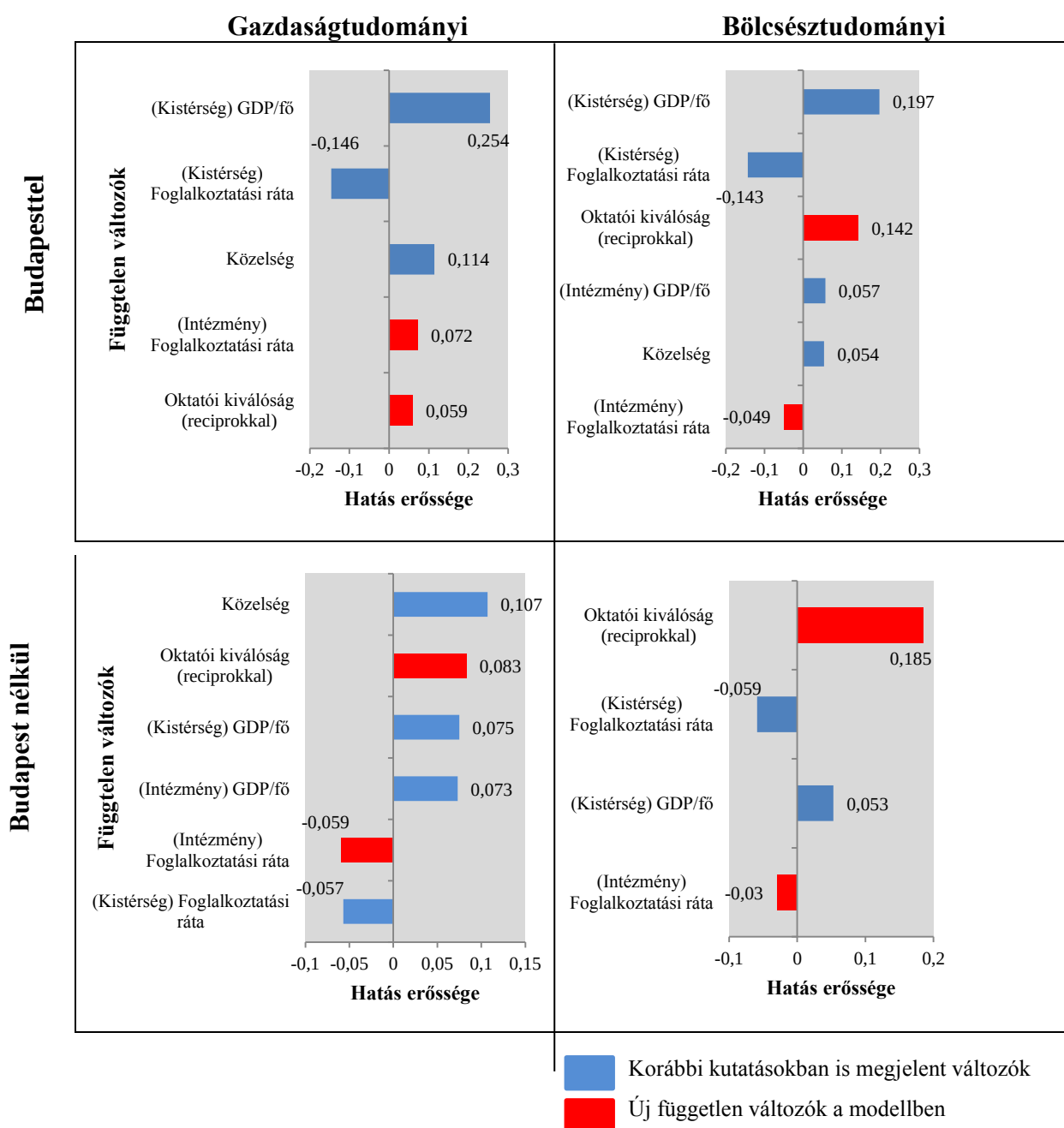
5.1.2. Intézmény megyéjének foglalkoztatási rátája

Egy intézmény környezetének gazdasági vonzereje meghatározó lehet a jelentkezéskor. Ha egy terület gazdasági potenciálja nagy, így sok más tényezőn keresztül is (mint például a kedvező elhelyezkedési lehetőségek, vagy magasabb jövedelmű állások) erős a vonzereje. A foglalkoztatási ráta jól szemlélteti egy adott területen az elhelyezkedési lehetőségeket. Magyarországon Budapest vonzereje e tekintetben is jelentősnek mondható. A változó elemzésekor érdekes lehet megvizsgálni azt is, hogy Budapest vonzereje milyen mértékben érvényesül a döntésre vonatkozóan.

Az új változók részletesebb bemutatása után ismertetem a gravitációs modell alkalmazásával kapott eredményeket.

5.2. A GRAVITÁCIÓS MODELL ALKALMAZÁSA SORÁN KAPOTT EREDMÉNYEK

A gravitációs modell segítségével megvizsgálhatjuk, hogy a térbeli pontok között milyen vonzások jelennek meg közgazdaságtani értelemben véve. Esetünkben a térbeli pontokat a jelentkezők kistérségének központja és a felsőoktatási intézmények székhelye képezi. Így a modell arra ad választ, hogy e térbeli pontok közötti áramlásokra mely tényezők vannak hatással a korábban ismertetett független változók közül. A vizsgálatot két tudományterületre, nevezetesen a gazdaságtudományi és bölcsész tudományi területre végeztem el. A kapott eredményeket a 7. ábra ismerteti.



7. ábra: A gravitációs modell eredménye

Forrás: Saját szerkesztés

Az ábra négy esetet mutat be. Egyrészt a vizsgálat eredményét láthatjuk a két tudományterületre vonatkozóan, másrészt e területeket úgy is megvizsgáltam, hogy kivettem a modellből a budapesti felsőoktatási intézményeket. Ennek oka a következő: A budapesti egyetemek komoly presztizzsel rendelkeznek, így érdemes megvizsgálni, hogy adódnak-e különbségek a modellben e két esetre vonatkozóan, és ha igen, akkor azok miben nyilvánulnak meg. Mielőtt azonban nekilátnánk a kapott eredmények értelmezésének, meg kell említeni, hogy egyes függő változók esetében transzformációt alkalmaztam a jobb átláthatóság végett. Ezt azért végeztem el, hogy egységesítsem a változók vizsgálatakor kapott eredményeket. Azokban az esetekben, ahol negatív kapcsolatot feltételezünk a függő és független változó között, a független változó reciprokával számoltam az elemzés során. Így az olvasó könnyebben átlátja a kapott eredményeket, hiszen a vizsgált változó kedvező irányba történő változása fog pozitív előjelű standardizált bétát eredményezni. A modellben ez két változót érint, egyrészt a térbeli pontok közötti távolságot (a továbbiakban így közelségként nevezem meg), másrészt pedig az oktatói kiválóság alapján meghatározott rangot. A táblázatban látható változók között új változók jelennek meg a korábbi kutatásokhoz képest, mégpedig az intézmény oktatói kiválóság alapján meghatározott rangja, valamint az intézmény megyéjének foglalkoztatottsági rátája. Ezeket a változókat az ábrán piros színnel emeltem ki.

A következőkben vegyük sorra elsőként a vizsgált független változók hatását általánosan, majd vizsgáljuk meg, milyen megállapításokat tehetünk, a budapesti egyetemek figyelembevételével, illetve anélkül.

5.2.1. Az egyes független változók hatásának bemutatása

A változók vizsgálatát az intézménynek a jelentkező kistérségéhez viszonyított közelségének interpretálásával kezdem. Az eredmények értelmezése során az érdekes megállapításokat dőlt betűvel emelem ki.

5.2.1.1. Közelség

Budapest figyelembevételével:

Ahogy az ábrán is jól látszik, a közelség erős befolyásoló tényező. A szignifikáns független változók között mindkét tudományterület esetében megjelenik. A standard β értékek mindkét esetben pozitívak, vagyis azt mondhatjuk, hogy minél közelebb helyezkedik el a két térbeli pont egymáshoz annál nagyobb a közöttük fennálló áramlás erőssége. Itt megjelenhetnek a gazdasági megfontolások is, hiszen a közelség csökkenésével növekednek az utazásra fordított kiadások is, ami szintén arra ösztönözheti a jelentke-

zőket, hogy jelöléskor a közelebb elhelyezkedő intézményt részesítsék előnyben. Azt is megfigyelhetjük, hogy tudományterületenként is különböző a standard β érték a változóra vonatkozóan. A gazdaságtudományi terület esetében értéke 0,114, míg a bölcsész-tudományi területre vonatkoztatott vizsgálat esetén értéke csak 0,054. azt mondhatjuk, hogy az eredmények alapján *a gazdaságtudományi területre jelentkezők számára nagyobb befolyásoló tényezőként jelenik meg a közelség, mint a bölcsész-tudományi területre jelentkezők esetében.*

Budapest figyelembevétele nélkül:

Itt különbség mutatkozik az előző esethez képest. A budapesti intézmények figyelembevétele nélkül azt mondhatjuk, hogy a két tudományterületet vizsgálva is eltérések jelennek meg a közelség vizsgálatakor. Míg a gazdaságtudományi területre jelentkező hallgatók esetében pozitív kapcsolat áll fenn a közelség és az adott intézménybe történő jelentkezés között, addig a bölcsész-tudományi területre jelentkezők esetén ez a változó nem szignifikáns. Feltételezhető, hogy Budapesttől távolodva csökken az ELTE vonzása, ami azt eredményezi, hogy még több Budapesti és Pest megyei jelentkezik bölcsészetre, míg távolabbról pedig kevesebben. Eközben pedig minden más bölcsész-képzést adó helyre a megmaradt érdeklődés viszonylag egyenletesen jelentkezik a távolságtól függetlenül.

5.2.1.2. Oktatói kiválóság által meghatározott rang

Budapest figyelembevételével:

Az oktatói kiválóság által meghatározott rang tekintetében azt mondhatjuk, hogy mindegyik esetben pozitív kapcsolatban áll a függő változóval. Erősségét tekintve előrébb helyezkedik el bölcsész-tudományi területre jelentkezők esetén, mint a gazdaságtudományi terület esetében. A gravitációs modell alapján azt mondhatjuk, hogy a bölcsész-tudományi területre jelentkezők esetében az oktatói kiválóság által meghatározott rang nagyobb befolyással bír a döntéskor, mint a gazdaságtudományi terület esetén, ha a budapesti egyetemeket is figyelembe vesszük. A változó általános hatása logikusnak tűnik, hiszen *minél magasabb az oktatók színvonala egy intézményben (tehát ez alapján előrébb kerül a hallgatói rangsorban), feltételezhetően annál nagyobb valószínűséggel fog a jelentkező az adott intézménybe jelentkezni.* Ne feledkezzünk meg róla, hogy a pozitív kapcsolat itt a kedvező irányba történő változásra vonatkozik az egységesítés miatt!

