

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. Prelucrarea datelor în modulul VISUAL DASHBOARD

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 corporală:** reprezintă indicele de masă corporală 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

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 înregistrării 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 și înregistrări din tabelul curent și un cursor pentru zoom.



Fig.1

Epi Info - Modulul de analiză 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 calculelor
- exportarea datelor în formate diferite

2. 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 posibilă afișarea tuturor valorilor (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 comandă
- **Chart** - afișează datele și rezultatele sub formă grafică. Sunt disponibile o paletă largă de tipuri de grafice

3. Aplicații

3.1 Să se importe în **DashBoard** fișierul RiscCardioVascular.xls.

Procedura

Se execută click-dreapta pe zona de lucru și se importă fișierul RiscCardioVascular.xls, urmărind pașii din figura 3.

3.2 Să se afișeze coloanele: Vârsta, Greutate, Talie, TAS, TAD.

Procedura

Se execută pașii din figura 3. Se observă în fereastra comenzii Line List, posibilitatea stabilirii numărului de înregistrări afișate. Indiferent de câte înregistrări sunt afișate în calculele ulterioare vor fi luate în considerare toate înregistrările variabile respective.

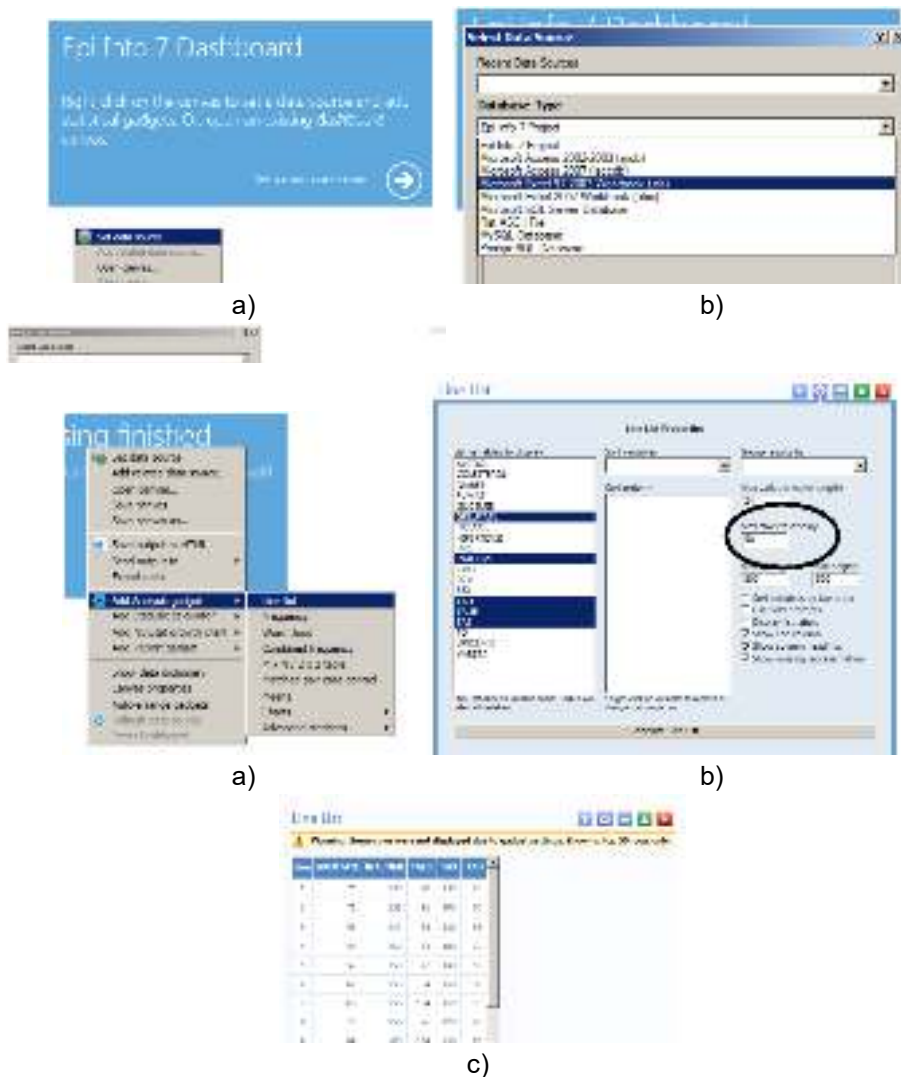


Fig.3

Afișarea unor variabile (coloane)

