

Csüllög Mária–Kiss Tibor

A FAKTORELEMZÉS LÉNYEGE ÉS ALKALMAZÁSI PROBLÉMÁI

A társadalmi-gazdasági tényezők egymás közötti kapcsolatát, kölcsönhatását nehéz, sőt elképzelhetetlen a teljesség igényével számszerűsíteni, matematikai függvénykapcsolatként kezelni pedig biztos tévedés. Az egy bizonyos tényező alakulását befolyásoló összes mutatót még felsorolni is nehéz, nemhogy számokkal kifejezni. A mutatók egymás közötti összefüggése, visszahatása, a közvetetten jelentkező hatások szintén megnehezítik a statisztikai elemzést. Ennek ellenére ma már nem vitás, ahogy elkerülhetetlen a kvantitatív módszerek alkalmazása a társadalmi-gazdasági folyamatok elemzésében. Ezért szükséges az említett problémák elemzésében olyan matematikaistatisztikai módszer alkalmazása, amely lehetővé teszi az összetett rendszer elemzését, belső bonyolult törvényszerűségeinek feltárását, valamint a vizsgált problémát lényegesen befolyásoló tényezők kiszűrését. A faktorelemzés bizonyos megszorításokkal eleget tesz az említett követelményeknek.

A faktorelemzés alapja az, hogy az eredeti változókat, illetve a vizsgált rendszerre vonatkozó kvantitatív adatokat sűrített formában szolgáltatja, lényegesen kevesebb számú hipotetikus változó, úgynevezett faktor segítségével.

A módszer külön érdekessége, hogy a pszichometriában alkalmazták még a század elején. Charles Spearman pszichológus az intellektuális képességek vizsgálatában állította fel az első faktorelemzési modellt.¹ Külön ki kell hangsúlyoznunk, hogy a faktorelemzési modellnek, illetve modellezésnek is, mint általában a matematikai-statisztikai modellezésnek, két oldala van. Az egyik maga a modell – jelen esetben a faktorelemzési modell – amelyben a vizsgált tényezőt, jelenséget nem valós, hanem sűrített, összevont mutatókkal írjuk le. A modellezés problémájának másik oldala pedig a matematikai-statisztikai apparátus, illetve az, hogy milyen matematikai-statisztikai eljárásokkal fogjuk előállítani az említett sűrített mutatókat. Így, igaz, hogy a faktorelemzési modell a pszichológiában született meg, de ennek feltétele, majd igénye lett a megfelelő matematikai-statisztikai módszerek jelenléte, továbbfejlesztése. Ennek kapcsán el kell mondani azt is, hogy a faktorelemzési módszer alkalmazására, illetve alkalmazási területének bővítésére, majd ebből kifolyólag a különböző algoritmusok kidolgozására és a meglévők tökéletesítésé-

sére döntő hatást gyakorolt a számítógépek megjelenése, fejlődése és alkalmazása. A matematikai rész lebonyolítása ma már kész számítógépes programcsomagokkal történik, ezért a problémát nem kívánjuk az egyes algoritmusok hatékonysága, alkalmazási feltétele vagy esetleg tökéletesítése szempontjából vizsgálni. Ezzel kapcsolatban szeretnénk felhívni a figyelmet arra, hogy az alkalmazó szempontjából szükséges a matematikai rész ismerete, vagyis tudni kell azt, mi történik a bemenő adatokkal, és mit jelentenek, hogyan, milyen kikötésekkel értelmezhetők az eredményként kapott mutatók. Füstös László és munkatársai² az adatelemzés sokváltozós módszereit elemezve nagyon találóan hangsúlyozzák ki, hogy „A gondolatmenetek végigszámolásától mindenkit megkímél a korszerű számítástechnika, s ez nagyon jó dolog...”, de, így folytatják, „... bőséges lehetőség nyílik arra is, hogy egy halom összegyűjtött adatot leadva egy számítóközpontba, onnan úgy kapjunk vissza kész számítási eredményeket, hogy az alkalmazott módszerek lényegéről, logikai háttéréről, feltételrendszeréről szinte alig tudunk valamit. A törvényszerű következmény a téves, torz vagy a legjobb esetben eredménytelen alkalmazások sora lesz”. Teljes egészében egyetértünk ezzel a megállapítással, és ezért szeretnénk rámutatni néhány problémára a faktorelemzést illetően.

Kezdjük azzal, hogy a faktorelemzés többféle különböző problémakör felölésére alkalmas, és többféle faktorelemzési eljárás ismeretes.

