

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

NOTIUNI GENERALE

Statistica este stiinta care se ocupa cu:

- Colectarea
- Gruparea
- Analiza
- Interpretarea datelor referitoare la anumite fenomene in scopul luarii unor decizii.

Clasificari:

1.

- a. **Statistica economica** (*Statistica pentru afaceri – Business Statistics*) este ramura statisticii care studiaza fenomenele si procesele economice cu aplicatii in asigurari, marketing, finante etc.
- b. **Statistica industrială** (*Engineering Statistics* sau *Industrials Statistics*) este ramura a statisticii care studiaza fenomenele industriale cum ar fi fiabilitatea sau controlul calitatii unor anumite produse.
- c. **Statistica sociala** este o ramura a statisticii care studiaza anumite fenomene sociale cu aplicati in demografie, recensaminte, analize politice, etc.
- d. **Statistica medicala** (sau *Biostatistica*) este o ramura a statisticii care studiaza fenomenele din medicina.

2.

- a. **Statistica descriptiva** se ocupa cu:
 - Colectarea
 - Inregistrarea
 - Gruparea datelor.
- b. **Statistica matematica** este o ramura a matematicii care asigura instrumentul matematic pentru analiza si interpretarea datelor oferite de *statistica descriptiva*.
- c. **Statistica inferentiala (inductiva)** se ocupa cu interpretarea datelor oferite de statistica descriptiva si folosirea acestora pentru a trage anumite concluzii si a lua decizii.

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

Definitia 1.

- a. Se numeste *populatia statistica* (*univers statistic, colectivitate statistica, lot*) o multime de obiecte despre care se colecteaza si inregistreaza date in scopul unei analize statistice. In statistica se presupune *populatia omogena* (unitatile populatiei sunt de acelasi tip)
- b. Se numeste *individ* (sau *unitate statistica*) orice element al unei populatii statistice.
- c. Se numeste *efectivul* sau *volumul* (notat frecvent cu N) unei populatii statistice numarul de indivizi ai populatiei respective (in cazul unei populatii finite).
- d. Se numeste *esantion statistic* (*selectie statistica* sau *sondaj statistic*) o submultime a unei populatii statistice.

Obsevatii.

1. Din diverse motive (economice, tehnice, etc.) in statistica nu se lucreaza cu intreaga populati ci cu *esantioane*.
2. Pentru ca analiza statistica sa fie concludenta alegerea *esantionului* trebuie sa respecte anumite *reguli de reprezentativitate* (structura si efectivul esantionului sa tina cont de structura si efectivul populatiei statistice).

Definitia 2.

Se numeste *variabila statistica* (sau *caracteristica*) o trasatura comuna a indivizilor unei populatii statistice, care face obiectul unei analize statistice.

Exemple:

- nota obtinuta de candidatii la un examen
- culoarea parului unui grup de persoane, etc.

Definitia 3.

Se numeste *varianta* sau *scor* al unei variabile statistice o valoare particulara a variabilei (rezultata, in general, in urma unei masuratori).

Exemple:

- nota 7 obtinuta de un candidat la un examen
- calitatea de promovat sau nepromovat a unui candidat care a paricipat la un examen, etc.

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

Obsevatii.

Variabilele statistice pot fi clasificate astfel:

1. Calitative (nu pot fi masurate si se exprima prin attribute)

Exemple:

- profesia;
- scoala absolvita, etc.

2. Cantitative (variabilele care pot fi masurate si exprimate numeric).

a. *Discrete* sau *discontinue* (variabilele care iau numai valori intregi).

Exemple:

- numarul copiilor dintr-o familie;
- numarul de locuitori ai unui oras, etc.

b. *Continue* (variabilele care iau valori intr-un intreg interval).

Exemple:

- media de absolvire;
- greutatea, etc.

3. Alternative (iau numai valorile 1 sau 0 dupa cum subiectii respectivi posedă o anumita însușire sau nu)- ex.:promovat/nepromovat, prezent/absent, etc.

Definitia 4.

