

KÉT CIGÁNYVIZSGÁLAT*

(Kritikai elemzés)

„Nem tudom, de nekem
ez az adat gyanús!”

1993 őszén, több mint húsz évvel az 1971. évi cigányvizsgálat után, egymással párhuzamosan két nagyobb adatfelvétel készült hazánkban a cigányság helyzetéről. Az egyik felvétel a Központi Statisztikai Hivatalban készült, a másikat az MTA Szociológiai Intézet munkatársai és az MTA Közgazdaságtudományi Intézet részéről jelen tanulmány szerzője készítették közösen¹. A KSH-felvétel mintájának alapja az Egységes Lakossági Adatfelvételi Rendszer (ELAR) mintája volt, amelybe – a rendszeres negyedéves munkaerő-felvételek 1993 őszi hullámának keretében – első ízben bekerült egy a megkérdezett háztartások cigány etnikai hovatartozására vonatkozó kérdés is. A KSH felvételéről – amely a későbbiekben *nemzetiségi felvételként* került be a köztudatba – a hivatal munkatársai 1994 nyarán rövid beszámolót adtak közre *A cigányság helyzete* címmel (Mészáros et al. 1994). Az Akadémia adatfelvételéről (a továbbiakban: Cigányvizsgálat) – mely a KSH nemzetiségi felvételétől eltérően célzottan a cigányság és csakis a cigányság felmérésére törekedett – szintén 1994-ben készült egy rövid összefoglaló jelentés (Kemény–Havas–Kertesi 1994).

Már az első összehasonlításokból kiderült, hogy az adatok között számos tekintetben lényeges eltérések vannak. Ezeknek az eltéréseknek az alaposabb elemzésére az adott alkalmat, hogy a Közgazdaságtudományi Intézet megvásárolta a KSH-tól az ELAR-mintán lekérdezett 1993. szeptember–novemberi munkaerő-felvétel és a hozzá kapcsolódó nemzetiségi felvétel teljes adatállományát. Az itt következő tanulmány abból a célból született, hogy részletesen beszámoljon a két adatfelvétel eredményeinek különbségeiről és e különbségek valószínű okairól. Érvelésemet a lehető legnagyobb részletességgel igyekszem dokumentálni. Az alábbi észrevételek kettős célt szolgálnak: egyrészt *ellenőrizhető* okfejtések segítségével igyekeznek meggyőzni az olvasót a szerző álláspontjáról (a Szociológiai Intézet adatfelvételének nagyobb megbízhatóságáról), másrészt igyekeznek eloszlatni a statisztikai adatfelvételekkel kapcsolatos igen

* Köszönettel tartozom Kőrösi Gábornak statisztikai jellegű tanácsaiért, valamint kutatási asszisztenseimnek – Ábrahám Árpádnak és Kézdi Gábornak – a számításokban nyújtott segítségükért és tanácsaikért.

általános bizalmatlanságot. Meggyőződésem, hogy a statisztikai adatokkal kapcsolatban *van helye a racionális diskurzusnak*. Ez az írás egy konkrét példán igyekszik bemutatni azt a sokféle – logikai és empirikus – lehetőséget, amely a kutató rendelkezésére áll, ha elfogja a kétely, vajon megbízhat-e ebben vagy abban az adatban.

1. A két minta belső eltérései

Mindenekelőtt a legalapvetőbb eltéréseket kell bemutatnunk. Sejtésünket, mely szerint a két minta – fontos társadalmi indikátorokat tekintve – jelentős mértékben különbözik egymástól, homogenitás-vizsgálatok segítségével próbáljuk igazolni. A társadalomstatisztikai gyakorlatban ilyen alapvető változónak számít a háztartásnagyság (vagy egy analóg jelentésű indikátor: a háztartásban élő, eltartott gyermekek száma), a lakóhely, az életkor, a nem és az iskolai végzettség. Az alábbiakban megvizsgáljuk, hogy az ELAR munkaerő-felvétel cigány almintája és a Szociológiai Intézet cigány mintája statisztikailag azonos populációt reprezentálnak-e a fent említett változók szerinti megoszlások alapján. A nullhipotézis az, hogy azonos populációt reprezentálnak. Vajon a homogenitás-vizsgálatok alapján el kell-e vetnünk a nullhipotézist?² A következő oldalakon közölt (1–6.) táblázatokban rendre elvégezzük az említett próbát.

1. táblázat

A két minta összehasonlítása: a háztartások megoszlása az eltartott gyerekek száma szerint (százalék)*

gyerekszám (fő)	ELAR súlyozatlan	CIGV	ELAR súlyozott	CIGV
0	42,12	26,40	39,82	26,40
1	18,28	22,10	23,19	22,10
2	20,00	23,40	23,46	23,40
3	14,17	16,20	11,39	16,20
4+	5,43	11,90	2,14	11,90
összesen	100,00	100,00	100,00	100,00
(esetszám)	(755)	(1931)	(120107)	(1931)
chi2(4)	74,603		923,423	
valószínűség	100%		100%	
nullhipotézis	elvetve		elvetve	
átlag	1,265	1,742	1,140	1,742
szórás	1,387	1,553	1,162	1,553

* Eltartott gyerek = ha: $KOR \leq 14$ vagy ($KOR \geq 15$ és nappali tagozatos tanuló)

2. táblázat

A két minta összehasonlítása: a háztartások megoszlása a háztartás létszáma szerint (százalék)

háztartáslétszám (fő)	ELAR súlyozatlan	CIGV	ELAR súlyozott	CIGV
1	14,30	6,90	14,04	6,90
2	20,79	12,00	20,21	12,00
3	16,03	15,60	20,87	15,60
4	21,72	22,20	24,12	22,20
5	14,97	18,60	13,58	18,60
6	7,02	10,50	4,89	10,50
7+	5,17	14,30	2,29	14,30
összesen (esetszám)	100,00 (755)	100,00 (1931)	100,00 (120107)	100,00 (1931)
chi2(6)		111,536		1411,953
valószínűség		100%		100%
nullhipotézis		elvetve		elvetve
átlag	3,483	4,426	3,278	4,426
szórás	1,811	2,228	1,541	2,228

3. táblázat

A két minta összehasonlítása: településtípus szerinti megoszlás (százalék)

településtípus	ELAR súlyozatlan	CIGV	ELAR súlyozott	CIGV
Budapest	5,09	9,60	10,49	9,60
város	29,99	30,00	30,29	30,00
község	64,93	60,40	59,21	60,40
összesen (esetszám)	100,00 (747)	100,00 (1931)	100,00 (118024)	100,00 (1931)
chi2(2)		15,107		1,846
valószínűség		100%		60%
nullhipotézis		elvetve		elfogadva

4. táblázat

A két minta összehasonlítása: az egyének megoszlása nemek szerint (százalék)

nem	ELAR súlyozatlan	CIGV	ELAR súlyozott	CIGV
férfi	49,81	49,39	49,95	49,39
nő	50,19	50,61	50,05	50,61
összesen (esetszám)	100,00 (2630)	100,00 (8537)	100,00 (393715)	100,00 (8537)
chi2(1)		0,145		1,052
valószínűség		30%		69%
nullhipotézis		elfogadva		elfogadva

5. táblázat

A két minta összehasonlítása: az egyének megoszlása korcsoportok szerint (százalék)

korcsoport (év)	ELAR súlyozatlan	CIGV	ELAR súlyozott	CIGV
–14	31,41	37,30	30,46	37,30
15–19	10,87	11,80	9,99	11,80
20–24	7,30	8,70	9,33	8,70
25–29	7,83	7,40	8,23	7,40
30–39	16,16	15,70	16,04	15,70
40–54	13,19	12,20	12,81	12,20
55–59	4,11	2,30	4,58	2,30
60–69	5,93	3,40	5,62	3,40
70–74	1,56	0,70	1,20	0,70
75+	1,63	0,50	1,74	0,50
összesen (esetszám)	100,00 (2630)	100,00 (8517)	100,00 (393715)	100,00 (8517)
chi2(9) valószínűség nullhipotézis	133,066 100% elvetve		426,633 100% elvetve	
átlag szórás	28,04 19,98	23,79 17,83	28,01 19,83	23,79 17,83

6. táblázat

A két minta összehasonlítása: a személyek megoszlása iskolai végzettség szerint (százalék)*

iskolai végzettség **	ELAR súlyozatlan	CIGV	ELAR súlyozott	CIGV
0 osztály	5,45	9,10	5,37	9,10
1–7 osztály	30,21	32,90	27,08	32,90
8 osztály	44,85	43,50	45,51	43,50
szakmunkásképző	11,95	12,20	12,99	12,20
középiskola	5,70	2,00	6,98	2,00
felsőfokú	1,84	0,30	2,09	0,30
összesen (esetszám)	100,00 (1632)	100,00 (4910)	100,00 (249917)	100,00 (4910)
chi2(5) valószínűség nullhipotézis	123,588 100% elvetve		439,958 100% elvetve	

* A 15–74 éves népességre (nappali tagozaton tanulók nélkül).

** Szakmunkásképző: szakiskolával együtt. Középiskola: gimnázium, technikum, szakközépiskola, polgári. Felsőfokú: egyetem, főiskola.

Mielőtt a próbák értékelésébe fognánk, egy előzetes megjegyzést kell tennünk. A KSH az ELAR-minta *egészének* reprezentativitását – mind háztartás, mind pedig egyéni szinten – átsúlyozási eljárás alkalmazásával biztosította.³ Erről a súlyozási eljárásról tanulmányunkban még sok szó esik. Itt most elegendő annyit megjegyeznünk, hogy mivel a súlyozás erőteljesen befolyásolja a próbák eloszlását, ezért a homogenitás-vizsgálatokat súlyozatlan és súlyozott adatokra egyaránt elvégezzük. Lássuk tehát a próbák eredményeit.

Az 1–6. táblázatok alapján egyértelműen megállapíthatjuk: a *nemek* szerinti (4. táblázat), illetve a *lakóhely* szerinti (3. táblázat) megoszlások *kivételével valamennyi változó esetében szignifikáns különbség van a két minta között*. A háztartáslétszámban, az eltartott gyerekek számában, a koreloszlásban és az iskolai végzettségben (1., 2., 5., 6. táblázat) megjelenő – súlyozástól függetlenül – magas próbaértékek egyértelmű bizonyítékát nyújtják annak, amiről a megoszlások alapján szabad szemmel is bárki könnyen meggyőződhet: a két minta belső szerkezete jellegzetesen eltér egymástól. A Cigányvizsgálat megoszlásaihoz képest az *ELAR-minta cigány almintájában lényegesen kisebb a háztartások mérete, illetve alacsonyabb a háztartásokban élő eltartott gyerekek átlagos száma; lényegesen nagyobb arányban szerepelnek az idősebb korcsoportok és lényegesen kisebb arányban a fiatalabb korcsoportok*; végül pedig az ELAR-felvételben szereplő cigány népesség *lényegesen iskolázottabb, mint a Cigányvizsgálaté*.

Nyilvánvaló, hogy nem mehetünk el szó nélkül ilyen mértékű strukturális különbségek mellett. A további fejezetekben megpróbálunk utána járni ezeknek az eltéréseknek. Először a mintavételi eljárások különbségeinek betudható eltéréseket próbáljuk meg számszerűsíteni; majd a megghiúsult kikérdezések eltérő pótlásának következményeit igyekszünk fölmérni; végül pedig az átlagos gyermekszám, illetve az iskolai végzettség adatok realitását egy külső adatforrás – a Művelődési Minisztérium iskolastatisztikái – alapján ellenőrizzük.

2. A mintavételi eljárások különbségeinek következményei

Vajon a két minta közötti, imént feltárt különbségek milyen mértékben vezethetők vissza a két adatfelvétel során alkalmazott mintavételi eljárások különbségeire? Mielőtt azonban e különbségek *következményeit* megpróbálnánk fölmérni, röviden ismertetnünk kell a két adatfelvétel során alkalmazott mintavételi eljárások elveit.

Az ELAR-felvétel mintavételi eljárása

A mintavétel ismertetésére a leghitelesebb forrást, magukat a minta készítőit hívjuk segítségül. Az alábbi három bekezdés idézet az ELAR munkaerő-felmérés 1993. évi adatait tartalmazó KSH-kiadványból:

„A munkaerő-felmérés mintája többletcsős, rétegzett minta. A 15000 vagy ennél több főt számláló településeken a népszámlálási számlálókörzetek voltak az elsődleges és a lakások a másodlagos mintavételi egységek, míg a kisebb településeken az elsődleges, másodlagos, illetve utolsó mintavételi egységek rendre a települések, a számlálókörzetek, illetve a lakások voltak. Az 1993 elején végrehajtott mintacsökkentés után a minta 8272 népszámlálási számláló-

körzetet tartalmaz az ország 555 településén, ami az 1990. január elsejei állapot szerint az összes település mintegy 18 százalékának felel meg.

A munkaerő-felmérés mintája egyszerű véletlen kiválasztással jött létre a címregiszterből, amely – az említett mintacsökkenés óta – a 8272 körzet együttes címanyagából, azaz összesen mintegy 460000 címből áll. A munkaerő-felmérés negyedéves mintája három-három címet tartalmaz minden egyes mintakörzetből, egy-egy hónapban azonban a mintakörzeteknek csupán egy-harmadát keresik fel az összeírók. A munkaerő-felmérés mintája alapján véve lakásminta; ez azt jelenti, hogy a mintába került lakásokban minden háztartásról és minden 15–74 év közötti életkorú személyről van gazdasági aktivitásukra vonatkozó információ. 1993-ban a 3*8272 cím általában mintegy 27000 háztartást és 47000 személyt eredményezett a munkaerő-felmérés negyedéves mintájában. A mintában alkalmazott egyszerű rotációs eljárás szerint bármely háztartás, amely valamilyen időpontban belép a mintába, hat egymást követő negyedévben szolgáltat adatokat, majd végleg elhagyja a mintát. Teljes háztartásnál történt meghíúsulás esetén – körzetenként legfeljebb kétszer – szabad pótcímeket használni, ezért két egymást követő időszak mintájának a közös része általában kisebb, mint 5/6.

