

## Könyvszemle

**Kovács Erzsébet:**

**Pénzügyi adatok statisztikai elemzése**

Negyedik bővített kiadás. Tanszék Kft. 2011. Budapest. 228 old.

Adatok sokasága vesz minket körül, és ezek rengetegében korántsem egyszerű feladat eligazodni. Gyakran megesik, hogy éppen a túl sok adat miatt nem könnyű észrevenni a közöttük levő összefüggéseket; hasznosításuk – tömegüknek és komplexitásuknak köszönhetően – egyre nehezebb, és a szakembereket is komoly kihívások elé állítja. Mindennek köszönhetően a többváltozós statisztikai módszerek jelentősége növekszik. Alkalmazásuk kapcsán azonban számos kérdés merül fel. Mely változókkal érdemes dolgozni, azoknak milyen előfeltételeknek kell megfelelniük, valamint melyik elemzési módszert válasszuk egy adott statisztikai probléma kapcsán? Ezek és sok egyéb kérdés megválaszolásához nyújt segítséget *Kovács Erzsébet* „Pénzügyi adatok statisztikai elemzése” című könyvének újabb kiadása.

Az előző kiadáshoz képest a kiadvány három új fejezettel és szerzővel bővült: *Vékás Péter* egy újabb modellezési eljárást, a túlélési modelleket mutatja be. Ezen felül *Szűle Borbála* és *Fliszár Vilmos* esettanulmánya gazdagítja a tankönyvet. Mindkét tanulmány a korábbi fejezetekben ismertetett eljárásokat felhasználva készít elemzést egy-egy problémára. A kötet a bővítés ellenére is megőrizte egyik – az egyetemi hallgatók számára – különösen fontos előnyét: rövideségét. A mindössze 228

oldalnyi tudástár könnyen tanulható, és méreténél fogva a hordozhatóság sem jelent akadályt.

Bár a könyv az alapvető statisztikai módszerek ismeretét feltételezi, azért röviden emlékeztet a fontosabb összefüggésekre, így a statisztikában kevésbé járatos olvasók számára is érdekes olvasmány lehet. Minden fejezet az adott módszert bemutató általános, matematikai leírással kezdődik, majd ennek specifikusabb változataival folytatódik, végezetül pedig a probléma SPSS-programban történő megvalósítására láthatunk példát, az eredmények értelmezésével. A futtatásokat bárki könnyedén megismételheti, hiszen a feladatok – néhány kivételtől eltekintve – a programban megtalálható mintafájlok alapján készültek.

A tankönyv elsőként a leíró és feltáró adatelemzést ismerteti. Külön felhívja a figyelmet arra, hogy az elemzés lehetőségei nagymértékben függenek attól, hogy milyen szintű mérési skála áll rendelkezésünkre. Kovács Erzsébet táblázatban foglalja össze, hogy melyik mutatószámhoz legalább milyen skálájú változóra van szükség, ami nagyban segíti a megértést. A fejezet további részében a leíró statisztika készítéséhez leggyakrabban használt SPSS-parancsok (frequencies, descriptives, explore) bemutatása található meg; nem csupán ezek futtatására és értelmezésére találhatunk példákat, a szerző röviden kitér a „paste” utasításra is. A gyakorlottabbakat pedig a gondolkodtató feladatok bizonyos összefüggések levezetésére ösztönzik.

A tankönyv a keresztábrás elemzésekkel folytatódik (amely még mindig kevésbé tekinthető többváltozós statisztikai módszernek, mi-

vel jellemzően két változóval dolgozik). A szerző már a fejezet elején világosan elhatárolja a kapcsolatvizsgálat ezen típusát más mérési szintű kapcsolatvizsgálatoktól, amit különösen fontosnak tartok, mivel statisztikával foglalkozó tanárként gyakran tapasztalom, hogy a hallgatók nincsenek tisztában az asszociáció, a vegyes kapcsolat és a korreláció közötti különbséggel. Később szó esik skálatranszformációról is, ami megmutatja, hogy milyen lehetőségei vannak az elemzőnek a különböző mérési szintű változók alkalmazásakor. A továbbiakban a fejezet a nominális és ordinális szintű mérési változók közötti kapcsolatvizsgálatra fókuszál: az SPSS-ben megtalálható mérőszámokat mutatja be. Ezek képleteit is tartalmazza a fejezet, valamint az előzőhöz hasonlóan itt is találhatunk értelmezésekkel kibővített példákat.

A kezdő statisztikusoknak már az eddigi fejezetek is bőséggel szolgáltattak elemzési lehetőségeket, az igazi többváltozós statisztikai elemzési módszerek bemutatása azonban (amennyiben az explore parancs almintákra bontásával nem számolunk) csak ezután kezdődik.

