

II.C

ANALIZA STATISTICĂ DESCRIPTIVĂ A DATELOR MEDICALE - Epi Info v7

1. Introducere

După colectarea datelor medicale, o prima etapă a analizei statistice este descrierea datelor, atât din punct de vedere numeric cu ajutorul indicatorilor (de tendință centrală și dispersie), cât și pictural cu ajutorul reprezentărilor grafice.

În continuare, vom exemplifica realizarea unei analize statistice descriptive cu ajutorul aplicației **Epi Info v7**.

2. Probleme



Problema 1

Se considera datele de la problema în care s-a creat baza de date cu privire la factorii ce influențează apariția unui accident vascular cerebral (AVC). Fișierul de date în format Excel se găsește în Anexă (AVC.xls).

Pe lângă informațiile referitoare la identitatea celor care răspund: nume, prenume, data nașterii, vârstă, sex, greutate, înălțime, s-au colectat și informații legate de starea de sănătate și stilul de viață: tensiunea arterială, obezitate, stil de viață, boli cardiovasculare, diabet zaharat, fumat, colesterol crescut, AVC antecedent.

Tipuri de variabile



Care este tipul fiecărei variabile incluse în baza de date?

Identificarea tipului unei variabile este o etapă esențială în analiza statistică descriptivă, precum și în cea inferențială. În Tabelul 1 sunt prezentate variabilele colectate, împreună cu tipul acestora și o scurtă descriere.

Tabel 1. Tipul variabilelor referitoare la identitatea respondenților, precum și la factorii ce influențează apariția unui AVC.

Variabila	Tipul variabilei	Observații
Nume, prenume	Nominal	-
Vârsta	Numeric, continuu	-
Greutate	Numeric, continuu	-
Înălțime	Numeric, continuu	-
Tensiunea arterială	Nominal, categoric	Se consideră categoria în care se încadrează tensiunea: normală, prehipertensiune, hipertensiune.
Obezitate	Nominal, dihotomic	Se consideră prezența/absența obezității (în special a celei abdominale).
Inactivitate	Nominal, dihotomic	Se consideră absența/prezența unui stil de viață activ.
Boli cardiovasculare	Nominal, categoric	Se consideră prezența unei boli cardiovasculare, precum fibrilație atrială, afecțiuni sau proteze valvulare, cardiopatie ischemică, cardiomiopatie dilatativă, insuficiență cardiacă.
Diabet zaharat	Nominal, dihotomic	Se consideră prezența/absența diabetului zaharat, indiferent de ce tip este.
Fumat	Nominal, dihotomic	Se înregistrează dacă persoana este fumător/fumătoare.
Colesterol crescut	Nominal, dihotomic	Se consideră prezența/absența unui colesterol crescut (în special LDL).
AVC antecedent	Nominal, dihotomic	Se considera prezența/absența unui AVC antecedent.
Atac ischemic tranzitor antecedent	Nominal, dihotomic	Se consideră prezența/absența unui atac ischemic tranzitor precedent.
Arterioscleroză	Nominal, dihotomic	Se consideră prezența/absența arteriosclerozei.

Variabilele de tip numeric-continuu sau cantitative pot fi descrise cu ajutorul tuturor indicatorilor de tendința centrală, și anume: media, mediana și moda. În același timp, indicatorii de dispersie ce pot și folosiți sunt varianta, deviația standard, amplitudinea.

Indicatori de tendință centrală și de dispersie

Indicatorii de tendință centrală și de dispersie pentru **variabilele de tip numeric-continuu / cantitative**.

Indicatorii de tendință centrală	Indicatorii de dispersie
<ul style="list-style-type: none"> • Media • Mediana • Moda 	<ul style="list-style-type: none"> • Variația • Deviația standard • Amplitudinea

Variabilele de tip ordinal pot fi descrise folosind doar doi indicatori de tendință centrală, și anume, mediana și moda, iar indicatorii de dispersie sunt intervalul interquartilic și intervalul interquantilic.

