

## A STATISZTIKA: TUDOMÁNY ÉS SZAKMA\*

DR. PUKLI PÉTER – VÉGVÁRI JENŐ

A tanulmány a statisztikával foglalkozik: egyrészt a statisztikával mint tudománnyal, másrészt mint szakmával. Előbb arra a kérdésre keresi a választ, hogy egyáltalán tudomány-e a statisztika, s ha igen, milyen tudomány. A magyar szakirodalomban előforduló ezzel kapcsolatos vitákat felidézve a szerzők úgy foglalnak állást, hogy a statisztika önálló tudományág, mégpedig módszertudomány. A szerzők bemutatják a hivatalos magyar statisztikában használatos tudományági osztályozásokat, amelyek „nem ismerik” a statisztika tudományát, egyúttal kifejtik nézetüket arról, hogy egy kidolgozandó új, korszerű osztályozásban hol lehetne a statisztika helye.

Ezt követően a tanulmány körülírja azt az ismeretanyagot, amely a statisztikus szakma gyakorlásához szükséges. Áttekinti a statisztikai adatok előállításának folyamatát annak fényében, hogy a személyi számítógépek tömeges elterjedésével és a felhasználóbarát szoftverek megjelenésével az informatikai tevékenység egy része integrálódott a statisztikai szakmába. Ily módon a szakma tartalma viszonylag rövid idő alatt átalakult, s ennek hatására a hivatalos statisztika és a statisztika tudománya („akadémiai statisztika”) között szakadék alakult ki.

TÁRGYSZÓ: Statisztika tudománya. Statisztikus szakma. Tudományágak osztályozása.

Ha valaki elolvassa a statisztika mibenlétével foglalkozó tanulmányokat és cikkeket, és fellapozza a statisztika tárgyú tankönyvek bevezetőit, a statisztika többféle definíciójával találkozik. A különböző definíciók abban azonban többnyire megegyeznek, hogy a statisztika egyfelől tudomány, másfelől gyakorlati tevékenység. A pontosabb megfogalmazások ezt a sajátosságot a statisztika szavunkra vonatkoztatják, azaz ugyanazzal a szóval illetünk egy tudományágot és egy – ezen a tudományon alapuló – tevékenységet. A tudomány és a gyakorlat más területein nincs feltétlenül szükség ilyen megkülönböztetésre, ugyanis más-más szóval jelölik az adott tudományt, illetve a hozzá kapcsolódó gyakorlatot. Elmondhatjuk tehát, hogy a különböző tudományágak körül háromféle tevékenység alakul ki: a kutatás, az oktatás és az alkalmazás. Nincs közöttük merev választóvonal, az oktatásban is folyhat kutatás és az alkalmazók is kutathatnak, oktathatnak. Sőt esetenként a hármat nem is lehet elválasztani egymástól (például az egyetemi klinikákon). Az alkalmazás rendszerint a tudo-

\* Tanulmányuk megírása során a szerzők a Központi Statisztikai Hivatal több munkatársával folytattak eszmecsere-t. Kollégáik értékes észrevételeiért, tanácsaiért, valamint a Magyar Tudományos Akadémia Kutatásszervezési Intézetének megértő támogatásáért ez úton mondanak köszönetet.

mányághoz kötődő foglalkozási ágak, szakmák kereteiben történik. A statisztikaalkalmazók tevékenysége, ennek folytán, egy szakma, a statisztikusszakma gyakorlása. A statisztika szavunk tehát tanulmányunkban két fogalmat jelöl: egy tudományágot és egy szakmát. Nem függetlenek egymástól, de nem is azonosak. Kísérletet teszünk mindkettő körülírására.

## A STATISZTIKA TUDOMÁNYA

*Ernst Engel*, ismert német statisztikus és közgazdász, a porosz statisztikai hivatal hajdani igazgatója szerint a statisztikának több mint 180 definíciója ismeretes (*Rédei* [1954]). Sietünk megnyugtatni az olvasót, nem vesszük sorra mind a 180-at. Csupán a második világháború utáni magyar szakirodalomban tallózunk, az ötvenes évektől máig. Legutóbb 1999-ben jelent meg egy tanulmány a *Statisztikai Szemlében* „Gondolatok a statisztikáról” címmel (*Hunyadi–Rappai* [1999]). Bár a *Szemle* a tanulmányt vitaindító-nak szánta, a cikk nem vert nagy hullámokat sem a tudományos közéletben, sem a statisztikustársadalomban. Ennek oka feltehetően az, hogy a szerzők fő megállapításával, tudniillik azzal, hogy a statisztika önálló tudományág, mégpedig módszertudomány, a statisztikusok többsége egyetért (másokhoz, például a matematikusokhoz és más potenciális vitapartnerekhez a folyóirat valószínűleg nem jut el).

A statisztika mibenlétéről valódi vita legutóbb az ötvenes évek első felében folyt. A statisztika fogalmáról és tárgyáról a *Statisztikai Szemle* 1953. évi II. félévi és az 1954. évi I. félévi számaiban jelentek meg tanulmányok (*Rédei* [1953], [1954]; *Gyulay* [1954]), részben kapcsolódva szovjet szerzők hasonló tárgyú tanulmányaihoz (*Plosko* [1953], *Kuvsinnyikov–Szmulevics* [1954]). A különböző, gyakran ellentétes vélemények ütköztetése végül 1954-ben a Központi Statisztikai Hivatal (KSH) által kezdeményezett, *Péter György*, a Hivatal akkori elnöke részvételével tartott vitaülésen került sor. A vitában egyetemi statisztikaoktatók, gyakorló statisztikusok, matematikusok és közgazdászok vettek részt. (*A statisztika fogalma...* [1954]) A publikált tanulmányokban és az ülésen a részt vevők által megfogalmazott nézetek négy, egymástól eltérő álláspontban foglalhatók össze.

*A statisztika nem tudomány.* E nézetet legerőteljesebben képviselő résztvevők szerint a statisztika nem tár és nem is tárhat fel statisztikai törvényszerűségeket, ilyenek ugyanis nincsenek. A statisztika feltárhat rendszeres ismétlődéseket, vagy másféle szabályszerűségeket; azt azonban, hogy törvényszerűségekké állunk-e szemben, csupán az illetékes szaktudomány tárhatja fel, a statisztikai módszer tehát nem önálló tudomány. A másik, hasonló felfogást képviselő nézet egy kivételt tesz: kivétel a matematikai statisztika, amely azonban a matematika (a valószínűség-számítás) része.

Ezen felfogás képviselői – leegyszerűsítve s meglehetősen eltúlozva – a statisztikát a statisztikai évkönyvvel azonosítják. Finomabban fogalmazva: csak a szakstatisztikákra (gazdaságstatisztika, népesedésszatisztika stb.) gondolnak. Ez utóbbiak valóban jórészt a szaktudományok fogalmaival dolgoznak és a feltárt változások megítélése valóban szaktudományos műveltséget (közgazdaságtan, demográfia stb.) igényel. Azonban kívül esik látókörükön vagy érdeklődésükön az az elméleti, módszertani háttér, amelyet a statisztikai nyelvezet általános statisztikának nevez és amelyhez képest a szakstatisztikák alkalmazásának minősülnek.

*A statisztika tudományos módszer, de nem önálló tudományág.* Ennek a felfogásnak a megítélése azért körülményes, mert hiányzik belőle a „tudományos módszer”-nek mint fogalomnak az egyértelmű definíciója. Tartalmaz egy „negatív” meghatározást, amely a tudomány fogalmából indul ki. Eszerint a tudomány a természetről, a társadalomról és a gondolkozásról felhalmozott, bizonyítható ismeretek összessége. Az azonban, hogy miként lehet az ismereteket összegyűjteni, összefoglalni, rendszerezni, önmagában nem tudomány, így a statisztika sem az. Ez, legalábbis, vitatható. Hogyan lehet egy új gyógyszermolekula hatását a fehér egerek halmazain vizsgálni, vagy miként lehet egy távoli galaxison a vöröseltolódás tényét igazolni, az éppen úgy része a tudománynak, mint maga az eredmény. (Emlékeztetnénk itt a „metodológia” fogalmára, amely – mint a tudományos módszerek tana – az általános tudományelmélet része).<sup>1</sup> A „tudományos módszer” felfogás képviselőinek szavaiból közvetve egy másik definíció is följellik, amikor elvétve a „statisztikai számvitel” kifejezést használják. Itt – akaratlanul is – a statisztikát valószínűleg a könyvvitellel hozzák kapcsolatba. Meglehet, a könyvvitel – mivel a közgazdaság-tudomány logikájára épül – valóban tudományos módszer. De a statisztikának saját logikája, önálló fogalmi rendszere van, a statisztika és a könyvvitel – bár a gazdaságstatisztika révén van közös területük – nem hasonlíthatók össze.

*A statisztika társadalomtudomány.* Ezen felfogás szerint a statisztika a gyakorlatban alkalmazott módszerek és eredmények (mutatószámok) összefoglalását képviselő tudományág. Tárgya a társadalmi jelenségek mennyiségi mozzanatainak a vizsgálata két kiegészítéssel: *a)* a mennyiségi mozzanatok mellett a statisztika képes a minőségi elemek megragadására is; *b)* másrészt a statisztika vizsgálja a társadalmi jelenségek „előfeltételeit alkotó” természeti jelenségeket is, hozzátéve, hogy az anorganikus természettudományokban a tudományos megállapítások legfontosabb formája a matematikai képlet, ebben a körben a statisztikai módszereknek csak kis szerepük van. Itt nem tudni, hogy „a statisztika által vizsgált természeti jelenségek” megszorítás („előfeltételeit alkotó”) pontosan mit jelent. De ha akkoriban, 1954-ben, esetleg indokolt is volt a statisztikával kapcsolatban a természeti jelenségek ilyen formában történő szűkítése, napjainkra a természettudományok, a műszaki, orvos- és agrártudományok olyan széles körben alkalmaznak statisztikai módszereket, hogy a statisztikát társadalomtudománynak minősíteni tarthatatlan.

*A statisztika módszertudomány.*<sup>2</sup> Ezen állítás gondolatmenetéből a következő tételeket kívánjuk kiemelni.

– A statisztika vizsgálódási köréből nem lehet kirekeszteni a természettudományok tárgyát képező jelenségeket.

– Ennélfogva a statisztika nem társadalomtudomány, hanem a társadalmi és természeti jelenségek széles körének vizsgálatára szolgáló módszertudomány.

– A statisztikai tudomány azonban nem egyszerűen módszerek gyűjteménye, nem receptkönyv, a statisztikai tankönyv nem „hasznos tudnivalók statisztikusok számára”. A statisztika egységes tudományág, saját alapfogalmai vannak: a sokaság és az ismérv.

– A statisztika rendszerezi a sokaságok és az ismérvek típusait és a sokaság tagjait jellemző ismérvek közötti kapcsolatokat vizsgálja. Az ismérvek közötti kapcsolatok feltárásának legegyszerűbb, de alapvető módja az osztályozás.

<sup>1</sup> Magyar Nagylexikon, XIII. köt. 52. old.

<sup>2</sup> Az itt felvázolt gondolatmenet túlnyomórészt *dr. Köves Pálnak* az említett 1954-es vitaülésen elhangzott felszólalását követi, illetve részben idézi.

– A statisztika tudományának középpontjában az általános statisztikának nevezett diszciplína áll, statisztikatudományon elsősorban ezt kell érteni, de a szakstatisztikák, mint az általános statisztika alkalmazásai, is részei a statisztika tudományának.