A klaszterelemzés fejezetben a szerző rátér az igazi többváltozós statisztikai módszerek bemutatására. A csoportosítási eljárás általános ismertetése után két klaszterezési technika részletesebb leírása következik: a hierarchikus, illetve a nemhierarchikus klaszterezésé. Ezek bemutatása szintén az SPSS nyújtotta kereteken belül történik: a leírásban többek között a program segítségével futtatható távolságmértékek képleteit is megtalálja az olvasó. A fejezet az SPSS-futtatással és annak elemzésével végződik. A szerző annak érdekében, hogy valóban szemléletes legyen a példa, ezúttal egy mesterségesen konstruált adatbázison mutatja be az elemzési módszereket.

A többváltozós regressziószámítás fejezetében az alapmodellt, a lineáris regressziót ismerteti Kovács Erzsébet: azon túl, hogy vilá-

gosan leírja a modell alkalmazhatóságának feltételeit, magát a modell építésének folyamatát, a reziduálisokra vonatkozó elnevezéseket is tisztázza.

A főkomponens- és faktorelemzés ugyan egy fejezetben szerepel, mégis az alfejezeteknek köszönhetően világosan elkülönül egymástól a két témakör. Az eljárások matematikai háttere meglehetősen részletes, és biztosan minden eleme helyére kerül, ha a bemutatott feladatokat végignézi az olvasó. Amellett, hogy a futtatások mellett szokás szerint megtalálhatók az értelmezések is, a főkomponens-elemzés elején külön táblázat figyelmeztet a modell alkalmazhatóságának azon mérőszámaira, melyeket az SPSS-ben is lehet kérni.

A diszkriminanciaelemzés kapcsán ismét a matematikai háttér leírásával találkozhatunk először, amit egy, az eddigieknél még részletesebb SPSS-futtatás és értelmezés követ. A szerző egy mesterségesen konstruált adatbázist használ fel ehhez, mivel az ilyen típusú elemzés alkalmazásának rendkívül szigorú előfeltételei vannak. Ezekről részletesebben a fejezet végén is szó esik; a könyv külön kiemeli azokat az eseteket, melyeket az SPSS segítségével nem lehet tesztelni.

A következő fejezetben a sokdimenziós skálázással találkozunk, mely a főkomponens-elemzéshez erősen hasonlító eljárás: alapvetően a változók számának redukciójára törekszik. A tankönyv párhuzamot von emellett a klaszterezési eljárásokkal is, mivel ezen elemzés elsősorban a megfigyelések térképezésével foglalkozik. Mind a klasszikus és az ordinális skálázás módszerét, mind ezek SPSS-beli megvalósításait megismerhetjük. Egy szemléletes példa után az egyéni különbségek skálázásával egészül ki a fejezet, melyet egy másik adatbázison futtatott elemzés zár le.

A logisztikus regressziós modell fogalmának bevezetésekor az olvasó egy illusztratív példának köszönhetően azonnal választ kaphat

arra a ki nem mondott kérdésre, hogy miért nem a lineáris regressziót alkalmazzuk a bináris függő változókra? A tankönyv világosan elhatárolja egymástól e két módszert, és szemléletes, életszerű példákat hoz a logisztikus regresszió alkalmazási területeire.

Az eddigi fejezetek mindegyike Kovács Erzsébet tollából született, és a tankönyv régebbi kiadásában is megtalálható volt. Felfedezhetők azonban kisebb-nagyobb különbségek: sok esetben bővültek az egyes témakörök. A gondos munkának köszönhetően a fejezetek mindegyike letisztultabbá, átláthatóbbá vált. Emellett a korábbi kiadásokhoz képest több példa is szerepel, illetve megjelentek a már korábban említett, az olvasót elmélyült gondolkodásra készítő feladatok is.

A tárgyalt elemzési módszerek köre a túlélési modellekkel bővült. Ezek a bizonyos eseményig eltelt időt, illetve az ehhez tartozó valószínűség-eloszlást modellezik. A probléma megértése érdekében Vékás Péter, egy példával indít, melyben hitelszerződés elemzését végzi, fiktív adatokon. Ez a rövid, egyszerű példa nemcsak az adatok bemutatásánál, de még a modellépítésnél is előkerül, aminek köszönhetően az olvasó számára könnyen érthetővé válik annak menete. A fejezet elsősorban a Kaplan–Meier-moddellel és a Cox-regresszióval foglalkozik, de ez utóbbi megértésének érdekében külön alfejezetet szentel a kockázati rátának is. A két alfejezet végén külön-külön ismerteti az SPSS által kínált elemzési lehetőségeket, majd mintapéldán is bemutatja az addig leírtakat. Mivel egy ilyen túlélési modellhez speciális adatbázisra van szükség, így a szerző saját adatállománnyal dolgozik.

A tankönyv két esettanulmánnyal végződik, melyek mindegyike egy adott problémát jár körül a korábbiakban bemutatott elemzési technikák széles köréből válogatva. A két fejezet jó példát szolgáltat a hallgatóknak arra, hogy miként épül fel egy adatsor statisztikai

elemzése. Kiválóan illusztrálja, hogy a statisztikusnak egy adott probléma esetén milyen elemzési szempontok alapján kell elindulniuk.

