



STATISTICĂ APLICATĂ ÎN ȘTIINȚA SPORTULUI ȘI EDUCAȚIEI FIZICE

Sorina CERNĂIANU

Facultatea de Educație Fizică și Sport, Craiova

s_cernaianu@yahoo.com

Cuprins

1. Definiția statisticii. Termeni uzuali.
2. Tipuri de date.
3. Noțiunea de variabilă
4. Scale de măsurare
5. Tipuri de analiza a datelor

1. Definiția statisticii

- recoltarea, descrierea și analiza datelor, în vederea extragerii unor concluzii pe baza acestora.

Statistica se preocupă de trei mari aspecte:

- 1) modul în care se face selecția observațiilor care urmează să fie analizate;**
- 2) descrierea datelor;**
- 3) explicarea și/sau predicția unor fenomene studiate.**

Dată - valoarea unei variabile în cazul unui element al populației sau eșantionului.

Ex: vârstă – **25 ani**

Set de date - valorile unei variabile în cazul unei populații sau a unui eșantion

Statistica descriptivă:

- descrierea sintetică a informației cuprinse într-un set de date;
- colectarea, clasificarea și prezentarea datelor numerice.

Statistica inferențială:

- generalizări despre caracteristicile unei populații, pe baza informațiilor culese de la un eșantion;
- interpretarea datelor oferite de statistica descriptivă și folosirea acestora pentru a formula concluzii și a lua decizii.

2. Tipuri de date

- a) **date primare** – sunt construite de către cercetător cu un scop bine precizat (ex. anchetele cantitative sau calitative);
- b) **date secundare** – date furnizate de către organisme oficiale naționale sau internaționale (ministere, federatii, organizații, institute, etc.)

Exemplu: indici din diferite sectoare de activitate, precum educație, sănătate, turism, sport etc.

3. Noțiunea de variabilă

Variabilitate - proprietatea unui obiect de studiu de a lua valori diferite

Variabilă – criteriu prin care elementele (obiectele, indivizii) sunt clasate în diverse categorii;

- caracteristică ce prezintă interes în cazul fiecărui element (individ) al unei populații.

Variabile latente și variabile observate

Variabile latente - care nu pot fi direct măsurate, ci pot fi aflate prin intermediul unor indicatori

Ex.:

- impactul marilor evenimente sportive asupra dezvoltării destinațiilor turistice – confort, atracții turistice, ambianță, intenția de a revizita locația etc.;
- impactul antrenamentului sportiv asupra rezultatelor în competiții, etc.).

Variabile latente și variabile observate

Variabile observate - pot fi observate și măsurate direct

Exemple: masa corporală, anvergura, timpul realizat, lungimea săriturii, etc.

Variable continue și variabile discrete

Variable continue – rezultatele numerice pot să ia orice valoare (eventual între limite inferioare și superioare).

Exemple: înălțimea, masa corporală, lungimea, etc.

Dacă măsurătorile se efectuează cu precizie mare, subiecții înregistrează valori diferite.

Variable continue și variabile discrete

Variable discrete – iau un număr finit de valori.

Exemple:

1. Numărul de copii dintr-o familie.
2. Numărul de persoane care au practicat o ramură sportivă, etc.

Variabilele factuale

care identifica:

- a) persoana;**
- b) mediul din care aceasta face parte;**
- c) comportamentul sociologic, politic, economic, și psihologic.**

a) Variabile factuale legate de persoană (exemple)

- vârstă;
- sexul;
- nivelul de educație;
- profesia;
- venitul.

b) Variabile factuale legate de mediu (exemple)

- situația familială;
- locul de muncă;
- condițiile de muncă;
- habitatul persoanei, tipul de locuință, etc.;
- relații familiale, de muncă, de vecinătate cu alte persoane etc.

c) Variabile factuale legate de comportamente

- **comportamente economice**, legate în particular de consumul unor alimente, medicamente, echipament sportiv, etc.);
- **comportamente politice** (vot, etc.);
- **comportamente sociologice și psihologice** (loisir, sport, activități, etc.).

Opiniile

- ceea ce o persoană gândește despre un subiect, un lucru etc.;

Cunoașterea opiniilor poate să explice și să prevadă anumite comportamente.

Măsurarea opiniilor

- **direcția :**
 - opinia poate sa fie pozitivă sau negativă;
- **intensitatea :**
 - opinia poate să se exprime pe mai multe grade, poate să fie slabă sau intensă.

Exemple:

Dezacord total/ acord total

Deloc interesat/ Foarte interesat

Variabile calitative:

- gen (feminin, masculin);
- fumător (da, nu);
- practicarea unei activități fizice (da, nu);

Variabile cantitative:

- numărul de mese pe zi;
- masa corporală;
- înălțimea;
- vârsta.

Variabile independente – reprezintă un factor cauzal, un element care este determinant în explicarea unui fenomen.

Variabile dependente – sunt cele care sunt influențate, deci cele la care observăm variații;

- sunt manipulate, explicate de către alte variabile.

Example:

- *Variabila dependentă*: masa corporală, în kg, cu *variabila independentă* mediul socio-economic.
- *Variabila dependentă*: masa corporală, în kg, cu *variabila independentă* numărul de mese pe zi.
- *Variabila dependentă*: numărul de mese pe zi, cu *variabila independentă* structura familială.

4. Scale de masurare

- 1) Nominală
- 2) Ordinală
- 3) De interval
- 4) De raport

Reguli:

1. Categoriile trebuie să țină cont de toate posibilitățile.
2. O persoană trebuie clasată într-o singură categorie.
3. Clasarea se va face cu minimum de erori posibile.

Aceste reguli se aplică la toate scalele.

1) Scala nominală

- clasează indivizii în categorii date;
- este cel mai scăzut nivel de măsură accesibil;
- enumerare de posibilități; indivizii se grupează în funcție de acestea;
- fiecare categorie este echivalentă cu celelalte.

1) Scala nominală - exemple

Gen:

- Feminin
- Masculin

Status marital:

- Căsătorit(ă)
- Necăsătorit(ă)
- Divorțat(ă)
- Altele

2) Scala ordinală

- o categorie poate să fie mai mică sau mai mare decât alta;
- există o ordine în categoriile utilizate;
- proprietăți:
 - a) clasează indivizii în categorii;
 - b) **stabilește o ordine între categorii.**

2) Scala ordinală – exemple

Clasamentul în competiții:

- **Locul 1**
- **Locul 2**
- **Locul 3**

Ordinea are sens (1 e mai bun decât 2), dar diferența dintre locuri nu este neapărat egală.

2) Scala ordinală – exemple

Utilitatea unui echipament, serviciu, exercițiu etc.:

- Inutil
- Mai puțin util
- Util
- Foarte util

Nivel de studii:

- Fără studii
- Școală primară
- Gimnaziu
- Școală profesională
- Liceu
- Școala postliceală
- Facultate / Licență
- Master
- Doctorat

2) Scala ordinală - exemple

Nivelul de pregătire al sportivilor

- Începător
- Intermediar
- Avansat
- Elite

Ordinea este clară, dar nu spune cât mai bun este un sportiv față de altul.

3) Scala de interval

- posedă proprietățile scalelor nominală și ordinală;
- apar intervale egale între diferitele niveluri ale scalei de măsură.

3) Scala de interval - exemple

Înălțimea pentru selecție:

- 170-174
- 175-179
- 180-184
- 185-189
- 190-194

3) Scala de interval - exemple

Cum a fost nivelul tău de energie comparativ cu nivelul obișnuit?”

- **-2 = mai scăzut**
 - **-1 = ușor mai scăzut**
 - **0 = normal**
 - **+1 = ușor mai ridicat**
 - **+2 = mai ridicat**
- valorile sunt **echidistante**;
 - există un **zero convențional** (reprezintă nivelul obișnuit / media), nu zero absolut;
 - diferențele pot fi interpretate numeric.

4. Scala de raport

- posedă aceleași proprietăți ca scalele precedente;
- apare **zero absolut**, are sens;
- proporțiile calculate au un sens.

4. Scala de raport - exemple

- un subiect are masa corporală de 60 kg, iar un altul de 120 kg;
- al doilea subiect este cu 60 kg mai greu decât primul;
- al doilea subiect este de două ori mai greu decât primul;
- ambele afirmații au sens.

4. Scala de raport - exemple

Viteza de deplasare:

- 10 km/h, 20 km/h, 30 km/h

Zero = nu există mișcare.

Numărul de repetări:

- Sărituri în înălțime, flotări,
genuflexiuni: 0, 10, 20, 30 repetări

Zero = nu s-a realizat nicio repetare.

Tabel centralizator – scale de masurare

Scale	Proprietăți			
	Clasament	Ordine	Distanță	Zero absolut
1) Nominală	Da	Nu	Nu	Nu
2) Ordinală	Da	Da	Nu	Nu
3) De interval	Da	Da	Da	Nu
4) De raport	Da	Da	Da	Da

Indicatori statistici în funcție de tipul de scală

Scală	Indicatori statistici utilizați	Teste statistice, coeficienți, indici
Nominală	– frecvență absolută și relativă – mod	– χ^2 – coeficient de contingență – coeficient phi – coeficient Lambda – regresie logistică
Ordinală	– frecvență absolută și relativă – mod – mediană – măsuri de poziție	Cele de la scala nominală, plus: – Corelația rangurilor – alte teste non parametrice – regresie logistică

Indicatori statistici în funcție de tipul de scală

Scală	Indicatori statistici utilizați	Teste statistice, coeficienți, indici
Interval	Cei utilizați la scalele precedente, plus: - Măsura tendinței centrale și dispersiei	Cele utilizate la scalele precedente, plus: – Analiza de varianță – Corelția Pearson – Regresia simplă și multiplă
Raport	Toți	Toate

Aceeași variabilă poate fi măsurată pe scări diferite (dar nu este totdeauna posibil pentru toate variabilele).

Exemplu:

- variabila «Gen – (M/F)» va fi totdeauna măsurată pe scară nominală.

Exemplu: Practicați exerciții fizice în timpul liber?

Scară nominală :

- 1. Da
- 2. Nu

Scară ordinală:

- 1. Niciodată
- 2. Foarte rar
- 3. Rar
- 4. Des
- 5. Foarte des

5. Tipuri de analiză a datelor

- **Analiză univariată** – studiul frecvențelor de apariție, măsurarea tendinței centrale, dispersiei, indicatori de poziție;
 - variabilele se analizează una câte una;
- **Analiză bivariată** – tabele de contingență, analiză de varianță, corelație, regresie simplă;
 - se studiază relația între două variabile;

5. Tipuri de analiză a datelor

- **Analiză multivariată** – analiza de corespondență, analiza în componente principale, analiza discriminantă, regresia multiplă;
 - se studiază relația dintre trei sau mai multe variabile.



EȘANTIONAREA

Sorina CERNĂIANU

Facultatea de Educație Fizică și Sport, Craiova

s_cernaianu@yahoo.com

Definitii

Populația statistică (N) - mulțimea tuturor obiectelor sau indivizilor care prezintă interes pentru studiu.

Eșantion (n) - un subset sau o submulțime a populației analizate, extras pe baza unor criterii stabilite anterior.

Exemplu:

- Totalitatea elevilor din liceele cu program sportiv din Romania, din care extragem un esantion

Tipuri de esantionare

A. Probabilista - fiecare obiect sau individ din populatie are sansa de a face parte din esantion

B. Neprobabilista – exista obiecte sau indivizi care nu au nicio sansa de a face parte din esantion

A. Esantionarea probabilista

1. Simpla aleatoare
2. Stratificata
3. Multistadiala
4. Cluster

A. Esantionarea probabilista

Criterii de bază pentru esantioane:

- 1) trebuie să fie reprezentative pentru populația aleasă studiată;
- 2) trebuie alese aleatoriu; fiecare membru al populației are șanse egale de a fi ales;
- 3) trebuie să fie suficient de mari pentru a nu denatura rezultatele.

1. Esantionarea simpla aleatoare

[Home](#) [Games](#) [Numbers](#) [Lists & More](#) [Drawings](#) [Web Tools](#) [Statistics](#) [Testimonials](#) [Learn More](#) [Login](#)

RANDOM.ORG

Search [RANDOM.ORG](#)

True Random Number Service

Random Integer Generator

This form allows you to generate random integers. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs.

Part 1: The Integers

Generate random integers (maximum 10,000).

Each integer should have a value between and (both inclusive; limits $\pm 1,000,000,000$).

Format in column(s).

