

.....

Róbert Péter

Nemzetiségi és etnikai csoportok társadalmi észlelése

Hiányos válaszok – hiányos eredmények

A Replika szerkesztősége ismét jó ötlettel állt elő: széles körben hozzáférhetővé tett egy (két) adatbázist, érdekes együtt-gondolkodásra, illetve elemzésre kínálva így alkalmat.

Köszönöm a szerkesztőknek, hogy az adatbázisokat tartalmazó floppyt, illetve felkérőlevelüket külön is eljuttatták hozzám, annak ellenére, hogy a témában (a nemzetiségi, illetve etnikai csoportokra vonatkozó kutatásban) semmiféle műltra nem tekintek vissza, és egyáltalán nem érzem magam szakértőnek. Ezt előre jelezniem kell, hogy a szerkesztők és az olvasók elvárásait minél inkább csökkentsem. Így a felkérőlevelben megfogalmazott kérdésekre vonatkozóan is inkább csak kósza ötleteim, s nem válaszaim vannak.

Bevallom, sosem gondolkoztam eddig azon a kérdésem, hogy "milyen módszerekkel becsülhető, illetve vizsgálható az egyes magyarországi nemzetiségi-kisebbségi csoportok létszáma?". Az a benyomásom, így lehetek ezzel a kérdéssel azok is, akik a mellékelt két adatfelvétel mintájába kerültek. Nemcsak a "nem tudom" válaszok (a létszámbebecslésre nem vállalkozók) aránya olyan magas, hogy az már megkérdőjelezi a probléma vizsgálhatóságát, hanem a 12 nemzetiségi-kisebbségi csoport többsége esetében az érvényes válaszokhoz is többnyire az átlagot meghaladó szórás tartozik. Valószínű, hogy a nagy tömegű "véleménynélküliségen" túlmenően a létszámbebecslésre vállalkozók nem csekély hányada sem gondolkodott soha a neki feltett kérdésem, rögtönzött választ adtak, amely mögött "rejtett véleménynélküliség" húzódik meg (vö. Angelusz Róbert írása¹).

Az alapvető gond az lehet, hogy az egyes nemzetiségi-kisebbségi csoportok létszámára vonatkozó kérdést a megkérdezettek nem mint e csoportok *észlelésére vonatkozó attitűdöt*, hanem mint *ismeretszint-kérdést* fogták fel. (Ismert probléma a kérdőíves adatgyűjtések-nél: a válaszolók bizonyos objektív adatokra vonatkozóan olykor nem tényszerű választ adnak, hanem aszerint felelnek, ahogy ők saját magukat tekintik az adott kérdésben, tehát attitűdjükről adnak számot. Máskor – mint itt – az attitűd-kérdés mögött vizsgáztatást sejtnek.) Ebben az összefüggésben viszont az jut eszembe, hogy vajon egy országos reprezentatív minta hány százaléka *tudja a helyes választ* arra, hogy mennyi Magyarország lakossága? Mit gondolhatnak erről azok a válaszolók, akik bizonyos nemzetiségi-kisebbségi csoportok létszámára vonatkozóan milliós nagyságrendű becslést tesznek? A létszám-becslések jelentős eltérései mögött tehát érzésem szerint az alacsony ismeretszint és a rejtett véleményélküliség kombinációja húzódik meg.

Azt, hogy a létszámbecslés milyen kevésbé képes betölteni az attitűd-kérdés funkcióját, jól jelzi a kérdés kapcsolata (pontosabban a kapcsolat *hiánya*) más kérdésekkel. Bár, hangsúlyozom, nem vagyok a terület szakértője, mégis megfogalmaztam azt a feltételezést, hogy

1. a magasabb létszámbecslés negatív attitűdöt kellene hogy jelentsen az adott csoporttal kapcsolatban: a válaszoló szerint mintegy "túl sokan vannak" az illetők;

2. azok, akik nem szeretnék, ha a gyerekiük egy csoport képviselőjével barátkozna, szintén rossz véleménnyel vannak az illetőkről.

