

# A MINTAVÉTEL, A NYERS ADATOK MEGBÍZHATÓSÁGA

---

*Kapitány Balázs*

## A MINTAVÉTEL

Tanulmányunkban a Népeségtudományi Kutató Intézetben „Életünk fordulópontjai” címmel indított vizsgálat mintájának kialakítását és a mintavétel főbb jellegzetességeit ismertetjük.

A mintavételi eljárást természetesen mindig a kutatás szándékainak és a lehetőségeknek a kettőssége határozza meg. Esetünkben – a kötetet indító írásban taglalt – kutatási koncepció különös nehézségeket támasztott.

– A szükséges *elemszám* meghatározásakor egyfelől arra kellett figyelemmel lennünk, hogy a demográfiai magatartás elemezni kívánt változása szempontjából lényeges társadalmi csoportok (pl. élettársi kapcsolatban élők, fiatal egyedül élők) előfordulása megfelelő nagy számú legyen.

Másfelől azonban arra is ügyelnünk kellett, hogy a panelfelvétel háromévenként ismétlődő hullámai között az elemzést lehetővé tévő gyakorisággal következzenek be demográfiai események (pl. gyermekvállalás, házasság, otthonról való elköltözés stb.)

A fenti megfontolásokat szem előtt tartó számítások után úgy döntöttünk, hogy a sikeres interjúk szükséges száma legalább 15 000 darab, azaz a minta minimális mérete 15 000 fő legyen.

– A *paneljelleg* miatt különösen fontos csökkenteni a meghiusulásokból ill. az elutasított interjúkból származó adattorzulást, hiszen ha most még súlyozással látszólag korrigálható is az adatbázis, a minta a 3 év múlva esedékes következő fordulóra valószínűleg nagyon eltorzulna. Ráadásul a követéses vizsgálat természetéből adódóan ugyanakkor számolnunk kellett a válaszmegtagadások eleve várható magasabb arányával.

Annak érdekében, hogy a *meghiúsult* és az *elutasított* interjúk számát minimalizáljuk egyfelől kérdezőtechnikai (erről lásd a kötet adatfelvételtől szóló tanulmányát) másfelől mintatechnikai lépéseket kellett tennünk.

Az utóbbi – tehát a minta-technikai lépések – kapcsán a legfontosabb kérdés: pótoljuk-e, s ha igen, hogyan a meghiúsult és elutasított kérdőív felvételeket. Valamiféle pótlás a panelfelvétel jelzett sajátosságaiból következően elkerülhetetlen volt. A Központi Statisztikai Hivatal erre alkalmazott hagyományos módszere az ún. „pótcímkezés”. E szerint „többszörös” mintát veszünk, azaz minden mintába került személy mögött áll egy (vagy akár több) hasonló nemű, korú, lakhelyű, (és esetleg családi állapotú, végzettségű) „póttag”, aki szükség esetén helyettesíti az eredeti interjúalanyt. A módszernek azonban van egy nagy hátránya: a tapasztalatok szerint a kérdező az alternatív megoldás tudatában nem tesz meg mindent a számára valamiért gondot jelentő személyek elérése és/vagy meggyőzése érdekében. Emiatt ez a megoldás a minta igen nagy rejtett, vagyis az elsődleges kemény demográfiai változók révén fel nem tárható torzulásával járhat.

Ebből kiindulva döntöttünk úgy, hogy inkább előzetesen „torzítjuk” a címlistát, azaz korábbi mintavételek tapasztalataira alapozva a szükséges mértékben eleve felülreprezentáljuk azokat a csoportokat, amelyek esetében, magasabb arányú meghiúsulás és interjú-elutasítás várható. Így érhetjük el azt, hogy a válaszadás után a tényleges adatbázis összetétele közelebb kerüljön az alapsokaságban kialakult arányokhoz. Ezt az eljárási módot nevezik „lemorzsolódós” mintavételnek. A módszer alkalmazásának sikerességét viszont az korlátozza, hogy előzetesen nem ismerhetjük az eredményes. Előnye viszont a pótcímkezéssel szemben, hogy itt minden mintaalany helyettesíthetetlennek számít, így a kérdezők nem szűrhetik ki a nehézséget okozó válaszadókat.