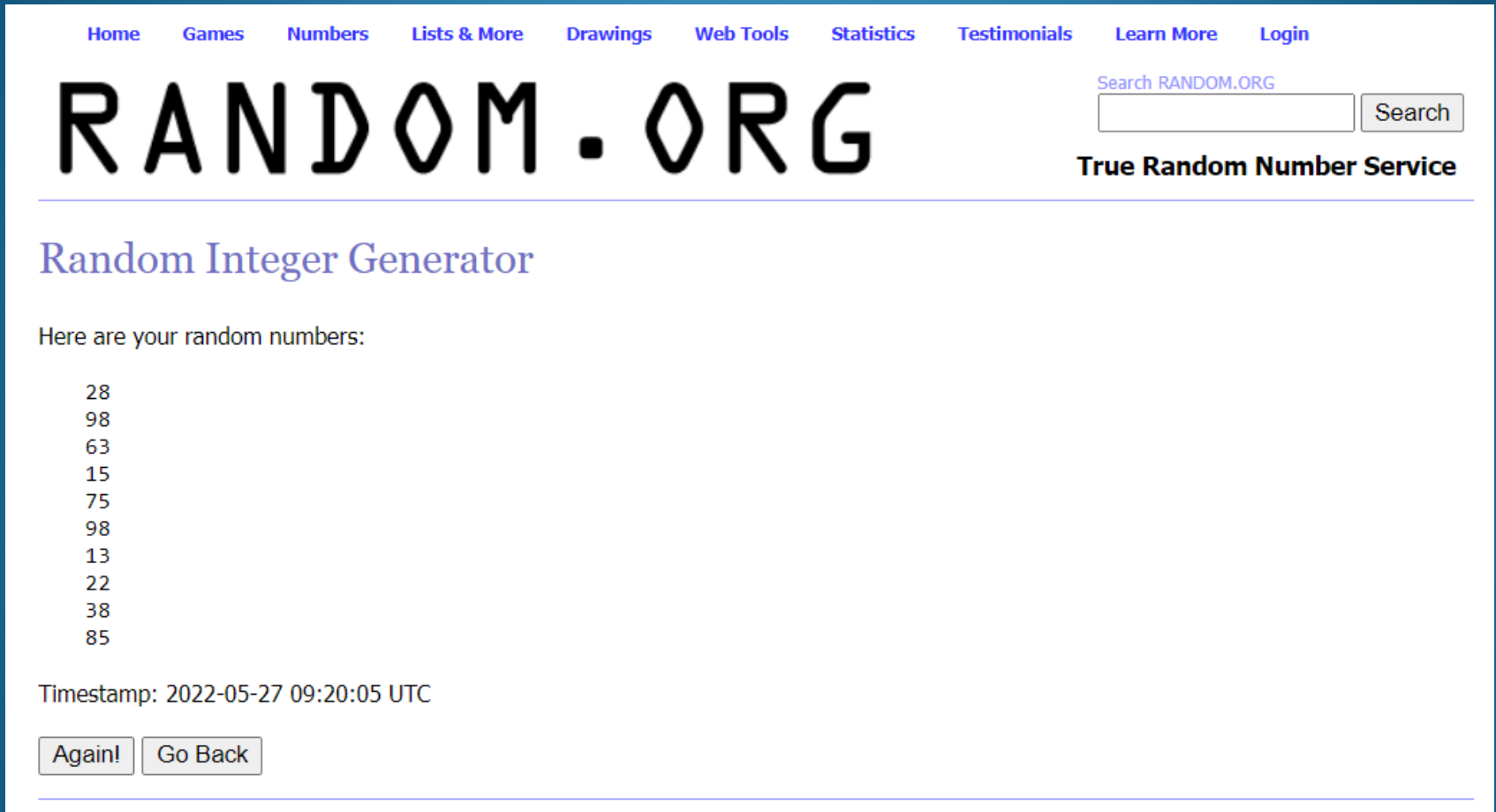
Part 2: Go!

Be patient! It may take a little while to generate your numbers...

Need more numbers than this form supports? Check out our [File Generation Service](#).

Note: The numbers generated with this form will be picked independently of each other (like rolls of a die) and may therefore contain duplicates. There is also the [Sequence Generator](#), which generates randomized sequences (like raffle tickets drawn from a hat) and where each number can only occur once.

1. Esantionarea simpla aleatoare



The screenshot shows the RANDOM.ORG website. At the top, there is a navigation bar with links: Home, Games, Numbers, Lists & More, Drawings, Web Tools, Statistics, Testimonials, Learn More, and Login. Below the navigation bar is the site's logo, "RANDOM.ORG", in large, bold, black letters. To the right of the logo is a search bar with the placeholder text "Search RANDOM.ORG" and a "Search" button. Below the search bar is the text "True Random Number Service".

Below the logo, the page title "Random Integer Generator" is displayed in a purple font. Underneath, the text "Here are your random numbers:" is followed by a list of ten random integers: 28, 98, 63, 15, 75, 98, 13, 22, 38, and 85. Below the list, the timestamp "Timestamp: 2022-05-27 09:20:05 UTC" is shown. At the bottom, there are two buttons: "Again!" and "Go Back".

Home Games Numbers Lists & More Drawings Web Tools Statistics Testimonials Learn More Login

RANDOM.ORG

Search RANDOM.ORG Search

True Random Number Service

Random Integer Generator

Here are your random numbers:

28
98
63
15
75
98
13
22
38
85

Timestamp: 2022-05-27 09:20:05 UTC

- sau in Excel

Calculul reprezentativității eșantionului

1. **Nivelul de încredere (N.I.)** – cât mai mare

Ex: 95%, 99%

2. **Eroarea standard (ES)** = $\sigma / \sqrt{n}$ pentru populație
= $S / \sqrt{n}$ pentru eșantion

3. **Marja de eroare** = $z \times ES = z \times \sigma / \sqrt{n}$

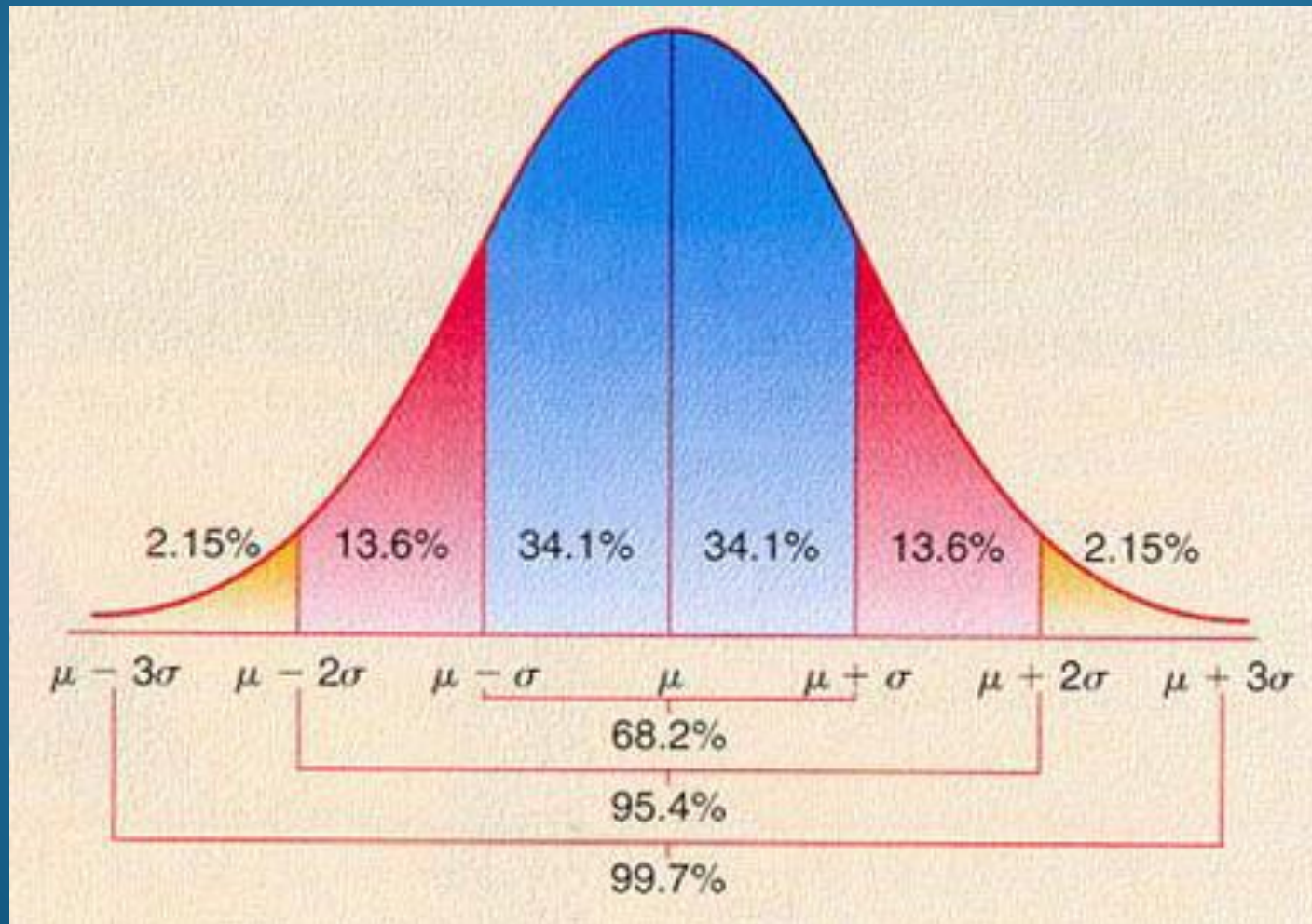
z – distanța în abateri standard față de medie

N.I. = 95%, $z = 1.96$

N.I. = 99%, $z = 2.58$

Ex. Marja = $1.96 \times 0.5 / \sqrt{1250}$
= 0.028
(± 2,8%)

Curba lui Gauss



2. Esantionarea stratificata

- Populatia nu e omogena
- Populatia se imparte in straturi (grupuri) omogene in interior si eterogene in exterior
- Definim straturile dupa anumite criterii: ex. varsta, sex, educatie, stare civila etc;

2. Esantionarea stratificata

- Se reprezinta straturile in esantion proportional cu reprezentarea lor in populatie (ex. fete-baieti)
- Se poate realiza o esantionare multipla – 2 sau mai multe criterii
- De obicei, reprezentativitatea este superioara comparativ cu esantionarea simpla aleatoare

3. Esantionarea multistadiala

- Organizare pe verticala a populatiei

Exemplu:

Universitati



Facultati



Ani de studii



Grupe



Studenti

Limite – nr. mic la un anumit nivel → straturi foarte mici
- se combina esantionarea multistadiala cu cea stratificata

Ex. universitati cu traditie – universitati mai noi

4. Esantionarea cluster (grupare)

- Este la fel ca esantionarea stratificata multistadiala, dar la ultimul nivel sunt intervievati toti subiectii (tot cluster-ul)
- Cand ultimul nivel exista in mod natural in populatie (ex studentii sau elevii)
- In loc sa esantionam studentii, esantionam grupele (care contin in medie 25 studenti) in cadrul carora intervievam toti studentii
- Cluster = grupa – se extrag aleator

Calculul marimii esantionului

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1-p)}{(N-1) \cdot e^2 + Z^2 \cdot p \cdot (1-p)}$$

n = marimea esantionului

N = populatia

Z = – distanța în abateri standard față de medie

Nivelul de incredere 90% -> Z=1,645

Nivelul de incredere 95% -> Z=1,96

Nivelul de incredere 99% -> Z=2,575

e = marja de eroare maxima tolerata (ex. 5%)

p = proportia pe care estimam sa o gasim (0.5 daca nu o cunoastem)

Calculul marimii esantionului

<https://www.calculator.net/sample-size-calculator.html>

Calculator.net

FINANCIAL

FITNESS & HEALTH

[home](#) / [math](#) / [sample size calculator](#)

Sample Size Calculator

Find Out The Sample Size

This calculator computes the minimum number of necessary samples to meet the desired statistical constraints.

Result

Sample size: **384**

This means 384 or more measurements/surveys are needed to have a confidence level of 95% that the real value is within $\pm 5\%$ of the measured/surveyed value.

Confidence Level:

Margin of Error:

Population Proportion: Use 50% if not sure

Population Size: Leave blank if unlimited population size.

Calculate



Clear

B. Esantionarea neprobabilista

1) De convenienta (de disponibilitate) – in cercetari experimentale sau exploratorii, preliminare

- ex.: cercetari studentesti, care nu au la dispozitie un buget,
- accesul la participanti este foarte dificil
- include persoane pe care le avem la dispozitie, cu care putem sa derulam cercetarea.

B. Esantionarea neprobabilista

Alte tipuri de esantionare - in cercetari calitative;

- nu ne intereseaza reprezentativitatea esantionului

Ex. ancheta pe baza de interviu:

2) Pe cote

- seamana cu cea stratificata,
- accesul la subiecti este lasat la alegerea operatorului de ancheta
- Indivizii se grupeaza pe straturi: varsta, sex, stare civila
- Cand identificam un anumit strat, indivizii nu se selecteaza aleator
- Se pot impune anumite conditii necesare – ex. un venit minim
- Nu e necesara reprezentativitatea

B. Esantionarea neprobabilista

3) Tip bulgare de zapada

Studii de caz – se specifica tehnica de esantionare

- este adesea folosită în populațiile ascunse, la care cercetătorii le este dificil accesul.
- exista si eşantionare virtuală tip bulgăre de zăpadă

4) Most similar/dissimilar cases

- Utilizata in studii comparative
- Selectia pe baza de asemanari sau diferentieri
- Folosita in studiile de caz
- Alegerea cazurilor, dupa anumite criterii, tine de obiectivele cercetarii si de specificul cazurilor

B. Esantionarea neprobabilista

5) Typical cases sampling

- Se foloseste in studii calitative: interviuri in profunzime, biografii, memoriale etc.
- Persoanele isi povestesc viata
- Se stabileste ce caracteristici are un caz tipic
- Selectia subiectilor trebuie sa urmeze criteriile stabilite

6) Critical cases sampling

- Selectarea cazurilor care fac diferenta
- De exemplu: votul la nivelul unui mare oras care poate sa prezica votul final
- Sunt cazuri care descriu mai bine decat altele ce urmeaza sa se intample intr-un anumit fenomen

INDICATORII TENDINȚEI CENTRALE

Principalii indicatori

1. media aritmetică ($\bar{x}$);
2. mediana (Me);
3. modul (Mo).

1. Media aritmetică, $(\bar{x})$

– reprezintă valoarea în jurul căreia se concentrează valorile individuale.

a) pentru seriile simple: considerăm caracteristica x , cu valorile $x_1, x_2, \dots, x_n$,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{în care :}$$

$$x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$$

n = numărul unităților;

1. Media aritmetică, $(\bar{x})$

b) pentru seriile cu distribuție de frecvențe (variantele caracteristicii se înregistrează de mai multe ori):

$$\bar{x} = \frac{\sum_{i=1}^m X_i \cdot f_i}{\sum_{i=1}^m f_i} = \sum_{i=1}^m x_i \cdot f_i^*$$

în care:

f_i = frecvența absolută;

$f_i^* = \frac{f_i}{\sum f_i}$ = frecvența relativă;

m - numărul de grupe ale caracteristicii x .

