

Anexa 1. CONCEPTE STATISTICE ESENȚIALE

„...gândirea statistică va deveni într-o zi la fel de necesară pentru un cetățean eficient la fel ca și abilitatea de a citi și a scrie”.

H.G. Wells (1866-1946)

- (Caracteristici) variabile
- Serii (șiruri) statistice
- Populație statistică
- Eșantion statistic
- Randomizare
- Inferențe statistice

„**Statistica** studiază mulțimi de observații efectuate asupra unor obiecte denumite **unități statistice** care prezintă anumite **caracteristici variabile**. Unitățile statistice pot fi *clasate*, *ordonate* sau *măsurate* în raport cu caracteristicile respective. Mulțimile de observații se numesc **serii statistice** sau **șiruri statistice**”. (L4)

Numărul de observații din seria respectivă se numește **volumul seriei**.

Exemplu: Într-un spital clienții (*unitățile statistice*) au, printre altele, următoarele caracteristici:

- grupa sanguină. Această caracteristică poate fi doar *clasată* (repartizată în clasele, variantele respective: A, B, 0, AB).
- nota sau calificativul pe care îl acorda calității asistenței medicale primite. Această caracteristică poate fi clasată și *ordonată* doar.
- greutatea la internare. Această caracteristică poate fi clasată, ordonată și *măsurată*.

Măsurarea presupune posibilitatea existenței unui „unități de măsură” cu semnificație invariantă în orice context. Astfel, în cazul greutății, diferența de 1 kg între greutatea de 70kg și cea de 69kg, înseamnă același lucru cu diferența de 1 kg între oricare alte două greutăți întregi consecutive. Mai mult, o greutate de 80 kg este de două ori mai mare decât una de 40 kg, adică poate fi obținută punând pe cântar două greutăți de 40 kg. Acest lucru nu este valabil pentru nota acordată calității asistenței medicale. Aici, diferența de un punct între un client care acordă nota 5 și unul care acordă nota 4 nu are aceeași semnificație cu diferența, tot de un punct, între notele 10 și 9. Mai evident, un client care acordă nota 10 nu înseamnă că este de două ori mai mulțumit decât unul care a acordat nota 5 ori că mulțumirea lui este egală cu suma mulțumirii celor doi care au acordat, fiecare, nota 5.

Variabilele care pot fi doar clasate se numesc variabile *calitative (nominale)*. Cele care pot fi și ordonate doar se numesc *ordinale*, iar cele care pot fi măsurate se numesc variabile *cantitative (dimensiuni)*.

Exemple de variabile

- calitative: grupa sanguină, RH-ul, sexul, diagnosticul medical, specialitatea medicală, patologia, tratamentul, tipul de serviciu medical, spitalul, medicul, județul;

- ordinale: nota sau calificativul acordat asistenței medicale, ierarhia medicilor într-un spital, gravitatea unei patologii, starea de sănătate, performanța unui cadru medical;
- cantitative: greutatea, înălțimea, numărul de clienți, timpul de spitalizare, durata de funcționarea a unui aparat, salariul unui cadru medical, costul unui tratament, numărul de zile de spitalizare.

Variabilele calitative au *variante*, cele ordinale, *ranguri* iar cantitativele au *valori* (numere).

Variabilele cantitative pot fi tratate drept ordinale, iar cele ordinale drept calitative.

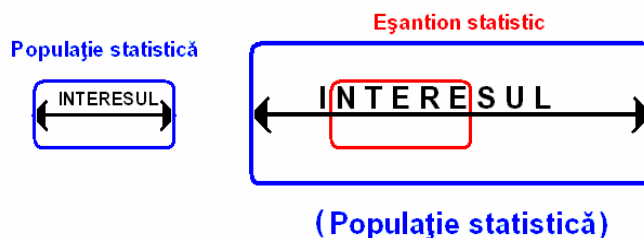
Seriile statistice pot fi univariate, bivariate ori multivariate, după numărul de variabile luate în considerație simultan.

Exemple de serii statistice

- {gripă, TBC, gripă, SIDA, TBC} serie statistică *univariată* de volum 5, a variabilei calitative „patologia”, în care au apărut doar variantele: gripa, TBC și SIDA
- $\left\{ \begin{array}{l} \text{slabă medie slabă bună} \\ \text{slabă slabă medie bună} \end{array} \right\}$ serie statistică *bivariată* de volum 4, a perechii de variabile ordinale „calitatea asistenței medicale în salonul X”, evaluată de 4 clienți la o săptămână de la internare (seria de sus), respectiv aceeași variabilă evaluată de aceeași 4 pacienți după încă o săptămână (seria de jos) în care s-a aplicat un program pentru creșterea calității asistenței medicale. În ambele variabile au apărut rangurile „slabă”, „medie”, „bună”.
- un chestionar de audit care conține variabilele sex, vârsta, diagnostic, medic curant, nota acordată calității îngrijirii medicale, aplicat mai multor clienți formează o serie *multivariată*.

