

Status = stat \Rightarrow Statistică

- Statistica = știința
colecției
organizării
analizării și
interpretării datelor în vederea luării unei decizii
- Date = informații provenite din observație, numărare, măsurare sau răspunsuri. Valori colectate.

Multiplu de date: populație vs. eșantion

Populație = colecția tuturor rezultatelor, răspunsurilor, măsurătorilor sau numărărilor de interes

Eșantion = o submulțime a populației

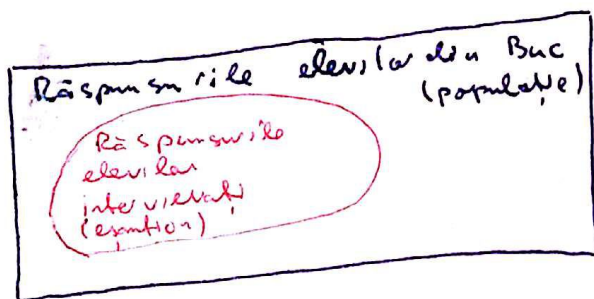
Exemplu: 3000 de tineri elevi din București au fost întrebați dacă obișnuiesc să lucreze zilnic la matematică.

600 au răspuns afirmativ. Descrieți datele!

Soluție: Populație = elevii din București (Răspunsurile elevilor din București)

Eșantionul = cei 3000 de elevi intervievați (Răspunsurile celor 3000 de elevi intervievați)

Multiplu de date: 600 de "Da!"
2400 de "Nu!"



Exercițiul 1: Pentru a stabili prețul mediu pentru 1kg de mere se efectuează un studiu la 600 de țărâșe. În septembrie 2017 prețul mediu a fost 29,15 lei. Identificați ce este populația și eșantionul. Care sunt datele?

Datele pot fi obținute de la întreaga populație sau doar de la un esantion: în primul exemplu datele puteau proveni de la întreaga populație a părerilor din București.

Parametru = descriere numerică a unei caracteristici a populației

Statistică = descriere numerică a unei caracteristici a unui esantion

Exemplul 2: Decideți dacă valoarea numerică este un parametru al populației sau o statistică a esantionului:

a) Un studiu recent efectuat pe 700 de absolvenți arată că absolvenții Oxford sau Cambridge debutează cu salarii mai mari cu o medie de 7600 £ pe an față de absolvenții altor universități.

b) Un studiu efectuat pe cei 7000 de absolvenți din universitățile din București din 2016 arată că un absolvent câștigă în primul an în medie $2.366 \cdot 12^{12} = 28.392^{12} \approx 5460$ £.

Soluție: (*)

a) Deoarece mărimea numerică de 7600 £ se referă pe un esantion al populației absolvenților este o statistică.

b) Deoarece mărimea numerică de 5460 £ se referă la populația absolvenților este un parametru.

Exercițiul 2: În 2016 Dacia a vândut 584.219 vehicule dintre care 450.000 cu aer condiționat. Decideți dacă această ultimă valoare numerică este un parametru al populației sau o statistică a unui esantion?

(*) Populație = absolvenți

esantionul = 700 de absolvenți

variabile statistice = salariul în primul an

variante = valori posibile ale salariilor

date = variante alese drept să știm

Ramuri ale statisticii: statistici descriptive vs. statistici inferentiale

Statistica descriptivă este ramura statisticii ce se ocupă cu organizarea, cantitizarea și prezentarea datelor

Statistica inferențială este ramura statisticii ce folosește esanțioanele pentru a trage concluzii despre populație folosind elemente de teoria probabilităților.

Exemplul 3:

Un studiu efectuat în 2009 a arătat că
11.900 oameni bărbați au murit în accidente de mașină și doar
4.900 oameni femei (în decursul aceluși an.)

Care parte a studiului reprezintă ramura de statistică descriptivă?
Ce concluzii se pot trage folosind statistica inferențială?

*

Exercițiul 3:

Un studiu efectuat în 2009 a arătat că rata mortalității
în accidente de mașină a fost de

2,5 morți / 100 milioane mile	în rândul bărbaților adulți
1,7 morți / 100 milioane mile	în rândul femeilor adulți
9,2 morți / 100 milioane mile	pt tineri de 16-19 ani
5,3 morți / 100 milioane mile	pt tineri de 16-19 ani

- Identificați aspectele descriptive ale studiului
- Ce previziuni se poate efectua?

În acest curs vom parcurge aplicații ale ambelor ramuri ale
statisticii. O temă importantă va fi folosirea statisticii
esantionelor pt a face inferențe despre parametrii populației

* Soluție:

Aici am afirmat "11.900 oameni bărbați au murit în accidente de mașină
în 2009" respectiv "4.900 oameni femei au decedat în accidente rutiere"
fac parte din domeniul de interes al statisticii descriptive.

Statistica inferențială se ocupă cu emiterea și justificarea unor
afirmații cum ar fi că "Femeile sunt oameni mai prudenți decât
măi buni ca bărbații."

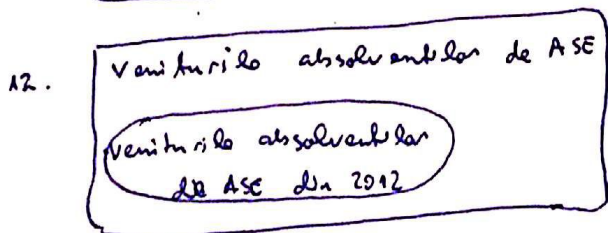
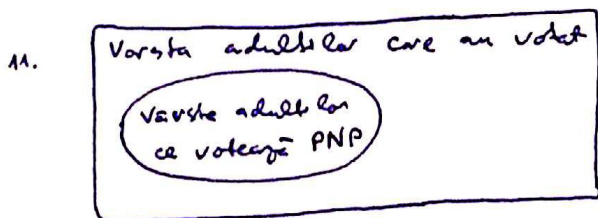
Exersarea:

1. Care-i legatură dintre esantion și populație?
2. De ce se efectuează mai des sondaje pe esantioane decât pe populații?
Statisticii valoarea de adevăr a proprietăților:
3. O statistică descrie o caracteristică a unei populații
4. Un esantion este o submulțime a unei populații
5. Un sondaj de opinie se realizează interogând întreaga populație a țării.
6. Statistica inferențială folosește esantioane pentru a trage concluzii despre întreaga populație.

În exercițiile de mai jos stabiliți dacă mulțimea de date este un esantion sau o populație:

7. Vârsta fiecărui elev din clasa XII-a din țară
8. Vârsta fiecărui al cincilea pensionat din HAPN
9. Salariul fiecărui al zecelea locuitor.
10. Un sondaj pe 200 de elevi din cei 800 ai unei școli

În exercițiile de mai jos identificați esantioanele / populațiile din diagramele Venn:



Identificați mulțimile de date: populație / esantion și parametrii / statisticele.

13. Un studiu efectuat pe 2000 de gospodării din SUA arată că în 2017 10% au renunțat la cablul TV.
14. Un studiu efectuat pe 300 de familii din București arată că plămișca ră chelburăscă în medie 1200 € pentru vacanța următoare
15. Venitul mediu anual pt. 100 din cei 2000 de angajați ai unei companii este de \$51.000.
16. Venitul mediu anual al angajaților Primăriei București este \$9300/an

Clasificarea datelor

- După tip : (numerice)
 - date calitative : atribute, denumiri, întrebări numerice
 - date cantitative : numerice : mărimi (numerice) sau numărări

Exemplul 1: În tabelul de mai jos sunt trecute prețurile biletelor dus-întors cu avionul din București către câteva orașe europene pe data de 26 oct 2017:

Oraș	Preț
Madrid	260
Brussels	153
Paris	568
Milano	100
Roma	118

Care date sunt calitative și care-s cantitative? De ce?

Soluție:

În tabel sunt prezentate două mulțimi de date:
un set de date cantitative numele destinațiilor
al 2-lea set cantitative prețurile biletelor

Prima mulțime de date conține nume de orașe ⇒ date calitative.

Prețurile sunt înregistrări numerice ⇒ date cantitative.

Exercițiul 1:

În tabelul de mai jos sunt prezentate prețurile de bază pentru-

Model	Preț
Logan	6950
Logan MCV	7900
Logan MCV Stepway	11.950
Sander	7150
Sander Stepway	7150
Duster	10900

Identificați datele calitative și datele cantitative. Justificați.

• Nivelul de măsurare (scala de măsurare)

- în scala nominală (categorică):
 - date nominale: nume, etichete, calități ⇒ doar date calitative ⇒ ~~operații matematice~~
 - ↳ în cazul datelor cantitative
 - date ordinale: pot fi ordonate dar diferențele dintre intervale nu-s semnificative. Pot fi atât date calitative cât și cantitative
 - ↳ în scala interval (cardinal)
 - date interval: pot fi ordonate, pot fi calculate diferențe semnificative
 - ↳ interarea multă reprezentată doar o poziție pe axă, nu semnifică zero absolut
 - ↳ sunt doar date cantitative
 - ↳ în scala de raport (proportională) (ratiaonală)
 - date rationale: ca la date interval dar zero este 0 absolut. Putem vorbi de raportul datelor ⇒ o dată poate fi multiplul alteia.

Exemplul 2: Clasificări următoarelor date în funcție de nivelul de măsurare:

a)

Top 10 seriale comedie (IMDB)	
1.	Rick and Morty (2013-)
2.	Suits (2011-)
3.	Shameless (2011-)
4.	Friends (1994-2004)
5.	The Tick (2017-)
6.	Ten Wolf (2011-)
7.	Orange Is The New Black (2013-)
8.	The Big Bang Theory (2007-)
9.	Modern Family (2009-)
10.	The Office (2005-2013)

b)

Președinții României post-revoluționari aparținând politicii	
Ion Iliescu (NPSD)	
Emil Constantinescu (PNL)	
Traian Băsescu (PS)	
Klaus Iohannis (PNL)	

c)

Real Madrid - Cupe Europene	
1956, 1957, 1958, 1959	
1960, 1966, 1998, 2000,	
2002, 2014, 2016, 2017.	

d)

Campionate europene Câștigate	
Real Madrid	12
AC Milan	7
Bayern Munich	5
Barcelona	5
Liverpool	5
Ajax	4
Internazionale	3
Man. United	3

soluție:

- a) - clasificări seriale de comedie și abia datele sunt ordinale: reprezentate fiind de n. de ordine pot fi ordonate. Diferența dintre rangul 1 și rangul 5 însă în cre semnificativă matematic.
- b) - enumeră alfabetică politică a tuturor președinților înainte de alegeri ⇒ date nominale.
- c) - Are sens să vorbim că între primul și ultimul titlu au trecut 2017-1956=61 ani dar raportul 2017/1956 nu are sens ⇒ date interval
- d) - R.M. are cu 100% mai multe titluri ca FC Barcelona ⇒ date rationale.

