

EPI INFO - Analiza și manipularea datelor -

Pentru analiza statistică a datelor în Epi Info sunt disponibile două module: **Classic** și **Visual Dashboard**. Ambele oferă aceleași rezultate dar modul de afișare și de accesare a comenzilor sunt diferite.

1. Modulul CLASSIC

Modulul **Classic** (figura 1) este organizat în trei zone:

- **Command Explorer**: un navigator ce conține toate comenzile disponibile grupate pe categorii de utilizare, fig1 - 1.

- **Output**: zona în care sunt afișate rezultatele obținute. Acestea sunt stocate în fișiere de tip *.html cu denumirea implicită outputX.html, unde X este un număr ce reprezintă sesiunea de lucru, fig1-2. Calea implicită de salvare este: C:\Users\Output - Storing Output.

- **Program Editor**: o consolă text unde pot fi introduse și rulate comenzi. Tot aici se regăsesc comenzile accesate prin navigator, fig.1-3. Conținutul consolei text poate fi salvat într-un fișier cu extensia *.pgm în vederea execuției ulterioare sau a creării de șabloane de lucru.



Fig.1

Epi Info - Modulul de analiză Classic

1 - navigator; 2 - zona de afișare rezultate; 3 - consola text

1.1 Prelucrarea datelor în modulul Classic

În vederea exemplificării se vor prelucra datele conținute în fișierul **RiscCardioVascular.xls**. Câmpurile conținute în fișier sunt:

- **Sex**: reprezintă sexul pacientului: F sau M
- **Diabet**: dacă pacientul suferă de diabet sau nu
- **Vârsta**: reprezintă vârsta pacientului
- **Hipertensiv** - dacă pacientul este hipertensiv sau nu
- **Alcool**: dacă pacientul consumă alcool în mod regulat
- **Fumat**: dacă pacientul este fumător
- **Greutate**: reprezintă greutatea pacientului, în kilograme
- **Înălțime**: reprezintă înălțimea pacientului, în centimetri
- **Talie**: reprezintă diametrul abdominal al pacientului, în centimetri
- **Indice de masă corporal** : reprezintă indicele de masă corporal al pacientului
- **Tensiune arterială sistolică** : reprezintă tensiunea arterială sistolică a pacientului
- **Tensiune arterială diastolică** : reprezintă tensiunea arterială diastolică a pacientului

- **Glicemie:** reprezintă glicemia pacientului
- **Colesterol:** reprezintă nivelul de colesterol al pacientului
- **Tireoglobulina:** reprezintă nivelul tireoglobulinei pacientului
- **Colesterol HDL:** reprezintă nivelul de lipoproteine cu densitate înaltă a pacientului
- **Uricemie:** reprezintă nivelul de acid uric din sângele pacientului
- **Indice de risc cardiovascular:** reprezintă nivelul de risc cardiovascular al pacientului

1.2 Importarea datelor în modulul de analiză Classic

În vederea prelucrării datelor este necesară datelor în modulul de analiză, utilizând pașii:

- se deschide modulul **Classic**
- se activează din navigator comanda **Data - Read**
- în fereastra deschisă se alege tipul de bază de date: Microsoft Excel 97-2003 Workbook (.xls)
- în câmpul **Data Source** se caută fișierul RiscCardioVascular.xls (D:\)
- se selectează foaia de lucru curentă (sheet 1\$)

Rezultatul este prezentat în figura 4.

Se remarcă, în consola text, sintaxa comenzii **READ**.

În zona de afișare a rezultatelor sunt specificate o serie de informații referitoare la baza de date deschisă: denumirea și locația, numărul de înregistrări, data ultimei accesări etc..



Fig.4

Rezultatul importului în modulul Classic

1.3 Prelucrarea datelor în modulul de analiză Classic

Se vor efectua următoarele operații:

-Afișarea datelor

Pentru afișarea datelor din baza de date importată se utilizează comanda: **Statistics - List**. În fereastra deschisă de comandă sunt disponibile două moduri de afișare: sub formă de tabel tip Grid (figura 5a) sau sub formă de tabel gata de imprimare (figura 5b).



a)

b)

Fig.5

Afișarea datelor din baza de date

a) sub formă de grid, b) sub formă de tabel printabil

Exportul datelor

În vederea exportării unor date din modulul **Classic** se utilizează comanda **Data - Write (Export)**

În fereastra activată de comandă se fac următoarele setări:

- **Output Mode:** Append
- **Output Format:** Microsoft Excel 97-2003 Workbook (.xls)
- **Connection Information:**
 - se apasă butonul Browse (...)
 - în fereastra deschisă se apasă butonul Browse (...)
 - În fereastra deschisă se alege directorul unde se dorește salvarea și se introduce numele dorit pentru fișierul final
- **Destination Table:** se introduce numele foii de lucru din fișierul destinație (Sheet1)

În fereastra de export date este posibilă alegerea variabilelor exportate.

