

### § 3.4. ▶ Alți indicatori de localizare

În afară de indicatori de tendință centrală, care au sens doar pentru distribuțiile unimodale, putem vorbi de indicatori de *localizare a tendințelor extreme* sau *intermediare*, valabili pentru orice distribuții. De pildă, *valorile minimă și maximă* dintr-un șir (notate în mod curent,  $x_{min}$ , respectiv  $x_{max}$ ) sunt exemple de indicatori de localizare a extremelor, cu sens pentru orice distribuție.

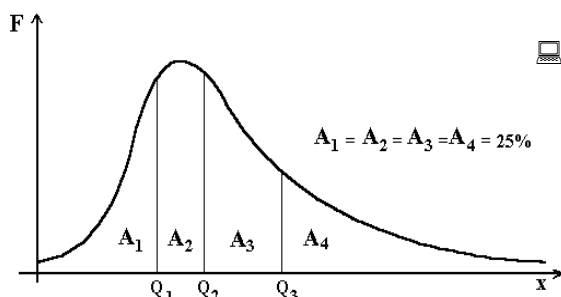
În continuare, generalizând modelul "geometric" al medianei, vom introduce o gamă foarte utilizată de indicatori de localizare (*cuartilele*, *decilele*, și *centilele*) și o generalizare a acestora (*cuantilele de ordinul m*). În finalul paragrafului se prezintă generalizarea conceptelor de *cuantilă inferioară*, respectiv *superioară de ordinul m*, și anume  $\alpha$ -*cuantila inferioară*, respectiv *superioară*. Ultimele două noțiuni vor fi extrem de importante în special pentru statistica inductivă.

#### 3.4.1. Cuartile

1° **Notății:**  $Q_1$ ,  $Q_2$ , respectiv  $Q_3$ .

2° **Definiții**

- În cazul unei *curbe de frecvențe* (distribuție continuă a unei variabile continue), **cuartilele** = cele 3 puncte care împart aria de sub curba de frecvențe în 4 arii egale (fiecare arie reprezentând 25% din întreaga arie de sub curbă).



*Observație:* Cuartila a doua,  $Q_2$ , este chiar mediana,  $Me$ , și este, deci, un indicator de tendință centrală, în timp ce cuartilele întâia și a treia,  $Q_1$ , respectiv  $Q_3$ , sunt indicatori de localizare a unor tendințe intermediare.

- În cazul *seriilor statistice*, **cuartile** = trei valori care împart seria statistică ordonată crescător în patru subserii de volume egale (volumele fiind măsurate în număr de unități statistice și, eventual, părți ale acestora).

$Q_1$  se numește **cuartilă inferioară** și lasă la stânga sa, în seria statistică ordonată crescător, 25% din termeni și, eventual, părți ale acestora.

$Q_2$  este mediana seriei.

$Q_3$  se numește **cuartilă superioară** și lasă la stânga sa, în seria statistică ordonată crescător, 75% din termeni și, eventual, părți ale acestora.

În acest caz, definiția poate conduce la valori unice doar cu convenții suplimentare, întocmai ca la definiția medianei în același context. Convențiile necesare pentru stabilirea unor valori unice vor rezulta din exemplul următor.

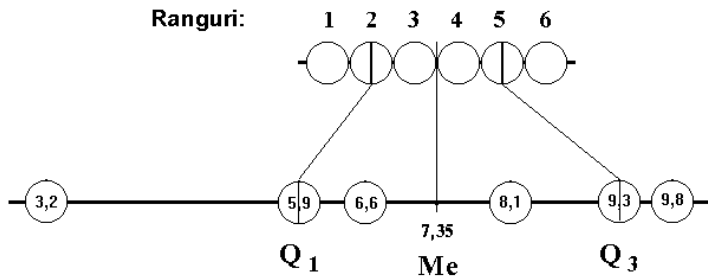
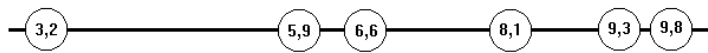
#### Exemplul 3.4.1.

