

# Lr 1 Rezumat

## Partea a II-a. INTRODUCERE ÎN STATISTICA DESCRIPTIVĂ

Statistica descriptivă sintetizează datele **grafic** - în *tabele statistice* și *reprezentări grafice* adecvate - și **numeric** - în *indicatori* (*parametri* în sens restrâns) specifici. Prin sinteză se renunță la o parte din informație pentru câștig în relevanță.

### Capitolul 3. STATISTICĂ DESCRIPTIVĂ UNIVARIATĂ

#### 3.1. Sinteza grafică univariată

Se face prin *tabele statistice simple* și reprezentări grafice adecvate tipului de variabilă: *diagrame circulare* pentru variabile calitative, *diagrame prin batoane* pentru variabile tip rang și *histograme* pentru variabile tip măsurătoare. Gruparea în tabele se poate face: *fără* sau *cu pierdere de informație* (în cazul grupării în clase a mai multor variante sau valori). Se obțin *distribuții de frecvențe*, respectiv *distribuții de frecvențe grupate*.

#### 3.1.1. Prezentarea unui șir univariat într-un tabel statistic simplu și distribuții de frecvențe corespunzătoare

1° Distribuții negrupate

Seria  $(x_i)_{i=1,2,\dots,N}$  se poate prezenta ca

- *distribuție de frecvențe absolute*

$$\begin{pmatrix} x_j \\ N_j \end{pmatrix}_{j=1,2,\dots,p}$$

sau, renunțând la o parte din informație, ca

- *distribuție de frecvențe relative*

$$\begin{pmatrix} x_j \\ F_j \end{pmatrix}_{j=1,2,\dots,p}$$

Sn **replicate** valorile, respectiv variantele care rezultă din repetarea măsurării, respectiv observării unei singure unități statistice.

## 2° Reprezentări grafice univariate

**Diagramă circulară** = cerc format din sectoare pentru fiecare variantă, valoare sau clasă\*  $x_j$  astfel încât *unghiul*, respectiv *aria* fiecărui sector să fie proporțional(ă) cu frecvența corespunzătoare.

**Diagramă prin batoane** = reprezentare carteziană plană, în care pe axa orizontală avem marcate valorile, variantele sau clasele variabilei, în fiecare fiind construit pe verticală un segment sau un baton (segment îngroșat cât un dreptunghi de lățime constantă și astfel încât dreptunghiurile să nu se alipească) de *lungime*, respectiv *înălțime* proporțională cu frecvența corespunzătoare.

**Poligon de frecvențe** = linie frântă care unește vârfurile segmentelor din diagrama în batoane, fără a mai reprezenta și segmentele.

**Curbă de frecvențe** (pentru o distribuție continuă, teoretică, a unei variabile continue), de exemplu, **curba erorilor de măsurare întâmplătoare (clopotul lui Gauss)**.

**Valori aberante** = valori care „ies din regula” celorlalte date. Pot fi detectate intuitiv și pot fi eliminate riguros doar dacă știm sau presupunem, pentru datele respective, o anumită formă teoretică de distribuție. Concept de statistică inductivă.

## 3° Distribuții grupate

Grupând variante, ranguri sau valori renunțăm la o parte din informație pentru câștig în relevanță. Grupurile astfel formate sînt *clase*. În cazul valorilor sînt și *intervale de grupare*. Distribuția de frecvențe ale claselor sînt **distribuție grupată**.

**Histogramă** = reprezentare carteziană plană a unei *distribuții grupate de valori*, formată din dreptunghiuri alipite, cu bazele plasate pe intervalele de grupare și cu *ariile* proporționale cu frecvențele claselor. Există histograma cu intervale de grupare egale, respectiv, inegale.

O variabilă tip măsurătoare (continuă) nu poate fi măsurată exact, ci doar aproximativ. Rigoarea științifică cere exprimarea, alături de valoarea aproximativă, a gradului de aproximare (erori de măsurare). De exemplu: 185 cm  $\pm$  0,5cm - înălțimea unui subiect uman.

+ 4° Aplicație specific ecologică a distribuțiilor negrupate - distribuția de abundențe

Forma sa de funcție strict descrescătoare cu o “coadă” lungă *provine din convenția de reprezentare* și din faptul că, de regulă, **într-o biocenoză numărul speciilor rare este mult mai mare decât cel al speciilor dominante** prin abundențe. Distribuția *uniformă* sînt în ecologie **distribuție echitabilă** sau **regulată**.

\* Am menționat și termenul "clasă" pentru ca definițiile să fie valabile și în cazul distribuțiilor grupate.

5° Observație



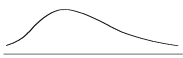
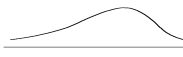
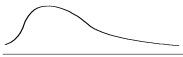
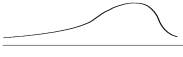
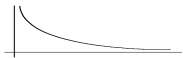
O constantă fizică sau biologică dintr-o măsurătoare (continuă) se poate descrie statistic prin două numere, unul descriind aproximativ valoarea reală, iar celălalt exprimând exact gradul de aproximare.