A létszámbecslésre és a barátkozásra vonatkozó kérdéseknek tehát – páronként az egyes nemzetiségi-kisebbségi csoportok vonatkozásában – össze kellene fűggniük, a legmagasabb létszámbecslés azoktól várható, akik ellenzik, hogy a gyerekiük barátkozzon az adott csoport képviselőjével. A statisztikai elemzés (*variancia-analízis*) azonban nem mutatott az e kérdésekre adott válaszok között szignifikáns kapcsolatot.

Ahhoz azonban, hogy a nemzetiségi-kisebbségi csoportok észlelésére vonatkozó bármely kérdést egyáltalán elemezni lehessen, kezelni kell valahogy a hiányzó válaszokat. (Ezek aránya nemcsak a kérdésekre 25–60 százaléka.) Ha ezeket az eseteket kihagyjuk az elemzésből, fennáll annak a veszélye, hogy eltorzítjuk az eredetileg vélhetően reprezentatív mintánkat. Tévedés lenne ugyanis azt hinni, hogy a nem válaszolók megoszlása *véletlenszerű és nem szisztematikus*. Ha a hiányzó értékek teljesen véletlenszerűen oszlanának el (azaz MCAR típusúak lennének: missing completely at random²), lehetne alkalmazni például az átlaggal való helyettesítést (mean substitution). (Ez a módszer persze mindenképpen csökkenti a válaszok variabilitását, vagyis a jelenség szórását alulbecsüljük.)

A valóság azonban többnyire más. A létszámbecslésre vonatkozóan összehasonlítva a minta és a nem válaszolók megoszlását, az utóbbiak között statisztikailag szignifikánsan felülreprezentáltak a nők; a 65 éven felüliek (s néha az egészen fiatalok); a vidékiek; az alacsony iskolázottságúak; a segéd és betanított munkások; az alacsony keresetűek. Ha ezen kérdezettek hiányzó adatait az átlaggal helyettesítjük, lényegesen torzítjuk a mintát.

A fentebb hivatkozott SAGE kötet szerzőinek egy példája talán kínál a megoldásra analógiát. Egy vizsgálatban az embereket az őket ért anyagi jellegű károkról, illetve ezek összegéről kérdezték. A pénzüsszegrre vonatkozó adatok hiányosak voltak. A kutatók megnézték, hogy a mintába került más kérdezettek jeleztek-e ugyanolyan vagy hasonló jellegű károkat, mint amilyenekre a hiányzó válaszok vonatkoztak, s azután az ott szereplő

pénzösszegekkel helyettesítették a hiányzó értékeket. Vagyis a hasonló jellegű anyagi károkat elszenvedőkre "ráfogták" (imputation), hogy a kár összege is hasonló nagyságú lehetett.

E módszer alkalmazásának eredményeit a létszámbecslés esetére az 1. táblázat foglalja össze. Látható, hogy ha csak az érvényes esetekkel dolgozunk, a minta több mint fele elvész. Lehetséges a hiányzó értékekhez kapcsolódó szociodemográfiai ismérvek alapján kiválasztani ezen válaszok pótlására szolgáló átlagokat, de ezek a jelen esetben csupán 2–5 megkérdezett válaszára alapulnának, ami szintén meglehetősen esetleges, s ezért itt vitatható eljárás. Annyi mindenesetre a táblázat alsó felében látható, hogy ez a megoldás, illetve az átlaggal való helyettesítés eltérő eredményhez vezet.

A hivatkozott szerzők, egyéb irodalmakra is támaszkodva, más módszert is javasolnak a hiányzó értékek helyettesítésére. Ismét egy egyszerű példával élve: egy iskolás gyerekekre vonatkozó adatbázisban adatokkal rendelkezünk a gyerekek nemére, életkorára és magasságára vonatkozóan. Ez utóbbi esetében azonban sok az adathiány. Azt feltételezzük, hogy a gyerekek neme és életkora (bizonyos mértékig) meghatározza testmagasságukat, tehát megkísérrelhető a testmagasság becslése egy regressziós egyenlettel, amelyben a nem és az életkor a magyarázó változók. Ezután lehetőség nyílik arra, hogy a hiányzó értékeket helyettesítsük az érvényes válaszok alapján végzett becslésből adódó becsült értékekkel.