Se numeste **serie statistica (serie de date)** un sir de valori ale unei variabile statistice.

Observatii.

1. Seriile statistice se pot clasifica in:

a. **Serii simple** – referitoare la siruri de date negrupate, necentralizate.

Exemplu: Zece elevi au obtinut la un test urmatoarele note 8,8,6,7,5, 9,7,5,6,8.

b. **Serii de frecventa** (*distributii, repartitii*) – referitoare la date grupate sau centralizate.

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

Exemplu: Zece elevi au obtinut la un test urmatoarele note:

Nota	Numar elevi (frecventa absoluta)
8	3
6	2
7	2
5	2
9	1
Total	10

- c. **Serii pe variante** – cand caracteristica este bine precizata (exemplele de la (a) si (b)).
- d. **Serii pe intervale** – cand caracteristica nu este precis exprimata ci este cuprinsa in anumite intervale.

Exemplu: 100 de elevi au obtinut la un test de cunostinte urmatoarele note:

Nota	Numar elevi (frecventa absoluta)
<5	5
5-5,99	10
6-6,99	15
7-7,99	30
8-8,99	20
9-9,99	15
10	5
Total	100

- e. **Serii unidimensionale, bidimensionale, multidimensionale** (*unicriteriale sau multicriteriale*) – dupa numarul de caracteristici considerate.
2. In cazul seriilor statistice pe intervale, in calculul statistic se utilizeaza frecvent valoarea din mijlocul intervalului.

Exemplu: 100 de elevi au obtinut la un test de cunostinte urmatoarele note:

Nota	Numar elevi (frecventa absoluta)	Nota medie
<5	5	3
5-5,99	10	5,50
6-6,99	15	6,50
7-7,99	30	7,50
8-8,99	20	8,50
9-9,99	15	9,50
10	5	10
Total	100	

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

Vom folosi de multe ori notatia $X = (x_1 \ x_2 \ . \ . \ . \ x_{n-1} \ x_n)$ pentru o serie statistica simpla sau notatia $X = \begin{pmatrix} x_1 & x_2 & . & . & . & x_{n-1} & x_n \\ f_1 & f_2 & . & . & . & f_{n-1} & f_n \end{pmatrix}$ pentru o serie statistica in frecventa, care inseamna unde x_1, x_2, \dots, x_n sunt valorile caracteristicii iar f_1, f_2, \dots, f_n sunt frecventele (absolute sau relative).

Definitia 5.

Fiind data seria statistica $X = \begin{pmatrix} x_1 & x_2 & . & . & . & x_{n-1} & x_n \\ f_1 & f_2 & . & . & . & f_{n-1} & f_n \end{pmatrix}$ si

$k \in \{1, 2, \dots, n\}$, vom numi:

- a. Frecventa *cumulata crescator* (absoluta, relativa) a valorii x_k numarul

$$\sum_{x_i \leq x_k} f_i;$$

- b. Frecventa *cumulata descrescator* (absoluta, relativa) a valorii x_k numarul

$$\sum_{x_i \geq x_k} f_i;$$

- c. Vom nota, de multe ori, *frecventa cumulata* a valorii x_k cu F_k , deducand din context daca este vorba de frecventa absoluta sau relativa cumulata crescator sau descrescator.

Propozitia 1

Fie $X = \begin{pmatrix} x_1 & x_2 & . & . & . & x_{n-1} & x_n \\ f_1 & f_2 & . & . & . & f_{n-1} & f_n \end{pmatrix}$ o serie statistica, astfel incat

$x_1 < x_2 < \dots < x_n$ si $k \in \{1, 2, \dots, n\}$. Atunci:

- a. Frecventa cumulata crescator a valorii x_k este $\sum_{i=1}^k f_i$.

- b. Frecventa cumulata descrescator a valorii x_k este $\sum_{i=k}^n f_i$.