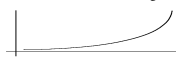

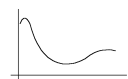

**3.1.2. Pentru ce grupăm măsurători sau „limbajul repartițiilor”**

1° Concluzii

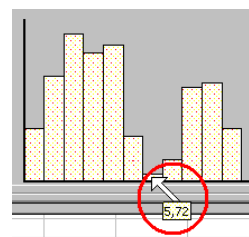
Grupăm ca să sesizăm una din următoarele:

2° Forme tip de distribuții

<i>unimodală</i> (o modă)  <i>- Exprimă omogenitate</i>	<i>simetrică</i>	<i>concentrată într-un punct (1)</i> 	- Exprimă <i>omogenitate absolută</i>
		<i>neconcentrată într-un punct (2)</i> 	- Exprimă cel mai bine o <i>tendință centrală</i> .
	<i>[asimetrică]</i>	<i>slab asimetrică</i>	<i>de stânga (3)</i> 
			<i>de dreapta (4)</i> 
		<i>puternic asimetrică</i>	<i>de stânga (5)</i> 
			<i>de dreapta (6)</i> 
		<i>extrem asimetrică</i>	<i>de stânga (7) (în formă de i)</i> 

			<b>de dreapta (8)</b> (în formă de j) 	
<b>Bimodală</b> (două mode) - <i>Exprimă eterogenitate</i>	<b>simetrică (9)</b> (de exemplu în formă de U) 	= Exprimă eterogenitate absolută		
	<b>asimetrică (10)</b>			
<b>Multimodală</b> (multimodală) - <i>Exprimă eterogenitate</i>	<b>multimodală propriu-zisă (11)</b> (n > 2. mode) 			
	<b>uniformă (12)</b> (numai mode) 			

3° Distribuțiile bimodale trebuie « decupate » în două distribuții unimodale prin *limită empirică de discriminare (identificare)*. Programul de la link-ul citat mai jos îndrumă și ajută utilizatorul să determine limite de discriminare (vezi figura alăturată).



+ 5° Principalele transformări de simetrizare (cvasigaussianizare, cvasinormalizare)

Distribuție:	Transformarea indicată:
Puternic asimetrică de stânga	$\sqrt[n]{x}$
Puternic asimetrică de dreapta	$x^n$
Extrem asimetrică de stânga	$\log_a x$ cu $a > 1$
Extrem asimetrică de stânga conținând și valoarea 0	$\log_a (x+b)$ cu $a > 1$ și $b > 0$

### 3. 1. 3. Cum grupăm măsurători

Utilizăm după specificul problemei și al datelor intervale de grupare egale sau inegale.

1° Grupare cu intervale de clasă egale

*Numărul de clase* se poate calcula după reguli empirice, de exemplu **formula lui Sturges**:  $nc \approx 1 + 10 / 3 \times \lg N$ , unde  $N$  = volumul seriei, de unde rezultă *lungimea intervalului de clasă*,  $ic = (x_{max} - x_{min}) / nc$ . La adresa <http://app.inthlerom.ro/histo/HistoSetup.zip> obțineți gratuit un program «ad-in EXCEL» care calculează histograme și vă îndrumă decizia.

#### Reguli de rotunjire a datelor în calcule

Un mod de grupare este și rotunjirea datelor. Prezentăm în continuare câteva *observații* asupra rotunjirii datelor în calculul manual [6].

1) Reguli de rotunjire în calculul manual:

- Cifrele 0,1,2,3,4 se șterg. De exemplu: 2,64 devine 2,6.
- Cifrele 6,7,8,9 se rotunjesc prin adaos de o unitate la zecimala superioară. De exemplu 2,68 devine 2,7.
- Cifra 5 se rotunjește prin adaos, respectiv lipsă, la valoarea pară cea mai apropiată. De exemplu 2,65 devine 2,6 iar 2,55 va deveni de asemenea 2,6. (Observăm că, în calculator, cifra 5 este tratată întotdeauna prin adaos. Și noi vom proceda în continuare în acest mod.)

2) Rotunjirea prin afectarea mai multor zecimale.

Numărul 1,959964 rotunjit:

- |                         |                        |
|-------------------------|------------------------|
| la 5 zecimale = 1,95996 | la 4 zecimale = 1,9600 |
| la 3 zecimale = 1,960   | la 2 zecimale = 1,96   |
| la 1 zecimală = 2       |                        |

3) Numărul de zecimale păstrate în calculul manual:

- Se recomandă păstrarea a 2, 3, maximum 4 zecimale,
- În calculele intermediare putem păstra o zecimală în plus care dispare în rezultatul final.
- **Exactitatea rezultatelor depinde doar de exactitatea măsurătorilor și nu de mărirea numărului de zecimale în cadrul calculelor.**

## **4 1 Teste, exerciții și probleme**

### **TG1. Durata 185'' pe calculator.**

Renunțarea la o parte din informație pentru câștig în relevanță este obiectivul:

1. statisticii moderne
2. statisticii descriptive
3. statisticii inductive
4. statisticii clasice

Construim histogramme pentru variabile de tip:

1. rang
2. pur calitative
3. măsurătoare
4. binar

Alegeți afirmația corectă:

1. O distribuție unimodală exprimă omogenitatea populației în raport cu variabila respectivă
2. O distribuție multimodală exprimă omogenitatea populației în raport cu variabila respectivă
3. O distribuție unimodală exprimă eterogenitatea populației în raport cu variabila respectivă

