

Capitolul 1

INTRODUCERE

Motto: "Există trei tipuri de minciuni: minciuni, minciuni gogonate și statistici."
Mark Twain

Statistica este un instrument de cunoaștere științifică, un mod de gândire complex și dificil de asimilat, deoarece cere experiență, fler, o cunoaștere aprofundată a multor subdomenii - cele mai multe aparținând matematicii - precum și o anumită etică profesională.

Yule & Kendall consideră în [32] că "în mâinile unui om lipsit de experiență metodele statistice sânt cele mai periculoase instrumente de lucru". Puține sunt metodele care au o aplicație mai larg răspândită, și nici o altă metodă nu necesită mai multă grijă în aplicarea sa. Statistica este una din acele științe ai cărei slujitori trebuie să dea dovadă de o adevărată artă a discernământului". (Sublinierile ne aparțin).

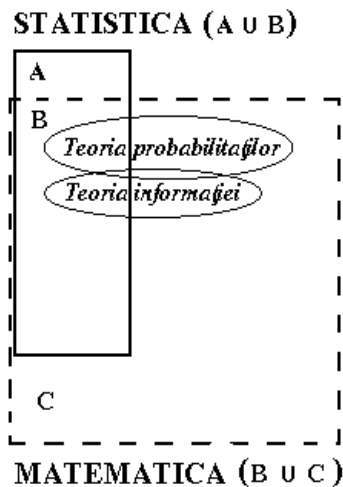
De aceea statistica, aflată pe mâna unor novici sau a unor statisticieni escroci, poate produce statistici în sensul din motto: minciuni hipergogonate, cu iz de susținere științifică.

§ 1.1. Statistica și matematica

Statistica se bazează pe:

- ◆ **reguli empirice și postulate** (plasate, în diagramă, în zona A);
- ◆ **discipline matematice** (care intersectează zona B):
 - *teoria probabilităților*
 - *teoria (matematică a) informației*
 - *geometrie euclidiană, algebră liniară, topologie, logică matematică*

Zona C conține discipline matematice care NU operează, în mod uzual, în cadrul statisticii, cum ar fi de exemplu: ecuațiile diferențiale, analiza funcțională, funcțiile complexe etc.



În diagramă am figurat numai *teoria probabilităților* și *teoria informației* deoarece, în cuprinsul acestei lucrări, se vor utiliza elemente aparținând doar acestor teorii matematice.

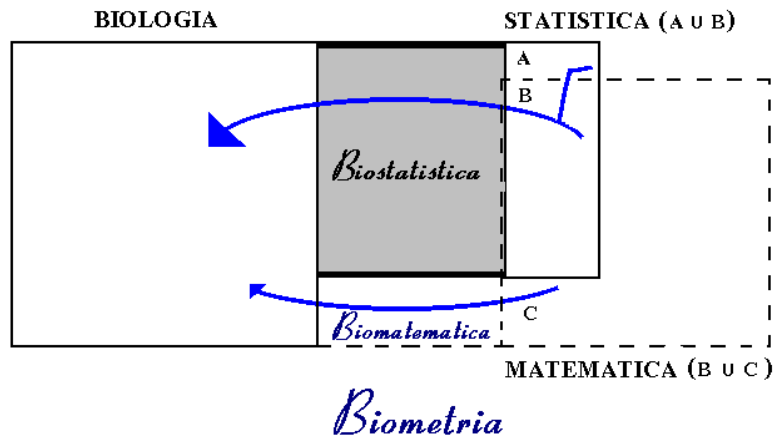
Desenul arată, totodată, că:

- ✓ statistică nu înseamnă numai teoria probabilităților;
- ✓ zona B este un subdomeniu al matematicii denumit "statistică matematică";
- ✓ zona A este mult mai mică decât zona B. De aceea zona A este adesea ignorată ajungându-se astfel la concluzia eronată că întreaga statistică (nu numai statistica matematică) este un subdomeniu al matematicii.

§ 1.2. Biostatistica, biomatematica, biometria

Biostatistica = știința aplicării în biologie a metodelor statistice

Biomatematica = știința aplicării în biologie a metodelor matematice care nu aparțin statisticii matematice.



Biometria = termen care include atât *biostatistica* cât și *biomatematica*.

