

# Biostatisztikai alapismeretek gyakorló orvosoknak (3.)

HAJTMAN BÉLA DR.

*Klinikai Biostatisztikai Társaság, Budapest*

E-posta: hajtman@stu.hu

A biostatisztika egyik legrégebb és leghatékonyabb módszere a varianciaanalízis, amelynek magyar elnevezése (szórás-elemzés) mindmáig nem tudott meghonosodni. Csaknem 90 éve dolgozták ki a módszert, de mind a mai napig gyakran és eredményesen alkalmazzák különféle vizsgálati és kísérleti adatok elemzésére.

A módszer lényege a következő: Az adatok közti eltérések mérőszámát, az úgynevezett varianciát bontják fel összetevőkre, és ennek segítségével igyekeznek meghatározni az eltérések okát.

Az adatok közti eltérések (különbségek) mérőszáma valójában a szórás. Ezt a fogalmat mindenki ismeri, és sokan tisztában vannak számítási és alkalmazási módjával is. A variancia a szórás négyzete. Tulajdonképpen formai okai vannak annak, hogy ezt és nem magát a szórást elemezzük. (A „négyzetgyök alatti” nem végezhető el semmiféle részekre bontás.)

Visszatérve a varianciaanalízisre, a legegyszerűbb esetet a következőképpen képzelhetjük el: Vannak csoportjaink (pl. betegcsoportok), akik különböző kezelést kapnak ugyanarra a betegségre. A statisztika szempontjából mindegy, hogy milyen kezelésekről van szó, de gondoljuk egyszerűség kedvéért azt, hogy mindegyik csoport más-más gyógyszert szed. Valamilyen alkalmasan megválasztott paramétert mérünk meg mindegyik csoport esetében, amely a „gyógyulás” mértékét mutatja; pl. a testhőmérsékletet, a pulzusszámot, esetleg egy vérvizsgálat során kapott laboratóriumi eredményt. Ezek lesznek azok az adatok, amelyeknek elemzését el kívánjuk végezni.

Az adatok ugyanazon csoporton belül sem egyformák: a betegség sem egyformán súlyos az egyes betegeknek és a gyógyulás mértéke is különbözik, bár a betegek ugyanazt a kezelést kapják – és egyáltalán: a különböző személyeken mért adatok (sőt az azonos személyen különböző időben mért adatok) mindig különböznek valamennyire. Ezeket a különbségeket szokás a statisztikában véletlen ingadozásnak nevezni. Véletlen, mert bár biztosan van valamilyen oka (sőt nem is egy oka), ezeket az okokat nem ismerjük és befolyásolni sem tudjuk.

Az adatok közti különbözőséget „méri” – vagy inkább: egyetlen számba sűrítve jellemzi – a variancia (és természetesen

annak négyzetgyöke, a szórás). Az összes adatból számolva kapjuk a „teljes” varianciát, és ezt bontjuk fel azután összetevőkre. Az említett egyszerű példa esetében két összetevőre: a csoporton belüli és a csoportok közti részre. A csoporton belül csak a „véletlen” működik: a csoport minden tagja ugyanazt a kezelést kapja. A csoportok közt azonban lehet különbség: mindegyik csoport más kezelést kap, és ha a kezelés hatásos, ez befolyásolja a gyógyulás mértékét jellemző, általunk mért paraméter értékét.

Végül a két komponenst – a két varianciát – összehasonlítjuk. Ha a csoportok közt nagyobb az eltérés, mint azt a véletlen ingadozás indokolná (ez utóbbiról a csoporton belüli variancia ad felvilágosítást), akkor bebizonyosodott (?), hogy a kezelések valamelyikének – vagy talán valamennyinek – hatása van a gyógyulásra.

De mennyivel kell ehhez a csoportok közti varianciának nagyobbak lennie? Hiszen az, hogy „kicsit nagyobb”, ismét lehet a véletlen műve! Ennek eldöntésére – az eljárás befejezésekképp – egy statisztikai próbát végzünk, amely megmutatja, hogy okozhat-e a két variancia közt ekkora eltérést a véletlen, vagy a kapott eltérések valamilyen oka van; ez az ok pedig mi lenne más, mint a kezelések különbözősége, hiszen ez az egyetlen különbség a csoportok között.

A „statisztikai próba” szintén közismert fogalom, és azt is tudjuk, hogy az eredményben egy p-vel jelölt valószínűség áll. Általában a  $p < 5\%$  a „kritikus” érték. Ez azt jelenti: kicsi (5%-nál kevesebb) annak valószínűsége, hogy a varianciák különbözőségét csupán a véletlen okozza; valószínű, hogy annak más oka van. Ezért tettünk korábban kérdőjelet a „bebizonyosodott” szó után: nem biztos, csak 95%-nál valószínűbb, hogy a kezelések közt valamelyik (vagy valamennyi) hatásos. (Hogy melyik és mennyire, ezt újabb statisztikai elemzésekkel kell majd ezek után kideríteni.)