Fontos azonban látni, hogy semmilyen mintavételi módszer, illetve utólagos súlyozási eljárás nem képes jóvátenni az abból következő potenciális torzítást, hogy a sikeres illetve -a bármely okból- sikertelen interjúk alanyainak szinte minden szempontból eltérhetnek egymástól. Ezért az elsődleges cél mindig a sikertelen interjúk arányának csökkentése.

*Összefoglalva a mintavétel fő jellemzőit: interjúk elutasításának és meghiúsulásának a lehetőségével számolva előzetesen torzított, régi-*

ónként, településnagyság-kategóriánként és életkori kategóriánként rétegzett többlépcsős mintát vettünk. A mintavételi egység a személy, az alapsokaság az 1926. január 1. és 1983. december 31-e között született magyar állampolgárok köre volt.

Az adatok két különböző forrásból származtak: az alapsokaságra vonatkozó adatokat a Központi Statisztikai Hivatal, a válaszadó személyek tényleges adatait a Belügyminisztérium Központi Adatfeldolgozó, Nyilvántartó és Választási Hivatala (BM KANYV) biztosította.

Az alapsokaságot a lakónépesség a 2000. január 1-i állapotra vonatkozóan becsült száma képezte, mivel a minta tervezésekor ez volt a legfrissebb információ, a 2001. évi továbbvezetett településsoros adatok ugyanis még nem álltak rendelkezésre, a 2001. évi népszámlálás előzetes eredményei pedig csupán 2001. decemberére születtek meg.

Sajnos nincsenek teljes körű adatok arról, hogy milyen a lakosság pontos korösszetétele a településeken (Ez a népszámlálásokból ismerhető meg), ezért a település teljes népességén belül a 18–75 évesek (tovább bontva a 18–29, illetve 30–75 évesek) számát becsülni kellett. Ezt a munkát a KSH 2000. IV. negyedévi és a 2001. I. negyedévi ún. ELAR felvételek adatállományából ( $N = 84\ 000$ ) a KSH Demográfiai és Módszertani Osztályán végeztél el. A nagyobb települések (Budapest és az azt követő 70 legnagyobb lélekszámú magyarországi település) mind szerepeltek az ELAR felvételekben, így ezek korstruktúrájára közvetlenül lehetett becslést adni. A többi kisebb település esetében közvetett eljárást alkalmaztak: a régiók szerinti településnagyság-kategóriák egészére határozták meg a korcsoportok becsült arányát és ezt vetítették rá egységesen az összes, abba a településnagyság-kategóriába tartozó településre.

A tapasztalatok szerint BM KANYVH által szolgáltatott személyi adatok hibás (pl. nem ott lakik a személy, nem létező a cím), így az interjú elkészítését elkerülhetetlenül megghiúsító tételeinek aránya ma az adatbázisban már 10–20 % között van. Ez igen magas hányad, de mivel a 1990. évi népszámlálás címlistájának aktualizálása nem történt meg, és a 2001. évi népszámlálás címlistái nem voltak még készen, a minta tervezéskor ezek

voltak a legmegbízhatóbb adatok Magyarországon<sup>1</sup>. Ráadásul az ebből következően megíúsult interjúk nem véletlenszerűen oszlanak el, hanem bizonyos régiókban és bizonyos társadalmi csoportokban (pl. a földrajzilag különösen mobil fiataloknál; vagy a legszegényebbeknél) lényegesebben gyakoribbak. Jelentős számú a nem teljesen hibás, de a lekérdezést igencsak megnehezítő adatok száma. Mint legjellegzetesebb példát említhetjük a nem a tartózkodási címén élők csoportjának azokat a tagjait, akikről a bejelentett tartózkodási címükön kideríthető tényleges tartózkodási helyük. Hogy ezen emberek ne vesszenek el a kutatás számára külön szabályoztuk azt, hogyan érhetőek ők is el a kérdezőink számára.

