Universitatea “OVIDIUS” din Constanţa

Facultatea de Farmacie

Specializare: Farmacie

Disciplina: Informatică

Referat

Variabile statistice

 Student: Cristu Mariana-Manuela

Anul: I

Grupa: 3

Constanţa

2014

Cuprins

[I. Variabilele statistice 4](#_Toc384679199)

[1. Caracteristica statistică 4](#_Toc384679200)

[I. Clasificare: 4](#_Toc384679201)

[1. După modul de exprimare există: 4](#_Toc384679202)

[2. După cardinalul mulţimii observaţiilor există: 4](#_Toc384679203)

[3. După structura algebrică (SA) cu care este înzestrată mulţimea valorilor individuale şi după tipul scalei de măsurare există: 5](#_Toc384679204)

[4. După conţinutul variabilelor, acestea pot fi de timp, de spaţiu şi atributive. 7](#_Toc384679205)

[II. Organizarea datelor 9](#_Toc384679206)

[III. Reprezentarea grafica a datelor 9](#_Toc384679207)

[1. Poligonul frecvenţelor 10](#_Toc384679208)

[2. Histogramele 10](#_Toc384679209)

[3. Graficul Pie 11](#_Toc384679210)

[IV. Tipuri de variabile 12](#_Toc384679211)

[1. Variabile calitative şi variabile cantitative 12](#_Toc384679212)

[2. Variabile discrete şi variabile continue 13](#_Toc384679213)

[V. Eşantionare şi tipuri de eşantioane 13](#_Toc384679214)

[1. Esantion probabilistic 13](#_Toc384679215)

[2. Esantion neprobabilistic 14](#_Toc384679216)

[3. Eşantion aleator simplu 14](#_Toc384679217)

[4. Eşantion sistematic 14](#_Toc384679218)

[5. Eşantion stratificat 14](#_Toc384679219)

[VI. Bibliography 15](#_Toc384679220)

# Variabilele statistice

## Caracteristica statistică

Desemnează însuşirea, proprietatea, trăsătura comună unităţilor unei colectivităţi statistice, reţinută în programul statistic pentru a fi înregistrată şi care capătă accepţiuni sau valori diferite de la o unitate la alta sau de la un grup de unităţi la altul. Exemple de caracteristici statistice pot fi: vârsta, greutatea, sexul, culoarea ochilor, statutul matrimonial, naţionalitatea, ocupaţia, cifra de afaceri, nivelul extrasului de cont etc.

Caracteristicile statistice se mai numesc variabile statistice deoarece au proprietatea de a-şi modifica valoarea în timp şi spaţiu de la o unitate la alta. Nivelul de dezvoltare (varianta) este valoarea observată a unei variabile la o unitate (element) statistică. Valorile înregistrate de aceeaşi caracteristică la unităţile colectivităţii statistice se numesc variante

# Clasificare:

Variabilele statistice se clasifică după diferite criterii şi în mod deosebit după natura spaţiului observaţiilor. Aceasta este determinată de numărul unităţilor (cardinalul lui P), structura algebrică a spaţiului observaţiilor, pe baza acestora selectându-se şi scala de măsurare a variabilelor.

## **După modul de exprimare există**:

* Variabile calitative. Acestea sunt exprimate prin cuvinte care desemnează apartenenţa la o categorie sau o modalitate dintr-o mulţime finită de observaţii a unităţilor populaţiei statistice.

## **După cardinalul mulţimii observaţiilor există**:

* Variabile alternative (binare). În cazul acestora, spaţiul lor de observaţii este compus din două valori numerice (de exemplu {0,1}) sau două modalităţi (de exemplu {masculin, feminin} sau {urban, rural} etc.)
* Variabile cu un număr finit de valori numerice. În această clasă se încadrează toatevariabilele calitative, pentru care spaţiul observaţiilor (mulţimea valorilor individuale) format dintr-un număr finit de modalităţi, precum şi variabilele cantitative discrete pentru care mulţimea valorilor individuale (spaţiul observaţiilor) este echivalentă cu mulţimea numerelor naturale (card M ≡ card N).
* Variabile cantitative continue. În cazul acestora mulţimea specifică a valorilor individuale este un interval de numere reale.

Dacă valorile variabilelor discrete sunt mărimi „exacte” (de multe ori ele contorizează elementele care aparţin unei grupe sau clase de echivalenţă) nu acelaşi lucru se poate spun despre valorile variabilelor continue, datorită impreciziei instrumentelor de măsurare şi a factorilor care influenţează procesul măsurării.