– „Az a körülmény, hogy minden szakstatisztikában ugyanolyan módszereket alkalmaznak, rámutat arra, hogy itt a jelenségeknek egy bizonyos jellegzetességéről van szó, ami mindezt lehetővé teszi. Van olyan tárgy, amely más tudománynak nem tárgy. De ez a tárgy nem a társadalmi jelenségek általában, vagy a társadalmi jelenségek számszerűsége, hanem a jelenségeknek az a vonása, hogy statisztikát lehet róluk készíteni. A statisztika tudományának a másik tárgya maga a statisztikai tevékenység.” („A statisztika fogalma és tárgya” című vitautülés anyaga. *Statisztikai Szemle*, XXXII. évf. 1954. 522. oldal).

Ezeket a nézeteket nem lehet közvetlenül személyekhez kötni (az utóbbi gondolatmenet kivételével) annál kevésbé, mert a vitában részt vevők egyike-másika a vita során változtatott felfogásán. (Az itt leírtak az érveléseket leegyszerűsítik, meglehetősen, esetenként önkényesen; egyes mozzanatokat kiemelnek, másokat elhanyagolnak. Ezért ajánljuk, hogy a *Szemle* érdeklődő olvasói olvassák el legalább a szóbeli vita összefoglalóját (*A statisztika fogalma...* [1954]). Ugyanakkor a jelen tanulmány szerzői nem kívánnak az olvasók előtt a vita résztvevőinek bölcs ítéseiként feltűnni, sőt véleményük szerint – bár a statisztika tudománya és alkalmazása az elmúlt ötven évben óriási mértékben fejlődött – ez volt a magyar statisztikai szakirodalomban a legszínvonalasabb, legmélyebbre hatoló vita (valódi vita) a statisztika mibenlétéről. Gyakorló statisztikusként pedig, tapasztalataink alapján kijelentjük, hogy (egyetértve Köves Pál okfejtésével) a statisztika önálló tudományág, de nem társadalom-, hanem módszertudomány. Ehhez a következő kiegészítéseket fűzzük.

– A statisztika tudományának középpontjában – az elterjedt szóhasználattal – az általános statisztika áll. A statisztikai gyakorlat viszont túlnyomórészt a különböző szakstatisztikák művelését jelenti.

– A szakstatisztikák háromféle fogalomkészlettel dolgoznak: az általános statisztika fogalmaival (például gyakoriság, középérték, szóródás, mintavétel, trend, korreláció stb.); a szakstatisztika tárgyát képező fogalmakkal (például ipari termelés, árváltozás, csecsemőhalálozás, középfokú oktatás, a háztartások fogyasztása stb.); informatikai fogalmakkal (adatbázis, internet-, tábla- és szövegszerkesztés, metaadat, elektronikus adatgyűjtés, fájl stb.).

– A szakstatisztika művelése olyan folyamat, amely az adatigény megfogalmazásától az évkönyv megjelenéséig tart. A folyamat fontosabb fázisai: a módszertan kidolgozása, a fogalmak alkotása, a kérdőív szerkesztés, a mintavétel, az elemi adatok felvétele, fogadása, ellenőrzése, pótlása, becslési algoritmusok kidolgozása és alkalmazása, aggregálás, az adatok tárolása, kiértékelése, elemzése, nyilvánosságra hozatala, a folyamatok dokumentálása. Mindez egy nagy méretű, bonyolult, szellemi és anyagi erőforrásokból, infrastruktúrális elemekből, statisztikai adatokból, egymásba épülő alrendszerekből álló, önszervező és tanulorendszer, amelyet a szaknyelv statisztikai információ-rendszernek nevez. (*Végyári* [1993]) A rendszer elmélete, kidolgozásának, megszervezésének, fenntartásának és fejlesztésének módja is szerves része a statisztika tudományának.

A statisztika tehát önálló tudományág, de hol van a helye a tudományok rendszerében?

## A STATISZTIKA A TUDOMÁNYÁGAK OSZTÁLYOZÁSÁBAN

A statisztika elhelyezése a tudományok rendszerében nem okozna nehézséget, ha létezne valamiféle tudományrendszertan, tudomány-rendszerezés, hasonlóan például a Mendelejev-féle periódusos rendszerhez. Ilyen rendszertan kidolgozását a tudományfilo-

zófiától, vagy a „tudományok tudománya” (tudománytan) nevű diszciplínától lehetne elvárni. Bár a tudományágak közötti kapcsolatok feltárására születtek bonyolult sémák, és „üres” mátrixok, mai ismereteink szerint a tudományos közvélemény által elfogadott tudomány-rendszertan nem létezik. (*Farkas* [1981], 63–108. old.)

A tudományok bármiféle rendszerezéséhez szükség lenne a tudományágak közötti határvonalak meghúzására, a határesetek megítélésére. A tapasztalatok szerint azonban a határvonalak rugalmasak, változékonyak, gyakran elmosódnak. Az a vélemény is megfogalmazódik, hogy a tudományt feldarabolni felesleges, a tudomány egy és egységes. (*Megérteni ...* [2003].) Ha egyetemességre igényt tartó rendszerezés nem is, de tudományági osztályozások, regiszterek léteznek. Egy részük melléktermék, például a tudományos szakemberek tudományterületek szerinti csoportosítása. Készültek osztályozások különböző enciklopédiák, lexikonok címszavai alapján is (*Farkas* [1981], 67–70. old.). Mindezek azonban inkább felsorolások, semmint osztályozások.

Kifejezetten az osztályozás rangjára igényt tartó, nemzetközileg is elfogadott és használt tudományági osztályozás a múlt század hetvenes éveiben készült, elsősorban a statisztika – konkrétan a tudományos kutatás és fejlesztés statisztikája – igényei alapján. Mielőtt erről szót ejtenénk, rövid kitérőt teszünk a különböző szakstatisztikák által kidolgozott és használt osztályozások egyes kérdései és problémái felé, elsősorban a hivatalos statisztika gyakorlata alapján.

Mint ismeretes, a szakstatisztikák szükségletei alapján kidolgozott osztályozások egy részét – elsősorban azokat, amelyeket több szakstatisztikában használnak, – a hivatalos statisztika szabványosítja, szabványos osztályozásként kezeli. Ilyenek például a tevékenységek (ágazatok), a termékek, a szolgáltatások, a betegségek, a jövedelmek, a beruházások, a balesetek stb. osztályozásai. Mindezek (és mások) nemzetközi szabványként is funkcionálnak, különböző nemzetközi szervezetek (ENSZ, Európai Unió, OECD, UNESCO stb.) ajánlásai és koordinációja alapján.

Az osztályozás során a statisztika az adott sokaság elemeit olyan egynemű részsokaságokra bontja, amelyek egyesítése kiadja a teljes sokaságot, s a részsokaságoknak nincs közös részük. Szükség van az osztályozás tárgyát képező sokaság, valamint az egyes osztályok (csoportok, alcsoportok) definiálására is, különös tekintettel a határesetekre; definíciók nélkül az osztályozások, illetve azok osztályai gyakran nem értelmezhetők. Az osztályozások kidolgozása közben a statisztikus a sokaság elemeit oly módon sorolja osztályokba, hogy felteszi a kérdést: mit mivel rokonítson? Mit mitől válasszon el? Eközben gyakran kisebb-nagyobb dilemmák elé kerül. Példaként bemutatjuk, hogy a rokonítással-elválasztással kapcsolatos néhány ismert, leegyszerűsített kérdést hogyan fogalmazzunk meg a termékosztályozásokkal kapcsolatban. A kitermelt terméskő bányászati termék vagy építőanyag? A versenyautó jármű vagy sportszer? A műanyag edény a fémedényekkel tartozik-e egy osztályba vagy a műanyagcikkkel? A fotópapírt a fényképezőgéppel rokonítsuk-e vagy a papírárúkkal? A takarmánykeverék malomipari termék-e (mint például a liszt) vagy állati eledel? A dinnye zöldségnövény (mint az uborka, a tök) vagy gyümölcs? Mind eme példákkal csupán azt kívántuk szemléltetni, hogy a statisztikai osztályozásokkal kapcsolatban nincs mindig kétséget kizáróan helyes megoldás, a gyakorlat esetenként kompromisszumokra kényszerül. Ilyen dilemmák és kompromisszumok a tudományági osztályozások körében is lehetségesek.

A már korábban említett, statisztikai célra készült tudományági osztályozás az UNESCO ajánlása alapján került nemzetközi alkalmazásra; a kilencvenes évek végéig ezt használták a hazai hivatalos statisztikában is (1. Függelék). A szóban forgó osztályozás már az első áttekintés alapján is elavultnak látszik. Az utóbbi évtizedekben a tudomány fejlődésének egyik jellegzetes vonása a tudományágak közötti határvonalak elmosódása, kialakultak az interdiszciplináris jellegű határtudományok, elsősorban a természettudományok körében (fizika, csillagászat, kémia, biológia). Más tudományágak terjeszkedése a szóban forgó tudomány jellegét is megváltoztathatja. A közelmúltban például véleménycsere folyt arról, hogy természettudomány-e a pszichológia (Garai [1994]). A kérdésre a pszichológusok korábban igennel válaszoltak (a pszichológusképzés az anatómiával, az élettannal, az etológiával kezdődik), a pszichológia a többi természettudománytól csak tárgyának nagyobb fokú bonyolultságában különbözik. Ez a felfogás – laikusként leegyszerűsítve – azáltal vált kérdésessé, hogy a pszichológiának olyan új ágai alakultak ki, mint például a szociálpszichológia, a szervezetpszichológia, a gazdaságpszichológia. Hasonló a helyzet a földrajzzal, amely természetföldrajzként a földtudomány egyik ága. A társadalomföldrajz kialakulásával azonban a földrajz elhelyezkedése a tudományágak osztályozásában kérdésessé válik. Az ilyen és hasonló fejlemények elhatárolási problémákhoz vezetnek. Kézenfekvő példával: a gazdaságpszichológia inkább gazdaságtan, vagy inkább pszichológia-e? Erre a kérdésre aligha lehet egyértelmű választ adni. Hacsak nem az *Ilf–Petrov* szerzőpárost idézzük, akik – arra a kérdésre, hogy miként tudják műveiket együtt írni – azt a komolytalan választ adták: egyikünk írja a magán-, másikunk a mássalhangzókat (Garai [2003]). A gazdaságpszichológia minden bizonnyal interdiszciplináris tudományág.

Az 1. Függelékben leírt osztályozás azért is elavult, mert hiányoznak belőle a már kidolgozása idején is létezett olyan tudományágak, mint például a hadtudomány, a politológia, a vallástudomány. Feltehetően az UNESCO által ajánlott osztályozás kritikájaként, hazánkban 2000-ben új tudományági osztályozás jelent meg egy kormányrendelet mellékleteként (2. Függelék). Ez az osztályozás kétség kívül korszerűbb megoldás az UNESCO-ajánlás számos problémájára. Legfontosabb vonásai, hogy 1. nemcsak a tudományágakat, hanem a művészeteket is osztályozza; 2. a társadalomtudományoktól elkülöníti a bölcsészettudományokat;<sup>3</sup> 3. minden tudományterület kiegészül egy „multidiszciplináris” sorral, az ilyen tudományágak mintegy helyet kapnak az osztályozásban; 4. megjelennek a másik osztályozásban hiányolt olyan tudományágak, mint például a műszaki tudományok között a katonai műszaki tudomány és az informatika, a természettudományok között a környezettudomány, a társadalomtudományok körében a hadtudomány, a politológia és a vallástudomány stb.

Nehezen magyarázható azonban, hogy egyes tudományágak – legalábbis önálló címszóként – eltűntek az osztályozásból, például a csillagászat, az energetika, a vízügyi tudomány, a talajtan, a demográfia.

