

Studying medicine is
like true love..
You suffer, fail, learn,
fail again, fight and grow,
but you never want to let go.

NOȚIUNI INTRODUCTIVE DE STATISTICĂ MEDICALĂ

Sorana D. Bolboacă

Cuprins ...

Obiective educaționale

- Definirea unor termeni statistici
- Identificarea și explicarea diferenței dintre statistica descriptivă și statistica inferențială
- Înțelegerea diferenței dintre dată și informație, respectiv dată și variabilă
- Recunoașterea și clasificarea diferitelor tipuri de variabile
- Definirea termenilor de populație și eșantion

Glosar

Statistica = domeniu științific care permite studiul fenomenelor, parametrilor și caracteristicilor a căror proprietate fundamentală este variabilitatea

Statistica medicală = știința colectării datelor, a prezentării lor într-o formă sistematică, a analizei acestora și a interpretării rezultatelor obținute

Se ocupă cu culegerea, sintetizarea, descrierea și interpretarea datelor medicale

Unitatea statistică = elementul constitutiv al populației statistice

Statistica descriptivă vs. inferențială

Statistica medicală se ocupă de aplicațiile biostatisticii în medicină și științele sănătății, inclusiv epidemiologia, sănătatea publică, criminalistica

utilizate pentru a generaliza rezultatele obținute prin studierea unui eșantion la populația din care acesta a fost extras



Date – Informații – Cunoștințe

În medicină, noțiunea de dată nu este echivalentă cu noțiunea de informație

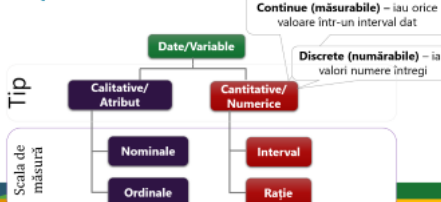
Ce?	Exemplu
Data	Semne vitale: ritm cardiac, ritm respirator, temperatura corpului, tensiune arterială
Informația	Semne vitale colectate de-a lungul timpului, plasate într-un context
Cunoștințe	Identificarea tiparelor și intervenția pe baza informațiilor
Înțelepciune	Acuratețea sintezei informației și selecția adecvată a intervenției

Data vs. Variabilă

Variabilă (denumire și unitate de măsură)					
ID	Gen (F/M)	Vârsta (Da/No)	Greutate (kg)	...	Domiciliu
001	F	Da	56	--	Ciuj
002	M	Da	85	--	Ciuj
003	M	Da	120	--	Săliș
...
199	M	Nu	75	--	Maramureș

Observație / Caz

Tipuri de variabile/date



Populația

= mulțime de elemente (obiecte sau subiecți) care au anumite însușiri (atribute sau caracteristici) comune, care formează obiectul unei analize statistice

- Talia sau volumul populației = numărul elementelor populației
- Unitatea statistică = un element al populației

Exemple:

- un grup de pacienți
- o mulțime de obiecte
- un grup de fenomene sau evenimente ...

Eșantionul

Eșantionul = o colecție (submulțime) de elemente din populație

Dependente	Independente
<ul style="list-style-type: none">membrii aleși pentru un eșantion determină automat care membri trebuie să fie incluși în al doilea eșantiondeterminarea multiplei pe aceeași itemi determinarea TAS la același pacient la includerea în studiu și respectiv la 6 luni de la începerea tratamentuluielementele unui eșantion sunt selectate pentru a se potrivește cu elementele celui de-al doilea	<ul style="list-style-type: none">Membrii aleși într-un eșantion nu determină ce membri sunt aleși pentru cel de-al doilea eșantionprobabilitatea ca un element să fie inclus în eșantion NU depinde de selecția unui alt element în cel de-al doilea eșantionsursete diferite de itemi/obiecte/subiecți

De reținut!



Obiective educaționale

- Definirea unor termeni statistici
- Identificarea și explicarea diferenței dintre statistica descriptivă și statistica inferențială
- Înțelegerea diferenței dintre dată și informație, respectiv dată și variabilă
- Recunoașterea și clasificarea diferitelor tipuri de variabile
- Definirea termenilor de populație și eșantion



Glosar

Statistica = domeniu științific care permite studiul fenomenelor, parametrilor și caracteristicilor a căror proprietate fundamentală este variabilitatea

Statistica medicală = știința colectării datelor, a prezentării lor într-o formă sistematică , a analizei acestora și a interpretării rezultatelor obținute

Se ocupă cu culegea, sintetizarea, descrierea și interpretarea datelor medicale

Unitatea statistică = elementul constitutiv al populației statistice



Glosar

Populația = mulțimea entităților (persoane, ființe vii, obiecte, măsurători) sau elementelor de aceeași natură care au caracteristici comune. Criterii de includere și excludere

Eșantionul = un grup reprezentativ al unei populații

Variabila = caracteristică care poate prezenta variație de la un element la altul al unei colectivități;

