

A TÖBBVÁLTOZÓS MATEMATIKAI STATISZTIKAI MÓDSZEREK FELHASZNÁLÁSA *

DR. ANDORKA RUDOLF

Az elektronikus számítógépek és a hozzájuk kidolgozott programok ugrásszerű változást hoztak a többváltozós matematikai statisztikai módszerek¹ felhasználási lehetőségeiben. Például az 1960-as évek közepén egy 5 változót és 24 megfigyelési egységet tartalmazó korrelációs és regressziós elemzés elvégzésekor a kézi számítógépekkel végrehajtott számítások elriasztó mennyiségű munkaidőt igényeltek (1). Két év múlva azonban, amikor ugyanezt a problémát 191 megfigyelési egység adatai alapján vizsgáltam (2), de már elektronikus számítógép segítségével, a számítások eredményei néhány nap alatt a kezemben voltak. Elemzési lehetőségeinkben tehát minőségi változás következett be.

Ez lényeges, mert a többváltozós matematikai statisztikai módszerek jellegzetesen olyan kutatási problémák kezelésére alkalmasak, amelyekkel a társadalmi jelenségek vizsgálatában lépten-nyomon találkozunk. Tipikus ilyen problémák a következők:

– valamilyen fajta megfigyelési egységekről – egyes emberekről, családokról, városokról, termelőegységekről, országokról – nagyszámú adat áll rendelkezésünkre, de éppen az adatgazdagság nehezíti meg, hogy tisztán lássuk, milyen alapvető dimenziók szerint rendeződnek el, különböznek egymástól a megfigyelési egységek, más szóval milyen a megfigyelt sokaság rejtett szerkezete (ilyen probléma merül fel például, amikor városokat akarunk az urbanizáltságuk szerint vizsgálni, amikor országok fejlettségét hasonlítjuk össze, amikor családok élet-színvonalát és életkörülményeit tanulmányozzuk);

– kutatjuk valamilyen társadalmi jelenség okait, és tudjuk, vagy feltételezzük, hogy azt több tényező befolyásolja; ki akarjuk választani, melyek a legerősebb hatótényezők és számszerűsíteni akarjuk a hatásukat (ilyen probléma merül fel a termékenység vagy a halandóság, a vándorlás vagy a társadalmi mobilitás, a bűnözés vagy az alkoholizmus okainak kutatásánál);

– típusokba akarjuk sorolni megfigyelési egységeinket, osztályozásokat akarunk kialakítani, amelyek a legjobban tükrözik a megfigyelt hasonlóságokat és különbségeket (ilyen probléma merül fel például a településtípusok vizsgálatánál, a társadalmi rétegek és csoportok kialakításánál);

– számszerűsíthető modellekkel akarjuk leírni a társadalom valamely részterületének, például az egészségügynek működését, vagy a társadalom fejlődésének egészét átfogó modellt akarunk felépíteni.

Megvannak tehát az állandóan jelentkező elemzési problémák. Tizenegynéhány éve rendelkezésünkre állnak az elektronikus számítógépek és a kidolgozott progra-

* Az MKT Statisztikai Szakosztálya Statisztikatörténeti Szakcsoportjának 1979. május 28–30-án Baján tartott XVII. vándorülésén megvitatott előadás.

¹ Többváltozós módszernek nevezem azokat a matematikai statisztikai módszereket, amelyekkel kettőnél több változó összefüggéseit elemezhetjük.

mok nyújtotta lehetőségek. A tanulmányban azt kívánom áttekinteni, hogy mennyire és hogyan éltünk ezekkel a lehetőségekkel a társadalom- és népesedéssziszti elemzésekben.² Az áttekintés alapját azok a tanulmányok képezik, amelyek a *Statisztikai Szemle*, a *Demográfia* és a *Területi Statisztika* hasábjain megjelentek, továbbá a Központi Statisztikai Hivatal munkatársainak néhány további munkája.

A különböző többváltozós módszerek tárgyalásánál követhetném a történeti sorrendet: kiindulhatnék a legtöbbet alkalmazott regressziós elemzésből, és befejezhetném az áttekintést a még nem alkalmazott módszerekkel. Mégsem ezt az eljárást választom, hanem a kutatási logikát követem: a faktoranalízissel kezdem, mert ez tulajdonképpen a legelső kutatási kérdés megoldására, az alapadatok közötti „rendteremtésre”, a közöttük levő összefüggések első áttekintésére szolgál, s az ok-okozati kapcsolatok kutatására használható módszerekkel folytatom, majd áttérek a tipizálásra és rangsorolásra használható módszerekre.

