



UMF
UNIVERSITATEA DE
MEDICINĂ ȘI FARMACIE
IULIU HAȚIEGANU
CLUJ-NAPOCA

METODE DE SUMARIZARE A DATELOR

Sumar curs anterior...VARIABLE

VARIABLE CANTITATIVE:

- continue
- discrete

VARIABLE CALITATIVE:

- dihotomiale
- nominale
- ordinale



Plan de curs ...

Prezentarea TABELARĂ a datelor

Prezentarea GRAFICĂ a datelor

Scenariu...

Id_pacient	Varsta (ani)	Sex	Judet	Dislipidemie	Alcool	Fumat	TAS (mmHg)	TAD (mmHg)	HTA	IMC (kg/m2)	Status ponderal (0=normal, 1= supraponderal, 2= obezitate clasa I, 3 = obezitate clasa II)
1	27	F	AB	nu	nu	nu	110	80	da	20.20	0
2	41	M	BH	da	nu	nu	130	70	da	34.22	2
3	67	M	BN	nu	nu	nu	170	100	da	31.49	2
4	60	F	CJ	nu	nu	nu	190	120	da	37.91	3
5	26	M	CV	nu	nu	nu	130	90	da	27.74	1
6	46	M	HR	nu	nu	nu	170	90	da	36.23	3
7	34	M	MM	nu	nu	da	110	70	nu	26.54	1
8	31	M	MS	nu	nu	nu	110	70	nu	25.25	1
9	45	M	SB	nu	da	da	100	70	nu	25.59	1
10	35	M	SJ	nu	nu	nu	130	90	da	33.80	2
11	64	F	SM	nu	nu	nu	130	100	da	28.20	1
12	64	F	AB	nu	nu	nu	170	100	da	32.89	2
13	34	F	BH	nu	nu	nu	110	80	da	23.81	0
14	35	M	BN	nu	nu	da	110	70	nu	19.05	0
15	41	M	CJ	nu	da	nu	110	79	nu	30.08	2
16	49	F	CV	nu	nu	nu	130	100	da	35.76	3
17	64	F	HR	nu	nu	nu	160	110	da	32.89	2
18	43	M	MM	nu	nu	nu	120	80	da	24.38	0
19	58	M	MS	da	nu	nu	145	85	da	30.30	2
20	44	M	SB	nu	nu	nu	150	100	da	26.64	1
21	45	F	SJ	nu	nu	nu	120	80	da	21.64	0
22	25	M	SM	nu	nu	da	130	80	da	22.66	0
23	62	F	AB	nu	da	nu	130	80	da	30.44	2
....

Obiectiv studiu: evaluarea potentialilor factori de risc pentru H'IA la subiectii adulti cu varsta pana în 50 ani.

Esantion de studiu: 62 pacienti adulți spitalizați

Variabile: ??? Tipuri de variabile: ?

Noțiuni statistice de bază: *populația țintă, accesibilă*

- **Populația**= ansamblu de obiecte/subiecți pe care dorim să o studiem

Exemple:

- ⊗ toți pacienții adulți spitalizați

Populația *țintă*: populația pe care dorim să o studiem, populația căreia dorim să extrapolăm rezultatele la sfârșitul studiului pe un eșantion

Populația *accesibilă*: populația la care avem acces: cabinet medical/stomatologic, spital, școală etc.

Populația țintă



Populația accesibilă

Pacienți adulți spitalizați



Pacienți adulți
urmărire într-un anumit spital

Variable statistice versus Date

Variabila: caracteristica (demografică, clinică etc.) studiată pe diferiți indivizi (aparținând unui eșantion)

- ✓ Variabile calitative (dihotomiale, ordinale, nominale) et cantitative (discrete, continue)
- ✓ Variables independente si variabile dependente

Data: „valoarea” (poate un număr sau un tip/categorie „da”/„nu”) a variabilei
una sau mai multe variabile pot fi măsurate pe un individ ->

Serie statistică de date = valorile variabilelor măsurate în timpul unui studiu



Notiuni de baza: Tipul unei serii statistice

Număr de variabile	Serie statistică
1	uni variata
2	bi variata
3	tri variata
>3	multi variata

Număr de variabile CANTITATIVE	Serie statistică
1	uni dimensională
2	bi dimensională
3	tri dimensională
>3	multi dimensională

Tipuri de serii statistice: EXEMPLE

- **unidimensională:** Vârsta (ani)

	x_1	x_2	x_3	x_4	x_5	x_6	x_{62}
Vârsta (ani)	27	41	45	50	26	46		33

- **bidimensională:** Varsta (ani) și TAS(mmHg)

	x_1	x_2	x_3	x_4	x_5	x_6	...	x_{62}
Vârsta (ani)	20	45	70	67	60	35	...	55
TAS (mmHg)	110	130	170	190	130	100	...	95

- **univariată:** Dislipidemie (da=1;nu=0)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_{62}
Dislipidemie	0	1	0	0	0	1	0		0

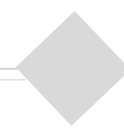
Reguli în prezentarea datelor

- Reguli în prezentarea tabelară
- Reguli în prezentarea grafică
- Prezentarea datelor pentru:
 - 1 variabilă calitativă
 - 1 variabilă cantitativă
 - 2 variabile calitative
 - 2 variabile cantitative
 - 1 variabilă calitativă & 1 variabilă cantitativă



Principii de sumarizare tabelară

1. Simple: de preferat 2/3 tabele mai mici în loc de unul încărcat
2. Informative prin ele însele
 - Abrevieri sau simboluri explicate la subsolul tabelului
 - Etichete de rând și coloană
 - Unități de măsură
 - Titlul: ce? când? unde?
 - Linii și/sau coloane de sinteză (total)
3. Dacă datele nu sunt originale trebuie să se menționeze sursa lor într-o notă de subsol



