

Statisztikai Szemle

A KÖZPONTI STATISZTIKAI HIVATAL
TUDOMÁNYOS FOLYÓIRATA

SZERKESZTŐBIZOTTSÁG:

DR. BOZSONYI KÁROLY, ÉLTETŐ ÖDÖN, DR. HARCSA ISTVÁN, DR. HUNYADI LÁSZLÓ,
DR. HÜTTL ANTÓNIA (főszerkesztő), DR. JÓZAN PÉTER, DR. LAKATOS MIKLÓS,
DR. MELLÁR TAMÁS, DR. RAPPAI GÁBOR, SÁNDORNÉ DR. KRISZT ÉVA,
DR. SIPOS BÉLA, DR. SPÉDER ZSOLT, SZABÓ PÉTER, DR. VARGHA ANDRÁS,
DR. VITA LÁSZLÓ, DR. VUKOVICH GABRIELLA (a Szerkesztőbizottság elnöke)

91. ÉVFOLYAM 5. SZÁM

2013. MÁJUS

*A Statisztikai Szemlében megjelenő tanulmányok
kutatói véleményeket tükröznek, amelyek nem esnek szükségképp egybe
a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.*

Utánnomás csak a forrás megjelölésével!

ISSN 0039 0690

Megjelenik havonta egyszer
Főszerkesztő: dr. Hüttl Antónia
Osztályvezető: Dobokayné Szabó Orsolya
Kiadja: a Központi Statisztikai Hivatal
A kiadásért felel: dr. Vukovich Gabriella
2013.67 – Xerox Magyarország Kft.

Szakreferensek: dr. Németh Zsolt, dr. Laczka Éva
Szerkesztők: Bartha Éva, dr. Kondora Cosette, Visi Lakatos Mária
Tördelőszerkesztők: Bartha Éva, Simonné Káli Ágnes

Szerkesztőség: Budapest II., Keleti Károly utca 5–7. Postacím: Budapest, 1525. Postafiók 51.

Telefon: 345-6908, 345-6546

Internet: www.ksh.hu/statszemle

E-mail: statszemle@ksh.hu

Kiadó: Központi Statisztikai Hivatal, Budapest II., Keleti Károly utca 5–7.

Postacím: Postafiók 51. Budapest, 1525. Telefon: 345-6000

Előfizetésben terjeszti a Magyar Posta Rt. Hírlap Üzlet (1089 Budapest, Orczy tér 1).

Előfizethető közvetlen a postai kézbesítőknél, az ország bármely postáján,
valamint e-mailen (hirlapelofizetes@posta.hu) és faxon (303-3440).

További információ: 06-80-444-444

Előfizetési díj: fél évre 6 000 Ft, egy évre 10 800 Ft

Beszerezhető a KSH Információs szolgálatán (Budapest II., Fényes Elek u. 14–18. Telefon: 345-6789)

Tartalom

Tanulmányok

Adatserével anonimizált mikroadatok használhatósága – Egy szimulációs vizsgálat tanulságai – <i>Bartus Tamás</i>	465
A bírósági ügyek munkaidőigényének mérése – <i>Dr. Örkényi László</i>	498

Műhely

Észrevételek Huszár Ákos tanulmányához – <i>Harcza István</i>	523
---	-----

Fórum

Beszélgetés S. Molnár Edit szociológus-demográfussal – <i>Dr. Lakatos Miklós</i>	527
Beszámoló az MTA Statisztikai és Jövőkutatási Tudományos Bizottságának Statisztikai Tudományos Albizottságának üléséről – <i>Dr. Hunyadi László – Dr. Kovács Péter</i>	535
Beszámoló a társadalomstatisztikai felhasználói fórumokról – <i>Keszler Ágnes</i>	539
Hírek, események	540

Szakirodalom

Folyóiratszemle

Musil, P. – Kramulová, J. – Cadil, J. – Mazouch, P.: Az egy főre jutó regionális makroaggregátumok becslése a regionális árszintek segítségével a PPS számításánál – (<i>Marton Ádám</i>)	542
Stadler, B. – Wiederhofer-Galik, B.: A fiatal felnőttek különböző migrációs ismérvek szerinti képzettségi és munkaerő-piaci jellemzői Ausztriában – (<i>Lakatos Judit</i>)	544
Rosenski, N.: Statisztikai elemzés Németország nonprofit szektorának gazdasági szerepéről – (<i>Nádudvari Zoltán</i>)	546
Kiadók ajánlata	548
Társfolyóiratok	550

Adatcsérével anonimizált mikroadatok használhatósága – Egy szimulációs vizsgálat tanulságai*

Bartus Tamás

PhD, a Budapesti Corvinus
Egyetem docense

E-mail: [tamas.bartus@uni-
corvinus.hu](mailto:tamas.bartus@uni-corvinus.hu)

A tanulmány áttekinti az adatok felfedés elleni védelmét szolgáló eljárások statisztikai következményeit, és részletesen elemzi az adatcsere kovariancia- és regressziós becslésekre gyakorolt hatását. Amellett érvel, hogy az adatcsere kitüntetett szerepet tölt be a felfedés elleni védelem során. Az adatcsere különböző módszereinek kovariancia- és regressziós becslésekre gyakorolt hatását szimulációval vizsgálja. E vizsgálatok eredménye szerint az esetek többségében az adatcsereből fakadó relatív torzítás mértéke 10 százalék alatt tartható. A torzítást egyrészt a donorok véletlenszerű kiválasztása, másrészt az adatcsere (magyarázó-) változók közötti megosztása minimalizálja. Az eredményeket a mérési hibák elméletére támaszkodva értelmezi.

TÁRGYSZÓ:
Mikroadatok.
Adatcsere.
Anonimizálás.

* A tanulmány az Új Magyarország Fejlesztési Terv Társadalmi Megújulás Operatív Program támogatási rendszeréhez benyújtott „Munkaerő-piaci előrejelzések készítése, szerkezetváltási folyamatok előrejelzése” című TÁMOP-2.3.2-09/1-2009-0001 kiemelt projekt keretében készült. Köszönettel tartozom *Daróczy Gergőnek* lelkiismeretes asszisztensi munkájáért. Szintén köszönet illeti *Cseres-Gergely Zsombort* a kutatást segítő megjegyzéseikért, valamint azért, hogy rendelkezésemre bocsátotta a KSH munkaerő-felvétel 2011. I. negyedéves anonimizált adatait. A tanulmány korábbi változatát „A mikroadatok hozzáféréseivel és az adatok felfedés elleni védelmével kapcsolatos kérdésekről” (Budapest, 2012. november 6.) szervezett műhelykonferencián prezentáltam. Hasznos észrevételeikért köszönettel tartozom a KSH Módszertani főosztálya munkatársainak: elsősorban *Szép Katalinnak* és *Vereczkei Zoltánnak*, valamint *Antal Lászlónak*, *Dobány Máténak* és *Nagy Beátának*.

A statisztikai törvény és annak végrehajtási rendelete az adatszolgáltató beleegyezéséhez kötötte az ún. egyedi, azaz az adatszolgáltatóval „kapcsolatba hozható” adatok továbbadását.¹ A jogszabályok miatt az adatgazdák – például a Központi Statisztikai Hivatal (KSH) – az adatokat csak anonimizálás után adhatják tovább. Az adatszolgáltatók védelmére számos eljárást dolgoztak ki (*Hundepool et al.* [2010]); ezekről magyarul is tájékozódhatnak az érdeklődők (*Bánszegi* [1997], *Erdei-Horváth* [2004], *Szép-Gadácsi* [2010]). A hatásos adatvédelem azonban korlátozza a felhasználók érdekeit (*Boudreau* [2005]), sőt akár ellehetetlenítheti a fontos kérdések empirikus kutatását. Kutatói szempontból nem az adatvédelem hatásossága, hanem az anonimizált adatok használhatósága a fő kérdés; az, hogy milyen mértékben veszélyezteti a hatásos adatvédelem az adatokból levont következtetések érvényességét. Ezzel a kérdéssel a hazai szakirodalom mindeddig alig foglalkozott. Tanulmányunk ezért az anonimizálásból fakadó torzítással foglalkozik.

A célunk ennél konkrétabb: a felfedés elleni védelem egyik eljárásának, az adatcserének a kovariancia- és regressziós becslésekre gyakorolt torzító hatását vizsgáljuk. Az adatcserére több okból esett a választás. Egyrészt az anonimizálási eljárások vagy technikák célja az, hogy ne lehessen az adatszolgáltatóval „kapcsolatba hozni” az adatbázisokban található információkat. Az adatszolgáltatókat ún. kváziazonosítók segítségével lehet azonosítani: ezek olyan könnyen megfigyelt változók (például lakóhely, nem, korcsoport, gyerekszám), melyek együttesen alkalmasak a megfigyelt egyén felfedésére. Az adatcsere pont a kategorikus változók védelmére szolgál. Másrészt – ahogy nemsokára látni fogjuk – a felhasználók szemszögéből az adatcsere számos kedvező tulajdonsággal rendelkezik. A többváltozós regresszió-elemzés során bekövetkező torzításokat azért érdemes vizsgálni, mert az alapkutatások és a hatásvizsgálatok talán legfontosabb adatelemzési módszere. Habár az adatvédelmi technikákat áttekintő publikációkban (*Domingo-Ferrer-Torra* [2001a], [2001b]; *Hundepool et al.* [2010]) számos, az információvesztés mérő általános mérőszámot találhatunk, a regressziós együtthatók torzulásával kapcsolatos konkrét eredmények hiányoznak.² A szakirodalom tárgyalta egyes technikák variancia- és kovarianciabecslésekre gyakorolt hatását, sőt azt is, hogyan lehet anonimizált állományokból torzítatlan variancia- és kovarianciabecsléseket végezni (*Kim* [1990], *Gouweleeuw et al.* [1998]). Kérdéses azonban, hogy az eredmények kiterjeszthetők a többváltozós becslések kontextusára.

¹ Lásd az 1993. évi XLV. I. törvény 17. paragrafusát, valamint a 170/1993. (XII. 3.) Korm. rendelet 16. paragrafusát. A hazai jogszabályok célja egybeesik például az amerikai gyakorlattal; lásd *Sullivan* [1992].

² A regresszióelemzés iránti érdektelenség valószínűleg azzal magyarázható, hogy az adatvédelem a hivatalos statisztika része, a statisztikai hivatalok munkatársainak feladata pedig nem a regresszióelemzés, hanem átlagok és szórások publikálása, illetve becslése.

A tanulmány újdonsága a szimulációs módszer használata. A többváltozós modellek legkisebb négyzeteken alapuló becslése ugyanis a magyarázóváltozók variancia-, kovarianciamátrixa inverzének és a függő és a magyarázó változók kovarianciamátrixának (pontosabban vektorának) szorzata. A mátrixalgebra miatt nehezen látható át, milyen mértékben torzulnak a becslések, ha az anonimizálás miatt módosul egyes magyarázóváltozók varianciája vagy kovarianciája. A probléma hasonló ahhoz, amikor mérési hiba folytán egy adott változó szórása nő, és ezáltal az összes változó együttthátójának regressziós becslése módosul (*Maddala* [2004]). Ez a nehézség indokolja a szimulációs módszer használatát. A szimulációs módszert eddig a mikroaggregálás regressziós becslésekre gyakorolt hatásainak elemzésére használták (*Liu–Little* [2003]; *Lenz et al.* [2006]; *Schmid–Schneeweiss* [2005], [2007], [2008]).

A tanulmány felépítése a következő. Először áttekintjük a mikroadatok anonimizálására széleskörűen használt eljárásokat és ezek átlag-, szórás- és kovarianciabecslésekre gyakorolt hatásait. Mivel az adatvédelmi szabályok a kváziazonosítók anonimizálására ösztönöznek, a kváziazonosítók – mint például a településkód, a foglalkozás kódja – pedig gyakran kategorikus változók, az adatvédelem egyik legfontosabb technikájának az adatcserének kell lennie. A második rész az adatcsere technikáit, illetve az adatcserével kapcsolatos eredményeket elemzi. Bemutatjuk, hogy az adatcsere kovarianciabecslésekre gyakorolt hatása a mérési hibák fogalmi keretén belül értelmezhető. Ez az eredmény azért fontos, mert a mérési hibák többváltozós regressziós becslésekre gyakorolt hatása analitikusan nehezen kezelhető (*Maddala* [2004]), az esetleges torzítások vizsgálata ezért szimulációra váró téma. A harmadik rész a szimulációs vizsgálat módszerét és az eredményeket értelmezi. A szimulációhoz a KSH munkaerő-felvételének 2011. első negyedéves (anonimizált) adatait használjuk. A regressziós becslésekkel kapcsolatos vizsgálatok során egy olyan kutató helyzetét vizsgáljuk, aki a lakóhely-azonosítót is tartalmazó adatbázist szeretné használni, de adatvédelmi okok miatt az adatgazda csak akkor bocsátja ezt rendelkezésre, ha az azonosításhoz szükséges egyéb változókat módosítják. A szimuláció során különböző feltevéseket fogalmazunk meg azzal kapcsolatban, hogy kik azok, akik védelemre szorulnak, és milyen változók módosításával garantálható az anonimitás. A tanulmány végén az eredményeket a mintavételi és a mérési hibák elméletére támaszkodva értelmezzük.

1. Mikroadatok anonimizálásának statisztikai következményei

A felfedés elleni védelem – különösen a hazai jogszabályok fényében – legegyszerűbb módja a kváziazonosítók visszatartása (törlése) vagy átkódolása. Ezek a

technikák nem torzítják, hanem lehetetlenné tesznek bizonyos becsléseket, továbbá jelentősen korlátozzák az adatbázis használhatóságát. Ebben a szakaszban áttekintjük az ennél kevésbé korlátozó, de az adatok módosításával járó technikák statisztikai következményeit. Konkrétan azt vizsgáljuk, befolyásolják-e az egyes eljárások az anonimizált – azaz valamilyen adatvédelmi technikával módosított – változók átlagát, szórását és más változókkal számolt kovarianciáját.³ Ezek a statisztikák alkotják az adatelemzés során leggyakrabban használt eljárások – például a regresszióelemzés, a faktorelemzés – inputjait. Az eljárások logikájával, alapjaival foglalkozunk, és figyelmen kívül hagyjuk az egyes eljárásokon belüli további technikai változatokat, melyek célja az adatvédelem hatásosságának fokozása. A fejezetben található képletek azt feltételezik, hogy a felfedés elleni technikákat a teljes adatbázison, nem pedig annak valamelyik részmintáján használják.

1.1. Adathiány-generálás

Az eljárás során az egyik kváziazonosító változó értékét a magas felfedési kockázatú egyéneknél adathiányra kódoljuk át, úgy, hogy a többi kváziazonosító már ne tegye lehetővé az azonosítást. A módszer kifinomultabb változata annak a „brutális” megoldásnak, amikor az egész esetet törlik az adatbázisból. A módszer nyilvánvaló hátránya az elemzéshez használható mintanagyság csökkentése és az anonimizált változó átlagának torzulása. Ha az n elemű mintában $k = pn$ megfigyelésnél töröljük az x változó értékét, akkor az anonimizált x^a változó átlaga

$$\bar{x}^a = \frac{\bar{x} - p\bar{x}_k}{1 - p}$$

lesz, ahol \bar{x}_k x átlaga a törléssel védett részmintában, p pedig az anonimizált megfigyelések relatív gyakorisága. A képlet súlyozott átlagbecslésekre is érvényes. Ha a súlyok normalizáltak, azaz a súlyok összege azonos a mintanagysággal, akkor az egyenletben a p paramétert a törölt megfigyelésekhez tartozó súlyok összegének és a mintanagyság hányadosaként kell értelmezni – azaz:

$$p = n^{-1} \left(\sum_{(k)} w \right),$$

³ A tanulmányban rendszeresen használjuk a változó anonimizálása, valamint az anonimizált változó terminusokat. Az előbbi a „változó adatvédelmi okok miatt végzett módosítása”, az utóbbi az „adatvédelmi megfontolások miatt módosított változó” kifejezést rövidíti.

ahol w_i az i esethez tartozó súly, $\sum w$ pedig a súlyok összege. A képlet világosan mutatja, hogy az átlagbecslés akkor torzul, ha kis mintából kiugró értéket törölünk.

A 0–1 kódolású indikátorváltozóknál még egyszerűbb a képlet. Ha törlésre csak az $x=1$ értékekénél kerül sor, $\bar{x}_k=1$ és az anonimizált indikátorváltozó súlyozatlan átlaga

$$\bar{x}^a = \frac{\bar{x} - p}{1 - p}.$$

Az anonimizálás okozta torzítás

$$\bar{x}^a - \bar{x} = \frac{p}{1 - p}(\bar{x} - 1),$$

tehát annál nagyobb, minél nagyobb p és minél nagyobb a változó átlaga. Az anonimizált indikátorváltozó szórásnégyzete:

$$Var(x^a) = \frac{\bar{x} - p}{(1 - p)^2 \bar{x}} Var(x).$$

Ha p értéke nulla, az anonimizált változó és annak varianciája azonos az eredetivel. Mivel a variancia sosem lehet negatív, $p \leq \bar{x}$. Mivel x indikátorváltozó, az egyenlőtlenség azt a triviális feltételt fogalmazza meg, hogy az anonimizált esetek aránya nem haladhatja meg az $x=1$ esetek arányát. A p növekedésével tehát az anonimizált változó varianciája csökken; az anonimizálás „elkoptatja” az eredeti változó varianciáját.

Szintén torzulhat az anonimizált indikátorváltozó egy tetszőleges másik változóval vett kovarianciája. Ha az adatvédelem az adatbázis 100 p százalékára terjed ki, és ismét csak $x=1$ esetekre, akkor az anonimizált indikátorváltozó és a tetszőleges anonimizálatlan y változó kovarianciája:

$$Cov(x^a, y) = \frac{Cov(x, y)}{(1 - p)^2} - \frac{p(1 - \bar{x})(\bar{y}_1 - \bar{y}_0)}{(1 - p)^2},$$

ahol \bar{y}_1 és \bar{y}_0 y átlaga az anonimizálatlan adatbázisban az $x=1$, illetve $x=0$ csoportokban. A kovarianciabecslés torzulása nyilvánvalóan p és a szóban forgó csoportátlagok közti különbség függvénye.

Az adathiány-generálásnál tehát egyszerű képletet kaptunk arra, milyen mértékben torzulnak az átlag- és varianciabecslések. A kovarianciabecslések torzulására kapott képlet viszont bonyolultabb.

1.2. Adatcsere

Az adatcsere (data swapping) során a felfedési kockázatot jelentősen növelő változó (vagy változók) értékeit cseréljük fel egyes megfigyelések között (Dalenius–Reiss [1982]). Képzelnék el, hogy egy adatbázisban magas a falusi egészségügyi dolgozók és a városi mezőgazdasági dolgozók felfedési kockázata. A felfedési kockázat csökkenthető, ha k számú falusi egészségügyi dolgozó foglalkozását mezőgazdasági dolgozóra, és ezzel párhuzamosan szintén k számú városi mezőgazdasági dolgozó foglalkozását egészségügyi dolgozóra módosítjuk – azaz a foglalkozási adatokat ki-cseréljük.

Legyen $\delta_{x_{ij}} = 1$, ha az x változó értékét az i és a j -edik megfigyelések között ki-cseréljük; különben $\delta_{x_{ij}} = 0$. Az adatcsere formális definíciója a következő (Boudreau [2005]):

$$x_i^a = (1 - \delta_{x_{ij}})x_i + \delta_{x_{ij}}x_j,$$

$$x_j^a = (1 - \delta_{x_{ij}})x_j + \delta_{x_{ij}}x_i.$$

A formális definíció – meglepő módon – semmilyen információt nem tartalmaz az i és j egyének felfedési kockázatáról. Az adatcsere céljának figyelembe vétele mellett triviális, hogy a két egyén közül az egyik – de csak az egyik – könnyen felfedhető.

Az adatcsere nem módosítja az átlagot és a szórást, de nem őrzi meg feltétlenül az együttes eloszlásokat. Példánkban a foglalkozás cseréje után mind a településtípus, mind pedig a foglalkozás peremeloszlása változatlan marad – ugyanakkor megváltozik a foglalkozás és településtípus együttes eloszlása, hiszen a csere révén csökken a falusi egészségügyiek és a városi mezőgazdaságiak száma (és értelemszerűen nő a falusi mezőgazdasági és a városi egészségügyi dolgozók száma). A probléma megoldására Dalenius és Reiss [1982] azt javasolta, hogy az adatcserét további megfigyelések bevonásával kell folytatni, mindaddig, amíg helyreáll a többdimenziós eloszlás. A sikerre azonban nincs garancia; ráadásul az újabb és újabb cserék megtalálása ropant időigényes. A gyakorlatban is könnyen megvalósítható adatcsere ezért csak a peremeloszlásokat őrzi meg tökéletesen – az együttes eloszlásokat viszont csak közelítőleg (Reiss [1984]). Ilyen könnyen kivitelezhető technika például az, amikor az

adatcserebe bevont esetek a kicserélt változóktól eltérő más változók szempontjából hasonlítanak egymásra (*Shlomo–Tudor–Groom* [2010]).

Az együttes eloszlás változásának két következménye van. Egyrészt torzulnak a súlyozott becslések, hiszen az adatcsere nem terjed ki a súlyváltozókra. Ha az x_j és x_k értékeket cseréljük ki, akkor az anonimizált és az eredeti változók súlyozott átlagainak különbsége

$$\frac{w_j(x_k - x_j) + w_k(x_j - x_k)}{\sum w_i} \quad /1/$$

lesz, ahol w_i az i -edik megfigyeléshez rendelt súly. A súlyozott becslésekre vonatkozó képletek rendkívül bonyolultak (*Boudreau* [2005]).

A másik következmény: torzulnak a (súlyozatlan) kovarianciák. Tegyük fel, hogy az adatcsere pn megfigyeléspárt érint. Az x változón végrehajtott adatcsere következményeit tekintve azonos az y változón végzett adatcserevel. Jelölje y_{01} és y_{10} azokat az y értékeket, melyeknél az x indikátorváltozót nulláról egyesre, illetve egyesről nullára cserélték. Az adatcsere egyetlen hatása: x^a és y szorzatösszegét $\sum y_{01}$ összeggel növeljük és a $\sum y_{10}$ összeggel csökkentjük. Ez alapján az adatcsereből fakadó torzítás

$$\text{Cov}(x^a, y) - \text{Cov}(x, y) = -p[\bar{y}_{10} - \bar{y}_{01}]. \quad /2/$$

Ha y várható értéke az $x = 1$ csoportban magasabb, és az anonimizált megfigyelések a minta véletlenszerűen kiválasztott mintája, akkor $\bar{y}_{10} > \bar{y}_{01}$ és így a /2/ egyenlet jobb oldalán szereplő különbség negatív. Ennek ellentéte igaz, ha y várható értéke az $x = 0$ csoportban magasabb. Az adatcsere tehát a csökkenti, „koptatja” a kovariancia abszolút értékét. A kopás mértéke annál nagyobb, minél több megfigyelésre terjed ki az adatcsere.

Mivel tetszőleges x indikátorváltozó és tetszőleges y változó kovarianciája x varianciájának és a $\bar{y}_1 - \bar{y}_0$ különbség szorzata, az anonimizált x^a indikátorváltozó és y kovarianciája a következő formára hozható:

$$\text{Cov}(x^a, y) = \text{Cov}(x, y) \left[1 - \frac{p}{\text{Var}(x)} \frac{\bar{y}_{10} - \bar{y}_{01}}{\bar{y}_1 - \bar{y}_0} \right].$$

A jobb oldalon a szögletes zárójelben szereplő mennyiséget érdemes külön jelöléssel ellátni:

$$Q_x(y) = 1 - \frac{p}{\text{Var}(x)} \frac{\bar{y}_{10} - \bar{y}_{01}}{\bar{y}_1 - \bar{y}_0}. \quad /3/$$

Ha az adatcsere véletlenszerű és teljesül az $\bar{y}_{10} - \bar{y}_{01} = \bar{y}_1 - \bar{y}_0$ egyenlőség, $Q_x(y)$ még egyszerűbben írható fel:

$$Q_x(y) = 1 - \frac{p}{\text{Var}(x)}. \quad /4/$$

A képlet üzenete világos: az adatcserével védett megfigyelések növekedésével egyre nagyobb mértékben torzul a kovariancia. A torulás mértéke azonban az adatcserével érintett változó varianciájától is függ. Ha az adatgazda nyilvánosságra hozza a p együttható értékét, a felhasználó a

$$\widehat{\text{Cov}(x, y)} = \frac{\text{Cov}(x^a, y)}{Q_x(y)} \quad /5/$$

képlettel becsülheti az anonimizálatlan állományban érvényes kovarianciát.

1.3. Utólagos randomizálás

Az *utólagos randomizálás* (post-randomization – PRAM) az adatcsere kifinomultabb változata: ez eljárás során adott változó értékeit egy előre meghatározott eloszlás szerint véletlenszerűen módosítják (Kooiman–Willenborg–Gouweleeuw [1997], Gouweleeuw et al. 1998). A módszert a randomizált válaszok technikája (Sarndal–Swensson–Wretman [1982]) inspirálta. Diszkrét, 0 és 1 értékeket felvevő változó esetén a módszer azt írja elő, hogy a nullákat adott $(1 - \theta_0)$ valószínűséggel 1-re, az egyeseket adott $(1 - \theta_1)$ valószínűséggel nullákra cseréljük. Többértékű változókra általánosítva: az adatcserét irányító valószínűség-eloszlást egy $k \times k$ dimenziójú \mathbf{P} (perturbációs) mátrix definiálja, melynek ij -edik eleme annak valószínűségét adja meg, hogy a változó i -edik értéke kicserélődik a j -edik értékre.

Az adatok cseréje tehát nem a védelemre szoruló egyént, hanem a véletlenszerűen kiválasztott egyénet érinti. Ha ezt a tényt az adatgazda nyilvánosságra hozza, a rosszindulatú felhasználó nem lehet biztos abban, hogy egy adott falu állatorvosa tényleg falusi állatorvos – hiszen lehetséges, hogy az utólagos randomizálás pont egy falusi mezőgazdasági segédmunkás foglalkozását cserélte fel egy városi állatorvos foglalkozására.

Az utólagos randomizálás után módosul a manipulált változó átlaga és szórása. Az utólag randomizált kétértékű változó átlaga

$$\bar{x}^a = (\theta_0 + \theta_1 - 1)\bar{x} + (1 - \theta_0), \quad /6/$$

szórásnégyzete pedig

$$Var(x^a) = (\theta_0 + \theta_1 - 1)^2 Var(x) \quad /7/$$

lesz (Gouweleeuw et al. [1998]). Az utólagos randomizálás tehát torzíthatja az átlagot és a varianciát. Az átlagok és a szórások azonban anonimizált állományokból is becsülhetők maradnak – feltéve, hogy az adatgazda nyilvánosságra hozza a randomizálás során használt \mathbf{P} perturbációs mátrixot. Ha a θ paraméterek ismertek a felhasználók számára, a /6-/7/ egyenletek alapján a korrigált átlagbecslés

$$\hat{\bar{x}} = \frac{\bar{x}^a - (1 - \theta_0)}{(\theta_0 + \theta_1 - 1)}, \quad /8/$$

a korrigált varianciabecslés pedig

$$\widehat{Var(x)} = \frac{Var(x^a)}{(\theta_0 + \theta_1 - 1)^2} \quad /9/$$

(Gouweleeuw et al. [1998]). A becslés természetesen csak akkor lehetséges, ha θ_0 és θ_1 összege nem azonos eggyel. Ha például úgy döntünk, hogy a diszkrét változó zérusait 5 százalékos eséllyel cseréljük egyre, akkor ezzel párhuzamosan nem dönthetünk úgy, hogy az egyes értékeket 95 százalékos eséllyel cseréljük nullákra.

A felhasználók érdekei akkor sérülnek legkevésbé, ha az utólagos randomizálás az esetszámra azonos (Bycroft–Merrett [2005] 126. old.). A \mathbf{P} perturbációs mátrix elemeit ekkor úgy választjuk ki, hogy az anonimizált és az eredeti változó átlagai azonosak legyenek. Például kétértékű változók esetén akkor esetszámra azonos az utólagos randomizálás, ha teljesül az alábbi egyenlőség:

$$\theta_1 = 1 + \frac{1 - \bar{x}}{\bar{x}}(\theta_0 - 1).$$

Indikátorváltozóknál az esetszámra azonos utólagos randomizálás nem torzíthatja az átlagok és a szórások becslését, viszont a /2/ egyenlet miatt torzíthatja a kovarianciabecsléseket, illetve a súlyozott átlagbecsléseket.

1.4. Mikroaggregálás

A mikroaggregálásnak számos technikai változata létezik (*Mateo-Sanz–Domingo-Ferrer* [1998], *Schmid–Schneeweiss* [2005]). Az eljárás logikája mégis egyszerű. Első lépésben az adatokat a védelemre kiszemelt változó vagy egy másik változó szerint sorba rendezzük. Ezután a megfigyeléseket előre rögzített k vagy az eljárás során – valamilyen statisztikai eljárással megállapított – változó nagyságú csoportokba soroljuk. Az egyéni megfigyeléseket végül a szóban forgó csoportátlagokkal helyettesítjük, amelyeket a rendezés miatt egymáshoz hasonló adatokból számolunk ki, az anonimizált és a valós értékek eltérése kicsi is lehet. Ez felveti azt a kérdést, vajon hatásosan védi-e a mikroaggregálás a személyes adatokat. Azonban a sorbarendezés nélküli csoportképzés sem garantálja automatikusan a hatásos védelmet: előfordulhat, hogy egy csoportspecifikus értékösszeget egyetlen megfigyelés dominál.⁴

Az eljárás nyilvánvalóan változatlanul hagyja a változó átlagát és csökkenti a varianciát: a szórásnégyzet-felbontás közismert képlete alapján az anonimizált változó varianciája a belső szórásnégyzettel, azaz a kategóriákon belüli szórásnégyzetek összegével lesz kisebb az eredeti változó varianciájánál. Az aggregálást megelőző sorbarendezés célja az, hogy a szórás csökkenése minimális legyen. A szórásbecslések mellett az eljárás a kovarianciabecsléseket is torzíthatja. Ha az eljárás során j darab k elemű aggregátumot alakítanak ki, akkor a torzítás – azaz az anonimizált és az anonimizálatlan állományokon számolt kovarianciák különbsége –

$$\text{Cov}(x^a, y) - \text{Cov}(x, y) = - \sum_{(j)} \sum_{(i=1)}^K y_{ki} (x_{ki} - \bar{x}_k).$$

Ha az egyes mikroaggregátumokon belül azonos lenne a megfigyelések csoportátlagoktól való eltérése, a torzulás mértéke a mikroaggregátumok méretének növekvő függvénye. A kovarianciák és a varianciák módosulása miatt a regressziós becslések is torzulnak – a torzítás konkrét mértékét számos szimulációs vizsgálatban elemezték (*Liu–Little* [2003]; *Lenz et al.* [2006]; *Schmid–Schneeweiss* [2005], [2007], [2008]).

1.5. Zajosítás

Az eljárás lényege: az egyedi vagy ritka adatokhoz egy véletlen zajt – azaz 0 átlagú, előre meghatározott szórással rendelkező ε véletlen számot – adunk. A zajosítás

⁴ A probléma ugyanaz, mint az aggregált adatoknál ismert dominanciaprobléma: adatvédelmi szempontból aggályos olyan értékösszegek publikálása, melyeknél az értékösszeget két adatszolgáltató dominálja, és ezért ők az értékösszeg ismeretében többé-kevésbé pontos becslést adhatnak a másik domináns adatszolgáltató értékére.

nem torzítja az átlagot, viszont torzítja a variancia- és kovarianciabecsléseket (*Brand* [2002]). A zaj véletlenszerűsége miatt az anonimizált változó varianciája a zaj varianciájával haladja meg az anonimizálatlan változó varianciáját:

$$\text{Var}(x^a) = \text{Var}(x) + \text{Var}(\varepsilon).$$

Ha a zaj független az adatbázisban szereplő változóktól, akkor a zajosított változó más változókkal vett kovarianciája várhatóan változatlan marad. Ha viszont a zajosítás mindkét változóra kiterjed, a zajosított változók kovarianciája az eredeti kovariancia és a zaj varianciájának az összege – feltéve, hogy a zajok szórása megegyezik.

A zajosítási eljárás dokumentálása és a zaj varianciájának publikálása lehetővé teszi a torzítatlan becsléseket (*Kim* [1990], *Brand* [2002]). A zaj varianciájának ismeretében a felhasználó a

$$\widehat{\text{Var}(x)} = \text{Var}(x^a) - \text{Var}(\varepsilon) \quad /10/$$

képlettel becsülheti az eredeti változó varianciáját és a

$$\widehat{\text{Cov}(yx)} = \text{Cov}(y, x^a) - \text{Var}(\varepsilon) \quad /11/$$

képlettel a zajosított változók kovarianciáját (*Kim* [1990]). A /10/–/11/ képletekkel természetesen a korrelációs együttható is becsülhető. Mivel a számítógépekkel szó szoros értelemben vett véletlenszámokat nem lehet létrehozni, egy konkrét mintában a zaj kismértékben korrelálhat a zajosítatlan változóval, így a /10/–/11/ egyenleten alapuló becslések torzíthatnak. E technikai tökéletlenségből fakadó esetleges torzítások minimalizálhatók a szisztematikus zajosítással (*Evans–Zayatz–Slata* [1996]).⁵ Az információvesztés azzal is csökkenthető, ha eljárást csak a felfedhető egyének részmintáján használják (*Fagan–Greenberg* [1988]).

A zajosítás az utólagos randomizáláshoz hasonlóan tehát rendelkezik azzal a kedvező tulajdonsággal, hogy az eljárás paramétereinek – konkrétan a zaj varianciájának – ismeretében a felhasználó torzítatlan becslést tegyen, még akkor is, ha az eredeti helyett csak a zajosított adatbázist használhatja.

⁵ Az eljárás lényege, hogy az adatbázist először a zajosításra váró változó szerint sorba rendezzük. Ezután felváltva adunk pozitív és negatív zajt a megfigyelésekhez. A pozitív és negatív értékeket két külön eloszlásból vesszük, melyek várható értékei szimmetrikusak, szórásaik pedig azonosak. Például: a pozitív értékeket generáló eloszlás átlaga 1, szórása 0,2; a negatív értékeket generáló eloszlás átlaga –1, szórása szintén 0,2. Ha a szórás az átlag abszolút értékéhez képest kicsi, akkor a normális eloszlás tulajdonságai miatt csak nagyon ritkán fordulhat elő, hogy a pozitív (illetve negatív) eloszlás a szándékotól eltérően negatív (illetve pozitív) zajt generál.

1.6. Kerekítés

Az eljárás során az adott változót előre meghatározott szabályok szerint kerekítik, hogy a pontos értékek visszatartásával az alanyok azonosítása nehezebbé, illetve lehetetlenné válik (Fischetti [1998]). A kerekítés során nem feltétlenül egész számra, de százaskra, ezresekre vagy akár tízezresekre történő kerekítés is előfordulhat, amennyiben az adatok érzékenysége azt kívánja, illetve a kerekítő algoritmus meghatározott eloszlásnak megfelelően véletlen számokkal is dolgozhat (Shlomo [2005]).

