

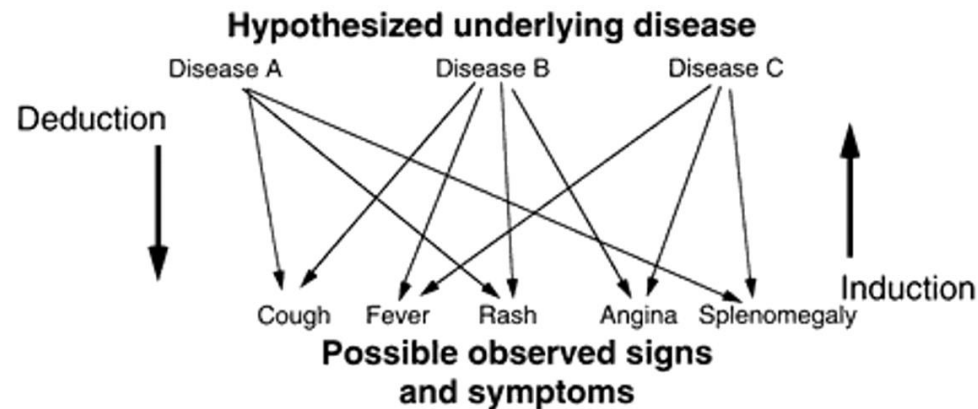
INTRODUCERE ÎN STATISTICA MEDICALĂ

Despre ...

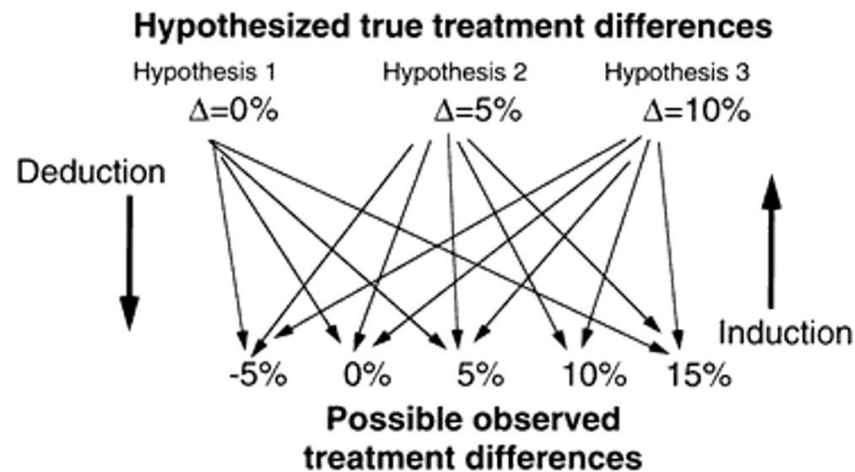
- Terminologie
- Stadii ale cunoașterii
- Măsurarea și acuratețea
- Tipuri de date medicale
- Populația și eșantionul. Eșantionarea

INFERENȚA MEDICALĂ VS INFERENȚA STATISTICĂ

Medical Inference



Statistical Inference



DEFINIȚII

- Statistica = domeniu științific care permite studiul fenomenelor, parametrilor și caracteristicilor a căror proprietate fundamentală este variabilitatea
- Se ocupă cu culegea, sintetizarea, descrierea și interpretarea datelor medicale
- Statistica medicală = știința colectării datelor, a prezentării lor într-o formă sistematică , a analizei acestora și a interpretării rezultatelor obținute
- Unitatea statistică = elementul constitutiv al populației statistice
- Variabila = caracteristică care poate prezenta variație de la un element la altul al unei colectivități;

DEFINIȚII

- Datele statistice = valorile observate și înregistrate ale unei variabile statistice
- Informațiile statistice = rezultă din prelucrarea și analiza datelor statistice
- Statistica descriptivă = Statistica descriptivă este ramura statisticii care se ocupa cu prezentarea unui set de date. Prin prezentare se permite înțelegerea facilă a caracteristicilor evenimentelor studiate.
- Estimarea = procesul utilizat pentru determinarea valorii unui parametru statistic asociat unei populații
- Estimatorul = funcție statistică aplicată asupra eșantionului pentru a estima un parametru necunoscut al populației. Valoarea obținută este o estimată a valorii populației.

DEFINIȚII

- Eșantionul = o colecție (submulțime) de elemente din populație
- Eșantioane perechi =
 - două eșantioane în care elementele sunt clar perechi (determinarea TAS la aceiași pacienți la includerea în studiu și respectiv la 6 luni de la inițierea tratamentului)
 - Elementele unui eșantion sunt selectate pentru a se potrivi cu elementele celui de-al doilea eșantion
- Eșantioane independente = probabilitatea ca un element să fie inclus în eșantion nu depinde de selecția unui alt element în cel de-al doilea eșantionul

CUNOAȘTERE/CUNOȘTINȚE

- Totalitatea noțiunilor, ideilor, informațiilor pe care le are cineva într-un domeniu oarecare.
- “KNOWLEDGE IS THE ENEMY OF DISEASE”
 - *“The application of what we know will have a bigger impact on health and disease than any single drug or technology likely to be introduced in the next decade.” - Sir Muir Gray*

SURSE

- Cercetarea clinică și preclinică (evidențe)
- Analiza datelor colectate în rutina medicală zilnică (date statistice)
- Experiența practicienilor.