Date = valorile observate și înregistrate ale unei variabile statistice

Informațiile statistice = rezultă din prelucrarea și analiza datelor statistice

Glosar

Populație

- Numărul elementelor
= volumul / talia /
dimensiunea
populației
- **N**

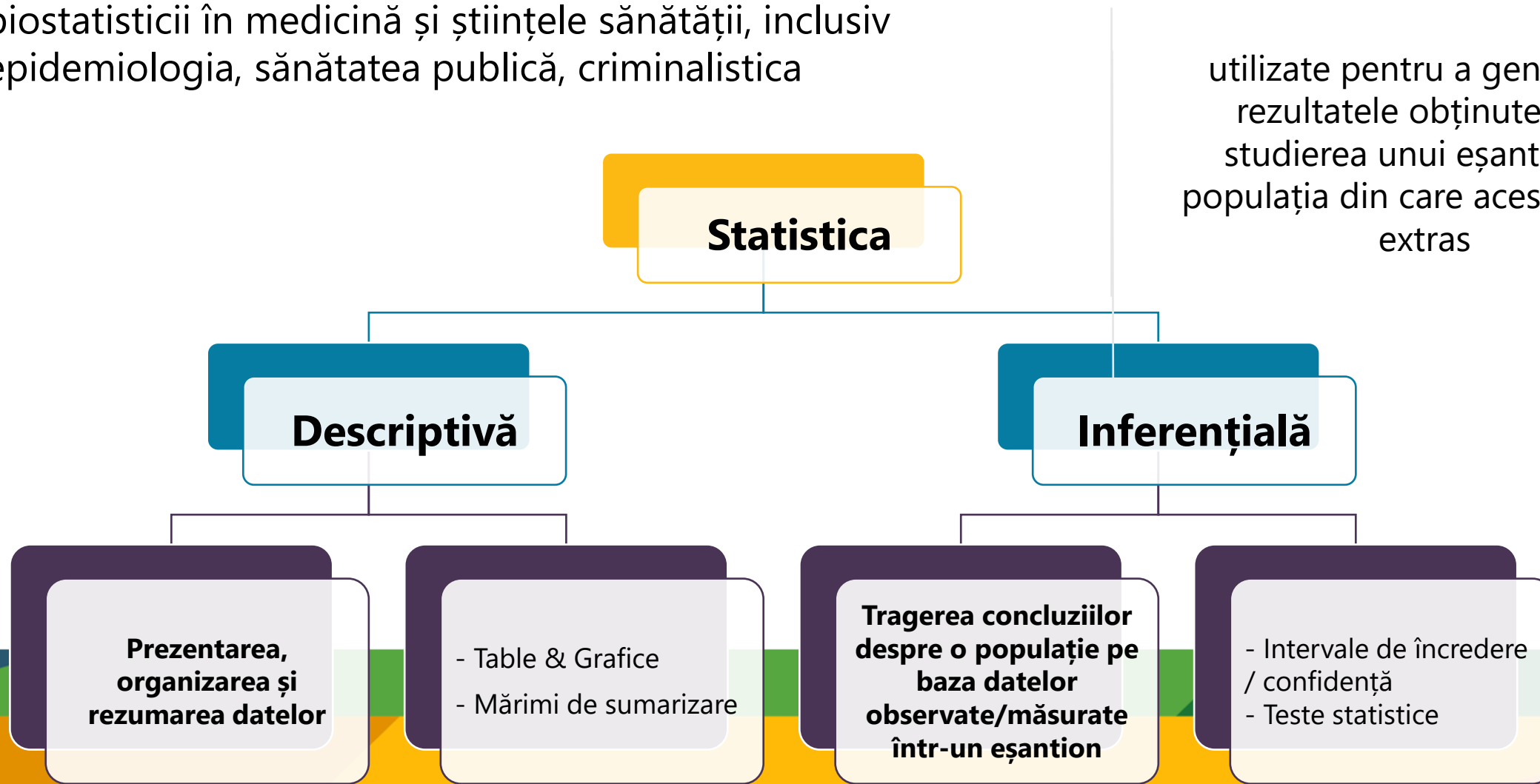
Eșantion

- Numărul elementelor
= volumul / talia /
dimensiunea
eșantionului
- **n**

Statistica descriptivă vs. inferențială

Statistica medicală se ocupă de aplicațiile biostatisticii în medicină și științele sănătății, inclusiv epidemiologia, sănătatea publică, criminalistica

utilizate pentru a generaliza rezultatele obținute prin studierea unui eșantion la populația din care acesta a fost extras



Descriptiv sau Inferențial?

Descriptiv		Inferențial
Ce ...	descrie	oferă informații asupra unei populații prin studierea unui eșantion.
Referă ...	subiecții evaluați	populația
Volum ...	mic	mare
Eșantionare ...	nu neapărat	obligatorie
Aplicat la ...	o situație particulară	extrapolarea la populație

Descriptiv sau Inferențial?

	Descriptiv	Inferențial
Cu ce ...	parametrii statistici: <ul style="list-style-type: none"> • centralitate (medie, mediană), • dispersie (cât de departe sunt datele de parametrii de centralitate) • asimetrie (distribuția datelor e simetrică/asimetrică) 	<ul style="list-style-type: none"> • Testarea ipotezelor statistice (teste statistice) (<i>sunt mediile similare sau diferite?</i>) • Intervale de încredere (<i>cât de mare/mică este diferența?</i> relevanța clinică) • Analiza de regresia (<i>care este relația dintre cel puțin 2 variabile</i>)
Forma ...	grafice tabele	comparații Probabilități
Face ...	Organizarea și prezentarea într-un mod organizat astfel încât rezultatele să poată fi interpretare	Compară Testează Estimează Prezice

Statistica descriptivă vs. inferențială

Descriptiv

The prevalence of HDV has not diminished further in Europe. In blood samples collected from HBsAg patients in 2006–2007 in Italy, the overall prevalence of anti-HD remained 8.1% with no further reduction compared with 1997. However, among incident cases, the prevalence was 14.3%, suggesting the arrival of a new wave of HDV-infected individuals.



Inferențial

... 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. ...



Descriptiv sau Inferențial?

?

„Prevalența insuficienței mitrale (IM) la sportivii de performanță nu este semnificativ diferită de cea a populației generale”



?

“Raportăm opt pacienți care au dezvoltat tulburări neuromusculare în săptămânile care au urmat vaccinării împotriva COVID-19, cu o etiologie presupusă mediată imun.”



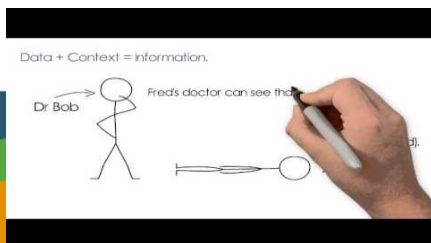
Date – Informații – Cunoștințe

În medicină, noțiunea de dată nu este echivalentă cu noțiune de informație

Ce?	Exemplu
Data	Semne vitale: ritm cardiac, ritm respirator, temperatura corpului, tensiune arterială
Informația	Semne vitale colectate de-a lungul timpului, plasate într-un context
Cunoștințe	Identificarea tiparelor și intervenția pe baza informațiilor
Înțelepciune	Acuratețea sintezei informației și selecția adecvată a intervenției

Date – Informații – Cunoștințe

Date	220, 89, femeie
Informații	O femeie de 89 ani (Sara) cu presiunea sistolică de 220 mmHg → Date + Context = Informație
Cunoștințe	Medicul lui Sara a observat că presiunea arterială sistolică (PAS) a crescut semnificativ și i-a administrat un medicament pentru scăderea acesteia. → Informație + Reguli = Cunoștințe Informațiile devin cunoștințe când aplicăm reguli. În acest caz, regulile cu privire la valorile normale ale presiunii arteriale sistolice au fost aplicate după identificarea creșterii PAS.



<https://www.youtube.com/watch?v=QsP5WGV0aQc>

Dată vs. Variabilă

Variabila (denumire
& unitatea de
măsură)

Data (plural: **date**) = o valoare
observată sau măsurată care
corespunde unei caracteristici