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

Propozitia 2

Fie $X = \begin{pmatrix} x_1 & x_2 & \cdot & \cdot & \cdot & x_{n-1} & x_n \\ f_1 & f_2 & \cdot & \cdot & \cdot & f_{n-1} & f_n \end{pmatrix}$ o serie statistica, astfel incat $x_1 > x_2 > \dots > x_n$ si $k \in \{1, 2, \dots, n\}$. Atunci:

c. Frecventa cumulata crescator a valorii x_k este $\sum_{i=k}^n f_i$.

d. Frecventa cumulata descrescator a valorii x_k este $\sum_{i=1}^k f_i$.

Exemplu:

Seria statistica a varstei unui esantion de 234 de pacienti:

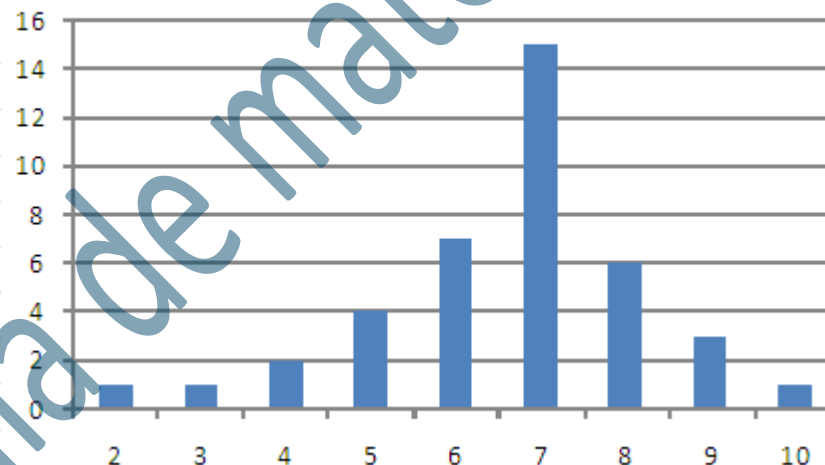
Varsta (caracteristica)	Frecventa absoluta (f_a)	Frecventa relativa (f_r)	Frecventa absoluta cumulata crescator (E_+)	Varsta (caracteristica)	Frecventa absoluta (f_a)	Frecventa relativa (f_r)	Frecventa absoluta cumulata crescator (E_+)
26	1	0.43%	1	53	10	4.27%	101
28	1	0.43%	2	54	11	4.70%	112
29	1	0.43%	3	55	14	5.98%	126
30	2	0.85%	5	56	11	4.70%	137
31	2	0.85%	7	57	9	3.85%	146
32	1	0.43%	8	58	19	8.12%	165
35	3	1.28%	11	59	5	2.14%	170
36	2	0.85%	13	60	9	3.85%	179
37	3	1.28%	16	61	13	5.56%	192
38	1	0.43%	17	62	5	2.14%	197
40	3	1.28%	20	63	4	1.71%	201
41	5	2.14%	25	64	4	1.71%	205
42	1	0.43%	26	65	6	2.56%	211
43	4	1.71%	30	66	2	0.85%	213
44	10	4.27%	40	67	4	1.71%	217
45	6	2.56%	46	68	3	1.28%	220
46	6	2.56%	52	69	4	1.71%	224
47	5	2.14%	57	70	1	0.43%	225
48	13	5.56%	70	71	2	0.85%	227
49	2	0.85%	72	72	2	0.85%	229
50	4	1.71%	76	74	1	0.43%	230
51	6	2.56%	82	77	2	0.85%	232
52	9	3.85%	91	78	2	0.85%	234
				Total	234		

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

EXEMPLE DE GRUPARE SI REPREZENTARE GRAFICA A DATELOR STATISTICE

1. Seria statistica a rezultatelor unui esantion de elevi la teza de matematica:

NOTA (valoarea caracteristicii)	Numar elevi (FA)
2	1
3	1
4	2
5	4
6	7
7	15
8	6
9	3
10	1