TABELUL DE FRECVENȚĂ

- **POTRIVIT PENTRU:**
- 1 variabilă calitativă
- 1 variabilă cantitativă (valori individuale / clase de frecvență)

Sumarizare tabelară: Cazul unei variabile CALITATIVE:

- Se determină frecvența fiecărei valori/categorie a variabilei
- Se includ valorile/categoriile distincte (disjuncte) și frecvențele lor într-un tabel pe două coloane:
 - **Frecvența absolută** = numărul de cazuri care îndeplinesc criteriul
 - **Frecvența relativă** = raportul dintre frecvența absolută și volumul eșantionului/populației (simbol = n).
 - Frecvențele relative se pot prezenta și procentual.
- Microsoft Excel: funcția [COUNTIF](#) ([video tutorial](#)); [Tabele Pivot](#) [Data - Pivot Table]

TABELUL DE FRECVENȚĂ

- Se pot alcătui tabele de frecvențe care să conțină pe lângă frecvențele absolute și cele relative și următoarele tipuri de frecvențe :

Tipul de frecvență	Definiție
Frecvența absolută cumulată crescător a categoriei (x_i)	numărul de indivizi din eșantion pentru care caracterul studiat poate lua o valoare mai mică sau egală cu x_i
Frecvența absolută cumulată descrescător a categoriei (x_i)	numărul de indivizi din eșantion pentru care caracterul studiat poate lua o valoare mai mare sau egală cu x_i
Frecvența relativă cumulată crescător a categoriei (x_i)	proporția de indivizi din eșantion pentru care caracterul studiat poate lua o valoare mai mică sau egală cu x_i
Frecvența relativă cumulată descrescător a categoriei (x_i)	proporția de indivizi din eșantion pentru care caracterul studiat poate lua o valoare mai mare sau egală cu x_i

TABELUL DE FRECVENȚĂ: 1 VARIABILĂ CALITATIVĂ

Suma frecvențelor relative ale tuturor valorilor seriei care sunt mai mici sau egale decât x/n

Suma frecvențelor absolute ale tuturor valorilor seriei care sunt mai mici sau egale decât x

Grad HTA	f_a	f_r	f_a cumulată crescător (\uparrow)	f_r cumulată crescător (\uparrow)
0	19	19	30.65	30.65
1	23	42	37.10	67.74
2	16	58	25.81	93.55
3	4	62	6.45	100.00
Total	62			

Notății: 0 = normală, 1= HTA grad 1; 2 = HTA grad 2; 3= HTA grad 3.

TABELUL DE FRECVENȚĂ: 1 VARIABILĂ CANTITATIVĂ

Procentul de subiecți care au avut valoarea IMC-ului în intervalul dat

Câți subiecți au avut valoarea IMC-ului în intervalul dat?

Care este procentul de subiecți cu IMC mai mic sau egal cu o valoare (ex. 24 kg/m²)

IMC (kg/m ²)	f_a	f_r	f_r cumulată ↑
≤ 24	13	20.97	20.97
(24; 29]	27	43.55	64.52
(29; 34]	16	25.81	90.32
(34; 39]	5	8.06	98.39
(39; 44]	1	1.61	100.00
Total	62	100	

TABELUL DE FRECVENȚĂ

E1: Zilele de spitalizare pentru ultimii 10 pacienți internați în salonul 13 al secției de Boli Infecțioase sunt 5, 6, 7, 7, 8, 8, 5, 7, 8, 7.

Căreia dintre valorile de mai jos îi corespunde frecvența relativă cumulată crescător de 0,7???

A. 8

B. 6

C. 5

D. 7

E. Nici un răspuns nu este corect

Zile de spitalizare	f_a	f_r	f_r cumulată ↑
5	2	$= 2/10 = 0,2$	$= 0,2$
6	1	0,1	$= 0,2 + 0,1 = 0,3$
7	4	0,4	0,7
8	3	0,3	1
Total	10		

Răspuns corect = D

TABELUL DE CONTINGENȚĂ: 2/X VARIABLE CALITATIVE

Tabel 1. Tabel de contingenta de 2×2

	HTA (da)	HTA (nu)	Total
Feminin	21	6	27
Masculin	24	11	35
Total	45	17	62

2 variabile **CALITATIVE DIHOTOMIALE**:
Sex (F/M) și HTA(da/nu)

Tabel 2. Tabel de contingenta de 6×4

	AB	BH	BN	CJ
Hepatită A	166	171	16	50
Hepatită B	13	14	9	9
Hepatită C	1	25	4	7
Alte tipuri de Hepatită	0	8	6	0
Hepatită cronică	0	0	12	9
Purtători cronici de HBsAg	21	53	14	2

2 variabile **CALITATIVE NOMINALE**:
Comorbiditati și Județ (AB, BH,...)

Principii de sumarizare grafică

- Orice reprezentare grafică trebui să aibă:
 - Titlul
 - Definirea axelor
 - Unități de măsură pentru fiecare axă (dacă este cazul)
 - Legendă (dacă este cazul)
- O reprezentare grafică trebuie să se “înțeleagă” singură!
 - Fără a se citi textul!!!

Principii de sumarizare grafică

- Scopul unei reprezentări grafice este de a transmite o informație
- Când construim o reprezentare grafică trebuie să răspundem la întrebarea: *Care este scopul acestei reprezentări?*
- Datele trebuie reprezentate grafic în așa fel încât să fie utile în înțelegerea fenomenului clinic
- Atenție la compoziția culorilor (nu puneți fundaluri colorate) și la dimensiunea caracterelor!

