
Puterea statistică și testarea ipotezelor

Florin Sava

Universitatea de Vest, Timișoara

E-mail: afsava@socio.uvt.ro

Abstract

Lots of psychologists tend to say that statistics it's nothing more than mathematics and therefore a lots of troubles. In this article I intend to fight against this vision, by resorting to situations in witch we can draw different conclusions (for example, rejecting or not the statistical hypothesis) although we use the same information set.

Key words: *statistical hypotesis, experimental research, statistical power*

Mulți psihologi tind să afirme că statistica nu este nimic altceva decât matematică, iar de aici, o mulțime de necazuri. În cadrul acestui articol îmi propun să combat această viziune, apelând la situații în care putem desprinde concluzii diferite (de exemplu, respingerea sau nerespingerea ipotezei statistice), deși utilizăm același set de date.

Principala cauză care poate duce la această situație stranie este dată de nivelul *puterii statistice* a unei cercetări. Imaginați-vă puterea ca o capacitate de a detecta vizual efectele. Dacă puterea statistică este mică, va fi dificil să observăm efectele, deoarece posibilitatea noastră de detectare este redusă. În schimb, dacă puterea statistică este foarte mare, vom fi capabili să detectăm chiar și efecte minuscule sau neimportante.

În paginile următoare voi explica importanța puterii statistice în testarea ipotezelor, iar apoi voi oferi o serie de măsuri menite să crească puterea statistică a unei cercetări. Acest lucru este necesar deoarece majoritatea studiilor suferă de o putere statistică redusă. De altfel, Tversky și Kahneman (1993) sau Smith (2000) reamintesc de un studiu mai vechi a lui Cohen, care identifica în cadrul studiilor publicate într-o revistă de specialitate, o valoare mediană a puterii statistice de .46. Menționăm că valoarea optimă este în jurul a .80. Aceste rezultate modeste s-au păstrat și după 25 de ani de la efectuarea aceluși studiu (Cohen, 1990).

Tehnic, lucrarea conține două secțiuni majore. Prima, prezintă limitele aplicării mecanice a ipotezei statistice, o meteahnă a majorității cercetătorilor. Ultima, cuprinde douăsprezece măsuri menite să îmbunătățească

aceste neajunsuri, prin creșterea puterii statistice a unei cercetări.

Limitele ipotezei nule

Ipoteza statistică sau ipoteza nulă (H_0) ne arată probabilitatea de a obține datele observate în condițiile în care ipoteza nulă este adevărată (Cohen, 1990). Ea nu exprimă probabilitatea ca ipoteza nulă să fie adevărată sau nu. Astfel, un rezultat semnificativ statistic la $p < .05$ semnifică mai puțin de 5% șanse de a obține astfel de date în condițiile în care ipoteza nulă este adevărată. Ea nu indică faptul că există sub 5% șanse ca ipoteza nulă să fie adevărată. Acest raționament care își are originea în ideile lui Fisher, considerat a fi părintele statisticii moderne, a condus la o fetișizare a deciziei statistice în jurul pragului statistic p de .05. Exagerând, am putea afirma că publicarea sau nepublicarea studiilor, evoluția lucrărilor de doctorat sau licență, fericirea sau nefericirea cercetătorilor a devenit dependentă de acest prag. Ce ne facem însă în situația în care realizăm deficiențele și subiectivitatea deciziei luate bazându-ne exclusiv pe acest prag? Răspunsul se află la îndemâna noastră, însă pe cât este de simplu, pe atât este de greu de acceptat de consumatorii de statistică, adepți ai unor măsuri mecaniciste de genul „ți s-a confirmat sau nu ipoteza?”. Avem nevoie de o planificare a cercetării și prin aceasta avem nevoie să cunoaștem puterea statistică a unei cercetări.