Magának a településmintának a kiválasztásakor két réteggépző ismérv alkalmazására került sor.

Az első réteggépzési szempont Magyarország regionális beosztása volt (Nyugat-Dunántúl, Közép-Dunántúl, Dél-Dunántúl, Központi régió, Észak-Magyarország, Észak-Alföld, Dél-Alföld). Az erre tekintettel levő rétegzés növeli a regionális adatok megbízhatóságát, s így megfelelő elemszám mellett előkészíti ezek jó színvonalú elemzésének a lehetőségét.

A második réteggépzési szempont a településnagyság-kategória volt. A minta tíz olyan településnagyság-kategóriára bomlik, amelyben az összlakosság közelíti az ország összlakosságának egytizedét, azaz az akkori adatok szerint a 1 004 322 főt. Fontos megjegyzés, hogy a lakosságszámával kiugró Budapest 23 kerületre bontva jelenik meg, az így lehetséges, hogy települések száma összesen: 3156.

A 6–10. kategóriába tartozik 70 önálló település és Budapest 23 kerülete, ezek ún. „önreprezentáló” települések, vagyis mindegyik a mintába került, és ezeken belül külön-külön történt a mintavétel.

Az 1–5. kategóriába azonban már 3036 település tartozik, esetükben így egyszerű szervezési és egyéb okokból sem oldható meg, hogy minden egyes településen külön mintavétel történjen. A KSH kérdezőbiztosai ugyanis szokásszerűen egy-egy településen dolgoznak csupán, ugyanis nemcsak munkaszervezési okokból kifolyólag, hanem a költségek és az adatminőség szempontjából sem szerencsés, ha 14 főnél kevesebb személyt interjúvolnak meg egy településen belül.

---

<sup>1</sup> Azóta a BM-KANYVH adatainak frissítése is megindult, az országgyűlési és önkormányzati választásokra tekintettel.

Erre gondolva maguknak a településeknek a kijelölése úgy történt, hogy a megkérdezettek száma régióként és településnagyság-kategóriánként megfeleljen a kiválasztási valószínűség arányának, vagyis mintha egy régió egy településnagyság-kategóriájának összes települése egy önreprezentáló település lenne. Végül is a kérdésbe bevont települések száma 441 (Budapestet egy településnek számítva), ebből 370 tartozik a nem önreprezentáló öt településkategóriába.

1. táblázat.  
Településnagyság-kategóriák a lakónépesség száma szerint  
(2000. január 1-i állapot)<sup>2</sup>

nagyság-kategória	településnagyság (fő)	települések száma a kategóriában (db.)	a kategóriához tartozók létszáma (fő)
1.	– 1214	1919	1 003 218
2.	1215– 2376	585	1 004 887
3.	2377– 4353	320	1 003 983
4.	4354– 9182	161	1 002 245
5.	9183– 17 816	78	993 484
6.	17 817– 32 169	42	1 007 243
7.	32 170– 63 337	21	956 499
8.	63 338– 85 877	14	1 023 488
9.	85 878–125 941	9	950 793
10.	125 942–203 648	7	1 097 384

Természetesen az általunk választott mintavételi mód esetén nemcsak a meghíúsulások, hanem a *válaszmegtagadások torzító hatását* is becsülnünk kellett. Az eddigi vizsgálatok eredményei szerint az interjú elutasításának aránya összefügg az életkorral (fiatal korban nagyobb<sup>3</sup>), a település típusával (nagyobb lélekszámú településen nagyobb), a válaszadó nemével (férfiaknál nagyobb) és sok egyéb tényezővel is.

Nekünk arra volt lehetőségünk, hogy településenként és életkori csoportokként más címkiválasztási arányt határozzunk meg, s ennek során ne csak az interjú meghíúsulásának, hanem a válasz megtagadásának gyakoriságával is kalkuláljunk.

