

# A következtető statisztikáról részletesebben

Ferenci Tamás  
tamas.ferenci@medstat.hu

2018. március 29.

# A sokaság és a minta fogalmáról

- „Mennyi a jelenlevők átlagos testtömege?” → véges sokaság ( $N = 30$ ), név szerint felsorolhatóak az elemei
- De: „Egy új vérnyomáscsökkentő gyógyszer-jelölt valóban csökkenti a vérnyomást?”
  - Mi itt a sokaság?
  - Ennek nem lehet név szerint felsorolni az elemeit, ez egy absztrakt konstruktum
  - Szokás fiktív sokaságnak hívni (vagy, szintén használt találó névvel, végtelen sokaságnak)
- Általában nem tudjuk az egész sokaságot megfigyelni → *mintavételes helyzet*
- Amit meg tudunk figyelni: *minta*

# A problémafelvetés

- Nagyon sok esetben tehát technikai okokból, vagy elvileg is lehetetlen a teljes sokaság megfigyelése
- Csak egy részét, a mintát ismerjük
- És itt jön a kulcsprobléma: *mi mégis a sokaságról akarunk nyilatkozni!*
- Lehet egyáltalán? Hogyan?
- Biztosat már nem tudunk mondani. . . de valószínűségi állítást igen!

# Mintavételi ingadozás

- Ha csak a sokaság egy részét (a mintát) ismerjük, akkor minden belőle számolt jellemző két dologtól fog függeni:
  - 1 a jellemző sokaságbeli értékétől (a gyógyszer *valódi* hatásától)
  - 2 attól, hogy konkrétan hogyan választottuk ki a mintát (kik kerültek be a gyógyszerkísérletbe)
- Mi értelemszerűen az elsőre vagyunk kíváncsiak
- ... csakhogy a – kikerülhetetlen – második („pont milyen mintát vettünk”) azt fogja okozni, hogy minden eredményünk mintáról-mintára változni fog
- A *szerencse*: ez az ún. mintavételi ingadozás követ valószínűségszámítási törvényeket, így valószínűségi állításokat meg tudunk fogalmazni!
- Hibázhatunk, de ennek természetéről tudunk nyilatkozni

# Mintavételi hiba

- Figyelem, ennél a hibázásnál nem arról van szó, hogy „rosszul” veszünk mintát: például a legtökéletesebben véletlenszerű mintavételnél is előfordulhat, hogy egy 100 fős sokaságból úgy becsüljük az átlagos testtömeget, hogy pont a 20 legkönnyebbet választjuk ki
- De: ennek a valószínűsége extrém kicsi! (egész pontosan  $1/\binom{20}{100} \approx 2 \cdot 10^{-19}\%$ )
- Az érzékeltetés kedvéért: ha ekkora valószínűséggel dobunk fejet egy cinkelt pénzérmével, és másodpercenként 1-szer fel tudjuk dobni, akkor az első fej várhatóan 17 ezer milliárd év múlva fog kijönni. . .
- (A trükkös rész: egy konkrét mintánál soha nem tudhatjuk, hogy az véletlenül nem pont az az 1-e, ami átlag 17 ezer milliárd évente egyszer fordul elő!)
- Az egész csak ilyen képzeletbeli ismételt mintavételi értelemben nyer értelmet; ezt a megközelítést szokás frekvencionista statisztikai iskolának nevezni
- Statisztikai apparátus: becslélmélet („konfidenciaintervallumok”) és hipotézisvizsgálat („p-értékek”)

- Becslőfüggvény (pl. a sokasági átlagot „tippeljük” a minta átlagával)
- Persze tévedhetünk, de reméljük, hogy vannak jó tulajdonságai a becslőfüggvényünkben (torzítatlanság, hatásosság)
- A fentiekkel egyetlen számot, „a” legjobb becslést adjuk vissza eredményként
- Nem adunk számot arról, hogy ebben mekkora a bizonytalanság. . .
- . . . pedig erről is tudunk nyilatkozni! („Kalkulálható bizonytalanság”)
- Tipikus szemléltetés: *konfidenciaintervallum* (CI): mi az a tartomány, amire igaz, hogy ha sokszor megismételnék a mintavételt, és mindegyik mintából megszerkesztenénk a CI-t, akkor ezen CI-k várhatóan 95%-a tartalmazná az igazi (sokasági) értéket (95% megbízhatóság melletti CI)

## A konfidenciaintervallumról bővebben

- „Fordított logika” (ezt később is látni fogjuk): nem azt mondjuk, hogy ha ez a minta, akkor hol lehet a valódi érték, hanem, hogy ha ez a valódi érték, akkor hihető-e még, hogy ez jöjjön ki mintaként
- Úgy is szokták mondani, hogy a CI megadja, hogy – ilyen értelemben – mi kompatibilis a mintával
- (Finom különbségnek tűnik, pedig hatalmas jelentősége van, lásd később, a hipotézisvizsgálatnál)
- A „hihetőségre” természetesen határt kell húzni – ez lesz az a bizonyos megbízhatósági szint
- Minden CI szerkesztő eljárásnak vannak előfeltevései (a több ilyen használók általánosságban szűkebb intervallumot adnak, de ha nem fennálló előfeltevésre építünk, akkor invalid lesz a CI, tehát nem 95% lesz a lefedése)
- Nagyobb megbízhatóság  $\leftrightarrow$  semmitmondóbb intervallum