Din această cauză, „greutatea exactă”, „înălţimea exactă” etc. nu mai sunt decât noţiuni abstracte, concretizabile numai sub forma unor aproximaţii din ce în ce mai bune.

Este posibil ca din diverse motive practice să nu fie necesară o precizie foarte mare şi atunci se apelează la „discretizare”, adică la aproximarea valorilor reale cu valori dintr-o mulţime discretă. Această discretizare nu trebuie realizată oricum (la întâmplare) ci în funcţie de natura originală a mulţimii valorilor individuale.

Acurateţea observaţiilor statistice depinde în mod decisiv de procesul de măsurare. Acesta nu poate fi, însă, aplicat în mod uniform tuturor variabilelor statistice. Gradele diferite de măsurabilitate sunt determinate de exprimarea cantitativă şi calitativă a variabilelor, de structura algebrică cu care este înzestrată mulţimea valorilor individuale (spaţiul observaţiilor M) şi care induc tipul scalei de măsurare.

## După structura algebrică (SA) cu care este înzestrată mulţimea valorilor individuale şi după tipul scalei de măsurare există:

Variabilele calitative nominale:

  Se caracterizează prin faptul că mulţimea specifică a valorilor individuale (M) nu este înzestrată cu structură algebrică şi se măsoară pe o scală nominală.

Exemple de astfel de variabile pot fi: categoria socioprofesională, starea civilă, tipologia şcolilor profesionale, ramura de activitate etc. Din analiza acestor exemple se observă următoarele:

Mulţimea de modalităţi M este finită, singurul criteriu distinctiv al acestora fiinddenumirea (cuvântul sau cuvintele prin care acestea se exprimă);

Mulţimea M nu posedă nici o structură, exceptând relaţia de identitate (=) – nonidentitate ( ≠ ) care asigură diferenţierea unităţilor. Nici în cazul în care se recurge la codificarea modalităţilor (prin atribuirea de coduri numerice) nu este posibil să fie indusă o structură.

Deci, singura operaţie obiectivă permisă de scala nominală este structurarea (divizarea) populaţiei statistice (eşantionului) în clase. În general, numărul claselor este identic cu numărul de modalităţi distincte. Nu este exclusă însă agregarea/dezagregarea claselor formate iniţial în clase mai mari sau în subclase, mai mult sau mai puţin compacte.

Variabilele calitative ordinale au mulţimea finită a valorilor individuale (M) exprimate prin modalităţi (sau coduri numerice asociate) înzestrată cu o structură de ordine totală (≤ ) şi se măsoară pe scala ordinală. Existenţa acestei structuri defineşte în M o ierarhie care, deseori, sugerează informaţii raţionale pentru luarea unor decizii.

Variabilele cantitative ordinale se caracterizează prin faptul că au o mulţimecontinuă de valori individuale înzestrată cu o structură de ordine şi se măsoară pe scala de interval. Pe scala de interval are sens definirea distanţei dintre valorile (numeric exprimate) ale unei variabile. Punctul zero al acestei scale şi unitatea de măsură se pot alege în mod arbitrar. Datorită caracterului relativ al localizării originii pe această scală, nu are sens suma a două valori precum şi raportul acestora. În schimb au sens, diferenţadintre două valori şi suma sau raportul diferenţelor. Pentru înţelegerea utilizării acestui tip de scală sunt clasice următoarele exemple:

Măsura rea timpului calendaristic. Punctul zero (de origine) al scalei este ales convenţional pentru a desemna de către unele popoare începutul erei creştine (nu toate popoarele au adoptat aceeaşi convenţie).

Temperatura se măsoară fie pe scala [Celsius](http://en.wikipedia.org/wiki/Celsius), fie pe scala Fahrenheit. Pe acestea, punctul zero şi unitatea de măsură sunt alese în mod arbitrar. Astfel, 00 pe Celsius nu înseamnă lipsa temperaturii, ci este un punct critic care desemnează schimbarea stării de agregare a apei. Aceeaşi semnificaţie o regăsim pe Fahrenheit la 320F.

Variabile cantitative măsurabile cardinal. Mulţimea valorilor numerice a acestor variabile este înzestrată cu o structură de corp ordonat (≤,+,× ), iar scala de măsurare corespunzătoare este scala de raport. Spre deosebire de scala de interval, scala de raport se caracterizează prin faptul că numai unitatea de măsură se poate alege arbitrar, punctulzero (de origine) este dat în mod natural, specifică absenţa fenomenului studiat.