Budapest figyelembevétele nélkül:

Ha kivesszük a modellből a budapesti egyetemeket, akkor érdekes észrevételt tehetünk. Mindkét tudományterületen magasabb standard β értéket kap ez a független változó, mint az előző esetben. Feltételezhetjük, hogy erre magyarázattal szolgál, *hogy a budapesti egyetemek esetén a jelentkezők kevésbé veszik figyelembe az oktatói kiválóságot. Ennek oka lehet, hogy egy adott jelentkező, akkor is szívesebben jelöl meg egy budapesti, általános megítélésben magasabb presztizzsel rendelkező intézményt, ha az oktatói kiválósági rangsorban egy vidéki egyetem után következne.*

5.2.1.3. Kistérség megyéjének foglalkoztatási rátája

Budapest figyelembevételével:

A kistérség megyéjének foglalkoztatási rátája mindegyik esetben negatív kapcsolatban áll a függő változóval. Véleményem szerint magyarázattal szolgálhat erre, a jelentkezőt ösztönző saját környezet. Feltételezésem szerint, ha egy egyén környezetében azt tapasztalja, hogy nehezen jut munkához, vagyis esetünkben azt mondhatjuk, hogy alacsony a foglalkoztatási ráta, akkor az ösztönzőleg hat rá, hogy magasabb végzettséget szerezzen a könnyebb állásszerzés végett.

Budapest figyelembevétele nélkül:

A budapesti intézmények figyelmen kívül hagyása után szintén negatív kapcsolatot figyelhetünk meg a vizsgált független és a függő változó között. Azt is megfigyelhetjük, hogy a standard β értékek alacsonyabbak ez esetben. Ennek egy lehetséges magyarázata az, hogy a jelentkező az alacsony foglalkoztatási rátával rendelkező környezet hatására szívesebben választ egy olyan intézményt, melynek környezetében a foglalkoztatási ráta magasabb. Ha nincsenek benne a modellben a budapesti egyetemek, akkor ez a hatás kevésbé érvényesül.

5.2.1.4. Kistérség megyéjének egy főre jutó GDP-je

Budapest figyelembevételével:

A kistérség megyéjének egy főre vonatkozó GDP-értéke komoly befolyásoló tényezőnek bizonyult a döntést befolyásoló tényezők közt. Mindegyik esetben azt mondhatjuk, hogy *minél magasabb a jelentkező kistérségének megyéjében az egy főre jutó GDP, annál nagyobb hajlandóságot mutat az adott felsőoktatási intézménybe való jelentkezésre.* Egy terület által termelt bruttó hazai össztermék szimbolizálja a térség gazdasági erejét. Így egy magasabb gazdasági színvonallal rendelkező területen jobb lehetőségek kínálóznak a továbbtanulásra.

Budapest figyelembevétele nélkül:

Ha kivesszük a modellből a budapesti egyetemeket, akkor is megmarad a két változó között fennálló kapcsolat iránya, de ez esetben már nem az első helyen szerepel a független változók között. Ennek oka lehet az, hogy *Budapest gazdasági potenciálja nagy, így sok más tényezőt keresztül is (mint például a kedvező elhelyezkedési lehetőségek, vagy magasabb jövedelmű állások) erős a vonzereje. Így Budapestet kivéve minden más egyetemi város messze alacsonyabb gazdasági erővel bír, ami azt eredményezi, hogy a többi szempont felértékelődik.* Ezt az is mutatja, hogy a bölcsészet esetén nem is jelenik meg szignifikánsan. Budapest a diákélet számára is vonzóbb város lehet, mint más csak egy-egy főiskolával rendelkező város, ez pedig indirekten hathat arra, hogy elhagyásával más szempontok kerüljenek előtérbe.

5.2.1.5. Az intézmény megyéjének foglalkoztatási rátája

Az intézmény megyéjének foglalkoztatási rátája esetén a budapesti egyetemek figyelembevételével, illetve anélkül is hasonló eredményeket kapunk a modell értelmezésekor. A vizsgált független változó a legtöbb esetben negatív kapcsolatban áll a függő változóval (kivételt csak a gazdaságtudományi terület vizsgálata képez a budapesti intézmények figyelembevételekor).

Budapesttel együtt a gazdasági területen pozitív a kapcsolat, míg bölcsészeknél negatív. Feltételezhető, hogy azért, mert, mint ahogy az látszik, más bölcsészképző helyekre például a oktatói kiválóság által meghatározott rangsor alapján jelentkezők, a gazdasági viszonyokat pedig kevésbé veszik figyelembe. Így lehetséges az, hogy Budapest gazdasági erejével szemben más várost preferálnak, amely alacsonyabb gazdasági ereje miatt lesz negatív a standard β érték. Ez alátámasztja azt, hogy a budapesti intézmények elhagyásával az oktatói kiválóság szerepe egészen az első helyre ugrik a bölcsészek esetében, miközben a gazdasági területen kiesik ez a változó.

5.2.1.6. Az intézmény megyéjének egy főre jutó GDP-je

Budapest figyelembevételével:

Budapest figyelembevételével ez a változó csak a bölcsész tudományi terület elemzésekor bizonyult szignifikánsnak. Ez esetben a standard β értéke pozitív, ami azt jelenti minél magasabb az intézmény megyéjében az egy főre jutó bruttó hazai össztermék értéke, annál többen jelentkezők az adott intézménybe. Itt is azzal a megközelítéssel élhetünk, hogy a magas GDP/fő érték magasabb gazdasági színvonalat jelent, ami ösztönzőleg hat a jelentkezőre.

Budapest figyelembevétele nélkül:

Ekkor a vizsgált változók kapcsolata megegyezik az előző esetben tárgyaltakkal. A különbség csak az, hogy itt nem a bölcsész tudományi, hanem a gazdaságtudományi területen bizonyult szignifikánsnak a vizsgált független változó. Ez arra enged következtetni, hogy a különböző érdeklődéskörbe tartozó jelentkezők preferenciái is eltérőek a jelentkezést illetően. A budapesti intézmények kivételekor megváltozott a helyzet, hiszen a másik tudományterületen jelent meg az ott korábban nem szignifikánsnak bizonyult változó. Ez két dologra enged következtetni: Az egyik az, hogy *a különböző érdeklődéskörbe tartozó jelentkezők preferenciái is eltérőek a jelentkezést illetően, a másik pedig, hogy Budapest olyan erőteret jelent a gravitációs modellben, hogy a budapesti intézmények figyelmen kívül hagyása képes erős változásokat előidézni a modellben.*

5.2.2. A gravitációs modellel kapott eredmények összefoglalása:

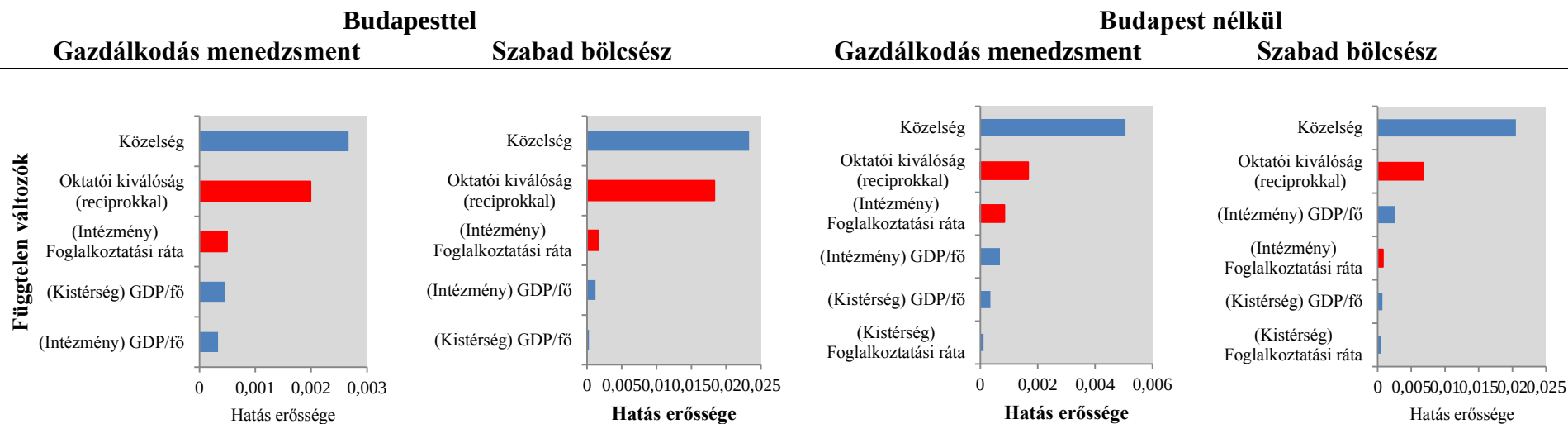
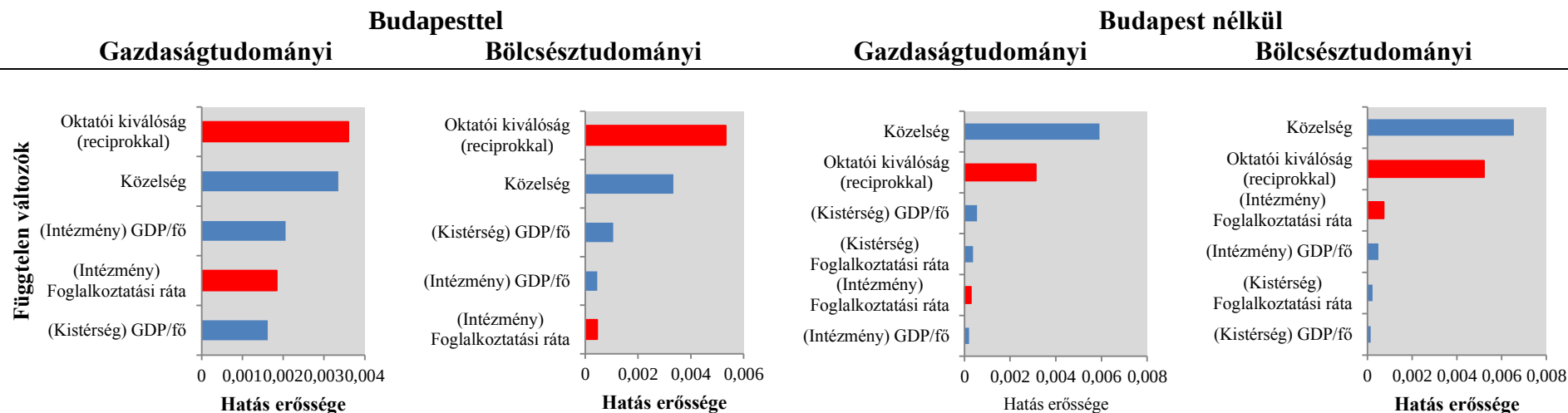
Összefoglalva az eddigi eredményeket a következő megfigyeléseket tehetjük:

- A jelentkezők esetében a közelség komoly befolyásoló tényező
- Az oktatói kiválóság által meghatározott rang a budapesti egyetem figyelembevételekor kisebb jelentőséget mutat a döntéshozatalkor
- Ha a kistérség megyéjének foglalkoztatási rátája alacsony, azt ösztönzőleg hat a jelentkezőre a tekintetben, hogy tanulmányait felsőoktatási intézményben folytassa.
- A jelentkező kistérségének gazdasági színvonalának növekedése növeli a jelentkezésre való hajlandóságot
- Az intézmény megyéjének foglalkoztatási rátája tekintetében nincs egyértelműen megmagyarázható kapcsolat
- Budapest olyan erőteret, amely a gravitációs modell eredményét nagymértékben befolyásolja.
- A különböző tudományterületekben érdekelt jelentkezők preferenciái is eltérőek a felsőoktatási intézményekbe történő jelentkezést illetően.

A gravitációs modell segítségével az egyes kistérségek és egyetemek közötti hallgatói (esetünkben még csak jelentkezők) áramlást vizsgáltam. Ehhez képest egy másfajta megközelítést jelent a logit modell alkalmazása, ahol azt vizsgáltam, hogy mely tényezők befolyásolják azt, hogy az egyén hányadik helyen jelöl meg egy adott intézményt. A következőkben a bináris logit modell eredményét ismertetem.

5.3. A BINÁRIS LOGIT MODELL ALKALMAZÁSA SORÁN KAPOTT EREDMÉNY ÉRTELMEZÉSE

A gravitációs modell esetén az vizsgáltam, hogy melyek azok a tényezők, amelyek befolyással vannak arra vonatkozóan, hogy egy adott kistérségből hányan jelentkeznek egy adott intézménybe. Ebben az esetben térbeli pont párok közötti áramlást modelleztem. A hallgatói preferenciák egy másik megközelítésben történő vizsgálata is érdekes. Ezért bináris logisztikus regressziós modell (bináris logit) alkalmazásával megvizsgáltam azt is, hogy mely tényezők vannak befolyással arra vonatkozóan, hogy egy jelentkező a leadott jelentkezésében melyik intézményt jelöli meg első helyen. Tehát míg a gravitációs modell arra alkalmas, hogy a pontok közötti áramlást modellezzük, addig a logit modell segítségével jelentkezői oldalról történő megközelítés alkalmazható és az is elemezhető, hogy melyik intézményt preferálja az adott jelentkező. A bináris logit modell esetén is mindkét kiválasztott tudományterületre vonatkozóan elvégeztem a vizsgálatokat, majd az elemzett jelentkezők körét tovább szűkítettem a gazdálkodás menedzsment és szabad bölcsész szakokra vonatkozóan. A 8. ábra szemlélteti a kapott eredményeket.



8. ábra: A bináris logit modell alkalmazása során kapott eredmény tudományterületekre, illetve szakokra vonatkozóan

Forrás: Saját szerkesztés

■ Korábban is vizsgált független változók
■ Új független változók a modellben

Az ábrán látható változók között a korábbi kutatásokhoz képest új változóként jelenik meg az intézmény oktatói kiválóság alapján betöltött rangja, valamint az intézmény megyéjének foglalkoztatottsági rátája. Ezeket a változókat az ábrán piros színnel emeltem ki.

A bináris logisztikus regressziót SPSS program segítségével végeztem el. A program outputjaként nem jelenik meg standardizált β érték (bár nagyon jól látszik, melyek azok a változók, amelyek szingifikánsnak bizonyulnak az adott esetben). Így erőssorrend meghatározása a standardizálás hiánya miatt ezek az értékek nem alkalmasak. Ezért döntési fákat alkalmaztam, hogy megkapjam az egyes esetekben a vizsgált változók erőssorrendjét. A döntési fa alkalmazása során CRT módszert alkalmaztam. A módszer segítségével mindegyik változóra egy értéket kapunk („importance”), ami a hatásának erősségét hivatott kifejezni.

A következőkben vegyük sorra az egyes változókat és vizsgáljuk meg az erőssorrendben elfoglalt helyüket! Mielőtt azonban rátérnék az eredmények interpretálására, emlékeztetőül megjegyzem, hogy a gravitációs modellnél ismertetett egységesítés (oktatói kiválóság alapján betöltött rang reciproka, illetve távolság reciproka) a független változókra vonatkozóan itt is, és a továbbiakban is érvényes. A távolság reciprokát itt is közelségnek nevezem meg.

5.3.1. Az egyes független változók hatásának bemutatása

A változók hatásának interpretálását a legnagyobb hatású független változóval, vagyis a közelséggel kezdem.

5.3.1.1. Közelség

Budapest figyelembevételével:

A közelség mindkét tudományterület esetén meghatározó tényező, az erőssorrendben második helyen szerepel. Tehát egy jelentkező annál előbbre sorol egy intézményt a választási lehetőségei között, minél közelebb van lakóhelyéhez. Feltételezhető, hogy a döntéskor pénzügyi megfontolások is szerepet játszanak, hiszen nagyobb távolság esetén magasabb kiadásokkal kell számolni (albérlet, utazási költség stb...) E szemponton kívül szerepet játszhat még a családtól, barátoktól való elszakadás problémája is.

Budapest figyelembevétele nélkül:

A budapesti intézmények figyelembevétele nélkül a közelség az erőssorrendben a változók között már az első helyre kerül, ami egy érdekes eredmény. Ennek egy lehetséges

magyarázata, hogy *a budapesti intézményeket (a már korábban is boncolgatott magasabb presztízs miatt) akkor is jobban preferálják a jelentkezők egy vidéki egyetemmel szemben, ha az éppenséggel távolabb helyezkedik el, mint a vidéki, vagyis akkor más független változó a távolság elé sorolható. Ha kivesszük a budapesti egyetemeket, akkor a távolság nagyobb szerepet játszik a döntéskor.*

5.3.1.2. Oktatói kiválóság által meghatározott rang

Budapest figyelembevételével:

Az oktatói kiválóság által meghatározott rang a közelség mellett a másik nagy hatással rendelkező tényező. A budapesti intézmények figyelembevétele esetén, az első helyen helyezkedik el a változók között. Az kiszámított $\exp(\beta)$ értéke nagyobb, mint egy, ez azt jelenti, hogy az oktatói kiválóság alapján betöltött rangban előbbre sorolt intézménybe nagyobb valószínűséggel adják le jelentkezésüket az érintettek az első helyen.

Budapest figyelembe vétele nélkül:

Ebben az esetben a vizsgált független változó a második helyen szerepel az erőssorrendben. A kiszámított $\exp(\beta)$ érték itt is nagyobb egynél, vagyis az oktatói kiválóság rangsorban, ha előrébb helyezkedik el egy intézmény, akkor azt a jelentkezők nagyobb hajlandóságot mutatnak annak első helyen történő megjelölésére. Érdekességgé jelenhet meg, hogy a gravitációs modell vizsgálatakor azt tapasztalhattuk, hogy az oktatói kiválóság alapján betöltött intézményi rang a budapesti egyetemek figyelmen kívül hagyásának esetén helyezkedett el előrébb az erőssorrendben, ellentétben a logit modellel. Itt meg kell, hogy jegyezzem, hogy ez a két modell különbségéből is fakad, hiszen ahogy a fentiekben erről szó esett a két esetben nem teljesen ugyanazt a jelenséget vizsgáljuk (Utóbbi esetben a jelentkező szempontjából vizsgáljuk, hogy első helyen jelöli-e meg az adott intézményt vagy sem, a korábbi esetben a térbeli pont-párokra vonatkoztatott jelentkezések száma a függő változó.).

5.3.1.3. Kistérség - Intézmény megyéjének egy főre vonatkoztatott DGP- értéke

E két változó hatását egyszerre ismertetem. Ennek oka a nagymértékű hasonlóság a két változó esetében. *Akár a budapesti intézményekkel, akár azok nélkül vizsgáljuk meg e két tényező hatását, láthatjuk, hogy mindkét esetben hatásuk igencsak elmarad a közelséghez vagy az oktatói kiválóság alapján meghatározott ranghoz képest. A modell eredményeképpen kapott $\exp(\beta)$ érték pontosan 1. Ez azt jelenti, hogy a két változó közül egyik sincs kü-*

lönösebb hatással arra, hogy a jelentkező első helyen jelöl-e meg egy adott intézményt vagy sem. Ez a döntési fa eredményében is látható, hiszen hatásuk erőssége (importance) jóval elmarad az oktatói kiválóság által meghatározott rangtól vagy a közelségtől.

5.3.1.4. Kistérség-intézmény megyéjének foglalkoztatási rátája

A hasonlóság ez esetben is fennáll, így e két változót is célszerű egyben értelmezni. A foglalkoztatási ráta esetén is azt mondhatjuk, hogy az $\exp(\beta)$ értékek ez esetben is 1 körül mozognak. Az eltérés azonban olyan csekély, hogy könnyelműség lenne kijelenteni azt, hogy ezek a változók egyértelműen befolyásolják a jelentkezők döntését. A kistérség megyéjének foglalkoztatási rátája nem is bizonyult szignifikánsnak abban az esetben, ha a budapesti intézményeket is benne hagyjuk a modellben. Megfigyelhetjük azt is, hogy a gazdaságtudományi terület esetén az intézmény megyéjének foglalkoztatási rátája jóval nagyobb hatást mutat a döntési fa eredményei alapján a többi esethez viszonyítva (budapesti egyetemek figyelembevételkor). Ennek egy lehetséges oka, hogy akik gazdaságtudományi területen szeretnének tovább tanulni, azokról nagyobb valószínűséggel állíthatjuk, hogy érdeklődőbbek a gazdasági mutatók iránt, vagy akár tájékozottabbak ezen a területen. Így nem lenne meglepő, ha egy olyan jelentkező, aki tájékozottabb az egyes régiók gazdasági, munkaerőpiaci helyzetéről, nagyobb figyelmet szentelne annak döntésekor.