Reprezentarea grafică

- **COLOANE/BARE:**

- Histograma pentru variabile cantitative (clase de frecvență): alegerea intervalelor este importantă (alegerea inadecvată a acestora poate influența modelul distribuției)
- Variabile calitative: scală ordinală sau nominală – modulul

- **PLĂCINTA (PIE):** permite identificarea proporțiilor

- Variabile calitative

- **LINIE:**

- Secvență de timp
- Ordinea evenimentelor



Reprezentarea grafică

- **GRAFIC DE TIP PUNCTE (SCATTER)**

- ✓ pune în evidență relația dintre **2 variabile CANTITATIVE**
- ✓ se reprezintă grafic o pereche de valori cantitative pentru fiecare pacient în parte
- ✓ indică dacă punctele sunt împrăștiate la întâmplare sau nu.

- Distribuția de frecvențe 3D

- Reprezentarea a mai multor caracteristici
- Frecvența fiecărei coloane este citibilă dar cititorul trebuie să extrapoleze înălțimea coloanei, extrapolarea putând fi distorsionată din cauza 3D



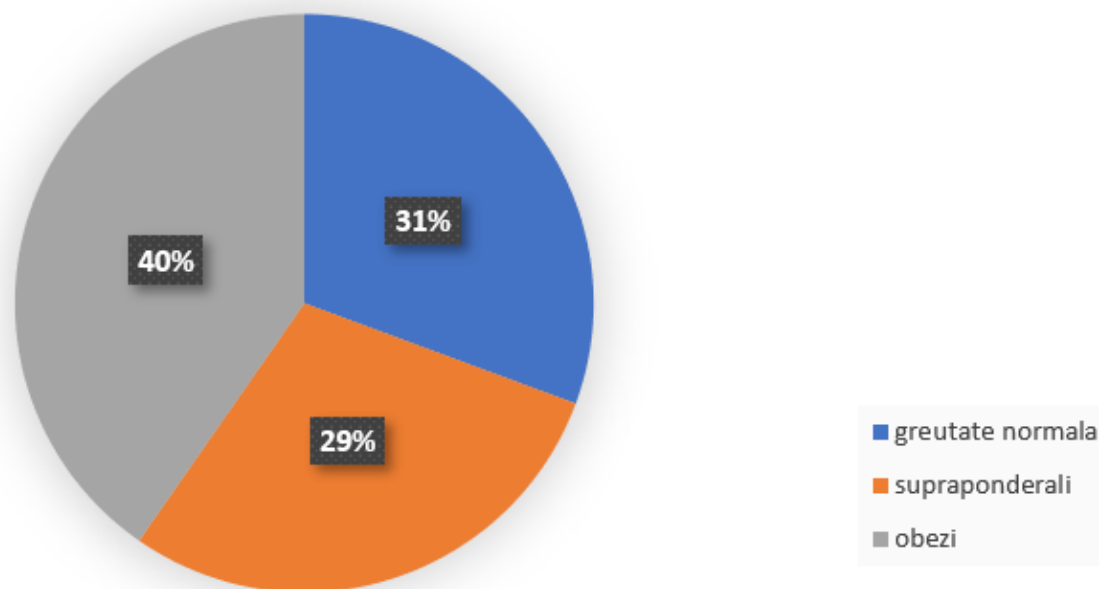
Reprezentarea grafică

VARIABLE CALITATIVE	VARIABLE CANTITATIVE
<p>Pentru descrierea unei SINGURE variabile calitative</p> <ul style="list-style-type: none">- grafic sectorial (pie)- grafic prin bare (coloane)	<p>Pentru descrierea unei SINGURE variabile cantitative</p> <ul style="list-style-type: none">- histograma- grafic cutie cu mustăți (box-plot – vezi cursul următor)- grafic liniar
<p>Relatia intre DOUA variabile calitative:</p> <ul style="list-style-type: none">- grafic prin bare/coloane	<p>Relatie intre DOUA variabile cantitative:</p> <ul style="list-style-type: none">- grafic nor de puncte

Grafic sectorial (Plăcintă)

- Eșantion de 62 subiecți, dintre care 19 au greutatea în limite normale, 18 sunt obezi și 25 sunt supraponderali.

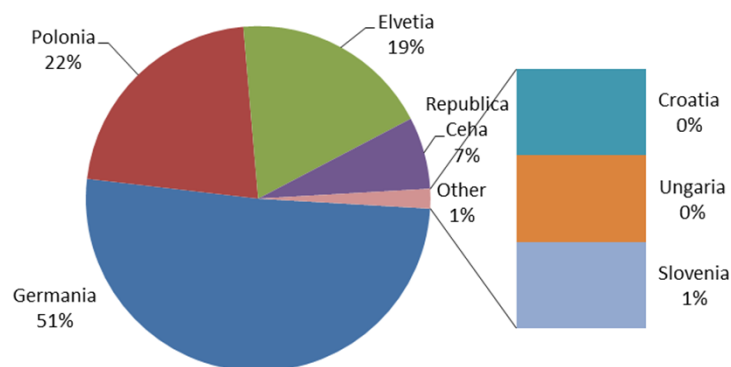
Repartiția pacienților după statusul ponderal



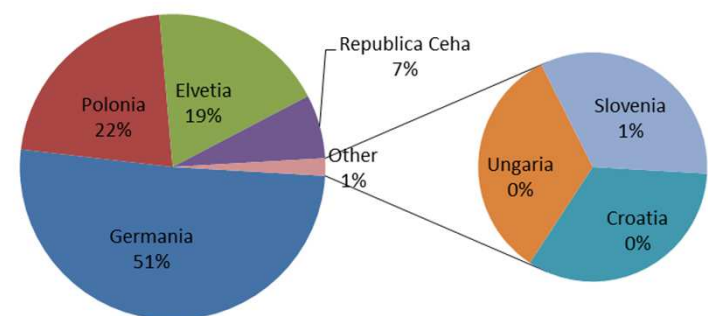
Grafic: Pie of Pie sau Pie of Bar

Țara	Număr de cazuri de pojar raportate în 2012
Germania	166
Polonia	71
Elveția	61
Republica Cehă	22
Croația	2
Ungaria	2
Slovenia	2