Exercițiul 2: Clasificați următoarele date în funcție de nivelul de măsurătoare:

- Clasamentul final în campionatul național de fotbal.
- Agenda de telefon.
- Temperatura corpului a unui atlet în timpul antrenamentelor.
- Pulsul în bpm al unui atlet — — — — —

Operații posibile în funcție de nivelul de măsurătoare.

Nivelul de măsurătoare	pot fi categorisite	pot fi ordonate	databele se pot nota	Pot determina dacă o valoare e multiplul altuia
Nominal	Da	Nu	Nu	Nu
Ordinal	Da	Da	Nu	Nu
Interval	Da	Da	Da	Nu
Rational	Da	Da	Da	Da

Rețineri:

Nivelul de măsurătoare	Exemplu de m. de date.	Operații posibile:
Nominal	Concursuri de excelență la mela: - ONM de mela, România, Est. - Adolf Haimovici - Olimpiada pt juniori - Olimpiada pt fete.	Așezare pe categorii
Ordinal	Restricții TV 16+ - audi video general 12+ - audi video parental 14+ - restricționat 18+ - interzis minorilor	Așezare pe categorii dar n ^o în ordine: 18+ → cea mai restricționată
Interval	Temperaturile maxime medii lunare în București, [°C] ian 5° feb 7 mar 12 apr 17 mai 24 iun 27 iul 30 aug 31 sep 27 oct 20 nov 10 dec 5	Așezare pe categorii ordonare diferențe cu semnificație: 7-5=2°C ⇒ feb este cu 2°C mai cald decât ian
Rational	Alte de precipitații medii Ro ian 30 feb 41 mar 49 apr 78 mai 50 iun 87 iul 50 aug 98 sep 24 oct 17 nov 31 dec 22	Așezare pe categorii ordonare diferențe semnificative raport la semnificație $\frac{98}{50} \approx 2 \Rightarrow$ aug de 2 ori mai plouos decât ian

Exercice:

1. La ce nivel de numărare putem întâlni date calitative? Dar cantitative?

Stabilim valoarea de adevăr a propozițiilor:

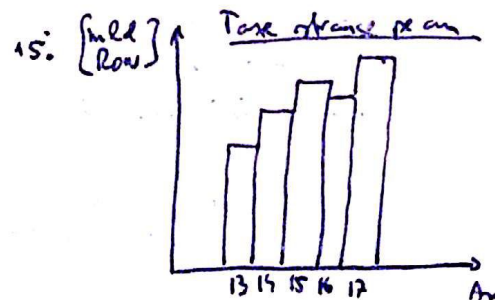
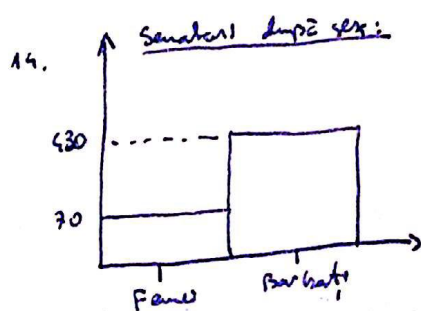
2. La nivel nominal se întâlnesc doar date calitative
3. La nivel ordinal datele sunt doar cantitative
4. La nivel rațional se pot calcula diferențe semnificative
5. La nivel interval se pot calcula diferențe semnificative.

Clasificăm următoarele date după tip:

6. Numerele de telefon din agenda
7. Maximale zilnice din mai.
8. Numerele jucătorilor echipei Steaua.
9. Numărul golurilor înscrise de jucătorii echipei Steaua.

Clasificăm următoarele date după nivel:

10. Cele mai importante turnee de tenis: Australian Open, French Open, The Lawn Tennis Championship (Wimbledon), U.S. Open
11. Nr de pești (la mie) prins de fiecare pescar:
36, 14, 47, 87, 23, 11, 17, 21, 112, 255, 385
12. Lungimea în țesut a peștelor dintr-un esanțon pescuit în țesut:
15, 27, 31, 15, 12, 12, 35, 30, 14
13. Departamentul de vânzări al unei companii este împărțit între:
București, Cluj, Iași, Timișoara, Pitești.



Dezvoltări

16. Care-s desigurile principale între nivelul interval respectiv rațional de numărare a datelor? Justificati cu exemple concrete.

Protecerea unui studiu statistic

Scopul unui studiu statistic este să colecteze date și să le folosească ulterior în luarea unor decizii. Dacă datele sunt culese defectuos concluziile ar putea fi eronate (cel mai probabil va fi iar studiul va fi compromis)

Când trebuie interpretate rezultatele unui studiu statistic este necesar să restituiem datele compromise pentru a obține concluzii corecte. Pentru aceasta este necesar să înțelegem inițial cum se protejează un studiu statistic.

Pași pentru protecerea unui studiu statistic:

1. Se identifică • variabilele de interes (focusul studiului)
• populația / esanșionul pe care se va face studiul
2. Se dezvoltă modalitatea de culegere a datelor. Trebuie să ne asigurăm că datele -s reprezentative pentru populație
3. Se colectează datele
4. Se descriu / centralizează datele folosind statistica descriptivă
5. Se trag concluzii / iau decizii folosind statistica inferențială
6. Se identifică posibilele erori.

Colectarea datelor:

Există mai multe metode de colectare a datelor. Alegerea metodei se face în funcție de caracteristicile datelor de colectat.

De exemplu, datele pot fi colectate:

- prin efectuarea unui experiment
- în urma unei simulări
- prin recensământ
- prin sondaje

Exemplul 1

Pentru studiile statistice enunțate mai jos indicați o metodă de alegere a datelor adecvată. Justificați răspunsul. Se studiază:

- Efectul impactului unui asteroid cu Pământul
- Efectul medicamentului homeopatc *onco coccinum* în dureri imposter va răceală
- înălțimea abacombilor eclipselor din divizia A de fotbal
- satisfacția cetățenilor față de rețeaua rutieră națională.

Soluție

- Acesta studiu nu poate fi creat. Dacă nu ar putea crea ar fi oricum foarte periculoasă prin urmare se folosesc o simulare.
- Va trebui verificat efectul asupra pacienților prin urmare se va efectua un experiment. Este foarte important să se folosească două grupuri similare: primul grup va primi medicamentul sau al doilea grup numit grup de control va primi un placebo. E important ca nici asistenții ce interacționează cu pacienții să nu știe căruia grup aparține fiecare pacient.
- Întrucât datele sunt disponibile înregistrate în parte se va organiza un recensământ.
- Este aproape imposibil să interogăm fiecare cetățean al populației așa că se va efectua un sondaj.

Exercițiul 1:

Indicați care o metodă adecvată de alegere a datelor pentru:

- Studiul efectului consumului în minimum 2 l de apă pe zi asupra elasticității pielii
- Studiul efectului exploziilor solare în perioade de maximă activitate solară în România.
- Studiul efectului defrișării munților Apuseni asupra variației prețului mulsului produs de IUGA
- Studiul variației înălțimii cu vârsta în rândul elevilor din CNGP.

Modalități de alegere a esanțonului pentru sondaj

Reamintim că datele alterate duc la concluzii indolabile prin urmare trebuie depuse eforturi pentru a obține date reprezentative. Pentru aceasta este foarte important să alegem corect metoda în esanțion din populație. Aceasta se poate realiza în mai multe moduri:

1. Esanțion aleator ^{simple (randomizat)}: Fiecărui membru al populației se atribuie un număr cu șanse egale de a fi selectat, urmând a fi alese aleator atâtea numere câți membri trebuie pentru esanțion

Exemplu 1: într-un liceu sunt 800 de elevi. Dăm să formăm un esanțion de 30 de elevi pentru a răspunde întrebărilor unui chestionar. Esanționul va fi aleator simplu

realiză:
Fiecărui elev se atribuie un nr. de la 1 la 800. Generăm aleator 30 de numere de la 1 la 800 și astfel obținem esanționul

2. Esanțion stratificat: membrii populației sunt repartizați în grupuri cu caracteristici comune: gen, vârstă, etnie după care sunt alese esanțioane randomizate din fiecare grup. Astfel se asigură prezența în esanțion a reprezentanților din fiecare strat (strat) al populației.

3. Esanțion cluster: formăm ^{unele mai multe} două sau mai multe grupuri naturale în populație

4. Esanțion sistemic: fiecare a k-lea membru al populației este ales
 $k = | \text{populație} | / | \text{esanțion} |$

5. Esanțion de conveniență: formăm din persoanele disponibile. Se de obicei conduce la studii cu concluzii eronate.

Exemplu 3: Darim să determinăm preferința pentru cola / apă / alte băuturi ale locuitorilor din Dâmbovița. Pentru această tehnică să alegem un esantion pentru completarea unui chestionar. Vom exemplifica diverse tehnici de alegere a esantionului:

a) Pentru a alege un esantion aleator putem să trimitem chestionare spre completare locuitorilor județului (cu scris de buletin DS) generând aleator numere de pașe cișe pentru numărul c.i. Unele numere nu corespund nici unei persoane și trebuie eliminate dar problema va fi aflarea adreselor. Pentru aceasta este necesară baza de date a evidenței populației. O altă problemă ar constitui-o locuitorii analfabeți, incapabili să completeze singuri un chestionar.

Altă variantă ar fi să abținem fiecare locuință ca un număr de ordine, să alegem aleator câteva numere câte elemente are esantionul și din fiecare gospodărie să notăm preferința unui membru ales aleator. Aceasta implică un efort financiar mare pt personal.

b) Pentru a alege un esantion stratificat putem împărți populația pe niveluri socio-economice și să alegem aleator indivizi din fiecare nivel

c) Pentru un esantion cluster putem împărți locuitorii județului în grupuri în funcție de codul postal. Ulterior să alegem aleator câteva coduri postale (un loc) și să interogăm locuitorii corresp.

d) Pentru un esantion sistematic putem alege fiecare al o mielea număr de buletin sistematic și să interogăm posesarii.

e) Pentru un esantion de conveniență putem efectua sondajul la raionul de băuturi să căutăm dintr-un supermarket local.