Ideea a fost intuită încă din anul 1928, de către Jerzy Neyman și băiatul lui Karl Pearson, Egon. Aceștia au propus alături de ipoteza nulă, o ipoteză alternativă, fiecare dintre ele având atribuite un risc alfa, respectiv

beta. Respingerea ipotezei nule când aceasta nu trebuia respinsă a fost denumită eroare de tip I, controlată de nivelul lui alfa, în timp ce eșecul de a respinge ipoteza alternativă când aceasta trebuia respinsă este o eroare de tip II, controlată de nivelul lui beta (Cohen, 1990). Din păcate, în parte datorită neînțelegerilor dintre Fisher și Pearson, continuate și în relația cu fiul acestuia, ideea a fost înăbușită, fiind reluată cu mai multă vigoare doar după cel de-al doilea război mondial. Totuși, această schemă a fost reținută și transferată în ideile clasice ale lui Fisher, fiind prezentată în tabelul următor.

	<i>Ho este respinsă</i>	<i>Ho nu este respinsă</i>
<i>Există NU</i>	Eroare de tip I	Decizie corectă
<i>Există DA</i>	Decizie corectă	Eroare de tip II

Spre exemplu, eroarea de tip I se produce atunci când respingem ipoteza nulă și afirmăm că există diferențe semnificative între două loturi de subiecți deși nu există, în timp ce eroarea de tip II apare atunci când concluzionăm că nu există diferențe semnificative între două grupe, deși acestea există și ar fi trebuit identificate. Eroarea de tip I este o consecință a unei puteri statistice reduse, în timp ce eroarea de tip II este o consecință a unei puteri statistice inadecvate, dar foarte mari. Dintre cele două, eroarea de tip I este mai cunoscută, inclusiv de către studenții și absolvenții de psihologie. Astfel, dacă vor fi întrebați după ce criteriu resping sau nu ipoteza nulă, cei mai mulți vor răspunde simplu, în funcție de pragul de semnificație al testului aplicat. Dacă valoarea acestuia este sub .05 ar respinge ipoteza nulă, iar dacă această valoare este mai mare de .05 ar lua o decizie contrară. Puțini sunt conștienți de limitele acestei decizii și de subiectivitatea ce derivă din ea, datorită posibilității apariției erorii de tip I. Despre eroarea de tip II se cunosc și mai puține informații, iar cei care își amintesc de ea din punct de vedere teoretic nu prea țin cont în practică, atunci când își planifică sau interpretează studiile empirice. Totuși, pe baza erorii de tip II, controlată prin nivelul lui beta, putem stabili puterea statistică a unei cercetări.

Dacă alfa (α) este stabilit a priori, de obicei la nivelul de .05, beta (β) este foarte rar luat în considerare în planificarea unui studiu. Atunci când este utilizat se alege frecvent valoarea de .20, care semnifică aproximativ 20% șanse de a comite eroarea de tip II. Raportul de 4:1 în favoarea lui alfa își are

originea în tendința cercetătorilor de a acorda o mai mare importanță erorii de tip I. Bineînțeles că valoarea lui β se poate modifica, în funcție de tema cercetată.

Puterea statistică a unei cercetări

Puterea statistică poate fi determinată foarte ușor dacă ținem seama de valoarea lui β , deoarece valoarea sa este egală cu $1 - \beta$. Astfel, dacă vom alege un β de .20, puterea statistică a cercetării respective va fi de .80, ceea ce semnifică 80% șanse de a detecta un efect dacă acesta există. Similar, un β de .05, conduce la 95% șanse de a identifica un efect, oricât de mic ar fi el, cu condiția ca el să existe. Importanța practică a acestor informații este capitală în construcția designului de cercetare, deoarece puterea statistică a unei cercetări poate fi modificată în funcție de trei parametri importanți: *mărimea efectului* („Cât de mare trebuie să fie un efect sau o relație dintre două variabile pentru a fi identificată?”); *mărimea eșantionului* („De câți subiecți avem nevoie pentru a observa fenomenul de interes?”); *valoarea α* („Care este pragul de semnificație de la care vom respinge ipoteza statistică?”). Pentru a calcula puterea statistică există programe, precum Gpower (Faul și Erdfelder, 1992), ce poate fi descărcat gratuit de pe internet la adresa <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/index.html>.

