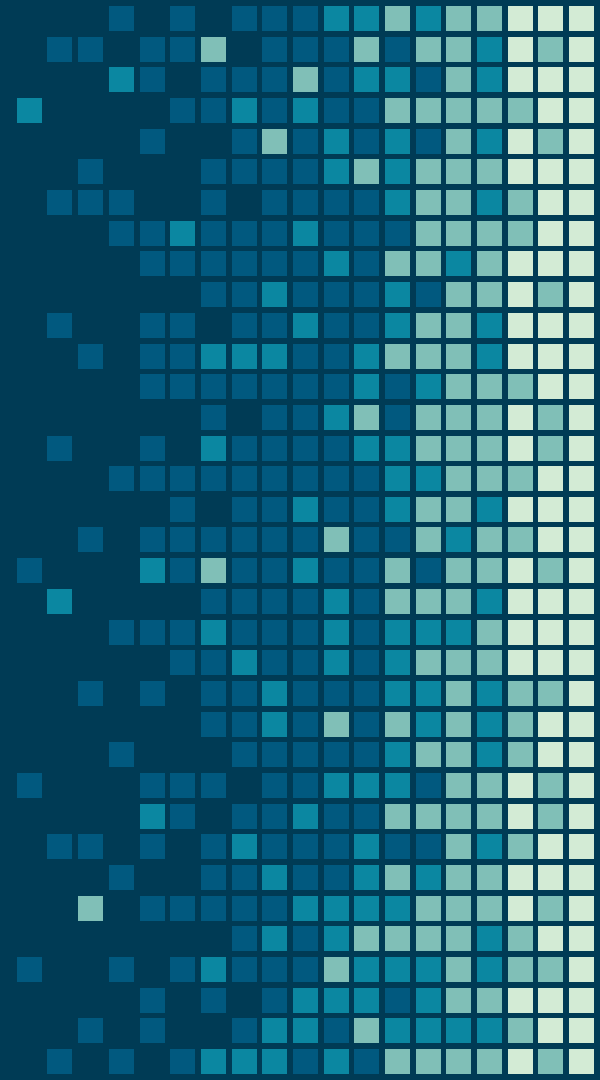


PROBABILITĂȚI ȘI STATISTICĂ ÎN SISTEME MEDICALE

Cursul 1, 16.09.2020

Elemente și noțiuni introductive



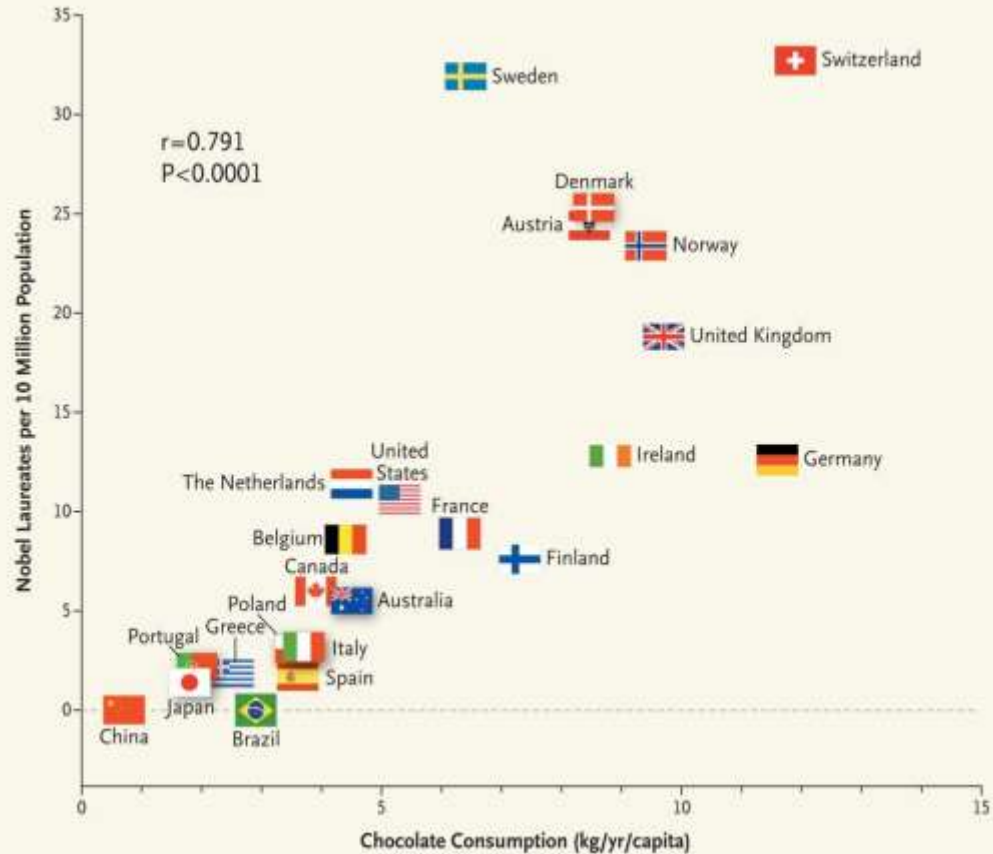
STRUCTURA CURSULUI

1. Tipuri de variabile utilizate în sistemele medicale. Transformarea variabilelor
2. Particularități ale datelor din sistemul medical
3. Proiectarea și tehnici de analiză pentru studii epidemiologice

BIBLIOGRAFIE CURS

1. **Drugan, T., Bolboacă, S.D., Leucuța, D., Bondor, C., Călinici, T., Văleanu, M., Colosi, H., Iancu, M., Istrate, D.** (2018). Curs de biostatistică medicală, Universitatea De Medicină și Farmacie "Iuliu Hațieganu" Cluj-Napoca, pag. 7-13
2. **Mărușter, M.** (2006). Noțiuni fundamentale de biostatistică- note de curs, University Press Tg. Mureș, pag. 6-16, pag. 115-134

Statistica, pe scurt! 😊



DEFINIRE

Statistica medicală
(**biostatistica**) se ocupă cu aplicarea metodelor statistice asupra datelor medicale/în sisteme medicale și/sau farmaceutice

Ramuri

Statistica descriptivă:

organizarea și sumarizarea datelor cu ajutorul estimatorilor, a tabelelor și a reprezentărilor grafice

Statistica inferențială:

utilizarea metodelor statistice care să permită generalizarea datelor obținute prin studierea eșantionului asupra populației din care s-a extras eșantionul prin testarea unor ipoteze statistice.

Se poate realiza doar dacă eșantionul este reprezentativ pentru populația de interes.

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