Cunoștințe explicite (formalizate)

- tratamente
- medicamente
- terminologia și clasificarea bolilor

ETAPE

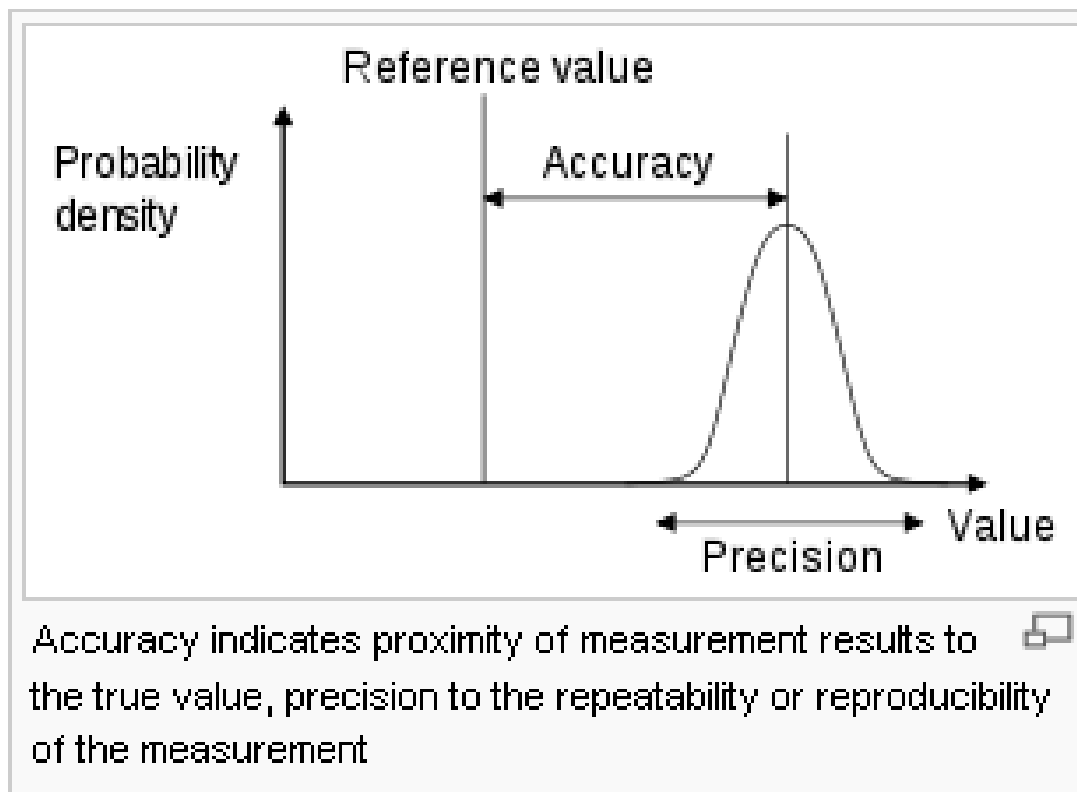
1. Descrierea ... evenimentului medical
 - Patologie nouă
 - Forme de manifestare ieșite din tipar
 - Studiul de caz sau serii de cazuri
2. Explicarea ... evenimentului medical
 - Realizată prin testarea unei ipoteze
 - Ipoteza clinică $\neq$ ipoteza statistică
3. Predicția ... prezicerea apariției unui eveniment
 - Model statistic utilizat în decizia medicală
 - Răspunde la întrebări de tipul: *"Dacă volumul tumoral este de 70 cm^3 și antigenul tumoral specific este pozitiv, care este probabilitatea ca formațiunea să fie malignă?"*

MĂSURAREA ȘI ACURATEȚEA

- Informațiile obținute din dat sunt mai informative dacă datele sunt cantitative
- Cuantificarea:
 - Cantitativ prin natura variabilei
 - Bazat pe măsurătoare

MĂSURAREA ȘI ACURATEȚEA

- Acuratețea: cât de mult valoarea măsurată reflectă valoarea reală
- Precizia: cât de aproape sunt valorile a două măsurători ale aceleiași variabile cu același instrument



MATRICEA DE DATE (DATE BRUTE)

Variabila (denumire & unitatea de măsură)