De asemenea, inversând demersul logic, putem calcula numărul de subiecți de care avem nevoie pentru a realiza o cercetare eficientă, cu o putere statistică suficient de mare pentru a reduce șansele de a comite eroarea de tip II. Vă reamintim, că mediana puterii cercetărilor publicate în anumite reviste de specialitate nu a depășit .50, ceea ce înseamnă că erau doar ½ șanse de a observa un efect atunci când acesta exista. Înainte de a trece la prezentarea unor modalități de creștere a puterii de cercetare, voi oferi două exemple concrete de utilizare a informațiilor prezentate mai sus, pentru a înțelege importanța utilizării conceptului de putere a cercetării.

În primul exemplu, vom presupune că un cercetător s-a arătat interesat de a studia în ce măsură există o corelație pozitivă între două probe de inteligență, Matricile Progressive Raven – varianta standard și scorul general obținut la WAIS. Să presupunem că el a testat 25 de persoane și a obținut o corelație r de .30. Această valoare nu este semnificativă statistic, deoarece valoarea corespunzătoare din tabel

este puțin mai mare, r tabelar fiind .33. În aceste condiții, cercetătorul nu va putea respinge ipoteza nulă și va afirma că cele două teste de inteligență nu corelează, după care, probabil, se va lansa în tot felul de speculații legate de domeniile distincte pe care cele două probe le măsoară. Totuși, aceste teste ar fi trebuit să coreleze într-o anumită măsură, deoarece, în principiu, fac referire la același construct. În astfel de cazuri ne putem pune întrebarea: chiar nu există o corelație între cele două probe sau rezultatul este o consecință a unei erori de tip II prea mari, fapt ce mărește șansele de a ne înșela și de a nu observa corelația dintre cele două teste. Pentru a răspunde la întrebare vom putea face un simplu calcul apelând la Gpower. Au fost testați 25 de subiecți, valoarea α prestabilită a fost de .05, ipoteza de tip unidirecțional (există corelație pozitivă?), iar mărimea efectului f de .30. Înlocuind în formulă, a rezultat o putere a cercetării de .45, ceea ce înseamnă mai puțin de jumătate șanse de a observa corelația între cele două variabile dacă ea există și are o valoare r de aproximativ .30.

În aceste condiții, concluziile legate de nerespingerea ipotezei nule și de afirmarea absenței corelației dintre cele două probe devin extrem de discutabile și atacabile. Dacă cercetătorul ar fi cunoscut importanța puterii statistice ar fi putut planifica mai bine studiul, pentru a ajunge la o valoare acceptabilă a riscului de a comite eroarea de tip II. Astfel, dacă dorea să testeze existența unei asocieri pozitive între cele două variabile, în condițiile în care se aștepta la o mărime a efectului în jurul a .30, se mulțumea cu un $p < .05$, iar puterea cercetării dorite era de .80 ($\beta = .20$), ar fi avut nevoie de 64 de subiecți pentru a testa relația dintre probe. Dacă ar fi procedat astfel, rezultatele obținute ar fi fost mai veridice, deoarece șansele de a comite eroarea de tip II ar fi fost mult reduse.

Cel de-al doilea exemplu ar putea servi drept model în realizarea unui design adecvat din perspectiva erorilor de testare a ipotezelor statistice. Să presupunem că un cercetător dorește să studieze rolul feedback-ului în rezolvarea unor sarcini noi. În acest scop, el grupează aleator participanții la studiu în patru grupe: unul de control, unul alcătuit din subiecți care primesc doar feedback negativ legat de sarcină; un al treilea grup care primește feedback pozitiv legat de sarcină, respectiv un al patrulea grup care primește atât feedback pozitiv, cât și negativ. Variabila dependentă este măsurată în

secunde, vizând rapiditatea rezolvării sarcinii de către subiecți. Subiecții din cele trei grupe experimentale primesc feedback-ul în timpul rezolvării sarcinii, în timp ce subiecții din grupul de control nu primesc nici un fel de feedback. Într-un astfel de caz, cercetătorul poate planifica caracteristicile designului astfel încât să aibă erori de tip I și II reduse.