A KSH jelenleg (és mindaddig, amíg be nem vezetik az Európai Unió statisztikai hivatalában kidolgozás alatt álló új, remélhetőleg korszerű nemzetközi ajánlást) egy olyan

<sup>3</sup> Bár a bölcsészettudományoknak – ismereteink szerint – nincs objektív definíciója (a Magyar Larousse Enciklopédia szerint a bölcsészettudományok azok a tudományágak, amelyeket a bölcsészkarokon oktatnak), a bölcsészettudományok elkülönült megjelenését nem kifogásoljuk, mindazonáltal szerencsésebb lett volna a „humán tudományok” kifejezést használni.

„vegyes” osztályozást használ a tudományos kutatási statisztikában, amely mind a nemzetközi együttműködésben még érvényes UNESCO-féle, mind a hazai gyakorlatba bevezetett új osztályozással kompatibilis.

Az 1. és a 2. Függelékben leírt két osztályozás – mint látható – sokban különbözik egymástól, egyben azonban teljesen azonosak: nem ismerik sem a módszertudományokat, sem a statisztikát. A hivatalos statisztika a megfelelő szakstatisztikában minden, vagy csaknem minden tudományt számon tart, kivéve a statisztika tudományát. Ha az illetékes szakstatisztikusok, felvételeik során, statisztikai kutatóhelyeket észlelnek, azok adatait a matematikai kutatóhelyek adataihoz sorolják.

Még két olyan osztályozásról teszünk említést, melynek közvetlen célja nem a tudományágak osztályozása, de összefügg vele. Az egyik a Magyar Tudományos Akadémia által használt rendszer az Országos Tudományos Kutatási Alapprogram (OTKA) céljaira. Az osztályozás szerkezetének vannak figyelemreméltó vonásai, például az, hogy a tudományokat három nagy csoportba sorolja: *a)* a társadalomtudományok, *b)* az élettelen természet tudományágai, és *c)* az élettudományok.

A részletezés viszont – céljának megfelelően – nem a tudományágakat, hanem a hazai tudományos műhelyeket, kutatási témákat képezi le. Az orvostudomány például 28 részletezést tartalmaz (példaként: anyagcsere-kutatások, endokrinológia, hematológia, járványügyi kutatások stb.). Szerepel viszont az osztályozásban a statisztika, mégpedig a társadalomtudományok, azon belül a gazdaság- és jogtudományok között a következőképpen (lerövidítve):

1. Társadalomtudományok
13. Gazdaság- és jogtudományok
1301. Demográfia
1302. Ipargazdaságtan
1303. Jövőkutatás
1304. Közgazdaságtan
1305. Politikatudomány
1306. *Statisztika*
1307. Szociológia
1308. Állam- és jogtudomány

A megoldás hasonló ahhoz, ahogy az MTA Statisztikai Bizottsága is az Akadémia Gazdaság- és Jogtudományok Osztályának égisze alatt tevékenykedik. (A *Magyar Tudományos Akadémia Almanachja* [1997]) Az a látszat, mintha az MTA szerint a statisztika társadalomtudomány lenne.

A másik itt említendő osztályozás az oktatási statisztika céljaira készült nemzetközi szervezetek együttműködésével; a magyar KSH 2003. június 1-jétől alkalmazza (*Képzési területek...* [2003]). Az osztályozás – céljának megfelelően – képzési főirányokat, képzési irányokat, s azokon belül képzési szakirányokat különböztet meg. Egyik képzési szakirány a statisztika, a főként természettudományokat magába foglaló képzési főirány keretei között, a következőképpen:

4. Matematika, számítástechnika, egyéb természettudományok
- 4.2. Élettani tudományok
- 4.2.1. Biológia és biokémia
- 4.2.2. Környezeti tudományok

- 4.4. Fizikai tudományok
  - 4.4.1. Fizika
  - 4.4.2. Kémia
  - 4.4.3. Földtani tudományok
- 4.6. Matematikai tudományok
  - 4.6.1. Matematika
  - 4.6.2. Statisztika
- 4.8. Számítástechnika
  - 4.8.1. Számítástechnikai tudományok
  - 4.8.2. Számítástechnikai alkalmazások

A rendszer a képzési szakirányok mindegyikét definiálja. A statisztikához a következő definíció tartozik. „A statisztika az adatok gyűjtésére, csoportosítására, elemzésére irányuló tanulmányok összessége. A valószínűség-számítás és a biztosítási matematika is e tudományterület része. A következő főbb képzési programok tartoznak ide:

- biztosítási matematika (aktuáriusképzés),
- matematikai (elméleti) statisztika,
- valószínűség-számítás,
- alkalmazott statisztika,
- felvételtervezés,
- adatgyűjtési (mintavételi) technikák.

Az elméleti matematikai statisztika is ide tartozik. A demográfiát viszont nem ebbe a tématerületbe, hanem a szociológia és kulturális tanulmányok kategóriába kell sorolni.” (*Képzési területek...* [2003])

Tanulmányunk ezen fejezete szerint a hivatalos statisztika által használt tudományági osztályozásokban nincs helye a statisztika tudománynak. Más, nem ilyen céllal készült osztályozások viszont eltérően ítélik meg a statisztikát: az egyik a társadalomtudományok, a másik, a matematikával és számítástechnikával együtt, a természettudományok közé sorolja.

Véleményünk szerint nem a statisztikusok dolga állást foglalni arról, hogy lehetséges-e egy univerzális tudományrendszerezés kidolgozása, s egyáltalán a tudomány „feldarabolására”, ágakra bontására magának a tudománynak szüksége van-e. Állítjuk azonban, hogy a statisztikai gyakorlatnak – ez idő szerint – szüksége van egy tudományági osztályozásra, nem öncélúan, hanem a statisztika felhasználói (tágabb értelemben a társadalom) számára.

Visszatérve az eredeti témánkhöz, föl kell tennünk azt a kérdést, hogy a statisztikusok – kutatók, oktatók, alkalmazók – hogyan foglaljanak állást a statisztika besorolásáról egy új, ismereteink szerint kidolgozás alatt álló osztályozásban? Egyáltalán, állást foglaljanak-e?

Az egyik lehetséges válasz, hogy ne foglaljanak állást. A statisztika nem attól tudomány, hogy szerepel valamilyen rendszerezésben, a statisztika tudománya és alkalmazása eddig is, ezután is fejlődött és fejlődik akár van, akár nincs helye egy tudományági osztályozásban. A kérdést rá lehet bízni a „tudományok tudománya” szakértőire. (No meg a kutatási statisztika művelőire, akik a statisztikai kutatóhelyek adatait akkor is be fogják sorolni valahová – legjobb tudásuk szerint – ha a „statisztika” címszó nem is szerepel az osztályozásban).



A másik lehetőség az állásfoglalás, vagyis annak vizsgálata, hogy a statisztika mivel rokonítható, mitől választandó el.

Arra a kérdésre, hogy a statisztika mivel rokonítható, kézenfekvő a válasz: a matematikával. A következő megfontolások szólnak mellette.

- A matematikának és a statisztikának van közös területe: a matematikai statisztika.
- „Kétségtelen, ... hogy a statisztikai módszertan jó része matematikai alapokra épül. Statisztikai állítás nem mondhat ellent matematikai állításoknak, a statisztikai állítások – ugyanúgy, mint a matematikaiak – bizonyítandók és gyakran matematikai eszközökkel bizonyíthatók.” (Hunyadi–Rappai [1999] 9. old.)
- Korunk tudományának egyik jellegzetessége a matematizálódás. Ugyanez elmondható a statisztikáról is: egyre több tudományágban – a csillagászatól a történettudományig – használnak statisztikai módszereket. Vergilius hexametereit is vizsgálták matematikai eszközökkel; költők szókincseről, rímeiről is készült statisztikai elemzés.
- A matematikával kapcsolatban föl lehet tenni azt a kérdést: a matematikai tételeket feltalálják-e (kitalálják), avagy felfedezik? (Davies [2000] 136–139. old.) A kérdésre a matematikusoknak kell válaszolniuk. Nekünk, statisztikusoknak, akik áhítattal szemléljük a matematika hatalmas építményét, a Ludolf-féle szám jut eszünkbe (a kör területének és átmérőjének viszonya), amelyet – gondoljuk – a természet „talált ki”, az ember pedig felfedezte. De tovább lehet lépni. A prímszámok például elvontan akkor is léteztek, mielőtt a matematikusok felfedezték volna őket? Ez már a filozófia világába vezet. Csupán arra kívánjuk felhívni a figyelmet, hogy ugyanezek a kérdések a statisztikával kapcsolatban is feltehetőek. Az a tény, hogy egy sokaság tulajdonságai egy alkalmas módon vett minta alapján is megbecsülhetők, felfedezés vagy találmány? Az indexszámok közötti összefüggések akkor is igazak voltak, mielőtt az első indexszámot leírták volna? E tanulmány szerzői nem vállalkoznak arra, hogy válaszoljanak az ilyen kérdésekre, s a kérdések fontosságát (avagy lényegtelen voltát) sem ítélik meg. Csupán jelezni kívánják: ezek rokonítható kérdések rokonítható tudományágakra vonatkozóan.

Az előbbiekkal nem azt kívántuk kifejezni, hogy a statisztika csupán a matematika egyik ága, netán azonos a matematikával. Csupán azt kívántuk érzékeltetni – amennyiben a statisztikát el kívánjuk helyezni a tudományágak osztályozásában – helyét a matematika közelében kereshetjük, például így:

1. Természettudományok
  - 1.1. Matematika, statisztika
    - 1.1.1. Matematika
    - 1.1.2. Statisztika
  - 1.2. Fizika

Lényegében ehhez hasonlóan – a természettudományi ágak közé – sorolja be a statisztikát a képzési szakirányok korábban említett osztályozása is. Ennek ellenére a statisztika természettudományként való besorolása éppúgy vitatható, mint a társadalomtudományi besorolás. Valamely sokaság elemeinek osztályokba sorolásakor – mint korábban már utaltunk rá – sajátos dilemma elé kerülhetünk. A versenyautó például az egyik szempont szerint jármű, egy másik szerint sportszer. Egyik besorolás sem jó, de nem is rossz.

Ilyen alapon a statisztika tudományágak közé sorolásakor egy másik logikát is követhetünk. A statisztika különböző ágait, a szakstatisztikákat, nemcsak a matematikával rokoníthatjuk, hanem ugyanolyan alapon más tudományágakkal is, a népesedéssziszti-  
kát a demográfiával, az igazságügyi statisztikát a jogtudománnyal, az egészségügyi statisztikát az orvostudománnyal stb. A statisztika tudományának azon ága pedig, amelynek tárgya az, hogy miként kell statisztikát „csinálni”, más terminológiával: a statisztikai

információrendszernek az elmélete, az informatikával és a rendszerelmélettel áll szoros kapcsolatban.

Amikor korábban a 2. Függelékben leírt osztályozást értékeltük, s kiemeltük, hogy – helyesen – minden tudományterületen (természettudományok, műszaki tudományok stb.) belül megjelentek az interdiszciplináris tudományágak, nem tettünk említést egy hiányosságról, arról, hogy nincs „helyük” azoknak az interdiszciplináris tudományágaknak, amelyek nem az egyes tudományterületeken belül, hanem azok határvidékein alakultak (alakulnak) ki. Ilyennek foghatjuk fel a földrajzot, vagy a meglehetősen széles összefüggésekkel operáló jövőkutatást, de a sok tudományággal érintkező statisztikát is. A hat nagy tudományterület mellett tehát helye lehetne egy hetediknek is: az „egyéb interdiszciplináris tudományok”-nak, vagy az „egyéb, máshová nem sorolható tudományágak”-nak. Ennek a megoldási elképzelésnek nincs sok köze valamilyen tudományrendszerezési filozófiához, inkább statisztikai szemléletet fejez ki (minden osztályozásnak kell tartalmaznia egy „egyéb” osztályt), de végül is egy statisztikai célú osztályozásról van szó.

