



UNIVERSITATEA DE MEDICINĂ,
FARMACIE, ȘTIINȚE ȘI TEHNOLOGIE
„GEORGE EMIL PALADE”
DIN TÂRGU MUREȘ

PROBABILITĂȚI ȘI STATISTICĂ ÎN SISTEME MEDICALE

Cursul 3, 30.09.2020

STATISTICA DESCRIPTIVĂ - continuare

prof. univ. dr. habil Manuela Rozalia GABOR

2.3. INDICATORI AI DISPERSIEI (ÎMPRĂȘTIERII) DATELOR ÎN JURUL VALORII MEDII

Varianța notată s_x^2 este un indicator de împrăștiere a datelor. Formula de calcul este:
$$S_x^2 = \frac{\sum x_i^2 - (\bar{x})^2}{n - 1} \quad (III.4),$$

Abaterea standard sau deviația standard reprezintă rădăcina pătrată din varianță (dispersie) :
$$S_x = \pm \sqrt{S_x^2} \quad (III.5),$$

Coeficientul de variație se calculează ca un raport procentual între abaterea standard și valoarea medie a șirului de valori.

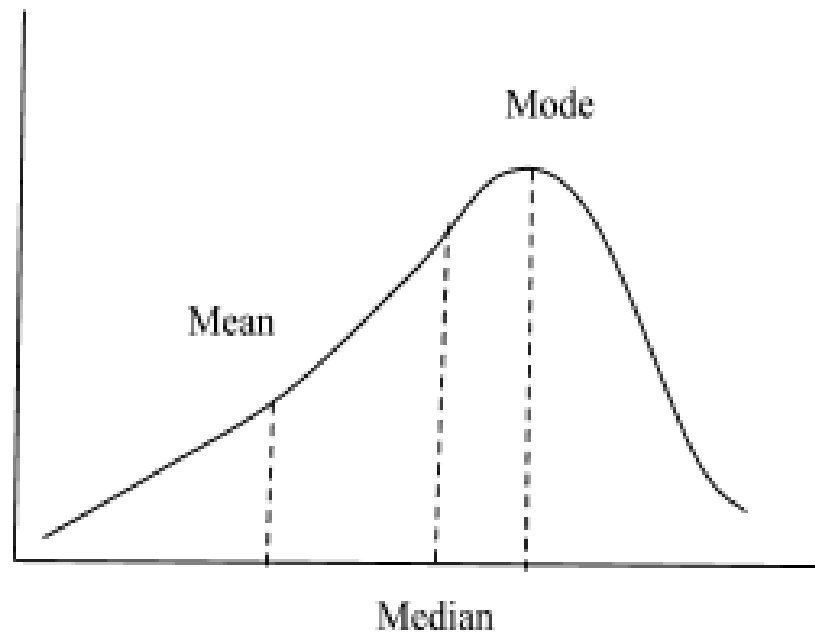
$$CV\% = \frac{S_x}{\bar{x}} \cdot 100 \quad (III.6),$$

Asimetria (skewness) caracterizează gradul de “asimetrie” a unei distribuții în jurul valorii medii, comparativ cu distribuția normală. Valori pozitive ale asimetriei indică o distribuție de frecvență ce prezintă o “coadă” (în engleză tail) asimetrică în zona valorilor “pozitive” ale distribuției (valori mai mari decât media). Similar, valori negative ale asimetriei indică o distribuție de frecvență ce prezintă o “coadă” (în engleză tail) asimetrică în zona valorilor “negative” ale distribuției (valori mai mici decât media).

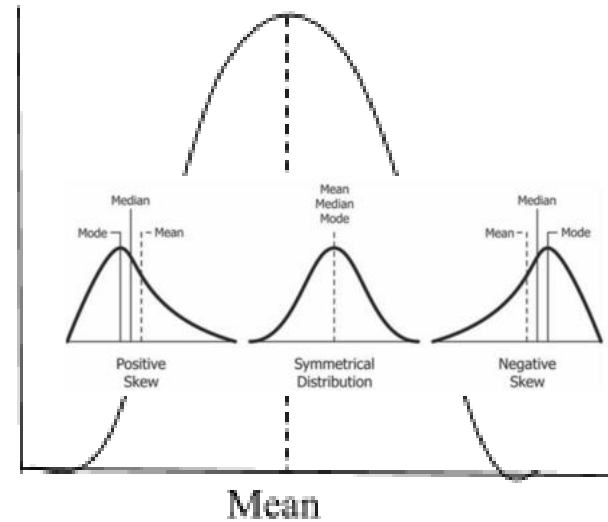
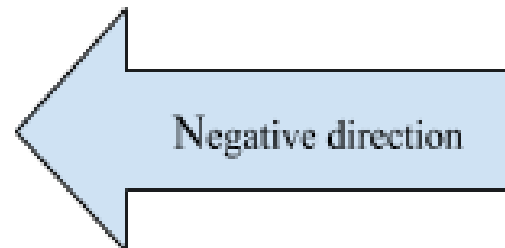
$$Skewness = \frac{n}{(n-1) \cdot (n-2)} \cdot \sum \left(\frac{x_i - \bar{x}}{S} \right)^3 \quad (III.7), \quad \text{unde } S = \text{abaterea}$$

