

DESPRE METODĂ Cum să interpretăm valoarea calculată a lui p ?

Prep. Drd. Robert Balazsi

Universitatea Babeş-Bolyai, Cluj-Napoca
Facultatea de Psihologie și Științe ale Educației
Gh. Bilașcu, Nr. 37
Robertbalazsi@psychology.ro

Demersul de Testare a Ipotezei Nule (DTIN) este cea mai frecvent utilizată metodă de analiză a datelor experimentale (Nickerson, 2000). Analiza atentă a rezultatelor publicate (comportament rar întâlnit în rândul consumatorilor de literatură de specialitate), respectiv a interpretărilor oferite de autori acestor rezultate, deseori relevă greșeli în modul de conceptualizare a demersului. Scopul acestui articol este de a puncta câteva aspecte esențiale în interpretarea corectă a valorii lui p .

În esență DTIN presupune un calcul probabilistic. În acest context p reprezintă o probabilitate condiționată, mai precis probabilitatea de a obține o anumită diferență (exprimată în cote standard) în condițiile în care H_0 (ipoteza de nul) este adevărată (faptul că mediile grupelor reprezintă fluctuații datorate selecției aleatoare de eșantioane dintr-o populație).

p nu reprezintă probabilitatea ca H_0 să fie adevărată, ci probabilitatea de a obține o anumită diferență în contextul în care H_0 este adevărată (Gigerenzer, 1998).

Ca urmare, valorile p mai mari decât pragul α setat (diferențe ne semnificative) nu reflectă absența vreunui efect, adică faptul că diferențele observate se datorează selecției aleatoare. Acest fapt poate fi inferat pe baza coroborării rezultatelor testului statistic cu calculele de mărime a efectului și de putere a testului statistic, nicidecum pe baza valorii lui p (Cohen, 1990).

Raționamentul DTIN presupune o decizie dihotomică, dacă valoarea lui p (ex. 0.06) este mai mare decât valoarea setată a lui α (ex. 0.05), diferența observată este considerată ne semnificativă. În acest sens nu își are locul interpretarea unor valori p ne semnificative, dar apropiate pragului, ca fiind indicii ale unei tendințe. Aceasta reflectă o conceptualizare a valorii lui p pe marginea unui continuum, asemenea mărimii efectului.

Comunicarea exactă a valorii calculate a lui p este utilă din perspectiva unor meta-

analize ulterioare, însă și în acest caz decizia va fi una dihotomică.

Termenul de tendință nu este argumentată nici pentru că valoarea lui p nu oferă informații asupra unei evoluții ulterioare a rezultatelor în cazul replicării studiului, cu toate acestea 60% din psihologi cred că $1-p$ reprezintă probabilitatea de a replica rezultatele studiului (Oakes, 1986). Replicabilitatea rezultatelor semnificative ale unui studiu depinde de puterea testului. Chiar și în cazul în care mărimea efectului este mare (și o considerăm constantă) rezultatul nu va fi replicat dacă puterea testului este redusă.

Infirmarea H_0 (p calculat mai mic sau egal alfa) nu înseamnă că testul statistic a pus în evidență o relație causală. Cauzalitatea nu derivă din demersul statistic ci din specificul design-ului experimental utilizat. În cazul unui cvasi experiment infirmarea H_0 nu este un argument al cauzalității deoarece design-ul nu permite controlul unuia din criteriile cauzalității - cauza întotdeauna precede efectul. Nu avem de unde să știm dacă într-adevăr variabila etichetă a dus la diferențele observate ale variabilei dependente sau variabila dependentă (sau alte variabile externe) au determinat valorile variabilei independente. Cauzalitatea ține de validitatea internă a studiului și nu de procedura statistică (fapt adevărat și pentru experiment) (Anderson, 2001).

BIBLIOGRAFIE

- Anderson, N., H. (2001) Empirical direction in design analysis. Lawrence Erlbaum Associates, London.
- Cohen, J. (1990) Things that I have learned (so far). American Psychologist, 50, 1103.
- Oakes, M. (1986) Statistical inference: a commentary for the social and behavioral sciences. Wiley, New York.
- Nickerson, (2000) Null hypothesis significance testing: a review of an old and continuing controversy. Psychological Methods, 5, 241.