Distribuția cazurilor de pojar în Europa Centrală:
2012

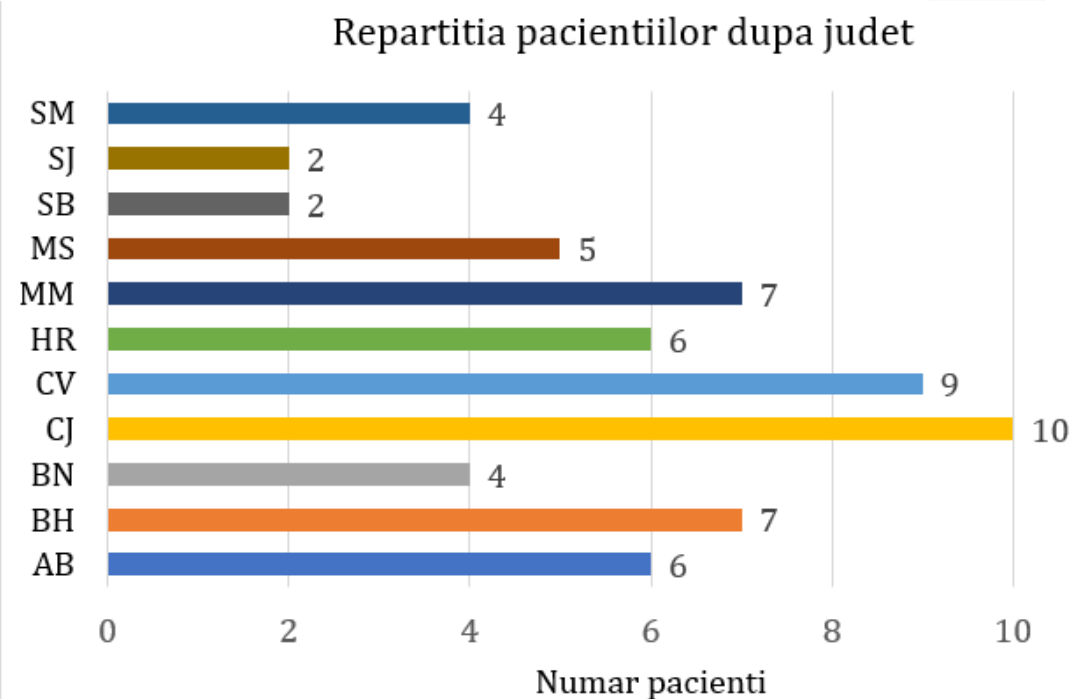


Distribuția cazurilor de pojar în Europa Centrală: 2012



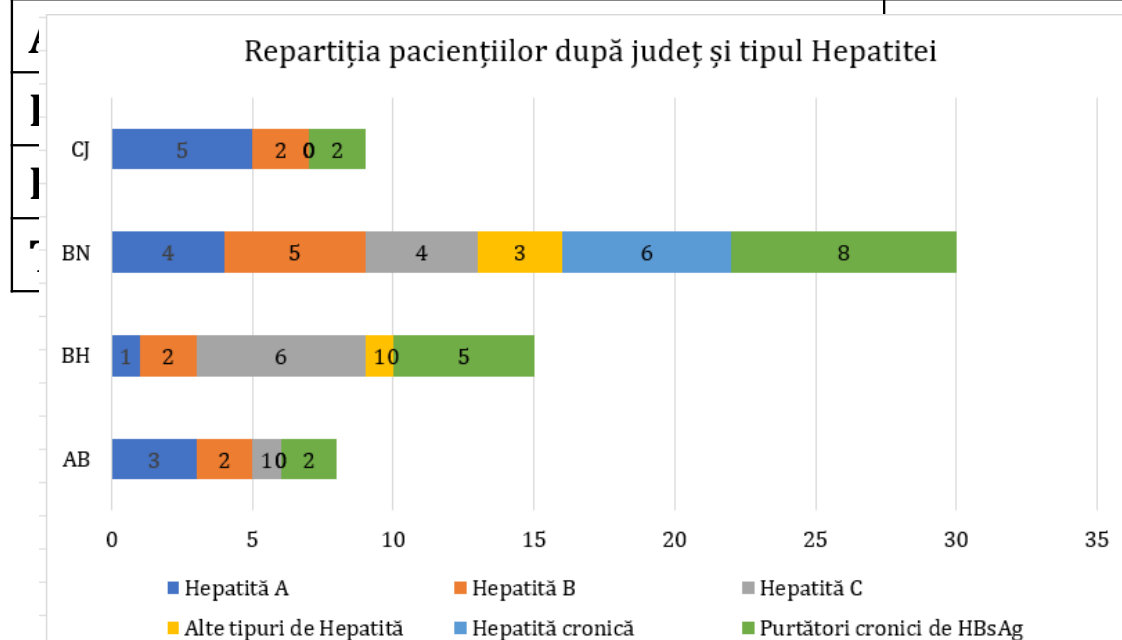
Grafic: Bare

Judet	Frecventa absoluta	Frecventa relativa
AB	6	9.68
BH	7	11.29
BN	4	6.45
CJ	10	16.13
CV	9	14.52
HR	6	9.68
MM	7	11.29
MS	5	8.06
SB	2	3.23
SJ	2	3.23
SM	4	6.45
Total	62	100



