

UNIVERSITATEA DE MEDICINA SI FARMACIE “Victor Babeș” TIMISOARA

DISCIPLINA DE INFORMATICA MEDICALA

www.medinfo.umft.ro/dim



BIOSTATISTICA

Curs pentru
Studii Doctorale

Bibliografie

1. Knapp RG, Clinton Miller M – Clinical Epidemiology and Biostatistics, NMS, 1992, Baltimore: Williams & Wilkins
2. Mihalaş GI, Lungeanu D – Biostatistică (seria Studii Doctorale), 2011, Timișoara, Ed. Victor Babeș
3. Colton T – Statistics in Medicine, 1974, Boston: Little, Brown & Co
4. Ingelfinger JA et al – Biostatistics in Clinical Medicine, 1983, New York: Macmillan Publ Co

Cursul 1

Plan CURSUL 1

- *Locul biostatisticii în cercetarea medicală*
- *Inferența statistică*
- *Statistica descriptivă*
- *Estimarea statistică*

1. Locul Biostatisticii în cercetarea medicală

1.1. Tipuri de studii în cercetarea științifică medicală”

- “Productia cunostintelor [explicite]” =
cercetare
- Metode de cercetare
 - Observationale
 - *anchete*
 - *trialuri clinice*
 - Experimentale
 - *studii de laborator – metode specifice*
 - Teoretice – modelare/simulare

1.2. Etapele activității de cercetare

- Stabilirea obiectivelor
- Documentare
- Organizarea cercetării
 - *Studii clinice (*)*
 - *Studii de laborator*
 - *Cercetarea teoretică, formalizarea cunoștințelor*
- Prelucrarea datelor
- Prezentarea rezultatelor

1.3. TIPURI DE STUDII CLINICE

- *[Studiul stării de sănătate a populației]*
- *[Studiul serviciilor de sanatare]*
- *Evaluarea unui procedeu diagnostic*
- *[Asistarea deciziei medicale]*
- *Evaluarea unei abordări terapeutice*
- *Cercetarea unor factori de risc și/sau prognostici*

2. Noțiuni introductive în biostatistică

2.1. Capitolele biostatisticii

- Inferenta statistica
- Statistica descriptiva - Parametrii statistici
- Estimarea statistica
- Teste statistice
- Corelatia si regresia

- Prelucrări avansate
 - Pattern recognition
 - Analiza clasificării
 - Analiza discriminanata
 - Regresia multipla, regresii neliniare

2.2. CONCEPTE GENERALE

- a) *populatie, individ*
- b) *definitie:*
 - Biostatistica = știința estimării caracteristicilor unei populații și/sau compararea populațiilor
- c) *metode:*
 - recensământ - toți indivizii; același moment
 - screening - număr mare; criterii de selecție
 - esantionare - submulțime a populației

- EXTINDEREA PROPRIETATILOR DETERMINATE PE UN ESANTION LA INTREAGA POPULATIE
- *e) ESANTION REPREZENTATIV*
 - CRITERII:
 - ECHIPROPBABILITATE
 - INDEPENDENTA
- *f) METODE DE SELECTIE*
 - SELECTIA SIMPLA
 - NUMERE ALEATOARE ASOCIATE
 - SELECTIA MULTISTRATIFICATA
 - SELECTIA MIXTA
 - CLUSTERE

• 2.3. **VARIABILE**

– a) **DEFINITIE:**

- o caracteristica a populatiei care este studiată și măsurată la toți indivizii din esanțion

– b) **TIPURI DE VARIABLE:**

- **NUMERICE**

- Scală raport (Ratio scale) – zero corespunde valorii 0
- Scala interval – zero arbitrar, intervale egale

- **ORDINALE** (rang, “rank”)

- **NOMINALE** (calitative, “count data”)