Fie următoarea serie de 6 concentrații de oxigen măsurate în mg/l, în apa din Delta Dunării și ordonate crescător:

3,2 5,9 6,6 8,1 9,3 9,8.

Numerele ordonate sunt ca niște mărgelile înșirate, la diverse distanțe, pe o ață. Strângem mărgelile unele lângă altele "desființând" distanțele. Astfel numerele devin ranguri. Tăiem apoi acest șirag în patru părți egale - de câte o mărgelă și jumătate, în acest caz - ca în mijlocul figurii următoare.

**Seria statistică ordonată ascendent:**  
(Valori)



Cuartila inferioară,  $Q_1$ , va “tăia” mijlocul “mărgelei” a 2-a, adică va fi 5,9. Mediana va cădea între cea de-a 3-a și cea de-a patra “mărgea”. Va fi semisuma acestora:  $Me = (6,6 + 8,1) / 2 = 7,35$ . Cuartila superioară  $Q_3$ , va “tăia” mijlocul mărgelei a 5-a. Va fi, deci, 9,3.

**În practică**, calculul cuartilei  $Q_l$ , pentru  $l = 1, 2, 3$ , se face astfel, conform convențiilor introduse mai sus:

- Ordonăm ascendent seria de volum  $N$ .
- Calculăm rangul cuartilei respective  $rg(Q_l) = N \cdot (l / 4)$ .
- Dacă  $rg(Q_l)$  este număr fracționar, îl rotunjim prin adaos și  $Q_l$  este termenul cu rangul rotunjit. În caz contrar,  $Q_l$  este semisuma dintre termenul cu rangul  $rg(Q_l)$  și următorul termen.

În cazul exemplului anterior, seria ordonată ascendent fiind:

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
3,2	5,9	6,6	8,1	9,3	9,8

$rg(Q_1) = 6 \cdot (1 / 4) = 1 \frac{2}{4}$ , rotunjit prin adaos = **2**  $\Rightarrow Q_1 = x_2$  din seria ordonată ascendent = 5,9.

$rg(Q_2) = 6 \cdot (2 / 4) = 3 \Rightarrow Q_2 = Me = \frac{x_3 + x_4}{2}$  ( $x_3, x_4$  din seria ordonată ascendent) =  $\frac{6,6 + 8,1}{2} = 7,35$ .

$rg(Q_3) = 6 \cdot (3 / 4) = 4 \frac{2}{4}$ , rotunjit prin adaos = **5**  $\Rightarrow Q_3 = x_5$  din seria ordonată ascendent = 9,3.

### Exemplul 3.4.1’.

Fie seria de volum 4 ordonată ascendent: 1; 2; 8; 8. Cuartilele vor fi numerele marcate cu bold sub seria ordonată:

1	2	8	8
	<b>1,5</b>	<b>5</b>	<b>8</b>

### Exemplul 3.4.1’’.

Să se calculeze cuartilele pentru seria 8; 7; 3; 1; 2.

*Rezolvare:*

Ordonăm ascendent seria și marcăm cu bold cuartilele:

1	2	3	7	8
	<b>2</b>	<b>3</b>	<b>7</b>	

## 3.4.2. Decile și centile

În mod analog, se introduc noțiunile de *decile* ( $D_1, D_2, \dots, D_9$ ) și de (*per*)centile ( $C_1, C_2, \dots, C_{99}$ ), respectiv de *decilă inferioară* ( $D_l$ ), *decilă superioară* ( $D_b$ ), *centilă inferioară* ( $C_l$ ) și *centilă superioară* ( $C_{99}$ ).

Definițiile decilelor, respectiv, ale centilelor se pot obține înlocuind în definiția cuartilelor cuvântul “trei” cu “nouă”, respectiv cu “nouăzeci și nouă” și cuvântul “pătrimi” cu “zecimi”, respectiv cu “sutimi”. Algoritmul de calcul al acestora se poate obține, de asemenea, înlocuind în algoritmul de

calcul al cuartilelor, expresia " $N \cdot (I / 4)$ " cu " $N \cdot (I / 10)$ ", respectiv cu " $N \cdot (I / 100)$ ". Cele mai utilizate centile sunt  $C_5$  și  $C_{95}$ .