5.3.2. Tudományterületek és szakok eredményének összehasonlítása

A bináris logit modell esetében vizsgálataimat nem csak tudományterületekre vonatkozóan végeztem el. Érdekes elemezni, hogyan változnak az eredmények akkor, ha a vizsgált sokaság nagyobb homogenitást mutat. Éppen ezért a tudományterületeken belül két konkrét szakra vonatkozóan is elvégeztem az elemzéseket. Ezek nevezetesen a gazdálkodás menedzsment és a szabad bölcsész szakok. A következőkben vegyük sorra, hogy az egyes változók esetében mutatkoznak-e különbségek a szakok vizsgálatakor az egész tudományterületekhez képest.

Első ránézésre az ábrán szembetűnik, hogy az összetartozó tudományterület-szak párok esetében ugyanazok a független változók bizonyultak szignifikánsnak.² A változás főként a független változók hatásának sorrendjében mutatkozik. Elsőként nézzük meg, milyen megállapításokat tehetünk a közelségre és az oktatói kiválóság által meghatározott rangra vonatkozóan!

5.3.2.1. Közelség - Oktatói kiválóság által meghatározott rang

A közelség esetében látható, hogy ha szakokra vonatkozóan végezzük vizsgálatainkat, akkor mindegyik esetben az első helyen áll ez a független változó, ellentétben a tudományterületek vizsgálatakor felállított modellel. Ugyanezt az oktatói kiválóság által meghatározott rang esetén is megállapíthatjuk, hiszen ez a változó egységesen a második helyen található a független változók között az erőssorrendben a szakok vizsgálatakor. Változás a tudományterületek-illetve szakok vizsgálatában a budapesti intézmények figyelembevételkor állt be. Lehet, hogy azért, mert sok olyan budapesti egyetem van, ahol meg lehet jelölni a választott két szakot, de azok nem budapesti egyetemek. Ezért az újonnan kapott, szakokra szűrt adatbázis arányaiban kevesebb budapesti intézményt tartalmaz a tudományterületihez képest. Ha ez így van, akkor pedig ezzel hasonló hatást érünk el, mint amikor a budapesti egyetemeket vettük ki a modellből. mivel korábban kiderült, hogy Budapest komoly erőternek számít, ezért a változások is ebben az esetben szembetűnőbbek.

5.3.2.2. Foglalkoztatási rátára, illetve egy főre jutó GDP-re vonatkozó változók

Míg a közelség és oktatói kiválóság rangszáma tekintetében azt mondhatjuk, hogy a szakokra végzett vizsgálatok során azonos képet mutat az erőssorrend az egyes esetekre vonatkozóan, az egy főre jutó GDP-re és foglalkoztatási rátára vonatkozó változók esetében ezt már nem mondhatjuk el. Innentől kezdve az erőssorrendben hátrafelé haladva a vizsgált független változók helyzete már nem egyforma a szakokra végzett vizsgálat egyes eseteiben. Ebből is látható, hogy az oktatói kiválóság által meghatározott rang, illetve a közelség

² Ez tulajdonképpen nem meglepő, hiszen, ha egy vizsgált változó egy adott adatbázisra vonatkozóan szignifikánsnak bizonyul a logisztikus regresszió során, akkor nagy az esély rá, hogy egy abból képzett, homogénebb sokaság esetében is szignifikáns lesz. Esetünkben pedig azt mondhatjuk, minél inkább szűkítjük a vizsgált jelentkezőket egy adott ismérv (szak) alapján, annál homogénebb lesz az újonnan kapott sokaság. Jelen esetben a homogenitás alatt azt értjük, hogy a konkrét szakra szűkített adatbázisban a jelentkezők preferenciáiról feltételezzük, hogy közelebb állnak egymáshoz, a még nagyobb hasonlóságot mutató érdeklődési terület miatt. Továbbá a szűkített adatbázisban szereplő intézmények is csak azok, amelyekben az adott szakra lehet jelentkezi, így az intézmények is feltételezhetően több hasonló jellemzővel bírnak egymáshoz képest.

olyan nagy befolyással rendelkező változók, melyek fontosságának megítélésében nagyobb az „egyetértés” a jelentkezők körében.

Azt is megfigyelhetjük, hogy *az intézmény megyéjének munkanélküliségi rátája és a kistérség megyéjének egy főre jutó GDP-je a szakok vizsgálatánál a négy esetből háromban azonos szinten helyezkedik el az erőssorrendben. A tudományterületekhez képest itt is nagyobb mértékű egyezést találunk.* A modell eredményeként kiszámított $\exp(\beta)$ -ra ez esetben is, csak úgy, mint a tudományterületek vizsgálatakor, 1 körüli értékek adódnak, vagyis azt mondhatjuk, hogy a feltehetően homogénebb sokaságban sem bírnak nagy befolyásoló hatással e változók.

Az összehasonlítás során így azt a következtetést vonhatjuk le, hogy az összetartozó tudományterület-szak párok esetén ugyanazok a változók szignifikánsak. A tudományterületek és szakok esetén a változók erőssorrendje különbségeket mutat, ugyanakkor a szakokra vonatkozó elemzésben nagyobb egyezést találunk (a szakok elemzésén belüli négy esetet összevetve) a vizsgált változók erőssorrendben elfoglalt rangját illetően.

A következőkben összefoglalom, melyek a főbb megállapítások a kapott eredmények értelmezése során.

5.3.3. A bináris logit modell alkalmazása során levont főbb következtetések

A fentiekben a bináris logisztikus regresszió eredményét mutattam be, tudományterületekre, valamint szakokra vonatkozóan is. Az eredmények értelmezésekor tett megállapításaimat a következő gondolatokkal foglalom össze:

- A közelség és az oktatói kiválóság alapján kapott rang érték rendelkezik a legnagyobb befolyással arra vonatkozóan, hogy a jelentkező hányadik helyen jelöl meg egy intézményt.
- A budapesti intézményeket akkor is előbbre sorolják a jelentkezők, ha azok egy adott jellemző alapján kedvezőtlenebb választásnak bizonyulnának.
- A jelentkező és az intézmény környezetére vonatkozó változók csekély befolyásoló hatással rendelkeznek (egyes esetekben egyáltalán nem befolyásol).
- A különböző tudomány- vagy szakterületen érdekelt hallgatók preferenciái is különböznek az adott intézménybe való jelentkezésre vonatkozó döntésnél.

- A szakokra vonatkoztatott vizsgálatkor még a különböző szakra jelentkezők esetén is nagyobb egyezést találunk a vizsgált változók erőssorrendben elfoglalt rangját illetően, mint a tudományterületek vizsgálatakor.

Az eddigiekben megvizsgáltam, hogy melyek azok a tényezők, melyek legnagyobb mértékben befolyásolják a jelentkezők döntését a felsőoktatási intézmény kiválasztásakor. A független változókra vonatkozó erőssorrendet láthattuk a bináris logisztikus regresszió és a gravitációs modell alkalmazása esetén is. Munkám során azt is vizsgáltam, hogy hogyan alakulnak az eredmények akkor, ha az egyes megjelölt helyekre vonatkozóan külön vizsgáljuk a modellt. Ehhez ordinális logit (Rank ordered logit) modellt alkalmaztam, melynek eredményeit mutatja be a következő alfejezet.

5.4. AZ ORDINÁLIS (RANK ORDERED) LOGIT MODELL EREDMÉNYÉNEK ISMERTETÉSE

A vizsgálatot SPSS program segítségével végeztem el. Multinomiális logisztikus regressziót alkalmaztam, de a következő beállításokkal: A referencia kategóriának nullát adtam meg, így az egyestől kezdve lépteti a program a függő változó értékét a maximálisig. A futtatás során a „Backward elimination” módszert használtam, amely fokozatosan kilépteti a modelltől a nem szignifikáns változókat. E módszer segítségével képet kaphatunk arról, hogy hogyan befolyásolják az egyes független változók szintről szintre haladva (1. helyen történő megjelölés, 2. helyen történő megjelölés, ... , 5. helyen történő megjelölés) a jelentkezést egy adott intézménybe.

Az elemzés eredménye látható az 1-es és 2-es táblázatokban. A változók közül a 0,05 alatti szignifikancia értékkel rendelkezőket tekintettem szignifikánsnak, ezeket szürke szín jelöli. Vastagon szedve a modellbe beépített új változókat emeltem ki. Narancssárga szín jelöli a táblázatokban az adott helyeken nem szignifikáns változókat.

A szükséges elemzéseket ez esetben is elvégeztem a két vizsgált tudományterületre és szakra is egyaránt. A következőkben vizsgáljuk meg a kapott eredményeket először a tudományterületekre vonatkozóan!

1. táblázat: Az ordiális (Rank ordered) logit modell eredményének összefoglaló táblázata tudományterületekre vonatkozóan