Să presupunem că cercetătorul va alege să testeze ipoteza apelând la ANOVA simplă, urmată în caz de succes de o serie de comparații post hoc. În aceste condiții, el poate face o serie de calcule pentru a determina volumul eșantionului în condițiile în care sunt patru grupe, α dorit este de .05, puterea statistică așteptată este de .80, iar efectul feedback-ului asupra duratei de rezolvare a sarcinii este unul mediu ($f = .25$). În urma calculului a rezultat că cercetătorul va avea nevoie de un eșantion de 180 de persoane, adică de 45 de subiecți în fiecare grupă pentru a putea testa optim ipoteza și a menține scăzute erorile. În cazul în care ar fi ales 100 de subiecți, câte 25 în fiecare grupă, ar fi obținut o putere a cercetării de .50, ceea ce ar fi însemnat $\frac{1}{2}$ șanse de a observa efectul feedback-ului asupra performanței, în cazul în care acest efect există și are o intensitate medie. Mărimea efectului poate fi stabilită, fie pe baza rezultatelor obținute din cercetări anterioare, fie pe bază teoretică, pornind de la niște convenții care împart mărimea efectului în: efect mic (influență redusă a VI asupra VD); efect mediu (influență moderată a VI asupra VD) și efect mare (influență puternică a VI asupra VD). Asupra mărimii efectului nu vom insista aici, însă cei interesați pot găsi mai multe informații în Hinkle, Wiersma și Jurs (1994), Pedhazur și Schmelkin (1991), Rosenthal, Rosnow și Rubin (2000) sau Sava (2001). Vom preciza doar, că mărimea efectului poate fi calculată prin două categorii de indicatori: bazați pe cantitatea de dispersie explicată (r^2 , r^2 ajustat, η , ω^2 ș.a.) sau bazați pe diferențele standardizate dintre grupe (d a lui Cohen, g a lui Hedges ș.a.) (Thompson, 1999).

Douăsprezece modalități de a crește puterea statistică

Ținând seama că majoritatea cercetărilor realizate suferă de o putere statistică redusă și implicit de șanse mari de a comite eroarea de tip II, voi prezenta în continuare o serie de măsuri menite să crească puterea statistică a unei cercetări. Aceste măsuri pot fi împărțite în trei categorii

mari: măsuri care țin predominant de designul cercetării (1-7), măsuri care privesc măsurările efectuate în cadrul cercetării (8-10) și măsuri exclusiv statistice (11-12). Aceste modalități se bazează pe lucrările lui Embertson și Reise (2000), Hinkle, Wiersma și Jurs (1994), Kirk (1982), Pedhazur și Schmelkin (1991), Smith (2000) sau West, Biesanz și Pitts (2000), cu toții interesați să continue munca depusă de Jacob Cohen, părintele puterii statistice. Majoritatea acestor măsuri au fost prezentate într-o altă lucrare (Sava, 2002).

