

Útmutató az 1971-es cigányfelvétel egyéni adatbázisához

2014. szeptember 11.

Kemény István vezetésével 1971-ben reprezentatív adatfelvétel készült a magyarországi cigányság életkörülményeiről. A felvett anyagból – melynek kérdőívei részben megsemmisültek – a későbbiekben számítógépes fájlok keletkeztek.

Az adatbázishoz felhasznált források

1) MTA Szociológiai Kutató Intézet (MTA SZKI). *Egyéni* adatbázis, a megfigyelt személyek száma 2608, a változósám 205. Az adatbázisból hiányoznak két megye (Pest és Baranya) és három nagyváros (Debrecen, Pécs és Győr) megfigyelései, anyaguk évtizedekkel ezelőtt elveszett. Az egyéni adatbázis a 14 éven felüli családtagok személyi adatait tartalmazza. Ez a file a fellelhető kérdőívek utólagos rögzítésével keletkezett az MTA Szociológiai Intézetében, Fehérvári Anikó irányításával.

2) TÁRKI. Egy 1120 háztartás adatait tartalmazó file, ahol a változók száma 2030. A háztartási adatbázis a gyerekekre és egyéb családtagokra vonatkozó – a megkérdezett személy által szolgáltatott – információkat is tartalmaz, és szerepelnek benne az MTA SZKI adatbázisból hiányzó megyék és nagyvárosok adatai. A megfigyelés egysége egy megkérdezett személy egy lyukkártyája, a rekordok száma 30 733. Egy-egy kérdőív (megkérdezett személy) anyagát maximálisan 16 lyukkártyára vitték fel, de nem minden megkérdezett személyhez tartozik 16 lyukkártya, és hiányzik az egyéni azonosító. Ezt az anyagot 1989-ben, a fellelhető lyukkártyákról vitték gépre az MTA Szociológiai Intézetében, onnét került a TÁRKI nyilvános adatbankjába *TDATA-B10: A magyarországi cigányság életkörülményei 1971. adatlap. Verzió: 2007-01-19* néven.

3) MTA Közgazdaságtudományi Intézet (MTA KTI). A TÁRKI-ból származó, fent leírt háztartási file-t 1993-ban Juhász Éva, az MTA KTI munkatársa javította. Egyéni azonosítót képzett, háztartások szerint rendezte az adatbázist, és nagyszámú duplumot távolított el.

Az egyéni adatbázis tisztítása és kiegészítése

Az MTA SZKI adatbázis változóinak ellenőrzését és tisztítását Máthé Judit és Köllő János végezték az eredeti kódutasítás és kérdőív alapján, az MTA (KRTK) Közgazdaságtudományi Intézetében, 2013-2014-ben.

- A javítás kiterjedt a változók és változóértékek címkézésére, új, jobban használható változók képzésére, a szöveghibák javítására. Kísérletet tettünk a válaszok logikai összeférhetetlenségeinek kiküszöbölésére, vagy ha erre nem volt lehetőség, jelzésére. A tisztítás során észlelt inkonzisztenciákat és a változókhoz fűzött magyarázatokat jelen dokumentum mellékletében foglaltuk össze. A megjegyzések a Stata 13 formátumú file jegyzeteiben is megtalálhatók és előhívhatók a *notes list* illetve változónként a *notes list [változónév]* paranccsal.
- Az MTA KTI file-ban generáltuk az MTA SZKI file változóit. Az így létrehozott file felhasználásával sikerült kiegészítenünk az anyagot Pest és Baranya megyei megfigyelésekkel. Debrecen, Pécs és Győr adatai azonban továbbra is hiányoznak. Az egyéni megfigyelések számát 2608-ról 2912-re sikerült emelni.
- Az MTA KTI file-ból számos, a lakásra, valamint a háztartás összetételére és életkörülményeire vonatkozó változót emeltünk át az egyéni file-ba (v1720-v2020

változók). Az egyéni file-ban szereplő 2912 egyénből 2628-hoz sikerült háztartási változókat rendelni.

- Az elkészült adatbázis 2912 tizennégy évesnél idősebb személy adatait tartalmazza, akik 1056 egy fedél alatt élő háztartáshoz tartoznak.

A rendelkezésre álló minta

Az 1971-es felvételtől készült beszámoló táblázatai szerint a minta 6395 személy adatait tartalmazta, akik közül 3510 fő volt 14 évesnél idősebb. (Kemény István: Beszámoló a magyarországi cigányok helyzetével foglalkozó, 1971-ben végzett kutatásról. Lásd itt: http://www.sulinet.hu/oroksegtar/data/magyarorszagi_nemzetisegek/romak/beszamolo_a_magyarorszagi_ciganyok/)

Az általunk összeállított minta tehát 17 százalékkal kisebb. A különbség egy része gyanúnk szerint a duplumok eltávolításából adódik, más része valós veszteség. Az ebből adódó torzítás elvileg felmérhető a fenti kutatási jelentésben közölt és az itt közreadott mintából számított megoszlások összehasonlításával, és a torzítás korrekciós súlyozással mérsékelhető. Szándékunkban áll az adatbázist ilyen irányban tovább fejleszteni. Reméljük, hogy a jövőben sikerül támogatókat találni ennek a munkának az elvégzéséhez.

Az egyéni adatbázis használata

Az adatbázis pontos értelmezéséhez ajánlott az eredeti kódutasítás és kérdőív használata, melyek a <http://www.kemenyistvan.hu> honlapon érhetők el. Az adatbázis javított változata kérhető Stata 13 formátumban az MTA KRTK Adatbankjától, a szabo.endre@krtk.mta.hu címre küldött e-mailben. Szerepeljen a levélben az adatkérő neve és telefonszáma. Amennyiben szervezeti háttérrel is rendelkezik, úgy a szervezet pontos megnevezését is tüntesse fel. Az Adatbank vállalja az adatbázis konvertálását régebbi Stata verziókba, valamint SPSS, SAS és Excel formátumba. Az anyaggal illetve a konvertálással kapcsolatban információ kérhető az alábbi címeken:

Köllő János: kollo.janos@krtk.mta.hu

Máthé Judit: majudit@yahoo.com

Szabó Endre: szabo.endre@krtk.mta.hu