Proprietățile mediei aritmetice

1. Valoarea mediei aritmetice este cuprinsă între valorile extreme ale seriei de date;

$$X_{min} < \bar{X} < X_{max}$$

2. Pentru seriile cu distribuții de frecvențe, aceasta oscilează în jurul intervalului cu frecvența cea mai mare.

Proprietățile mediei aritmetice

3. Suma abaterilor valorilor individuale de la media aritmetică este egală cu 0.

Pentru serii simple:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Pentru serii cu distribuție de frecvențe:

$$\sum_{i=1}^m (x_i - \bar{x}) \cdot f_i = 0$$

Media aritmetică este sensibilă față de valorile extreme, astfel că devine nerepresentativă dacă termenii seriei sunt prea împrăștiați.

Calculul mediei aritmetice simple în *Excel*

Funcția AVERAGE (media aritmetică) - calculează media aritmetică a argumentelor.

Sintaxă: AVERAGE(number1;number2;...)

- number1;number2;... - argumente numerice pentru care se calculează media aritmetică.
 - Dacă într-o celulă este introdusă **valoarea zero**, atunci aceasta intra în calculul mediei.
 - Dacă celula este vidă, atunci aceasta nu intră în calcul.

EXCEL – FUNCȚII

Funcțiile în *Excel* sunt compuse din:

- 1) *nume*;
- 2) *sintaxă* - specifică modul în care trebuie introduse diferitele părți ale funcției și locul în care trebuie introduse ghilimele, paranteze sau alte semne de punctuație.
- 3) *argumente* - numere, text, referințe ale celulelor utilizate de către funcție pentru a calcula o valoare; unele argumente sunt opționale.

Exemplu:

AVERAGE(*number1;number2;...*)

Inserarea unei funcții în *Excel*

- 1) clic în celula în care se dorește să se insereze funcția;
- 2) clic pe *bara de formule*  *fx*
- 3) se selectează o categorie de funcții și funcția dorită;
- 4) clic pe butonul OK;
- 5) se selectează domeniul de celule pentru care se calculează funcția;
- 6) clic pe butonul OK.

Calculul mediei aritmetice simple în Excel

Exemple: Funcția **AVERAGE**

a. Să se calculeze media aritmetică simplă pentru notele obținute de un subiect la trei discipline sportive

D5 ▾ fx =AVERAGE(A5:C5)				
	A	B	C	D
1				
2		NOTE		
3				
4	ATLETISM	VOLEI	FOTBAL	
5	8	9	7	=AVERAGE(A5:C5)
6				
7				

Calculul mediei aritmetice simple în Excel

Exemple: Funcția AVERAGE

b. Să se determine media aritmetică simplă pentru următorul șir de date.

	A	B	C	D	E
23		Media aritmetică simplă	=AVERAGE(B25:B38)		
24				10,55	
25		18,05			
26		18,17		Valoarea 0 este introdusă în calcul.	
27		8,87		Celulele vide nu intră în calcul.	
28		8,12			
29		12,76		AVERAGE(plaja de celule)	
30		12,41			
31		0,68			
32		0,42			
33		18,85			
34		6,77			
35		0			
36					
37		18			
38		14			
39					

2. Mediana (Me)

– reprezintă valoarea centrală a unei serii ordonate.

a) Când seria are **număr impar de termeni**, valoarea din mijlocul seriei;

$$\text{rang}(Me) = \frac{n+1}{2}$$

b) Când seria are **număr par de termeni**, mediana este egală cu media aritmetică a celor două valori din mijlocul seriei.

$$Me = \frac{\frac{x_n}{2} + \frac{x_{n+1}}{2}}{2}$$

Calculul medianei în *Excel*

Funcția MEDIAN – calculează mediana unui șir de date.

Sintaxă: MEDIAN(number1,number2,...)

number1,number2,... – reprezintă numerele pentru care se calculează mediana.

Exemplu: Pentru un grup de elevi s-a măsurat lungimea săriturii de pe loc, în metri.
Să se calculeze mediana șirului de date.

165					
166	A	B	C	D	E
		Nr. crt.	Subiecți	Săritura în lungime (m)	Săritura în înălțime (m)
167					
168		1	S1	5,6	1,55
169		2	S2	5,3	1,65
170		3	S3	4,9	1,58
171		4	S4	5,8	1,65
172		5	S5	5,6	1,7
173		6	S6	5,2	1,65
174		7	S7	5,6	1,6
175		8	S8	6,3	1,55
176			MEDIAN	5,6	1,6
177				↓	
178				=MEDIAN(D168:D175)	
179					

3. Modul (Mo)

– reprezintă termenul cu frecvența cea mai mare din șirul de valori.

Calculul modului în *Excel*

Funcția MODE - întoarce valoarea care apare cel mai frecvent într-un șir de date.

Sintaxă: `MODE(number1;number2;...)`

number1;number2;... - reprezintă de la 1 până la 30 de numere pentru care se calculează valoarea modală.

Calculul modului în *Excel*

Exemplu: Opt subiecți au fost evaluați la săritura în lungime. Să se calculeze modul pentru înregistrările efectuate.

151					
152	A	B	C	D	E
		Nr. crt.	Subiecți	Săritura în lungime (m)	Săritura în înălțime (m)
153					
154		1	S1	5,6	1,55
155		2	S2	5,3	1,65
156		3	S3	4,9	1,58
157		4	S4	5,8	1,65
158		5	S5	5,6	1,7
159		6	S6	5,2	1,65
160		7	S7	5,6	1,6
161		8	S8	6,3	1,55
162		MODE		5,6	1,65
163				↓	
164				=MODE(D154:D161)	
165					

Exemplu: calculul indicatorilor tendinței centrale

Notele obținute de către cinci studenți la examenul de atletism sunt următoarele: **10, 6, 7, 10, 4.**

Pentru a analiza pe ansamblu situația celor cinci studenți se calculează cei trei indicatori.

1. **Media aritmetică** (nota medie) se determină ca raport între suma notelor obținute și numărul studenților:

$$\bar{x} = \frac{10+6+7+10+4}{5} = 7,4$$

Exemplu: calculul indicatorilor tendinței centrale

2. **Mediana** (nota mediană) este valoarea care împarte studenții în două părți egale: 50% se situează sub nota mediană, 50% se situează peste nota mediană; se determină ca valoare (notă) centrală, după aranjarea valorilor seriei în ordine crescătoare sau descrescătoare.

Aranjăm valorile în ordine crescătoare:

4, 6, 7, 10, 10

↑

Me = 7 (50% dintre studenți au luat note sub 7, iar 50% peste 7).

Exemplu: calculul indicatorilor tendinței centrale

3. **Dominanta, modul** (nota dominantă) este nota care se înregistrează la cei mai mulți studenți:

Mo = 10 (pentru ca această notă apare la un număr de doi studenți, în timp ce notele celelalte apar la un singur student).

Ca urmare s-au calculat cei trei indicatori ai tendinței centrale, care caracterizează seria statistică respectivă:

$$\bar{x} = 7,4$$

$$Me = 7$$

$$Mo = 10$$

Atunci când distribuția este absolut simetrică, cei trei indicatori ai tendinței centrale: media, mediana și modul sunt egali.

În acest caz distribuția datelor este normală.

INDICATORII VARIAȚIEI

1. Indicatorii simpli ai variației

Amplitudinea variației

a) **amplitudinea absolută (A_a)**, reprezintă diferența între valoarea maximă ($x_{\max}$) și valoarea minimă ($x_{\min}$) a caracteristicii; are aceeași unitate de măsură ca și caracteristica studiată;

$$A_a = x_{\max} - x_{\min}$$

b) **amplitudinea relativă (A_r)**, reprezintă raportul procentual între amplitudinea absolută a variației și valoarea medie a caracteristicii ($\bar{x}$); se exprimă în procente.

$$A_r = \frac{A_a}{\bar{x}} \cdot 100$$

Cu cât amplitudinea este mai mare, cu atât colectivitatea este mai eterogenă, iar cu cât aceasta este mai mică, cu atât colectivitatea este mai omogenă.

Amplitudinea depinde numai de valorile extreme ale caracteristicii.

1. Indicatorii simpli ai variației

Abaterea valorii individuale față de medie

a) **abaterea individuală în valoare absolută (d_i)**, ca diferență dintre valoarea variabilei și media aritmetică:

$$d_i = x_i - \bar{x}$$

b) **abaterea individuală relativă (d_i^r)**, ca raport procentual între abaterea individuală absolută și nivelul mediu al caracteristicii:

$$d_i^r = \frac{d_i}{\bar{x}} \cdot 100$$

Calculul amplitudinii în *Excel*

Funcții necesare:

Funcția MAX - găsește cea mai mare valoare într-o listă dată de argumente.

Sintaxă: MAX(*number1;number2;...*)

number1;number2;... - reprezintă de la 1 până la 30 de numere pentru care se calculează valoarea maximă.

Exemplu: Un subiect este evaluat la trei discipline sportive: atletism, volei și fotbal. Să se determine nota cea mai mare obținută de acesta.

D5	fx =MAX(A5:C5)			
	A	B	C	D
1				
2		NOTE		
3				
4	ATLETISM	VOLEI	FOTBAL	
5	8	9	7	=MAX(A5:C5)
6				

Calculul amplitudinii în *Excel*

Funcția MIN - găsește cea mai mică valoare într-o listă dată de argumente.

Sintaxă: MIN(*number1;number2;...*)

number1;number2;... - reprezintă de la 1 până la 30 de numere pentru care se calculează valoarea minimă.

Exemplu: Un subiect este evaluat la trei discipline sportive: atletism, volei și fotbal. Să se determine nota cea mai mică obținută de acesta.

D5	fx =MIN(A5:C5)			
	A	B	C	D
1				
2		NOTE		
3				
4	ATLETISM	VOLEI	FOTBAL	
5	8	9	7	=MIN(A5:C5)
6				

Lucrul cu formule în *Excel*

Reguli pentru scrierea unei formule:

- formulele în *Excel* încep întotdeauna cu semnul „=”;
- semnul „=” trebuie introdus totdeauna în celula în care doriți să apară rezultatul;
- semnul „=” arată că ceea ce urmează este o parte a unei formule și nu doar un nume sau un număr.

Exemplu: =7+3

Referința unei celule într-o formulă - reprezintă o combinație între litera corespunzătoare coloanei și numărul corespunzător unei linii, de exemplu A1, B5 etc.

Lucrul cu formule în *Excel*

Pași pentru scrierea unei formule:

1. Introducerea datelor
2. Adăugarea semnului (=)
3. Adăugarea referințelor celulelor

Greșeli în scrierea unei formule:

- neintroducerea semnului (=) la începutul formulei;
- inserarea unui spațiu înaintea semnului (=);
- datele sunt scrise într-un format greșit (de exemplu, ca text în loc de număr);
- selectarea greșită a datelor.

Se calculează amplitudinea absolută prin introducerea formulei în *Excel*:

$$A_a = x_{max} - x_{min}$$

= MAX - MIN

2. Indicatorii sintetici ai variației

- oferă o imagine a dispersiei valorilor caracteristicii.

a) **abaterea medie pătratică sau abaterea standard** (σ) se calculează ca media pătratică a abaterilor valorilor variabilei de la media aritmetică.

Pentru o serie simplă:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Pentru o serie cu distribuție de frecvențe:

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^m f_i}}$$

Calculul abaterii standard în Excel

Funcția STDEV - estimează abaterea standard a unui șir de valori pe baza unui eșantion.

Sintaxă: STDEV(number1;number2;...)

number1;number2;... - reprezintă de la 1 până la 30 de argumente numerice corespunzătoare unui eșantion dintr-o populație.

Exemplu: Opt sportivi au fost evaluați la săritura în lungime.

Să se calculeze abaterea standard pentru înregistrările efectuate.

195					
196	A	B	C	D	E
197		Nr. crt.	Numele și prenumele	Săritura în lungime (m)	Săritura în înălțime (m)
198		1	S1	5,6	1,9
199		2	S2	5,3	1,65
200		3	S3	4,9	1,6
201		4	S4	5,8	1,85
202		5	S5	5,6	1,7
203		6	S6	5,2	1,65
204		7	S7	5,6	1,6
205		8	S8	6,3	1,55
206			STDEV	0,42	0,12
207				↓	
208				=STDEV(D198:D205)	
209					

2. Indicatorii sintetici ai variației

b) dispersia sau varianța (σ^2) este abaterea standard ridicată la pătrat.

Pentru serii simple:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Pentru serii cu distribuție de frecvențe:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^n f_i}$$

Calculul varianței în *Excel*

Funcția VAR - estimează dispersia (varianța) unui șir de valori pe baza unui eșantion.

Sintaxă: VAR(number1;number2;...)

number1;number2;... - reprezintă de la 1 până la 30 de argumente numerice corespunzătoare unui eșantion dintr-o populație.

Exemplu: Opt sportivi au fost evaluați la săritura în lungime.

Să se calculeze dispersia pentru înregistrările efectuate.

181					
182	A	B	C	D	E
		Nr. crt.	Numele și prenumele	Săritura în lungime (m)	Săritura în înălțime (m)
183					
184		1	S1	5,6	1,9
185		2	S2	5,3	1,65
186		3	S3	4,9	1,6
187		4	S4	5,8	1,85
188		5	S5	5,6	1,7
189		6	S6	5,2	1,65
190		7	S7	5,6	1,6
191		8	S8	6,3	1,55
192			VAR	0,18	0,02
193				↓	
194				=VAR(D184:D191)	
195					

2. Indicatorii sintetici ai variației

c) **coeficientul de variație (C_v)** se calculează ca raport procentual între abaterea medie pătratică (σ) și media aritmetică ($\bar{x}$).

$$C_v = \frac{\sigma}{\bar{x}} \cdot 100 \text{ (%)}$$

Cu cât C_v are o valoare mai mică, cu atât seria este mai omogenă și deci media mai reprezentativă.