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

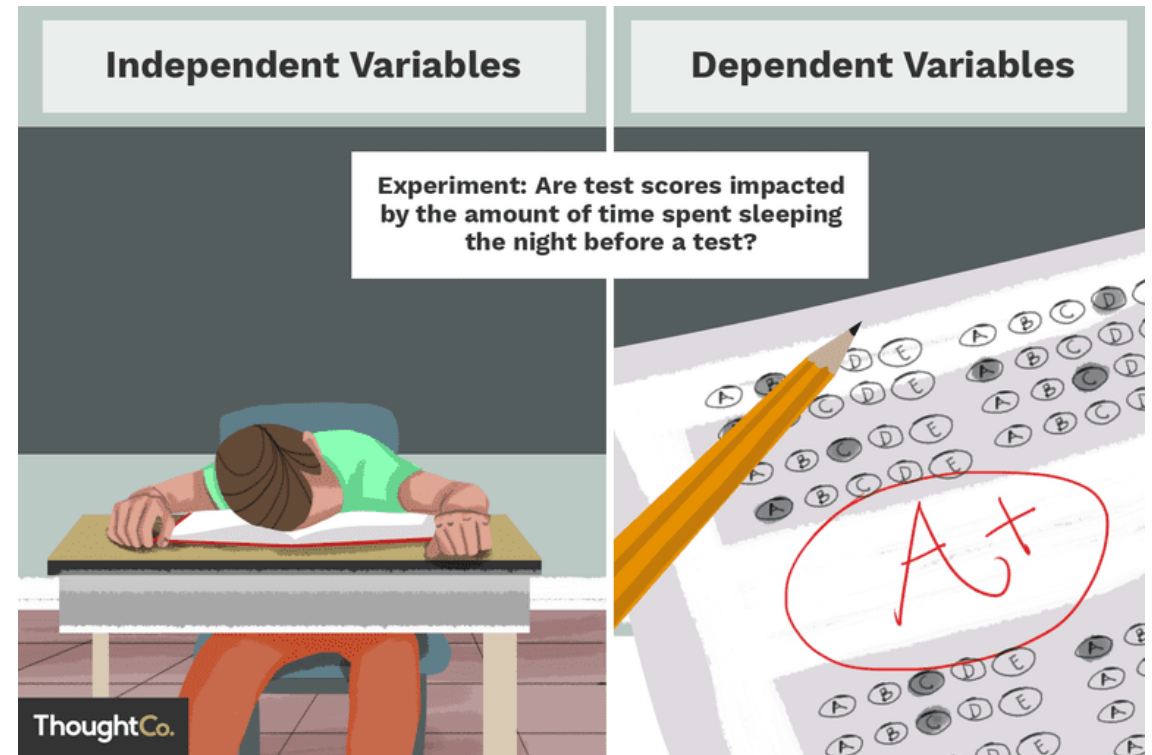
Dată

Variabile

Independente: factorul care este schimbat de persoana care face experimentul

Dependente: factorul care este măsurat/observat în experiment

Constante: factorul care nu-și schimbă valoarea



Variabile & date

Care este procentul de studenți fumători în cadrul Facultății de Medicină?

- Variabila: FUMATOR
- Data: răspunsul de tip DA/NU dat de fiecare student în parte

Studiile de cercetare și analizele statistice: studiul variabilelor



Variabila vs. data

Variabila

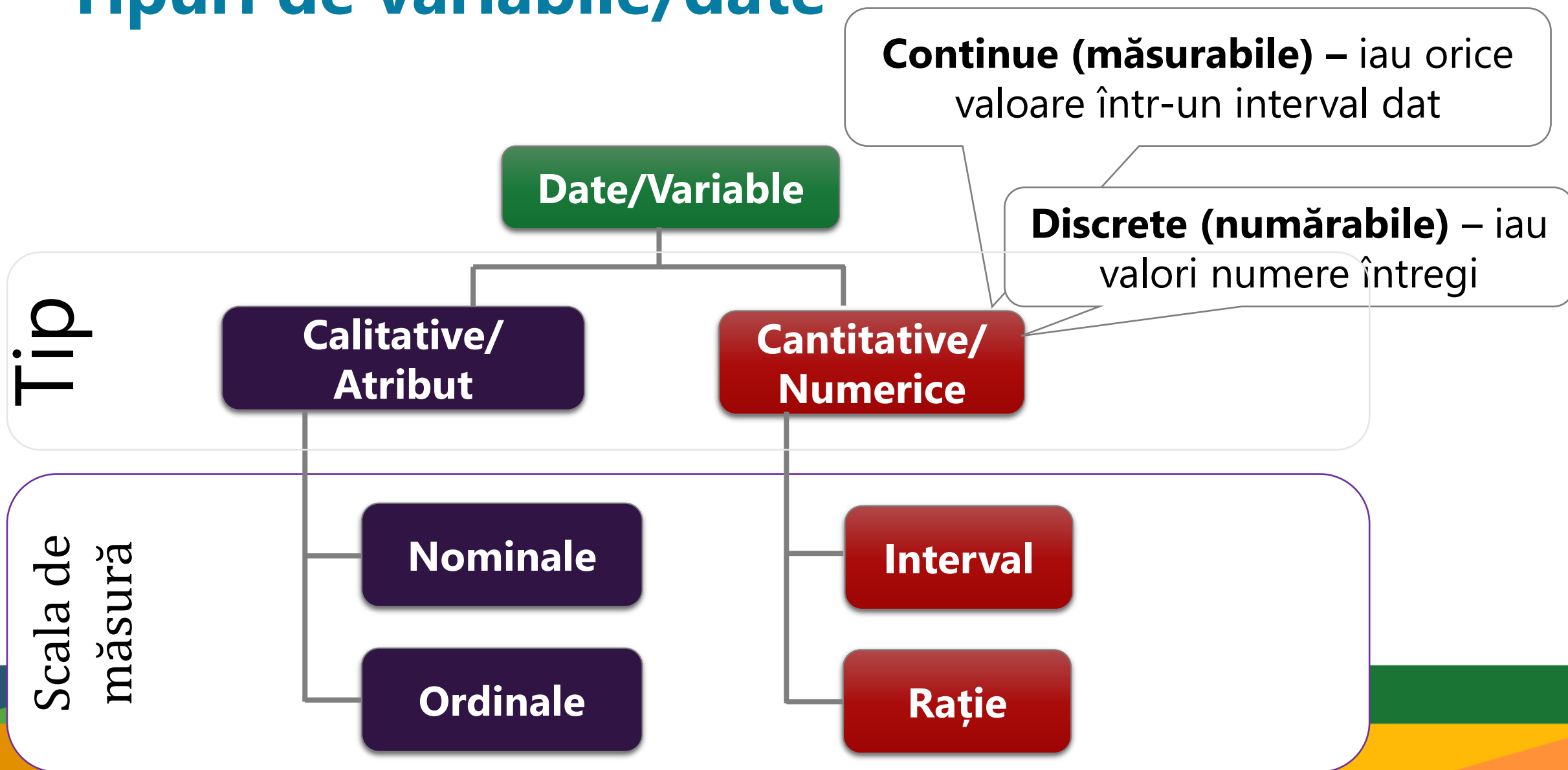
	A	B	C	D	E	F	G
	ID	Vârsta (ani)	Sex (F/M)	Greutate (kg)	Înălțime (cm)	RC /min	PAS (mmHg)
1							
2	1	87	F	81	187	63	115
3	2	82	F	76	168	57	130
4	3	64	M	97	170	54	98
5	4	37	M	78	174	70	127
6	5	62	F	76	172	60	114
7	6	43	M	86	172	61	166
8	7	83	M	75	175	69	119
9	8	64	M	84	176	65	148
10	9	24	M	78	172	67	151
11	10	93	M	67	170	66	145
12	11	56	M	73	166	63	133
13	12	77	M	93	166	61	89
14	13	76	M	80	159	82	95

Dorim să identificăm dacă hipertensiunea arterială sau diabetul zaharat sunt responsabile de cele mai multe internări în secția interne a spitalului din Dej și care sunt costurile asociate acestor internări. Am colectat din baza de date a spitalului următoarele variabile pentru toți pacienții internați în luna martie 2022 (disponibile în fișierul BD-DateSpitalizare.xlsx): Vârsta (ani), Sex (F/M), Greutate (kg), Înălțime (cm), ritmul cardiac (RC/min), presiunea arterială sistolică (PAS (mmHg)), presiunea arterială diastolică (PAD (mmHg)), Glicemia (mg/dL), Colesterol (mg/dL), trigliceride (TG (mg/dL)), HDL (mg/dL), durata spitalizării (DS (zile)).