<sup>2</sup> Forrás: Mihályffy 2001

<sup>3</sup> Természetesen ezek az összefüggések sem nem ilyen egyszerűek sem nem lineárisak.

A 2. táblázat azt mutatja, hogy egy interjú bizonyosan sikerrel járó elkészítéséhez becslésünk szerint hány címre van szükség. (A sikertelenség mindkét csoportjával számolva.)

2. táblázat.  
A mintavétel biztonsági kvóciensei

településnagyság-kategória	19-29 év közötti válaszadó	30-75 év közötti válaszadó
1.	1,44	1,2
2-3.	1,56	1,3
4-5.	1,68	1,4
6.	2,16	1,8
7-8.	2,4	1,8
9-10 <sup>5</sup> .	2,4	2

A becsült alapsokaság 7 454 196 fő.<sup>6</sup> A kiválasztási arány: 15 000 / 7 454 196, tehát megközelítőleg 0,2 %. Ha az elméletileg nyert pontos kiválasztási arányt megszorozzuk a II. táblázat értékeivel megkapjuk a településnagyság-kategóriánként és korcsoportonként szükséges tényleges kiválasztási arányt. Eszerint a felhasznált címek száma 25 510 volt, ebből 7247 tartozott a 18-29 évesek kategóriájába.

A mintakiválasztás és a kérdezés megszervezésének lépései arra irányultak, hogy a kutatás eredményei a lehető legteljesebb mértékben megbízhatóak legyenek még az olyan társadalmi alcsoportokra nézve is, amelyek a kérdőíves adatfelvételekkel általában nehezen közelíthetők meg. Ezt szolgálta például annak kontrollálása, hogy a nem ön-

<sup>4</sup> Ebbe a csoportba soroltuk Budapest néhány kis lélekszámú belső kerületét is, ahol a sikeresen felvett kérdőívek aránya hagyományosan alacsony (I., VI., VIII., XII. kerület.), az általában gondot okozó többi kerület magas lakosság száma miatt eredendően ide tartozott.

<sup>5</sup> Ebbe a csoportba soroltuk Budapest néhány kis lélekszámú belső kerületét is, ahol a sikeresen felvett kérdőívek aránya hagyományosan alacsony (I., VI., VIII., XII. kerület.), az általában gondot okozó többi kerület magas lakosság száma miatt eredendően ide tartozott.

<sup>6</sup> A 2001. évi népszámlálás eredményei szerint az alapsokaság 7 618 280 fő. A különbség a továbbvezetett illetve a népszámlálás során nyert adatok mintegy 200 000 fős eltérésének az alapsokaságra vetített része.

reprezentáló települések közül a jelentős roma lakossággal rendelkezők arányuknak megfelelően kerüljenek a mintatelepülések sorába.<sup>7</sup>

Úgynevezett címkövetéssel elértük, hogy legalább részben felleljük azokat a mintaalanysokat, akik nem a bejelentett lakcímükön laknak. Ennek révén reményeink szerint az átlagnál valamivel jobb arányban sikerült bevonnunk a vizsgálatba a fiatal mobil rétegeket. (A kérdezéstechnikai részleteket jelen kötetünk egy másik tanulmánya ismerteti.)

A fenti megfontolásokat követve 25 510 címet kértünk a BM Központi Adatnyilvántartó Hivatalából, abban bízva, hogy becslésünknek megfelelően legalább 15 200 sikeres interjúhoz jutunk.

#### A NYERS ADATOK MEGBÍZHATÓSÁGA

A tényleges adatok a következőképpen alakultak: A kért 25 510 cím 95,7% -a (24 417 cím) került a kérdezőbiztosoknak kiosztásra, a címkártyát 24 138 főről (94,7%) töltötték ki. A meghíúsult és az elutasított interjúk aránya az összes címhez képest 30,4%, az elkészült interjúké 64,3%, 16 394 darab. (Tehát a sikeres interjúk aránya a kitöltött címkártyák – a kérdezőbiztosok által felkeresett címek 67,9%-a.) Az eredmény értelmezése során számításba kell venni a nehezebben elérhető csoportok irányában történt előzetes torzítást, ami már önmagában „nehezítette” a lekérdezést.