A kerekítés felfogható a zajosítás inverzének: a *védelemre szoruló* változó olyan, mintha az *anonimizált* változóhoz hozzáadnánk egy véletlen számot, a zajt. Persze ez a „zaj” nem normális, hanem egyenletes eloszlást követ a $[-h, +h]$ intervallumon, ahol h az *anonimizált* változó nagyságrendjének a fele. Az ezresekre kerekítés például annak az eljárásnak az inverze, hogy az ezresekre kerekített számokhoz a $[-500, 500]$ intervallumból véletlenszerűen kiválasztott számot adunk. Ha ez az analógia helyes, a zajosításból fakadó torzításokat definiáló képletek a kerekítésből adódó torzításokat is leírják – feltéve, hogy az anonimizált és az anonimizálatlan változókat felcseréljük az egyes képletekben. Az anonimizált változó és a zaj varianciáinak ismeretében az anonimizálatlan változó varianciájának becslőfüggvénye

$$\widehat{Var}(x) = Var(x^a) + \frac{h^2}{3},$$

a korrelációs együttható becslőfüggvénye pedig

$$\widehat{\rho^2} = \frac{Var(x^a)}{Var(x^a) + h^2/3} r^2,$$

ahol r a korrelációs együttható az anonimizált adatbázisban. A képletekben szereplő $h^2/3$ hányados a $[-h, +h]$ intervallumon értelmezett egyenletes eloszlású változó varianciája.

1.7. Újra-mintavételezés

Az újra-mintavételezés (resampling) során a módosítandó változó eredeti értékei szerint sorba rendezzük az adatbázist, majd a változóból almintákat hozunk létre a *bootstrap* vagy a *jackknife* eljárással. Az almintákat szintén sorba rendezzük, majd hozzáfűzzük az eredeti adatbázishoz. Az anonimizált és nyilvánosságra hozható változó az almintákból számolt átlag lesz.

A *bootstrap* eljárás során n elemű mintából újabb n elemű, előre meghatározott (a mai számítógépes kapacitásokhoz mérten általában magas, minimum 10 000) számú almintát generálunk visszatevéses, véletlen mintavétel segítségével. Az anonimizált változó értéke az i -edik megfigyelésnél

$$x_i^a = \frac{\sum_{s=1}^S x_{is}}{S}$$

lesz, ahol S az alminták száma, s az alminta sorszáma, x_{is} pedig az i -edik megfigyelés az s almintában. Mindegyik almintában igaz az, hogy x értékei olyan sorrendben követik egymást, mint az eredeti adatbázisban. (Ha tehát x eredeti értékeit növekvő sorrendbe állítjuk, akkor ugyanezt kell tenni mindegyik almintában is.)

A *bootstrap* eljárás nem zárja ki annak lehetőségét, hogy egy adott almintába ugyanaz az érték többször is bekerül, míg más értékek egyáltalán nem kerülnek be. Sőt, elvileg az is előfordulhat – igaz, elenyészően kis valószínűséggel⁶ –, hogy egy adott alminta kizárólag egyetlen esetet tartalmazza n duplikátumban. E probléma kezelésére alkalmas a *jackknife* eljárás. Ennek során az n elemű mintáinkból n számú, $n-1$ elemszámú almintát generálunk, minden egyes alminta esetében egy tag elhagyásával. Az elhagyott elem lehet minden esetben más és más vagy véletlenszerűen kiválasztott. A hagyományos *jackknife* az anonimizálás során nem használható további megkötések nélkül, hiszen minden egyes almintában lesz 1 pótlólagos adathiány. Ha a statisztikai szoftver az adathiányt végtelenként értelmezi, akkor az adathiány mindig a növekvő sorrendbe rendezett alminták utolsó megfigyeléséhez tartozik, tehát az utolsó megfigyelésnél adathiányt generálnánk. E probléma elvi megoldása az lehet, ha valamilyen technikával úgy rendezzük növekvő sorrendbe az almintákat, hogy az adathiány egy véletlenszerűen kiválasztott sorba kerüljön.

1.8. Összegzés

Ebben a részben áttekintjük, milyen mértékben torzítják az anonimizálási eljárások a – súlyozott, illetve súlyozatlan – átlag-, szórás- és kovarianciabecsléseket. Az átlagbecsléseket az eljárások döntő többsége torzítatlanul hagyja. A varianciabecslések torzítatlanságát már csak az adatcsere és az esetszámra azonos utólagos randomizálás garantálja. A mikroaggregálás okozta torzítás elvileg kismértékű, a zajosítással és az esetszámra nem azonos utólagos randomizálás védett állományokból pedig torzítatlanul becsülhető a variancia, ha az adatgazda publikálja az anonimizálási eljárás releváns paramétereit. A kovarianciabecsléseket szinte mind-

⁶ A szóban forgó valószínűség n^{-n} .

egyik módszer torzítja, ám itt is érvényes az, hogy az adatvédelem során használt releváns paraméter (vagy paraméterek) publikálása lehetővé teszi a felhasználók számára a becslések korrigálását.

Az áttekintett anonimizálási módszerek közül kiemelt szerepet játszik az adatcsere. Egyrészt az adatcsere során használt paraméterek publikálása lehetővé teszi a felhasználók számára a torzítatlan becsléseket. Másrészt az adatgazdák tipikus célja a legtöbbször kategorikus kváziazonosítók (például településkódok) anonimizálása, a kategorikus változók kifinomult védelmére pedig csak az adatcsere – valamint annak továbbfejlesztett változata, az utólagos randomizálás – alkalmas. Végül: elvileg semmi akadálya, hogy adatcserevel folytonos változókat is anonimizáljunk – míg a folytonos változók védelmére kidolgozott technikák kategorikus változókra történő alkalmazása nem magától értődő.⁷ Érdeemes ezért az 1.2. alfejezetben bemutatott adatcseret és annak statisztikai következményeit alaposabban szemügyre venni.

2. Az adatcsere statisztikai következményei: további eredmények

Ebben a fejezetben egyrészt a mérési hibák elméletének kontextusába helyezzük az adatcsere statisztikai következményeire vonatkozó eredményeinket. Másrészt azt vizsgáljuk, hogy a szóban forgó eredmények robusztusak maradnak-e, ha az adatcsere nem teljesen véletlenszerű.

2.1. Az adatcsere okozta torzítás mint mérési hiba

Az előző alfejezetben láttuk, hogy az adatcsere torzítja a kovarianciákat: a torzítás mértéke pedig a

$$Q_x(y) = 1 - \frac{p}{Var(x)} \frac{\bar{y}_{10} - \bar{y}_{01}}{\bar{y}_1 - \bar{y}_0}$$

mennyiség függvénye. Mivel az egyváltozós regressziós becslés egy kovariancia és egy variancia hányadosa, az anonimizált állományból számolt egyváltozós becslés az anonimizálatlan állományból számolt becslés és $Q_x(y)$ szorzata. Ez az eredmény nagyon hasonlít arra, amely a mérési hibák regressziós becslésekre gyakorolt hatásá-

⁷ Nem világos például, hogy a kedvező tulajdonságokkal rendelkező zajosítást hogyan lehetne kategorikus változókra alkalmazni, hiszen ekkor a zaj normális eloszlására vonatkozó feltevést módosítani kell. A mikroaggregálás és a kerekítés kategorikus analógiája az átkódolás, melynek statisztikai következményeit nehéz elemezni.

ra vonatkozik. Képzeld el, hogy x^a nem adatcserevel anonimizált, hanem u mérési hibával mért változó! A mérési hibák becslésekre gyakorolt hatása ismert (Fuller [1987], Maddala [2004]): a mérési hiba elköptatja a regressziós együtttható abszolút nagyságát, mivel

$$\hat{\beta} = \frac{\text{Cov}(y, x^a)}{\text{Var}(x^a)} = \frac{\text{Cov}(x, y)}{\text{Var}(x) + \text{Var}(u)} = \beta \frac{\text{Var}(x)}{\text{Var}(x) + \text{Var}(u)} = \beta R_x. \quad /12/$$

R_x az anonimizált változó megbízhatósági együttthatója.

A /12/ és az /5/ egyenlet hasonlósága alapján a Q mennyiséget érdemes *relatív megbízhatósági együttthatónak* nevezni. A relatív jelző arra utal, hogy Q értéke függ attól a változótól, amivel kovarianciát számolunk. A megbízhatósági együtttható terminus alkalmazása indokolt, mert az adatcsere következményeit tekintve mérési hiba: ahhoz hasonlóan koptatja a regressziós becsléseket. Az adatcserevel anonimizált változó relatív megbízhatósága annál nagyobb, minél nagyobb az anonimizálásra váró változó szórása és minél kisebb az adatcserevel érintett megfigyelések aránya. Sőt, a regresszioelemzés kontextusában az adatcsere olyan eljárásnak tekinthető, mintha a magyarázóváltozót u mérési hibával mérnénk, a fiktív mérési hiba varianciája pedig az $R = Q$ azonosság alapján:

$$\text{Var}(u) = \frac{p \text{Var}(x)}{\text{Var}(x) - p}.$$

Ahhoz, hogy Q is normalizált legyen és a fiktív mérési hiba varianciája ne lehessen negatív, az indikátorváltozó varianciájának kisebbnek kell lennie a p paraméternél, azaz teljesülnie kell a

$$p \leq \bar{x} - (\bar{x})^2$$

egyenlőtlenségnek. Ha például az indikátorváltozó eloszlása szimmetrikus, azaz az egyesek és nullák száma azonos, akkor az adatcsere elvileg a teljes mintára is kiterjedhet, a megfigyeléspárok relatív gyakorisága tehát $1/2$, mégis az egyenlőtlenség ennek felét szabja meg korlátként.

2.2. Adatcsere-technikák és a torzítás várható mértéke

Az adatcsere kovarianciabecsléseket torzító hatása a

$$Q_x(y) = 1 - \frac{p}{\text{Var}(x)} \frac{\bar{y}_{10} - \bar{y}_{01}}{\bar{y}_1 - \bar{y}_0}$$

menntiség függvénye. Az egyszerű kifejtés kedvéért mindeddig azt feltételeztük, hogy a cserepartnereket véletlenszerűen választjuk ki és teljesül az

$$\bar{y}_{10} - \bar{y}_{01} = \bar{y}_1 - \bar{y}_0 \quad /13/$$

egyenlőség.

A gyakorlatban azonban ez a feltevés nem teljesül szükségszerűen. Egyrészt a véletlenszerű kiválasztás nyilvánvalóan nem garantálja, hogy a /13/ egyenlet minden mintában teljesül. A pontosság elérésének egyik eszköze a rétegzés lehet. Ha a rétegző ismérvek korrelálnak az y változóval, és adatcserére az adott rétegeken belül kerül sor, akkor nagy mintákban és tömeges mértékű adatcserénél /13/ egyenletnek teljesülnie kell. A rétegzésnek viszont az a mellékkövetkezménye, hogy egy adott rétegen belül az adatcsere kevés megfigyelésre terjedhet ki, az esetszám csökkenése viszont veszélyezteti a /13/ egyenlőség fennállását. A rétegzés módszerét *Shlomo–Tudor–Groom* [2010] használták, és célzott adatcserének (targeted data-swapping) nevezték.

A „nem véletlenszerűség” szándékosan is előidézhető. Az egyik az irányítottág – abban az értelemben, hogy az $x=1$ megfigyeléseket az $x=0$ megfigyelések egyik részalmazából választjuk ki. Képzeld el, hogy szükség van a legmagasabb iskolai végzettség anonimizálására, mert egyes diplomások más ismérvekkel együtt beazonosíthatók. Tegyük fel, hogy a diplomásokat a hozzájuk leginkább hasonló érettségizettekkel akarjuk felcserélni. Ebben az esetben az x indikátorváltozó a diplomásokat azonosítja, az $x=0$ feltétel a nem diplomásokat jelöli. Az adatcsere azonban nem terjedhet ki az $x=0$ halmaz minden elemére – csak azokra, akik érettségizettek. Ha a kutatás során vizsgált y változó korrelál az iskolai végzettséggel, akkor y átlaga magasabb az érettségizettek körében, mint az összes nem diplomás körében. Emiatt $\bar{y}_{01} > \bar{y}_0$ és ezért $\bar{y}_{10} - \bar{y}_{01} < \bar{y}_1 - \bar{y}_0$, feltéve, hogy $\bar{y}_{10} > \bar{y}_1$. Ebben a példában a diplomásokat a hozzájuk y szempontból leginkább hasonló nem diplomásokra cseréltük. A hasonló megfigyeléseket célzó irányított adatcsere növeli a Q megbízhatósági együtthatót.

3. Szimulációs vizsgálatok

Az adatcsere kovarianciabecslésekre gyakorolt hatása – a /13/ egyenletben megfogalmazott feltevés mellett – ismert. Nem világos azonban, hogy az eredmények kiterjeszthetők a többváltozós becslések kontextusára. A többváltozós modellek legkisebb négyzeteken alapuló becslése ugyanis a magyarázóváltozók variancia-kovarianciamátrixa inverzének és a függő és a magyarázóváltozók kovarianciamátrixának (pontosabban vektorának) szorzata. A mátrixalgebra miatt nehezen látha-

tó át, milyen mértékben torzulnak a becslések, ha például az egyik magyarázóváltozó varianciája az adatvédelem miatt megnő. A probléma hasonló ahhoz, amikor mérési hiba folytán egy adott változó szórása nő, és ezáltal az összes változó együtthatójának regressziós becslése módosul (Maddala [2004]). Az analitikus eredmények hiánya vagy értelmezési nehézségei indokolják a szimulációs módszerek használatát. Az anonimizálási eljárások statisztikai következményeinek szimulációs vizsgálata bevett gyakorlat. A regressziós becslésekre gyakorolt hatások vizsgálata azonban eddig főleg a mikroaggregálásra korlátozódott (Liu–Little [2003]; Lenz *et al.* [2006]; Schmid–Schneeweiss [2005], [2007], [2008]).

Tekintsük az egyszerű

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

többsváltozós modellt. Ha az adatbázis nem szorul védelemre, az együtthatókat az alábbi képletekkel számoljuk ki:

$$\beta_1 = \frac{\frac{\text{Cov}(yx_1)}{\text{Var}(x_1)} - \frac{\text{Cov}(yx_2)\text{Cov}(x_1x_2)}{\text{Var}(x_1)\text{Var}(x_2)}}{1 - \frac{\text{Cov}^2(x_1x_2)}{\text{Var}(x_1)\text{Var}(x_2)}},$$

$$\beta_2 = \frac{\frac{\text{Cov}(yx_2)}{\text{Var}(x_2)} - \frac{\text{Cov}(yx_1)\text{Cov}(x_1x_2)}{\text{Var}(x_1)\text{Var}(x_2)}}{1 - \frac{\text{Cov}^2(x_1x_2)}{\text{Var}(x_1)\text{Var}(x_2)}}.$$

Ha viszont az első magyarázóváltozót adatcserével védik, az anonimizált állományban értelmezett

$$y = \beta_0 + \beta_1 x_1^a + \beta_2 x_2 + \varepsilon$$

modell becsléséhez az alábbi becslőfüggvényeket kell használni:

$$\hat{\beta}_1 = \frac{Q_1(y) \frac{\text{Cov}(yx_2)}{\text{Var}(x_2)} - Q_1(x_2) \frac{\text{Cov}(yx_1)\text{Cov}(x_1x_2)}{\text{Var}(x_1)\text{Var}(x_2)}}{1 - Q_1(x_2) \frac{\text{Cov}^2(x_1x_2)}{\text{Var}(x_1)\text{Var}(x_2)}}, \quad /14/$$

$$\hat{\beta}_2 = \frac{\frac{Cov(yx_2)}{Var(x_2)} - Q_1(y)Q_1(x_2) \frac{Cov(yx_1)Cov(x_1x_2)}{Var(x_1)Var(x_2)}}{1 - Q_1(x_2) \frac{Cov^2(x_1x_2)}{Var(x_1)Var(x_2)}}.$$

Az anonimizálás mindkét együtthatót érinti, tehát az anonimizálás által nem érintett változó együtthatója is torzul. A torzítás irányát és nagyságát nehéz előre jelezni: hiába ismerjük az egyes Q értékeket, a torzítás nagysága a többi kovariancia és variancia nagyságától is függ. A nehézség analóg azzal a problémával, hogy a mérési hibákból fakadó torzításokat is nehéz a vonatkozó képletek alapján előre jelezni (Maddala [2004]). A tanulmány hátralevő részében ezért szimulációs módszerrel vizsgáljuk az analitikusan nem előre jelezhető torzítások mértékét.⁸

3.1. Adatok és módszerek

A szimulációs vizsgálattal egy olyan kutató helyzetébe képzeljük magunkat, aki arra kíváncsi, milyen mértékben befolyásolja az iskolai végzettség és a településtípus a munkaerő-piaci aktivitást és a kereseteket. A kutató egy olyan nagymintás felméréshez szeretne hozzáférni, mely e változók mellett a munkaerő-piaci siker alternatív okairól – például az életkorról, a háztartásban élő gyermekek számáról és a nemről – is tartalmaz információkat. Az anonimizálatlan adatokhoz való hozzáférés azonban lehetetlen – feladatunk annak vizsgálata, mennyiben torzítja az adatcserével végzett adatvédelem a kutató becslési eredményeit.

A szimulációhoz a KSH munkaerő-felvétel 2011. első negyedéves adatait használjuk. Az adatbázis valójában már anonimizált; jelen kutatás keretében azonban úgy teszünk, mintha az anonimizálatlan adatbázis lenne birtokunkban. (Az anonimizálatlan adatbázishoz nem férhettünk hozzá.) Az adatbázis 47 162 egyénről tartalmaz adatokat; közülük 23 783 dolgozott a megkérdezés idején. Az iskolázottságot három indikátorváltozóval mérjük: szakmunkás végzettség (ISK_2), érettségi (ISK_3) és diploma (ISK_4). A településtípus ($TELTIP$) kategorikus változó, melynek definíciója: 1 = Budapest, 2 = megyei jogú város, 3 = egyéb város és 4 = község. Az „egyéb város” helyett a továbbiakban kisváros elnevezést használjuk. A nem olyan indikátorváltozó, melynek 1-es értékei a férfiakra vonatkoznak. Használni fogjuk a

⁸ Az eredményeket az utólagos randomizálásra is érvényesnek tekintjük. Egy szimulációs vizsgálat kontextusában az adatcseré és az utólagos randomizálás ekvivalens. A gyakorlatban e két technika csak abban tér el, hogy az előbbinél tudatosan, az utóbbinál véletlenszerűen dől el, melyik megfigyeléssel cserélünk fel egy adott megfigyelést. Szimulációs vizsgálatokban azonban csak véletlenszerű cserék léteznek, tudatosan kiválasztott cserék nem.

gyermek jelenléte indikátorváltozót is, melynek értéke akkor 1, ha van a háztartásban 0–6 éves gyermek.

A munkaerő-felvételben nincsenek béradatok. Pótlásként szimulált jövedelemváltozót hozunk létre a rendelkezésre álló változókra támaszkodva. A szimulált logaritmus jövedelem definíciója:

$$\log \text{kereset} = 9,67 + 0,1ISK_2 + 0,2ISK_3 + 0,6ISK_4 - 0,1(TELTIP - 1) + \\ + 0,5KOR - 0,0002KOR^2 + 0,2NEM + e,$$

ahol e standard normális eloszlást követő véletlen szám. Az együtthatókat *Kertesi-Köllő* [2001] eredményei inspirálták (lásd az idézett tanulmány F2. táblázatát). A reziduum szórása egységnyi, a determinációs együttható (R^2) értéke így durván 25 százalék.

A fiktív kutató célja tehát egyrészt a logaritmus jövedelem, másrészt a munkaerőpiaci aktivitás modellezése, előrejelzése. A kutató azonban csak az anonimizált adatbázishoz férhet hozzá. Szimulációs vizsgálatunk lényege: a szimulált jövedelmet is tartalmazó munkaerő-felvétel adatbázist anonimizálatlannak tekintjük, a képzeletbeli felhasználónak pedig különböző eljárásokkal anonimizált állományokat bocsátunk rendelkezésre. Ezután azt vizsgáljuk, milyen mértékben térnek el az anonimizált állományokból becsült eredmények az anonimizálatlan(nak tartott) állományokban becsült (valós) eredményektől.

Az anonimizálásra 16 eljárást dolgoztunk ki. Mindegyik eljárásban közös az a feltételezés, hogy az anonimizálatlan(nak tekintett) állományban a diplomás falusi megkérdezettek egy része felfedhető – annak ellenére, hogy az adatbázis már nem tartalmaz olyan kváziazonosítókat, mint például a település neve vagy kódja, a foglalkozás neve vagy kódja. Az egyes eljárások három dimenzióban térnek el.

1. A diplomás falusiak védelmét vagy csak az iskolai végzettség, vagy csak a lakóhely, vagy mindkét ismérvt együttes, vagy a két ismérvt megosztott anonimizálásával oldjuk meg. Az utóbbi azt jelenti, hogy az azonosíthatónak tekintett egyének véletlenszerűen kiválasztott felénél a diplomás végzettséget, a másik felénél a falusi lakóhely változót anonimizáljuk.

2. A donorokat egyszerű véletlen vagy rétegzett kiválasztással választhatjuk ki. A rétegzett kiválasztásnál adatcserére csak a rétegeken belül kerülhet sor. A rétegeket a nem és a korcsoport kombinációi definiálják.⁹

⁹ A korcsoportváltozónak 5 kategóriája van, melyek rendre a 16–25, 26–35, 36–45, 46–55 és 56–65 éves egyéneket azonosítják. A rétegek száma tehát $2 \times 5 = 10$.

3. A donorok – akár egyszerű, akár rétegzett – kiválasztása lehet irányítatlan vagy irányított: irányítatlan kiválasztásnál bárki lehet donor, aki nem diplomás (illetve nem falusi), míg az irányított kiválasztásnál csak a diplomásokra, illetve falusiakra leginkább hasonlító egyének – tehát az érettségizettek, illetve a kisvárosokban lakók – lehetnek donorok.

Mindegyik módszernél azt feltételeztük, hogy a falusi diplomások ($ISK_4 = 1$ és $TELTIP = 4$) p százaléka felfedhető. A szimuláció során p a 10, 25 és 50 értékeket vette fel. A munkaerő-felvétel mintájában az aktív megkérdezettek durván 6 százaléka falusi diplomás. A három értékkel tehát olyan helyzetet modellezünk, amikor egy adatbázisban a megfigyelések rendre 0,6, 1,5 és 3 százaléka fedhető fel. A kísérletet a 16 módszer és a p paraméter mindegyik kombinációjánál ezer alkalommal ismételtük meg.

A 16 módszer és a p paraméter értékei által definiált anonimizált adatbázisokban különböző becsléseket végzünk, és ezeket összehasonlítjuk az anonimizálatlan(nak tekintett) állományban végzett becsléssel. Az összehasonlításokat relatív torzítás formájában prezentáljuk. Egy adott s statisztika relatív torzítását a következőképpen számoljuk ki. Adott módszer és p paraméter mellett az anonimizálatlan (vagy annak tekintett) adatbázist egy adott eljárással R alkalommal anonimizáljuk (vizsgálatainkban $R = 1000$). Az r -edik replikációban a statisztika értéke s_r . Az anonimizálatlan állományban a statisztika értéke S . A relatív torzítás képlete:

$$s \text{ relatív torzítása} = 100 \frac{\sum_{r=1}^R s_r - RS}{RS}. \quad /15/$$

Ha a /13/ képletben megfogalmazott feltétel teljesül, a kovarianciabecslések relatív torzítása a /2/ képlet alapján:

$$\frac{\hat{\beta} - \beta}{\beta} = -\frac{p}{Var(x)}. \quad /16/$$

A relatív torzítás tehát egyenesen arányos az adatcserével védett megfigyelések arányával. A /14/ egyenlet bonyolultsága miatt a többváltozós regressziós becslések relatív torzítását a /16/ egyenlet segítségével sem lehet előre jelezni.

3.2. Eredmények

A kovarianciabecslések torzulásai. A képzeletbeli kutató célja a jövedelmekben és a munkaerő-piaci aktivitásban mérhető egyenlőtlenségek elemzése. Mivel a több-

változós regressziós becslések kovarianciák és varianciák függvényei, érdemes először a kovarianciabecslések relatív torzításait elemezni. Az 1. és 2. táblázat egyrészt a diplomás iskolai végzettség és a szimulált logaritmus jövedelem, másrészt a diplomás végzettség és a falusi lakóhely indikátorváltozók kovarianciáinak relatív torzításait mutatja.

1. táblázat

A diplomás indikátorváltozó és a szimulált logaritmus jövedelem kovarianciájának relatív torzulásai

Adatsere módszere	$p = 10$ százalék		$p = 25$ százalék		$p = 50$ százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Iskolai végzettség anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-1,600	0,906	-4,000	1,365	-8,192	1,799
véletlenszerűen kiválasztott közeli donorokkal	-1,924	1,016	-4,764	1,469	-9,622	1,902
rétégzéssel kiválasztott donorokkal	-2,142	0,770	-5,413	1,159	-11,043	1,596
rétégzéssel kiválasztott közeli donorokkal	-2,089	0,889	-4,991	1,341	-10,188	1,703
Településtípus anonimizálása:	0	0	0	0	0	0
Iskolai végzettség és településtípus együttes anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-1,598	0,892	-4,035	1,385	-8,227	1,831
véletlenszerűen kiválasztott közeli donorokkal	-1,916	1,022	-4,748	1,521	-9,476	1,863
rétégzéssel kiválasztott donorokkal	-2,138	0,764	-5,407	1,210	-10,970	1,610
rétégzéssel kiválasztott közeli donorokkal	-2,081	0,871	-5,045	1,299	-10,168	1,580
Iskolai végzettség és településtípus megosztott anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-0,829	0,647	-1,963	1,026	-3,979	1,343
véletlenszerűen kiválasztott közeli donorokkal	-0,941	0,726	-2,445	1,092	-4,775	1,528
rétégzéssel kiválasztott donorokkal	-1,060	0,562	-2,693	0,866	-5,354	1,204
rétégzéssel kiválasztott közeli donorokkal	-1,051	0,617	-2,488	0,942	-5,130	1,273

Megjegyzés. A kovariancia valós nagysága 0,08. A településtípus anonimizálása nem torzítja a vizsgált kovarianciát.

2. táblázat

A diplomás indikátorváltozó és a falusi településtípus indikátorváltozó kovariációjának relatív torzításai

Adatsere módszere	$p = 10$ százalék		$p = 25$ százalék		$p = 50$ százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Iskolai végzettség anonimizálása: véletlenszerűen kiválasztott donorokkal	8,224	0,807	20,672	1,256	41,921	1,819
véletlenszerűen kiválasztott közeli donorokkal	11,731	0,770	29,212	1,177	58,391	1,676
rétégzéssel kiválasztott donorokkal	8,110	0,797	20,366	1,271	41,472	1,786
rétégzéssel kiválasztott közeli donorokkal	11,724	0,787	29,222	1,254	58,614	1,724
Településtípus anonimizálása: véletlenszerűen kiválasztott donorokkal	8,219	0,810	20,661	1,266	41,915	1,715
véletlenszerűen kiválasztott közeli donorokkal	15,600	0,684	38,818	1,011	77,630	1,408
rétégzéssel kiválasztott donorokkal	8,091	0,795	20,365	1,284	41,368	1,725
rétégzéssel kiválasztott közeli donorokkal	15,236	0,666	37,990	1,056	75,870	1,401
Iskolai végzettség és településtípus eggyüttes anonimizálása: véletlenszerűen kiválasztott donorokkal	7,505	0,951	18,475	1,479	36,024	2,148
véletlenszerűen kiválasztott közeli donorokkal	7,438	1,031	18,044	1,569	34,800	2,159
rétégzéssel kiválasztott donorokkal	7,344	0,948	18,161	1,542	35,216	2,097
rétégzéssel kiválasztott közeli donorokkal	7,157	1,044	17,292	1,612	32,834	2,163
Iskolai végzettség és településtípus megosztott anonimizálása: véletlenszerűen kiválasztott donorokkal	8,141	0,791	20,371	1,264	40,957	1,777
véletlenszerűen kiválasztott közeli donorokkal	13,591	0,738	33,829	1,119	67,208	1,587
rétégzéssel kiválasztott donorokkal	8,044	0,810	20,206	1,271	40,392	1,775
rétégzéssel kiválasztott közeli donorokkal	13,439	0,755	33,426	1,166	66,286	1,576

Megjegyzés. A kovariancia valós nagysága $-0,031$.

Az eredmények megfelelnek annak a várakozásnak, miszerint a torzítás mértéke egyenesen arányos az anonimizált megfigyeléspárok (p) arányával (lásd a /16/ egyenletet). Durván két és félszer akkora torzításokat tapasztalunk a $p = 25$ oszlopokban, mint a $p = 10$ oszlopokban, és kétszer akkora torzítást a $p = 50$ oszlopokban, mint a $p = 25$ oszlopokban. A 2. táblázatban lényegesen nagyobb torzításokat tapasztalunk, mint az 1. táblázatban. A szimulált logaritmus jövedelem és a diploma kovariációjának torzítását nagyon alacsonyan lehet tartani, ha az anonimizált megfigyeléspárok aránya 10 vagy 25 százalék.

Vegyük ezután szemügyre, hogy az adatsere melyik módszere minimalizálja a kovarianciabecslések torzítását. A szimulált logaritmus jövedelem és a diplomás in-

dikátorváltozó kovarianciájának torzítását a véletlenszerű és irányítatlan kiválasztás minimalizálja a donorok kiválasztásának négy módszere közül. Talán meglepő, de sem a rétegzés, sem az irányítás – azaz a donorok halmazának szűkítése és a közeli donorok kiválasztása – nem javít az eredményeken. A munkaerő-felvételhez hasonló állományokban a donorok kiválasztásakor tehát érdemes a véletlenre hagyatkozni, és nem érdemes sem tudatos szűkítéssel, sem rétegzéssel a donorok kiválasztásába beavatkozni. A torzítások legkisebb mértékben a rétegzett kiválasztásnál szóródnak, de a szórások között jóval kisebb a különbség, mint az átlagos torzítások között.

A falusi településtípus és a diplomás indikátorváltozó kovarianciájának torzítását ezzel szemben a rétegzett kiválasztás minimalizálja. A véletlenszerű és a rétegzett kiválasztás közötti különbség azonban elenyésző mértékű, és jóval kisebb annál, amit a jövedelem és a diplomás végzettség kovarianciájának az elemzésekor találtunk. Tehát továbbra is fenntartható az a következtetés, hogy a donorok kiválasztásakor a rétegzésnek nincs hozzáadott értéke.

Egyváltozós lineáris regressziós becslések torzulásai. Képzeltbeli kutatónk egyik fő célja a (szimulált) jövedelmi különbségek elemzése. Kutatását a diplomások és érettségizettek, illetve kisvárosiak és falusiak közötti jövedelmi különbségek leírásával kezdi. Az egyváltozós regressziós becslésekben bekövetkező relatív torzításokat a 3. és 4. táblázatok mutatják.

A torzítás mértéke ismét egyenesen arányos a p paraméterrel. A legfontosabb eredmény az, hogy az adatvédelemben bevont megfigyeléspárok arányától függetlenül mindig található olyan eljárás, amely a relatív torzítást 10 százalék alatt tartja. Az adatok alapján a torzítás akkor minimális, ha az adatcserét megosztjuk az iskolai végzettség és a településtípus között, a donorokat pedig véletlenszerűen választjuk ki. Tehát sem a rétegzés, sem az irányítás – azaz a donorok halmazának szűkítése és a közeli donorok kiválasztása – nem javít az eredményeken. Egyik eredmény sem meglepő. A megosztott adatcsere sikere annak tulajdonítható, hogy egy adott változónál az adatcserében részt vevő megfigyelések – és ezáltal a p paraméter effektív nagysága – a felére csökken. A véletlenszerű és irányítatlan kiválasztás sikere pedig annak tudható, hogy az egyváltozós regressziós becslések kovarianciák és varianciák hányadosai, az adatcsere pedig csak az előbbit torzítja.

Többváltozós lineáris regressziós becslések torzulásai. Képzeltbeli kutatónk tudja, hogy az egyváltozós regressziós együtthatók torzan mérik az oksági hatásokat, ezért az egyváltozós elemzések után olyan többváltozós lineáris regressziós modellt becsül, melynek magyarázóváltozói: az iskolai végzettséget mérő három indikátorváltozó, a településtípust mérő három indikátorváltozó, a nem, az életkor és annak négyzete. Az iskolázottságnál és a településtípusnál az alapfokú végzettség, illetve Budapest a referenciakategória. A kutató arra kíváncsi, mennyivel haladja meg a diplomások ceteris paribus keresete az érettségizettét, illetve mennyivel haladja meg a „kisvárosiak” (azaz a nem megyei jogú városok) lakóinak ceteris paribus keresete a falusiak keresetét. A többváltozós regressziós becslések relatív torzításait az 5. és 6. táblázatok mutatják.

Az egyváltozós becslések elemzésekor azt találtuk, hogy 1. a torzítást a véletlenszerű adatcsere minimalizálja, 2. ha lehet, érdemes a falusi diplomások védelmét megosztott adatcserevel biztosítani, és 3. a torzítás mértéke még viszonylag tömeges adatcsere esetén is 10 százalék alatt tartható. A többváltozós együtthatók relatív torzításaira e következtetések közül csak az első illik maradéktalanul: most is a donorok véletlenszerű és irányítatlan kiválasztása a legjobb módszer. A 2. következtetés most csak a diplomások kereseti előnyére vonatkozó becslésekre igaz – a kisvárosiak kereseti előnyére vonatkozó becslések torzulásait ezzel szemben az együttes anonimizálás minimalizálja. Végül: a többváltozós együtthatók nagyobb mértékben torzulnak, mint az egyváltozós együtthatók.

3. táblázat

Diplomások és érettségizettek közti nyers bérkülönbség relatív torzítása

Adatcsere módszere	$p = 10$ százalék		$p = 25$ százalék		$p = 50$ százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Iskolai végzettség anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-1,686	0,926	-4,174	1,483	-8,306	1,931
véletlenszerűen kiválasztott közeli donorokkal	-2,590	1,318	-6,378	2,018	-12,720	2,559
rétegzéssel kiválasztott donorokkal	-2,158	0,849	-5,377	1,278	-10,844	1,737
rétegzéssel kiválasztott közeli donorokkal	-2,632	1,146	-6,797	1,748	-13,522	2,099
Településtípus anonimizálása:	0	0	0	0	0	0
Iskolai végzettség és településtípus együttes anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-1,628	0,960	-4,203	1,483	-8,384	1,941
véletlenszerűen kiválasztott közeli donorokkal	-2,558	1,327	-6,315	1,965	-12,601	2,600
rétegzéssel kiválasztott donorokkal	-2,176	0,852	-5,322	1,310	-10,864	1,707
rétegzéssel kiválasztott közeli donorokkal	-2,647	1,174	-6,832	1,786	-13,642	2,217
Iskolai végzettség és településtípus megosztott anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-0,837	0,705	-2,087	1,068	-4,161	1,501
véletlenszerűen kiválasztott közeli donorokkal	-1,322	0,968	-3,139	1,477	-6,252	1,999
rétegzéssel kiválasztott donorokkal	-1,095	0,593	-2,685	0,971	-5,313	1,289
rétegzéssel kiválasztott közeli donorokkal	-1,298	0,827	-3,450	1,279	-6,859	1,761

Megjegyzés. Bérkülönbségen a szimulált logaritmus bérben mutatkozó különbséget értjük. A bérkülönbség valós nagysága 0,464. A településtípus anonimizálása nem torzítja a vizsgált bérkülönbséget.

