

A 2022 évi népszámlálásból származó oktatási célú mikroadat-állomány¹ módszertani dokumentációja

A KSH célja, hogy az új felhasználói igényeket folyamatosan kövesse, és a lehető leggyorsabban, ill. a legnagyobb mértékben kielégítse azokat. Manapság egyre szélesebb körből jelentkezik igény a felvételeinkből származó mikroadat-állományok 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 mikroadat-állományokat kell olyan mértékben anonimizálni, hogy azok a széles körű elérhetőségük ellenére se tegyék lehetővé az adatszolgáltatóink beazonosítását, illetve hozzájuk kapcsolódó új információk felfedését.

A KSH 2007 óta biztosít kutatási lehetőséget tudományos célból részletes, közvetlen azonosításra alkalmatlan mikroadat-állományokon a KSH központi kutatószobájában. A kutatás megkezdéséhez kutatási tervvel és tudományos háttérrel kell rendelkezni. Egyedi igényre előkészített állományok rendelkezésre bocsátása esetén a KSH költségtérítést számíthat fel². A kutatók számára a kutatószobai környezetben rendelkezésre áll a 2011-es és a 2022-es népszámlálási állomány 10%-os mintája.

A szélesebb felhasználói kör felé való nyitás első lépcsőjeként 2014 márciusában közzétettünk a 2011-es népszámlálási állományokból készült ún. *tesztállományokat*³ és 2024-ben a 2022-es népszámlálásból kialakított tesztállományt is. Noha ezek az állományok bárki számára szabadon letölthetőek, a megcélzott felhasználók körét azok a kutatók jelentik, akik biztonságos környezetben (kutatószobai hozzáférés, távoli végrehajtás) kívánják elemezni a népszámlálási állományokat. A tesztállományok ugyanis a programfájlok előzetes megírását, tesztelését hivatottak elősegíteni, elemzési, kutatási célra alkalmatlanok.

1. Az oktatási célú népszámlálási mikroadat-állomány

Az **oktatási célú** mikroadat-állományt elsősorban oktatási intézmények hasznosíthatják a statisztikai kurzusok során, de ajánljuk mindazoknak, akik a statisztikai módszerekkel szeretnének a gyakorlatban is megismerkedni.

A 2022-es népszámlálásból létrehozott oktatási állomány

- 10 000 rekordot – kb. 1‰ (egy ezrelék)-es kiválasztási arány – és
- 13 változót⁴

tartalmaz. A változók definíciója és a nomenklatúrák a '*Rekordleírás_Népszámlálás 2022_Oktatási mikroadat-állomány.xlsx*' mikroadat-állomány' állományban találhatóak meg.

Az oktatási célú népszámlálási mikroadat-állomány részletezettsége jelentősen alacsonyabb a KSH kutatószobájában biztonságos környezetben kutatható 10%-os személyi- és lakásmintához képest. Az

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

² A részletekért ld. https://www.ksh.hu/kutatoknak_kutatooszoba

³ https://www.ksh.hu/kutatoknak_nyilvanos_mikroadat_allomanyok

⁴ 10 változó személyi jellemző, 3 pedig lakásjellemző

oktatási célú mikroadat-állomány 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 (9 603 634 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 2016. évi CLV. törvénnyel, az ennek végrehajtásáról szóló 184/2017. (VII. 5.) Korm. rendelettel, továbbá az információs önrendelkezési jogról és az információszabadságról szóló 2011. évi CXII. törvénnyel összhangban történnek. Annak érdekében, hogy a mikroadat-állományban szereplő személyeket ne lehessen azonosítani, a következő intézkedéseket tettük:

- *Alacsony (kb. 1‰) kiválasztási arányt* alkalmaztunk. Ha egy felhasználó egy olyan rekordot talál az állományban, melynek – 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.*
- *A közvetett azonosítónak⁵ (pl. pontos születési dátum, lakóhely települése, korév) minősülő vagy a különleges adat⁶ (pl. vallás, fogyatékoság, nemzetiség) fogalmába tartozó változókat eltávolítottuk vagy a kérdőívben megadotthoz képest csak kevésbé részletes formában közöltük (pl. lakhely csak NUTS2 régió szinten jelenik meg).*

4. A népszámlálási mikroadat-állomány felhasználásának korlátai

Mivel az oktatási célú népszámlálási mikroadat-állomány egy kb. 1‰-es egyszerű véletlen minta, ezért az állományból készített becslések (pl. átlagos iskolai végzettség) *mintavételi hibával* terheltek. 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 sokaságban megfigyelhető értékeket⁷, amennyiben valaki tudományos célból kutatná a Népszámlálásból kapott válaszokat, javasoljuk, hogy fontolja meg a biztonságos környezetben történő hozzáférést. Ebben az esetben, ahogy már korábban is említettük, jóval részletesebb adatokat tudunk rendelkezésre bocsátani (10%-

⁵ Mikroadat-állomány olyan változója, mely kulcs részeként segíthet a statisztikai egységek azonosításában

⁶ A különleges adat definícióját ld. a 2011. évi CXII. törvény 3. § 3.) pont alatt

⁷ Ld. 1. Függelék.

os minta a személyi, lakás-, család- és háztartásállományból, nem aggregált változók – pl. korév korcsoport helyett –, a kérdőíveknek megfelelő teljes változókészlet).