Colectare & Sintetizare & Prezentare &
Descriere

Inferențială

Comparare

DESCRIPTIV vs INFERENTIAL



RESULTS: HCV prevalence of 0.38% (95% confidence interval: 0.22-0.54) was found. The level of knowledge of the natural history, risk behavior, and prevention and treatment of HCV was poor among the participants. The lowest level of knowledge for all participants pertained to its prevention and treatment.

RESULTS: ... All patients had 88% engagement in primary care during 2008-2013. HCV mono-infected and HCV/HIV co-infected patients had similar treatment (6% vs. 5%, $p=0.6622$) and HCV SVR (1% vs. 0.5%, $p=0.1737$) rates in the interferon era. However, more HCV/HIV co-infected patients were engaged in care (93% vs. 87%, $p=0.0044$), accessed HCV treatment (36% vs. 23%, $p<0.0001$), and achieved HCV SVR (31% vs. 21% $p=0.0002$) compared to mono-infected patients in the DAA era.

...

CONCEPTE DE BAZĂ

Unitatea statistică:	elementul constitutiv al populației
Variabila:	caracteristică care poate prezenta variație de la un element la altul al unei colectivități
Datele statistice:	valorile observate și înregistrate ale unei variabile statistice
Informațiile statistice:	rezultă din prelucrarea și analiza datelor statistice
Statistica descriptivă:	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 prin evaluarea unui eșantion
Estimatorul :	Funcție statistică aplicată asupra eșantionului pentru a estima un parametru necunoscut al populației. Valoarea obținută este o estimată a valorii populației

