

MTA Statisztikai Bizottság ülése
2011. március 7. KSH Keleti Károly terem

E m l é k e z t e t ő

Az MTA Statisztikai Bizottsága márciusi ülésének napirendjén a mintán alapuló felvételek alapvető kérdése, a rétegzés szerepelt a KSH lakossági felvételeihez kapcsolódóan. Az előadó Fraller Gergely, a KSH szakmai tanácsadója volt, a korreferátumot Hunyadi László, a Statisztikai szemle főszerkesztője tartotta. Ezt követte a Bizottság jövőjét, feladatait érintő kérdések megvitatása. Az ülést Besenyei György, a bizottság elnöke vezette.

Fraller Gergely előadásának alapját a KSH-ban 2008-2010 időszakban a lakossági felvételek településrétegzése terén végzett kutatása adta. A kutatás a lakossági adatgyűjtések mintáinak harmonizálását célozta.

A KSH lakossági felvételeire jellemző a többlépcsős rétegzett mintavétel, ahol az első lépcső a településminta, az utolsó lépcső a mintába választott településen belül a címek kiválasztása. A különböző lakossági felvételek jelenlegi mintavételi gyakorlatában közös elem a megye és településnagyság szerinti rétegzés, de mindegyikben a jelentős egyedi megoldások a jellemzők.

A rétegzett mintavétel lényege, hogy a teljes sokaságot felosztjuk nem átfedő részsokaságokra, melyek uniója kiadja a teljes sokaságot. Az egyes részsokaságok a mintavételi rétegek. Egy adott rétegben a többi rétegtől függetlenül választunk mintát. Megfelelő rétegzéssel és allokációval akár jelentősen csökkenthető a mintából számított becslések szórása. Ha egy heterogén sokaságot sikerül homogén rétegekre bontani, akkor a homogén rétegeken belül egy kisebb mintával is pontos, az adott rétegre jellemző becslés adható. A pontos, rétegenkénti becslések felhasználásával pedig a teljes sokaságra vonatkozóan is pontosabb becslést kapunk.

Fraller Gergely először áttekintette, példával szemléltette a legfontosabb optimális rétegzési technikákat, ismertette a rétegzés hatékonyságát jellemző rétegzési indexet, mely megmutatja, hogy az összegbecslés szórásnégyzete hány százalékkal csökken a rétegzés hatására arányos allokáció esetén. A kutatás során a különböző rétegző változók és technikák tesztelését a 2001. évi népszámlálás adatai alapján végezte, ez lehetővé tette az elméleti szórások meghatározását. A különböző változók bevonásával, különböző módszerekkel és különböző rétegszámokkal jellemezhető több, mint 200 vizsgált tesztet eredményeit (rétegzési index értéke a munkanélküliek, foglalkoztatottak, felsőfokú végzettségűek, 60 évnél fiatalabb felsőfokú végzettségű háztartásfők száma változók esetében) a munkaerőfelvétel jelenlegi mintájával hasonlította össze. Az eredmények alapján a kanadai munkaerőfelvételnél használt, Drew, Bélanger és Foy (1985) nevéhez köthető eljárás bizonyult a legjobbnak, bebizonyosodott, hogy az egyes megyékben célszerű egymástól függetlenül kialakítani a rétegeket úgy, hogy még a rétegek száma sem egységes. A rétegek kialakításába a becslési kívánt jellemzővel összefüggésben lévő, más forrásból rendelkezésre álló változókat, konkrétan a regisztrált munkanélküliek számát és jövedelemváltozókat is célszerű bevonni. A kétlépcsős mintából számított becslés szórásnégyzetének csak az egyik (kisebb) hányada csökkenthető az első lépcsőben alkalmazott településrétegzés segítségével, a másik része a településrétegzéstől gyakorlatilag független, de különböző módszerekkel ez is csökkenthető. Így a vizsgálatok kiterjedtek (különböző rétegzések mellett) az optimális mintaallokációra is, azaz rétegenként a mintába kerülő településszámok, és a településeken

belüli mintaelemszámok meghatározására. Összeírás-szervezési szempontból cél volt a minél kevesebb település mellett biztosítani a becslések pontosságának adott szintjét. Az eredmények alapján megtervezésre került a jelenleginél hatékonyabb mintavételi terv, a jelenleginél jelentősen kevesebb település bevonásával. Ennek időbeli hatékonyságát a 2011. évi népszámlálás eredményei alapján lehet majd tesztelni, s megfelelő eredmények esetén ez lehet az alapja a KSH lakossági felvételei új, harmonizált mintavételi tervének.

Hunyadi László korreferátumában elismerően nyilatkozott a bemutatott kutatásról, kiemelve, hogy a szerző alaposan körbejárta a témát, mindvégig valóságkövető kutatást végzett, és az elméleti mellett gyakorlatban használható eredményeket állított elő. Más szóval a kutatás három kiemelkedő erőssége a szakirodalmi megalapozottság, a kiterjedt konkrét elemző munka valamint a gyakorlatban hasznosítható eredmény. Másik oldalról, negatívumként említette, hogy a 2001-es népszámlálási adatok – bár nincs újabb teljeskörű forrás – elég régiek. Célszerű lett volna a kutatásban kitérni az utólagos rétegzés kérdéskörére valamint a költségelemzésre, továbbá a szakirodalmi előzmények között a magyar források (ajánlható az MTMT adatbázis vagy az interneten elérhető Statisztikai Szemle) nem szerepeltek kellő súllyal. A szervezés hiányosságaként róható fel, hogy a Statisztikai Bizottság tagjai nem kapták meg előzetesen a számítások eredményeit bemutató táblázatokat. Végül kifejtette, hogy a Statisztikai Bizottságnak nem a KSH belügyeivel kell foglalkozni, de ez a kutatás – bár a KSH lakossági mintáival foglalkozott – tudományosan releváns, általánosan használható eredményeket hozott, így üdvözli, hogy a bizottság napirendjére tűzte. A kutatás eredményeit bemutató cikk a Statisztikai szemle 2011. áprilisi számában jelenik meg.