Valori:

- dacă $C_v < 10\%$, atunci eșantionul este omogen;
- dacă $10\% \leq C_v < 20\%$, atunci eșantionul este relativ omogen;
- dacă $C_v \geq 20\%$, atunci eșantionul este neomogen.

Pentru calculul coeficientului de variație în *Excel* este necesară introducerea formulei de calcul.

În prealabil se vor determina **media aritmetică** și **abaterea standard** a șirului de date.

Calculul diferenței a două valori. Progresul.

Un grup de elevi este evaluat de două ori, obținând următoarele note la gimnastică.

Să se determine, în *Excel*, progresul realizat de către fiecare elev.

Rezolvare

1. Se calculează diferența dintre testarea finală și cea inițială, introducând formula în Excel:

$$D = T_f - T_i.$$

2. Pentru a calcula progresul în procente, se introduce formula:

$$D(\%) = \frac{(T_f - T_i)}{T_i} \times 100(\%)$$

Nr. crt.	Subiecți	Note gimnastica			
		Testare inițială	Testare finală	Diferența	%
1	A.B.	7	7	0	0,00
2	A.T.	6	8	2	33,33
3	C.R.	6	9	3	50,00
4	D.S.	8	9	1	12,50
5	E.G.	6	7	1	16,67
6	E.M.	7	8	1	14,29
7	F.T.	9	10	1	11,11
8	M.S.	10	10	0	0,00
9	N.L.	7	8	1	14,29
10	Z.C.	8	9	1	12,50



STATISTICĂ APLICATĂ ÎN ȘTIINȚA SPORTULUI ȘI EDUCAȚIEI FIZICE

SORINA CERNĂIANU

FACULTATEA DE EDUCAȚIE FIZICĂ ȘI SPORT, CRAIOVA

Cuprins

1. Tipuri de teste
2. Tipuri de ipoteze
3. Normalitatea distributiei variabilelor cantitative. Teste
4. Teste t parametrice pentru comparare de medii
 - 1) Testul t pentru un esantion (One-Sample T Test)
 - 2) Testul t pentru doua esantioane independente (Independent-Samples T Test)
 - 3) Testul t pentru doua esantioane perechi (Paired-Samples T test)

1. Analiza bivariata - Tipuri de teste pentru date numerice

- **Parametrice** – presupunem ca datele provin dintr-o populație cu distribuție normală, Gaussiană (ex. Student, ANOVA);
- **Non-parametrice** – nu se tine cont de distributia normala; valorile variabilei sunt ordonate si se compara grupurile (ex. Wilcoxon, Mann-Whitney și Kruskal-Wallis)

2. Tipuri de ipoteze

- **Unidirectionale** – specifica exact directia relatiei dintre doua variabile
- **Bidirectionale** – nu specifica exact directia relatiei dintre doua variabile

Pragul de semnificatie prezentat in SPSS este bidirectional (two-tailed).

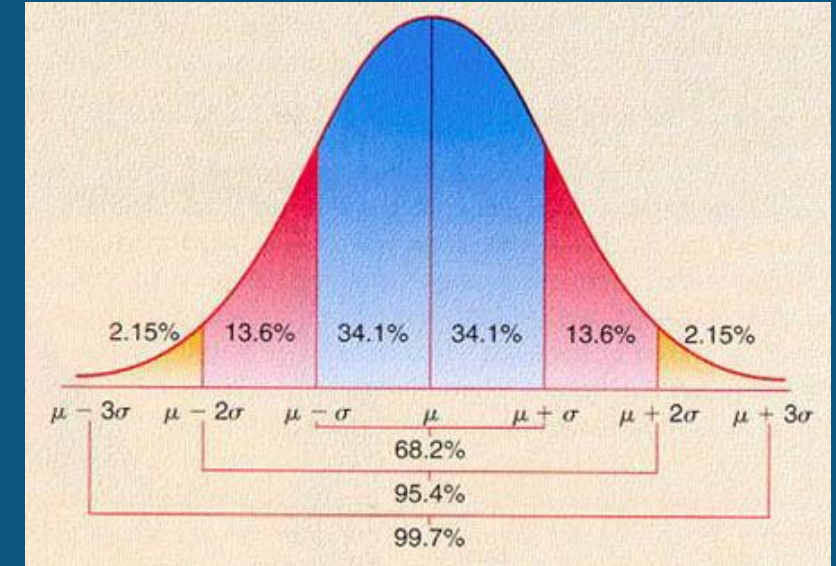
Pentru unidirectional (one-tailed) se imparte la 2 cel bidirectional si se precizeaza in paranteza.

Ex. $t(101) = 1,828, p=0.07$

$t(101) = 1,828, p=0.035$ (one-tailed)

3. Normalitatea distributiei variabilelor cantitative

- Distributia valorilor acestor variabile nu difera semnificativ de distributia normala



Daca aceasta conditie nu este indeplinita:

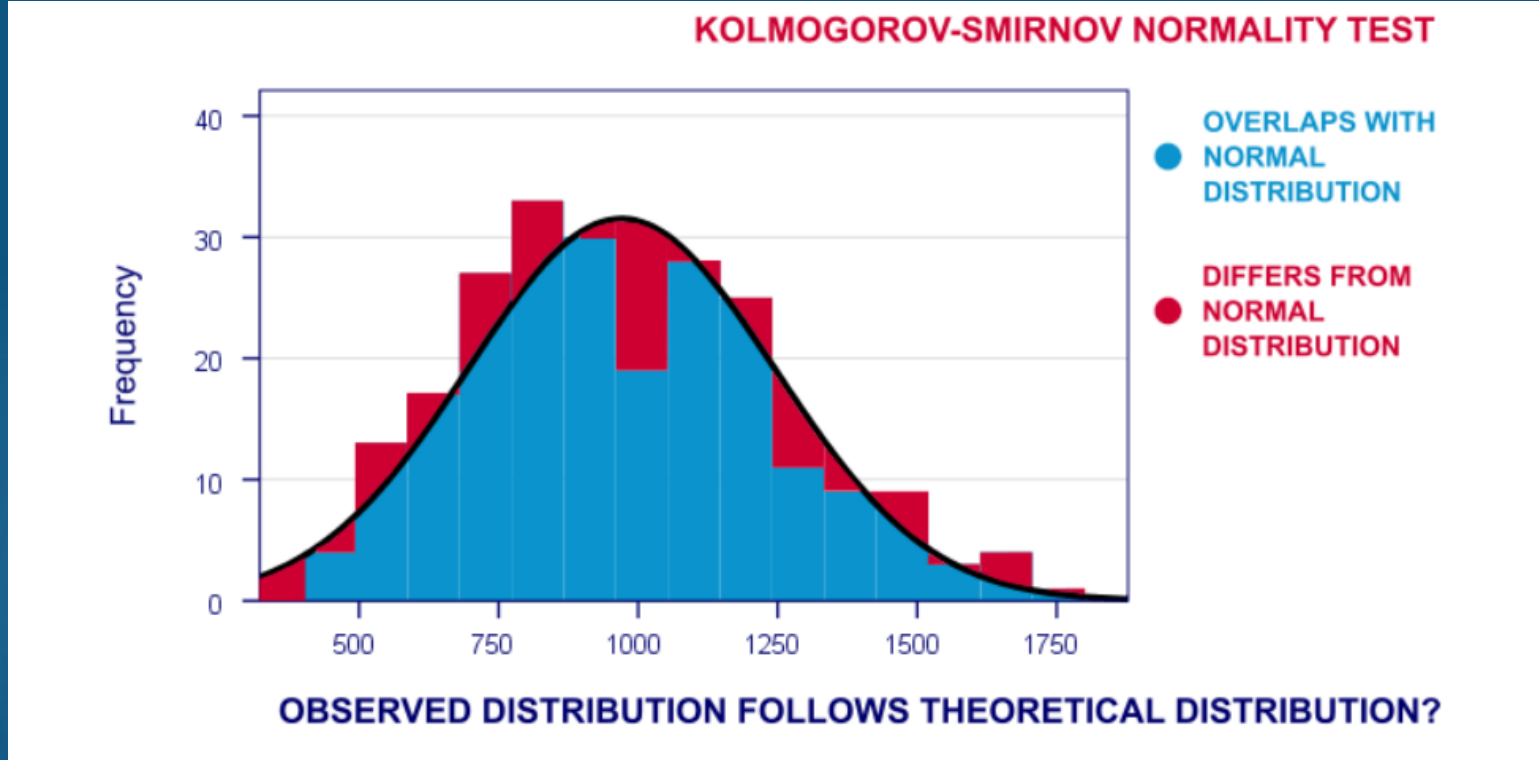
- Aplicarea testelor neparametrice;
- Transformarea variabilei prin diferite operatii (ex. logaritmare, extragerea radicalului, etc.;
- Categorizarea variabilei

Verificarea normalitatii

- Grafic – utilizarea histogramei
- Statistic:
 - testul Kolmogorov-Smirnov,
 - testul Shapiro-Wilk

Testul Kolmogorov-Smirnov

– compara distributia valorilor variabilei cu cea normala avand aceeasi medie si abatere standard



Daca $p > 0.05$ (test nesemnificativ) – nu exista diferente, variabila este normal distribuita

Testul Shapiro-Wilk este mai puternic

4. Teste t parametrice pentru comparare de medii

Tipuri de teste t

- se folosesc pentru variabile cantitative

- 1) Testul t pentru un esantion (One-Sample T Test)
- 2) Testul t pentru doua esantioane independente (Independent-Samples T Test)
- 3) Testul t pentru doua esantioane perechi (Paired-Samples T test)

1) Testul t pentru un esantion (One-Sample T Test)

- Compararea mediei unei variabile cu o valoare standard, fixata anterior

Conditii:

- 1) Variabila cantitativa, masurata pe scala de interval sau de proportii

- 2) Variabila normal distribuita

Exemplu: Notele obtinute la titularizare de studentii FEFS Craiova difera semnificativ sau nu de media pe tara.

1) Testul t pentru un esantion (One-Sample T Test)

Rezultate

- t (nr. grade de libertate) = valoare t , p

Exemplu: $t(29) = 3.81, p \leq 0,001$

- Se calculeaza marimea efectului, conform formulei lui Cohen (indicele de marime al efectului)

$$d = \frac{\mu - m_e}{\sigma}$$

in care:

μ – media populației;

m_e – media eșantionului;

σ – abaterea standard a populației (daca se cunoaste) sau a esantionului.

1) Testul t pentru un esantion

(One-Sample T Test)

Tabelul 4.1. Interpretarea mărimii efectului

Mărimea efectului	d Cohen	r
foarte puternic	$\geq 1,00$	$\geq 0,70$
puternic	0,80	0,50
mediu	0,50	0,30
slab	0,20	0,10

(Cohen, 1988, citat de Labăr, 2008)

2) Testul t pentru doua esantioane independente (Independent-Samples T Test)

- Compararea mediilor variabilei dependente a doua grupuri – design intergrup (between groups)

Conditii:

- 1) Independenta grupurilor (fiecare subiect se regaseste intr-un singur grup)
- 2) Variabila dependenta este **cantitativa**, masurata pe scala de **interval** sau de **proportii**
- 3) Variabila dependenta este **normal distribuita**

2) Testul t pentru doua esantioane independente (Independent-Samples T Test)

Conditii (continuare):

4) **Omogenitatea variantelor** – grupurile trebuie sa faca parte din populatii cu variante egale → **testul Levene**

Rezultat test Levene:

- Daca $p > 0,05$ (nesemnificativ), atunci variantele sunt egale → se va citi valoarea lui t
- Daca $p \leq 0,05$ (semnificativ), atunci variantele sunt inegale → se va citi valoarea ajustata a testului t

Ex. $F(56) = 0.123$, $p=0.402$, → variante egale

2) Testul t pentru doua esantioane independente (Independent-Samples T Test)

Se calculeaza **marimea efectului**, conform formulei lui Cohen:

$$d = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

→

$$d = \frac{t(n_1 + n_2)}{\sqrt{df} \sqrt{n_1 n_2}}$$

In care:

- M_1, M_2 – mediile celor doua esantioane
- σ_1, σ_2 – abateri standard
- n_1, n_2 – volume esantioane
- df – grade de libertate

3) Testul t pentru doua esantioane perechi (Paired-Samples T test)

- Compararea mediilor a doua variabile pereche
 - design intragrup (within groups)
 - cu masuratori repetate (repeated measures)

3) Testul t pentru doua esantioane perechi (Paired-Samples T test)

Patru demersuri:

- 1) Compararea valorilor unei variabile in conditii pre-test, post-test
- 2) Compararea valorilor unei variabile intre doua esantioane corelate (ex. valori-parinti, valori-copii)
- 3) Compararea acelorasi subiecti: doua aspecte diferite, masurate pe aceeasi scala – ex. interesul pentru doua materii diferite)
- 4) Verificarea fidelitatii unor instrumente de masurare:
 - Fidelitatea test – retest sau in timp;
 - Fidelitatea sau echivalenta formelor paralele

3) Testul t pentru doua esantioane perechi (Paired-Samples T test)

Conditii:

- 1) Variabila dependenta este **cantitativa**, masurata pe scala de interval sau de proportii
- 2) Variabila dependenta este **normal distribuita**
- 3) Variabila independenta este dihotomica, iar nivelurile (sau grupele) sunt pereche, sau legate intr-un anumit mod (pretest-posttest, copii-parinti, etc.)
- 4) **Diferenta** dintre valorile fiecarui subiect sau caz trebuie sa fie **normal distribuita** (ex. atat pretest, cat si posttest)

3) Testul t pentru doua esantioane perechi (Paired-Samples T test)

Exemplu:

Eficienta unui program de antrenament

VI – momentul temporal (testare initiala, testare finala)

VD – timpul realizat pe 100m (initial si final) – variabile pereche

- Se calculeaza t si p, se interpreteaza
- Se calculeaza marimea efectului si se interpreteaza

$$d = \frac{\bar{D}}{s_D}$$

$$d = \sqrt{\frac{t^2}{df}} = \sqrt{\frac{t^2}{N-1}}$$

$\bar{D}$ - media diferentelor de scoruri

s_D – abaterea standard a diferentelor de scoruri



<https://www.gnu.org/software/pspp/>



CORELATIA

Sorina CERNĂIANU

Facultatea de Educație Fizică și Sport, Craiova

s_cernaianu@yahoo.com

Tipuri de analiză a datelor

- **Analiză univariată** – studiul frecvențelor de apariție, măsurarea tendinței centrale, dispersiei, indicatori de poziție;
 - variabilele se analizează una câte una;
- **Analiză bivariată** – tabele de contingență, analiză de varianță, **corelație, regresie simplă**;
 - se studiază relația între două variabile;

Tipuri de analiză a datelor

- **Analiză multivariată** – analiza de corespondență, analiza în componente principale, analiza discriminantă, regresia multiplă;
 - se studiază relația dintre trei sau mai multe variabile.

