Capitolul 2

CONCEPTE STATISTICE DE BAZĂ

Conceptele statistice fundamentale, prezentate în acest capitol, sunt:

- ♦ *unitate statistică*
- ♦ *caracteristică variabilă* sau *variabilă*,
- ♦ *şir statistic* sau *serie statistică* şi *distribuţie de frecvenţe*
- ♦ *populaţie statistică*
- ♦ *eşantion*
  - *eşantioane independent prelevate*
  - *eşantioane de observaţii perechi*

Statistica studiază mulţimi de observaţii efectuate asupra unor obiecte denumite **unităţi statistice** care prezintă anumite **caracteristici variabile**. Unităţile statistice pot fi *clasate, ordonate* sau *măsurate* în raport cu caracteristicile respective. Mulţimile de observaţii se numesc **şiruri** sau **serii statistice**.

Exemplul 2.

Într-un acvariu, peştii (*unităţile statistice*) prezintă, printre altele, următoarele *caracteristici*:

- Specia. Datele acestea se *clasează*.
- O notă de frumuseţe a exemplarelor. Se *ordonează*.
- Lungimea sau greutatea. Se *măsoară*.

§ 2.1. Clasificări ale variabilelor

2.1.1. Clasificarea grosieră

1. **Variabile (pur) <u>calitative</u>** sau **nominale** sau **categoriale** = variabile ale căror *<u>variante</u>* pot fi (doar) clasate (nu ordonate sau măsurate).
   Exemple:
   Variabila culoarea ochilor cu *variantele* negri, căprui, albaştri, verzi.
   Variabila sex cu *variantele* masculin, feminin.
2. **Variabile <u>cantitative</u>** = variabile ale căror *<u>valori</u>* pot fi ordonate sau chiar măsurate. Cele care pot fi doar ordonate se numesc şi **semicantitative** sau **ordinale**, iar valorile respective se numesc **ranguri**.
   Exemple:
   Variabile semicantitative: ierarhia taxonomică, ierarhia într-un grup de cimpanzei.
   Variabile cantitative: greutatea, înălţimea, circumferinţa toracică la oameni.

2.1.2. Clasificarea duală a lui Anderberg

Anderberg a clasificat variabilele după (1°) *mulţimea* („cardinalul mulţimii") *de reprezentare* şi după (2°) *scala de reprezentare* („scala de măsurare") [1].

1° După mulţimea de reprezentare variabilele pot fi:

- ♦ *continue*. Acestea iau, evident, numai un număr *infinit şi nenumărabil* de valori într-un interval (a, b) sau o reuniune de intervale.
- ♦ *discrete* sau *discontinue*:
  - *finite.* Mulţime de forma $\{a_1, a_2, \dots, a_n\}$.

- *infinite și numărabile.* Mulțime de forma $\{a_1, a_2, \dots, a_n, \dots\}$.

♦ *binare* sau *dihotomice.* Mulțime de forma $\{a_1, a_2\}$.

2° Scalele de reprezentare

„Fie două obiecte (unități statistice n.t.) A şi B şi o variabilă X care are, pentru cele două obiecte. scorurile x_A, respectiv, x_B.

a. O **scală nominală** face doar distincție între clase. Astfel, pentru A şi B se poate spune doar că $x_A = x_B$ ori $x_A \neq x_B$.
b. O **scală ordinală** induce o relație de ordine între obiecte. Astfel, în plus față de diferențierea între $x_A = x_B$ şi $x_A \neq x_B$, în cazul inegalității se poate spune dacă $x_A > x_B$ ori $x_A < x_B$.
c. O **scală interval** atribuie o măsură semnificativă a diferenței între două obiecte. Se poate spune, nu numai că $x_A > x_B$ ci şi că A este **cu** $x_A - x_B$ unități mai mare ca *B*.
d. O **scală raport** este o scală interval cu un punct zero semnificativ. Dacă $x_A > x_B$ se poate spune că A este **de** x_A / x_B ori mai mare ca *B*.

