

PREFAȚĂ LA EDIȚIA I

Începând cu anul 1990, mi s-a deschis posibilitatea valorificării în învățământul superior a experienței acumulate în biometrie și informatică aplicate în cercetarea biomedicală, experiență dobândită în 10 ani de lucru nemijlocit în Centrul de Antropologie și Institutul Victor Babeș, precum și prin colaborări cu Muzeul Grigore Antipa, Institutul de Biologie, Institutul de Igienă, etc. În cadrul Facultății de Biologie a Universității din București, am construit cursuri de "Tehnică de calcul și programare", "Biostatistică", "Ecostatistică", "Modelare matematică în ecologie", etc., primele două concretizându-se și în manualele [2] și [3].

Pentru că învățământul modern în domeniul științelor vieții presupune asimilarea biostatisticii la nivel de mod de gândire și utilizarea calculatoarelor moderne la nivel de obișnuință, este nevoie de un colectiv puternic specializat în acest sens, așa cum se întâmplă în învățământul superior de excelență din Europa de Vest sau din S.U.A. De aceea, am început să formez ca asistenți patru absolvenți de matematică, politehnică (secția bioinformatică) și biologie. Din cauza situației actuale din țara noastră, în prezent trei lucrează în specialitate în S.U.A., pe timp nelimitat, iar cel de-al patrulea este mai bine motivat în țară, la o universitate particulară. Deoarece aceste pierderi au devenit, la un moment dat, în mod previzibil "pierderi planificate", am fost nevoit, pe de o parte, să construiesc "asistenți automați", iar pe de alta, să antrenez cei mai buni studenți la formarea celorlalți. Așa s-au născut (1) o serie de **instrumente informatice originale**, care automatizează diverse faze ale procesului didactic precum și (2) **probleme de statistică aplicată în biologie, biochimie sau ecologie**. Să le explicităm pe rând.

(1) Instrumente informatice

Două instrumente informatice se observă încă din volumul [3]. Este vorba despre programe care desenează exact diverse entități statistice (de exemplu distribuția normală, diverse reprezentări grafice, etc.) și de programe de demonstrație, cum ar fi de exemplu cel dedicat distribuției mediei de eșantionaj, etc. În acest volum sunt prezentate încă trei instrumente realizate pe calculator. Primul este un sistem automat de construcție și aplicare de teste grilă (detalii în Anexa 7). Fiecare lucrare practică din acest volum începe cu un exemplu de Test Grilă (notat TG) din porțiunea respectivă de materie sau din materia până în acel moment. Studenții cursului de "Biostatistică" al Facultății de Biologie a Universității din București susțin acest test pe calculator, contra cronometru, la fiecare lucrare practică, precum și la verificarea finală. Al doilea instrument este un sistem de construcție și aplicare de Teste de Completare (notate TC). De asemenea, fiecare lucrare conține câte un exemplu pentru materia respectivă. Din păcate, acest test nu se poate susține pe calculator, pentru că durează mult mai mult iar numărul de calculatoare este insuficient. Al treilea instrument este format din mai multe programe care "știu" să rezolve anumite exerciții tip, pentru orice set de date. Am denumit produsele acestor programe Exerciții Generate Automat (EGA). Și aceste programe se utilizează *off line*, din același motiv.

Am enumerat toată această infrastructură software pentru că este bine să se știe că, practic, există pentru biostatistică un set modern de instruire asistată de calculator: volumul [3], prezenta lucrare și toate programele enumerate. Prin urmare, este gata o lucrare modernă cu dischetă atașată (conținând două sisteme de autotestare a teoriei, un sistem de generare de exerciții și programe demonstrative pentru aspecte delicate ale gândirii statistice, cum ar fi distribuții de eșantionaj, randomizarea, etc.) Din păcate însă, nu cred că piața noastră este capabilă să absoarbă, în prezent, o asemenea lucrare.

(2) Exerciții și probleme

Am evitat, în mod intenționat, date reale în probleme pentru că acestea fie au volume prea mari, fie necesită prelucrări preliminare și/sau multiple. De aceea, toate problemele sunt cu date fictive dar verosimile și au fost construite cu scopuri țintite. Dezvoltarea capacității de rezolvare a acestui tip de probleme este obiectivul maximal care se poate cere pregătirii începătorilor. Rezolvarea problemelor reale și mai ales problematizarea situațiilor concrete necesită cel puțin parcurgerea a încă două niveluri, cel intermediar și cel pentru avansați.