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

3.3 Să se calculeze, pentru variabila Greutate, următoarele valori:

- Media aritmetică (Mean)
- Variația (Variance) sau dispersia (pătratul abaterii standard) adică media pătratelor 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 frecventă valoare dintr-o mulțime de valori

Procedura

Se execută pașii prezentați în figura 4.

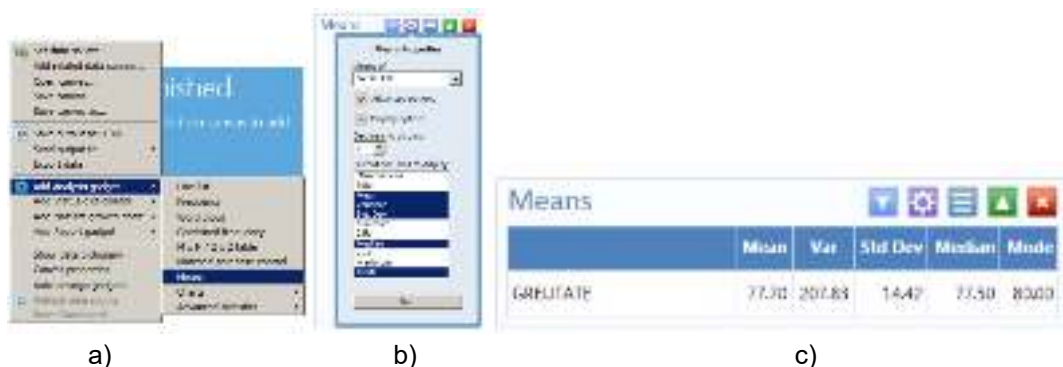


Fig.4

Calcul statistic - valori medii

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

3.4 Se observă că la diferite variabile unele valori se repetă, adică sunt mai mulți pacienți care au aceeași greutate, aceeași înălțime etc.. Să se calculeze, pentru variabila Înălțime valoarea frecvenței (numărul de repetări pentru aceeași valoare). Să se afișeze valoarea numerică (**Frequency**), valoarea procentuală (**Percent**) și reprezentarea grafică (**Percent Bars**).

Procedura

Se execută pașii prezentați în figura 5

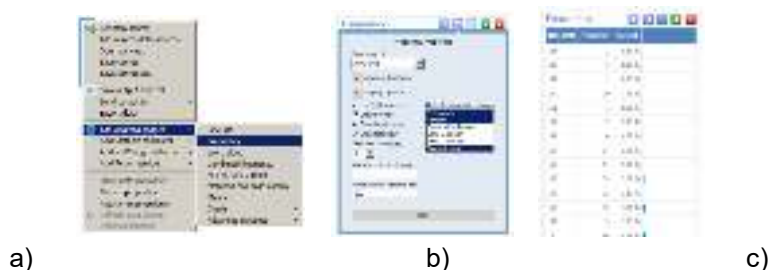


Fig.5

Calcul statistic - frecvențe

- a) click-dreapta - Add Analysis Gadget - Frequency; b) se alege variabila de afișat, numărul de zecimale și valorile de interes (în display Options); d) se vizualizează rezultatul