Hasonló eljárásra mód nyílik a nemzetiségi-kisebbségi csoportok észlelésével kapcsolatban is. Emögött persze az a feltételezés húzódik meg, hogy a jelenség kapcsolatban áll a különféle társadalmi-demográfiai ismérvekkel. Ez a gondolat Enyedi Zsolt és Fleck Zoltán cikke szerint nem alaptalan, amennyiben ők megrajzolják a cigánysághoz és a zsidókhoz fűződő viszony társadalmi profilját.³ A cigányellenesség magyarázatában a társadalmi tényezők szerepet játszanak Erős Ferenc és Fábán Zoltán tanulmánya szerint is.⁴ A szociológiai adatfelvételek szokásos változói, a nem, az életkor, a lakóhely, az iskolázottság, a foglalkozás, a jövedelem tehát valószínűleg befolyásolják a nemzetiségi-kisebbségi csoportok észlelését.

Kísérletképpen elvégeztem a hiányzó adatok pótlását a létszámbecslés esetére, ami a nemzeti-kisebbségi csoportok száma alapján 12 regressziós egyenleten alapuló becslést jelentett a következő képlet szerint:

$$N = B_1 * NEM + B_2 * KOR + B_3 * LAKÓHELY + B_4 * ISKOLA + B_5 * FOGLALKOZÁS + B_6 * JÖVEDELEM + KONSTANS,$$

ahol N egy adott nemzetiségi-kisebbségi csoport létszáma (az érvényes válaszokon alapuló becsült érték), és a B értékek a sztenderdizálatlan regressziós együtthatók. Az SPSS lehetővé teszi a kapott becsült értékek változóba írását és elmentését, s ezután a hiányzó értékek behelyettesíthetők a becsült értékekkel. A 2. táblázatban szerepelnek a létszám-becsléseknek a regressziós egyenletből adódó becsült értékei, valamint azok az értékek, amelyeket akkor kaptam, amikor a hiányzó értékeket ezekkel a becsült értékekkel helyettesítettem. A kapott átlagok, illetve szórások ismét eltérnek a korábbiaktól.

Ez az eljárás természetesen szintén felvet bizonyos problémákat. Ha a magyarázó változók sem mentesek a hiányzó értékektől (valaki nem dolgozik, nem mondja meg a keresetét stb.), nincs lehetőség becsült értékek számítására. A szakirodalom a módszert *determinisztikusként* minősíti, hiszen a figyelembe vett magyarázó változók meghatároz-

zák a feltételezett válaszokat, s az azonos ismérvekkel rendelkező személyek válaszai azonosak lesznek. Ezzel a vizsgált jelenség szórását ismét alábecsüljük. Meg kell jegyez- nem továbbá – miközben az olvasókat megkímélem attól a sok számtól, amely egy regresszió elemzés SPSS outputjában megjelenik –, hogy a 12 modell esetében a magya- rázó erő (R^2) sehol sem haladta meg a 6 százalékot, ami igen alacsony érték, illetve hogy a magyarázó változók közül rendre csak egy vagy kettő volt szignifikáns, többnyire az iskolázottság, helyenként a kor vagy a foglalkozás, egyszer-egyszer a jövedelem, illetve a lakóhely. Hivatkozhatnék persze a magyarázó változók közötti nyilvánvaló korrelációra, ami oka lehet annak, hogy ezek együtt és egyszerre nem szignifikánsak. Lehetséges lenne a becslési modellt finomítani nem lineáris, illetve interakciós hatások figyelembevételével. Egyáltalán nem biztos azonban, hogy ezzel ebben az esetben sokkal előbbre jutnánk. Szó sincs tehát arról, hogy az üdvöztető, hanem sokkal inkább a kényszermegoldások körében mozgunk.

Végezetül, hogy teljes legyen a kép, egy táblázatban összefoglaltam, hogy milyen kapcsolat áll fenn a létszámbecslés és a társadalmi-demográfiai ismérvek között (3. táblázat – nulladrendű korrelációk). Ez a táblázat megmutatja, milyen kutatási eredményre jutunk annak függvényében, hogy a hiányzó értékeket sehogyan nem, avagy valamelyik jelzett módszerrel helyettesítjük. Az eredményeket két szinten foglalhatjuk össze.

Először:

1.a. A nemi hovatartozás ritkán számít a létszámbecslésben, a legfontosabb talán, hogy a nők szerint több a cigány, mint a férfiak szerint.