# A hipotézisvizsgálat alapfogalmai

- Feladat: sokaságra vonatkozó állítás eldöntése minta alapján
- Lényegében az intervallumbecslés ikertestvére, de hatalmas gyakorlati jelentősége miatt külön eszköztára van
- Alapeszköze a *statisztikai próba* (vagy *teszt*), mely a mintaelemek alapján kiszámol egy ún. *tesztstatisztikát* (próbafüggényt)
- Vizsgált állításaink: nullhipotézis – ellenhipotézis
- Egy tipikus példa:

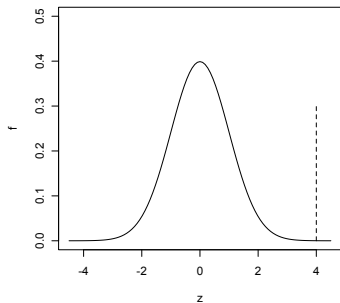
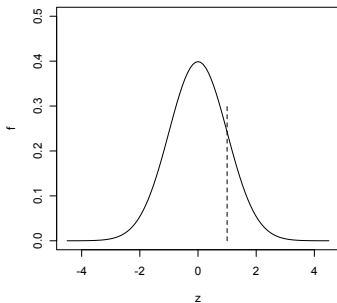
$$H_0 : HR = 1$$

$$H_1 : HR \neq 1$$



# Döntés a hipotézisvizsgálatban

- Hihető-e, hogy az empirikus (adott, konkrét mintából kapott) érték ebből az eloszlásból származik?
- Biztos döntés nincs! De: mennyire hihetők ezek?



# Döntés a hipotézisvizsgálatban

- Megint csak, valahol „határt kell húznunk” → szó szerint is!
- Azt mondjuk, hogy a nagyon kis valószínűségű területekre esést már nem hisszük el
- Pedig az *nem* lehetetlen, sőt: az is tudható, hogy az oda esés (azaz a fenti logikával történő hibázás) valószínűsége épp ez a nagyon kis valószínűség
- Tipikus, hogy a felső és alsó szélén is 2,5-2,5 % valószínűségű területet jelölünk ki ( $\alpha = 5\%$ , ez a szignifikanciaszint), határai: a  $c_a$  alsó és a  $c_f$  felső kritikus értékek (példában:  $\pm 1,96$ )

- Vagy: Mennyi lenne az a szignifikanciaszint, ami mellett a mintából kapott (empirikus) tesztstatisztika-érték épp az elfogadás és az elutasítás határára kerülne?
- (Ez nem más, mint az empirikus értéktől extrémebb helyeken vett integrálja a mintavételi eloszlásnak)
- A neve:  $p$ -érték
- „Az olvasó is tud dönteni”: ha a választott szignifikanciaszint nagyobb, mint a  $p$ -érték, akkor elutasítunk, különben elfogadunk
- Ez természetesen ugyanaz, mint a korábban látott definíció: ha fennáll a nullhipotézis (pl. nem hat a gyógyszer), mekkora valószínűséggel kapunk olyat – vagy még extrémebbet – mint amit ténylegesen kaptunk is
- Számos félreértés, lásd mindjárt és filozófiai kérdőjel, lásd később. . .

# Döntés a hipotézisvizsgálatban

- „Fordított logika” (megint csak) a döntéshez: *ha* fennáll a nullhipotézis – nem hat *valójában* a gyógyszer – *akkor* mennyire hihető, hogy a mintában *mégis* hat
- A mintavételi hiba miatt nem nyilvánvaló kérdés: ha a kezelt csoportban 10% halt meg, a kontrollcsoportban pedig 20, attól még lehet, hogy nem hat a gyógyszer – mert a valóságban mindkét csoportban 15% hal meg, de mi pont ilyen mintát vettünk (pusztán a véletlen ingadozás miatt!)
- Ennek a valószínűségét méri a  $p$ -érték (precízen: hogy ilyen vagy ennél is extrémebb mintát kapunk, feltéve, hogy fennáll a nullhipotézis)
- Logika: ha ez nagyon kicsi, akkor azt mondjuk, hogy hat a gyógyszer (mert a másik magyarázat, hogy egy nagyon valószínűtlen esemény történt)
- Vigyázat, hipotézisvizsgálatban mindig az elutasítás az erős döntés (nem véletlen, hogy úgy fogalmazunk, hogy „nem tudtuk elvetni” a nullhipotézist, nem úgy, hogy „elfogadtuk”)
- A hiány bizonyítéka nem a bizonyíték hiánya!

- Elvetjük  $H_0$ -t, pedig fennáll (elsőfajú hiba,  $\alpha$ ): pontosan szabályozható valószínűségű
- Elfogadjuk  $H_0$ -t, pedig el lehetne vetni (másodfajú hiba,  $\beta$ ): általánosságban nem ismert, függ a valóságtól
- $1 - \beta$ : próba ereje („mennyire ismeri fel az eltérést, ha tényleg van”)
- Mi két dologgal tudjuk befolyásolni a próba erejét, mindkettőhöz egy-egy tételmondat:
  - 1 Választott próba: „mindig annyi előfeltevésre építő próbát használjunk, amennyit tudunk, se többet se kevesebbet” (több előfeltevésre építő próbák erősebbek ugyan, de ha szükséges előfeltevés nem teljesül, a próba nem lesz valid)
  - 2 Mintanagyság: „kis hatáshoz nagy minta kell, nagy hatáshoz elég a kisebb minta is” → ha feltételezünk egy eltérést, akkor kiszámítható, hogy annak adott valószínűségű – erejű – kimutatásához hány beteget kell bevonni!