

A 2011. évi népszámlálásból származó oktatási célú mikroadatfájl¹ módszertani dokumentációja

1. Az oktatási célú népszámlálási mikroadatfájl

A KSH célja, hogy az új felhasználói igényeket folyamatosan kövesse, és azokat a lehető leggyorsabban, illetve a legnagyobb mértékben kielégítse. Manapság egyre szélesebb körből jelentkezik igény a felvételeinkből származó mikroadatfájlok iránt. Ilyen állományok „nyers” közlése adatvédelmi okokból nem lehetséges, ezért vagy a felhasználók körét és a felhasználás helyét kell nagyon pontosan behatárolni, majd a kutatási eredményeket adatvédelmi ellenőrzésnek alávetni, vagy a mikroadatfájlokat kell olyan mértékben anonimizálni, hogy azok széles körű elérhetőségük ellenére se tegyék lehetővé az adatszolgáltatóink beazonosítását, illetve valamely jellemzőjük felfedését.

A KSH 2007 óta biztosít kutatási lehetőséget tudományos célból részletes, közvetlen azonosításra alkalmatlan mikroadat-állományokon a KSH kutatószobájában. A kutatás megkezdéséhez kutatási tervvel és tudományos háttérrel kell rendelkezni. Egyedi igényre előkészített állományok rendelkezésre bocsátása esetén a KSH költségtérítést számíthat fel.²

A szélesebb felhasználói kör felé való nyitás első lépcsőjeként 2014 márciusában közzétettünk a 2011. évi népszámlálási állományokból készült ún. *tesztállományokat*.³ Noha ezek az állományok bárki számára szabadon letölthetőek, a megcélzott felhasználók körét azok a kutatók jelentik, akik biztonságos környezetben (kutatószobai hozzáférés,⁴ távoli hozzáférés,⁵ távoli végrehajtás⁶) kívánják elemezni a népszámlálási állományokat. A tesztállományok ugyanis a programfájlok előzetes megírását, tesztelését hivatottak elősegíteni, elemzési, kutatási célra alkalmatlanok.

Az **oktatási célú** mikroadatfájl is a 2011. évi népszámlálás adatállományyaiból készült. Elsősorban oktatási intézmények hasznosíthatják a statisztikai kurzusok során, de ajánljuk mindazoknak, akik a statisztikai módszerekkel szeretnének a gyakorlatban is megismerkedni.

¹ Rekordok sorozatából álló állomány, megfigyelési egységek (itt: személyek) adatait tartalmazza.

² A részletekért ld. a http://www.ksh.hu/adatigenyles_nyito oldalon az *Ingyenes és térítésköteles szolgáltatások* menüpontot.

³ <http://www.ksh.hu/nepszamlalas/tesztallomanyok>

⁴ http://www.ksh.hu/kutatooszobai_hozzaferes

⁵ http://www.ksh.hu/tavoli_hozzaferes

⁶ http://www.ksh.hu/tavoli_vegrehajtás

Az állomány

- 10 000 rekordot – kb. 1‰ (egy ezrelék)-es kiválasztási arány – és
- 13 változót⁷

tartalmaz. A változók definíciója és a nomenklatúrák a „*Rekordleírás_Népszámlálás 2011_Oktatási mikroadatfájl.xlsx*” állományban találhatóak meg.

Az oktatási célú népszámlálási mikroadatfájl részletezettsége sokkal kisebb a KSH kutatószobájában biztonságos környezetben kutatható 10%-os személyi és lakásmintához képest. Az oktatási célú mikroadatfájl nagy előnye viszont az, hogy könnyebben hozzáférhető: a felhasználási feltételek elfogadása után a KSH honlapjáról bárki szabadon letöltheti.

2. Mintavétel

A 10 000 elemű személyi minta a teljes népszámlálási lakónépességből (9 937 628 fő) választott valószínűségi minta. A személyek kiválasztása rétegzés nélkül, egy lépcsőben, visszatevés nélküli egyszerű véletlen módon történt a SAS *surveysselect* eljárása segítségével. A kiválasztás egyszerű módjának köszönhetően a mintából számított becslések pontossága könnyedén jellemezhető a jól ismert képlet szerinti szórás segítségével.