Aceste definiții de scale sunt ordonate ascendent de la cea nominală la scala raport. Fiecare scală are toate proprietățile tuturor scalelor inferioare ei. Prin urmare, renunțând la o parte din informație se poate reduce o scală la o oricare scală inferioară…

Variabilele de pe o scală nominală sau ordinală sunt denumite, în mod frecvent, variabile *categoriale* sau *calitative,* adeseori cu ambiguitate asupra faptului dacă există sau nu o relație de ordine. Pentru contrast, variabilele pe scale interval sau raport sunt adesea referite drept variabile *cantitative*." [1] (Sublinierile ne aparțin.)

Observăm că o scală interval are o origine arbitrară (zero) şi, deci, poate permite valori negative. În schimb, o scală raport are un zero „absolut" şi, deci, nu permite valori negative.

Exemplele 2.1.2.

a. Rasa, sexul, culoarea părului, specia, varietatea, tratamentul.

b. Scala durității mineralelor a lui Mohs: "mineralul *A* este mai dur decât *B* dacă *A* zgârie pe *B*", ierarhia militară, nota (pentru că putem spune că un student care a primit nota 10 este mai pregătit decât unul cu 9, dar nu putem spune care este diferența. Dacă nota ar fi reprezentabilă pe o scală interval ar însemna că am putea afirma că diferența de pregătire între cei doi – de un punct – este egală cu diferența de pregătire între un student cu nota 5 şi unul cu nota 4 – tot de un punct. Aceste „puncte", „intervale" nu sunt egale.

c. Anul calendaristic după Cristos (pentru că există şi ani înainte de Cristos), temperatura în grade Celsius sau în grade Fahrenheit (pentru că nu au 0 absolut).

d. Greutatea, înălțimea, temperatura în grade Kelvin, (pentru că au 0 absolut şi, deci, nu pot avea valori negative). Putem spune că $20°K$ indică o temperatură dublă față de cea de$10°K$, deoarece aceste valori exprimă dublarea gradului de agitație moleculară (ceea ce nu se poate afirma în cazul gradelor Celsius sau Fahrenheit).

Observăm că **a.** şi **b.** sunt utilizate cu mulțimi de reprezentare discrete iar **c.** şi **d.** pot utiliza atât mulțimi discrete cât şi continue.

Măsurătorile reprezentate prin numere întregi (0, 1, 2, … , n), altfel spus, **contorii**, aparțin scalei raport. În biologie, *contorii* se numesc **variabile meristice**.

Exemple: număr sepale, număr petale.

c. şi **d.** se numesc **scale metrice** sau **parametrice**, iar variabilele respective se mai numesc şi **parametri** (<u>în sensul larg al cuvântului</u>, deoarece vom vedea că există şi un sens restrâns). În contrast, **a.** şi **b.** se numesc **scale nonmetrice** sau **neparametrice**.

+3° Transformări permise în cadrul fiecărei scale

a. Permutarea şi redenumirea.

Exemplu: Sex *M*, *F* sau *F*, *M* (*permutare*) sau 1, 2 (*redenumire*).

b. Orice funcție $f(x)$ strict crescătoare.

Exemple: $\log_a x$ cu $a > 1$; reținerea rangurilor în locul valorilor.

c. $f(x) = m \cdot x + n$, cu $m \neq 0$. $m > 0$ păstrează aceiaşi ordine; $m < 0$ inversează ordinea.

Exemplu: $T[°K] = t[°C] + 273{,}15$ (deci $m = 1$ şi $n = 273{,}15$)

d. $f(x) = m \cdot x$, cu $m > 0$.

Exemple: $g[\mu g] = 1000\ G[mg]$; $L[inch] = 2{,}54\ [cm]$.

2.1.3. Clasificarea pentru această lucrare

Clasificarea pentru această lucrare este prezentată în prima coloană a tabelului 2.1.

Tabelul 2.1. Clasificarea pentru această lucrare

Variabilă	conține			reprezentabilă pe o scală cel mult	scalele fiind
♦ **calitativă**	*variante*	–	în nr. finit	*nominală*	*nonmetrice, neparametrice*
♦ **ordinală**	*ranguri*	–	discontinue	*ordinală*	
♦ **cantitativă (** • discontinuă: **numărătoare** sau **contor** [31] • continuă: **măsurătoare** sau **dimensiune)**	*valori*	–	discontinue continue	*interval* sau *raport*	*metrice, parametrice*

Variabila ordinală are o situație specială deoarece *rangurile* pot proveni fie din *variante* înzestrate cu o relație de ordine, fie din *valori* pentru care ignorăm proprietățile scalei interval.