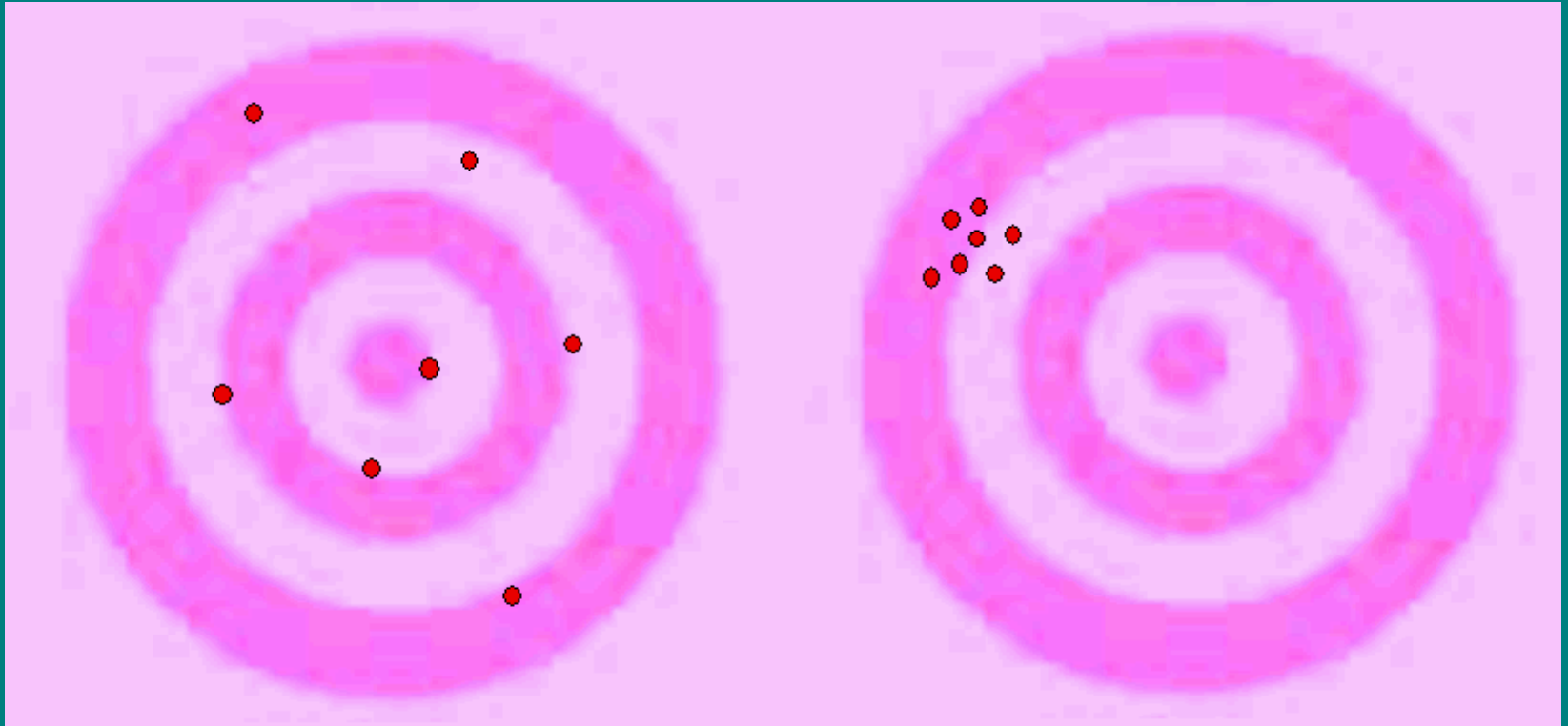
c) Surse de variație

- Variabilitate biologică intrinsecă
- Variabilitatea condițiilor de determinare
- Variabilitate datorată operației de măsurare
 - Erori întâmplătoare
 - Erori sistematice

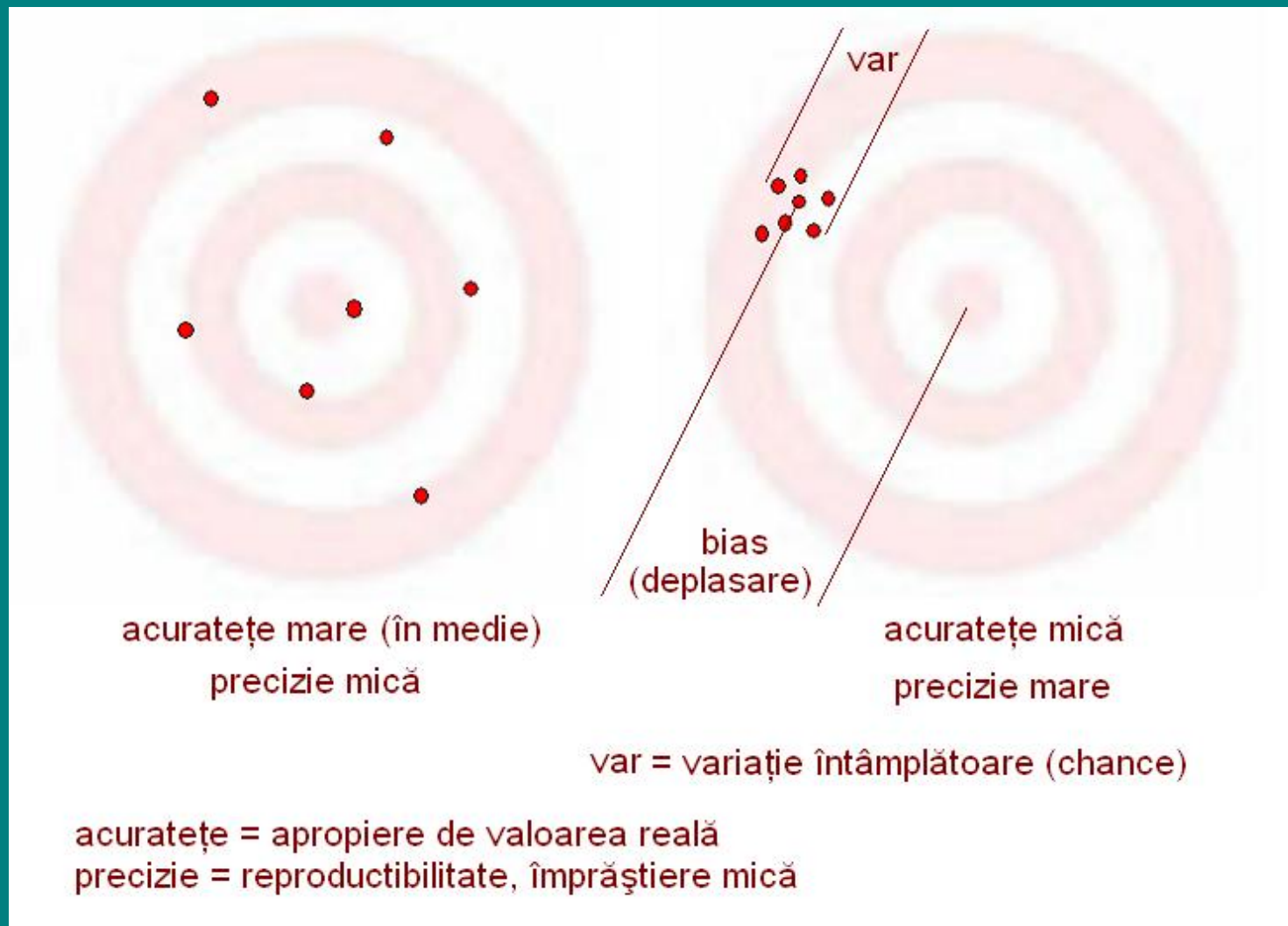
d) Tipuri de variații

- Variații intrapacient
- Variații interpacienți

e) Acuratete – Precizie



e) Acuratete – Precizie



2.4. PROIECTAREA UNUI STUDIU (I)

– STABILIREA SCOPULUI STUDIULUI (ipoteza, obiectiv, plan de studiu)

- selectia variabilelor
- acuratetea masurarilor
- dimensiunea esantionului
- alegerea metodei de prelucrare si a limitelor admise