4. táblázat

Falusiak és kisvárosiak közti nyers bérkülönség relatív torzítása

Adatcsere módszere	$p = 10$ százalék		$p = 25$ százalék		$p = 50$ százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Iskolai végzettség anonimizálása:	0	0	0	0	0	0
Településtípus anonimizálása: véletlenszerűen kiválasztott donorokkal	0,903	1,667	2,182	2,625	4,295	3,497
véletlenszerűen kiválasztott közeli donorokkal	2,911	2,597	7,172	3,831	14,065	5,015
rétegzéssel kiválasztott donorokkal	1,689	1,450	4,043	2,303	8,120	3,099
rétegzéssel kiválasztott közeli donorokkal	4,576	2,325	11,639	3,593	22,922	4,496
Iskolai végzettség és településtípus együttes anonimizálása: véletlenszerűen kiválasztott donorokkal	0,781	1,636	2,224	2,587	4,215	3,416
véletlenszerűen kiválasztott közeli donorokkal	2,894	2,624	7,257	3,902	14,088	5,124
rétegzéssel kiválasztott donorokkal	1,611	1,484	4,046	2,336	8,012	3,112
rétegzéssel kiválasztott közeli donorokkal	4,533	2,380	11,527	3,530	22,851	4,613
Iskolai végzettség és településtípus megosztott anonimizálása: véletlenszerűen kiválasztott donorokkal	0,427	1,137	1,083	1,873	2,019	2,568
véletlenszerűen kiválasztott közeli donorokkal	1,412	1,894	3,763	2,946	7,154	3,971
rétegzéssel kiválasztott donorokkal	0,821	1,036	1,997	1,663	4,112	2,326
rétegzéssel kiválasztott közeli donorokkal	2,267	1,671	5,770	2,573	11,460	3,547

Megjegyzés: Bérkülönségen a szimulált logaritmus térben mutatkozó különbséget értjük. A bérkülönség valós nagysága 0,161. Az iskolai végzettség anonimizálása nem torzítja a vizsgált bérkülönséget.

5. táblázat

Diplomások és érettségizettek közti korrigált bérkülönbség relatív torzítása

Adatsere módszere	$p = 10$ százalék		$p = 25$ százalék		$p = 50$ százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Iskolai végzettség anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-3,081	0,993	-7,647	1,592	-15,190	2,075
véletlenszerűen kiválasztott közeli donorokkal	-4,117	1,351	-10,052	2,026	-19,796	2,685
rétegzéssel kiválasztott donorokkal	-3,021	1,011	-7,504	1,524	-15,098	2,100
rétegzéssel kiválasztott közeli donorokkal	-3,922	1,350	-9,969	2,068	-19,642	2,552
Településtípus anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-0,440	0,105	-1,072	0,185	-2,038	0,301
véletlenszerűen kiválasztott közeli donorokkal	-0,587	0,078	-1,350	0,174	-2,308	0,367
rétegzéssel kiválasztott donorokkal	-0,442	0,113	-1,064	0,192	-2,023	0,287
rétegzéssel kiválasztott közeli donorokkal	-0,580	0,080	-1,336	0,181	-2,271	0,369
Iskolai végzettség és településtípus együttes anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-3,049	1,042	-7,740	1,648	-15,489	2,103
véletlenszerűen kiválasztott közeli donorokkal	-3,800	1,379	-9,432	0,200	-18,642	2,664
rétegzéssel kiválasztott donorokkal	-3,052	1,028	-7,506	1,622	-15,257	2,111
rétegzéssel kiválasztott közeli donorokkal	-3,661	1,402	-9,371	2,147	-18,686	2,686
Iskolai végzettség és településtípus megosztott anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-1,752	0,730	-4,350	1,134	-8,576	1,571
véletlenszerűen kiválasztott közeli donorokkal	-2,398	0,978	-5,688	1,546	-10,975	2,068
rétegzéssel kiválasztott donorokkal	-1,747	0,705	-4,270	1,157	-8,391	1,559
rétegzéssel kiválasztott közeli donorokkal	-2,233	0,970	-5,703	1,519	-11,090	2,099

Megjegyzés: Bérkülönbségen a szimulált logaritmus bérben mutatkozó különbséget értjük. A bérkülönbség valós nagysága 0,4.

6. táblázat

Falusiak és kisvárosiak közti korrigált bérkülönbség relatív torzítása

Adatcsere módszere	$p = 10$ százalék		$p = 25$ százalék		$p = 50$ százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Iskolai végzettség anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-4,611	0,694	-11,002	1,131	-20,423	1,649
véletlenszerűen kiválasztott közeli donorokkal	-5,836	0,575	-13,539	0,984	-23,925	1,574
rétegzéssel kiválasztott donorokkal	-4,506	0,713	-10,828	1,099	-20,194	1,649
rétegzéssel kiválasztott közeli donorokkal	-5,906	0,566	-13,593	1,003	-24,024	1,487
Településtípus anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-2,490	2,665	-6,229	4,145	-12,746	5,701
véletlenszerűen kiválasztott közeli donorokkal	-4,637	4,119	-11,528	6,049	-23,129	8,183
rétegzéssel kiválasztott donorokkal	-2,437	2,554	-6,480	4,205	-13,083	5,489
rétegzéssel kiválasztott közeli donorokkal	-4,916	4,146	-11,635	6,479	-23,443	8,330
Iskolai végzettség és településtípus együttes anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-1,161	2,814	-1,646	4,513	-1,132	6,004
véletlenszerűen kiválasztott közeli donorokkal	3,965	4,359	10,386	6,481	21,544	8,476
rétegzéssel kiválasztott donorokkal	-1,120	2,841	-2,081	4,556	-1,939	5,951
rétegzéssel kiválasztott közeli donorokkal	3,556	4,400	9,884	6,501	21,085	8,516
Iskolai végzettség és településtípus megosztott anonimizálása:						
véletlenszerűen kiválasztott donorokkal	-3,571	1,917	-8,617	3,074	-16,178	4,235
véletlenszerűen kiválasztott közeli donorokkal	-5,249	2,959	-12,026	4,755	-22,414	6,460
rétegzéssel kiválasztott donorokkal	-3,488	1,921	-8,455	3,043	-15,947	4,325
rétegzéssel kiválasztott közeli donorokkal	-5,344	3,068	-12,420	4,770	-22,749	6,456

Megjegyzés. Bérkülönbségen a szimulált logaritmus bérben mutatkozó különbséget értjük. A bérkülönbség valós nagysága 0,089.

Többváltozós probit regressziós becslések torzulásai. Képzeltbeli kutatónk másik kérdése az, mekkora a munkaerő-piaci részvételben mért előnye a diplomásoknak és a kisvárosiaknak az érettségizettekhez, illetve a falusiakhoz képest. Mivel a munkaerő-piaci részvétel diszkrét, az előrejelzés kézenfekvő módszere a többváltozós

probit regresszió. A magyarázóváltozók: az iskolai végzettséget mérő három indikátorváltozó, a településtípust mérő három indikátorváltozó, a nem, az életkor és annak négyzete, valamint a 6 éves vagy annál fiatalabb gyermek jelenléte a háztartásban.

A lineáris regressziós becslésekkel kapcsolatos eredmények szerint a torzítást akkor minimalizálhatjuk, ha a donorokat véletlenszerűen – tehát irányítás, illetve rétegzés nélkül – választjuk ki. Azt is láttuk, hogy a torzítás leginkább az adatsere megosztásával mérsékelhető. A lineáris regresszióra vonatkozó kvalitatív következtetések érvényesek a probit modelleknél bekövetkező torzításokra is. Ennek oka az, hogy a probit modell valójában egy olyan lineáris regressziós modellel ekvivalens, melynek folytonos függő változója nem megfigyelhető, és a megfigyelt diszkrét kimenet birtokában csak annyit tudunk, hogy a nem megfigyelhető függő változó nagyobb nullánál vagy sem. Ha a látens függő változó teljes mértékben megfigyelhető lenne, a probit becsléseket úgy számolnánk ki, hogy a lineáris regressziós becsléseket elosztjuk a becült reziduum szórásával. Probit modelleknél a reziduum szórása sem ismert, mégis, a probit együtthatók lineáris regressziós együtthatók és nem megfigyelhető reziduális szórások hányadosai. Emiatt a lineáris regressziós becslések torzulásaira vonatkozó kvalitatív következtetések a probit becslésekre is érvényesek.

A szimulált jövedelmekre vonatkozó lineáris regressziós becslések alapján viszont nem lehet megjósolni a probit becslések torzításainak nagyságát. A 7. táblázat az előbbi többváltozós probit modell becsléseiben bekövetkező relatív torzításokat mutatja. A táblázat lényegesen egyszerűbb az előzőknél. Az előző vizsgálatban használt 16 módszer helyett most csak kettőt használunk: a donorok véletlenszerű és irányítatlan kiválasztásán alapuló együttes, illetve a megosztott adatszerét.

7. táblázat

Probit együtthatók relatív torzítása

Adatsere módszere	<i>p</i> = 10 százalék		<i>p</i> = 25 százalék		<i>p</i> = 50 százalék	
	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Diploma – érettségi különbség együtthatója:						
Iskolai végzettség és településtípus együttes anonimizálása	-5,215	1,124	-12,893	1,674	-25,867	2,267
Iskolai végzettség és településtípus megosztott anonimizálása	-2,673	0,804	-6,645	1,198	-13,335	1,649
Egyéb város – falu különbség együtthatója:						
Iskolai végzettség és településtípus együttes anonimizálása	1,813	2,520	5,325	3,825	13,799	5,364
Iskolai végzettség és településtípus megosztott anonimizálása	-2,824	1,606	-6,667	2,448	-12,007	3,696

Megjegyzés. Az anonimizálatlan adatbázisban a diploma-érettségi különbség probit együtthatója 0,467, az egyéb város-falu különbség együtthatója 0,086.

A diplomások előnyét mérő probit együtthatók torzulásai durván másfélszeresei a diplomások előnyét mérő többváltozós OLS-együtthatók torzulásainál. A probit együtthatók torzulásai most is a megosztott adatcserénél kisebbek. A kisvárosiak előnyét mérő probit együtthatók értelmezése nehezebb, mivel a torzítás együttes adatcserénél pozitív, megosztott adatcserénél negatív.

4. Következtetések

A tanulmányban áttekintettük a felfedés elleni védelem statisztikai következményeit, és szimulációs módszerrel vizsgáltuk az adatcsere kovariancia- és regressziós becslésekre gyakorolt hatását. Amellett érveltünk, hogy az adatcserének kitüntetett szerepe van a felfedés elleni védelemben. Egyrészt az adatvédelmi jogszabályok leginkább a kváziazonosítók módosítására ösztönzik az adatgazdákat, és ezek a változók legtöbbször kategorikusak. A kategorikus változók anonimizálására alkalmas technikák közül pedig csak az adatcsere (illetve az utólagos randomizálás) az, amely a változók átlagait és szórásait nem, a kovarianciabecsléseket pedig ismert mértékben módosítja. Az adatcsere többváltozós regressziós becslésekre gyakorolt hatását viszont nehéz pontosan megjósolni. Ezért az adatcsere statisztikai következményeinek elemzését szimulációs módszerrel folytattuk.

A szimulációhoz a KSH munkaerő-felvételének 2011. első negyedéves adatait használtuk. Azt vizsgáltuk, az adatcsere egyes technikai milyen mértékben torzítanak egy- és többváltozós lineáris regressziós, valamint többváltozós probit becsléseket. A szimuláció során olyan adatvédelmi problémát elemzünk, amikor a szokásos kváziazonosítók – például a településkód – anonimizálása nem nyújt tökéletes védelmet, és az anonimitás garantálásához még olyan változókat is anonimizálni kell, melyeket regressziós elemzésekben magyarázóváltozóként fogunk használni. Vizsgálatunkban az iskolai végzettséget és a településtípust használtuk: azt feltételeztük, hogy a falusi diplomások egy része felfedhető maradt.

Azt találtuk, hogy a becslések akkor torzulnak minimális mértékben, ha 1. az adatvédelem terhét megosztják a felfedési kockázatot növelő változók – példánkban az iskolai végzettség és a településtípus – között, és 2. ha a védelemre szoruló megfigyeléseknél a kiválasztott ismérveket teljesen véletlenszerűen – tehát rétegzés és irányítás nélkül – cseréljük ki alacsony felfedési kockázatú egyének ismérveivel. A megosztott adatcsere természetesen nem mérsékli szükségszerűen a felfedési kockázatot, és lehetséges, hogy az adatszolgáltatók anonimitása csak a változók együttes anonimizálásával érhető el. Eredményeink szerint az együttes anonimizálás sem jár számottevő pótlólagos torzító hatással.

A szimulációs vizsgálatokban konkrét számokkal tudtuk kifejezni, milyen mértékben torzulnak a becslések az adatcserével anonimizált állományokban. Ha a falusi diplomások 25 százaléka szorul további védelemre, a regressziós becslések relatív torzítása az esetek döntő többségében 10 százalék alatt marad. Vajon elhanyagolható vagy jelentős a 10 százalékos torzítás? A survey statisztika irodalma különbséget tesz mintavételi és a nemmintavételi hibák között (*Sarndal–Swensson–Wretman* [1992], *Biemer–Lyberg* [2003]). Nyilvánvaló, hogy az anonimizálásból fakadó torzítás értelmezhető a mérési hibák elméletén belül.¹⁰ Az anonimizáláshoz adatcserét használtunk, amely alulbecsülheti a kovarianciákat – feltéve, hogy a változók között pozitív az összefüggés. A regresszióelemzés kontextusában ugyanezzel a következménnyel – tehát a (pozitív) együtthatók alulbecslésével – járna, ha a változókat mondjuk zajosítással védenénk, hiszen a zajosítás növeli a varianciákat és ezáltal csökkenti a parciális korrelációkat. A magyarázóváltozók szórása inflálódásának következményeit hagyományosan a méréselmélet tárgyalja; az elmélet szerint a regresszióelemzés kontextusában a magyarázóváltozók mérési hibája koptatja a regressziós együtthatókat (*Fuller* [1987], *Maddala* [2004]). Az anonimizálásból fakadó torzítás következményeit tekintve mérési hiba. Valószínű, hogy az adatgyűjtés során keletkező mérési hibák jóval nagyobbak az anonimizálásból fakadó, nem számszerűsíthető mérési hibáknál. Ráadásul az anonimizálás releváns paramétereinek publikálása lehetővé teszi ez utóbbi hibaforrás kiküszöbölését (*Gouweleeuw et al.* [1998], *Shlomo* [2010]). Az anonimizálásból fakadó 10 százalékos relatív torzítás vélhetően sokkal kisebb az adatgyűjtés során keletkező hibáknál. Ám az anonimizálásból fakadó mérési hiba hozzáadódik ezekhez a hibákhoz. Az anonimizálásból fakadó becslési hibát tehát akkor tolerálhatják a felhasználók, ha az anonimizálatlan adatbázisban már eleve kismértékű a mérési hibák szórása.

Következtetéseink természetesen nem tekinthetők véglegesnek, hiszen azok csak egyetlenegy adatbázis szimulációs elemzésén alapulnak. Lehetséges, hogy a munkaerő-felvételtől eltérő adatokon másfajta eredményeket kaptunk volna. Például érdeemes lenne más adatokon ellenőrizni annak a következtetésnek az érvényességét, miszerint a rétegzés nem javít a becslési eredmények megbízhatóságán. További kutatásoknak kell tisztázniuk, hogy másfajta adatbázisokon milyen mértékű (lehet) az adatcseréből fakadó torzítás.

Az elméleti következtetések ideiglenes jellege mellett a tanulmánynak van egy gyakorlati szempontból fontos tanulsága: a felhasználók érdekei akkor sérülnek a

¹⁰ A mintavételi hibák elméletének használata nehezebb. Abból indulunk ki, hogy egy adott anonimizálási eljárás alulbecsli a sokasági paramétert. Ezzel párhuzamosan a konfidenciaintervallumokat definiáló alsó és felső értékeket is alulbecsüljük. A mintavételi hibák elmélete azt sugallja, hogy az anonimizálás ugyanolyan következményekkel jár, mint a mintanagyság csökkenése. A 10 százalékos relatív torzítás értelmezéséhez tehát azt a relatív mintanagyság-csökkenést kell kiszámolni, aminek hatására a konfidenciaintervallum alsó határa ugyanolyan mértékben csökken, mint a torzítás hatására. A nehézség abból fakad, hogy a kérdésre adott válasz nem független a sokasági paraméter értékétől.

legkevésbé, ha a felfedés elleni védelmet adatcserével vagy utólagos randomizálással biztosítják az adatgazdák. Ezen technikák egyrészt számos kedvező statisztikai tulajdonsággal rendelkeznek, másrészt eleve a kategorikus kváziazonosítók módosítására szolgálnak – ezek azok a változók, melyek tipikusan lehetővé teszik a felfedést. Az adatcsere és az utólagos randomizálás elterjedését azonban más érdekek és megfontolások korlátozhatják. Egyrészt ezek nem védik mindig tökéletesen a személyes adatokat: az eljárások a valóságban nem létező attribútumkombinációkat hozhatnak létre, melyek ismeretében a rosszindulatú felhasználó rájöhet arra, mely esetek védelme miatt használták a módszert (Gouweleeuw *et al.* [1998], Boudreau [2005] 9. old.). Másrészt ezek a technikák kifinomulttan ugyan, de „manipulálják” az adatokat, ami megrendítheti az adatokat kezelő intézményekkel szembeni bizalmat (Evans–Zayatz–Slata [1996]). Ezzel szemben az olyan egyszerű technikák, mint egyes kváziazonosítók visszatartatása vagy néhány egyszerű eset törlése „manipulációktól” mentes anonimizált állományokat hoznak létre. Az egyszerű törlésnél vagy adatvisszatartásnál kifinomultabb technikák elterjedésére tehát akkor lehet számítani, ha már eleve bíznak a statisztikával foglalkozó intézményekben, vagy ha a felhasználók megértik, és nem hamisításként értékelik az anonimizálási technikákat.

Irodalom

- BÁNSZEGI K. [1997]: Felfedést akadályozó módszerek a statisztikai tájékoztatásban. *Statisztikai Szemle*. 75. évf. 12. sz. 1039–1046. old.
- BIEMER, P. P. – LYBERG, L. E. [2003]: *Introduction to Survey Quality*. John Wiley & Sons. Hoboken.
- BOUDREAU, J.-R. [2005]: *Data Swapping is Not the Panacea*. Proceedings of Statistics Canada's Symposium 2005. "Methodological Challenges for Future Information Needs." Statistics Canada.
- BRAND, R. [2002]: Microdata Protection through Noise Addition. In: *Domingo-Ferrer, J. (ed): Inference Control in Statistical Databases. From theory to practice*. Springer. Berlin. pp. 97–116.
- BYCROFT, C. – MERRETT, K. [2005]: *Experience of Using a Post Randomisation Method at the Office for National Statistics*. Monographs of Official Statistics. UNECE and Eurostat Work Session on Statistical Data Confidentiality. Geneva.
- COMMISSION OF EUROPEAN COMMUNITIES [2002]: Commission Regulation (EC) No 831/2002. http://eur-lex.europa.eu/pri/en/oj/dat/2002/l_133/l_13320020518en00070009.pdf
- DALENIUS, T. – REISS, S. P. [1982]: Data-swapping. A Technique for Disclosure Control. *Journal of Statistical Planning and Inference*. Vol. 6. No. 1. pp. 73–85.
- DOMINGO-FERRER, J. – TORRA, V. [2001a]: Disclosure Control Methods and Information Loss for Microdata. In: *Doyle, P. – Lane, J. – Theeuwes, J. – Zayatz, L. (eds.): Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*. North-Holland, Amsterdam. pp. 93–112.

- DOMINGO-FERRER, J. – TORRA, V. (2001b): A Quantitative Comparison of Disclosure Control Methods for Microdata. In: *Doyle, P. – Lane, J. – Theeuwes, J. – Zayatz, L. (eds.): Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*. North-Holland, Amsterdam. pp. 113–134.
- ERDEI V. – HORVÁTH R. [2004]: Az adatfeldedés elleni védelem statisztikai eszközei. *Statisztikai Szemle*. 82. évf. 8. sz. 705–727. old.
- EVANS, T. – ZAYATZ, L. – SLANTA, J. [1996]: *Using Noise for Disclosure Limitation of Establishment Tabular Data*. US Bureau of the Census. www.census.gov/prod/2/gen/96arc/iaevans.pdf
- FAGAN, W. T. – GREENBERG, B. [1985]: *Algorithms for Making Tables Additive: Raking, Maximum Likelihood, and Minimum Chi-Square*. US Bureau of the Census. www.amstat.org/sections/srms/proceedings/papers/1988_086.pdf
- FISCHETTI, M. – SALAZAR, J. J. [1998]: Experiments with Controlled Rounding for Statistical Disclosure Control in Tabular Data with Linear Constraints. *Journal of Official Statistics*. Vol. 14. No. 4. pp. 553–565.
- FULLER, W. A. [1987]: *Measurement Error Models*. John Wiley & Sons. New York.
- GOUWELLEEUW, J. – KOOIMAN, P. – WILLENBORG, L. – DE WOLF, P.-P. [1998]: The Post Randomization Method for Protecting Microdata. *Qüestiió*. Vol. 22. No. 1. pp. 145–156.
- HUNDEPOOL, A. – DOMINGO-FERRER, J. – FRANCONI, J. – GIESSING, S. – LENZ, R. – NAYLOR, J. – SCHULTE NORDHOLT, E. – SERI, G. – DE WOLF, P.-P. [2010]: *Handbook on Statistical Disclosure Control*. ESSNet SDC. On-line: http://neon.vb.cbs.nl/casc/%5CSDC_Handbook.pdf
- KERTESI G. – KÖLLŐ J. [2001]: A gazdasági átalakulás két szakasza és az emberi tőke átértékelődése. *Közgazdasági Szemle*. 48. évf. 11. sz. 897–919. old.
- KIM, J. J. [1990]: Subpopulation Estimation for the Masked Data. In: *Proceedings of the ASA Section on Survey Research Methods*. Alexandria. pp. 303–308.
- KOOIMAN, P. – WILLENBORG, L. C. R. J. – GOUWELLEEUW, J. M. [1997]: *PRAM: A Method for Disclosure Limitation of Microdata*. Research paper 9705. Statistics Netherlands. Voorburg, Heerlen.
- LENZ, R. – ROSEMAN, M. – VORGRIMLER, D. – STURM, R. [2006]: *Anonymising Business Micro Data – Results of a German Project*. Statistisches Bundesamt. Berlin.
- LIU, F. – LITTLE, R. J. A. [2003]: SMIKe vs. Data Swapping and PRAM for Statistical Disclosure Control in Microdata: A Simulated Study. In: *Proceedings of the ASA Section on Survey Research Methods*. Alexandria. pp. 2497–2502.
- MADDALA, G. S. [2004]: *Bevezetés az ökonometriába*. Nemzeti Tankönyvkiadó. Budapest.
- MATEO-SANZ, J. M. – DOMINGO-FERRER, J. [1998]: A Comparative Study of Microaggregation Methods. *Qüestiió*. Vol. 22. No. 3. pp. 511–526.
- REISS, S. P. [1984]: Practical Data-swapping: The First Steps. *ACM Transactions on Database Systems*. Vol. 9. No. 1. pp. 20–37.
- SARNDAL, C. E. – SWENSSON, B. – WRETMAN, J. [1992]: *Model Assisted Survey Sampling*. Springer. New York.
- SCHMID, M. [2006]: Estimation of a Linear Model under Microaggregation by Individual Ranking. *Allgemeines Statistisches Archiv*. Vol. 90. No. 3. pp. 419–438.
- SCHMID, M. – SCHNEEWEISS, H. [2005]: The Effect of Microaggregation Procedures on the Estimation of Linear Models: A Simulation Study. In: *Pohlmeier, W. – Ronning, G. – Wagner,*

- J. (eds): *Econometrics of Anonymized Micro Data. Jahrbücher für Nationalökonomie und Statistik*. Vol. 225. No. 5. pp. 529–543.
- SCHMID, M. – SCHNEEWEISS, H. – KUCHENHOFF, H. [2007]: Estimation of a Linear Regression under Microaggregation with the Response Variable as a Sorting Variable. *Statistica Neerlandica*. Vol. 61. No. 4. pp. 407–431.
- SCHMID, M. – SCHNEEWEISS, H. [2008]: *Estimation of a Linear Model in Transformed Variables under Microaggregation by Individual Ranking*. Manuscript. University of Munich. Munich.
- SCHMID, M. – SCHNEEWEISS, H. [2009]: The Effect of Microaggregation by Individual Ranking on the Estimation of Moments. *Journal of Econometrics*. Vol. 153. No. 2. pp. 174–182.
- SHLOMO, N. [2010]: *Measurement Error and Statistical Disclosure Control*. S3RI Methodology Working Papers, M10/05. Southampton Statistical Sciences Research Institute, University of Southampton. Southampton.
- SHLOMO, N. – TUDOR, C. – GROOM, P. [2010]: *Data Swapping for Protecting Census Tables*. S3RI Methodology Working Papers, M10/06. Southampton Statistical Sciences Research Institute, University of Southampton. Southampton.
- SULLIVAN, C. M. [1992]: *An Overview of Disclosure Principles*. U.S. Bureau of the Census. www.census.gov/srd/papers/pdf/rr92-09.pdf
- SZÉP K. – GADÁCSI K. [2010]: *Adatok anonimizálása, hozzáférés a mikroadatokhoz, archiválás*. Előadás a Fényes Elek Műhelyben. Május 26.
<http://fenyeselekegyesulet.wordpress.com/eloadasok/2010-2/adatok-anonimizalasa-hozzaferes-a-mikroadatokhoz-archivalasi-tevekenysegek/>

Summary

The paper reviews the statistical implications of several methods of disclosure control. The effects of data swapping on covariance and linear regression estimates are examined in details. It is argued that data swapping has several advantages over alternative methods. The effect of various data swapping techniques on covariance and linear regression estimates is examined using simulations. The simulation study uses anonymized data from the HCSO Labor Force Survey. The author finds that the relative bias associated with data swapping might be held under 10 percent, and the bias can be minimized by selecting the pairs to be swapped randomly and by manipulating several explanatory variables simultaneously. The results are interpreted within the framework of measurement errors.

A bírósági ügyek munkaidőigényének mérése*

Dr. Örkényi László

bíró, az Országos Bírósági
Hivatal osztályvezetője

E-mail: orkenyil@obh.birosag.hu

A bírói munkateher mérése új törvényi kötelezettség, ám ennek hazánkban nincs kiértelt módszertana. Tudjuk, hogy a bírói munkatehernek fontos tényezője az érkező ügyek száma, de valószínű, hogy nagy jelentősége van az ügyek közötti – a befejezésükhöz szükséges munkaidő tekintetében bizonyosan meglévő – különbségeknek is. Jelenleg azonban nem ismerjük pontosan a szükséges munkaidő mennyiségét, nem tudjuk, hogyan lehet ezt mérni, ismeretlen a különböző munkaidő-igényű ügyek eloszlása, és nem tudjuk, hogy a szükséges munkaidő az ügyek milyen más jellemzőivel, milyen jellegű és erősségű összefüggésbe hozható. A tanulmány alapját képező kutatás célja ezeknek a kérdéseknek a megválaszolása, valamint a várható munkaidőigényre vonatkozó, az ügyek kezdőiratainak bizonyos jellemzőiből kiinduló predikció lehetőségének meghatározása volt befejezett bírósági ügyek iratainak retrospektív vizsgálata során nyert adatokból kiindulva, azok statisztikai módszerekkel való elemzése útján.

TÁRGYSZÓ:

Bírósági statisztika.

Bírósági eljárás.

Bírói munkateher.

* Ezúton is köszönetet nyilvánítok *dr. Handó Tiendének*, az OBH elnökének, aki engedélyezte a kutatás alapját képező iratvizsgálatot, a vizsgálat megszervezésében és lebonyolításában közreműködő OBH Munkateher Munkacsoport tagjainak, minden együttműködő, a munkát szakmai tanácsaikkal és észrevételeikkel segítő bírósági vezetőnek és bírónak, valamint az adatrögzítést végző bírósági titkároknak.

Az Országos Bírósági Hivatal (OBH) elnökének a bíróságok szervezetéről és igazgatásáról szóló 2012. évi CLXI. törvény (Bszj) 76. §-ának (4) bekezdés a) pontján alapuló kötelezettsége, hogy meghatározza a bírói munkateher mérésére szolgáló módszereket, valamint a peres és nemperes eljárások átlagos országos munkaterhét ügyszakonként és (ítélkezési) szintenként. Mindez nem önmagáért való, hanem a bírói munkateher egyenletes elosztását szolgálja.¹ Ez utóbbi azért fontos, mert ismert (Örkényi [2010], [2012]), hogy a – pusztán az érkezett ügyek darabszáma alapján mért – munkateher hosszabb időn át fennálló egyenlőtlen elosztása drámai különbségeket okoz az egyes szervezeti egységek között az eljárások időszerúsége tekintetében, az érkezett ügyek aránytalan eloszlása lényegében az elsődleges oka az időszerrütlenségnek.² Ez pedig közvetlenül és rendkívül hátrányosan érinti a túlterhelt bíróságokon pereskedni kénytelen jogkereső állampolgárokat, és természetesen az ott dolgozó bírákat is. Össztársadalmi érdek tehát a bírói munkateher egyenletes elosztása.³

A bírói munkateher fogalmát a törvény nem definiálja, azon azonban az egyes bírókra időegység alatt érkező (kiosztott) ügyek eljárási szabályok szerinti, gyors befejezéséhez szükséges összes érdemi munkavégzés idejének összegét értjük (Örkényi [2012]). A munkateher tehát lényegében az érkezett ügyek számától és azok egyenkénti időigényétől (összetettségétől, komplexitásától, bonyolultságától, „súlyától”) függ. Ehhez képest országosan vagy egy szervezeti egységben az átlagos munkateher az oda időegység alatt érkezett összes ügy jelentette munkateher (a befejezésükhöz szükséges érdemi munkaidő összege) és az ott lévő bírói létszám hányadosa, azaz az egységnyi létszámra jutó munkateher.

A bírósági statisztikai rendszer azonban hagyományosan „darab-darab” szemléletű, alapvetően az ügyforgalmi adatokat (így például az érkezett ügyek számát) gyűjti,

¹ Ennek érdekében a Bszj. 76. §-ának (4) bekezdés a) pontja azt is kimondja, hogy az OBH elnökének a peres és nemperes eljárások átlagos országos munkaterhének mutatói alapján meg kell határoznia a bíróságok szükséges – a törvényszékek esetében az illetékességi területükön működő közigazgatási és munkaügyi bírósággal, valamint járásbíróságokkal együttes – bírói létszámát. A jogalkotó egyéb kiegészítő eszközöket is biztosít az arányosság megvalósítása érdekében.

² A bíróságok működése egyensúlyi állapot esetén (az időegység alatt érkezett ügyek száma egyenlő a befejezett ügyek számával, a folyamatos ügyek száma állandó) is instabil, ha egy szervezeti egység ebből az állapottól jelentősebb mértékben kilendül (például megnő az érkezett ügyek száma), egy bizonyos határon túl saját erőből (változatlan létszámmal) nem képes az eredeti egyensúlyi állapotba visszatérni.

³ A Bszj. indoklása alapján úgy tűnik, hogy immár a jogalkotó is úgy véli, a bíróságok működésének időszerúségében mutatkozó különbségek oka a munkateher egyenlőtlen eloszlása. Ez a triviálisnak tűnő felismerés fontos paradigmaváltást jelent, mert korábban még a bíróságok központi igazgatásán belül is az volt az uralkodó álláspont, hogy az egyes szervezeti egységeknél jelentkező időszerrütlenség az ott dolgozó bírák és igazgatási vezetők tevékenységének hiányosságaira vezethető vissza.

az egyes ügyek munkaidőigényére vonatkozóan jelenleg semmilyen adatunk nincs. Ezért a munkatehermérést jelenleg kizárólag az érkezett ügyek számára alapíthatjuk, ami viszont csak akkor vezethet helyes eredményre, ha az alapját képező feltevések legalább egyike igaz: minden ügy befejezéséhez szükséges munkaidő mennyisége azonos (átlagos), vagy ennek hiányában, a különféle munkaigényű ügyek eloszlása minden tekintetben (földrajzilag, ítélkezési szinttől, ügyszaktól függetlenül) egyenletes. Tapasztalati tények alapján azonban erősen valószínű, hogy ezek közül egyik sem igaz: ügyszakokon belül is lennie kell különbségnek az egyes ügyek között, csakúgy, mint a különböző ügyszakokba tartozók esetén, hiszen jelentősen eltér az egyes ügyszakokban időegység alatt befejezett ügyek száma.⁴ Elképzelhetők különbségek ügyszakon belül, földrajzi alapon is.⁵ Ezeket a kétségeket minden vonatkozásban megerősítik az Országos Igazságszolgáltatási Tanács (OIT) korábbi vizsgálatai.⁶ Márpedig, ha az egyes ügyek között a várakozásoknak megfelelően van különbség, és a különböző munkaidő-igényű ügyek eloszlása is egyenlőtlen, akkor a munkatehermérés körében az ügyek darabszáma mellett azok időigénye tekintetében meglévő különbségeket is figyelembe kell venni.

Azt sem tudjuk azonban jelenleg biztosan, hogy milyen tényezők befolyásolják az egyes ügyek munkaidőigényét. Ez pedig azért fontos, mert ha a munkaidőigény mint függő változó összefüggésbe hozható az ügyek kezdőiratának bizonyos jellemzőivel mint független változókkal, akkor esetleg felépíthető egy olyan predikciós modell, amely segítségével az ügy várható munkaidőigénye már előre, az ügy érkezésekor kielégítő biztonsággal megbecsülhető. Ha ez lehetséges, akkor olyan munkatehermérési rendszert kell kiépíteni (*Örkényi* [2012]), amelynek keretében az ügy érkezésekor a későbbi munkaidőigény szempontjából releváns kezdőirati jellemzők is rögzítésre kerülnek, és a rögzített adatokból egy szoftver az előzetesen felépített predikciós modellből származtatott algoritmus segítségével becslést ad a várható munkaidőigény mennyiségére, majd ez alapján egy súlyszámot kapcsol minden ügy

⁴ Az egy tárgyaló tanács által egy hónap alatt befejezett peres ügyek száma az egyes ügyszakokban 2002 és 2011 között átlagosan a következő volt. Helyi polgári: 25,37, gazdasági: 21,77, munkaügyi: 29,25, büntető: 17,68; törvényszéki elsőfokú polgári: 16,55, gazdasági: 12,17, közigazgatási: 22,25, büntető: 2,76; törvényszéki másodfokú polgári: 38,93, gazdasági: 32,91, munkaügyi: 36,4, büntető: 32,81.

⁵ Ezt valószínűsítik a 2011. július 13. napján hatályba lépett, az egyes eljárási és igazságszolgáltatást érintő egyéb törvények módosításáról szóló 2011. évi LXXXIX. törvény által kiemelt jelentőségű üggyé minősített, egyes súlyosabb bűncselekmények tárgyában folyó, a büntető és a 400 millió Ft-ot meghaladó pertárgyértékű polgári perek 2011. második félévi érkezési adatai: a törvényszékeken a kiemelt ügyek aránya az összes elsőfokú peres érkezésen belül 1,3 és 6,56 százalék között ingadozik. Abból a megalapozottnak tűnő feltételezésből kiindulva, hogy a kiemelt ügyek munkaigénye általában nagyobb, mint egy átlagos ügyé, és az aránytalanság nemcsak a kiemelt, hanem a kiemeltnek nem minősülő, de az átlagot meghaladó munkaigényű ügyek esetében is fennáll, levonhatjuk azt a következtetést, hogy az egyes törvényszékek között az érkezett ügyek számának azonossága esetén is jelentős különbségek lehetnek az összes munkaigény tekintetében.

⁶ A vizsgálatok a szervezeti egységek közötti esetleges munkateher-különbségek kimutatását célozták, de az alkalmazott pontozásos módszer miatt csak relatív összehasonlítást tettek lehetővé. Bár ezek a különbségek meglétét igazolták, az eredményeiket végül nem használták fel.

höz.⁷ Végül soron a munkatehermérés rendszerében az ügyek darabszámának funkcióját ez a súlyszám vehetné át. Ennek segítségével az egyes szervezeti egységek munkaterhének esetleges változását már előre észlelni lehetne, és ezáltal időben megtehetőek a szükséges intézkedések.