Mindkét besorolási megoldás mellett és ellen megfontolásra érdemes érvek állíthatók fel. E tanulmány szerzőinek a szemléletéhez a második besorolási megoldás áll közelebb.

#### A STATISZTIKUSSZAKMA

A statisztika szó, ahogy ezt a bevezetőben is említettük, nemcsak tudományágat, hanem gazdasági tevékenységet is jelent. Az utóbbi értelemben a statisztika a valóság, – a népesség, a gazdaság, a társadalom, a környezet, a természet – egyedeit, jelenségeit, folyamatait figyeli meg, a megfigyelés során információkat gyűjt, azokat feldolgozza és hasznosítja. E tevékenységek folytatásához szükséges tudományos (elméleti) ismeretek, gyakorlati tapasztalatok és készségek összessége a statisztikusszakma.

A szakma gyakorlásához szükséges tudományos ismeretek teremtik meg a kapcsolatot a statisztikatudomány és a statisztikusszakma között. Az előző részben ezen ismeretek körét úgy határoztuk meg, hogy a statisztika tudományának a középpontjában az általános statisztikai ismeretek állnak, de magában foglalja a szakstatisztikákat is, mint az általános statisztika alkalmazásait (*Szilágyi* [1975], 726. old.), és a statisztikai tevékenység (statisztikai információrendszer) megszervezésének, fenntartásának és fejlesztésének ismereteit is.

A statisztikatudomány alkalmazása, vagyis a statisztikusszakma gyakorlása az esetek egy részében főtevékenységként, másik részében segédtevékenységként jelenik meg a gyakorlatban. Főtevékenységként végzik azokon a területeken, ahol a fő cél maguknak a statisztikai adatoknak a létrehozása és nyilvánosságra hozatala, az itt dolgozó statisztikus szakembereknek ez a főhivatásuk. Ebben a körben a statisztikusszakma művelőinek ismeretanyaga egyfelől maga a statisztikai tudomány, másfelől az a tudományág, amely a vonatkozó szakstatisztika műveléséhez szükséges. Az ismert példákkal: közgazdaságtan (gazdaságstatisztika), szociológia (társadalomstatisztika), demográfia (népesedésszatisztika), orvostudomány (egészségügyi statisztika), jogtudomány (igazságügyi statisztika), környezettudomány (környezeti statisztika), földrajz (területi statisztika), politológia (közvélemény-kutatás) stb.

A statisztika alkalmazásának másik ágán a statisztika egy adott tudományágon belül a kutatási célok elérésére, az eredmények értékelésére szolgál. Ilyenkor a statisztika az

adott tudományág művelésében a segédtudomány szerepét tölti be. Amikor például egy gazdaságtörténész a magyarországi gabonaárak alakulását kívánja megítélni a napóleoni háborúk idején, kutatásai során statisztikai módszereket (is) alkalmaz, de számára a statisztika csupán nélkülözhetetlen segédeszköz. (Rácz [1977], 1127. old.)

A statisztikusszakma tartalmát a következőkben első szerepkörében vizsgáljuk, a hivatalos statisztikára összpontosítva. Ebben a körben a statisztikai szakmát alkotó ismereteket három csoportba sorolhatjuk: 1. a statisztikai tudomány általános ismeretei, 2. a szakstatisztikák speciális ismeretei, 3. az információgyűjtéshez és -feldolgozáshoz kapcsolódó ismeretek, beleértve az informatikai eszközök és termékek működtetéséhez szükséges tudást is.

A definíciók szintjén nincs alapvető tartalmi eltérés a statisztika szó két jelentése között, a jelenlegi magyar és nemzetközi gyakorlat azonban azt mutatja, hogy a statisztika mint „tudományág” és a statisztika mint „szakma” az ismeretek körét tekintve különbözik egymástól. A nemzetközi szakirodalom terminológiáját használva „a hivatalos statisztika” (official statistics) és a statisztikatudomány (academic statistics) között szakadék van. A tartalmi eltérést az okozza, hogy a statisztikatudomány, és a hozzákapcsolódó oktatás, jelenlegi gyakorlata nem, vagy csak részben (mintavétel), fedi le az előző felsorolás harmadik csoportját, az információgyűjtéshez és -feldolgozáshoz kapcsolódó ismereteket, ugyanakkor a statisztikusszakmához, a statisztikai munkához ezek az ismeretek elengedhetetlenek.

A statisztikai tudomány által nyújtott ismeretek a statisztikai munkafolyamaton belül szinte kizárólag csak az elemzéshez kapcsolódnak, azaz a gyakorlatban az előzőkben ismertetett „statisztikai tudomány” definíciónak egy szűkített értelmezése érvényesül, mely nem tartalmazza a statisztikai tevékenység (statisztikai információ-rendszer) megszervezésének, fenntartásának és fejlesztésének ismereteit.

A statisztikai szakma „kézműves” időszakában, az információgyűjtéshez és -feldolgozáshoz szükséges ismeretek kizárólagos forrásai a gyakorlati tapasztalatok voltak. A huszadik század közepétől azonban fokozatos változás következett be az ez irányú szakmai ismeretek körében, mivel megjelent a termelési folyamatban, a nagy tömegű információ gyors feldolgozásához igen hatékony eszközrendszert biztosító informatika (számítástechnika). Kezdetekben a statisztikai tevékenység informatikai támogatása jól elkülöníthető, önálló tevékenységként kapcsolódott a statisztikához, így a statisztikai szakmai ismeretek bővülése inkább csak az informatikai és rendszerelméleti fogalmak átvételét jelentette. A korszerű személyi számítógépek és a felhasználóbarát szoftverek megjelenése azonban új helyzetet teremtett. A statisztikusok egyre több adatfeldolgozó rendszer esetében jutottak olyan alkalmazások birtokába, ahol saját maguk végezhetik el az adatelemző, adatjavító és -összesítő munkákat. Fogalmazhatunk úgy is, hogy az informatikai szakmai tevékenység egy része integrálódott a statisztikai szakmába. „Az iparszerű termelési jegyeket felmutató adatgyártással megszűnt a megrendelő-végrehajtó kapcsolat a statisztikus és az informatikus között. A statisztikai termelési folyamatban ma már nagyobb részben rendszerszemléletű informatikai és menedzsment tudásra van szükség.” (Papp [2003]).

A „kézműves” időszakban a tudomány és a szakma közötti különbség nem okozott problémákat a statisztikai tevékenység megfelelő színvonalú végzésében, segítettek a gyakorlati tapasztalatok. Az utóbbi tíz évben azonban olyan jelentős fejlődés következett be az információgyűjtés és -feldolgozás területén, az ismeretanyag olyan szerteágazóvá

és bonyolulttá vált, hogy a tudományos szintű rendszerezés, és különösen a megfelelő szintű oktatás hiánya már veszélyezteti a statisztikai szolgáltatás minőségét, eredményességét. A statisztikai elemzések iránti egyre növekvő társadalmi igény arra készíti a hivatalos statisztikai szolgáltatókat, hogy nagyobb erőket összpontosítsanak a modern információtechnológia által támogatott adat-előállítási rendszerek fejlesztésére. Kézenfekvőnek tűnik az a megoldás, hogy a statisztikai tudomány erőforrásait, elsősorban a statisztika oktatóit, ezen rendszerek elméleti-módszertani hátterének megerősítése érdekében, a korábbiaknál nagyobb mértékben vonják be a statisztikai szolgáltató munkájába. Ez a törekvés jól érzékelhetően jelenik meg a nemzetközi statisztikai szervezetek és egyes nemzeti statisztikai hivatalok kutatási programjaiban. Az Eurostat által meghirdetett kutatások között túlsúlyban vannak azok a témák, amelyek a statisztikai információs rendszer fejlesztéséhez kapcsolódnak. (Lásd a 3. Függelék.) Több országban kialakultak a hivatalos statisztika és a kutatók együttműködésének konkrét formái. Az Egyesült Királyságban és az Egyesült Államokban az akadémiai statisztikát képviselő statisztikai társaságok (Royal Statistical Society – RSS, American Statistical Association – ASA) külön szekciókat hoztak létre a hivatalos statisztikával való kapcsolatok ápolására. Az ASA „Kormányzati statisztika” szekciójának működési körében olyan témákat találunk, mint a statisztikai termelési folyamatban alkalmazható módszerek és eljárások kifejlesztésében való részvétel, képzési lehetőségek teremtése a hivatalos statisztika előállítói és felhasználói számára. Az RSS „Hivatalos statisztika” szekciója évente más és más területekre koncentrálja a hivatalos statisztikát támogató tevékenységét. Ez évi konferenciájának fő témái: a statisztikai integráció, a tájékoztatási tevékenység, az adatvédelem, az adatszolgáltatói terhek.

Magyarországon az információgyűjtéshez és -feldolgozáshoz kapcsolódó kutatási és fejlesztési tevékenység meglehetősen belterjesen, lényegében a hivatalos statisztikai szolgáltatóhoz tartozó szervezetekben, szinte a statisztikai termelési munkafolyamat napi teendőihez kötődően folyik. A kutatási munka eredményei csak néhány tanulmány, illetve konferenciaelőadás formájában jutottak el a szakmai közvéleményhez. A kutatási-fejlesztési munka kiszélesítése a tudomány és a szakma hazai képviselőinek közös érdeke, ugyanis ezáltal a statisztikatudomány és -oktatás jelenlegi fehér foltja felszámolható, a hivatalos statisztika pedig biztosabb elméleti-módszertani hátteret kap napi munkájához.

A tanulmány hátralévő részében e stratégiai törekvéshez azzal szeretnénk hozzájárulni, hogy vázlatosan, a statisztikai termelési folyamat mentén rendszerezve, áttekintjük az információgyűjtéshez és -feldolgozáshoz szükséges ismeretek körét. Leírjuk azt a termelési rendszert, termelési folyamatot, melynek során információkat, konkrétan statisztikai adatokat állítanak elő, és tesznek közzé statisztikai táblák és statisztikai elemzések formájában.

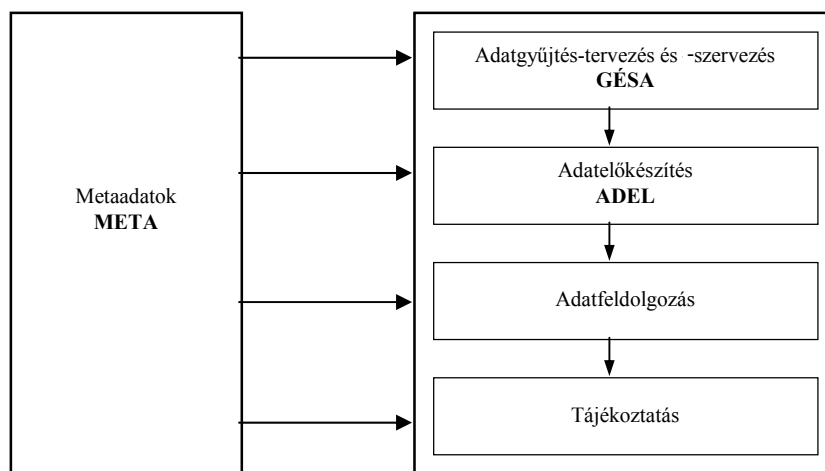
## A STATISZTIKAI INFORMÁCIÓ-RENDSZER TERVEZÉSE ÉS MŰKÖDTETÉSE

A statisztikai termelési folyamat kiindulópontja a felhasználói igények felmérése és lefordítása a statisztikai fogalmak nyelvezetére. Az igényfelmérés utáni tevékenységek a következő, időrendben egymást követő, csoportokba rendezhetők.