- Variabila?
- Data?

Data

Tipuri de variabile/date



Date și variabile: Tipuri și scale de măsură

Tip	Scala	Definiție
Calitative	Nominală	Valorile sunt incluse în categorii sau grupuri pe baza unor caracteristici comune. Ordonarea nu are sens.
	Dicotomială	Poate lua doar 2 valori (ex., prezent/absent, da/nu, etc.)
	Ordinală	Valori ordonate după ranguri dar distanța dintre două valori nu este aceeași pentru oricare două valori. Se aplică expresii de genul mai mare decât sau mai mic decât.

Date și variabile: Tipuri și scale de măsură

Tip	Scala	Definiție
Cantitative	Interval	Intervale ordonate de lungime egală. Valoarea zero are semnificați. (nu are zero absolut) Valorile pot fi discrete sau continue.
	Raport	Intervalele ordonate au lungime egală și raportul dintre numere are semnificație. Există zero absolut. Valorile pot fi discrete sau continue.

Variabile cantitative

Interval

Intervale ordonate

Lungime egală

Zero arbitrar

- 30° este cu 20° mai cald decât 10° ,
dar nu este de 3 ori mai cald decât
 10°

- 0° e temperatura de îngheț

Raport/Rație

Intervale ordonate

Lungime egală

Zero absolut

- 40 kg este de două ori mai mult
decât 20 kg și de 4 ori mai mult
decât 10 kg

- 0 kg ?

Observația, variabilă, și date

Gender (F/M)	Ethnicity	Living place (R/U)	Weight (g)	Length (cm)	Cranial perimeter (cm)	Thoracic perimeter (cm)
F	Rrom	R	900	38	28	27
F	Romanian	R	950	40	24	25
M	Romanian	R	1150	47	32	31
...
F	Hungarian	U	1200	48	28	27
F	Rrom	U	980	40	25	26

Observație

dată

variabilă

? Calitativ/Cantitativ

distrofie corneeană afectează de obicei ambii ochi și progresează lent

Grade of Fuchs dystrophy (Krachmer Grading Scale):

- Grade 0 (*No apparent disease*)
- Grade 1 (*Definitive onset of the disease*)
- Grade 2 (*A zone of confluent central guttae 1 to 2 mm in horizontal width*)
- Grade 3 (*A zone of confluent central guttae 2 to 5 mm wide*)
- Grade 4 (*A zone of confluent central guttae greater than 5 mm wide*)
- Grade 5 (*A zone of confluent central guttae greater than 5 mm wide plus edema of the corneal stroma and/or corneal epithelium*)

? Calitativ/Cantitativ

APGAR (**A**ppearance, **P**ulse, **G**rimace, **A**ctivity, **R**espiration) score: a quick test performed on a baby at 1 and 5 minutes after birth introduced by Apgar in 1952



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.

