

Jerzy Neymannak (1894–1981)

*Jerzy Neyman Oroszországban született, Lengyelországban és
Angliában dolgozott, mielőtt 1938-ban az USA-ba érkezett.
Korunk nagy statisztikusainak egyike volt.*

Előszó

*Hogy milyen dalt énekeltek a szirének, s hogy mi volt
Akhilleusz neve, amikor asszonyok között bújkált - rejtelmes
kérdések, sejtésünk mégis van róluk.*

SIR THOMAS BROWNE (ANGLIA, 1605–1682)

AZ OLVASÓHOZ

Be szeretnénk mutatni néhány érdekes problémát, melyeket statisztikai módszerek segítségével tanulmányoztak, továbbá azt is, hogyan lehet ezeket a módszereket önállóan alkalmazni. Igyekszünk elmagyarázni, melyik módszer mitől működik, s mire jó odafigyelni, ha mások használják őket. Úgy tűnik, hogy a matematikai írásmód sokak előtt kifejezetten eltakarja a lényegét – ez a könyv ezért szavakra, ábrákra és táblázatokra épül; elvéve fordulnak csak elő benne x -ek és y -ok. Amikor szakkönyvet olvas, még a hivatásos matematikus is gyakran átsiklik a képleteken. Voltaképpen arra vágyik, hogy legyen mellette egy megértő barát, aki elmagyarázza, milyen elgondolás vagy milyen elképzelt tevékenység áll az egyenletek mögött. Arra törekszünk, hogy a könyvünk olvasóinak ilyen barátai legyünk.

MI A STATISZTIKA?

A statisztika annak mestersége, hogyan lehet bizonyos rejtelmes kérdésekkel kapcsolatosan számszerű következtetésekre – vagy inkább sejtésekre – jutni.

- Mik egy új orvosi beavatkozás hatásai?
- Mi okozza a szülők és gyermekeik közötti hasonlóságot – és mennyire erős ez a hatás?
- Mitől van a ruletten haszna a kaszinónak?
- Ki nyeri a soron következő választásokat? Mennyivel?
- Mennyi a foglalkoztatottak száma? Mennyi a munkanélkülieké?

Nehéz kérdések – ha van kedve ilyesmiken gondolkodni, abban a statisztikai módszerek sokat segíthetnek. Ezeket a módszereket olyan emberek dolgozták ki az évszázadok során, akik különféle, őket foglalkoztató problémákra keresték a megoldást. Közülük is megismerkedünk majd néhányal.

VÁZLAT

Az I. rész a kísérletek megtervezésével foglalkozik. Ha jó a kísérleti terv, akkor az adatokból megbízható következtetéseket lehet levonni. Néhány rosszul tervezett kísérletet is görcső alá veszünk, hogy kiderüljön, hol lehetnek buktatók, és hogy megmutassuk, milyen kérdéseket érdemes feltenni, mikor az ember vizsgálatokról olvas. Az egész könyvből talán éppen ez, a vizsgálatok tervezése a legfontosabb – azért is kezdjük ezzel. Esetleg úgy látszik majd, hogy minden egyszerű ebben a részben – de a látszat sokszor csal: az I. részben sok a mélység.

Egy vizsgálat során általában olyan sok szám keletkezik, hogy feltétlenül összeítani kell őket. A II. részben a leíró statisztikával – az adatok összesítésének mestersegével – fogunk ismerkedni. Hisztogramokról, átlagról, szórásról és a normálgörbéről esik szó. Ezt a témakört folytatja a III. rész, amelyben az összefüggések elemzéséről fogunk beszélni: például, hogy miképpen függ a jövedelem az iskolázottságtól. A legfontosabb címszavak: korreláció és regresszió.

A statisztikai gondolatmenetek jelentős része a IV. részben tárgyalt valószínűségszámításon alapul; közöttük az V. részben kifejtésre kerülő sorsolós modellek teremtik meg a kapcsolatot. Érmék, dobókockák, rulettkerekek a fő példák. Megismerkedünk a várható értékkel és a standard hibával; elméleti hisztogramokat készítünk, és tárgyaljuk a normális görbéhez való konvergenciát.