- ✓ Termenul de *biometrie* este utilizat de către biologi și în sensul de știință a măsurării variabilelor biologice. Aceasta este conturată pe domenii biologice specifice: *antropometrie, dendrometrie, etc.*
- ✓ *Biomatematica* conține, de exemplu, știința aplicării anumitor structuri algebrice în genetică sau a anumitor ecuații diferențiale sau cu derivate parțiale în ecologie etc. La prima Conferință Europeană de Biometrie (Budapesta 1990) s-a constatat că biometricienii capătă abilitățile necesare după circa 10 ani de practică și că, de regulă, sunt dublu specializați într-o subdisciplină matematică sau în statistică și una biologică.

1.2.1. Comentarea definițiilor

◆ Fiind știința aplicării statisticii și matematicii în biologie, *biometria* este un domeniu interdisciplinar plasat între biologie, pe de o parte, și statistică și matematică, pe de alta - a se vedea figura anterioară. În consecință, *biometria* nu face parte nici din biologie și nici din statistică sau din matematică. De aceea, denumirea de *Statistică biologică* ni se pare neprotrivită, căci induce în planul logicii noastre ideea că biostatistica ar fi o specie (o ramură) a statisticii. Această confuzie poate produce consecințe grave asupra calității învățământului și cercetării, prin apelul la specialiști nepotriviți (statisticieni, nu biostatisticieni). Analog, denumirea de *Statistică ecologică* ar trebui readusă la denumirea de *Ecostatistică*, în România, botezată inspirat de A. Vădineanu, la începutul anilor 90.

Biologia studiază fenomenele vieții. *Statistica și matematica* construiesc procedee generale de gândire, iar *biometria* este știința utilizării adecvate și optime a acestor procedee, în biologie. Rezultă că învățarea biometriei trebuie începută prin cunoașterea procedeelelor respective, adică prin elemente de statistică și de matematică strict necesare.

Cerința de mai sus creează adesea impresia că biometria este inclusă în statistică și în matematică, când, de fapt, biometria doar utilizează elemente din aceste discipline. Specificul său este capacitatea de a le utiliza în biologie. Acest lucru se obține foarte greu, doar după o îndelungată practică și bazat pe un talent specific. De aceea, o definiție provizorie mai bună a biometriei ar putea fi:

Biometria este “arta” de a utiliza în biologie, în mod adecvat, cât mai aproape de optim sau chiar inspirat, metode statistice sau matematice.

În desenul anterior se observă că săgeata care vizualizează aplicarea metodelor statistice pornește din ambele zone, *A* și *B*. Prin această reprezentare am vrut să subliniem faptul că nici-o metodă statistică nu poate fi aplicată în biologie fără a face apel la cel puțin o regulă empirică sau postulat epistemologic - epistemologia este teoria cunoașterii științifice. Reguliile sau postulatele pot fi generale, specifice biologiei ori chiar unui anumit subdomeniu al acesteia.

Săgeata corespunzătoare biostatisticii este mult mai mare decât săgeata pentru biomatematică, pentru a exprima faptul că metodele statistice sunt mult mai mult utilizate decât cele matematice nestatistice. Aceasta și din cauza faptului că orice metodă matematică nestatistică presupune, pentru verificarea sa în practică, apelul la metode statistice.

◆ Definiția “clasică” dată biometriei de K. Pearson este “știința *aplicării* metodelor matematice (statistice) la studiul multiplelor aspecte ale vieții” [2]. Deși este concisă și sugestivă, această definiție poate lăsa impresia greșită că întregul demers statistic este inclus în cel matematic.

Vom vedea în subparagraful 1.2.4., că o anumită direcție a biostatisticii este strict necesară oricărui biolog, în timp ce biomatematica este adresată aproape în exclusivitate “elitelor”.

1.2.2. Mecanismul apariției și dezvoltării biometriei

Construcția matematicii s-a produs și se produce în continuare în următoarele etape care se reiau fie prin *aprofundare*, fie *schimbând domeniul experimental* de inspirație și/sau aplicare: (1) pornește de la cerințe ale unui domeniu experimental, apoi (2) devine “artă pentru artă” - dezvoltată de matematicieni - și în final (3) se aplică spectaculos în domeniul experimental respectiv. De cele mai multe ori, mai devreme sau mai târziu, noile metode sau teorii se aplică și în alte domenii experimentale.