Generalități

Corelație = asociere, legătură, interdependență
- Se măsoară două caracteristici cantitative (variabile) pentru același eșantion.

Exemple:

- Există o relație între numărul de ore de antrenament și rezultatul obținut în competiție?
- Există o relație între înălțime și masa corporală?

Generalități

Valori cuprinse între -1 și 1

- 1 - corelație perfect negativă
- +1 - corelație perfect pozitivă
- 0 - nu există corelație (asociere aleatoare)

Coeficient de corelație mare → legătură puternică

- Coeficient **pozitiv** – corelație directă

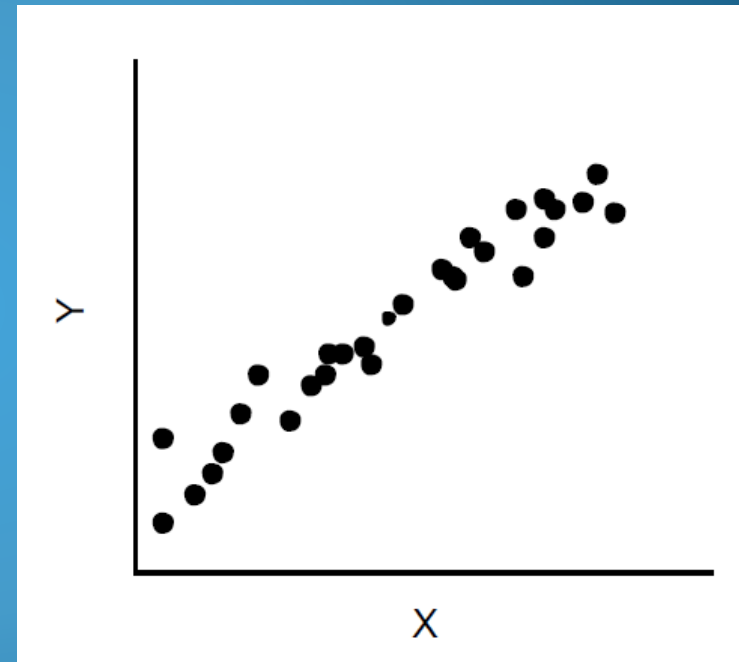
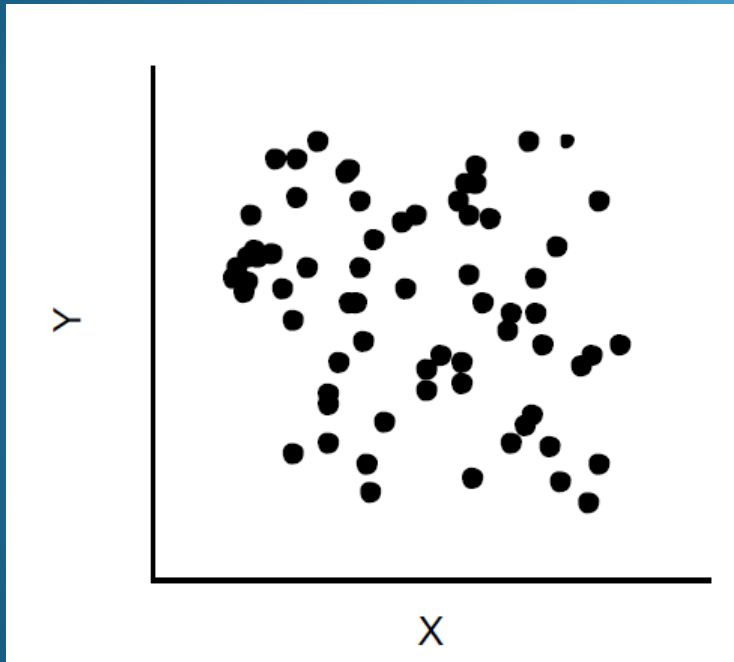
Exemplu: IMC și masa corporală

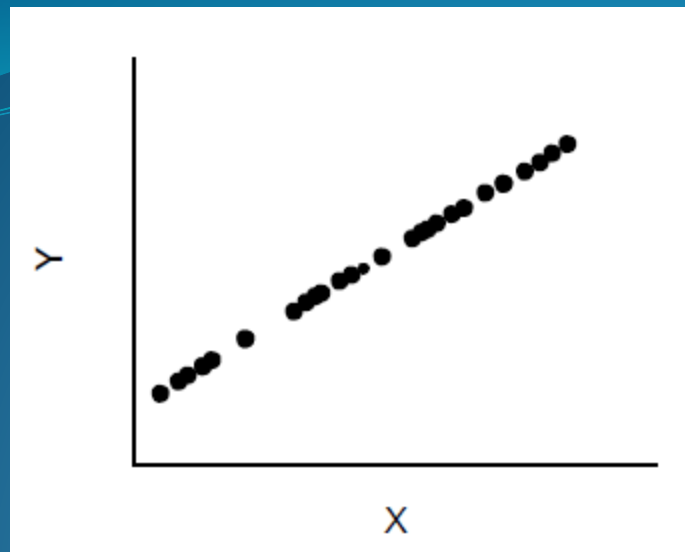
- Coeficient **negativ** – corelație inversă

Exemplu: IMC și înălțime

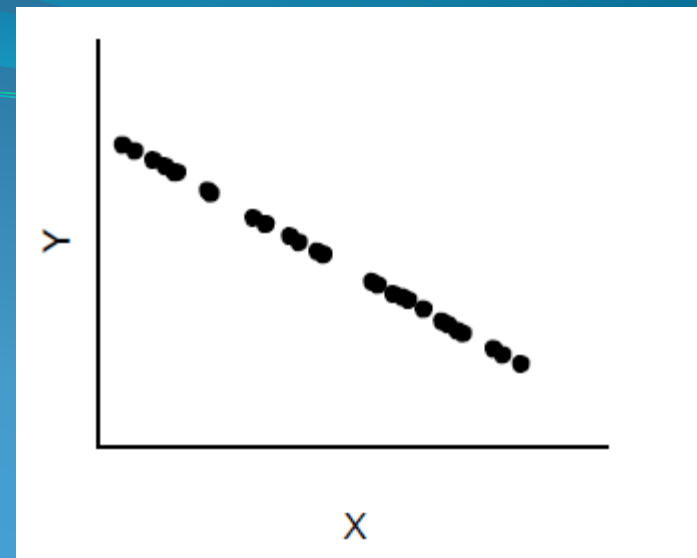
$$\text{IMC} = \text{greutatea corporala (kg)} / (\text{inaltimea in m})^2$$

Diagrama de dispersie

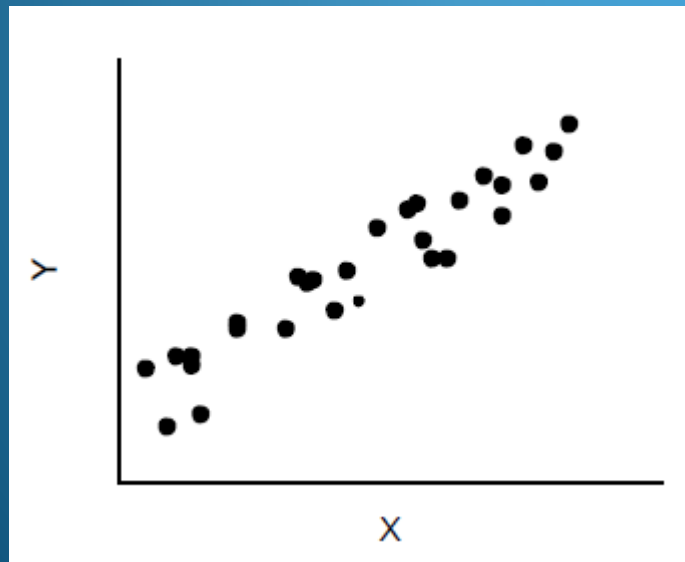




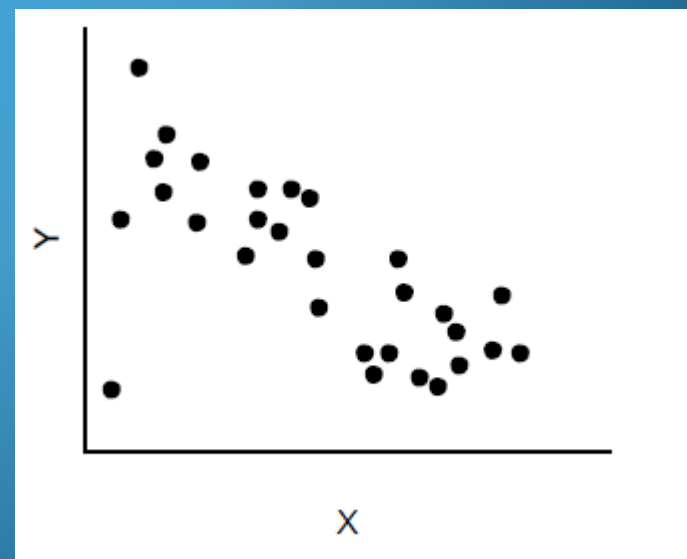
Corelatie perfect pozitiva



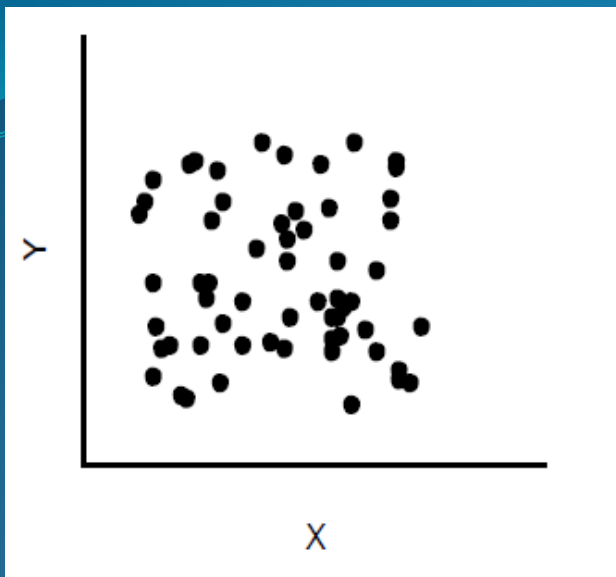
Corelatie perfect negativa



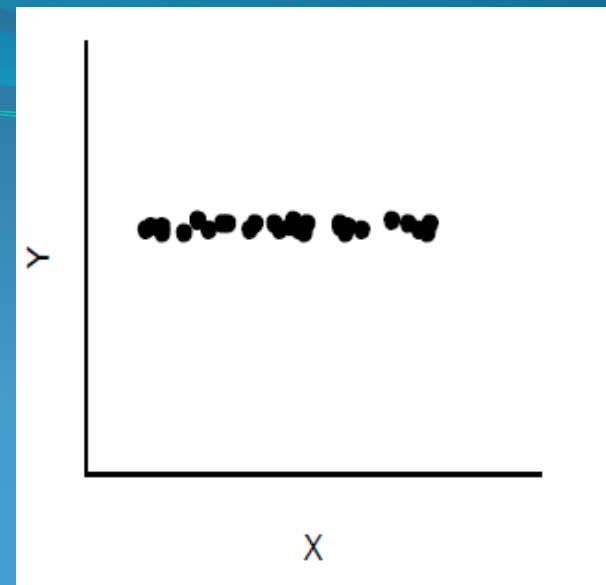
Corelatie puternic pozitiva



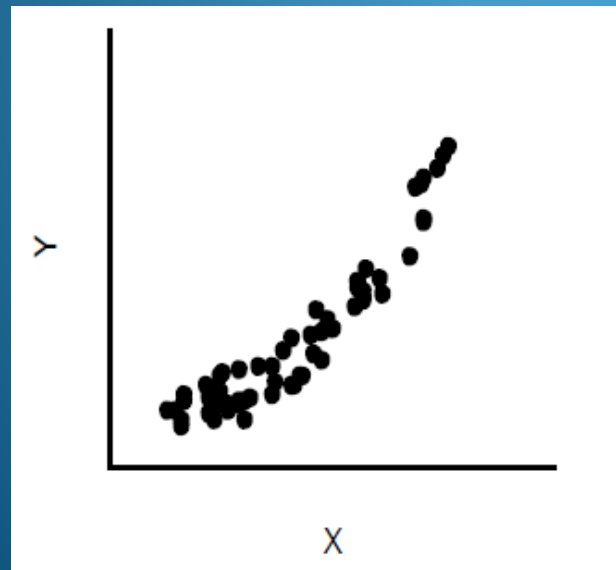
Corelatie puternic negativa



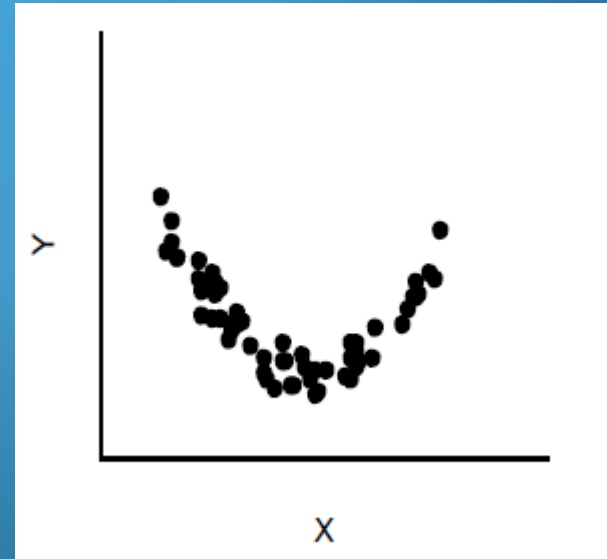
Nu exista corelatie



Nu exista corelatie



Relatie non-liniara



Relatie non-liniara

Caracteristici

- 1. Direcția:
 - pozitivă
 - negativă
- 2. Forma:
 - liniară
 - neliniară
- 3. Gradul de asociere:
 - Între -1 și 1

Tipuri de coeficienți de corelație:

1. Pearson, r
2. Spearman, ρ (rho)
3. Kendall, tau-b
4. Coeficientul de corelație parțială

1. Coeficientul de corelație Pearson r

Condiții:

- 1) Cele două variabile sunt cantitative
- 2) Variabilele sunt normal distribuite
- 3) Relația dintre variabile este liniară

Cazurile extreme (outlier) pot distorsiona liniaritatea relației dintre variabile.