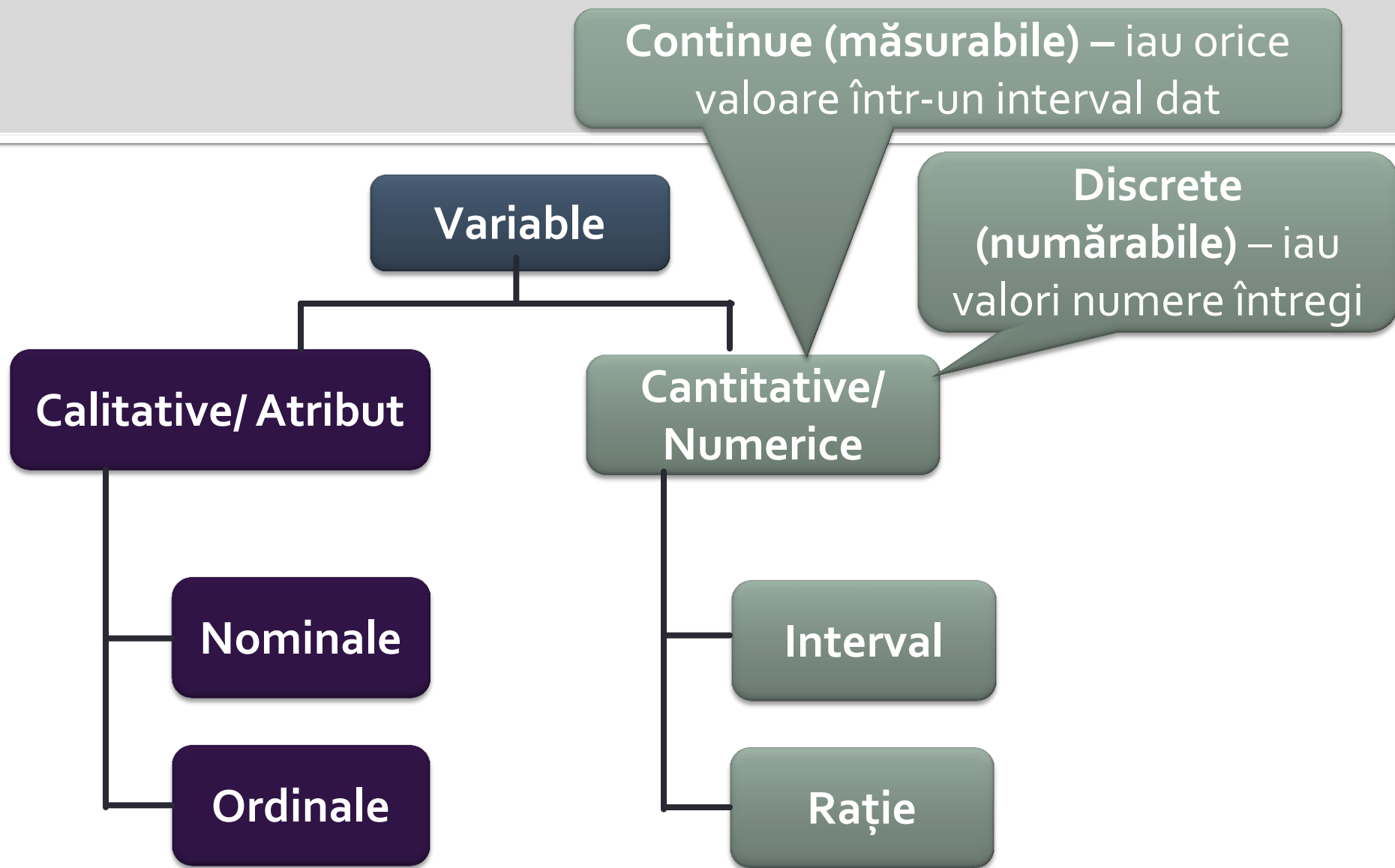
Valoarea mare indică extindere mai mare

ID	Gen (F/M)	Fumat (Da/Nu)	Greutate (kg)		Domiciliu	Stadiul patologiei maligne
001	F	Da	56	...	Cluj	1
002	M	Da	85	...	Cluj	0
003	M	Da	120	...	Sălaj	4
...	...	...	...	...	...	...
199	M	Nu	75	...	Maramureș	2

Observație / Caz

VARIABLE & DATE

- Care este procentul de studenți fumători în cadrul Facultății de Medicină?
 - Variabila: FUMATOR
 - Data: răspunsul de tip DA/NU (sau numărul de țigări) dat de fiecare student în parte
- Studiile de cercetare și analizele statistice: studiul variabilelor.



SCALA DE MĂSURĂ

Nominal Variabile clasificate în grupuri discrete pe baza unor caracteristici particulare Grupurile nu pot fi ordonate	Ordinal Clasificare ordonată după ranguri (de la mare la mic ...) Nu se specifică care este distanța dintre ranguri
Interval Intervalul (sau distanța) între două puncte pe scală are semnificație precisă	Rație Diferă față de scala de tip interval doar prin existența punctului de zero cu semnificație

CALITATIVE / ATRIBUT

Scala de măsură: proprietăți

Nominal

- Identitatea (exprimă apartenența elementelor la o categorie)
- Presupune o clasificare a variabilei fără a indica o anumită ordine ori cantitate
- Pot fi notate cu cifre (0-feminin; 1-masculin) însă nu pot fi procesate în termeni de cantitate sau ordine.

Ordinal

- Elementele sunt clasate conform unei ordini, preferințe
- Entitățile pot fi comparate între ele în termeni de *mai mult, mai puțin sau egal*.

CANTITAIVE

Scala de măsură: proprietăți

Interval

- Pentru variabile cantitative
- Identitate și ordine
- Intervalul între numere are un sens (permite compararea diferențele între numere).
- Punctul 0 (zero) este ales arbitrar.
 - Ex: temperatura 0 nu e lipsa temperaturii, e doar punctul de îngheț al apei.