1. A kérdések

Mielőtt elkezdjük kiépíteni a munkatehermérés törvény által megkívánt rendszert, mindenképpen válaszolnunk kell a következő kérdésekre. Van-e különbség a bírósági ügyek között a befejezésükhöz szükséges érdemi munkaidő mennyisége tekintetében? Ha van ilyen különbség, akkor egyenletes-e a különböző munkaigényű ügyek eloszlása? Ha nem egyenletes az eloszlás, akkor összefüggésben van-e az ügyek munkaidőigénye valamely kezdőirati jellemzővel? Felépíthető-e egy megbízható predikciós modell, amelynek segítségével az érkező ügyek várható munkaidőigénye már előre, megfelelő biztonsággal megbecsülhető?

2. A módszer

Az 1. pontban szereplő bármelyik kérdés megválaszolásának alapfeltétele, hogy minél nagyobb számú ügy esetében ismerjük pontosan a befejezéshez szükséges, érdemi munkaidő-ráfordítást. Ilyen adatbázis nincs, ezért az első lépésben a felmerült munkaidőigényt kell meghatározni. Ennek egyik lehetséges és általunk is választott módja, hogy azt utólag, a befejezett ügyek iratainak bizonyos mérhető jellemzőiből

⁷ A vázolt munkateher-mérési rendszer ebben az elemében lényegében egy becslési eljárás, amely – hasonlóan minden más, jövőre vonatkozó prognózis készítését célzó eljáráshoz – annál megbízhatóbb, minél szélesebb múltbéli adatbázison alapul, és minél inkább igénybe veszi a modern matematika és statisztika eszköztárát.

Egyébként minden, jelenleg általam ismert, igényesebb külföldi bírói munkateher-mérési rendszer (például Németországé, Hollandiáé) is az ügy érkezésekor, az akkor ismert információk alapján próbálja előre becsülni a várható munkaigényt. Tehát sohasem utólag, bizonyos paraméterek rendszeres mérése alapján, természetesen határozzák meg az egyes ügyekben elvégzett munkát, hanem azt egy előre rögzített módszerrel, az eljárás kezdetén jósolják meg több-kevesebb biztonsággal. A becslési eljárások között elvi szinten csak annyi a különbség, hogy az ügy indulásakor azok a rendelkezésre álló információk milyen széles körére támaszkodnak.

Megjegyzem, egy szervezeti egységen belül – automatikus szignálás hiányában – a szignálást végző igazgatói vezető most is nap mint nap ugyanezt teszi: a kezdőiratot áttekintve megkísérli megítélni az adott ügy tárgyát, illetve várható összetettségét, és értékítélete alapján az egyes bírák leterheltségét, képességét figyelembe véve dönt, hogy azt kire ossza.

kiindulva tegyük. Ezzel gyorsan, standardizált módszerekkel és körülmények között, a bírák további igénybevétele nélkül juthatunk adatokhoz. Kétségtelen, hogy az eljárás számos lépésében becslést tartalmaz, és ezzel puha adatokat kapunk, a társadalmi jelenségekkel kapcsolatos statisztikai munka gyakorlatában azonban nem egyedülálló, hogy a közvetlen megfigyelés helyett csupán az adatokkal kapcsolatba hozható egyéb változók mérésére van lehetőség. A jelen esetben alkalmazott becslések emellett javarészt hosszú évek tapasztalatain, széles körű szakmai egyeztetést követően létrejött szakmai konszenzuson alapulnak.⁸ Az első feladat tehát a munkaidőre vonatkozó becslés módszertanának meghatározása volt.

Tapasztalati tény, hogy az első fokú peres büntetőügyek befejezéséhez szükséges összes munkaidő négy tételből áll össze.

1. A tárgyalások (és egyéb, felek részvételével zajló eljárási cselekmények, így tanácsülés, előkészítő ülés, személyes meghallgatás, tanúmeghallgatás) időtartama a tárgyalóteremben töltött időt jelenti. Ez utólag is, óra/perc pontossággal meghatározható az eljárási cselekményekről készült jegyzőkönyvekben rögzített kezdő és befejező időpontokból kiindulva, és a tárgyalások időtartamát összegezve.

2. A tárgyalásra való felkészülés ideje, ami lényegében a nyomozati iratok, azok mellékleteinek és a vádirat tanulmányozását („megtanulását”), illetve az alkalmazandó joganyag áttekintését jelenti a tárgyalás előtt, tehát legbiztosabban a nyomozati iratok oldalban mért terjedelméből kiindulva becsülhető meg (oldalszám \times olvasási idő/oldal). Szakmai tapasztalatok szerint a nyomozati iratokhoz esetenként tartozó számozatlan mellékletek egyharmad részét kell áttanulmányozni (az ehhez szükséges idő = melléklet cm-ben mért terjedelme \times 0,33 \times (a cm-ben mért terjedelem és az oldalszám közötti átváltást biztosító) konstans \times olvasási idő/oldal). A két részüdőt össze kell adni. Azokban az ügyekben, ahol nem tartottak tárgyalást, a felkészülési idő nulla.⁹

3. A fellebbezhető határozatok szerkesztésének időtartama a döntés meghozatalának, megszövegezésének, leírásának, lediktálásának, javí-

⁸ Egy másik lehetőség az lett volna, ha egy megfelelő hosszúságú jövőbeli időintervallumon (például egy hónapon, éven) belül érkező ügyek (vagy az azokból képzett reprezentatív minta) esetében az eljáró bírák azok befejezéséig pontosan (órával) mérik, hogy velük kapcsolatban milyen tevékenység folytán, mekkora munkaidő-ráfordítás merült fel. Bár ezzel a módszerrel kaphattuk volna a legpontosabb eredményeket, használata ellen szólt egyebek mellett bizonyos ügyszakok eljárásainak hosszú időtartama. Ezeket mérlegelve a másik megoldást választottuk, azzal hogy ez utóbbi lehetőség alkalmazását az erőforrások rendelkezésre állása esetén, a rövidebb (néhány hónap) eljárási időtartamokkal jellemezhető (nemperes ügyek, felülbírálati szint stb.) vagy jellemzően nagy tömegű iratokkal dolgozó (törvényszéki büntető első fok) ügyszakokban, illetve a becsült adatok kontrollja érdekében nem vetjük el.

⁹ Elképzelhető, hogy a felkészülési idő növekedése nem lineárisan követi a nyomozati iratok terjedelmét.

tásának időigénye. Mennyisége a határozatok terjedelme alapján becsülhető (határozatok összes oldalszáma*időigény/oldal).¹⁰

4. A kurrenciális tevékenység időszükségletének mérése a legnehezebb, mert ebbe beletartozik minden, az előző pontokban nem nevesített időráfordítás. Ennek időszükséglete azonban a beadványok terjedelmével van a legszorosabb összefüggésben, mert először el kell olvasni azokat. Mivel a beadványok teszik ki a bírósági iratok tömegének túlnyomó részét (a határozatok, jegyzőkönyvek terjedelme ehhez képest nem számottevő), így az időszükséglet az irat terjedelmének egyszerű (cm-ben való) mérésével jól megbecsülhető. Ezt az értéket kettővel szorozzuk meg, mert a beadványokat nemcsak elolvasni kell, hanem általában intézkedést is kell tenni velük kapcsolatban (kurrenciális munkaidő = $2 \times$ a bírósági irat cm-ben mért terjedelme \times konstans \times olvasási idő/oldal).

A számítások során alkalmazott szorzószámokat (olvasási sebesség/oldal, és határozatszerkesztés ideje/oldal) a bírák között, 2012 márciusában végzett, nem reprezentatív, kérdőíves vizsgálat (12196/2012. OBH) alapján határoztuk meg, tehát szakmai konszenzuson alapulnak, melynek eredményei szerint a határozatszerkesztés időigénye helyi büntető ügyszakban 38,79 perc/oldal, az olvasási sebesség 76,03 perc/50 lap.¹¹ A megadott 50 lap körülbelül 10 százalékban tartalmaz kétoldalas lapokat is (tapasztalatok szerint bírósági iratok esetében ez a jellemző), tehát az 50 lap 55 oldalnak felel meg. Az iratok cm-ben mért vastagságának oldalszámra való átváltását biztosító konstans értékét tapasztalati úton állapítottuk meg: 1 cm vastagságú bírósági iratketeg körülbelül 86 lapot tartalmaz, ami a 10 százalékos kétoldalaslaparányt figyelembe véve nagyjából 95 oldalt tesz ki.

A szakma szerint a helyi büntető peres ügyek munkaidő-mennyiségével mint függő változóval összefüggésbe hozható kezdőirati jellemzők mint független változók a következők: a nyomozati iratok mennyisége és minősége, a vádirat terjedelme, a bírósági eljárás során alkalmazandó eljárásjogi szabályok, az eljárásban szereplő személyek száma és egyéb ismérvei, a vád tárgyává tett cselekmények száma és jellege. A munkaidő-kalkuláció alapját képező és a kezdőirati jellemzők rögzítésére szolgáló, a vizsgálat során használt adatlap mintáját a tanulmány Mellékletének M1.

¹⁰ Az összefüggés linearitása itt is kérdéses lehet, mivel például egy tízszer nagyobb terjedelmű határozat általában magasabb szintű szintetizálási folyamat végeredménye, amely belső komplexitása és összetettsége folytán nem tízszer annyi, hanem valószínűleg ennél több időt igényel.

¹¹ A vizsgálat során elektronikus úton kérdőíveket juttattunk el valamennyi bíró részére, amelyeket önkéntes alapon tölthettek ki. Összesen 590 kitöltött adatlap érkezett vissza (a teljes bírói létszám 21,53 százaléka választott), ebből 179 adatlap helyi büntető ügyszakban dolgozó bíraktól (az ügyszakok közül itt volt a legnagyobb érdeklődés).

táblázata tartalmazza, ami a *Statisztikai Szemle* honlapján (<http://www.ksh.hu/statszemle>) található.¹²

A munkamennyiség meghatározását, valamint a munkaidő és a kezdőirati jellemzők közötti összefüggés feltárását célzó empirikus vizsgálat tárgyául (alapsokaság) a 2010-ben befejezett helyi első fokú peres büntetőügyeket választottuk,¹³ logisztikai szempontokra tekintettel pedig a Fővárosi Törvényszék, a Balassagyarmati Törvényszék és a Miskolci Törvényszék helyi bíróságait vontuk be.

Ténylegesen a Fővárosi Törvényszék területéről 232 (hiányzik 21), a Miskolci Törvényszék területéről 228 (hiányzik 3), a Balassagyarmati Törvényszék területéről 278 (hiányzik 14), összesen tehát 738 (hiányzik 38) véletlenszerű mintavétellel választott ügyet vizsgáltunk meg, és töltöttünk ki róluk adatlapot.

Első lépésként a már ismertetett módon minden egyes ügy esetében kiszámítottam a munkaidő mennyiségét. Ezt korrigáltam az újraindulások számával (munkaidő/(1+újraindulások száma)). Erre azért volt szükség, mert az iratvizsgálat során azoknak az ügyeknek az esetében, amelyek előzményeikkel, illetve befejezett utóirataikkal együtt érkeztek, nemcsak a mintába került ügyek iratait vizsgáltam meg, hanem az előzményeket és az utóiratokat is. Az adatokat MS Excel és Statistica programokkal elemeztem.

3. Eredmények

A számítások eredményeképpen adódó munkaidő-mennyiségek leíró statisztikáit az egész minta vonatkozásában az 1. táblázatban foglaltam össze.

Először is fontos annak a kérdésnek a megválaszolása, hogy mennyire megbízhatók a kapott munkaidőadatok, különös tekintettel arra, hogy azok jelentős részben utólagos becslésen alapulnak. Azt természetesen semmilyen módon nem tudjuk ellenőrizni, hogy az egyes ügyekben mennyire adott pontos eredményt a kidolgozott becslési módszer, mert a felmerült munkaidő vonatkozásában tényadatok nincsenek. Ha azonban a korrigált átlagos munkaidőadatot (10,54 óra)¹⁴ és a helyi büntető ügyszakban 2010-ben egy

¹² A nevesített körülményeken túl számos más kezdőirati jellemző fontossága is felmerült, de ezeket a szakma többsége nem támogatta. Hasonló szempontok alapján történt az ügyek pontozása is a már említett OIT-vizsgálat során.

¹³ Az eredeti terveknek megfelelően az adatrögzítésbe a törvényszéki első fokú ügyeket is bevontuk, de ezek helyi ügyekkel együttes statisztikai elemzésére eltérő jellemzőik miatt nem volt lehetőség. Számuk viszont alacsony volt egy külön statisztikai vizsgálathoz. Mégis, bár a tanulmány csak a helyi büntető peres ügyekre vonatkozik, a kidolgozott módszertan valamennyi bírósági ügyszakra adaptálható.

¹⁴ A korrekció azért szükséges, mert a ténylegesen megvizsgált ügyek törvényszékek szerinti összetétele az objektív okok következtében nem megvizsgálható ügyek miatt eltér a minta eredetileg tervezett és országosan is reprezentatív tekintett összetételétől. Így az összes megvizsgált ügy munkaidőigényének átlaga nem feltét-

tárgyaló tanács által havonta befejezett ügyek számának országos átlagát (15,4)¹⁵ összeszorozzuk, akkor havi 162,31 órát kapunk, ami napi 7,38 órának (162,31/havi körülbelül 22 munkanap) felel meg. Ez összhangban áll a havi 176 (közelítőleg 22 munkanap × napi 8 óra munkaidő) és a napi 8 óra törvényes munkaidővel.¹⁶ Levonható tehát a következtetés, hogy a munkateher mérésére kidolgozott, az eljárás során keletkezett iratokra épülő utólagos becslésen alapuló módszer alkalmas a bírósági ügyek munkaidőigényének nemcsak relatív, hanem abszolút értelemben való meghatározására is.¹⁷

1. táblázat

A munkaidőadatok leíró statisztikái

Ismérv	Érték
Elemszám	738
Átlag (óra)	10,5279
Konfidencia-intervallum -95% (óra)	9,4937
Konfidencia-intervallum +95% (óra)	11,5621
Medián (óra)	6,4640
Minimum (óra)	0,0219
Maximum (óra)	231,1443
Alsó kvartilis (óra)	2,9066
Felső kvartilis (óra)	12,8627
Szórás (óra)	14,3115
Átlag standard hibája (óra)	0,5268
Ferdeség	6,5537
Csúcsosság	80,9821

A kapott leíró statisztikai adatok alapján megállapítható, hogy a legkisebb érték 0,022, a legnagyobb 231,14 óra, a terjedelem 231,118 óra, a maximális és a minimá-

lenül lesz egyenlő az országos átlaggal, ehhez az egyes törvényszékek átlagos munkaidőadatából (Fővárosi Törvényszék: 11,4 óra, Miskolci Törvényszék: 10,5 óra; Balassagyarmati Törvényszék: 9,83 óra) az eredetileg tervezett mintaelemszámokkal (Fővárosi Törvényszék: 253; Miskolci Törvényszék: 231; Balassagyarmati Törvényszék: 292; összesen 776) súlyozott átlagot (10,54 óra) kell számítani. Az eltérés a három törvényszék átlagához (10,53 óra) képest minimális. Ismételten hangsúlyozom, hogy az országos átlagra vonatkozó adatot – a minta országosan bizonytalan reprezentativitására tekintettel – óvatosan kell kezelni.

¹⁵ A 2010-ben befejezett összes helyi büntető peres ügy (77 699) /a tárgyaló tanácsok összes száma helyi büntető peres ügyekben (420,51) /12=15,4.

¹⁶ Az eredmények értékelésekor azt is figyelembe kell venni, hogy a helyi büntető bírák általában nemcsak peres, hanem büntető nemperes, szabálysértési, idegenrendészeti, esetleg további más ügyeket is intéznek.

¹⁷ Amennyiben az eredmények abszolút értékét a további vizsgálatok vagy a munkaidőigény prospektív mérésének eredményei cáfolják (azok magasabb vagy alacsonyabb átlagos munkaidőértékeket adnak), úgy egy korrekciós tényező beiktatásával az itt kapott adatok érvényessége fenntartható.

lis érték hányadosa 10 506. A minimális érték (1,31 perc)¹⁸ szakmai szempontok alapján kissé kevésnek tűnik, de elvileg nem zárható ki; fontos megjegyezni, hogy a 738 ügyből mindössze 8-nál merült fel a szakmailag elfogadhatóbb, 20 percnél rövidebb munkaidőigény.¹⁹ Az adatok teljes terjedelmen belüli eloszlását ábrázolja az M1. ábra. Ebből látható, hogy egy maximális munkaidő-igényű ügy²⁰ ugyan erősen kilóg a többi közül (az utána következő ügy munkaidőigénye már csak 92,99 óra volt), a 80–90 munkaóra-időigény már nem tekinthető rendkívülinek. Ha a legnagyobb és a 20 percnél²¹ rövidebb munkaidő-igényű ügyeket figyelmen kívül hagyjuk, a terjedelem még akkor is 92,64 óra (92,99–0,35), a maximális és a minimális érték hányadosa pedig 265,69.

Az 1. táblázatból az is leolvasható, hogy az ügyek 50 százalékában a munkaidő-igény a 2,9 és 12,86 óra értékeken kívül esett, a szórás 14,31 óra. Az átlagos munkaidőigény 10,53 óra, átlagos ügy azonban szinte alig van: csak 105 (az esetek 14,22 százaléka) esik a 8,42–12,63 (átlag \pm 20 százalék) közötti sávba. Az adatokból levonható tehát a következtetés, egy úgyszakon belül is igen jelentős különbség van az ügyek között a munkaidőigény tekintetében, még a szakmailag elfogadható és az igazán kiugró értéktől mentes tartományon (20 perc–92,99 óra) belül is. A megválaszolni kívánt első kérdésre tehát igen a válasz.

A leíró statisztikák alapján megállapítható, hogy az eloszlás erősen jobbra ferdült, és egyúttal csúcsosabb is a normálisnál. (Lásd az 1. táblázatot.) Ehhez képest nem meglepő, hogy az adatok ábrázolását követően elvégzett számítások is azt igazolják, a bírósági ügyek munkaidőigénye nem mutat normál eloszlást. A Kolmogorov–Smirnov-teszt eredménye: d (Kolmogorov–Smirnov-távolság) = 0,2314, p (valószínűség) < 0,01; a Lilliefors-teszté: p < 0,01; a Shapiro–Wilk-teszté: W = 0,5715, p = 0,0000 volt.

Erre tekintettel a későbbi hipotézisvizsgálatok során a paraméteres eljárások nem voltak alkalmazhatók. Fontos azonban, hogy az eloszlás az adatok egyszerű logaritmikus transzformálásával (log(újraindulással korrigált munkaidő+2)) közelítőleg normalizálható. (A Kolmogorov–Smirnov-teszt eredménye d = 0,03708, p > 0,2; a Lilliefors-teszté p < 0,5; a Shapiro–Wilk-teszté W = 0,98227, p = 0,0000).

¹⁸ A minimális munkaidő-igényű ügy egy 1. sorszám alatt befejeződött magánvadás ügy volt, amiben az ismeretlen elkövető kilétének felderítése érdekében nyomozást rendeltek el.

¹⁹ Ha a munkaidő mennyiségének meghatározniánk egy minimális értéket (például 20 percet), az erre vonatkozó szakmai aggodalom elosztható lenne.

²⁰ A maximális munkaidőigény egy, a Budapesti IV. és XV. Kerületi Bíróság előtt folyamatban volt ügyben merült fel. Ebben a nyomozati irat terjedelme 1 339, a mellékleteké 650, a vádiraté 7,4 oldal volt (az átlag 2,58 oldal). A három vádlott ellen 19 rendbeli cselekmény miatt emeltek vádat, 27 tanú kihallgatását indítványozták; 38 óra 50 percnyi tárgyalást tartottak, 12,3 oldal fellebbezhető végzést hoztak (átlag: 0,77 oldal), és 53 oldalas ítélet született (átlag: 3,17 oldal), 206. sorszám alatt (átlag: 12. sorszám).

²¹ Pontosan 0,3451 óra = 20,7 perc.

2. táblázat

A munkaidőadatok leíró statisztikái az egyes városi bíróságokon

Szervezeti egység	Elem- szám	Átlag (óra)	Konfi- dencia- inter- vallum -95% (óra)	Konfi- dencia- inter- vallum +95% (óra)	Medián (óra)	Mini- mum (óra)	Maxi- mum (óra)	Alsó kvar- tilis (óra)	Felső kvar- tilis (óra)	Szórás (óra)	Átlag standard hibéje (óra)	Ferde- ség	Csu- csosság
Sátoraljnújehelyi Városi Bíróság	17	11,3814	4,2052	18,5576	6,3182	1,2608	56,9165	3,0542	11,4482	13,9574	3,3852	2,4758	6,9554
Ózdi Városi Bíróság	22	6,3907	3,5609	9,2205	4,2454	1,0841	28,3225	2,2542	7,5216	6,3824	1,3607	2,3306	6,1257
Szerencsi Városi Bíróság	13	6,3832	4,3387	8,4277	6,2253	0,8345	12,4442	5,1358	8,8271	3,3833	0,9384	0,0569	-0,2552
Székesúti Városi Bíróság	7	15,8837	0,7532	31,0142	8,1104	1,7503	43,5773	5,0633	34,7378	16,3601	6,1835	1,1989	-0,3221
Encsi Városi Bíróság	12	5,7940	2,0428	9,5451	4,1057	0,8752	22,5136	2,1372	6,9630	5,9039	1,7043	2,3346	6,3278
Kazincbarcikai Városi Bíróság	30	9,4608	5,3176	13,6040	5,9392	1,3805	54,1954	3,5021	10,5542	11,0957	2,0258	3,0332	10,0802
Tiszaujvárosi Városi Bíróság	18	5,7999	3,0717	8,5281	4,1117	0,7045	18,6588	2,1412	7,3492	5,4861	1,2931	1,4499	1,0213
Mezőkövesdi Városi Bíróság	9	12,9794	2,6120	23,3468	10,2300	0,5022	46,1591	7,4224	13,3427	13,4874	4,4958	2,1562	5,5726
Miskolci Városi Bíróság	100	12,9096	10,0089	15,8102	8,4959	0,0438	92,9890	4,2595	17,1477	14,6184	1,4618	3,1480	13,1113
Salgótarjáni Városi Bíróság	114	9,6178	7,4731	11,7625	6,0906	0,0438	84,8051	3,5479	11,2526	11,5583	1,0825	3,4594	17,0652
Balassagyarmati Városi Bíróság	140	10,2252	8,2836	12,1648	7,3788	0,0875	79,8467	2,7313	13,1159	11,6072	0,9810	2,8597	11,5436
Pásztói Városi Bíróság	24	8,4970	3,5454	13,4487	5,7291	0,0438	57,5372	2,3284	8,8281	11,7264	2,3936	3,4548	14,0105
Budapesti II. és III. Kerületi Bíróság	19	13,2632	7,3289	19,1975	8,4814	1,3277	44,1877	4,4397	18,3782	12,3122	2,8246	1,4022	1,1840
Budai Központi Kerületi Bíróság	18	8,6796	4,6581	12,7011	6,2827	0,0438	26,8881	3,7288	12,5005	8,0869	1,9061	1,1381	0,4109
Budapesti XVIII. és XIX. Kerületi Bíróság	11	7,9613	2,6093	13,3133	5,9799	0,3451	27,4502	1,4262	9,9417	7,9666	2,4020	1,6858	3,0425
Budapesti IV. és XV. Kerületi Bíróság	21	19,0585	-3,4650	41,5820	5,9680	0,4376	231,1443	1,7503	12,6717	49,4810	10,7976	4,3278	19,2998
Budapesti XX., XXI. és XXIII. Kerületi Bíróság	19	5,6486	1,7240	9,5732	2,7051	0,0438	27,7697	0,7635	6,2475	8,1426	1,8680	2,1474	3,9459
Pesti Központi Kerületi Bíróság	144	11,3928	9,3578	13,4279	7,2579	0,0219	81,0327	2,8124	17,1662	12,3543	1,0295	2,2673	7,6849

3. táblázat

A munkaidőadatok leíró statisztikái a székhelyi és nem székhelyi városi bíróságokon

Szervezeti egység	Elem- szám	Átlag (óra)	Konfi- dencia- inter- vallum -95% (óra)	Konfi- dencia- inter- vallum +95% (óra)	Medián (óra)	Mini- mum (óra)	Maxi- mum (óra)	Alsó kvar- tilis (óra)	Felső kvar- tilis (óra)	Szórás (óra)	Átlag standard hibája (óra)	Ferde- ség	Csú- csosság
Három székhelyi bíróság	384	11,3621	10,0842	12,6401	7,5926	0,0219	92,9890	3,1577	15,0516	12,7365	0,6500	2,8027	11,3117
Összes nem székhelyi bíróság	354	9,6230	7,9701	11,2758	5,7186	0,0438	231,1443	2,7540	10,9526	15,8126	0,8404	8,6451	110,7650
Miskolci Városi Bíróság	100	12,9096	10,0089	15,8102	8,4959	0,0438	92,9890	4,2595	17,1477	14,6184	1,4618	3,1480	13,1113
Többi Borsod-Abaúj-Zemplén megyei helyi bíróság	128	8,6157	6,8639	10,3676	5,1996	0,5022	56,9165	2,9918	10,0717	10,0161	0,8853	2,8638	9,2572
Balassagyarmati Városi Bíróság	140	10,2232	8,2856	12,1648	7,3788	0,0875	79,8467	2,7313	13,1159	11,6072	0,9810	2,8597	11,5436
Többi Nógrád megyei helyi bíróság	138	9,4229	7,4783	11,3675	6,0562	0,0438	84,8051	2,8147	10,9781	11,5524	0,9834	3,4132	15,9516
Pesti Központi Kerületi Bíróság	144	11,3928	9,3578	13,4279	7,2579	0,0219	81,0327	2,8124	17,1662	12,3543	1,0295	2,2673	7,6849
Többi fővárosi kerületi bíróság	88	11,4018	5,9868	16,8168	5,7987	0,0438	231,1443	1,5540	12,4490	25,5571	2,7244	7,5201	64,2798

4. táblázat

A munkaidőadatok leíró statisztikái az három törvényszéken

Szervezeti egység	Elem- szám	Átlag (óra)	Konfi- dencia- inter- vallum -95% (óra)	Konfi- dencia- inter- vallum +95% (óra)	Medián (óra)	Mini- mum (óra)	Maxi- mum (óra)	Alsó kvar- tilis (óra)	Felső kvar- tilis (óra)	Szórás (óra)	Átlag standard hibája (óra)	Ferde- ség	Csú- csosság
Miskolci Törvényszék	228	10,4990	8,8801	12,1178	6,4870	0,0438	92,9890	3,5493	12,8299	12,4051	0,8215	3,2487	14,4350
Balassagyarmati Törvényszék	278	9,8269	8,4614	11,1925	6,5081	0,0438	84,8051	2,7779	11,4664	11,5661	0,6937	3,1103	13,3735
Fővárosi Törvényszék	232	11,3962	9,0093	13,7831	6,4490	0,0219	231,1443	2,3005	15,3722	18,4521	1,2114	7,8378	87,7522

A következő kérdés, hogy az előzők szerint jelentősen különböző munkaidőigényű ügyek eloszlása az egyes szervezeti egységek között egyenletes-e. E körben először az egyes városi bíróságokat hasonlítottam össze, az erre vonatkozó leíró statisztikai adatokat a 2. táblázatban mutatom be. Az átlagokat összehasonlítva látható, hogy a maximális érték 19,06 óra (Budapesti IV. és XV. Kerületi Bíróság), a második legnagyobb érték 15,88 óra (Szikszói Városi Bíróság). Ha e két értéket figyelmen kívül hagyjuk (mivel az elsőt a maximális időigényű ügy torzítja felfelé, az utóbbi bíróság esetén pedig mindössze 7 ügy került a mintába, és ezeknél tapasztalhatók a legnagyobb szóráserkékek is), a következő legnagyobb adat (13,26 óra Budapesti II. és III. kerületi Bíróság) még így is 2,3-szerese a minimális értéknek (5,65 óra, Budapesti XX., XXI. és XXIII. Kerületi Bíróság).

A különbségek jól érzékelhetők az M2. ábrán. Az erre alapozott várakozásokat igazolta a Kruskal–Wallis-teszt is, szignifikáns különbséget mutatva (df (szabadságfok)= 17, $N = 738$, $H = 30,22246$, $p = 0,0248$) a bíróságok között. Azt a nullhipotézist tehát, hogy az egyes városi bíróságok között az ügyek munkaidőigénye szempontjából nincs különbség, elvettem.

A szervezeti egységek közötti eltérések másik lehetséges dimenziója az a különbség a székhelyi és a többi városi bíróság között, ami a hozzájuk érkező ügyek eltérő összetételére vezethető vissza.²² A várakozásokat igazolták a 3. táblázatban közölt eredmények, ugyanis a vizsgálatba vont három székhelyi bíróság előtti ügyek munkaidőátlaga 11,36 óra volt szemben a többi városi bíróság 9,62 órás átlagával. Ezt mutatja az átlagok 95 százalékos konfidencia-intervallumát ábrázoló M3. ábra is.

A különbségek valódiságával kapcsolatos várakozásokat igazolta a kérdés eldöntése érdekében végzett Mann–Whitney-féle U-teszt eredménye: $U = 59787,5$, $p = 0,0047$. (Lásd az M2. táblázatot.) Tehát a székhelyi és a többi bíróság közötti különbség hiányára vonatkozó nullhipotézist elvetve egyértelműen kijelenthető, hogy a helyi bíróságok kétféle csoportja között az előttük befejeződött ügyek munkaidőigénye tekintetében különbség van, az átlagok eltérése szignifikáns.

Ha törvényszékenként vizsgáljuk a három székhelyi városi bíróság és a többi városi bíróság adatait, akkor az látható, hogy csak a Miskolci Törvényszék esetében mutatkozik szembeötlő különbség. (Lásd a 3. táblázatot.) Az egyetlen kiugró értéket (19,06 óra, Budapesti IV. és XV. Kerületi Bíróság) leszámítva a Fővárosi Törvény-

²² A büntetőeljárásról szóló 1998. évi XIX. törvény (a továbbiakban Be.) 17. §-ának (5) és (6) bekezdése sorolja fel azokat a bűncselekményeket, amelyek ügyében a törvényszék székhelyén működő helyi bíróságok kizárólagosan illetékesek (jelenleg például közveszélyokozás (a büntető törvénykönyvről szóló 1978. évi IV. törvény (a továbbiakban Btk.) 259. §), közérdekű üzem működésének megzavarása (Btk. 260. §), visszaélés radioaktív anyaggal (Btk. 264. §), visszaélés nukleáris létesítmény üzemeltetésével (Btk. 264/A. §), visszaélés atomenergia alkalmazásával (Btk. 264/B. §), a gazdasági bűncselekmények (Btk. XVII. Fejezet) – kivéve a számvitel rendjének megsértését (Btk. 289. §), valamint a pénzügyi bűncselekményeket (Btk. XVII. Fejezet III. Cím); a kizárólagos illetékesség alá eső bűncselekmények köre az utóbbi időben szűkült, tehát jelenleg kisebb az eltérés a székhelyi és a többi helyi bírósághoz érkező ügyek összetétele között, mint 2010-ben volt).

széken szintén úgy tűnik, hogy számottevő eltérés van az ügyek munkaidőigénye között. A Balassagyarmati Törvényszéken viszont nincs szignifikáns különbség. A várakozásokat jól igazolták az elvégzett Mann–Whitney-féle U-teszt eredményei: a különbség csak a Miskolci Törvényszék esetén szignifikáns: $U = 4856$, $p = 0,0018$. (Lásd az M2. táblázatot.)

Végül a vizsgálatba vont törvényszékek területén befejezett ügyek átlagos munkaidőigényét vettem össze egymással. A leíró statisztikai adatokat a 4. táblázat tartalmazza, melyek alapján látható, hogy az ügyek átlagos munkaidőigénye a Fővárosi Törvényszék területén a legmagasabb, a Balassagyarmati Törvényszékén a legalacsonyabb, a Miskolci Törvényszékén pedig a három törvényszék átlagához közeli. Ezek az átlagok 6–8 százalékos eltérést mutatnak egymáshoz képest, a Fővárosi és a Balassagyarmati Törvényszékek közötti különbség 1,57 óra, azaz 14,91 százaléka a három törvényszék munkaidőátlagának. Kérdés azonban, hogy ezek az eltérések statisztikai értelemben valós különbségek-e.

Az M4. ábrán látható, hogy a törvényszékek átlagai egy kivétellel – a Fővárosi Törvényszék átlaga a Balassagyarmati Törvényszékhez képest – a többi törvényszék átlagának 95 százalékos konfidencia-intervallumán belül vannak. A három törvényszék munkaidőátlagai közötti szignifikáns különbség igazolására – a munkaidőadatok normálistól eltérő eloszlására tekintettel – a Kruskal–Wallis-tesztet végeztem el: $df = 2$, $N = 738$, $H = 0,6977558$, $p = 0,7055$, mely alapján a nullhipotézist megtartva kijelenthető, hogy a tapasztalt különbségeket csak a véletlen okozza, a törvényszékek átlagainak eltérése nem szignifikáns. A kimutatott és szakmai értelemben egyértelműen releváns különbség tényleges meglétét nagyobb mintán lehetne igazolni.²³

Mivel az előzők szerint bebizonyosodott, hogy az egyes ügyek munkaidőigénye között jelentős különbségek vannak, és a különböző munkaidő-igényű ügyek eloszlása bizonyos dimenziók szerint különböző, érdemes megvizsgálni, hogy mely tényezők befolyásolják az ügyek munkaidőigényét, és ez megfelelő biztonsággal prognosztizálható-e azokból. Az ennek megválaszolása érdekében végzett exploratív vizsgálatok a munkaidőigény kapcsolatának feltárását célozták a vizsgálatba bevont kezdőirati változókkal. Első lépésként kiszűrtem azokat a kezdőirati változókat, amelyek a vizsgált mintában nem, vagy csak egyszer fordultak elő. A többiek a mérési szinttől függően lehettek nominális (például megye, székhelyi/nem székhelyi bíróság), ordinális (például cselekmények rendűsége, kérdőív 40–80. kérdései) és skaláris (például vádirat oldalszáma), az általuk felvehető értékek számától függően pedig diszkrét vagy kvázi folytonos változók. A diszkrét binomiális változók egy része (például magán-/nem magánvadás eljárás) eleve csak két értéket vehetett fel (elvileg

²³ Korábbi vizsgálatok alapján tudjuk, hogy pusztán az ügyek darabszáma tekintetében fennálló ilyen mértékű, tartós különbségek drámai eltéréseket okoznak az egyes szervezeti egységek között az eljárások időszereiségében (Örkényi [2012]).

is binomiális változók), voltak azonban olyanok is, amelyek elvileg többet is felvehettek volna, de a vizsgált ügyekben csak kétféle értékük fordult elő (gyakorlatilag binomiális változók)²⁴. A binomiális változók harmadik csoportját azok a változók alkotják, amelyeket több – az alacsony esetszám miatt – önállóan nem vizsgálható binomiális változó összevonásával, vagy ordinális változók binomiálissá alakításával képeztem (képzett változók).

A vizsgálatba vont kezdőirati jellemzők közül a binomiális változóknak a munkaidő mennyiségére vonatkozó hatását szemléltető leíró statisztikák kivonatát az M3. táblázat tartalmazza. Az adatok alapján látható, hogy a vizsgált binomiális változók többsége jelentősen befolyásolja a munkaidőigényt. Ugyanakkor az is egyértelmű, hogy a számos esetben alacsony esetszám miatt a megfigyelt eltérések tényleges jelentősége bizonytalan. Ez jól látszik az egyes binomiális változók különböző értékei esetén megfigyelhető munkaidőértékek átlagainak box plot diagramjai alapján is, mivel a konfidencia-intervallumok erősen fedték egymást.²⁵ (Lásd az M5–M17. ábrákat.)