A faktoranalízist³ akkor célszerű használni, ha sok különböző adatunk van, nagyszámú megfigyelési egységről (az utóbbiak számának nagyobbak kell lennie a mutatók vagy változók számánál), és feltételezzük, hogy e sok mutató valójában kisebb számú lényeges jellemzőt mér. Például sok adatunk van az ország településeiről a népszámlálások és területi statisztikák alapján, és feltehetjük, hogy ez a nagyszámú mutató a települések néhány alapvető jellegzetességét fejezi ki: a városi funkciók meglétét, az iparosodottságot, a lakosság demográfiai összetételét. A faktoranalízis segítségével feltérképezhetjük e nagy adathalmazt. A faktoranalízis programja ugyanis először kiszámítja a változók közötti páros korrelációs együtthatókat, majd a korrelációs matrix alapján néhány faktort határoz meg. A faktorok tulajdonképpen új mesterséges változók, amelyek az adatainkban meglévő rejtett tendenciákat, dimenziókat fejezik ki. A program ezután meghatározza az eredeti változók és a faktorok közötti kapcsolat erősségét kifejező faktorsúlyokat. Az utóbbiak alapján értelmezzük az egyes faktorokat, vagyis megkísérelünk választ adni arra a kérdésre, hogy e faktorok milyen dimenziókat (a városiasodottságot stb.) fejeznek ki. Az egész elemzésnek csak akkor van értelme, ha ténylegesen meg tudjuk mondani, mit képviselnek a faktorok. Végül kiszámítja a program a faktorpontszámokat, amelyek rangsorolják a megfigyelési egységeket az egyes faktorok dimenziójában, például sorokba rakják a településeket a városiasodottság, az ipari fejlettség, a népesség korösszetétele alapján. A faktoranalízis alkalmazásához könyvtári programok állnak rendelkezésre. Ezek a főkomponens-elemzés nevű változatot használják.⁴

A faktoranalízis alkalmazásai előtt szeretném megemlíteni a probléma hasonló megoldásának egy korai előfutárát, Cseh-Szombathy László tanulmányát az életszínvonal nemzetközi összehasonlításáról (8). A kutatási alapprobléma ugyanaz volt, mint a faktoranalízis által kezelhető esetekben: sok fogyasztási mutató áll rendelkezésre, és ki kell választani közülük azokat, amelyek az életszínvonalat a legjobban jellemzik. A kutatás első lépése is a faktoranalízishez hasonló volt: Cseh-Szombathy László korrelációs együtthatókat számított az egy főre jutó nemzeti jövedelem és a fogyasztási színvonal különböző mutatói között, és az utóbbiak közül azokat választotta ki az életszínvonal jellemzésére és összehasonlítására, amelyek az egy főre jutó jövedelemmel a legszorosabb korrelációt mutatták, de ugyanakkor a fogyasztásnak különféle oldalait fejezték ki.

² Nem foglalkozom itt a többváltozós matematikai statisztikai módszerek gazdaságstatisztikai alkalmazásaival. Ezen a téren mind a világirodalomban, mind a magyar szakirodalomban sokkal nagyobb fejlődést értek el az elmúlt évtizedekben.

³ A faktoranalízist ismertette Vita László (3), Magyar nyelven rendelkezésre áll a (4) is.

⁴ A rendelkezésre álló könyvtári programokról szólva egyrészt az (5) munkára, másrészt Gombosiné Gárdos Eszter tanulmányára (6) támaszkodom, amelyek leírják, milyen programok állnak rendelkezésre a Központi Statisztikai Hivatal számítógépeihez. A faktoranalízis-programokat ezen kívül leírta Zágan Csaba (7).

Bár a faktoranalízist csak nagy elektronikus számítógépek segítségével lehet elvégezni, a módszer néhány év óta gyorsan elterjedt a statisztikai elemzésekben.

A legtöbb alkalmazás esetében országrészek, települések gazdasági fejlettségét, városiasodottságát és egyéb jellemzőit elemezték (9)–(21). Levonhatjuk azt a következtetést, hogy a területi egységek különféle adatainak elemzésére igen jól alkalmazható e módszer, mert általában sok népszámlálási és területi statisztikai adatunk van, és e módszer segítségével sikerül azokat „feltérképezni”, a tendenciákat kimutatni és a területi egységeket különféle szempontok szerint sorba rendezni. Például ki lehet mutatni, hogy sok adat figyelembevételével mely települések vagy mikrorégiók különösen elmaradottak, vagy mely települések a leginkább városiasodottak. Nem szabad azonban elfelejtenünk, hogy a faktoranalízis eredményei teljes mértékben a bevitt adatoktól függenek, tehát a változókat jól kell megválasztani ahhoz, hogy jó eredményeket kapjunk.

Alkalmazták ezenkívül a faktoranalízist Magyarország demográfiai régióinak körülhatárolására (22), a budapesti gyermekek testméreteinek elemzésére (23), a szocialista országok gazdasági fejlettségének és szerkezetének vizsgálatára (24).

Megemlítem, hogy a faktoranalízis problémája megfordítható, és az ún. Q-faktor analízissel kezelhető, ha több változónk van, mint megfigyelési egységünk. Ebben az esetben a számításban a változók veszik át a megfigyelési egységek szerepét, és megfordítva, a megfigyelési egységek a változókét. Tehát a faktorsúlyok a megfigyelési egységek és a faktorok közötti kapcsolat erősségét mérik, a faktorpontszámok pedig a változókat rendezik sorba. A faktorok nyilvánvalóan mást fejeznek ki, mint az eredeti faktoranalízis esetén, de az abban kapott faktorokkal bizonyos rokonságot mutatnak. Hazai alkalmazását nem ismerem, de egy nemzetközi konferencián végzett számításokból a megyék sorrendjére vonatkozóan érdekes információkat kaptunk.⁵

Azt mondhatjuk tehát, hogy a faktoranalízist technikailag elsajátítottuk. Az eredmények értelmezését és felhasználását illetően azonban bizonyos problémák még látszanak, de nemcsak a magyar, hanem a nemzetközi szakirodalomban is.