A sikeres 16 394 interjú (noha a felhasználható adatok száma az ellenőrzés nyomán még valamelyest csökkent) olyan arányt képviselt, amely megfelelt várakozásainknak, sőt csekély mértékben felül is múlta azokat. Összesen négy megyében nem sikerült teljesíteni az előzetes becslések alapján kalkulált szintet. A legnagyobb elmaradás Borsod-Abaúj-Zemplén megyében mutatkozott, ahol a várt 1338-nál 92-vel kevesebb interjú készült el. Mind abszolút mértékben, mind arányában extrém módon túlteljesített Fejér megye, ahol a kalkulált 589-nél 213-mal több sikeres kérdőívfelvétel történt. Fontos azonban felhívni a figyelmet arra, hogy ezek a különbségek nem csupán a megyék egyéni teljesítményével, hanem a településnagyságtól és a fiatalok arányától függően kalkulált elvárási szintek eltéréseivel is magyarázhatók.

---

<sup>7</sup> A roma lakosság arányának becslését Kertesi-Kézdi (1998) alapján végeztük el.

Az igazán lényeges kérdés azonban az, hogy mennyire volt sikeres a kiválasztott címek előzetes torzításának a becslése, vagyis a kapott eredmények mennyire közelítik meg a tényleges eloszlást. Ennek a tisztázására leginkább a teljes adatbázis és a 2001-es népszámlálás végleges adatainak összevetése alkalmas. Első közelítésben logikusabbnak tűnne, ha a mintavételi alapsokaság (2001. január 1.) eloszlását vetnénk össze a kapott adatokkal. De ha így járnánk el, akkor egy technikai kényszerből elfogadott bázishoz hasonlítanánk az adatainkat, amelyekről közben kiderül hogy nem is pontos teljesen. További hátránnyal járna, hogy ebben az esetben le kellene mondanunk olyan jellemzők (például az iskolai végzettség) egybevetéséről, amelyekre nincs továbbvezetett adat. Emiatt a pontosabbnak tekinthető, és a vizsgálatunk eszmei időpontjához lényegesen közelebb eső népszámlálást vesszük alapul, amikor a kapott eredményeink nyers, súlyozatlan eloszlásait értelmezzük. Természetesen így sem kerülhető el némi pontatlanság, mivel a népszámlálás és a mi adatfelvételünk ideje között csaknem egy év telt el, és ezalatt valamelyest változott az alapsokaság eloszlása.

A nemek szerinti eloszlás az adatbázisban 45,4% férfi, 54,6% nő. Noha a férfiak magasabb halandósága folytán az alapsokaságban is a nők vannak többen (52,4 %), realizált mintánkban ez az arány tovább torzult amiatt, hogy a férfiakat nehezebb mind elérni, mind szóra bírni.

A korcsoportok szerinti összevetést a következő eredményeket adja. (3. táblázat)

3. táblázat.  
Generációk szerinti összetétel az adatbázisban

	1926–1941	1942–1951	1952–1961	1962–1971	1972–1983	összesen
népszámlálás 2001	1 574 583	1 346 282	1 510 999	1 323 724	1 862 692	7 618 280
népszámlálás %	20,7 %	17,7 %	19,8 %	17,4 %	24,5 %	100 %
Életünk fordulópontjai	3495	2882	3091	2605	4321	16 394
Életünk fordulópontjai %	21,3 %	17,6 %	18,9 %	15,9 %	26,4 %	100 %

Feltűnő, hogy mintavétel során szándékoltn felülreprezentált fiatalok aránya elérte, sőt csaknem két százalékkal meghaladta a népszámlálás szerinti arányukat. A többletet azonban nem egyenletesen ellentételezi a többi korcsoportban mutatkozó hiány, hanem ez a fiatal közép-

korúaknál összpontosul. Úgy látszik a mintavétel során- bár csekély mértékben ennek a – némi pontatlansággal 30 és 40év közötti – korcsoportnak a felülreprezentálása is indokolt lett volna.