A statisztikai következtetés lesz a VI-VIII. rész témája: hogyan lehet minták alapján érvényesen általánosítani. A VI. rész a becslésekről szól. Hogyan jelzi például előre a választási eredményt a Gallup-felmérés? Mitől jobb az egyik mintavételi módszer, mint a másik? A VII. rész sorsolós modelleket használ a mérési hibák elemzéséhez és a genetika elméletének kialakításához. A VIII. részben a szignifikancia-próbákkal ismerkedünk – ezek segítségével lehet eldönteni, összeegyeztethető-e mintánk a populációra vonatkozó feltevéseinkkel. A VI-VIII. részből ki fog derülni, hogy a statisztikai következtetés sorsolós modelleken nyugszik: ha rossz a modell, az eredményül kapott következtetés nagyon ingatag lehet.

A szakmabeliek számára mostanában a következtetések elmélete a statisztika legérdekesebb szakterülete. Akik viszont nem statisztikusok, azok a leíró statisztikát többnyire jobban tudják használni, és könnyebben is értik. Emiatt a leíró statisztikát a következtetések előtt tárgyaljuk. Mondanivalónk pusztá váza az 1-től 6., 13., 16-tól 21., 23. és 26. fejezetben található meg. Az olvasó, ha ezeket elolvasta, utána kedvére tallózhat. Folytatásként mi talán a 8., 10., 27. és 29. fejezeteket ajánlanánk.

Feladatok

Többé-kevésbé minden fejezet minden szakaszához csatlakozik egy feladatsor (megoldások a könyv végén). Ha az olvasó sorra, ahogy következnek, megoldja őket, és utána ellenőrzi a megoldást, ezzel némi gyakorlatra tehet szert az újonnan elsajátított készségekben – és megtudja, milyen mértékben sikerült elsajátítania őket. Minden fejezet (az 1. és a 7. kivételével) ismétlő feladatokkal zárul. Ezeknek a megoldása nem szerepel a könyvben.

A feladatok megoldása során esetleg majd kísértést érez, hogy visszalapozzon, hátha feltűnik valahol az idevágó képlet. A könyv ilyenfajta visszafeléolvasása igen frusztráló lehet. Az ismétlődő feladatokhoz egy-egy képletnél sokkal több kell. Hozzávetőleges becslésekre és kvalitatív ítéletekre egyaránt szükség lehet. Más szóval: intuitíve is érteni kell, hogy miről van szó. Ahhoz, hogy ez a megértés kialakuljon, érdemes a könyvet az elejétől a vége felé haladva olvasni.

Miért van ebben a könyvben ennyire sok olyan feladat, amiket nem lehet a megfelelő képletbe helyettesítéssel megoldani? Ennek főként az az oka, hogy a valódi világban sem sok probléma oldható meg ezzel a módszerrel. Viszont sokszor származik zűrzavar abból, amikor emberek úgy alkalmaznak statisztikai formulákat, hogy valójában fogalmuk sincsen arról, amit csinálnak. Ebben a könyvben egy másik megközelítést javasolunk: gondolkozzunk.

MI VÁLTOZOTT A HARMADIK KIADÁSBAN?

A sok könyvek írásának nincs vége.

PRÉDIKÁTOR KÖNYVE

Többé-kevésbé ugyanazok a témák, és ugyanabban a sorrendben követik egymást most is, mint a második kiadásban. Az idő azonban a számok fölött sem múlik el nyomtalanul. (Havi 50 dollárért például a negyvenes években nagyszerű lakást lehetett bérelni.) Ezért az adatok egy részét a 90-es éveknek megfelelően korszerűsítettük. A jövedelmi és iskolázottsági statisztikák az amerikai Rendszeres Népeségfelmérés (Current Population Survey) 1993. márciusi adatfelvételéből valók; a harmadik HANES-vizsgálat (*National Health and Nutrition Examination Survey*, Országos Egészség- és Táplálkozásfelmérés) 1994-es felvételének előzetes eredményei is többször szerepelnek.