– ESANTION REPREZENTATIV

- *Definiție*
- *Criterii:*
 - Echiprobabilitate
 - Independență

2.4. PROIECTAREA UNUI STUDIU (II)

– ***COLECTAREA DATELOR***

– ***PREZENTAREA PRELIMINARA***

- **tabele**

- **grafice:**

- *histograme,*

- *sectoare [“pie” (%)]*

- *linii (poligon de frecvențe)*

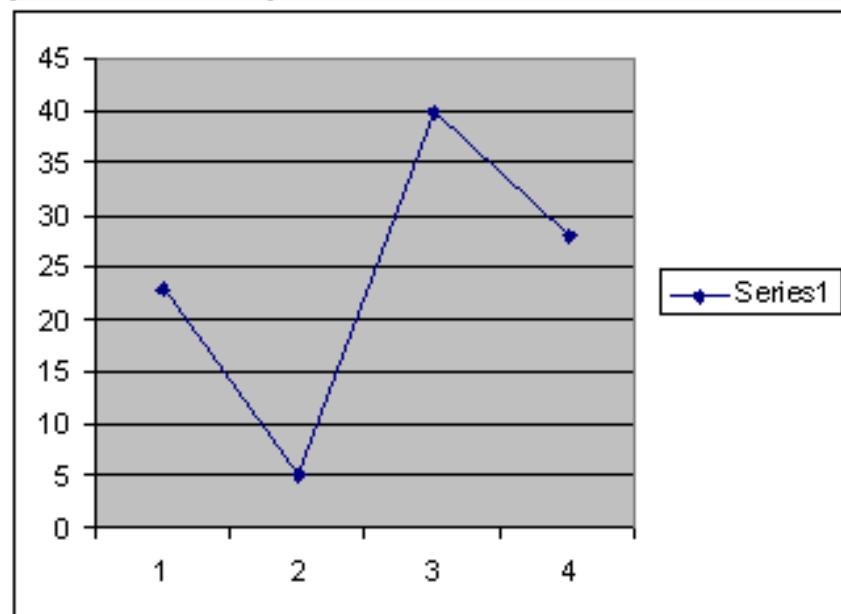
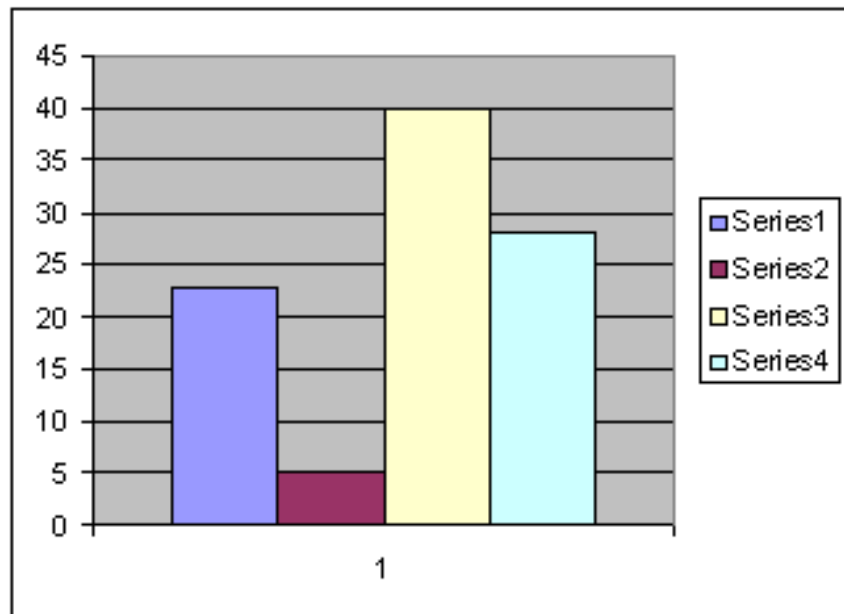
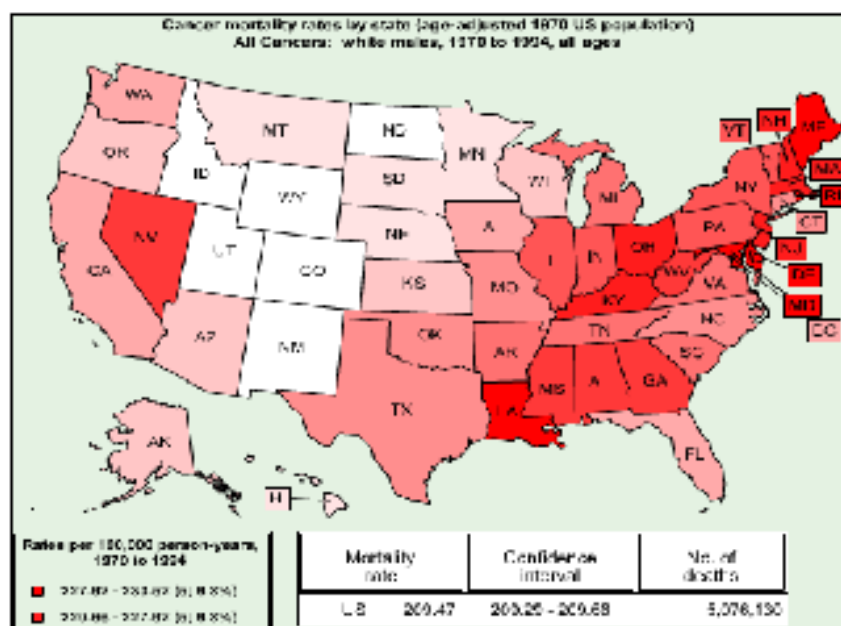
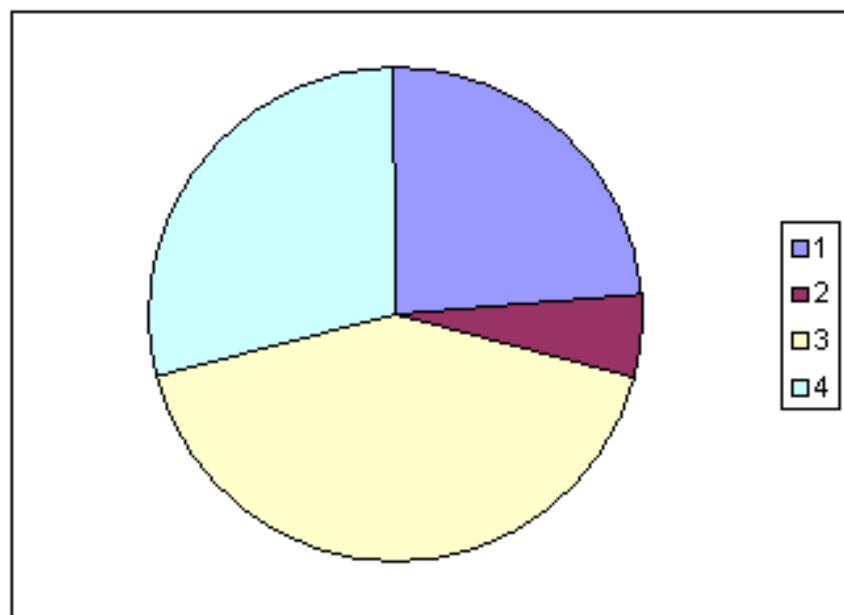
- *puncte (scatter)*