Történetileg a faktoranalízist megelőzve terjedt el a kétváltozós számítás kiterjesztéseképpen kidolgozott többszörös korrelációs és regresszió-elemzés. Logikailag azonban a faktoranalízis után következik. Az utóbbival ugyanis feltérképezhetjük a rendelkezésünkre álló sokaságot, és az így nyert felismerések alapján választhatjuk ki a függő változót és a magyarázó változókat a regresszióelemzéshez. Sőt eljárhatunk úgy is, hogy magukat a faktorpontszámokat alkalmazzuk változókként.⁶

Több könyvtári program is rendelkezésre áll, közülük a lépésenkénti regressziószámítás alkalmazása a legcélszerűbb, mert sorban egymás után veszi fel az egyenletbe a legerősebb, majd a második legerősebb, majd a harmadik legerősebb stb. magyarázó változót, végül elhagyja azokat, amelyek a már figyelembe vett változóktól függetlenül alig vagy egyáltalán nem befolyásolják a függő változót. Ez az eljárás lényegesen megkönnyíti az eredmények értelmezését.

A kétváltozós korreláció- és regressziószámítás évtizedek óta ismert és használt módszer a statisztikai elemzésekben. Mégis érdemes hangsúlyozni, hogy ez a nagyon egyszerű – kézi számításokkal is alkalmazható – módszer igen tömören és jól jellemzi két jelenség összefüggését. A korrelációs együttható megadja annak erősségét, a regressziós együttható az irányát, a regressziós egyenes ábrázolása a szó-

⁵ A Nemzetközi Szociológiai Társaság Társadalomökológiai Kutatási Bizottságának 1976. évi konferenciája alkalmával számításokat végeztünk a Ljubljana-i Egyetem számítógépén. A számítás alapadatai a konferenciára benyújtott dolgozatomban szereplő magyarországi megyei fejlettségi mutatók voltak.

⁶ A faktoranalízis és regresszióelemzés ilyen összekapcsolódására a hazai statisztikai szakirodalomban lásd Szabady Balázs munkáját a termékenységre ható tényezőkről (21).

rási diagramon pedig jól szemlélteti annak természetét és a regressziós egyenes által kifejezett tendenciától erősen eltérő megfigyelési egységeket. A gyakorlott olvasó számára ez a néhány adat és ábra többet mond hosszadalmas leírásoknál.

A kétváltozós módszer alkalmazására csak példákat sorolhatok fel, az összes ilyen munka felsorolása hatalmas bibliográfiát igényelne. Az egyik probléma, ahol e módszert sokszor alkalmazták, a művi abortuszok, a koraszülések, a csecsemőhalandóság és a termékenység közötti összefüggések kutatása. A felmerült módszertani kérdések miatt érdemes ezeket röviden ismertetni.

Barsy Gyula és Sárkány Jenő a megyék és megyei városok adataival végzett regresszióelemzéssel bizonyították, hogy a 0–6 napos csecsemők halandósága egészen más (erősen negatív) kapcsolatban áll az élveszületési arányszámmal, mint a csecsemőhalandóság általában, és különösen mint a késői csecsemőhalandóság (mert az utóbbinak kapcsolata erősen pozitív). Ezért egészen más okokat kell keresnünk a korai és a késői csecsemőhalandóság mögött. Ennek alapján vonták le azt a következtetést, hogy a korai csecsemőhalandóság okai között lényeges szerepet játszik a művi abortuszok gyakorisága (25). Ugyanebben a tanulmányban pozitív korrelációs kapcsolatot mutattak ki a száz élveszületésre jutó spontán vetélések száma és az ezer 15–44 éves nőre jutó művi vetélések száma között.

Miltényi Károly az utóbbi megállapítást szintén korrelációszámítással cáfolta, mondván, hogy nem a száz szülésre jutó spontán vetélések száma, hanem az ezer propagatív korú nőre jutó spontán vetélések száma és a művi vetélések száma között kell korrelációt számítani, és az ebben az esetben a zérushoz közeli együtthatót mutat ki (26).

A módszertani viták még élesebbek és tanulságosabbak voltak, amikor *Mikolás Miklós* közel egy évtizeddel később ismét felvetette azt, hogy a művi abortuszok gyakoriságának megnövekedése és a születési átlagsúly csökkenése között erős pozitív korreláció van. Ezt az 1955–1970. évek idősoros adatain igazolta (27). *Andorka Rudolf, Cseh-Szombathy László és Vukovich György* erre válaszolva utaltak arra, hogy a területi keresztmetszeti adatok alapján számított korrelációk megbízhatóbbak az idősorok alapján számítottaknál. Ők a megyék 1965 és 1971 közötti átlagos koraszülési aránya és a száz szülésre jutó abortuszok száma között számítottak korrelációt, és a kapott együttható alapján, amely sokkal alacsonyabb volt a Mikolás által számított hasonló együtthatóknál, arra a következtetésre jutottak, hogy a művi abortuszok gyakorisága és a koraszülések gyakorisága között van ugyan bizonyos pozitív korrelációs kapcsolat, de annak szorosságát nem tudjuk megnyugtatóan megítélni (28). *Szokolczai György* viszont az utóbbi tanulmánnyal vitatkozva úgy végezte el a területi adatokon alapuló számításokat, hogy egyes területi egységeket elhagyott, külön-külön számolt az egyes országrészek adataival, és így más, egyes esetekben lényegesen erősebb korrelációs együtthatókat kapott (29).⁷ Az érdemi vitát itt nem dönthetem el, csak arra hívom fel a figyelmet, hogy a vita során a korrelációszámítással kapcsolatos több fontos módszertani probléma vált világossá.