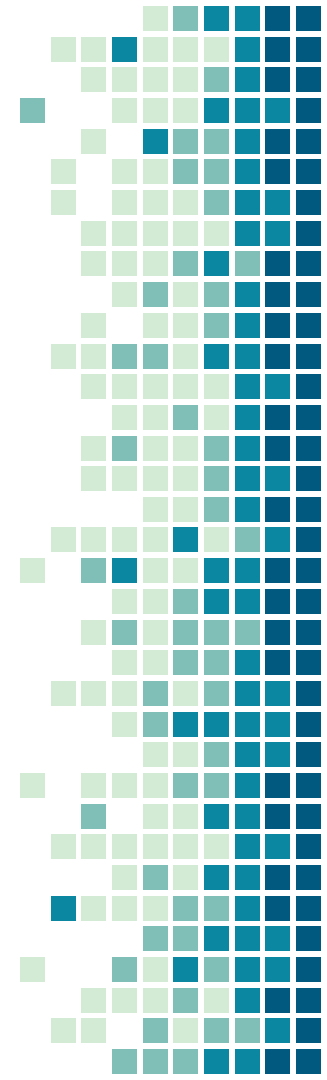
CONCEPTE DE BAZĂ

Eșantionul :	o colecție(submulțime) de elemente din populație
Eșantioane perechi (dependente)	<ul style="list-style-type: none">• două eşantioane în care elementele sunt perechi (determinarea PAS – Presiunea Arterială Sistolică - la aceiași pacienți la includerea în studiu și respectiv la 6 luni de la inițierea tratamentului)• elementele unui eşantion sunt selectate pentru a se potrivi cu elementele celui de-al doilea eşantion
Eșantioane independente	Eșantioane independente = probabilitatea ca un element să fie inclus în eşantion nu depinde de selecția unui alt element în cel de-al doilea eşantion
Parametrul (media aritmetică, deviația standard, mediana, etc.)	este o sumarizare numerică obținută pe populație. E simbolizat cu litere grecești (ex μ = media populației, σ = deviația standard a populației).
Statistica	Statistica este o sumarizare numerică obținută prin investigarea unui eşantion. E simbolizat cel mai frecvent cu litere latine (ex s = deviația standard a eşantionului)

PARAMETRU vs STATISTICA



- Hepatita C este o patologie infecțioasă cauzată de virusul hepatitic C care are o prevalență de ~2-3% la nivelul populației lumii. → Valoarea prevalenței este un parametru deoarece este o sumarizare a ceea ce se întâmplă la nivelul populației.
- Se investighează un eșantion format din 100 persoane ale corpului medical care lucrează în spitalele de boli infecțioase din Bulgaria; 4.3% dintre aceștia au prezent virusul hepatitic C. → Această valoare reprezintă o statistică deoarece e o sumarizare obținută pe un eșantion.



Surse ale cunoașterii medicale

- Cercetarea clinică și preclinică (evidențe/dovezi)
- 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?



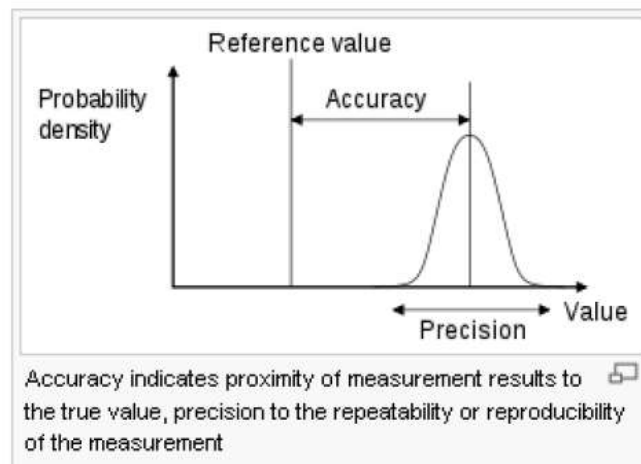
FORME ALE CUNOAȘTERII

- Observarea → caracter constatativ-existențial cel mult descriptiv-calitativ:
 - Indispensabilă măsurării și experimentului
- Măsurarea (cunoaștere experimentală) → caracter descriptiv-cantitativ
 - Atribuirea unei valori cantitative (numerice)
 - Trebuie să se caracterizeze prin precizie
 - Efectuarea măsurărilor într-un număr adecvat de repetiții (eroarea de măsurare este invers proporțională cu radicalul numărului repetițiilor – $e = 1/\sqrt{n}$)
- Experimentul → intervenția directă a cercetătorului în desfășurarea naturală a evenimentului de interes
- Modelarea → transformarea în modele (teoretice, matematice ...) a fenomenelor de sănătate