Raymond Queneau, citat în [24], observă că orice știință experimentală trece prin patru faze de aprofundare a raportului cu matematica: “faza **empirică**, atunci când faptele se *enumără*, faza **experimentală**, când se *măsoară*, **analitică**, când se *calculează* și **axiomatică**, când se *deduc*”.

Până acum cca. 30 de ani principalul domeniu experimental de inspirație și aplicare a matematicii a fost fizica, în special mecanica generând, de exemplu, calculul diferențial și integral. Datorită vechimii acestui “mariaj”, precum și a simplității fenomenelor fizice în raport cu cele biologice, fizica a parcurs toate cele 4 faze, fizica teoretică ilustrând etapa axiomatică. În ultimele decenii, asistăm la “UN NOU MARIAJ”, după cum se afirmă în literatură, și anume între matematică și biologie [26], generând biometria. “Cele mai interesante probleme sunt un amestec de biologie cu matematică neconvențională”, spune un biometrician american. Revenind la desenul anterior observăm că deoarece biometria conține și demersul invers vor trebui marcate săgețile și în sens invers. Adică va trebui modificată definiția biometriei astfel:

Biometria este *știința aplicării* în biologie a metodelor statistice sau matematice precum și cea a *descoperirii / inventării* de noi metode statistice sau matematice generate de probleme puse de biologie.

Revenind la etapizarea de mai sus, biologia se poate plasa, în momentul de față, doar prin anumite subdiscipline ale sale, cel mult în faza analitică.

Pe de altă parte însă, există transferuri reciproce între matematică și biologie și câștiguri ale aplicării în alte științe a matematicii generate de biologie.

De exemplu "teoria catastrofelor", o matematică generată de biologie, din probleme de morfogenează, explică în orice domeniu "catastrofele". Prin "catastrofă" se înțelege producerea unui efect major, catastrofal cu un efort minor, infinitezimal. Un alt exemplu este "taxonomia numerică" domeniu matematic generat de taxonomie și sistematică care în prezent are aplicații fundamentale în inteligența artificială.

1.2.3. Caracterul interdisciplinar al biometriei și abilitățile unui biometrician

◆ Caracterul interdisciplinar al biometriei înseamnă că între statistică, matematică și biologie există transferuri conceptuale cu câștiguri metodologice reciproce, nu numai o simplă aplicare a unei discipline într-o altă disciplină. Vom ilustra acest lucru în mai multe locuri, unul din cele mai importante fiind o propunere inspirată din biologie, de mod de abordare statistică a variabilității caracteristicilor calitative. Acestea vor fi tratate, în §3.8., atât cantitativ cât și calitativ – distincție insuficient conturată în literatură.

În practică, este foarte important să se observe că pentru aplicarea adecvată a metodelor statistice sau matematice în biologie este mai mult decât recomandabil apelul la biometricieni, iar rezolvarea eficientă* - eventual prin noi matematici - se poate obține numai prin intermediul biometricienilor. În acest caz, aceștia vor constitui o interfață între specialiștii în științele vieții și matematicienii care vor dezvolta matematica necesară.

* "**Acțiune efectivă** = acțiune cu efect. **Acțiune eficace** = acțiune *efectivă* în sensul dorit. **Acțiune eficientă** = acțiune *eficace* optimă (raport (efect / efort) maxim)." V. Săhleanu. Observăm că aceste diferențieri, în limba engleză sunt valabile.

◆ Pentru a nu se confunda matematicienii sau statisticienii cu biometricienii putem face următoarele analogii. Primii pot fi asimilați cu inginerii care construiesc aparatură chirurgicală, iar biometricienii cu chirurgii. Oricine poate realiza ce se poate întâmpla dacă în locul chirurgilor vom apela la ingineri sau invers. Continuând analogiile, biologul – în sensul cel mai larg, de specialist în științele vieții – este medicul generalist, iar problemele biologice constituie pacienții.