Az egyes csoportok között észlelt különbségek vonatkozásában elvégzett hipotézisvizsgálatok (Mann–Whitney-teszt) adatait az M4. táblázat tartalmazza. Az adatokból látható, hogy a kérdőívre felvett 25 elvileg vagy gyakorlatilag binomiális változó közül 10 esetében van jelentős és szignifikáns különbség a munkaidő mennyiségében. További két változó esetén tapasztalható jelentős, a szignifikancia határán álló különbség. A fennmaradó 13 változó közül 8 esetében az összminta átlagához (10,53 óra) képest jelentős ($\pm > 2$ óra) volt az eltérés, a szignifikancia hiánya tehát jobbra a rendkívül alacsony esetszám (az ittas járművezetés vétsége kivételével $N < 10$) következménye volt.

Az átlagos munkaidőt önmagukban nem szignifikáns mértékben emelő változók összevonásával és binomiálissá alakításával egy ún. „binomiális nehezítő” változót hoztam létre, amire nézve már szignifikáns különbség volt a várható munkaidőigényben. Az egyik önmagában is szignifikáns hatású változót (eredetileg tárgyalás mellőzésével indult-e az ügy) összevontam a szakmai tapasztalatok és a mért adat alapján szintén a munkaidőt csökkentő, de csak egy esetben előforduló másik binomiális változóval (eredetileg a Be. 569., 570., 571. §-a szerinti ügyként indult-e): a hatás együttesen is szignifikáns mértékben munkaidőt csökkentő. Ez alapján tehát végeredményben a vizsgálatban előforduló 25 binomiális változó közül 17 esetében bizonyítható értékelhető hatás a későbbi munkaidőigény vonatkozásában. Ez azt igazolja, hogy a szakma alapvetően helyesen határolta be a releváns vagylagos kezdőirati jellemzők körét.

²⁴ Főleg bűncselekményfajták, amelyek elvileg korlátlan rendűségben fordulhatnak elő.

²⁵ Terjedelmi okokból a diagramokat csak néhány változó vonatkozásában közlöm, elsősorban ott, ahol az eredmény bizonytalanak tűnt: jelentős esetszám mellett nem volt túl jelentős a különbség az egyes csoportok munkaidőátlagai tekintetében, vagy fordítva, kis esetszám mellett nagy különbség volt a munkaidőigény átlagai között.

Megvizsgáltam a munkaidőigény és az egyes binomiális kezdőirati jellemzők közötti korrelációs kapcsolatot is. Az adatok normalitásának hiányára tekintettel választott Kendall-féle rangkorrelációs próba eredményeit szintén az M4. táblázat tartalmazza, amelyből látható, hogy a vizsgált binomiális változók egyenként általában gyenge vagy igen gyenge kapcsolatot mutatnak a munkaidőigénnyel, de számos esetben biztosra vehető, hogy azt nem a véletlen okozza. Ez arra utal, hogy a munkaidőigény szempontjából e változóknak egyenként nincs jelentős szerepe.

Az elemzés során azokat a változókat tekintettem ordinális változóknak, amelyek esetén az ismérv értéke szerinti csoportok egyértelműen sorba rendezhetők voltak, és az azokba tartozó ügyek száma elegendő volt a csoportok közötti statisztikai összehasonlításhoz. Abból indultam ki, hogy ilyen esetekben egyértelmű sorrendiség van az ismérv értékei szerint különböző csoportokba tartozó ügyek munkaidőigénye között, mert például a több szakvélemény értékelése, több tanú kihallgatása, több rendbeli cselekmény elbírálása nehezebb. Az ordinális változók egyes csoportjaiban mért átlagos munkaidőigények közötti különbségek megítélése érdekében a hipotézisvizsgálatot Kruskal–Wallis-tesztel végeztem. Az eredmények alapján megállapítható, hogy csak a szakkérdések, a kihallgatni indítványozott szakértők, a polgári jogi igények, a fiatalok terheltek és a vád tárgyává tett gazdasági bűncselekmények, valamint a rablások számának változása okoz szignifikáns különbséget a csoportok munkaidőátlagai között. (Lásd az M5. táblázatot.) A többi változó összevonva sem bizonyult jelentősnek.

Megvizsgáltam az egyes ordinális változók és a munkaidőigény közötti korrelációs összefüggéseket is a Kendall-féle rangkorrelációs próba segítségével. Az eredmények szerint a perben érintett szakkérdések számától eltekintve, egyenként valamennyi ordinális kezdőirati változó gyenge vagy igen gyenge korrelációs kapcsolatban van a munkaidőigénnyel, a kapcsolat azonban a polgári jogi igények száma esetében szignifikáns, a szakkérdések, a kihallgatni indítványozott szakértők, a fiatalok terheltek és a vád tárgyává tett gazdasági bűncselekmények, valamint a rablások száma esetében erősen szignifikáns.

A kvázi folytonos (folytonos) kezdőirati változók esetében első lépésként leíró statisztikákat készítettem, melynek eredményeit az M6. táblázat tartalmazza. Az értékek eloszlására nézve normalitásvizsgálatot végeztem, az eredmény a Kolmogorov–Smirnov-teszt szerint $p < 0,01$, a Shapiro–Wilk-teszt alapján pedig $p = 0,0000$ volt valamennyi változó esetében, tehát nem mutatott normál eloszlást. Erre tekintettel a folytonos kezdőirati változók nyers adatai és a munkaidőigény közötti korrelációs kapcsolatok Kendall-féle rangkorrelációs próba segítségével voltak vizsgálhatók. Az eredményeket az 5. táblázat tartalmazza. Ez alapján megállapítható, hogy a vád tárgyává tett egyéb bűncselekmények kivételével ($p = 0,04235$) valamennyi folytonos kezdőirati változó erősen szignifikáns ($p < 0,000651$), és a binomiális vagy ordinális változók többségénél lényegesen erősebb (abszolút értékben $0,04994 < \tau < 0,508610$) korrelációs kapcsolatban áll a munkaidőigénnyel.

5. táblázat

Az egyes kvázi folytonos kezdőirati változók és a munkaidő igény közötti korrelációs vizsgálatok eredményei

Kezdőirati változó	Kendall-féle próba			
	jellege	Kendall τ	Z	p
8. Vádirat terjedelme	kvázi folytonos	0,406918	16,54237	1,82E-61
9. Nyomozati iratok terjedelme	kvázi folytonos	0,508610	20,67642	5,65E-95
10. A nyomozati iratok okirati mellékleteinek terjedelme	kvázi folytonos	0,083869	3,40952	0,000651
15. Hány szakvéleményt csatoltak?	kvázi folytonos	0,254864	10,36094	3,73E-25
26. Az eljárásban kihallgatni indítványozott tanúk száma	kvázi folytonos	0,421209	17,12332	9,94E-66
30. Bűnjelek száma	kvázi folytonos	0,230205	9,35847	8,09E-21
31. Vádoltak/fejjelentettek száma	kvázi folytonos	0,273463	11,11703	1,04E-28
37. A vád tárgyává tett cselekmények száma, folytatólágosság esetén a részcselekmények számának figyelembevételével	kvázi folytonos	0,316585	12,87007	6,63E-38
45. Csálás büntetési alakzatai (Btk. 318. §. (4)-(7) bekezdés)	kvázi folytonos	0,157405	6,39896	1,56E-10
46. Egyéb	kvázi folytonos	-0,049940	-2,030080	0,042348
63. Közbizalom elleni bűncselekmények (Btk. XVI. fejezet III. cím)	kvázi folytonos	0,085056	3,457784	0,000545
74. Vagyyonelleni bűncselekmények vétségi alakzatai (Btk. XVIII. fejezet)	kvázi folytonos	0,101297	4,118004	3,82E-05

Megjegyzés: Kendall τ a korreláció szoroságát jellemző szám.

A kapott eredmények megerősítése érdekében a folytonos kezdőirati változók és a munkaidőadatok (az összefüggés lineárisra való visszavezetése érdekében) egyszerűen vagy többszörösen logaritmikusan transzformált értékeinek scatter plot diagramjait is elkészítettem, és görbeillesztéseket végeztem. Azt, hogy az egyes kezdőirati változók esetében hányszoros logaritmikus transzformációt végeztem, az döntöttem el, hogy melyik transzformációnál kaptam a leginkább lineáris összefüggést. (Lásd az M7. táblázatot és az M8–M25. ábrákat.) Bár az így számolt Pearson-féle korrelációs együtthatók direkt módon nem hasonlíthatók össze a Kendall-féle τ értékekkel, mégis jól összhangban álltak azokkal.

Mindezek alapján levonható a következtetés, hogy vizsgált kezdőirati változók jelentős része erősebb vagy gyengébb, de egyértelműen szignifikáns kapcsolatba hozható a később felmerült munkaidőigénnyel. Bizonyos jellemzők ritkán fordulnak elő, vagy hatásuk kevésbé jelentős, illetőleg csak nagy szórással érvényesül, ezeknél a pontos összefüggések feltárásához nagyobb mintára lenne szükség.

Ezután már csak egyetlen nyitott kérdés maradt megválaszolatlanul: milyen számszerű kapcsolatban vannak a munkaidővel mint függő változóval a vele szignifikáns korrelációt mutató kezdőirati változók mint független változók, és milyen biztonsággal jelezhető előre a kezdőirati jellemzők összessége által a munkaidő várható mennyisége. Ennek tisztázása érdekében regressziós számításokat végeztem. A munkaidő mennyiségének meghatározásában szerepet játszó kezdőirati változók számára, a folytonos kezdőirati változóknak a munkaidővel fennálló, többé-kevésbé linearizálható kapcsolatára és a logaritmikusan transzformált munkaidőadatok normálisához közeli eloszlására tekintettel megkísértem a többszörös lineáris regressziós módszer alkalmazását, amelyre a Statistica program általános lineáris modelljét (GLM) választottam. Az ebben szereplő független változók körét, a függő változót, a független változók program szerinti besorolását, az alkalmazott szűrési feltételeket, valamint a kiszűrt ügyek számát az M7. táblázat tartalmazza.

A regressziós modell alapját 731, a szűrési feltételeknek megfelelő ügy képezte. A modell alapbeállításokkal történt futtatása után kapott eredmények szerint a bevont kezdőirati jellemzők együttesen nagyon erős ($R = 0,7812$) kapcsolatban vannak a munkaidő mennyiségével, összességében a munkaidő változékonyságát 58,7 százalékban magyarázzák ($\bar{R}^2 = 0,5870$), az összefüggés erősen szignifikáns ($p = 0,00$). Az összefoglaló adatokat és az egyes kezdőirati jellemzők szerepét a modellben az M8. és M9. táblázat mutatja be. A változók vonatkozásában elvégzett t -próbák eredménye szerint a modell kialakításában mindössze kilenc kezdőirati jellemzőnek van bizonyosan jelentős szerepe ($p < 0,05$).

A regressziós modell ellenőrzése érdekében – a program megfelelő funkciója segítségével – minden modellbe bevont ügyre meghatároztam a regressziós modell által becsült munkaidőt (illetőleg annak $\log(y+2)$ értékét). A becsült és a tényadat különbsége a hibatag, ami lényegében a tévedés mértékét mutatja az egyes ügyek ese-

tében. A hibatagok (logaritmikusan transzformált értékeinek) leíró statisztikái szerint a maradékok átlaga a lineáris regressziós módszer sajátossága folytán 0, a szórása 0,2118, standard hibája 0,0078. Az eloszlás nagyjából szimmetrikus (ferdeség = -0,2721), azonban a normálisnál csúcsosabb (csúcsosság = 2,0154) és a P-P plot diagram alapján várható, közelítőleg normális eloszlással szemben a próbák alapján nem normális (a Kolmogorov–Smirnov-teszt eredménye: $d = 0,07802$, $p < 0,01$; a Lilliefors-teszté: $p < 0,01$; a Shapiro–Wilk-teszté: $W = 0,96827$, $p = 0,00000$).

Megvizsgáltam a változók varianciájának állandóságát (homoszkedaszticitását) is. Az elvégzett Levene-teszt eredménye a nyomozati iratok terjedelme kivételével minden más független változó vonatkozásában szignifikáns (p minden esetben $< 0,0067$) volt, tehát a homoszkedaszticitás nullhipotézisét elvetve megállapíthatjuk, a változók heteroszkedasztikusak. Mindezekből az következik, hogy a modell ellenőrzésére használt, az eloszlással szemben a normalitás és a homoszkedaszticitás követelményét támogató próbák eredményeit óvatosan kell kezelni.

A nagy számú független változóra tekintettel ellenőriztem a modellben jelenlévő multikollinearitás mértékét is. Ennek érdekében kiszámoltam az egyes független változók vonatkozásában a varianciát infláló faktort (variancianövelő tényezőt, *VIF*-et) és annak reciprokát, a toleranciamutatót (*T*). A multikollinearitás mértéke akkor tekinthető erősnek, zavarónak, ha a $VIF \geq 2$, és a $T \leq 0,5$, nagyon erősnek, károsnak pedig akkor, ha a $VIF \geq 5$, illetve a $T \leq 0,2$. Ez alapján a regressziós modellben szignifikáns szerepet játszó független változók közül kizárólag az ügy magánvádas jellegénél volt tapasztalható káros mértékű multikollinearitás ($VIF = 55,5336$, $T = 0,0180$). Erős volt a multikollinearitás a vád tárgyává tett cselekmények száma ($VIF = 2,9019$, $T = 0,3445$), a nyomozati iratok terjedelme ($VIF = 2,6977$, $T = 0,3706$), a vádirat terjedelme ($VIF = 2,3633$, $T = 0,4231$) és az eljárásban kihallgatni indítványozott tanúk száma ($VIF = 2,0665$, $T = 0,4838$) esetében is. A többi függő változónál a multikollinearitás mértéke gyenge volt ($VIF < 2$, $T > 0,5$).

Erre tekintettel a végleges regressziós modellből az ügy magánvádas jellegére utaló független változót kihagytam. Az így felépített újabb modell összefoglaló adatait a 7. táblázatban közlöm. Látható, hogy a modellben bennmaradt, immár csak hét releváns kezdőirati jellemző együttesen továbbra is nagyon erős ($R = 0,7695$) kapcsolatban van a munkaidő mennyiségével, összességében a munkaidő változékonyságát 57 százalékban magyarázzák ($\bar{R}^2 = 0,5703$), az összefüggés erősen szignifikáns ($p = 0,00$). Az egyes kezdőirati jellemzők szerepét a modellben a 6. táblázat foglalja össze. A modell már mentes a káros mértékű multikollinearitástól (VIF max. $< 2,89$, T min. $> 0,3460$), és multikollinearitás mértéke csak három releváns változó esetén erős, a többinél gyenge.

6. táblázat

Az egyes kezdőirati változók szerepe a szűrt adatok alapján felépített végleges modellben és a multikolinearitás jelenléte

Kezdőirati változó	Koefficiens	t	p	β	Tolerancia-mutató (T)	Variánciamutató faktor (V/F)
Tengelymetszet	-0,139885	-0,70713	0,479726			
Vádirat terjedelme	14,897845	2,60111	0,009490	0,09623	0,429978	2,32570
Nyomozási iratok terjedelme	0,213402	10,61047	0,000000	0,37925	0,460640	2,17089
A nyomozási iratok okirati mellékleteinek terjedelme	0,182800	1,51836	0,129381	0,04119	0,799710	1,25045
Hány szakterés tekintetében csatlakoztak szakvéleményt?	0,719310	0,10098	0,919597	0,01097	0,049847	20,06127
Hány szakvéleményt csatlakoztak?	0,590676	0,09381	0,925284	0,01022	0,049589	20,16580
Az eljárásban kihallgatni indítványozott tanúk száma	0,238040	5,46448	0,000000	0,18541	0,511174	1,95628
Az eljárásban kihallgatni indítványozott szakértők száma	0,268014	1,51135	0,131155	0,03892	0,887586	1,12665
Büntetések száma	0,123188	0,80031	0,423803	0,02310	0,706289	1,41585
Vádlottak fejlemének száma	1,139714	3,03583	0,002489	0,09206	0,639976	1,56256
Hány, a vádirat szerinti fiatalokot terhelte vád?	-0,054352	-0,36160	0,717759	-0,01015	0,746981	1,33872
A vád tárgyává tett cselekmények száma, folytatlagosság esetén a részselekmények számának figyelembe vételével	5,134829	2,20895	0,027505	0,09109	0,346048	2,88977
Csalás büntetői alakzatai (Btk. 318. §. (4)-(7) bekezdés)	1,457613	1,26979	0,204586	0,04036	0,576833	1,73361
Egyéb	-0,027460	-0,36291	0,716782	-0,01159	0,577186	1,73254
Garzósági büntéselekmények (Btk. XVII. fejezet)	0,000000				0,000000	
Közbizalom elleni büntéselekmények (Btk. XVI. fejezet III. csm)	-0,20735	-2,23098	0,026001	-0,06338	0,724518	1,38023
Rablás (Btk. 321. §.)	0,307829	1,61933	0,105832	0,04380	0,804422	1,24313
Vagyronelleni büntéselekmények vétségű alakzatai (Btk. XVIII. fejezet)	0,024185	0,22903	0,818910	0,00684	0,660297	1,51447
Hatalyon kívül helyezés utáni-e az ügy?	0	1,31123	0,190213	0,07654	0,172711	5,79002
Csatlak-e digitális melléklet?	0	0,39607	0,692172	0,06305	0,023222	43,06187
Keil-e határozni kényszerítéskérésről?	0	0,51715	0,605219	0,03349	0,055003	18,18077
A vádlott(ak) valamennyi cselekmény elkövetését beismerék (teljes beismerés)?	0	-0,07103	0,943394	-0,01033	0,027830	35,93205
Hitelserzés (Btk. 330 §.)	-0,140696	-2,49136	0,012958	-0,06121	0,974811	1,02584
Hűtlen kezelés büntetői alakzatai (Btk. 319. §. (3) bekezdés)	0	-1,25440	0,210121	-0,03472	0,768185	1,30177
Eredetileg tárgyalásmeilőzéses vagy a Be 569., 570., 571. §-a szerinti ügykénti indult-e?	0	-0,91310	0,361507	-0,08119	0,074443	13,43318
Összevont neheztől bűnműködés + zsarolás (Btk. 323. §.) (bűnműködés)	0	0,72832	0,466541	0,08060	0,048084	20,79696

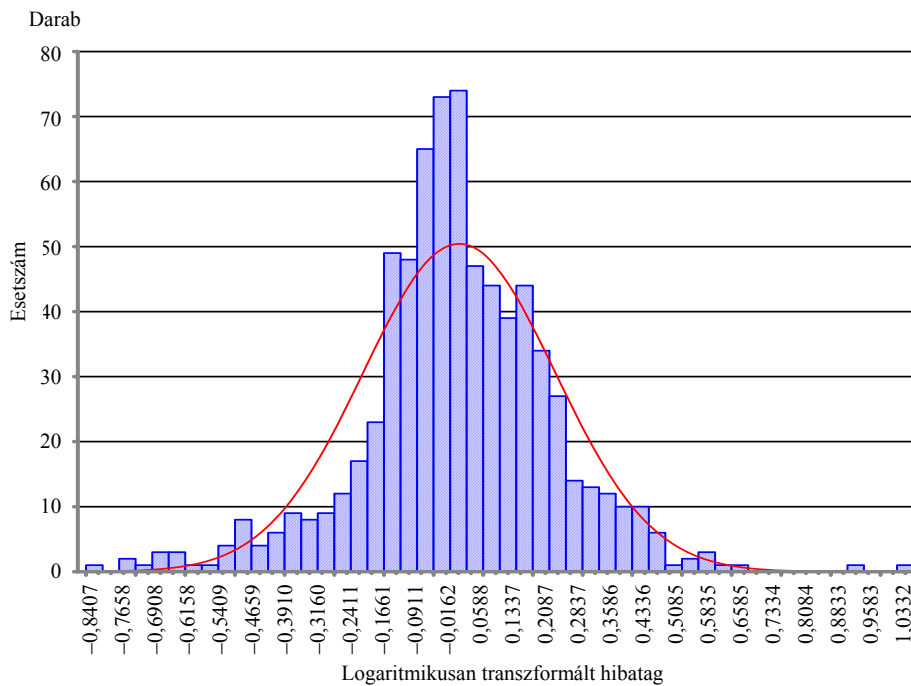
Megjegyzés. A „p” valószínűséget jelent.

7. táblázat

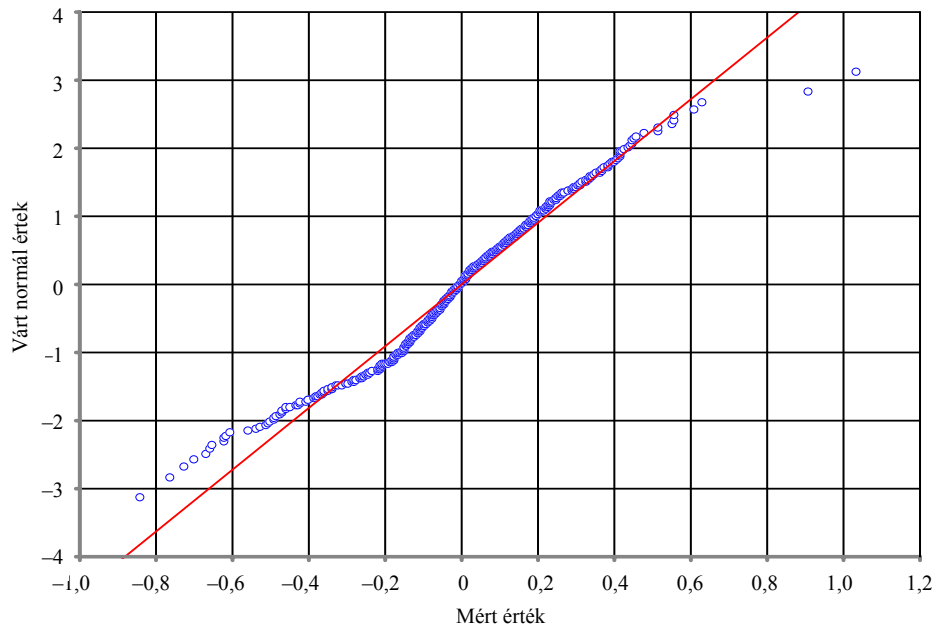
*A munkaidő és a kezdőirati jellemzők közötti többszörös lineárisregresszió-számítás
végleges eredménytáblája
(szűrt adatok)*

Ismérv	Érték
Többszörös korrelációs együttható (R)	0,769521
Többszörös determinációs együttható (R^2)	0,592162
Korrigált többszörös determinációs együttható (\bar{R}^2)	0,570387
Négyzetösszeg (regresszió)	49,777620
Szabadságfok (regresszió)	37
Átlagos négyzetösszeg (regresszió)	1,345341
Négyzetösszeg (maradék)	34,283190
Szabadságfok (maradék)	693
Átlagos négyzetösszeg (maradék)	0,049471
F	27,194710
p	0,00

1. ábra. A végleges regressziós modell (logaritmikusan transzformált) hibatajának hisztogramja



2. ábra. A végleges regressziós modell (logaritmikusan transzformált) hibatagjának normál P-P plot diagramja



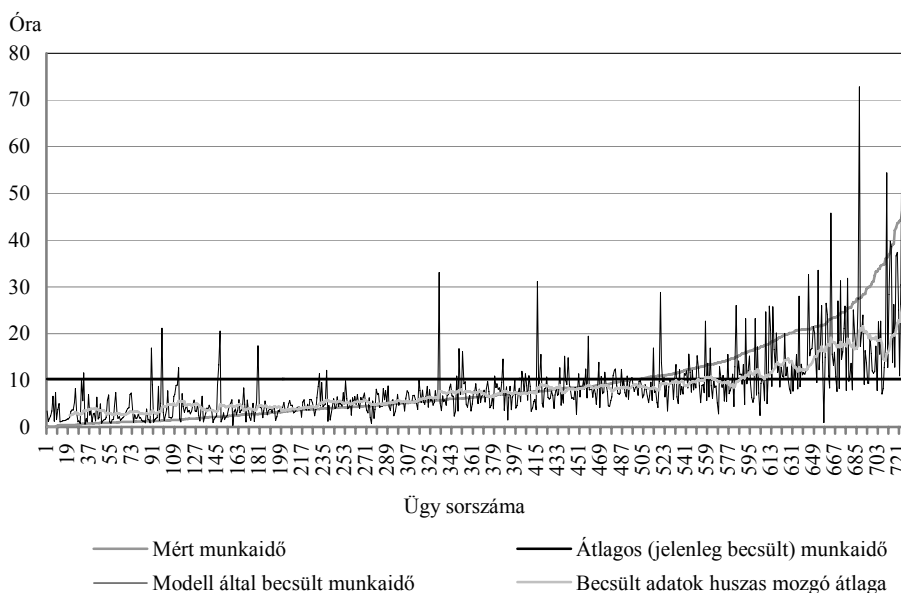
A hibatagok (logaritmizált) szórása 0,2167, standard hibája 0,0080, az eloszlás azonban – az 1. ábrán látható hisztogram, és P-P plot diagram ellenére – továbbra sem normális (a Kolmogorov–Smirnov-teszt eredménye: $d = 0,073104$, $p < 0,01$; a Lilliefors teszté: $p < 0,01$; Shapiro–Wilk-teszté: $W = 0,972678$, $p = 0,00000$) és heteroszkedasztikus (a Levene-teszt eredménye a vádirat terjedelme és a kihallgatni indítványozott tanúk száma kivételével minden más független változó vonatkozásában szignifikáns ($p < 0,0009$)).

Összefoglalva tehát a regressziós modell bár alkalmasnak tűnik a kezdőirati változók és a várható munkaidőigény közötti összefüggésrendszer leírására, alkalmazási feltételeinek maradéktalan teljesülése hiányában jelenlegi formájában előrejelzésre és elemzésre nem kielégítően alkalmas. A nagy számú – és egy nagyobb mintán elvégzett újabb vizsgálat eredményeképpen a munkaidő mennyiségével vélhetően egyre jelentősebb számban szignifikáns összefüggésbe hozható – függő változóra tekintettel főkomponens-elemzés, vagy ennek sikertelensége esetén, nem lineáris modell alkalmazása jobb eredményt hozhat.

A modellt – annak nem kielégítő volta ellenére – azért érdemes összehasonlítani a jelenleg alkalmazott munkaidő-becslési módszerrel. Ennek eredményeit a 3. ábrán foglaltam össze, ahol a végleges regressziós modell kialakításában figyelembe vett ügyek (731 db) szerepelnek a kisebb munkaidő-igényűektől a nagyobbak felé halad-

va, sorba rendezve a vízszintes tengelyen. Az ügyekhez tartozó mért munkaidőértékeket a függőleges tengelyen ábrázoltam. Eloszlásukat az egyes ügyek értékeit összekötő sötétebb szürke vonal mutatja. A jelenlegi munkateher-mérési rendszerben nem vesszük figyelembe az ügyek munkaidőigénye közötti különbségeket. Minden ügy munkaidőigényét egyenlőnek, vagyis átlagosnak tekintjük (egalizációs módszer). Ezt szimbolizálja a diagramon a 10,27-es értéknél²⁶ húzott vastag fekete vízszintes vonal. Az adott ügyben a regressziós modell által becsült munkaidőigény értékét a vékony fekete vonal jelzi. Az utóbbi lefutása alapján látható, hogy a becsült értékek valamennyi tartományban igen nagy szórást (9,38 óra) mutatnak, azaz a modell az egyes ügyek munkaidejét pontosan megjósolni nem mindig tudja (a tényleges és a jósolt érték közötti átlagos eltérés 4,42 óra). Szemmel látható azonban az is, hogy az erős hullámzás ellenére a vékony fekete vonal lényegesen jobban követi a mért értékeket ábrázoló sötétebb szürke vonalat, mint a jelenleg alkalmazott módszer által eredményezett vízszintes vonal (szórás 0, a tényleges és a jósolt érték közötti átlagos eltérés 7,95 óra). A modell által becsült és a mért értékek közötti egyezés az átlagos körüli vagy attól kisebb munkaidő-igényű ügyek esetében egészen jó, ezeknél a „hullámzás” is kevésbé jelentős, bár a jósolt érték általában magasabb a ténylegestől.

3. ábra. Az ügyek mért, átlagos (jelenleg becsült) és a végleges regressziós modell által becsült munkaidőigénye (szűrt adatok)



²⁶ Az érték az adatok szűrése miatt nem egyezik a teljes minta 10,53-os átlagával.

Az előrejelzés biztonsága a nagyobb munkaidő-igényű ügyek esetében romlik, ebben a tartományban erősebb a hullámzás, és a jószolt érték is egyre inkább elmarad a ténylegestől. Ez összhangban áll azzal, hogy a független változók és a hibatagok értékei között pozitív korreláció volt megfigyelhető. Ezzel együtt a mért értékekhez összességében mégis sokkal közelebb vannak a modell által becsültek, mint az átlag (jelenlegi becslés). Ez vizuálisan is érzékelhető, ha az egyes ügyekhez tartozó modell által becsült értékek helyett azok huszas mozgó átlaga által kiadott világosabb sűrű vonalat hasonlítjuk a mért vagy az átlagos értékek vonalához. Vagyis a matematikai szempontból, abszolút értelemben nem kielégítő regressziós modell összességében minden fogyatékosága ellenére mégis jobb eredményt képes produkálni, mint a jelenlegi módszer. Kijelenthetjük tehát, hogy az utolsó kérdésre is igen a válasz: a helyi bírósági peres büntető ügyek várható munkaidőigénye pusztán a kezdőirat bizonyos jellemzői alapján ésszerű – de legalábbis a jelenlegitől lényegesen nagyobb – bizonyossággal előre becsülhető.

Összefoglalva megállapíthatjuk, hogy a helyi bírósági peres büntető ügyek munkaidőigénye – a befejezett ügyek iratai alapján – abszolút és relatív értelemben is mérhető, és ez a mérési módszer valamennyi más bírósági ügyszakra adaptálható. A mért adatok szerint az ügyek munkaidőigényében jelentős eltérések vannak, és a különböző munkaidő-igényű ügyek területi eloszlása sem egyenletes. A mért munkaidőigény az ügyek kezdőiratának számos jellemzőjével laza összefüggésbe hozható, melyek együttesen azonban már erős korrelációt mutatnak azzal. Az ennek alapján felépített, de jelenlegi formájában még nem végleges többszörös lineáris regressziós modell segítségével pusztán a kezdőirat alapján elfogadható, bár egyelőre nem kielégítő biztonsággal előrejelezhető az ügyek várható munkaidőigénye, mellyel jelentős mértékben javulhat a becslés pontossága ahhoz képest, mintha minden ügyet egyenként tekintenénk. Erre tekintettel érdemes megismételni a vizsgálatot nagyobb mintán, valamint – a módszert megfelelően adaptálva – más bírósági ügyszakokban is, és tovább kell dolgozni egy megbízhatóbb predikciós modell felépítésén. A kiépítendő munkateher-mérési rendszerben az érkezett ügyek száma mellett várhatóan az iratok lényeges jellemzői által ügyszakonként, szervezeti egységenként vagy egyedileg meghatározott munkaidőigényt is figyelembe lehet majd venni.

Függelék

Beadvány: A bírósági ügyek nem a bíróság tevékenysége folytán keletkező (jegyzőkönyvek, végzések, ítéletek, átiratok, megkeresések stb.), hanem külső forrásból származó iratai (felek írásbeli kérelmei, indítványai, előkészítő iratai, jogorvoslati kérelmei, megkeresések stb.).

Bírói létszám: az igazságügyi igazgatásban számos eltérő tartalmú bírói létszám-fogalom (engedélyezett, tényleges, működő, tárgyaló) létezik, a munkateher mérésére a tárgyaló tanácsok szá-

mát célszerű használni. Ez egy olyan tisztított létszámadat, amelyet a tisztán ítélkezési tevékenységet végző bírák ténylegesen ítélkezéssel töltött (távollétekkel csökkentett) munkaideje alapján határoznak meg a bírósági statisztikai adatok gyűjtéséről és feldolgozásáról szóló 2003. évi 8. számú OIT-szabályzat alapján.

Érdemi munkavégzés: minden olyan, a bírósági ügy elintézése érdekében végzett önálló intézkedő, döntés-előkészítő, döntéshozó tevékenység, amelyre jogszabály alapján csak bíró (bírósági titkár, bírósági ügyintéző, bírósági fogalmazó, végrehajtási ügyintéző) jogosult.

Ítélezési szint: az igazságszolgáltatási szervezet horizontális tagolása folytán létrejött, általában igazgatásilag is elkülönülő szervezeti egység (Kúria, ítélőtáblák, törvényszékek, járásbíróságok, közigazgatási és munkaügyi bíróságok, első és másodfokú ítélkezés).

Kezdőirat: irat, amelynek a bíróságon nincs előzménye vagy amely törvény értelmében érdemben, illetve ügyviteli szempontból befejezett ügyben új eljárás lefolytatását, valamint a már befejezett ügygel kapcsolatban önálló intézkedést, valamely kérelem önálló elintézését igényli (a bírósági ügyvitel szabályairól szóló 14/2002. (VIII. 1.) IM-rendelet 2. § 12. pont). Lényegében olyan beadványt, iratot jelent, amely a bíróságon új ügyként kerül lajstromozásra. Az eljárások kezdetén kizárólag ez az irat áll rendelkezésre.

Kurrenciális tevékenység: a beérkező új ügy tanulmányozása, tárgyalásra kitűzése, a tárgyalások között érkező utóiratok elolvasása, tanulmányozása, elintézése, felvilágosítás adása, bármilyen átírat, feljegyzés, nem fellebbezhető végzés készítése, jegyzőkönyv javítása, utasítás adás stb.

Nyomozati iratok: a bíróság elé kerülő büntető eljárásoknak a nyomozó hatóság által lefolytatott nyomozati szakasza során keletkezett iratokat összefoglaló neve. (A büntető eljárásnak nyomozati, ügyészségi és bírósági szakasza van.)

Rendűség: A vád tárgyává tett tényállás elkövető(k) általi megvalósításának száma (például egy vagy két rendbeli súlyos testi sértés volt a vád tárgyává tett cselekmény).

Székhelyi városi bíróság: a megye törvényszékének városában működő városi bíróság, Budapesten a Pesti Központi Kerületi Bíróság. A többi városi bíróságtól való megkülönböztetést a székhelyi városi bíróságok speciális hatáskörei indokolják.

Újrainszertelt ügy: A bírósági ügyek nem minden esetben fejeződnek be véglegesen. Előfordul, hogy a befejezés csak ügyviteli, azaz az ügy érdemében nem születik végleges döntés. Ez történik, ha az eljárás félbeszakad (valamelyik fél halála miatt), szünetel (felek kérelmére átmenetileg nem folyik az eljárás), vagy a bíróság a tárgyalást felfüggeszti (valamely előzetes kérdés más bírósági eljárásban való eldöntéséig elhalasztja az ügy tárgyalását). Ilyen esetekben később általában folytatódik az eljárás (új ügyszámon) az érdemi döntés meghozataláig. Az újra kezdődő ügyek a bírósági ügyvitelben új(ainszertelt) ügyekként jelennek meg. Ezek iratait a bírósági ügyvitel szabályai szerint véglegesen az előzményi iratokhoz kapcsolják, azokkal együtt kezelik és mozgatják, és a különböző ügyszámon keletkezett iratok az eljárás befejeződése után sem válnak el egymástól.

Ügycsoport: a bírósági ügyek tárgyuk, valamint az alkalmazandó anyagi és eljárási szabályok azonossága vagy specialitása alapján kialakított halmazai (a bírósági ügyvitel szabályairól szóló 14/2002. (VIII. 1.) IM-rendelet 2. § 18. pont).

Ügyszak: az egyes szinteken belüli, az egyes ügycsoportokat intéző szakmai alapon szerveződő egység, egy-egy ítélkezési szint vertikális tagolása.

Vádírat: A bíróság elé kerülő büntető eljárásoknak az ügyészség által lefolytatott szakaszát befejező irat. A bíróságon az első fokú büntető peres ügyek kezdőirata.

Irodalom

ÖRKÉNYI L. [2010]: Az ítékezés időszerűtlenségének okai és a lehetséges megoldások a Fővárosi Bíróság polgári első fokú csoportjában. *De Jure*. 5–6. sz. 27–44. old.