További példák a kétváltozós korreláció- és regresszió-számítással kezelt problémákra:

- az egyes szakmák, munkakörök átlagbére és a kérdéses területen a nők aránya közötti negatív kapcsolat (31), (32);
- a hírközlő eszközök elterjedtsége és a gazdasági fejlettség közötti kapcsolat, az irodalom és művészet iránti érdeklődés és az iskolázottsági színvonal közötti kapcsolat (33);

⁷ A csecsemőhalandóság és a művi vetélések között számított korrelációt *Lits Józsefné* is területi adatok alapján. Közepes erősségű pozitív korrelációt kapott (30). Ez is arra enged következtetni, hogy a művi abortuszokon kívül más tényezők is erősen befolyásolják a csecsemőhalandóság nagyságát.

- a külterületekről való elköltözési szándék és az adott külterületek különböző jellemzői közötti kapcsolat (34);
- az egy főre jutó jövedelem, valamint a mezőgazdaságban, illetve az iparban foglalkoztatottak aránya közötti kapcsolat (35).

Noha a kétváltozós korreláció- és regressziószámítás ennyire általánosan használatos volt a statisztikai elemzésekben, a többváltozós elemzés viszonylag kevésbé terjedt el. Leginkább területi keresztmetszeti adatok alapján végeztek ilyen elemzéseket. Vizsgálták vele többször is a termékenység szintjét meghatározó gazdasági és társadalmi tényezőket (1), (2), (21). Saját korábbi ilyen elemzéseim legérdekesebb eredménye nem annyira a figyelembe vett magyarázó változók által magyarázott termékenységi különbségek voltak, mint inkább a megmagyarázatlanul maradt viszonylag nagy rész, amely országrészenként meglehetősen következetes képet mutatott végig a XX. század folyamán. Ezt a regionális különbségek valamilyen történeti okaival próbáltam magyarázni. Szabady Balázs (21) lényegesen több változóval végzett számításai alapján jobban meg tudta magyarázni a termékenység alakulását.

Pápai Béla többváltozós regressziós egyenletek segítségével a budapesti agglomerációt határolta körül olyan módon, hogy azokat a településeket, ahol a Budapestre ingázók, illetve a nem mezőgazdasági keresők tényleges aránya nagyobb volt, mint amekkora a regressziós egyenletekből következett volna, az agglomeráció részének tekintette (36). Szauter Edit a belső vándorlást meghatározó tényezőket kutatta regresszióelemzéssel (37). Vizsgáltuk azt, hogy az alkoholizmus elterjedtségét a városiasodottság vagy a bortermelés magyarázza-e inkább, és az előbbi hatását sokkal erősebbnek találtuk (38).

A közelmúltban végezték el az első többváltozós regressziószámításokat nem területi adatok, hanem mintavételi adatfelvételekből származó egyéni és családi adatok alapján (39), (40). A módszernek ilyen adatok elemzésére való felhasználása ennél sokkal nagyobb lehetőségeket kínál, és a mintavételi felvételek elterjedésével párhuzamosan az eddiginél nagyobb mértékben kellene felhasználni. Problémát okoz, hogy a Központi Statisztikai Hivatal olyan nagy mintákkal dolgozik, amelyeknek regresszióelemzéssel való kezelése túlságosan nagy számítógépóra-igénnyel jár. A kiút valószínűleg az, amelyet (39) szerzői alkalmaztak, nevezetesen a mintán belül egy kisebb mintát választottak.

A többszörös regresszióelemzésnek a nemzetközi szakirodalomban az utóbbi időben gyakran alkalmazott változata az *útelemzés*. Mivel nincsenek útelemzés elnevezésű könyvtári programjaink, érdemes a módszer lényegét itt ismertetni. Közöséges többváltozós regresszióelemzést kell végezni, és az abban kapott standardizált regressziós együtthatókat vagy béta-együtthatókat használják fel útegyütthatóként, amelyeket az útábrára felírnak. Az útábra nem más, mint egy rekurzív egyenletrendszer ábrázolása. Az útábra nyilai mutatják, hogy az egyes mutatókat mint független változókat mely más változók magyaráznak meg. Például a társadalmi mobilitás alapvető útmodelljében:

1. az összeírt jelenlegi társadalmi helyzetét magyarázza első foglalkozáskori helyzete, iskolai végzettsége, az apa társadalmi helyzete és iskolai végzettsége;
2. az összeírt első foglalkozáskori társadalmi helyzetét magyarázza iskolai végzettsége, valamint az apa társadalmi helyzete és iskolai végzettsége;
3. az összeírt iskolai végzettségét magyarázza az apa társadalmi helyzete és iskolai végzettsége.

Útelemzési eredményeket publikáltak a pályakezdő fiatalok adatai alapján *Surányi Bálint* és *Vita László* (41). *Kulcsár Rózsával* mi is végeztünk kísérleti jelleggel

hasonló számításokat. Ezeknek eredményeit összehasonlítottuk a szocialista és a tőkés országokban végzett hasonló számítások eredményeivel. Igazolódni láttuk azt a tételt, hogy a szocialista országokban az apa társadalmi helyzetének közvetlen (az iskolai végzettségtől független) hatása a fia társadalmi helyzetére gyengébb, mint a tőkés országokban, viszont az apa társadalmi helyzetének hatása a fia iskolai végzettségére és ezen keresztül közvetve a fia társadalmi helyzetére erősebb. Az útlelemzést nemcsak, sőt nem is elsősorban a társadalmi mobilitás elemzésében lehet használni, hanem minden olyan társadalmi jelenség vizsgálatánál, ahol az ok-okozati mechanizmust egy rekurzív modellel írhatjuk le, és ahol a változókat számszerűsíteni tudjuk.