Lucrarea conține 120 de exerciții și probleme de biostatistică originale, cu excepția a 5 probleme: problema 8 din Lp 9 preluată din [1] cu enunț modificat (pentru a evita contradicția semnalată de noi în observația 2 din rezumatul aceleiași lucrări practice) și problemele 2, 3, 6 și 7 din Lp 8, inspirate de lucrarea [4]. Majoritatea exercițiilor au fost generate cu programele amintite mai sus. Problemele sunt fie unicate, fie "probleme model", fie probleme care imită, pentru consolidare, problemele model. Ultima categorie cuprinde un număr important de probleme construite de către foști studenți. Am rezolvat demonstrativ fiecare problemă din această ultimă categorie, ameliorând enunțul ori de câte ori a fost cazul. În toată lucrarea există unele **mici inovații** menite să apropie studentul de situațiile reale. Menționăm în acest sens problemele nerezolvabile (vezi problema 2 din Subiecte finale), probleme pentru care se cere doar rezolvarea de principiu, fără calcule (vezi problema 3 din Subiecte finale), probleme cu rezolvări "capcană", greșite (de exemplu, 5 din Lp 11) și probleme cu rezolvări model cu accentuarea aspectelor care, adesea, se tratează greșit (de exemplu, 4 și 5 din Lp 7).

Toate exercițiile și problemele sunt rezolvate. Cu toate acestea, ele sunt partajate în "exerciții sau probleme rezolvate" și "exerciții sau probleme propuse". Deosebirea constă în faptul că fiecare problemă din prima categorie are plasată rezolvarea imediat după enunț, în timp ce rezolvările așa-ziselor "exerciții sau probleme propuse" sunt grupate către finalul lucrării într-un capitol dedicat acestora și sunt, eventual, mai puțin detaliate. 117 exerciții și probleme sunt plasate la sfârșitul lucrărilor practice acolo unde le este locul în funcție de cunoștințele și deprinderile necesare rezolvării lor, iar 3 probleme formează un subiect de verificare finală plasat în capitolul "*Subiecte finale*".

Din punct de vedere al conținutului biostatistic, problemele conțin în esența sau rezolvarea lor aspecte mai mult sau mai puțin originale. De pildă, în cazul problemelor care necesită aplicarea de teste statistice, am introdus diferențierea "decizie statistică - decizie de specialitate" precum și conceptul de "ipoteză logică" asociat celui de "ipoteză statistică"

(pentru detalii, vezi Anexa 8). Am algoritmizat metodologia stabilirii eficacității unui tratament atât în variantă parametrică (8 și 19 din Lp 9), cât și neparametrică (4 și 11 din Lp 11). Am construit modelul statistic al rezultatelor unui examen "bine calibrat" (1 din Lp 4) și am verificat statistic acest lucru pe date reale din propria experiență (1 și 6 din Lp 9). Problemele 2, 3, 6 și 7 din Lp 8 explică experimentatorului care este probabilitatea repetării respingerii ipotezei nule, de unde rezultă importanța publicării nivelului p de respingere, oricât de mic ar fi acesta. Oriunde a fost posibil, am utilizat așa-numitele formule de calcul în locul celor teoretice (de exemplu, pentru dispersie și χ^2).

Rezumate ale teoriei

Consultând lucrarea [1] într-o fază avansată a elaborării acestui volum, am preluat ideea introducerii la începutul fiecărei lucrări practice a unui **rezumat teoretic**. Cu acest prilej am adăugat, ameliorat sau corectat diverse aspecte din volumul [3].

Pentru o sesizare rapidă a acestor actualizări, le-am introdus într-un chenar ca aici,

sau le-am subliniat cu o linie dublă.

Menționez, în continuare, unele ameliorări introduse în rezumatele din partea de statistică inductivă. De exemplu, în Lp 7, definind conceptele de "reprezentativitate în sens comun" și cel de "reprezentativitate sub control probabilist", am arătat că reprezentativitatea din statistică diferă esențial de conceptul intuitiv, comun. Tot acolo am "demască" eroarea "volum mare \Rightarrow eșantion reprezentativ" provenită din neînțelegerea reprezentativității din statistica inductivă.

În Lp 8 am schimbat modul de formulare al ipotezei nule în cazul testelor statistice unilaterale, explicând în Anexa 8 motivul acestei decizii, prin introducerea conceptului de "ipoteză logică" atașată unei ipoteze statistice.