- Az adatgyűjtések tervezése és szervezése.
- Az adatok ellenőrzése, javítása és feldolgozása.
- Az adatok közzététele, tájékoztatás.

Magyarország statisztikai tevékenységet folytató legnagyobb szervezete, a Központi Statisztikai Hivatal, esetében a termelési rendszer sematikus ábrája a következő.

*A statisztikai termelési folyamat sémája*



A séma, az általános folyamatleírás mellett, a statisztikai munka két újkeletű jellegzetességét is tükrözi, melyek kiváltója az információ-technológia robbanásszerű fejlődése volt, és jelentős változásokat hozott a szakmai ismeretek körében. Az egyik ilyen új vonás a metaadatok megjelenése a termelési folyamatban, mégpedig jól láthatóan úgy, hogy hatókörük valamennyi szakaszra kiterjed. A másik változást az egyes termelési szakaszok mellett megjelenő fantázianevek érzékeltetik, melyek általánosan használható alkalmazásokat takarnak, támogatva és ösztönözve a statisztikai munka integrációs folyamatát. Egyre több statisztikai hivatal hajt végre jelentős változásokat a statisztikai termelési folyamat szervezésében. Az eddigi gyakorlatban domináns, ún. „kályhacső” („stove-pipe”) típusú adat-előállítási folyamatot, amikor lényegében adatgyűjtésenként külön-külön termelési rendszerek dolgoztak egymás mellett, felváltja a valamennyi adatgyűjtést felölelő, a termelési folyamat minden egyes szakaszában egységes eljárások szerint működő, azonos metaadatok által vezérelt, integrált rendszer. Mielőtt rátérnénk a termelési folyamat egyes szakaszainak vázlatos ismertetéséhez, szólni kell az egész folyamatot vezérlő és irányító metaadatbázisról.

#### *A statisztikai metaadatbázis kiépítése*

Az általánosan elfogadott definíció szerint a statisztikai metaadatok olyan információk a statisztikai adatokról és az előállításukról, melyek elősegítik megfelelő értelmezésüket és felhasználásukat. A statisztikai metaadatbázis a metaadatok logikai struktúra szerint rendezett összessége.

A statisztikai metaadatbázisnak több funkciója van.

1. Dokumentálja a statisztikai információ-rendszer tartalmát, módszereit, a termelési folyamat egyes műveleteit.
2. Menedzseli, koordinálja a teljes termelési folyamatot, biztosítja integrációját.
3. Informálja a felhasználókat a statisztikai adatok tartalmáról és minőségéről.
4. Támogatja a felvételek tervezését.

A metaadatbázis „vezérlő, menedzselő” funkciója abban nyilvánul meg, hogy az adat-előállítási folyamaton belül az egyes programok kiválasztását és elindítását adott metaadatok határozzák meg. A metaadatok felhasználásával egy-egy feladattípus megoldására általános számítástechnikai program írható, amely adott, konkrét esetre történő alkalmazása során a metaadatokat paraméterként használja a feladat elvégzéséhez. Szabványok és névkonvenciók segítségével biztosítja az egységes, áttekinthető fejlesztést és működést.

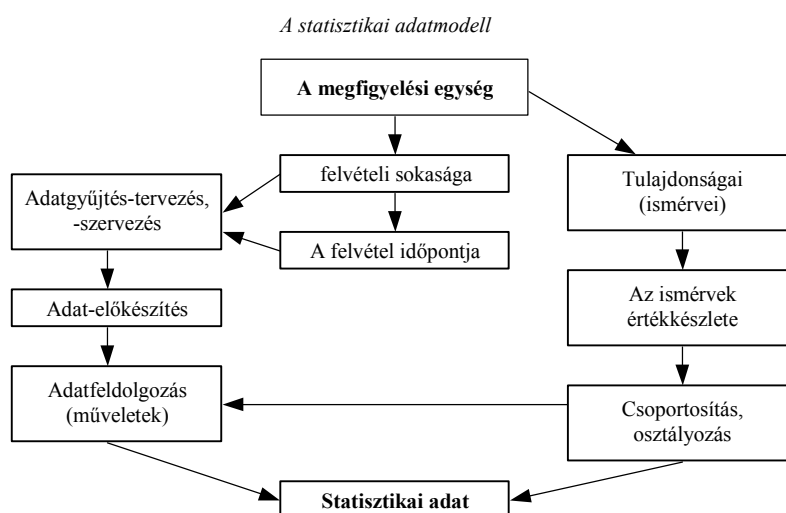
A metaadatbázis „tájékoztató” funkciója mögött az a felismerés áll, hogy a statisztikai adatok felhasználói viszonylag keveset tudnak az adatok előállítási folyamatáról, sőt gyakran az adatok tartalmát leíró mutatók, ismérvek pontos jelentését sem ismerik. Mindezen ismeretek nélkül azonban a statisztikák helyes értelmezése nem könnyű feladat, és az ismeretek hiánya gyakran vezet félreértésekhez, helytelen következtetések levonásához. A fenti problémák elkerülése végett nem elég csak az információt (statisztikai adatot) eljuttatni a felhasználóhoz, hanem információkat is kell adnunk az információkról. A statisztikai kiadványokban fellelhető „módszertani megjegyzések”, illetve a külön kiadványokként megjelenő „módszertani füzetek” ezt a célt szolgálják. Az utóbbi években egyre szélesebb körben terjedő, elektronikus úton elérhető adattárházak azonban új formát igényeltek az adatok tartalmára és előállítására vonatkozó módszertani ismeretek megjelenítésére is. A statisztikai metaadatrendszer biztosítani tudja a statisztikai módszertani ismeretek tárolásának ezen új formáját.

A hagyományos statisztikai módszertani dokumentációkkal szemben a statisztikai metaadatbázisnak nagy előnye, hogy rendszerszemléletű, azaz megköveteli, hogy a statisztikai módszertanban megtestesülő információkat egymással összefüggésben (egy általános modell alapján) dokumentálják. Az általános modell biztosítja azt, hogy

- a statisztikai módszertan dokumentálása során ne maradjanak fehér foltok (a statisztikai adat tartalmára és előállítására vonatkozó valamennyi információ dokumentált legyen),
- az egyes módszertani elemek összefüggése mindenki számára átlátható és érthető legyen,
- az egyes szakstatisztikák azonos struktúrában rögzítsék módszertani ismereteiket.

A statisztikai metaadatokkal kapcsolatos kutatások kezdete a 70-es évekre tehető, arra az időszakra, amikor egyre több kutató definiálta a statisztikát információs rendszerként, és rendszerelméleti megközelítésben próbálta leírni tartalmát és összefüggéseit. A vezető szerepet a skandináv országok statisztikai hivatalai töltötték be, elsősorban a svéd statisztikai hivatal (*Sundgren* [1980]), de a rövid idő alatt nemzetközi szintűvé váló kutató munkában jelentős szerepet töltöttek be a lengyel (*Olenski* [1978]), a csehszlovák (*Soltés* [1979]) és a magyar statisztikusok és informatikusok (*Dörnyei* [1983]) is. A kutatások szervezettebbé tételében jelentős lépés volt, amikor az ENSZ Statisztikai Bizottsága és Európai Gazdasági Bizottsága, az Európai Statisz-

tikusok Konferenciájának munkaprogramja keretében, létrehozta a METIS Group-ot a metaadatokkal kapcsolatos nemzetközi kutatások fórumaként. A kutatómunka eredményeit összefoglaló dokumentum, a statisztikai adatok és metaadatok modellezésére vonatkozó irányelvek (*Guidelines ...* [1995]), máig is meghatározó alapjai a statisztikai metainformációs rendszerrel foglalkozó kutatásoknak. Röviden összefoglalva az irányelvek szerint a jó statisztikai adatmodell egy részről leírja az adat keletkezési folyamatát, más részről azokat a jellemzőit, melyek egyértelműen meghatározzák tartalmát, felvilágosítást adnak jelentéséről.



Ma már a kutatások olyan szerteágazók, hogy harmonizálásukra az EU 5. Kutatási és fejlesztési programjának keretében (MetaNet) az Edinburghi Egyetem koordinációjával és irányításával külön projektet hoztak létre. A magyar informatikusok és statisztikusok a kutatási eredmények gyakorlatba való átültetése során arra törekedtek, hogy egységes, a teljes statisztikai információ-rendszert lefedő, metaadatbázist hozzanak létre. A Központi Statisztikai Hivatalban fejlesztés alatt álló statisztikai metaadatbázis négy alrendszert tartalmaz: 1. mutatók, 2. csoportosítások, osztályozások, 3. általános metaadatok, fogalmak, 4. Adat-előállítási folyamatra vonatkozó adatok.

A mutatók alrendszere tárolja a statisztikai mutatók megnevezését, definícióját, érvényességi időszakát, értékkészletét (az ismérvértékeket). Kapcsolat teremthető az ismérvértékekből képzett csoportosítást vagy osztályozást tartalmazó alrendszerhez, a mutatóhoz kapcsolódó adatgyűjtésekhez, adat-előállítási folyamathoz, a mutató definíciójában szereplő fogalmakhoz. A csoportosítások és osztályozások alrendszere tárolja a statisztikai rendszerben előforduló csoportosítások és osztályozások elemeit, különböző változatait, és kapcsolataikat. Az általános metaadatok alrendszerében található a statisztikai információrendszer általános fogalmai, a statisztikai fogalmak, a gyakorlatban alkalmazott statisztikai módszerek és műveletek általános leírása. A statisztikai adatok minőségére és előállítására vonatkozó információk képezik a metaadatbázis negyedik alrendszerét.

### *Az adatgyűjtések tervezése és szervezése*

A termelési folyamatot vezérlő metaadatbázis vázlatos leírása után térjünk rá az információ-rendszer tervezésének és működésének egyes szakaszainak ismertetésére. Az első szakaszban, mint említettük, az adatgyűjtés tervezéséhez és szervezéséhez szükséges alapinformációkat a felhasználói igények felméréséből nyerjük. A több évtizedes tapasztalatok alapján az alábbi főbb felhasználói csoportok igényeit kell kielégíteni: a lakosság; a média; az államigazgatási és államhatalmi szervek; az üzleti körök és a társadalmi szervezetek; a tudományos kutatás és oktatás, valamint a nemzetközi szervezetek.

Az adatigények felmérésének formái különbözők lehetnek felhasználói csoportoktól függően, de van néhány általánosítható szempont, melyet célszerű figyelembe venni. A statisztikusnak támogatnia, segítenie kell a felhasználókat az igények világos, egyértelmű megfogalmazásában, ezért célszerű, ha vannak ismeretei arról a témakörrel, melyhez a felhasználó az információkat alkalmazni kívánja. A felhasználókat szembesíteni kell azokkal a nehézségekkel, akadályokkal, költségekkel, amelyek az igények kielégítését korlátozzák, és tudatosítani kell velük, hogy a részletesség, a pontosság és a gyorsaság követelménye között fontossági sorrendet kell felállítani.

A felhasználói igények sok esetben nem közvetlenül megfigyelhető jelenségekre, folyamatokra vonatkoznak. A statisztikusnak ezeket az igényeket le kell fordítani a statisztikai fogalmak nyelvére, hogy információkat tudjon gyűjteni róluk. E munkafolyamat végpontja a statisztikai felvétel eredményének, az elemzési munkát segítő munkatábláknak és a közlési tábláknak a megtervezése, ami a statisztikai munka kiemelt fontosságú fázisa, annak egész folyamatára kihat. A kijelölt célsokaság és a vizsgálni kívánt gazdasági vagy társadalmi folyamatokat leíró mutatók a felvételek tervezésének kiinduló szempontjai.

