

# Egy régi probléma újra előtérben: a nullhipotézis szignifikancia-teszt téves gyakorlata

Bárdits Anna, Németh Renáta, Terplán Győző

[barditsanna@gmail.com](mailto:barditsanna@gmail.com)

[nemethr@tatk.elte.hu](mailto:nemethr@tatk.elte.hu)

[terplangyozo@caesar.elte.hu](mailto:terplangyozo@caesar.elte.hu)

ELTE Társadalomtudományi Kar

Mottó:

*„A Basic and Applied Social Psychology (BASP) 2014-es szerkesztőségi állásfoglalása hangsúlyozta, hogy a nullhipotézis szignifikancia teszt eljárása (NHSzTE) érvénytelen, ezért szerzőinktől a továbbiakban nem követeltük meg annak használatát (Trafimow, 2014). Egyúttal türelmi időt jelöltünk ki szerzőinknek, mely idő alatt szerkesztőségünk nem tiltotta az NHSzTE-t. Szerkesztőségünk ezennel bejelenti, hogy a türelmi idő lejárt. Mától a BASP betiltja az NHSzTE használatát.”*

BASP, szerkesztői bevezető, 2015. február

# A szignifikancia-teszt problémája

A gyakorlati alkalmazás hibái (mások: inherens probléma)

## **tisztán tudományos problémák**

### *matematikai hibák*

- A tesztek erejének figyelmen kívül hagyása
- a szignifikanciateszt használata kifejezetten nagy mintáknál
- a p-érték azonosítása a nullhipotézis valószínűségével
- ragaszkodás az ötszázalékos küszöbhez
- a teszt feltételeinek figyelmen kívül hagyása

### *interpretációs hibák*

- a szubsztantív szakmai fontosság összetévesztése a statisztikai szignifikanciával
- a p-érték, mint egyetlen mutató azonosítása a hipotézissel kapcsolatos bizonyítékkal
- a hatásnagyság vizsgálatának elmaradása, dichotóm döntés  $p < 5\%$  alapján

# A szignifikancia-teszt problémája (folyt.)

## **tudományszociológiai háttérű problémák**

- tudományos konklúziók vagy közpolitikai/üzleti döntések nem alapozhatók  $p < 5\%$  kérdésre. Teljes átláthatóság és részletes riport kellene.
- elemzési torzítás (szignifikanciavadászat)
- publikációs torzítás

# A problémák okai

- “Tanítjuk, mert használjuk, s használjuk, mert tanítjuk.”
- Szoftverek hibája
- Statisztikailag nem eléggé képzett személyek végzenek elemzéseket.  
(Prognózis: adatgazdag világban még kifejezettebb lesz)
- Pszichológiai ok: leegyszerűsítés vágya, A „szürkét” hajlamosak vagyunk feketébe vagy fehérbe transzformálni.
- Tudományszociológiai okok:
  - a folyóiratok a „pozitív” eredményeket preferálják
  - az egyéni kutatók érvényesülési vágya
  - puha tudományok kisebbségi érzése

# A probléma története

- 1988 International Committee of Medical Journal Editors (ICMJE) “önmagában az NHSZT nem elégséges, p-érték különösen nem”
- 1980-as évek vége *Epidemiology* Rothman: konfidencia-intervallumok propagálása
- 2014 Nature-cikk a p-értékekről
- 2014 Basic and Applied Social Psychology (BASP) állásfoglalás nem követelik meg a szignifikancia tesztek alkalmazást
- 2015 BASP főszerkesztői állásfoglalás mindenfajta következtetési statisztika használatának tiltása
- 2015 Royal Statistical Society körkérdése a tiltásról
- 2016 American Statistical Association állásfoglalás a p-értékekről

# Elterjedtség és mérése

- Tévhitek elterjedtsége a kutatók között
- Rossz gyakorlatok elterjedtsége folyóiratokban
  - Orvostudomány
  - Pszichológia
  - Közgazdaságtan
- Szociológiai Szemle

# Ajánlások

szerkesztőségi instrukciók, szakírók: a p-érték helyett...

- leíró statisztikák, pontbecslés és megbízhatósági intervallum
- hatásnagyság-mutatók
- bayes-i módszerek

American Statistical Association friss állásfoglalása (*The American Statistician*, 2016. június)

- szokatlan gesztus (Még sosem foglaltak állás statisztikai gyakorlat kérdésében!)
- célközönség: nem-statisztikus kutatók, alkalmazók és szakírók
- praktikus használati tanácsok, rámutatás téves gyakorlatokra
- egyszerű megközelítés (pl. statisztikai erőre nem hivatkozik)



# Konklúzió

A probléma nem csak a kutatási módszertant érinti, hanem

- a teljes tudományos intézményrendszer (kutatásfinanszírozás, publikálás, karrierlehetőségek, oktatás)
- és annak eredményeit, hasznosulását (“How the standard error costs us jobs, justice and lives”)

A statisztikai következtetés jóval összetettebb,

- mind statisztikailag (megfelelő design, erőszámítás stb),
- mind a szubsztantív szakmai háttérrel (a vizsgált jelenség ismerete, megértése) tekintve,

mint a  $p < 5\%$  dichotóm döntés.

Magyar vonatkozások?