1.b. Életkor szerint bizonyos nemzetiségi csoportokat, amelyekkel szemben talán kevésbé fogalmazódnak meg előítéletek (németek, szlovákok, szerb-horvátok), inkább az idősek, másokat viszont, amelyekkel szemben gyakrabban fogalmazódnak meg előítéletek (cigányok, zsidók, arabok, négerék), inkább a fiatalok érzékelnek nagyobb számban.

1.c. A lakóhely megint kevésbé számít, de azért a zsidók, az ázsiaiak, kisebb mértékben az erdélyi magyarok a városiak számára vannak többen.

1.d. A státusz: iskolázottság, foglalkozás, illetve jövedelem tekintetében általában inkább az alacsony szint valószínűsíti a magasabb létszámbecslést, kivéve talán akkor, ha szov- jetekekről, zsidókról, arabokról, esetleg négerekről van szó, ahol érdekes módon a magasabb státusz növeli a létszámbecslést.

A másik szint (hogy az előbbi eredmények tekintetében óvatosságra intsek):

2.a. Úgy tűnik, a kapott eredmények (szignifikáns vagy nem szignifikáns a kapcsolat) gyakran múlnak azon, hogyan manipuláltam az adatokat, mi történt a hiányzó értékekkel. A "legjobb" eredményeket talán akkor kaptam, amikor a regresszió elemzésen alapuló becslésekkel dolgoztam, ami érthető, hiszen a társadalmi-demográfiai hatásokat így jobban bevittem az adatokba.

2.b. Ezért talán azokat az eredményeket lehet egy kicsit komolyabban venni, amelyek minden esetben megjelentek, függetlenül attól, hogy az adatok milyen módosításon mentek keresztül.

2.c. Az egész eljárás arra a – szokásos és legegyszerűbben tesztelhető – feltételezésre épül, hogy a létszámbecslés és a szociodemográfiai ismérvek közötti összefüggés lineáris. Ez persze egyáltalán nem biztos. Érdemes lenne nem-lineáris kapcsolatokat is vizsgálni, ez azonban még további elágazásokat, illetve kombinációkat vitt volna az elemzésbe.

Mint írásom elején jeleztem, nem vagyok a téma szakértője, s ezért inkább csak egy módszertani játékkal próbálkoztam meg, amely elvégezhető lenne az adatbázisban szereplő más vélemény-kérdések esetében is.

Osszefoglalva: az a benyomásom, hogy a következőkkel kell tisztában lennünk a nemzetiségi-kisebbségi csoportok észlelésének vizsgálatakor:

1. A téma kutatása valószínűleg sokkal nagyobb körültekintést igényel, mint a más közvéleménykutatási eredmények kiértékelése. Az emberek egy jelentős részének egyszerűen nincs véleménye az adott kérdés(ek)ről, vagy a sztenderd kérdőíves adatgyűjtési módszer nem alkalmas az esetleges "spontán attitűdök" megismerésére.

2. Ha mégis a teljes népességre vonatkozóan (annak reprezentatív mintája alapján) akarunk állításokat megfogalmazni, akkor a hiányzó válaszokat valamilyen módszerrel "imputálnunk" kell. A választott módszer hatással lesz a kapott eredményekre.

3. Ezért a kutatásokat az egyes kérdésekben talán kisebb, "kompetensebb" csoportokra kell irányítani. A témában nagyobb teret kell engedni a probléma-feltáró vizsgálatoknak. Ellenkező esetben megvan a veszélye annak, hogy olyan adatbázist hozunk létre és elemzünk, amelyben a hiányos (torz) válaszok sajnálatos módon hiányos (torz) kutatási eredményekhez vezetnek.

Jegyzetek

1. Angelusz Róbert: *Kommunikáló társadalom*, Gondolat, Budapest, 1983.
2. Bourque, Linda B. – Clark, Virginia A.: *Processing Data: The Survey Example*, *SAGE Series Quantitative Applications in the Social Sciences*, No. 85., SAGE Publications, Inc. 1992.
3. Enyedi Zsolt – Fleck Zoltán: *Számháború? Beszélő*, 1995. március 2.
4. Erős Ferenc – Fábíán Zoltán: *Autoritarizmus és társadalmi tényezők a cigányellenesség magyarázatában*, 1995. (Megjelenés alatt.)