Ariile sunt proporționale cu frecvențele:

1. numai la histograma
2. numai la histograma și diagrama în batoane
3. numai la diagrama circulară
4. la histograma, diagrama în batoane și diagrama circulară

O distribuție de abundențe se figurează în ecologie ca o repartiție:

1. unimodală asimetrică la stânga
2. unimodală extrem asimetrică la stânga
3. unimodală simetrică
4. multimodală

Forma unei distribuții de abundențe reprezentate în ordine descrescătoare se datorează:

1. convenției de reprezentare
2. faptului că într-o biocenoză, de regulă, se găsesc câteva specii dominante ca efectiv și multe specii rare
3. convenției de reprezentare și faptului că, de regulă, într-o biocenoză se găsesc câteva specii dominante ca efectiv și multe specii rare

O distribuție puternic asimetrică de stânga poate fi simetrizată prin transformarea:

1.  $x^n$ \*
2.  $\log a(x)$  cu  $a > 1$
3.  $\log a(x+b)$  cu  $a > 1$  și  $b > 0$
4. decupare în două distribuții unimodale
5.  $x^{1/n}$

O distribuție extrem asimetrică de stânga care nu conține valoarea 0 poate fi simetrizată, cel mai simplu, prin transformarea:

1.  $\log a(x)$  cu  $a > 1$
2. identica ( $x$ )
3.  $x^n$
4.  $x^{1/n}$
5.  $\log a(x+b)$  cu  $a > 1$  și  $b > 0$
6. decupare în două distribuții unimodale

## TC1. Durata 5'.

1. \_\_\_\_\_ este o reprezentare carteziană plană a unei distribuții grupate, formată din dreptunghiuri alipite, cu \_\_\_\_\_ plasate pe intervalele de grupare și cu ariile proporționale cu \_\_\_\_\_ claselor.
2. O distribuție de \_\_\_\_\_ are o formă de grafic de funcție \_\_\_\_\_, obținută prin îmbogățirea artificială a variabilei calitative taxon cu o relație de ordine.
3. Dacă numărul total de indivizi s-ar distribui „echitabil” între toate speciile dintr-o biocenoză, s-ar obține o distribuție \_\_\_\_\_ sau \_\_\_\_\_ sau \_\_\_\_\_.
4. Eterogenitățile care apar ca un amestec de două sau mai multe omogenități, adică sub formă de distribuții bi sau multimodale, vor fi \_\_\_\_\_ în două, respectiv  $n$  distribuții \_\_\_\_\_.

---

\*  $x^n$  înseamnă  $x^n$ .

8 **Dragomirescu L.**, 2003, *Lucrări practice de biostatistică*. Ediția a III-a revăzută și adăugită, Editura "Agronomia", București, 264p. ISBN 973-86668-0-5.

5. Distribuțiile unimodale asimetrice vor fi cvasigaussianizate prin \_\_\_\_\_ de cvasigaussianizare adecvate, deoarece ideea de tendință centrală este cel mai bine exprimată de distribuțiile \_\_\_\_\_ și \_\_\_\_\_.

6. Transformarea indicată pentru o distribuție unimodală puternic asimetrică de stânga este \_\_\_\_\_, iar pentru cea puternic asimetrică de dreapta este \_\_\_\_\_.

## Exerciții sau probleme rezolvate

### 1.

S-au determinat grupele sangvine ale unui lot de donatori și a rezultat seria statistică:

A A B O O B AB B AB O.

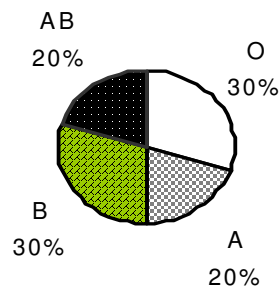
(a) Din ce tip de variabilă este alcătuită seria statistică ? (b) Realizați tabelul statistic simplu corespunzător seriei și (c) reprezentați distribuția de frecvențe prin tipul de reprezentare grafică adecvat datelor.

*Rezolvare:*

(a) Variabilă calitativă.

(b) Tabelul statistic simplu și (c) diagramă circulară.

variante $x_j$	frecvențe absolute $N_j$	frecvențe relative $F_j = N_j / N$
O	3	3 / 10
A	2	2 / 10
B	3	3 / 10
AB	2	2 / 10
<i>Totaluri:</i>	10	1



### 2.

S-au măsurat temperaturile unor copii sănătoși, cu o precizie de 2 zecimale și a rezultat seria statistică:

36,52 36,57 36,60 36,65 36,68 36,69 36,71  
36,76 36,79 36,83 36,84 36,87 36,92 36,98.