Exemple

- Pentru prima situație, variantele "mic, mediu, mare" primesc în mod natural o relație de ordine. „Glosarul IIS[1] de termeni statistici" numeşte aceste date „categorizare ordonată".
- Pentru a doua situație putem considera situația primilor trei clasați la un concurs sportiv. Ei se plasează astfel pentru că au obținut cele mai mari scoruri, fără a conta însă valorile propriu-zise ale scorurilor, mai precis, intervalele dintre acestea - proprietatea definitorie a scalei interval. „Glosarul IIS de termeni statistici" numeşte această procedură „statistică de ordine".

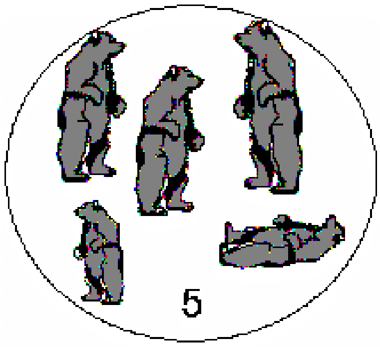
Variabila cantitativă poate fi atât o *numărătoare* cât şi o ***măsurătoare***. O **numărătoare, un contor,** o variabilă meristică, conține numai valori discontinue, discrete, rezultatul unei numărări fiind întotdeauna un număr întreg pozitiv. În mod contrar, o **măsurătoare** produce ca rezultat o măsură în sensul comun cunoscut - "valoare a unei mărimi, determinată prin raportare la o unitate dată" [18]. În consecință, poate lua atât valori întregi cât şi fracționare. De aceea, spunem că o măsurătoare conține valori continue, ca o **dimensiune** din fizică.

Exemple

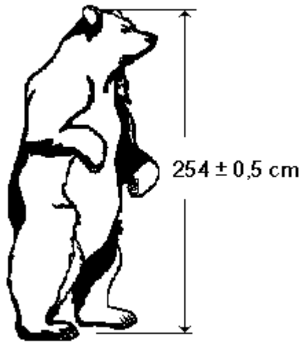
- Numărul de petale este o numărătoare. Rezultatul este întotdeauna un număr natural (întreg pozitiv).
- Înălțimea sau greutatea unui om sunt măsurători. Ambele produc numere întregi sau fracționare prin raportarea la unități arbitrare şi generale stabilite convențional, de exemplu centimetru, respectiv kilogram.

[1] Institutul Internațional de Statistică

O *variabilă cantitativă* produce o *valoare* denumită şi *mărime.* De exemplu, în limbaj curent, se vorbeşte atât de *mărimea unei populații* de animale, cât şi de *mărimea unui animal.* În primul caz - fiind vorba de numărul de animale din populația respectivă - avem de a face cu o numărătoare, deci cu o variabilă discretă, iar în al doilea caz este vorba de o măsurătoare, adică de o variabilă continuă.



Mărimea acestei populații este de **5** exemplare.



Mărimea acestui exemplar este de **254 ± 0,5 cm**.

Deci, o *numărătoare* produce un rezultat exact – un număr întreg pozitiv. În schimb, o *măsurătoare* produce un rezultat aproximativ, eventual cu grad de aproximare cunoscut, adică o *aproximare exactă.* În ultimul caz, rezultatul va fi format din două numere, întregi sau fracționare, primul reprezentând **rezultatul aproximativ al măsurării**, iar al doilea, **mărimea erorii de (aproximare prin) măsurare.** Acesta din urmă este, întotdeauna, un număr real strict pozitiv.

Exemplu

Înălțimea unui anumit exemplar din specia *Ursus arctos* este 254 cm ± 0,5 cm.

Mărimea erorii de măsurare (± 0,5 cm în exemplul nostru) exprimă în mod invers **precizia măsurării** şi este numită şi **marjă de eroare** (**de măsurare**). Fără specificarea marjei de eroare rezultatul unic este neştiințific.