ÖRKÉNYI L. [2012]: Arányos ügyteher elosztás az igazságszolgáltatásban – elvek és teendők. *Debreceñijogimuhely.hu*. IX. évf. 1. sz. http://www.debreceñijogimuhely.hu/archivum/1_2012/aranyos_ugyteher_elosztas_az_igazsagszolgalattasban_elvek_es_teendok/

Summary

At present we do not know how much time is needed to complete individual court cases, how to measure it, we do not have information on the distribution of cases that require different working time, as well as on the kind – and strength – of correlation between working time and the other features of cases. The aim of this study was to answer these questions based on data gained from the retrospective examination of the files of closed court cases, by means of statistical analysis and a prediction model. According to the results, the working time required for the completion of litigious criminal cases of first instance ranges between 0.022 and 231.14 hours at local level, the average is 10.53 hours. There is a significant difference between the averages of the district court of the county seats and of other district courts, just like between these of individual district courts. The difference between the regional courts is not significant. The characteristics of the starting documents are in diverse connection with the required working time; it is weak or very weak with regard to the individual features; but in total the connection is very strong, and the total rate of the determination of these characters is $\bar{R}^2 = 0,57$. Consequently, the working time can be estimated on the basis of the documents of closed cases, the working time required for individual cases is different, and the territorial distribution of the cases with various working time needs is irregular. This way the expected working time can be estimated with an acceptable appraisal relying on the starting document of the case.

Harcza István,
a KSH statisztikai főtanácsadója
E-mail: istvan.harcza@yahoo.com

Észrevételek Huszár Ákos tanulmányához*

Az írást igen fontos elméleti alaptanulmánynak tekintem a hosszabb ideje esedékes rétegződésvizsgálat elindítása előtt. Úgy vélem, hogy egy kiterjedt rétegződéskutatást további hasonló jellegű tanulmányokkal lehet kellően megalapozni, amelyek számos kérdésben hozzájárulhatnak a szakmai konszenzus kialakításához.

1. A dolgozat legfőbb erényei

Az anyag első része világos, lényegre tapintó kritikai értékelését adja a fontosabb elméleti irányzatoknak, bemutatva azok előnyeit és hátrányait.

A szerző abból indul ki, hogy egy-egy elméleti koncepció érvényességét csak annak alapján lehet megítélni, hogy az operacionalizálás mennyire volt konzisztens a kiinduló elméleti alapokkal. Egy-egy elméleti megközelítés releváns kérdéseinek a megválaszolása azonban még nem jelenti azt, hogy a társadalom viszonylag valóság-hű tagozódásáról kellő képet kapunk. Hogy mit tekintünk viszonylag valóság-hű tagozódásnak, az sok tekintetben „szakmai konszenzus” kérdése, de nem csak az.

A szerző nyitott a különböző elméleti megközelítések párhuzamos alkalmazására. Erről tanúskodik bevezetőjének következő mondata: „A dolgozat akkor éri el a célját, ha hozzásegít a magyar társadalom rétegződésének vizsgálatára alkalmas osztálymodellek kidolgozásához.” E nyitottság jó alapot ad a későbbiekben szakmai konszenzussal elfogadott, relevánsnak tekinthető modellek kidolgozásához és teszteléséhez. E tekintetben a kulcskérdés az, hogy mit tekintünk releváns modellnek? Erre az a válasz adható, hogy akkor releváns egy modell, ha az elméleti mód-

* Hozzászólás Huszár Á. [2013]: Foglalkozási osztályszerkezet (I.) – Elméletek, modellek, valamint Foglalkozási osztályszerkezet (II.) – Az osztályozás problémái (*Statisztikai Szemle*. 91. évf. 1. sz. 31–56. old. és 2. sz. 117–131. old.) tanulmányához.

szertani és empirikus eszköztárát tekintve kellően illeszkedik a vizsgálni kívánt kérdésekhez.

2. További tisztázásra váró kérdések

Nincs kellően kifejtve és hangsúlyozva, hogy a társadalom sokszínű tagozódását nem lehet egyetlen univerzális modellel leképezni. Ha elfogadjuk a különböző megközelítések jogosságát, akkor azt is ki kell fejteni, hogy mi az, amit e megközelítések a sokszínű társadalmi tagozódás ábrázolása érdekében relevánsnak tekintünk, továbbá, hogy a különböző rendezőelvek alapján felépített modellek segítségével végül is hogyan jutunk közelebb – párhuzamos megközelítésekre alapozva – egy valóságot kellően tükröző társadalmi rétegződés megalapozásához.

Feltehető, hogy ez egy következő szakasz az alapozó tanulmányok sorában, nevezetesen azon alapvető rendezőelvek csokorba szedése, amelyek párhuzamos (többféle) modellben való alkalmazása teljesebb képet adhat a magyar társadalom rétegzettségéről.

Foglalkozási osztályszerkezetről beszél a szerző, azonban nincs kifejtve, hogy a mai magyar társadalmat milyen mértékben lehet osztályszempontok alapján ábrázolni, illetve milyen mértékben nem. Az osztályfogalom tartalmi relevanciáját mindenképpen be kellene mutatni és ezzel együtt annak korlátait is. Vajon például az EGP-sémában¹ minden csoportot osztálynak tekinthetünk? Nem valószínű.

Az elméletek, majd azok operacionalizálása mennyiben értékfüggő? Ezt a kérdést azért is tartom fontosnak, mert a szerzővel való előzetes „írásváltás” során kiderült, hogy ebben eltérő véleményen vagyunk.

Maga az elméleti eszköztár megválasztása is értékfüggő, amit azért célszerű hangsúlyozni, mert minden elméleti megközelítés – részben a kutató értékrendje, részben célszerűségi szempontok miatt – eleve lehatárolja érvényességi körét, és amikor az ezen elméleti alaphoz illeszkedő, empirikus információk alapján levont következtetéseket értékeljük, akkor csak a kijelölt érvényességi körön belüli jelenségvilág esetében kérhetjük számon a relevanciát.

Előrebocsátható, hogy valóságközeli értelmezését alapvető mértékben meghatározza az életvilágról vallott elképzelés, jelen esetben a kutatóban megfogalmazódó társadalomkép. A társadalomkép megformálódásában iránytűként „viselkedik” a kutató értékrendje, sőt azt is mondhatnánk, hogy a megismerési folyamatban ez az értékrend egyfajta szemüveget kölcsönöz a kutatónak, aki a továbbiakban azt látja,

¹ A nemzetközi gyakorlatban használatos, és a szerző is használja az Erikson – Goldthorpe – Portocarrero nevével fémjelzett rétegződési sémát.

amit ezen a szemüvegen keresztül észlel. Ennek megfelelően, az észlelt jelenség értelmezésére koncepcionális keretet alakít ki, hipotéziseket fogalmaz meg, majd empirikus eszközökkel próbálja verifikálni az elméleti megközelítést. Ha a kutató mindezt konzekvens módon hajítja végre, akkor a teljes folyamat minden szakasza értékfüggő, hiszen a közbenső, eldöntésre váró kérdéseket is egy koherens értékrend szelvében fogalmazza meg.

Ez a tevékenység némi túlzással hasonlatos a „valóságshow-khoz”, mert ezekben a produkciókban is azt láttatják a nézővel, amit egy előzetesen kialakított koncepció alapján láttatni akarnak. Miután sokféle változatban jöhetnek létre értékrendek, ezért „sokféle valóság” létezik, ami természetes is, hiszen az életvilág és ezen belül a társadalmi újratermelődés rendszere sokféle, egymást erősítő vagy gyengítő, esetleg egymással semleges viszonyban levő értékek és érdekek alapján szerveződik. A tükrötetés során pedig kiemelt fontossága van a mérési módszerek konzisztens használatának. Ebben az értelmezésben a társadalom tagozódása sokféle metszetben jelenhet meg, ha ezeket párhuzamosan állítjuk egymás mellé, akkor inkább egymás mellett élő „társadalmak” összességéről beszélhetünk. Ebből az összességből egy-egy modell csak a számára releváns elemek, illetve rendezőelvnek megfelelően tud releváns információkat adni, de nem magáról a tagozódás teljességéről.

3. Néhány korábbi tanulság

A korábbi Társadalomstatisztikai főosztályon² végzett rétegződési vizsgálatok alapján elmondható, hogy kellően megalapozott rétegződésvizsgálatot csak teljes körű népszámlálási, valamint célzott, nagymintás társadalmirétegződés-vizsgálatok adatainak együttes alkalmazása alapján célszerű megtervezni. A népszámlálás előnye, hogy a teljeskörűségből fakadóan részletes csoportbontásokat tesz lehetővé, ám a modellekbe bevonható információk (változók) köre meglehetősen szerény. A sokszínű tagozódás bemutatásához szükséges információk körét csak a célzott rétegződésvizsgálat alapján lehet elfogadható szinten biztosítani.

A tapasztalatok azt mutatják, hogy *Grusky*³ javaslatának megfelelően „az elemzés fókuszát vigyük lejjebb, az egyes foglalkozásokat átfogó mikroosztályok szintjére”. Ez a korábbi munkálatok során azt jelentette, hogy első szinten az alacsonyabb szinten megragadható, sok közös vonást felmutató foglalkozási csoportok rendszerét

² A hivatkozott szervezeti egység 1990 és 2006 között működött a KSH-ban, és jogutóda volt az 1970-ben, ugyanezen néven szereplő főosztálynak.

³ *Weeden, K. A. – Grusky, D. [2005]: The Case for a New Class Map. The American Journal of Sociology. Vol. 111. No. 1. pp. 141–212.*

próbáltuk összeállítani, majd ezt követően, bizonyos kompromisszumok alkalmazásával, különböző aggregációs szinteket alakítottunk ki.

4. A fogalmak időtállósága versus a változó viszonyok alapján való „aktualizálása”

Az előrelépés lehetséges útjainak feltérképezésére gondolva fontosnak vélem az egyes elméleti-módszertani megközelítések empirikus alapjainak „tesztelését”, ezen belül elsősorban a sémaképzésnél használt fogalmak áttekintését, sőt a közöttük megteremthető „átjárhatóságot”. Ehhez kapcsolódva kulcskérdésnek tekintem a fogalmak időbeli relevanciáját. Statisztikai szempontból releváns, hosszú távú trendeket csak kellően koherens fogalmi rendszerek alapján lehet kimutatni. Ilyen megközelítés esetén lehetőség nyílik a tartós, illetve a meghatározott ideig jellemző tendenciák kimutatására.

Véleményem szerint a fogalmak időbeli relevanciáját figyelembe kellene venni és biztosítani kellene más, a rétegződés szempontjából fontos adatgyűjtések esetében is. Alapvető cél, hogy e fogalmak egyidejűleg legyenek alkalmasak a korábbi (20–40 évvel ezelőtti) adatgyűjtések (például népszámlálások, ismétlődő társadalmirétegvizsgálatok, munkaerő-felmérések stb.) mai fogalmi keretek közötti újrafeldolgozására, valamint egy-egy kiemelt megközelítési szemponton (metszeten) belül a közelmúlt hasonló tematikájú vizsgálatainak a korábbi viszonyokkal való összehasonlítására.

Mіндеzt csak egy integrált társadalomstatisztikai rendszer keretében lehet megvalósítani, amelyben kiemelt szerepet kap az egységes kulcsváltozók alkalmazása, illetve az ehhez szükséges elemi információk azonos módon való begyűjtése.

Beszélgetés S. Molnár Edit szociológus-demográfussal

Filozófia szakos tanári diplomával és pszichológiai tudományos fokozattal a zsebében *S. Molnár Edit* demográfiai és családszociológiai kutatásoknak szentelte életét. Népesedési, családpolitikai, attitűdvizsgálati téren egyaránt a családi-életforma-választás, a házassági-élettársai kapcsolatok, a gyermekvállalási motívumok, majd később a nyugdíj előtt álló korosztály magatartásvizsgálata foglalkoztatta. Mindezekben épp úgy, mint a kérdőíves kérdezéstechnika módszertani kérdéseinek tisztázásában jelentős eredményeket ért el. Hosszú ideig tagja volt a Magyar Tudományos Akadémia Demográfiai Bizottságának. Nemzetközi összehasonlító kutatásai mellett kutatásszervezési, oktatói munkája is jelentős. Kutatási eredményei elsősorban a KSH Népeségtudományi Kutatóintézet sorozataiban, a Demográfia és a Statisztikai Szemle folyóiratokban láttak napvilágot.

Beszéljünk először a családi körülményeiről, a szülői házról, iskoláiról.

Szüleim polgári iskolai tanárok voltak, apám matematika és fizika, anyám rajz és kézimunka szakos. Családi könyvtárunk elég jó színvonalú volt, különösen sok, két világháború között született szociográfiai mű sorakozott a polcokon. Már középiskolás, főként pedig egyetemista koromban nagy hatással voltak rám például *Szabó Zoltán*, *Györffy István*, *Nagy Lajos*, *Féja Géza*, *Németh László*, *Kodolányi János* művei – ezek a mai napig kedves olvasmányaim. Lenyűgözött e tudósoknak a helyismerete, empátikus kérdezési technikája, a valóság iránti olthatatlan érdeklődésük. Alapvetően humán érdeklődésű lettem, ezért a középiskola után filozófiát tanultam, minthogy ebben az időben még nem volt szociológusképzés. 1957-ben filozófia szakos tanári diplomával egy budai gimnáziumban kezdtem pszichológiát és logikát tanítani. A kísérletpszichológia-tankönyv szerzője *Lé-*

nárd Ferenc pedagógus-pszichológus volt, később az ő ösztönzésére kértem felvételemet hároméves ösztöndíjas pszichológiai aspirantúrára (ez a mai doktori iskolának felel meg). *Lénárd Ferenc* lett az aspiránsvezetőm, és így részt vehettem az ő kutatómunkájában is. A vele való együttműködés terelte figyelmemet a kérdésszövegek – esetünkben a tankönyvi, a tanári, valamint a tanulók által feltett kérdések megértése, értelmezése felé. Azt vizsgáltuk, milyen gondolkodási műveletek zajlanak a tanulók fejében a kérdések megértését igénylő feladat-megoldásaikban, és ezek a gondolkodási műveletek milyen pedagógiai módszerekkel fejleszthetők. Kandidátusi értekezésem témája is ez volt; „A szépirodalmi szöveg megértésének neveléslélektani vizsgálata” címen megjelent könyvemmel 1963-ban nyertem el a pszichológiai tudomány kandidátusa fokozatot. Így aztán egy életre eljegyeztem magam annak a módszertani problémának a vizsgálatával, hogyan lehet biztosítani az ún. „puha”, „lágyművelet” adatok minőségét, az ezek begyűjtését célzó

kérdőívkérdések megfelelő szintű megérttetésével.

A pedagógusi pálya elhagyása után fordult a közvélemény-kutatás felé?

Igen, az 1960-as évek második felében már Magyarországon is terjedőben voltak az ún. survey jellegű kutatások, felvételek. Az aspiráns évek befejezése után módomban volt a Magyar Rádió és Televízió (ez akkoriban még közös elnökség irányításával működött) Közvélemény-kutató osztályán dolgozni, mely később, 1968-ban Tömegkommunikációs Kutatóközpont néven intézetté alakult. Ez az időszak izgalmas tanulási korszakom volt, hiszen mivel még ebben az időben sem volt szociológusképzés, nemzedékem tagjainak maguktól kellett elsajátítaniuk a szociológia alapvető ismereteit. Ezért volt jelentősége annak, hogy ad hoc jelleggel önképzőköröket, tanulóköröket szerveztünk magunknak, és közösen megvitattuk ismereteinket, olvasmányainkat, kezdeti tapasztalatainkat. Jól emlékszem például a Központi Statisztikai Hivatalban dolgozó *Éltesítő Ödönnek* a mintavétel rejtelmeibe bevezető stúdiómára, vagy *Szalai Sándornak* a szociológia alapvető kérdéseivel foglalkozó, alkalmankénti délutáni összejöveteleire. Érdeklődésemet e foglalkozásokon is elsősorban az a kérdés mozgatta, hogy az ún. kemény adatok mellett megfelelő kérdezéstechnikai módszerekkel miként lehet legjobban biztosítani az ún. lágy, szubjektív jellegű adatok minőségét. Úgy találtam, hogy ennek vizsgálatára különösen alkalmasak azok a kérdőíves adatfelvételek, amelyek az emberek ún. „demográfiai magatartását” – a párválasztást, a gyermekvállalást, a párkapcsolat megszűnését – befolyásoló döntések értékvilágára is kiterjednek, az ízlések, vélemények, attitűdök, normák szerepét is vizsgálják. Ilyen felmérések akkoriban már a Központi Statisztikai Hivatalban is meg-

indultak a longitudinális termékenységi és családtervezési vizsgálatok keretei között. Érdeklődésem ezért az 1960–1970-es évek egyik izgalmas tudományos diszciplínája, a demográfia tudománya felé fordult.

Hogyan kapcsolódott be a demográfiai kutatásokba, kik voltak a tanítómesterei?

Fontos fejleménynek tartom, hogy 1963-ban megjelent az első demográfia tankönyv, a „Bevezetés a demográfiába”, amely nemcsak a szűk szakma, hanem az érdeklődők szélesebb köre számára is használható ismereteket nyújtott. Ugyanilyen fontos volt, hogy 1963-ban megalakult a KSH Népeségtudományi Kutatóintézete (NKI) és megindult a „*Demográfia*” című folyóirat. Ebben az időben a közvélemény is érdeklődni kezdett a gyermekvállalást, a családi életet, gyermeknevelést segítő intézkedések iránt, és a hatalom is fontosnak tartotta az ehhez kapcsolódó problémák kutatását, hogy tisztázza, mit lehetne tenni a születésszám csökkenésének lassítása, megállítása érdekében. Mindez szerencsésen találkozott azzal, hogy a kérdés iránt elkötelezett, nagy tudású szakemberek csoportja foglalkozott demográfiával. E nagy generáció tagjai közül meg kell említeni *Andorka Rudolf*, *Cseh-Szombathy László*, *Dányi Dezső*, *Klinger András*, *Pallós Emil*, *Tamássy József*, *Valkovics Emil* és *Vukovich György* nevét. Amikor 1973–1974-ben a KSH Népeségtudományi Kutatóintézetébe kerültem, személyes kapcsolatba léphettem velük, és sokat tanultam tőlük. De szembesülnöm kellett azzal is, hogy van mit pótolnom a demográfiai ismeretek terén. Nagy szerencsémnek tartom, hogy Valkovics Emil elfogadott magántanulónak. Emil kifejezetten nekem „magánórákat tartott”, a készülő egyetemi tankönyvéhez kapcsolódó házi feladatokkal látott el, így mondhatni, szinte „egyetemi képzést” kaptam tőle a demográfia tudomá-

nyából. „Közkatonaként” részt vehettem abban a folyamatban is, amelynek során a demográfia a magyar tudományos közéletben is fokozatosan önálló diszciplína lett. Az 1970-es évektől sorra alakultak meg különféle, a demográfia önállóságát biztosító bizottságok, testületek. Megalakulásától kezdve hosszú évekig tagja voltam a Magyar Tudományos Akadémia Demográfiai Bizottságának, valamint a Tudományos Minősítő Bizottság Szociológiai és Demográfiai Szakbizottságának, amely korábban a tudományos (kandidátusi és nagydoktori) fokozatok odaitélésével foglalkozott. Különösen sokat tett a demográfia tudományának önállóságáért Andorka Rudolf, az Országos Tudományos Kutatási Alap (OTKA) megalapítója és egykori elnöke, akivel ugyancsak módomban volt együttműködni mint az OTKA demográfiai zsűrijének elnöke, társelnöke, majd később egy ideig az OTKA Társadalomtudományi Kollégiumának tagja. E különböző tudományos-közéleti szerepeket főképp az 1980-as évektől az ezredfordulót követő első évtizeddel bezárólag töltöttem be.

Amikor elkezdte a demográfia területén a kutatásokat, mely témakörök érdekelték leginkább?

Kutatásaimat nem egyedül végeztem, hanem évtizedeken keresztül kollégámmal, munkatársammal, többszörös szerzőtársammal, *Pongrácz Tiborné Hüttl Mariettával* közösen. Jól kiegészítettük egymást. Mariettának mindig is hihetetlenül jó érzéke volt a problémák megragadásához, társadalmi jelentőségük, nemzetközi összefüggéseik gyors átlátásához. Én „szöszmötölőbb” típus vagyok, szeretek hosszan rágódni az adatok jelentésén, interpretálhatóságuk variációin, és ez szerencsés együttműködésnek bizonyult.

1973 és 1974 között népesedési és népese-
déspolitikai témákról kezdtünk kismintás köz-

vélemény-kutatásokat, lakossági megkérdezéseket készíteni, ettől kezdve ezek többé-kevésbé rendszeressé váltak az NKI-ban.

Speciális demográfiai problémákhoz kapcsolódott két olyan, régebben – az 1980-as és 1990-es években – folytatott kutatásunk, amelyek a mai napig a „kedvenceim” közé tartoznak. (Talán azért is olyan kedvesek ezek számomra, mert velük egy időben, 1992-ben Fényes Elek Emlékéremben részesültem, amely a KSH egyik legszebb kitüntetése.) Az első kedvenc a serdülőkorú nők terhességeinek, illetve szülei vizsgálatára volt, amelynek az adta érdekességét, hogy a felvételt követően tíz év múltán, ugyanazon személyekre vonatkozóan megismételtük a vizsgálatot. Az érintettek mintegy 80 százalékát megtaláltuk és újrakerdezhettük őket. Meglepő volt számunkra – és véleményem szerint ez még ma is probléma – a tizenévesek hiányos szexuális ismerete, a szülők és az iskola vonakodása a szinte gyermekkorú lányok felvilágosításától. Adataink jól mutatták azt is, hogy a fiatalok gyermekszülés esetében az átlagosnál jóval nagyobb a kis súllyal és az idő előtt született gyermekek aránya, ami később – ez tíz év után már mérhető volt – a gyerekek iskolai tanulmányaira sem volt hatástalan. Másik kedvenc témánk a házasságon kívüli születések vizsgálata volt. Különleges módszertani feladatnak bizonyult a felvétel mintájának kiválasztása, amit még a gyermekszülések megtörténte előtt kellett elkészíteni. A minta kialakításához és a válaszadók felkereséséhez igénybe vettük az akkori igen hatékonyan működő védőnői hálózatot. Az 1998-ban lezárult kutatás demográfiai háttérének előkészítésében *Kamarás Ferenc*, az adatok matematikai-statisztikai elemzésében *Hablicsek László* segítettek munkánkat. Eredményeink jól mutatták, hogy a házasságon kívül szült nők tipikus csoportjai között kisebbségben maradtak azok, akik további életüket (is) házasságon kívül, élettársi kapcsolat-

ban, esetleg egyedül, önállóan kívánják folytatni. Persze, abban az időben a házasságon kívüli gyermekszülés – ha a tendencia növekedése már látszott is – korántsem volt olyan gyakori, mint manapság. Azt gondolom, mégis van történelmi érdekessége annak, hogy a mintegy másfél évtizeddel ezelőttihez képest miként módosultak a házasságon kívüli gyermekvállalás motívumai.

Ez valóban érdekes, hiszen a társadalom véleménye is rohamosan változott ezekről a kérdésekről. A témakörben megjelenő KSH NKI-kiadványok segítenek megérteni a közvélemény értékítéletének változásait az 1980-as évek elejétől.

Így van. Hiszen amikor ezeket a kutatásokat végeztük, jóllehet már kevésbé ítélték el a házasságon kívüli teherbe esést, de a többség azt elvárta, hogy a gyermek feltétlenül a szülők házasságában jöjjön világra. Manapság ebben a közvélemény jóval megengedőbb.

Szívesen emlékezem arra, hogy az 1990-es években Pongrácz Tibornéval együtt először vehettünk részt egy kisgyermekes szülőkre kiterjedő nemzetközi összehasonlító vizsgálatban, amely Magyarország mellett Lengyelországban, Oroszországban és az akkor még „két” Németországban (NDK-ban és NSZK-ban) folyt. Némileg bennünket is, de a szélesebb szakmát még inkább meglepte, hogy a magyarok gyermekvállalásról, családi munkamegosztásról, női munkavállalásról vallott nézetei jóval hagyományosabbnak bizonyultak, mint a másik három, illetve négy országban. Később ezt más összehasonlító vizsgálatunk is alátámasztotta. Személyes érdeklődésem szempontjából ez a vizsgálat azért volt érdekes, mert ekkor szembesültem először azzal, hogy mit jelent egy angol nyelvű kérdőív hazai terepen történő alkalmazása. Nem elég az, hogy a kérdőív korrekt módon legyen magyar-

ra fordítva. Meg kell találni azokat az apró kérdezéstechnikai és nyelvi megoldásokat, amelyek a magyar nyelven beszélő és gondolkodó lakosság számára jól képesek közvetíteni azt a kontextust, amit az eredeti kérdőív sugall. Itt igazán otthonos közegben mozogtam, hiszen egész életemben izgatott a kérdésszövegek megértésének problémája. Ebben a tekintetben jelentős fejlődést tapasztalok, mert fiatal kutató társaim már gyakorlottak a nemzetközi kutatások világában, jól tudnak nyelveket, és jó érzékük van az eredetileg nem magyar nyelven fogalmazott kérdőívkérdések hazai terepre adaptálásához.

Tudom, hogy az összeírások kommunikációjával is sokat foglalkozott, melyek voltak e munkáinak fontosabb állomásai?

Ehhez a munkához a KSH Népszámlálási főosztálya mellett szerveződő, hajdani ún. ÖSZKO, az Összeírási és Kommunikációs Osztály nyújtott kereteket. Az ÖSZKO, Dobossy Imre osztályvezető és Virágh Eszter irányításával, elsősorban a KSH összeíróinak elméleti és gyakorlati felkészítésével foglalkozott egy igen alaposan kidolgozott tanterv alapján. Az általános képzési program keretében több tankönyv is készült, közülük én az „Adatgyűjtési módszerek, kérdezéstechnikai ismeretek” című tananyag voltam a szerzője, Virágh Eszterrel és Síklaki Istvánnal közösen pedig a „Nem lehet nem kommunikálni” című tankönyvet állítottuk össze. A tankönyvek fogadtatását, használhatóságát folyamatosan ellenőriztük az összeírók foglalkozásain. A képzési program az ezredfordulót követő évekre már közel állt ahhoz, hogy azt a kérdezőbiztosi szakma országosan is akkreditálható tananyagának javasoljuk, de végül az események másképp alakultak.

Az elméleti munkánál talán még többet jelentett számomra az összeírókkal kiépített sze-

mélyes kapcsolat, amikor egy-egy adatfelvétel alkalmából a kérdőívhasználatot a gyakorlatban is kipróbáltuk; a jó és a rossz kérdőí megoldásokat a bemutatkozástól a válaszadók részvételének megnyerésén keresztül a kérdőívkérdések kényesebb részeinek gyakoroltatásáig, és e fázisokat közösen elemeztük. Ez a munka engem mint kutatót is igen értékes tapasztalatokhoz juttatott. Tevékenységemet megkönnyítette, hogy Dobossy Imrével és Virágh Eszterrel már a korábbi években is szoros szakmai kapcsolatban voltunk, közös kutatásokat is végeztünk, amint ezt szerzőtársaként megjelent publikációink is tanúsítják.

A szakirodalomból látom, hogy igen tevékenyen telt a rendszerváltozás utáni két évtized. A KSH Népeségtudományi Kutatóintézete több kutatásában részt vett, melyek voltak ezek?

Az NKI jelenlegi igazgatója, *Spéder Zsolt* vezetése alatt az empirikus vizsgálatoknak két – Andorka Rudolf szóhasználatával élve – „királyi útja” vált egyre intenzívebbé. Az egyik az Intézet nemzetközi kapcsolatainak fellendítése, kiadványainak, valamint a kutatómunkának bekapcsolása a szakma nemzetközi életébe és az összehasonlító vizsgálatok egész sorába. Utóbbiak közül nekem két összehasonlító vizsgálatba volt módom bekapcsolódni. Európában már a 90-es években megkezdődött a „A népesedéspolitika fogadtatása” (Population Policy Acceptance – PPA) című kutatás, amelynek a 2000 és 2003 között végzett kérdőíves adatfelvételében tizennégy európai ország között Magyarországot az NKI képviselte. A kérdőív a családi élettel, a gyermekvállalással, a háztartási és munkahelyi feladatok összhangjának alakulásával, az embereknek a családpolitikával szembeni elvárásaival foglalkozott. A magyar kérdőív kialakítása és a nemzetközi eredmények összehasonlítása jó

terepet adott számomra, annak vizsgálatára, hogy az európai szinten felmerülő problémák milyen módon fogalmazódnak meg a magyar közönség gondolkodásában, értékvilágában. A másik hasonló témákkal foglalkozó vizsgálat-sorozat a „A társadalmi nem és a generációk” (Gender and Generation) címet viseli, és szorosan kapcsolódik az NKI másik jelentős vállalkozásához.

Az intézet kutatásainak másik „királyi útja” ugyanis az „Életünk fordulópontjai” című demográfiai panel megszervezése, elindítása volt. Ez a „követés”, longitudinális vizsgálat-sorozat 2001-ben indult a 18–75 éves népességet mintegy tizenhatezer fővel reprezentáló országos mintán. Az akkor felkeresett személyeket három-négy évenként újra kérdeztük, 2012–2013-ban a negyedik kérdezési hullám zajlott le. (Az öregedés, a halálozások miatti mintavesztéséget természetesen újabb és újabb „fiatal minták” pótolják.)

Így egy igen értékes adatállomány van kialakulóban az NKI-ban. Ez már jelenleg is több mint egy évtizedes időtávra visszanzve tudósít arról, hogy az élet olyan jelentős fordulópontjai, mint a párkapcsolat létrejötte, a családalapítás, gyermekvállalás, a család felbomlása, a munkavállalás, a nyugdíjba lépés, az egészségi állapot alakulása milyen változásokat hoz az emberek életútjában, továbbá, hogy közérzetük, véleményeik változásai hogyan követik mindezeket a folyamatokat. Az egyes kérdezési hullámok tartalmi kialakításában és szervezésében – különösen az első és a második hullám esetében – magam is aktívan részt vettem. Az egyik fő feladatom – koromnak megfelelően – a nyugdíjba lépés, az idős, illetve öregkor sajátosságainak, életmódváltozásainak vizsgálata volt. Talán ez vezetett ahhoz is, hogy szívesen vállaltam néhány évig annak a projektnek a gondozását, amelynek keretében a KSH két évente jelentette meg az ún. „Ezüstkor”-kiadványokat, részben zsebkönyv formá-

ban, részben pedig az ugyancsak kétévente megrendezett, öregedéssel foglalkozó konferenciák előadásait tartalmazó tanulmánykötetekben.

Beszéljünk arról a kutatásról, melyet 2008 és 2011 között Pongrácz Tibornéval közösen végeztek, ahol – úgy látom – módjukban volt összefoglalni a több évtizedes demográfiai magatartások vizsgálatának eredményeit.

Pongrácz Tiborné 2008-ban elnyert OTKA-pályázata azt célozta meg, hogy egy adatfelvételen a lehetőség szerint változatlan formában vagy az interpretációt csekély mértékben befolyásoló, kissé módosított formában újra feltegyük a 18–50 éves korú férfiak és nők 1600 fős országos reprezentatív mintáján az ún. családi értékekkel kapcsolatos számos témakör – 1970-es évek óta alkalmazott – kérdőívkérdéseit. Célunk annak bemutatása volt, hogyan követhető a főbb demográfiai tendenciák és a közmegítélés változása, ezek „együtt járása” több évtizedes vizsgálatainkban. A 2009-ben lebonyolított adatfelvétel eredményeit a korábbi vizsgálatok 18–50 éves korúak almintáinak eredményeivel vetettük egybe.

Három témakörre összpontosítottunk. Közülük egyik a párkapcsolat létrejöttét/megszűnését érintő magatartási normák változásának vizsgálata volt, a másik a közvélemény gyermekszám-preferenciáinak alakulása, míg a harmadik a nemi szerepmegosztásról, a családi élet és a munka összhangjáról alkotott vélemények változásainak elemzése. Én főképp az első és a második témakörrel foglalkoztam, míg Pongrácz Tibornéval közösen a harmadik témát, a nemi szerepmegosztásról, a családi élet és a munka összhangjáról alkotott vélemények változását elemeztük. Megjegyzem, hogy a pályázat végrehajtásában más kollégák is részt vállaltak, a megszületett tanulmányok az NKI Kutatási Jelentések kiadvány-

sorozatának 91. számában jelentek meg. A pályázat ugyanis kitért egy – legalábbis nálunk újszerű – téma, a családi férfiszerepek, apaképek vizsgálatára is (ezzel Spéder Zsolt foglalkozott). Ugyancsak érintette a pályázat a kisgyermekes anyák munkába lépésének problémáját (ezt *Blaskó Zsuzsa* vizsgálta), és egy másik témaként bemutatta azt is, hogy miként ítéli meg a közvélemény az ország népesedési helyzetét, milyen lehetőségeket látnak az emberek a népességfogyás csökkentésére, illetve miként vélekednek a családpolitikai támogatásokról (ezt a kérdéskört *Kapitány Balázs* elemezte).

Kutatásaik eredményei közül érdekesnek tartom annak vizsgálatát, hogy a demográfiai magatartások változásában a közvélemény értékítéletének vagy az egyéni élethelyzetek alakulásának van-e nagyobb szerepe?

Ez valóban fontos kérdés. Vizsgálataink egy része a magánélet eseményeiről szóló véleményeket, az ún. közmegítélés megismerését tekintette céljának. Felfogásunk szerint a magánéletet (mint például a párkapcsolat létesítése, a családforma megválasztása, a gyermekvállalás) bizonyosfajta társadalmi presszióként működő – közmegítélés kíséri. Ez arra készíti az embereket, hogy valamiképpen igazodjanak az elvárásokhoz. A környezet, az ún. mikrocsoportok pozitív kontrollal fogadhatják a közmegítélés szerint helyeselt normák követését, vagy rosszhallhatják az attól eltérő magatartásokat. Ez azonban nem csak utólag fejt ki hatását. Az emberek többnyire ismerik a cselekedeteikre várható választ, és döntéseikben eleve mérlegelik azokat. A demográfiai trendek azt mutatják, hogy a közmegítélés kontrollszerepe némileg csökkent, de ahhoz nemigen fér kétség, hogy az ember tevékenységét, döntéseit, értékítéleteit a mai modern világban is a közvélekedés, a szomszédok, a munkatár-

sak, a barátok helyeslése vagy rosszallása irányítja. Az egyéni döntések, magatartások valamint a közmegítélés normái között mai tudásunkkal nem lehet megállapítani egyértelmű oksági kapcsolatot vagy rangsort. Már csak azért sem, mert e két terület valójában még az intézményes hatásokkal is kiegészül. Ilyen például az oktatási rendszer, a munkaerőpiac, a családtámogatások, az egészségügy szerepe. Így a radikális demográfiai változások okainak valójában nem kettős, hanem hármas egységét lehet feltételezni. Publikációinkban azonban számos példát, elemzést mutattunk be arról, hogy a tényleges demográfiai trendek (mint például az élettársi kapcsolatok, a nem házas születések terjedése, a gyermekvállalás mérséklődése, illetve későbbi életkorra halasztása) hosszabb időtávon gyakorlatilag „együtt mozognak” a hagyományos közmegítélés lazulásával, engedékenyebbé válásával.

Az is meglepő eredmény, hogy a felnőttek szerint a gyermekek sem tulajdonítanak nagyobb jelentőséget annak, hogy a szüleik házasok-e, vagy élettársi kapcsolatban élnek. Valóban így van ez?