Indicatorii de tendință centrală și dispersie pentru **variabile de tip ordinal**.

Indicatorii de tendință centrală	Indicatorii de dispersie
<ul style="list-style-type: none"> • Mediana • Moda 	<ul style="list-style-type: none"> • <i>Intervalul interquartilic</i> • <i>Intervalul interquantilic</i>

Variabilele de tip nominal sau calitativ pot fi descrise cu ajutorul unui singur indicator de tendință centrală și anume, moda, iar indicatorii de dispersie sunt procentul și deviația standard a procentului.

Indicatorii de tendință centrală și de dispersie pentru **variabilele de tip nominal/calitativ**.

Indicatorul de tendință centrală	Indicatorii de dispersie
<ul style="list-style-type: none"> • Moda 	<ul style="list-style-type: none"> • Procentul • Deviația standard a procentului



Soluții 1

Se deschide aplicația EpiInfo v7. În Figura 1 este prezentată fereastra principală a aplicației. Modulele oferite de aplicație sunt dispuse vertical, începând cu modulul **Create Forms**, care facilitează parcurgerea primului pas în colectarea datelor, și anume, crearea câmpurilor în care vor fi colectate datele sub

forma unor formulare (chestionare). Modulul **Enter Data** permite introducerea datelor prin completarea formularului sau formularelor create.

Modulul **Analyse Date** include instrumentele necesare realizării analizei statistice a datelor, atât la nivel descriptiv, cat si inferențial.

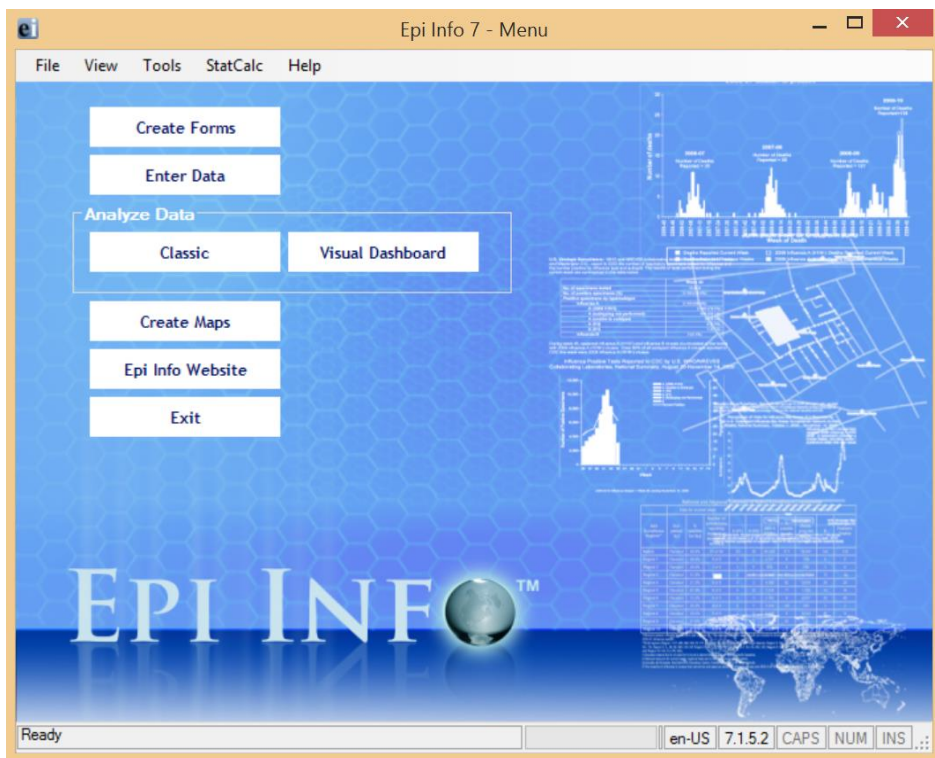


Figura 1. Fereastra principală a aplicației EpiInfo v7.