- *harti*

– ***PRELUCRAREA DATELOR - CONCLUZII***

– ***REDACTAREA STUDIULUI***

(+ estimarea erorilor)



3. STATISTICA DESCRIPTIVA: PARAMETRII STATISTICI

- **3.1. EXEMPLU:**

- studiul dezvoltarii copiilor

- populatia: copii de 10 ani, din Timisoara, in 2007

- dimensiunea: 500 copii

- date colectate : inaltimea, in cm

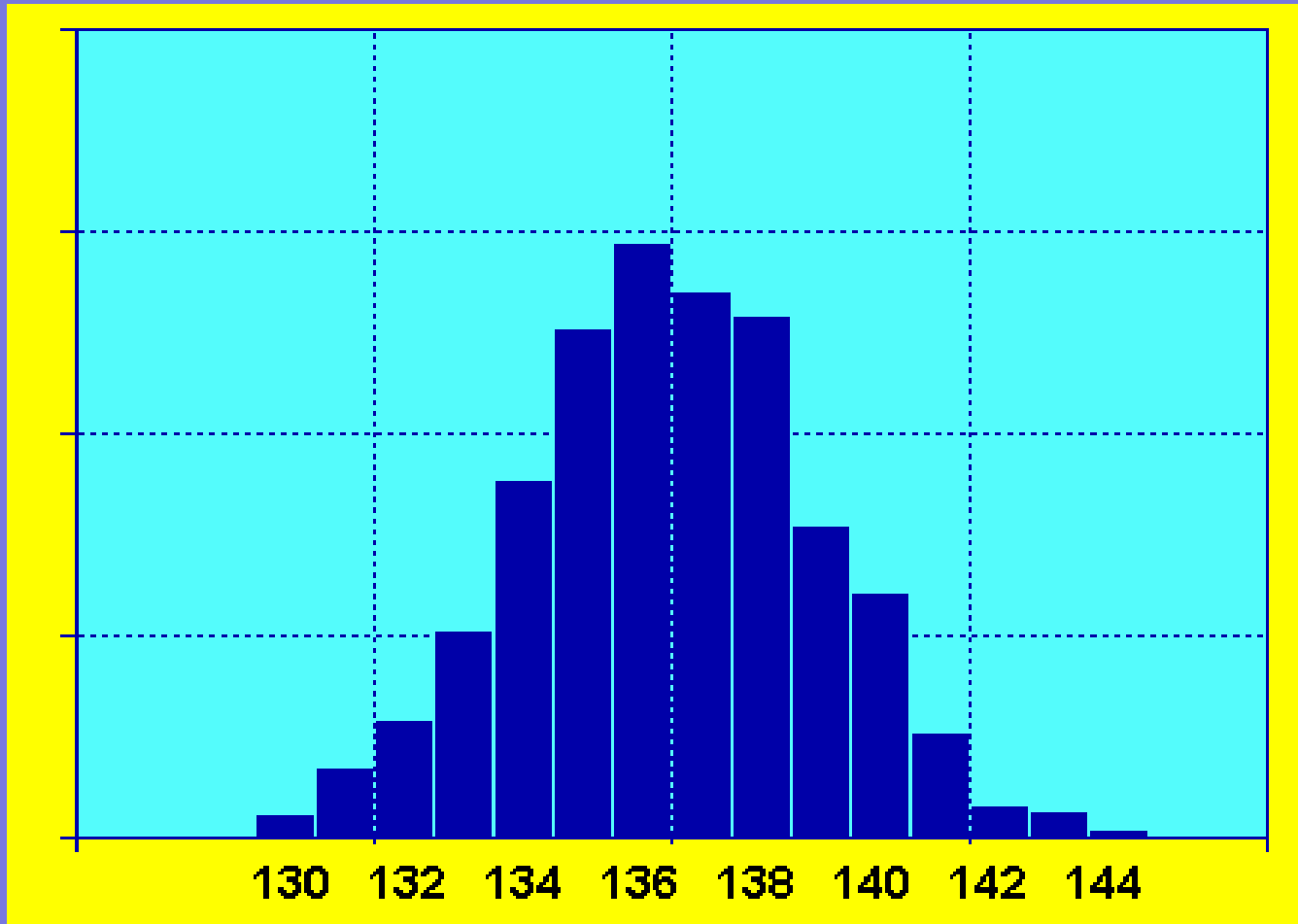
- Precizia de exprimare a măsurării: 1 cm

- Cea mai mică diviziune a instrumentului

- **tabel de frecvențe si histograma**

3.2. TABEL si HISTOGRAMA

INALTIME (CM)	NR. COPII
130	3
131	9
132	17
133	28
...	...