Rație

- Folosită pentru variabile cantitative
- Are un o absolut care înseamnă lipsa caracteristicii sau proprietății respective
- Ex.: De exemplu un venit de 0 lei înseamnă inexistența unui venit

EXAMPLE

Nominal Culoarea părului; Educația; Starea civila (căsătorit, divorțat, văduv, necăsătorit) ... Variabile dihotomiale: Gen	Ordinal Locul ocupat de un concurent la o competiție (primul, al doilea, ..., ultimul).
Interval Temperatura (diferența dintre 50 °C și 60°C este aceeași ca și diferența dintre 100 °C și 110 °C);	Rație Bani (venit); Masa (greutate); Lungimea (cm); Volumul; Timpul de reacție (sec) ...

CALITATIV VS. CANTITATIV

Cantitative (măsurabile)

- Continue
 - Exisă un număr infinit de valori posibile care se găsesc în intervalul a două valori observate
- Discrete
 - Valori numere întregi

Atribut (calitative)

- Nemăsurabile
- Pot lua un număr finit de valori
- Calculul mediei nu are sens
- Statistica descriptivă: frecvențe absolute și relative

TRANSFORMAREA VARIABILELOR

- Este posibilă transformarea scalelor de măsură tip interval și rație în scale de tip ordinal sau nominal dar întotdeauna se face cu pierdere de informație
 - transformarea scalei asociată variabilei vârstă în scală ordinală “clase de vârstă”
- Nu este posibilă transformarea scalei de tip nominal sau ordinal în scală de tip interval sau rație chiar dacă atribuim valori diferitelor clase
 - Sex: $M = 1$, $F = 0$

VARIABLE DEPENDENTE VS INDEPENDENTE

- Când două variabile au o legătură una cu cealaltă, ele se numesc variabile asociate sau dependente
 - Asocierea poate să fie pozitivă sau negativă
- Dacă două variabile nu sunt asociate ele se numesc independente.

SERII STATISTICE

- » Unidimensională: șir de valori numerice ($X_1, X_2, \dots, X_n$): TAS (mmHg)

X_1	X_2	X_3	X_4	X_5	X_6	X_7
120	195	110	100	160	115	185

- » Bidimensională: TAS (mmHg) și TAD (mmHg)

X_1	X_2	X_3	X_4	X_5	X_6	X_7
120	195	110	100	160	115	185
70	110	65	60	90	70	105

- » Multidimensională:

X_1	TAS (mmHg)	TAD (mmHg)	Puls	G (kg)	Î (cm)
X_2	195	110	95	200	200
X_3	100	65	56	62	158
X_4	110	60	64	45	150
X_5	160	90	60	85	175

VARIABILE CALITATIVE/ATRIBUT ± CANTITATIVE

- » Univariată: o singură variabilă
- » Bivariată: 2 variabile
- » Multivariată: > 2 variabile

Gen	F	M	M	F	F	M	F
-----	---	---	---	---	---	---	---

Gen	F	M	M	F	F	M	F
TAS (mmHg)	120	160	180	210	150	100	110

Bivariată &
Unidimensională

Gen	F	M	M	F	F	M	F
TAS (mmHg)	120	160	180	210	150	100	110
Zile spitalizare	5	4	3	5	1	2	4
TAD (mmHg)	75	50	65	100	80	52	70

Multivariată &
Multidimensională

- Care este tipul variabilei 'număr de bacterii identificate pe frotiu'?
- Variabila 'nivelul educațional' (gimnazial, liceal, postliceal, studii de licență, studii de master, studii de doctorat, studii postdoctorale) este:
 - A. Numerică, continuă
 - B. Numerică, discretă
 - C. Dicotomială
 - D. Calitativă, ordonată
 - E. Calitativă, nominală
- Seria statistică formată din genul studentului și nota obținută la examenul de anatomie este:
 - A. Univariată
 - B. Bidimensională
 - C. Unidimensională
 - D. Bivariată
 - E. Multidimensională



- Care este tipul variabilei număr de bacterii identificate pe frotiu?
 - Care este tipul variabilei nivelul educațional (gimnazial, liceal, postliceal, studii de licență, studii de master, studii de doctorat, studii postdoctorale)?
 - A. Numeric, continuu
 - B. Numeric, discret
 - C. Dicotomial
 - D. Calitative, ordonat
 - E. Calitativ, discret
- » Seria statistică formată din genul studentului și nota obținută la examenul de anatomie este:
- A. Univariată
 - B. Bidimensională
 - C. Unidimensională
 - D. Bivariată
 - E. Multidimensională

76% din studenți au
avut note de
promovare

Un număr semnificativ
statistic mai mare de
fete au avut note de
promovare ($p < 0,01$)