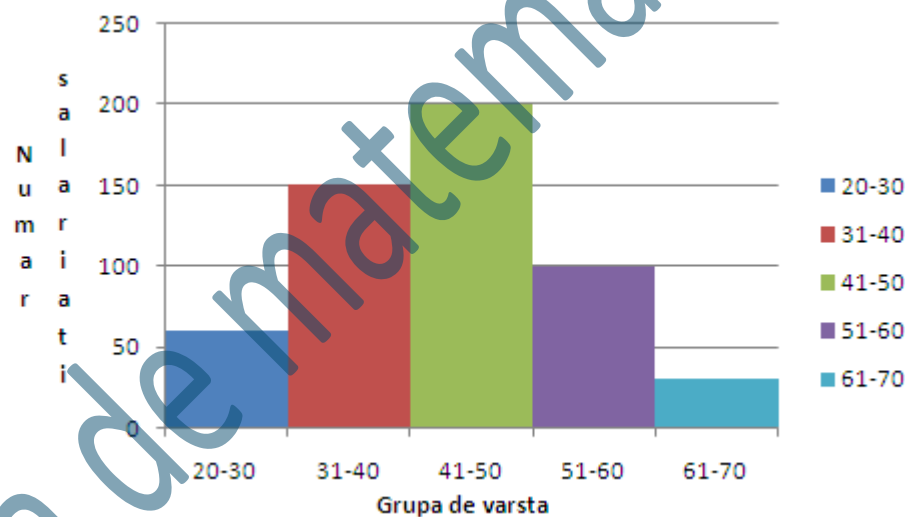
Comentarii:

- tabel cu o singura intrare: nota;
- populatia: elevii clasei;
- esantionul: intreaga populatie;
- caracteristica: nota la teza;
 - cantitativa;
 - discreta;
- **digrama cu coloane**

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

2. Seria statistica in frecventa absoluta a structurii de varsta a a salariatilor unei societati comerciale:

Varsta in ani	Numar salariati
20-30	60
31-40	150
41-50	200
51-60	100
61-70	30
Total	540



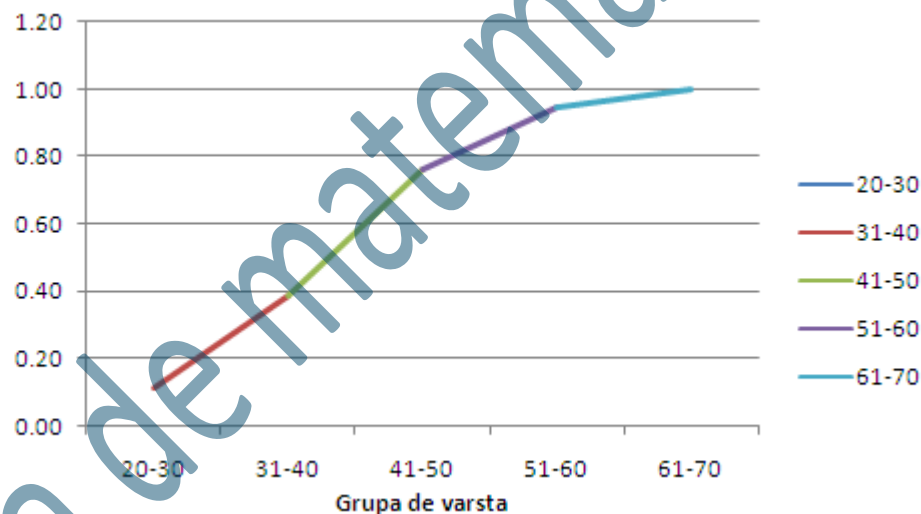
Comentarii:

- tabel cu o singura intrare: grupa de varsta;
- populatia: salariatii societatii;
- esantionul: intreaga populatie;
- caracteristica: grupa de varsta;
 - cantitativa;
 - continua;
- **histograma (coloane apropiate).**

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

3. Seria statistica in *frecventa cumulata* a structurii de varsta a a salariatilor unei societati comerciale:

Varsta in ani	Frecventa relativa
20-30	0.11
31-40	0.39
41-50	0.76
51-60	0.94
61-70	1.00
Total	540