Jelentkezés helye	Vizsgált változó	Budapesttel				Budapest nélkül			
		Gazdaságtudományi		Bölcsész tudományi		Gazdaságtudományi		Bölcsész tudományi	
		Szignifikancia	Exp β	Szignifikancia	Exp β	Szignifikancia	Exp β	Szignifikancia	Exp β
Jelenetkezés 1. helyen	Oktatói kiválóság rang (reciprok)	0,000	1,372	0,000	2,141	0,000	1,34	0,000	3,327
	Közelség	0,000	2,482	0,000	2,117	0,000	2,038	0,000	1,7
	(Intézmény) Foglalkoztatási ráta	0,000	0,966	0,000	0,898	0,000	0,904	0,000	0,879
	(Intézmény) GDP/fő	0,000	1	0,000	1	0,000	1,001	0,009	1
	(Kistérség) Foglalkoztatási ráta	-	-	0,006	1,048	0,029	0,983	0,048	0,95
	(Kistérség) GDP/fő	0,090	1	0,000	1	0,000	1	0,003	1
Jelenetkezés 2. helyen	Oktatói kiválóság rang (reciprok)	0,000	1,224	0,000	1,628	0,000	1,177	0,000	2,415
	Közelség	0,000	2,224	0,000	2,113	0,000	1,836	0,000	1,702
	(Intézmény) Foglalkoztatási ráta	0,415	0,986	0,007	0,924	0,105	0,961	0,008	0,913
	(Intézmény) GDP/fő	0,000	1	0,000	1	0,009	1	0,285	1
	(Kistérség) Foglalkoztatási ráta	-	-	0,975	1,001	0,676	0,993	0,126	0,96
	(Kistérség) GDP/fő	0,000	1	0,000	1	0,000	1	0,002	1
Jelenetkezés 3. helyen	Oktatói kiválóság rang (reciprok)	0,794	1,012	0,000	1,554	0,299	1,052	0,000	2,557
	Közelség	0,000	2,095	0,000	2,024	0,000	1,731	0,000	1,627
	(Intézmény) Foglalkoztatási ráta	0,757	1,005	0,576	0,985	0,221	0,968	0,190	0,961
	(Intézmény) GDP/fő	0,003	1	0,001	1	0,110	1	0,037	1
	(Kistérség) Foglalkoztatási ráta	-	-	0,926	1,002	0,379	0,984	0,017	0,941
	(Kistérség) GDP/fő	0,001	1	0,037	1	0,000	1	0,153	1
Jelenetkezés 4. helyen	Oktatói kiválóság rang (reciprok)	0,943	0,995	0,000	1,38	0,000	1,7	0,049	1,784
	Közelség	0,000	2,045	0,000	2,18	0,711	1,028	0,000	1,77
	(Intézmény) Foglalkoztatási ráta	0,971	1,001	0,507	0,966	0,054	0,928	0,847	0,989
	(Intézmény) GDP/fő	0,000	1	0,001	1	0,001	1,001	0,608	1
	(Kistérség) Foglalkoztatási ráta	-	-	0,514	1,021	0,065	0,953	0,444	0,962
	(Kistérség) GDP/fő	0,041	1	0,010	1	0,109	1	0,337	1

Forrás: Saját szerkesztés

 Szignifikánsnak bizonyult változók az adott helyen
 Nem szignifikáns változók az adott helyen
Vastag betű Új változók a modellben

2. táblázat: Az ordinális (Rank ordered) logit modell eredményének összefoglaló táblázata konkrét szakokra vonatkozóan

Jelentkezés helye	Vizsgált változó	Budapesttel				Budapest nélkül			
		Gazdálkodás menedzsment		Szabad bölcsész		Gazdálkodás menedzsment		Szabad bölcsész	
		Szignifikancia	Exp β	Szignifikancia	Exp β	Szignifikancia	Exp β	Szignifikancia	Exp β
Jelentkezés 1. helyen	Oktatói kiválóság rang (reciprok)	0,001	1,022	0,000	1,828	0,000	1,666	0,000	2,712
	Közelség	0,000	4,815	0,000	6,394	0,000	3,789	0,000	3,864
	(Intézmény) Foglalkoztatási ráta	0,000	1,028	0,000	0,959	0,001	1,02	0,000	0,93
	(Intézmény) GDP/fő	0,000	1	0,000	1	0,000	1	0,000	1
	(Kistérség) Foglalkoztatási ráta	0,789	1,001	0,080	1,005	0,000	0,982	0,000	0,983
	(Kistérség) GDP/fő	0,000	1	0,000	1	0,000	1	0,000	1
Jelentkezés 2. helyen	Oktatói kiválóság rang (reciprok)	0,000	0,93	0,000	1,527	0,332	0,966	0,000	2,173
	Közelség	0,000	3,581	0,000	5,483	0,000	3,107	0,000	3,088
	(Intézmény) Foglalkoztatási ráta	0,000	1,166	0,000	0,96	0,000	1,222	0,000	0,943
	(Intézmény) GDP/fő	0,405	1	0,000	1	0,000	0,999	0,000	1
	(Kistérség) Foglalkoztatási ráta	0,000	1,018	0,000	1,019	0,918	1,001	0,890	0,999
	(Kistérség) GDP/fő	0,246	1	0,093	1	0,000	1	0,000	1
Jelentkezés 3. helyen	Oktatói kiválóság rang (reciprok)	0,000	0,909	0,000	1,436	0,079	1,06	0,000	2,021
	Közelség	0,000	4,01	0,000	5,639	0,000	3,332	0,000	3,348
	(Intézmény) Foglalkoztatási ráta	0,000	1,128	0,000	0,966	0,000	1,142	0,000	0,944
	(Intézmény) GDP/fő	0,101	1	0,000	1	0,045	1	0,000	1
	(Kistérség) Foglalkoztatási ráta	0,000	1,019	0,000	1,034	0,905	0,999	0,078	1,012
	(Kistérség) GDP/fő	0,387	1	0,003	1	0,000	1	0,000	1
Jelentkezés 4. helyen	Oktatói kiválóság rang (reciprok)	0,045	0,96	0,000	1,442	0,679	1,021	0,000	2,107
	Közelség	0,000	3,717	0,000	5,028	0,000	3,126	0,000	3,123
	(Intézmény) Foglalkoztatási ráta	0,000	1,13	0,000	0,944	0,000	1,141	0,000	0,905
	(Intézmény) GDP/fő	0,433	1	0,000	1	0,262	1	0,000	1,001
	(Kistérség) Foglalkoztatási ráta	0,217	1,008	0,000	1,038	0,741	0,997	0,464	1,008
	(Kistérség) GDP/fő	0,637	1	0,000	1	0,001	1	0,003	1

Forrás: Saját szerkesztés

 Szignifikánsnak bizonyult változók az adott helyen
 Nem szignifikáns változók az adott helyen
Vastag betű Új változók a modellben

5.4.1. Az ordinális (Rank ordered) logit modell eredménye tudományterületekre vonatkozóan

A táblázatokon első ránézésre is szembeűnő, hogy a szűrke színnel jelölt részek nagysága egyre csökken, ahogy haladunk az első helyen történő jelentkezéstől a 4. helyen történő jelentkezésig. A modellben nem kapott helyet az 5. helyen történő jelentkezés. Ennek oka, hogy az 5. helyen már olyan alacsony volt a minták száma, hogy abból már nem tudtam szignifikáns statisztikai eredményeket produkálni.

Megfigyelhetjük, hogy minél magasabb értékű helyen vizsgáljuk a jelentkezést befolyásoló tényezőket, annál több olyan változót találunk, ami nem szignifikáns. Tehát ez arra enged következtetni, hogy *a jelentkezők tudatossága a választáskor sokkal magasabb (vagyis több szempontot mérlegelnek) amikor arról hoznak döntést, hogy melyik intézményt jelöljék meg első helyen, mint amikor már a negyedik helyre választanak egy intézményt.* Szépen látszik az is, hogy a negyedik helyre történő választás felé haladva fokozatosan csökken ez a feltételezett tudatosság. A bennmaradt változókat megvizsgálva azt tapasztaljuk, hogy az első helyre vonatkozó jelentkezés esetén a kistérség megyéjének egy főre jutó GDP értékén kívül mindegyik változó szignifikánsnak bizonyul, míg a negyedik helyen történő jelentkezésnél, már szinte csak a közelség, illetve az oktatói kiválóság rangértéke az, ami szignifikáns marad. Ebből arra következtethetünk, hogy bár a jelentkezők „tudatossága” egyre csökken a soron következő jelentkezések helyeire vonatkozóan, *a közelség és az oktatók kiválósága olyan erős befolyásoló hatással bírnak, hogy e változókat még akkor is szem előtt tartják, amikor a negyedik helyen jelölnek meg egy intézményt.* Az egy főre jutó GDP-re vonatkozó független változók esetén az $\exp(\beta)$ értékekre mindegyik helyen 1 adódik, ez azt jelenti, hogy hatásuk igen csak csekély.

5.4.1.1. Budapest hatása az ordinális logit modellben

A fentiekben tárgyalt jelenség (ahogy az a táblázatból is látszik) a budapesti egyetemek figyelembevételével és a nélkül is fenn áll. Azonban azt megfigyelhetjük, hogy erőteljesebb a hatás akkor, ha a budapesti egyetemeket figyelmen kívül hagyjuk. A negyedik helyre vonatkozó jelölés esetén szemmel láthatóan több értéket találunk a 0,05 feletti szignifikanciaszinttel a budapesti egyetemek figyelmen kívül hagyása esetén. Ahogy a korábbiakban is kifejtettem Budapest erős gazdasági vonzerővel bír, így hozzá képest minden más egyetemi város messze alacsonyabb gazdasági vonzerővel rendelkezik, ami azt eredményezi, hogy Budapestet kivéve a modellből felértékelődik a többi szempont.

Ez nagyon jól látszik az eredményekben is, hiszen például *a bölcsész képzés esetén a gazdasági jellegű változók közül Budapest figyelembevétele után már egy sem szignifikáns. Ezzel szemben a közelség és az oktatói kiválóság alapján betöltött rang, még mindig fontos szempontoknak mutatkoznak a jelentkezéskor.*

A módszer jelen esetben is igazolta Budapest torzító hatását a jelentkezők döntésének elemzésére vonatkozóan.

5.4.1.2. Tudományterületek közötti eltérés az ordinális logit modellben

Nem csak az okoz eltéréseket a modellben, ha kivesszük a budapesti egyetemeket. Az eltérő tudományterületeken más-más változók maradnak benn a modellben. Például az első helyen történő jelentkezés esetén (budapesti egyetemek figyelembevételekor), a gazdaságtudományi területen csak a kistérség megyéjének egy főre jutó GDP értéke nem bizonyul szignifikánsnak. Ezzel szemben bölcsész tudományi terület összes változója szignifikáns. A negyedik helyre vonatkozó jelöléskor itt, mindkét esetben 2-2 változó van mely nem szignifikáns. Eltérést fedezhetünk fel, ugyanis gazdaságtudományi területnél ezek az oktatói kiválósági rang és az intézmény megyéjének foglalkoztatási rátája, a bölcsész tudományi terület esetén pedig az intézmény, illetve a kistérség megyéjének foglalkoztatási rátája. Egy érdekes megállapítás lehet azonban, hogy *bár sok eltérés fedezhető fel a két területre jelentkező preferenciáira vonatkozóan, ennek ellenére a közelség az egyetlen olyan változó, amely a legtöbb esetben szignifikáns mindkét tudományterületen. Tehát azt mondhatjuk, hogy bár a különböző tudományterületekre jelentkező különböző szempontok alapján döntenek, a közelség mindegyikük számára nagyon fontos jellemző.*

5.4.2. Az ordinális (Rank ordered) logit modell eredménye szakokra vonatkozóan

A tudományterületekre vonatkozó elemzésekhez képest láthatjuk, hogy kevesebb olyan változóról beszélhetünk, amely kiesik a modellből a magasabb sorszámú jelölési lehetőségek felé haladva. Ennek oka lehet az a homogenitás, amelyre már a korábbiakban is kitértünk. A szakok esetében azonban egy érdekes dolgot figyelhetünk meg. A budapesti egyetemek figyelembevételekor a szabad bölcsész szakra jelentkezők vizsgálatakor a 4. helyen történő jelölés esetén minden változó szignifikáns. Ez a szűk terület eltérő jellegzetes mutat a korábbiakban megállapítottakhoz képest. Ennek egy lehetséges oka, hogy Magyarországon a szabad bölcsész szakon magas pontszámokkal kell számolni. Ez pedig előidézheti, hogy egy ilyen szakra jelentkező egyén nagyobb valószínűséget tulajdonít annak az eseménynek a bekövetkezésére, hogy nem vesz fel az *i*-edik helyen

megjelölt intézménybe a szigorú ponthatárok miatt. Ennek tudatában körültekintéssel fog döntést hozni az $i+1$ -edik helyen történő jelentkezésre, mint az, akinek az esetben nem kell ilyen magas ponthatárokkal számolni.