Exemplul 4:

Intr-un liceu se efectuează un studiu privind opinia elevilor referitor la abordarea orilor de matematică din partea profesorilor. În funcție de esanționul ales, prezentați tipul acestuia:

- a) s-a selectat o clasă la întâmplare și fiecare elev a fost interogat
- b) elevii au fost împărțiți în grupuri în funcție de specializare și din fiecare grup au fost interogați elevi aleși aleator.
- c) fiecărui elev i s-a atribuit un număr și au fost generate aleator numere de ordine.

soluție:

- a) clasele reprezintă grupuri formate natural prin urmare acesta este un esanțion cluster
- b) esanțion stratificat deoarece elevii au fost împărțiți în grupuri cu caracteristici comune.
- c) grup aleator simplu deoarece elevii au șanse egale de a fi aleși.

Exercițiul 4:

Se efectuează studiul din exemplul 4 însă esanționul este format astfel:

- a) sunt selectați elevii înscriși la cursul de statistică
- b) fiecărui elev i se atribuie un număr de ordine, se alege un prim număr și fiecare al trezeilea elev este interogat
- c) sunt selectați elevii claselor de filologie.

Exersare:

Stabiliti valoarea de adevar a propozitiilor:

1. Folosirea unui esantion sistematic garanteaza interogarea unora membri din fiecare grup socio-economic al populatiei.

2. Pentru a alege un esantion cluster populatia este impartita in grupuri si sunt interogati toti membrii din cel putina un grup (dar nu din toate)

Alegeti tipul esantionului pentru fiecare din studiile de mai jos:

3. Se studiază efectul lamivudinului asupra bolnavilor infectati cu HIV

4. Se studiază profitabilitatea producției în cantități mari spre distribuție a unui nou medicament asupra bolnavilor infectati cu HIV

5. Se studiază viteza răspândirii unui virus nou în mediul urban.

6. Se studiază salarii la ministrii la noul guvern.

Identificati tipul esantionului in fiecare din cazurile de mai jos:

7. Folosind o centrală telefonică de apelare aleatoare au fost sunati 1000 de cetățeni români și intervievați în legatură cu marca de mașină condusă

8. Elevii au fost interogati la intrarea din bibliotecă despre obiceiul de a ieși în cluburi.

9. Profesorii au fost interogati la intrarea din cancelarie despre obiceiul de a folosi metode medicale în predare

10. Un nou soi de porumb a fost plantat. Câmpul a fost împărțit în zone de câte un morgen și din fiecare zonă au fost recoltate probe în vederea estimării recoltei.

11. După o furtună membrii fiecărei a 50-a gospodării din zonă afectată au fost interogati pentru a evalua necesitățile de ajutorare.

12. Pentru controlul calității fiecărui al 10-lea fier de calcat produs este testat.

Stabiliti dacă întrebările sunt potrivite pentru un sondaj. Cum ați reformula?

13. De ce este rău că consumul de mașini naturale?

14. De ce sunt de câteva ori mai periculoși șoferii ce schimbă benzile?

Apofundare.

15. Enumerati avantajele și dezavantajele întrebărilor deschise respectiv închise.

16. Enumerati av. & dezav. unui sondaj în care membrii populației trebuie să apeleze un membru de la pt. a răspunde la întrebări. Ce tip de esantion?

Ce am învățat în Capitolul 1?

- 1.1. Descrierea datelor populației și esanțion (Ex 1-4)
- 1.1. Descrierea datelor parametrilor și statistice. (Ex 5-8)
- 1.2. Descrierea datelor calitative și date cantitative
- 1.2. Clasificarea datelor în funcție de nivelul de măsurătoare (nominal, ordinal, interval, rațional)
- 1.3. Moduri de colectare a datelor (experiment, simulare, recensământ, sondaj)
- 1.3. Moduri de a forma un esanțion (aleator simplu, stratificat, cluster, sistematic, de conveniență)

De ce am învățat? Folosire & ponoare?

Folosire:

Statistica ajută în luarea unor decizii informate. Cunoștințele despre statistică și protecția sondajelor permit identificarea erorilor în studii existente. Un studiu protejat necesită o investiție poate fi înșelător și poate abuză resurse financiare. Este important să deosebim tipurile distincte de date și să obținem dacă putem efectua calcule matematice semnificative.

Ponoare:

Cel mai mare pericol în statistică este reprezentat de folosirea esanționelor contaminate. Vom învăța că dacă un esanțion este nefișent de numeros și aleator va fi reprezentativ pentru populația pe care o reprezintă. De multe ori însă este oneros să alegem un esanțion nefișent de mare sau aleator. De exemplu când se efectuează un sondaj apelând aleator numere de telefon acesta NU va fi reprezentativ pentru partea din populație

- cu telefonul nelistat
- ce muncă în timpul zilei când este efectuat sondajul
- ce muncă să răspundă interviurilor

In exercitiile 1-4 identificați populația și esanșionul:

1. Un studiu efectuat de Vanderclaud, pe 2392 adulți a devaluit că:
 - 27% au identificat "gigabyte" drept "o insectă din Sud America"
 - 42% cred că "mother board" este "puntea unui vas de croazieră"
 - 23% cred că "MP3" e un Robot din Star Wars
 - 18% — "Blue Ray" e un animal marin
 - 15% — "Software" este "imprăcâmintele confortabile"
 - 12% — "USB" este un avion pt. tură din Europa
 - 11% — "HTML" este o boală cu transmitere sexuală.
2. Un studiu efectuat în 2013 de Public Religion Research Institute prin apelarea telefonică aleatoare a 1033 de persoane a devaluit că 27% din americani cred că Dumnezeu are un rol important în determinarea câștigătorului unui meci.
3. Un sondaj din 2012 efectuat pe 2200 de cetățeni ai US arată că:
 - 26% din americani cred că Soarele se învârt în jurul Pământ
 - 50% din americani cred că antibioticele sunt eficiente împ. virusurilor
4. Un sondaj din 2014 efectuat online arată că 54% din americani se consideră mai inteligenți ca americanul obișnuit.

In exercitiile 5-8 decideți dacă valoarea numerică este parametru sau statistică

5. Psihologul David Buss a efectuat un studiu pe studenții Universității din Texas, Austin și a aflat că 75% dintre aceștia erau gândiți la a comite o crimă.
6. Psihologul David Buss a efectuat un sondaj pe 5000 de americani și a aflat că:
 - 91% dintre bărbați n-au gândit cum ar fi să omoreze pe cineva
 - 84% dintre femei —
7. Într-un studiu efectuat pe 3000 tineri britanici arată că:
 - 58% cred că Sherlock Holmes a existat.
 - 47% cred că Richard the Lionheart e un personaj fictiv
 - (77% au declarat că nu citesc cărți istorice iar 61% scriu pe post de început un documentar istoric)
8. Studiul Eylert efectuat pe copii cu vârste între 5 și 13 ani ai mentorialelor cu soldați americani după WWII a arătat că IQ-ul copiilor albi și al celor negri este similar. Concluzia: IQ-ul depinde de gene și de mediul social.

In exercitiile 9-12 stabiliți care date sunt cel mai ușor și care-s
cum trebuie:

9. Salariile angajaților dintr-o firmă de protecție
10. Nr. de cont ale angajaților dintr-o firmă de protecție
11. Vârstele angajaților unei firme
12. Codurile poștale ale angajaților firmei.

In exercitiile 13-16 identificați nivelul de măsurătoare al datelor

13. Tipurile de pești din acvarii din câmin: gușt, namtar, neoni.
14. Lungimile peștelor — — — — : 21 cm, 12 cm, 13 cm, 23 cm, 11 cm
15. Temperaturile înregistrate în timpul zilei de nămățat pentru
apa din acvariu: 19° 22° 23° 15° 21°
16. Pești pentru acvariu sunt de 3 categorii: 1. foarte colorați;
2. colorați; 3. lipsiți de culoare.

In exercitiile 17-20 alegeți o metodă de colectare a datelor potrivite
pentru:

17. Un studiu despre activitățile de voluntariat ale elevilor din CNES
18. Un studiu despre activitățile de voluntariat ale elevilor din România
19. Un studiu despre impactul activităților de voluntariat asupra
împăduririi.
20. Un studiu despre influența aerului de munte asupra elevilor
cu probleme respiratorii.

In exercitiile 21-24 identificați modul de alegere al esantionului:

21. Un studiu despre mișcarea la adolescenți a rețetelor de
10. creșterea din care au fost intervievați toți adolescenții
22. Polițiștii din Buftea au măsurat viteza tuturor șoferilor
de pe principală în ziua de duminică.
23. Un jurnalist interviuează primii 100 de elevi ce ies de la
examenul de BAC
24. Câte 15 elevi au fost selectați din fiecare clasă dintr-un
colegiu pentru a participa dorințele pentru programul în paștel.

Text

Grundelemente der Statistik

Aufgabe 1: In einer Klasse von 30 Schülern wurden die Anzahl der richtigen Antworten bei einer Prüfung mit 10 Fragen erfasst. Die Ergebnisse sind in der folgenden Tabelle dargestellt:

0	0	1	2	3	4	5	6	7	8	9	10
4	4	5	5	5	5	6	6	6	7	7	7
8	8	10	10	10	10	10	10	10	10	10	10
12	12	12	12	12	12	12	12	12	12	12	12
16	16	16	16	16	16	16	16	16	16	16	16
20	20	20	20	20	20	20	20	20	20	20	20
24	24	24	24	24	24	24	24	24	24	24	24
28	28	28	28	28	28	28	28	28	28	28	28
32	32	32	32	32	32	32	32	32	32	32	32
36	36	36	36	36	36	36	36	36	36	36	36

Wsklichkeit $P(X=x)$ für $x=0,1,2,3,4,5,6,7,8,9,10$:

Klasse	Häufigkeit
0 - 10	30
10 - 20	15
20 - 30	12
30 - 40	12
40 - 50	6

Das ergibt ein Diagramm (Histogramm):



Die mittlere Anzahl der richtigen Antworten ist:

$$\text{Mittelwert} = \frac{0 \cdot 4 + 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 5 + 4 \cdot 5 + 5 \cdot 6 + 6 \cdot 6 + 7 \cdot 6 + 8 \cdot 7 + 9 \cdot 7 + 10 \cdot 7}{30} = \frac{282}{30} = 9,4$$

Distribuția frecvențelor. Reprezentări grafice

Când lucrăm cu date numerice este dificil să observăm modelul. Este util să grupăm datele în intervale (numite clase) și să numărăm datele din fiecare interval.

definiție: frecvența unei clase (a unui interval) este nr. datelor din acel interval. Distribuția frecvențelor este tabelul în care notăm frecvențele corespunzătoare fiecărui interval.