A varianciaanalízisnek csak a legegyszerűbb esetét említettük, szemléltető példaként. Nemcsak egy-, hanem többféle kezelésben is részesülhetnek a betegek (pl. gyógyszerek mellett – vagy nélkülük – műtét, kórházi vagy ambuláns kezelés stb.), és ezek hatását egyenként is megismerhetjük a variancia most már több komponensre (vígyázat: ezek a komponensek nem „részek”,

nem összeadandók!) bontásával. Mérhetünk nem egy, hanem több paramétert, és vizsgálhatjuk viselkedésüket együttesen is, nemcsak külön-külön. Mindez – és még sok más finomítás, általánosítás – a varianciaanalízishez tartozik, és a modern számítógépek igénybevételével pillanatok alatt elvégezhető. Csak éppen azt kell tudni, hogy esetünkben a gép által felkínált modellek közül melyik a megfelelő, és mit jelentenek a kapott – sokszor tömegesen ránk zúdított – eredmények.

A fenti bekezdésekben több olyan statisztikai fogalom is szerepelt (pl. a szórás), amelyet nem magyaráztunk meg, vagy pedig csak hevenyészett, nehezen érthető magyarázatát adtuk (pl. a statisztikai próbák). Nem meglepő, ha az olvasóban kérdések merülnek föl, vagy ha értetlenül áll ennek az – egyébként a megértés könnyítését célzó – kis eszmefuttatásnak láttán.

Kérjük, hogy ha lehet, ne vágják dühösen sarokba az egészet, hanem tegyék fel kérdéseiket, és mi igyekszünk azokra a következő lapszámokban képességeink szerint válaszolni. Biztos, hogy nem ez a kérdezgetősi, kiegészítősi a legjobb módszer. De jobbnak tűnt, mint az a szokásos eljárás, hogy az egyetemen tanultakat már elfelejtett, de érdeklődő orvosok számára a biostatistika rövid összefoglalóját adjuk, módszeresen, az alapfogalmakkal kezdve. Úgy gondoltuk, hogy mire pl. a varianciaanalízisig eljutnánk (nem jutnánk!), olvasó és szerző egyaránt rég ráunna az egészre, és nem lenne belőle más, mint eggyel több hamvába holt próbálkozás.

Bizakodva bocsátjuk az olvasó elé ezt a próbálkozást, és várjuk a visszajelzéseket! Azokból majd kiderül, hogy érdemes-e folytatni, és hogyan.

*Bősze Péter*

## **GONDOLATOK A MAGYAROS ÍRÁSMÓDRÓL**

A magyaros írásmód az idegen szavak magyar helyesírás szerinti írásformája: az idegen szavak szokásos kiejtését a magyar hangjelölés szabályai szerint tükröztetjük (*mikroszkóp, izotóp, hisztamin, digitális*). Ennek értelmében az alapszó magánhangzójának kiejtés szerinti megnyúlását a magyarosan írt idegen szóban – ellentétben az idegesen írttal – jelöljük (*incurabilis/inkurábilis, incretio/inkrécio*). Ám kivételek előfordulnak: egy-egy magyarosan írt alakban, a hagyomány szerint, a magyar kiejtéstől eltérő elemek is megmaradhatnak (*millió, futball*). A magyarosítás az idegen szavak magyar formájává öltöztetése. A magyarosan írt idegen szavak és a magyar szavak írásmódjának, toldalékolásának, képzésének és az elő-, utótagok csatlakoztatásának helyesírási szabályai megegyeznek.

Magyarosan csak a latin betűs idegen szavakat írhatjuk, a nem latin betűs szavak a magyar nyelvben nem ismert és nem is értelmezhető betűkből állnak (görög, japán, kínai, cirill stb. betűk). Ezek nem építhetők be nyelvünkbe, ezért mindegyik szavát – a közsavakat és a tulajdonneveket is – átírjuk a magyar ábécé betűivel. Néhány idegen, nem latin betűs nyelvből – elsősorban a görög ábécéből ( $\alpha$ ,  $\beta$ ,  $\delta$  stb.) – származó betűt nyelvünk mégis megtartott; ezeket az orvosi nyelvben többnyire kiegészítő betűként (l. tartozékbetűk) alkalmazzuk. Hasonlóan megmaradtak a római számok is.