Comentarii:

– Tabelul de frecvențe

- Numărul intervalelor de clase (8 - 20); depinde și de N
- Marginile (endpoints = “bin”)
[cu o zecimală în plus față de exprimarea măsurării]

– Exprimarea frecvențelor

- Absolute (nr)
- Relative
 - Proporție – față de 1
 - Procent – față de 100 %
- Cumulative (uzual %)

Tabel corect

INALTIME (CM)	NR. COPII
130	3
131	9
132	17
133	28
...	...

	Frecv	Fr.Rel.	Fr.Rel.%	Fr.Cum.%
129.5 - 130.4	3	0.006	0.6	0.6
130.5 - 131.4	9	0.018	1.8	2.4
131.5 - 132.4	17	0.034	3.4	5.8
132.5 - 133.4	28	0.056	5.6	11.4
133.5 - 134.4	48	0.096	9.6	21.0
134.5 - 135.4	69	0.138	13.8	34.8
135.5 - 136.4	82	0.164	16.4	51.2
136.5 - 137.4	75	0.150	15.0	66.2
137.5 - 138.4	71	0.142	14.2	80.4
138.5 - 139.4	42	0.084	8.4	88.8
139.5 - 140.4	34	0.068	6.8	95.6
140.5 - 141.4	14	0.028	2.8	98.4
141.5 - 142.4	4	0.008	0.8	99.2
142.5 - 143.4	3	0.006	0.6	99.8
143.5 - 144.4	1	0.002	0.2	100.0
	500	1.000	100.0	

Concluzii:

- valori extreme - rare
- valori centrale - mai des
 - **INDICATORI AI TENDINTEI CENTRALE**
- variabilitatea
 - **INDICATORI DE DISPERSIE**

3.3. INDICATORII TENDINTEI CENTRALE

– a) MEDIA ARITMETICA (MEAN):

$$\bar{X} = \left(\sum_{i=1}^N X_i \right) / N$$

- ***b) MEDIANA (MEDIAN)***

- **VALOAREA CARE IMPARTE
ESANTIONUL IN DOUA PARTI
EGALE**

- **Valorile se ordonează înainte de calcul**
- **Ex: pentru un numar par sau impar de elemente**
- **Recomandata pentru variable ordinale**

c) MODA (MODE)

- VALOAREA CEA MAI FRECVENTA
- CLASE MODALE
- DISTRIBUTII UNI~, BI~, MULTIMODALE
 - distribuțiile uzuale sunt unimodale
 - distribuții multimodale apar la amestec de populații
 - recomandate pentru variabile nominale

3.4. INDICATORI DE DISPERSIE

- A) Pentru variabile numerice***
- B) Pentru variabile ordinale***
- C) Pentru variabile calitative***

A) Variabile numerice

- a) Domeniul de valori:

$$R = (X_{\min} , X_{\max}) , \text{ ex: varsta}$$

- b) Eroarea absoluta:

- Abatererea centrala: $\varepsilon_i = X_i - \bar{X}$

- Eroarea medie absoluta:

$$\varepsilon = (1 / N) \sum | \varepsilon_i |$$

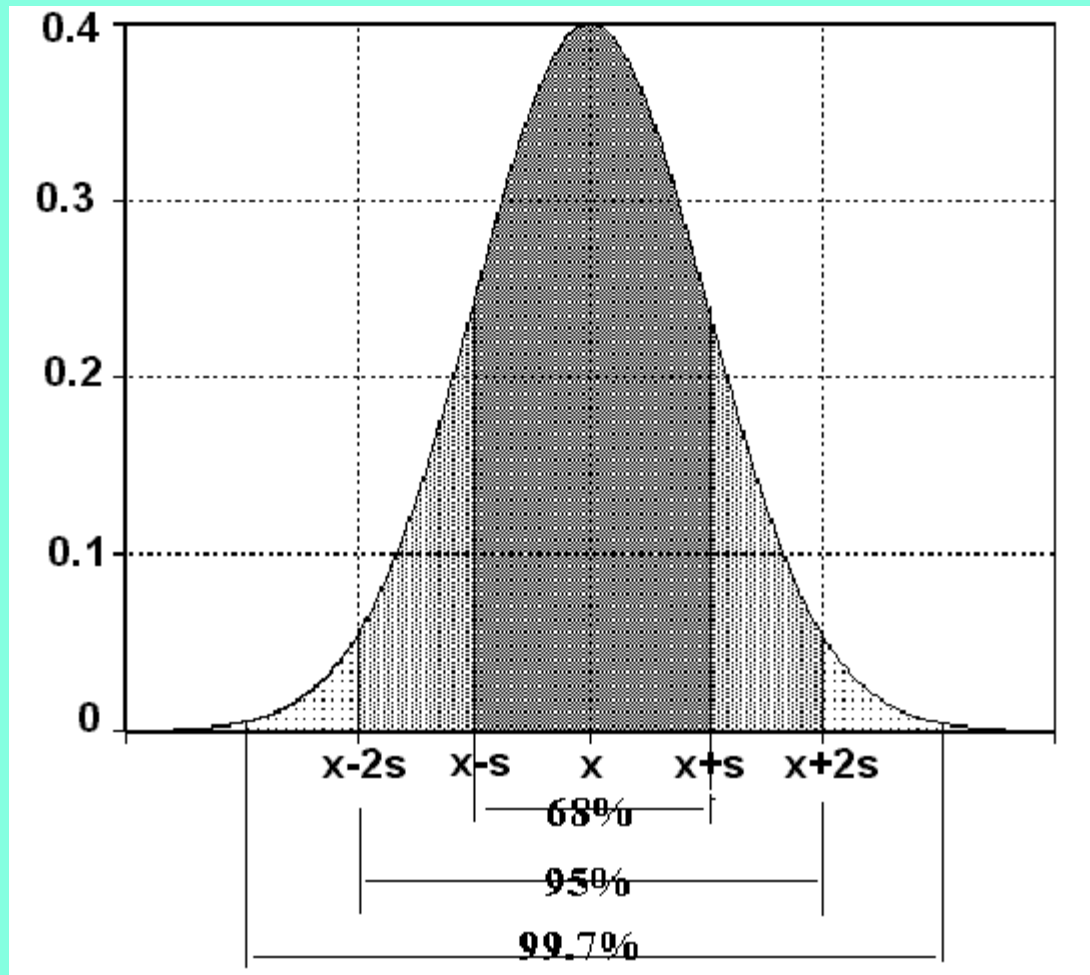
c) DEVIATIȚIA STANDARD

d) COEFICIENTUL DE VARIATIE

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

$$cv = s / \bar{x}$$

3.5. DISTRIBUTIA NORMALA (GAUSS) REPREZENTARE GRAFICA



e) Intervale definite cu “s”