? Calitativ/Cantitativ

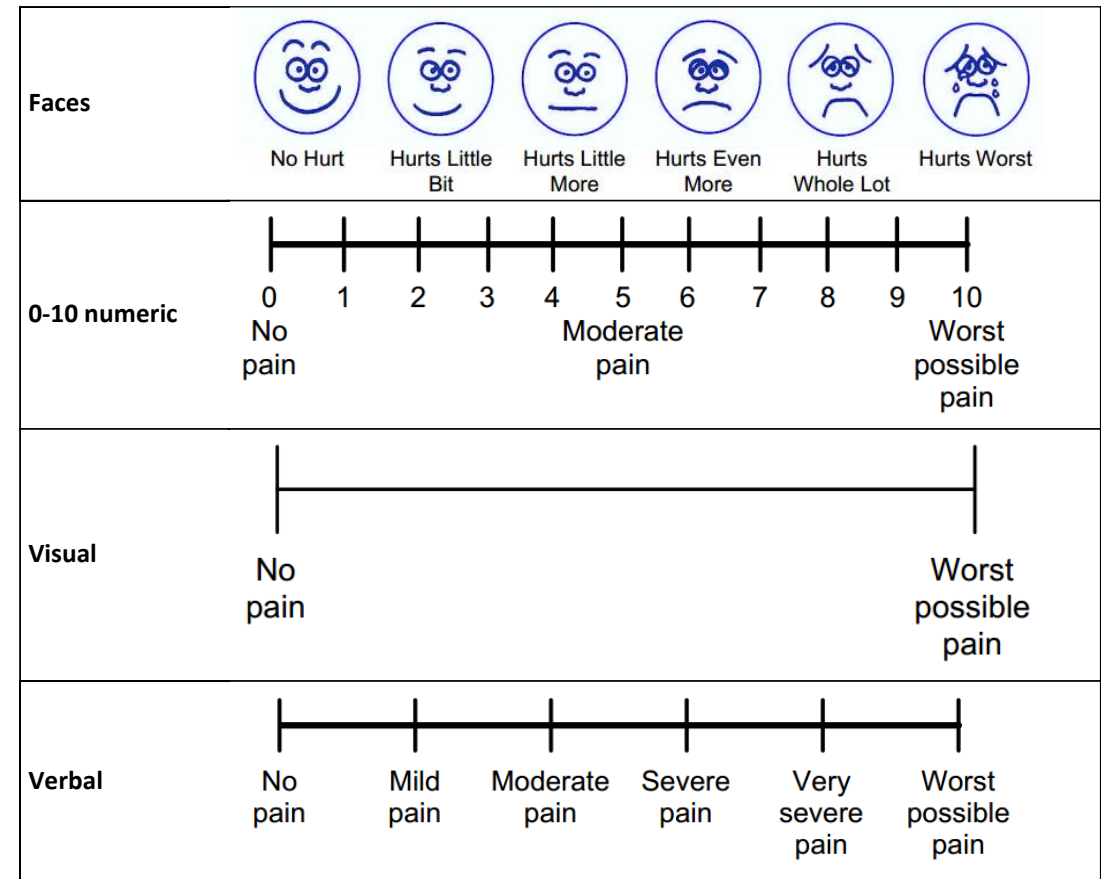
Durerea: diferite scale de evaluare

Wong-Baker FACES

Numerică 0–10

Vizual

Verbal



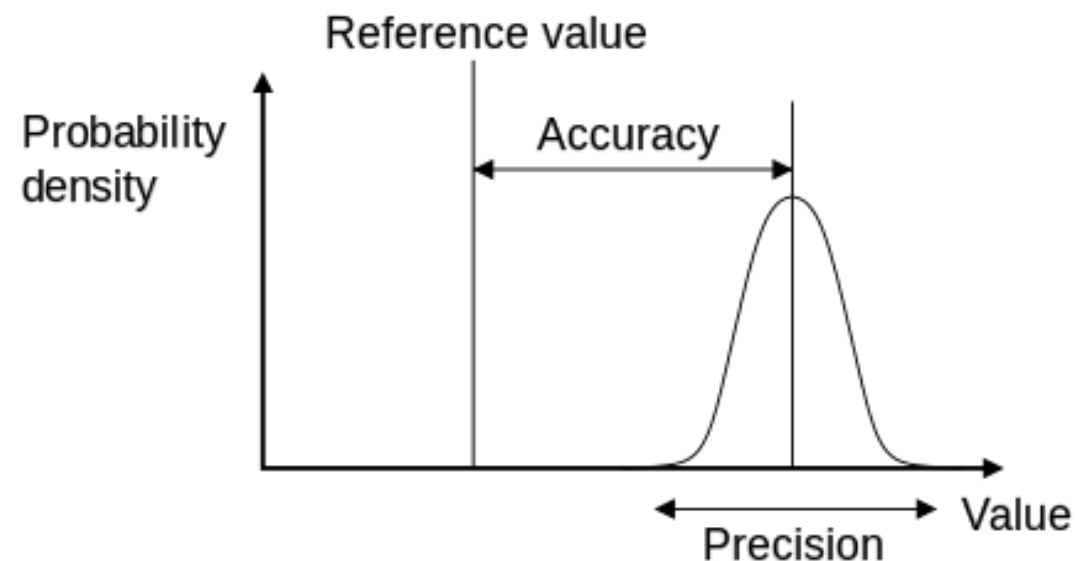
? Calitativ/Cantitativ

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

Date cantitative măsurate: precizie vs. acuratețe

Acuratețe	Precizie
Gradul în care valoare măsurată a standardului se apropie de valoarea de referință cunoscută	Gradul de apropiere a măsurătorilor repetate
Eroare sistematică	Repetabilitate Reproductibilitate
Cât de aproape este valoarea măsurată de valoarea adevărată (reală)?	Cât de aproape sunt valorile rezultate din măsurători repetate?



https://en.wikipedia.org/wiki/Accuracy_and_precision (citată în 06/10/2022)

Date cantitative măsurate

2 (Ioana și Dan) studenți măsoară independent perimetrul cranian al aceleiași nou născut de gen feminin. Fiecare student face trei măsurători utilizând același instrument

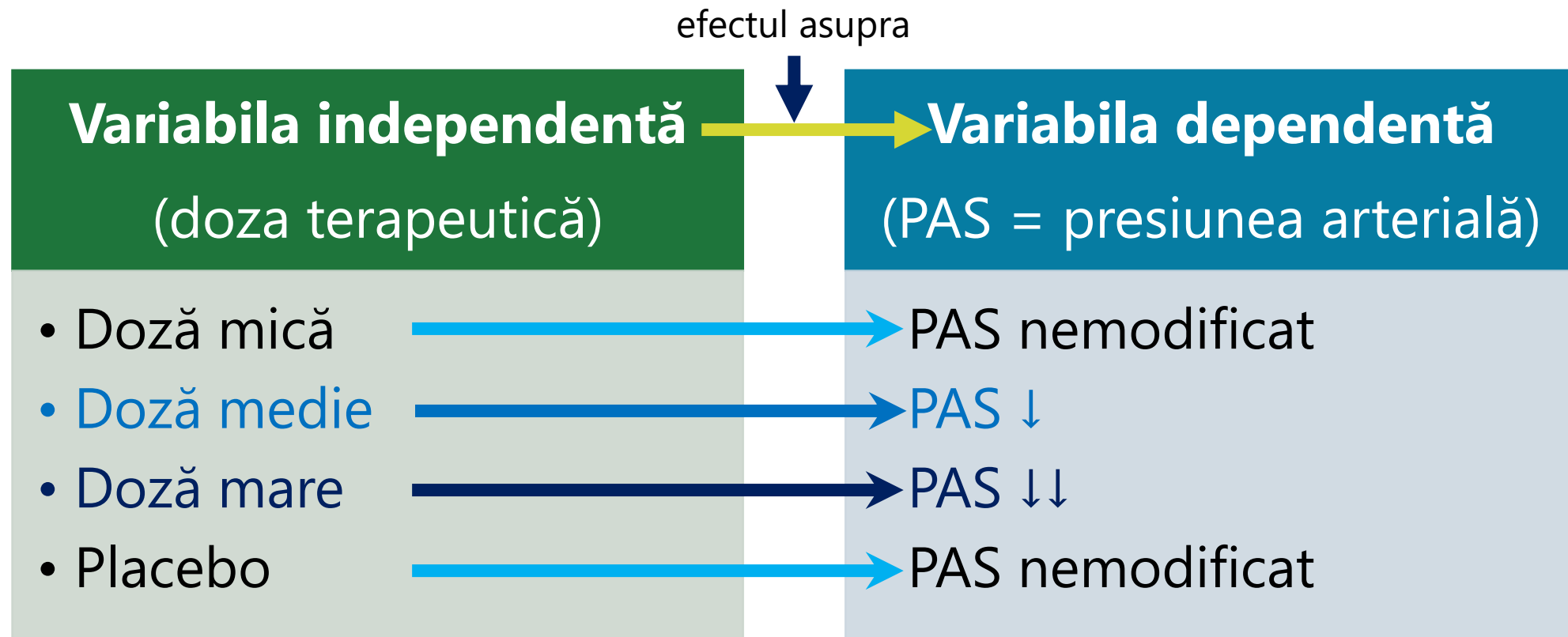


Ioana	Dan
31,7	32,3
32,0	32,0
31,8	31,9

Care din cei doi studenți au raportat măsurători mai precise?

Știm că nou născutul are un perimetru cranian egal cu 32 cm. Indicați studentul cu măsurătoarea cea mai acurată.

Variabile dependente vs. independente



Aplicăm diferite doze de tratament

Măsurăm efectul terapeutic asupra PAS

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: $M = 1$, $F = 0$



Variabile derivate

Cantitative

PAM (Presiune Arterială Medie)

- $= PAD + (0.33 + (RC * 0.0012)) * PP$
- PAD = presiunea arterială diastolică
- RC = rata pulsului
- PP = PAS – PAD, unde PAS = presiunea arterială sistolică

Calitative

- HTA (HiperTensiune Arteriala):
- $PAM < 90$ indică normotensiune;
- $90 \leq PAM < 92$ indică TA de graniță;
- $92 \leq PAM < 96$ indică HTA grad I;
- $PAM > 96$ indică HTA grad II.