A 18–50 éves korúakat kérdezve nem kétséges, hogy ez meglehetősen gyors és alapos változás a közmegítélésben. Míg 1991-ben csak 20, 1997-ben 31, 2009-ben már 52 százalékuk értett egyet azzal az állítással, hogy „*ma már a gyerekeknek mindegy, szüleik házasokban, vagy élettársi kapcsolatban élnek-e*”. Ez az álláspont természetesen nem független a házasságon kívüli gyermekszülések dinamikus növekedésétől, de attól sem, hogy napjainkra a közgondolkodásból határozottan eltűnt az ún. „törvénytelen gyerek” stigmája. Jellemző, hogy ezt két évtizeddel ezelőtt még másképp tapasztalták. A sokgyermekes családok 1991-ben lezárt vizsgálatában tapasztalhattuk, hogy az élettársi kapcsolatban élő három- és több-

gyermekes anyák egy része házasságkötését akkorra időzítette, amikor a legidősebb gyermek óvodába, iskolába lépett, mégpedig azért hogy megvédjék őt az intézményes szinten is érzékelhető diszkriminációtól. Gondolom, ilyen magatartással a gyermekintézményekben ma már nemigen találkozunk.

Áttérve a gyermekvállalás kérdésére, a szakirodalomból látom, hogy sokat foglalkoztak azzal a világszerte vizsgált kérdéssel, hogy vajon a gyermekvállalást mennyire befolyásolják az egyének, családok anyagi körülményei, illetve egyéb élethelyzeteik? Milyen változásokat tapasztaltak ebben a kérdésben?

Az nem vitatható, hogy az anyagi körülmények (kibővítve ezt a lakáshelyzettel, a munkahely meglétével, azzal, hogy mindkét szülő rendelkezék valamilyen jövedelemmel) alapvetően meghatározók. Azt azonban megfigyelhettem, hogy az egyéni törekvések, az ún. individuális értékek előtérbe kerülése csökkenően hat a „még vállalható” gyermekszám meghatározására. A társsal, partnerrel való harmónia, a sikeresség a munkában, az elismertség, a tisztelet kivívása a családon kívül, a lehetőség egyéni céljaink, hobbijaink megvalósítására, a havi egy-két alkalommal megvalósítható szórakozás fontossága voltak azok, amelyeket posztmateriális értékeknek nevezünk. Már 2000 és 2009 között is mérhető volt, hogy ezek értéke valamelyest nőtt az emberek gondolkodásában, jóllehet az ún. materiális értékek súlyát nem előzték meg. Mégis észrevehető volt, hogy akiknél az ún. posztmateriális értékek súlya növekedett, egyértelműen kevesebb számú gyermekkel vélték e célokat megvalósíthatónak, mint egy évtizeddel korábban. Rá tudtam mutatni arra is, hogy azok a feszültségek, amelyek a tervezett és tényleges gyermekszámnak a családtervezési vizsgálatok révén kimutatott különbségéből adódnak, a köz-

gondolkodás egészében is tetten érhetők. A családok számára általánosságban ideálisnak tartott gyermekszám átlaga, amely mindig magasabb a ténylegesen megvalósított gyermekszámnál, több évtized óta folyamatosan csökken – követve a termékenység tényleges csökkenését. Újabb a közgondolkodásban még a gyermekenélküliség tudatos választása iránt is bizonyos megértés mutatkozik.

Szintén a gyermekvállalás kérdéséhez kapcsolódik a nők és a munka világának témája, a női és férfi szerepek megítélése a magyar társadalomban. Mik a tapasztalatai erről a témáról?

Ez volt az egyik fő témája a korábban már említett, ún. PPA-vizsgálatnak, amelyben tizenny európai ország között („Népesedés és család” címmel) Magyarország is részt vett. Világosan kimutatható volt, hogy a tradicionális családi értékek elfogadását vagy elvetését vizsgáló kérdésekben a magyar lakosság gondolkodása erősen hagyományos. Jellemzően ilyen például a családi munkamegosztás kérdése. Magyarországon vallották legtöbben, hogy „a férj feladata a pénzkeresés, a feleségé a háztartás és a család ellátása”. Pongrácz Tibornéval közösen végzett elemzésünk szerint azonban az ezredfordulóhoz képest az azt követő egy évtized során figyelemre méltó változások tapasztalhatók. Modernebb, munkaorientáltabb lett az értékrend, az egyértelmű család- és gyermekközpontúság némileg visszaszorult. Elfogadottabbá vált, hogy az anya kereső tevékenysége nem feltétlenül negatív hatással a gyermek fejlődésére, sőt a kétkeresős családmódel a család stabilitása, biztonsága szempontjából elengedhetetlen követelmény. A tradicionális szerepmegosztás ideologikus töltésű témakörében ugyanakkor ma még erőteljes a hagyományos szemléletmód továbbélése. Megjegyzem, Magyarország esetében némiképp vitatható az, hogy a nemzetközi in-

terpretációknak megfelelően a női munkavállalást, a kétkeresős családmódel preferálását, mint a modernizálódás felé elmozdulást értelmezzük. Azt gondolom, hogy a mai munkaerőpiaci helyzetben az egyébként tradicionális értékeket vallók nagy része inkább kényszernek, semmint egy modernebb szemlélet elfogadásának érzi azt, hogy mindkét szülő rendelkezzen önálló keresettel.

Az internetes kérdezéstechnika terjedésével jelentős módszertani változásnak vagyunk tanúi. Szakmai munkájának jelentős része a szövegértés problémája körül mozgott. Vajon ennek a paradigmaváltásnak melyek lesznek a hatásai?

Számomra ez valóban új helyzet. A racionalitás, az ökonomikusság nagyon is megkívánja, hogy megszabaduljunk a papír alapú kérdőívek óriási tömegének terhétől, szállításától, ellenőrzésétől, tárolásától. Az út egyértelműen a kérdőívek internetes kitöltése, illetve amennyiben személyes kérdésre van szükség, a válaszok laptopon történő rögzítése. Joggal számolhatunk ma már azzal is, hogy maga a vizsgálandó közeg is sokat változott, fejlődött e téren, hiszen nemcsak a fiatalok, hanem a középkorúak, sőt még a nyugdíjas korosztály körében is egyre terjed az internethasználat. Valószínűnek tartom, hogy az ún. kemény adatok internetes összegyűjtése terén könnyebb a helyzet. Keveset tudunk azonban arról, hogy milyen konkrét élethelyzetekben történik a vélemények, attitűdök iránt érdeklődő kérdések webes megválaszolása. Vajon valóban a minta szerint kiválasztott személy válaszait kapjuk-e meg, vagy valamiféle „kollektív családi kitöltés” („mit is válaszoljunk erre a kérdésre?”) eredményét? Egyelőre arról is nehéz meggyőzni a leendő válaszadókat, hogy az interneten beküldött véleményeik a feldolgozás és az összesítés után névtelenné

válnak. Laptopos kérdés esetén pedig a megkérdezett személy nem látja maga előtt a kérdőívet, mint korábban, a papírkérdőívek idején, és nehéz követnie, hogy az összeíró mi mindent és hogyan rögzít a kis gépezetébe. Ezek mind vizsgálatokra, tapasztalatszerzésre várnak. Negyven évvel fiatalabban – azt gondolom – szívesen folytatnám a régi „hobbimat”, mivel nagyon is érdekel, hogy a szövegmegértés nehézségei miként hatnak az in-

ternetes kérdőív-kitöltéssel szerzett adatok minőségére. Erre azonban már csak egy fiatalabb generáció tud majd válaszolni.

Köszönöm a beszélgetést, jó egészséget kívánok!

Dr. Lakatos Miklós,

a KSH szakmai főtanácsadója

E-mail: Miklos.Lakatos@ksh.hu

Beszámoló az MTA Statisztikai és Jövőkutatói Tudományos Bizottságának Statisztikai Tudományos Albizottságának üléséről

Az MTA Statisztikai és Jövőkutatói Tudományos Bizottságának Statisztikai Tudományos Albizottsága (STAB) 2013. február 28-án tartott ülést a KSH Keleti Károly-termében.

Először *Hunyadi László*, a STAB elnöke ismertette a napirendet, és javasolta, hogy tekintettel azok fontos és közérdekű vonatkozásaira, a résztvevők először az Egyéb napirendi pontokat tárgyalják. Három ilyet említett. Először beszámolt arról a levélváltásról, ami *Vukovich Gabriella*, a KSH elnöke és közötté zajlott le. Ebben Hunyadi László felajánlotta a STAB segítségét a készülő új statisztikai törvény előkészítése során a törvényalkotóknak azzal, hogy a statisztikai gyakorlat, a hivatalos statisztika és a statisztikai tudomány viszonyában, valamint a kutatói adathozzáférés kérdésében szeretnék közvetíteni a statisztikai tudomány álláspontját. A hivatal elnöke köszönettel vette a segítséget, és közölte, hogy a törvényalkotás megfelelő fázisában igénybe fogják azt venni. A következő egyéb napirendi pont a szűk körű szakmai rendezvények kérdése volt. A Pécsi Tudományegyetem Közgazda-

ságtudományi Karának Módszertani Intézete folytatja az ökonometriai tárgyú vitanapokat, *Rudas Tamás* egy idén szervezendő, szociológiai-statisztikai kiskonferenciát említett, *Marton Ádám* pedig egy szakmai beszélgetést kezdeményezett a fogyasztói árindexek és az infláció statisztikai kérdéseiről. A STAB elnöke ezt követően beszámolt arról, hogy az albizottság honlapjának készítése jól halad, egy próbaváltozata rövidesen már tesztelhető lesz.

A második (a kiadott program szerinti első) napirendi pont a statisztika nyelvéről szóló kerekasztal-beszélgetés volt. Ehhez a STAB elnöke rövid vitaindítót készített, amelyben exponálta a kérdés fontosságát, javaslatot tett e széleskörű téma szűkítésére, és felvetette egy ezzel a kérdéssel tartósan foglalkozó munkabizottság létrehozásának ötletét.

A vita a következő témákat ölelte fel. Kell-e, érdemes-e a statisztika nyelvével foglalkoznunk? A statisztika alkalmazóinak (statisztikusoknak, kutatóknak, hétköznapi embereknek stb.) milyen szintű statisztikai nyelvismeretre van szüksége? A különböző csoportoknak mit

jelent a statisztikai nyelv? Mit lehet tenni a statisztikai nyelv korrekt használatáért? Egyáltalán mit értünk a statisztikai nyelvezet korrekt használatát alatt? Fontos-e a statisztikai kifejezések egységes helyesírása? Mi a résztvevők véleménye a közlés lehetséges módozatairól, van-e újabb ötletük erre? Milyen más nyelvi problémák merülnek fel a statisztikával kapcsolatosan? Milyen szakterületek bevonását tartják fontosnak és milyen sorrendben? Helyesnek tartják-e a munkabizottság létrehozását, és kik szeretnének részt venni egy ilyen munkabizottságban? Mi lehet a munka eredménye: egy fogalomtár, szinonimaszótár vagy valami hasonló, ami lehetővé teszi a közlekedést a különböző területek és műhelyek által használt nyelvek közt?

A vitában először *Herman Sándor* kérdés-ként vetette fel, hogy mi lehet a magyar statisztikai nyelv gyökere: az angol, a német, esetleg a latin? Beszélt a magyar statisztikai nyelv, valamint az angol és a német nyelvek megfelelő kapcsolatáról. Példákkal illusztrálta a felmerülő nehézségeket (egyes kifejezésekre nincs megfelelő magyar nyelvű terminus, míg egyes idegen nyelvi kifejezéseknek a magyarban több változata létezik, illetve a magyarban néha egy kifejezést használunk különböző idegen nyelvi kifejezésekre) és taglalta, hogy miként lehetne (értelmes fordítással, szűkítéssel, az eredeti meghagyásával) a magyar szaknyelvet egyértelművé és kommunikációra alkalmassabbá tenni. *Pukli Péter* üdvözölte a kezdeményezést és ennek kapcsán kitért a statisztikai kultúra általános kérdéseire, a statisztikai alfabetizmus felszámolásának fontosságára, az oktatás szerepére. Hangsúlyozta a hivatalos statisztika és a KSH szerepét ebben a kérdésben, említvén az ott már régóta folyó munkát (metaadatbázis), az informatikai háttérrel, valamint a fejlett infrastruktúra szerepét. Véleménye szerint a hivatalos és az akadémiai statisztika összefogása, együttműködése, egy kö-

zös projekt lehet a követendő út. A munka eredménye több kellene, hogy legyen, mint egy szinonimaszótár, illetve mint egy ajánlás, szerepet kell, hogy kapjon benne az akadémiai és a hivatalos statisztika, valamint a média is. Kitért arra, hogy a középiskolai oktatásban részt vevő tanárok sokat tehetnek a program sikeréért. A statisztikaalkalmazás szintjeit, illetve a kapcsolódó nyelvismeretet megpróbálta különböző szinteken definiálni (1. szint – tudunk statisztikákat olvasni, 2. szint – értő hozzáállás, azaz véleményt tudunk formálni, 3. a professzionális szakma szintje). Fontosnak tartja a gazdaságstatisztika nyelvi problémáinak kezelését, és kinyilvánította részvételi szándékát egy ilyen munkacsoport tevékenységében. *Szilágyi György* hozzászólásában azt emelte ki, hogy a statisztika magyar nyelve elég jó és precíz, ennek ellenére továbbfejlesztése fontos kérdés. Mivel a statisztika mértékadó nemzetközi nyelve az angol, törekedni kell a kétnyelvűség, az egyértelmű megfeleltetések megteremtésére. Ennek ellenére határozottan kiállt az önálló magyar szaknyelv ápolása mellett, jóllehet fontosnak tartja, hogy az oktatásban a hallgatók minél előbb és alaposabban megismerjék az angol szaknyelvet is. *Éltető Ödön* ugyancsak fontosnak tartja a statisztika nyelvének egységesítésére irányuló törekvéseket, hiszen ez szükséges ahhoz, hogy a különböző szakterületek szakemberei jól megértsék egymást. *Kerékgyártó Györgyné* a fogalmi tisztázás fontosságáról beszélt. Példákkal illusztrálta, hogy néhány fontos statisztikai szakkifejezés az egyes szakterületeken, a sajtóban és a köznyelvben mennyire más értelmezést kap. Megemlítette a statisztikai kifejezések helyesírása körüli bizonytalanságot, problematikát. Ő is hangsúlyozta a középiskolák szerepét, valamint azt, hogy a nyelvi kérdések tisztázása a középiskolai oktatás fejlesztéséhez is segítséget nyújthat. *Szép Katalin* arra tért ki, hogy a KSH Módszertani főosztá-

lyán jelenleg is létezik egy kétnyelvű metafogalomtár, ami egyes statisztikai módszerekre is kiterjed; a jelen erőfeszítéseknek ezért elsősorban a módszertani fogalmak és elnevezéseik dokumentálására kellene irányulnia. Problémaként említette, hogy még a módszertani tankönyvek szakkifejezései is gyakran eltérnek egymástól. Javasolta a STAB készülő honlapjának felhasználását a munka eredményeinek bemutatására.

Besenyei Lajos egyetértett a téma megvitatásával. A nyelvhasználat négy szintjéről (a professzionális szakemberek, az oktatásban statisztikát tanulók, a közvetítő szerepet betöltő média, illetve a hétköznapi emberek szintje) beszélt, valamint arról, hogy ezeken a szinteken más és más, eltérő kezelést igénylő problémák jelentkeznek. Hangsúlyozta, minden nyelvet, így a statisztika nyelvét is tanulni kell, és ez a folyamat természetesen a szakma tanulását és jobb elsajátítását is jelenti. *Sándorné Kriszt Éva* szintén kiemelte a nyelv és a statisztikatanulás szoros kapcsolatát, illetve a felhasználói oldal fontosságát. Úgy vélekedett, nem szabad irreális célokat kitűzni: ha eljutunk egy jól használható fogalomtárig, már az is komoly eredmény kiváltképp akkor, ha figyelembe vesszük, hogy a statisztika a legfontosabb, és legszélesebb körben elterjedt kvantitatív diszciplína. *Lencsés Ákos* szerint nem csupán szótárat, de magyarázó szótárat kellene összeállítani. Kiinduló pontként javasolta a Nemzetközi Statisztikai Intézet soknyelvű szótárát. Fontosnak tartja, hogy az eredmény nyilvános legyen, melyre a STAB honlapját megfelelő fórumnak tekintené, de szerinte egy ilyen kezdeményezés, amennyiben eredményekig jut, ennél többet is érdemel: a tanulóhoz is el kellene juttatni, illetve akár önálló kötet tárgya is lehet. Véleménye szerint a KSH könyvtárban többen nyitottak lennének a közreműködésben. *Kondora Cosette* utalt arra, hogy a *Statisztikai Szemle*ben is sok gond van

a fogalmakkal és az egyes kifejezések helyesírásával, ezért a folyóirat is érdekelt egy ilyen munkában. Gondot okoz az új fogalmak idegen nyelvi használatának terjedése, és nagyon kellene a szakma állásfoglalása azok magyartásában. Véleménye szerint azonban az ISI szószeretetének magyar változata hiányos. *Lencsés Ákos* meggyőződése, hogy helyesírási kérdésekben az MTA nyelvészeinek véleményét el kell fogadni és problematikus (például egy fogalom félreértéséből fakadó) állásfoglalások esetén párbeszédet kell kezdeményezni. *Vargha András* egyetértett az alapkérdés fontosságával, hiszen több tudományág használja lényegileg ugyanazt az eszköztárat, és gyakran ugyanarra a fogalomra más kifejezéssel. Hozzászólásában további kérdéseket fogalmazott meg. Szótár legyen vagy enciklopédia? Hány nyelvű legyen ez az összeállítás? Nyitott legyen, mint a wikipédia, vagy szűken szerkesztett? Ki legyen az összeállítás célcsoportja? Szükséges-e, legalább az ajánlás szintjén, az egységesítés? A felszólaló utalt a forgalomban lévő szoftverek nyelvformálásban betöltött fontos szerepéről, és megemlítette a rövidítésekkel kapcsolatos problémákat is.

Marton Ádám a konkrét példákat hiányolta a korábbi hozzászólásokból. A problémát fontosnak tartja, szerinte az angol nyelvű szakkifejezésekből kellene egy ilyen munka során kiindulni, bár az angol nyelv sem egységes és nem mindig egyértelmű. Felhívta a figyelmet arra is, hogy az egyes kifejezések értelmezéséhez gyakran komoly felhasználói érdekek fűződnek, amik nehezítik a helyes és szabatos fogalom-, illetve szóhasználatot. *Kovács Péter* nem lát arra esélyt, hogy valódi egységesítés legyen e téren, de a fogalmi szótár összeállítását mindenképpen támogatja. Hangsúlyozta, hogy a különböző szinteken más és más szótárra van szükség, és az angol kifejezésekkel a magyar megfelelőket mindenképpen célszerű lenne összehasonlítani. Titkári minőségében

felolvasta a távol lévő *Zádor Márta* írásos hozzászólását, aki a mindennapi tudás szintjéhez, az alapintelligenciához tartozónak tartja a statisztikai alapfogalmak ismeretét, és a korai, középiskolai statisztikai alapozás javítását tartja kiemelkedőnek. Már középiskolában is szükségesnek véli az olyan alapfogalmak (például a GDP összetevőinek, a bért, a foglalkoztatást, az árszínvonalat meghatározó kategóriák) oktatását, amelyek a hétköznapi tájékozódáshoz, az újságolvasói szintű, állampolgári tájékozódáshoz nélkülözhetetlenek.

A hozzászólások után a levezető elnök összefoglalta az elhangzottakat, ám dr. Pukli Péter hiányolta abból annak megfogalmazását, hogy a STAB ebben a kérdésben szoros együttműködésre törekedjék a KSH-val. *Laczkó Éva*, aki mind a STAB-nak, mind a KSH elnökségének tagja, vállalta, hogy az együttműködés kérdését és lehetséges módját felveti a KSH legközelebbi elnökségi ülésén. *Telegdi László* némi éllel megjegyezte, hogy az akadémiai bizottság nem alárendelt szerve a KSH-nak, az, ahogy a STAB értelmezi a statisztikát, messze túlnyúlik a hivatal területén és kompetenciáján. Ezt követően többen úgy vélekedtek, hogy felesleges a hivatalos és az akadémiai statisztika között kiélezní a viszonyt, hiszen a STAB-nak szüksége van a KSH-ra, a hivatalnak pedig szüksége van a tudomány támogatására. Bár a múltban a KSH kétségtelenül dominálta, sőt olykor részleteiben is befolyásolni akarta az akkori statisztikai bizottság tevékenységét, mára a STAB függetlenné vált a KSH-tól, sem tagsága, sem vezetősége nem kötődik szorosan a hivatalos sta-

tisztikához, így jó esély van arra, hogy az együttműködés egyenlő felek kölcsönösen előnyös tevékenysége lesz ebben a témában.

Utolsó napirendi pontként Kovács Péter tartott előadást a közgazdaság- és gazdálkodástudományban alkalmazott statisztikai eszköztárról. Ez az első eleme volt egy olyan tervezett sorozatnak, ami a különböző tudományterületeken alkalmazott statisztikai módszereket mutatja be. Az előadó a tudományág jellemző vonásaiból, területeiből, irányzataiból vezette le az alkalmazott, elsősorban modellezési célú statisztikai eszközöket. Előadását, melynek anyagát a rövidesen elkészülő honlapon bárki megtekintheti, színes és érdekes példákkal illusztrálta.

Az ülés zárásaként a levezető elnök köszönetet mondott a résztvevőknek aktivitásukért. Megkérte az albizottság tagjait, hogy akik részt kívánnak venni a megalakítandó munkacsoportban, elektronikusan jelezzék a vezetőség felé, ami gondoskodni fog a munkacsoport felállításáról. Végül jelezte, hogy a következő általános ülésre várhatóan 2013 májusában-júniusában kerül sor, amelynek fő témája a statisztikai publikációk és a minősítés feltételei lesznek, segítendő ezáltal a STAB tudományos fokozatra aspiráló tagjainak felkészülését.

Hunyadi László,

a STAB elnöke

E-mail: hunyadi44@gmail.com

Kovács Péter,

a STAB titkára

E-mail: pepe@eco.u-szeged.hu

Beszámoló a társadalomstatisztikai felhasználói fórumokról

Dr. Németh Zsoltnak, a Központi Statisztikai Hivatal (KSH) elnökhelyettesének a kezdeményezésére a hivatal a területi főosztályok bevonásával társadalomstatisztikai felhasználói fórumokra invitálta az érdeklődőket. A tavaly november és idén március között lezajlott rendezvénysorozat célja az volt, hogy a potenciális felhasználók, azon belül is kiemelten az egyetemek és a kutatók, betekintést nyerjenek a KSH egyedülálló társadalomstatisztikai adatgyűjteményébe, s ezen keresztül képet kapjanak hazánk társadalmi folyamatairól, a magyar társadalom állapotáról, szokásairól.

A fórumokon a hivatal munkatársai 10-10 perces előadásokban mutatták be a társadalomstatisztika főbb területeit és az ezekhez kapcsolódó legfontosabb adatgyűjtéseket. A résztvevők tájékoztatást kaptak arról is, hogy milyen hozzáférési formák segítségével használhatják a hivatal adatbázisát.

A rendezvényekhez kapcsolódóan a KSH együttműködési lehetőséget kínált az egyetemek számára annak érdekében, hogy szakembereik, hallgatóik kutatásokat és elemzéseket végezhesenek az adatbázis felhasználásával. A fórunsorozat keretében a hivatal együttműködési megállapodást kötött a Miskolci Egyetemmel és a Szegedi Tudományegyetemmel.

A helyszíneken nagy érdeklődés fogadta a hivatal munkatársait, megkönnyítve ezzel a szakértőkből álló csapat számára a korai indulással és a többórás utazással járó feladatot. A szakértői gárda és a témák a következők voltak: Módszertani kérdések, kutatószoba, távoli hozzáférés (dr. Németh Zsolt); Népszámlálási tájékoztató koncepció (*Nagy Beáta, Vereczkei Zoltán* (Módszertani főosztály)); Népmozgalmi adatok (*Kovács Marcell* (Népszámlálási főosztály)); Munkaerő-felmérés (*Branyczikiné Géczy*

Gabriella, Kamarás Ferenc (Népesedési és szociális védelmi statisztikai főosztály)); Oktatás (*Lindnerné dr. Eperjesi Erzsébet, Váradi Leventéné, Kadlecik Roland* (Életmód-, foglalkoztatás- és oktatásstatisztikai főosztály)); Háztartási Költségvetési és Életkörülmények Adatfelvétel (*Janák Katalin, Lukácsné Szabó Gerda, Branstetter Tímea* (Életmód-, foglalkoztatás- és oktatásstatisztikai főosztály)); Időmérleg (*Ménesi Éva, Huszár Ákos* (Életmód-, foglalkoztatás- és oktatásstatisztikai főosztály)); A lakosság utazási szokásai (*Grábics Ágnes* (Miskolci főosztály)); A lakosság utazási szokásai (*Kovács Csaba, Gilyán Csaba* (Külkereskedelemszatisztikai főosztály)); Európai lakossági egészségfelmérés (*Tokaji Károlyné, Boros Julianna* (Népesedési és szociális védelmi statisztikai főosztály)); Nonprofit adatfelvétel (*Nagy Renáta, Sebestény István* (Népesedési és szociális védelmi statisztikai főosztály)); Lakáspiaci tranzakciók (*Székely Gáborné, Lehoczki-Krsjak Adrienn* (Népesedési és szociális védelmi statisztikai főosztály)); Tájékoztató adatbázis (dr. *Bódiné Vajda Györgyi* osztályvezető (Tájékoztatósi főosztály), *Antoni Soma* osztályvezető (Informatikai főosztály)).

A sorozat 2013. március 12-én megtartott budapesti záróeseményét dr. *Kertesi Gábor* előadása színesítette, aki a kutatószobában hozzáférhető adatok felhasználásával készült kutatási eredményeiről számolt be a közönségnek. Dr. Németh Zsolt a rendezvény zárásaként röviden értékelte a felhasználói fórunsorozatot, és megköszönte a meghívottak részvételét, valamint a munkatársaknak a kiváló előadásokat.

Keszler Ágnes,

a KSH osztályvezetője

E-mail: Agnes.Keszler@ksh.hu

Hírek, események

Kinevezés/megbízás. *Dr. Vukovich Gabriella*, a Központi Statisztikai Hivatal elnöke 2013. március 28-ai hatállyal *Kocsis-Nagy Zsoltot* a Szegedi főosztály főosztályvezetőjévé, míg 2013. április 1-jétől *Pál Tamarát* – osztályvezetői beosztásának változatlan betöltése mellett – az Elnöki főosztály főosztályvezető-helyettesévé nevezte ki; 2013. április 1-jei hatállyal, öt éves határozott időtartamra *Fülöp Ágnes*t a KSH Könyvtár főigazgatói vezetői teendők ellátásával bízta meg.

2013. április 1-jei hatállyal *dr. Laczka Éva* gazdaságstatisztikai elnökhelyettes *Imre Magdolnát*, *Cech Vilmost*, *Gilyán Csabát* és *Lovászné Skach Editet* a hivatal Szolgáltatás- és külkereskedelemszatisztikai főosztálya osztályvezetőivé, *dr. Németh Zsolt* társadalomstatisztikai elnökhelyettes pedig *Rácz Attilát* a Szegedi főosztály osztályvezetőjévé nevezte ki.

A 2011. évi népszámlálás végleges adatairól tartott sajtótájékoztatót a Központi Statisztikai Hivatal 2013. március 28-án a Keleti Károly-teremben. *Dr. Vukovich Gabriella*, a KSH elnöke a népszámlálás során gyűjtött adatokat egyedülálló értéknek nevezte, mert azok ismerete egyaránt fontos az üzleti élet képviselőinek, a közvélemény- és piackutatóknak, az elemzőknek, valamint a politikusoknak. Az adatok kétharmadát a kérdezőbiztosok vették fel, de minden ötödik kérdőívet a lakosság már online töltött ki, a fennmaradó hányadot pedig a levélben válaszolók tették ki. Ez az ún. „többszörös módszer” megnevezette az adatfeldolgozást, a papíralapú válaszok összesítése egy évig tartott.

A kínai Jiangsu tartomány statisztikai hivatalának főigazgató-helyettese és vezető munkatársai folytattak konzultációt a KSH-ban

2013. április 16-án. A tárgyalásokat a magyar fél részéről *dr. Laczka Éva* gazdaságstatisztikai elnökhelyettes vezette, nyitóelőadásában bemutatva a az adatgyűjtésekkel kapcsolatosan nemrég bevezetett, hivatali stratégiai fejlesztéseket. A szakértői előadásokat a Vállalkozásstatisztikai, a Központi adatgyűjtő, az Életmód-, foglalkoztatás- és oktatásstatisztikai, az Informatikai, illetve a Tájékoztatósi főosztály munkatársai tartották.

„Infláció régen és ma” címmel került sor a Magyar Statisztikai Társaság Gazdaságstatisztikai Szakosztályának és az MTA Statisztikai és Jövőkutatási Tudományos Bizottság Statisztikai Albizottságának szakmai rendezvényére 2013. április 25-én a KSH Árvay János-teremben. *Dr. Marton Ádám* „Infláció és pénzromlás – hosszú távon” címmel, *Mináry Borbála* pedig napjaink inflációs számításai kihívásairól tartott előadást. Részletes beszámolót olvasóink a következő számban olvashatnak.

A KSH Könyvtár és a T-Systems Magyarország Zrt. rendezett teadélután a könyvtár Bibó István-teremben 2013. április 23-án „Digitális, mobil és okos rendszerek – könyvtár a zsebünkben” címmel. A rendezvényen először *Fülöp Ágnes*, a KSH Könyvtár főigazgatója mondott köszöntőt, majd a következő előadások hangzottak el: 1. QR-kód a könyvtárban és egyéb „mobilságok” (*Horváth Zoltánné*, a T-Systems Magyarország Zrt. szakértékesítési menedzsere); 2. Infonear az Origón. Olvassuk vagy hallgassuk? (*Lesku Gergő*, a T-Systems Magyarország Zrt. szakértője); Mobilibri (*Kótai Katalin*, a T-Systems Magyarország Zrt. senior projektvezetője); ADVISE (*Habók Lilla*, a T-Systems Magyarország Zrt. junior elemző-tervezője); Elektro-

nikus és digitális szolgáltatási rendszer technológiai és integrációs háttere (*Szemereki Viktor*, a T-Systems Magyarország Zrt. kompetencia-központ-vezetője); Könyvtár és levéltár – rokonságok és azonosságok egy nagyprojekt tük-

rében (*Kótai Katalin*); Mobil mánia – beszámoló az OCLC EMEA strassbourgi konferenciájáról (*Horváth Zoltánné*). A moderátor *Horváth Zoltánné* volt.

A Nemzetközi Statisztikai Intézet (International Statistical Institute – ISI) fontosabb konferenciaajánlatai

(A teljes ajánlatlista megtalálható a <http://isi.cbs.nl/calendar.htm> honlapon.)

Prága, Csehország. 2013. augusztus 25–29.

7. Globális energiavezérlési és -optimalizálási konferencia. (7th Global Conference on Power Control and Optimization.)

Telefon: (+420) 57 603 5192

GSM: (+420) 775 161 965

E-mail: ivan.zelinka@ieee.org,

ivan.zelinka@vsb.cz

Honlap: www.ivanzelinka.eu

München, Németország. 2013. augusztus 25–29.

A Nemzetközi Klinikai Biostatistikai Társaság 34., éves konferenciája. (34th Annual Conference of the International Society for Clinical Biostatistics.)

E-mail: iscb2013@interplan.de

Honlap: <http://www.math.sc.chula.ac.th/sae2013>

Bangkok, Thaiföld. 2013. szeptember 1–4.

A Nemzetközi Statisztikai Intézet első ázsiai szatellit ülése a kisterületi becslésről. (First Asian ISI (International Statistical Institute) Satellite Meeting on Small Area Estimation.)

Honlap: <http://www.math.sc.chula.ac.th/sae2013>

Tokyo, Japán. 2013. szeptember 2–4.

A Bernoulli Társaság „Aszimptotikus statisztika és ehhez kötődő témák: elméletek és módszertanok” című szatellit ülése. (“Asymptotic Statistics and Related Topics: Theories

and Methodologies” Bernoulli Society Satellite Meeting.)

Honlap: <http://www.sigmath.es.osaka-u.ac.jp/~tokyo2013/>

Nürnberg, Németország. 2013. szeptember 9–11.

3. Európai Létesítménystatisztikai Műhelykonferencia. (3rd European Establishment Statistics Workshop.)

Információ: *Wilhelmus Kloek*

Wilhelmus.Kloek@cec.eu.int

Telefon: (+352)-4301-34963

Honlap: www.enbes.org

Amszterdam, Hollandia. 2013. szeptember 16–18.

8. Nemzetközi műhelykonferencia a kritikus informatikai infrastruktúrák biztonságáról. (8th International Workshop on Critical Information Infrastructures Security.)

Információ: *Cristina Alcaraz*

E-mail: critis2013@gmail.com

Honlap: www.critis2013.nl

Isztambul, Törökország. 2013. szeptember 19–21.

Fiatál üzlet- és iparstatistikusok közös ülése. (Joint Meeting of Young Business and Industrial Statisticians.)

Információ: *Paulo Canas Rodrigues*

E-mail: paulocanas@gmail.com

Honlap: ybis13.msgsu.edu.tr/

Folyóiratszemele

Musil, P. – Kramulová, J. – Cadil, J. – Mazouch, P.:

Az egy főre jutó regionális makro-aggregátumok becslése a regionális árszintek segítségével a PPS számításánál

(Application of Regional Price Levels on Estimation of Regional Macro-Aggregates Per Capita in PPS.) – *Statistika*. 2012. 49. évf. 4. sz. 4–13. old.

A tanulmány letölthető: [http://www.czso.cz/csu/2012edicniplan.nsf/eng/EC004449B2/\\$File/e-180212q4k01.pdf](http://www.czso.cz/csu/2012edicniplan.nsf/eng/EC004449B2/$File/e-180212q4k01.pdf)

Az országok gazdasági fejlettségének leggyakrabban használt mutatója az egy főre jutó GDP vásárlóerő-standardon (purchasing power standard – PPS) számítva. Az EU országokban a PPS-t az országos átlagából számítják, figyelmen kívül hagyva az országon belüli regionális árkülönbségeket. Az utóbbi időben az országok közötti különbségek mérése mellett igény mutatkozott bizonyos régiók összehasonlítására, az országokon belüli különbségek, esetleg eltérő tendenciák feltárására is.

Csehországban a régiónkénti árszínvonal-különbségek vizsgálatával eddig nem foglalkoztak. A tanulmány célja, hogy Csehországot a NUTS 3 régiókra bontva, megvizsgálja a regionális áreltéréseket és javaslatot tegyen az Eurostat/OECD-módszertan ilyen irányú módosítására. (A NUTS 3 szintnél részletesebb adatok nem állnak rendelkezésre.)

Az egyes valuták vásárlóerő-paritásának vizsgálata, az International Comparison Program (ICP – Nemzetközi Összehasonlítási Program) keretében, 1968-ban kezdődött 10 ország részvételével. (Az ICP-munkákban Magyarország is kezdettől fogva részt vett.) 2011-ben már 200 országra végeztek számításokat, amelyek között minden földrész képviselve volt.

Az ICP európai részprojektjéhez Csehország 1993-ban csatlakozott. A szerző felsorolja azt a 47 (tíz nem európai OECD-) országot, amely részt vesz a programban. Minden ország Ausztriával végzett bilaterális összehasonlítása képezi a további elemzések alapját. Az OECD és az Eurostat közös módszert alakítottak ki a PPS kiszámítására, míg az európai országokban a PPS-sel, az OECD-országokban az OECD „dollárral” számoltak.

A GDP kiadási oldalából kiindulva a részletes csoportokra, ún. basic headings-re (BH) ismertek a fogyasztás szerkezetét jellemző súlyok. A vásárlóerő-paritást minden BH-ra kiszámítják, a partnerországgal összehasonlítva az ún. Laspeyres- és Paasche-formulákkal, majd azok geometriai átlagából adódik a jól ismert EKS-indexek mátrixa.