Statistica

Descriptivă

Inferențială

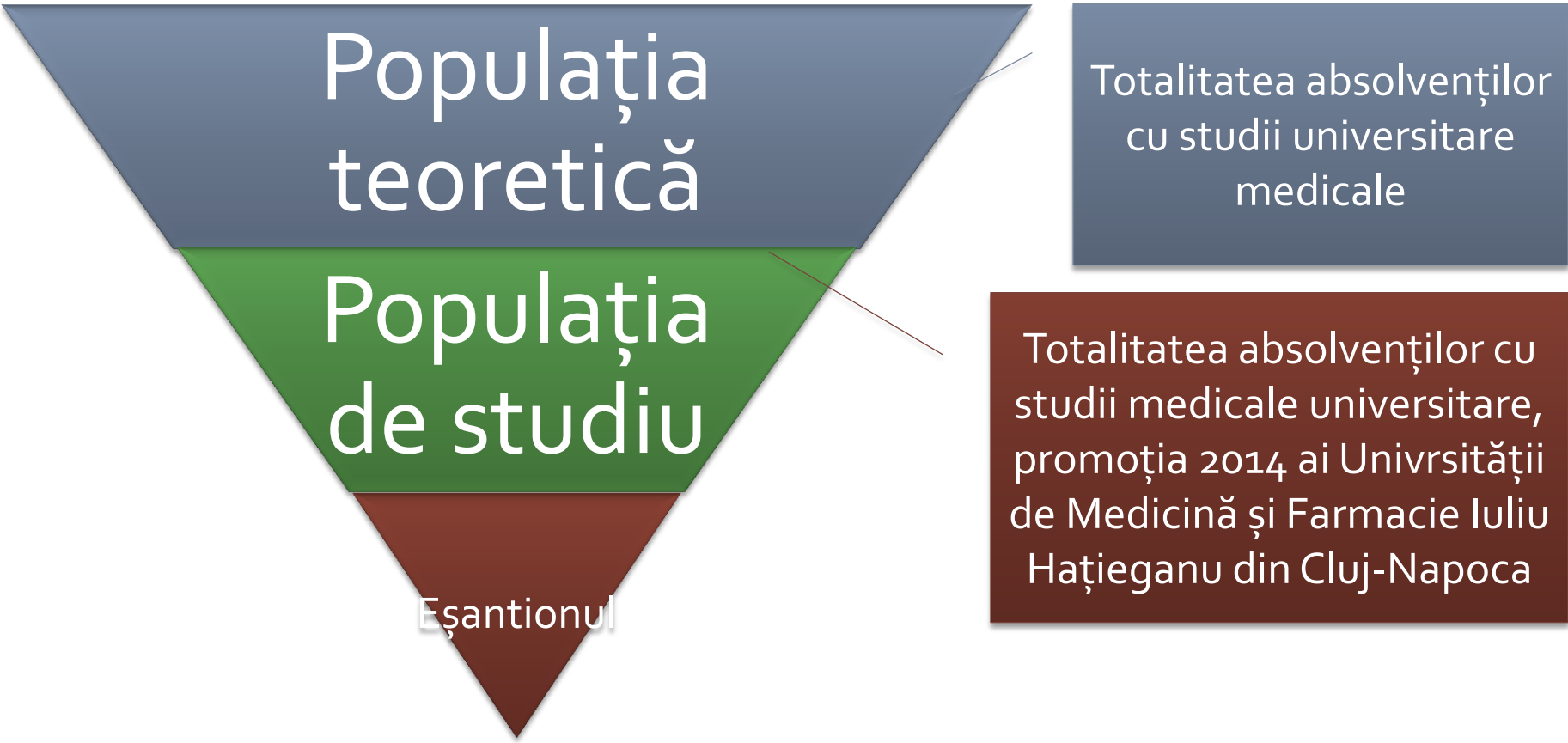
Colectare & Sintetizare & Prezentare & Descriere

Comparare

DOMENII

POPULAȚIA STATISTICĂ

- mulțime de elemente (obiecte sau subiecți) care au anumite însușiri (atribute sau caractere) comune, care formează obiectul unei analize statistice
- numărul elementelor populației se numește ***volumul sau talia populației***
- Exemple:
 - un grup de pacienți
 - o mulțime de obiecte
 - un grup de fenomene sau evenimente ...



Populația
teoretică

Totalitatea absolvenților
cu studii universitare
medicale

Populația
de studiu

Totalitatea absolvenților cu
studii medicale universitare,
promoția 2014 ai Univrsității
de Medicină și Farmacie Iuliu
Hațieganu din Cluj-Napoca

Eșantionul

EȘANTIONUL: DE CE?

- De cele mai multe ori volumul populației nu permite investigarea exhaustivă a acesteia
- Restricții: timp – bani – personal
- Studiul întregii populații poate determina distrugerea ei
- Avem acces doar la o parte din populație
- Procesul și tehnicile de măsurare sunt mai precise la nivelul eșantionului decât la nivelul populației

Prin extragerea corectă a participanților la studiu dintr-o populație specifică, cercetătorul poate analiza eșantionul și realiza inferențe despre caracteristica studiată a populației.

EȘANTIONUL:

- Reprezentativ pentru populație:
 - Talie
 - Caracteristici
- Calculul taliei eșantionului:
 - Riscul de a respinge ipoteza nulă dacă ea e adevărată (alfa, $\alpha = 5\% = 0,05$)
 - Puterea studiului (probabilitatea de a respinge ipoteza nulă când nu este adevărată)

VOLUMUL EȘANTIONULUI

- **Acuratețe:** valoarea reală + eroarea (cu cât volumul eșantionului e mai mare cu atât probabilitatea de eroare e mai mică)
- **Cost** (cu cât volumul eșantionului e mai mare cu atât costul cercetării e mai mare)
- **Omogenitatea** populației (membrii populației sunt similari în ceea ce privește caracteristica de studiat): cu cât variabilitatea în populație e mai mare cu atât volumul eșantionului trebuie să fie mai mare
- Alți **factori** care pot influența: (a) există variabile pe care nu le putem controla; (b) se dorește împărțirea eșantionului în subgrupe; (c) se prevede un număr mare de pierduți din vedere; (d) se dorește o putere statistică înaltă.