1. táblázat

Az egyes nemzetiségi-kisebbségi csoportok létszámbecslésére vonatkozó válaszok átlagai és szórásai. A hiányzó értékek helyettesítése a mintaátlaggal, illetve szociodemográfiai tényezőkön alapuló feltételezett válaszokkal.

Hány él Magyarországon	Teljes minta: a 0 értékek kihagyva a mintából (csak az érvényes esetek)			Alminta: nő, 65 év feletti, nem budapesti, 6-8 osztályt végzett, segéd- és betanított munkás, az alsó két jövedelmi kvintilisbeli		
Változó	Átlag	Szórás	N	Átlag	Szórás	N
német	286,60	474,92	401	401,33	527,64	3
szlovák	152,3	250,88	371	21,33	18,04	3
szerb-horvát	110,20	211,07	352	348,33	564,41	3
jugoszláv	59,43	107,12	328	11,00	8,49	2
román	43,97	96,11	357	8,50	9,19	2
romániai magyar	86,62	201,00	391	12,50	10,61	2
szovjet	21,28	37,53	319	6,00	1,41	2
cigány	815,59	770,13	499	800,00	685,57	5
zsidó	402,79	587,46	378	125,00	183,76	4
arab	32,1	123,09	331	2,00	1,14	2
ázsiai	42,46	118,74	353	4,00	5,20	3
néger	15,58	81,12	333	1,50	0,71	2

Hiányzó érték a

Hány él Magyarországon	mintaátlaggal helyettesítve			a feltételezett átlaggal helyettesítve		
Változó	Átlag	Szórás	N	Átlag	Szórás	N
német	286,60	303,41	981	354,43	308,62	981
szlovák	152,31	154,15	981	70,86	166,74	981
szerb-horvát	110,20	126,32	981	262,89	170,34	981
jugoszláv	59,43	61,88	981	27,19	65,96	981
román	43,97	57,92	981	21,41	60,39	981
romániai magyar	86,62	126,80	981	42,04	131,90	981
szovjet	21,28	21,40	979	10,98	22,57	979
cigány	815,59	548,99	981	807,93	549,05	981
zsidó	402,79	364,55	980	232,15	388,84	980
arab	32,11	71,43	981	12,16	72,83	981
ázsiai	42,46	71,16	981	17,84	73,52	981
néger	15,58	47,24	980	6,28	47,71	980

2. táblázat Az egyes nemzetiségi-kisebbségi csoportok létszámbecslésére vonatkozó válaszok átlagai és szórásai. (A hiányzó értékek helyettesítése egy regressziós egyenletből származó becstelt értékkel.)

Hány él Magyarországon	A regressziós egyenletből származó becstelt értékek			Hiányzó értékek helyettesítve		
	Átlag	Szórás	N	Átlag	Szórás	N
német	322,27	132,06	821	313,85	338,56	868
szlovák	95,88	28,79	821	118,42	168,22	863
szerb-horvát	127,92	40,39	821	124,27	138,85	860
jugoszláv	64,75	18,72	821	63,94	67,90	856
román	44,69	9,31	821	44,72	62,23	861
romániai magyar	90,84	24,71	821	89,59	136,22	866
szovjet	19,94	4,86	821	20,06	23,25	855
cigány	850,57	106,94	821	837,97	583,97	880
zsidó	393,48	125,69	821	396,75	398,56	860
arab	31,41	22,32	821	30,68	78,33	863
ázsiai	40,20	25,13	821	40,64	78,23	862
néger	15,02	15,28	821	14,71	51,77	863

3. táblázat Az egyes nemzetiségi-kisebbségi csoportok létszámbecslésére vonatkozó válaszok összefüggése társadalmi-demográfiai változókkal. (Korrelációs kapcsolatok a magas létszám irányába.)