În Lp 9 am propus o formulare mai generală și mai sugestivă a condiției care produce respingerea ipotezei nule, utilizând exprimarea "statistica testului este *mai excentrică* decât ...". și am explicat expresia pentru fiecare tip de test considerat în funcție de modul de plasare a zonei de respingere. Am descris mai detaliat ce înseamnă teste parametrice, respectiv neparametrice. Am prezentat testarea concordanței nu numai cu o distribuție normală ci și cu o distribuție uniformă, precum și cu o distribuție produsă de legea Hardy-Weinberg pentru două alele pe același locus. Am indicat care este instrumentul de statistică parametrică corect pentru stabilirea eficacității unui tratament (vezi observația 1). Am plasat sugestia din [4] de a se utiliza în testele unilaterale nivelul $\alpha/2$, în locul nivelului α , în contextul metodologiei parametrice de stabilire a eficacității unui tratament, astfel fiind mai ușor de înțeles motivația acestei sugestii, prin evidențierea contradicției "inefectiv dar eficace", care s-ar produce dacă s-ar lucra cu nivelul α și în testele unilaterale (vezi observația 2 din Lp 9).

În sfârșit, în *Lp 11* am introdus formula de calcul rapid și exact al lui χ^2 cu corecția lui Yates pentru tabele 2×2 și am indicat metodologia neparametrică de testare a eficacității unui tratament.

În cadrul capitolului "*Subiecte finale*" am considerat util să prezint un exemplu de probe de verificare finală și testele de biostatistică utilizate în selecția finală a loturilor naționale olimpice de biologie în anii 1998 și 1999. Probele de verificare finală au făcut parte chiar din sistemul de notare analizat statistic în problemele menționate mai sus, în raport cu referențialul propus în problema amintită de asemenea mai sus, 1 din *Lp 4*.

Lucrarea se încheie cu cele 6 anexe din [3] pentru a putea fi consultată și independent. În aceste anexe sunt unele mici modificări, subliniate conform convenției amintite cu linie dublă. Menționăm doar marcarea, în Anexa 3, a nivelului 2,5 % pentru testele Student unilaterale dreapta. La aceste 6 anexe se adaugă încă două.

Anexa 7 este special dedicată studenților care urmează cursul de Biostatistică în cadrul Facultății de Biologie a Universității din București și conține o prezentare a sistemului nostru de aplicare de teste grilă prin utilizarea calculatorului *on line*. Am proiectat acest sistem pe care îl exploatăm de mai bine de 8 ani pentru că numai studiul continuu poate asigura o învățare eficientă. În SUA, de exemplu, studiul este impus printr-o verificare continuă. Studentul este obligat să trimită până la cursul următor rezolvarea unei teme propuse la cursul curent. Tema este corectată, iar nota primită intră în calculul notei finale. Din păcate, această metodă nu este adecvată la noi atât din cauza numărului redus de cadre didactice care nu pot corecta așa de multe teme, cât, mai ales, din cauza mentalității studenților de înțelegere greșită a "întrajutorării". De aceea, am găsit de cuviință că o soluție de impunere a învățării continue poate fi verificarea la fiecare lucrare practică a asimilării materiei la zi printr-un test grilă pe calculator. Sistemul este astfel conceput încât cu un număr redus de calculatoare să poată susține testul toți studenții, iar cei care-l susțin concomitent nu pot să se "ajute" între ei. Notele obținute automat într-un timp standardizat au o pondere foarte mare în nota finală. Recomand viitorilor studenți să studieze această anexă înainte chiar de prima lucrare practică.

Anexa 8 prezintă conceptul statistic de nivel de semnificație, punând în evidență ceea ce corespunde intuiției comune în opoziție cu ceea ce este specific statisticii inductive, în particular introducând conceptul de "cuplu de ipoteze logice corespunzătoare unui cuplu de ipoteze statistice".

Scurt istoric și mulțumiri

Acum nouă ani Prof. Dr. *Angheluță Vădineanu*, înființând o secție de ecologie cu un nou plan de învățământ, mi-a propus să încep construcția unui colectiv de biometrie și informatică pentru Facultatea de Biologie, după modelul universităților de performanță din SUA și Europa de Vest, fiind dovedit că modelul alternativ al solicitărilor de acest tip la alte

departamente este net inferior. Am acceptat cu greu ambițioasa provocare și pentru că astfel a trebuit să părăsesc excelentul colectiv în care m-am format ca biometrician și m-am dezvoltat ca informatician, și anume Centrul de Antropologie condus de mult regretatul Prof. Dr. Doc. *Victor Săhleanu*.