A munkaerő-felmérés mintájának kialakításánál a rétegeképző ismérvek a következők voltak: földrajzi egységek, települések nagysága szerinti kategóriák, továbbá lakóövezetek; ez utóbbiakra példák a (történelmi) városközpontok, külvárosok, lakótelepek stb. Az értékösszegek a munkaerő-felmérés mintájából teljeskörűsítéssel kerülnek becslésre: a mintában megfigyelt értékek alkalmas mintasúlyal történő szorzása és összegezése után. A jelenlegi mintavételi terv szerint – az 1993. évi mintacsökkentés következtében – a minta 114 utólagos rétegre vonatkozóan kell mintasúlyokat meghatározni, éspedig minden egyes utólagos rétegben a teljes körű lakásszám és a mintában megfigyelt lakások számának hányadosaként. A mintasúlyok képzésénél csak a lakott lakások vehetők figyelembe, viszont az építésügyi statisztikák az üres lakásokat is tartalmazzák. Ezért a lakott lakások számát a teljes körű lakásszámokon belül mintainformáció alapján becsülni kell. A teljes mintához tartozó mintavételi arány lakásszámokra vonatkoztatva 0,66 százalék. A nem-válaszolásból származó torzítás ellensúlyozására az adatok a továbbvezetett népességszámokkal is korrigálásra kerültek” (*Munkaerő-felmérés* 1994: 8–9).

Idáig az idézet. Ehhez hozzá kell tennünk a következőket. A *szóban forgó* ELAR-adatfelvétel úgy állt elő, hogy az 1993. évi III. negyedéves hullámból a szeptemberi, a IV. negyedéves hullámból pedig az októberi és novemberi megkérdezetteket vonták be. Továbbá: a szeptemberi-októberi-novemberi minta összesen 26604 felkeresett háztartásból állt, ebből 22418 esetben lett sikeres kikérdezés. A meghíúsult kikérdezések száma 4186 volt. A *cigányság helyzete* című KSH-kiadvány minden adata ennek a 22418 háztartásnak, illetve e háztartásokban lakó egyéneknek a súlyozott ismérveiből állt elő.

Ami a cigány al minta reprezentativitását illeti, a kutatási stratégia világos. Mivel a minta területileg (településszinten és annál kisebb aggregációs szinten is) jól lefedi az ország egészét, továbbá kellően nagy elemszámú, arra számíthatunk, hogy a népesség viszonylag kis hányadát kitevő kisebbség esetében is kellő elemszámú, reprezentatív mintát biztosít. Képletesen szólva: a sűrű szövésű háló a kis halakat is befogja. A rövidség kedvéért nevezzük el azt az elvet, amelyen a szóban forgó vizsgálat mintavételi eljárása nyugszik, a „*sűrű szövésű háló*” elvének.

A Cigányvizsgálat mintavételi eljárása

A Szociológiai Intézet cigányfelvételének – mely a dolog természetéből adódóan, kizárólag az adott kisebbség felmérésére irányult – ettől gyökeresen eltérő mintavételi stratégiát kellett követnie. A legnagyobb problémát természetesen az jelentette, hogy sem a kérdéses sokaság nagyságáról, sem térbeli eloszlásáról nem áll rendelkezésre megbízható, teljes körű, népszámlálási információ. Ennek hiányában a hagyományos mintavételi eljárástól eltérő utat kellett követnünk. Az eljárás alapelve a következő volt: az általános iskolai cigány tanulók száma alapján kiválasztott mintavételi körzetekben összeírtuk a fellelhető összes cigány háztartást, majd belőlük választottunk véletlen mintát. A rövidség kedvéért nevezzük el azt az elvet a „*célzott vizsgálat*” elvének. A mintavétel és a kikérdezés 1993 októbere és 1994 februárja között történt. A mintavételi eljárásról részletesebben a következőket mondhatjuk.

A minta a magyarországi cigány háztartások 2 százalékát fogja át. A háztartásokat többlépcsős, rétegzett mintavétel útján választottuk ki. Első lépésben megkülönböztettük egymástól a községeket és a városokat. A budapesti és miskolci mintavétel külön eljárás szerint történt.

Ami a községeket illeti, a kiindulópontunk a Művelődésügyi Minisztérium 1989/90. tanévi általános iskolai statisztikái voltak, amelyek általános iskolai szinten (az ország kb. 3600 általános iskolájára) tartalmaztak *teljes körű* cigány tanulószámadatokat. Ezeket az adatokat először községi szintre aggregáltuk. Majd, hogy összehasonlítható adatokhoz juthassunk – hiszen a községek mintegy egyharmadában nincs általános iskola –, olyan községbokrokat hoztunk létre, ahol van legalább egy teljes (alsó és felső tagozatot egyaránt tartalmazó) általános iskola, s amelyekre igaz az, hogy az adott községbokorhoz tartozó iskola vagy felső tagozatos iskola nélküli településekről kizárólag az adott bokorhoz tartozó más települések általános iskoláiba járnak be tanulni az általános iskolás gyerekek. Ezekről a településbokrokról föltételeztük, hogy az általános iskolás cigány tanulókra vonatkozó létszám adataik megfelelően tükrözik a teljes cigány népesség eloszlását. Ezt követően a szóban forgó községbokrok cigány tanuló létszám adatai alapján osztályoztuk az említett község-

bokrokat: megkülönböztettünk cigányok által *ritkán*, *közepesen*, illetve *sűrűn* lakott településbokrokat. Azokat a községbokrokat, ahol egyetlen cigány tanuló sem fordult elő, kihagytuk a mintából. Ezután pedig eltérő kulcsok szerint mintát vettünk az érintett községekből: takarékosági okokból a cigányok által sűrűbben lakott településekből relatíve többet, a ritkábban lakottakból relatíve kevesebbet.

A megyeszékhelyek mindegyike belekerült a mintába, a városok közül véletlenszerűen minden ötödik. A városi településkategóriában a városok kisebb övezetekre bontása úgy történt, hogy a helyi közigazgatás szakembereivel osztályoztattuk – saját *ad hoc* kritériumaik alapján – a szóban forgó városok önkormányzati választókörzeteit aszerint, hogy azok (megítélésük szerint) cigányok által sűrűn, közepesen vagy ritkán lakottnak számítanak-e. Majd hogy ennek az *ad hoc* kategorizálásnak az önkényességét kiküszöböljük, mind a megyeközpontok, mind pedig az egyéb városok kategóriáján belül külön-külön megkevertük a sűrű, a közepes és a ritka körzeteket. A keveréssel randomizáltuk a kategorizálás önkényes szempontjait, s így viszonylag homogén elvek alapján kiválasztott választókörzeti mintából választottuk ki – a ritka, közepes és sűrű kategórián belül eltérő kulcsok szerint – a városokon belüli kisebb mintakörzeteket. Budapest és Miskolc esetében hasonló elvek szerint jártunk el, azzal a különbséggel, hogy mindkét város esetében rendelkezésünkre álltak egy 1992. évi adatfelvételből⁴ *teljes körűen* az alsó tagozatos általános iskolás cigány gyerekek *lakóhely* szerinti adatai. Ezeket először Budapest esetében városrendezési körzet, Miskolc esetében pedig iskolakörzet szintre aggregáltuk, majd az így kapott városrendezési, illetve iskolakörzeteket ritka, közepes, sűrű típusba soroltuk, végül pedig a községi mintavételhez hasonlóan eltérő kulcsok alapján mintát vettünk a körzetekből.

A munka oroszlánrészre azonban csak ezután következett: a kiválasztott mintakörzetekben (községek esetén: a kiválasztott községbokrokon belül; a városok esetén: a kiválasztott önkormányzati választókörzeteken belül; Budapesten és Miskolcon a kiválasztott városrendezési, illetve iskolakörzeteken belül) a helyi közigazgatás (önkormányzat) és a terepen működő szociális, egészségügyi, oktatási intézmények (iskolák, óvodák, családsegítők, védőnők, szociális célú civilszervezetek), valamint cigányszervezetek segítségével *teljes körű összeírást*⁵ hajtottunk végre a területen lakó cigány háztartásokról. Ezután meghatározott kulcsok szerint véletlen mintát vettünk az összeírt háztartásokból. Háztartás szintű mintánkat végül is úgy kaptuk meg, hogy a területi egység szintű mintavétel, illetve a területi egységeken belüli mintavétel arányszámainak szorzata mindenütt egységesen 0,02, azaz 2 százalék lett. Budapest esetében 4 százalékos, Miskolc esetében pedig 8 százalékos mintát választottunk annak érdekében, hogy e két város mintája önmagában is elegendő elemszámú, reprezentatív mintát biztosítson. A minta *egésze* természetesen csak országosan

és településtípusok szerint reprezentatív. Területileg legfőljebb *nagyterség* szinten az (de már megyei szinten nyilvánvalóan nem az).

A két minta összevetése az általános iskolás gyerekszám által definiált körzetsűrűség kategóriái szerint

Felmerülhet a gyanú, hogy a két minta között a homogenitás-vizsgálatok révén feltárt különbségek a mintavételi eljárások különbségéből adódnak. Ebben a részben ezt a hipotézist fogjuk közelebbről szemügyre venni. Hogy ezt megtehessük, a két mintát *azonos mintavételi szempontok szerint* kell megvizsgálnunk. Minthogy az ELAR-felvétel mintavételi kritériumai elvileg sem alkalmazhatók a Cigányvizsgálat mintájára – az ugyanis a *teljes* magyarországi népesség területi elhelyezkedése szerinti rétegzést alkalmazott –, csak a fordított eljárással kísérletezhettünk: *a Cigányvizsgálat rétegzési szempontjait alkalmaztuk az ELAR-felvétel cigány almintájára.*

Mindenekelőtt látnunk kell, hogy a Cigányvizsgálatban alkalmazott rétegzési kritériumnak (az általános iskolás gyerekek területi eloszlásának) az alkalmazása távolról sem problémamentes. Hiszen, ha igaz az – márpedig igaz (lásd a későbbiekben bemutatott 8–9. táblázatokat) –, hogy a körzetsűrűség pozitív korrelációban áll a háztartáslétszámmal és a háztartásokban élő eltartott gyermekek számával, akkor a gyermekeknek (g) a teljes népességhez (n) viszonyított aránya ($r=g/n$) bármely aggregált szinten maga is pozitív függvénye lesz a körzetsűrűség folytonos, elméleti mérőszámának (n -nek). Tömören kifejezve: $\partial r/\partial n > 0$. Ennek folytán, ha a gyerekszám alapján soroljuk *ritka*, *közepes* és *sűrű* kategóriákba a településeket, akkor az osztályközök *határain* előfordulhatnak olyan esetek, hogy egy-egy települést a gyerekszám alapján a *sűrű* kategóriába sorolunk, holott teljes népessége alapján a *közepes* kategória felső sávjába kerülne; vagy a gyerekszám alapján a *közepes* kategóriába sorolunk, holott teljes népessége alapján a *ritka* kategória felső sávjába kerülne. Vajon mérőműszerünknek ez a hibája milyen mértékben felelős a két minta társadalmi-demográfiai indikátorainak nagymérvű eltéréséért? Operacionálisan ez két kérdés tisztázását jelenti: 1. Ha mindkét mintára *ugyanazt* a mintavételi eljárást alkalmazzuk, mekkora különbséget találunk? 2. A mintavételi eljárások különbségének betudható különbség milyen mértékben magyarázza a homogenitás-vizsgálatok által feltárt jelentős strukturális különbségeket a két minta társadalmi-demográfiai indikátorai között? A lehetséges 3. kérdést – azt ugyanis, hogy a mintavételi eljárások különbségének betudható eltérések milyen mértékben vezethetők vissza külön-külön a két mintavételi eljárás *hibáira* – nem tudjuk megvizsgálni. A hatások ugyanis aligha választhatók szét. Ennek hiányában *föltesszük* – bár nem engedjük meg –, hogy a mintavételből adódó

eltérések egyedül a Cigányvizsgálat rétegzési kritériumának torzító hatásaira vezethetők vissza.

A Cigányvizsgálat rétegzési szempontját oly módon alkalmazhatjuk az ELAR-felvétel mintájára, hogy az általános iskolai statisztikákban szereplő cigány gyerekek eloszlása alapján elkészített településtipológiának – mely cigányok által *ritkán*, *közepesen* sűrűn, illetve *sűrűn* lakott településeket különböztet meg – a kódjait alkalmazzuk az ELAR-felvételben szereplő településekre. Minthogy a Cigányvizsgálat mintavételi eljárása során ezt a kritériumot csak a *községi almintára* alkalmaztuk, a tesztet is erre a részsokaságra⁶ szűkítettük le.⁷ Elsőként (ez a fent említett 1. kérdés) arra vagyunk kíváncsiak, hogy vajon a háztartások eloszlása érdemleges módon eltér-e egymástól a két mintában a szóban forgó rétegtípusok ismérve szerint.

A következőket találtuk. Az ELAR-minta 755 (súlyozva: 120107) cigányháztartásából 484 (súlyozva: 69886) háztartás került a szóban forgó községi almintába. A Szociológiai Intézet Cigányfelvételének 1931 háztartásából pedig 1161 háztartás. Az ELAR esetében ez a súlyozott háztartásminta 58,19 százaléknak, a Cigányvizsgálat esetében pedig a minta 60,12 százaléknak felel meg. Az eltérés nem mondható túlságosan nagy. Lényegesen fontosabb szempont azonban a két minta – mondjuk így: körzetsűrűség szerinti – belső összetétele. A teszt ugyanis erre irányul. A 7. táblázat mutatja a háztartások eloszlását az általunk használt településkategóriák szerint.