standard.

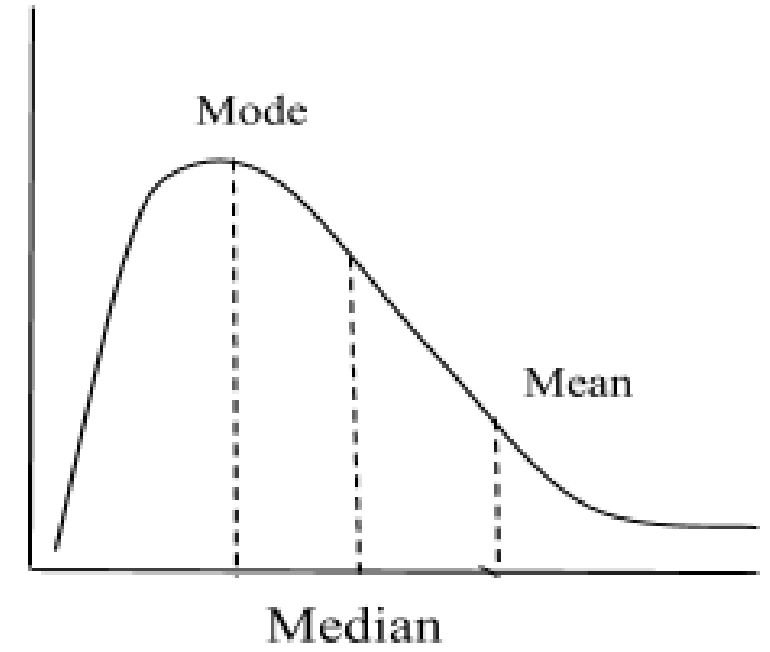
EXAMPLE



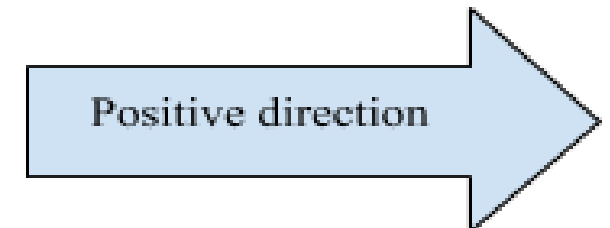
$\text{mean} < \text{median} < \text{mode}$



