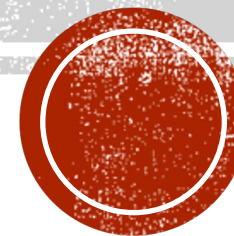


CURS 4 — INDICATORII SINTETICI AI DISTRIBUTIEI

Lect. univ. dr. Adrian Gorbănescu



PREZENTARE GENERALĂ

- ***statistica descriptivă ne oferă un set de indicatori numerici cu rolul de a descrie distribuția.***
- ***Analiza de frecvențe are utilitate limitată și prezintă o serie de dezavantaje (devine tot mai complicată pe măsură ce valorile sunt tot mai numeroase).***
- ***Indicatorii sintetici conturează o imagine a distribuției.***
- ***indicator sintetic este un „descriptor numeric care condensează într-o valoare unică o anumită caracteristică a întregii distribuții de valori” ([Popa, 2008, p. 58](#)).***



PREZENTARE GENERALA

- *Tipuri de indicatori sintetici:*
 1. *Indicatori ai tendinței centrale – reprezentativi pentru distribuție.*
 2. *Indicatori ai variabilității (împrăstierii) – ne informează despre cât de împrăstiate (diversificate) sunt valorile din distribuție.*
 3. *Indicatori ai formei distribuției – oferă informații despre forma distribuției.*



INDICATORI AI TENDINTEI CENTRALE

1. Media aritmetică

- ***Lb engleză – mean.***
- ***La nivel de eşantion se notează cu m .***
- ***La nivelul populației se notează cu μ .***

$$m = \frac{\sum X}{N}$$



INDICATORI AI TENDINTEI CENTRALE

Exemplu



Profesorului de Statistică îi place foarte mult înghețata și obișnuiește să mănânce cel puțin câte una pe zi. La sfârșitul unei săptămâni, într-un moment în care calculează bugetul familiei, se gândește „*Oare nu am cheltuit cam mulți bani pe înghețată? Câți bani am alocat, în medie, pe zi pentru consumul de înghețată în ultima săptămână?*”. Mai jos sunt prezentate sumele cheltuite în fiecare zi din săptămână.

Luni	Marți	Miercuri	Joi	Vineri	Sâmbătă	Duminică
10	15	13	13	12	15	13

$$m = \frac{\sum X}{N} \rightarrow m = \frac{10+15+13+13+12+15+13}{7} \rightarrow m = \frac{91}{7} \rightarrow m = 13.$$



INDICATORI AI TENDINTEI CENTRALE - MEDIA

▪ ***Proprietățile mediei aritmetice:***

- 1. Suma abaterii tuturor valorilor față de medie este egală cu 0.***
- 2. Suma pătratelor abaterilor tuturor valorilor de la medie este mai mică decât suma pătratelor abaterilor față de orice valoare din distribuție.***
- 3. Adăugarea/scăderea unei constante la fiecare valoare din distribuție determină creșterea/scăderea mediei cu acea constantă.***
- 4. Multiplicarea/împărțirea fiecărei valori a distribuției cu o constantă generează multiplicarea/împărțirea mediei cu acea constantă.***



INDICATORI AI TENDINTEI CENTRALE - MEDIA

- ***Media aritmetică este ușor influențată de valorile extreme (outliers).***
- ***valoarea extremă este o valoare foarte mică sau foarte mare comparativ cu celelalte scoruri din distribuție.***



INDICATORI AI TENDINTEI CENTRALE - MEDIA

- Pentru a calcula media aritmetică în R vom folosi funcția **mean**.
- `mean(bdcap4$ACE)`, unde:
- `mean` — este funcția din R care calculează media aritmetică
- `bdcap4` — reprezintă numele bazei de date
- `ACE` — este numele variabilei analizate.

```
mean(bdcap4$ACE)
```

```
[1] 55.45905
```

```
mean(bdcap4$ACE [bdcap4$Genu1==2])
```

```
[1] 56.2967
```