Se obține fișierul prezentat în figura 6.

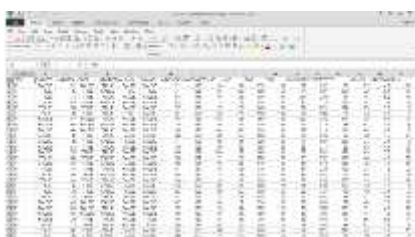


Fig.6

Date exportate din modulul Classic în eXCEL

Exerciții

1. Să se creeze un director nou **D:\Laborator EPIinfo** și să se seteze ca director implicit de salvare a rezultatelor obținute în modulul **Classic**
2. Să se importe în modulul **Classic** conținutul fișierului **RiscCardioVascular.xls**
3. Să se afișeze valorile sub formă de tabel printabil
4. Să se exporte într-un nou fișier tip Excel valorile variabilelor Vârsta, Alcool, Fumat și Indicele de masă corporală. Denumirea fișierului va fi **FactoriRisc.xls**
5. Să se importe în modulul **Classic** conținutul fișierului **FactoriRisc.xls**

2. Modulul VISUAL DASHBOARD

Modulul **Visual Dashboard** (figura 1) este organizat în trei zone:

- **Header:** antetul, ce conține butoane pentru comenzile uzuale, cum ar fi: indicarea sursei de date, deschiderea unei foi noi de lucru sau salvarea acesteia. Tot aici se regăsesc afișate informațiile referitoare la sursa de date și numărul înregistrărilor curente.
- **Canvas:** zona de lucru, în care pot fi afișate comenzile specifice, sub forma unor "gadget-uri", disponibile în meniul contextual (click-dreapta mouse). După încărcarea unui set de date, pe lateralele acestei zone sunt două gadgeturi permanente cu care se pot manipula (Defined Variables) și filtra (Data Filters) datele.
- **Status bar:** bara de stare, unde sunt afișate numărul gadgeturilor din zona de lucru, numărul de variabile înregistrate din tabelul curent și un cursor pentru zoom.



Fig.1

Epi Info - Modulul de analiz Visual Dashboard

1 - antet (header); 2 - zona de lucru (canvas); 3 - bara de stare (status bar)

Facilitățile oferite de Modulul Visual Dashboard sunt:

- afișarea datelor din proiecte Epi Info, foi de lucru Excel, baze de date Acces etc.
- relaționarea datelor din surse diferite
- filtrarea datelor
- definirea unor noi variabile
- afișarea datelor
- afișarea rezultatelor calculului
- exportarea datelor în formate diferite

2.1 Comenzi disponibile

În modulul **Visual Dashboard** comenzile sunt disponibile sub forma unor "gadgeturi" - elemente vizuale, flotante în zona de lucru, care pot fi accesate din meniul contextual, figura 2a.

Fiecare gadget dispune de o serie de butoane (figura 2b) ce declanșează acțiuni specifice asupra gadgetului.



a)

- Set Gadget Filter** - aplica filtre datelor
- Change Gadget Properties** - selectează înregistrarea care se prelucrează
- Set Description** - permite inserarea unor comentarii
- Collapse/Expand Gadget Output** - minimizează sau maximizează rezultatul gadgetului
- Close** - închide gadgetul

b)

Fig.2

a) comenzi disponibile în Visual Dashboard; b) acțiuni asupra gadgetului

Printre cele mai utilizate gadgeturi sunt:

- **Line List** - afișează datele sub formă de tabel. Este posibil să afișezi toate valorile (CTRL-A în lista cu variabile din fereastra deschisă de comandă) sau doar a celor selectate.
- **Frequency** - afișează un tabel cu distribuția datelor și intervalele de încredere
- **Means** - afișează o serie de parametri specifici statisticii descriptive ce pot fi selectați în fereastra deschisă de comandă
- **Chart** - afișează datele și rezultatele sub formă grafică. Sunt disponibile o paletă largă de tipuri de grafice

2. 2 Aplicații

2.2.1 Se se importe în **DashBoard** fi ierul RiscCardioVascular.xls.

Procedura

Se execut click-dreapta pe zona de lucru i se import fi ierul RiscCardioVascular.xls, urm rind pa ii din figura 3.