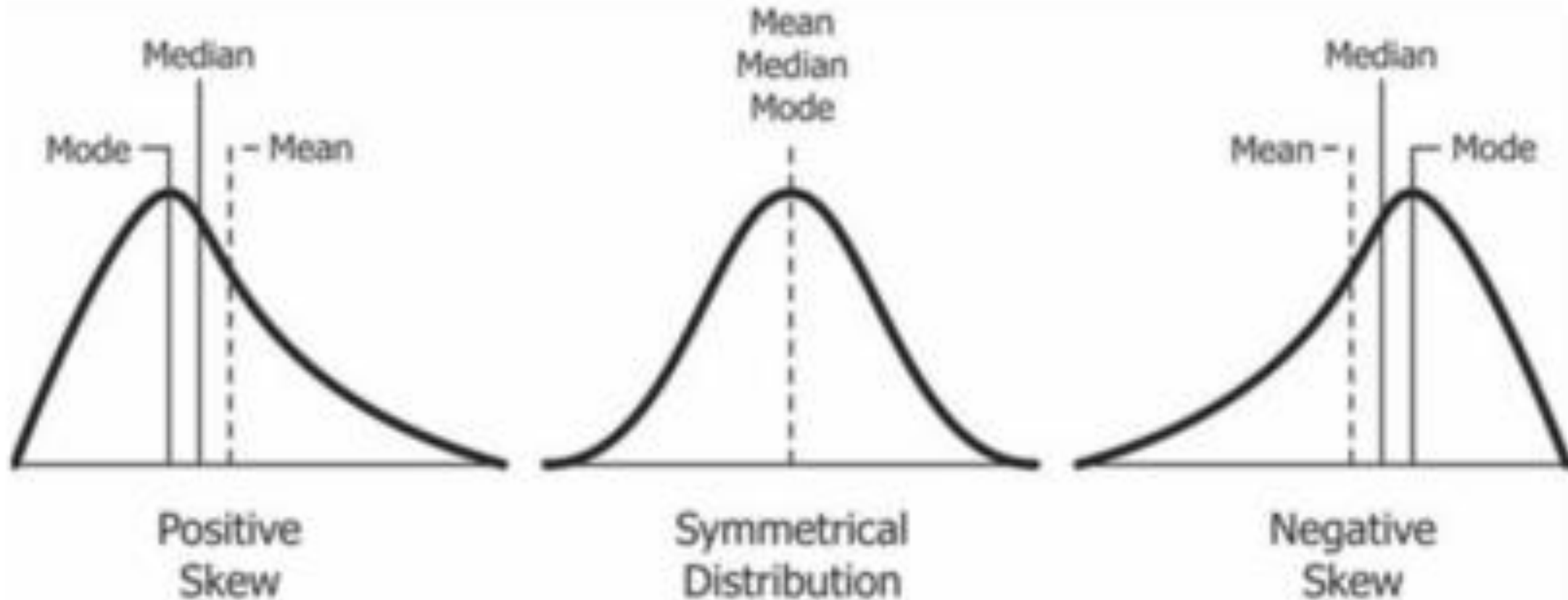
Symmetrical data
 $\text{mean} = \text{median} = \text{mode}$



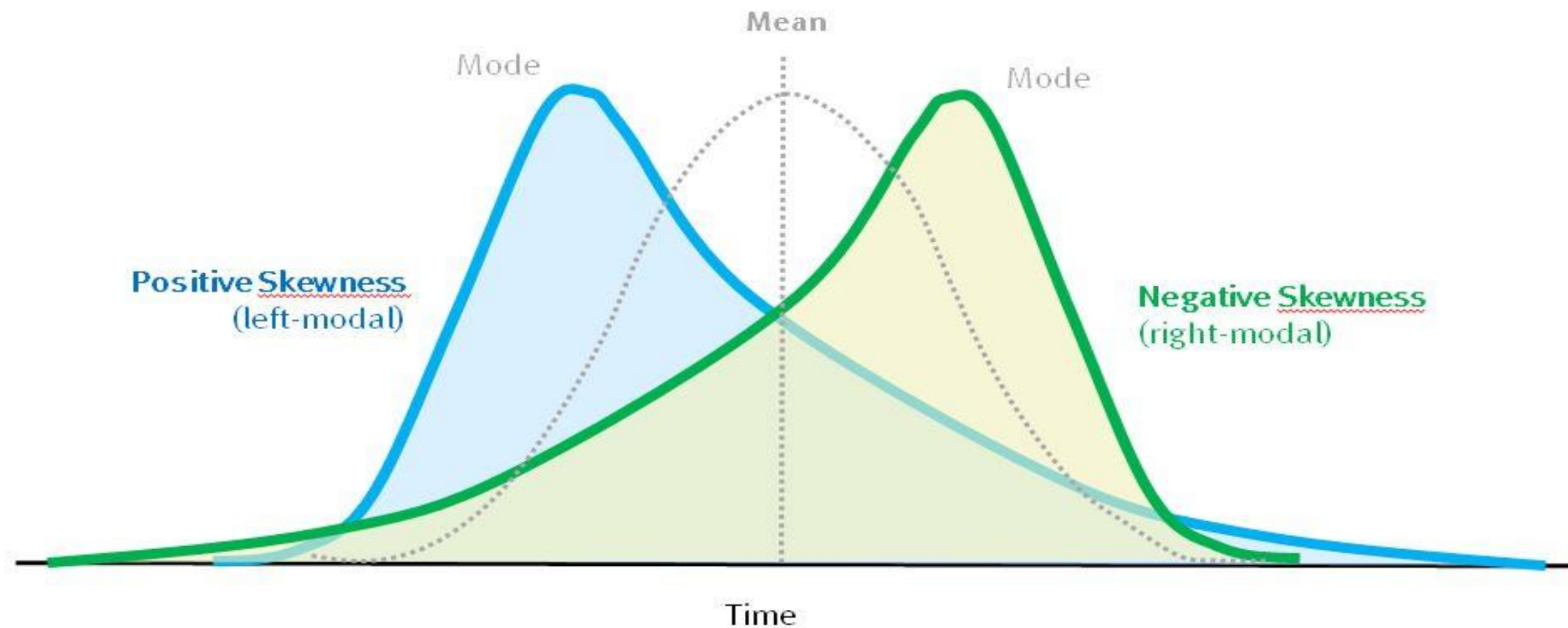
$\text{mode} < \text{median} < \text{mean}$