5.4.2.1. Budapest hatása

A budapesti egyetemek figyelmen kívül hatása itt is változást eredményez a modellben, csak úgy, mint az előző esetben. Itt is megfigyelhetjük, hogy nem ugyanazok a változók maradnak szingifikánsak a budapesti egyetemek figyelembevételével és anélkül. Következésképpen *Budapest torzító hatása akkor is érződik, ha a vizsgált sokaság homogénebb jelleget mutat a korábbihoz képest.*

5.4.2.2. Különböző szakok közötti eltérések

Különböző szakok esetén is látszódik, hogy a jelentkezők döntésére különböző változók gyakorolnak hatást. A két szak közötti különbségre magyarázattal szolgálhat a ponthatárokra visszavezetett elmélet, miszerint, ha a szabad bölcsész szakra magasabb a felvételi ponthatár, akkor a jelentkező tartva az elutasítás lehetőségétől, a magasabb értékkel rendelkező jelentkezési helyeken is nagyobb körültekintéssel fog dönteni, mint az átlagosan alacsonyabb ponthatárra számító jelentkező. Ez pedig már eltéréseket fog eredményezni a két szakra jelentkezők preferenciái között.

A következőkben összefoglalom az ordinális (Rank ordered) logit modell eredményeinek értelmezése során tett megállapításaimat.

5.4.3. A fontosabb megállapítások az ordinális logit modell eredményeinek interpretálása során

Ebben a fejezetben arra kerestem a választ, hogy milyen eredményeket kapok, ha azt vizsgálom, hogy milyen kapcsolat állítható fel a függő és független változók között külön figyelembe véve az első, második, harmadik, illetve negyedik helyen történő jelöléseket. Az eredmények értelmezése során a főbb megállapítások a következők:

- A jelentkezők tudatossága a választáskor sokkal magasabb (vagyis több szempontot mérlegelnek) amikor arról hoznak döntést, hogy melyik intézményt jelölik meg első helyen, mint amikor már a negyedik helyre választanak egy intézményt.
- A közelség és az oktatók kiválósága olyan erős befolyásoló hatással bírnak, hogy e változókat még akkor is szem előtt tartják, amikor a negyedik helyen jelölnék meg egy intézményt.
- A módszer jelen esetben is igazolja Budapest kiemelkedő gazdasági vonzerejének hatását a jelentkezők döntésére vonatkozóan.

6. A GRAVITÁCIÓS ÉS LOGIT MODELLEK ÖSSZEHASONLÍTÁSA

A korábbiakban már említettem, hogy a gravitációs, illetve logit modellek a jelentkezésre vonatkozó döntést, más-más megközelítésben vizsgálják. Következésképpen a kapott eredmény sem jelenti ugyanazt mindkét esetben. Az egyes változókra kapott standard béta illetve $\exp(\text{béta})$ értékek összehasonlítása igencsak nehézkes volna, mivel előbbi esetben a térbeli pontok között fennálló áramlásra vonatkoznak, míg utóbbiban arra, hogy a jelentkezők hányadik helyen jelölik meg az adott intézményt. Azt azonban érdemes megvizsgálni, hogy a modellek az egyes hatásokra, mint például Budapest kiemelkedő vonzereje, vagy az eltérő tudományterületek preferenciái hogyan reagálnak. Az alábbi táblázat mutatja összesítve a gravitációs és a logit (bináris) modellek eredményeit. A korábbiakban ezeket az eredményeket már interpretáltam, így ebben a részben a két modell eredményeinek egymáshoz történő összehasonlítására koncentrálok.

Budapesttel

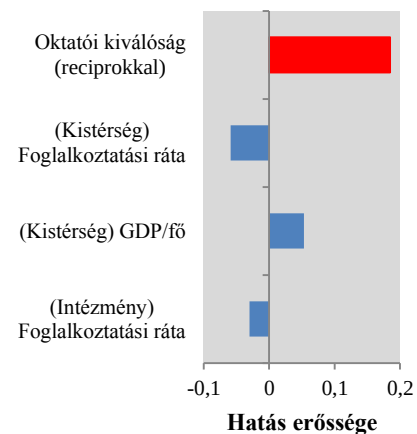
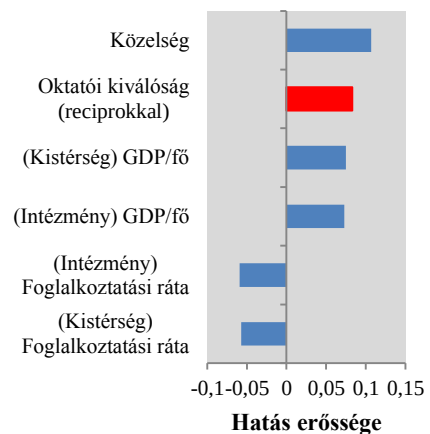
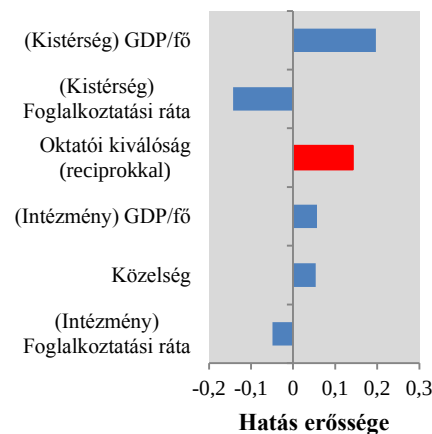
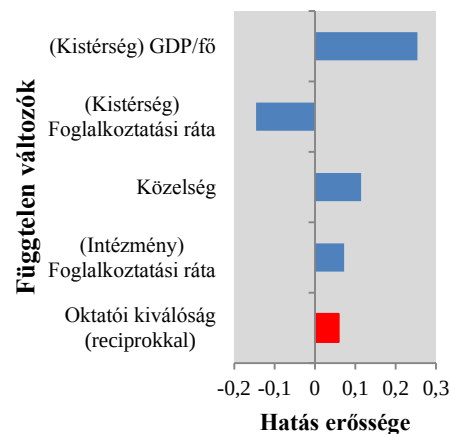
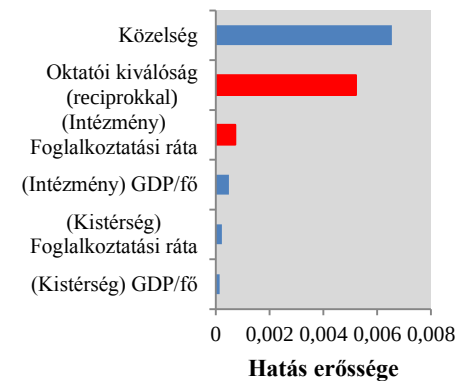
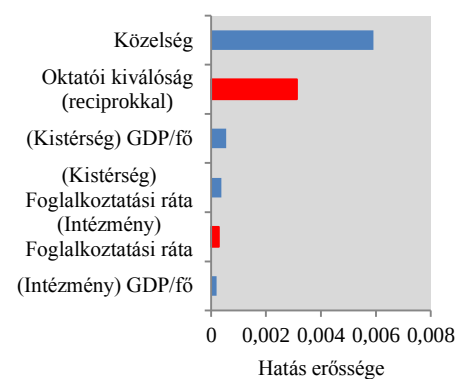
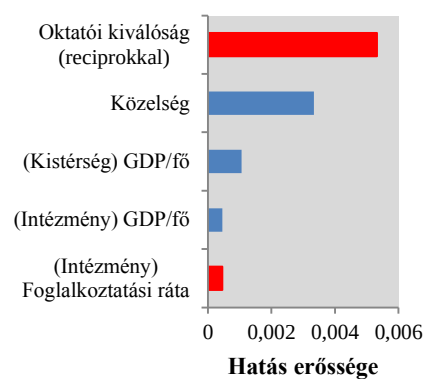
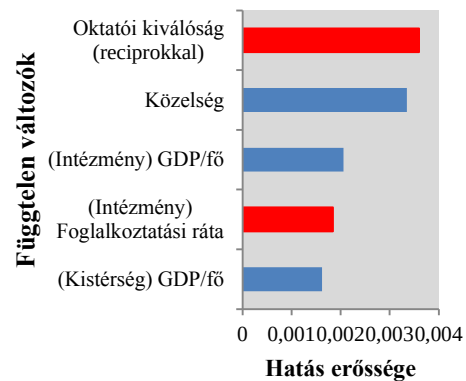
Budapest nélkül

Gazdaságtudományi

Bölcsész tudományi

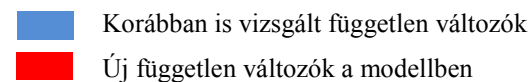
Gazdaságtudományi

Bölcsész tudományi



9. ábra: A gravitációs és bináris logit modellek eredményeinek összehasonlítása

Forrás: Saját szerkesztés



A 9. ábrán piros szín jelöli azokat a változókat új változóként lettek beépítve a modellbe.

A két modellt a következő szempontok alapján fogom összehasonlítani:

- Budapest kiemelkedő vonzerejének modellre gyakorolt hatása
- Az eltérő tudományterületek preferenciáinak különbözősége

Az összehasonlítást kezdjük Budapest kiemelkedő vonzerejének hatásával!