Szüle Borbála esettanulmánya a kvalitatív változók közötti összefüggések elemzésével foglalkozik. Változatos eszköztárat vonultat fel az SPSS mintafájlaiban is megtalálható adatbázis elemzéséhez. A leíró és feltáró elemzésekben használt módszerek után a többdimenziós skálázás, valamint a hierarchikus klaszterelemzés segítségével vizsgálja a változókat, az eredmények teszteléséhez pedig keresztábrák elemzéseket alkalmaz. Mindezeket az SPSS-beli futtatás eredményeivel is illusztrálja. A tanulmány kifejezetten olvasmányosnak mondható, a gondolatmenet világos, könnyedén követhető, részletes összefoglalást nyújt a kvalitatív változók elemzéséről.

A második esettanulmány egy pénzügyi, banki alkalmazásra mutat példát: az ügyfelek hitelminősítésének statisztikai módszereivel foglalkozik. Mivel a jó és rossz ügyfelek azonosítása létfontosságú a bankok számára, így a legjobban becsülő modell megtalálása különösen fontos feladat számukra. A fejezetben egy hitelminősítő-modell kialakítását követheti nyomon az olvasó, amely szintén SPSS-adatbázisra készült. A szerző, Fliszár Vilmos nem csupán az alkalmazható módszereket mutatja be, de rövid összefoglalást is ad azok előfeltevéseiről, és külön felhívja a figyelmet az adattisztítás jelentőségére. A fejezetben szó esik továbbá arról, hogy milyen módon érdemes a meglévő modelleket összehasonlítani, hogyan lehet azok helyességéről megbizonyosodni. Ez utóbbihoz több, a leíró statisztikából ismert ábrát használ fel, ami a modellek közötti választások esetében nagy segítséget nyújthat az elemzőnek, egyúttal illusztratív példával is szolgálhat az olvasónak.

Összességében úgy gondolom, hogy a tankönyv újabb kiadása nemcsak értékes fejeze-

tekkel gazdagodott, de a meglévők csiszolásának köszönhetően, még több segítséget nyújthat annak, aki a többváltozós statisztikai elemzés világában próbál eligazodni. A könyv logikailag jól felépített, precíz, igényes munka,

és sok-sok példájának köszönhetően tudatosan segíti a megértést.

#### Rétallér Orsolya

PhD-hallgató, Budapesti Corvinus Egyetem  
E-mail: retaller.orsolya@gmail.com

## Folyóiratszemle

**Nöthen, M.:**

### A korösszetétel a fő oka a német egészségügy magas költségeinek?

(Hohe Kosten im Gesundheitswesen: Eine Frage des Alters?) – *Wirtschaft und Statistik*. 2011. No. 7. pp. 665–675.

A szerző a hivatalos német egészségügyi statisztika fontosabb eredményeire támaszkodva, több nézőpontból elemzi a kórházi kezelések költségeinek alakulását. A demográfiai hatásokat figyelembe véve, kísérletet tesz a várható költségalakulás 2030-ig terjedő előrejelzésére is.

Németországban 2008. év folyamán összesen 254,3 milliárd eurót költöttek a 82,1 millió főnyi népesség egészségének megőrzésére és a betegségek elhárítására. Az összeg közel felét (123,1 milliárd eurót) a 65 éves és idősebb lakosok – akik a népességnek csupán egyötödét képviselték – gyógykezelésére fordították. Az egy főre jutó gyógyítási költségek ugyanis a magasabb életkorokban gyorsuló arányban nőnek: az átlagos 3 100 euróval szemben a 65–84 éves korosztályban 6 520 eurónak, a 85 éves vagy idősebbek körében pedig 14 840 eurónak feleltek meg.

Emiatt általános az a vélekedés, hogy az idősödő korösszetétel folytán az egészségügyben „költségrobbanás” fog bekövetkezni. A demográfiai adatok első látásra alátámasztják ezt a félelmet, hiszen a 65 éves és idősebb korúak száma 2002 és 2008 között 16,7 százalékkal emelkedett, szemben az össznépesség számának enyhe (0,4 százalékos) csökkenésével. Ugyanebben az időszakban az egészségügyi ráfordítások összesen 16,7, a 65 éves és idősebb korosztályban viszont 30,1 százalékkal nőttek.

A várható költségalakulás modellezésénél további tényezőként azt is figyelembe kell venni, hogy a betegellátás költségei az emberi élet utolsó szakaszában aránytalanul megnövekednek. A kórházi adatforrások lehetőséget nyújtanak a halált megelőző utolsó kezelések gyakoriságának, időtartamának és költségeinek elemzésére. Több érv is szól amellett, hogy a vizsgálat során a kórházakra koncentráljanak. Az egészségügyi ellátás költségeinek csaknem negyede (66,7 milliárd euró) itt merül fel. Az éves halálozások közel felére (2008-ban 47 százalék) a kórházakban kerül sor, s a tapasztalatok szerint a halált megelőzően igen intenzív a kórházi ellátás igénybevétele, valamint a kórházi szektorról állnak rendelkezésre legrészletesebben az elemzéshez szükséges adatok.

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