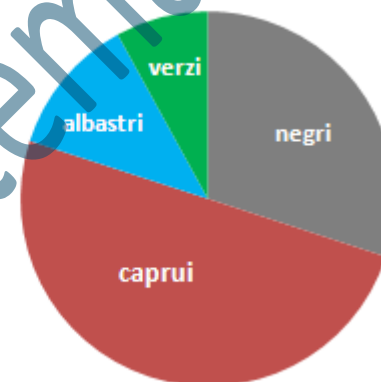
Comentarii:

- tabel cu o singura intrare: grupa de varsta;
- populatia: salariatii societatii;
- esantionul: intreaga populatie;
- caracteristica: grupa de varsta;
 - cantitativa;
 - continua;
- ogiva (poligonul sau curba frecventelor cumulate).

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

4. Seria statistica a culorii ochilor unui esantion de 50 de persoane:

CULOAREA OCHILOR (valoarea caracteristicii)	Numar persoane (FA)
negri	15
caprui	25
albastri	6
verzi	4



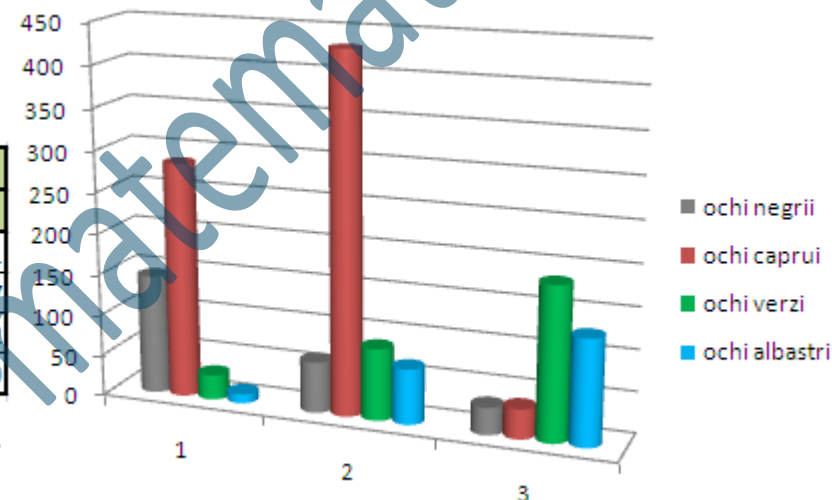
Comentarii:

- tabel cu o singura intrare: culoarea ochilor;
- caracteristica: culoarea ochilor;
 - calitativa;
- **digrama de structura (cu discuri)**

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

5. Seria statistica a culorii ochilor si a parului unei populatii de 1500 de copii:

		Culoarea ochilor				Total
		Negri	Caprui	Verzi	Albastri	
Culoarea parului	Negru(1)	145	285	30	11	471
	Castaniu(2)	62	431	87	67	647
	Blond(3)	33	36	185	128	382
	Total	240	752	302	206	1500



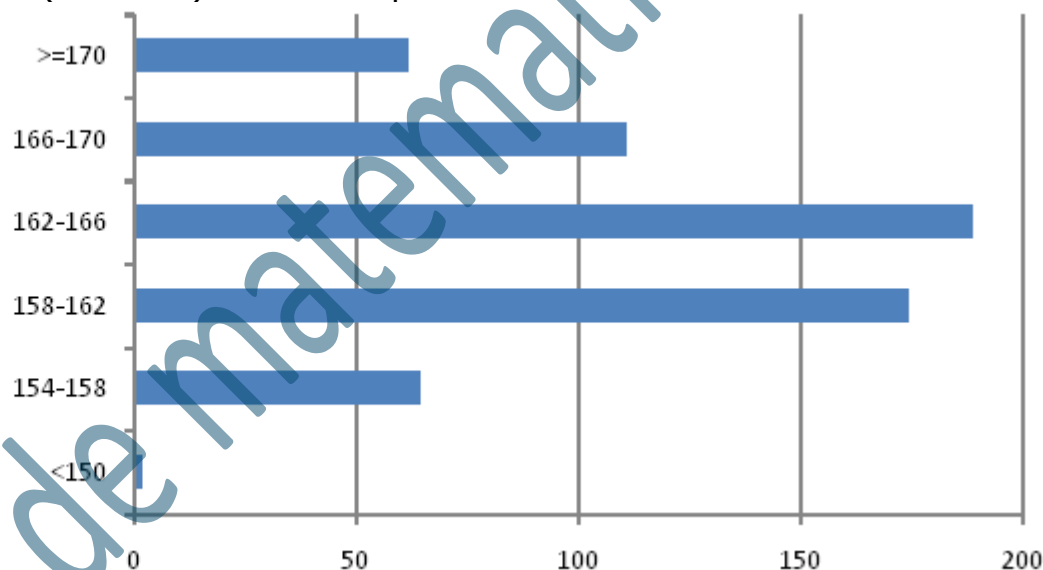
Comentarii:

- tabel cu o doua intrari: culoarea ochilor, culoarea parului;
- caracteristica: culoarea ochilor, culoarea parului;
 - calitativa;
 - serie statistica *bivariata* (cu doua caracteristici);
- **diagrama cu cilindri.**

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

6. Seria statistica a inaltimii indivizilor unei populatii (esantion) de 604 de persoane.

Talia(cm) (valoarea caracteristicii)	FA
<150	2
154-158	65
158-162	175
162-166	189
166-170	111
>=170	62
Total	604



Comentarii:

- tabel cu o singura intrare: intervalul de inaltime de forma $[a,b)$;
- caracteristica: intervalul de inaltime de forma $[a,b)$;
 - cantitativa;
 - continua;
 - intervalele (*clasele*) au lungimi egale;
 - **diagrama cu bare (benzi)**

Obsevatie.

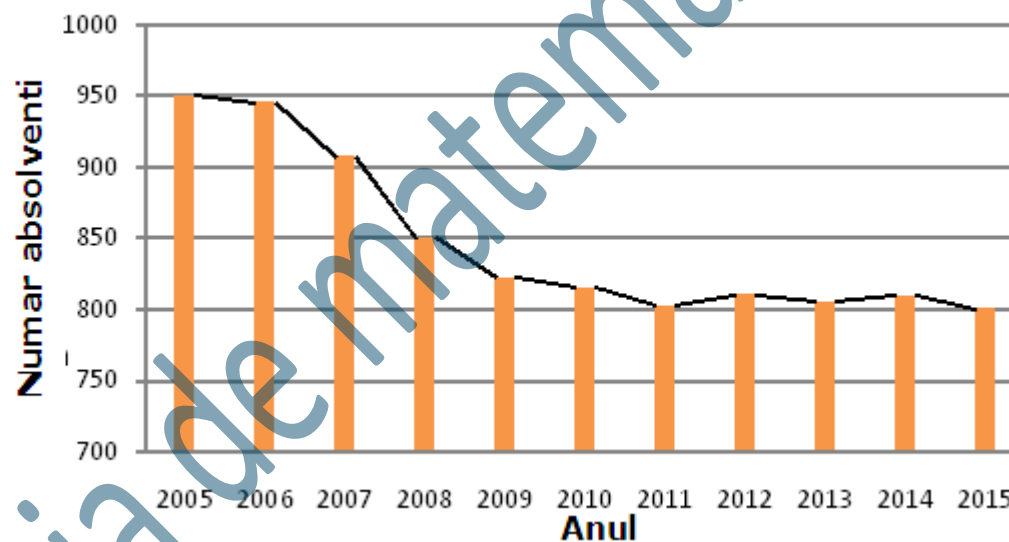
In cazul caracteristicilor continue, daca *numarul de intervale* k nu rezulta dintr-o logica specifica analizei statistice, se poate folosi de *formula lui Sturges*:

$$k \cong 1 + 3,322 \lg N, \quad N = \text{volumul populatiei}$$

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

7. Seria statistica a numarului de absolventi ai unei scoli intr-o anumita perioada de timp:

Perioada (caracteristica)	FA
2005	950
2006	945
2007	907
2008	850
2009	822
2010	815
2011	802
2012	810
2013	805
2014	809
2015	800