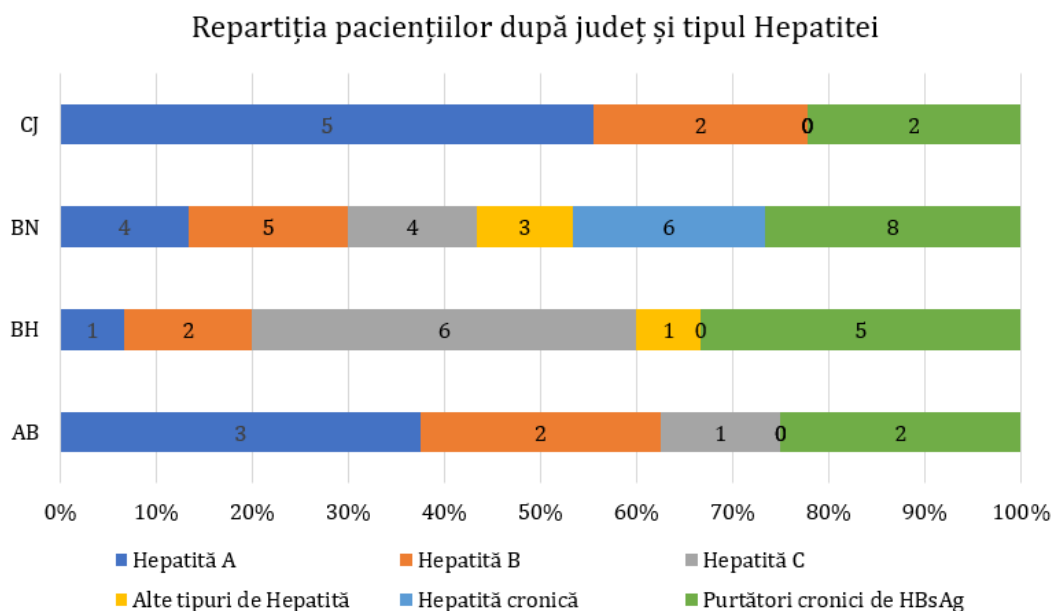
Grafic: Stacked Bar (bare suprapuse)

	AB	BH	BN	CJ	Total
Hepatită A	3	1	4	5	13
Hepatită B	2	2	5	2	11
Hepatită C	1	6	4	0	11
		1	3	0	4
		0	6	0	6
		5	8	2	17
		15	30	9	62



Grafic: 100% Stacked Bar

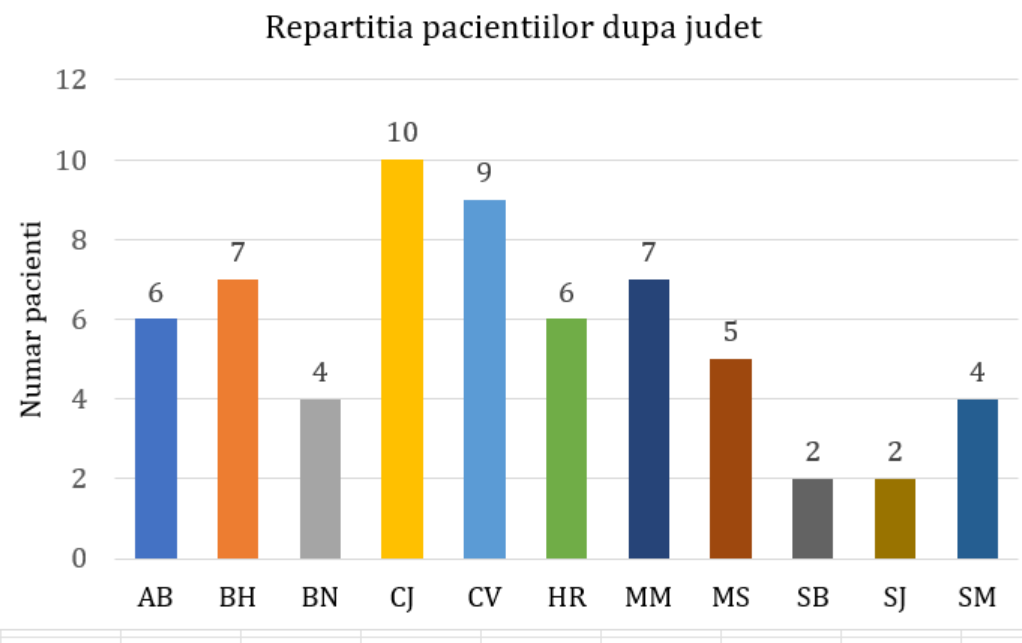
	AB	BH	BN	CJ	Total
Hepatită A	3	1	4	5	13
Hepatită B	2	2	5	2	11
	1	6	4	0	11
	0	1	3	0	4
	0	0	6	0	6
	2	5	8	2	17
	8	15	30	9	62



- Graficul cu bare suprapuse prin procente oferă posibilitatea de a vizualiza atât frecvențele absolute (numărul de cazuri) cât și frecvențele relative (vezi axa Ox)

Grafic: Coloane

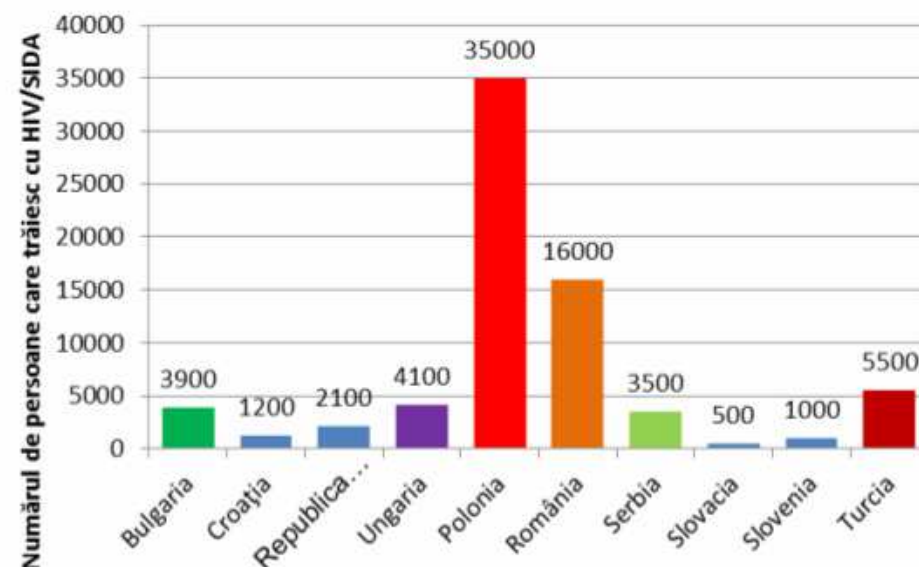
Județ	Frecvența absolută	Frecvența relativă
AB	6	9.68
BH	7	11.29
BN	4	6.45
CJ	10	16.13
CV	9	14.52
HR	6	9.68
MM	7	11.29
MS	5	8.06
SB	2	3.23
SJ	2	3.23
SM	4	6.45
Total	62	100