Pentru a construi distribuția frecvențelor unei mulțimi de date:

1. stabilim numărul claselor. De obicei alegem între 5 și 20.

(altfel ar deveni dificil să observăm un model)

Putem folosi relația lui Sturges: $k = 1 + 3,322 \ln n$

2. determinăm lățimea clasei:

$$\text{Lățimea clasei} = \frac{\text{val. maximă} - \text{val. minimă}}{\text{nr. clasele}}$$

3. determinăm limitele claselor:

Limita inferioară a unei clase = valoarea minimă pe care o pot lua datele din acea clasă

Limita superioară a unei clase = valoarea maximă pe care o pot lua datele din acea clasă.

de ex. limita inf. a primei clase este data minimă din mulțime

limita inf. a celei de-a doua clase = limita inf. + lățimea clasei

4. determinăm frecvența fiecărei clase numărând datele cu valori între limita inferioară și cea superioară.

∇ Este recomandat ca toate clasele să aibă o lungime comună

Exemplul 1: S-au notat perioadele necesare efectuării temei pentru cei 35 elevi ai unei clase. În minute, acestea au fost:

18 43 42 35 57 45 51
 19 25 23 27 36 35 35
 40 41 53 51 57 41 40
 35 15 19 51 52 50 45
 46 45 39 48 63 12 59

Să se realizeze o distribuție a frecvențelor cu șapte clase.

Soluție:

- 7 clase
- lățimea clasei = $\frac{63 - 12}{7} = \frac{51}{7} \approx 7,3$ rotunjim la 8
- limitele inferioare: 12, 20, 28, 36, 44, 52, 60
- Numărăm datele din fiecare clasă:

Clasă	Număratoare	Frecvență
12-19		5
20-27		2
28-35		5
36-43	IIII	8
44-51	IIII II	9
52-59		5
60-67		1

$\sum f_i = 35 \checkmark$ ← VERIFICATI!

- Atenție:
- Este util să împărțim de la început în grupe de 5: ||||
 - Verificati dacă suma frecvențelor coincide cu numărul datelor.

Exemplul 1: Realizati o distribuție a frecvențelor pentru vârstele conștăeniilor lui Andrei.

Pentru a înțelege mai bine tendința datelor vom introduce câteva noțiuni noi:

definiție:

- mediana unei clase este mediana aritmetică a limitelor inf. resp. sup.
- frecvență relativă a unei clase = $\frac{\text{frecvența clasei}}{\text{cardinalul eșantionului}}$
- frecvență cumulată a unei clase este suma frecvențelor clasei respective cu frecvențele tuturor claselor inferioare.
frecvență cumulată a ultimei clase este cardinalul eșantionului.

Observație cu excepția primei clase, mediana fiecărei clase se poate obține adunând lățimea clasei la mediana clasei precedente.

Exemplul 2: Găsiți medianele, frecvențele relative și frecvențele cumulate pentru fiecare clasă din distribuția de frecvențe din exemplul 1. Ce observați?

rezultat:

Clasă	f	mediana	frecvență relativă	frecvență cumulată
12-15	5	15,5	$\frac{5}{35} = \frac{1}{7}$	5
20-27	2	23,5	$\frac{2}{35}$	5+2=7
28-35	5	31,5	$\frac{1}{7}$	7+5=12
36-43	8	39,5	$\frac{8}{35}$	20
44-51	9	47,5	$\frac{9}{35}$	29
52-59	5	55,5	$\frac{1}{7}$	34
60-67	1	63,5	$\frac{1}{35}$	35

$$\sum f = 35$$

Se observă că cei mai mulți elevi au avut nevoie între 52 și 59 pt. lucr...

Exercițiul 2: Găsiți medianele, frecvențele relative și frecvențele cumulate pt. fiecare clasă din distribuția de frecvențe din exercițiul 1. Ce observați?

Reprezentarea distribuției frecvențelor

Reprezentarea grafică prin bare a distribuției de frecvențe n.n. histograma de frecvențe. Proprietăți ale histogrammei:

1. scala orizontală este cantitativă (cubine valori la detaliu)
2. scala verticală măsoară frecvența claselor.
3. Barele consecutive se ating.

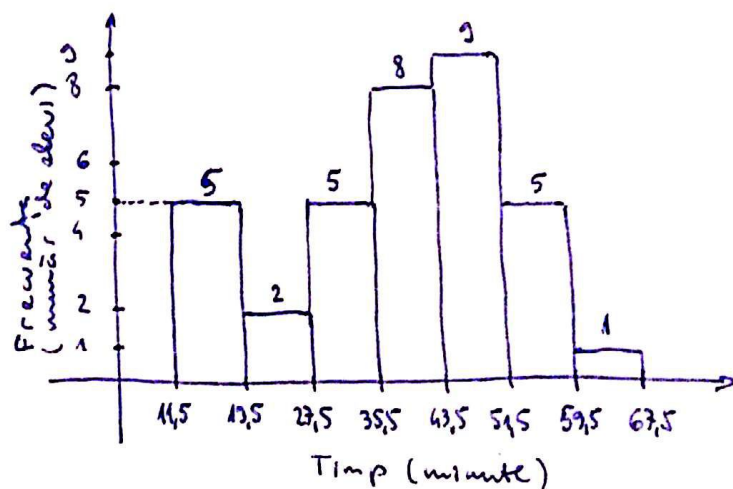
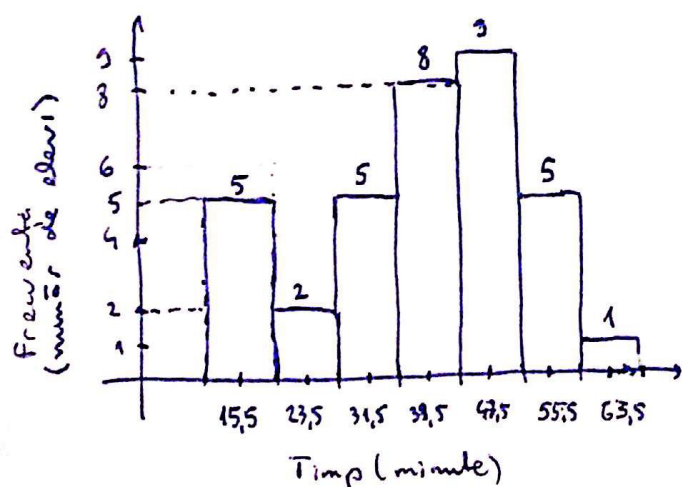
Dacă barele ar începe cu limitele inferioare și n-ar termina cu limitele superioare ale claselor nu n-ar mai atinge. La mijlocul distanței dintre limita superioară a unei clase și limita inferioară a următoarei vom considera frontiera clasei ce reprezintă reprezentările a două clase consecutive. Pe scala orizontală vom marca fie mediana frecinței clase fie frontiera sale.

Exemplul 3:

Vom reprezenta histograma distribuției de frecvențe din Exemplul 2. Care vom observa tendința de comportament noi?

soluție:

Clasă	frontiere	frecvențe
12-19	11,5 - 19,5	5
20-27	19,5 - 27,5	2
28-35	27,5 - 35,5	5
36-43	35,5 - 43,5	8
44-51	43,5 - 51,5	9
52-59	51,5 - 59,5	5
60-67	59,5 - 67,5	1



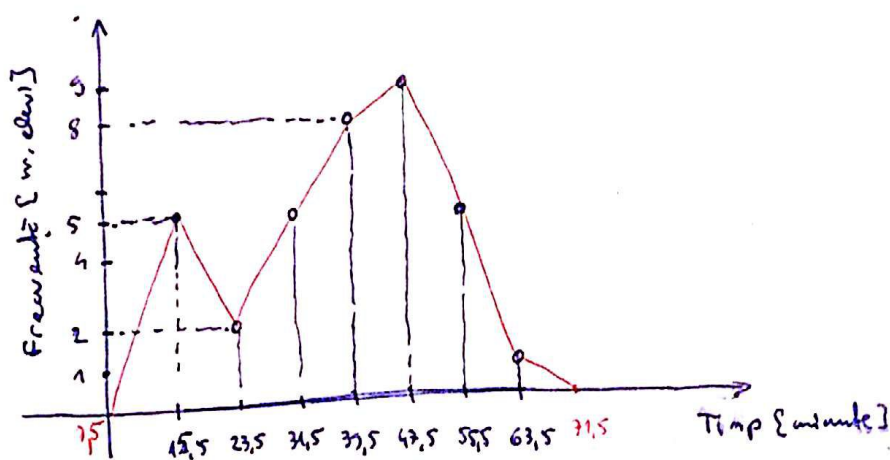
Observăm cu ușurință că majoritatea elevilor au alocat între 35,5 și 51,5 minute timp

Exercițiul 3: Folosiți distribuția de frecvențe din exemplul 1 pentru a trasa histograma ce reprezintă vârstele consilierilor băușilor lui Andrei. Ce tendințe observați?

definiție: Altă modalitate de a reprezenta o distribuție de frecvențe este prin linia frântă ce unește punctele de pe histogramă corespunzătoare medianelor claselor. Aceasta este poligonul de frecvențe și reprezintă variabilă frecvențelor cu datele.

Exemplul 4 Trasați poligonul de frecvențe corespunzător distribuției din exemplul 3.

soluție:



Pași pt. trasarea poligonului de frecvențe:

- folosim cretă din histogramă
- reprezentăm punctele de coordonate (mediană, frecvență clasă)
- extindem linia poligonului cu o linie de clasă spre st / din lăta de prima / ultima mediană

De obicei trăsăm histograma și poligonul de frecvențe împreună.