Analogiile pot merge mai departe, căci biometricienii au, ca și chirurgii, câte o dublă specializare. Una legată de subdomeniul matematic iar cealaltă de subdomeniul de aplicare. De exemplu, statistică multivariată în antropologie, sau analiza varianței în ecologie, algebră în genetică, statistică neparametrică în cercetarea clinică etc. La chirurgi dubla specializare provine din instrumentarul folosit și din zonele din organism operate. De exemplu, chirurgie laser pe encefal, chirurgie laparoscopică pe esofag etc.

1.2.4. Importanța biostatisticii pentru biolog

Biostatistica este pentru biolog - în sens larg - un instrument:

- 1) **strict necesar** pentru verificarea ipotezelor de specialitate (prin *statistica inductivă* - singurul instrument științific de generalizare a observațiilor) și
- 2) *util* și *modern* pentru sugerarea de ipoteze de specialitate (în special prin *statistica descriptivă multivariată* - "unul dintre rarisimele instrumente de descoperire" [1]).

Altfel spus, statistica inductivă este un instrument obligatoriu pentru orice biolog, în timp ce statistica multivariată este un domeniu opțional. În ambele situații elementul cheie este capacitatea de colaborare cu biostatisticianul.

- ✓ Capitolele moderne ale statisticii multivariate fiind de o mare complexitate, nu pot fi abordate în acest cadru.
- ✓ Viitorul biolog trebuie să rețină că în carieră va colabora cu biostatisticieni, iar pentru acesta are nevoie de o pregătire elementară de biostatistică. Continuând analogiile anterioare, putem spune că orice viitor medic generalist are nevoie de un stagiu de chirurgie.

- ✓ Învățarea bio/eco-statisticii presupune un dublu efort desfașurat în timp suficient pentru rezultate de calitate: (1) asimilarea **modului de gândire specific statisticii** [ca o combinație între (a) statistica matematică și (b) regulile empirice proprii] și (2) învățarea unei „**inginerii**” **specifice aplicării gândirii statistice în bio/eco-logie**. În deceniul trecut se alocău două semestre ecostatisticii și s-au format absolvenți competitivi mondial (vezi http://www.liviu-dragomirescu.ro/Eco_num/Eco_num_files/prefata.pdf)

§ 1.3. Schema metodologică a apelului la biostatistică

Putem formula următoarele reguli cu privire la aplicarea biostatisticii (folosind, eventual, și unele programe de calculator):

REGULA 1: Metodele de prelucrare a datelor se stabilesc înaintea efectuării observației sau experimentului, deși se aplică după acestea.

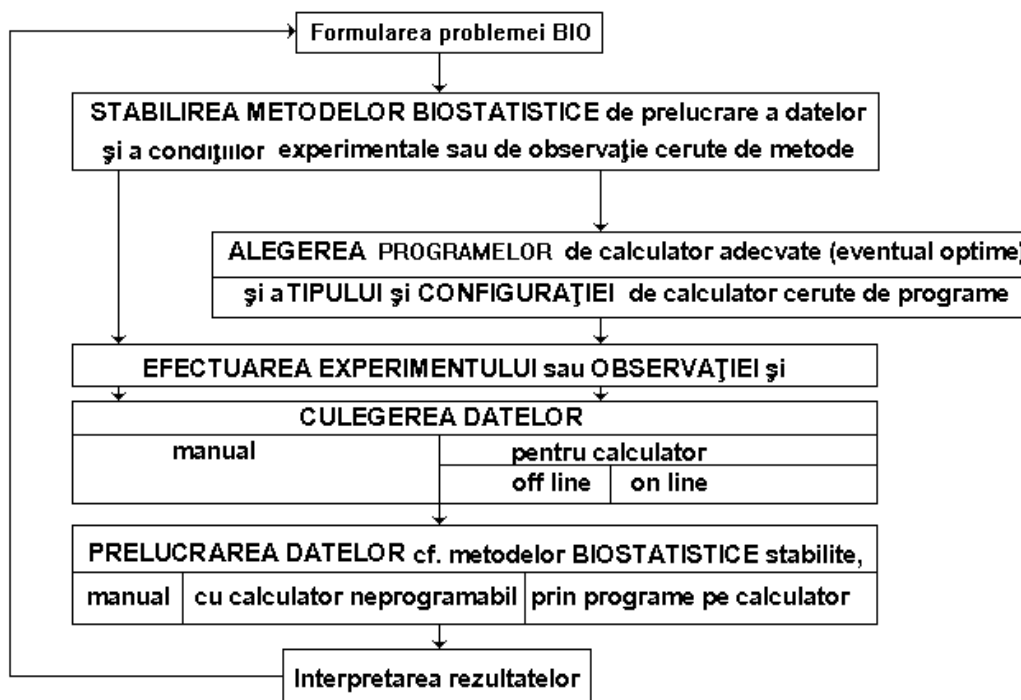
Trebuie procedat astfel deoarece orice metodă are anumite exigențe asupra datelor, exigențe ce trebuie respectate în timpul observației sau experimentului.