7. táblázat

A 15 000 főnél kisebb lélekszámú községekben élő háztartások eloszlása az általános iskolás cigány gyerekek előfordulási gyakorisága alapján képzett településkategóriák (körzetsűrűség) szerint az ELAR-felvétel cigány almintájában és a Szociológiai Intézet Cigányfelvételében

a település típusa (körzetsűrűség)*	ELAR (súlyozott)		Cigányfelvétel	
	gyakoriság	százalék	gyakoriság	százalék
ritka	5 408	7,74	91	7,84
közepes	21 546	30,83	324	27,91
sűrű	42 932	61,43	746	64,25
összesen	69 886	100,00	1161	100,00
chi2(2)			4,666	
valószínűség			90%	
nullhipotézis			bizonytalan	

* *Ritka*, ha az általános iskolás cigány gyerekek száma: 1–10, *közepes*, ha: 11–40, *sűrű*, ha 41 vagy annál több.

A megoszlások – ahhoz képest, hogy a két adatfelvétel tökéletesen különböző mintavételi eljárással készült – meglepően hasonlítanak egymásra. A *ritka* településkategórián belül gyakorlatilag nincs eltérés, ahol pedig különbségeket találunk, ott annak mértéke nem éri el a 3 százalékot. A Cigányvizsgálat községi almintájában 2,8 százalékkal nagyobb arányban fordulnak elő a *sűrű* körze-

tekben lakó cigány háztartások, az ELAR-minta községi almintájában pedig 2,9 százalékkal nagyobb valószínűséggel fordulnak elő a *közepes* településkategóriába tartozó háztartások. Intuitív is látszik, hogy ilyen mérvű különbségek-ből aligha adódhatnak olyan mértékű strukturális különbségek, mint amilyeneket az 1. rész homogenitás-vizsgálatai révén megállapítottunk.

Ez utóbbi intuitív megállapításunkat azonban empirikusan is tesztelnünk kell. Ez természetesen nem más, mint a fentebb említett 2. számú tisztázandó probléma. A minták körzetsűrűség szerinti eltérései azért befolyásolhatják a kritikus társadalmi-demográfiai indikátorok (háztartásnagyság, gyerekszám, életkor, iskolai végzettség) átlagos értékeit, illetve megoszlásait, mivel várhatóan *szoros korrelációban* állnak velük. Mint az alábbiakban megmutatjuk, valóban ez a helyzet. Nem közömbös azonban, hogy a két minta körzetsűrűség szerinti összetételének eltérései – az említett szoros korrelációk folytán – *milyen mértékben* változtatnák meg az egyik vagy a másik minta társadalmi-demográfiai indikátorainak átlagos értékeit vagy megoszlásait.

A tesztet természetesen továbbra is csak a 15000 főnél kisebb lélekszámú községek kategóriájára tudjuk elvégezni. A teszt abból áll, hogy megvizsgáljuk, milyen következményekkel jár az, ha az ELAR-minta körzetsűrűség szerinti átlagos háztartásnagyság, gyerekszám, életkor adatait, valamint iskolai végzettség szerinti megoszlásait a Cigányvizsgálat körzetsűrűség szerinti megoszlásaival súlyozzuk, illetve, ha a Cigányvizsgálat mintájának hasonló, körzetsűrűség szerint bontott átlagértékeit az ELAR-minta körzetsűrűség szerinti megoszlásaival súlyozzuk (lásd 8–12. táblázat).

8. táblázat

A háztartások átlagos létszáma a települések körzetsűrűsége szerint az ELAR-minta és a Cigányvizsgálat mintájának községi almintájában (fő/háztartás)

körzetsűrűség		ELAR (súlyozott)	CIGV
ritka		3,13	3,73
közepes		3,32	4,17
sűrű		3,55	4,58
a főátlag	saját * súlyokkal	3,45	4,40
értéke		(1,55)	(2,27)
súlyozva	a másik ** súlyjaival	3,46	4,39

* A körzetsűrűség szerinti átlagértékek a *saját* (az ELAR esetében az ELAR, a CIGV esetében a CIGV körzetsűrűség szerinti) megoszlásaival (lásd 7. táblázat) súlyozva (zárójelben: a szórások).

** A körzetsűrűség szerinti átlagértékek a *másik* (az ELAR esetében a CIGV, a CIGV esetében az ELAR körzetsűrűség szerinti) megoszlásaival (lásd 7. táblázat) súlyozva.

9. táblázat

A háztartásokban élő gyermekek átlagos száma a települések körzetsűrűsége szerint az ELAR-minta és a Cigányvizsgálat mintájának községi almintájában (fő/háztartás)

körzetsűrűség		ELAR (súlyozott)	CIGV
ritka		0,97	1,15
közepes		1,25	1,49
sűrű		1,28	1,79
a főátlag	saját * súlyokkal	1,25	1,65
értéke		(1,18)	(1,54)
súlyozva	a másik ** súlyjaival	1,25	1,64

* A körzetsűrűség szerinti átlagértékek a *saját* (az ELAR esetében az ELAR, a CIGV esetében a CIGV körzetsűrűség szerinti) megoszlásaival (lásd 7. táblázat) súlyozva (zárójelben: a szórások).

** A körzetsűrűség szerinti átlagértékek a *másik* (az ELAR esetében a CIGV, a CIGV esetében az ELAR körzetsűrűség szerinti) megoszlásaival (lásd 7. táblázat) súlyozva.

10. táblázat

A községi almintában található személyek megoszlása és átlagéletkora a községek körzetsűrűsége szerint az ELAR-mintában és a Cigányvizsgálat mintájában

körzet-sűrűség		ELAR súlyozott		CIGV	
		megoszlás	átlag	megoszlás	átlag
ritka		7,03	33,09	6,66	26,68
közepes		29,82	28,14	26,38	25,39
sűrű		63,15	26,67	66,96	23,94
összesen		100,00	–	100,00	–
(esetszám)		(242856)		(5103)	
chi2(2)			32,330		
valószínűség			100%		
nullhipotézis			elvetve		
a főátlag	saját * súlyokkal	–	27,56	–	24,50
értéke		–	(20,12)	–	(18,19)
súlyozva	a másik ** súlyjaival	–	27,49	–	24,50

* A körzetsűrűség szerinti átlagértékek a *saját* (az ELAR esetében az ELAR, a CIGV esetében a CIGV körzetsűrűség szerinti) megoszlásaival (lásd 7. táblázat) súlyozva (zárójelben: a szórások).

** A körzetsűrűség szerinti átlagértékek a *másik* (az ELAR esetében a CIGV, a CIGV esetében az ELAR körzetsűrűség szerinti) megoszlásaival (lásd 7. táblázat) súlyozva.

11. táblázat

A községi almintában található személyek* iskolai végzettség szerinti megoszlása a községek körzetsűrűsége szerint az ELAR-mintában és a Cigányvizsgálat mintájában (százalék)

iskolai végzettség	ritka	körzetsűrűség közepes	sűrű	összesen
		ELAR		
0 osztály	3,51	3,79	8,87	7,04
1–7 osztály	24,58	23,38	38,04	32,91
8 osztály	47,57	50,63	41,08	44,26
szakmunkásképző	17,26	12,66	9,27	10,82
középiskola	7,08	5,58	2,26	3,56
felsőfok	–	3,95	0,48	1,42
összesen	100,00	100,00	100,00	100,00
esetszám	11 209	42 065	96 198	149 472
(százalék)	(7,50)	(28,14)	(64,36)	(100,00)
		CIGV		
0 osztály	11,87	9,77	10,29	10,27
1–7 osztály	30,14	34,49	34,81	34,38
8 osztály	38,81	42,40	41,78	41,73
szakmunkásképző	15,53	11,62	12,16	12,26
középiskola	3,20	1,48	0,81	1,16
felsőfok	0,46	0,25	0,15	0,20
összesen	100,00	100,00	100,00	100,00
esetszám	219	809	1982	3010
(százalék)	(7,28)	(26,88)	(65,85)	(100,00)
chi2(5)		134,994		
valószínűség		100%		
nullhipotézis		elvetve		

12. táblázat

A községi almintában található személyek* iskolai végzettség szerinti megoszlása az ELAR-mintában és a Cigányvizsgálat mintájában a körzetsűrűség szerinti megoszlásokkal súlyozva (százalék)

iskolai végzettség	ELAR-minta		CIGV-minta	
	ELAR súly	CIGV súly	CIGV súly	ELAR súly
0 osztály	7,04	7,12	10,27	10,26
1–7 osztály	32,91	33,12	34,38	34,37
8 osztály	44,26	44,12	41,73	41,73
szakmunkásképző	10,82	10,76	12,26	12,26
középiskola	3,56	3,50	1,16	1,18
felsőfok	1,42	1,38	0,20	0,20
összesen	100,00	100,00	100,00	100,00
chi2(5)		3,825		0,006
valószínűség		43%		0%
nullhipotézis		elfogadva		elfogadva

* 15–74 éves korcsoport (nappali tagozatos tanulók nélkül).

A 8–12. táblázatok alapján jól látszik: a mintavételi eljárások eltéréséből adódó – valóban meglevő – különbségek *teljesen elhanyagolható mértékben érintik* a minták legfontosabb társadalmi-demográfiai indikátorait. Noha minden kétséget kizáróan igaz az, hogy a Cigányvizsgálatban alkalmazott területi rétegzési elv (az általános iskolás cigány gyerekek száma alapján meghatározott körzetsűrűség) szoros korrelációban áll az említett indikátorokkal, a községi alminták körzetsűrűség szerinti eloszlásai – a mintavételi eljárások gyökeres különbözősége ellenére is – oly kis mértékben különböznek egymástól az ELAR-felvétel és a Cigányvizsgálat esetében, hogy függetlenül attól, melyik minta körzetsűrűség szerinti súlyait alkalmazzuk, lényegében ugyanazokat a megoszlásokat, illetve átlagokat kapjuk vissza, mint amelyekkel az egyik vagy a másik minta esetében eredetileg rendelkezünk.

A különbségek okát tehát másutt kell keresnünk. A következő fejezetben az adatfelvételi eljárások egy másik, fontos elemét – a *meghiúsult kikérdezésekből adódó torzítások* problémáját – vesszük szemügyre.

3. A két adatfelvétel különbségeinek egyik lényeges oka: a meghiúsult kikérdezésekből adódó torzítások eltérő kezelése

A továbbiakban megpróbálunk amellet érvelni, hogy a két adatfelvétel belső szerkezeti különbségei (legalábbis részben) abból adódnak, hogy míg a Cigányvizsgálatban a meghiúsult kikérdezéseket megfelelő pótcímekkel⁸ pótoltuk, addig az ELAR-adatfelvételben az eredeti háztartásminta mintegy 15–16 százalékában⁹ nem került sor a meghiúsult kérdőívek pótlására. Az ELAR-minta esetében a népszámlálás nem, életkor, lakóhely szerint bontott, továbbvezetett népességszámainak megfelelő belső arányokat a sikeres kikérdezéseket jelentő megfigyelések átsúlyozásával biztosították.¹⁰ Ez elvileg teljesen korrekt eljárás a minta *egészére* nézve. Egy *részpopulációra* nézve azonban már korántsem ilyen egyértelmű a helyzet. Ha feltesszük, hogy a teljes népességre reprezentatív, szakmailag korrekt módon megtervezett minta – a nagy esetszámból adódóan – viszonylag kis létszámú kisebbség esetében is reprezentatív mintát biztosít számunkra, akkor a meghiúsult kikérdezések miatt bekövetkező mintatorzításokat csak abban az esetben korrigálhatjuk megfelelő módon a szóban forgó *részpopuláció* esetében a *teljes* mintára kidolgozott átsúlyozási eljárás segítségével, ha joggal feltételezhetjük, hogy a meghiúsult kérdőívek a részpopuláció esetében nagyjából ugyanolyan arányban kerülnek ki a társadalom különböző rétegeiből, mint a minta egészének esetében. Az alábbiakban amellet érvelünk, hogy az ELAR-minta cigány almintájának esetében empirikusan egészen biztosan nem állt fenn ez az eset. Mi több: a Cigányvizsgálat felvételének meghiúsult kikérdezéseiről összegyűjtött információink alapján jó okunk van feltételezni, hogy a meghiúsulások a minta egésze, illetve a cigány részmin-

ta esetében jellegzetesen különböző arányban kerültek ki a társadalom különböző rétegeiből. A részminta megfelelő arányait ezért a teljes mintára kidolgozott átsúlyozási eljárás segítségével nem lehet helyreállítani. Amint az alábbiakban kimutatjuk: a súlyozási eljárás ebben a konkrét esetben nemhogy javítaná, hanem éppenséggel tovább rontja a cigány al minta reprezentativitását.

Az ellenőrzést két kulcsváltozó – a háztartáslétszám és a lakásban lakó gyermekek száma – alapján végezzük el. Mint a két minta belső eltéréseit taglaló fejezetben láttuk, e tekintetben igen nagyok a két adatfelvétel különbségei. A probléma tehát ugyancsak releváns. Az ellenőrzés során indirekt bizonyítást alkalmazunk. Föltesszük, hogy a súlyozás *mind a teljes populáció, mind pedig a cigány részp populáció esetében reprezentatív mintát eredményezett*. Ennek a feltevésnek az alapján megvizsgáljuk a teljes mintában és a cigány részmintában az átlagos háztartáslétszám és az átlagos gyerekszám súlyozás nélküli, illetve súlyozott értékeit. Ha igaz az, hogy a súlyozás megfelelően helyreállítja a minták reprezentativitását, akkor a súlyozott és súlyozatlan átlagértékek *különbségei*nek egyértelmű bizonyítékát kell adniuk annak, hogy a meghiúsulások átlagosan milyen irányban torzították el az eredetileg helyes mintákat. Lássuk tehát mindenekelőtt a kérdéses számadatokat.