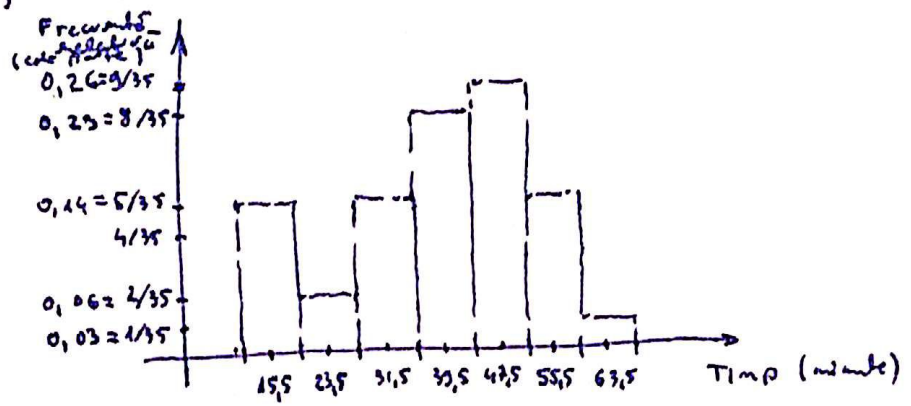
Exercițiul 4: Trasați poligonul de frecvențe pt a reprezenta vârstele părinților din satul lui Andrei.

definiție histograma frecvențelor relative se obține reprezentând prin bare verticale unite variatia frecvențelor relative pe clase de frecvențe.

Histograma frecvențelor relative are aceeași formă cu histograma frecvențelor (absolute) iar în unele convenții alare pe axa verticală histogramale sunt identice

Exemplul 5 Desenăm histograma frecvențelor relative pentru datele din exemplul 2

interval	Clasa	frecvența relativă
12-19	12-19	5/35
20-27	20-27	2/35
28-35	28-35	5/35
36-43	36-43	8/35
44-51	44-51	7/35
52-59	52-59	5/35
60-67	60-67	1/35

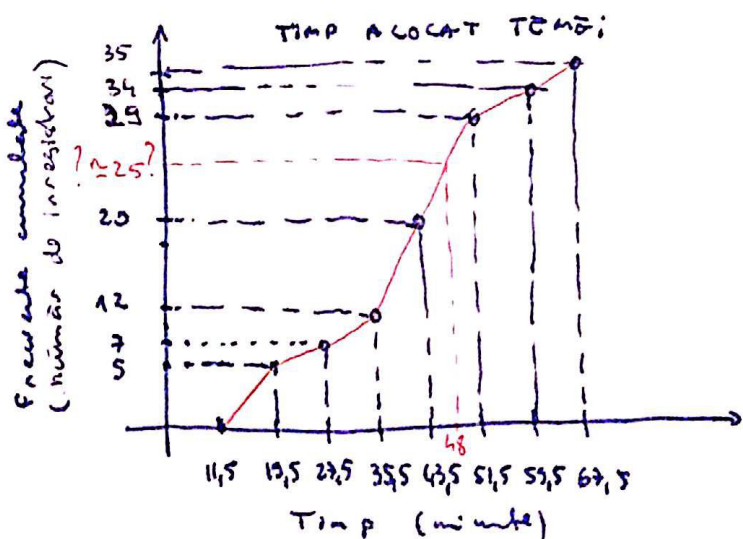


Exercițiul 6: Reprezentăm histograma frecvențelor relative pt vârstele copiilor din satul lui Andrei.

definiție Ogiva (curba frecvențelor cumulate) este o linie poligonală ce reprezintă frecvența cumulată a fiecărei clase în punctul frontierei superioare. Punctele de frontieră superioară se reprezintă pe axa orizontală iar frecvențele cumulate pe axa verticală.

Exemplul 6 Trasăm ogiva pt. distribuția de frecvențe din Exemplul 2. Estimăm câți elevi acordă mai puțin de 48 minute rezolvării temei.

Clasa	frontiere	f_i	frecvențe cumulate
12-19	11,5-19,5	5	5
20-27	19,5-27,5	2	7
28-35	27,5-35,5	5	12
36-43	35,5-43,5	8	20
44-51	43,5-51,5	7	27
52-59	51,5-59,5	5	32
60-67	59,5-67,5	1	33



Atenție:

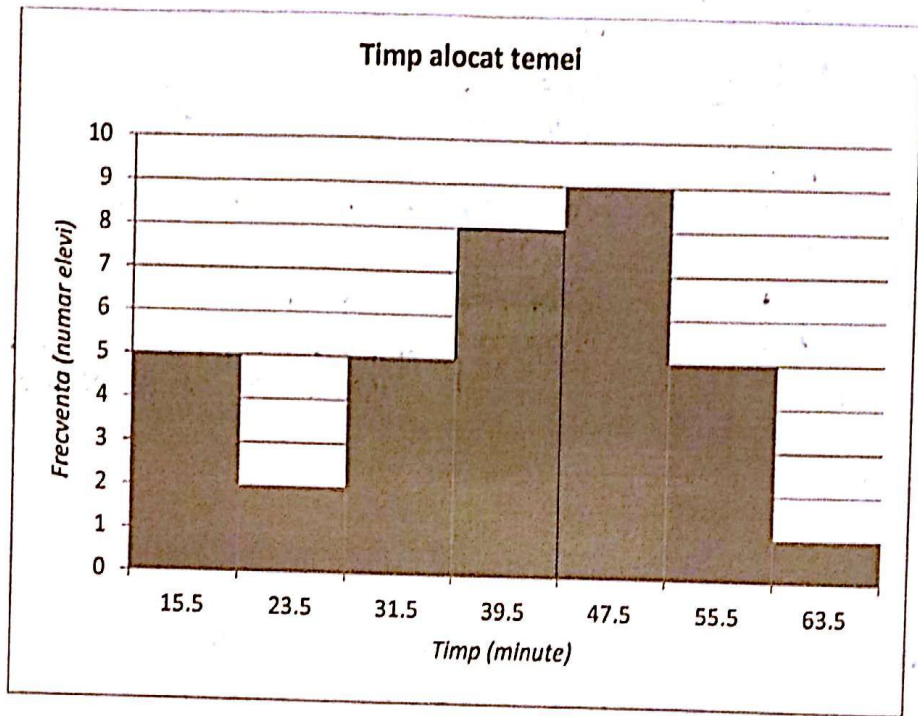
- primul punct al ogivei se află pe axa orizontală și are abscisa frontierei inferioare a primei clase
- ultimul punct al ogivei are abscisa frontierei superioare a ultimei clase și ordonata cardinalului esanțonului

Din ogivă aproximăm că 25 de elevi au acordat mai puțin de 48 de minute rezolvării temei. Cea mai mare creștere a nr. de elevi se observă în clase 35,5' - 43,5' (atci punct e năruimă!)

Exemplul 6: Trasați oțiva pentru tabelul lui Andrei (exemplul 2.)

Exemplul 7: Folosiți calculatorul pentru a trasa histograma pentru frecvențele de distribuție din Exemplul 2.

Soluție: am folosit Microsoft Excel și am obținut:



Exemplul 7: Folosiți calculatorul pentru a trasa histograma pentru frecvențele din Exemplul 2.

Exersare:

1. Care-s avantajele folosirii distributiei de frecvente? Dan al reprezentarii acestora?

Statistica valorarea de adevăr a propozițiilor:

2. Media clasei este rădăcina pătrată a produsului limitelor inferioare respectiv superioare

3. Frecvența relativă a unei clase este raportul frecvenței clasei cu cardinalul esanionului.

En exercitiile de mai jos identificați din mansimea clasei de frecvențe, frontierele claselor și medianele. Calculați frecvența cumulată și frecința clasei.

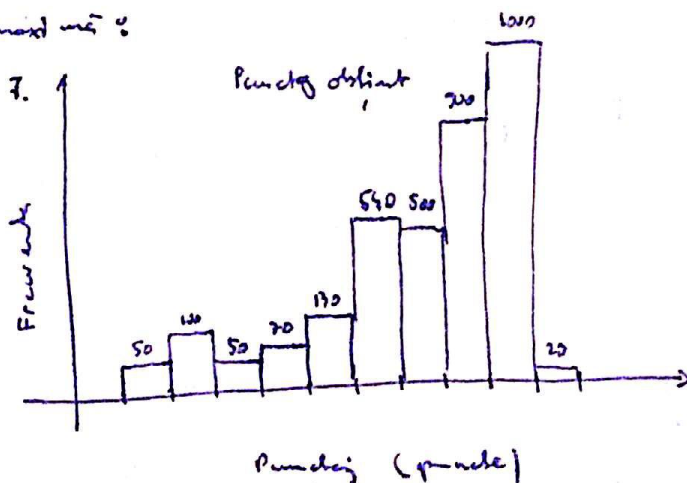
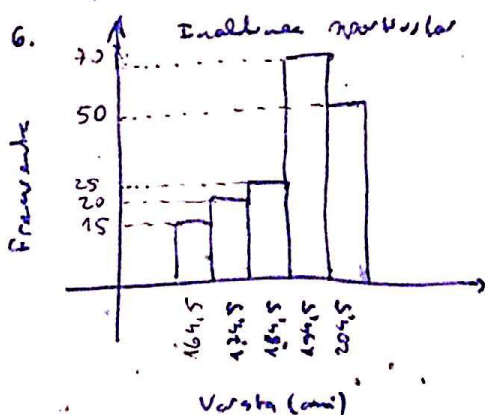
4. Înălțimea sportivilor [cm]

clasa	Frecvența
160-169	15
170-179	20
180-189	25
190-199	70
200-209	50

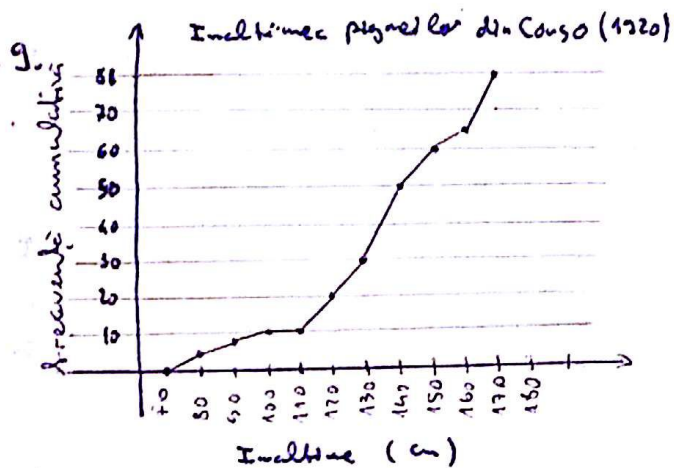
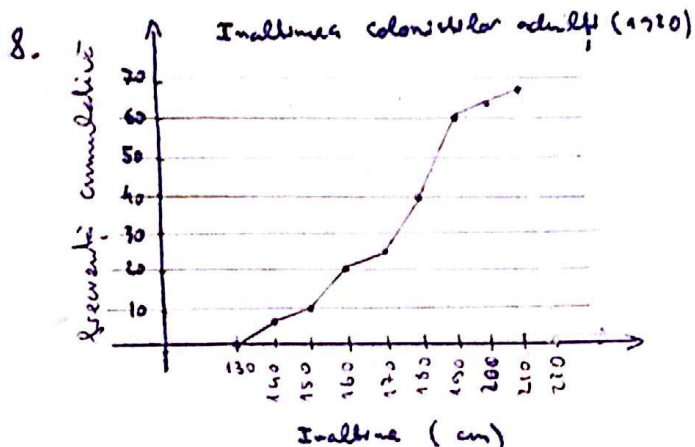
5. Punctajul obținut

clasa	Frecvența
100-199	50
200-299	100
300-399	50
400-499	70
500-599	130
600-699	540
700-799	500
800-899	900
900-999	1000
1000-1099	20