O *variabilă cantitativă* este denumită şi **parametru (în sens larg)**.

§ 2.2. Clasificări ale şirurilor statistice

2.2.1. În funcție de ordinea elementelor în şir

există:

- şiruri în care ordinea elementelor nu contează;
- şiruri, serii în care ordinea elementelor este conformă unei succesiuni:
  - temporale. Acestea se numesc **serii temporale** sau **serii cronologice**.
  - spațiale.

În această lucrare nu vom trata decât prima categorie de şiruri. Deoarece ordinea elementelor nu are importanță, şirul este - din punct de vedere matematic - o mulțime finită. De aceea, vom utiliza, pentru descrierea unui şir, notația tradițională a unei mulțimi definite sintetic: două acolade care cuprind elementele şirului scrise în orice ordine. Elementele sunt separate prin virgulă sau punct şi virgulă, după caz.

2.2.2. În funcție de numărul de variabile luate în considerație simultan

există:

a. şiruri statistice *univariate*;

b. şiruri statistice *bivariate*;

c. şiruri statistice *multivariate* (cu n variabile, $n > 2$).

Exemplele 2.2.2

Revenind la exemplul 2,

a. {crap, caras, crap, crap}, {7, 9, 6, 8} şi {1,2 Kg; 0,3 Kg; 0,7 Kg; 1,5 Kg} sunt trei şiruri statistice de specii, note, respectiv, greutăţi;

b. $\left\{\begin{array}{cccc} \text{crap} & \text{caras} & \text{crap} & \text{crap} \\ 7 & 9 & 6 & 8 \end{array}\right\}$ este un şir statistic de specii şi note;

c. $\left\{\begin{array}{cccc} \text{crap} & \text{caras} & \text{crap} & \text{crap} \\ 7 & 9 & 6 & 8 \\ 1{,}2\text{Kg} & 0{,}3\text{ Kg} & 0{,}7\text{Kg} & 1{,}9\text{Kg} \end{array}\right\}$ este un şir statistic de specii, note şi greutăţi.

Dacă un şir nu conţine decât o dimensiune, două dimensiuni sau $n > 2$ dimensiuni, se poate numi şir **unidimensional**, respectiv **bidimensional**, respectiv **multidimensional**.

Exemplu

Şirul care conţine caracteristicile (specie, notă de frumuseţe, lungime şi greutate) pentru peştii observaţi şi măsuraţi este un şir multivariat, nu multidimensional, deoarece doar lungimea şi greutatea sunt dimensiuni.

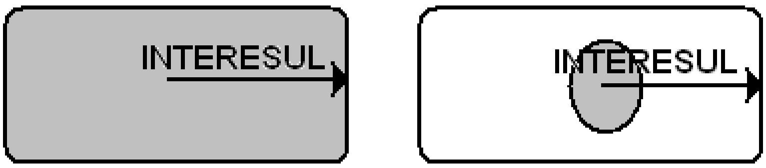
Putem considera că, în gândirea statistică actuală, coexistă două moduri fundamental diferite de abordare. Le-am denumit [11] *statistică clasică*, respectiv, *statistică modernă*.

Statistica clasică este preponderent uni şi bivariată şi se bazează pe teoria probabilităţilor. **Statistica modernă** este esenţialmente multivariată, se bazează pe geometrie, algebră şi logică formală, mai mult decât pe teoria probabilităţilor şi se dezvoltă puternic datorită informaticii.

Prezenta lucrare va aborda numai modul clasic în care este esenţială următoarea clasificare.

§ 2.3. Clasificarea mulţimilor de unităţi statistice şi structura statisticii clasice

În funcţie de orizontul nostru de interes o mulţime de unităţi statistice poate fi considerată fie *populaţie statistică*, fie *eşantion* dintr-o anumită populaţie statistică.



Populaţie statistică   Eşantion statistic

Populaţie statistică, **colectivitate statistică**, **univers statistic** sau, uneori, **lot** = o mulţime de unităţi statistice la care se limitează orizontul nostru de INTERES.