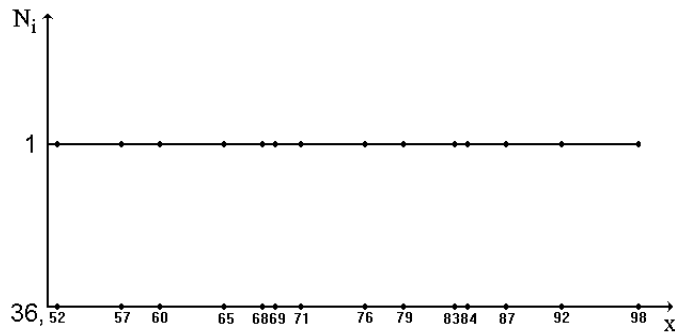
(a) Să se grupeze într-o distribuție de frecvențe și să se reprezinte printr-un poligon de frecvențe. (b) Ce formă tip de distribuție a rezultat ? (c) Alcătuiți o nouă serie statistică, prin micșorarea preciziei de măsurare a temperaturii



(rotunjirea datelor la o zecimală). (d) Alcătuiți tabelul statistic simplu al acestei serii și (e) reprezentați seria printr-un poligon de frecvențe. (f) Ce formă tip de distribuție a rezultat ? (g) Care dintre cele două distribuții reflectă mai corect fenomenul și de ce ?

**Rezolvare:**

(a) Deoarece seria este formată numai din valori distincte ordonate ascendent, distribuția de frecvențe este formată din valorile seriei fiecare având frecvența 1. Poligonul de frecvențe pentru distribuția dată este de forma alăturată.



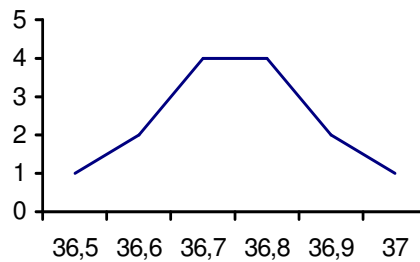
(b) Distribuție uniformă.

(c) Seria statistică cu datele rotunjite la o zecimală:

36,5 36,6 36,6 36,7 36,7 36,7 36,7  
36,8 36,8 36,8 36,8 36,9 36,9 37,0.

(d) Distribuția de frecvențe și (e) poligonul de frecvențe pentru datele rotunjite:

valori $x_j$	frecvențe absolute $N_j$	frecvențe relative $N_j/N$
36,5	1	1/14
36,6	2	2/14
36,7	4	4/14
36,8	4	4/14
36,9	2	2/14
37,0	1	1/14
<i>Totaluri:</i>	14	1



(f) Distribuție unimodală și simetrică (cvasigaussiană). (g) Distribuția cvasigaussiană deoarece un lot omogen are o distribuție gaussiană, nu uniformă.

### 3.

S-a studiat radiosensibilitatea bulbilor de lalele la iradiere cu radiații Röntgen, pe un eșantion de 59 indivizi. Radiosensibilitatea a fost exprimată prin doza critică ce caracterizează indivizii și a rezultat seria statistică:

560 560 573 578 580 581 583 584 586 586 587  
 588 589 593 595 598 601 603 604 606 606 607  
 608 608 608 609 610 615 616 616 618 619 622  
 624 625 629 630 633 636 637 637 639 640 641  
 644 645 645 649 653 658 658 662 667 669 675  
 677 680 683 699.

(a) Realizați o grupare cu intervale de clasă egale și (b) reprezentați-o prin tipul de reprezentare grafică adecvat. (c) Ce formă tip de distribuție s-a obținut ?

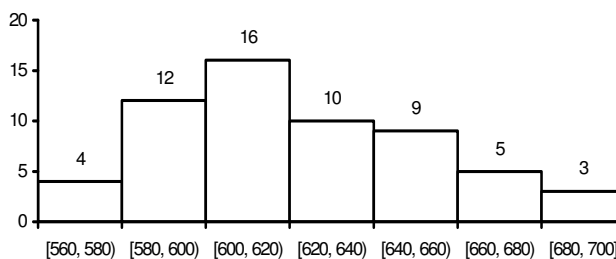
*Rezolvare:*

(a)  $N = 59 (> 50)$ ;  $nc = 1 + (10 \cdot \lg 59) / 3 = 1 + 10 \cdot 1,77 / 3 = 6,9 \cong 7$ ;

$ic = (x_{max} - x_{min}) / nc = (699 - 560) / 7 = 19,86 \cong 20$ .

Tabela statistică simplă cu intervale de grupare egale și (b) histograma:

intervale de clasă	$N_j$
[560, 580)	4
[580, 600)	12
[600, 620)	16
[620, 640)	10
[640, 660)	9
[660, 680)	5
[680, 700]	3
Total:	59



(c) Distribuție unimodală asimetrică de stânga.

### 4.

Se dă șirul ordonat de concentrații de amoniu în  $\mu\text{g} / \text{l}$  din Delta Dunării (date reale ale Colectivului de Ecologie al Universității din București):

37 43 57 62 99 118 120 123 124 129 132  
 135 137 137 142 156 157 174 176 176 181 181  
 183 188 199 219 226 →228 242 256 256 272 284

304 304 318 336 341 361 364 →427 475 510 590  
 599 608 →629 651 665 723 →1013

(Săgețile marchează anticipat limitele claselor pe care le vom calcula la rezolvare.)

(a) Să se grupeze folosind intervale de grupare egale și, formula lui Sturges orientativ (fără a o calcula exact). (b) Să se deseneze histograma distribuției grupate. (c) Să se caracterizeze statistic distribuția. (d) Ce transformare trebuie aplicată pentru cvasigaussianizarea distribuției ?