VOLUMUL EȘANTIONULUI

- Empiric:

Volumul/talia populației (N)	Volumul/talia eșantionului (n) (%)
0 – 100	100
101 – 1000	10
1001 – 5000	5
5001 – 10000	3
> 10000	1

- Formule de calcul (în funcție de variabila urmărită)

METODE DE EȘANTIONARE

Eșantionare probabilistică:

fiecare unitate din populație are o șansă mai mare de zero de a fi selectată și inclusă în eșantion

- Simplu randomizată
- Sistematică
- Stratificată
- Cluster
- ...

Eșantionare non-probabilistică:

unele unități ale populației nu au nici o șansă de a fi selectate și incluse în eșantion

- > Nu permite estimarea erorii de eșantionare

Ex:

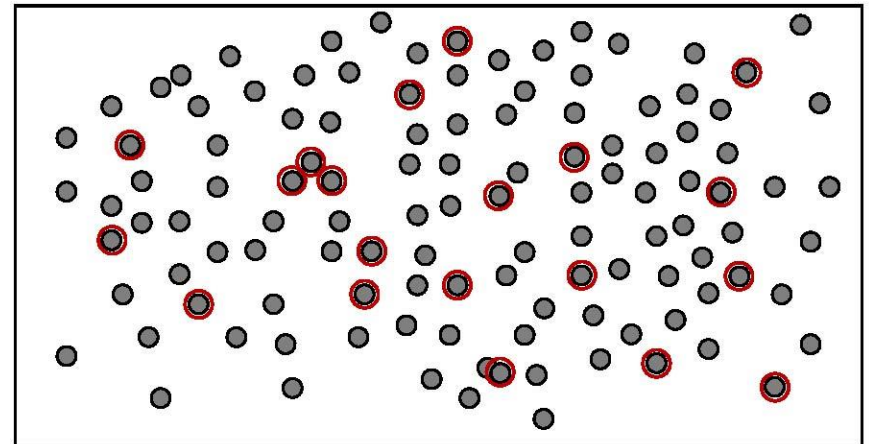
- > Eșantionul conveniet
- > Eșantionul de tip bulgăre de zăpadă
- > Eșantionul deliberat

EȘANTIONARE NON-PROBABILISTICĂ

- Convenient:
 - Participanții sunt selectați deoarece sunt accesibili
 - Relativ ușor
- Bulgărele de zăpadă:
 - Subiecții incluși în studiu vor aduce alți potențiali participanți (ex. membrii ai aceluiași grup, activități comune, etc.)
- Deliberat
 - Grup de tehnici de eșantionare care au la bază gândirea cercetătorului (ex. eșantionarea cu variație maximă, eșantionarea omogenă, eșantionarea cazurilor comune, eșantionarea cazurilor extreme, eșantionarea realizată de experți)

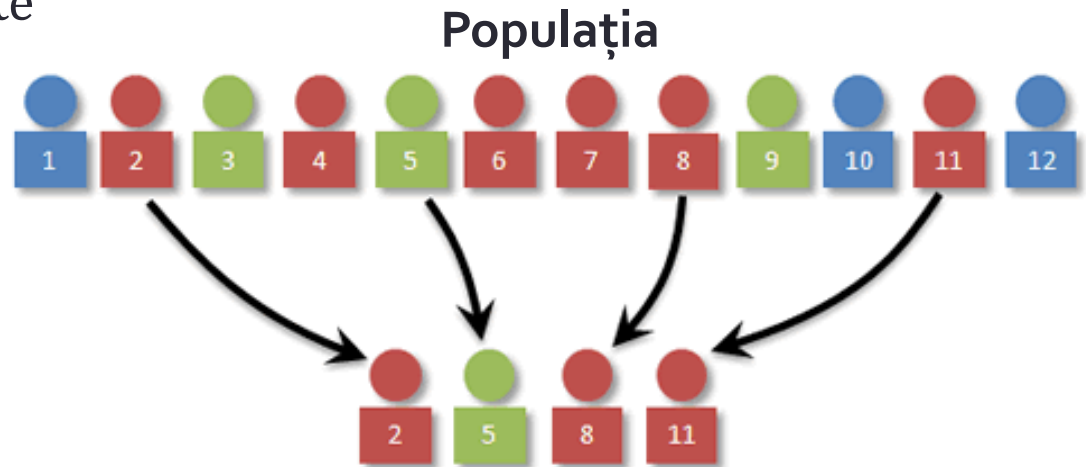
EȘANTIONAREA SIMPLU RANDOMIZATĂ

- Subiecți extrași la întâmplare din populația statistică
- Fiecare subiect are aceeași șansă de a fi inclus în eșantion
 - [Tabele cu numere randomizate](#)
 - [Generarea aleatorie a unui set de numere discrete](#)
 - Folosirea funcțiilor Excel pentru generarea numerelor randomizate (RANDBETWEEN)