A 13. táblázat tanúsága szerint a szóban forgó teszt szempontjából a minta egésze és a cigány al minta teljesen eltérő módon viselkedik. A minta egészében azt látjuk, hogy a súlyozott átlagértékek kis mértékben meghaladják a megfelelő súlyozatlan értékeket, ezzel szemben a súlyozatlan cigány átlagadatok rendre alatta maradnak (ráadásul lényegesen nagyobb mértékben) a súlyozott átlagértékeknek. Ebből a különbségből *logikailag az következne* – feltéve persze, hogy nem kérdőjelezzük meg azt, hogy a súlyozás mind a teljes minta, mind pedig a cigány al minta esetében reprezentatív mintát biztosít –, hogy az *ELAR-minta egészében* a kikérdezések némileg nagyobb arányban hiúsultak meg a nagyobb, mint a kisebb létszámú háztartások esetében, a *cigány al mintán* belül viszont épp ellenkezőleg, a kikérdezések nagyobb arányban hiúsultak meg a kisebb, mint a nagyobb méretű háztartások esetében. Ha most figyelembe vesszük azt is, hogy a háztartáslétszám minden empirikus tapasztalat szerint szignifikáns negatív összefüggésben áll a háztartások egy főre jutó átlagjövedelmével és társadalmi státusával, akkor rögtön világossá válik, milyen különös következményekhez vezet az, ha az ELAR-minta egészének reprezentativitásáról gondoskodó súlyozásról feltesszük, hogy az a cigány al minta esetében *is* reprezentatív mintát biztosít. Azt jelenti ugyanis, hogy a cigány al minta esetében – amely al minta köztudott módon zömében nagy létszámú, alacsony státusú és szegény háztartásokból áll – a kikérdezések mégis nagyobb arányban hiúsultak meg a kis létszámú, magasabb státusú, jobb anyagi helyzetű háztartások esetében.

13. táblázat

Az átlagos háztartásnagyság és az átlagos gyerekszám értékei az ELAR-felvétel teljes mintájában és cigány almintájában (fő/háztartás)

megnevezés *	háztartáslétszám				gyermekszám			
	minta		cigány		minta		cigány	
	μ	σ	μ	σ	μ	σ	μ	σ
súlyozatlan	2,68	1,36	3,48	1,81	0,64	0,94	1,26	1,39
súlyozott	2,71	1,29	3,27	1,54	0,69	0,90	1,14	1,16
t-próba (μ)	-3,387		-9,443		3,646		2,920	
szignif.szint	0,000%		0,000%		0,000%		0,004%	
nullhipotézis	elvetve		elvetve		elvetve		elvetve	
absz. változás	+0,03	.	-0,21	.	+0,05	.	-0,12	.
rel. változás	+1,1%	.	-6,0%	.	+7,8%	.	-9,5%	.
absz.táv.vált.	-0,24				-0,17			
rel.táv.vált.	-30,0%				-27,4%			

* μ = átlag, σ = szórás, u = súlyozatlan, w = súlyozott, c = cigány almintá, m = teljes minta.

Abszolút változás = $\mu_w - \mu_u$.

Relatív változás = $100.0 * [(\mu_w - \mu_u) / \mu_u]$.

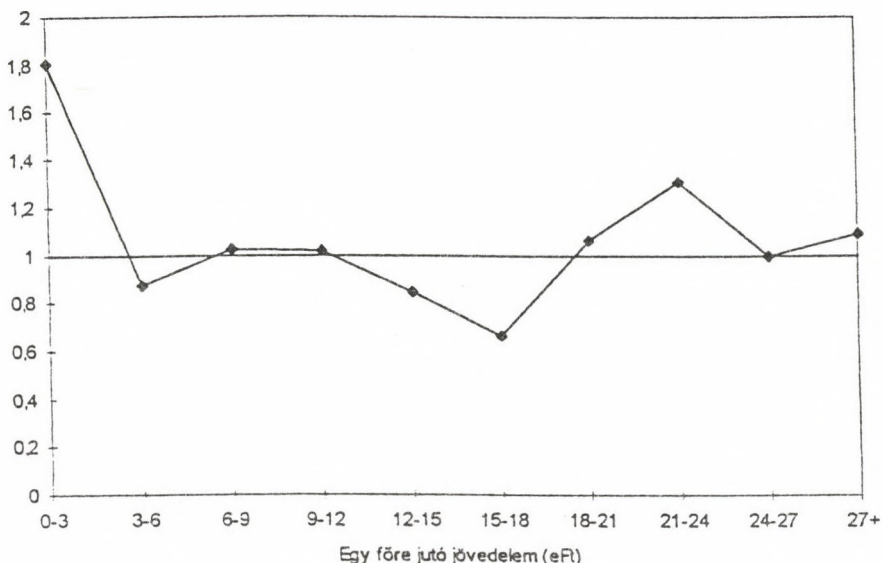
Az abszolút távolság változása = $[(\mu_{cw} - \mu_{mw}) - (\mu_{cu} - \mu_{mu})]$.

A relatív távolság változása = $100.0 * [(\mu_{cw} - \mu_{mw}) - (\mu_{cu} - \mu_{mu})] / (\mu_{cu} - \mu_{mu})$.

Az empirikus adatfelvételek egybehangzó tapasztalata ugyanakkor az, hogy a megíúsult kikérdezések halmazán belül többnyire a társadalmi hierarchia *szélein* levő háztartások vannak fölülreprezentálva. Vagyis: a magas és az alacsony társadalmi státusúak. Azok a háztartások, melyeknek megbízható részvételére lehet számítani a reprezentatív adatfelvételekben, mintabeli súlyuknál nagyobb valószínűséggel kerülnek ki a társadalom középrétegeiből. Ezt a megállapításunkat jól szemléltethetjük a TÁRKI 1993. évi háztartáspanel-felvételének megíúsult kikérdezéseivel, hiszen a szóban forgó vizsgálat paneljellegéből adódóan (az *előző* évi kikérdezés alapján) megbízható ismereteink vannak a megíúsult esetekről. Az 1. ábra összegzi ezeket a tapasztalatokat. Az ábra vízszintes tengelyén az egy főre jutó jövedelemkategóriákat tüntettünk föl, a függőleges tengelyen pedig azt ábrázoltuk, hogy az adott egy főre jutó háztartási jövedelemkategória esetében milyen mértékben vannak fölül-, illetve alulreprezentálva az 1992. évi mintabeli súlyukhoz képest az 1993. évben megíúsult kikérdezéseket jelentő (de az 1992. évi adatok alapján ismert egy főre jutó jövedelmű) háztartások. Az ábrából világosan kivehető a szegény (alacsony egy főre jutó jövedelmű) és a gazdag (magas egy főre jutó jövedelmű) háztartások relatív túlreprezentációja a megíúsult kikérdezések esetén.

1. ábra

Az 1993-ban megküldött kikérdezéseket jelentő háztartásminta fölül-, illetve alul-reprezentációjának mértéke az 1992. évi teljes minta megoszlásához viszonyítva (TÁRKI háztartáspanel, az egy főre jutó jövedelem szerinti megoszlások hányadosai)



Minthogy a cigányság zöme a társadalmi és jövedelmi hierarchia alján helyezkedik el, joggal számíthatunk rá, hogy a cigány megkérdezettek körében a megküldött kikérdezések zöme – ellentétben azzal, ahová a cigány al minta reprezentativitásának feltételezéséből logikai úton eljutottunk – az alacsony státusú, szegény (és ezért *nagyobb létszámú*) háztartások közül kerül ki. Ha pedig ez így van, akkor *ellentmondáshoz* jutottunk. Abból a feltételezésből, hogy a teljes mintára kidolgozott súlyozási eljárás megfelelően gondoskodik a cigány al minta reprezentativitásáról, azt a következtetést vezettük le, hogy akkor a cigány megküldések zömének a *kis létszámú* háztartásokból kell kikerülnie. A megküldések jövedelem és társadalmi státus szerinti eloszlásának általános tapasztalataiból, valamint a cigányság alacsony jövedelmi és társadalmi helyzetéből, továbbá a társadalmi és jövedelmi helyzet és a háztartásnagyság közti negatív korrelációból viszont ezzel épp ellentétes konklúzióra jutottunk: a cigány megküldések zömének a *nagy létszámú* háztartásokból kell kikerülnie. Ha tényadatokkal sikerül alátámasztanunk azt, hogy az utóbbi logikai levezetés empirikusan is helytálló, akkor a kis létszámú háztartások dominanciájára vonatkozó következtetés egyedül azért nem állhatja meg a helyét, mert a *premissza volt hibás: a minta egészének tulajdonságai alapján kidolgozott súlyozási eljárás nem biztosítja a cigány al minta reprezentativitását.*

A Cigányvizsgálat megghiúsult kikérdezéseiről rendelkezésünkre álló információk egyértelműen ezt az értelmezést támasztják alá. A megghiúsulásokról a kérdezőbiztosoknak adatlapot kellett kitölteniük, amely a megghiúsulás oka mellett a megadott cím közvetlen lakóhelyi környezetének megítélését (a lehetséges kódok: feltűnően slumos, lepusztult; szegényes, de viszonylag rendezett; átlagos; az átlagosnál jobb színvonalú), valamint a háztartás társadalmi helyzetének megítélését is tartalmazta (mármint abban az esetben, ha ez a háztartásról szerzett benyomások, a lakás képe és egyéb külsőségek alapján egyáltalán eldönthető volt). Az utóbbi eset lehetséges kódjai a „jómódú”, „átlagos”, „szegény” voltak. A megghiúsult kikérdezések száma a *teljes* mintában 405, azaz a minta (2222 háztartás) 18,4 százaléka volt. Ebből 52 olyan megghiúsulás volt, melyben a megkérdezettek vagy cigány mivoltukra hivatkozva, vagy éppenséggel azért, mert nem tartották magukat cigánynak, utasították vissza a válaszadást. Ez a két esemény elvileg sem fordulhatott elő az ELAR-adatfelvételben, hiszen a megkérdezett cigány háztartások tagjait nem informálták arról, hogy őket mint cigány háztartások tagjait is fölmérik. A fennmaradó 353 megghiúsulás a minta 16,4 százalékát tette ki. Összehasonlításképpen: ugyanez az arány az ELAR-felvételben 15,7 százalék (4186/26604) volt.

Ha tehát beigazolódnak az a sejtésünk, hogy a cigányok körében tapasztalható megghiúsulások zöme az alacsony társadalmi státusú rétegekből kerül ki, akkor az a tény, hogy a cigányok körében nagyobb valószínűséggel hiúsultak meg a kikérdezések, összességében még jobban megnöveli a kieső alacsony státusú (véltetően nagy létszámú) háztartások súlyát, melyeknek megfelelő súlyú mintabeli szerepeltetését a teljes populáció tulajdonságai alapján kidolgozott átsúlyozási eljárás nem képes megoldani. A 14–15. táblázat egyértelműen ezt a megállapításunkat támasztja alá.

Ha az ELAR-mintavétellel való kompatibilitás kedvéért figyelmen kívül hagyjuk a cigányságra hivatkozó válaszmegtagadás esetét, akkor azt látjuk, hogy a Cigányvizsgálatban a kikérdezések nagyobb részében – mintegy 54–58 százalékban – *slumos vagy szegényes környéken élő, illetve rossz anyagi helyzetű háztartások esetében hiúsultak meg*. Aligha feltételezhetjük, hogy ez másként történt az ELAR-minta cigány almintájának esetében. Vagyis arra számíthatunk, hogy a megghiúsulások következtében szignifikánsan nagyobb arányban hullhattak ki – az eredetileg reprezentatív – mintából a rossz körülmények között élő, szegény, nagyobb létszámú háztartások.

14. táblázat

A közvetlen lakóhelyi környezet kérdezőbiztosi megítélése a Szociológiai Intézet
Cigányvizsgálatának meghíúsult kikérdezései esetében (háztartások száma és százaléka)*

a meghíúsulás típusa	a lakóhelyi környezet színvonala a kérdezőbiztos megítélése szerint			összesen
	átlagosnál rosszabb	átlagos	átlagosnál jobb	
cigányságra hivatkozó válaszmehtagadás	18 (35,3)	25 (49,0)	8 (15,7)	51 (100,0)
egyéb válaszmehtagadás	38 (12,2)	36 (11,6)	9 (2,9)	83 (26,7)
meghíúsulás egyéb okból	130 (41,8)	88 (28,3)	10 (3,2)	228 (73,3)
összesen	168 (54,0)	124 (39,9)	19 (6,1)	311 (100,0)
mindösszesen	186 (51,4)	149 (41,2)	27 (7,4)	362 (100,0)

* 43 meghíúsulás esetében a kérdezőbiztos nem tudta megítélni a lakóhelyi környezet színvonalát.

15. táblázat

A háztartás anyagi helyzetének kérdezőbiztosi megítélése a Szociológiai Intézet
Cigányvizsgálatának meghíúsult kikérdezései esetében (háztartások száma és százaléka)*

a meghíúsulás típusa	a háztartás a kérdezőbiztos szerint			összesen
	szegény	átlagos	jómódú	
cigányságra hivatkozó válaszmehtagadás	14 (29,2)	17 (35,4)	17 (35,4)	48 (100,0)
egyéb válaszmehtagadás	32 (16,8)	29 (15,3)	15 (7,9)	76 (40,0)
meghíúsulás egyéb okból	79 (41,6)	28 (14,7)	7 (3,7)	114 (60,0)
összesen	111 (58,4)	57 (30,0)	22 (11,6)	190 (100,0)
mindösszesen	125 (52,5)	74 (31,1)	39 (16,4)	238 (100,0)

* 167 meghíúsulás esetében a kérdezőbiztos nem tudta megítélni a háztartás anyagi helyzetét.