1. Creșterea numărului de subiecți testați. Cea mai simplă și comună metodă de a crește puterea statistică a unei cercetări constă în includerea mai multor participanți în studiu. Acesta poate fi făcută, fie prin creșterea numărului de subiecți din fiecare lot testat, fie prin mărirea numărului de subiecți din anumite grupe, mai importante din perspectiva ipotezelor testate. Spre exemplu, avem un design unifactorial intergrup în care, prin intermediul a trei grupe de subiecți (T1-grup de control; T2- terapie individuală; T3 – terapie de grup), dorim să testăm efectul terapiei cognitiv-comportamentale în reducerea gradului de timiditate. În condițiile în care fiecare grup este format din 10 subiecți avem la dispoziție două posibilități de a crește puterea cercetării prin mărirea volumului eșantionului. Prima constă în dublarea numărului de subiecți din fiecare grupă, ajungând la 60 de subiecți testați. Metoda merită utilizată când toate grupele au o importanță aproximativ egală. Dacă grupele au o importanță diferită, cum ar fi situația în care dorim să aflăm dacă intervenția de orice fel (individuală sau în grup) este semnificativ mai eficientă decât non-intervenția, vom apela la o strategie diferită. De exemplu, vom adăuga doar 5 subiecți în grupele experimentale, respectiv 20 de subiecți în grupul de control. Similar, dacă este mai important să determinăm care dintre formele de intervenție (individuală vs. în grup) este mai eficientă, vom putea crește grupul de control cu 6 subiecți, în timp ce vom adăuga 12 subiecți în fiecare grup experimental. Se observă că numărul de subiecți adăugați se menține constant la valoarea de 30, însă strategia este diferită în funcție de ipotezele urmărite.

2. Utilizarea unui design intragrup (cu măsurători repetate). În cazul în care avem același număr de subiecți, un design intragrup are o putere statistică mai mare decât designul de tip intergrup. Acest fapt se datorează scăderii dispersiei erorii (neexplicate) în primul caz, datorită unei variabilități mai reduse a

rezultatelor, fiind vorba de aceeași subiecți testați în situații diferite și nu de subiecți diferiți testați în condiții diferite. Aproape orice test statistic are la numitor, dispersia erorii și în consecință, va duce la un scor t sau F mai mare, măbind șansele de a obține un rezultat semnificativ statistic.

3. Introducerea unei covariabile eficiente. În situații în care nu este posibilă creșterea volumului eșantionului din măsuri de economie de timp și resurse sau în care designul de tip intragrup nu este potrivit se poate ține seama de efectul unei terțe variabile, denumită *covariabilă*. Spre exemplu, cunoscând importanța încrederii în terapeut, un cercetător poate măsura și controla influența acestei variabile atunci când testează efectul intervenției cognitiv-comportamentale în reducerea timidității. Prin introducerea unei covariabile care influențează relația dintre alte variabile, fie ele VI și VD, se reduce dispersia erorii, și astfel, cresc șansele obținerii unui rezultat semnificativ statistic, fără a recurge la creșterea numărului de subiecți testați. Acest raționament poate fi vizualizat în Figura 1:

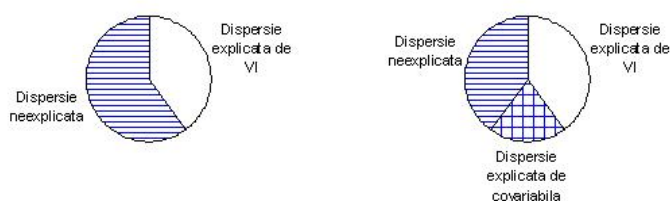


Figura 1. Sursa dispersiei în ANOVA (stânga), respectiv ANCOVA (dreapta)

4. Utilizarea unui design de cercetare cât mai simplu. Acest principiu filosofic, cunoscut sub numele de briciul lui Ockham, exprimă primatul unor explicații simple și corecte în dauna unor modele complicate și dense, cu prea multe variabile. În statistică, acest principiu a fost etichetat de Cohen (1990) sub forma „mai puțin este mai bine”. În studii cu foarte multe variabile crește șansa de a comite eroarea de tip I, adică de a obține anumite rezultate semnificative din întâmplare. Vom ajunge în cealaltă extremă, întrebându-ne care dintre rezultatele semnificative obținute sunt reale? Prin aplicarea unor corecții, vom putea reduce distorsiunile create, dar vom crește șansele de a comite o eroare de tip II. De exemplu, prin corecția Bonferroni, aplicată la o matrice de corelații între 16 variabile (de exemplu, factorii primari de la 16PF a lui

Cattel), pragul de semnificație pentru respingerea ipotezei nule se va modifica de la .05 la .0004 (.05/60). Astfel, o corelație a cărei prag de semnificație obținut este de .001 nu va putea respinge ipoteza nulă. Corecția este absolut necesară, deoarece în absența ei, în condițiile în care *nu* există asocieri reale între variabile, șansa de a obține, din întâmplare, unul sau mai multe rezultate semnificative este de 99,3%.