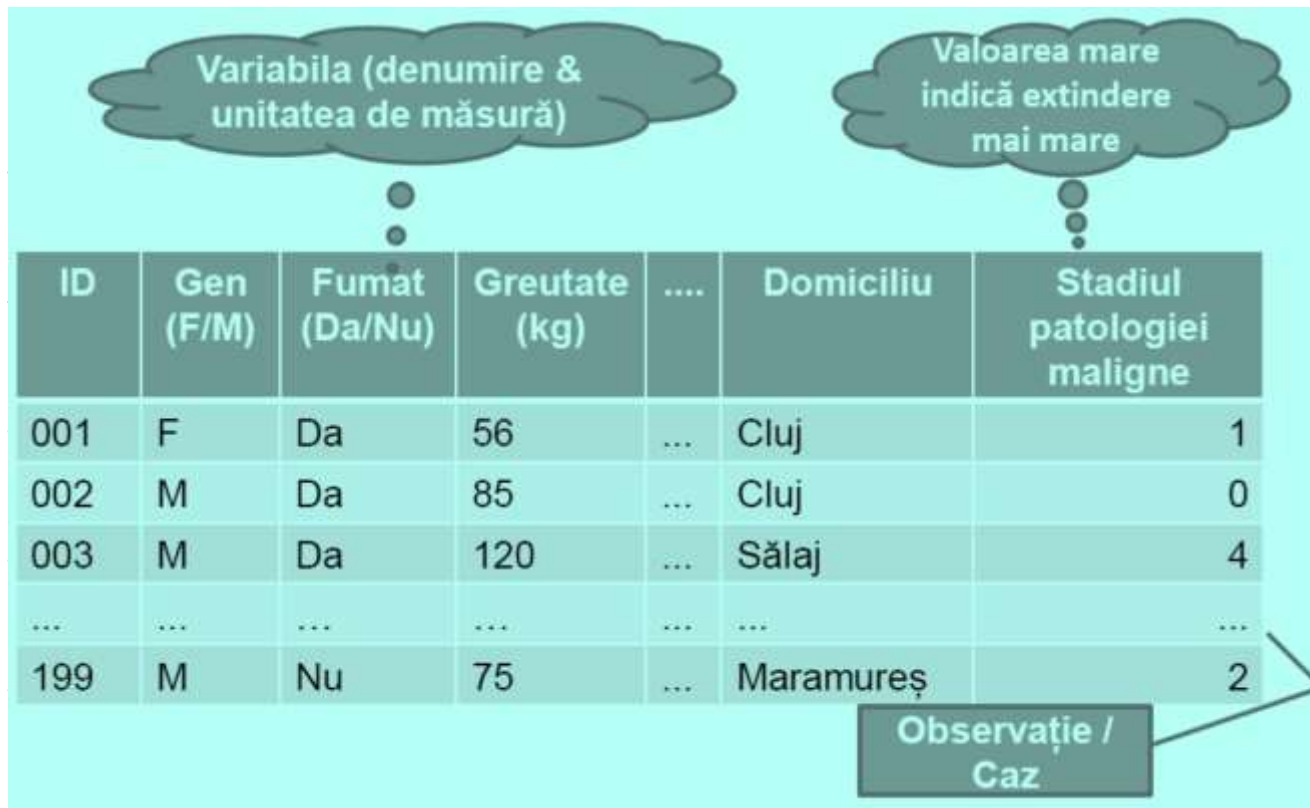


MĂSURAREA ȘI ACURATEȚEA

- Date cantitative → nivel informațional mare
- Cuantificarea:
 - Cantitativ prin natura variabilei
 - Bazat pe măsurătoare
- 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)