2.2.2 Se se afi eze coloanele: Vârsta, Greutate, Talie, TAS, TAD.

Procedura

Se execut pa ii din figura 4. Se observ în fereastra comenzii Line List, posibilitatea stabilirii num rului de înregistr ri afi ate. Indiferent de câte înregistr ri sunt afi ate în calculele ulterioare vor fi luate în considerare toate înregistr rile variabile respective.

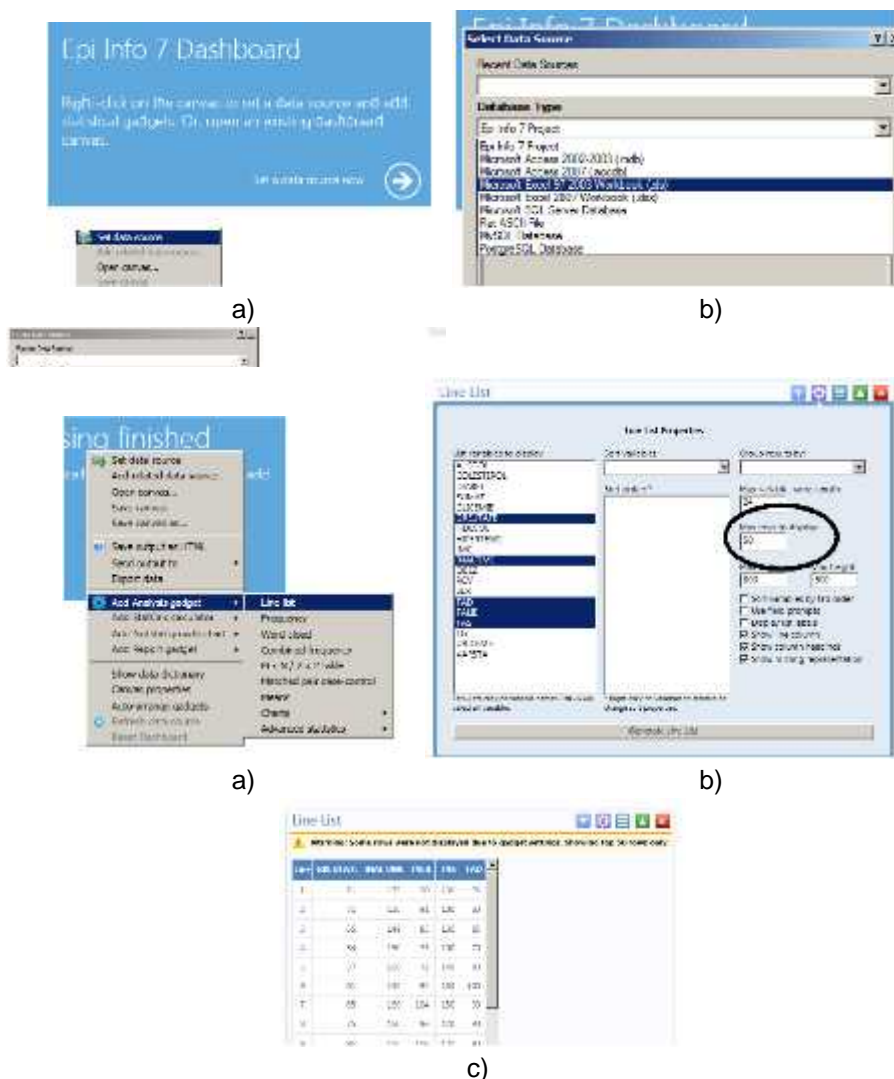


Fig.4

Afi area unor variabile (coloane)

- a) click-dreapta - Add Analysis Gadget - Line List;
- b) se aleg variabilele de afi at cu tasta CTRL
- d) se vizualizeaz rezultatul

2.2.3 S se calculeze, pentru variabila Greutate, urm toarele valori:

- Media aritmetic (Mean)
- Variația (Variance) sau dispersia (p tratul abaterii standard) adic media p tratelor abaterilor numerelor de la media lor.
- Deviația standard (Std. Dev.), este rădăcina medie pătrată (RMS) a mulțimii abaterilor fiecărui element de la media mulțimii.
- Mediana (Median), valoarea de mijloc a unei distribuții
- Modulul (Mode), este cea mai frecvent valoare dintr-o mulțime de valori

Procedura

Se execut pașii prezentați în figura 5.