Az adatgyűjtési munkafolyamat megtervezése sorrendben az alábbi munkaszakaszokból áll:

- a kérdőív és az adatgyűjtés-szervezés tervezése,
- a próbafelvétel,
- az adatelőkészítés tervezése,
- az adatfeldolgozás tervezése,
- a költségterv elkészítése.

A tervezési munka mintegy tükörképe annak a termelési rendszernek, amely keretet ad a statisztikai tevékenység szervezett végzésének. Az adatgyűjtés-tervezés és -szervezés munkaszakasz leírásakor elsőként a kérdőív-tervezésről kell szólni. Az adatgyűjtés módjához jól illeszkedő, az adatszolgáltatók érdekeit is figyelembe vevő, jól szerkesztett kérdőívek nagymértékben hozzájárulhatnak a sikeres adatgyűjtéshez. A kérdőív, vagyis a megválaszolendő kérdéseket tartalmazó űrlap funkciója kettős, egyrészt megismerteti az adatszolgáltatóval az adatszolgáltatási kötelezettséget, másrészt mint az adatszolgáltatás teljesítésének közege, adathordozó. Megjelenhet papírra nyomtatva vagy mágneses adathordozón (floppyn), képernyőn. A kérdőív szerkesztése előtt a következőket kell pontosan meghatározni:

- a felvétel megfigyelési körét,
- az adatszolgáltatók potenciális belső nyilvántartásait,



- az adatgyűjtés módszerét és
- az adatfeldolgozás módját.

A kérdőívek kialakítására és tesztelésére célszerű létrehozni egy olyan munkacsoportot, melyben a kérdőív tervezői, a feldolgozás szervezői, az adatszolgáltatók képviselői közösen egyeztetik álláspontjaikat a legmegfelelőbb kérdőív kialakításához. Néhány általános szabály e területen is megfogalmazható.

Kerülni kell az idegen, félreértésre okot adható kifejezéseket, terminológiákat, rövidítéseket, valamint az olyan kérdéseket, amelyekre két vagy több válasz is adható egyszerre. A kérdések sorrendjét úgy kell kialakítani, hogy az adatszolgáltató számára értelmesnek tűnjék, a kérdések logikai összefüggésben kövessék egymást, a hasonló tartalmú kérdések egy csoportba kerüljenek. A kérdőív kitöltési útmutatója lehetőleg legyen rövid és közérthető. A kérdőív egyes rovataihoz tartozó magyarázat könnyen és gyorsan megtalálható legyen. A kitöltési útmutatóban (esetleg külön felkérő levélben) az adatszolgáltatót tájékoztatni kell a felmérés céljáról, az adatok felhasználási módjáról, bizalmas kezeléséről.

A nemzetközi szakirodalomban számos, a kérdőívtervezés szakmai kérdéseivel foglalkozó tanulmány található. A kérdéskör fontosságát jelzi, hogy a Nemzetközi Statisztikai Intézet (ISI) 47. konferenciáján (Párizs, 1989) külön szekciót szerveztek e témakörben.

A kérdőív tervezéséhez szorosan kapcsolódik annak a munkafolyamatnak a megtervezése, amelynek eredményeként a megfelelő kérdőív eljut a megfelelő adatszolgáltatóhoz, illetve visszaérkezik az adatgyűjtőhöz. Ez a munkafolyamat az alábbi részfeladatokat foglalja magába.

- Az adatgyűjtések adatszolgáltatói körének, a felvételi keretnek, a reprezentatív adatgyűjtések esetén a mintavételi keretnek és a mintavétel módjának meghatározása, a mintaelemek kiválasztása.
- Az adatgyűjtés módjának, útjának meghatározása és az ehhez kapcsolódó szervezési feladatok rögzítése.
- Az adatgyűjtési dokumentumok (kérdőív, megszemélyesített kérdőívek, kitöltési utasítás, segédlet) előállítás, eljuttatása az adatszolgáltatókhoz, azaz az expedálás.
- A kérdőívek begyűjtése, az ún. érkeztetés, a szükséges információk regisztrálása az adatszolgáltatást nem teljesítőkről.

Az adatszolgáltatói kör meghatározásához szükség van egy olyan nyilvántartásra (regiszterre), mely azonosítható formában tartalmaz minden potenciális adatszolgáltatót, és azok legfontosabb ismérveit. A regiszterek segédinformációkat tartalmaznak azokról az egységekről, melyekről adatokat akarunk gyűjteni, és ezzel elősegítik az adatgyűjtés megszervezését és az ellenőrzést. A gazdasági szervezetek esetében ezek az információk a következők:

- azonosítási paraméterek (név, cím, azonosító szám),
- rétegtípus paraméterek (tevékenység, méret, gazdálkodási forma),
- demográfiai paraméterek (keletkezés, megszűnés),
- kapcsolati paraméterek (telefon, e-mail).

A pontos és naprakész nyilvántartás jelentős hatással van az adat-előállítási folyamat további tevékenységeire, és közvetve a statisztikai adatok minőségére. Ezért a regiszterek folyamatos karbantartása a statisztikai információ-rendszer egyik fontos funkciója. A regiszterek és általában a felvételi keretek meghatározásához és karbantartásához kapcsolódó szakmai ismeretek fontosságát bizonyítja, hogy 1986-ban, kanadai kezdeményezés-

re, több statisztikai hivatal nemzetközi munkacsoportot hozott létre (International Roundtable on Business Survey Frames), melynek munkaülésein évente egyszer áttekinthetik az e területen jelentkező új feladatokat, illetve az egyes országok eredményeit.

A KSH-ban használt legfontosabb regiszterek: a lakossági címjegyzék, a gazdasági szervezetek regisztere, a kiskereskedelmi regiszter, a gépjármű-nyilvántartás, a földnyilvántartás és a farmregiszter.

Az adatgyűjtés-szervezés egyes munkaműveleteit áttekintve talán nem meglepő, hogy ez volt az a munkaszakasz, ahol a szervezatlenség elkerülése, a hatékonyság növelése érdekében, először jelent meg az egyes felvételek közötti integráció igénye. A KSH-ban a GÉSA adatgyűjtés-szervezési rendszer kifejlesztése szolgálta ezt a célt. Jelenleg ugyan csak a gazdaságstatisztikai adatgyűjtéseket foglalja magában, de ebben a körben általános, konkrét adatgyűjtéstől független rendszer, mely metaadatokat használ valamennyi munkaművelet vezérléséhez.

Az e területen végzett kutatási-fejlesztési munka további eredménye a KSH adatgyűjtés-szervezési rendszerének legújabb alrendszere az elektronikus adatszolgáltatás, amely az adatgyűjtések kapcsolati közegét az élet minden területét átszövő Internetre helyezi át. A modern informatikai technológiák alkalmazásával lehetőséget ad arra, hogy nyomon lehessen követni az adatszolgáltatás teljes folyamatát. Az adatszolgáltató szemszögéből nézve az elektronikus adatszolgáltatás teljes folyamata leegyszerűsítve a következő.

- Az adatszolgáltatók a KSH weboldalán, Interneten keresztül, jelentkezhetnek elektronikus adatszolgáltatásra.
- Szabályozott, adminisztrált úton a KSH-tól lehetőséget kapnak arra, hogy a KSH elektronikus adatgyűjtési rendszerén keresztül küldjék statisztikai adatszolgáltatásaikat.
- A rendszer eljuttatja az adatszolgáltató részére azokat az információkat (megszemélyesített kérdőív, határidő), melyek segítségével az a statisztikai adatszolgáltatási kötelezettségének eleget tud tenni.
- Az adatszolgáltató kitölti a kérdőív adatait, s védett vonalon elküldi a KSH fogadó rendszerének.
- Az adatelőkészítési fázis után az adatszolgáltató értesítést kap az adatok hibáiról, lehetőséget teremtve az adatjavításra.

### *A próbafelvétel*

Minden új adatgyűjtés bevezetése előtt próbafelvételt kell végrehajtani, aminek célja az adatgyűjtés egyes munkaszakaszainak tesztelése, az adatfelvétel költségigényének becslése, az adatszolgáltatói fogadókészség vizsgálata. A próbafelvétel adatszolgáltatói körét a célsokaság jellemzőitől, elsősorban heterogenitásától függően, szakmai szempontok alapján kell kiválasztani. A kérdőív és a kitöltési útmutató tesztelése során értékelni lehet:

- a kérdések érthetőségét;
- azt, hogy a kérdőív felépítése mennyire segíti a helyes és egyszerű kitöltést;
- a kitöltési útmutató teljességét, áttekinthetőségét;
- a kérdőív kitöltésének időigényét.

A próbafelvétel során információkat lehet szerezni arról is, hogy az adatszolgáltató nyilvántartási rendszere lehetővé teszi-e a kérdőív határidőre történő kitöltését. Osztott mintás tesztelés végzésére is van mód, ha a kérdőív két vagy több változata közül a leg-

jobbat akarják kiválasztani. A kitöltött kérdőívek alapján az adatfeldolgozási programokat is tesztelni lehet.

#### *Az adatok ellenőrzése, javítása és feldolgozása*

A statisztikai adatgyűjtések kérdőíveinek rögzítését, ellenőrzését és javítását adatelőkészítésnek nevezik. E tevékenység előfeltétele az ellenőrzési szempontok meghatározása. Az általánosítható szempontok a következők.

- Vizsgálni kell az adatszolgáltató, illetve a megfigyelési egység azonosítójának helyességét és érvényességét, a kódolt adathelyek kódérvényességét és a kódolt adathelyek közti összefüggéseket.
- Vizsgálni kell az értékek számszaki összefüggéseit, valamint a kérdőíven belül az adathelyek egymáshoz viszonyított nagyságrendi összefüggéseit.
- Célszerű vizsgálni az adat előző időszaki értékétől, illetve a hasonló adatszolgáltatók adataitól való eltérés nagyságrendjét.
- Minden ellenőrzési szemponthoz kell rendelni egy hibazonosítót és egy hibaüzenetet.
- Az előforduló hibákat súlyossági fokozatokba kell sorolni a figyelmeztetés, a jelzéssel továbbvihető hiba, a javítandó hiba kategóriák szerint.

Az adatrögzítés, -ellenőrzés és -javítás megvalósítására többféle lehetőség áll rendelkezésre.

- Rögzítés bizonylatolvasóval, javítás automatikus hibajavítással.
- Rögzítés bizonylatolvasóval, javítás intelligens javító alkalmazással.
- Hagyományos rögzítés, javítás intelligens adatbevitellel.
- Intelligens adatbevitel az összeírók vagy az adatszolgáltatók által.

A statisztikai munkában tapasztalható integrációs törekvések az adatelőkészítési munkafázisban is egyre nagyobb teret nyernek. Ma már a fenti tevékenységeket olyan eszközökkel végzik, melyek általánosan használhatók a különböző adatgyűjtésekhez. A KSH-ban ilyen általános adat-előkészítő rendszer az ADEL. Elnevezése az *ADatELőkészítés* szóból képzett mozaikszó. A rendszer tervezésekor és megvalósítása során a fő célkitűzés az volt, hogy egy olyan rendszert alakítsanak ki, mely

- jelentésektől függetlenül biztosítja az egységes működést az adatelőkészítési munka során,
- könnyen kezelhető, jól átlátható szabványos felülettel,
- szervesen integrálódik a meglévő rendszerekhez (META, GESA),
- alkalmas az elektronikusan érkező adatok fogadására.