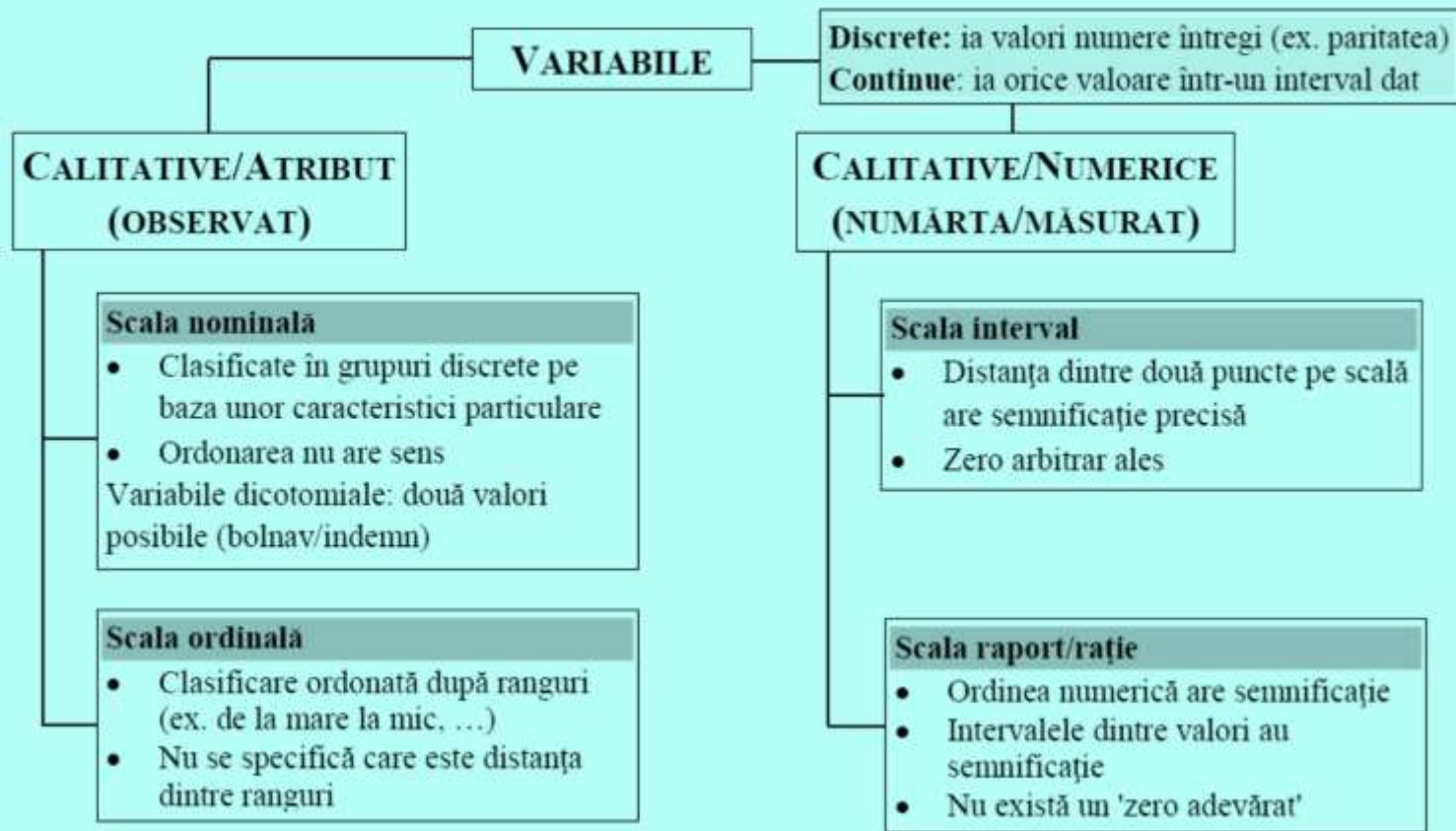


The diagram illustrates a data matrix structure. Two clouds at the top provide context: the left cloud, 'Variabila (denumire & unitatea de măsură)', points to the column headers; the right cloud, 'Valoarea mare indică extindere mai mare', points to the numerical values in the last column. A callout box labeled 'Observație / Caz' points to the last column header.

<div>Variabila (denumire & unitatea de măsură)</div> <div>Valoarea mare indică extindere mai mare</div>						
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

TIPURI DE VARIABLE/ DATE



EXEMPLE DE VARIABLE

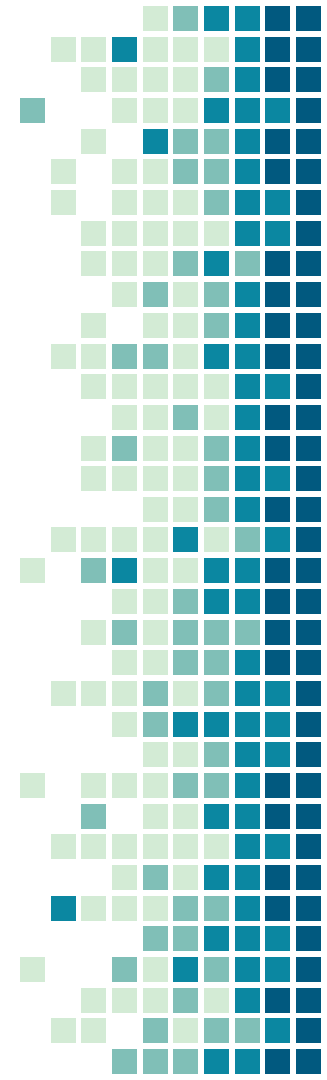


Nominal	Ordinal
Culoarea părului; Educația; Starea civila (căsătorit, divorțat, văduv, necăsătorit) ... Variabile dicotomiale: Gen (F/M)	Starea de sănătate: slabă / medie / bună / foarte bună Starea la externare: vindecat / ameliorat / agravat / staționar / decedat
Interval	Rație/Raport
Temperatura (diferența dintre 50 °C și 60°C este aceeași ca și diferența dintre 100 °C și 110 °C)	PAS (mmHg); Greutatea (kg); Lungimea (cm); Volumul (cm ³); Timpul de reacție (sec)