Csehországban belül NUTS 3 szinten végeztek vásárlóerő-paritáson számított (purchasing power parity – PPP) becsléseket az elfogadott gyakorlat szerint. Ilyen részletes GDP-adatok azonban nincsenek. (Az export/import számbavétele nem oldható meg.) A háztartások végső fogyasztása a GDP kiadási oldalának mintegy 50 százalékát teszi ki. A régiónkénti árszínvonal-

Megjegyzés. A Folyóiratszemelet a KSH Könyvtár (*Lencsés Ákos*) állítja össze.

kat kiszámítva számottevő különbségek adódtak. A számításokat a 148 BH-nél részletesebben, reprezentáns szinten (700) végezték el. A hiányzó adatok kezelését úgy oldották meg, hogy a többi régió árainak számtani átlagát vették. Tekintettel a hazai csoportosításra, a háztartási kiadások tekintetében is szükség volt bizonyos korrekcióra.

Az adatok összegyűjtése sok nehézséggel járt, amit országonként eltérő módszerekkel oldottak meg. Csehországban a regionális árszínvonal eltéréseinek vizsgálatához az árak a fogyasztói árstatistika havi gyakoriságú felvételeiből származtak. Bizonyos tételek (például lakbérek) esetében becslésekre volt szükség.

*A régióinkénti összehasonlítás eredményei
a hivatalos és a korrigált módszerrel
(százalék)*

Régió	PPP	Nettó jövedelem	Nettó jövedelem PPS-ben
Prága	119,7	132,2	110,5
Středočeský	101,9	107,3	105,4
Jihočeský	97,9	96,4	98,5
Plzeňský	97,1	99,1	102,1
Karlovarský	101,4	89,5	88,3
Ústecký	94,9	87,7	92,4
Liberecký	101,4	93,5	92,2
Královéhradecký	96,4	96,9	100,5
Pardubický	98,2	94,8	96,6
Wysoká	95,6	95,0	99,4
Jihomoravský	103,4	98,2	94,9
Olmoucký	96,9	92,1	95,1
Zlínský	100,8	96,7	95,9
Moravskoslezský	96,7	90,1	93,2
Csehország	100,0	100,0	100,0

Mint várható volt, a fővárosban a legmagasabbak az árak. Meglepetést okozott, hogy két kis régióban (Lyberek és Zlín) átlag fölött alakult az árszínvonal, valószínűleg a városi jellegű települések viszonylag nagy súlya miatt.

A legfontosabb eredmény az, hogy régióinként a vásárlóerővel korrigált mutatószámok sokkal kisebb szórást mutatnak. A nettó jövedelem „szórása” folyó áron az átlag százalékában: 90,1–132,2 százalék között van, míg az a vásárlóerő-paritással korrigálva sokkal kisebb: 93,2–110,5 százalék.

A régióinkénti adatok elemzésénél gondolni kell arra is, hogy a népesség régióinként eltérő arányban nem ott lakik, ahol dolgozik. A népesedési statisztikák további elemzéseket tesznek lehetővé.

A tanulmány célja az volt, hogy ráirányítsa a figyelmet az árszínvonal területi eltéréseinek vizsgálatára. A nemzetközi összehasonlításokban használt országonként homogénnek tekintett árak torzíthatnak, mivel országon belül (például Prágában), az árak jóval magasabbak, mint az ország többi részében. A tanulmány feltárta, hogy a területi GDP-mutatók is jelentős szórást mutatnak. Adatok hiányában azonban visszamenőleges számítások, idősor előállítás nem lehetséges. A jövőben célszerű lenne további kutatásokat végezni. Továbbá, szükség lenne a fogyasztói árindexek (consumer price index – CPI) területi alakulását is vizsgálni. Végül érdekes lenne, ha más országok is elemeznék a PPP országon belüli regionális eltéréseit.

Marton Ádám

kandidátus, a KSH ny. osztályvezetője
E-mail: Adam.Marton@ksh.hu

Stadler, B. – Wiederhofer-Galik, B.:

A fiatal felnőttek különböző migrációs ismérvek szerinti képzettségi és munkaerő-piaci jellemzői Ausztriában

(Bildungs- und Erwerbspartizipation junger Menschen in Österreich unter besonderer Berücksichtigung des Migrationshintergrundes.) – *Statistische Nachrichten*. 2012. 67. évf. 12. sz. 957–975. old.

2011-ben az Ausztriában élő 15–34 évesek 23,2 százalékának, összesen 481 ezer főnek volt valamilyen migrációs háttere. Ebből 359 500 fő volt első generációs migráns. Mivel ez a generáció meghatározó tényezője lesz a jövő munkaerőpiacának, fontos tudni, hogy a migrációs háttér milyen képzettségi, illetve munkaerő-piaci hátrányt jelent. Ausztriában az elmúlt évtizedekben jelentősen változott a bevándorlók összetétele is. A nyolcvanas évekig az akkori Jugoszlávia, illetve Törökország voltak a fő kibocsátó országok, melyet egy-egy, a keleti blokk országaiból kiinduló migrációs hullám színezett (1956-ban a magyar, 1968-ban a cseh, majd ezt követően a lengyel eseményekre visszavezethetően). A vasfüggöny lebontása után inkább a ma már uniós tag, volt szocialista és – egyre nagyobb arányban – Európán kívüli országokból (Kína, Fülöp-szigetek) jöttek a bevándorlók. Emiatt az első és második generációs migránsok rétege meglehetősen eltérő, ezért az elemzésben a migrációs háttérrel rendelkezők nem kezelhetők egységesen, sőt az utóbbi csoporton belül is meg kell különböztetni azokat, akiknek csak az egyik szülője nem ausztriai születésű.

A tanulmány a 2004 óta folyamatos és negyedévente a népesség mintegy 0,6 százalékára kiterjedő mikrocenzus 2011. évi adatainak felhasználásával készült. Ezt az tette lehetővé, hogy a felvétel kérdései között 2008 óta az apa, valamint az anya születési helyére vonat-

kozó is szerepel, így az kombinálható a célszemély képzettségi, illetve foglalkoztatási jellemzőivel.

A fiatal migránsok vizsgálatánál az első szembeszökő sajátosság az, hogy minél fiatalabb korcsoportról van szó, annál magasabb a második generációs migránsoknak az összes migránson belüli aránya. Ennek oka, hogy nagyon sok, korábban gyermekként bevándorolt felnőtté válva Ausztriában alapított családot. A teljes 0–4 éves kohorszból 21,3 százalékuk mindkét, további 15,5 százalékuk az egyik szülője külföldön született, ugyanez az 5–9 évesek esetében 17,9 és 13,2 százalék volt, míg a 19 éven felüliek egyetlen korcsoportjában sem haladta meg a második generációs migránsok aránya a 7 százalékot. Ezzel szemben a 25–29 évesek 21,4, a 30–34 évesek 24,5 százaléka külföldön született. A 20–24 évesek 15,5 százaléka szintén első generációs migráns, viszont az ennél fiatalabb korosztályokban arányuk már csekély. Elmondható tehát, hogy a 19 éven felüli fiatal migránsok jellemzően első, míg az ennél fiatalabbak második generációs migránsok.

A második generációs migráns fiatalok többségének szülei a volt Jugoszláviából vagy Törökországból vándoroltak be, de a legfiatalabbak között mind nagyobb arányt képviselnek a más EU-országból, illetve az Európán kívüli országokból jött szülők gyermekei. A 15–19 éves korosztály sajátos származási ország szerinti összetétele a jugoszláv háború miatti menekülthullám eredménye.

Az első generációs fiatal migránsok későbbi munkaerő-piaci boldogulásánál kulcskérdés, hogy milyen életkorban érkeztek Ausztriába. Azok, akik itt váltak tankötelezetté, nyelvi gond nélküli iskolai karrier útra számíthatnak, míg azoknak, akik tanulmányukat más országban kezdték egyszerre kell megbarátkozniuk a nyelvvel és egy más típusú iskolarendszerrel. A 2011-ben vizsgált első

generációs fiatal migránsoknak alig hetede volt 7 évnél fiatalabb, mikor szüleivel áttelepült Ausztriába, és harmaduk 7–15 év közötti volt. A tankötelezettségi korban Ausztriába érkező fiatalok 30,2 százaléka Jugoszláviában, 31,8 százaléka Törökországban született.

Befejezett iskolai végzettség tekintetében a migránsok a két szélső képzettségi szint – az általános iskolát be nem fejezők és a diplomások – esetében alulreprezentáltak. A második generációs migránsok képzettségi szintje magas, itt jóval alacsonyabb a csak alapfokú végzettséggel rendelkezők aránya, mint az első generációs fiatalok között. Jelentős, bár minél fiatalabb korosztályról van szó, annál kisebb a nők végzettségbeli lemaradása. A „szintiszta” osztrákokra jellemzőnél is jobb azoknak a fiataloknak az iskolai végzettség szerinti összetétele, akiknek csak az egyik szülőjük volt külföldi, míg azoké, akik külföldi szülőtől Ausztriában születtek lényegesen rosszabb. Ez utóbbi azt jelzi, hogy a migrációs hátrány, mint ahogy erre az OECD vonatkozó tanulmánya is rámutatott, öröklődik.

Az uniós tagországokból, illetve nem a két nagy kibocsátó országból (Jugoszláviából, illetve Törökországból) származó fiatal felnőtt (25–34 éves) migránsok iskolai végzettség szerinti összetétele jobb az osztrákokénál, a legalább érettségivel rendelkezők aránya 15, illetve 18 százalékponttal magasabb az osztrákokra jellemző 42 százalékos aránynál. Ezzel szemben a török születésűek mindössze 14,7 százalékanak van érettségije vagy ennél magasabb végzettsége, 51,6 százalékuk pedig csak a kötelező alapfokú oktatásban vett részt. Mivel a második generációs migránsok nagy százalékban jugoszláv, illetve török származásúak, ez jelentősen befolyásolni fogja a jövő munkaerő-kínálatát.

Az elemzés kitér arra is, hogy milyen kapcsolat van aközött, hogy a fiatal felnőtt hány éves korban érkezik Ausztriába, illetve milyen

típusú végzettséget szerez, továbbá bemutatja azt, hogy a különböző migrációs csoportok tagjai milyen arányban vesznek részt az oktatásban.

A nemzetközi összehasonlítás céljaira kidolgozott legfontosabb oktatási mutató a korai iskolaelhagyók, vagyis az iskolarendszerből képzetlenül kilépő 18–24 éveseknek a teljes korosztályhoz viszonyított aránya. A mutató országosan 8,3 százaléknak felelt meg 2011-ben, de a migrációs háttérrel nem rendelkező többség esetében ennél 3 százalékponttal alacsonyabb volt. Az első generációs migránsokra jellemző érték 20,6 százalék, míg a második generációsoké ennél kedvezőbb, 15 százalék. Az első generációs migránsokra jellemző igen kedvezőtlen mutató részben azzal függ össze, hogy közülük sokan már tanköteles korban kerültek Ausztriába, s ha sikerült is az alapfokú végzettséget megszerezni, továbbtanulásukat nyelvi nehézségek korlátozták, viszont segéd munkásként azonnal fizető munkához tudtak jutni.

A másik nemzetközileg használt képzettségi indikátor az OECD NEET (Not in Education, Employment or Training – az oktatási, a foglalkoztatási és a szakképzési rendszerben nem lévők aránya), melynek vonatkozási korcsoportját a 15–29 évesek jelentik. Ez a mutató az előbbihez hasonló méretű migrációs hátrányt jelez, bár a második generációs migránsok relatív helyzete ennek alapján egy árnyalattal kedvezőbb. A 15–34 éves korosztályra kiterjesztett mutató esetén viszont igazán drámai különbségek tapasztalhatók. Az ilyen korú migráns fiatalok 20,3 százaléka se nem tanult, se nem dolgozott 2011-ben, ami közel háromszorosa az osztrák születésűekre jellemző aránynak. Ennek egyik magyarázata, hogy az első generációs migráns nők jelentős arányban háztartásbeliként élik életüket.

A képzettségihez hasonló deficit jellemzi a migránsokat a munka világában is. A 25–34

éves osztrák születésűek foglalkoztatási rátája 2011-ben 87,5 százalék volt, a második generációs migránsoké 75,3 százalék, az első generációhoz tartozóké pedig mindössze 71,6 százalék. Az osztrák fiatalok (15–24 évesek) munkanélküliségi rátája ugyan ebben az évben 6,2, a migránsoké 15,5 százaléknak felelt meg. Az inaktívak aránya a 25–34 éves korcsoportból országosan 13,1 százalék volt, ezen belül az első generációs migránsokat 22,4, a második generációs migránsokat 18, míg a többieket 9,4 százalékos mutató jellemezte. Az első generáció magas inaktivitását a nők alacsony foglalkoztatási mutatói magyarázzák, míg a második generációhoz tartozó nem dolgozó nők között sokan munkanélküliek. A tanulmányban származási ország és nem szerinti bontásban is megtalálható a 25–34 éves korosztály foglalkozási rátája.

A munkaerő-piaci integráció fontos jelzőszáma a munkavégzés tartóssága, az, hogy az adott csoportba tartozók mióta dolgoznak jelenlegi munkájukban. 2011-ben a 25–34 évesek átlagosan 4,4 évet dolgoztak összeírásuk munkájukban, viszont ebből a migrációs háttérrel rendelkezők mindössze 3,2 évet. Az átlagot esetükben az első generációra jellemző 2,8 év rontja le.

A migrációs háttérű és az osztrák születésű fiatalok által végzett munka kvalifikációs igénye is különböző, az előbbi csoportba tartozók az utóbbiaknál lényegesen nagyobb arányban végeznek segéd- és betanított munkát, de a kvalifikációs igény szerinti megoszlás származási ország szerint is jelentősen differenciálódik.

A tanulmány szerzői a befejező részben külön is vizsgálták a második generációs migráns fiatalokat, megadva a fontosabb képzettségi és aktivitási mutatóiknak a teljes népességre jellemzőtől való eltérését.

Lakatos Judit

E-mail: Judit.Lakatos@ksh.hu

Rosenski, N.:

Statisztikai elemzés Németország nonprofit szektorának gazdasági szerepéről

(Die wirtschaftliche Bedeutung des Dritten Sektors.) – *Wirtschaft und Statistik*. 2012. 3. sz. 209–217. old.

A tanulmány letölthető:

https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/UnternehmenGewerbeanzeigen/WirtschaftlicheBedeutung3Sektor032012.pdf?__blob=publicationFile

A cikk Németország nonprofit szektorát értékeli, a statisztikai regiszterek és a nemzeti számlákba illeszthető gazdasági elszámolások alapján. A szektor felmérései, elszámolásai a nyilvános információs rendszerben elérhetők (http://www.stifterverband.info/statistik_und_analysen/zivilgesellschaft_in_zahlen/ziviz_modul_1_abschlussbericht.pdf).

A szerző a háromrészes kiadvány első, statisztikai modulja adatai alapján elemzi a német gazdaság 2007. évi teljesítményében a harmadik (nonprofit) szektor intézményeit, a becsült bruttó hozzáadott érték arányát, valamint a fontosabb munkaügyi adatsorokat. Nemzetgazdasági ágak, valamint szövetségi tartományok szerint is értékeli a nonprofit szektor országos összefoglaló adatait.

Az intézményi szektor fogalmi meghatározása megfelel az ENSZ „Handbook on Non-Profit Institutions in the System of National Accounts” (Kézikönyv a nonprofit intézményekről a nemzeti számlarendszerben) című dokumentumában szereplő öt ismérvnek (http://unstats.un.org/unsd/publication/seriesf/seriesf_91e.pdf):

- formális szervezethez tartozás, bizonyos mértékig intézményesítve;
- magánjellel a kormányzat szektorától elhatároltan;

– nem elsődleges cél a profitmaximum, tulajdonosa azt a szervezet rendeltetésének megfelelően ismételten befekteti, és a megszerzett jövedelemből a tagok nem részesülnek;

– tevékenységet más nem ellenőrzi, önigazgató a működés;

– a részvétel önkéntes.

A harmadik szektorba sorolható a bejegyzett és szervezeten működő intézmény, ha teljesíti az előbbi öt ismérvtől mindegyikét. A cikk részletes módszertani leírást ad arról, hogy miként alkalmazták a németországi statisztikai regiszterek azonosító rendszerét a harmadik szektorba sorolásokhoz, előbb automatikus, majd a minta esettanulmánya alapján, kézi eljárással.

A szerző a 2007. tárgyév adatait ismerteti a statisztikai regiszter 2009. szeptember 30-i állománya alapján. A vizsgálatok alapsokaságába a nem agrár és nem kormányzati intézmények tartoznak (a B–N és P–S nemzetgazdasági ágakban). A teljes vonatkozási kör mintegy 3,6 millió vállalata összesen 25 millió főt foglalkoztat. A regiszter adatai szerint ezek a vállalatok összesen 4,9 milliárd euró adóköteles értékesítési árbevételt jelentettek a tárgyévben.

A cikk ismerteti a statisztikai programra alapozott elhatárolás eljárásait, az erre alkalmazott, adminisztratív adatforrásból származó regiszteradatokat. A gépi csoportosítás paramétere a jellegzetes cégnév is, például a „közhasznú” gazdálkodási forma a harmadik szektorba sorolható. A projektben közreműködő szakértők felsorolják a „harmadik szektor” gazdálkodási formáit (például egyesület, alapítvány). Egyértelmű, hogy profitcélú vállalat (például részvénytársaság, kft, pénztintézet) nem tartozhat a harmadik szektorba.

A gépi csoportosítás az alapsokaság mintegy 70 százalékaról egyértelműen meg-

állapította a harmadik szektorhoz tartozását, illetve annak kizáró ismérvét (például a kormányzati tulajdont, fenntartást). A sokaság 30 százalékát (1,1 millió regiszteregységet) érintették a további vizsgálatok, ezek „potenciálisan” részei a harmadik szektornak. Mintavétel alkalmazásával határozták meg az ilyen statisztikai egységek alapvető kategóriáit és összesen 813 ezer olyan tételt találtak, amely nem tartozhat a harmadik szektorba.

A nemzeti számlához csatlakozó becslés érdekében a 11 ezer vállalatot tartalmazó mintát mélyebben is elemezték, és meghatározták a bruttó hozzáadott érték és a foglalkoztatottak számának tárgyévi adatait, majd rétegek szerint, modellezéssel végezték a teljes körre számítást.

A cikk ismerteti a harmadik szektor összesen 62 ezer egységet tartalmazó mintájának tételes felmérését és az adatelemzés lépéseit. Ez a vizsgálat sorozat öt hónapig tartott és vagy a harmadik szektorba sorolták az egységet, vagy egyértelmű lett, hogy nem oda tartozik. Ahol az anyavállalat adata ismert, ott a leányvállalat „örökölte” az irányítójának besorolását. A vizsgálat során közel 100 ezer anyavállalatot, illetve ezek csoportjaiban 260 ezer irányított egységet soroltak be. A nem közszférába tartozó anyavállalat és így leányvállalatai is a harmadik szektorba kerültek. A közszférába sorolt anyavállalat minden leányvállalata kizárható a harmadik szektorból.

A szerző összefoglalja a németországi statisztikai felmérés eredményeit, a harmadik szektor összesen 105 ezer vállalatának adatai alapján. Németországban a harmadik szektor 2007-ben 2,3 millió főt foglalkoztatott, és további mintegy 300 ezer főt alkalmaztak a csekély keresetű kategóriákban (a juttatás mértéke nem haladja meg a 450 euró/hó mértékét és csak nyugdíjjárulék-fizetés terheléssel). A nonprofit szektor részesedése eszerint a regiszter egységeinek számából 3, a foglal-

kozottakból 9, illetve a csekély keresetűekből 7 százalék. A társadalombiztosításban regisztrált foglalkoztatottak 40 százalékát részmunka-időben foglalkoztatták a harmadik szektor egységei.

A cikk összefoglaló értékelést tartalmaz a 2007. év gazdasági tevékenységek ágazati szerkezetéről. Vannak olyan szolgáltató ágazatok például a szociális ellátásban, ahol a foglalkoztatottak több mint 80 százalékát a harmadik szektor nonprofit intézményei alkalmazzák.

Részletes adatsorok mutatják be a harmadik szektor részesedését a fontosabb ágazatok foglalkoztatottjainak 2007. évi számából. A harmadik szektorban foglalkoztatottak 20-20 százaléka tartozik az egészségügyi és szociális ellátás, valamint az érdekvédelem és az egyházi tevékenység ágazataihoz, az oktatás részesedése 14 százalék. A szerző megállapítja, hogy viszonylag nagyok az említett szolgáltató ágazatok intézményei, részesedésük a foglalkoztatottak számában 20-20 százalék, viszont az egészségügyi intézmények, illetve a szociális otthonok számának aránya csak 2,4, illetve 6,5 százalék.

A cikk bemutatja a harmadik szektor adatait a szövetségi tartományok szerint is.

Vannak több telephelyet fenntartó intézmények, azonban az egységek 97 százaléka csak egy tartományban működik, viszont a foglalkoztatottak számában csak 68 százalékos ez az arány. A harmadik szektor alkalmazta a Berlinben (fővárosban) foglalkoztatottak 14 százalékát 2007-ben, az országos arány ennél kisebb, 9,2, Hamburgban kerekén 7 százalék.

A Szövetségi Statisztikai Hivatal 2011 szeptemberében jelentette meg a nemzeti számlák nagy felülvizsgálatának eredményeit, ezen belül a szektorszámolókat és ebben a nem piaci termelők bruttó hozzáadott értékét. A harmadik szektor adta Németországban a bruttó hazai termék 2007. évi értékének (2 181 milliárd euró) 4,1 százalékát, ez mintegy 89 milliárd euró (csak az árbevétel jelentő, a statisztikai regiszterbe felvett, működő intézményeket véve számításba, így kimaradtak a csekély foglalkoztatási (és adózási) hatású (nem adózó) intézmények). Az említett teljesítmény megfelel például az építőipar vagy a gépjárműgyártás bruttó hozzáadott értékének.

Nádudvari Zoltán,

a KSH ny. főtanácsosa

E-mail: nadyzol@freemail.hu

Kiadók ajánlata

HÄRDLE, W. K. ET AL. [2013]: *Basics of Modern Mathematical Statistics*. (A modern matematikai statisztika alapjai.) Wiley. New York.

A mai statisztikai adatigények összetettsége modern matematikai eszközöket kíván meg. A tudomány számos terén a matematikai statisztika használatára és a statisztikai módszerek folyamatos frissítésére van szükség. Gyakorlat teszi a mestert, ugyanis használatuk elsajátításával meg is értjük ezeket az eszközö-

ket. A könyv, ami gyakorlatokat és azok megoldásait tartalmazza, az alkalmazások széles körét és számtani megoldásait kínálja R-nyelven.

A modern matematikai statisztikaoktatás célja, hogy a statisztikát hallgató diákok számára számos alapfeladat álljon rendelkezésre, és azok megértsék, hogy az elmélet miként alkalmazható valódi problémák esetén.

A szerzők az alkalmazási szempontokra is nagy hangsúlyt helyeznek, mivel a korábban

kiadott könyvek főleg elméleti levezetéseket közölnek. A kötetben szerepel néhány, mai kutatási témákkal kapcsolatos probléma is.

FUEST, C. – ZODROW, G. R. (EDS.) [2013]: *Critical Issues in Taxation and Development*. (Kritikus adózási és fejlesztési kérdések.) The MIT Press. Cambridge.

Számos fejlődő ország számára nehéz bevételeket teremteni olyan alapvető közszolgáltatások biztosításához, mint az oktatás, az egészségügyi ellátás és az infrastruktúra. A fejlett országoknak nem kell szembe nézniük a fejlődő országok adópolitikai kihívásait bonyolító kérdésekkel, többek között a széles körű korrupcióval, az adócsalással és -elkerüléssel, valamint a hatástalan politikai struktúrákkal. Ebben a tanulmánykötetben a szakértők a fejlődő országok bevételeinek emelésében fontos szerepet játszó feladatokat vizsgálják.

Az áttekintést követő fejezetekben modern empirikus módszerekkel történik minden egyes kritikus kérdés vizsgálata, melyek nélkülözhetetlenek az adók fejlődésre gyakorolt hatásának megértésében. A tárgyalt témák között van az adózás hatása a külföldi tőkebefektetésre; a korrupció formái, az adócsalás és -elkerülés, amik a fejlődő országok sajátosságai; a politikai struktúrára vonatkozó kérdések, beleértve a pénzügyi decentralizáció negatív hatásait a fejlesztési segélyek hatékonyságára; valamint a demokrácia és az adózás közötti kapcsolat azokban az ázsiai, latin-amerikai és EU-országokban, amelyekben nemrégiben ment végbe a politikai és gazdasági átmenet.

HALLE, D. – BEVERIDGE, A. A. (EDS.) [2013]: *New York and Los Angeles – The Uncertain Future*. (New York és Los Angeles – a bizonytalan jövő.) Oxford University Press. Oxford.

A könyvben az Egyesült Államok két legnagyobb városáról, New Yorkról és Los Angelesről, illetve térségükről szóló, részletes összehasonlító tanulmányok szerepelnek. A vezető szakértők által írt, és a népszámlálás, illetve egyéb források legfrissebb információin alapuló fejezetek a politikai, a gazdasági kilátásokat és a pénzügyi válságot, valamint számos társadalmi kérdést tárgyalnak és hasonlítanak össze. A kötet az oktatási reformmozgalmakat, a kisebbségi politikát, a költségvetési megszorításokat, a bűncselekményellenes stratégiákat, az infrastruktúra fejlesztését és politikai összefüggéseit, a növekvő egyenlőtlenséget, a bevándorlást és a bevándorlók közösségeit, a szegények és a kisebbségek szegregációját, a gazdasági elit újszerű elkülönülését, a környezeti hatásokat és az ezek kezelésére tett próbálkozásokat, a két város és a régiók arculatát a mozifilmekben, az építészeti tendenciákat, valamint a pénzügyi válság hatását és a rá való reagálást vizsgálja, beleértve az ingatlanárverési formákat is.

Ez az összehasonlítást nyújtó keret bemutatja, hogy az olyan régi paradigmák, mint a városok „hanyatlása” vagy „újjáéledése” már nem megfelelők azoknak az új kihívásoknak a megragadására, amelyekkel Amerika két legnagyobb világvárosa szembenéz. Ezek néha hasonló, néha eltérő módon reagálnak az elmúlt évtized két meghatározó eseményével járó feladatokra: a 2011. szeptember 11-ei támadásra és következményeire, valamint a pénzügyi válság további hatásaira. Az, hogy milyenek ezek az események, intézmények és tendenciák a New York-i és Los Angeles-i régiókban, nemcsak a két város számára fontos, hanem más egyesült államokbeli városok, az egész nemzet, illetve a világ más városai számára is előhírnök. A könyv nélkülözhetetlen útmutatót nyújt számos, városaink jövőjét meghatározó erő megértésére.

FINKELSTEIN, M. O. – LEVIN, B. [2013]: *Statistics for Lawyers*. (Statisztika jogászok számára.) Springer. New York.

A könyv a jogászhallgatókat, a tanárokat, a gyakorló jogászokat és a bírákat a jogban alkalmazott valószínűség-számítás és statisztika alapjaiba vezeti be. A részletes ismertető részeket az életből vett példák leírása és olyan esettanulmányok követik, amelyekben statisztikai adatok is szerepet kaptak. Az olvasóknak az elméletet a tények vonatkozásában is használniuk kell, ki kell számolniuk az eredményeket, és meg kell vizsgálniuk a mennyiségi eredményekből eredő jogi kérdéseket. A szerzők számításai és megjegyzései a könyv hátsó

részében található. Az ügyek és esettanulmányok a jogi témák széles körét lefedik, többek között a diszkrimináció tilalmát, az adózást, az iskolai pénzügyeket, a bizonyítékok értékelését, az előzetes letartóztatást, a kézirati vitákat, a szavazást, a környezetvédelmet, a szabad versenyt és a halálbüntetést. A kötet első, 1990-ben megjelent kiadását jogi, statisztikai és társadalomtudományi kurzusokon használták. 1991-ben a Michigani Egyetem *Law Review* folyóirata az év egyik legfontosabb jogi kötetének választotta. A második kiadás számos új problémát is tárgyal, beleértve egy új járványtani fejezetet is, amelyek a jog új-donságait tükrözik.

Társfolyóiratok



A FRANCIA GAZDASÁGI ÉS PÉNZÜGYMINISZTERIUM, VALAMINT A STATISZTIKAI ÉS GAZDASÁGKUTATÓ INTÉZET FOLYÓIRATA

2011. ÉVI 454. SZÁM

Alet, É. – Bonnal, L.: A gyakorlati képzés hatása az „5. szintű” diákok iskolai teljesítményére egy francia példa alapján.

Guergoat-Larivière, M. – Marchand, O.: A foglalkoztatás minőségének meghatározása és mérése európai összehasonlítások segítségével.

Mordier, B.: A foglalkoztatás minősége a szolgáltatási szektorban – stabilitás és munkajövedelem 1997 és 2007 között.

Baguelin, O.: Rövid távú foglalkoztatáspolitikai és a helyi munkaerő-piaci szerkezetek –

a 2009. évi foglalkoztatástámogatási szerződés eredményei helyi szinten.



AZ EGYESÜLT NEMZETEK EURÓPAI GAZDASÁGI BIZOTTSÁGÁNAK FOLYÓIRATA

2012. ÉVI 3–4. SZÁM

Forbes, S. – Brown, D.: Konceptcionális gondolkodás a nemzeti statisztikai hivataloknál.

Hoekstra, R. – ten Bosch, O. – Harteveld, F.: Automatikus adatgyűjtés webes forrásokból a hivatalos statisztika számára – első tapasztalatok.

Marske, R. A.: A válaszadás javítása az Egyesült Államok gazdasági összeírásában a

kommunikáció, a koordináció és a kapcsolatok kiterjesztése útján.

Baffour, B. – Valente, P.: Az összeírások minőségének értékelése.

Smaill, K.: A longitudinális nemválaszolás modellezése üzleti felvételek esetén.

Conn, S.: Az adatháló tágítása – a nemzeti statisztikai hivatalok vezető szerepe.

Höninger, J.: Morpheus – a távoli adat-hozzáférés innovatív megközelítése.



AZ EGYESÜLT ÁLLAMOK
MATEMATIKAI STATISZTIKAI INTÉZETÉNEK
FOLYÓIRATA

2012. ÉVI 4. SZÁM

Bach, F. et al.: Strukturált szórt reprezentánsok konvex optimalizálással.

van der Geer, S. – Müller, P.: Kvázilikelihood és/vagy robusztus becslés nagy dimenziószámánál.

Huang, J. – Breheny, P. – Ma, S.: Csoporthelyezés többdimenziós modelleknél – szakirodalmi áttekintés.

Giraud, C. – Huet, S. – Verzelen, N.: Többdimenziós regresszió ismeretlen szórásnégyzettel.

Lafferty, K. – Liu, H. – Wasserman, L.: Szórt nemparaméteres grafikus modellek.

Negahban, S. N. et al.: Egységes keretrendszer M-estimator többdimenziós vizsgálatához.

Rigollet, P. – Tsybakov, A. B.: Szórt becslések exponenciális súlyozáshoz.

Zhang, C. – Zhang, T.: Konkáv megkötések általános elmélete többdimenziós szórt becsléseknél.

McElroy, T. S. – Holan, S. H.: Beszélgetés David Findley-vel.

2013. ÉVI 1. SZÁM

Friendly, M. – Monette, G. – Fox, J.: Elliptikus észlelés – a statisztikai módszerek megértése az elliptikus geometria segítségével.

Pfeffermann, D.: Új, jelentős fejlemények a kisterületi becslések esetén.

Lin, M. – Chen, R. – Liu, J. S.: Előretekintő stratégiák szekvenciális Monte Carlo-módszerhez.

Dutta, R. – Ghosh, J. K.: Bayesi modellválasztás útmintavételezésnél – faktormodellek és további példák.

Reimherr, M. – Nicolae, D. L.: A mennyiségi függés – értelmezhető intézkedések/mértékek kialakítási kerete.

Taqqu, M. S.: Benoît Mandelbrot és a fraktális Brown-mozgás.

Statistische Nachrichten

AZ OSZTRÁK KÖZPONTI STATISZTIKAI
HIVATAL FOLYÓIRATA

2013. ÉVI 2. SZÁM

Nemzetközi vándorlás Ausztriában 2011-ben.

A háztartások és családok 2012. évi előre-számítása Ausztriában és tartományaiban 2060-ig.

Fogyasztói árindex 2012 decemberében.

Gyümölcsstermesztés 2012-ben.

Zöldségstermesztés 2012-ben.

Idegenforgalom a 2012. évi nyári szezonban.

Társaságiadó-statisztika 2008-ban.

2013. ÉVI 3. SZÁM

A több lakcímmel rendelkező osztrák állampolgárok területi és társadalmi-demográfiai jellemzői.

A munkanélküliségi arány új kiegészítő indikátorai európai szinten.

Foglalkoztatási adatok a nemzeti számlák és a munkaerő-felmérés alapján – eltérő nézőpontok és adatok.

Fogyasztói árindex 2012-ben.

Fogyasztói árindex 2013 januárjában.

Szántóföldi növények termesztése 2012-ben.

A 2010. évi jövedelemadó-adatok értéklése.

WIADOMOŚCI STATYSTYCZNE

A LENGYEL STATISZTIKAI FŐHIVATAL
FOLYÓIRATA

2013. ÉVI 2. SZÁM

Bal-Domańska, B.: A munkatermelékenység konvergenciája az EU-régiókban.

Kasprzyk, B.: Modell az anyagi életkörülmények megítéléséről.

Jefmański, B.: A tanácsadói javaslatok hatása a továbbtanulásról, a mérések eredménye.

Fihel, A. – Mészé, F. – Vallin, J.: Az 1996 és 2002 közötti lengyelországi halálesetek becslése okok szerint.

Dolata, M. – Lira, J.: A vidéki járáások vizsgálata a vízgazdálkodási eszközök és a szennyvízkezelési infrastruktúra alapján.

Mongiolo, D.: A lengyel EU-csatlakozás hatása a szolgáltatások kereskedelmére.

Gorczyca, M.: Lakáshelyzet Fehéroroszországban.

Majsterek, M.: A Lengyel Statisztikai Hivatal Tudományos Statisztikai Tanácsának jelentése – 2012. december 10.

Lengyelország társadalmi-gazdasági helyzete 2012-ben.

Wirtschaft und Statistik

A NÉMET SZÖVETSÉGI STATISZTIKAI
HIVATAL FOLYÓIRATA

2013. ÉVI 1. SZÁM

Räth, N. – Braakmann, A.: Bruttó hazai termék 2012-ben.

Hirner, S. – Stiglmayr, S.: A 2011. évi népszámlálás hivatkozási adatkészlete.

Körner, T. – Meinken, H. – Puch, K.: Kik a kizárólag kis mértékben foglalkoztatott személyek? Elemzés a társadalmi körülmények alapján.

Heinig, S.: A sportszektor vállalkozásszerkezete és gazdasági jelentősége.

Árak 2012 decemberében.

2013. ÉVI 2. SZÁM

Pötzsch, O.: Hogyan befolyásolja a gyermekvállalás későbbre tolódása az egyes kohorszok termékenységét Németország keleti és nyugati felén?

Hohmann, K. – Scharfe, S.: A teljes állami költségvetés pénzügyi eszközei 2011. december 31-én.

Reim, U. – Reichel, B.: Tömegközlekedés buszon és vasúton 2011-ben.

Duschek, K. – Buhtz, C.: Lakástámogatás Németországban 2011-ben.

Günther, R.: A 2010. évi kereseti felvétel szerkezetének módszertana.

Schindler, S.: Nyitási folyamatok a középfokú oktatásban és az oktatási egyenlőtlenségek alakulása.