A faktoranalízis és a regresszióelemzés kombinációjának tekinthetjük a *kanonikus korrelációs számítás*át. Olyan problémák elemzésére használhatjuk, ahol nemcsak a magyarázó változók oldalán, hanem a függő változók oldalán is több változónk van. Egy pszichológiai példa szemlélteti az ilyen típusú problémákat: az intelligencia és a személyiség között akarnak korrelációt számítani, és mind az intelligenciát, mind a személyiséget több teszt eredményeivel jellemzik. A módszer mind a függő, mind a független változók oldalán faktorokat számít ki, és a faktorok közötti korrelációt méri. Könyvtári programunk van, de az alkalmazásra eddig csak kísérletek történtek a mobilitáselemzés területén.

A többváltozós regresszióelemzéshez hasonló módszer a többszörös *varianciaelemzés* azzal a különbséggel, hogy míg az előbbit akkor használhatjuk, ha minden változót számszerűsíteni tudunk, a varianciaelemzés akkor is alkalmazható, ha csak nominális skálákon (például férfi–nő, különböző társadalmi csoportok) tudjuk mérni a magyarázó változókat. Valójában az utóbbi típusú probléma talán gyakrabban fordul elő a társadalmi és népesedésszociológiai elemzésekben, mint a teljesen számszerűsíthetők. Mégis a többszörös varianciaelemzést szinte alig alkalmazták (42). Ugyanezt az elemzési gondolatmenetet követik *Éltető Ödön* és *Frigyes Ervin* nagyhatalmú számításai, amelyekben kimutatták, hogy a háztartás demográfiai összetétele (a gyermekek száma, a családanya kereső vagy eltartott volta) erősebben befolyásolja a háztartás egy főre jutó jövedelemszintjét, mint az aktív keresők által végzett munka és az ő keresetük (43), (44).

A regresszió- és varianciaanalízis kombinációja a *kovarianciaelemzés*. Ennek segítségével egy vagy több számszerűsíthető változónak és egy, esetleg több nominális változónak hatását vizsgálhatjuk a függő változóra. Úgy járunk el, hogy külön-külön számítunk regressziós egyenleteket a nominális változó szerint elkülönített megfigyelési egységekre, és azután kiszámíthatjuk mennyivel nőtt a függő változó varianciájából megmagyarázott rész azáltal, hogy külön-külön végeztük el a számítás. Ezt a logikát követtem, amikor külön-külön számítottam regressziós egyenleteket a termékenység és különböző gazdasági és társadalmi változók között Magyarország négy régiója számára, és megállapítottam, hogy ezek az egyenletek jobban illeszkedtek az adatokhoz, mint amikor az egész országra vonatkozóan egyetlen egyenletet számítottam (2). Ténylegesen azonban nem végeztem el a varianciafelbontást a változók között.

A típusok definiálása és elkülönítése kezdettől fogva foglalkoztatta a szociológia tudományát. Minden statisztikai osztályozás kidolgozásánál ugyanez a probléma merül fel. Amikor például társadalmi osztályokat és rétegeket különböztetünk meg, arra törekszünk, hogy olyanokat soroljunk egy kategóriába, akiknek társadalmi helyzete a leginkább hasonló. Amikor településtípusokat határozunk meg, akkor a leginkább hasonló városokat és falvakat igyekszünk összefogni. A számítógépeken elvégezhető matematikai statisztikai elemzési módszerek lehetővé teszik, hogy ne

intuíciónk, hanem számítások alapján és nemcsak néhány, hanem igen nagyszámú ismérv figyelembevételével határozzuk meg a típusokat, és soroljuk oda az egyes megfigyelési egységeket. Az egyik erre használható módszer a *fürtös elemzés* (cluster analysis).⁸ Ennek segítségével kisebb vagy nagyobb számú „csomóba” (klaszterbe) sorolhatjuk a megfigyelési egységeket, azután megkereshetjük az egyes „csomók” értelmezését. Ha sok mutatónk van a megfigyelési egységekről, akkor célszerű előbb faktoranalízist végezni, és a megfigyelési egységeknek az első néhány faktor dimenziójában számított faktorpontszámait alapján végezni el a fürtös elemzést. Így járt el *Institórisz Andrásné* a Nógrád megyei települések tipizálása céljából (19). Fürtös elemzéssel csoportosította a megyéket az 1964–1976. évi termékenységi jellemzők és változások alapján *Dányi Dezső* (47). A módszert alkalmazta nemzetközi struktúra-összehasonlítások céljára *Szilágyi György* (48). A módszer közérthetősége, a könyvtári program megléte és a faktoranalízissel való kombinálási lehetőség alapján talán nem elhamarkodott következtetés, hogy alkalmazása gyorsan elterjedhet a statisztikai osztályozási és típusalkotási feladatok megoldásában.

Tipizálásra használhatjuk a *diszkriminancia-analízist* is. Ez a módszer meghatároz egy diszkriminancia-függvényt, amelyben a különböző figyelembe vett ismérvek, változók szerepelnek. Azután a diszkriminancia-függvény alapján számítja ki, hogy az egyes megfigyelési egységeket milyen típusokban, csoportokban kell elhelyezni. Könyvtári program áll rendelkezésünkre, de eddig nem alkalmazták társadalmi és népesedési statisztikai elemzésekben.