LET'S PRACTICE

- M / F
- 1 = Masculin; 2 = Feminin; 3 = Nedefinit
- Fumător/ Nefumător
- Grupa sangvină: 0 / A / B / AB
- Indexul tratamentului ortodontic: 1-5 unde 1 = necesitatea cea mai mică de tratament, 5 = necesitatea cea mai mare a tratamentului
- Numărul de vizite la medicul de familie într-un an
- Numărul de episoade dureroase
- Excelent / Bun / Rezonabil / Prost (igiena orală)
- Mică / Moderată / Severă (durere)
- Înălțime (cm)
- Greutate (kg)
- Presiunea arterială sistolică (mmHg)



• **Scorul APGAR:** test utilizat la măsurarea semnelor vitale ale nou-născutului



1. Heart Rate:
 - a. Absent heartbeat = 0.
 - b. Slow heartbeat (less than 100 beats per minute) = 1.
 - c. Adequate heartbeat (more than 100 beats per minute) = 2.
2. Respiration:
 - a. Not breathing = 0.
 - b. Weak cry, irregular breathing = 1.
 - c. Strong cry = 2.
3. Muscle Tone:
 - a. Limp, flaccid = 0.
 - b. Some flexing or bending = 1.
 - c. Active motion = 2.
4. Response to Stimulation (also called Reflex Irritability):
 - a. No response = 0.
 - b. Grimace = 1.
 - c. Vigorous cry or withdrawal = 2.
5. Color:
 - a. Pale or blue = 0.
 - b. Normal color body but blue extremities = 1.
 - c. Normal color = 2.

TRANSFORMAREA VARIABILELOR

- Este posibilă transformarea scalelor de măsură tip interval și raport/rație în scale de tip ordinal/nominal dar această transformare 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/ordinal în scală de tip interval/raport chiar dacă atribuim valori diferitelor clase
 - Sex: M = 1, F = 0

VARIABILE 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.

EXEMPLE DE TRANSFORMĂRI



Cholesterol	Value (mg/dL)
LDL (low density lipoprotein cholesterol)	
Optimal	< 100
Near/Above optimal	100-129
Borderline high	130-159
High	160-189
Very high	≥ 190

HDL (high density lipoprotein cholesterol)	
Low HDL cholesterol	< 40
High HDL cholesterol	≥ 60
TG (Triglycerides)	
Normal	< 150
Borderline high-TG	150-199
High	200-499
Very high	≥ 500

SERII STATISTICE (variabile cantitative)

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

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

» Bidimensională: PAS (mmHg) și PAD (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

SERII STATISTICE (variabile calitative)

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

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

Bivariată &
Unidimensională

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

Multivariată &
Multidimensională

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

POPULAȚIE 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 2020 ai UMFST
GE Palade din Tîrgu Mureș

Eșantion

Eșantionul - utilizare

- 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



Eșantionul - utilizare

- 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:

- (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.

EXEMPLU

Determinarea empirică a volumului eșantionului



Volumul populației (N)	Volumul eșantionului (n) - %
0 -100	1
101 – 1 000	3
1 001 – 5 000	5
5 001 – 10 000	10
> 10000	100

Mai multe detalii în:

Gabor, M.R. – *Prospectarea pieței prin metoda statistică*, Ed. CH Beck, 2013, pag. 106 - 209

Metode de eșantionare



Metode de eșantionare

Eșantionare probabilistică

Fiecare unitate statistică din populație are o șansă mai mare de 0 de a fi selectată în eșantion

Permite estimarea erorii de eșantionare

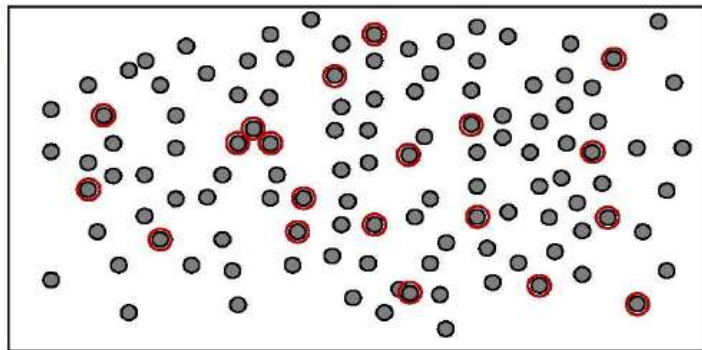
Unele unități statistice din populația teoretică/de studiu nu au nicio șansă de a fi selectate și incluse în eșantion.