### 1° Metodă de calcul rapid

Pentru calculul rapid al centilelor propunem utilizarea următoarei metode. Aceasta presupune trei pași.

**Pasul 1.** Pornim de la distribuția de frecvențe relative procentuale – a se vedea primele două coloane din tabela următoare. În coloana 1 (notată Col. 1) sunt trecute, ordonat ascendent, valorile distincte al seriei, iar în coloana 2 (Col. 2) frecvențele relative procentuale ale valorilor din prima coloană - denumite pe scurt "procente".

**Pasul 2.** Construim coloana 3, care cuprinde frecvențele relative procentuale cumulate – pe scurt, "procentele cumulate". Modul de calcul al acestora este descris de săgețile din tabelă:

- primul element se transcrie din coloana de procente,
- oricare din următoarele elemente este suma dintre:
  - elementul din aceeași linie din coloana de procente și
  - elementul precedent din propria coloană (de procente cumulate).

### Exemplul 3.4.2.

S-a măsurat greutatea, în kilograme, pentru 103 băieți de 17 ani (date Dr. Cristiana Glavce). Calculându-se procentele **valorilor distincte** și *procentele cumulate*, s-a obținut tabela următoare, în care numerele din coloanele 2 și 3 sunt procente.

Col. 1	Col. 2	Col. 3	Col. 1	Col. 2	Col. 3	Col. 1	Col. 2	Col. 3
	%	%		%	%		%	%
<b>44</b>	1,0	1,0	<b>57</b>	2,8	32,8	<b>66</b>	2,9	78,6
<b>46</b>	1,9	2,9	<b>58</b>	1,0	33,8	<b>67</b>	1,9	80,5
<b>47</b>	1,9	4,8	<b>59</b>	4,9	38,7	<b>68</b>	1,0	81,5
<b>49</b>	2,8	7,6	<b>60</b>	6,8	45,5	<b>69</b>	6,8	88,3
<b>51</b>	1,0	8,6	<b>61</b>	7,8	53,3	<b>70</b>	2,9	91,2
<b>52</b>	1,9	10,5	<b>62</b>	5,8	59,1	<b>71</b>	3,9	95,1
<b>53</b>	1,0	11,5	<b>63</b>	1,0	60,1	<b>72</b>	1,0	96,1
<b>54</b>	6,8	18,3	<b>63,5</b>	1,0	61,1	<b>75</b>	1,9	98,0
<b>55</b>	3,9	22,2	<b>64</b>	7,8	68,9	<b>77</b>	1,0	99,0
<b>56</b>	7,8	30,0	<b>65</b>	6,8	75,7	<b>80</b>	1,0	100,0

**Pasul 3.** Determinăm centila dorită. Pentru aceasta căutăm în coloana 3 - de procente cumulate - cel mai apropiat procent mai mare sau egal cu indicele centilei respective. Dacă procentul cumulat, astfel determinat, este mai mare strict decât indicele centilei, valoarea din coloana 1 de pe aceeași linie va fi centila căutată. În caz de egalitate, centila va fi semisuma dintre valoarea din coloana 1 de pe aceeași linie și valoarea de pe linia următoare.

### Exemplul 3.4.2'.

Pentru centila  $C_3$  găsim procentul cumulat 4,8 care este pe linia valorii 47. Deoarece  $4,9 > 3$  rezultă  $C_3 = 47$ .

În mod analog, pentru centila  $C_{30}$  găsim procentul cumulat 30, care este pe linia valorii 56. Procentul cumulat fiind egal cu indicele centilei, rezultă  $C_{30} = (56+57)/2 = 56,5$ .