INDICATORI AI TENDINTEI CENTRALE - MEDIANA

- *Mediana (lb engleză = median) este valoarea care împarte distribuția în două jumătăți (50% din scoruri sunt mai mici decât ea, iar 50% din valori sunt mai mari).*
- *mediana corespunde rangului percentil 50 sau quartilei 2 (Q_2).*
- *În concluzie, mediana este quartila 2 (Q_2).*
- *Mediana nu este afectată de valori extreme.*



MEDIANA

■ *10 12 16 16 17 18 20*



INDICATORI AI TENDINTEI CENTRALE - MEDIANA

- *Mediana unei distribuții cu număr par de valori*

8 10 10 12 13 14 14 17



INDICATORI AI TENDINTEI CENTRALE - MEDIANA

- **median()** este funcția care permite calcularea medianei.
- `median(bdcap4$ACE)`, unde:
- `median` — este funcția din R care calculează mediana;
- `bdcap4` — reprezintă numele bazei de date;
- `ACE` — este numele variabilei analizate.

```
median(bdcap4$ACE)
```

```
[1] 54
```

```
median(bdcap4$ACE [bdcap4$Genu1==1])
```

```
[1] 53
```



INDICATORI AI TENDINTEI CENTRALE - MODUL

- ***Modul (lb engleză = mode) este valoarea cu frecvența absolută cea mai mare.***
- ***Dacă sunt două scoruri cu ceva mai mare frecvență de apariție vom spune că distribuția este bimodală.***
- ***Atunci când sunt cel puțin trei valori cu cea mai mare frecvență distribuția este multimodală.***
- ***10 12 13 13 13 15 15 Modul = 13***
- ***În distribuția 5, 6 , 8 , 8, 7, 5, 10, 8, 7, 5 modul este reprezentat de valorile 5 și 8 (apar de câte trei ori) – $M_o = 5, 8$.***



INDICATORI AI TENDINTEI CENTRALE - MODUL

```
getmode <- function(ACE)
{
  mod <- unique(ACE)
  mod[which.max(tabulate(match(bdcap4$ACE, mod)))]
}

getmode(bdcap4$ACE)
```

```
[1] 53
```



INDICATORI AI IMPRASTIERII

- ***Magazinele din Statistics Mall promovează o amplă campanie de reduceri pentru toate produsele. Într-un week-end managerul complexului comercial se hotărăște să înregistreze sumele cheltuite de clienți. În acest sens, la ieșirea din Mall clienții care doresc să participe la studiu pot completa o bază de date electronică în care sunt rugați să specifice genul și suma cheltuită. Sumele cheltuite de bărbați sunt cuprinse între 55 și 85 de lei, iar sumele cheltuite de femei variază între 55 și 75 de lei. În medie, bărbații au cheltuit 68 de lei, în timp ce femeile au cheltuit 65 de lei. În concluzie, persoanele înregistrate în baza de date au cheltuit sume cuprinse între 55 și 85 de lei, cheltuindu-se în medie 66,50 lei de persoană.***



INDICATORI AI ÎMPRĂSTIERII - AMPLITUDINEA

- ***amplitudinea (lb. engleză = range) exprimă diferența dintre valoarea cea mai mare și valoarea cea mai mică din distribuție.***

$R = \text{valoarea maximă} - \text{valoarea minimă}$

$$***$R = 85 - 55 \rightarrow R = 30.$***$$

- ***Amplitudinea poate fi foarte ușor afectată de valorile extreme***



INDICATORI AI ÎMPRĂSTIERII - AMPLITUDINEA

- amplitudinea poate fi calculată cu ajutorul funcției **range()**.
- `range(bdcap4$Stres)`, unde:
- `range` — este funcția din R care calculează amplitudinea;
- `bdcap4` — reprezintă numele bazei de date;
- `Stres` — este numele variabilei analizate.

```
range(bdcap4$Stres)
```

```
[1] 0 21
```