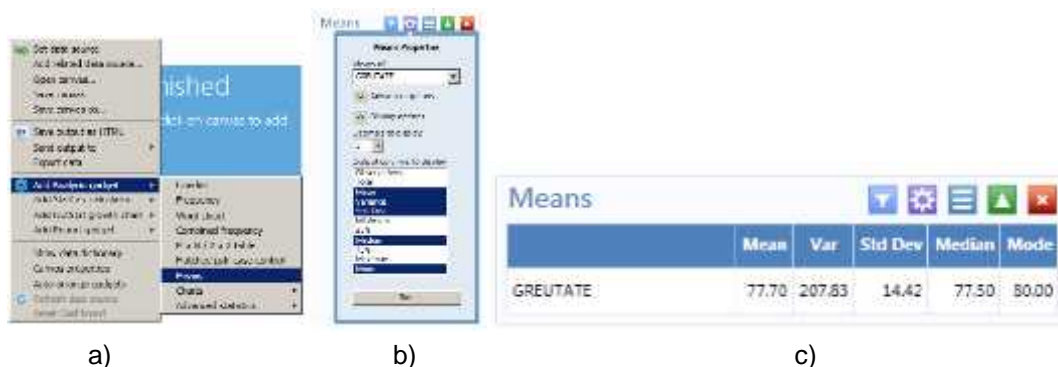


Fig.5

Calcul statistic - valori medii

- a) click-dreapta - Add Analysis Gadget - Means; b) se alege variabila de afi at, num rul de zecimale i valorile de interes (în Display Options); c) se vizualizeaz rezultatul

2.2.4 Se observ c la diferite variabile unele valori se repet , adic sunt mai mulți pacienți care au aceeași greutate, aceeași înălțime etc.. Să se calculeze, pentru variabila Înălțime valoarea frecvenței (num rul de repet ri pentru aceea i valoare). S se afi eze valoarea numeric (**Frequency**), valoarea procentual (**Percent**) i reprezentarea grafic (**Percent Bars**).

Procedura

Se execut pașii prezentați în figura 6

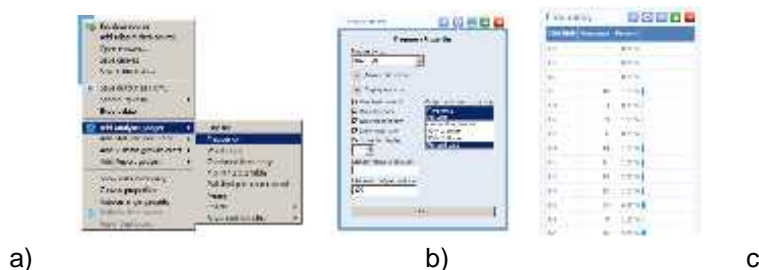


Fig.6

Calcul statistic - frecvențe

- a) click-dreapta - Add Analysis Gadget - Frequency; b) se alege variabila de afi at, num rul de zecimale i valorile de interes (în display Options); d) se vizualizeaz rezultatul