✓ Metoda poate fi utilizată și pentru calculul decilelor, cuartilelor și al medianei:

### Exemplul 3.4.2''.

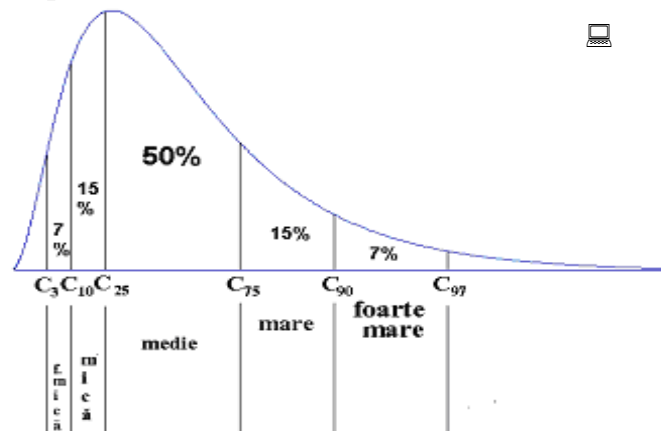
Decilele se calculează ținând cont că  $D_1 = C_{10}$ ,  $D_2 = C_{20}$ , etc. Analog, cuartilele:

$Q_1 = C_{25} = 56$ , deoarece 30,1 - procentul cumulat superior cel mai apropiat de 25 - se află pe linia lui 56. Analog,  $Q_2 = Me = C_{50} = 61$ , iar  $Q_3 = C_{75} = 65$ .

- ✓ În practică vom calcula centile doar pentru serii de volum mare ( $n > 200$ ), căci o regulă de bun simț ne cere să "împărțim" în o sută de părți doar serii cu număr mult mai mare de unități. Exemplul 3.4.2. cu volum 103 are doar rol didactic.

## 2° Scări de clasificare bazate pe centile

Cu ajutorul unor centile se alcătuiesc scări de clasificare. De exemplu, în practica antropologică se pot utiliza centilele  $C_3$ ,  $C_{10}$ ,  $C_{25}$ ,  $C_{75}$ ,  $C_{90}$  și  $C_{97}$  pentru stabilirea categoriilor unor dimensiuni (*foarte mică, mică, medie, mare și foarte mare*) [14].



Deci o scară de clasificare cu 5 trepte, bazată pe centile este următoarea:

Interval de variație a dimensiunii $x$ :	Diagnostic (clasă) dimensiune:
$C_3 \leq x < C_{10}$	foarte mică
$C_{10} \leq x < C_{25}$	mică
$C_{25} \leq x < C_{75}$	medie
$C_{75} \leq x < C_{90}$	mare
$C_{90} \leq x < C_{97}$	foarte mare

Valorile mai mici decât  $C_3$  sau mai mari decât  $C_{97}$  sunt excluse din serie.

### Exemplul 3.4.2''''.

Calculând pentru distribuția din exemplul 3.4.2. - după modelul anterior - centilele necesare acestei scări de clasificare, obținem  $C_3 = 47$ ,  $C_{10} = 52$ ,  $C_{25} = 56$ ,  $C_{75} = 65$ ,  $C_{90} = 70$  și  $C_{97} = 75$ .

Intervalele și denumirile vor fi:

$47 \leq x < 52$	foarte mic ca greutate	Valorile mai mici decât 47 kg și cele mai mari sau egale cu 75 kg sunt excluse din clasificare.
$52 \leq x < 56$	mic ca greutate	
$56 \leq x < 65$	mediu ca greutate	
$65 \leq x < 70$	mare ca greutate	
$70 \leq x < 75$	foarte mare ca greutate	

- ✓ Recomandăm utilizarea scărilor bazate pe centile în cazul distribuțiilor asimetrice. Acestea au aceeași logică de construcție ca și mediana  $Me$  - care este centila  $C_{50}$  - și care exprimă mai bine tendința centrală pentru distribuțiile unimodale asimetrice. În cazul distribuțiilor simetrice recomandăm scările sigmatice - vezi punctul 1° din 3.7.5.