Az előadásokat hozzászólások követték. Mihályffy László a gyakorlati megvalósítással kapcsolatban vetett fel kérdéseket: milyen hatása lesz a meghíúsulásnak az eredményekre, kalibrálással vagy más módokon tervezik-e kezelni, hogy kezelhetők a kistelepülések, melyek nem teszik lehetővé 30 mintaelem kiválasztását, a településnagyság hogy esett ki a rétegző változók közül? Telegdi László egyrészt rámutatott, hogy a rétegzés nagy részben meghatározott a publikálandó adatok alábontásának megfelelően, és sok esetben nincs már lehetőség további statisztikai szempontok figyelembevételére. Más oldalról a faktoranalízis esetleges alkalmazására kérdezett rá a potenciális rétegző változóknak lévő információk faktorokba tömörítése érdekében. Herman Sándor a költségoptimalizálásra hívta fel a figyelmet. Marton Ádám emlékeztetett, hogy a meghíúsulások ellensúlyozására a pótcímezt nemzetközi szinten nem tartják elfogadható megoldásnak.

Fraller Gergely válaszában elmondta, hogy számunkra nem ismert a költségkalkuláció módja, így azt az optimalizációs modellbe nem tudtuk beépíteni, de az eredményekhez tartozó költségkalkulációt kezdeményeztük. A településenkénti 30 cím és a pici települések ellentmondása okán a legegyszerűbb megoldás a pici települések figyelmen kívül hagyása, ami nincs jelentős hatással az eredményekre. A településnagyság, mint potenciális rétegző változó szerepelt a vizsgálatban, de nem hozott jelentős eredményt, valószínűleg a mérettel arányos valószínűségű kiválasztás már kellőképpen képviseli ezt a szempontot. A meghíúsulás kezelésére jelenleg is több eljárást követ a KSH, az egyik az arányosan nagyobb mintaválasztás, azonban az új mintavételi tervvel kapcsolatban ezzel a kérdéssel még nem foglalkoztunk, ez egy következő feladat. A kutatás kezdetén az utólagos rétegzés szerepelt a vizsgálandó technikák sorában, azonban képletszerű megoldást nem sikerült találni a nagysággal arányos valószínűséggel történő kiválasztás mellett, ezért nem képezte részét a tényleges kutatásnak. A nemválaszolás egy egyre nagyobb jelentőségű probléma nemcsak nálunk, de nemzetközi szinten is, mértéke már a kezelhetőség határait feszegeti, így nem véletlen, hogy számos kutatás foglalkozott a kérdéssel. A téma olyan szerteágazó, hogy Szép Katalin javasolta, hogy majd egy külön ülésen tárgyaljuk ezt a témát.

Az ülés szakmai részét követte az egyéb ügyek megbeszélése.

Besenyei Lajos beszámolt arról, hogy az Akadémia kezdeményezésére a tudományos bizottságok számát csökkenteni akarják, többek között a Statisztikai Bizottságot a Jövőkutatási Bizottsággal tervezik összevonni. A javaslatot tárgyaló osztályülésen az összevonásuk terve a többi bizottság részéről is ellenkezést váltott ki. Az összevonás névleges is lehet, albizottságként működhetnének tovább a korábbi bizottságok. A Statisztikai Bizottság tagjainak hozzászólásai azt a javaslatot támogatták, hogy a Bizottság elnöke levélben forduljon a MTA elnökéhez, egyidejűleg a IX osztály elnökéhez, tiltakozzon az összevonás ellen, amiben érvként szerepel a tudomány jellege, a KSH-nak nyújtott tudományos háttér, a 150 éves múlt, de az is felmerült, hogy összevonás esetén mely intézmény fogja biztosítani a működéshez szükséges termet, adminisztrációs költségeket? A javaslatot a Bizottság egyhangúlag elfogadta.

A Statisztikai Bizottság 2010. évről szóló beszámolóját megküldtük a IX osztálynak. Ebben kétféle létszámkimutatás készült, az egyik a 2008. évi választások eredményei szerint, egy pedig a jelenlegi szabályozásnak megfelelően az MTA köztestületi nyilvántartása szerint. Tekintettel arra, hogy az ez évi választások a köztestület nyilvántartás szerint történnek majd a hatályos akadémiai ügyrendnek megfelelően, fontos, hogy a SB tagjai ellenőrizzék, és ha kell, módosítsák adataikat. Szép Katalin felhívta a figyelmet, hogy egy köztestületi tag csak egy bizottságot tüntethet fel. A jelenlegi tagjaink közül többen más bizottságokhoz tartoznak a köztestületi nyilvántartás szerint. Ennek a nyilvántartásnak része a korábban is említett Magyar Tudományos Művek Tára, ahova minden köztestületi tagnak be kell vinnie publikációit. Ez azért is fontos, mert ez szolgál számos tudományos értékelés alapjául és a statisztikai folyóiratok megítélése is ez alapján történik.

Besenyei Lajos tájékoztatta a Bizottság tagjait, hogy a Bizottság társszervezője a magyar EU elnökség végén, júniusban tervezett konferenciának, az információk a KSH honlapján lesznek elérhetőek.

Budapest, 2011. március 22.

Szép Katalin