

A statisztika alapelvei, hipotézisek

Félév beosztása

- ▶ Hipotézisek, skálatípusok.
- ▶ Eloszlások, szórás.
- ▶ Korrelációs számítás.
- ▶ Normális eloszlás, standard normális eloszlás.
- ▶ Valószínűség, mintavétel, konfidenciatartomány, szignifikanciaszintek, p -értékek.
- ▶ Variancia és átlag összehasonlítása: F -próba, t -próba.
- ▶ Nem parametrikus próbák: khi-négyzet, Wilcoxon, Mann-Whitney, Kruskal-Wallis.

Ha marad rá idő, kedv, kitartás:

- ▶ Lineáris regresszió.
- ▶ Varianciaanalízis (ANOVA).
- ▶ Kevert modellek.
- ▶ Gépi tanulás: klaszterelemzés, döntési fák.

Irodalom

Órák anyaga:

phon.nytud.hu/mady/courses/statistics/2022

Általános, bevezető részekhez:

Reiczigel, J., Harnos, A. & Solymosi, N. (2010): Biostatisztika nem statisztikusoknak. Nagykovácsi: Pars.

Nyelvészeti kérdések, feladatok:

Baayen, R. H. (2008): Analyzing linguistic data: a practical introduction to statistics using R. Cambridge: University Press.

Az R működésének jobb megismeréséhez:

Peter Dalgaard (2008): Introductory statistics with R. New York: Springer.

További modellek, főleg a pszichológiában

Field, Andy, Miles, Jeremy, & Field, Zoë (2012): Discovering statistics using R. London: SAGE.

Statisztikailag tesztelhető állítások

Statisztikailag tesztelhető állítások

- ▶ Ha négy hétig fogyasztó koktélokat eszünk, soványabbak leszünk.
- ▶ A kétnyelvű hatévesek kognitív teljesítménye jobb, mint az egynyelvűeké.
- ▶ Juli néni egy nap alatt több időt tölt beszéléssel, mint Feri bácsi.
- ▶ Az angolok többet olvasnak, mint a franciák.
- ▶ Az éghajlat ma melegebb, mint 100 évvel ezelőtt.

Statisztikailag nem tesztelhető állítások

Statisztikailag nem tesztelhető állítások

- ▶ A keserűcsokoládé finomabb, mint a tejsoki.
- ▶ A patkány a legrondább állat, a pók szorosan követi.
- ▶ Az emberek régen sokkal szebben beszéltek, mint ma.
- ▶ Az Ödön nevű emberek ellenszenvesek.

Miért?

Statisztikailag nem tesztelhető állítások

- ▶ A keserűcsokoládé finomabb, mint a tejsoki.
- ▶ A patkány a legrondább állat, a pók szorosan követi.
- ▶ Az emberek régen sokkal szebben beszéltek, mint ma.
- ▶ Az Ödön nevű emberek ellenszenvesek.

Miért?

Ezek az állítások szubjektívek, nem tesztelhetőek számszerű adatokkal, azaz nem mérhetőek.

Kvantitatív és kvalitatív leírás

Kvantitatív adatok: megszámlálható vagy mérhető egységek.

A kvantitatív adatgyűjtést gyakran megelőzi a kvalitatív adatgyűjtés, ami alapján kiválaszthatóak a releváns, azaz tesztelendő változók.

Kvalitatív adatok: megfigyelések részletes leírása, pl. a csokoládétípusok közötti különbségek leírása, az emberek undora a patkányoktól és a pókoktól, a nők és háziasszonyok szociális helyzetének leírása.

A kvalitatív adatok gyakran kvantifikálhatóak a kérdés átfogalmazásával. Például: a keserűcsokoládé ízének megítélése egy 5-ös skálán. A múltra és egyes személyekre vonatkozó kijelentések nehezen kvantifikálhatóak, sőt, objektívvá sem mindig tehetőek.

Kezdeti megfigyelések

Egy kísérlet gyakran egy sejtésen vagy megfigyelésen alapul.

- ▶ Idén sokkal több a szúnyog a Balatonnál, mint az előző években.
- ▶ A nők több verbális és nonverbális visszajelzést adnak egy társalgásban, mint a férfiak.
- ▶ Az emberek mostanában ritkábban tüszentenek nyilvánosan, mint korábban.

Az elmélet felállítása

A tudományos kísérletekben a meglévő eredmények alapján feltételezhetőek bizonyos magyarázatok, amiket tesztelhetünk. Vegyük a szúnyogokat.

Potenciális magyarázatok:

- ▶ A szúnyogok szaporodási időszakában ideálisak voltak az időjárási viszonyok, ezért sokan életben maradtak.
- ▶ Kevesebb volt a szervezett irtás.
- ▶ Immunissá váltak egy gyakran használt irtószerre.

Kísérleti dizájn

Az elméletek tesztelésére összehasonlítható adatokat kell gyűjtenünk:

- ▶ Szúnyogok száma egy adott területen, ahol a szaporodási időszakban napos, nedves, hideg vagy meleg volt az idő.
- ▶ Irtási akciók időtartama.
- ▶ Az egyes irtószerek használatának mértéke korábbi évekkel összehasonlítva.

A kísérletező által kontrollált változók (napos, hideg terület stb.) neve **független változó**. Egyéb elnevezések: faktor, tényező, magyarázó változó, fix hatás.