$$x_i \in (\bar{x} - s, \bar{x} + s); p = 68\%$$

$$x_i \in (\bar{x} - 2s, \bar{x} + 2s); p = 95\%$$

$$x_i \in (\bar{x} - 3s, \bar{x} + 3s); p = 99.7\%$$

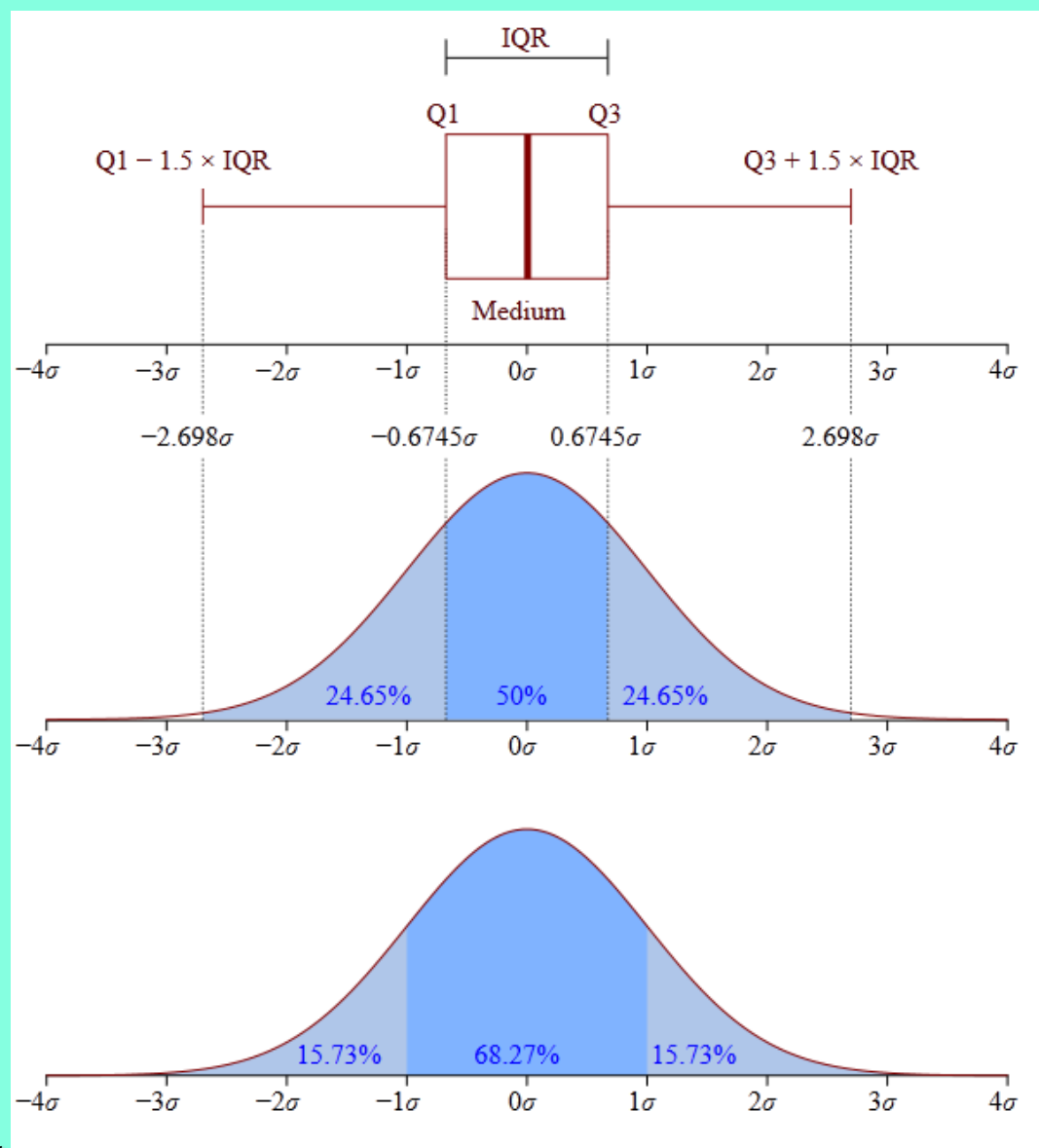
- p = proportia indivizilor care au marimea in intervalul respectiv
- p = probabilitatea ca un individ sa aiba marimea in intervalul respectiv

EX: studiu inaltimea copiilor: n=25 copii
m=137 cm (ac=1)
s=5 cm

B) Variabile ordinale (rang)

- QUANTILE – impartirea in “n” intervale
- CAZURI PARTICULARE
 - MEDIANA $n = 2$
 - QUARTILE $n = 4$
 - DECILE $n = 10$
 - CENTILE $n = 100$
 - PROMILE $n = 1000$
- Calcul: dupa ordonare!