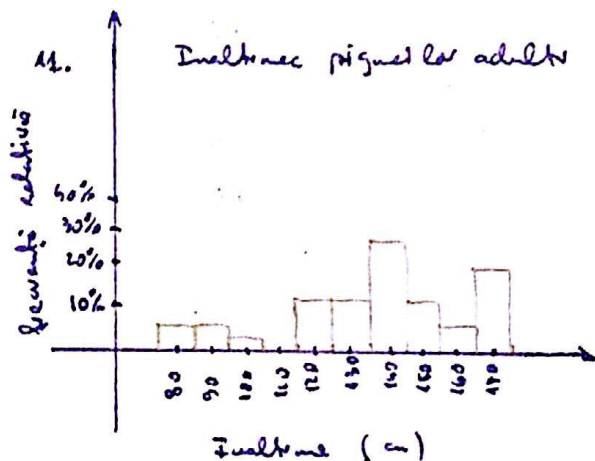
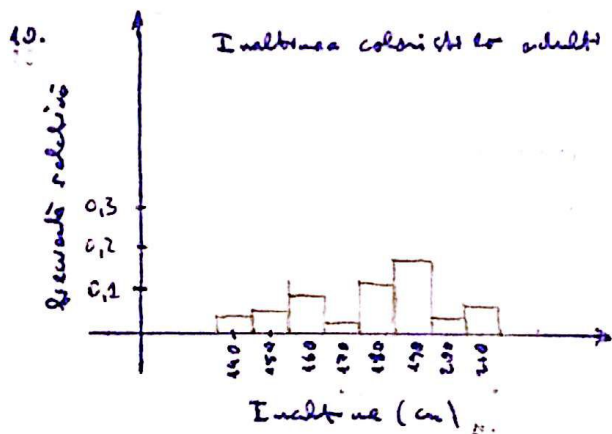
Citiți histograma de frecvențe pentru a identifica cel puțin unele clasele cu frecvențe minime respectiv maxime:



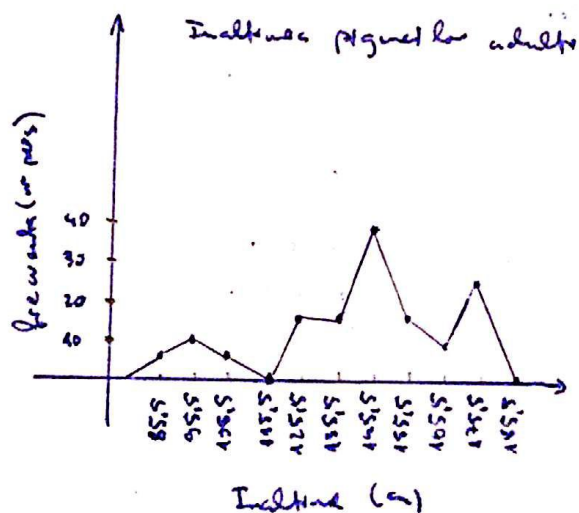
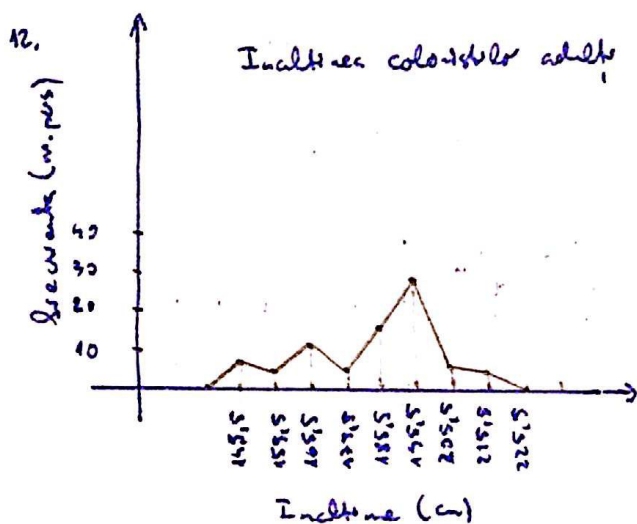
- Citeste ogivella de mai jos si identificați:
- a) coordonatul esantionului;
 - b) venăarul claselor;
 - c) clasa cu cea mai mare crestere in inaltime;
 - d) frecventa cumulata pentru inaltimea de 1,45 m;



- In exercitiile urmatoare foloseste histogramele de frecventa relativa pentru a gasi:
- a) clasele cu frecventele relative minime/maxime;
 - b) frecventele relative minime/maxime.



- In exercitiile de mai jos foloseste poligonul de frecventa pentru a identifica clasele cu frecventele minime/maxime.



Reprezentarea ramii frunze

Similare histogramei, reprezentarea ramii frunze păstrează datele inițiale.
A fost concepută de John Tukey în 1977 ca o formă facilă de a.
surta datele.

Exemplul 1: Putem reprezenta datele din Exemplul 1/21 într-o

ramificație neordonată:

1	89592
2	597
3	565559
4	32501105658
5	713171209
6	3

ramificație ordonată:

1	25899
2	579
3	555569
4	00112355568
5	01112379
6	3

De asemenea putem scrie frecvențe ram de două ori: pe primul rând
de la început și pe al II-lea 5,6,7,8,9 astfel:

ramificație neordonată

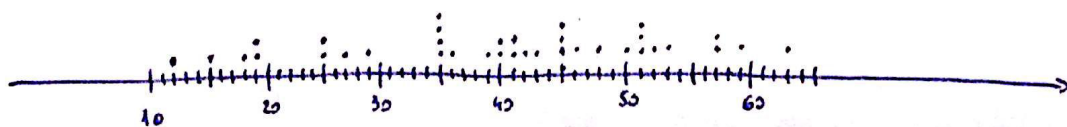
1	2
1	8959
2	5975
3	
3	565559
4	320110
4	55658
5	131120
5	779
6	3
6	

ramii frunze ordonate

1	2
1	5899
2	5579
3	
3	555569
4	001123
4	55568
5	011123
5	779
6	3
6	

Exercițiul 1: Reprezentați într-o diagramă ramii frunze variabilele
valabile din Exercițiul 1/21.

Exemplul 2: Reprezentați într-o diagramă cu puncte aceluși date ca în Exp. 1.



Exercițiul 2: Reprezentați prin puncte datele din Exemplul 1.

Reprezentarea grafică a esențelor de date calitative

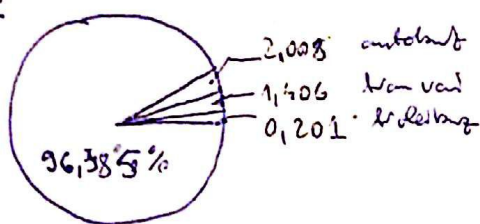
Diagrama de structură sau

Pie chart este foarte des întâlnită: pentru reprezentarea relației între părțile esanționului și chiar între acestea și întregul esanțion

Exemplul 5:

S-a estimat că zilnic în 2017 480.000 bucureșteni au mers pe metrou în timp ce aproximativ 18.000 folosesc mijloacele de transport public la dispoziția de RATB: 10.000 folosesc autobuzul, 7000 tramvai de iar 1000 troleibuzul. Organizați datele utilizând pie chart.

soluție:



$$\begin{array}{rcl}
 480k & \dots & 100 & \dots & 360^\circ \\
 480k & \dots & x = 96,385 & \dots & \alpha \approx 347^\circ \\
 10k & \dots & y = 2,008 & \dots & \beta \approx 7^\circ \\
 7k & \dots & z = 1,406 & \dots & \gamma \approx 5^\circ \\
 1k & \dots & t = 0,201 & \dots & \delta \approx 1^\circ
 \end{array}$$

Din diagramă deducem că foarte puțini călători folosesc RATB. De fapt nu e chiar atât de puțin că pare aproape ireal. De fapt, o concluzie plauzibilă ar fi că la metrou accesul este mai bine controlat în timp ce RATB-ul este utilizat gratuit.

Exercițiul 3:

Numărul zilnic al călătorilor pentru fiecare mijloc de transport în București, anul 2000 este reprezentat în tabelul de mai jos:

Mijloc de transport	Număr Călători (mii)
autobuz	30
tramvai	25
troleibuz	45
metrou	200

- Calculați frecvența relativă pentru fiecare dată.
- Calculați unghiul la centru corespunzător fiecărei date (cu regula de trei simple sau înmulțind frecvența relativă cu 360°)
- Comparați datele din 2000 cu cele din

Reprezentare prin grafic Pareto

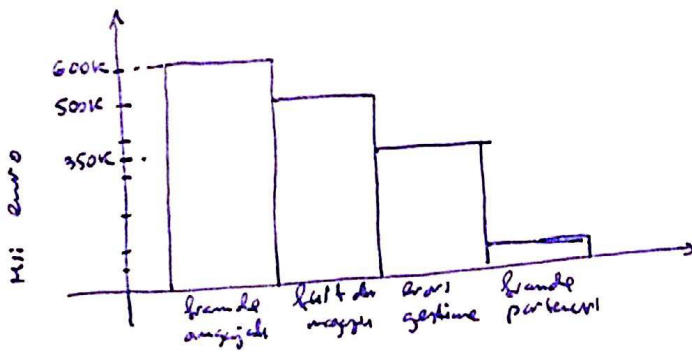
Un grafic Pareto este format din bare verticale ce reprezintă frecvențe sau frecvențe relative și sunt ardate descrescător.