Pentru realizarea analizei descriptive a datelor, se va lucra cu modulul **Analyze Data**, varianta **Classic** (Figura 1).

Se citește fișierul de date (anonimizat) cu comanda **Read**. În fereastra de citire, se vor utiliza următoarele opțiuni: opțiunea implicită **Epi Info 7 Project** pentru **Database Type**, se alege fișierul **AVC.prj** pentru **Data Source**, iar la **Data Source Explorer->Epi Info Forms** se selectează **AVCChestionar** și se finalizează citirea cu OK. Dacă citirea s-a făcut cu succes, pe ecran ar trebui să fie afișate informațiile referitoare la locația fișierului citit, numărul de înregistrări din fișier, precum și data (și ora) la care s-a făcut citirea.

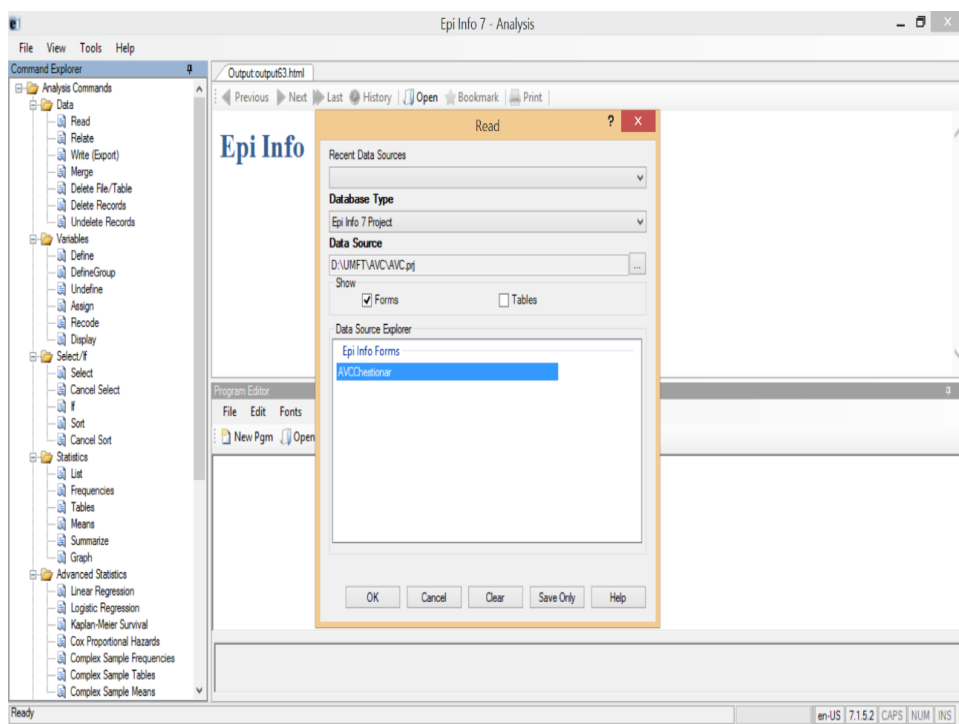


Figura 2. Citirea fișierului de date în modulul **Analyze Data, Classic**.

Înainte de a începe analiza statistică, se poate afișa conținutul fișierului citit cu ajutorul comenzii **Statistics->List**. În cazul în care se păstrează selecția implicită * la opțiunea **Variables**, atunci toate variabilele prezente în fișier vor fi listate. În caz contrar, se pot selecta separat variabilele care nu vor fi afișate.

În continuare, vom analiza variabilele în funcție de tipul lor.



Variabile de tip numeric

Pentru descrierea variabilelor de tip numeric, vom utiliza comanda **Statistics->Means**. Astfel, dacă dorim să descriem variabila **Vârsta**, vom selecta **Means of->Varsta** în fereastra **Means**, și vom finaliza comanda cu OK (Figura 3).