O *populaţie statistică* poate fi alcătuită din obiecte, indivizi umani ori dintr-o altă specie, fenomene, evenimente, idei, opinii, numere etc. Poate fi *finită* sau *infinită*. Poate fi *reală* ori *ipotetică*.

O populaţie ipotetică este numită şi **populaţie teoretică** deoarece este concepută, de regulă, prin intermediul unei teorii *matematice*, *statistice*, *biologice* etc.

Exemple

- Mulțimea numerelor reale, utilizată în statistică ca suport al rezultatelor posibile ale procesului de măsurare, este o mulțime teoretică *matematică*.
- Mulțimea tuturor submulțimilor cu un număr fixat de unități statistice posibil a fi extrase dintr-o anumită mulțime de unități statistice este o mulțime teoretică *statistică*.
- Mulțimea exemplarelor posibile dintr-o anumită specie este o mulțime teoretică *biologică*.

✓ Conceptul de populație statistică este diferit de cel de populație biologică de organisme vii care reprezintă un grup de indivizi din aceeaşi specie locuind în acelaşi areal [6]. Astfel, orice populație biologică poate fi considerată populație statistică, dar nu şi invers.

Studiul complet, exhaustiv al *populațiilor statistice reale* este obiectivul **statisticii descriptive**.

Studiul *populațiilor statistice teoretice* este un obiectiv al **teoriei probabilităților**.

Populațiile statistice reale sunt în marea majoritate a cazurilor foarte mari iar cele mai multe populații teoretice sunt infinite. Deoarece nu este economic să studiem exhaustiv populațiile reale şi este practic imposibil să studiem toate unitățile statistice ale unei populații infinite se recurge la *eşantioane*.

Eşantion, **mostră**, **probă**, **colectivitate de selecție** sau **lot** = o mulțime de unități statistice la care NU se limitează orizontul nostru de INTERES, fiind o parte dintr-o populație statistică, pe care o studiem cu INTERESUL de a obține informații cu privire la întreaga populație.

Un studiu complet (exhaustiv) asupra unei populații statistice produce rezultate certe în privința populației respective. Dimpotrivă, un studiu care utilizează doar un **eşantion** oferă **rezultate** care sunt **incerte** în privința populației respective.

Altfel spus, deosebirea fundamentală între studiile exhaustive şi cele prin **eşantioane** este aceea că primele produc rezultate cu grad de certitudine unitar (1 sau 100%), în timp ce ultimele aduc **rezultate** despre populația respectivă **cu grad de certitudine** strict subunitar (**probabilitate** < 1 sau < 100%).

Extrapolarea rezultatelor obținute pe eşantioane la populații - ceea ce presupune un proces logic de inducție[2] - se poate face:

- empiric, fără a putea exprima numeric gradul de certitudine;
- ştiințific, exprimând numeric gradul de certitudine.

Exprimarea numerică a gradului de certitudine, inclusiv cu posibilitatea verificării sale obiective, nu se poate face decât prin *inferențe statistice* în cadrul *statisticii inferențiale*, numită şi *statistică inductivă*, *statistică analitică* sau *statistică matematică*. Aceasta se bazează *pe teoria probabilităților* şi poate exprima obiectiv gradul de certitudine numai pentru aşa-numitele *eşantioane probabiliste*. (Un eşantion este denumit **eşantion probabilist** dacă i se poate calcula probabilitatea de a fi obținut.)

Studiul incomplet al *populațiilor statistice* prin intermediul eşantioanelor probabiliste este scopul statisticii inductive. Aceasta se bazează pe *statistica descriptivă* şi pe *teoria probabilităților.*

În concluzie, **statistica clasică** este structurată în trei demersuri specifice*: statistică descriptivă*, o parte a *teoriei probabilităților* şi *statistică inductivă.*

[2] Inducția este procesul logic de trecerea de la particular la general, aici, de la eşantion la populație.