Se elimină și se recalculează r .

1. Coeficientul de corelatie Pearson, r

$$r_{xy} = \frac{\sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}}$$

În care:

- x_i = valorile individuale ale variabilei x
- y_i = valorile individuale ale variabilei y
- $\bar{x}$ = media aritmetică a tuturor valorilor x
- $\bar{y}$ = media aritmetică a tuturor valorilor y

Valori coeficient r

$-0,25 < r < 0,25$	Fara corelatie
$-0,5 < r < -0,25$ sau $0,25 < r < 0,5$	Corelatie slaba
$-0,75 < r < -0,5$ sau $0,5 < r < 0,75$	Corelatie moderata
$r < -0,75$ sau $r > 0,75$	Corelatie puternica

(Colton, 1974)

r value	Classification of Relationship/Correlation Strength
0.10 until 0.29	Weak
0.30 until 0.49	Moderate
0.50 until 1.0	Strong

(Cohen, 1988)

Coeficientul de corelatie Pearson r

- test normalitate: Kolmogorov-Smirnov sau Shapiro-Wilk;

Mărimea efectului:

- doi indicatori de bază: r , respectiv r^2
- efect scăzut ($r = .10$, $r^2 = .01$);
- efect mediu ($r = .30$, $r^2 = .09$);
- efect puternic ($r = .50$, $r^2 = .25$).

(Cohen, 1988)

Coeficientul de determinare (r^2)

- cât din variația unei variabile este explicată prin variația celeilalte variabile;
- pătratul coeficientului Pearson
- indică procentul din variația totală a variabilei dependente, care este explicată de variabila independentă

Exemplu: $r = 0,70$

$r^2 = 0,49$, adică

49% din variația unei variabile este explicată prin variația celeilalte variabile.

2. Coeficientul Spearman ρ (rho)

Condiții:

- 1) cele două variabile sunt cantitative, însă **nu sunt** normal distribuite;
- 2) ambele variabile sunt ordinale sau una este cantitativă, cealaltă ordinală;
- 3) ambele variabile sunt cantitative, însă eșantionul este de volum redus (ex. 7 subiecți);
- 4) valorile unei variabile sunt în mod monoton legate de valorile celeilalte variabile, nu neaparat linear (cresc sau descresc) – **inspectăm** **norul de puncte**.

2. Coeficientul Spearman ρ (rho)

- test neparametric
- corelație a rangurilor
- e mai puțin puternic decât coeficientul Pearson

Cazurile extreme (outlier) pot distorsiona liniaritatea relației dintre variabile.

Se elimină și se recalculează r .

Semnificația coeficientului de corelație

$$t_{calc} = r \sqrt{\left(\frac{n-2}{1-r^2} \right)}$$

- n – numărul de perechi de valori

Semnificație statistică dacă $p < 0,05$

- **one-tailed** – când se presupune tipul de corelație (pozitivă sau negativă)
- **two-tailed** – nu se presupune tipul de corelație.

Interpretarea valorilor lui p

- $p < 0.05$, legătura statistică este **semnificativă** (încredere 95%);
- $p < 0.01$, legătura statistică este **semnificativă** (încredere 99%);
- $p < 0.001$, legătura statistică este **înalt semnificativă** (încredere 99.9%);
- $p > 0.05$, legătura statistică este **nesemnificativă**.

3. Coeficientul τ (tau) al lui Kendall

Condiții:

1. ambele variabile sunt ordinale;
2. variabilele au un număr redus de categorii (variante de răspuns la item) (ex. 3);
3. datele ale căror ranguri egale depășesc 25% din numărul lor.

Se interpretează și mărimea efectului.

4. Corelația parțială

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

Condiții:

- 1) cele două variabile sunt cantitative;
- 2) variabilele sunt normal distribuite;
- 3) relația dintre variabile este liniară.

Def. - Corelația dintre două variabile în condiția menținerii constante a efectului altei (altor) variabile asupra celor două variabile analizate.

4. Corelația parțială

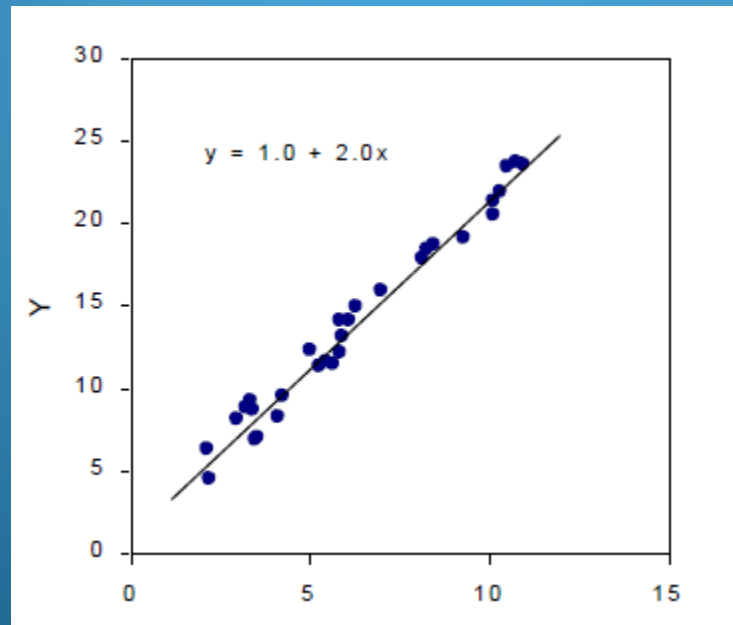
$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

Care este relația dintre X și Y , dacă ținem constant Z ?”

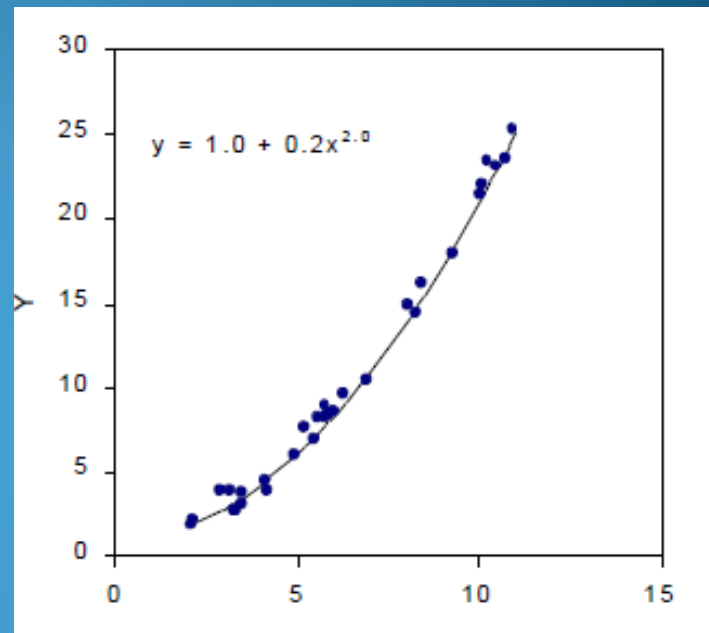
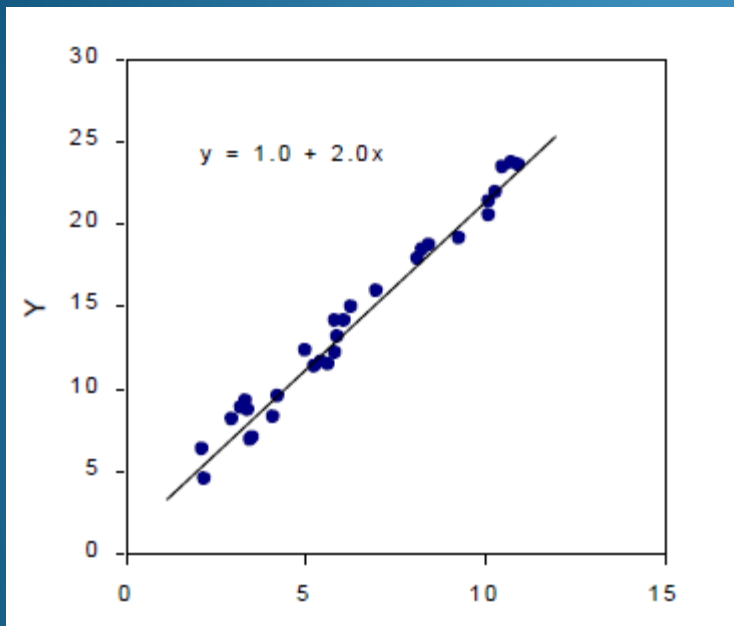
- r_{XY} = corelația simplă (Pearson) dintre X și Y
- r_{XZ} = corelația dintre X și Z
- r_{YZ} = corelația dintre Y și Z
- $r_{XY \cdot Z}$ = **corelația parțială** dintre X și Y , controlând Z

Regresia

- descrie relația dintre cele două variabile;
- ajută la efectuarea de estimări;
- variabile dependente și independente.



Tipuri de regresie



- **liniară** $y=a+bx$

- **neliniară** $y=a+bx^c$

- Simplă sau bivariată (cu o singură VI)
- Multiplă (cu două sau mai multe VI)

Regresia liniara simpla

Se aplică:

- **În scop predictiv**

Ex.: prezicerea (prin intermediul unei ecuații de regresie) rezultatului (valorii VD), cunoscând valoarea VI

- **În scop explicativ**

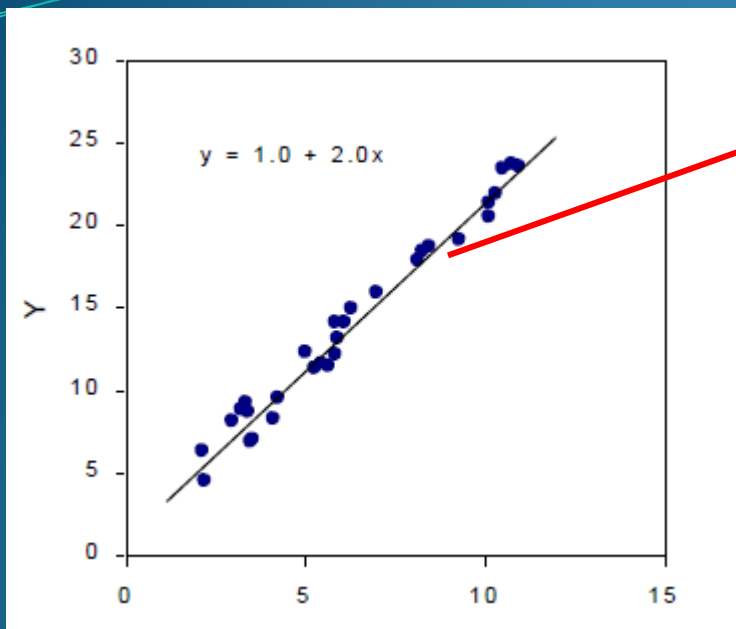
Ex.: aflarea procentului din varianța VD explicat de varianța VI

Regresia liniară simplă

Condiții:

- 1) VI și VD sunt cantitative;
- 2) VI și VD sunt normal distribuite;
- 3) relația dintre VI și VD este liniară;
- 4) erorile sunt normal distribuite;
- 5) evitarea cazurilor influente și a celor extreme.