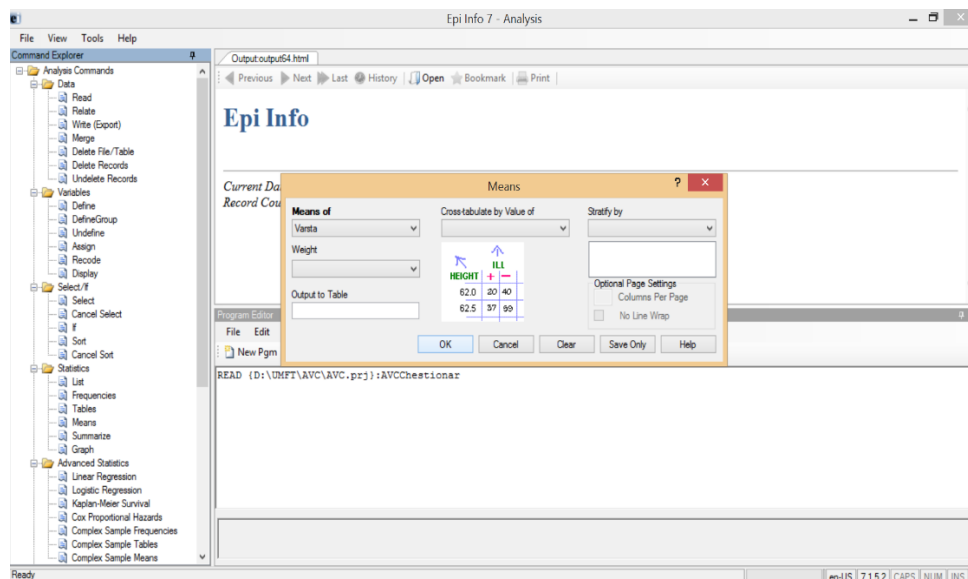


Figura 3. Aplicarea comenzii **Statistics->Means**.

Pe ecran vor fi afișate următoarele informații:

MEANS Varsta

Obs	Total	Mean	Variance	Std Dev
25.0000	1312.0000	52.4800	127.0933	11.2736
Minimum	25%	Median	75%	Maximum
24.0000	48.0000	55.0000	61.0000	73.0000
				Mode
				55.0000

- **Obs** indică numărul de observații/înregistrări ale variabile vârstă prezente în fișier. În cazul nostru, sunt 25 de înregistrări.
- **Total** indică suma valorilor de vârstă.
- **Mean** indică media valorilor de vârstă. În cazul nostru, cei 25 de participanți au media de vârstă de 52 de ani.
- **Variance** indică varianța valorilor de vârstă. În cazul nostru este 127.09.
- **Std. Dev.** indică deviația standard a valorilor de vârstă. În cazul nostru este 11.27.
- **Minimum** indică valoarea minimă de vârstă.
- **25%** indică prima quartila (Q1) care marchează limita celor mai mici 25% dintre valorile de vârstă.

- **Median** indică valoarea medianeii fiind totodată și a doua quartilă (Q2), care împarte setul de valori în două părți egale. În cazul nostru mediana este 73.
- **75%** indică a treia quartilă (Q3) care marchează limita celor mai mari 25% dintre valori.
- **Maximum** indică valoarea maximă de vârstă.
- **Mode** indică moda valorilor de vârstă, adică vârstă ce apare cel mai des în setul de date. În cazul nostru este 55 de ani.



Variabile de tip nominal

Pentru variabilele de tip nominal, indicatorul de tendință centrală se va calcula cu comanda **Statistics** → **Means** → **Mode**. Astfel, pentru variabila **Obezitate** obținem că cele mai multe persoane sunt obeze, valoarea **True** având frecvența cea mai mare.

Tabele de frecvență

De asemenea, putem determina procentul celor care sunt sau nu obezi cu ajutorul comenzii **Statistics** → **Frequencies** → **Frequency of Obezitate** (Figura 4).