Grafic: Coloane

Țara	Persoane care trăiesc cu HIV/SIDA
Bulgaria	3900
Croația	1200
Republica Cehă	2100
Ungaria	4100
Polonia	35000
România	16000
Serbia	3500
Slovacia	500
Slovenia	1000
Turcia	5500

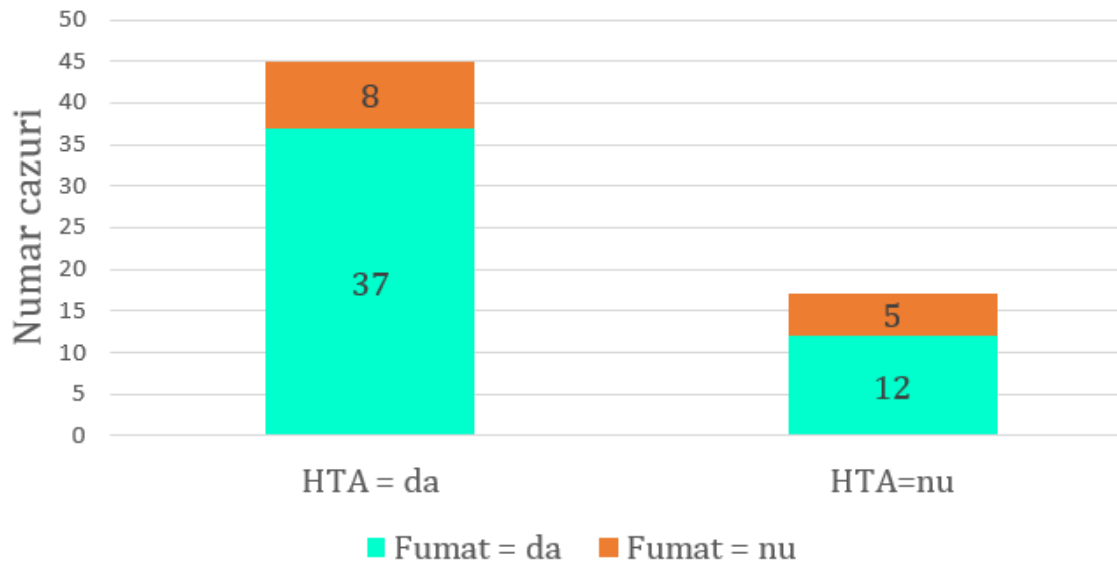
Statistici ale anului 2011



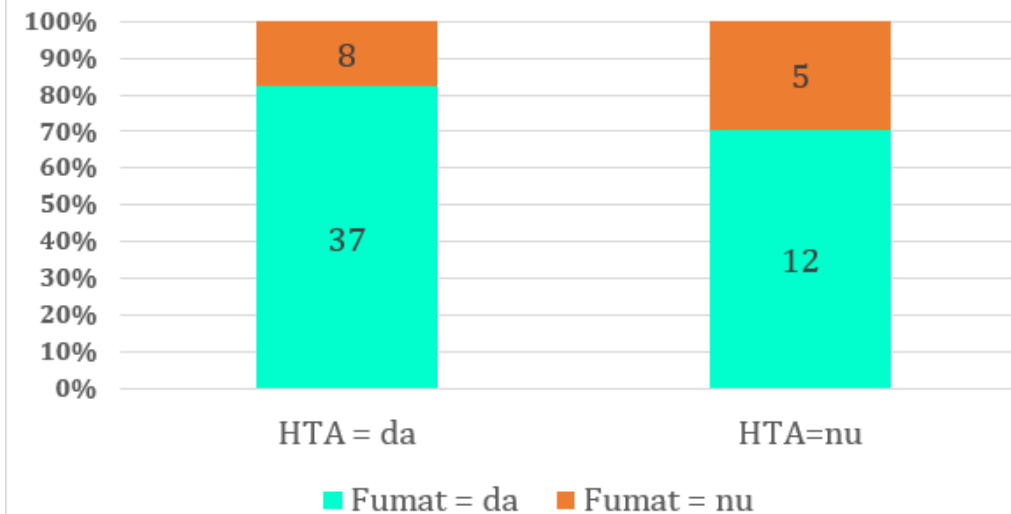
Grafice coloane: Stacked Column versus 100% Stacked Column

	Hipertensiune = da	Hipertensiune = nu	Total
Fumat = da	8	5	13
Fumat = nu	37	12	49
Total	45	17	62