Două valori măsurate pe această scală, indiferent de unitatea de măsură folosită, se află înacelaşi raport. Deci, prin trecerea de la o unitate de măsură la alta raportul dintre cele două valori nu se schimbă. Scala de raport permite cel mai înalt grad de măsurabilitate, deoarece valorilor precizate pe această scală li se pot aplica toate operaţiile aritmetice permise de structura de corp ordonat. Pe această scală au sens pe lângă relaţia de ordine şi operaţia de diferenţă (preluate de scalele anterioare) şi suma şi raportul a două valori ale variabilei. Prin urmare, scalele de măsurare prezentate oferă accesul la un conţinut informaţional care creşte de la scala nominală la cea de raport, treptele superioare integrând şi informaţia disponibilă în treptele inferioare.

## După conţinutul variabilelor, acestea pot fi de timp, de spaţiu şi atributive.

Variabilele de timp se caracterizează prin faptul că sunt exprimate prin funcţii de timp (u:T → M), adică valorile lor individuale aparţin unor momente de timp sau intervale de timp.

1. Variabilele de spaţiu (teritoriale) sunt definite ca funcţii de spaţiu (u:S → M), în sensul că fiecare valoare individual aparţine unei unităţi teritoriale care aparţine unui anumit nomenclator.
2. Variabilele atributive sunt definite printr-o funcţie (u:P →M) asociată fiecărei unităţi din populaţia statistică (eşantionul) investigat(ă). Valorile individuale ale acestor variabile (calitative sau cantitative) şi care formează spaţiul observaţiilor M se exprimă printr-un atribut (numeric sau nenumeric) asociat unităţilor observate.

Sinteză a tipurilor de variabile studiate într-o populaţie statistică (eşantion), potrivit scopului cercetării, se prezintă în figura următoare:



# Organizarea datelor

Datele variabilelor pot fi prezentate fie simplu, fie grupat. Primul tip de organizare constă în stabilirea frecvenţei de apariţie a fiecărei valori. Pentru aceasta este necesară o ierarhizare iniţială a valorilor în funcţie de mărimea lor.

Sa presupunem că au fost obţinute urmatoarele date ale variabilei studiate:

X = (7, 5, 7, 8, 4, 9, 8, 10, 5 3, 8, 10, 8, 7, 9, 6, 4, 7, 6, 1, 8, 6, 8, 7, 5, 7, 4, 7, 1, 9, 5, 8, 6, 7, 7). În total sunt 35 de date strânse.

Pentru a organiza datele utilizând o distribuţie simplă a frecvenţei sunt necesari urmatorii paşi:

1. Se caută valorile extreme din şirul de date (valoarea cea mai mare şi cea mai mică);
2. Se scriu toate valorile cuprinse între cele două extreme într-o ordine descendenta pe o coloană;
3. Se numară de câte ori apare fiecare valoare în şirul de date;
4. Se trece apoi în tabel, frecvenţa de apariţie a fiecărui numar.
5. În cazul de fata vom avea:

|  |  |
| --- | --- |
| Valoarea X | Frecvenţa f |
| 1 | 2 |
| 2 | 0 |
| 3 | 1 |
| 4 | 3 |
| 5 | 4 |
| 6 | 4 |
| 7 | 9 |
| 8 | 7 |
| 9 | 3 |
| 10 | 2 |

N=35

Mult mai utilizată, este gruparea datelor pe intervale. Pentru aceasta vom ţine cont de distribuţia grupată a datelor, fiind necesară împarţirea valorilor în clase de intervale egale. Există două metode principale de împarţire a datelor pe intervale.

Prima este propusă de Spatz (1997) şi are în vedere patru paşi de urmat iar o a doua posibilitate de a grupa datele pe interval pleacă de la formula matematică propusă de Sturges:

# Reprezentarea grafica a datelor

Desenele din statistică se numesc grafice. Acestea sunt modalitaţi imagistice de reprezentare a datelor, fiind cu atât mai sugestive şi mai utile cu cât numărul de date colectate creşte.

Există mai multe tipuri de reprezentare grafică. Putem aborda urmatoarele forme grafice adecvate statisticii descriptive: poligonul frecvenţelor, histogramele - pentru variabile cantitative de tip interval ori rapoarte;histogramele şi cercurile de tip "placintă" - pentru variabile calitative, nominale.