6.1. BUDAPEST KIEMELKEDŐ VONZEREJÉNEK MODELLRE GYAKOROLT HATÁSA:

Ahogy azt már több esetben is tapasztalhattuk a modellek eredményeinek bemutatásakor, Budapest olyan gazdasági potenciállal rendelkezik, amely komoly vonzerőt jelent a többi egyetemi várossal szemben. Így Budapestet a modellből elhagyva más preferenciák kerülnek előtérbe az egyes intézmények kiválasztásakor.

A budapesti felsőoktatási intézmények elhagyásával mindkét modell esetén megváltoznak az egyes változókra vonatkozó értékek. Például vegyük szemügyre a gazdaságtudományi területet! A logit modell esetén Budapest elhagyásával az oktatói kiválóság alapján meghatározott rang helyett a közelség kerül az első helyre, a gravitációs modellben pedig a kistérség megyéjének egy főre jutó GDP-je helyét veszi át a közelség az erőssorrendben. Ráadásul mindkét modell esetén olyan változók válnak szignifikánssá, melyek a változtatás előtt még nem bizonyultak annak (logit modellnél a kistérség megyéjének egy főre jutó GDP-je, gravitációsnál az intézmény megyéjének egy főre jutó GDP-je).

Ugyanígy a bölcsész tudományi területnél is megfigyelhetők az erőssorrendet ért változások. Budapest elhagyásával a logit modellnél az erőssorrendben az oktatói kiválóság alapján meghatározott rang helyét a közelség veszi át, míg a gravitációs modellnél a kistérség megyéjének egy főre jutó GDP-je helyett az oktatói kiválóság alapján meghatározott rang kerül az első helyre.

Ez alapján azt mondhatjuk, hogy Budapest kiemelkedő vonzerejének hatása mindét modell esetén megmutatkozik.

A következőkben vizsgáljuk meg, hogy a két modell a különböző tudományterületek vizsgálatakor milyen eltéréseket mutat!

6.2. AZ ELTÉRŐ TUDOMÁNYTERÜLETEK PREFERENCIÁINAK KÜLÖNBÖZŐSÉGE

Az eltérő tudományterületeken érdekelt jelentkezők eltérő preferenciái akárcsak a fentiekben vizsgált szempont, mindkét modell megmutatkozik. A jelenség erőteljesebben látszik azonban a gravitációs modell esetén. Itt a budapesti egyetemek elhagyásakor a gazdaságtudományi területen mind a hat független változó szignifikáns, míg a bölcsész-tudományi területnél már csak 4 szignifikáns független változónk marad (közelség, illetve intézmény megyéjének egy főre jutó GDP-je esik ki a modellből) a gravitációs modellnél. A ugyanitt a logit modellnél az első két helyen szereplő független változó (Oktatói kiválóság alapján meghatározott rang, illetve közelség) erőssorrendben elfoglalt helye megegyezik mindkét tudományterület esetén.

Ebből azt a következtetést vonhatjuk le, hogy mindkét modell kimutatja a különböző tudományterületen érdekelt jelentkezők preferenciájának különbözőségét, ugyanakkor megállapíthatjuk, hogy a gravitációs modell nagyobb különbségeket mutat a két területre vonatkozóan.

A 9. ábrán kívül egy érdekes dolgot még megfigyelhetünk a két modellel kapcsolatban. A 3. táblázat az egyes változókra számított $\exp(\beta)$ illetve standard β értékeket tartalmazza a két modell esetében.

3. táblázat: A gravitációs és bináris logit modellek eredményeinek összefoglaló táblázata

Modell	Független változók	Budapesttel		Budapest nélkül	
		Gazdaságtu- dományi $\exp(\beta)$	Bölcsésztu- dományi $\exp(\beta)$	Gazdaságtu- dományi $\exp(\beta)$	Bölcsésztu- dományi $\exp(\beta)$
Bináris logit	Oktatói kiválóság (reciprokkal)	1,057	1,797	1,702	2,601
	(Intézmény) Foglalkoztatási ráta	1,011	0,951	1,024	0,932
	(Intézmény) GDP/fő	1,000	1,000	-	1,000
	(Kistérség) Foglalkoztatási ráta	-	-	0,982	0,983
	(Kistérség) GDP/fő	1,000	1,000	1,000	1,000
	Közelség	4,009	4,073	2,848	2,686
Gravitációs modell		standard β	standard β	standard β	standard β
	Oktatói kiválóság(reciprok)	0,059	0,142	0,083	0,185
	(Intézmény) Foglalkoztatási ráta	0,072	-0,049	-0,059	-0,030
	(Intézmény) GDP/fő	-	0,057	0,073	-
	(Kistérség) Foglalkoztatási ráta	-0,146	-0,143	-0,057	-0,059
	(Kistérség) GDP/fő	0,254	0,197	0,075	0,053
	Közelség	0,114	0,054	0,107	-

Forrás: Saját szerkesztés

 Küldő helyre (kistérség megyéje) vonatkozó változók

Vastag betű – Új változók a modellben

A táblázatban külön figyelmet érdemel a narancssárga színnel jelölt rész. Megfigyelhetjük, hogy a logit modellnél a kistérség megyéjére vonatkozott változók esetén az $\exp(\beta)$ értéke a legtöbb esetben pontosan 1, vagy pedig nem is szignifikáns az adott változó. Ezzel szemben a gravitációs modell esetén jól interpretálható eredményeket kapunk a vizsgált változókra vonatkozóan.

Ebből azt a következtetést vonhatjuk le, hogy a küldő hely (esetünkben a kistérség megyéje) változóinak hatását a gravitációs modell jobban szemlélteti (könnyebben interpretálható eredményeket ad).

Az eddigi megállapításokat a következő gondolatokkal foglalhatjuk össze:

- Budapest vonzerejének kiemelkedő nagysága a gravitációs és a logit modell esetén is egyaránt megmutatkozik.
- A különböző tudományterületeken érdekeltek preferenciái között nagyobb különbségeket fedezünk fel abban az esetben, ha a térbeli pont párok közötti áramlást vizsgáljuk (gravitációs modell), mint amikor az intézmények megjelölésének sorszáma a függő változó (logit modell).
- A gravitációs modell segítségével a küldő hely változóinak hatása jobban szemléltethető (és interpretálható) mint a logit modell esetén.

7. ÖSSZEFOGLALÁS

Dolgozatom első részében ismertettem a statisztikai feltárás, a gravitációs-, és logit modellek és döntési fák szakirodalmi hátterét. Munkám során irodalomkutatót is végeztem, és megvizsgáltam külön a gravitációs, illetve logit modell esetén, hogy az eddigi kutatások milyen változók hatását vizsgálják a felsőoktatási intézményekbe való jelentkezésekre vonatkozóan. Az irodalomkutatás eredményeképp megállapítottam, hogy lehetőség nyílik még új változók beépítésére a megjelölt intézményre, illetve annak környezetére vonatkozó változók tekintetében.

Munkám során célul tűztam ki a magyarországi felsőoktatásba jelentkezők preferenciáinak elemzését a gravitációs és logit modellek segítségével. Dolgozatom második részében bemutattam a kutatási célokat, illetve kutatási kérdéseket fogalmaztam meg.

A továbbiakban bemutattam a kutatás során vizsgált változókat, az elemzésekhez szükséges adatbázisok szerkezetét és felépítésének menetét. Ezután interpretáltam elemzéseim eredményét külön a gravitációs, illetve bináris logit modellekre vonatkozóan, új függő változóként beépítve a gravitációs modellbe az intézmény megyéjének egy főre jutó GDP-jét, a logit modellbe pedig e változón kívül még az intézmény oktatói kiválóság alapján meghatározott rangját is. Dolgozatomban azt is megvizsgáltam, hogy a jelentkezők preferenciái hogyan alakulnak attól függően, hogy hányadik helyre tartozó jelentkezésről döntenek. Ehhez ordinális (rank ordered) logit modellt alkalmaztam.

Az egyes modellek eredményeinek értelmezése után már elegendő információ áll rendelkezésemre ahhoz, hogy megválaszoljam a feltett kutatási kérdéseket, melyek a következők voltak:

Első kutatási kérdés:

Melyek a jelentkezők döntését legnagyobb mértékben befolyásoló szempontok?

Első kutatási kérdésre adott válasz:

A közelség és az oktatói kiválóság alapján kapott rang érték rendelkezik a legnagyobb befolyással arra vonatkozóan, hogy a jelentkező hányadik helyen jelöl meg egy intézményt. E változók olyan nagy befolyással rendelkeznek a jelentkezésre vonatkozóan, hogy a bináris logit modell esetén erőssorrendben elfoglalt helyük még a különböző tudományterületek vizsgálatakor sem változott.

Második kutatási kérdés:

Budapest kiemelkedő gazdasági vonzereje gyakorol-e hatást az egyes intézmények megítélésére a jelentkezők döntésekor?

Második kutatási kérdésre adott válasz:

A hatás mindkét tudományterületen megmutatkozik. Budapest olyan gazdasági potenciállal rendelkezik, amely komoly vonzerőt jelent a többi egyetemi várossal szemben. Így Budapestet a modellekből elhagyva más preferenciák kerülnek előtérbe az egyes intézmények kiválasztásakor. Ez a jelenség a gravitációs és logit modellek esetén is egyaránt megmutatkozik.

Harmadik kutatási kérdés:

Mutatkoznak-e különbségek az eltérő tudományterületeken tovább tanulni kívánók preferenciájában a felsőoktatási intézmény(ek) megjelölésekor?

Harmadik kutatási kérdésre adott válasz:

A különböző tudomány- vagy szakterületen érdekelt hallgatók preferenciái különböznek az adott intézménybe való jelentkezésre vonatkozó döntésnél. Jó például szolgál erre a bináris logit modell alkalmazásakor a gazdaságtudományi terület esetén az intézmény megyéjének foglalkoztatási rátája. Ez a változó jóval nagyobb hatást mutat a döntési fa eredményei alapján a többi esethez viszonyítva (budapesti egyetemek figyelembevételekor). Ennek egy lehetséges oka, hogy akik gazdaságtudományi területen szeretnének tovább tanulni, azokról nagyobb valószínűséggel állíthatjuk, hogy érdeklődőbbek a gazdasági mutatók iránt, vagy akár tájékozottabbak ezen a területen. Így nem lenne meglepő, ha egy olyan jelentkező, aki tájékozottabb az egyes régiók gazdasági, munkaerő-piaci helyzetéről, nagyobb figyelmet szentelne annak a döntéskor. E különbségek miatt érdemes a tudományterületeket külön vizsgálni, ugyanis egy aggregáció érdekes és fontos tulajdonságokat összemos.