*Rezolvare:*

(a) Deoarece  $\lg 51$  este cuprins între  $\lg 10 (=1)$  și  $\lg 100 (=2)$ ,  $nc$  va fi cuprins între 4 și 8:

$$nc = 1 + 10 / 3 \cdot \lg 10 \approx 1 + 10 / 3 \cdot 1 = 1 + 3,3 = 4,3 \approx 4$$

$$nc = 1 + 10 / 3 \cdot \lg 100 \approx 1 + 10 / 3 \cdot 2 = 1 + 6,6 = 7,6 \approx 8.$$

Volumul  $N$  fiind mic ( $= 51$ ) vom prefera un număr mic de clase, 4 sau 5. În aceste cazuri intervalul de clasă  $ic$  va avea valorile:

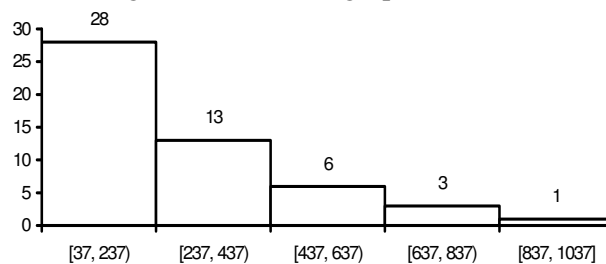
$$ic = (x_{max} - x_{min}) / nc = (1013 - 37) / 4 = 976 / 4 \approx 244 \approx 250$$

$$ic = (x_{max} - x_{min}) / nc = (1013 - 37) / 5 = 976 / 5 \approx 195 \approx 200.$$

Pentru ușurința calculului vom prefera  $ic = 200$ . Astfel, suntem mai aproape și de formula lui Sturges  $nc = 5$ , nu  $nc = 4$  care ar corespunde lui  $N = 10$ . Grupând astfel datele, obținem distribuția grupată următoare, reprezentată în histograma alăturată distribuției.

Tabelul statistic simplu și (b) histograma distribuției grupate:

intervale de clasă	$N_j$
[ 37, 237)	28
[237, 437)	13
[437, 637)	6
[637, 837)	3
[837, 1037]	1
Total:	51



(c) Distribuție unimodală extrem asimetrică de stânga. (d) Logaritmare cu o bază supraunitară.

## 5.

Fie distribuția grupată (cu clase de lungimi inegale) din primele două coloane ale tabelului următor:

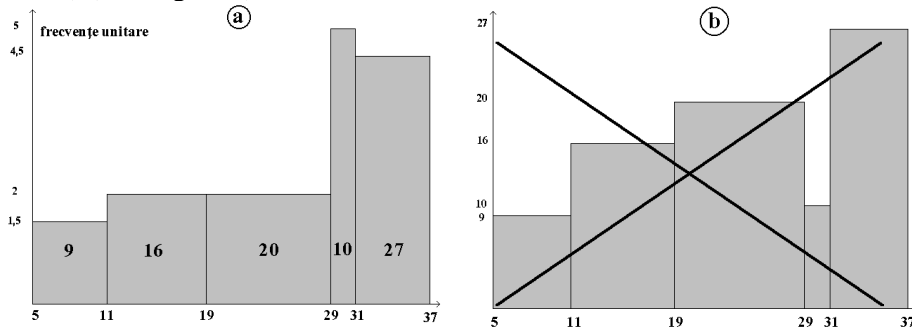
<i>intervale de clasă</i>	$N_j$	<i>lungime interval (<math>l_j</math>)</i>	<i>frecvențe unitare (<math>N_j / l_j</math>)</i>
[5, 11)	9	6	$9 / 6 = 1,5$
[11, 19)	16	8	$16 / 8 = 2$
[19, 29)	20	10	$20 / 10 = 2$
[29, 31)	10	2	$10 / 2 = 5$
[31, 37]	27	6	$27 / 6 = 4,5$

Să se calculeze frecvențele unitare și să se reprezinte distribuția sub formă de histogramă.

*Rezolvare:*

Se calculează ultimele două coloane de mai sus și se reprezintă histograma din prima figură ce urmează, în care bazele dreptunghiurilor sunt intervalele din prima coloană, iar înălțimile sunt frecvențele unitare, din ultima coloană.

*Observație:* S-a obținut o distribuție unimodală, clasa modală fiind intervalul [29,31) (vezi figura a).

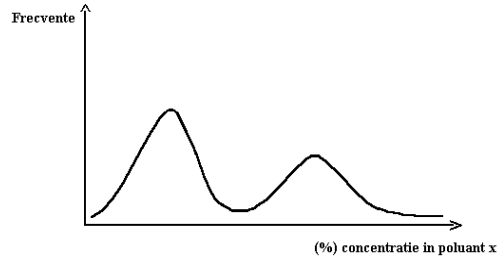


*Rezolvare eronată:*

Dacă, în mod greșit, s-ar lua drept înălțimi ale dreptunghiurilor frecvențele din coloana a doua, s-ar obține figura b, care reprezintă o distribuție bimodală.