### 3.4.3. Generalizare - cuantilele de ordinul $m$

- În cazul unei curbe de frecvențe (distribuție continuă a unei variabile continue), **cuantilele** sau **fractilele de ordinul  $m$**  = cele  $m - 1$  valori care împart aria de sub curba de frecvențe în  $m$  arii egale. Evident  $m = 2, 3, \dots, t$ .
- În cazul seriilor statistice, **cuantile** sau **fractile de ordinul  $m$**  =  $(m - 1)$  valori care separă seria statistică ordonată crescător în  $m$  subserii de volume egale, volumele fiind măsurate în număr de unități statistice și eventual fracțiuni ale acestora. Evident  $m = 2, 3, \dots, t$ . Pentru simplificarea enunțului următor vom nota **fractilele** prin  $F^m_1, F^m_2, \dots, F^m_b, \dots, F^m_{m-1}$ .

**În practică**, calculul celei de-a  $l$ -a cuantile de ordinul  $m$ , pentru  $l = 1, 2, 3, \dots, m - 1$  se face astfel:

- Ordonăm ascendent seria de volum  $N$ .
- Calculăm rangul cuantilei respective  $rg(F^m_l) = N \cdot (l / m)$ .
- Dacă  $rg(F^m_l)$  este număr fracționar, îl rotunjim prin adaos și cea de-a  $l$ -a cuantilă de ordinul  $m$  este termenul cu rangul rotunjit. În caz contrar, este semisuma dintre termenul cu rangul  $rg(F^m_l)$  și următorul termen.

Se observă că:

- ✓ *mediana este cuantila de ordinul 2,*
- ✓ *cuartilele sunt cuantilele de ordinul 4,*
- ✓ *decilele sunt cuantilele de ordinul 10,*
- ✓ *centilele sunt cuantilele de ordinul 100.*

Următoarele definiții generalizează noțiunile de *cuantilă superioară* și *cuantilă inferioară*.

---

**Cuantila superioară de ordinul  $m$**  = cea mai mare cuantilă de ordinul  $m$ .

**Cuantila inferioară de ordinul  $m$**  = cea mai mică cuantilă de ordinul  $m$ .

---

- ✓ *Cuantila superioară de ordinul  $m$  lasă la dreapta sa  $1 / m$* 
  - din aria distribuției (în cazul distribuțiilor continue), respectiv
  - din termenii seriei și eventual fracțiuni ale acestora (în cazul seriilor statistice).
- ✓ *Cuantila inferioară de ordinul  $m$  lasă la stânga sa  $1 / m$* 
  - din aria distribuției (în cazul distribuțiilor continue), respectiv
  - din termenii seriei și eventual fracțiuni ale acestora (în cazul seriilor statistice).

Pentru distribuțiile continue (teoretice) este utilă generalizarea noțiunilor de *cuantilă superioară*, respectiv *inferioară*, de ordinul  $m$ . Este vorba de a lucra cu orice proporții  $\alpha$  din arii (cu  $0 < \alpha < 1$ ), nu numai cu valori particulare de forma  $1 / m$ .

### 3.4.4. $\alpha$ -cuantile unilaterale și bilaterale

#### 1° $\alpha$ -cuantile unilaterale

Fie o *curbă de frecvențe* (distribuție continuă a unei variabile continue),

---

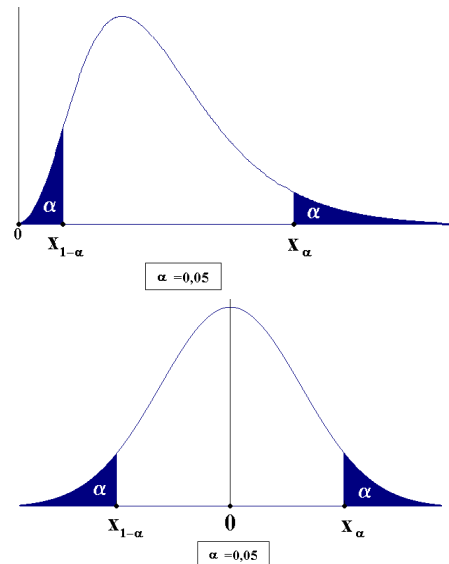
**$\alpha$ -cuantila unilaterală superioară** = punctul care lasă la DREAPTA sa proporția  $\alpha$  (respectiv procentul  $\alpha \cdot 100$  %) din aria distribuției. Se notează  $x_\alpha$ .