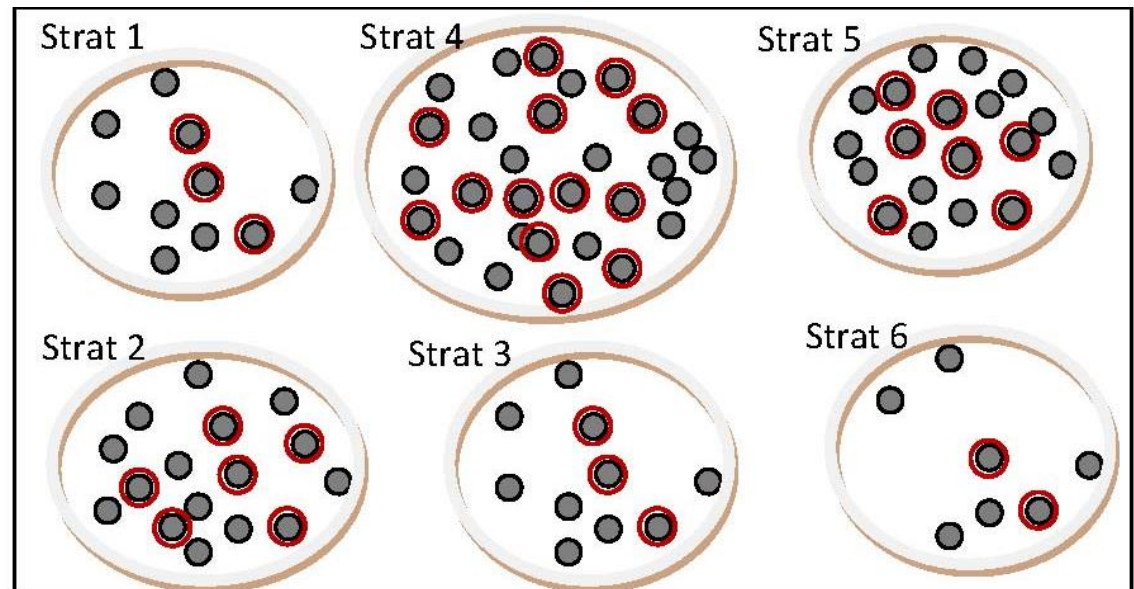
EȘANTIONAREA SISTEMATICĂ

- Este selectat pentru a fi inclus în eșantion fiecare al k-lea element din structura de eșantionare
- Numărul k se obține împărțind talia populației la talia dorită a eșantionului
 - Populație de 10000 – eșantion de 1000: $k = 10$
 - Se selectează prin randomizare punctul de start (0 - 9)
- Nu este indicat să fie folosită atunci când în structura de eșantionare ar putea apare o periodicitate



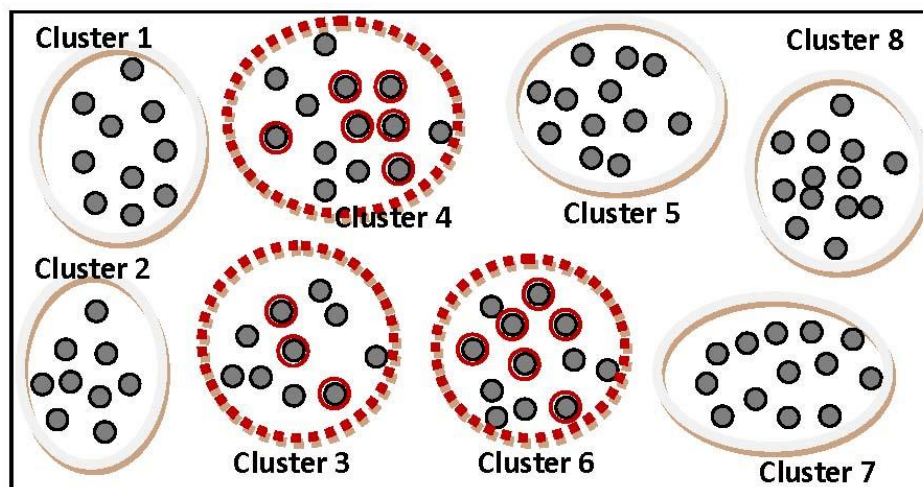
EȘANTIONARE STRATIFICATĂ

- Se împarte populația în mai multe subgrupe relevante numite **straturi**
- Se constituie eșantionul prin extrageri aleatoare din straturi
- Fiecare strat trebuie să fi reprezentat în eșantion în funcție de importanța sa în populație