EXAMPLE

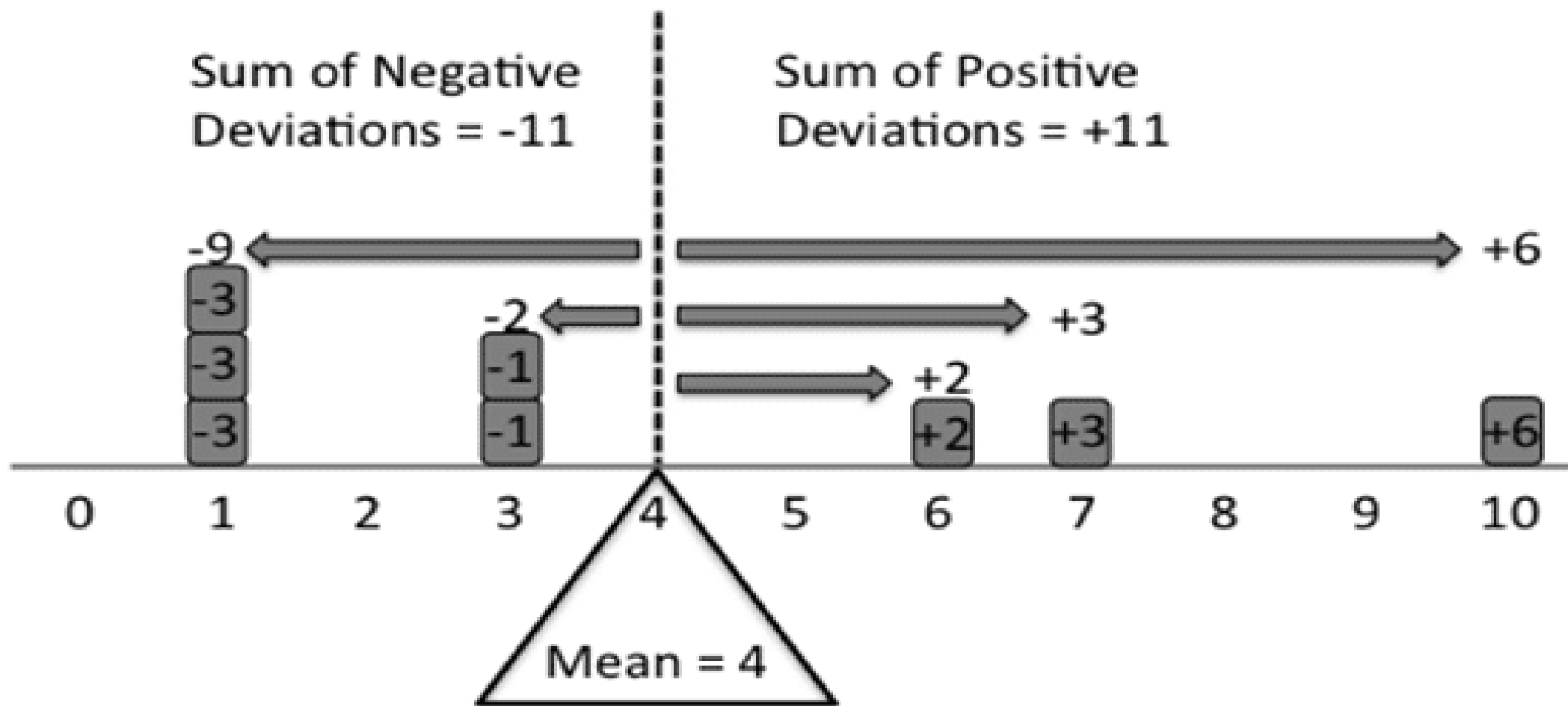


EXAMPLE



Sum of Negative
Deviations = -11

Sum of Positive
Deviations = +11



Aplatizarea (kurtosis) caracterizează gradul de “aplatizare” a unei distribuții, comparativ cu distribuția normală. Valorile pozitive ale acestui indicator indică o distribuție cu un “vârf” mai înalt decât distribuția normală. Similar, valori negative ale kurtosisului indică o curbă relativ aplatizată, comparativ cu distribuția normală.

$$Kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n+3)} \cdot \sum \left(\frac{x_i - \bar{x}}{S} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

(III.8), unde S = abaterea standard.