## 6.

S-au recoltat mai multe probe dintr-un lac acoperindu-se relativ uniform suprafața sa. Măsurându-se conținutul în poluantul  $X$  din fiecare probă și construindu-se o histogramă, s-a constatat că „în spatele” acesteia se află o distribuție de forma alăturată.



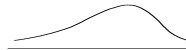
Studiindu-se repartitia stațiilor de prelevare a probelor s-a observat că valorile mai mari de concentrație se află, de regulă, în apropierea malurilor pline de stufăriș. (a) Propuneți o explicație ecologică. (b) Cum vor trebui prelucrate ulterior datele ?

*Rezolvare:*

- Poluantul este reținut de stufăriș sau provine de pe uscat din jurul lacului și încă nu s-a răspândit în tot lacul, etc.
- Distribuția trebuie „decupată” în două distribuții unimodale.

## 7.

a. Bifați numai afirmațiile pe care le considerați corecte și complete, în legătură cu următoarea distribuție:

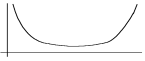
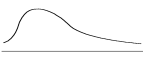




- Desenul reprezintă o distribuție unimodală asimetrică de dreapta.
  - Distribuția este bimodală, asimetrică de dreapta.
  - Distribuția este unimodală.
  - Distribuția din desen este asimetrică de dreapta.
  - Distribuția este unimodală, asimetrică de stânga.
  - Repartiția din desen este unimodală și asimetrică la dreapta.
  - Repartiția este simetrică.
- b. Bifați numai afirmațiile pe care le considerați corecte și complete:
- Scopul general pentru care grupăm fără pierdere de informație este câștigul în relevanță.
  - Motivul general pentru care grupăm, cu sau fără pierdere de informație, este câștigul în relevanță care se obține astfel.

14 **Dragomirescu L.**, 2003, *Lucrări practice de biostatistică. Ediția a III-a revăzută și adăugită*, Editura "Agronomica", București, 264p. ISBN 973-86668-0-5.

- Scopul general pentru care grupăm cu pierdere de informație este câștigul în relevanță.
- Obiectivul statistic urmărit prin grupare este sesizarea „în spatele” distribuției empirice a unei forme tip de distribuție.
- Scopul general al grupării este relevarea clopotului lui Gauss.
- Scopul medical al grupării este relevarea unei forme tip de distribuție.

c. Bifați numai formele tip de distribuție care exprimă ideea de omogenitate:

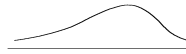
- 
- 
- 
- 

d. Bifați afirmațiile corecte:

- O distribuție unimodală și simetrică exprimă cel mai bine ideea de omogenitate absolută.
- O distribuție unimodală și simetrică exprimă cel mai bine (față de distribuțiile unimodale neconcentrate într-un punct) ideea de tendință centrală.
- O distribuție concentrată într-un punct exprimă ideea de omogenitate maximum posibilă, absolută.
- O distribuție uniformă ilustrează ideea omogenității maxime, absolute.
- O distribuție concentrată într-un punct exprimă ideea de tendință centrală.
- O distribuție uniformă ilustrează ideea eterogenității maxime, absolute.

*Rezolvare :*

a. Bifați numai afirmațiile pe care le considerați corecte și complete, în legătură cu următoarea distribuție:



✓ **Desenul reprezintă o distribuție unimodală asimetrică de dreapta.**

- Distribuția este bimodală, asimetrică de dreapta.

*Răspunsul este greșit, distribuția fiind unimodală (are o singură „cocoașă” nu două). În plus, tipul de asimetrie nu se definește decât pentru distribuții unimodale.*

- Distribuția este unimodală.

*Răspunsul este greșit fiind incomplet. Distribuția este, într-adevăr unimodală, dar nu s-a precizat tipul de asimetrie.*

✓ **Distribuția din desen este asimetrică de dreapta.**

Răspunsul este corect și complet prin deducție, deoarece tipul de asimetrie (stânga, dreapta) este definit numai pentru distribuțiile unimodale, iar distribuția este unimodală. Neintrând în discuție tipul de asimetrie și pentru distribuții bi sau multimodale este suficient să precizăm doar tipul de distribuție, deși este de preferată descrierea detaliată: „unimodală asimetrică de dreapta”.

Distribuția este unimodală, asimetrică de stânga.

*Răspunsul este greșit. Este unimodală, dar este asimetrică de dreapta, „coada” (partea lungă a) distribuției fiind în stânga, deci „capul” este în dreapta.*

✓ **Repartiția din desen este unimodală și asimetrică la dreapta.**

Răspunsul este corect, distribuția fiind unimodală și asimetrică de dreapta, deoarece „capul” distribuției este în dreapta. Termenul repartiție este sinonim cu cel de distribuție, iar sintagma „la dreapta” este acceptată alături de sintagma „de dreapta”.

Repartiția este simetrică.

*Răspunsul este greșit, deoarece repartiția sau, altfel spus distribuția, nu admite o axa de simetrie verticală pentru a fi considerată simetrică.*

---

b. Bifați numai afirmațiile pe care le considerați corecte și complete:

Scopul general pentru care grupăm fără pierdere de informație este câștigul în relevanță.