Exemplul 3:

S-a estimat că lanțul Carrefour a înregistrat în 2016 pierderi de 1.500.000€ din care 350.000 datorate erorilor în gestiune, 600.000 datorate furtului angajaților 500.000 — furtului din magazin respectiv 50.000 datorat fraudelor partenerilor. Ce cauză ar trebui eliminată imediat?

Soluția:

Dacă trasăm un grafic Pareto observăm ușor că în primul rând ar trebui rezolvată problema furtului din partea angajaților.



Exemplul 4

În fiecare an poliția interzice prin acte restrictive privind contravențiile în trafic. În 2017 pe raza municipiului București au fost înregistrate plângeri privind: tamponare (25.500), neacordare de prioritate biciclistilor (5600), neacordării de prioritate la trecerea pietonilor (2300), intrarea cu apă de pe carosabil (300).

- Reprezentare datele într-un grafic Pareto.
- Interpretare datele.

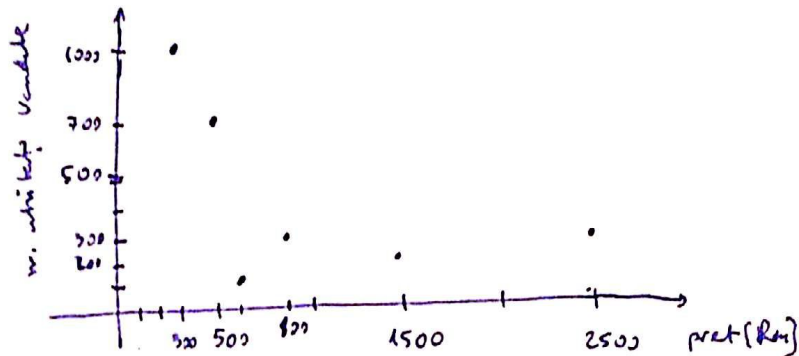
Reprezentarea perechilor de date

Dacă pentru două seturi de date există același număr de valori fiecarei date din primul set corespunzându-i o dată din al doilea avem perechi de date.

Exemplul 5:

Reprezentate datele din tabelul de mai jos:

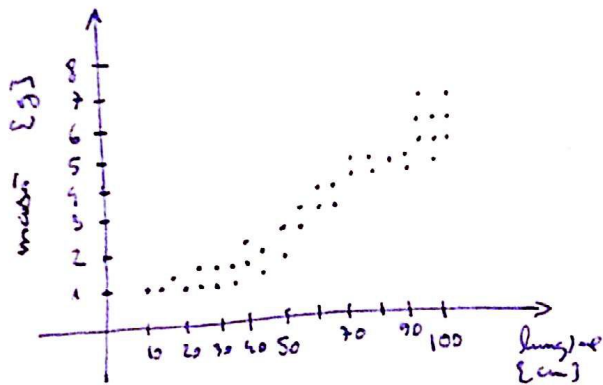
pret (RON)	număr de unități vândute
300	1000
500	700
600	100
800	300
1500	200
2500	300



În acest caz putem folosi cu succes o reprezentare cu coloane.
Pentru aceste date reprezentarea în perechi nu reprezintă avantaj.

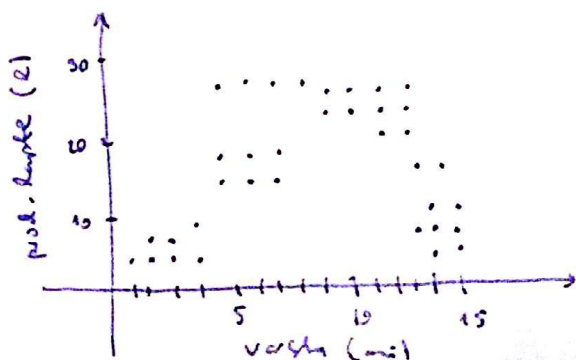
Exemplul 6

An fost măsurate mai multe caracteristici ale unei specii de roman și
a fost făcută reprezentarea de mai jos. Ce concluzii se poate trage?



Exemplul 6

An fost înregistrate vârsta și producția de lapte a vacilor de la o fermă și
obținut graficul de mai jos. Ce concluzii se pot trage?

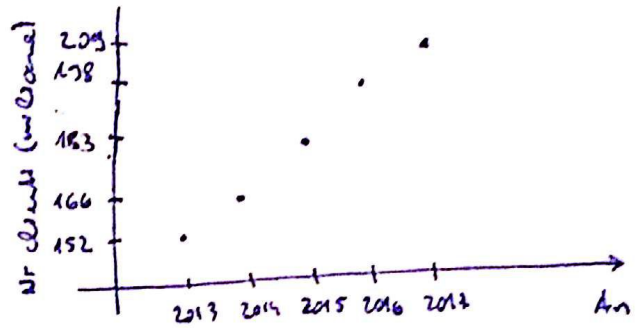


O serie temporală este formată dintr-o serie de date menurate la intervale regulate de timp cum ar fi temperatura zilnică de-a lungul anului sau nivelul precipitațiilor măsurate zilnic în raportul ploii.

Exemplul 7: În tabelul alăturat sunt notate numărul clientilor Vodafone pentru fiecare din anii 2013-2017 în India respectiv UK.

Construiți un grafic temporal.

An	UK	India
2013	19	152
2014	19,5	166,6
2015	18,4	183,8
2016	18,2	178
2017	17,7	209



Exercițiul 7: Construiți o reprezentare grafică a seriei temporale pentru evoluția nr. abonatilor în UK.



Măsurarea tendinței centrale: Medie, mediană, mod

Tendința centrală este reprezentată de mărimea tipică sau centrală = datelor.

media unei mulțimi de date = media aritmetică a valorilor datelor:

$$\text{media populației: } \mu = \frac{\sum x}{N}$$

$$\text{media eșantionului: } \bar{x} = \frac{\sum x}{n}$$

Exp 1: Știind că prețurile frigiderelor dintr-un eșantion de 10 unități

830, 1100, 1000, 950, 930, 1530, 1200, 1330, 1330, 1300

să se determine media prețurilor

sol:

$$\bar{x} = \frac{\sum x}{10} = \frac{830 + 1100 + \dots + 1300}{10} = \dots$$

Așadar prețul mediu pentru un frigider este ...

Exercițiul 1: Găsiți timpul mediu pt. efectuarea unei pt. eșantionul din Exp 1/21, respectiv pt. vârstele din Ex 1/21

mediana unei mulțimi de date = valoarea din mijloc când datele s-au ordonat (crescător sau descrescător).

Exp 2: Găsiți mediana prețurilor frigiderelor din exemplul 1/33.

Cât va fi mediana dacă frigiderul de 830 RON nu este considerat?

sol:

a) 830, 930, 950, 1000, 1100, 1200, 1300, 1330, 1330, 1530

$$\text{mediana} = 1150$$

b) 930, 950, 1000, 1100, 1200, 1300, 1330, 1330, 1530

$$\text{mediana}$$

Exercițiul 2: Găsiți mediana perioadelor din Ex 1/22. Dar dacă renunțăm

la înregistrarea de 15^h cât va fi mediana? Dar mediana vârstelor din

Ex 1/21?

modul unei multimi de date = data cu frecvență maximă.

Dacă frecvența maximă îi corespunde două valori ambele sunt moduri iar m. de date s.r. bi-modale. Există și mulțimi de date plurimodale.

Exp 3: Găsiți modul prețurilor fizicilor din ex 1/37

sol: întrucât frecvența absolută maximă este 2 modul este 1330

Ex 3: Găsiți modul vârstelor din Ex 1/21.

Exp 4: La o întrunire, studenții au fost întrebați ce grup politic simpatizează. Răspunsurile au fost grupate în tabelul de mai jos:

Grupuri Politice	Frecvențe
PNL	54
PVR	72
PD	15
PNP	20
PRM	10
USR	49

⇒ modul este PNP

Ex 5: Studenții unei serii au fost întrebați ce echipă de fotbal din campionatul intern al României merită să aibă răspuns pozitiv la întrebare.

Echipe preferate	Frecvențe
Steaua	5
Dinamo	6
Rapid	7
Astra	8
Viitorul	10
Spicul Brașov	9
Niduna	47

Găsiți modul rezultatelor de date.

Compararea indicatorilor tendinței centrale

Atât mediat cât și mediana sau modul descriu tendința centrală a unei mulțimi de date dar fiecare are atât avantaje cât și dezavantaje mai ales când mulțimea conține date exagerate

Exemplul 6: Vârstele studenților dintr-o grupă sunt:

19, 19, 19, 19, 19, 20, 20, 20, 20, 20, 21, 21, 21, 21, 22, 22, 22, 22, 78

Calculați indicatorii tendinței centrale!

sol:

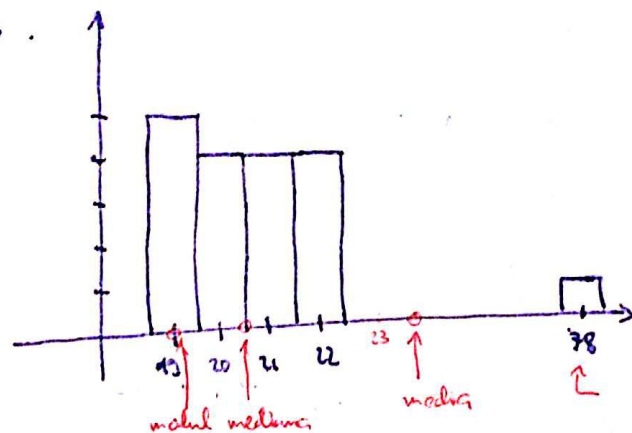
media: $\bar{x} = 23,61$

mediana: $\frac{20+21}{2} = 20,5$

modul: 19

În acest caz particular nici modul și nici mediat nu par să reflecte tendința centrală a vârstei studenților, mediana părând mai potrivită.

În realitate nu există un răspuns corect la întrebarea "Ce indicator reflectă mai fidel tendința centrală în acest caz?". De această dată histograma ar putea fi utilă în identificarea unei tendințe centrale:



Exercițiul 6: Calculați indicatorii tendinței centrale în cazul vârstelor studenților din Exp 6 după ce s-a înregistrat înregistrarea de 65 ani. Ce observați?

Media ponderată și media datelor grupate

$$\bar{x} = \frac{\sum x \cdot p}{\sum p} \quad \text{unde } p \text{ este ponderea fiecărei înregistrări } x$$

Exemplul 7: Nota finală a examenului de Analiză I, UTCB se calculează

luând în considerare media testelor din timpul semestrului (50%)

partajul de la jumătatea semestrului (10%)

laboratoarele (10%)

seminarurile (10%)

lecțiile (5%)

examenul final (15%)

Dacă Andrei a obținut punctajele 90, 100, 30, 0, 0 respectiv 100

care va fi nota finală?