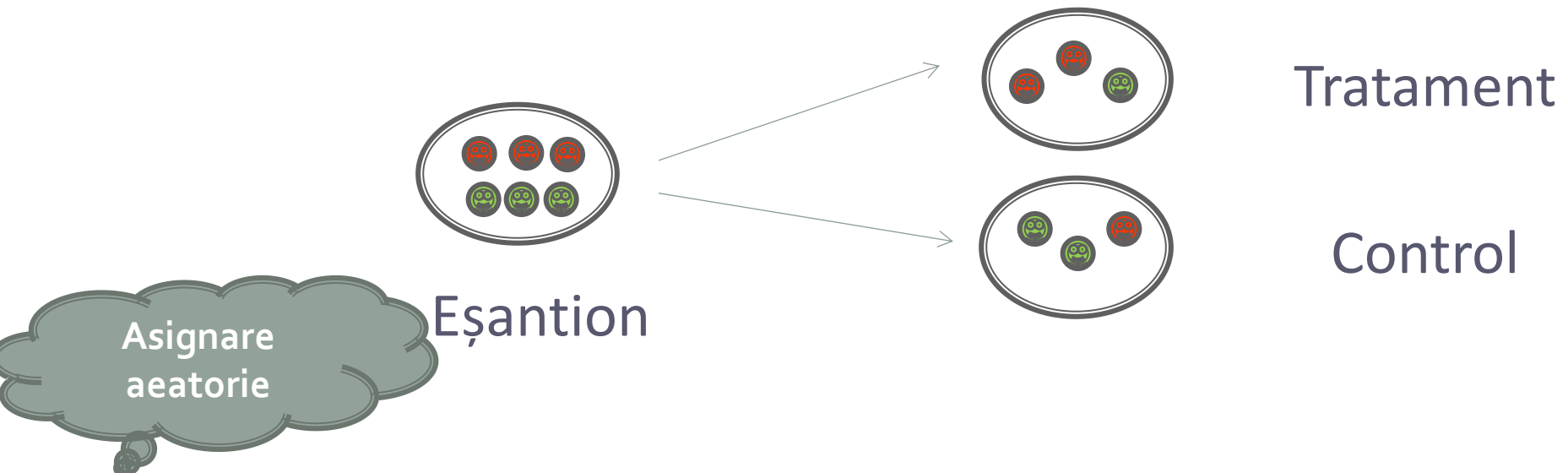
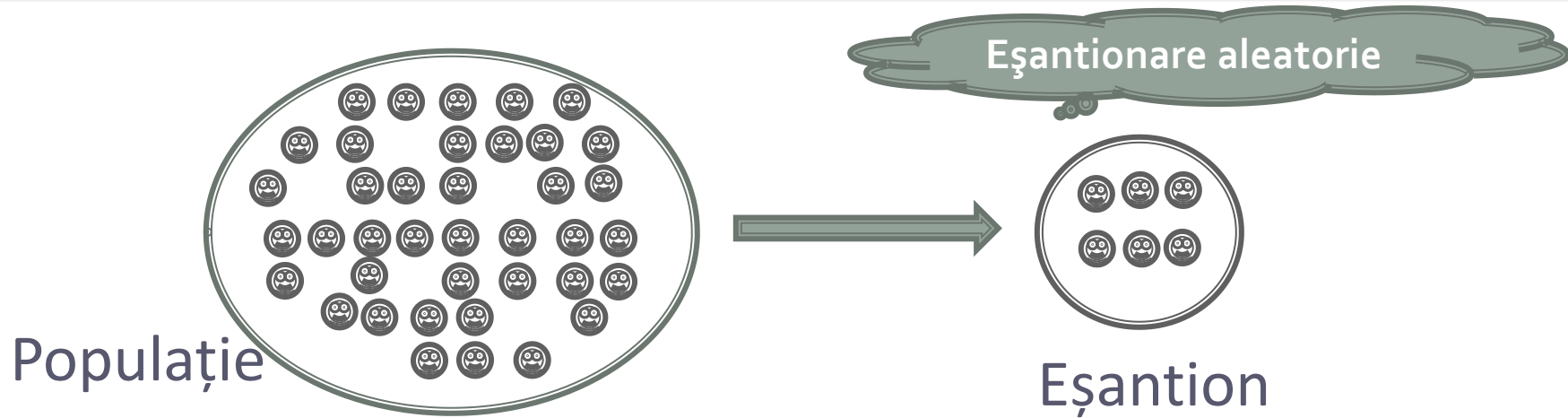


EȘANTIONARE DE TIP CLUSTER

- Procedeu în două etape
 - se împarte populația în clusteri
 - se selectează aleator o submulțime de clusteri
 - se selectează prin metoda simplu randomizată un număr de itemi din fiecare cluster selectat anterior
- În mod obișnuit alegerea clusterelor se bazează pe criterii geografice, iar acest procedeu este aplicat în special în studiile epidemiologice.



EȘANTIONARE ALEATORIE/RANDOMIZATĂ VS. ASIGNARE ALEATORIE



DE REȚINUT!

- Identificarea corectă a variabilelor investigate permite analiza statistică corectă.
- Dacă e posibil colectăm date cantitative
- Transformarea datelor cantitative în calitative se face cu pierdere de informație
- Scopul statisticii descriptive este de a prezenta realitatea statistică.
- Scopul statisticii inferențiale este de a concluziona existența sau nu a diferenței semnificative statistic.

DE REȚINUT!

- În cercetarea medicală evaluăm eșantionul pentru a identifica ce se întâmplă în populație.
 - Doar eșantionul corect selectat din populație (eșantion reprezentativ) permite extrapolarea rezultatelor obținute pe eșantion la nivelul populației.
- Evaluarea întregii populații nu se realizează deoarece este constisitoare, consumatoare de timp, și de cele mai multe ori imposibilă.
 - Poate duce inclusiv la distrugerea populației.