Amplitudinea - este diferența dintre valoarea maximă și cea minimă

$$A = A_{max} - A_{min}$$

Amplitudinea relativă - notată $A\%$ este raportul dintre amplitudinea absolută și media aritmetică a seriei de date. Atunci când avem foarte multe date se recomandă includerea lor în clase egale ca mărime, ceea ce ușurează mult prelucrările statistice ulterioare. Spre exemplu sortăm pacienții pe grupe de vârstă: 21-24 de ani, 25-30 ani, etc... În acest caz apare noțiunea de **frecvență a clasei**.

Indicatori statistici pentru serii de date cu apariții frecvente ale aceleiași valori

Dacă datele pe care le studiem conțin valori care se repetă des, se obișnuiește să se grupeze datele care au aceeași valoare . Numărul de apariții ale unei valori anume se numește frecvența de apariție și se notează cu f_i . Presupunem că în urma măsurărilor am obținut șirul de valori: x_1 cu frecvența f_1 , x_2 cu frecvența f_2 , ... x_n cu frecvența f_n

Indicatorii statistici se calculează conform noilor formule:

Media aritmetică

$$\bar{x} = \frac{\sum_{i=1,n} x_i \cdot f_i}{\sum_{i=1,n} f_i} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} \quad (III.10)$$

Mediana – este x_k (III.11,) unde $k = \frac{\sum_{i=1,n} f_i + 1}{2}$

Dispersia (varianța) : $s_x^2 = \frac{\sum_{i=1,n} (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1,n} f_i - 1} \quad (III.12)$

■
Calcul online medie
aritmetică

[https://www.calculat.org/ro/
medie/medie-
aritmetica.html](https://www.calculat.org/ro/medie/medie-aritmetica.html)

Calcul online medie
aritmetică ponderată

[https://www.calculat.org/ro/
medie/medie-
ponderata.html](https://www.calculat.org/ro/medie/medie-ponderata.html)

Calculator statistici

[https://www.calculator.net/s
tatistics-calculator.html](https://www.calculator.net/statistics-calculator.html)

[https://www.calculator.net/s
tandard-deviation-
calculator.html](https://www.calculator.net/standard-deviation-calculator.html)

[https://www.socscistatistics
.com/descriptive/](https://www.socscistatistics.com/descriptive/)

EROARE STANDARD

Este cunoscut faptul că determinările biologice sunt supuse influenței a două tipuri de erori: cele care influențează precizia determinării și cele care influențează exactitatea determinării. Pentru a afla exactitatea cu care s-au făcut o serie de determinări, trebuie să se calculeze abaterea medie a valorilor medii obținute sau, altfel spus, media erorilor ce se pot comite într-o determinare. Această abatere a fost denumită eroare standard, notată cu E. Calcularea ei se face cu ajutorul formulei:

$$E^2 = \frac{\sum (x_1 - \bar{x})^2}{n(n-1)}$$

Știind că, în cazul distribuției normale gaussiene, împrăștierea în jurul mediei colectivității a unei medii de șantion este n ori mai mică decât împrăștierea rezultatelor individuale, eroarea standard este dată și de formula:

$$E = \frac{S}{\sqrt{n}}$$

Ea reprezintă formula clasică a erorii standard. Rezultatele experimentărilor biologice trebuie să fie însoțite întotdeauna de eroarea standard sau de abaterea standard, utilizându-se formulări de tipul $M \pm S$ sau $M \pm E$, pentru a permite o justă interpretare a lor.

COEFICIENT DE VARIAȚIE

Coeficientul de variație se calculează ca un raport procentual între abaterea standard și valoarea medie a șirului de valori.

$$CV\% = \frac{S_x}{\bar{x}} \cdot 100$$

De remarcat că valoarea coeficientului de variație nu are unitate de măsură, se exprimă procentual. Acest fapt permite folosirea indicatorului la compararea a două sau mai multe serii de date, indiferent de ordinul de mărime al variabilelor (variantelor) și de unitățile de măsură folosite. Se poate considera că un coeficient de variație sub 10% indică o dispersie mică (o împrăștiere) a datelor, adică seria este omogenă. Un coeficient între 10% și 30% indică dispersie mijlocie, iar peste 30% indică dispersie mare. *Dacă dispersia este mare, media nu este un indicator reprezentativ.*