Repartitia pacientilor dupa Fumat si HTA



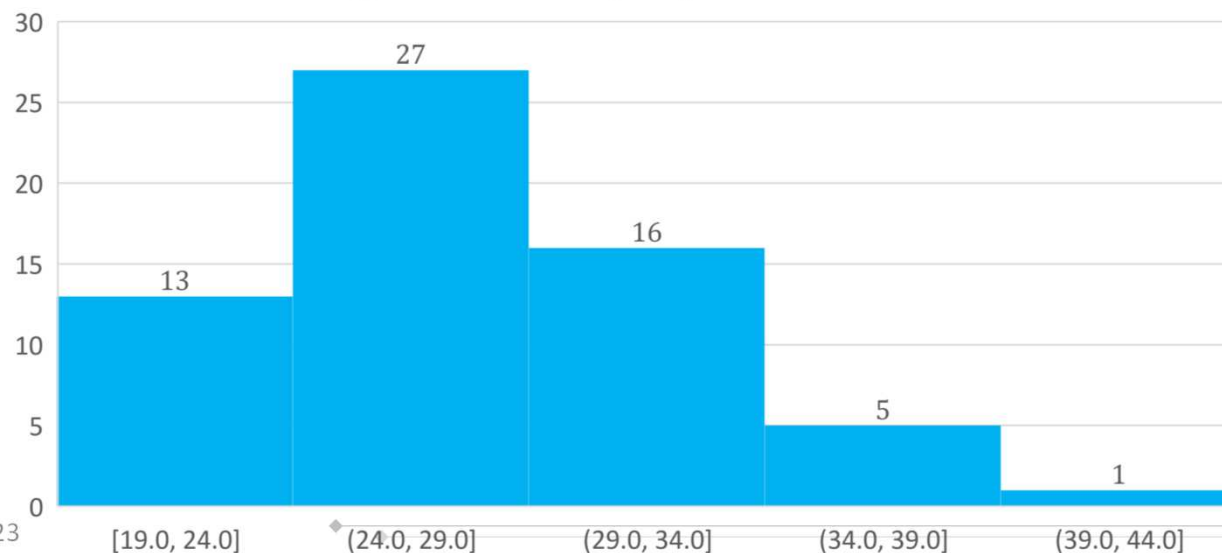
Repartitia pacientilor dupa Fumat si HTA



GRAFIC : HISTOGRAMA

<i>Clase de frecvență</i>	<i>Frecvența absolută</i>
≤ 24	13
(24; 29]	27
(29; 34]	16
(34; 39]	5
(39; 44]	1

Histograma IMC-ului pe eșantionul de studiu

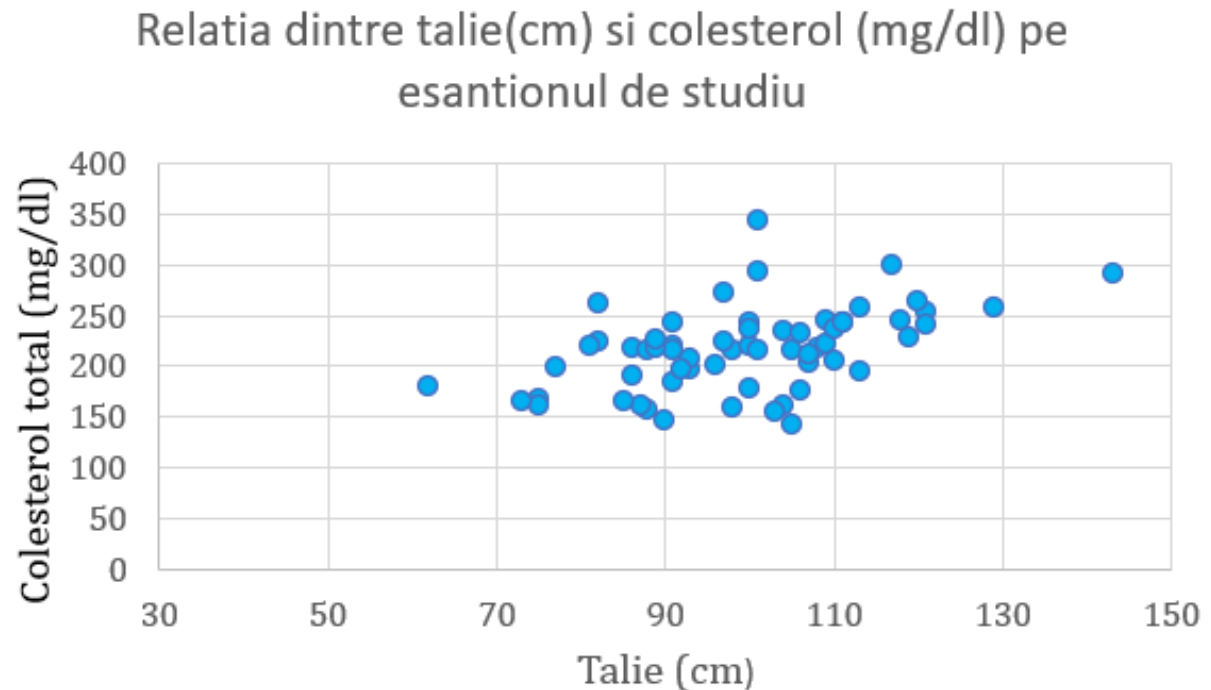


Histograma presupune:

- Impărțirea valorilor unei variabile cantitative în intervale de clasă (sau clase de frecvențe) –vezi axa Ox
- Intervalele de clasă sunt echidistante
- Înălțimea unei coloane a histogramei corespunde cu frecvența absolută sau frecvența relativă a clasei de frecvență
- Cu cât este mai mare înălțimea coloanei, cu atât este mai mare frecvența datelor

Nor de puncte (scatter): Relația între 2 variabile CANTITATIVE

Talie (cm)	Colesterol total (mg/dL)
75	168
101	343
119	229
129	257
104	161
143	292
108	218
98	159
97	272
113	195
91	220
109	246
90	147
88	157
106	175
113	257
109	223
91	184
118	245
100	242
75	162
98	215
101	293
93	197