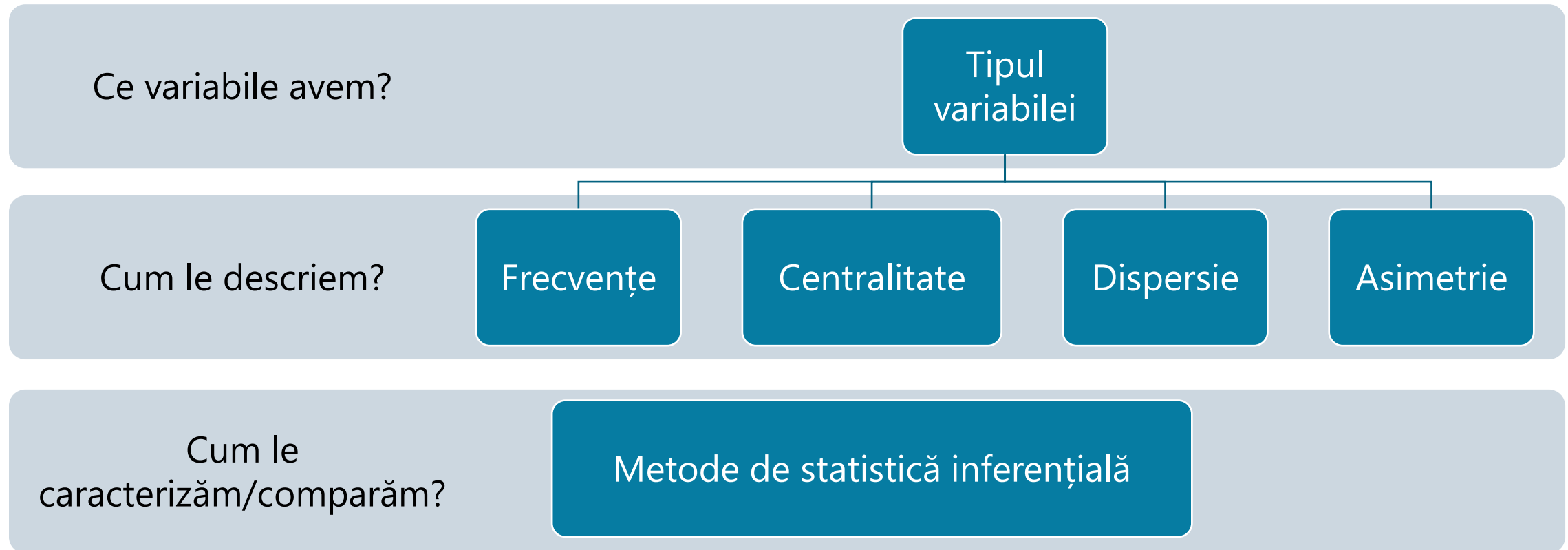
Variabile derivate

	A	F	G	H	I	J
1	ID	RC /min	PAS (mmHg)	PAD (mmHg)	PAM	HTA
2	1	63	115	75	88	normotensiune
3	2	57	130	65	87	normotensiune
4	3	54	98	70	79	normotensiune
5	4	70	127	61	83	normotensiune
6	5	60	114	57	76	normotensiune
7	6	61	166	76	106	HTA grad II
8	7	69	119	68	85	normotensiune
9	8	65	148	84	105	HTA grad II
10	9	67	151	78	102	HTA grad II
11	10	66	145	68	94	HTA grad I
12	11	63	133	60	84	normotensiune
13	12	61	89	83	85	normotensiune
14	13	82	95	92	93	HTA grad I

$=H2+(0,33+(F2*0,0012))*(G2-H2)$

$=IF(I2>96,"HTA grad II",IF(I2>=92,"HTA grad I",IF(I2>=90,"TA de graniță","normotensiune"))))$

De ce e important tipul variabilei/datelor




Populația

= mulțime de elemente (obiecte sau subiecți) care au anumite însușiri (atribute sau caracteristici) comune, care formează obiectul unei analize statistice

- Talia sau volumul populației = numărul elementelor populației
- Unitatea statistică = un element al populației

Exemple:

- un grup de pacienți
 - o mulțime de obiecte
 - un grup de fenomene sau evenimente ...
- 

Eșantionul

Eșantionul = o colecție (submulțime) de elemente din populație

Dependente

- membrii aleși pentru un eșantion determină automat care membri trebuie să fie incluși în al doilea eșantion
- determinări multiple pe aceeași itemi: 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

Independente

- Membrii aleși într-un eșantion nu determină ce membri sunt aleși pentru cel de-al doilea eșantion
- probabilitatea ca un element să fie inclus în eșantion **NU** depinde de selecția unui alt element în cel de-al doilea eșantionul
- seturi diferite de itemi/obiecte/subiecți

Eșantionul: De ce?

Eșantionul se poate evalua mai repede decât populația.

Evaluarea unui eșantion este mai ieftină comparativ cu studierea întregii populații.

Studiul întregii populații poate determina distrugerea ei.

Rezultatele obținute pe eșantion sunt adesea mai precise comparativ cu rezultatele obținute pe populație.

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

Eșantionul: dependent / independent?

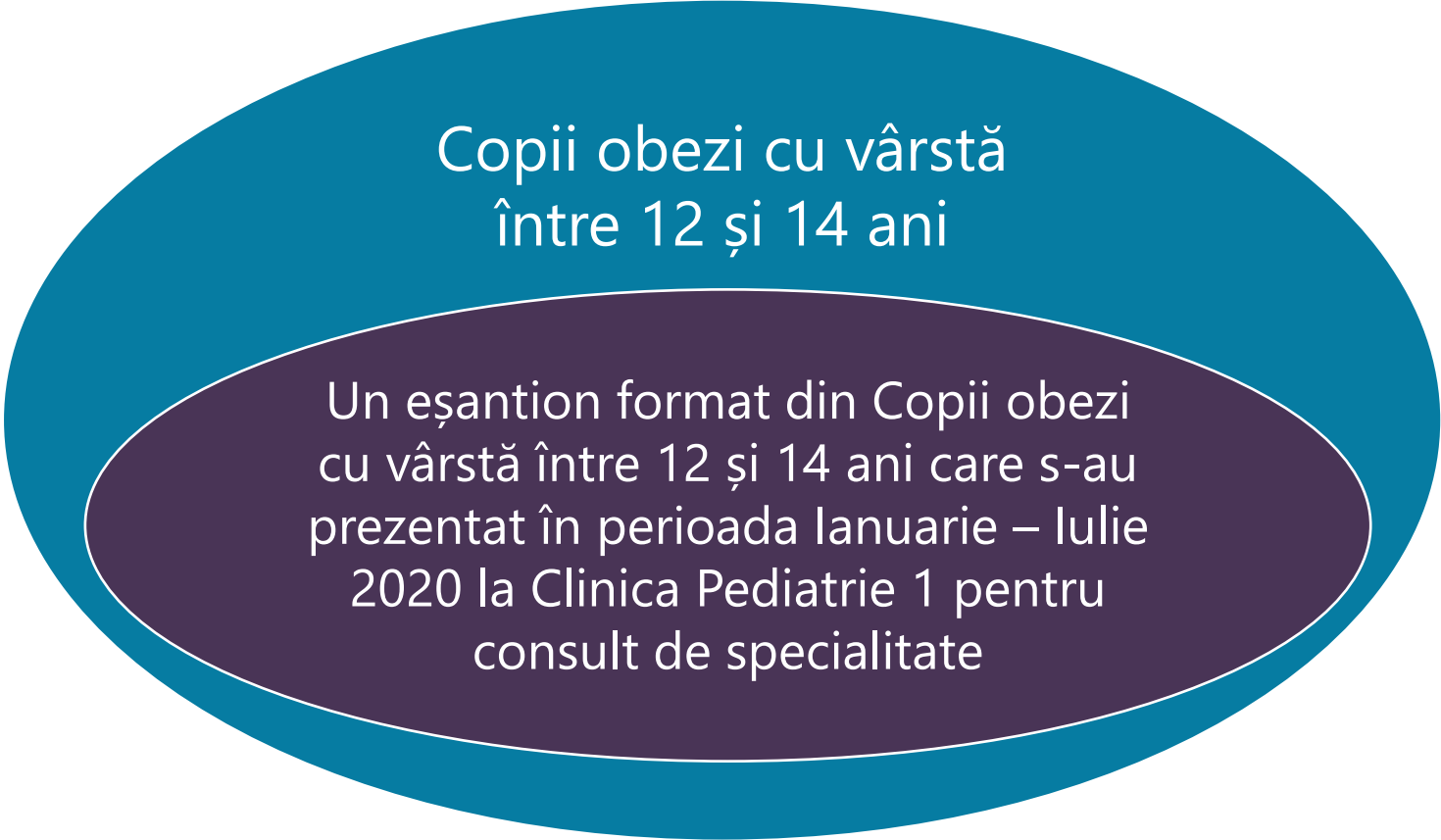
	A	B	C	D
1	ID	Grup	Sex	Varsta (ani)
2	1	diabet	m	63
3	2	diabet	m	70
4	3	diabet	f	63
79	78	diabet cu complicatii	f	79
80	79	diabet cu complicatii	f	59
81	80	diabet	f	64
82	81	diabet cu complicatii	f	69
83	82	diabet cu complicatii	m	69

Vârsta pacienților cu diabet cu complicații (“diabet cu complicatii” coloanal B / variabila **Grup**) este diferită de cea a pacienților cu diabet fără complicații (“diabet” coloanal B / variabila **Grup**)?