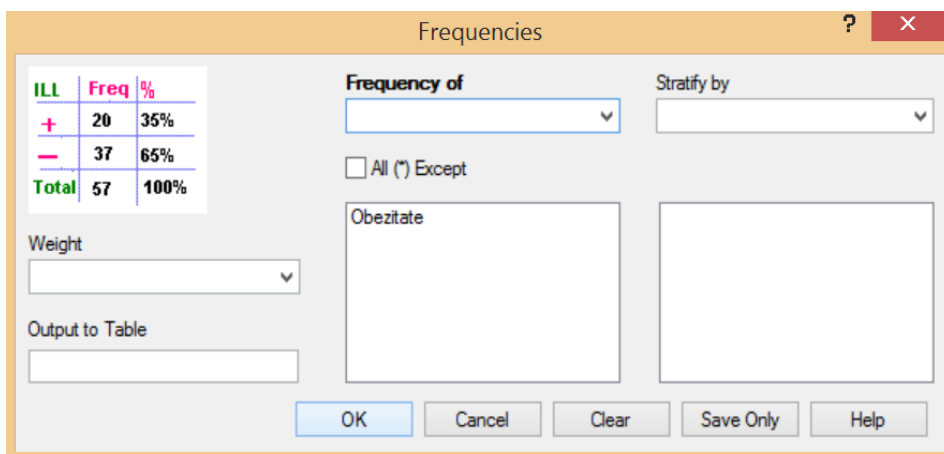


Figura 4. Aplicarea comenzii **Statistics** → **Frequencies**.

Astfel aflăm că 6 persoane nu sunt obeze, adică 24% din numărul de persoane și 19 persoane sunt obeze, adică 76%. Tabelul de frecvențe obținut este următorul:

FREQ Obezitate

OBEZITATE	Frequency	Percent	Cum. Percent	
No	6	24.00%	24.00%	<div style="width: 24%;"></div>
Yes	19	76.00%	100.00%	<div style="width: 76%;"></div>
Total	25	100.00%	100.00%	<div style="width: 100%;"></div>

Exact 95% Conf Limits

No 9.36% 45.13%

Yes 54.87% 90.64%



Problema 2

Calculați indicatorii de tendință centrală și de dispersie pentru variabilele **Greutate** și **Înălțime**.



Problema 3

Calculați indicatorii de tendință centrală și de dispersie pentru variabilele **Inactivitate**, **Diabet zaharat**, **Fumat**, **AVC antecedent**. Reprezentați grafic aceste variabile.



Problema 4

Determinați tabelul de frecvență pentru variabila **Boli cardiovasculare** și **Obezitate**. Reprezentați grafic cele două variabile.



Indicații:

Pentru variabila nominală de tip categorie **Boli cardiovasculare**, tabelul de frecvențe este

FREQ Bolicardiovasculare

BOLICARDIOVASCULARE	Frequency	Percent	Cum. Percent	
afecțiuni sau proteze valvulare	2	13.33%	13.33%	
cardiomiopatie dilatativă	5	33.33%	46.67%	
cardiopatie ischemică	1	6.67%	53.33%	
fibrilație atrială	4	26.67%	80.00%	
insuficiență cardiacă	3	20.00%	100.00%	
Total	15	100.00%	100.00%	

Exact 95% Conf Limits

afecțiuni sau proteze valvulare	1.66%	40.46%
cardiomiopatie dilatativă	11.82%	61.62%
cardiopatie ischemică	0.17%	31.95%
fibrilație atrială	7.79%	55.10%
insuficiență cardiacă	4.33%	48.09%

Grafice

Reprezentarea grafică a frecvențelor de apariție a categoriilor de boli cardiovasculare se realizează cu comanda **Statistics** → **Graph**. Vom alege un grafic de tip coloană, **Graph Type** → **Column**, iar la **Main Variables** vom selecta variabila **Boli cardiovasculare**.