2.2.5 S se calculeze, pentru variabila Greutate, media aritmetic pe fiecare clas de în lțimi.

Procedura: Se execut pașii prezentați în figura 7

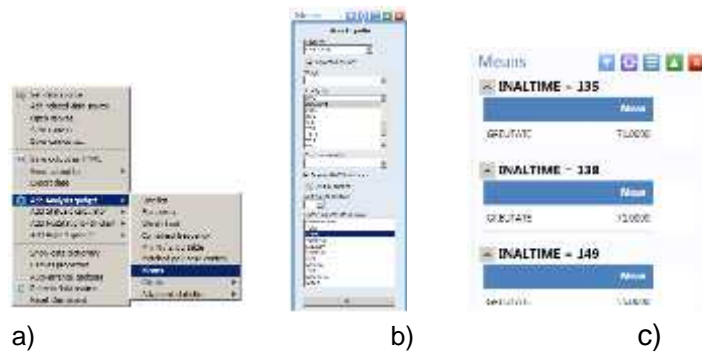


Fig.7

Calcul statistic - frecvențe

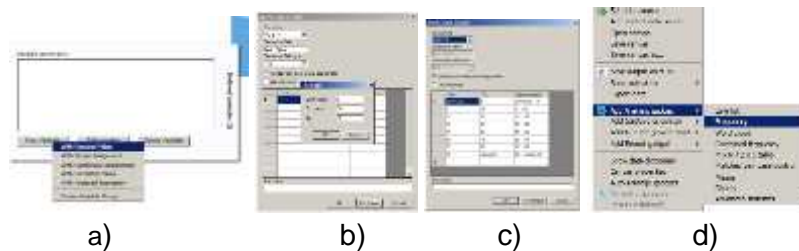
a) click-dreapta - Add Analysis Gadget - Means; b) se alege variabila de afișat, variabila grupată (Stratify By, din Advanced Options) numărul de zecimale și valorile de interes (în display Options); c) se vizualizează rezultatul

2.2.6 Se organizează variabila Vârsta pe clase de valori (0.....90, din 15 în 15) și se calculează frecvențele respective.

Procedura

Operația constă în crearea unei noi variabile, pornind de la cele existente. Se execută pașii prezentați în figura 8. Se activează fereastra **Define Variables** (stânga Dashboard) și se activează comanda **New Variable - With Recoded Value**. În fereastra deschisă semnificațiile sunt:

- **Source field:** variabila ce trebuie împărțită pe categorii (Vârsta)
- **Destination field:** numele noii variabile (Grup_varsta)
- **Destination field type:** tipul câmpului ce conține numele noii variabile (text)
- **Fill Ranges:** domenii de variație: Start value: valoare minimă (0); End value: valoarea maximă (90); By: incrementul (15)



Grup_Varsta	Frequency	Percent	Cum. Percent	Mean 95% CI	Median 95% CI
00 - 150	100	22.42 %	22.42 %	23.24 %	14.04 %
15 - 30	111	24.73 %	47.15 %	31.20 %	20.01 %
30 - 45	140	31.00 %	78.15 %	27.20 %	24.44 %
45 - 60	155	34.42 %	112.57 %	28.77 %	22.18 %
60 - 75	133	29.78 %	142.35 %	27.89 %	23.8 %
TOTAL	449	100.00 %	142.35 %		


e)

Fig.8

Calcul statistic - Reorganizarea datelor

a) activarea ferestrei Add Recode Variable; b) se specifică opțiunile; c) fereastra cu toate câmpurile completate; d) se activează comanda Frequency; e) vizualizarea rezultatelor.

3. Filtrarea datelor

Modulul Dashboard permite selectarea datelor prin aplicarea unor "filtre" - condiții logice impuse datelor. Filtrarea se poate aplica întregii baze de date sau doar datelor cuprinse într-un gadget, caz în care accesarea comenzii se face cu ajutorul butonului  din gadgetul respectiv.

Accesarea comenzii respective pentru întreaga bază de date se face prin meniul derulant lateral din dreapta ferestrei de lucru (figura 9): se ține cursorul pe meniu (fără apăsare) până la derulare. Cu ajutorul acestei comenzi se pot selecta numai valorile unor variabile ce îndeplinesc anumite condiții.

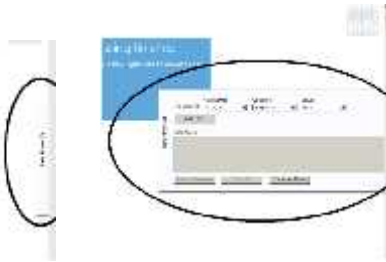


Fig.9

Modulul de analiză Visual Dashboard - activarea filtrelor

Comanda are două interfețe:

Guided (cea implicită) care conține trei câmpuri în care se specifică variabila, condiția dorită și valoarea de filtrare.

Advanced oferă posibilitatea introducerii condițiilor în linie de comandă într-o fereastră text.

Condițiile de filtrare disponibile se modifică în funcție de tipul variabilei: pentru cele booleene (da-nu) sunt disponibile doar trei în timp ce pentru cele numerice sunt disponibile nouă condiții.