*Răspuns incomplet. Câștigul în relevanță este într-adevăr scop sau motiv general, dar acesta se poate atinge de cele mai multe ori și prin gruparea cu pierdere de informație.*

✓ **Motivul general pentru care grupăm, cu sau fără pierdere de informație, este câștigul în relevanță care se obține astfel.**

Răspuns corect. Câștigul în relevanță este motivul general al grupării.

Scopul general pentru care grupăm cu pierdere de informație este câștigul în relevanță.

*Răspuns incomplet. Câștigul în relevanță este într-adevăr scop sau motiv general, dar acesta se poate atinge uneori și prin gruparea fără pierdere de informație.*

✓ **Obiectivul statistic urmărit prin grupare este sesizarea „în spatele” distribuției empirice a unei forme tip de distribuție.**

Răspuns corect. Relevarea unei forme tip este, într-adevăr, obiectivul statistic, dar nu medical, nici final, al grupării.

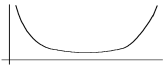
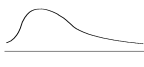


Scopul general al grupării este relevarea clopotului lui Gauss.

*Răspuns incorect. Relevarea unei forme tip definite în statistică, în particular a clopotului lui Gauss nu poate fi scop general. Scopul general este câștigul în*

relevantă. În plus, este o mare greșeală să se „vadă” în spatele oricărei distribuții un clopot al lui Gauss. Acesta apare relativ rar și rar în mod direct.

Scopul medical al grupării este relevarea unei forme tip de distribuție.  
*Răspuns incorect. Relevarea unei forme tip definite în statistică nu poate fi scop medical. Poate fi doar un mijloc tehnic intermediar pentru un scop medical.*

c. Bifați numai formele tip de distribuție care exprimă ideea de omogenitate:

<input type="checkbox"/>		<input checked="" type="checkbox"/>		<input type="checkbox"/>		<input checked="" type="checkbox"/>	
	<i>Răspuns incorect. Distribuție bimodală. „Cine aleargă după doi iepuri nu prinde nici-unul.”</i>		<i>Răspuns corect. Distribuție unimodală, deci omogenă.</i>		<i>Răspuns incorect. Distribuție omni-modală. Exprimă maximum de eterogenitate, contrariul omogenității.</i>		<i>Răspuns corect. Distribuție unimodală, concentrată într-un punct. Exprimă maximum de omogenitate, omogenitatea absolută.</i>

d. Bifați afirmațiile corecte:

O distribuție unimodală și simetrică exprimă cel mai bine ideea de omogenitate absolută.

*Răspuns greșit deoarece, deși ideea de omogenitate este exprimată în „limbajul repartițiilor” printr-o distribuție unimodală, pentru realizarea omogenității absolute este nevoie ca distribuția să se concentreze într-un singur punct.*

**O distribuție unimodală și simetrică exprimă cel mai bine (față de distribuțiile unimodale neconcentrate într-un punct) ideea de tendință centrală.**

*Răspuns corect. Distribuția concentrată într-un punct realizează sub formă extremă, „perfectă” tendința centrală. De aceea practic nu mai putem vorbi de o „tendință” ci de „atingerea perfectă a țintei”.*

**O distribuție concentrată într-un punct exprimă ideea de omogenitate maximum posibilă, absolută.**

*Răspuns corect.*

O distribuție uniformă ilustrează ideea omogenității maxime, absolute.

*Răspuns greșit deoarece distribuția având toate valorile posibile realizate de același număr de ori, exprimă, dimpotrivă, eterogenitatea maximă.*

O distribuție concentrată într-un punct exprimă ideea de tendință centrală.

*Răspuns greșit deoarece distribuția având toate valorile concentrate în același punct este impropriu să mai folosim termenul de „tendință”.*

**O distribuție uniformă ilustrează ideea eterogenității maxime, absolute.**



Răspuns corect.

(Testul a fost conceput după modelul de la link-ul [MCQs section](#) din site-ul [www.medstatsaag.com](http://www.medstatsaag.com) asociat lucrării [11].)

## Exerciții sau probleme propuse

### 8.

S-a notat gravitatea (gradul) arsurilor la pacienții tratați într-un spital și a rezultat următoarea serie statistică: I I I III III II II I II III I I I II II II I. (a) Din ce tip de variabilă este alcătuită seria statistică de mai sus ? (b) Realizați tabelul statistic simplu corespunzător seriei și reprezentați distribuția de frecvențe prin tipul de reprezentare grafică adecvată datelor.

### 9.

În cadrul unui studiu privind impactul poluării cu cupru asupra populației umane, s-au prelevat probe de urină de la un eșantion de indivizi ce lucrează în industria cuprului. Rezultatele obținute (exprimate în  $\text{mg}/100 \text{ cm}^3$ ) au fost următoarele: 0,41 0,52 0,58 0,63 0,64 0,67 0,69 0,71 0,72 0,72 0,74 0,76 0,78 0,79 0,80 0,89. (a) Deși volumul este  $< 50$  realizați gruparea cu intervale de clasă egale și reprezentați seria statistică prin tipul de reprezentare grafică adecvată tipului de variabilă. (b) Ce formă tip de distribuție s-a obținut?