Visszatérve indirekt bizonyításunk gondolatmenetéhez, következtetésünket az alábbiakban összegezhetjük. A cigány al minta esetében a minta reprezentativitását a súlyozás egyedül abban az esetben állíthatta volna helyre, ha a meghíúsulások következtében épp a kisebb háztartások hullottak volna ki szignifikánsan nagyobb arányban az eredetileg reprezentatív cigány al mintából. Ennélfogva az a körülmény, hogy józan empirikus megfontolások alapján ennek

épp az ellenkezője következhetett be, azt bizonyítja: *a súlyozás a cigány alminta esetében nem biztosítja a minta reprezentativitását*. Mi több: minthogy a cigány alminta esetében minden valószínűség szerint a nagyobb háztartások hullottak ki nagyobb valószínűséggel az eredeti mintából, a súlyozási eljárás pedig a cigány háztartások átlagos létszámát és gyerekszámát *csökkenti* (ahelyett, hogy növelné), *a súlyozás ahelyett, hogy javítaná, még tovább rontja a cigány alminta reprezentativitását*. Komoly érvek szólnak tehát amellett, hogy az ELAR-minta szisztematikusan alábecsüli a cigány háztartások átlagos létszámát és gyerekszámát. A 13. táblázat adatai alapján elmondhatjuk, hogy *e szisztematikus torzítás mértéke meglehetősen nagy, különösen ha figyelembe vesszük, hogy az eredeti (súlyozatlan) adatok maguk is szisztematikusan torzított becslései az elméleti adatoknak*.

4. Az átlagos gyermekszám adatok megbízhatóságának ellenőrzése országos iskolastatisztikai adatok alapján

A soron következő két fejezetben megpróbáljuk a két adatfelvétel megbízhatóságát egy egyértelműen objektívnek tekinthető külső adatforráson is lemérni. Ez a külső adatforrás nem más, mint a Művelődésügyi Minisztériumnak az 1970/71. és az 1992/93. tanév között a cigány gyerekekről is folyamatosan vezetett iskolastatisztikái. Mielőtt továbbmennénk, célszerű néhány szót ejtenünk a szóban forgó statisztikákról.

A Művelődésügyi Minisztérium az ötvenes évek végétől vezet évenkénti iskolaszintű nyilvántartást az ország valamennyi oktatási intézményéről (általános iskolájáról, közép-, illetve felsőfokú oktatási intézményéről). Ezekben az évről évre megismétlődő iskolaszintű adatfelvételekben az 1957/58. tanévtől kezdve több-kevesebb rendszerességgel, 1980/81-től pedig rendszeresen szerepel az egyes évfolyamokon tanuló cigány gyermekek számára vonatkozó kérdés. Az aggregált iskolastatisztikák alapjául szolgáló kérdésre mindig az adott iskola megfelelő osztályát tanító tanár (osztályfőnök) válaszolt – mégpedig ugyanannak a szociológiai jellegű kritériumnak az alapján, mint amit az 1993. évi ELAR-adatfelvétel és a Szociológiai Intézet 1993. évi Cigányvizsgálata is alkalmazott: „cigány az, akit a nem cigány környezet cigánynak tekint”. Az iskolastatisztikákban is ez az elv érvényesült: cigány tanulónak minősült az a diák, akit osztályfőnöke a gyermek, illetve a gyermek családjának *ismerete* alapján annak tekintett. Ezeket az információkat aggregálták azután először iskolaszintű adatokká (a kérdőíveket iskolaszinten töltötték ki), majd pedig további aggregáció útján ezekből készültek el az országos szintű iskolastatisztikai adatok. Minthogy az iskolastatisztikák alapjául szolgáló kérdőíveket mindig az adott tanév *októberében* töltötték ki, a cigány etnikai hovatartozásra vonatkozó kérdést olyan tanárok válaszolták meg, akik az első osztályosok esetében legalább

egy hónapja, a többiek esetében egy vagy több év óta *ismerték* őket és családjait. Nyilvánvaló tehát, hogy a hasonló kritériumot alkalmazó szociológiai jellegű adatfelvételekkel összevetve – ahol a kérdezőbiztosok *csak alkalmyszerűen, a megkérdezés időtartamára* álltak kapcsolatban a megkérdezettekkel – az iskolastatisztikák adatai megbízhatóbban mérik a fenti kritériumnak megfelelő cigány etnikai hovatartozást. Az iskolastatisztikák további előnyös tulajdonsága az, hogy *teljes körűek*: azáltal, hogy lefedik az adott oktatási intézménytípus valamennyi iskoláját, az adott iskolába járó korosztályok egészét, teljes körűen jellemzik. Minthogy a cigányság esetében az ily módon vezetett iskolastatisztikák jelentik a rendelkezésünkre álló *egyetlen, viszonylag megbízható, teljes körű adatbázist*, egy reprezentativitásra törekvő adatfelvételtől elvárható, hogy összhangban legyen vele.

A szóban forgó iskolastatisztikák mindenkori évfolyamszintű aggregált országos adatai rendelkezésünkre állnak az általános iskolák esetében az 1970/71-es tanévtől az 1992/93. tanévig – néhány év (1971/72, 1972/73, 1975/76, 1978/79, 1979/80) kivételével – folyamatosan, a szakmunkásképzők és szakiskolák esetében az 1980/81-es tanévtől, a gimnáziumok, szakközépiskolák és technikumok esetében pedig az 1979/80-as tanévtől egészen az 1992/93-as tanévig. Az egyetemre felvett cigány tanulókról nem készültek statisztikák.¹¹

A reprezentatívnak tekintett adatfelvételek gyermekszám, illetve iskolai végzettség adatainak megbízhatóságát tehát van min lemérnünk. Először a gyermekszám adatok megbízhatóságát teszteljük, majd a következő fejezetben sort kerítünk az iskolai végzettség adatok ellenőrzésére is.

A teszt természetesen csak a nappali tagozaton tanuló (normál és kisegítő) általános iskolás gyermekekre terjedhet ki. Az iskolastatisztikák szerint a nappali tagozatos általános iskolás cigány tanulók száma az 1992/93-as tanévben 88182 fő volt.¹² A 16. táblázatban bemutatunk néhány összefoglaló adatot az ELAR cigány almintájából és a Cigányvizsgálat mintájából. Mindkét esetben közöljük mind a súlyozatlan, mind pedig a súlyozott adatokat. A Cigányvizsgálat esetében a súlyozás mindössze azt jelenti, hogy a mintavételi körzetekre jellemző mintavételi arányok reciprokának felhasználásával teljeskörűsítettük adatainkat.

16. táblázat

A két minta megbízhatóságának összehasonlítása az általános iskolás gyerekek száma és a Művelődésügyi Minisztérium megfelelő teljes körű adata alapján

változók		ELAR		CIGV	
		súlyozatlan	súlyozott	súlyozatlan	súlyozott
háztartások száma	(1)	755	120107	2222	95421
személyek száma	(2)	2630	393715	9908	424645
nappali tagozatos általános iskolások száma	(3)	444	58226	2010	83176
(3)/MűvMin-adat*:					
1992/93-as tanév, százalék	(4)	–	66,03	–	94,32
általános iskolás arány, százalék (3)/(2)	(5)	16,88	14,79	20,28	19,59
MűvMin-adat alapján becsült** népesség, fő	(6)	–	596273	–	450203

* A Művelődési Minisztérium 1992/93. évi iskolai adatfelvétele szerint az általános iskolai cigány tanulók száma 88182 fő (74 441 normál általános iskolás + 13741 kisegítő iskolás) volt. *Forrás:* MKM Statisztikai Osztály. A (4)-es sorban szereplő adat a két adatfelvétel becsült általános iskolás létszám adatát az MKM teljes körű adatának százalékában adja meg.

** Becsült népesség = 88182 fő (MKM becsült adata) osztva a súlyozott általános iskolás aránnyal [(5)-ös sor].

A teszt egyértelműen igazolja azt a sejtésünket, hogy az ELAR-felvétel cigány almintájában *igen jelentős mértékben alul vannak reprezentálva az általános iskolás gyerekek*, s ennek következtében általában is alul vannak reprezentálva a gyerekek. A súlyozott adatok tanúsága szerint az ELAR cigány almintája mindössze 58000 körüli számban tartalmaz általános iskolás gyereket. Ha összevetjük ezt az MKM 1992/93-as adatával, országos szinten mintegy 34 százalékos elmaradást tapasztalunk a teljes körű adat alsó becsléséhez képest. Az *ELAR-felvétel cigány almintája nem tud elszámolni az országban élő nappali tagozatos általános iskolás cigány gyerekek 34 százalékával!* A Cigányvizsgálat ezzel szemben igen közel jutott az elméleti adathoz. Az elmaradás mindössze 5,7 százalékos. Ha legjobb becslésnek elfogadjuk az iskolastatisztikák 88000 körüli adatát, és erre az adatra vetítjük a két adatfelvételtől kiszámítható általános iskolás/teljes népesség arányt, akkor az alábbi igen meglepő eredményhez jutunk: amíg a Cigányvizsgálat becslése (450000 fő) nagyjából a két adatfelvétel által becsült 400–430 ezres népességszám közelében van, addig ez az eljárás az ELAR cigány almintájából becsült teljes cigány népességet csaknem 600000-es lélekszámúra növeli. Az ELAR saját becsléséhez (394000 fő) képest ez olyan mértékű eltérés, ami alapvetően megkérdőjelezi a szóban forgó adatbázis gyermekszám, illetve háztartáslétszám adatának realitását.

5. Az iskolai végzettség adatok megbízhatóságának ellenőrzése aggregált iskolai kiáramlási adatok alapján

Az országos iskolastatisztikai adatok azonban arra is alkalmasak – minthogy viszonylag hosszú idősor áll rendelkezésünkre –, hogy a *felnőtt cigány népesség iskolai végzettség szerinti megoszlását* ellenőrizzük velük. Az elgondolásunk az volt, hogy az iskolastatisztikák éves adatai alapján – alkalmasan megválasztott hipotézisek segítségével – megbecsüljük az adott évjáratból szakmunkásképzőt, illetve középiskolát (gimnáziumot, szakközépiskolát, technikumot) végzettek arányát, majd e kiáramlási (*flow*) adatok és az ellenőrizni kívánt adatfelvételek koreloszlásainak együttes figyelembevételével becslést adunk a szóban forgó iskolai végzettségű emberek jelenlegi állományáról (*stock*). A mintabeli állomány adatoknak természetesen nagyjából összhangban kell lenniük a kiáramlási adatok alapján *becsült stock* adatokkal. Eljárásunkat pontosabban az alábbi modell írja le.

A modell

Legyen N_{ijk} azon emberek száma, akik az i -edik évben szereztek k -adik iskolai végzettséget, és életben vannak a j -edik ($j > i$) évben. Ekkor természetesen N_{ij} azoknak a száma, akik az i -edik évben töltötték be a k -adik iskolai végzettség eléréséhez szükséges életkort (akár lett, akár nem lett k -adik iskolai végzettségű), és a j -edik évben is életben vannak. Továbbá N_{jk} azoknak a k -adik iskolai végzettséget szerzett embereknek a száma, akik életben vannak a j -edik évben. N_j a j -edik évben élő teljes populáció. A feladat az, hogy az adott (k -adik) iskolai végzettségi fokozatra nézve, az éves iskolai kiáramlási adatok, valamint a megfigyelés évének (a t -edik évnek) a koreloszlása segítségével megbecsüljük a t -edik évben élő népességben belül a k -adik iskolai végzettséggel rendelkező emberek arányát (N_{tk}/N_t).

Ha ismernénk azoknak az embereknek az arányát a t -edik időpontban élő teljes népességben belül, akik rendre az $i=1,2, \dots, t$ években szereztek k -adik iskolai végzettségi fokozatot (N_{itk}/N_i , $i=1,2, \dots, t$), akkor triviális módon előállíthatnánk ezeknek az adatoknak a felhasználásával a kívánt *stock* becslést. Ugyanis:

$$(1) \quad \frac{N_{tk}}{N_t} = \sum_{i=1}^t \frac{N_{itk}}{N_t}.$$

Az egyenlet jobb oldalán szereplő kiáramlási adatokat azonban nem ismerjük. Tudjuk viszont egyfelől az aggregált iskolastatisztikai idősorokból, hogy azok közül az emberek közül, akik életkoruknál fogva az i -edik évben szerezhettek k -adik iskolai végzettséget (N_{ii}), valójában hányan szereztek k -adik iskolai végzettséget (N_{iik}), másfelől pedig az 1993. évi adatfelvételekből ismerjük a jelenleg élő népesség korcsoportok szerinti eloszlását (N_{it}/N_t , $i=1,2, \dots, t$). Azt állítjuk, hogy e két hányados segítségével jó becslést adhatunk az (1)-es egyenlet jobboldalán szereplő összeg komponenseire, s így az egyenlet baloldalán szereplő hányadosra is. Az alábbi módon:

$$(2) \quad \sum_{i=1}^t \frac{N_{iik}}{N_{ii}} \cdot \frac{N_{it}}{N_t}.$$

Ha ugyanis a (2)-es kifejezésben szereplő összeg minden egyes tagját beszorozzuk N_{iik}/N_{ii} hányadossal (vagyis: eggyel), akkor – a megfelelő átrendezések elvégzése után – az alábbi formulát kapjuk:

$$(3) \quad \sum_{i=1}^t \frac{N_{iik}}{N_{ii}} \cdot \frac{N_{it}}{N_t} = \sum_{i=1}^t \left[\frac{N_{it} / N_{ii}}{N_{iik} / N_{iik}} \right] \cdot \frac{N_{iik}}{N_t}.$$

Látható, hogy a (3)-as egyenlet jobboldalán szereplő összeg minden tagja úgy áll elő, hogy az (1)-es képlet jobboldalán szereplő, elméletileg helyes, ám ismeretlen tagot a szögletes zárójelben szereplő szorzótényezővel módosítjuk. Hogyan értelmezzük, és mekkorára becsüljük ezt a torzító tényezőt?