Metaforic vorbind, **statistica clasică** este ca un „sandvici” care se construieşte pornind de la felia de pâine inferioară (*statistica descriptivă*), continuând cu partea cea mai consistentă (*teoria probabilităților*) şi terminând cu felia superioară care se sprijină pe cele două straturi anterioare (*statistica inductivă*):

+Observații

- ✓ Un şir statistic poate reprezenta fie o populație statistică fie un eşantion. De aceea, adeseori, clasificarea populație statistică / eşantion este atribuită şirurilor, precum în ediția I a acestei lucrări.
- ✓ Orice demers de statistică inductivă începe cu studiul complet al eşantionului. Altfel spus, seria extrasă ca eşantion este tratată prin statistică descriptivă ca o populație statistică, apoi este, tratată, din nou, prin statistică inductivă, drept eşantion. De exemplu, dacă dispunem de un şir statistic cu date ale unui eşantion de studenți din România, acesta va fi considerat mai întâi populație statistică atunci când sintetizăm statistic datele respective, apoi va fi reconsiderat drept eşantion al populației tuturor studenților români, când vom extrapola rezultatele la toți studenții români.
- ✓ Pe de altă parte, dualitatea populație-eşantion pentru o serie statistică este exprimată şi de faptul că termenul “lot” este atribuit uneori şi unei populații statistice, nu numai eşantioanelor. De exemplu, în cadrul „controlului statistic de calitate” se prelevează un eşantion dintr-un aşa-numit “lot de fabricație” pentru a se accepta sau respinge întregul lot.

§ 2.4. Eşantioane prelevate independent şi eşantioane de observații perechi

În imensa majoritate a situațiilor reale se studiază populații statistice prin eşantioane provenite din acestea. Eşantioanele pot fi produse de diverse fenomene naturale ori pot fi selectate sau generate de noi. Un exemplu al primei situații este populația biologică umană la un moment dat. Aceasta poate fi considerată un eşantion produs de natură din populația statistică a speciei umane. A doua situație cuprinde cazul general al studiilor de observație, respectiv cel al studiilor experimentale.

În toate aceste cazuri două sau mai multe eşantioane se pot produce ori pot fi prelevate în două moduri: dependent sau independent. În continuare vom considera doar cazul a două eşantioane.

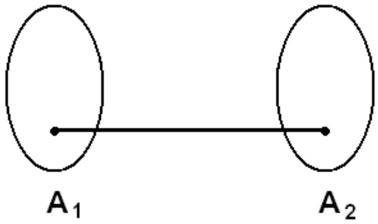
O situație în care două eşantioane pot fi considerate prelevate dependent este cea a observațiilor perechi.

Spunem despre două *eşantioane* că sunt **eşantioane de observații perechi** dacă au acelaşi volum şi selectarea unei unități într-un eşantion impune selectarea unei anumite unități, unei *perechi*, în celălalt eşantion.

Noțiunea de **eşantioane independent prelevate** reprezintă exact ceea ce specifică denumirea lor. În acest caz volumele pot fi atât egale cât şi diferite.

Să luăm în considerare în continuare ideea *eşantioanelor de observații perechi.*

- O situație generală de acest tip este cea în care unitățile statistice dintr-un eşantion A sunt observate sau măsurate de două ori (A_1, A_2 în desenul alăturat) fie de către (a) doi operatori sau (b) două aparate, fie (c) la două momente de timp diferite, (c_1) eventual după aplicarea unui tratament.



Exemplele 2.4.

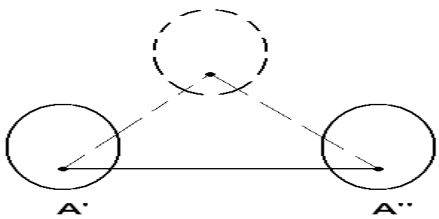
a. Compararea unui operator atestat cu unul neatestat, pentru eventuala atestare a celui de-al doilea.

b. Compararea unui aparat de măsură neetalonat cu un aparat etalonat se poate face apelând la cuplul de măsurători executate asupra aceluiaşi eşantion cu primul, respectiv, cu cel de-al doilea aparat.

c. Aşa numitele "studii longitudinale" antropologice care urmăresc probleme de creştere-dezvoltare prin câte două eşantioane de observații perechi primul eşantion fiind format din n copii la o anumită vârstă v, iar al doilea din aceiaşi n copii, un an mai târziu.

c_1. Cuplul de eşantioane produse de orice experiment clasic de studiu al eficacității unei substanțe medicamentoase. Se ia un lot de n subiecți cărora li se măsoară o anumită caracteristică (de exemplu tensiunea arterială) înaintea tratării cu substanța respectivă (un hipotensiv în cazul exemplului nostru) şi după. Primul eşantion va fi format din valorile măsurate înaintea tratamentului, iar cel de-al doilea din valorile de după tratament.