INDICATORI AI ÎMPRĂSTIERII – ABATEREA

INTERQUARTILA

- ***abatere interquartilă (lb. engleză, interquartile range) este diferența dintre valoarea corespunzătoare quartilei 3 și valoarea corespunzătoare quartilei 1.***
- ***nu este afectată de valori extreme***



INDICATORI AI ÎMPRĂSTIERII – ABATEREA INTERQUARTILA

- Quartilele pot fi afișate în R folosind funcția **quantile()**.
- `quantile(bdcap4$Stres)`, unde:
- `quantile` – este funcția din R care calculează quartilele;
- `bdcap4` – reprezintă numele bazei de date;
- `Stres` – este numele variabilei analizate.
- Abaterea interquartilă = $14 - 6$
- Abaterea interquartilă = 8

```
quantile(bdcap4$Stres)
```

0%	25%	50%	75%	100%
0	6	10	14	21



INDICATORI AI ÎMPRĂSTIERII – ABATEREA MEDIE (ABATEREA DE LA MEDIE)

- *Diferența dintre un scor din distribuție și medie se numește abatere de la medie.*

$$\sum (X_i - m) = 0$$



INDICATORI AI ÎMPRĂSTIERII – ABATEREA MEDIE (ABATEREA DE LA MEDIE)

X	$\underline{X_i} - m$
55	-13
60	-8
65	-3
55	-13
57	-11
70	2
75	7
85	17
80	12
78	10
$\sum X = 680$	$\sum(\underline{X_i} - m) = 0$
$m = 68$	



INDICATORI AI ÎMPRĂSTIERII – DISPERSIA

- ***dispersie (varianță; lb. engleză, variance).***
- ***Indică variabilitatea valorilor din distribuție***
- ***este notată cu simbolul σ^2 la nivelul populației***
- ***La nivel de eșantion dispersia este simbolizată prin s^2***

$$s^2 = \frac{\sum (X_i - m)^2}{N - 1}$$

(formula 4.5)



INDICATORI AI ÎMPRASTIERII – DISPERSIA

X	$\underline{X_i} - m$	$(\underline{X_i} - m)^2$
55	-13	169
60	-8	64
65	-3	9
55	-13	169
57	-11	121
70	2	4
75	7	49
85	17	289
80	12	144
78	10	100
$\sum X = 680$	$\sum (\underline{X_i} - m) = 0$	$\sum (\underline{X_i} - m)^2 = 1118$
$m = 68$		

$$s^2 = \frac{\sum (X_i - m)^2}{N - 1} \rightarrow s^2 = \frac{1118}{9} \rightarrow s^2 = 124,22.$$



INDICATORI AI ÎMPRĂSTIERII — DISPERSIA

- Funcția **var()** este utilizată pentru a calcula dispersia (varianța).
- `var(bdcap4$Stres)`, unde:
- `var` — este funcția din R care calculează dispersia;
- `bdcap4` — reprezintă numele bazei de date;
- `Stres` — este numele variabilei analizate.

```
var(bdcap4$Stres)
```

```
[1] 29.74847
```



INDICATORI AI ÎMPRĂSTIERII – ABATEREA STANDARD

- *abaterea standard (lb. engleză, standard deviation).*
- *Aceasta este notată cu simbolul σ (la nivelul populației) și cu s (la nivelul eșantionului).*
- *Din această notație putem intui că abaterea standard se obține prin extragerea radicalului din dispersie.*

$$s = \sqrt{\frac{\sum (X_i - m)^2}{N - 1}}$$

(formula 4.6)