Planul de învățământ al secției de ecologie a fost inițial foarte consistent pentru formația biometrică a viitorilor ecologi. Astfel, cu primele serii de absolvenți am fost în contact direct 4 sau 6 semestre, după ce au beneficiat de 2 semestre de matematică oferite cu profesionalism de facultatea de profil. În prezent, aproape toți doctoranzii în ecologie în departamentul de profil al Universității din București provin din aceste serii. În plus, am susținut 4 foști studenți care au fost acceptați în SUA pentru studii de master și de doctorat în biostatistică, alături de absolvenți de matematică.

Perioada în care ne aflăm a adus mult prea dese schimbări, acestea împiedicând cristalizarea unei formule de plan de învățământ și obligând la eforturi deosebite cadrul didactic care dorește ca fiecare serie de studenți să primească pregătirea cea mai adecvată și cât mai profundă. Formula actuală este de doar trei semestre de contact direct cu viitorii ecologi (1 semestru de biostatistică și 2 de ecologie numerică), 1 semestru cu ceilalți și doar studenții de la biochimie mai urmează un curs de matematică. Pentru compensarea parțială a acestei severe restrângeri am elaborat cartea [3], prezentul volum și întreg sistemul informatic punctat mai sus. Desigur, toate acestea mi-au limitat posibilitatea de a răspunde la alte solicitări de biometrie sau modelare matematică, venite atât din partea foștilor mei colegi antropologi cât și a noilor colegi. Tuturor le mulțumesc, pe această cale, pentru înțelegere.

În mod special trebuie să mulțumesc colegilor Prof. Dr. *Angheluță Vădineanu*, Conf. Dr. *Dan Gabriel Manoleli*, Dr. *Constantin Ciubuc*, Dr. *Sergiu Cristofor*, Dr. *Gheorghe Ignat* și Lect. Dr. *Dan Cogălniceanu* care mi-au oferit suportul logistic necesar. O bună colaborare am avut cu Prof. Dr. *Radu Meșter* și Conf. Dr. *Veronica Stoian*, în cadrul pregătirii loturilor naționale olimpice de biologie din anii 1998 și 1999. Pentru pregătirea lotului din 1998 am primit sprijin direct și din partea Prof. Dr. Ing. *Wanzer Drane* de la Universitatea Carolina de Sud, SUA.

De un real folos au fost diverse discuții profesionale cu Dr. *Ovidiu Ionescu*, Șeful Laboratorului de Biologia Vânatului și a Salmonidelor la Institutul de Cercetări și Amenajări Silvice, cu *Florin Albeanu*, medaliat cu argint la Olimpiada Internațională de Biologie 1998, în prezent student la Institutul de Tehnologie din Massachusetts, SUA (celebrul "MIT") și cu fostul meu student, *Alexandru Ionuț Petrișor*, proaspăt absolvent de master în biostatistică la Universitatea Carolina de Sud, SUA.

În sfârșit, studenții care au răspuns emulației create și care s-a finalizat și prin această lucrare sunt: *Cristina Cotroceanu*, *Dana Ghioca*, *Nicoleta Gheorghe*, *Mirela Tulbure*, *Gabriela Marcu* și *Raluca Băgăian*. *Cristina Cotroceanu*, absolventă de biochimie și de ecologie, a lucrat cu mine 6 semestre, ceea ce i-a deschis posibilitatea de a participa la construcția de probleme, asigurând "fabulația" biochimică și ecologică la un număr

important dintre acestea. *Dana Ghioca*, preluând modelul de construcție a unui test de completare, a propus imensa majoritate a întrebărilor acestor teste. *Nicoleta Gheorghe* a "furat" de asemenea o parte din meserie și a propus o parte semnificativă de teste grilă. Absolventa cea mai bună la biostatistică din anul I Biologie, *Mirela Tulbure*, a verificat benevol toate lucrările practice de statistică descriptivă, practic testând o jumătate din carte. *Gabriela Marcu* și *Raluca Băgăian* au participat la corecții ale materialului, în diverse stadii, alături de soția mea, *Nicoleta Dragomirescu*.

În încheiere mulțumesc anticipat tuturor celor care vor avea amabilitatea să facă observații pertinente, în vederea unei eventuale noi ediții.

Septembrie 1999