Exemplu

Să se afișeze pacienții care au diabet din fișierul RiscCardiovascular.xls

Rezolvare

- Se deschide meniul de filtrare
- Se alege variabila Diabet în câmpul Field Name
- Se alege condiția "is equal to" în câmpul Operator
- Se alege valoarea Yes în câmpul Value
- Se apasă butonul Add Filter

Rezultat

Ca urmare a aplicării filtrului, sub header apare o indicație cu condiția impusă. Prin afișarea valorilor din fișier (folosind comanda Line List) vor fi afișate doar acelea care corespund valorilor adevărate ale variabilei Diabet, figura 10.

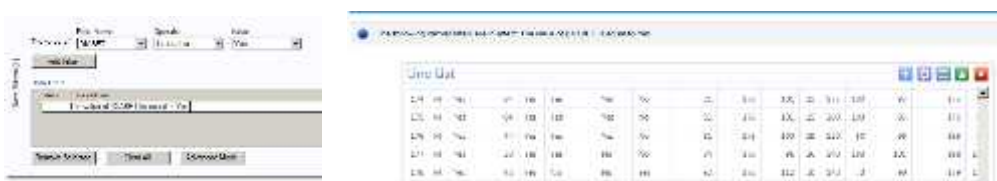


Fig.10

Modulul de analiză Visual Dashboard - aplicarea filtrelor

3.1 Combinarea filtrelor

Condițiile de filtrare pot fi combinate, astfel încât să se obțină rezultatele dorite.

Exemplu

Să se afișeze o listă cu pacienții fumători, bolnavi de diabet și alcoolici.

Rezolvare

- Se deschide meniul de filtrare
- Se apasă butonul Clear All pentru a elimina filtrele existente
- Se alege variabila Fumat în câmpul Field Name
- Se alege condiția "is equal to" în câmpul Operator
- Se alege valoarea Yes în câmpul Value
- Se apasă butonul Add Filter
- Se alege variabila Diabet în câmpul Field Name
- Se alege condiția "is equal to" în câmpul Operator
- Se alege valoarea Yes în câmpul Value
- Se apasă butonul Add Filter, din meniul nou apăsut se alege: "Add Condition with AND"
- Se alege variabila Alcool în câmpul Field Name
- Se alege condiția "is equal to" în câmpul Operator
- Se alege valoarea Yes în câmpul Value
- Se apasă butonul Add Filter, din meniul nou apăsut se alege: "Add Condition with AND"
- Se obține combinația prezentată în figura 11a

Rezultat

Ca urmare a aplicării filtrului, sub header apare o indicație cu condițiile impuse. Prin afișarea valorilor variabilelor (folosind comanda Line List) se obțin doar valorile corespunzătoare valorilor adevărate ale Fumat, Diabet și Alcool, figura 11b.



Fig.11
a) fereastra de filtrare; b) rezultat

3.2 Generarea variabilelor condiționate

Filtrele sunt disponibile și în procedura de recodare a datelor (generare de variabile noi).

Exemplu

Să se genereze o nouă variabilă cu denumirea "**Risc major**" care să includă pacienții fumători alcoolici și cu diabet, să se determine numărul acestora și să se specifice câți sunt hipertensivi.

Rezolvare

- Se deschide meniul Defined Variables
- Se creează o nouă variabilă ("New Variable") cu opțiunea "With conditional assignment"
- În fereastra comenzii (figura 12a) se completează :
 - Assign field: Risc major
 - Assign field type: text
- Se apasă butonul Create/Edit Condition

- În fereastra nou ap rut (similar cu cea de la filtrare) se introduc condițiile:
 - Fumat - yes; Alcool - yes; Diabet - yes
 - În câmpul Assign value se introduce textul: **Pacient cu risc crescut.**
 - Se afi eaz lista variabilei **Risc major**, folosind comanda Line List
 - Pentru afișarea numărului de pacienți cu risc se calculeaz frecvența variabilei **Risc major**, folosind comanda **Frequency** unde, în fereastra **Display Option**, zona **Output columns to display**, se alege doar: **Frequency**
 - Pentru afișarea numărului de pacienți cu risc care sunt și hipertensivi se calculeaz frecvența variabilei **Risc major** stratificat dup **Hipertensiune** (în zona **Advanced Option - Stratify by**), folosind comanda Frequency, cu Display Option: Frequency
- Rezultat:** Se obțin rezultatele prezentate în figura 12b,c,d.