A nem és az életkor szerinti összetételt együttesen tekintve megállapíthatjuk, hogy az adatbázisban a fiatal középkorú férfiak vannak leginkább alulreprezentálva, hiszen míg ideális esetben e csoport aránya 8,7% lenne, a mintánkban csupán 7,4% .

Lehetőségünk van az alapsokaság és a nyers adatbázis a családi állapot, illetve a kor és családi állapot szerinti összevetésére is. A válaszadók családi állapot szerint a következőképpen oszlottak meg: nőtlen, hajadon 25,2%, házas 56,7%, elvált 9%, özvegy 9,1%. A népszámlálási eloszlások a következők: nőtlen, hajadon 25,7%, házas 55,7%, elvált 9,9%, özvegy 8,7%. Az arányokat tekintve tehát a legkevésbé az elváltakat sikerült elérni és válaszra bírni, de a kategóriára vetített eltérés itt is csak 10 százaléknyi.

A családi állapot korcsoportos mintázatának a torzulásában szerencsére nincs olyan koncentráció, mint amilyent a fiatal középkorú férfiak képeztek.

Az iskolai végzettséget tekintve a minta 13,5%-a rendelkezett felsőfokú képesítéssel, míg ugyanaz az arány a népszámlálás esetében 12,6% volt. Az érettségizettek hányadát 31,9%-osnak mértük, az alapsokaság 28,6%-ával szemben. Az eltérés egyik oka az lehet, hogy a népszámlálás eszmei időpontja és az adatfelvétel között lezárult egy tanév, a minta tagjai közül tehát azóta sokan leérettségizhettek, illetve diplomát szerezhettek. (ezzel a problémával, és a megoldásával a súlyozásról szóló tanulmány foglalkozik részletesen). Ugyanakkor ezen felül az is felmerülhet, hogy a magas iskolai végzettségű rétegeket valamivel nagyobb arányban sikerült szóra bírni. Mindazonáltal ez a közelítés – összehasonlítva egyéb vizsgálatok eredményeivel – meglehetősen jónak mondható.

Az iskolai végzettséget és korcsoportot együtt tekintve, tovább pontosítható az előbbi kép. Megfigyelhető, hogy a nyers adatbázis a felsőfokúak arányát minden korcsoportra nézve elég pontosan képezi le, eltolódás az érettségizett, nem érettségizett kategóriák között mutatható ki, az előbbi javára. Ezt a torzulást felerősíti az a körülmény, hogy középkorúak általában is kevesebben vannak a mintában, így aztán míg

a 30 és 50 év közötti nem érettségizett lakosság az alapsokaság mintegy 21%-ára rúg, a nyers adatbázisban csupán 17,9%-ot tesz ki.

Természetesen nincs lehetőség egyenként áttekinteni valamennyi illeszkedési szempont minden kombinációját, és ha meg is tennénk, akkor sem állíthatnánk teljes bizonyossággal, hogy az össze nem vethető szempontok szerint is ilyen jó az illeszkedés. Egy rész minta vizsgálata azonban még elengedhetetlen, ez pedig a fővárosi populáció.

Ismert (pl. Waffenschmidtné, 2001), és a rendszerváltás óta egyre súlyosabbá vált az a probléma, hogy Budapesten (főleg annak is néhány belső, illetve hegyvidéki kerületében) kórosan magas mind a válaszmegtagadások, mind a nem bejelentett címükön lakók aránya. Ez a körülmény súlyosan veszélyezteti a kérdőíves felvételek budapesti adatainak megbízhatóságát. A „lemorzsolódós” mintavétel tervezésekor ezért Budapesten számoltunk a legalacsonyabb sikerességi aránnyal. Becslésünk többé-kevésbé sikeresnek is bizonyult, így a fővárosiak aránya elég pontosan megfelelt az elvárásainknak, (a kalkulált 18% helyett 17,5% fővárosi került a nyers mintába). Ennek ellenére azonban kérdéses a belső arányok pontossága.