	A	B	C	D	E
1	Tipul interventiei	Gen	BMIInitial	BMI3luni	BMI1an
2	Dieta 1	F	46	42	43
3	Dieta 1	F	34	28	29
4	Dieta 1	F	36	35	37
5	Dieta 1	F	36	30	30
6	Dieta 1	F	39	36	43
7	Dieta 1	F	48	46	47
8	Dieta 1	F	40	43	43

A avut Dieta 1 (Coloana A, “Tipul intervenției”) efect asupra Indicelui de Masă Corporală? (BMI = Bogy Mass Index = Indice de masă corporală)

Populație vs. eșantion

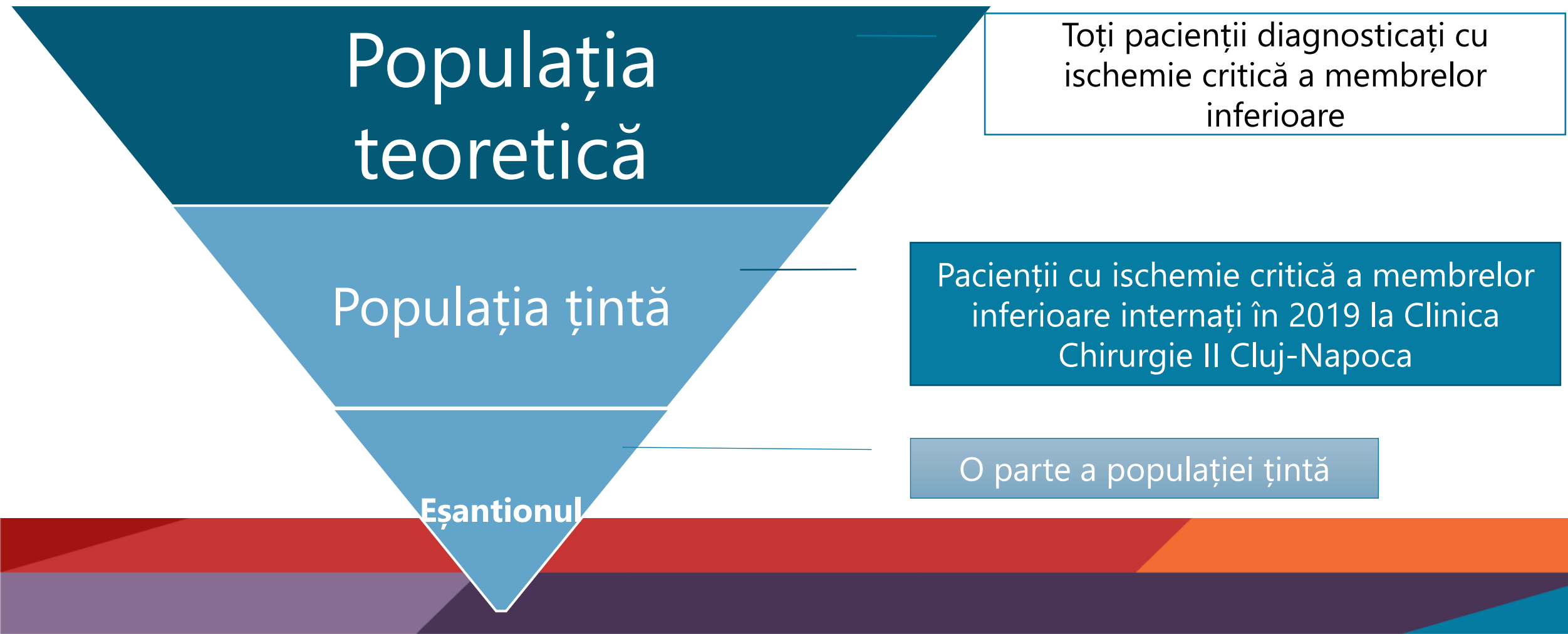


Copii obezi cu vârstă
între 12 și 14 ani

The diagram consists of two nested ovals. The outer oval is teal and contains the text 'Copii obezi cu vârstă între 12 și 14 ani'. The inner oval is dark purple and contains the text 'Un eșantion format din Copii obezi cu vârstă între 12 și 14 ani care s-au prezentat în perioada Ianuarie – Iulie 2020 la Clinica Pediatrie 1 pentru consult de specialitate'. The bottom of the slide features a decorative horizontal band with geometric shapes in red, orange, purple, and teal.