5. *Creșterea numărului de trepte pentru VI sau / și creșterea amplitudinii factorilor.* Altfel spus, sunt mai mari șansele de a obține un rezultat semnificativ statistic în condițiile în care factorul variază mult. Astfel, după doar trei ședințe de terapie, am putea concluziona că intervenția nu a condus la ameliorarea situației clientului. După doisprezece ședințe există șanse mai mari de evoluție pozitivă a clientului. De aceea, este important să acordăm o atenție deosebită nivelului intensității VI asupra VD. De cele mai multe ori acest nivel trebuie ales astfel încât să corespundă cât mai bine realității, pentru a avea o validitate ecologică sporită. Nu întotdeauna trebuie să mărim intensitatea VI pentru a observa efectul său asupra VD. Altfel, avem nevoie de o analiză detaliată a evoluției VI pe toată gama amplitudinii sale. Acest fapt poate fi observat și în legea Yerkes-Dodson. Dacă am dori să vedem care este efectul nivelului de arousal (motivare) asupra performanței în sarcină și am avea doar două grupe de subiecți (slab motivați, respectiv puternic motivați) am putea ajunge la concluzia că nu există diferențe între cele două grupe și implicit, că nu există un efect al nivelului de motivare asupra performanței subiecților. Dacă în loc de două grupe, am utiliza trei loturi, corespunzătoare a trei trepte ale VI (slab motivați, optim motivați, puternic motivați) am ajunge la rezultate diferite și am admite influența nivelului de arousal asupra performanței subiecților în sarcină.

6. *Tratarea variabilelor cauză drept factori ficși.* Din fericire, de cele mai multe ori, VI sunt tratate automat ca factori ficși în cadrul tehnicilor ANOVA. Deși tratarea VI ca factor aleator poate facilita generalizarea rezultatelor obținute la alte trepte ale VI, tehnica suferă de o putere statistică redusă. Într-un experiment tipic unifactorial, vom avea 2-3 trepte ale VI și aproximativ 90 de subiecți. Acest lucru este normal dacă VI este tratată ca factor fix. Dacă VI este considerată un factor aleator, scade puterea statistică de a detecta un efect, deoarece apare o problemă. Deși în privința

subiecților nu sunt dificultăți, numărul lor fiind suficient de mare, va fi greu să generalizăm concluziile la întreaga populație de trepte posibile, din care cele două sau trei nivele au fost extrase. Dacă este greu să generalizăm, extrapolând rezultatele pornind de la două, trei cazuri particulare, va fi la fel de dificil să generalizăm concluziile pentru întreg registrul VI, pornind de la două, trei trepte.

7. *Utilizarea unor ipoteze unidirecționale (unilaterale).* Ipotezele de cercetare emise pot avea două forme: bidirecționale („există o corelație între variabile”; „există diferențe între loturi”) sau unidirecționale („există o corelație pozitivă”; „există o corelație negativă”; „grupul A este mai bun decât grupul B”; „grupul A este mai slab decât grupul B”). În condițiile utilizării unei ipoteze de tip unidirecțional se dublează șansele de a obține rezultate semnificative statistic, deci crește puterea statistică a cercetării. Astfel, dacă efectuăm o corelație pe 25 de subiecți între două probe și obținem o valoare de -.30 vom interpreta diferit rezultatele în funcție de tipul de ipoteză avută. Datele ar fi sprijinit o ipoteză unidirecțională („persoanele cu un nivel crescut de depresie tind să petreacă mai puțin timp în oraș?”), $r = -.30$, $p < .05$, însă ar fi infirmat o ipoteză bidirecțională („există vreo asociere între nivelul de depresie a unei persoane și perioada de timp petrecută în oraș?”), $r = -.30$, $p > .05$.