Regresia liniară simplă



dreapta de regresie

$$y = a + bx$$

În care:

x = variabilă independentă

y = variabilă dependentă

a, b – coeficienți de regresie

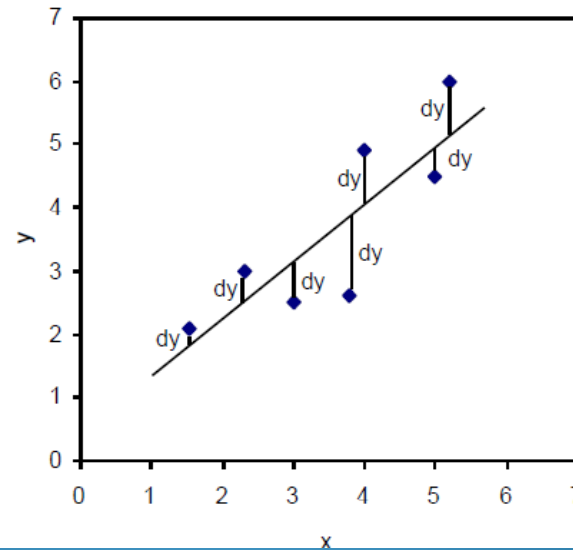
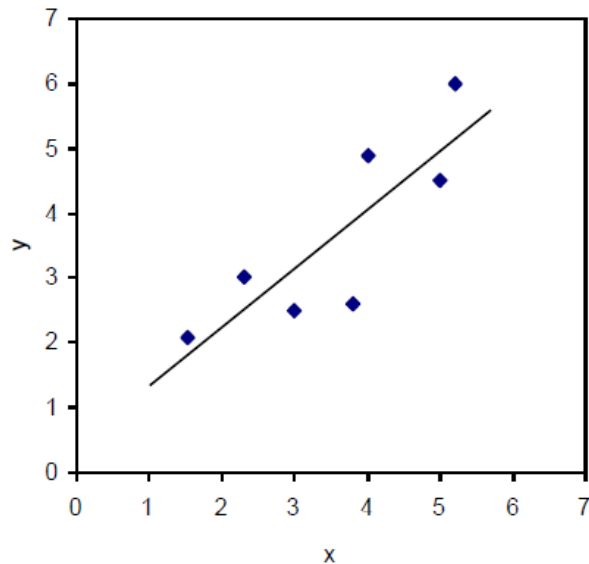
$a \rightarrow$ **intercept**: valoarea lui y când x este zero (intersecția cu OY)

$b \rightarrow$ **panta dreptei de regresie**: cantitatea cu care valoarea y se modifică în momentul în care modificăm valoarea lui x cu o unitate.

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Metoda celor mai mici pătrate

- Calculul coeficienților de regresie a și b , astfel încât suma pătratelor distanțelor până la dreapta de regresie să fie **minimă**.



$$y = a + b \cdot x$$

$$\sum dy = 0$$

$$\sum dy^2 = \text{minimum}$$

$$dy = y_i - y_i^{est}$$

Metoda celor mai mici pătrate

$$n \cdot a + b \sum x = \sum y$$

$$a \sum x + b \sum x^2 = \sum (x \cdot y)$$

In care:

n - mărimea eșantioanelor

$\sum x$ - suma tuturor valorilor x

$\sum y$ - suma tuturor valorilor y

$\sum x^2$ - suma tuturor pătratelor valorilor x

$\sum xy$ - suma tuturor produselor $x \cdot y$.

- se rezolvă sistemul de ecuații;
- se determină a și b ;
- se trasează dreapta de regresie;
- se testează semnificația coeficienților de regresie – testul t pentru pantă și pentru intercept.

Exemplu

Verificam legatura dintre înălțime și masa corporală.

VI – înălțimea (m)

VD – masa corporală (kg)

Verificarea normalității variabilelor

NPART TEST
/KOLMOGOROV-SMIRNOV (NORMAL) = Inaltime.

One-Sample Kolmogorov-Smirnov Test

		Inaltime
N		28
Normal Parameters	Mean	1.65
	Std. Deviation	.09
Most Extreme Differences	Absolute	.10
	Positive	.10
	Negative	-.08
Kolmogorov-Smirnov Z		.52
Asymp. Sig. (2-tailed)		.949

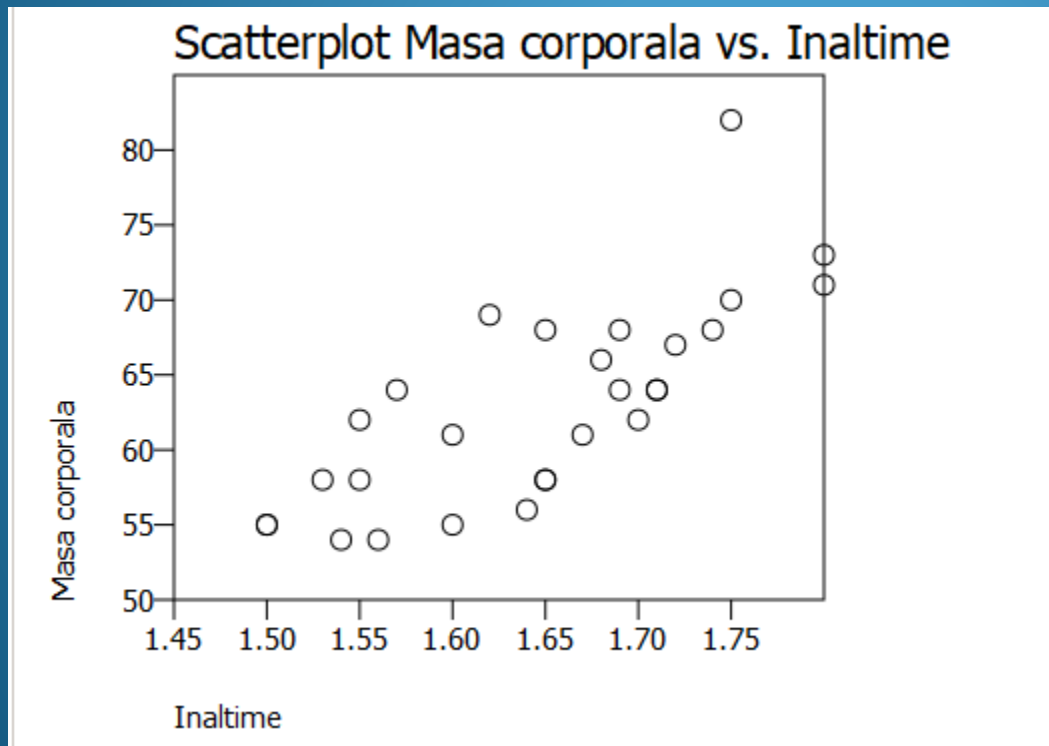
NPART TEST
/KOLMOGOROV-SMIRNOV (NORMAL) = Masa_corporala.

One-Sample Kolmogorov-Smirnov Test

		Masa corporala
N		28
Normal Parameters	Mean	63.04
	Std. Deviation	6.74
Most Extreme Differences	Absolute	.13
	Positive	.13
	Negative	-.09
Kolmogorov-Smirnov Z		.69
Asymp. Sig. (2-tailed)		.735

Exemplu

- Verificarea norului de puncte



Exemplu - rezultate

Model Summary (Masa corporala)

R	R Square	Adjusted R Square	Std. Error of the Estimate
.77	.59	.57	4.42

ANOVA (Masa corporala)

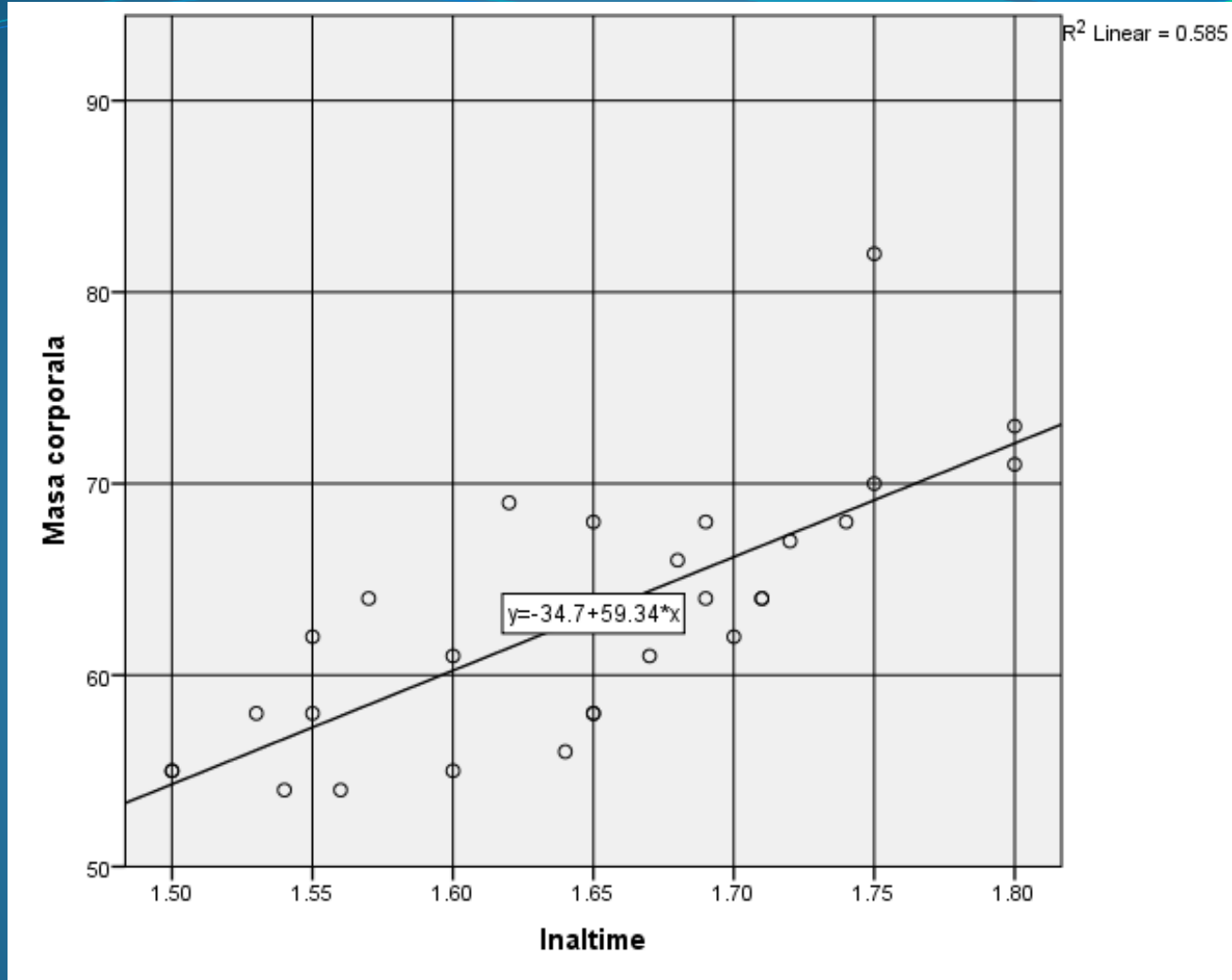
	Sum of Squares	df	Mean Square	F	Sig.
Regression	718.14	1	718.14	36.70	.000
Residual	508.83	26	19.57		
Total	1226.96	27			

Coefficients (Masa corporala)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-34.70	16.16	.00	-2.15	.041
Inaltime	59.34	9.80	.77	6.06	.000

$F(1, 26) = 36.7, p < 0.001$

Exemplu - rezultate



Ecuția de regresie:

$$\text{Masa corporală} = -34.7 + 59.34 \times \text{Înălțime}$$

Ancheta prin chestionar ce vizeaza frecventarea salilor de fitness

1. Item ce necesita raspuns unic

Cu precadere, unde practicati activitati sportive?

Rata de raspuns 98.5%

Variante de raspuns	Nr.	% raspunsuri
La domiciliu	16	12
Intr-o sala de sport a unui club sau a unei asociatii sportive	28	21,1
Intr-o sala de sport privata	66	49,6
Intr-o sala de sport, la serviciu	21	15,8
Altele	2	1,5
Total	133	100

Procentul este calculat din **numarul de raspunsuri**.

Analiza:

A) Fara non-raspunsuri :

Raspunsul X reprezinta% din ansamblul raspunsurilor care au fost date.

Din cei 98,5% respondenti, 49.6% practica o activitate sportiva intr-o sala de sport privata, 21,1% practica intr-o sala a unui club sau a unei asociatii sportive, in timp ce numai 1,5% practica activitati sportive in alte locuri.

B) Cu non-raspunsuri :

Variante de raspuns	Nr.	% raspunsuri
Non-raspuns	2	1,5
La domiciliu	16	11,9
Intr-o sala de sport a unui club sau a unei asociatii sportive	28	20,7
Intr-o sala de sport privata	66	48,9
Intr-o sala de sport, la serviciu	21	15,6
Altele	2	1,5
Total	135	100

48,9% din ansamblul persoanelor interogate practica o activitate sportiva intr-o sala privata.

Comparatie

Variante de raspuns	Nr.	% raspunsuri	Nr.	% raspunsuri
Non-raspuns	2	1,5	-	-
La domiciliu	16	11,9	16	12
Intr-o sala de sport a unui club sau a unei asociatii sportive	28	20,7	28	21,1
Intr-o sala de sport privata	66	48,9	66	49,6
Intr-o sala de sport, la serviciu	21	15,6	21	15,8
Altele	2	1,5	2	1,5
Total	135	100	133	100

2. Itemi ce necesita raspuns multiplu

Dintre criteriile urmatoare, care sunt cele care contribuie la o imagine buna a salilor de sport?

Variante de raspuns	Nr.	% din nr. total de respondenti
Diversitatea activitatilor sportive propuse	54	40
Personalul competent	52	38.5
Ambianta buna	52	38.5
Aparatura de calitate	59	43.7
Orar convenabil	43	31.9
Igiena locatiei	42	31.1
Traineri de calitate	52	38.5
Tarife atractive	21	15.6
Altele	9	6.7
Total	135	-

Procentele sunt calculate din numarul total de respondenti.

Analiza:

X% dintre indivizi considera ca, gandesc ca.....

43,7% dintre respondenti considera ca o aparatura de calitate contribuie la crearea unei bune imagini a salii de sport, in timp ce numai 6,7% sunt atrasi datorita altor criterii.