Negyedik kutatási kérdés:

Befolyásolja-e döntésének szempontjait, hogy a jelentkező hányadik helyre vonatkozó megjelölésről dönt a jelentkezéskor (abban az esetben, ha több intézményt jelöl meg)?

Negyedik kutatási kérdésre adott válasz

A kapott eredmények rávilágítanak arra, hogy minél magasabb értékű helyen vizsgáljuk a jelentkezést befolyásoló tényezőket, annál több olyan változót találunk, ami nem szignifikáns. tehát ez arra enged következtetni, hogy a jelentkezők a választáskor több szempontot mérlegelnek, amikor arról hoznak döntést, hogy melyik intézményt jelöljék meg első helyen, mint amikor például már a negyedik helyre választanak egy intézményt.

Az egyes modellek eredményeinek interpretálása után összehasonlítottam a gravitációs és logit modellek által adott eredményeket. Itt azt vizsgáltam, hogy a két modell hogyan reagál az olyan hatásokra, mint például Budapest kiemelkedő gazdasági vonzereje, vagy a különböző tudományterületenként jelentkező eltérő preferenciák. Az összehasonlítás során a következő megállapításokat tettem:

- Budapest vonzerejének kiemelkedő nagysága a gravitációs és a logit modell esetén is egyaránt megmutatkozik.
- A különböző tudományterületeken érdekeltek preferenciái között nagyobb különbségeket fedezünk fel abban az esetben, ha a térbeli pont párok közötti áramlást vizsgáljuk (gravitációs modell), mint amikor az intézmények megjelölésének sorszáma a függő változó (logit modell).
- A gravitációs modell segítségével a küldő hely változóinak hatása jobban szemléltethető (és interpretálható) mint a logit modell esetén.

E kutatás a magyarországi felsőoktatásba jelentkezők preferenciáinak elemzésére vonatkozóan a korábbi kutatásokhoz képest új változók bevonásával (intézmény megyének foglalkoztatási rátája, oktatók kiválósága alapján meghatározott rang) újszerűnek mondható.

Munkám során a kitűzött célt sikerrel teljesítettem. További kutatási irányvonalat jelentene azonban, a hallgatói preferenciák neurális hálókval történő elemzése, illetve potenciálmodell létrehozása a magyarországi felsőoktatási intézményekre vonatkozóan.

IRODALOMJEGYZÉK

Magyar nyelvű szakirodalom

- Bodon, F. (2010): Adatbányászati algoritmusok. *Free Software Foundation*, pp. 140-147.
- Dusek T. (2003): A gravitációs modell és a gravitációs törvény összehasonlítása. *Tér és Társadalom*, 17. évf. 1. sz., pp. 41-58.
- Hámori G. (2001): CHAID alapú döntési fák jellemzői, Statisztikai Szemle, 79. évfolyam, 8. szám, pp. 703-710.
- Izsó L., Kecskeméty L. (2011): Bevezetés az SPSS programrendszerbe. *Artéria Stúdió*, ISBN: 9789630811002
- Nagy G. (1996): A gravitációs modell alkalmazási lehetőségei a településen belüli mozgások tanulmányozására. *Tér és Társadalom*, 10. évf. 2. sz., pp. 149-156
- Neumanné V. I. (2010): A gravitációs modell, MEB 2010 – 8th International Conference on Management. *Enterprise and Benchmarking*, Budapest, Június 4-5.
- Neumanné V. I.: A külkereskedelem vizsgálata gravitációs modellel. „HITEL, VILÁG, STÁDIUM” Tudományos konferencia, Sopron, 2010. november 3, ISBN 978963 988373 4
- Sramó A. (1999): Adatbányászat és statisztika. *Statisztikai Szemle*, 1999, pp. 350-359.
- Telcs A., Kosztyán Zs. T., Török Á. (2013): Hallgatói preferencia sorrendek. *Rangsorok az átalakuló felsőoktatási viszonyok között konferencia*. Budapest, 2013. december 10.

Idegen nyelvű szakirodalom

- Ahmet G., Büsra K., Ferda M., Çetin D., Kübra G., Hasan Y. H. (2011): Determining the relationship between students' choice of profession and mission and vision of their high school. *Procedia Social and Behavioral Sciences*, Issue: 15, pp. 2595–2598.
- Alm, J., John V. Winters, J. V.(2009): Distance and intrastate college student migration. *Economics of Education Review*, Issue: 28, pp. 728–738.
- Asari, F. F. A. H, Idris, A R, Daud, N. M.(2011): Modelling Education Tourism Using Gravity Model in Malaysian Public Higher Education Institutions. *Australian Journal of Basic and Applied Sciences*, Issue: 5, pp. 1257-1274.
- Bruno, G., Improta, G.(2008): Using gravity models for the evaluation of new university site locations: A case study. *Computers & Operations Research*, Issue: 35., pp. 436 – 444.
- Coelli, M. B.(2011): Parental job loss and the education enrollment of youth. *Labour Economics*, Issue: 18, pp.: 25–35.
- Cohen-Zada D., Justman M.(2012): Affinity and tension between religious denominations: Evidence from private school enrolment. *Regional Science and Urban Economics*, Issue: 42., pp.:950-960.
- DesJardins, S. L., Dundar, H., Hendel, D. D.(1999): Modeling the College Application Decision Process in a Land-Grant University. *Economics of Education Review*, Issue: 18, pp. 117-132.
- Dijk, B., Fok, D., Paap, R.(2012): A Rank-Ordered Logit Model with Unobserved Heterogeneity in Ranking Capabilities. *Journal of Applied Econometrics*, Volume: 27, Issue: 5., pp.: 831–846.
- Duchessi, P., Lauría, E.J.M.(2013): Decision tree models for profiling ski resorts' promotional and advertising strategies and the impact on sales. *Expert Systems with Applications*, Issue: 40. pp. 5822–5829.
- Fayyad, U M., Shapiro, P. G., Smyth, P.(1996): From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Pres, pp. 1–34.
- Flannery, D., Kennelly, B., Doherty, E., Hynes, S., Considine, J.(2013): Of mice and pens: A discrete choice experiment on student preferences for assignment systems in economics. *International Review of Economics Education, forescoming*.
- Germeijs, V., Verschueren, K. (2007): High school students' career decision-making process: Consequences for choice implementation in higher education. *Journal of Vocational Behavior*, Issue: 70, pp. 223–241.
- Gibbons, S., Vignoles, A (2012).: Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, Issue: 42, pp. 98–113.
- Han, J., Kamber, M. (2006): Data mining: concepts and techniques (Second Edition). *Morgan Kaufmann Publisher*, ISBN: 13:978-1-55860-901-3, pp. 1-5.
- Horstschräer, J. (2012).: University rankings in action? The importance of rankings and an excellence competition for university choice of high-ability students. *Economics of Education Review*, Issue: 31, pp. 1162– 1176.
- Horstschräer, J.(2012): University rankings in action? The importance of rankings and an excellence competition for university choice of high-ability students. *Economics of Education Review*, Issue 31., pp. 1162– 1176.
- Joachim Z, Prathibha J.(2005): Academic choice behavior of high school students: economic rationale and empirical evidence. *Economics of Education Review*, Issue:24, pp. 297–308.

- Jung, J., Diane M. Hall, H., Rhoads, T.(2013): Does the availability of parental health insurance affect the college enrollment decision of young Americans? *Economics of Education Review*, Issue: 32, pp.: 49-65.
- Larose, S. Cyrenne, D., Garceau, O., Harvey, M., Guay, F., Deschênes, C.(2009): Personal and social support factors involved in students' decision to participate in formal academic mentoring. *Journal of Vocational Behavior*, Issue: 74, pp. 108-116.
- Leung, A., Hou, Z., Gati, I., Li, X.(2011): Effects of parental expectations and cultural values orientation on career decision-making difficulties of Chinese University students. *Journal of Vocational Behavior*, Issue: 78, pp. 11-20.
- Long, B. T(2004).: How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, Issue: 121, pp. 271 – 296.
- Montgomery, M.(2000): A nested logit model of the choice of a graduate business school. *Economics of Education Review*, Issue: 21., pp. 471-480.
- Montmarquette, C., Cannings, K., Mahseredjian, S.(2002): How do young people choose college majors? *Economics of Education Review*, Issue: 21., pp. 543-556.
- Niu, S. X., Tienda, M.(2008): Choosing colleges: Identifying and modeling choice sets. *Social Science Research*, Issue: 37, pp. 416–433.
- Ordovensky J. F.(1995): Effects of Institutional Attributes on Enrollment Choice: Implications for Postsecondary Vocational Education. *Economics of Education Review*, Vol. 14, Issue. 4, pp. 335-350.
- Reynolds, L. C.(2012): Where to attend? Estimating the effects of beginning college at a two-year institution. *Economics of Education Review*, Issue: 31., pp. 345-362.
- Rochat, D., Demeulemeester, J.(2011): Rational choice under unequal constraints: the example of Belgian higher education. *Economics of Education Review*, Issue: 20, pp. 15–26.
- Saisana, M., d'Hombres, B., Saltelli, A.(2011):Rickety numbers: Volatility of university rankings and policy implications. *Research Policy*, Issue: 40, pp. 165–177.
- Schwartz, B.(1985): Student Financial Aid and the College Enrollment Decision: the Effects of Public and Private Grants and Interest Subsidies. *Economics of education Rewiew*, Volume. 4, Issue. 2, pp. 129-144.
- Shtatland, E. S., Kleinman, K.,Cain, E. M.(2008): One more time about R^2 measures of fit in logistic regression,NESUG 15, Statistics, *Data Analysis & Econometrics*, pp. 222-226.
- Toutkoushian, R. K.(2001): Do parental income and educational attainment affect the initial choices of New Hampshire's college-bound students? *Economics of Education Review*, Issue: 20, pp. 245–262.
- Vrontis, D., Alkis Thrassou, A., Melanthiou, Y.(2007): A contemporary higher education student-choice model for developed countries. *Journal of Business Research*, Issue: 60, pp. 979–989.
- Weiler, W. C. (1989).: A Flexible Approach to Modelling Enrollment Choice Behavior. *Economics of Education Review*, Volume. 8., Issue. 3., pp.:227-283.
- Weiler, W. C. (1986): A Sequential Logit Modell of the Access Effects of Higher Education Institutions. *Economics of Education Review*, Volume. 5. Issue. 1., pp. 49-55.

MELLÉKLETEK

A LOGIT ÉS GRAVITÁCIÓS MODELLEK EREDMÉNYEINEK ÖSSZEFOGLALÓ ÁBRÁJA