Még két olyan módszert említek, amelyeknek könyvtári programjuk nem áll rendelkezésünkre, és amelyekkel nincsenek semmiféle tapasztalataink. A nemzetközi szakirodalomban azonban újabban gyakran lehet e módszerek statisztikai alkalmazásáról olvasni.

A *többdimenziós skálázás* (multidimensional scaling) segítségével meg lehet határozni nominális skálákon mért megfigyelési egységek egymástól való távolságát, számszerű értékeket lehet nekik adni, és így meg lehet állapítani hierarchikus sorrendjüket. Például egy mobilitási táblázat alapján meg lehet határozni a társadalmi osztályok és rétegek egymástól való távolságát és hierarchiáját szükség esetén több dimenzióban.⁹

A *log-lineáris elemzés* segítségével úgy lehet összehasonlítani két táblázatot – például két mobilitási táblázatot –, hogy a széleloszlások eltéréseinek hatását különválasztjuk az ún. belső interakciók különbségének hatásától. Két ország vagy két korszak társadalmi mobilitásának összehasonlításakor ez azt jelenti, hogy különválaszthatjuk a társadalmi–foglalkozási szerkezet eltéréseinek (például a kisebb mezőgazdasági népességnek és nagyobb értelmiségi rétegnek) hatását az osztályok és rétegek közötti közeledés vagy távolodás hatásától.¹⁰ Hangsúlyozni kell, hogy bár mindkét utoljára említett módszer alkalmazási példáit a társadalmi mobilitás elemzéséből vettem, azokat igen sok más társadalmi jelenség vizsgálatában is használják.

Végül meg kell említeni, hogy elvileg nincs akadálya annak, hogy a társadalmi folyamatokat is olyan szimultán egyenletrendszerből álló modellekkel írjuk le, mint amilyenekkel az ökonometria a gazdasági folyamatokat vizsgálja. Ezek a modellek

⁸ A fürtös elemzés programját *Csicsman József* írta le (45) és *Futó Péter* közölte róla egy módszertani cikket (46). A *Sigma* 1977. évi 3. számát e módszernek szentelte.

⁹ A többdimenziós skálázás módszerét ismerteti a külföldi szakirodalomban többek között *J. B. Kruskal* (49), (50), valamint az (52) munka. A magyar adatok alapján kísérleti számításokat végeztünk a Mannheimi Egyetem számítógépén.

¹⁰ A log-lineáris elemzés módszerét ismerteti a külföldi szakirodalomban többek között *S. E. Fienberg* (51). Magyar adatok alapján kísérleti számításokat végeztünk a Torontói Egyetem számítógépén a VIII. Szociológiai Világkongresszus alkalmával.

kiterjedhetnek egy-egy társadalmi jelenségre (például az egészségi állapotra és az egészségügy működésére), több társadalmi jelenségre együttesen vagy akár a társadalom valamennyi fő folyamatára. Az ilyen modellek felépítésének akadályait nem a módszertan területén látom, a nehézségek inkább a legmegfelelőbb társadalmi mutatók, jelzőszámok kiválasztásánál és az összefüggések elméleti tisztázásánál jelentkeznek.

Megkíséreltem bemutatni a többváltozós módszerek alkalmazásának fejlődését a társadalmi és népesedési statisztikai elemzésekben. Az áttekintésből, úgy gondolom, azt a következtetést vonhatjuk le, hogy az utóbbi években nagy előrehaladás történt annak köszönhetően, hogy rendelkezésünkre áll és állandóan fejlődik a szükséges számítástechnikai hardware és software. Mégis azt hiszem, hogy a lehetőségeknek csak viszonylag kis részét használjuk ki. Ennek fő oka, hogy a statisztikai elemző munkát végzők nagy többsége közép- és felsőfokú iskolai tanulmányai során nem tanulta – többnyire nem is tanulhatta – meg azokat a matematikai, matematikai statisztikai és számítástechnikai ismereteket, egyszóval azt a „számítástechnikai kultúrát”, amely a módszerek alkalmazásához szükséges. Ezért szinte az alapoktól kezdve kellene elsajátítaniok egy új egyetemi szintű ismeretanyagot. Bármennyire nagy erőfeszítéseket is igényel azonban ez, „kifizetődik” a statisztikai munkában, mert a korszerű matematikai statisztikai módszerek alkalmazása a számítógépek segítségével nagy mértékben megkönnyíti az elemzések munkaterhét, és ugyanakkor elmélyítheti az elemzést.