A rendszer segítségével biztosítható a felvételek adatainak a központi adatbázisba való közvetlen bevitele, az adatok ellenőrzése, javítása. A felderített hibákról, a bevitt adatokról a felhasználó igénye szerinti listák készíthetők. Működése során a rendszer egységes hibakezelésével és megjelenésével a felhasználók számára is biztonságot nyújt. A rendszer biztosítja, hogy az adat-előkészítési munka addig nem fejezhető be, amíg az adatok a megfogalmazott ellenőrzési szempontoknak eleget nem tesznek. Az ADEL rendszerre is érvényes, hogy a metaadatok nemcsak a rendszer elemeinek leírását szolgálják, hanem az alkalmazások működését is vezérlik. Ennek eredményeként a rendszer szinte öndokumentált. A rendszer segítségével a statisztikusok figyelemmel kísérhetik a teljes adat-előkészítési folyamatot. Bármikor lekérdezhetik a bevitt adatokat, összesítőket

kérhetnek a hibákról, azok indoklásairól és tájékoztatást kapnak a munkafolyamat előrehaladtáról.

A jelentések adatainak bevitelét, javítását végző képernyős alkalmazások egységes szemléletben készülnek. Az ellenőrzések megvalósítása a külső könyvtárakban tárolt rutinok segítségével történik. Így biztosítható, hogy mind bevitelkor, mind ellenőrzéskor ugyanaz az ellenőrző rutin hajtódik végre. Az adatok minőségbiztosítását szolgálja, hogy

- adatbevitel és -javítás csak az engedélyezett időszakban történhet,
- utólagos módosítás csak engedélyezett adatszolgáltatókra történhet,
- minden ellenőrzést kötelező minden adatszolgáltatóra elvégezni,
- nem maradhat indoklás nélküli elfogadható hiba,
- nem maradhat *súlyos* kategóriába sorolt hiba.

Az adat-előkészítés eredményeként előállított ellenőrzött, javított statisztikai adatokból az adatfeldolgozás során állítjuk elő az adatgyűjtés céljaként meghatározott outputot. Ez lehet a felhasználók számára elérhető adatbázis és/vagy hagyományos statisztikai táblák.

Az adatfeldolgozás funkcióhoz tartozik: a hiányzó adatok/kérdőívek pótlása, a reprezentatív minta alapján a teljes körű adatok becslése, az aggregátumok képzése, indexszámítás, az utólagos javítás, az adatbázis előállítása.

E területen a kutatások elsősorban a különböző adatpótlási (imputálási) módszerek kidolgozására, új becslési eljárások kifejlesztésére, a mintavételi hiba csökkentésére irányulnak.

#### *A statisztikai adatok közzététele, tájékoztatás*

A statisztikai termelési folyamat végpontjának, a statisztikai információs rendszernek célja a felhasználói igények kielégítése, az adatok közzététele statisztikai táblák és elemzések formájában. A statisztikai tevékenységnek ez az a területe, ahol a tudomány támogatása jól érzékelhető, az elemzési eszközök egyre bővülő tárháza új igények kielégítését teszi lehetővé, emeli az elemző munka színvonalát. De a jó, tudományosan megalapozott elemzések és adat-összeállítások elkészítésével a statisztikus munkája nem ér véget. A statisztikai adatokhoz való hozzáférés reális esélyének megteremtése a különböző felhasználói körök felkészültségéhez, igényéhez és lehetőségéhez igazodó tájékoztatási eszköz és forma megválasztását és létrehozását is jelenti. Ebből következik, hogy a szoros értelemben vett statisztikai munka mellett ebben a szakaszban nélkülözhetetlen a megfelelő közkapcsolat (PR) és marketingtevékenység is, vagyis mindaz, ami elősegíti, hogy a statisztikai információ eljuthasson a felhasználókhöz. A tájékoztatás eszköztárszerében az utóbbi években megnőtt az informatika szerepe, az adattárház-technológia és az Internet megjelenése alapvető változást eredményezett.

Az adattárház a felhasználók szempontjai szerint előkészített, rendszerezett adathalmaz. Az adatok összeállításánál bizonyos kész aggregált állományok mellett lehetőség van arra, hogy a felhasználó aktívan részt vegyen a statisztikai output elkészítésében, maga határozza meg az igényelt statisztikai tábla vagy grafikon tartalmát és formáját. Ez úgy biztosítható, hogy az adattárház homogén adatkörei („kockái”) nem hagyományos, hanem multidimenziós statisztikai táblák, melyek kialakítása gondos és előrelátó

statisztikai tervezőmunkát igényel. Az adattárház építése is a metarendszer által vezérelt, ezért, és a felhasználók tájékoztatása érdekében, különös gondot kell fordítani az adatok metarendszerben történő pontos leírására.

#### *A statisztikai adatok minőségének mérése*

A statisztikai szakma ismeretanyagának áttekintése nem lenne teljes, ha befejezésül nem említenénk meg egy olyan kutatási területet, mely különösen az utóbbi években nyerte el fontosságát. Ez a statisztikai adatok minőségének vizsgálata, a statisztikai utóvizsgálatok. Az ipari termelési, technológiai folyamatok végrehajtása után végzett felülvizsgálatok és azok eredményeinek visszacsatolása közismerten elengedhetetlen része a termelési ciklusnak. Így van ez a statisztikai termelési folyamat esetében is, azzal a különbséggel, hogy a felülvizsgálandó tevékenységek egy költséges adatgyűjtési folyamat részei. Az utóvizsgálat során tulajdonképpen arra keressük a választ, hogy az előállított statisztikai adatok milyen hibákat tartalmaznak, más szóval milyen a minőségük.

Visszatekintve a statisztikai adat-előállítási folyamat előzőekben bemutatott szakaszaira, számos olyan munkafázist találunk az adatgyűjtéstől a tájékoztatásig, ahol hibák torzíthatják az adatok pontosságát. A nemzetközi irodalomban a hibafajták körét és csoportosítását tekintve különböző megközelítésekkel találkozhatunk. A leggyakrabban alkalmazott csoportosítás a következő:

- mintavételi hiba,
- nemválaszolási hiba,
- lefedettség hiba,
- adatgyűjtési hiba,
- adatfeldolgozási hiba.

Tanulmányunk kereteit meghaladná az egyes területeken végzett kutatások, alkalmazott eljárások és módszerek akár csak vázlatos ismertetése is. A témakörrel és irodalmáról jó összefoglaló ismereteket nyújt magyar nyelven a KSH Könyvtár és Dokumentációs Szolgálat „Statisztikai módszerek” sorozatában megjelent *dr. Marton Ádám*: „A reprezentatív felvételek megbízhatósága” című tanulmánya, angol nyelven az Egyesült Államok Szövetségi Statisztikai Módszertani Bizottsága (Federal Committee on Statistical Methodology) által készített jelentés a statisztikai felvételek hibaforrásainak méréséről.

## KÖVETKEZTETÉSEK

Jelen tanulmány a statisztika szónak két jelentést tulajdonít. Első jelentése szerint a statisztika egy önálló tudományág, a társadalmi és természeti jelenségek széles körének vizsgálatára szolgáló módszertudomány. A statisztika tudományának középpontjában az általános statisztikának nevezett diszciplína áll, és emellett szerves részei az általános statisztika alkalmazásainak minősülő szakstatisztikák, valamint az az elméleti és gyakorlati ismeretanyag, amelyet a szaknyelv statisztikai információ-rendszernek nevez. A tudományágak új nemzetközi szabványos osztályozási rendszerében – ha készülne ilyen – a statisztikát vagy a matematika közelében (tehát a természettudományok között), vagy néhány hasonlóan széles összefüggésekkel operáló interdiszciplináris tudományággal

együtt a természet- és társadalomtudományoktól elkülönítve kellene elhelyezni. Mindkét megoldás mellett és ellen megfontolásra érdemes érvek hozhatók fel.

A statisztika szó másik jelentése szerint a statisztika szakma, amelynek tartalma a statisztikai tevékenység folytatásához szükséges tudományos ismeretek, gyakorlati tapasztalatok és készségek összessége. A statisztikusszakma gyakorlása az esetek egy részében főtevékenység; ilyenkor a cél maguknak a statisztikai adatoknak a létrehozása és nyilvánosságra hozatala. Az esetek másik részében a statisztika egy adott tudományágon belül a kutatási célok elérésére, az eredmények értékelésére szolgál; ilyenkor a statisztikát, az adott tudományon belül, segédtevékenységként művelik.

A statisztikusszakma műveléséhez szükséges ismeretanyag három csoportba sorolható: az általános statisztika; a szakstatisztikák speciális ismeretanyaga, ide értve a vonatkozó szakterületek tudományos alapjait (közgazdaságtan, demográfia, szociológia, jog, földrajz, politológia stb.); az információgyűjtéshez és adatkezeléshez kapcsolódó szervezési és módszertani ismeretanyagot, beleértve az informatikai ismereteket is. Az utóbbi szakmai ismeretkörrel illetően a közelmúltban mélyreható változások következtek be. Korábban a statisztikusszakma művelői (a gyakorló statisztikusok) és az informatikusok között a megrendelő-szolgáltató kapcsolat volt jellemző. Mára – a személyi számítógépek és a felhasználóbarát szoftverek megjelenésével – a statisztikusok nem megrendelik, hanem elvégzik az adatelemző, csoportosító, adatjavító, összesítő munkák egy jelentős részét; számukra a személyi számítógép mindennapi munkaeszközzé vált. Az informatikai ismeretek egy része a statisztikai szakmába integrálódik.

„A hivatalos statisztika” (official statistics) és a statisztikatudomány (academic statistics) közötti tartalmi eltérést az okozza, hogy a statisztikatudomány, és a hozzá kapcsolódó oktatás jelenlegi gyakorlata még nem alkalmazkodott ezekhez a változásokhoz, ugyanakkor a statisztikusszakmához, a statisztikai munkához ezek az ismeretek elengedhetetlenek. A megoldás a statisztikai tudomány erőforrásainak, a korábbiaknál nagyobb mértékű bevonása a statisztikai szolgálat munkájába. Ez a folyamat kifejezésre jut a különböző nemzeti és nemzetközi statisztikai intézmények kutatási programjaiban is, melyek jelentős szerepet szánnak az információ-technológiák kutatásának és fejlesztésének.

## 1. FÜGGELÉK

### *Tudományági osztályozás<sup>4</sup>* (az UNESCO ajánlása)

1. Természettudományok
  - 1.1. Matematika
  - 1.2. Fizika
  - 1.3. Kémia
  - 1.4. Csillagászat
  - 1.5. Földtudomány
    - 1.5.1. Földrajz
    - 1.5.2. Földtan
  - 1.6. Biológia
  - 1.7. Nem részletezhető természettudományi kutatások

<sup>4</sup> A kutatási és fejlesztési statisztika céljaira. Az osztályozás a műszaki tudományok többségét három számjegy mélységben tovább bontja; a kohászati tudományt például vas-, alumínium- és színesfém-kohászati ágakra. Ezeket a részletezéseket a jobb áttekinthetőség érdekében elhagytuk.