(GANASSALI, Stephane (2009). *Les enquêtes par questionnaire avec Sphinx*, Pearson Education: Paris.)

Reprezentarea rezultatelor

1. Analiza univariata

A. Itemi cu raspuns unic

a) Intre 2-5 variante de raspuns :

- diagrama cu structura radiala
- se insereaza rata de raspuns; nu se afiseaza nr. de non-raspunsuri

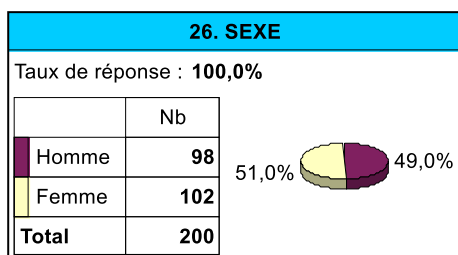


Figura 1. Diagrama cu structura radiala

b) Mai mult de 5 variate de raspuns :

- diagrama tip bara sau histograma
- se insereaza rata de raspuns si se ordoneaza variantele de raspuns in functie de frecventa de aparitie

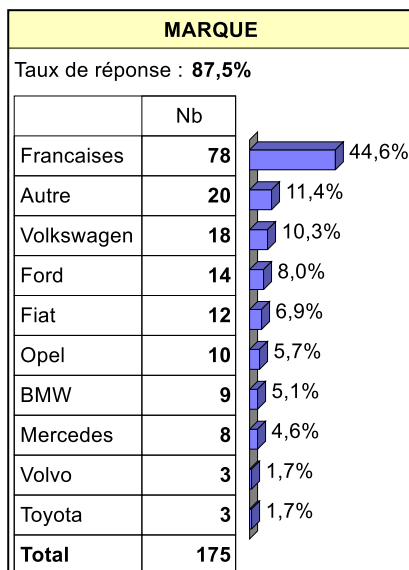


Figura 2. Diagrama tip bara

B. Raspunsuri de tip numeric

- raspunsurile se grupeaza in clase de valori
- mai putin de 5 clase – diagrama cu structura radiala
- 5 si mai mult de 5 clase – diagrama tip bara
- se insereaza media aritmetica, abaterea standard pentru a indica dispersia raspunsurilor

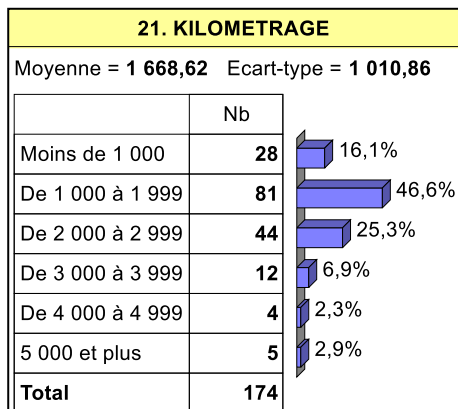


Figura 3. Gruparea in clase de valori a unei variabile numerice si reprezentarea grafica de tip bara

C. Itemi cu raspuns multiplu

- mai multe raspunsuri posibile
- daca se face suma frecventelor de aparitie, aceasta este mai mare decat 100
- **IMPORTANT : les fréquences sont calculées sur les observations ;**
- diagrama tip bara
- se insereaza rata de raspuns si se ordoneaza variantele de raspuns in functie de frecventa de aparitie

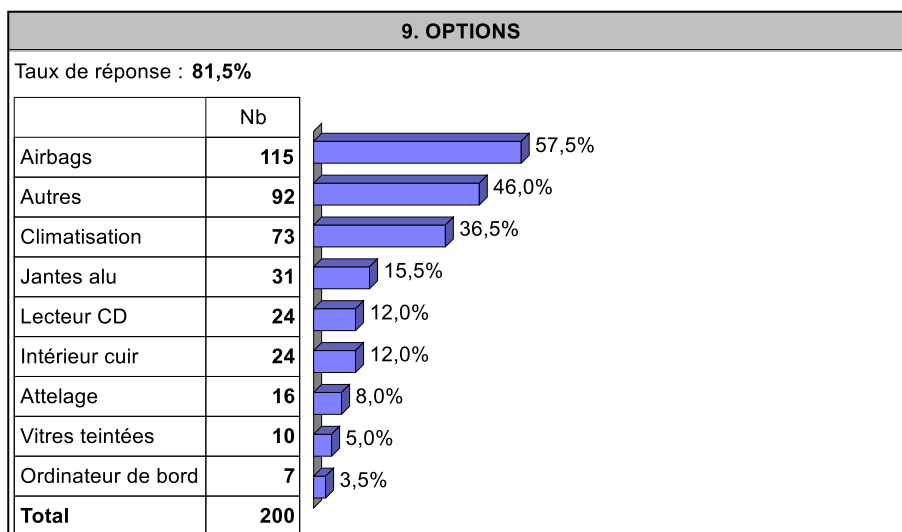


Figura 4. Reprezentarea rezultatelor pentru un item cu raspuns multiplu

D. Scale Likert

- grup de itemi de tip Likert, vizand satisfactia, acordul sau importanta
- diagrama tip bara, incluzand procentajul
- se insereaza o legenda langa grafic
- se reprezinta raspunsurile sub forma numerica, furnizand pentru fiecare item din grup media calculata in functie de baremul asociat

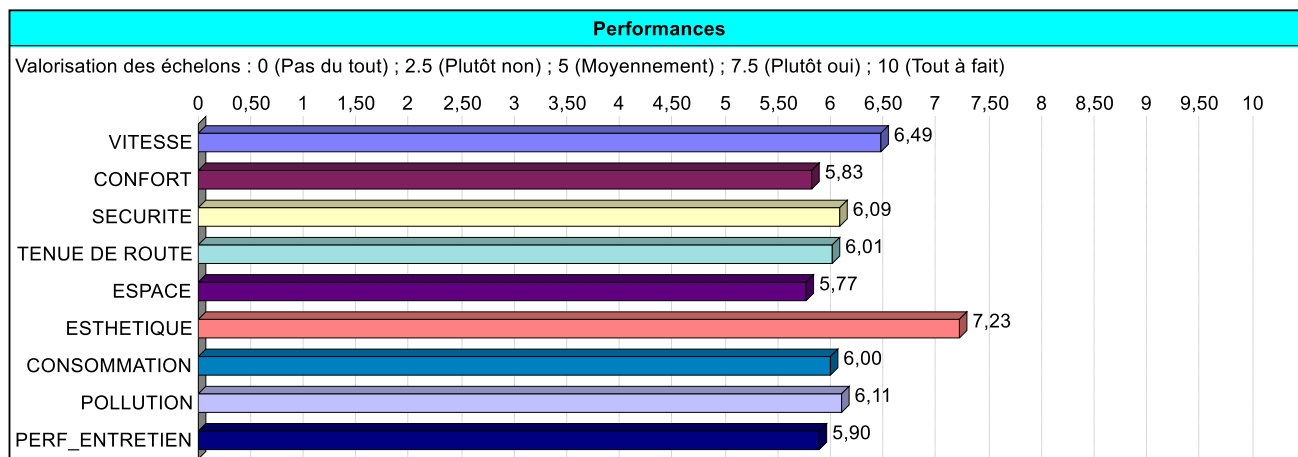


Figura 6. Reprezentare a rezultatelor pentru un grup de itemi – media aritmetica

E. Analiza bivariata

1. Asocierea raspunsurilor a doi itemi cu raspuns unic

- variabila explicativa pe linie ;
- - variabila explicata pe coloana ;
- calculati procentele pe linie

Exemplu – ancheta in turism:

- profesia pe linie (pp. ca influenteaza tipul de cazare)
- 9,1% dintre agricultori se cazeaza la hotel
- diagrama tip bara - repartitia diferitelor categorii ale variabilei explicative (profesia) in functie de diferitele modalitati de raspuns ale variabilei cazare.

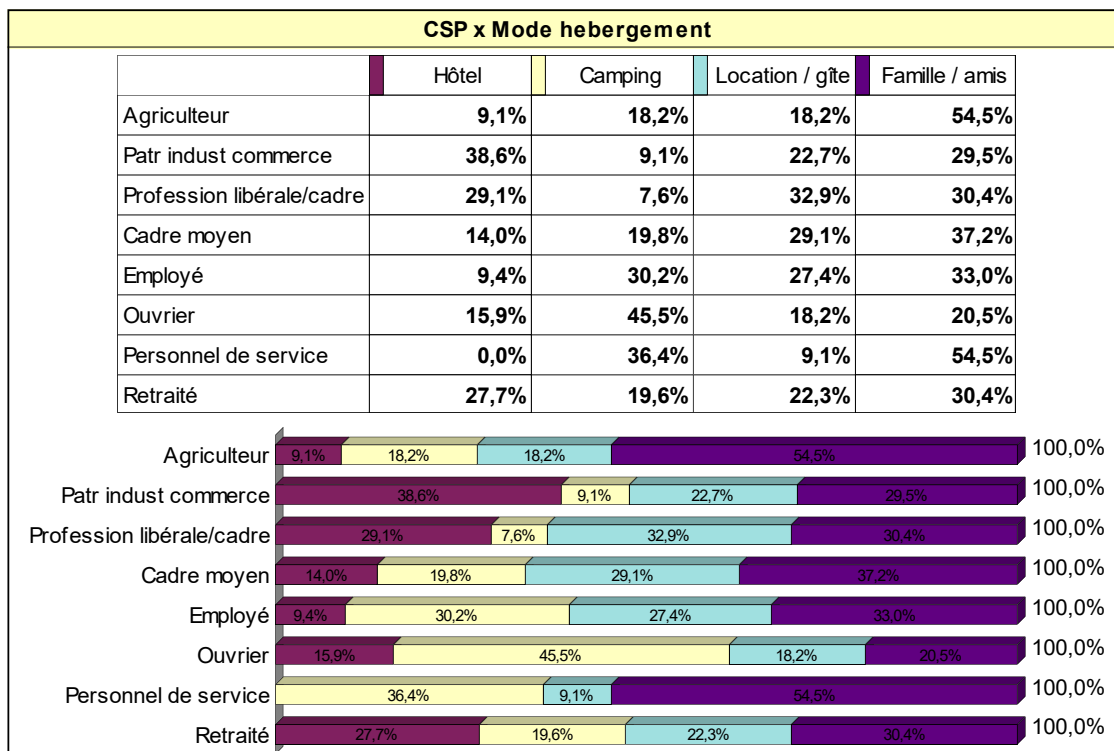


Figura 7. Reprezentare grafica a asocierii dintre doua variabile cu raspuns inchis

2. Raspuns inchis x raspuns numeric

- plasati variabila explicativa in linie
- plasati media variabilei numerice explicate pe coloana
- diagrama tip bara sau curbe daca trebuie sa reprezentam evolutia in timp

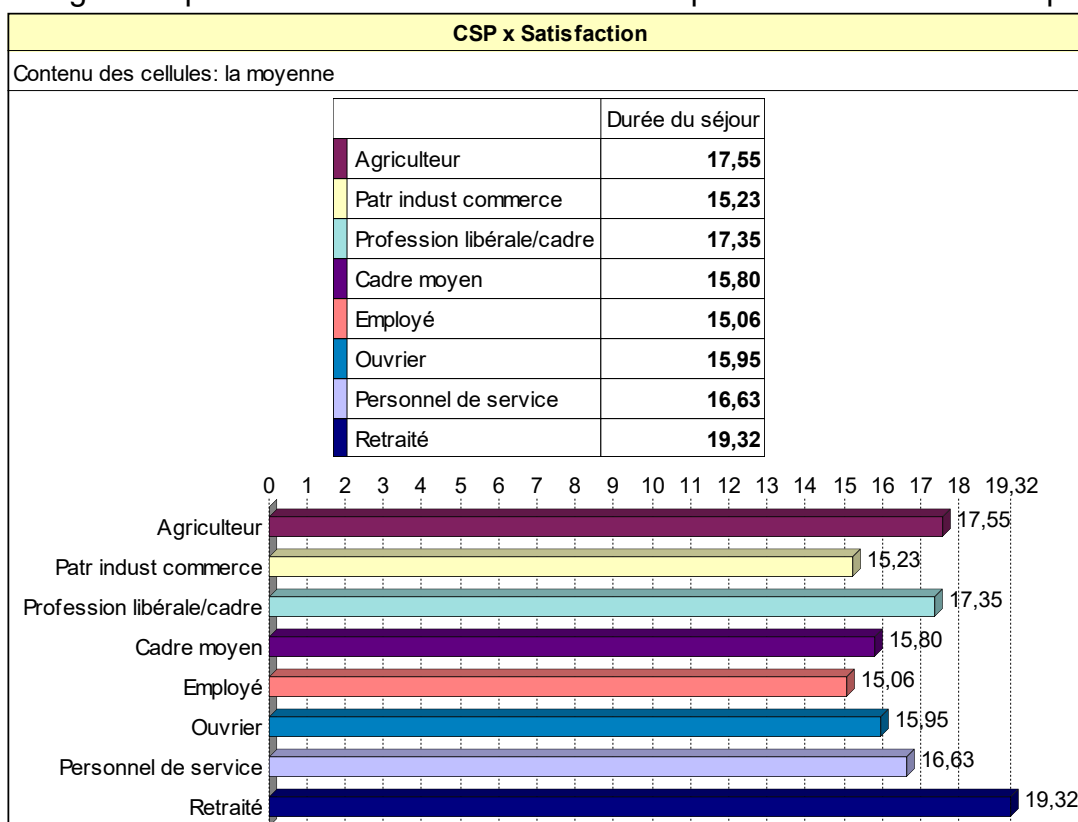


Figura 8. Reprezentarea rezultatelor asocierii dintre o variabila cu raspuns inchis si una numerica – diagrama tip bara