IRODALOM

- (1) Dr. Andorka Rudolf: A magyar népesség termékenységének alakulását befolyásoló gazdasági és társadalmi tényezők. *Demográfia*. 1967. évi 1. sz. 87–102. old.
- (2) Dr. Andorka Rudolf: A regionális termékenységi különbségeket befolyásoló gazdasági és társadalmi tényezők. *Demográfia*. 1969. évi 1–2. sz. 114–124. old.
- (3) Vita László: A faktoranalízis közgazdasági alkalmazásának lehetőségeiről. *Sigma*. 1970. évi 2. sz. 127–152. old.
- (4) Jahn, W. – Vahle, H.: A faktoranalízis és alkalmazása. Közgazdasági és Jogi Könyvkiadó. Budapest. 1974. 231 old.
- (5) Matematikai statisztikai alkalmazások kézikönyve. Szerk.: Gombosi Tamásné. KSH Számítástechnikai Igazgatóság. (Kézirat.)
- (6) Gombosiné Gárdos Eszter: Programrendszerek matematikai statisztikai elemzésekhez. *Statisztikai Szemle*. 1974. évi 7. sz. 671–683. old.
- (7) Zágon Csaba: A faktoranalízis alkalmazása a statisztikai gyakorlatban. *Statisztikai Szemle*. 1979. évi 11. sz. 1105–1128. old.
- (8) Cseh-Szombathy László: Az életszínvonal nemzetközi összehasonlításánál alkalmazott mutatószámok kiválasztása. *Statisztikai Szemle*. 1960. évi 7. sz. 678–694. old.
- (9) Dr. Végső Zoltán: Adalékok az Alföld gazdaságfejlődéséhez. *Területi Statisztika*. 1974. évi 6. sz. 630–694. old.
- (10) Abonyi Gyuláné – Móricz Ferenc: Az infrastruktúra területi vizsgálata. *Területi Statisztika*. 1975. évi 1. sz. 82–90. old.
- (11) Dr. Francia László: A faktoranalízis alkalmazása a lakosság életkörülményei és az infrastrukturális ellátottság közötti összefüggések területi elemzésében. *Területi Statisztika*. 1975. évi 3. sz. 245–253. old.
- (12) Dr. Laczkó László: A kedvezőtlen feltételekkel rendelkező területek fontosabb jellemző vonásai. *Területi Statisztika*. 1975. évi 4. sz. 352–362. old.
- (13) Dr. Laczkó László: Az ország elmaradott területeinek vizsgálata. *Területi Statisztika*. 1975. évi 5. sz. 474–485. old.
- (14) Dr. Kőszegfalvi György: Városaink infrastrukturális ellátottsági viszonyainak színvonala és a fejlesztés néhány problémája. *Területi Statisztika*. 1975. évi 3. sz. 263–275. old.
- (15) Dr. Enyedi György: A falusi életkörülmények területi típusai Magyarországon. *Területi Statisztika*. 1976. évi 3. sz. 217–225. old.
- (16) Andorka Rudolf: A faktoranalízis alkalmazása társadalom-ökonometriai vizsgálatokban. *Sigma*. 1976. évi 3. sz. 159–177. old.
- (17) Dr. Andorka Rudolf: A faktoranalízis felhasználása a regionális vizsgálatokban. *Területi Statisztika*. 1979. évi 1. sz. 8–17. old.
- (18) Kerekesné dr. Varga Éva: Az urbanizációs folyamat statisztikai vizsgálata a Pest megyei példák alapján. *Területi Statisztika*. 1979. évi 3. sz. 261–271. old.
- (19) Dr. Institorisz Andrásné: Az urbanizációs folyamatok statisztikai vizsgálata Nógrád megyei adatok alapján. *Területi Statisztika*. 1979. évi 3. sz. 282–297. old.
- (20) Dr. Lukács Pál: Az urbanizáció és a lakosság kulturális színvonalának összefüggései az alföldi városokban. *Területi Statisztika*. 1979. évi 2. sz. 131–150. old.