2. Műszaki tudományok
  - 2.1. Általános mérnöki tudomány
  - 2.2. Építéstudomány
  - 2.3. Bányászati tudomány
  - 2.4. Kohászati tudomány
  - 2.5. Energiagazdálkodás
  - 2.6. Vegyipari tudomány
  - 2.7. Gépipari és villamosipari tudomány
  - 2.8. Elektronikai ipar és számítástechnika
  - 2.9. Könnyűipari tudomány
  - 2.10. Élelmiszeripari tudomány
  - 2.11. Szállítási, hírközlési tudomány
  - 2.12. Vízgazdálkodási tudomány
  - 2.13. Nem részletezhető műszaki kutatások
3. Orvostudományok
  - 3.1. Elméleti orvostudomány
  - 3.2. Klinikai orvostudomány
  - 3.3. Társadalom-orvostudomány
  - 3.4. Gyógyszerészet és gyógyszerkutatás
4. Agrártudományok
  - 4.1. Növénytermelés
  - 4.2. Kertészet
  - 4.3. Erdészet és vadgazdálkodás
  - 4.4. Állattenyésztés
  - 4.5. Mezőgazdaság gépesítése
  - 4.6. Növényvédelem
  - 4.7. Állatorvos-tudomány
  - 4.8. Talajtan
  - 4.9. Mezőgazdasági üzemtan
  - 4.10. Nem részletezhető agrártudományi kutatások
5. Társadalomtudományok
  - 5.1. Filozófia
  - 5.2. Pszichológia
  - 5.3. Demográfia, szociológia
  - 5.4. Pedagógia
  - 5.5. Állam- és jogtudomány
  - 5.6. Kommunikáció
  - 5.7. Közgazdaságtudomány
  - 5.8. Szervezéstan
  - 5.9. Történelem, régészet, néprajz
  - 5.10. Nyelvészet és irodalom
  - 5.11. Művészet
  - 5.12. Nem részletezhető társadalomtudományi kutatások
6. Tudományáganként nem részletezhető kutatások

## 2. FÜGGELÉK

*Az egyes tudományterületekhez tartozó tudományágak, valamint a művészeti ágak*  
Melléklet a 169/2000. (IX.29.) Korm. rendelethez

1. Természettudományok
  - 1.1. Matematika- és számítástudományok
  - 1.2. Fizikai tudományok
  - 1.3. Kémiai tudományok

- 1.4. Földtudományok
- 1.5. Biológiai tudományok
- 1.6. Környezettudományok
- 1.7. Multidiszciplináris természettudományok
2. Műszaki tudományok
  - 2.1. Építőmérnöki tudományok
  - 2.2. Villamosmérnöki tudományok
  - 2.3. Építésmérnöki tudományok
  - 2.4. Anyagtudományok és technológiák
  - 2.5. Gépészeti tudományok
  - 2.6. Közlekedéstudományok
  - 2.7. Vegyészmérnöki tudományok
  - 2.8. Informatikai tudományok
  - 2.9. Agrár műszaki tudományok
  - 2.10. Katonai műszaki tudományok
  - 2.11. Multidiszciplináris műszaki tudományok
3. Orvostudományok
  - 3.1. Elméleti orvostudományok
  - 3.2. Klinikai orvostudományok
  - 3.3. Egészségtudományok
  - 3.4. Gyógyszertudományok
  - 3.5. Multidiszciplináris orvostudományok
4. Agrártudományok
  - 4.1. Növénytermesztési és kertészeti tudományok
  - 4.2. Állatorvosi tudományok
  - 4.3. Állattenyésztési tudományok
  - 4.4. Élelmiszertudományok
  - 4.5. Erdészeti és vadgazdálkodási tudományok
  - 4.6. Multidiszciplináris agrártudományok
5. Társadalomtudományok
  - 5.1. Gazdálkodás- és szervezéstudományok
  - 5.2. Közgazdaságtudományok
  - 5.3. Állam- és jogtudományok
  - 5.4. Szociológiai tudományok
  - 5.5. Politikatudományok
  - 5.6. Hadtudományok
  - 5.7. Multidiszciplináris társadalomtudományok
6. Bölcsészettudományok
  - 6.1. Történelemtudományok
  - 6.2. Irodalomtudományok
  - 6.3. Nyelvtudományok
  - 6.4. Filozófia tudományok
  - 6.5. Nevelés- és sporttudományok
  - 6.6. Pszichológiai tudományok
  - 6.7. Néprajz és kulturális antropológiai tudományok
  - 6.8. Művészeti és művelődéstörténeti tudományok
  - 6.9. Vallástudományok
  - 6.10. Média- és kommunikációs tudományok
  - 6.11. Multidiszciplináris bölcsészettudományok
7. Művészetek
  - 7.1. Építőművészet
  - 7.2. Iparművészet
  - 7.3. Képzőművészet
  - 7.4. Színházművészet



- 7.5. Film- és videoművészet
- 7.6. Zeneművészet
- 7.7. Tánc- és mozdulatművészet
- 7.8. Multimédia
- 8. Hittudomány

### 3. FÜGGELÉK

*Az Eurostat statisztikai kutatási programja*  
(Research in Official Statistics IST 2000-V.1.8 CPA8: Statistical tools, methods,  
indicators & applications for the Information Society) (Kivonat)

- 1. Módszertani kérdések
  - 1.1. Fogalomalkotás, osztályozások
  - 1.2. Nemzetközi harmonizáció
  - 1.3. A sokaság felvételi lefedettségének növelése
- 2. Adatgyűjtési technológiák
  - 2.1. Automatikus kódolás
  - 2.2. Mintavétel
  - 2.3. Automatikus adatbevitel
  - 2.4. Elektronikus adatgyűjtés
- 3. Minőség
  - 3.1. Minőségbiztosítás a statisztikai termelési folyamatban
  - 3.2. Módszerek az adatminőség mérésére és javítására
    - 3.2.1. A nemmintavételi hibák mérése és csökkentése
    - 3.2.2. A nemválaszolási hibák mérése és csökkentése.
    - 3.2.3. Az adatelőkészítési eljárások javítása.
    - 3.2.4. Imputálási technikák kidolgozása.
    - 3.2.5. Módszerek a mintavételi hiba becslésére.
- 4. Statisztikai elemzés, statisztikai modellek.
  - 4.1. Idősorelemzés, előrejelzés.
  - 4.2. Modellek a kockázat és a bizonytalanság mérésére.
- 5. A különböző forrásokból származó adatok integrációja.
  - 5.1. Adminisztratív adatforrások.
  - 5.2. Statisztikai metaadatok.
  - 5.3. Osztott adatbázisok kezelése.
  - 5.4. IT infrastruktúra.
- 6. Tájékoztatás, adatvédelem.

### IRODALOM- ÉS FORRÁSJEGYZÉK

- A Magyar Tudományos Akadémia Almanachja, 1997* [1997]. Magyar Tudományos Akadémia. Budapest.
- „A statisztika fogalma és tárgya” című vitautulés anyaga [1954]. *Statisztikai Szemle*. XXXII. évf. 6–7. sz. 519–535. old.
- A statisztikai tudomány helyzete és perspektívái [1999]. *Statisztikai Szemle*. 77. évf. 2–3. sz. 101–111. old.
- DAVIES, P. [2000]: *Isten gondolatai. Egy racionális világ tudományos magyarázata*. Vince Kiadó Kft., Budapest.
- DÖRNYEI J. [1983]: *The role of metainformation in statistical integration*. International Statistical Institute, Proceedings of the 44<sup>th</sup> Session, Madrid.
- FARKAS J. [1981]: *A modern tudomány szerkezete*. Akadémiai Kiadó. Budapest, 1981.
- GARAI L. [1994]: Természettudomány-e a pszichológia? *Magyar Tudomány*. 39. évf. 1. sz. 62–73. old.
- GARAI L. [2003]: Magánhangzók és mássalhangzók. *Élet és irodalom*. 2003. július 25. 24. old.
- Guidelines for the modelling of statistical data and metadata* [1995]. United Nations, New-York és Geneva.
- GYÖRKI I.–RÓNAI M.: *Metadata management*. Statistical Commission and Economic Commission for Europe, Conference of European Statisticians, Work Session on Statistical Metadata, Geneva. Kézirat.
- GYÖRKI I.: *Survey control as a subsystem of the statistical information system*. Seminar on integrated statistical information systems and related matters (ISIS'96). Bratislava (UN/ECE). Kézirat.
- GYULAI F. [1954]: A statisztika fogalma, tárgya és módszere kérdésében a szovjet statisztikusok között folyó vita jelenlegi állása. *Statisztikai Szemle*. XXXII. évf. 3. sz. 179–192. old.

- HUNYADI L. – RAPPAL G. [1999]: Gondolatok a statisztikáról. *Statisztikai Szemle*. 77. évf. 1. sz. 5–15. old.
- IST 2000-V.1.8 CPA8: Statistical tools, methods, indicators and applications for the Information Society, <http://europa.eu.int/comm/eurostat/research/>
- Képzési területek egységes osztályozási rendszere* [2003]. Központi Statisztikai Hivatal. Budapest.
- KOMÁR E. [2000]: *Adatfeldolgozás a minőség szolgálatában*. A Magyar Statisztikai Társaság vándorgyűlésén elhangzott előadás – A statisztika minősége. Gyöngyös. 2000. szept. 28.
- KUVSINNYIKOV, P. A. – SZMULEVICS, B. J. A. [1954]: A statisztika tárgya és módszere. *Statisztikai Szemle*. XXXII. évf. 1. sz. 3–9. old.
- OLENSKI, J. [1978]: *Structure of statistical information system and methodology of statistical surveys in terms of computerization*. ISIS seminar 1978 Proceedings, ECE/CES Geneva. Kézirat.
- PAPP Z. DR. [2003]: Új adatgyűjtési technológiák felé. *Gazdaság és Statisztika*. 15 (54.) évf. 2. sz. 25–65. old.
- PLOSKO, B. [1953]: A statisztikai tudomány tárgya. *Statisztikai Szemle*. XXXI. évf. 8. sz. 615–623. old.
- PUKLI P. [1997]: Fundamental stages in designing procedure of statistical survey. *Statistical Review. Special issue*. 28–35. old.
- RÁCZ A. [1977]: A statisztika tudomány helyzete. *Statisztikai Szemle*. 57. évf. 11. sz. 1126–1138. old.
- RATZSCH, D. [2002]: *Miből lesz a tudomány? Rövid bevezetés a tudományfilozófiába*. Harmat Kiadó. Budapest.
- RÉDEI J. [1953]: A statisztika fogalma és tárgya. *Statisztikai Szemle*. XXXI. évf. 8. sz. 624–627. old.
- RÉDEI J. [1954]: A statisztika fogalma. *Statisztikai Szemle*. XXXII. évf. 3. sz. 179–192. old.
- SOLTÉS, D. [1979]: *Metadata bases-data dictionary/directories*. ISIS seminar 1979 Proceedings, ECE/CES Geneva.
- SUNDGREN, B. [1980]: Detailed description of metadata base and dictionary management. Statistics Sweden Stockholm.
- SZILÁGYI GY. [1975]: A statisztika tudomány köre és ágazatai. *Statisztikai Szemle*. 55. évf. 7. sz. 724–733. old.
- TAMÁS I. [2003]: Megérteni a természet törvényeit. Oláh György Nobel-díjas kémikus párbeszéde diákokkal. *Népszabadság*. 2003. június 30. 8. old.
- VÉGVÁRI J. [1993]: A hivatalos statisztikai szolgálat: integráció és együttműködés. *Statisztikai Szemle*. 71. évf. 7. sz. 534–552. old.
- YULE, G. U. – KENDALL, M. G. [1964]: *Bevezetés a statisztika elméletébe*. Közgazdasági és Jogi Könyvkiadó. Budapest.

## SUMMARY

In the line of the international discussions on the gap between academic statistics and official statistics the authors try to define the difference in the content of the two concepts on the bases of the Hungarian practice. In the first part, looking back to the former discussions they clarify their view. Accordingly, statistics is an individual field of science, in spite of the fact that there is no item or class for statistics in the Hungarian classifications of science. The authors presents their ideas about the place of statistics in a new, more advanced classification. In the second part of the paper they give a general overview of the statistical production process. They lay emphasis on the new development resulted by the strong influence of the advanced information technologies in order to clarify and help to bridge the gap between official and academic statistics.