$$s = \sqrt{124,22} \rightarrow s = 11,14.$$



INDICATORI AI ÎMPRĂSTIERII – ABATEREA STANDARD

- ***Atunci când dorim să calculăm abaterea standard a unei distribuții de valori trebuie să parcurgem următorii pași:***
 - 1. Calculăm media aritmetică a valorilor din distribuție.***
 - 2. Calculăm abaterea de la medie a fiecărei valori ($X-m$).***
 - 3. Ridicăm la pătrat fiecare abatere de la medie.***
 - 4. Adunăm rezultatele obținute în Pasul 3.***
 - 5. Împărțim suma pătratelor obținută în Pasul 4 la numărul total de valori minus 1 ($N-1$). Rezultatul obținut în această etapă reprezintă dispersia (varianța).***
 - 6. Extragem radical din rezultatul obținut la Pasul 5 și aflăm abaterea standard.***



INDICATORI AI ÎMPRĂSTIERII – ABATEREA STANDARD

- ***Proprietățile abaterii standard:***

- 1. Abaterea standard este ≥ 0 ; poate fi egală cu 0 doar atunci când valorile din distribuție sunt identice.***
- 2. Dacă se scade/adună o constantă la fiecare valoare a distribuției, abaterea standard rămâne neschimbată. Această proprietate este explicată prin faptul că adăugarea/scăderea unei constante la fiecare scor crește/scade media cu acea constantă. În concluzie, abaterea față de medie rămâne neschimbată.***
- 3. Dacă fiecare valoare este multiplicată/împărțită cu o constantă, abaterea standard se multiplică/se divide cu acea valoare.***
- 4. Abaterea standard față de medie este mai mică decât abaterea standard față de orice altă valoare din distribuție.***



INDICATORI AI ÎMPRĂSTIERII – ABATEREA STANDARD

- Pentru a calcula abaterea standard vom folosi funcția **sd()**.
- **sd(bdcap4\$ED)**, unde:
- **sd** – este funcția din R care calculează abaterea standard;
- **bdcap4** – reprezintă numele bazei de date;
- **ED** – este numele variabilei analizate.

```
sd(bdcap4$ED)
```

```
5.131521
```



INDICATORI AI FORMEI DISTRIBUTIEI - SIMETRIA

- *O distribuție este simetrică atunci când valorile ei se împart în mod egal de o parte și de a alta a mediei.*
- *Distribuțiile în care cele mai multe valori se află de o parte sau alta a mediei sunt asimetrice (lb. engleză, skewed).*
- *Atunci când cele mai multe valori se află în partea stângă a mediei (în zona scorurilor mici) distribuția este pozitiv asimetrică.*
- *Dacă cele mai multe scoruri sunt în partea dreaptă a mediei (în zona scorurilor mari) distribuția este negativ asimetrică. Indicatorul numeric al simetriei se numește skewness.*
- *În cazul unei distribuții perfect simetrice, skewness este egal cu 0.*