8. *Utilizarea scalelor de măsurare numerice, de tip interval sau proporții.* Acest gen de scale de măsurare apelează frecvent la tehnici statistice parametrice. Din perspectiva puterii statistice este cunoscut faptul că tehnicile statistice parametrice au un nivel mai ridicat decât metodele statistice neparametrice. În consecință, ar fi de dorit ca cel puțin variabilele dependente sau criteriul să fie măsurate prin scale de tip interval sau proporții și să se evite, pe cât posibil, măsurarea lor prin scale de tip nominal sau ordinal.

9. *Transformarea datelor pentru a îndeplini condițiile diferitelor tehnici statistice.* Chiar dacă datele sunt măsurate prin scale numerice, majoritatea tehnicilor statistice necesită alte condiții (asumpții) suplimentare pentru a putea fi aplicate corect. Nerespectarea asumpțiilor majore, în situații de aplicare mecanică a testelor statistice, poate conduce la erori în testarea ipotezelor, fie de tip I, fie de tip II. Spre exemplu, o distribuție complet anormală a variabilei dependente ar presupune renunțarea la aplicarea testului t pentru două eșantioane independente, în

favoarea unei metode neparametrice corespunzătoare, în acest caz, testul U a lui Mann-Whitney. Puterea statistică a metodei este însă la nivelul a 69% din capacitatea testului t de a detecta diferențe între cele două grupe. În asemenea condiții pot fi adoptate o serie de artificii statistice de transformare a datelor brute, menite să transforme datele inițiale ale VD și să corecteze distribuția obținută, astfel încât să nu fie probleme cu aplicarea testelor parametrice. Există o multitudine de transformări posibile, cele mai întâlnite fiind ridicarea la pătrat, extragerea rădăcinii pătrate sau utilizarea logaritmulor naturali. Iată, spre exemplu, valorile indicatorilor de asimetrie și boltire, înainte și după transformare, dintr-o bază de date fictivă:

Variabilă	Asim.	Er.std.	Boltire	Er.std
X	-.978	.314	.082	.618
X-pătrat	-.439	.314	-.952	.618

Înainte de transformarea datelor se observă o deviere importantă de la normalitate, fiind o distribuție puternic asimetrică negativ. De altfel, anormalitatea distribuției este susținută și de rezultatul testului Kolmogorov-Smirnov ($z = 1,50$, $p < .05$). După ridicarea la pătrat a datelor se reduce foarte mult asimetria datelor, însă intervin unele modificări la nivelul boltirii. Totuși, după transformarea datelor, distribuția obținută nu diferă semnificativ de una normală ($z = 1,13$, $p > .05$), fapt ce va permite aplicarea unor teste parametrice. Astfel, scad șansele de a comite erori de tip I sau II.

10. *Selectarea unor probe cu caracteristici psihometrice ridicate, mai ales în măsurarea VD.* Itemii care au un nivel de eroare scăzut și care reușesc să reflecte cu sensibilitate orice variații ale variabilei latente corespunzătoare conduc la o putere statistică mai mare decât cei care nu îndeplinesc aceste condiții. Spre exemplu, o scală care are o consistență internă de .65, are un nivel ridicat al erorii, fapt ce afectează relaționarea sa cu alte probe. În corelarea scalei cu alte probe se vor obține valori scăzute, sub nivelul a .65. Astfel, relația sa cu o altă variabilă ar putea suferi (ex. am putea obține o corelație de .25, care ar fi fost mai mare dacă ambele probe ar fi avut o consistență internă ridicată, în jurul a .90). În privința sensibilității, teoria răspunsului la itemi oferă un bun mijloc de a testa și selecta acei itemi care răspund cel mai bine acestei cerințe. Nu în ultimul rând, scalele care corelează moderat sau puternic cu probele de

dezirabilitate socială pot conduce la scăderea puterii statistice.