NU permite estimarea erorii de eșantionare

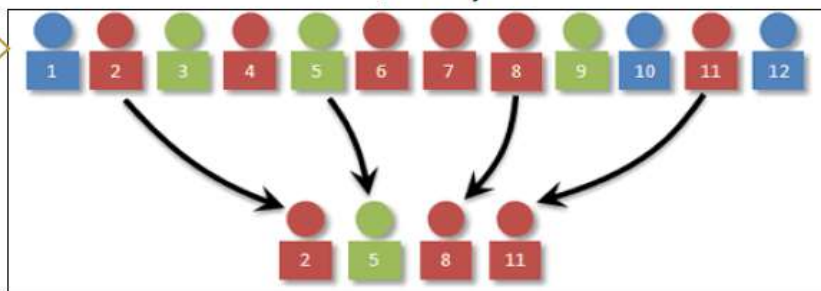
Eșantionare non-probabilistică

Eșantionare probabilistică

Simплу randomizată

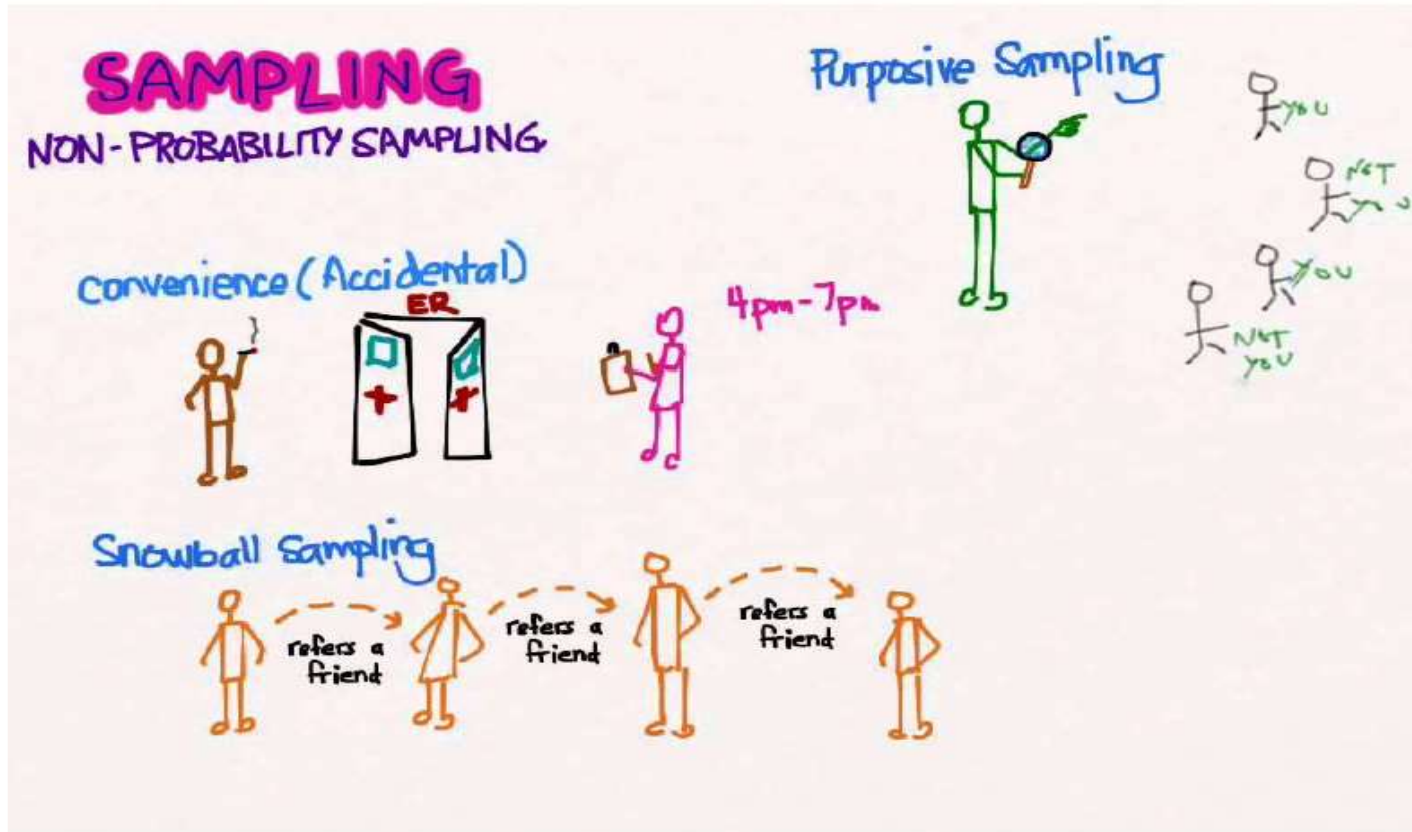


Eșantionarea sistematică

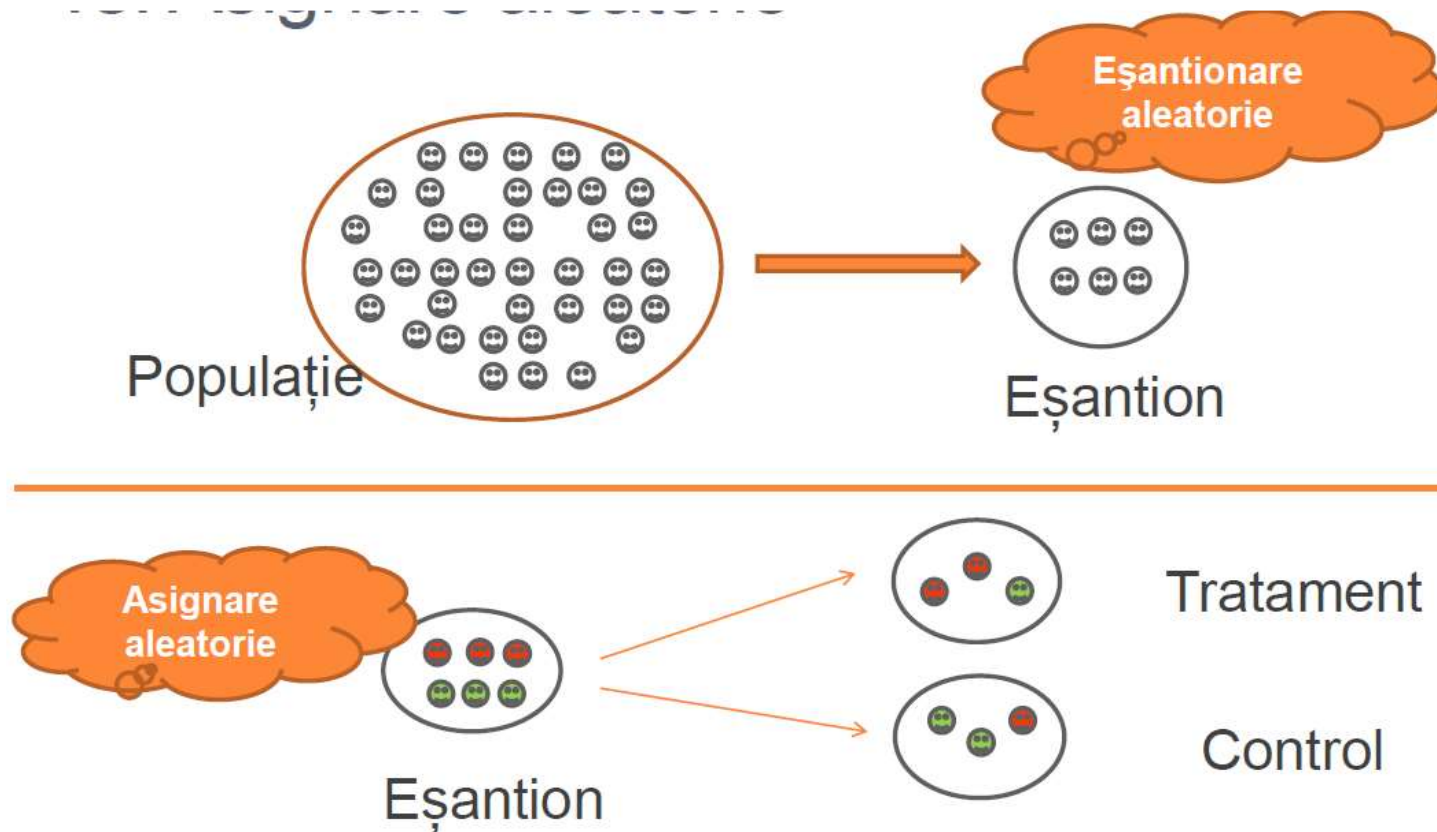


Eșantionul (tot al treilea individ)

Eșantionare non-probabilistică



Eșantionarea probabilistică/aleatoare vs alocare (atribuire/asignare) aleatorie





DE REȚINUT!

- Identificarea corectă a **VARIABLELOR** investigate permite analiza statistică corectă.
- Întotdeauna când este 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 **INFERENȚA** rezultatelor obținute pe eșantion la nivelul populației.
- Evaluarea întregii populații **NU** se realizează deoarece este costisitoare, consumatoare de timp, și de cele mai multe ori imposibilă. Poate duce inclusiv la distrugerea populației.

Nu cred în statistici. Cred în calcule.

citat din Ben Horowitz

THANKS!

Any questions?

You can find me at:
manuela.gabor@umfst.ro