INDICATORI AI FORMEI DISTRIBUTIEI - SIMETRIA

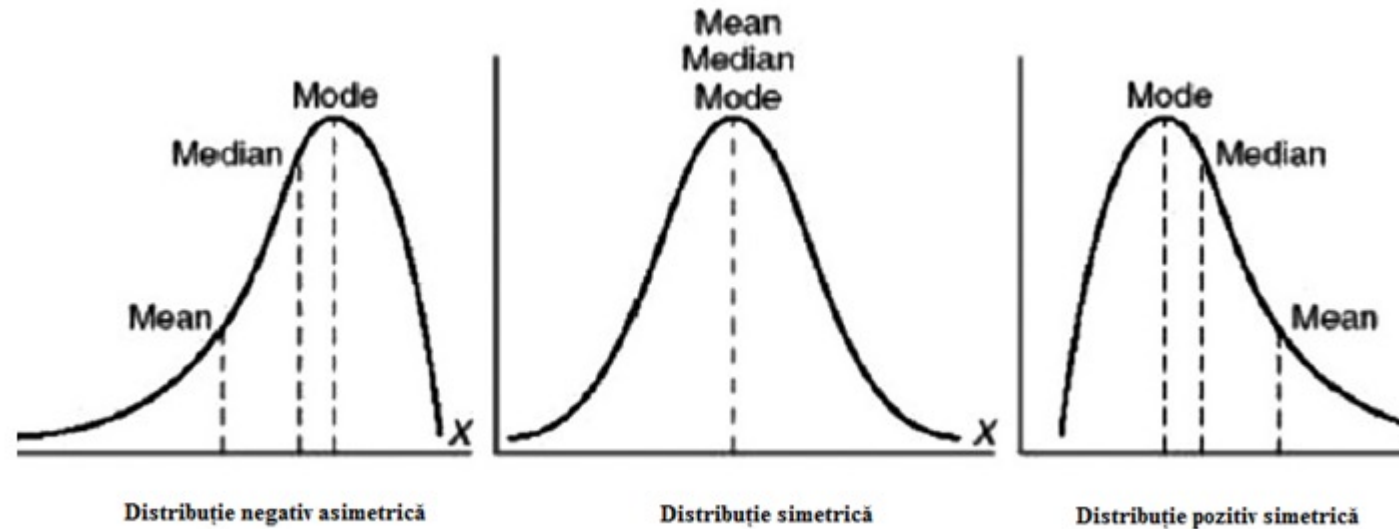
- în R vom folosi funcția **skew()**.
- `skew(bdcap4$ACE)`, unde:
- `skew` — este funcția din R care calculează skewness;
- `bdcap4` — reprezintă numele bazei de date;
- `ACE` — este numele variabilei analizate.

```
skew(bdcap4$ACE)
```

```
[1] 1.419507
```



INDICATORI AI FORMEI DISTRIBUTIEI - SIMETRIA



INDICATORI AI FORMEI DISTRIBUTIEI - APLATIZAREA

- Indicele de applatizare (kurtosis) ne oferă informații despre applatizarea/boltirea distribuției.
- Atunci când acesta are o valoare pozitivă distribuția are o formă înaltă, iar valorile tind să se apropie de medie (distribuție leptokurtică).
- Valorile negative ale lui kurtosis se traduc într-o distribuție cu o formă applatizată, în care valorile au tendința de a se îndrepta spre extreme (distribuție platikurtică) ([deCarlo, 1997](#)).
- O distribuție mezokurtică are o boltire și o îndepărtare a extremelor față de medie moderate, indicele kurtosis fiind egal cu 0.



INDICATORI AI FORMEI DISTRIBUTIEI - APLATIZAREA

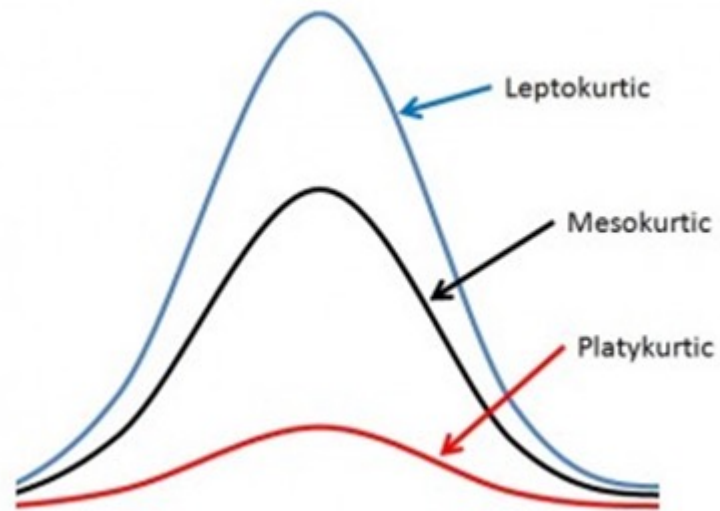
- funcția din R care ne ajută în acest sens este `kurtosi()`
- `kurtosi(bdcap4$Alexitimie)`, unde:
- `kurtosi` — este funcția din R care calculează kurtosis;
- `bdcap4` — reprezintă numele bazei de date;
- `Alexitimie` — este numele variabilei analizate.

```
kurtosi(bdcap4$Alexitimie)
```

```
[1] 0.1817175
```