A hányados értelmezése meglehetősen egyértelmű: a számlálóban az i -edik korcsoport t -edik időpontbeli ($t > i$) túlélési valószínűsége, a nevezőben pedig az i -edik korcsoport k -adik iskolai végzettséget elérő részének t -edik időpontbeli túlélési valószínűsége szerepel. Ha most már a k -adik iskolai végzettséget a szakmunkásképzőt, illetve középiskolát végzettek kategóriájára konkretizáljuk, akkor megállapíthatjuk: a cigányok körében, ahol az utóbbi két iskolai végzettségi kategória képezi a „magasabb iskolai végzettségűek” döntő többségét, standard emberi tőke elméleti megfontolások¹³ és iskolai végzettség specifikus aggregált mortalitási¹⁴ adatok alapján is erősen valószínűsíthetjük, hogy egy adott korcsoport magasabb iskolai végzettségű tagjainak túlélési esélyei jobbak, mint a – természetesen átlagos iskolai végzettségű – teljes korcsoporté. A szóban forgó hányados tehát empirikusan minden bizonnyal egynél kisebb szám lesz. Másfelől azt is figyelembe kell venni, hogy a szóban forgó torzító tényező nagysága a t -edik időpontbeli aktuális életkor függvénye is: minél alacsonyabb ez az életkor, annál valószínűbb, hogy az iskolai végzettség várható életkort növelő hatása nem érvényesül még. A fiatalabb korcsoportok esetében ezért a torzító tényező közel esik egyhez, vagyis a torzítás is kicsi. Mivel azon-

ban, minden empirikus tapasztalat szerint a magasabb iskolai végzettségűek körében a fiatalabb korcsoportok erősen fölülreprezentáltak – ez a gazdasági fejlődés egyenes következménye –, a torzítás mértéke az összeg legnagyobb súlyú tagjainál a legkisebb, és a legkisebb súlyú – idősebb – korcsoportjainál a legnagyobb. A torzítás tehát összességében nem lehet túlságosan nagy. Ami a torzítás *irányát* illeti, elmondhatjuk, hogy az ily módon *becsült stock* adat *alacsonyabb* lesz az elméletileg helyes adatnál, *de nem lesz sokkal alacsonyabb*. Következésképp a (2)-es formulát – melynek mérését empirikusan képesek vagyunk megoldani – jó közelítésként elfogadhatjuk az elméletileg helyes (1)-es kifejezés becslésére. Vagyis:

$$(4) \quad \frac{N_{tk}}{N_t} \approx \sum_{i=1}^t \frac{N_{iik}}{N_{ii}} \cdot \frac{N_{it}}{N_t}.$$

A becsült stock adat kiszámításának gyakorlati lépései

A fenti eljárás gyakorlati megvalósításának lépései a következők voltak:

1. Első lépésben létrehoztuk az iskolai kiáramlási adatokat a rendelkezésünkre álló abszolút számokból. Az eljárás nyilvánvalóan az, hogy a végzett szakmunkásképzősök esetében a 11 évvel korábban, a középiskolások esetében pedig a 12 évvel korábban beiskolázott általános iskolai elsős korosztály számának százalékában adjuk meg a mindenkori relatív kiáramlási adatokat.

2. A relatív kiáramlási százalékok kiszámításakor kétféle becslési eljárást követtünk. Az egyikben változtatás nélkül magukat a nyers létszámadatokat fogadtuk el mérvadónak, vagyis *teljesen figyelmen kívül hagytuk* azt a nyilvánvaló tényt, hogy a túlkoros, illetve évismétlő általános iskolai elsősök közül minden bizonnyal sokkal kisebb, a normál korú elsősök közül pedig sokkal nagyobb eséllyel jutnak el a gyerekek a sikeresnek tekinthető, szakmunkásképzős, illetve középiskolai pályák végpontjaira. Ez az eljárás természetesen csak a kiáramlási százalékok *alsó becsléseként* fogadható el. A másik eljárás során ezzel éppen ellentétes elvet követtünk: *maximálisan figyelembe vettük* ezt a tényezőt. A szóban forgó általános iskolai elsős korosztályoknak csak a normál korú részét vettük be a nevezőbe, a 11, illetve 12 évvel később szakmunkásképzőt, illetve középiskolát végzettek nyers adatait pedig változtatás nélkül szerepeltettük a számlálóban. Ezzel azt feltételeztük, hogy az adott generációból a sikeres pályák végpontjaira kizárólag normál korú diákoknak sikerül eljutni. Ez az eljárás természetesen csak a kiáramlási százalékok *felső becsléseként* fogadható el. Az elméleti adat valahol az alsó és felső becslés között lehet.

További nehézséget jelentett, hogy az általános iskolák első osztályába beiskolázott cigány gyerekek közül a normál korúak arányára csak öt megfigyelés-

sel rendelkezünk.¹⁵ A hiányzó adatokat lineáris interpoláció, illetve extrapoláció által becsült adatokkal töltöttük ki: vagyis azt feltételeztük, hogy a megfigyelések közti változások időben egyenletesen mentek végbe.

3. Mint korábban említettük, a szakmunkásképzőt, illetve középiskolát végzett cigány gyerekek számáról csak a nyolcvanas évtized elejétől vannak folyamatosan adataink, az általános iskolás cigány gyerekek számáról pedig az 1969/70. tanévtől folyamatosan (bár néhány tanév esetében adathiányokkal). Ennek következtében a 2. pontban jelzett számítások eredményeit extrapolálnunk kellett azokra a tanévekre (vagy, ami ugyanaz: korcsoportokra), melyekre vonatkozóan adathiány folytán nem álltak rendelkezésünkre megfigyelések. Ahol két megfigyelés közti év adatát kellett becsléssel pótolni, ott lineáris interpolációt alkalmaztunk, ahol pedig adatsorainkat az 1980 előtti évekre kellett meghosszabbítani (szakmunkásképző esetében csak 1960-ig¹⁶), ott az utolsó három, illetve öt egymást követő megfigyelés átlagát vetítettük vissza a múltba. Ez utóbbi eljárás jogosultságát mérési céljaink igazolják. Az a körülmény, hogy az ELAR-adatbázisból a középiskolát végzettek arányára vonatkozó stock adat realitásában kételkedünk – túlságosan magasnak tartjuk –, indokoltá teszi, hogy az adathiányos időszakra a valóságosnál minden bizonnyal *magasabb* flow adatokat vetítsünk vissza, hiszen ezzel a kiáramlási adatok alapján becsült stock adat is magasabb lesz a valóságosnál. Ez pedig bizonyításunkat nemhogy megkönnyítené, hanem éppenséggel megnehezíti.

4. Mivel a kiáramlási adataink csak a *nappali tagozaton* végzett diákokra vonatkoznak, mind az ELAR-mintából, mind pedig a Cigányvizsgálat mintájából ki kellett válogatnunk a 15 éves, illetve annál idősebb korosztályokból azokat, akik legmagasabb iskolai végzettségüket nappali tagozaton szerezték. Ennek a halmaznak a korcsoport-súlyaival súlyoztuk be az iskolastatisztikák alapján becsült kiáramlási adatokat. Ennek eredményeként kaptunk egy alsó és egy felső becsült stock értéket a szakmunkásképzőt és középiskolát végzettek arányára, amelyet összehasonlítottunk a Cigányvizsgálat és az ELAR cigány részmintájának megfelelő módon definiált tényadataival. Ezek után lássuk a számítások részletes menetét.

A becslési eljárás menete

Mindenekelőtt az ellenőrizni kívánt adatokat kellett létrehozunk. Ez a halmaz – mint említettük – csak azokból állhat, akik vagy egyáltalán nem jártak iskolába, vagy legmagasabb iskolai végzettségüket nappali tagozaton szerezték. A nappali tagozaton végzettek meghatározására csak közvetett módszert tudunk alkalmazni, mivel az ELAR-felvételben nem volt olyan változó, amely erre a konkrét kérdésre felelt volna. Rendelkezésünkre állt azonban a legmagasabb iskolai végzettség megszerzésének éve (bár az ELAR-adatbázisban

ez is csak a nyolc általánosnál magasabb végzettségűekre volt meg.) A Cigányvizsgálatban természetesen ugyanazt a kritériumot kellett alkalmaznunk. Módszerünk a következő volt. Az 1–7 osztályt végzettekről föltettük, hogy azt nappali tagozaton végezték el. A szakmunkásképzőt végzettek esetében nappali tagozaton szerzett végzettségnek tekintettük a 15 és 20 éves életkor között megszerzett végzettséget (ide tartoznak a kétéves szakiskolák is; ez indokolja az alacsony alsó életkorhatárt). A középiskolát végzettek esetében a 18 és 21 éves életkor között végzetteket, a felsőfokú végzettséggel rendelkezők közül pedig a 20 és 26 éves életkor között végzetteket számítottuk a nappali tagozaton végzettek közé. Akik nem nappali tagozaton végezték el a szakmunkásképzőt vagy a középiskolát, azokat automatikusan a nyolc általánost nappali tagozaton végzettek közé soroltuk. Analóg módon jártunk el azokkal, akik nem nappali tagozaton szerezték meg a felsőfokú végzettségüket: őket automatikusan a nappali tagozaton szerzett középiskolai végzettségűekhez soroltuk. Bonyolultabb volt a helyzet azokkal, akiknek a legmagasabb iskolai végzettsége a nyolc osztály volt. Az ő esetükben csak a Cigányvizsgálat adataira támaszkodhattunk. Akik a nyolc általánost 13 és 17 éves koruk között végezték el, azokat tekintettük nappali tagozaton végzetteknek. Akik 17 évesnél idősebb korukban végezték el a nyolc általánost, azokat automatikusan az 1–7 osztályt végzettek közé soroltuk be. Az ELAR-minta esetében pedig úgy jártunk el, hogy a nyolc osztályt nappali tagozaton végzettek számát a Cigányvizsgálat megfelelő arányszámának¹⁷ felhasználásával becsültük meg az ELAR-minta nyolc általános végzettségű összes emberének adatából.

Mivel az ELAR-minta iskolai végzettség adataival a széleken van vitánk, az ellenőrzést a legkritikusabb helyeken: a cigányok körében magas iskolai végzettségnek számító szakmunkásképző, illetve középiskola esetében végezzük el (a felsőfokú végzettségűek adatának ellenőrzése kontrolladatok híján nem lehetséges). 17. táblázatunkban az ellenőrizni kívánt, nappali tagozaton megszerzett legmagasabb iskolai végzettség adatokat foglaltuk össze (félkövér betűtípussal jelöltük az egész ellenőrzési eljárás célját jelentő szakmunkásképző, illetve legalább középiskolai végzettség adatokat). Minthogy iskolatípusonkénti kiáramlási (flow) adatokból becsülünk stock adatot, ezért a középiskolai végzettségűek állományába kell tartoznia mindenkinek, aki legalább középiskolai végzettséget szerzett (vagyis annak is, akinek a legmagasabb iskolai végzettsége az egyetemi vagy főiskolai diploma lett).

17. táblázat

A nulla osztályt végzetek vagy legmagasabb iskolai végzettségüket nappali tagozaton szerzett emberek nappali tagozaton megszerzett legmagasabb iskolai végzettség szerinti megoszlása a két mintában (százalék)

iskolai végzettség	ELAR	CIGV
0 osztály	5,64	9,82
1–7 osztály	33,02	37,92
8 osztály	41,26	39,57
szakmunkásképző	12,16	11,47
középiskola	6,38	1,48
felsőfokú	1,65	0,10
összesen	100,00	100,00
(esetszám)	(256752)	(4794)
chi2(5)	17231,473	
valószínűség	100%	
nullhipotézis	elvetve	
szakmunkásképző	12,16	11,47
legalább középiskola *	8,03	1,58

* Középiskolai + felsőfokú végzettségűek együtt.

A továbbiakban az iskolastatisztikák flow adataiból való stock becslés lépéseit követjük nyomon. A becsléshez fölhasznált cigány tanuló adatok két forrásból származnak. Egyrészt a hetvenes és nyolcvanas években készült öt részletesebb iskolai adatfelvétel a cigány gyerekekről (az 1970/71., 1974/75., 1977/78., 1981/82. és az 1985/86. tanévben). Ezeknek a felvételeknek az eredményeit* a Művelődésügyi Minisztérium belső kiadványokban adta közre (*Cigány tanulók* 1971; 1978; 1982; 1986). Általános iskolai statisztikáink egy része, valamint a szóban forgó évek normálkorú általános iskolai cigány tanulóinak részarányára vonatkozó adatok innen származnak. A nyolcvanas évek elejétől folyamatosan regisztrált általános iskolai adatokat, valamint a középiskolákba és szakmunkásképzőkbe járó cigány gyerekek valamennyi adatát a Művelődési és Köznevelési Minisztérium Statisztikai Osztálya bocsátotta a rendelkezésünkre. Alapadatainkat a Függelék F1., F2. és F3. táblázatában közöljük.

18. táblázat

A nyers kiáramlási adatok (alsó becslések)

a végzettség éve	szakmunkásképző		középiskola	
	korév	kiáramlási ráta (százalék)	korév	kiáramlási ráta (százalék)
1992/93	18	6,15	19	0,79
1991/92	19	6,21	20	0,73
1990/91	20	7,80	21	–
1989/90	21	–	22	–
1988/89	22	–	23	0,82
1987/88	23	7,26	24	0,88
1986/87	24	6,71	25	–
1985/86	25	–	26	0,64
1984/85	26	7,20	27	0,94
1983/84	27	7,37	28	–
1982/83	28	–	29	–
1981/82	29	–	30	0,59
1980/81	30	–	31	0,64

A 18. táblázat tartalmazza a nyers kiáramlási adatokat, melyeket úgy kaptunk meg, hogy az adott évben szakmunkásképzőt, illetve középiskolát végzetek számát elosztottuk a 11, illetve 12 évvel korábban beiskolázott általános iskolai elsősök számával. Ez a fentebb említett *alsó becslésnek* felel meg. A táblázatban a korév oszlop mindkét esetben azt jelenti, hogy az ELAR-felvétel és a Cigányvizsgálat időpontjában (1993) átlagosan hány éves lehetett az adott évben végzett egyén. Beiskolázott elsős adatok hiányában néhány év esetében nem tudjuk megadni a szóban forgó kiáramlási rátát.