Fig.12

- a) crearea variabilei Risc major; b) afișarea pacienților cu risc; c) numărul (frecvența) pacienților cu risc; d) numărul de pacienți hipertensivi cu risc major

4. Afișarea grafică a datelor

Pentru afișarea grafică a datelor este disponibilă comanda **Charts** din **meniul Add analysis gadget**. Sunt disponibile mai multe tipuri de grafice, fiecare dintre acestea corespunzând la diferite tipuri de variabile. De exemplu, graficul de tip Line nu este disponibil decât pentru variabile cu valori numerice, în timp ce graficul de tip Column poate fi utilizat pentru orice tip de variabil .

Exemplu

Să se reprezinte grafic variația **Greutății pacienților**.

Rezolvare

- Se activează comanda Add Analysis Gadget - Chart - Line
- Se alege în câmpul Main Variable variabila Greutate
- Se personalizează graficul din sub-fereastra Display, figura 13a

Rezultat

Se obține graficul din figura 13b.

Observație

În funcție de specificul variabilelor afișate grafic (binare - da, nu, numerice etc) unele variante de grafice pot fi mai sugestive. De exemplu, la un grafic tip coloană, cursorul mouse-ului permite afișarea valorii și a numărului de înregistrări pentru variabila respectivă în timp ce la graficul de tip linie această informație este disponibilă doar pe varfurile graficului.

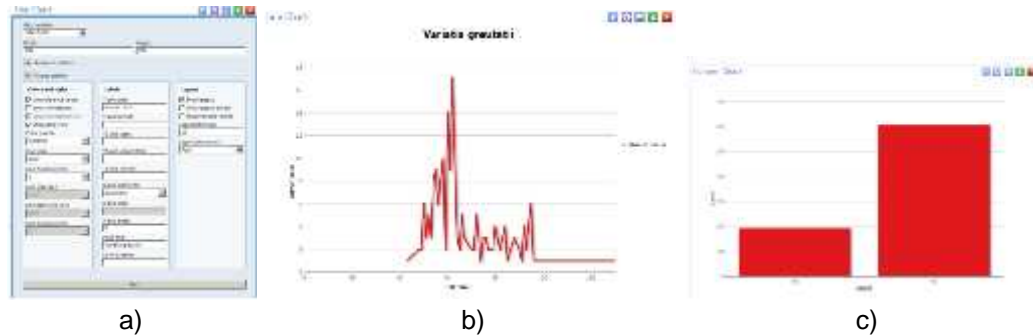


Fig.13

a) fereastra comenzii Chart; b) rezultat c) distribuția bolnavilor de diabet

Exemplu

S se reprezinte grafic pacienților cu diabet.

Rezolvare

- Se activează comanda Add Analysis Gadget - Chart - Column
- Se alege în câmpul Main Variable variabila Diabet
- Se personalizează graficul din sub-fereastra Display

Rezultat: Se obține graficul din figura 13c.

4.1 Exportul graficelor

Graficele obținute în Dashboard pot fi exportate în format html sau în documente word. În acest scop sunt disponibile comenzile: **Send output to word** și **Save output to html**. Prima comandă deschide un document word în care inserează toate gadgeturile active, cea de-a doua salvează gadgeturile active într-un fișier html. Din fișierul html graficele pot fi salvate ca imagini folosind comanda **Save image as**.

5. Exerciții

1. S se afișeze lista pacienților fumători cu glicemia între 90 și 100 și înălțimea între 165 cm și 185 cm.

2. S se creeze o nouă variabilă (**B rbați fumători hipertensivi**), care să reprezinte pacienții de sex masculin, fumători și hipertensivi, să se afișeze lista acestora și să se calculeze frecvența lor. Pentru condițiile îndeplinite variabila va avea valoarea: **B rbați cu risc cardiac**.

3. S se creeze o nouă variabilă (**Pacienți supraponderali**), care să reprezinte pacienții cu indicele IMC mai mare de 40, înălțimea mai mică de 175 cm și greutatea peste 78 kg. Să se afișeze lista acestora și să se calculeze frecvența lor simplă și stratificată după hipertensiune.

Pentru condițiile îndeplinite variabila va avea valoarea: **Pacient supraponderal**.

4. S se creeze o nouă variabilă (**Pacienți diabetici, hipertensivi, vârstnici - PDHV**), care să reprezinte pacienții cu diabet și hipertensiune, cu vârsta mai mare de 60 ani. S se calculeze frecvența acestora, grupată după sex. S se afișeze grafic rezultatul utilizând un grafic de tip "Pie".