## Poligonul frecvenţelor

Este un grafic utilizat în cazul variabilelor cantitative. Un poligon al frecvenţelor presupune două axe (una orizontală Ox-abscisă, alta verticală Oy-ordonatată). Pe abscisă sunt trecute clasele (intervalele) ori direct valorile şirului de date. Pe ordonată sunt trecute frecvenţele grupate sau simple, corespunzatoare fiecărei clase (valori).

Utilizând un şir de date prezentat anterior (n=48) care a fost împarţit în prealabil în 6 clase, vom obţine următorul poligon al frecvenţelor ca reprezentare grafică a şirului de date.



Poligon de frecvenţe 1

## Histogramele

Acest tip de grafic poate fi utilizat pentru variabile exprimate atât cantitativ cât şi calitativ.

Diferenţa dintre cele două tipuri de histograme constă în faptul că pentru datele cantitative sunt trecute pe ordonată frecvenţele grupate, în timp ce pentru datele calitative frecvenţele sunt specificate paralel cu abscisa. Pe baza exemplelor date, vom construi două histograme (una pentru date cantitative – şirul de date împarţit în 10 clase; cealaltă pentru date calitative – şirul de date prezentat în cazul psihologului intersat de publicitate.



Histograma 1



Histograma 2

## Graficul Pie

În cazul datelor calitative se utilizeaza frecvent un alt tip de reprezentare grafica sub forma unui cerc "placinta" – în engleza "pie".



Pie 1

# Tipuri de variabile

Metodele statistice asigură o modalitate de a gestiona variabilitatea. Variaţia apare între oameni, între şcoli, între oraşe, între diverse obiecte care ar putea să constituie obiectul nostru de interes în viaţa de zi cu zi..O variabilă este o caracteristică ce poate varia între subiecţii unui eşantion sau a unei populaţii. Fiecare subiect are anumite valori pentru o variabilă, însă subiecţi diferiţi pot avea valori diferite. Exemple de variabile: genm vârsta, religia, numărul de copii în famili, preferinţa politică.

## Variabile calitative şi variabile cantitative

Informaţiile sunt numite calitative atunci când pentru măsurarea lor se foloseşte un set de categorii care nu sunt ordonate în nici un fel. Exemple de date calitative: statusul marital (necăsătorit, căsătorit, divorţat, văduv), localitatea de reşedinţă, apartenenţa religioasă, preferinţa politică, etc. Pentru variabilele calitative, categoriile diferă între ele prin calitate şi nu prin cantitate sau magnitudine.

Atunci când posibilele valori ale unei variabile diferă diferă în magnitudine, variabila este numită cantitativă. Fiecare valoare posibilă a unei variabile cantitative este mai mare sau mai mică decât orice altă valoare posibilă. Astfel de comparaţii sunt cu putinţă pentru variabile măsurate pe o scală numerică. Exemple de variabile cantitative: venitul anual, numărul de ani de educaţie, numărul de fraţi, de câte o ori o persoană a fost condamnată, etc.

Setul de categorii pentru o variabilă calitativă este numit scală nominală, iar setul de valori numerice pentru o variabilă cantitativă se numeşte scală interval. Scalele interval au o anumită distanţă sau un anumit „interval” între fiecare pereche de niveluri. Venitul lunar este măsurat pe o astfel de scală interval; intervalul dintre 2000 si 3000 RON, de exemplu, este de 1000 RON. Astfel se pot face comparaţii între niveluri diferite ale venitului, comparaţii care nu au nici un sens pentru o scală nominală. Într-un anumit sens există şi un al treilea tip de scală, situată între scala nominală şi cea interval. Este vorba despre o scală categorică, având o ordine naturală a valorilor, dar fără a putea identifica distanţa dintre valori. Aceasta este scala ordinală. De exemplu: poziţia în spectrul politic (stânga, centru-stânga, centru, centru-dreapta, dreapta), opinia faţă de mărimea cheltuielilor cu protecţia socială (prea mici, normale, prea mari), etc.

Principalul motiv pentru care se face distincţia dintre datele cantitative şi cele calitative este că, aşa cum s-a mai spus, pentru fiecare tip de date se aplică metode statistice diferite.

## Variabile discrete şi variabile continue

O altă modalitate de a clasifica variabilele ţine cont numărul de valori cuprinse în scala de măsurare. O variabilă este numită discretă dacă ea poate lua doar un număr finit de valori şi este numită continuă dacă poate lua ca valori un număr infinit de numere reale.