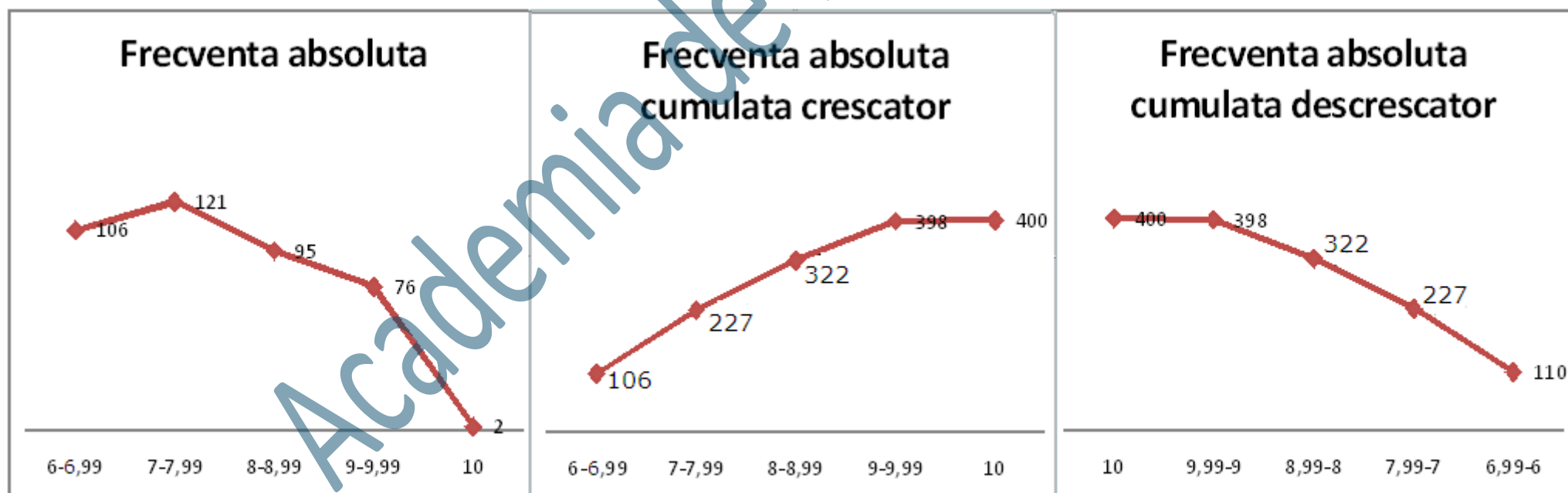
Comentarii:

- este o *serie cronologica* (prezinta evolutia in timp a unui fenomen)
- tabel cu o singura intrare: anul absolvirii;
- caracteristica: anul absolvirii;
 - cantitativa;
 - discreta;
- **diagrama cu coloane si poligonul frecventelor;**

CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

8. Seria statistica a mediilor elevilor unui liceu promovati la examenul de BACALAUREAT:

Transe de medii (caracteristica)	Frecventa absoluta (FA)	Frecventa absoluta cumulata crescator (FACC)	Frecventa absoluta cumulata descrescator (FACD)	Fecventa relativa (FR)	Fecventa relativa cumulata crescator (FRCC)	Fecventa relativa cumulata descrescator (FRCD)
6-6,99	106	106	400	0.265	0.265	1
7-7,99	121	227	398	0.3025	0.5675	0.995
8-8,99	95	322	322	0.2375	0.805	0.805
9-9,99	76	398	227	0.19	0.995	0.5675
10	2	400	106	0.005	1	0.275
Total	400					



CULEGEREA, GRUPAREA SI REPREZENTAREA GRAFICA A DATELOR STATISTICE

9. Diverse reprezentari grafice de serii statistice.