A normálkorúak aránya a beiskolázott általános iskolai elsősök körében öt tanévre állt rendelkezésünkre. A közbűlső évekre vonatkozó adatokat lineáris interpolációval becsültük meg. Az így kapott becsléseket és az eredeti adatokat a 19. táblázat tartalmazza.

A 19. táblázat második oszlopának számaival módosítottuk a 17. táblázat adatait oly módon, hogy az ott található számokat elosztottuk a becsült normálkorú arány adatok 11, illetve 12 évvel késleltetett értékével. Ezzel az abszolút kiáramlási adatokat a normálkorú beiskolázottakra vetítettük. Eredményül pedig a kiáramlási ráták fölső becsléséhez jutottunk. Az alsó és felső becslések eredményeit a 20. táblázatban foglaltuk össze.

19. táblázat

A normálkorúak aránya az általános iskola első osztályosainak körében (százalék)

tanév	a normálkorúak aránya	
	tényadat	becslés
1969/70	–	41,82
1970/71	43,90	43,90
1971/72	–	45,97
1972/73	–	48,05
1973/74	–	50,12
1974/75	52,20	52,20
1975/76	–	58,03
1976/77	–	63,87
1977/78	69,70	69,70
1978/79	–	69,00
1979/80	–	68,30
1980/81	–	67,60
1981/82	66,90	66,90
1982/83	–	65,50
1983/84	–	64,10
1984/85	–	62,70
1985/86	61,30	61,30

A tényadatok forrása: *Cigány tanulók* 1978: 14–16.; 1986: 53.

20. táblázat

Kiáramlási ráták a szakmunkásképzőkből és a középiskolákból (az adott korosztály általános iskolás elsőseinek százalékában)

tanév	szakmunkásképző		középiskola	
	alsó becslés	felső becslés	alsó becslés	felső becslés
1980/81	–	–	0,64	1,53
1981/82	–	–	0,59	1,34
1982/83	–	–	–	–
1983/84	7,37	14,70	–	–
1984/85	7,20	13,79	0,94	1,88
1985/86	–	–	0,64	1,23
1986/87	6,71	10,51	–	–
1987/88	7,26	10,42	0,88	1,38
1988/89	–	–	0,82	1,18
1989/90	–	–	–	–
1990/91	7,80	11,54	–	–
1991/92	6,21	9,28	0,73	1,08
1992/93	6,15	9,39	0,79	1,18

A következő lépésben a fenti négy adatoszlop számait visszavetítettük az időben, és a hiányzó adatokra is becsléseket adtunk. Minthogy pontjainkra nem illeszkedett igazán sem lineáris, sem másmilyen függvény, ezért úgy döntöttünk, hogy a véletlen ingadozások egyenletlenségeit kiegyenlíttjük. Ez a szak-

munkásképzőt végzettek esetében úgy történt, hogy akik az 1991/92. és 1992/93. tanévben végeztek, azok e két év átlagos értékét kapták meg; akik az 1987/88. és 1990/91. tanévek között fejezték be iskolájukat, azok e két tanév átlagát kapták meg, akik pedig a 1986/87. tanévben vagy korábban végeztek, azok a maradék három érték átlagát kapták meg. A középiskolát végzettek esetében a következő kiigazítást tettük: azok, akik az 1988/89. tanév után végeztek, megkapták az 1988/89. tanév és a két utolsó érték átlagát; akik az 1983/84. és 1987/88. tanévek között végeztek, azok e két időpont közti három megfigyelés átlagos értékét vették föl; akik pedig az 1982/83. tanév előtt végeztek, megkapták az első két adat átlagát. Az így kapott eredményeinket ezután ötéves korcsoportokra átlagoltuk, a legutolsó ötéves korcsoportra számított adatunkat pedig visszavetítettük az időben (a szakmunkásképzők esetében csak a jelenleg ötvenéves korcsoportig). A korcsoportok természetesen tökéletesen megfeleltethetők az ötéves történelmi időintervallumoknak: az 1993-ban 20–24 éves korcsoport például megfelel az 1986/87. és 1990/91. között szakmunkásképzőt végzetteknek, illetve az 1987/88. és 1991/92. között középiskolát végzetteknek. A 21. táblázatban foglaltuk össze *flow* becsléseink végső eredményeit.

21. táblázat

Korcsoportonkénti átlagos kiáramlási ráták a szakmunkásképzőkből és a középiskolákból (az adott korosztály általános iskolai elsőseinek százalékában)

korcsoport (éves)	szakmunkásképző		középiskola	
	alsó becslés	felső becslés	alsó becslés	felső becslés
15–19	6,18	9,34	0,78	1,15
20–24	7,44	11,38	0,79	1,22
25–29	7,09	13,00	0,78	1,49
30–34	7,09	13,00	0,62	1,43
35–39	7,09	13,00	0,62	1,43
40–44	7,09	13,00	0,62	1,43
45–49	7,09	13,00	0,62	1,43
50–54			0,62	1,43
55–59			0,62	1,43
60–64			0,62	1,43
65–69			0,62	1,43
70+			0,62	1,43

A becsült stock adatainkat végül úgy kapjuk meg, hogy a 21. táblázatban található korcsoportos kiáramlási adatokat a nappali tagozaton végzett sokaságok hasonló módon definiált ötéves korcsoportjainak megoszlási viszonysszámaival súlyozzuk. Természetesen itt ugyanazokkal a halmazokkal kell dolgoznunk, mint amelyeket az ellenőrizni kívánt iskolai végzettség állományi adatok meghatározásánál használtunk (lásd 17. táblázat). A szóban forgó halmaz korcsoportok szerinti megoszlása a 22. táblázatban található.¹⁸

22. táblázat

A nulla osztályt végzettek vagy legmagasabb iskolai végzettségüket nappali tagozaton szerzett emberek korcsoportok szerinti megoszlása (százalék)

korcsoport (éves)	ELAR	CIGV	ELAR	CIGV
15–19	12,41	17,80	9,28	14,94
20–24	18,36	17,18	13,72	14,41
25–29	16,88	14,79	12,61	12,41
30–34	15,35	16,33	11,47	13,70
35–39	17,57	15,44	13,13	12,95
40–44	10,83	10,89	8,10	9,14
45–49	8,59	7,58	6,42	6,36
50–54			5,12	4,76
55–59			7,03	3,71
60–64			4,94	3,15
65–69			3,68	2,48
70+			4,50	1,98
összesen (esetszám)	100,00 (191885)	100,00 (4023)	100,00 (256752)	100,00 (4794)

Becslési eljárásunk utolsó lépéseként a 21. táblázatban szereplő korcsoportos kiáramlási rátákat (illetve a megfelelő ötéves történelmi intervallumok kiáramlási rátáit) az ELAR és a Cigányvizsgálat korcsoport szerinti megoszlásaival (a 22. táblázat adataival) súlyozzuk, s így az *iskolastatisztikai flow adatokból alsó és felső stock értékeket becsülünk* a nappali tagozaton végzettek, illetve iskolázatlanok közül a szakmunkásképzőt, illetve középfiskolát végzettek 1993. évi arányáról. A becslések eredményeit és az ellenőrzés tárgyát jelentő, *megfigyelt stock* adatokat a 23. táblázatban foglaltuk össze.

23. táblázat

A nulla osztályt végzettek, illetve végzettségüket nappali tagozaton szerzett emberek hány százalékának van szakmunkásképző vagy legalább középfiskolai végzettsége 1993-ban? (a becsült és a mintabeli adatok összevetése)

az ELAR- és a CIGV-minta korcsoport-súlyaival becsült, ill. mintabeli adatok		szakmunkásképzőt végzettek aránya (százalék)	legalább középfiskolát végzettek aránya (százalék)
becsült adatok	ELAR-súly	7,04 – 12,25	0,69 – 1,39
	CIGV-súly	6,99 – 12,07	0,68 – 1,37
mintabeli adatok	ELAR-minta	12,16	8,03
	CIGV-minta	11,47	1,58
t-próba*		1,449	16,393
szignifikanciaszint		7,35%	0,00%
nullhipotézis		bizonytalan	elvetve

* A próba rendre az ELAR- és a CIGV-minta adatait veti össze.

Megfogalmazhatjuk becslési eljárásunk végső következtetését. Először is: a becült adatok a súlyozásra nem érzékenyek. Ez természetes következménye annak, hogy a 30 évesnél fiatalabb korcsoportok esetében a flow adatok igen szűk sávban mozognak, illetve hogy a korcsoportos kiáramlási adatokat a 30 évesnél idősebb korcsoportok esetében állandónak tekintettük. Ismételten hangsúlyozzuk azonban: az idősebb generációk *elméleti* kiáramlási adatai minden józan empirikus tapasztalat alapján *alacsonyabbak* a modellben feltételezetténél: egy 1993-ban 43 éves cigány ember egész biztos, hogy kisebb valószínűséggel jutott el a középiskolai érettségiig az 1968/69-es tanévben (hiszen az iskolastatisztikai idősorok tanúsága szerint szignifikánsan kisebb eséllyel jutott el egyáltalán oda, hogy akár a nyolc általánost elvégezze), mint – mondjuk – egy 1993-ban 24 éves ember az 1985/86-os tanévben.¹⁹ Ha ismer-nénk ezeket a számokat, becült intervallumaink *lejjebb csúsznának* (és természetesen a két minta eltérő súlyai is jobban érvényesülnének). Mármost a 23. táblázat alapján világosan látszik: ez *még a jelenleginél is kedvezőtlenebb* megvilágításba helyezné az ELAR-minta középiskolai végzettség adatát.

Másodszor: jól látható az is, hogy a szakmunkásképző adatok mindkét minta esetében a flow adatok révén becült stock intervallumértékek közé esnek, habár inkább a felső határhoz állnak közel. Ha a ma 30 évesnél idősebb korosztályok elméleti kiáramlási rátáit használhatnánk, akkor elképzelhető, hogy mind az ELAR, mind pedig a Cigányvizsgálat mintabeli szakmunkásképző adata kívül kerülne a szóban forgó intervallumon, jóllehet a intervallum valódi határait fölfelé nyomja az a tény, hogy a becült stock érték kiszámításánál alkalmazott közelítő formulában (lásd (3)-as képlet) implicite benne rejlő torzító tényező (voltaképpen egy túlélési esélyráta) *némileg* lefelé torzítja mérőeszközünket. Összességében: a szakmunkásképző esetében mindkét adatot nagyjából kompatibilisnek tekinthetjük az aggregált iskolastatisztikákkal.

Végül, ami a középiskolai adatokat illeti, a következőket mondhatjuk: bár a legalább középiskolát végzettek aránya a Cigányvizsgálatban is enyhén túlmegy az iskolastatisztikákkal kompatibilis stock érték felső határán, az igazán súlyos problémát az ELAR-minta megfelelő adata jelenti. A felső határértéket 100-nak tekintve, ez az eltérés a Cigányvizsgálat esetében 15 százalékosnak, az ELAR-minta esetében azonban 578 (!) százalékosnak adódik. Ez a hatalmas eltérés – becslési eljárásunk minden tökéletlensége mellett – sem tekinthető esetlegesnek. A próba tehát arról tanúskodik: amíg a Cigányvizsgálat adata enyhén fölülbecsüli az iskolastatisztikák alapján reálisnak tekinthető középiskolai végzettség adatot, de nagyjából kompatibilis vele, addig az ELAR-minta adata olyan mértékben megy túl a megengedett határon, hogy az aggregált iskolastatisztikákkal való kompatibilitás kritériumának egészen biztos, hogy nem tesz eleget.

Összegzés

Foglaljuk össze érvelésünk tanulságait. Kimutattuk, hogy a KSH nemzetiségi felvételének cigány almintája és a Szociológiai Intézet Cigányvizsgálatának mintája – fontos társadalmi indikátorok tekintetében – lényeges mértékben különbözik egymástól. Ezeknek az eltéréseknek az okai után nyomozva, három problémát vizsgáltunk meg tüzetesebben. Számot vetettünk a mintavételi eljárások különbségeinek következményeivel, és kimutattuk: noha a két felvétel mintavételi eljárásai tökéletesen eltérő alapelveken nyugszanak, ezekből az eltérésekből mégsem magyarázhatjuk a szóban forgó jelentős mintabeli különbségeket. A továbbiakban megvizsgáltuk azokat a különbségeket, amelyek a meghíúsult kikérdezések eltérő kezeléséből adódnak. Beláttuk, hogy az a körülmény, hogy a KSH adatfelvétele a cigány almintára esetében – a dolog természetéből adódóan – nem tudta a meghíúsult cigány kikérdezéseket megfelelő cigány pótcímekkel helyettesíteni, jelentős mértékben hozzájárult a KSH cigány almintájának torzulásaihoz. Beláttuk azt is, hogy a KSH által alkalmazott súlyozási eljárás – a cigány almintára esetében – nemhogy javítaná, hanem inkább rontja a minta reprezentativitását. Végül megpróbáltuk a két minta jószágát, pontosabban az általános iskolás tanulók számának, illetve a felnőtt népesség iskolai végzettségi adatainak a realitását egy külső adatforráson – a Művelődési Minisztérium iskolastatisztikai idősorain – ellenőrizni. Ez a teszt is egyértelműen a Cigányvizsgálat adatainak nagyobb megbízhatóságát igazolta.