Soluție:

$$\bar{x} = \frac{90 \cdot 50 + 100 \cdot 10 + 30 \cdot 10 + 0 + 0 + 100 \cdot 15}{100} = 73$$

Exercițiul 7: În urma contestărilor s-a constatat că Andrei avea 100 p pe
lecție și 50 p la seminar. Care este punctajul final în condițiile
actuale?

Media distribuției de frecvențe

$$\bar{x} = \frac{\sum x f}{\sum f} = \frac{\sum x f}{n} \quad \text{unde } x = \text{valorile mediere ale claselor,}$$

Ex 3: Calculați media distribuției de frecvențe din exp 2.

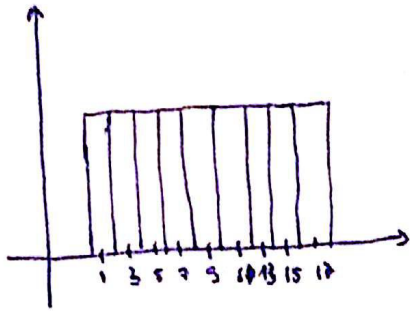
Soluție: Avem:

clasă	f	mediu
12-19	5	15,5
20-27	2	23,5
28-35	5	31,5
36-43	8	39,5
44-51	9	47,5
52-59	5	55,5
60-67	1	63,5

$$\begin{aligned} \bar{x} &= \frac{5 \cdot 15,5 + 2 \cdot 23,5 + \dots + 1 \cdot 63,5}{5 + 2 + \dots + 1} = \\ &= \frac{\dots}{35} \approx 39,04. \end{aligned}$$

Ex 8: Calculați media distribuției variabilei considerate din Exp 1/21

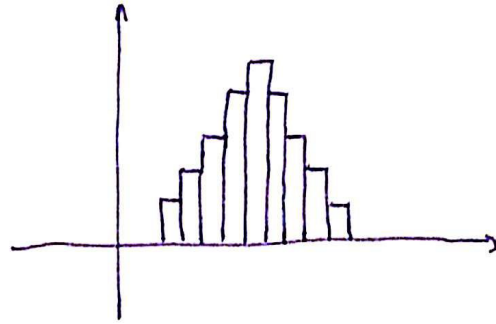
Forma distribuției



distribuție uniformă

(frecvențe egale pt toate clasele)

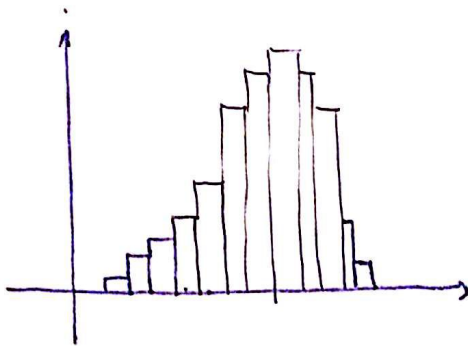
$$\text{medie} = \text{mediană} = \text{mod}$$



distribuție simetrică

(dreapta verticală de $x = \text{mediană}$
este axă de simetrie)

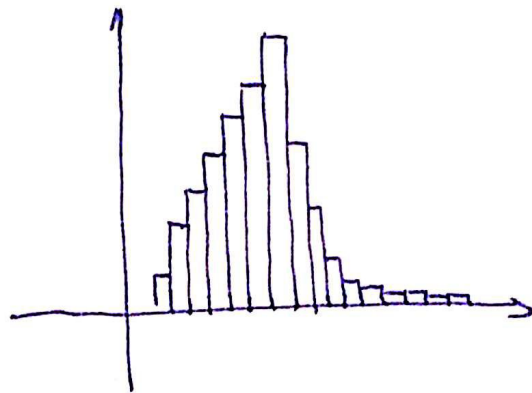
$$\text{medie} = \text{mediană} = \text{mod}$$



distribuție asimetrică

de stânga:

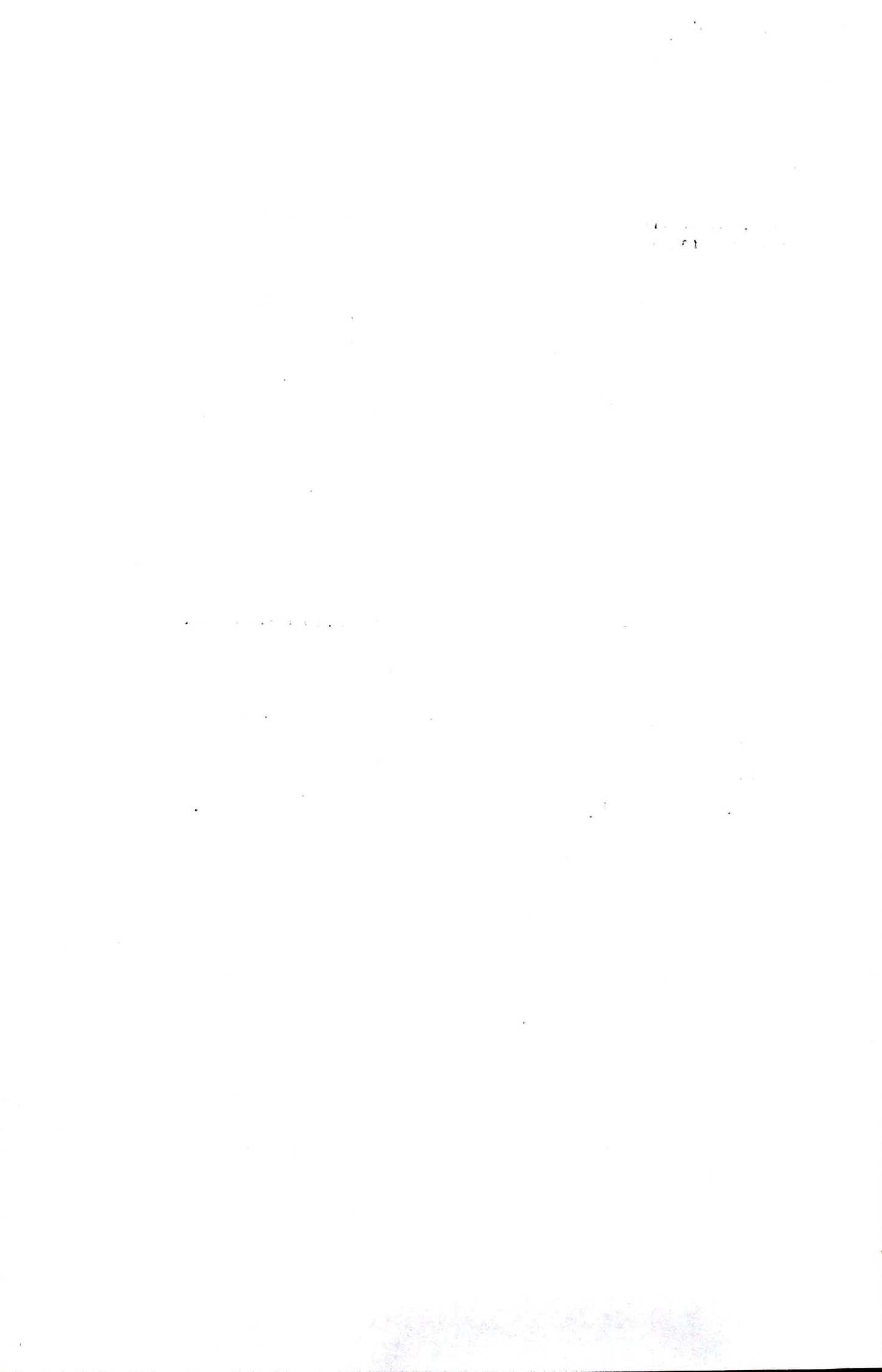
$$\text{medie} < \text{mediană} < \text{mod}$$



distribuție asimetrică

de dreapta:

$$\text{mod} < \text{mediană} < \text{medie}$$



Măsurarea variabilei

După cum am dedus din Exp 6 și Ex 6 de la pg 35, pe lângă medie, mediană și mod este important și cunoașterea și cât de împrăștiat sunt datele. Astfel apare noțiunea de amplitudine a unui set de date.

Amplitudinea unui set de date = diferența dintre valoarea maximă și valoarea minimă a înregistrărilor.

Exp 1: Amplitudinea în cazul vârstei studenților din Exp 6/35 este $78 - 19 = 59$ iar în cazul Ex 6/35 este doar $22 - 19 = 3$. Nu este de mirare că în ultimul caz mediana reflectă tendința centrală a datelor.

Ex 1: Calculați amplitudinea în cazul Ex 1/21.

Exp 2: Două firme au angajat câte 10 jurnaliști cu salariul anual în mii de euro după cum reiese din tabelul de mai jos:

Firma 1	4	8	12	16	20	26	30	34	38	42
Firma 2	14	16	18	20	22	24	26	28	30	32

În ambele cazuri medie = mediană = mod = 23

dar amplitudinea₁ = $42 - 4 = 38$ în timp ce

amplitudinea₂ = $32 - 14 = 18$ ceea ce arată că în primul

caz salariile sunt mai variate.

Totuși amplitudinea, deși ușor de calculat, caracterizează doar două valori de înțelegere ale datelor: valoarea maximă și valoarea minimă. În continuare vom introduce doi indicatori ce caracterizează gradul de împrăștiere al datelor luând în considerare toate datele. Acesta sunt dispersia respectiv abaterea medie standard.

Abateri individuale absolute arată împărțirea unei valori față de val. medie:

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

Abateri medii Linare este medie aritmetică a abaterilor individuale absolute.

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}$$

Abateri medii pătratice este

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

iar σ^2 m. dispersia populației

Exemplul 3: Determinați abaterile individuale absolute, abateri medii Linare și abateri medii pătratice pentru Firma 1 din exemplul 1.

soluție

individul x_i	\bar{x}	abs. ind. abs $d_i = x_i - \bar{x}$	Ab. medie linară $\frac{\sum d_i}{n}$	$(x_i - \bar{x})^2$	Ab. medie pătratică σ
4	23	-19			
8		-15			
12		-11			
16		-7			
20		-3			
28		3			
30		7			
34		11			
38		15			
42		19			