Seriile statistice univariate se pot sintetiza grafic în reprezentări grafice adecvate: *diagrame circulare* pentru cele calitative și *diagrame în batoane* și *poligoane de frecvențe* pentru cele ordinale și cantitative. Pentru cele cantitative grupate pe intervale de grupare se utilizează *histograme* (vezi 2.1 în anexa 2.)

Spunem despre o serie statistică că reprezintă o populație statistică dacă interesul nostru se limitează la seria respectivă (de fapt, la mulțimea respectivă de unități statistice). Spunem că reprezintă un eșantion statistic dacă este o parte a unei populației statistice la care se extinde interesul nostru (L4).



O populație statistică poate fi studiată exhaustiv, complet (caz relativ rar sau chiar imposibil) sau incomplet prin intermediul unui eșantion (cazul cel mai frecvent). În primul caz, rezultatele obținute sunt certe, iar în al doilea caz sunt incerte (cu privire la populația statistică respectivă). Extrapolarea rezultatelor de la eșantion la populația de origine este, din punct de vedere logic, o inducție (o „generalizare”). Inducția poate fi

completă ori incompletă. Numai în cadrul științelor formale (ca matematica, logica) se pot face inducții complete♦. Prin acestea se obțin rezultate certe sau, altfel spus, cu un grad de certitudine unitar (sau 100%). În cadrul științelor, cercetărilor experimentale sau de observație nu pot fi realizate decât inducții incomplete, deci incerte sau, altfel spus, cu grad de certitudine subunitar (sau < 100%). Gradul de certitudine al unei generalizări din științele experimentale sau de observație poate fi nespecificat ori specificat. Desigur, sunt preferate cele cu grad specificat (ca un plus de științificitate). Gradul de certitudine specificat poate fi o probabilitate subiectivă ori una obiectivă. Probabilitățile subiective sunt acordate subiectiv de experți ai domeniului experimental respectiv. În cele mai multe situații concrete, aceste probabilități subiective nu pot fi verificate practic. În contrast, probabilitățile obiective pot fi verificate experimental. Aceste probabilități obiective se pot obține numai (a) dacă informațiile din eșantioane sunt prelucrate prin metodele statisticii inductive și (b) eșantioanele s-au obținut ca și când ar fi fost extrase dintr-o „urnă bine amestecată” (pentru ca să se asigure aceeași șansă oricărui eșantion de volum fixat). Asemenea eșantioane sunt numite eșantioane aleatoare (simple) și se obțin printr-o procedură numită *randomizare*. Se acceptă și așa numita *pseudorandomizare* care înseamnă generarea cu ajutorul rutinelor specializate ale calculatorului a n numere pseudoaleatoare mai mici sau egale cu N , în care n este volumul eșantionului iar N este volumul populației statistice. După care elementele populației se numerotează de la 1 la N și sunt selectate în eșantion elementele cu numerele stabilite de calculator.

Condiția ca extragerea eșantionului să fie aleatoare, întâmplătoare este esențială deoarece numai în acest fel rezultatele practice se suprapun perfect cu teoria. O extragere aleatoare se obține la fel de greu ca și o sterilizare. Cel mai corect ar fi să spunem că eșantionul trebuie extras « prin randomizare ». Aceasta înseamnă că trebuie să apelăm la o procedură care ne asigură că în selectarea unităților cuprinse în eșantion intervine doar pura întâmplare și, teoretic, nici-o intervenție umană. Altfel, este ca și cum am fi extras eșantionul dintr-o « urnă măsluită », dar într-un fel necunoscut de nimeni. Adică, « ne furăm singuri căciula ». Continuând analogia, extragerea eșantionului fără a se face randomizare va avea consecințe analoge unei operații făcute fără sterilizare.

Prin a și b de mai sus se obțin așa numitele inferențe statistice (testele statistice și estimările statistice) care sunt metodele de inducție, ce-i drept, incompletă, dar care sunt singurele ce pot asigura maximum de științificitate demersului de generalizare în științele experimentale.

Rezumând cele de mai sus, putem spune că așa cum în științele formale gradul superior de științificitate este asigurat de inducția completă, în științele experimentale sau de observație gradul superior de științificitate este dat de aplicarea corectă (eșantioane prin randomizare) a gândirii statistice (inferențele statistice).

Metodele statistice care se opresc la orizontul seriei de date disponibile formează așa numita statistică descriptivă. Cele care se ocupă de extrapolarea (generalizarea, inducția) rezultatelor obținute pe datele empirice la populații statistice din care au fost extrase sau presupunem că au fost extrase alcătuiesc statistica inductivă. Aceasta este denumită și *statistică inferențială, analitică sau matematică*.

♦ „Gustul” inducției complete se poate simți ușor dacă ne reamintim una din celebrele butade ale lui Moisi: „Orice om are dreptul la un pahar cu vin. Dar după ce bei un pahar cu vin ești un alt om.” Deci ai dreptul la încă un pahar cu vin și așa mai departe. Rezultă, în final, că orice om are dreptul la o infinitate de pahare cu vin.