	Nem	Életkor	Lakóhely	Iskola	Foglalkozás	Jövedelem
Német	n.szig.	idősek	n.szig.	alacsony	n.szig.	n.szig.
	n.szig.	idősek	n.szig.	alacsony	n.szig.	alacsony
	n.szig.	idősek	n.szig.	alacsony	alacsony	alacsony
	N.SZIG	IDŐSEK	N.SZIG	ALACSONY	ALACSONY	ALACSONY
Szlovák	n.szig.	idősek	n.szig.	alacsony	n.szig.	n.szig.
	n.szig.	idősek	n.szig.	alacsony	n.szig.	n.szig.
	férfi	n.szig.	n.szig.	n.szig.	n.szig.	n.szig.
	N.SZIG	IDŐSEK	N.SZIG	ALACSONY	ALACSONY	ALACSONY
Szerb-horvát	n.szig.	n.szig.	n.szig.	alacsony	alacsony	n.szig.
	n.szig.	n.szig.	n.szig.	alacsony	alacsony	alacsony
	nő	idősek	vidéki	alacsony	alacsony	alacsony
	N.SZIG	IDŐSEK	N.SZIG	ALACSONY	ALACSONY	ALACSONY
Jugoszláv	n.szig.	n.szig.	n.szig.	n.szig.	alacsony	n.szig.
	n.szig.	n.szig.	n.szig.	n.szig.	alacsony	n.szig.
	n.szig.	n.szig.	n.szig.	magas	n.szig.	magas
	NŐ	N.SZIG	VIDÉKI	N.SZIG	ALACSONY	N.SZIG
Romániai román	n.szig.	n.szig.	n.szig.	n.szig.	n.szig.	n.szig.
	n.szig.	n.szig.	n.szig.	n.szig.	n.szig.	n.szig.
	n.szig.	n.szig.	városi	n.szig.	n.szig.	n.szig.
	N.SZIG	FIATAL	N.SZIG	N.SZIG	ALACSONY	N.SZIG

Romániai magyar	n.szig n.szig férfi N.SZIG	n.szig n.szig n.szig IDŐSEK	n.szig n.szig városi VÁROSI	n.szig n.szig n.szig. ALACSONY	n.szig n.szig n.szig. ALACSONY	n.szig n.szig magas N.SZIG
Szovjet	n.szig n.szig férfi FÉRFI	n.szig n.szig n.szig. FIATAL	n.szig n.szig n.szig. N.SZIG	n.szig n.szig magas MAGAS	n.szig n.szig n.szig. N.SZIG	n.szig n.szig magas MAGAS
Cigány	nő nő nő NŐ	fiatal fiatal fiatal FIATAL	n.szig n.szig n.szig N.SZIG	n.szig n.szig n.szig N.SZIG	n.szig n.szig n.szig N.SZIG	n.szig n.szig n.szig N.SZIG
Zsidó	n.szig n.szig n.szig N.SZIG	fiatal fiatal fiatal FIATAL	városi városi városi VÁROSI	n.szig. n.szig. magas N.SZIG	n.szig. n.szig. magas N.SZIG	magas magas magas MAGAS
Arab	n.szig. n.szig. n.szig. N.SZIG	fiatal fiatal fiatal FIATAL	n.szig n.szig n.szig N.SZIG	n.szig n.szig magas MAGAS	n.szig n.szig n.szig MAGAS	n.szig n.szig n.szig N.SZIG
Ázsiai	n.szig n.szig n.szig NŐ	n.szig. n.szig. n.szig. FIATAL	városi városi városi VÁROSI	n.szig. n.szig. n.szig. N.SZIG	n.szig. n.szig. n.szig. N.SZIG	n.szig. n.szig. magas N.SZIG
Néger	n.szig. n.szig. n.szig. N.SZIG	fiatal fiatal fiatal FIATAL	n.szig n.szig n.szig N.SZIG	n.szig n.szig n.szig MAGAS	n.szig n.szig n.szig N.SZIG	n.szig n.szig magas MAGAS

n.szig: statisztikailag nem szignifikáns. Minden más kapcsolat legalább .05 szinten szignifikáns.

Betűtípusok: Normál: A 0 értékek kihagyva a mintából (csak az érvényes esetek). *Kurzív: A hiányzó érték a mintaátlaggal helyettesítve.* **Bold: A hiányzó érték a feltételezett átlaggal helyettesítve.** NAGY BETŰK: A HIÁNYZÓ ÉRTÉK A REGRESSZIÓS EGYENLETBŐL ADÓDÓ BECSÜLT ÉRTÉKKEL HELYETTESÍTVE.