A faktorelemzési módszerek legfontosabb alkalmazási területei a következők:

1. Korrelálatlan változók előállítás – A regressziós elemzés egyik problémája, hogy a magyarázó változók között korrelációs kapcsolat áll fenn, ami torzítja a regressziós egyenlet paramétereinek becslését, és ezen keresztül torzul a megállapított regressziós egyenlet, valamint az összefüggések értelmezése is. Jelentős multikollinearitás jelenléte esetén nehéz egyértelműen meghatározni a lényeges magyarázó változók halmazát, mivel az egyes tényezők a valóstól eltérő jelentőséget kapnak. A multikollinearitás kiküszöbölésére szolgáló módszerek közül a faktorelemzés is hatékony segédeszköznek bizonyul. A faktorelemzés módszereinek alkalmazásával korrelálatlan faktorokat állíthatunk elő, s ezekkel becsülhetjük a regressziós egyenletet.

2. Lényegkiemelés – A faktorelemzés alkalmas az elemzett rendszer alakulását befolyásoló mutatók kiválasztására is. Egyrészt választ kapunk arra a kérdésre, hogy melyik eredeti mutatókat szükséges egyáltalán bekapcsolni az elemzésbe, másrészt az is kiderül, hogy egy-egy faktor kialakításához hány eredeti változó járult hozzá.

3. Mérőszámként való felhasználás – Ha egy úgynevezett főfaktorral sikerül a rendszer varianciájának lényeges részét megmagyarázni, akkor ezt a főfaktor rangsorolási kritériumként használhatjuk. A főfaktoroknak mérőszámként való alkalmazása különösen azokban az esetekben indokolt, amikor a rangsorolást csak több változóval jellemezhető közgazdasági kategória alapján kell elvégezni. Ilyen kategóriák például a fejlettségi szint, a hatékonyság, a munkaszervezetek nagysága vagy szervezettsége stb. Ha egy rangsorolásnál csak egy mutatót veszünk figyelembe – mert ez a megszokott gyakorlat – de bizonyított tény, hogy a vizsgált folyamatot vagy az adott mérhető helyzetet több változó

jellemzi, faktorelemzéssel leellenőrizhetjük – a jövőbeni korrigált hozzáállás érdekében – hogy helyesen jártunk-e el egyetlenegy mutató alkalmazásával. A már említett fejlettségi szint szerinti rangsorolást is szokásos egy mutató, az egy főre eső nemzeti jövedelem szerint végezni, azonban világos, hogy számtalan társadalmi-gazdasági tényező határozza meg a valós fejlettségi szintet.

4. Megfigyelések osztályozása és típusainak kialakítása – Az adott megfigyelések halmazát osztályokba sorolhatjuk az eredeti változókból képzett korrelálatlan faktorok segítségével. A faktorértékekre osztályokat alakíthatunk, s egy csoportba kerülnek majd azok a megfigyelések, amelyek megfelelő faktorértékei egy adott intervallumba tartoznak. Az egy csoportba tartozó faktoroknak megfelelő megfigyelésértékek alapján állapíthatók meg az adott csoportra jellemző típusjegyek.

Az említett alkalmazási területek bármelyikét is választjuk, a faktorelemzés lebonyolítása bizonyos problémákkal jár. Olyan problémák ezek, amelyeket a számítógépes program semmiképpen sem old meg: a megoldásokhoz mély szaktudományi ismeretekre alapozott minőségi elemzés szükséges.

A munka első szakasza a cél meghatározása, valamint ezzel összefüggésben a megfelelő mutatók kiválasztása. Ezzel kapcsolatban fontos elemzési feladat a felhasználandó mutatók, illetve változók meghatározása. Továbbá el kell dönteni, hogy milyen szerepük lesz a faktorelemzési modellben az egyes változóknak; ki kell vizsgálni e változók mérhetőségét, a rendelkezésre álló adatbázist, az adatbázis bővíthetőségét s ennek költségeit, a megfigyelés objektumait stb. Ilyen jellegű problémákkal találjuk magunkat szemben például területi vagy nemzetközi vizsgálatoknál: egyes statisztikai mutatók csak országos szinten állnak rendelkezésre, kisebb területi egységekre már nem; ha több ország szintjén végzünk elemzést ismernünk kell az egyes statisztikai mutatók meghatározási módját, mert gyakran ugyanolyan elnevezés mögött más belső tartalom van.

Tekintetbe kell venni azt is, hogy a faktorelemzésnél a változók számát elvben korlátlanul növelhetjük, de a mérések száma, azaz a megfigyelt objektumok száma lényegesen nagyobb kell hogy legyen a változók számánál. Egyesek szerint ez az arány 1:5, de csak abban az esetben, ha legalább húsz változót szerepeltetünk a modellben. A kész programcsomagokban feltétlenül meg van adva milyen dimenziójú lehet az adatmátrix, s annál nagyobb méretű nem kezelhető az adott programmal.