Exemple de variabile discrete: numărul de copii ai fiecărei familii, numărul de infracţiuni într-un an, numărul de vizite la medic în ultimul an, etc. Oricare dintre variabilele anterioare numite „numărul de...” este o variabilă discretă, întrucât poate lua doar valori din mulţimea {0, 1, 2, 3, 4, …}.

Exemple de variabile continue: înălţimea, greutatea, vârsta, venitul, etc. Este imposibil de precizat toate valorile posibile ale unei variabile continue. Greutatea unei persoane poate fi 73,82035... kg, funcţie de precizia cu care este făcută măsurătoarea.

În cazul variabilelor discrete nu se poate subdiviza unitatea de măsură. Numărul de copii într-o familie poate fi 2 sau 3, dar în nici un caz, 2,57. Pe de altă parte, o colecţie de valori ale unei variabile continue poate fi oricând redefinită între două valori posibile. Orice vârstă cuprinsă între 20 şi 20,5 ani, de exemplu, poate fi rotunjită la 20 de ani, iar orice vârstă cuprinsă între 20,5 şi 21 de ani poate fi rotunjită la 21 de ani.

Variabilele calitative sunt discrete, ele având un set finit de categorii. Variabilele cantitative pot fi atât discrete cât şi continue.

# Eşantionare şi tipuri de eşantioane

 Statistica inferenţială utilizează eşantioanele pentru a face predicţii despre parametrii populaţiilor din care acestea au fost extrase. Calitatea inferenţei depinde în mod esenţial de modul în care eşantionul reprezintă populaţia.

## Esantion probabilistic

Un eşantion probabilistic este acela în care subiecţii sunt aleşi pe baza unei probabilităţi cunoscute. Eşantioanele probabilistice trebuie utilizate ori de câte ori este posibil, deoarece doar ele asigură o corectă inferenţă statistică de la eşantion la populaţie. Există patru tipuri de eşantioane probabilistice: eşantionul aleatoriu simplu, eşantionul sistematic, cel startificat şi cel de tip cluster. Aceste tipuri diferă între ele prin cost, acurateţe şi complexitate.

## Esantion neprobabilistic

Un eşantion neprobabilistic este acela în care elementele componente sunt alese fără să se ţină cont de probabilitatea apariţiei lor. Pentru anumite studii, eşantioanele neprobabilistice, aşa cum sunt cele pe cote, cele intenţionate sau cele convenabile, sunt suficiente. Aceste eşantioane au unele avantaje faţă de cele probabilistice, cum ar fi uşurinţa şi viteza cu care pot fi construite, precum şi costul scăzut. Pe de altă parte, ele au două dezavantaje majore: lipsa de acurateţe şi imposibilitatea generalizării. Din aceste motive, eşantioanele neprobabilistice pot fi utilizate doar atunci când dorim o aproximare grosieră la un cost scăzut pentru a ne satisface o curiozitate personală sau atunci când dorim să realizăm un studiu pilot, care va fi urmat mai târziu de o cercetare mult mai riguroasă.

## Eşantion aleator simplu

Un eşantion aleator simplu este acela în care fiecare element din cadrul de eşantionare are aceeaşi şansă de a fi selectat. Pentru a putea utiliza un astfel de eşantion este necesar, în primul rând, să existe o listă completă a populaţiei. Apoi de pe această listă se poate extrage eşantionul dorit utilizând metoda loteriei sau un tabel cu numere aleatoare sau un calculator care să genereze numere aleatoare.

## Eşantion sistematic

 Deşi sunt uşor de utilizat, eşantionarea aleatorie simplă şi eşantionarea sistematică sunt în general mai puţin eficiente decât alte metode mai sofisticate şi nu se poate şti dacă eşantionul este într-adevăr reprezentativ. În cazul eşantioanelor sistematice, posibilitatea de eroare este chiar mai mare, în cazul în care cadrul de eşantionare este organizat după un anumit model.

## Eşantion stratificat

 Această metodă de eşantionare este mai eficientă decât cele anterioare, deoarece asigură reprezentarea indivizilor din întreaga populaţie, ceea ce oferă o mai mare precizie în estimarea parametrilor populaţiei.

# Bibliography

(n.d.). Retrieved from http://prelucraristatistice.wordpress.com/2013/04/13/variabilele-statistice/

(n.d.). Retrieved from http://statisticasociala.tripod.com/variabile2.htm

Duguleana, L. (2012). *Bazele statisticii economice.* CH Beck.

Viorel Gh. Voda, D. V. (1977). *Statistica si tolerante.* Tehnica.

Virgil, S. (n.d.). *Statistică aplicată în ştiinţele politice.*