A korábbi kiadásokhoz hasonlóan továbbra is sok számítógépes grafikát használunk az adatok bemutatásához. Viszont a vázlatok mind szabadkézi rajzok; ezzel szerettük volna az olvasót arra ösztönözni, hogy maga is készítsen hasonlókat – nem szerettük volna, hogy a túlzott precizitás elvegye ettől a kedvét. Újra gyönyörködhetünk Dana Fradon (*The New Yorker*) rajzaiban – van közöttük néhány új is.

Néhány döntő ponton új kifejtő részek is szerepelnek; új szakasz foglalkozik például az összemosódás, illetve a kísérleti kontrasztokra vonatkozó próbák kérdésével. A közel 200 új feladatban számos friss vizsgálatot szerepeltetünk. Egy-egy „Nagy ismétlődő feladatsor” zárja a 6., 15., 23. és 29. fejezetet; ezek a teljes addigi anyagrészt átfogják. A bennük szereplő feladatok megoldásában nem igazít el, hogy mi a környező fejezetek témája. Ezekhez a fejezetekhez „Összefoglalás és áttekintés” rész is tartozik, amely összefoglalja az előző részt, és átvezet a rákövetkezőhöz.

KÖSZÖNETNYILVÁNÍTÁSOK A HARMADIK KIADÁSHOZ

Dana Fradon rajzolta a karikatúrákat. Dale Johnson és Laura Southworth az ábrákat. A szöveget az Integre szedte, TEX-ben. A számítógépes grafikát PostScriptbe kódolva integráltuk a végső nyomtatási állományokba; a kódolás jelentős részben Charles Everett (A.A.E.C., Pennington, New Jersey), továbbá Richard és Adele Cutler (Utah State University) munkája.

Segítőkész megjegyzéseket sok helyről kaptunk. A harmadik kiadás tekintetében külön köszönetet mondunk Mike Andersonnak (Berkeley), Dick Berknek (UCLA), Jeff Fehminek (Berkeley), David Kaye-nek (Arizona State University), Steve Kleinnek (RAND), Russ Lyons-nak (Indiana), Mike Ostlandnek (Berkeley), Erol Pekoznak (UCLA), Diana Pettitnek (Kaiser Permanente), Juliet Shaffernek (ETS), Bill Simpsonnak (Winnipeg), Terry Speednek (Berkeley), Philip Starknak (Berkeley) és Allan Stewart-Oatennek (Santa Barbara). Végül, itt mondunk köszönetet azoknak, akik olvasták az első két kiadást (vagy a harmadik kiadás számos előkészítő verzióját), s azoknak is, akik ezekből tanultak.

KÖSZÖNETNYILVÁNÍTÁSOK A KORÁBBI KIADÁSOKHOZ

A könyv megírását a Ford Foundations (1973-1974) és a Regents of the University of California (1974-75) támogatta. Nehéz időkben segítséget és bátorítást kaptunk Earl Cheittől és Sanford Elbergtől.

Segítőkész hozzászólásokat kaptunk a következőktől: Frank Anscombe (Yale), Diccon Bancroft (Yale), Leo Breiman (Berkeley), John Cairns (Oxford), Merrill Carlsmith (Stanford), Persi Diaconis (Cornell), Fred Katz (Berkeley), David Lane (Modena), Richard Light (Harvard), Peter McCullagh (Chicago), Ludolf Meester (Delft), Gerald Mendelsohn (Berkeley), Lincoln Moses (Stanford), Bill Navidi (Colorado School of Mines), James Robins (Harvard), Thomas Rothenberg (Berkeley), Bruce Rotschild (UCLA), Bernard Saffran (Swarthmore), Shanna Swan (California State Department of Health Services), Amos Tversky (Stanford), Geoff Watson (Princeton), és Hans Zeisel (Chicago). A második kiadás készítésében részt vett Ani Adhikari (Stanford).

Külön köszönet illeti szerkesztőnket, Donald Lammet azért, hogy az egyre alakuló kéziratból valahogyan mégis könyv lett.