A 4. táblázat azt foglalja össze, hogy a budapesti lakosság egyfelől a népszámlálás, másfelől a nyers adatbázis alapján hogyan oszlik meg nem, életkor és iskolai végzettség szerint.

Az összehasonlításból kiderül, hogy az alapsokaság és a nyers adatbázis illeszkedése a nők esetén pontosabb, mint a férfiakéban. Az előbbieknél csak két csoportja mutat számottevő eltérést: az idős, felsőfokú végzettségű fővárosi nőből az alapsokasági arányukhoz képest lényegesen több, fiatal nem érettségizettből pedig kevesebb került a mintába. A férfiakat illetően szintén a fiatal iskolázatlanok kategóriájában hiány, többlet pedig, a nem középkorú, nem iskolázatlan csoportoknál jelentkezik. Az alulreprezentáltságot részben ismét az indokolhatja, hogy a népszámlálás és az adatfelvétel eszmei időpontja között eltelt időben emelkedett az alapsokaság iskolai végzettsége (például egy kohorsz leérettségizett).

Összességében megállapítható, hogy nyers adatok megfelelően közelítik az alapsokaság eloszlását, súlyozás segítségével a főbb dimenziókban a peremeloszlások pontos illesztése különösebb problémák nélkül elérhető. Az előzetesen torzított „lemorzsolódó” mintavétel bevált, a minta az alapsokaság vizsgált szempontok szerinti eloszlását megfelelő pontossággal leképezte.

4. táblázat.  
A budapestiek generációk és képzettség szerinti megoszlása nemenként

Nem	Végzettség	Adatbázis	1926-1941	1942-1951	1952-1961	1962-1971	1972-1983	össz.
férfi	felsőfok	Népszámlálás %	2,6	2,5	2,2	2,1	1,5	11,0
		Életünk fordulópontjai %	3,5	3,3	2,7	2,2	3,9	15,6
	érettségi	Népszámlálás %	2,1	2,6	2,5	2,8	5,8	15,7
		Életünk fordulópontjai %	2,9	2,3	2,2	2,5	6,9	16,9
	érettséginnél alacsonyabb	Népszámlálás %	3,8	3,3	3,4	3,4	5,6	19,5
		Életünk fordulópontjai %	3,5	3,3	2,7	2,2	3,9	15,6
nő	felsőfok	Népszámlálás %	1,7	2,4	2,6	2,5	2,0	11,1
		Életünk fordulópontjai %	2,6	3,0	3,4	2,8	2,4	14,3
	érettségi	Népszámlálás %	3,8	4,6	4,1	3,7	6,8	23,0
		Életünk fordulópontjai %	4,2	4,7	3,9	3,8	7,4	23,9
	Érettséginnél alacsonyabb	Népszámlálás %	7,2	3,4	2,8	2,3	4,0	19,7
		Életünk fordulópontjai %	7,4	3,6	2,7	1,4	2,9	18,0
össz.		Népszámlálás %	21,1	18,8	17,7	16,7	25,7	100,0
		Életünk fordulópontjai %	23,6	20,0	16,8	14,4	25,1	100,0

HIVATKOZOTT IRODALOM

- Kertesi Gábor-Kézdi Gábor (1998): A cigány népesség Magyarországon. Dokumentáció és adattár. Budapest. Socio-typo.  
 Mihályffy László (é.n.): A Demográfiai Panel mintájának terve. Kézirat.  
 Waffenschmidt Jánosné (2001): Adatgyűjtés és az adatok minősége. Statisztikai Szemle, 2001/9: 741-751.