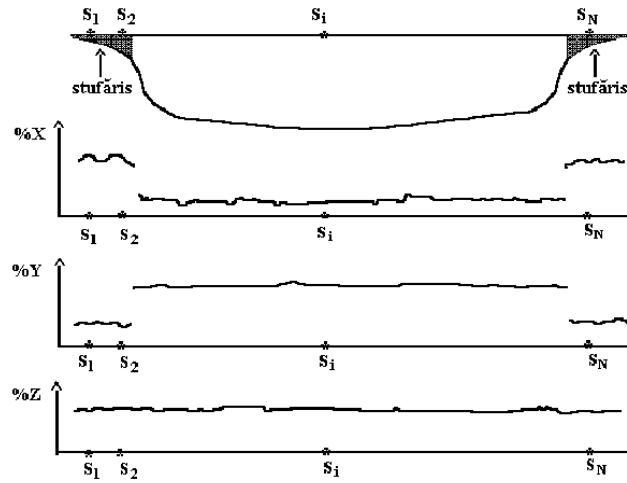
### 10.

Logaritmați în baza 10 datele din problema rezolvată 4, rotunjindu-le la prima zecimală. Apoi (b) grupați-le folosind intervale de grupare egale și, orientativ, formula lui Sturges. (b) Desenați histograma distribuției grupate și (c) caracterizați statistic distribuția. (d) Explicați de ce s-a obținut o distribuție de tipul respectiv.

## 11.

Dintr-o baltă s-au recoltat probe de apă de suprafață de-a lungul transectului reprezentat la începutul figurii alăturate, în care  $S_1, S_2, \dots, S_i, \dots, S_N$  sunt stațiile de recoltare.

Analizându-se probele s-au identificat poluanții  $X, Y, Z$  în concentrațiile marcate pe desenele plasate după transect.

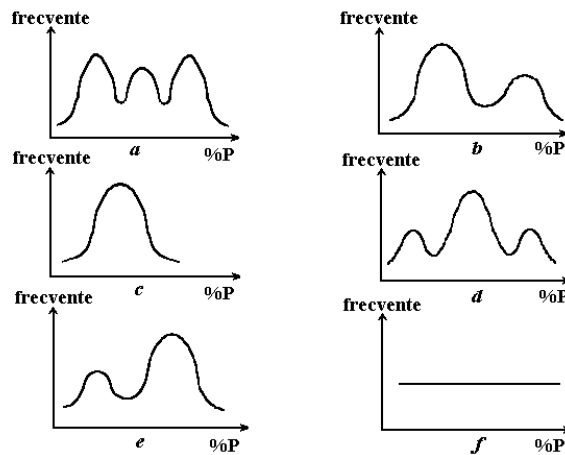


Datele au fost grupate, construindu-se câte o histogramă.

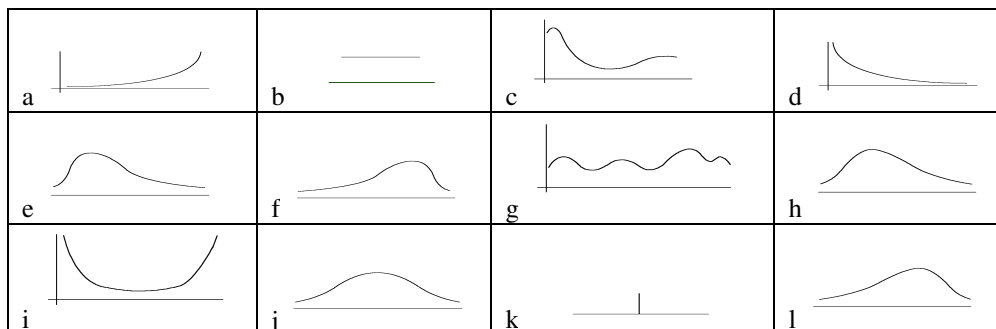
Să se precizeze pentru fiecare poluant în parte

$$P = X, Y, Z$$

care din curbele de frecvență alăturate exprimă forma histogramei obținute.



12.



1. Ordoneți ascendent după eterogenitate distribuțiile: g, c, b, k, i, j.
2. Ordoneți ascendent după omogenitate distribuțiile: g, c, b, k, i, j.
3. Enumerați distribuțiile simetrice din tabloul anterior.
4. Care distribuții sunt asimetrice (inclusiv extrem asimetrice) ?
5. Care sunt distribuțiile pentru care nu are sens calculul mediei ?
6. Cum trebuie tratate statistic descriptiv distribuțiile c, g, i ?
7. Denumiți tipul fiecărei distribuții din tablou.
8. Cum se pot trata statistic descriptiv distribuțiile a, d, e, f, h, l pentru a exprima cel mai bine ideea de tendință centrală ?
9. Care este cea mai grosolană eroare statistică în tratarea descriptivă a unei serii de măsurători ?
10. Cine trebuie să facă gruparea măsurătorilor: a. calculatorul; b. biometricianul singur; c. medicul singur; d. medicul împreună cu biometricianul cu sau fără calculator.
11. Pentru gruparea măsurătorilor statistica oferă medicului: a. o teorie a grupării; b. doar unele reguli empirice, orientative; c. proceduri automatizate.
12. Precizia rezultatelor finale depinde de: a. numărul de zecimale utilizat în calculele intermediare; b. precizia măsurătorilor inițiale; c. ambele; d. nici-una.