Függelék: Az oktatási célú népszámlálási mikroadat-állomány reprezentativitása

Illusztratív céllal ismertetjük a mikroadat-állományból származtatott egyváltozós megoszlásokat, és összevetjük a népszámlálás teljeskörű állományából számított megoszlási jellemzőkkel⁸.

- Régió (REGIO)

REGIO	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
11	17,4	17,5	0,1
12	13,6	13,9	0,3
20	11,0	11,0	0,0
30	10,4	10,2	0,2
40	8,8	8,9	0,1
50	11,6	11,4	0,2
60	14,9	14,6	0,3
70	12,3	12,5	0,2

- Nem (NEME)

NEME	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	48,3	48,1	0,2
2	51,7	51,9	0,2

- Korcsoport (KORCSOP)

KORCSOP	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	9,4	9,7	0,3
2	10,1	9,9	0,2
3	12,0	11,2	0,8
4	12,2	12,9	0,7
5	16,8	16,4	0,4
6	13,4	13,5	0,1
7	12,5	12,7	0,2
8	9,0	9,3	0,3
9	4,6	4,5	0,1

- Családi állapot (CSPOT)

CSPOT	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	44,3	44,1	0,2
2	36,9	37,1	0,2
3	8,4	8,4	0,0
4	10,4	10,5	0,1

⁸ A megoszlások összege a kerekítések miatt nem minden esetben pontosan 100%.

- Élve született gyermekek száma (EGYER)

EGYER	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
0	43,0	43,0	0,0
1	17,9	18,1	0,2
2	27,1	27,0	0,1
3	8,6	8,8	0,2
4	2,1	2,1	0,0
5	0,8	0,6	0,2
6	0,4	0,4	0,0

- Iskolába jár-e? (TANUL)

TANUL	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
0	83,6	84,0	0,4
1	16,4	16,0	0,4

- Iskolai végzettsége (ISKOLA)

ISKOLA	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	15,9	16,0	0,1
2	18,8	18,6	0,2
3	17,3	18,1	0,8
4	28,6	28,3	0,3
5	19,4	19,1	0,3

- Gazdasági aktivitás (GAKT)

GKAT	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	49,3	49,1	0,2
2	2,5	2,5	0,0
3	23,5	23,6	0,1
4	24,8	24,8	0,0

- Jelenlegi foglalkozása, munkaköre (FOGLKOD)

FOGLKOD	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	12,1	12,0	0,1
2	12,0	12,0	0,0
3	7,1	6,9	0,2
4	1,1	1,1	0,0
5	5,8	5,9	0,1
6	11,2	11,2	0,0
	50,7	50,9	0,2

- Családi összetétel (CSOSSZ)

CSOSSZ	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	24,8	25,0	0,2
2	48,0	48,2	0,2
3	12,9	12,7	0,2
4	14,4	14,0	0,4

- Lakás építési éve (EPEV)

EPEV	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	13,5	14,2	0,7
2	11,3	11,0	0,3
3	33,0	33,3	0,3
4	24,1	24,1	0,0
5	10,3	9,9	0,4
6	6,0	5,7	0,3
	1,7	1,7	0,0

- A lakás alapterülete (LAT)

LAT	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	4,2	4,6	0,4
2	6,0	6,0	0,0
3	13,1	13,0	0,1
4	20,4	19,9	0,5
5	20,5	20,5	0,0
6	34,2	34,2	0,0
	1,7	1,7	0,0

- Lakásban lakó személyek száma (LLAKO)

LLAKO	Oktatási fájl megoszlása, %	Sokasági fájl megoszlása, %	Különbség, százalékpont
1	14,8	14,8	0,0
2	22,7	23,4	0,7
3	21,3	20,8	0,5
4	19,5	19,5	0,0
5	10,5	10,5	0,0
6	9,4	9,3	0,1
	1,7	1,7	0

Amennyiben a mikroadat-állomány használatával, a róla közreadott információkkal kapcsolatban kérdése merül fel vagy észrevételt kíván tenni, úgy kérjük, forduljon hozzánk bizalommal.

Módszertani főosztály
2024. június 6.