REGULA 2: Consultăm un biometrician înainte de a începe cercetarea și nu după efectuarea ei.

REGULA 3: După stabilirea modului de organizare a experimentului sau observației și a metodelor de prelucrare a datelor, alegem cu consultația biometricianului **softul** (programele de calculator) **adecvat**.

Acesta este proiectat pentru un anumit tip de calculator și o anumită configurație a acestuia. Cumpărăm, deci, calculatorul și softul necesare înainte de culegerea datelor și numai după stabilirea metodelor prin care vor fi prelucrate datele.

Schema CORECTĂ metodologică a utilizării biostatisticii (și, eventual, a programelor de calculator) este următoarea:



Schema metodologică cel mai des aplicată este însă următoarea succesiune:

- Formularea problemei BIO.
- Efectuarea experimentului sau observației.
- Culegerea datelor.
- Prelucrarea datelor prin metode biostatistice.
- Interpretarea rezultatelor.

În acest caz prelucrarea datelor se face adesea prin una din următoarele variante de nedorit:

1. prin metode adecvate, dar simpliste \Rightarrow rezultate slabe, insuficiente;
2. metode adecvate, dar care cu doar câteva observații în plus ar fi putut aduce rezultate semnificative \Rightarrow rezultate neoptimale și insuficiente;
3. metode inadecvate, datele experimentale sau de observație neîndeplinind condițiile cerute de metodă (de exemplu, pentru că:
 - a. pentru aceste metode avem programe de calculator și/sau
 - b. nu se cunosc condițiile metodei) \Rightarrow rezultate eronate.

Cititorul care va studia această carte va fi capabil să găsească singur exemple concrete care să illustreze aceste afirmații, din păcate prea des adevărate.

+§ 1.4. Observații cu privire la softul de biostatistică

Pe piața de programe de calculator există programe care:

- ◆ sunt greșite;
- ◆ sunt relativ corecte* și
 - aplică doar metoda cerută de noi:
 - unele și atunci când datele nu se pretează la prelucrarea respectivă,
 - altele atrăgând însă atenția asupra acestui lucru;
 - aplică metoda cerută sau se orientează singure aplicând metode complementare adecvate unor particularități ale setului de date sau aplică mai multe metode relativ paralele, și atunci utilizatorul:
 - trebuie să știe să o aleagă pe cea potrivită,
 - este atenționat asupra condițiilor pe care trebuie să le îndeplinească datele pentru a se putea aplica fiecare metodă paralelă;
 - au suplimentar HELP-uri (texte de ajutor) cu succinte explicații statistice;
 - au suplimentar subprograme de învățare a utilizării produsului (denumite adesea TUTORIAL), iar acestea conțin unele elemente de metodologie statistică.

* În legătură cu *corectitudinea unui program* observăm că din nefericire (sau mai de grabă din fericire) nu pot exista programe corecte în sens absolut. Aceasta, deoarece nu există și nu poate exista o teorie matematică care să demonstreze corectitudinea unui program. Singurul lucru care se poate face este o testare cât mai bună. (Testarea poate evidenția eventuale erori, dar nu poate demonstra inexistența oricărei erori.) Deci "cele mai corecte programe" sunt programele cele mai bine și mai mult testate din punct de vedere logic și aritmetic doar, căci adecvarea la problema biologică nu poate fi niciodată evaluată doar prin testare.

Nici un program de calculator, oricât de performant ar fi, nu poate suplini nici gândirea statistică elementară strict necesară biologului și nici consultanța unui biostatistician.