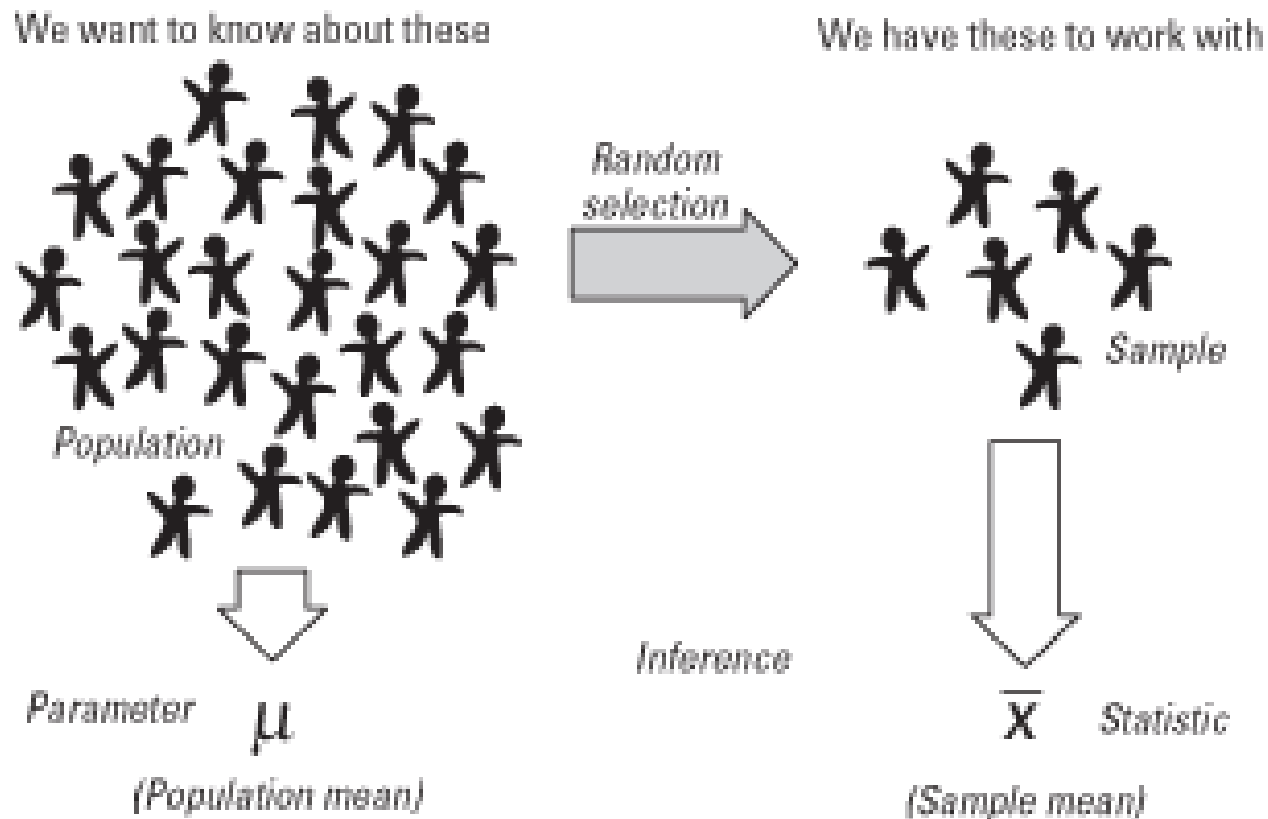
Un eșantion format din Copii obezi
cu vârstă între 12 și 14 ani care s-au
prezentat în perioada Ianuarie – Iulie
2020 la Clinica Pediatrie 1 pentru
consult de specialitate

Măsurarea nivelului seric de cotinină la pacienții cu ischemie critică a membrelor inferioare (*critical limb ischemia*)



Populație vs. Eșantion

Parametru vs. Statistică



Parametrul
populației – ex.
media
populației μ

Statistica
eșantionului –
media
eșantionului m

Obținut prin măsurarea tuturor
indivizilor **populației**

Obținut prin măsurarea
indivizilor din **eșantion**

Parametru vs. statistica

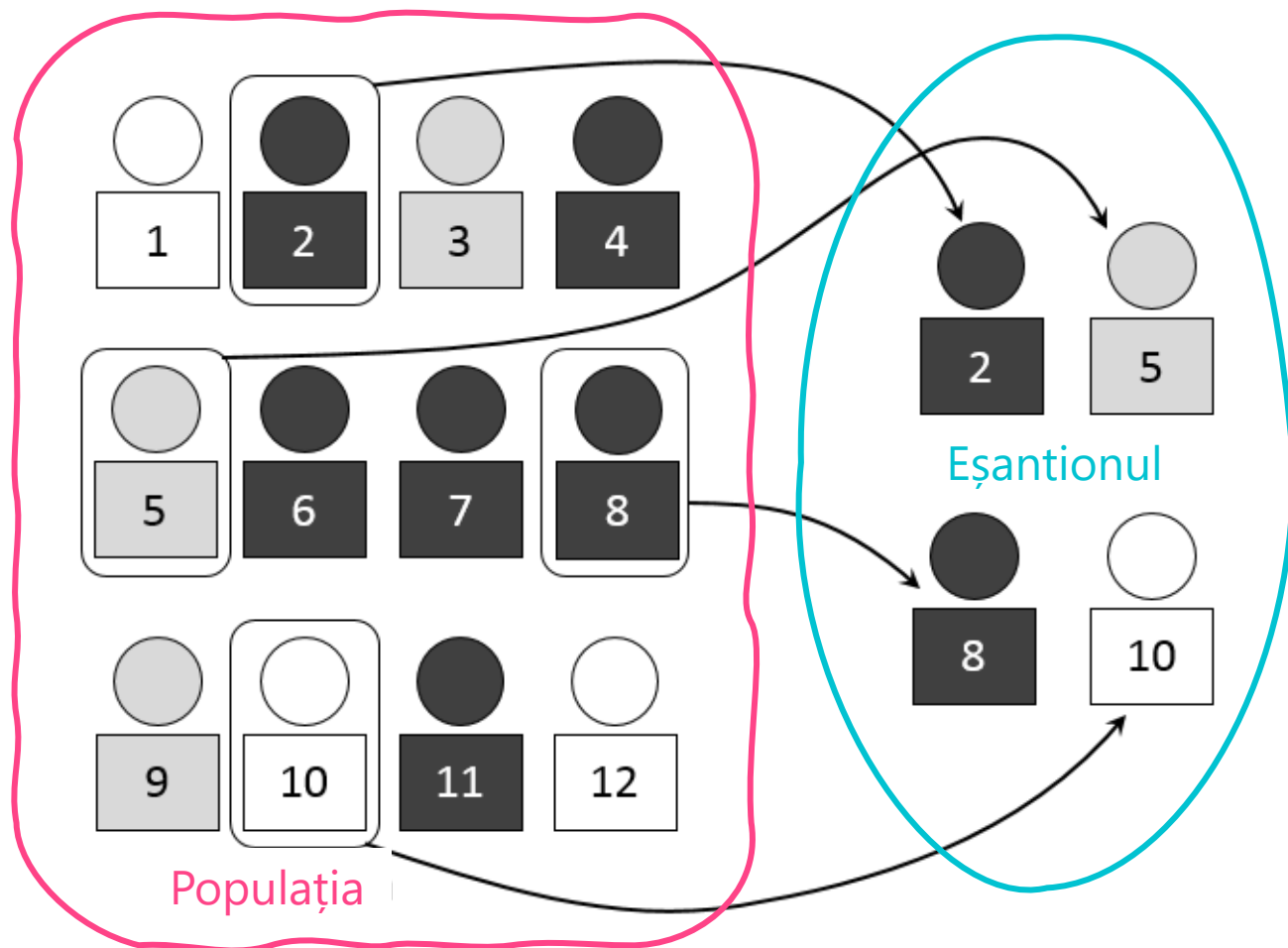
Parametru

- *Hepatitis C is an infectious disease caused by the hepatitis C virus (HCV) with a prevalence of ~2-3% of the world's population.*

Statistica

- *Suppose a sample of 100 health care professionals is obtained and from this sample it is found that 4.3% has HCV.*

Populația vs. eșantionul



Se dorește evaluarea consumului de cocaină în rândul studenților Universității de Medicină și Farmacie Iuliu Hațieganu Cluj-Napoca.

- Populația?
- Eșantionul?

De reținut!

Populația \neq
Eșantionul

Parametrul \neq
Statistica

Data \neq Informația

Data \neq Variabila

Studiem variabile NU
date

Statistica descriptivă
 \neq Statistica
inferențială

Avem 2 tipuri de
variabile și 4 scale de
măsură

O variabilă cantitativă
se poate transforma
în una calitativă dar
nu și vice-versa

Tipul de variabilă va
dicta analiza statistică
descriptivă și
inferențială

Dependent vs.
Independent: variabile

Dependent vs.
Independent:
eșantioane