A kísérlet során gyűjtött adatok neve (pl. szúnyogok száma a tesztelt területen) **függő változó**, mert függ az adatgyűjtés körülményeitől, azaz a független változóktól.

Háttér

Fallibilizmus (Popper): egy állítás igazolásából nem következik az igazság.

Alaptézisek:

- ▶ Ismételt megfigyelésből nem lehet levezetni, hogy valami törvényszerű. Ha csak fehér hattyút találok, abból nem következik, hogy minden hattyú fehér.
- ▶ Falszifikáció: keress fekete hattyút! Amíg az eredmény negatív (nem láttunk fekete hattyút), addig nem dőlt meg az a hipotézis, hogy minden hattyú fehér.
- ▶ Alapfeltétel: megdönthetőség. A kísérleti módszert úgy kell megválasztani, hogy a hipotézis, amennyiben helytelen, megdönthető legyen.

Követelmények

Operacionalizálás: kérdésfeltevés úgy, hogy empirikusan megfigyelhető adatok alapján megválaszolható legyen → körültekintő kísérlettervezés.

Állítás: a mai emberek igénytelenebbül beszélnek, mint a régiek.
Mérőszámok? Összehasonlíthatóság?

Reprodukálhatóság: a kísérleti dizájn és a felhasznált módszerek alapján az eredményeknek megismételhetőeknek kell lenniük.
Előfeltétel: módszerek részletes leírása.

Objektivitás: függetlenség a kísérletvezetőtől és a kísérlet körülményeitől.

Az Ödönöket reggel 6-ra hívjuk be tesztelni, a más nevéük választhatnak időpontot 😊

Alapfogalmak

Populáció vagy sokaság: a vizsgálandó elemek összessége, véges vagy végtelen. A teljes populáció vizsgálata többnyire lehetetlen.

Reprezentatív mintavétel: fontos tulajdonságainak arányában megfelel a populáció megfelelő tulajdonságainak. Véletlenszerű: minden kiválasztott elem egyforma valószínűséggel kerülhet bele a mintavételbe, pl. Magyarország összes 1. éves egyetemistája.

Irányítottan reprezentatív mintavétel: populáció eloszlásának leképezése, csoportokon belül véletlenszerű kiválasztás. Pl. egyetemisták nem, szakirány, kor szerint súlyozva.

Hipotézisek

Hipotézis: feltételezés. Itt: előzetes megfigyelésen alapuló válasz tudományos kérdésfeltevésekre.

Kísérleti hipotézis: a változók viszonyára vonatkozó állítás, pl. a nők csoportjában több verbális visszajelzés hangzik el, mint a férfiakéban.

Statisztikai vagy sztochasztikus hipotézis: egy adott esemény bizonyos körülmények között bizonyos valószínűséggel bekövetkezik. Ha egyik nap süt a nap, 70% a valószínűsége, hogy következő nap is sütni fog.

Hipotézisállítás

Cél: szeretnénk bizonyítani, hogy az általunk forgalomba hozott Szerecsen kávé hosszabb időn keresztül frissen tartja a fogyasztóit, mint a Hagyományos kávé.

Kiindulási vagy nullhipotézis (H_0): A Szerecsen és a Hagyományos kávé fogyasztása után a tesztalanyok **azonos** ideig maradnak frissek.

Ellenhipotézis vagy alternatív hipotézis (H_1): a Szerecsen kávéét fogyasztók később érzik magukat újra fáradtnak.

Miért? „Keress fekete hattyút!” Amíg nem találsz, addig a korábbi hipotézis (minden hattyú fehér) marad érvényben.

Vagyis: azt próbáljuk bizonyítani statisztikai modellekkel, hogy NINCS különbség két vagy több mintára vonatkozó mérések között. Ha a modell azt mutatja, hogy nagyon valószínűtlen, hogy a minták egyazon populációhoz tartozzanak, akkor elvetjük a nullhipotézist, és fenntartjuk az alternatív hipotézist.

A csoport közös kísérlete

Egy külföldi kolléga megfigyelése szerint a magyarban van egy olyan szabály, hogy a *viszontlátásra* köszönésre *viszlát* a válasz. Ha viszont valaki a *viszlát* kifejezést használja, a beszélgetőpartner válasza *viszontlátásra* lesz.

Hogyan lehet ezt tesztelni?

- ▶ Kikből áll a populáció?
- ▶ Hogyan gyűjtünk adatokat?
- ▶ Hogyan érhetjük el, hogy kiegyensúlyozott legyen a mintavétel?
- ▶ Mire van szükség a reprezentatív mintavételhez?
- ▶ Mik az adatgyűjtés buktatói?

Adatok formátuma lehet xls, xlsx, csv vagy txt. Előállítható MS Excelben, LibreOffice Calcban vagy szöveges fájlként, tabulátorokkal elválasztva. Ezt tartalmazza: (1) adatfelvevő, (2) milyen köszönést használt, (3) milyen választ kapott.

Így nézzen ki a táblázat, ASCII-karakterekkel, ékezetek nélkül:

| adatfelvevo | koszones | valasz |
|-------------|-----------------|-----------------|
| MK | vizsontlatastra | vizsontlatastra |
| MK | vizslat | vizsontlatastra |
| MK | vizslat | hello |
| MK | vizsontlatastra | vizslat |

Házi feladat február 14., hétfő estig: adatok elküldése emailben: mady@nytud.hu.