Jegyzetek

- ¹ A kutatást Kemény István, Havas Gábor (MTA Szociológiai Intézet) és Kertesi Gábor (MTA Közgazdaságtudományi Intézet) vezette.
- ² A homogenitás-vizsgálatról – amely tulajdonképpen nem más, mint a χ^2 -próbák egyik alelete – lásd például: Prékopa 1962: 371–372.
- ³ Az ELAR-minta átsúlyozását a KSH kétfajta eljárással oldotta meg. A korábbi gyakorlat – melyet egyébként a korábbi és az 1994. év első két negyedévében lefolytatott munkaerő-felvételek során mindvégig alkalmaztak – az volt, hogy a népszámlálás nem, életkor és lakóhely szerint bontott, továbbvezetett népességszámai alapján, a 15–74 éves korosztály egyéneire számítottak teljeskörűsítést és reprezentativitást egyszerre biztosító súlyokat. Nevezzük ezt a fajta – a minta egyéneihez rendelt – súlyt w_{kor1} -nek. 1994 júliusában a KSH másfajta – háztartás- és egyéni szintű – súlyozási eljárásra tért át. Ebben a tanulmányban, amikor súlyozott adatokról beszélünk, ez utóbbi (w_{kor2}) súlyokat használjuk. Megjegyezzük, hogy az általunk vitatott publikációban (*A cigányság helyzete*) a KSH szakértői is ezt a súlyozási eljárást alkalmazták.
- ⁴ A szóban forgó adatfelvételek file-jait az adatfelvétel készítőjétől, Ladányi Jánostól vásároltuk meg.
- ⁵ Természetesen nem állíthatjuk, hogy a „teljes körű” összeírást száz százalékosan sikerült megoldanunk. Mindazonáltal – mint majd azt a 4. pontban látni fogjuk – ehhez az ideális eredményhez meglehetősen közel jutottunk. Az ország egészét tekintve 90 százaléknál

minden bizonnyal nem sokkal rosszabb összeírási arányt érthetünk el. Lásd a 16. táblázat (4)-es sorát.

- ⁶ Pontosabban: a 15000-nél kisebb lélekszámú községek képezik az összehasonlítható településhalmazt. Ezek a települések az ELAR-mintába is véletlen kiválasztással kerültek be. A 15000 fős, illetve annál nagyobb lélekszámú települések teljes körben képviseltetik magukat az ELAR-mintában.
- ⁷ A tesztet a városokra sajnálatos módon nem terjeszthetjük ki, hiszen a két felvétel esetében a kisebb mintavételi egységek más-más aggregátumok: a Cigányvizsgálat esetében az önkormányzati választókörizetek, az ELAR-felvétel esetén pedig a népszámlálási számlálókörzetek voltak. Itt meg kell jegyeznünk azt, hogy az önkormányzati választókörizeteket mint a városokban történő mintavétel kisebb területi egységeit kényszerűségből alkalmaztuk. Jobban szerettünk volna mi is népszámlálási számlálókörzeteket használni. Ehhez ismernünk kellett volna a kiválasztott (városokkal, községekkel együtt) mintegy 300 településen belüli számlálókörzetek pontos határait. Az ehhez szükséges térképek azonban egyedül a KSH-nak állnak rendelkezésre, ahonnan ésszerű költségekért fénymásolatokat nem sikerült beszerezni. Így kényszerűségből kisebb területi egységekként az önkormányzati választókörizeteket használtuk, melyeknek területi határait valamennyi település önkormányzatánál pontosan ismerik. Ezt az információt kivétel nélkül minden érintett önkormányzat ingyen bocsátotta a rendelkezésünkre. Nagyvonalúságukat ezúton szeretnénk – a kutatás valamennyi irányítójának nevében – megköszönni.
- ⁸ A pót címeiket a falvak esetében az adott településről, a többi településkategória (Budapest, Miskolc, megyeszékhely, egyéb város) esetében pedig az adott település megfelelő sűrűségű (ritka, közepes, illetve sűrű) alkörzetéből választottuk. Az adott településrész pót címeit ezután megkevertük, majd a megkevert halmaz véletlenszerűen egymás után következő elemeit sorszámmal láttuk el, és a kérdezőbiztos számára előírtuk, hogy meghíúsulás esetén mindig e sorszám szerint sorban következő pót címet vegye igénybe. A keveréssel a községen, illetve az alkörzeten belüli pót cím-mintát randomizáltuk, a kötelező sorrendiség előírásával pedig e véletlenszerűség követésére készítettük kérdezőbiztosainkat. A pót címek száma település-alkörzetenként megegyezett az eredeti címek számával.
- ⁹ Az 1993. évi ELAR szeptemberi–októberi–novemberi hullámában a 26 604 fölkeresett háztartás közül a kikérdezés 4186 esetben különböző okoknál fogva meghíúsult. A meghíúsulási arány az eredeti minta 15,73 százalékát tette ki. A súlyozatlan háztartásmintaelemszám 22 418 volt.
- ¹⁰ Az Éltető és Marton (1995) által hivatkozott irodalom szerint a meghíúsulások által okozott mintatorzítások korrigálására sok szempontból jobb megoldást nyújt az utólagos átsúlyozás, mint a pótminta használata. A cigány minta esetében azonban ez a dilemma föl sem merülhet, ez esetben ugyanis nem áll rendelkezésre olyan teljes körű információ, amelyre az átsúlyozást alapozhatnánk.
- ¹¹ Az említett statisztikai kiadványok a következők: *Cigány tanulók* 1971; 1978; 1982; 1986. Az 1980/81. tanévtől az 1992/93. tanévig vezetett iskolastatisztikai adatok forrása a Művelődésügyi Minisztérium Statisztikai Osztályának nyilvántartása.
- ¹² Ez minden valószínűség szerint inkább alsó, mint felső becslésnek tekinthető. Aligha lehet gyakori esemény ugyanis az, hogy a környezetük által tipikusan nem cigánynak tekintett gyerekeket általános iskolai tanáraik cigány származásúnak minősítsenek. Inkább a fordított eset a valószínűbb: az ti., hogy az asszimilálódott családok gyermekeit tanáraik nem tekintik cigány származásúaknak.
- ¹³ Lásd például Grossman 1976.
- ¹⁴ Lásd például Orosz 1993: 238.

- ¹⁵ Lásd *Cigány tanulók* 1971; 1978; 1982; 1986.
- ¹⁶ Mai értelemben vett szakmunkásképző tanintézmények csak a hatvanas évek óta léteznek.
- ¹⁷ Ez az arányszám a nyolc osztályt nappali tagozaton végzettek és a nyolc általános végzettségűek hányadosaként állt elő; számszerűen pedig $1898/2101=0,901$ volt. Az ELAR-mintha korrigálatlan adata a nyolc általános iskolai végzettségűekre 117 585 fő volt. A nappali tagozaton végzettek becsült adata így az előbbi szám 90,1 százaléka, azaz: 105944 fő lett. A két szám különbözetét, azaz azt a 11 641 embert, aki nyolc osztályos iskolai végzettségét nem nappali tagozaton szerezte meg, átsoroltuk az 1–7. osztályos kategóriába.
- ¹⁸ A korcsoportok szerinti megoszlást az 1–7. és a 8. osztály közötti, ELAR-mintán belüli átsorolások természetesen nem befolyásolják.
- ¹⁹ Hogy egy összehasonlítást mégis megkockáztassunk: míg az 1957/58-as tanévben (ekkor lehetett elvileg nyolcadikos az 1993-ban 43 éves emberünk) az ország általános iskoláiban tanuló összes cigány gyereknek 11,7 százaléka járt felső tagozatú osztályba (lásd *Cigány tanulók* 1971: 5.), addig az 1981/82-es tanévben (ekkor lehetett elvileg nyolcadikos az 1993-ban 24 éves ember) a felső tagozatosok aránya már 40,6 százalék volt (lásd *Cigány tanulók* 1982: 12.).

Hivatkozások

- Cigány tanulók 1971. *Cigány tanulók az általános iskolában (1970/1971. tanév)*. Budapest: Művelődésügyi Minisztérium Statisztikai Osztály (csak belső használatra)
- Cigány tanulók 1978. *Cigány tanulók az általános iskolában (1977/1978. tanév)*. Budapest: Egyetemi Számítóközpont (csak belső használatra)
- Cigány tanulók 1982. *Cigány tanulók az alsó és középfokú oktatási intézményekben (1981/1982. tanév)*. Budapest: Tudományszervezési és Informatikai Intézet (csak belső használatra)
- Cigány tanulók 1986. *Cigány tanulók az alsó és középfokú oktatási intézményekben (1985/1986. tanév)*. Budapest: Tudományszervezési és Informatikai Intézet (csak belső használatra)
- Éltető Ödön–Marton Ádám 1995. „A mintanagyság és a meghíúsulások kapcsolata a reprezentatív adatfelvételekben”. *Statisztikai Szemle*, október
- Grossman, Michael 1976. The Correlation between Health and Schooling. In: N. Terleckyj (ed.), *Household Production and Consumption*. New York: Columbia University Press
- Kemény István–Havas Gábor–Kertesi Gábor 1994. *Beszámoló a magyarországi roma (cigány) népesség helyzetével foglalkozó, 1993 októbere és 1994 februárja között végzett kutatásról*. Budapest: ILO-programiroda részére készült tanulmány. Kézirat.
- Kertesi Gábor 1995. Cigány gyerekek az iskolában, cigány felnőttek a munkaerőpiacon. *Közgazdasági Szemle*, január
- Mészáros Árpád–Czibulka Zoltán–Fóti János–Nagy Orbán–Szűcs Zoltán 1994. *A cigányság helyzete, életkörülményei 1993*. Budapest: KSH
- Munkaerő-felmérés 1994. *Munkaerő-felmérés, 1993 (adattár)*. Budapest: KSH
- Orosz Éva 1993. A magyar egészségügy területi egyenlőtlenségei. In: Enyedi Gy. (szerk.), *Társadalmi-területi egyenlőtlenségek Magyarországon*. Budapest: KJK
- Prékopa András 1962. *Valószínűségelmélet műszaki alkalmazásokkal*. Budapest: Műszaki Könyvkiadó

FÜGGELÉK

F1. táblázat

Az általános iskolai cigány tanulók száma évfolyamonként

tanév	I.	II.	III.	IV.	V.	VI.	VII.	VIII.	összesen
1969/70	14908	10024	9659	7693	7769	5215	3899	2695	61862
1970/71	15016	9763	9086	7356	7442	4976	3548	2409	59596
1971/72	.	.	.	.	.	.	.	.	.
1972/73	.	.	.	.	.	.	.	.	.
1973/74	12927	9347	8535	7569	7457	5592	4033	2823	58283
1974/75	13833	9306	8607	7860	7870	5834	4429	3250	60989
1975/76	.	.	.	.	.	.	.	.	.
1976/77	13516	10018	8957	7910	8042	6642	5122	3752	63959
1977/78	13833	10232	9441	8531	8397	6795	5739	4118	67081
1978/79	.	.	.	.	.	.	.	.	.
1979/80	.	.	.	.	.	.	.	.	.
1980/81	13198	10296	9127	8827	9410	7670	6397	4774	69699
1981/82	14015	10345	9564	8849	9675	7829	6657	5075	72009
1982/83	13948	10921	9662	9157	9605	8064	6810	5290	73457
1983/84	13802	10933	10097	9467	9548	7849	6820	5087	73603
1984/85	13726	10676	10043	10020	9932	7958	6676	5236	74267
1985/86	13366	10732	10018	10188	10379	8498	6797	5190	75168
1986/87	13321	10382	9923	10157	10201	8774	6960	5123	74841
1987/88	11552	10483	9732	10175	10327	8801	7316	5514	73900
1988/89	12076	9790	9988	9991	10258	8794	7231	5734	73862
1989/90	12400	9870	9629	10087	10015	8833	7294	5697	73825
1990/91	12206	9989	9388	9423	10158	8455	7160	5594	72373
1991/92	12539	10068	9611	9590	9563	8711	7078	5557	72717
1992/93	13104	10285	9797	9694	9970	8409	7403	5779	74441

Forrás: 1970/71, 1974/75. és 1977/78. tanév végi adatok: *Cigány tanulók az általános iskolában*. Egyetemi számítóközpont, Budapest 1978. 20. o.

1969/70, 1973/74. és 1976/77. tanév végi adatok: becslült adatok forrása Kertesi (1995) B Függelék

1980/81. és 1992/93. közötti adatok: Művelődési és Köznevelési Minisztérium, Statisztikai Osztály

*F2. táblázat**A szakmunkásképzőbe járó cigány tanulók száma évfolyamonként*

tanév	I.	II.	III.	összesen
1981/82	1573	1113	853	3539
1982/83	1693	1166	996	3855
1983/83	1644	1066	953	3663
1984/85	1606	1157	996	3759
1985/86	1745	1092	944	3781
1986/87	1765	1200	907	3872
1987/88	2009	1285	1004	4298
1988/89	2071	1327	1060	4458
1989/90	1925	1278	1134	4337
1990/91	1682	1237	1030	3949
1991/92	1455	1092	871	3418
1992/93	1488	990	858	3336

Forrás: Művelődési és Köznevelési Minisztérium, Statisztikai Osztály*F3. táblázat**A gimnáziumi, szakközépiskolai és technikumi cigány tanulók száma évfolyamonként*

tanév	I.	II.	III.	IV.	összesen
1980/81	163	144	104	96	507
1981/82	181	147	111	89	528
1982/83	146	142	109	105	502
1983/84	160	128	138	91	517
1984/85	150	127	98	122	497
1985/86	158	136	104	89	487
1986/87	149	140	126	95	510
1987/88	169	137	128	119	553
1988/89	188	159	119	113	579
1989/90	181	141	123	91	536
1990/91	166	140	118	99	523
1991/92	166	157	116	96	535
1992/93	203	148	120	110	581

Forrás: Művelődési és Köznevelési Minisztérium, Statisztikai Osztály