A második lépésben el kell dönteni, hogy szükséges-e a faktorelemzés, esetleg többszörös faktorelemzés, és ki kell választani a megfelelő faktorelemzési modellt. Ehhez nem csak a különböző faktorelemzési modellek sajátosságait kell ismerni, hanem a többi többváltozós statisztikai módszert is. Például, ha az egyes területek fejlettségi szintjét akarjuk kivizsgálni, és az eredeti, a mért mutatóértékek eléggé áttekinthetők – ez főleg abban az esetben fordul elő, ha a mutatók száma aránylag kicsi – nem szükséges faktorelemzéssel összevonni a mutatókat. Továbbá, ha regressziós elemzést végzünk, és nem jelentkezik a multikollinearitás negatív hatása, szintén felesleges a faktorok előállítását, hiszen a regressziós egyenlet paraméterei ilyen esetben az eredeti változók segítségével is megfelelő pontossággal becsülhetők fel!

A következő probléma a faktorszám meghatározása. Ez általában, kész programcsomagok alkalmazása esetében is, az alkalmazóra van bízva. Ha túl nagy a faktorok száma az eredeti változók számához mérten, akkor a faktorértékek szintén áttekinthetetlenek, és nem felelnek meg az elemzés céljainak. Ilyen esetben célszerű az adatmátrix bővítése vagy a faktorszám meghatározására alkalmazott kritérium felülvizsgálása.

A faktorok előállításához több különböző jellegű, sőt különböző mértékegységben kifejezett eredeti változó járul hozzá, ami aztán gyakran megnehezíti a kapott faktorok, illetve faktorértékek értelmezését. Sok esetben éppen a kézenfekvőbb értelmezés biztosításához szükséges a kapott faktorok rotációja, transzformálása. A rotáció szükségességéről szintén a programcsomagok használója dönthet. Mivel a programok úgy vannak felállítva, hogy a rotált és a nemrotált faktorértékeket is megadják, e döntés nem jelent komolyabb gondot, ha eltekintünk a feldolgozás gépidejétől és költségétől. Az elemzésnél feltétlenül figyelembe kell venni az eredeti adatmátrixot, a nemrotált és a rotált faktorérték-mátrixot is.

A felhasználó dönt az elemzés más minta, bővített adatbázis, más faktorszám alapján való megismérlésének létjogosultságáról is.

Végül szeretnénk kihangsúlyozni, hogy a faktorelemzéssel kapott faktorértékek nem közvetlenül megfigyelhető értékek, hanem az eredeti változók alapján előállított értékek. Ez semmiképpen sem csökkenti a módszer jelentőségét és alkalmazásának fontosságát a felsorolt célok elérésében, olyan bonyolult rendszerek elemzésében, amelyeket csak rendkívül sok, stochasztikus változóval tudunk jellemezni.

Jegyzetek

- ¹ Charles Spearman: General Intelligence, Objectively Determined and Measured. American Journal of Psychology, 1904.
- ² Füstös László, Meszéna György és Simonné Mosolygó Nóra: Bevezetés az adatelemzés sokváltozós módszereibe. Tankönyvkiadó, Budapest, 1983.

Rezime

Suština i problemi primene faktorske analize

Faktorska analiza predstavlja skup metoda analize zasnovanih na matematičko-statističkim postupcima, razvijeni su ciljem ispitivanja pojava u svim onim područjima gde se pojavljuje veliki broj promenljivih sa složenim međusobnim vezama i različitim stepenima međuzavisnosti. Pomoću faktorske analize konstruiše se u odnosu na broj analiziranih pojava manji broj faktora koji predstavljaju hipotetičke promenljive ali koji mogu da zamene originalne promenljive uz zadržavanje najvećeg dela početnih informacija. Različiti su osnovni problemi koji se na zadovoljavajući način mogu tretirati primenom faktorske analize. Najčešće oblasti primene faktorske analize su izbor bitnih či-

nilaca formiranja pojave, utvrđivanje mernog broja karakteristike nekog sistema, formiranje skupa nekorelisanih promenljivih, klasifikacija i tipizacija.

Korišćenje elektronskih računara od bitnog je značaja za razvoj i primenu metoda faktorske analize. Autori ističu, da je i pored upotrebe gotovih programskih paketa faktorske analize, nezaobilazna uloga korisnika u postavljanju ciljeva analize, izboru skupa relevantnih pokazatelja i načina primene faktorske analize.

Summary

The Essence and Problems of Factor Analysis Application

The factor analysis represents the complex of methods based on mathematical and statistical procedures developed with the aim to investigate phenomenas in all those fields, where a great number of variables with complex interrelations and different interdependence levels appears. In relation to the number of the analysed phenomena a smaller number of factors is being constructed by means of the factor analysis, representing the hypothetical variables, which may replace the original ones, keeping the greatest part of the initial information. The basic problems, which may be treated satisfactorily by the application of the factor analysis, are different. The most frequent fields of its application are the choice of essential factors creating the phenomena, determination of measuring numbers of the system characteristics, developing the complex of uncorrelated variables, classification and standardization.

The use of electronic computers is of essential importance for the development and application of the factor analysis methods. The authors emphasize that beside the use of their ready-made program packages of the factor analysis, the role of the user is unavoidable in formulating the analysis purposes, choice of the complex of relevant indicators and ways of the factor analysis application.