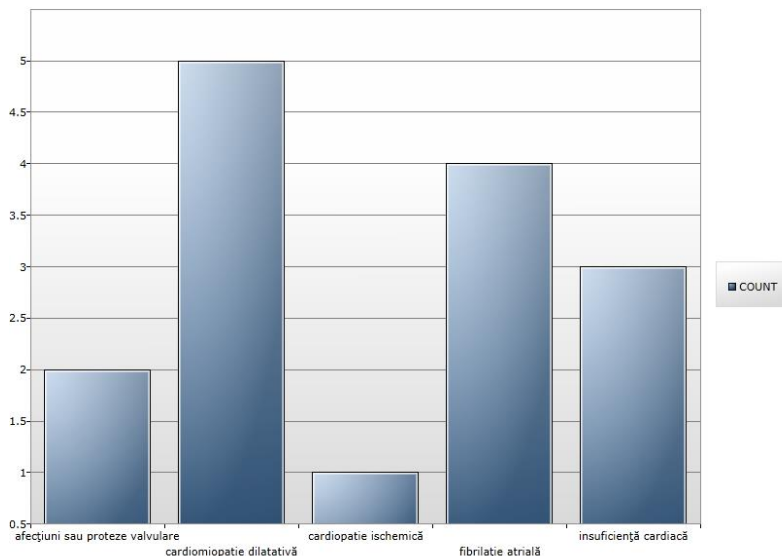


Figura 5. Reprezentarea grafica de tip **Column** a frecvențelor de apariție a categoriilor de boli cardiovasculare.

Iar pentru variabila **Obezitate**, vom realiza un grafic de tip **Pie**.

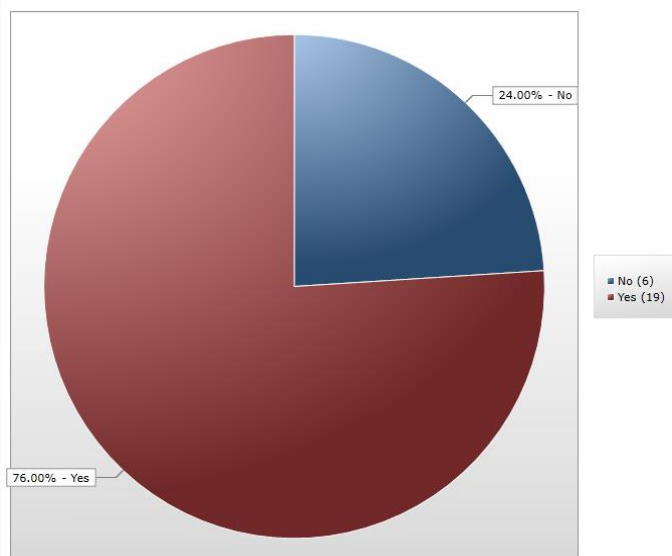


Figura 6. Reprezentarea grafica de tip **Pie** a procentelor celor care prezintă sau nu prezintă obezitate

3. Concluzii

În acest material ați învățat noțiunile de baza privind analiza descriptivă a datelor biomedicale cu ajutorul aplicației EpiInfo v7:

- ✚ Recunoașterea tipurilor de variabile;
- ✚ Calcularea indicatorilor de tendință centrală și dispersie pentru diferite tipuri de variabile;
- ✚ Determinarea tabelului de frecvențe;
- ✚ Realizarea de reprezentări grafice.

Referințe

- Vernic CV, Apostol SA, Frandes M, Mada L, Lungeanu D. APLICATII PRACTICE DE INFORMATICA SI BIOSTATISTICA MEDICALA IN NURSING. Editura Eurostampa, ISBN 978-606-32-0487-6, Colectia Derzelas, 2017:1-216.
- Mihalas GI, Lungeanu D. Introducere in informatica medicala si biostatistica, Editura „Victor Babes”, 2009: 1-225.
- Rosner B. Fundamentals of biostatistics (5th Edition). Pacific Grove: Duxbury. Thomson Learning, 2000.
- Knapp RG, Clinton Miller M: Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins, 1992.