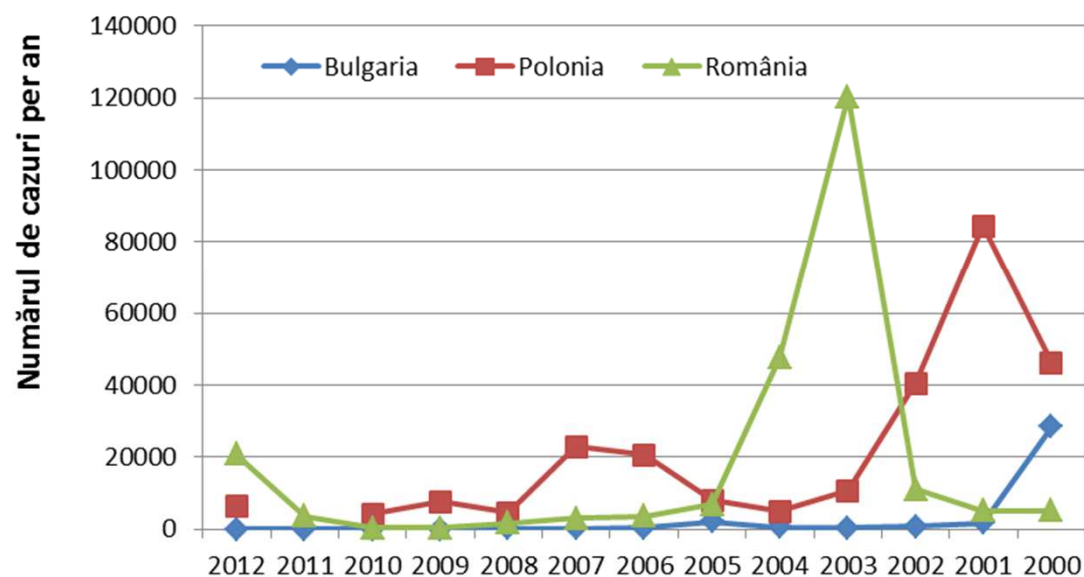
- Se folosește pentru a evidenția posibilă relație între două variabile cantitative
- Punctele de pe grafic reprezintă măsurătorile celor două variabile în raport cu fiecare unitate statistică a eșantionului
- În funcție de modul în care sunt dispersate punctele, există posibilitatea testării unui trend în privința valorilor celor două variabile

Grafic de tip Linie

Țara	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Bulgaria	18	41	39	44	58	88	247	1968	474	252	805	1655	28449
Polonia	6263	0	4196	7586	4598	22890	20668	7946	4857	10588	40518	84419	46181
România	20812	3494	350	343	1746	2958	3563	6801	47444	120377	11079	5076	5125

- Se folosește pentru a reprezenta, de regulă, date calendaristice (serii de timp)

Trendul cazurilor de rubeolă



Grafice: de reținut!

- Reprezentarea grafică trebuie să fie auto-explicativă
- Titlul și detaliile cu privire la reprezentarea grafică aparțin legendei – nu le includeți în figură.
- Figurile trebuie numerotate consecutiv în ordinea citării în text.
- Dacă figura a fost publicată anterior specificați sursa de unde ați luat-o și obțineți de la autori sau de la casa de editură acordul scris de acceptare al utilizării.
- Explicați clar în legendă semnificația simbolurilor utilizate precum și a altor semne (de exemplu săgeți, numere, litere, etc.)
- Evitați reprezentările 3D!

Grafice: de reținut!

- **Plăcintă (PIE):** 1 variabile calitativă
- **Bare:**
 - Variabile calitative
- **Coloane**
 - Variabile calitative
- **Histograma**
 - Variabila cantitativă – Clase de frecvență - Histograma
- **Linie:** evoluție în timp ...
- **Grafic de tip nor de puncte (SCATTER):**
 - 2 variabile cantitative

Tabele: de reținut!

- Informații concise, afișare eficientă.
- Permit prezentarea nivelului de detaliu și de precizie dorit.
- Numerotați tabelele consecutiv în ordinea în care sunt citate în text.
- Fiecare coloană și rând trebuie să aibă o denumire și să prezinte unitățile de măsură.
- Dacă utilizați abrevieri pentru denumirea coloanelor sau a rândurilor prezentați la subsolul tabelului definiția abrevierilor.
- Dacă utilizați date care nu sunt ale dvs., fie că aceste date sunt publicate sau nu, trebuie să obțineți permisiunea de folosire și să specificați sursa acestora.

Prezentarea tabelară: articol științific

Table 1 - Socio-demographic sample characterization

Variable	Medical students	Non-medical students	Total	Correlation/significance (χ^2 test)
n (absolute value)	512	238	750	---
Age (average)	21.69	20.39	21.28	---
Gender (%)				
Male	34.4%	46.2%	38.1%	Pearson Chi-Sq 9.660; $p = 0.002^{**}$
Female	65.6%	53.8%	61.9%	
Marital Status (%)				
Single	97.3%	100.0%	98.1%	Pearson Chi-Sq 6.632; $p = 0.01^*$
Married	2.7%	0.0%	1.9%	
Working students (%)	6.6%	8.8%	7.3%	Pearson Chi-Sq 1.244; $p = 0.265$
Previous/present psychiatric/psychologist follow-up (%)	20.1%	21.4%	20.5%	Pearson Chi-Sq 0.112; $p = 0.738$

* significant value ($p < 0.05$); ** significant value ($p < 0.01$)

Întrebare:
Conform recomandărilor din acest curs, considerați că tabelul prezentat a fost corect realizat???

Sursa: Moreira de Sousa J, Moreira CA, Telles-Correia D. Anxiety, Depression and Academic Performance: A Study Amongst Portuguese Medical Students Versus Non-Medical Students. Acta Med Port. 2018 Sep 28;31(9):454-462.

Link la articol: <https://pubmed.ncbi.nlm.nih.gov/30332368/>

De reținut...

- Este foarte important să faceți diferența dintre variabilă /serie statistică
- Trebuie să identificați corect tipul de variabilă (calitativă / cantitativă)
- Prezentarea datelor (grafică sau tabelară) se va face în funcție de tipul variabilei



Vă mulțumesc!



Contact: Conf.univ.dr. Mihaela Iancu
Email: miancu@umfcluj.ro