- (21) Szabady Balázs: A termékenység területi különbségeinek és változásainak okai. *Demográfia*, 1977. évi 4. sz. 413–467. old.
- (22) Szabady Balázs: A körzetesítés módszertani kérdései és Magyarország demográfiai körzetesítése. *Demográfia*, 1973. évi 2. sz. 197–215. old.
- (23) Vargáné Teghze-Gerber Zsuzsanna – Gombosiné Gárdos Eszter: A budapesti gyermekek testméreteinek kiértékelése faktoranalízissel. *Demográfia*, 1976. évi 2–3. sz. 184–211. old.
- (24) Dr. Szilágyi György: A gazdasági színvonal és struktúra összehasonlítása faktoranalízissel. *Statistikai Szemle*, 1978. évi 2. sz. 142–161. old.
- (25) Dr. Barys Gyula – Dr. Sárkány Jenő: A művi vetélések hatása a születési mozgalomra és a csecsemőhalandóságra. *Demográfia*, 1963. évi 4. sz. 427–467. old.
- (26) Dr. Miltényi Károly: A művi vetélések hatásainak kérdéséhez. *Demográfia*, 1964. évi 1. sz. 73–87. old.
- (27) Dr. Mikolás Miklós: Az abortuszlegalizáció népegészségügyi hatása és egyes társadalmi kísérőjelenségei Magyarországon. *Demográfia*, 1973. évi 1. sz. 70–113. old.
- (28) Dr. Andorka Rudolf – Cseh-Szombathy László – Dr. Vukovich György: Módszertani megjegyzések az abortusz hatásainak vizsgálatához. *Demográfia*, 1974. évi 1. sz. 63–73. old.
- (29) Dr. Szokolczai György: Az abortuszlegalizáció hatásai: módszertani problémák. *Demográfia*, 1976. évi 1. sz. 55–69. old.
- (30) Lits Józsefné dr.: A csecsemőhalandóság alakulása és befolyásoló tényezői Szolnok megyében. *Területi Statisztika*, 1977. évi 5. sz. 504–521. old.
- (31) Dr. Lengyel László – Olajos Árpád: Diploma és kereset. *Statistikai Szemle*, 1967. évi 5. sz. 387–408. old.
- (32) Somogyi Miklós: A nő- és a férfi dolgozók bérarányai a szakképzett alkalmazottak körében. *Statistikai Szemle*, 1975. évi 2. sz. 147–160. old.
- (33) Pártos Judit: Magyarország kulturális helyzetének nemzetközi összehasonlítása. *Statistikai Szemle*, 1972. évi 8–9. sz. 919–931. old.
- (34) Dr. Sántha Józsefné: A külterület vizsgálatának statisztikai módszerei. *Területi Statisztika*, 1970. évi 1. sz. 23–40. old.
- (35) Dr. Lengyel László: A foglalkoztatottság területi alakulása. *Statistikai Szemle*, 1968. évi 3. sz. 235–254. old.
- (36) Pápai Béla: A budapesti agglomeráció. *Demográfia*, 1967. évi 1. sz. 69–86. old.
- (37) Szauter Edit: A belföldi vándormozgalom alakulásának néhány jellegzetessége a felszabadulás után. *Területi Statisztika*, 1974. évi 3. sz. 295–310. old.
- (38) Dr. Andorka Rudolf – Cseh-Szombathy László – Dr. Vavró István: Társadalmi eltérés alá eső magatartások előfordulásainak területi különbségei. *Statistikai Szemle*, 1968. évi 1. sz. 43–54. old.; 1968. évi 2. sz. 145–158. old.
- (39) Fonyódi Valéria – Dr. Gombosiné Gárdos Eszter – Harsányi László: Bér- és keresetelemzés regresszióanalízissel. *Statistikai Szemle*, 1977. évi 2. sz. 163–174. old.
- (40) Éltető Ödön – Vita László: A családi jövedelmek becslése regressziós módszerrel. *Statistikai Szemle*, 1978. évi 6. sz. 596–608. old.
- (41) Surányi Bálint – Vita László: A pályakezdők vertikális társadalmi mobilitására ható tényezők vizsgálata az útelemzés módszerével. *Szociológia*, 1973. évi 1. sz. 68–82. old.; 1973. évi 2. sz. 244–263. old.
- (42) Dr. Melega Tiborné: A munkabérekre ható tényezők statisztikai vizsgálata. Megjelent: *Korszerű statisztikai törekvések Magyarországon*. Szerk.: Mád Aladárné. Akadémiai Kiadó, Budapest, 1968. 569–577. old.
- (43) Dr. Frigyes Ervin: A munkás és alkalmazotti jövedelmi rétegződés legfontosabb tényezői. *Statistikai Szemle*, 1964. évi 7. sz. 748–766. old.
- (44) Éltető Ödön: A jövedelemeloszlások egyenlőtlensége Magyarországon. Megjelent: *Korszerű statisztikai törekvések Magyarországon*. Szerk.: Mád Aladárné. Akadémiai Kiadó, Budapest, 1968. 533–543. old.
- (45) Csicsman József: A klaszter-elemzés módszerei és alkalmazási lehetőségei a statisztikában. *Statistikai Szemle*, 1979. évi 2. sz. 137–145. old.
- (46) Dr. Futó Péter: Hipergráf modellen alapuló klaszter-elemzés és alkalmazása. *Statistikai Szemle*, 1979. évi 2. sz. 130–136. old.
- (47) Dr. Dányi Dezső: Az 1964–1976. évi termékenység elemzése. Kísérlet. *Demográfia*, 1978. évi 2–3. sz. 221–243. old.; 1978. évi 4. sz. 452–466. old.
- (48) Dr. Szilágyi György: Nemzetközi struktúra-összehasonlítások klaszter-elemzéssel. *Statistikai Szemle*, 1979. évi 10. sz. 955–972. old.
- (49) Kruskal, J. B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964. évi 1. sz. 1–27. old.
- (50) Kruskal, J. B.: Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 1964. évi 2. sz. 115–129. old.
- (51) Fienberg, S. E.: The analysis of cross-classified categorical data. MIT Press, Cambridge, Mass. 1977. 151 old.
- (52) Multidimensional scaling. Theory and application, Szerk.: R. N. Shephard, A. K. Romney, S. B. Nerlove. 1. köt. Theory. 2. köt. Application. Seminar Press, New York, 1972.

РЕЗЮМЕ

Автор останавливается на применении многопеременных математико-статистических методов в общественно-статистических и демографико-статистических анализах. Возможность для применения этих методов возникла благодаря развитию вычислительной техники и разработке соответствующих программ.

В дальнейшем автор рассматривает, в какой мере используются существующие возможности. Отмечает, что, хотя в течение истекших лет в этой области был достигнут значительный прогресс, эти методы можно было бы применять в более широкой сфере деятельности.

В дальнейшем разделе своего очерка автор занимается следующими многопеременными методами: факторный анализ, многократный корреляционный и рег-

рессивный анализ, курсовой анализ, канонический корреляционный расчет, многократный дисперсионный анализ, анализ ковариантности, кластерный анализ, дискриминационный анализ, многомерный шкальный анализ, логарифмическолинейный анализ, симультантные системы уравнений.

SUMMARY

The study discusses the application of multivariate methods of mathematical statistics in the analyses of social and vital statistics. The use of these methods was rendered possible by the development of computing science and by the elaboration of programs serving this purpose.

The author investigates to what extent the possibilities afforded are utilized. He points out that, although the development was notable in the years past, the methods might have been used in a wider sphere than they are actually used at present.

In the subsequent part of the study the author discusses the following multivariate methods: factor analysis, multiple correlation and regression analysis, path analysis, canonical correlation, multiple variance analysis, co-variance analysis, cluster analysis, discriminant analysis, multi-dimensional scaling, log-linear analysis, simultaneous equation systems.