INDICATORI AI FORMEI DISTRIBUTIEI - APLATIZAREA



INDICATORI SINTETICI - R

- funcția `describe()` afișează toți indicatorii statistici importanți.
- `describe(bdcap4$ACE)`, unde:
- `describe` — este funcția din R care calculează indicatorii sintetici;
- `bdcap4` — reprezintă numele bazei de date;
- `ACE` — este numele variabilei analizate.

```
describe(bdcap4$ACE)
```

```
vars   n  mean   sd median trimmed  mad min max range skew kurtosis   se
1  525 55.46 8.02    54   54.65  4.45  25  96   71 1.42    4.95 0.35
```



INDICATORI SINTETICI - R

- dacă dorim să obținem indicatorii statistici pentru variabilele *Stres* și *ED* putem folosi linia de cod de mai jos:
- `describe(bdcap4[c("Stres", "ED")])`

```
describe(bdcap4[c("Stres", "ED")])
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Stres	1	525	10.12	5.45	10	10.11	5.93	0	21	21	0.03	-0.73	0.24
ED*	2	525	157.28	101.06	164	157.38	130.47	1	330	329	-0.08	-1.23	4.41



INDICATORI SINTETICI - R

- **describe(bdcap4)** afișează indicatorii statistici pentru toate variabilele din baza de date.

```
describe(bdcap4)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Genul	1	525	1.69	0.46	2	1.74	0.00	1	2	1	-0.84	-1.30	0.02
ACE	2	525	55.46	8.02	54	54.65	4.45	25	96	71	1.42	4.95	0.35
Alexitimie	3	525	45.63	12.62	44	45.04	13.34	20	94	74	0.48	0.18	0.55
Stres	4	525	10.12	5.45	10	10.11	5.93	0	21	21	0.03	-0.73	0.24
ED*	5	525	157.28	101.06	164	157.38	130.47	1	330	329	-0.08	-1.23	4.41



INDICATORI SINTETICI - R

- **describeBy(bdcap4, bdcap4\$Genul)** afișează indicatorii statistici pentru toate variabilele din baza de date separat în funcție de genul participanților.

```
describeBy(bdcap4, bdcap4$Genul)
```

```
group: 1
      vars  n  mean    sd median trimmed  mad min max range skew kurtosis  se
Genul    1 161  1.00  0.00     1    1.00   0.00  1  1    0  NaN      NaN 0.00
ACE      2 161 53.57  7.09    53   53.19   4.45 29 85   56  0.97    4.73 0.56
Alexitimie 3 161 46.55 12.84    46   45.87  13.34 24 94   70  0.67    0.88 1.01
Stres    4 161  8.39  5.09     9    8.22   4.45  0 21   21  0.25   -0.55 0.40
ED*      5 161 56.07 37.57    56   55.40  47.44  1 123  122  0.06   -1.20 2.96
-----
group: 2
      vars  n  mean    sd median trimmed  mad min max range skew kurtosis  se
Genul    1 364  2.00  0.00     2    2.00   0.00  2  2    0  NaN      NaN 0.00
ACE      2 364 56.30  8.27    54.0   55.35   4.45 25 96   71  1.52    4.76 0.43
Alexitimie 3 364 45.23 12.53    44.0   44.67  13.34 20 80   60  0.39   -0.22 0.66
Stres    4 364 10.89  5.44    11.0   10.97   5.93  0 21   21 -0.09   -0.71 0.29
ED*      5 364 125.96 79.63   128.5  126.20 103.04  1 261  260 -0.07   -1.27 4.17
```