3. Adatvédelem

A KSH maximálisan elkötelezett az adatszolgáltatóktól begyűjtött egyedi adatok védelme mellett. Az adatközléseink a statisztikáról szóló 1993. évi XLVI. törvénnyel, az ennek végrehajtásáról szóló 170/1993. (XII. 3.) Korm. rendelettel, továbbá az információs önrendelkezési jogról és az információszabadságról szóló 2011. évi CXII. törvénnyel összhangban történnek. Annak érdekében, hogy a mikroadatfájlban szereplő személyek anonimek maradjanak, a következő intézkedéseket tettük:

- *Alacsony (kb. 1‰) kiválasztási arányt* alkalmaztunk. Ha egy felhasználó egy olyan rekordot talál az állományban, amelynek – bizonyos jellemzői alapján – ő maga vagy egy általa ismert személy megfelel, akkor elhanyagolható a valószínűsége annak, hogy valóban önmagát vagy az általa ismert személyt azonosította be. Ennek oka az, hogy a beazonosítást lehetővé tevő kulcsokra legalább három előfordulás található a sokaságban. Ezért egyáltalán nem biztos, hogy a beazonosítani vélt személy az, akire a felhasználó gondol, és nem egy másik, hasonló tulajdonságokkal rendelkező személy.
- *Közvetlen azonosítót nem tartalmaz az állomány.*

⁷ 10 változó személyi jellemző, 3 pedig lakásjellemző

- A közvetett azonosítónak⁸ (pl. pontos születési dátum, lakóhely települése, korév) minősülő vagy a különleges adat⁹ (pl. vallás, fogyatékoság, nemzetiség) fogalmába tartozó változókat eltávolítottuk vagy a kérdőíven megadotthoz képest csak kevésbé részletes formában közöltük (pl. lakóhely csak a NUTS2 régió szintjén jelenik meg).

4. A népszámlálási mikroadatfájl felhasználásának korlátai

Mivel az oktatási célú népszámlálási mikroadatfájl egy kb. 1 ezrelékes egyszerű véletlen minta, ezért az állományból készített becslések (pl. átlagos iskolai végzettség) mintavételi hibával terhelték. Az elemzések eredményeinek értelmezésekor, ill. a sokasági jellemzőre való következtetések levonásakor mindig figyelemmel kell lenni a mintavételből eredő hibára.

Bár az egyváltozós eloszlások elég nagy pontossággal visszaadják a sokaság értékeit,¹⁰ amennyiben valaki tudományos célból kutatná a népszámlálásból kapott válaszokat, annak azt javasoljuk, hogy fontolja meg a biztonságos környezetben történő hozzáférést. Ebben az esetben, ahogy már korábban is említettük, jóval részletesebb adatokat tudunk rendelkezésre bocsátani (10%-os minta a személyi, lakás-, család- és háztartásállományból, nem aggregált változók – pl. korév korcsoport helyett –, a kérdőíveknek megfelelő teljes változókészlet).

Amennyiben a mikroadatfájl használatával, a róla közreadott információkkal kapcsolatban kérdése merül fel vagy észrevételt kíván tenni, úgy kérjük, forduljon hozzánk bizalommal a velemeney@ksh.hu elérhetőségen keresztül.

⁸ A mikroadatállomány olyan változója, amely a kulcs részeként segíthet a statisztikai egységek azonosításában.

⁹ A különleges adat definícióját ld. a 2011. évi CXII. törvény 3. § 3.) pont alatt

¹⁰ Lásd a Függelékét.

Függelék: Az oktatási célú népszámlálási mikroadatfájl reprezentativitása

Illusztratív céllal ismertetjük a mikroadatfájlból származtatott egyváltozós megoszlásokat, és összevetjük a népszámlálás teljes körű állományából számított megoszlási jellemzőkkel.¹¹

- Régió (REGIO)

REGIO	Oktatási fájl megoszlása, %	Sokaság megoszlása, %	Különbség, százalékpont
1	29,8	29,7	0,1
2	11,1	10,9	0,2
3	9,7	9,9	0,2
4	9,2	9,4	0,2
5	12,5	12,1	0,4
6	14,8	15,0	0,2
7	13,0	13,1	0,1

- Nem (NEME)

NEME	Oktatási fájl megoszlása, %	Sokaság megoszlása, %	Különbség, százalékpont
1	47,8	47,5	0,3
2	52,2	52,5	0,3