Interval inter-quartile: $IQR = Q_3 - Q_1$



3.6. INDICATORI DE DISPERSIE PENTRU VARIABILE NOMINALE

- **Proporția Clasei:**

$$p_i = N_i / N \text{ (procentul } \dots \times 100)$$

- **Eroarea standard a proporției:**

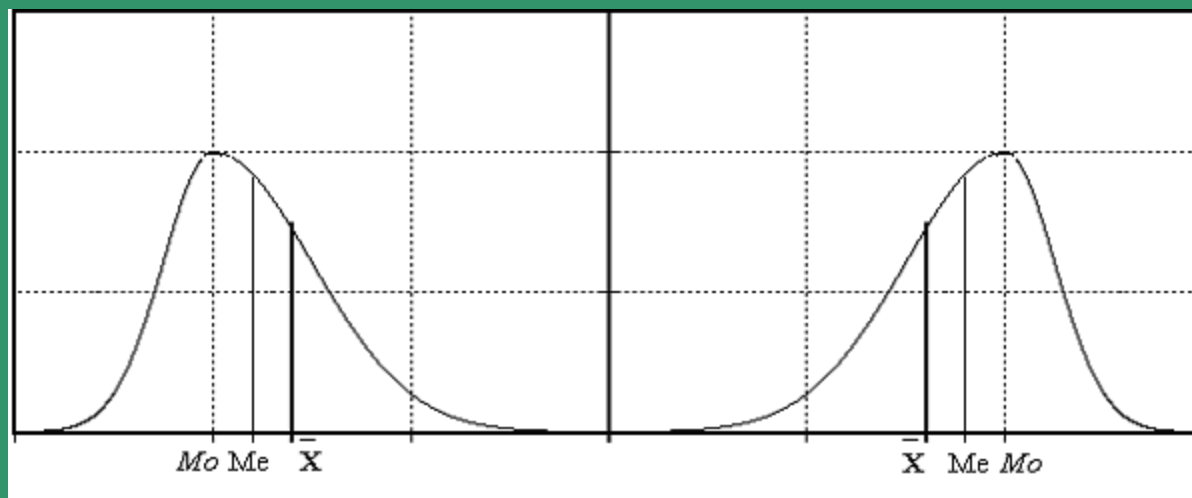
$$s_p = \sqrt{p_i q_i / N}$$

$$q_i = 1 - p_i$$

- p_i – probabilitatea de a aparține clasei
- q_i – probabilitatea de a nu aparține clasei

3.7. ASIMETRIE (SKEWNESS)

- Coeficientul lui Pearson: $\alpha = (X - Mo)/s$
- “coada” (tail) la dreapta sau la stanga



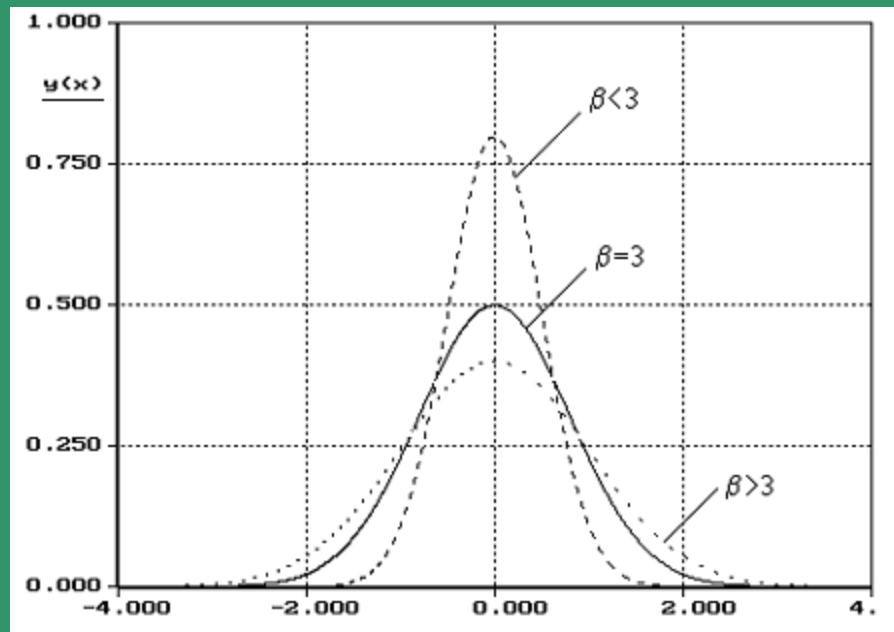
POZITIA RELATIVA

Medie – Mediană - Modă

- **DISTRIBUTII SIMETRICE :**
- **$X = Me = Mo$**
- **DISTRIBUTII ASIMETRICE (skew):**
- **$X =$ cea mai sensibila valoare**
- **$Mo =$ cea mai robusta (putin sensibila)**

3.8. *EXCES* (*KURTOSIS*)

- Exces: (β) – aplatizare, boltire
– (distributie leptokurtica, platikurtica)