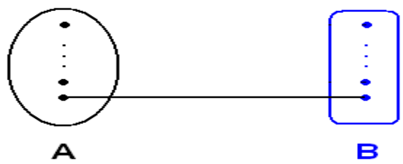
- O altă situație generală poate fi dată de un eşantion format din n unități pentru care s-a notat un anumit caracter (A' în desenul alăturat) şi un alt eşantion format din aceleaşi n unități studiat după alt caracter (A'') care admite, însă, aceeaşi unitate de măsură.



Exemplul 2.4'.

d. Dacă ne interesează să comparăm circumferința toracică cu cea abdominală într-o anumită populație umană sau animală vom reține într-un eşantion circumferințele toracice ale celor n indivizi selectați în lot şi în cel de-al doilea eşantion circumferințele abdominale ale aceloraşi n indivizi.

- În fine, cea mai generală situație cuprinde un eşantion format din n unități (A în desenul alăturat) şi alte n unități (B) relaționate cu primele una câte una în sensul interesului exprimat în prelevarea şi compararea eşantioanelor.



Exemplul 2.4''.

e. Pentru a determina care dintre cei doi soți ai unui cuplu are în medie un IQ (coeficient de inteligență) mai mare, primul eşantion va conține valorile coeficientului la cei n soți selectați în lot, iar al doilea eşantion va conține cei n coeficienți ai soțiilor acestora.

Pentru anumite probleme din cadrul exemplelor anterioare se pot imagina şi perechi de *eşantioane independent prelevate.*

Exemplele 2.4'''.

a'. Problema *a* nu poate fi rezolvată decât ca mai sus.

b'. Analog, nici problema *b* nu poate f rezolvată prin eşantoane independent prelevate.

c'. Se poate studia creşterea şi prin compararea a două eşantioane de volume egale sau diferite, prelevate independent din populația de copii de v ani, respectiv din cea de $v+1$ ani.

c_1'. Eficacitatea unei substanțe medicamentoase poate fi testată utilizând un eşantion de n_1 indivizi tratați cu substanța respectivă şi un al doilea eşantion, numit lot martor, format din n_2 indivizi care primesc placebo. Placebo este o substanță fără efect, care se administrează ca şi când ar fi substanța cu efect, pacientul fiind convins că primeşte substanța activă. În acest mod se înlătură din efectul măsurat influența autosugestiei. Mai mult chiar, se poate organiza alocarea pacienților în cele două loturi astfel încât nici personalul medical care tratează direct pacienții să nu ştie cui administrează placebo şi cui substanță activă. Asemenea studii se numesc în engleză "double blind", adică "dublu orb".

d'. Diferența între circumferința toracică şi cea abdominală poate fi testată, de asemenea, prin eşantioane independent prelevate.

e'. Problema *e* nu poate fi rezolvată decât ca mai sus, căci dacă un eşantion va fi format din coeficienții de inteligență ai n_1 soți, iar cel de-al doilea din coeficienții provenind de la n_2 soții, vom determina eventuala diferență de inteligență între bărbați însurați şi femei măritate şi nu cea existentă în cadrul cuplurilor.

✓ O greşeală metodologică gravă este amestecarea eşantioanelor de observații perechi, cu cele prelevate independent. De exemplu în studiile antropologice longitudinale se întâmplă adesea să se piardă indivizi de la un an la altul. Dacă se "completează" observațiile lipsă cu date de la noi indivizi, obținându-se aşa numitele studii "mixt-longitudinale", avem de a face cu eşantioane care nu sunt nici independente, nici dependente, pentru care nu există aparat statistic de inferență.

În volumul II, de biostatistică inductivă, vom arăta că aceleaşi date pot conduce la decizii complet diferite după cum au fost tratate ca dependente sau independente.