---

Pentru distribuțiile teoretice cu care se lucrează în teoria probabilităților și în statistica inductivă,  *$\alpha$ -cuantilele unilaterale superioare* importante în aplicații sunt tabelate. Tabelele conțin fie  *$\alpha$ -cuantile* în funcție de  $\alpha$ , fie invers. Modul de consultare a acestor tabele va fi prezentat și exersat în paragraful dedicat distribuției normale (3.7.) și în cadrul capitolelor de statistică inductivă.

Definiția  *$\alpha$ -cuantilelor unilaterale inferioare* se obține înlocuind în definiția anterioară cuvântul DREAPTA cu STANGA. Se notează  $x_{1-\alpha}$  deoarece lasă la dreapta aria  $1 - \alpha$ .

- ✓ Se observă că  $x_{1-\alpha} = -x_{\alpha}$  pentru orice  $\alpha$  dacă și numai dacă distribuția este simetrică față de axa  $x = 0$  (simetrie denumită, pe scurt, "față de zero") ca în cazul distribuției normale standard și al distribuțiilor Student (care vor fi prezentate în cadrul statisticii inductive).



În caz contrar (adică în cazul distribuțiilor asimetrice sau simetrice dar nu față de zero), egalitatea de mai sus nu mai este valabilă. Acesta este, de exemplu, cazul distribuțiilor  $\chi^2$  (care vor fi prezentate, de asemenea, în cadrul statisticii inductive).

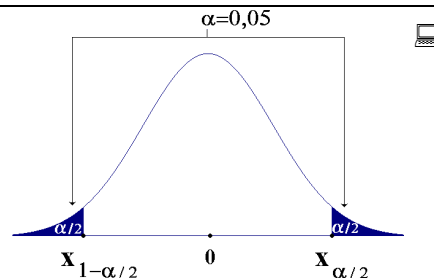
- ✓ În general, prin  $\alpha$ -cuantilă (*inferioară* sau *superioară*) vom înțelege  $\alpha$ -cuantilă unilaterală (*inferioară* sau *superioară*).

## 2° $\alpha$ -cuantile bilaterale

În cazul  $\alpha$ -cuantilelor *unilaterale*, aria  $\alpha$  a fost plasată într-o singură parte a distribuției. În cadrul paragrafului 10.2. dedicat testelor binare precum și în cadrul statisticii inductive se va vedea utilitatea plasării unui  $\alpha$  (dat) în ambele capete ale distribuției prin divizare în două arii egale (cu  $\alpha / 2$ ). În acest caz, vom vorbi de  $\alpha$ -cuantile *bilaterale*.

- **$\alpha$ -cuantila bilaterală superioară** = punctul care lasă la DREAPTA sa proporția  $\alpha / 2$  (respectiv procentul  $\alpha / 2 \cdot 100 \%$ ) din aria distribuției. Se notează  $x_{\alpha / 2}$ .
- **$\alpha$ -cuantila bilaterală inferioară** = punctul care lasă la STANGA sa proporția  $\alpha / 2$  (respectiv procentul  $\alpha / 2 \cdot 100 \%$ ) din aria distribuției. Se notează  $x_{1 - \alpha / 2}$ .

0,05-cuantilele *bilaterale* pentru distribuția normală standard sunt figurate în desenul alăturat.



- ✓  $\alpha$ -cuantila *bilaterală superioară* este  $\alpha / 2$ -cuantila *unilaterală superioară*,
- ✓  $\alpha$ -cuantila *bilaterală inferioară* este  $\alpha / 2$ -cuantila *unilaterală inferioară*.
- ✓ Cele două  $\alpha$ -cuantile *bilaterale* sunt egale și de semne contrare  $\Leftrightarrow$  distribuția este simetrică față de origine (conform ultimei proprietăți de la  $\alpha$ -cuantile *unilaterale* și ținând cont de observațiile anterioare).