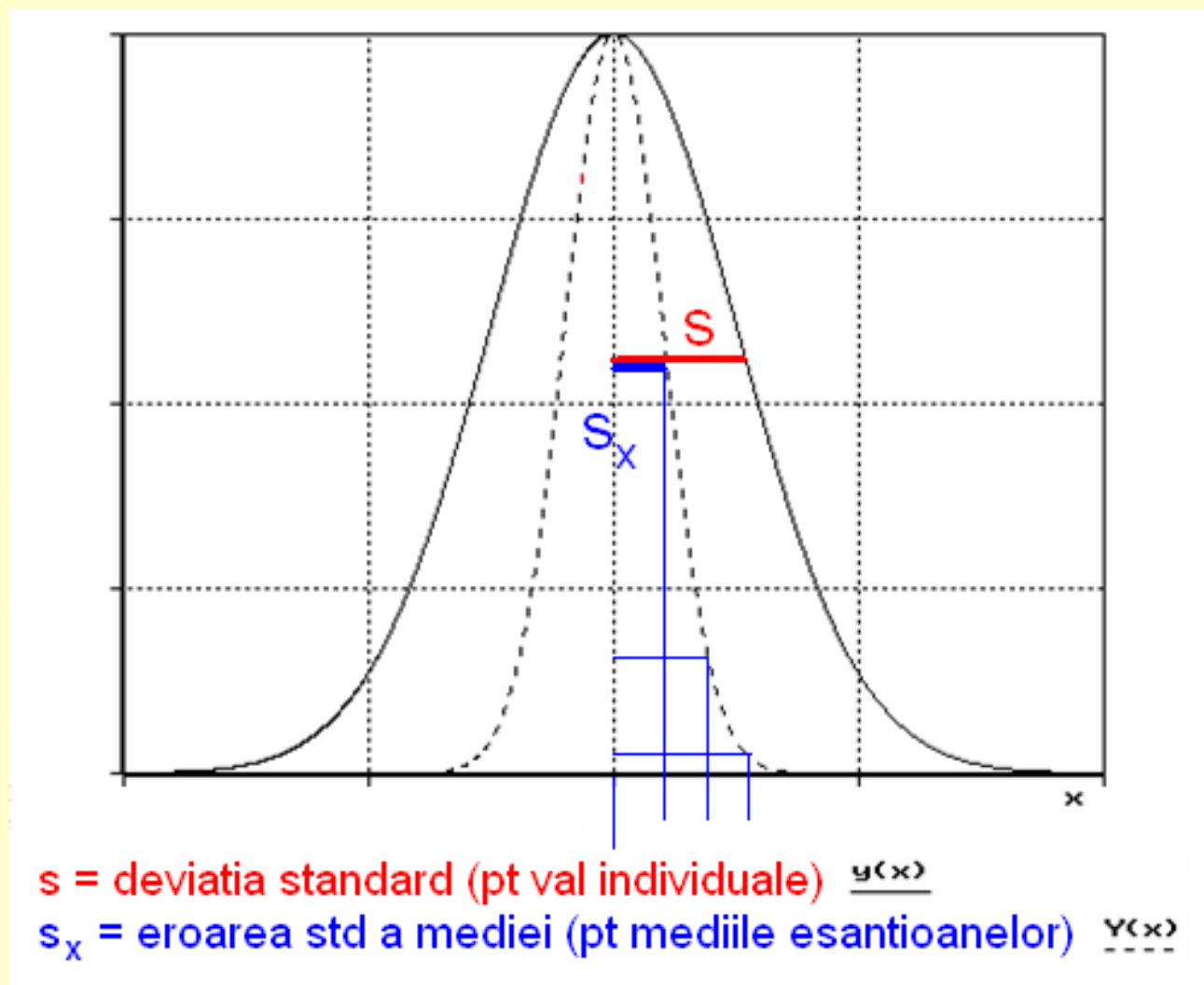
4. ESTIMAREA STATISTICA

4.1. Conceptul de estimator

Ex.

- Pentru POPULATIE
- Ex: “Care este inaltimea medie a copiilor din Timisoara?”
- Caracteristicile populatiei –
 - prin “inferenta statistica”
 - Elemente centrale:
 - NU “ce valoare”, ci “în ce interval”
 - INTOTDEAUNA avem un nivel limitat de incredere (exprimat probabilistic)

4.2. Distributia mediilor esantioanelor Reprezentare grafica



4.3. Eroarea standard a mediei: $s_x = s / \sqrt{n}$

$$\mu \in (\bar{x} - s_x, \bar{x} + s_x); p = 68\%$$

$$\mu \in (\bar{x} - 2s_x, \bar{x} + 2s_x); p = 95\%$$

$$\mu \in (\bar{x} - 3s_x, \bar{x} + 3s_x); p = 99.7\%$$

- p = probabilitatea ca media populatiei sa fie situata in intervalul respectiv
(*se alege conventional $p = 95\%$*)
- Precizia estimarii = $1 / s_x$

- **4.4. DEFINITII**

- a) **DEVIATIA STANDARD** =

- ***INDICATOR DE DISPERSIE CARE ARATA IMPRASTIEREA VALORILOR INDIVIDUALE IN JURUL MEDIEI ESANTIONULUI***

- b) **EROAREA STANDARD A MEDIEI** =

- ***INDICATOR DE DISPERSIE CARE ARATA IMPRASTIEREA MEDIILOR ESANTIOANELOR IN JURUL MEDIEI POPULATIEI***

4.5. Generalizare

TIPURI DE ESTIMARE STATISTICA

- ***PENTRU***
 - **MEDII**
 - **PROPORTII**
 - **DIFERENTE (intre medii, proportii)**

- **A. ESTIMAREA MEDIEI**

- a) **ESANTIOANE MARI $N > 30$**

- **\bar{X} = ARE DISTRIBUTIE NORMALA**

- ***(INDIFERENT DE DISTRIBUTIA INDIVIDUALA)***

<u>$1 - \alpha$</u>	<u>$z_{\alpha/2}$</u>
--------------------------------	----------------------------------

- | | | |
|---------|---|------|
| • 68% | - | 1 |
| • 90% | - | 1.65 |
| • 95% | - | 1.96 |
| • 95.4% | - | 2 |
| • 99% | - | 2.58 |
| • 99.7% | - | 3 |

$$\hat{X} \in \left(\bar{X} \pm z_{\alpha/2} \cdot s_x \right)$$

- **b) ESANTIOANE MICI $N < 30$**
 - **X - ARE DISTRIBUTIE “t”**
 - **GRADE DE LIBERTATE**

- **B. PENTRU PROPORȚII**

$$\hat{P} \in \left(P \pm z_{\alpha/2} \cdot S_p \right)$$

4.6. Dimensiunea esantionului

a) pentru variabile numerice

- se alege “precizia” incadrarii mediei populatiei (eroare admisa)
- se estimeaza eroarea standard “ s_x ” (er.adm. $L = s_x \cdot z_{\alpha/2}$)
 - [Pt $1-\alpha = 95\%$ avem uzual $z_{\alpha/2} \sim 2$, deci putem lua $L = 2 \cdot s_x$]
- deviatia standard trebuie cunoscuta:
 - din studii anterioare
 - studiu peliminar
 - metoda celor 6s intre **Max** si **min**
- se calculeaza “n”:

$$n = (s / s_x) ^ 2 = (2 s / L) ^ 2$$

b) pentru proportii

- acelasi rationament
- se calculeaza “n” pentru $1-\alpha = 95\%$:

$$n = 4 p q / L^2$$

- In general, in loc de 4 va fi z^2 corespunzator nivelului de incredere ales
- cand nu stim p putem lua clase egale; pt 2 clase $p=0,5$ si atunci $n = 1 / L^2$

~ Sfârșit ~