11. *Salvarea unor grade de libertate.* Deși am amintit implicit de acest aspect la punctul patru, salvarea unor grade de libertate, cu scopul de a spori puterea unei cercetări, poate fi făcută și într-un design complex (de exemplu, ANOVA 3x3x3). Într-un asemenea caz am putea calcula șapte indicatori F, câte trei pentru efectele simple (FA, FB, FC), pentru interacțiunile de ordinul doi (FAB, FAC, FBC) și unul pentru interacțiunea complexă de ordinul trei (FABC). Nu are rost să calculăm toate aceste efecte în condițiile în care sunt interacțiuni care nu ne interesează sau în cazul în care interpretarea rezultatelor ar fi dificilă sau lipsită de sens (ex. FABC). Renunțând la calcularea acelor indici vom salva un anumit număr de grade de libertate, ceea ce ar conduce la o creștere ușoară a puterii statistice a testelor F rămase a fi efectuate. În situații limită, rezultatele ne semnificative pot deveni semnificative (de exemplu, un p de .07 se schimbă într-un p de .04).

12. *Utilizarea tehnicilor statistice cu putere ridicată.* În practică, putem întâlni situații în care pot fi aplicate mai multe tehnici statistice. De exemplu, în cadrul comparațiilor post hoc din ANOVA, pentru grupe egale de subiecți și omogenitatea dispersiei am putea alege între procedurile Tukey și Bonferroni. Totuși, primele au o putere mai mare de a detecta diferențe între grupe în comparație cu ultimele. Similar, într-un design intragrup putem apela la ANOVA pentru măsurători repetate, în variantă corectată sau putem alege MANOVA. Prima opțiune are o putere statistică mai mare decât MANOVA, de aceea ar fi preferabilă utilizarea sa.

Toate aceste măsuri pot fi utilizate în sprijinul creșterii puterii statistice a unei cercetări. Voi încheia precizând că problema prezentată nu este doar un moft al statisticienilor, menit să îndepărteze și mai mult comunitatea psihologică de la o abordare științifică a problemelor. APA (2001) oferă dovada cea mai clară în acest sens și face recomandarea de a se ține cont de puterea statistică în stabilirea designului și în interpretarea rezultatelor unei cercetări.

Bibliografie

APA (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

-
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.
- Embertson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates, Publishers.
- Faul, F., & Erdfelder, E. (1992). *GPOWER: A priori, post hoc, and compromise power analyses for MS-DOS [Computer program]*. Bonn, Germany: Bonn University, Dept. of Psychology.
- Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1994). *Applied statistics for the behavioral sciences* (3rd ed.). Boston: Houghton Mifflin Company.
- Kirk, R.E. (1982). *Experimental design: procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks / Cole Publishing.
- Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design, and analysis: an integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Rosenthal, R., Rosnow, R.L., & Rubin, D.B. (2000). *Contrasts and effects sizes in behavioral research. A correlational approach*. Cambridge: Cambridge University Press.
- Sava, F. (2001). *Pagina de statistică socială*. Disponibilă la <http://statisticasociala.tripod.com>
- Sava, F. (2002). Ten ways to increase statistical power. *Analele Universității de Vest din Timișoara. Seria Sociologie, Psihologie, Pedagogie și Asistență Socială*, 8, 89-92.
- Smith, E.R. (2000). Research design. In H.T. Reis & C.M. Judd (eds.). *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press, 17-39.
- Thompson, B. (1999). If statistical significance tests are broken / misused, what practices should supplement or replace them? *Theory & Psychology*, 9(2), 165-181.
- Tversky, A., & Kahneman, D. (1993). Belief in the law of small numbers. In G. Keren & C. Lewis (eds.). *A handbook for data analysis in the behavioral sciences: methodological issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- West, S.G., Biesanz, J.C., & Pitts, S.C. (2000). Causal inference and generalization in field settings: experimental and quasi-experimental designs. In H.T. Reis & C.M. Judd (eds.). *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press, 40-84.