- Korcsoport (KORCSOP)

KORCSOP	Oktatási fájl megoszlása, %	Sokaság megoszlása, %	Különbség, százalékpont
1	10,0	9,6	0,4
2	11,0	10,9	0,1
3	12,4	12,4	0,0
4	15,9	15,9	0,0
5	13,1	13,2	0,1
6	14,1	14,5	0,4
7	11,9	11,8	0,1
8	7,5	7,6	0,1
9	4,1	4,0	0,1

¹¹ A megoszlások összege a kerekítések miatt nem minden esetben pontosan 100%.

- Családi állapot (CSPOT)

CSPOT	Oktatási fájlmegoszlása, %	Sokaságmegoszlása, %	Különbség, százalékpont
1	42,6	42,4	0,2
2	38,0	37,9	0,1
3	9,8	9,8	0,0
4	9,5	9,9	0,4

- Élve született gyermekek száma (EGYERUJ)

EGYERUJ	Oktatási fájlmegoszlása, %	Sokaságmegoszlása, %	Különbség, százalékpont
0	42,4	42,2	0,2
1	17,8	17,8	0,0
2	27,9	27,9	0,0
3	8,7	8,6	0,1
4	2,1	2,2	0,1
5	0,7	0,7	0,0
6	0,5	0,6	0,1

- Iskolába jár-e? (TANUL)

TANUL	Oktatási fájlmegoszlása, %	Sokaságmegoszlása, %	Különbség, százalékpont
0	77,9	78,2	0,3
1	22,1	21,8	0,3

- Legmagasabb befejezett iskolai végzettsége (ISKOLA)

ISKOLA	Oktatási fájlmegoszlása, %	Sokaságmegoszlása, %	Különbség, százalékpont
1	18,8	18,3	0,5
2	23,1	23,3	0,2
3	18,2	18,1	0,1
4	25,2	25,7	0,5
5	14,7	14,5	0,2

- Gazdasági aktivitás (GAKT)

GAKT	Oktatási fajtól megoszlása, %	Sokaságtól megoszlása, %	Különbség, százalékpont
1	39,9	39,7	0,2
2	5,6	5,7	0,1
3	29,5	29,7	0,2
4	25,0	24,9	0,1

- Jelenlegi foglalkozása, munkaköre (FOGLKOD)

FOGLKOD	Oktatási fajtól megoszlása, %	Sokaságtól megoszlása, %	Különbség, százalékpont
1	8,6	8,3	0,3
2	9,4	9,5	0,1
3	6,3	6,4	0,1
4	1,2	1,1	0,1
5	10,7	10,3	0,4
6	3,8	4,1	0,3
üres	60,1	60,3	0,2

- Családi összetétel (CSOSSZ)

CSOSSZ	Oktatási fajtól megoszlása, %	Sokaságtól megoszlása, %	Különbség, százalékpont
1	21,4	21,7	0,3
2	54,0	53,3	0,7
3	11,8	11,8	0,0
4	12,9	13,2	0,3

- A lakás építési éve (EPEV)

EPEV	Oktatási fajtól megoszlása, %	Sokaságtól megoszlása, %	Különbség, százalékpont
1	15,7	16,3	0,6
2	10,5	10,6	0,1
3	13,4	13,2	0,2
4	20,9	21,0	0,1
5	17,8	17,6	0,2
6	8,6	8,3	0,3
7	10,6	10,5	0,1
üres	2,5	2,4	0,1

- A lakás alapterülete (LAT)

LAT	Oktatási fajt megoszlása, %	Sokaság megoszlása, %	Különbség, százalékpont
1	7,3	7,3	0,0
2	6,5	6,3	0,2
3	14,7	14,5	0,2
4	20,9	21,0	0,1
5	21,4	21,3	0,1
6	29,2	29,5	0,3

- A lakásban lakó személyek száma (LLAKO)

LLAKO	Oktatási fajt megoszlása, %	Sokaság megoszlása, %	Különbség, százalékpont
1	11,5	11,7	0,2
2	22,2	22,8	0,6
3	23,3	23,2	0,1
4	21,5	21,1	0,4
5	10,5	10,4	0,1
6	11,1	10,8	0,3