3.5 Să se calculeze, pentru variabila Greutate, media aritmetică pe fiecare clasă de înălțimi.

Procedura: Se execută pașii prezentați în figura 6

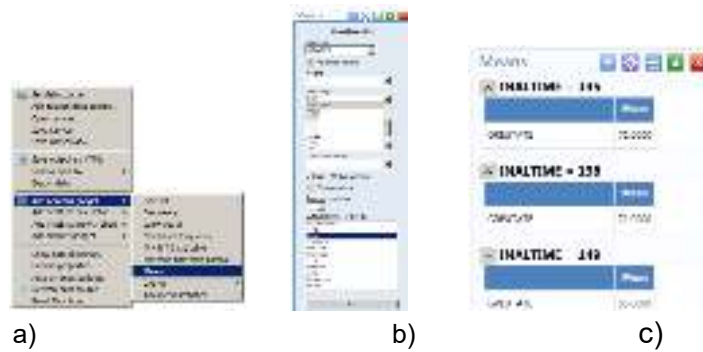


Fig.6

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

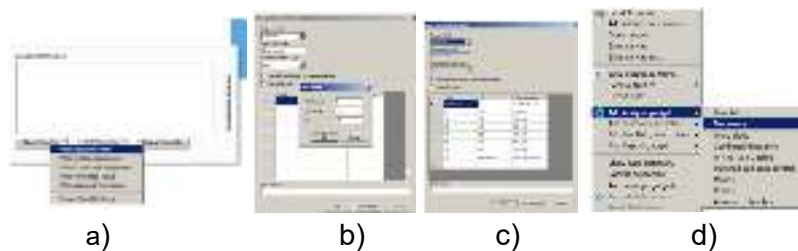
3.6 Să se organizeze variabila Vârsta pe clase de valori (0.....90, din 15 în 15) și să se calculeze frecvențele respective.

Procedura

Operația constă în crearea unei noi variabile, pornind de la cele existente. Se execută pașii prezentați în figura 7. 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)

(90); By: incrementul (15)



Exp. Value	Frequency	Percent	Cum. Percent	Relat. WC (C)	Relat. R/S (R)
10-15	11	16.7%	16.7%	1.27	1.47%
15-20	13	20.0%	36.7%	1.33	2.00%
20-25	19	28.8%	65.6%	1.52	2.88%
25-30	12	18.2%	83.8%	1.33	2.18%
30-35	5	7.6%	91.4%	1.07	1.15%
TOTAL	65	100.0%	100.0%		


e)

Fig.7

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.

4. Filtrarea datelor

Modulul Dashboard permite selectarea datelor prin aplicarea unor "filtre" - condiții logice impuse datelor. Filtrarea se poate aplica întregii bază 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 8): 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.

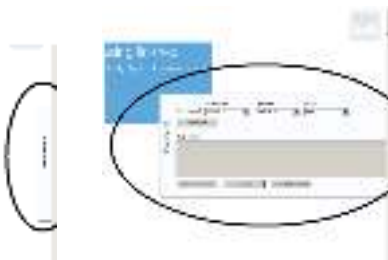


Fig.8

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 corespunzătoare valorilor adevărate ale variabilei Diabet, figura 9.

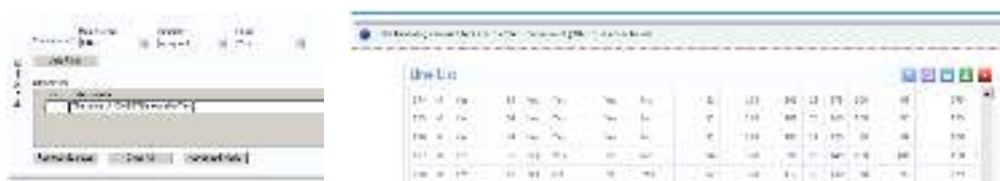


Fig.9

Modulul de analiză Visual Dashboard - aplicarea filtrelor

4.1 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 apărut 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 apărut se alege: "Add Condition with AND"
- Se obține combinația prezentată în figura 10a

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 10b.

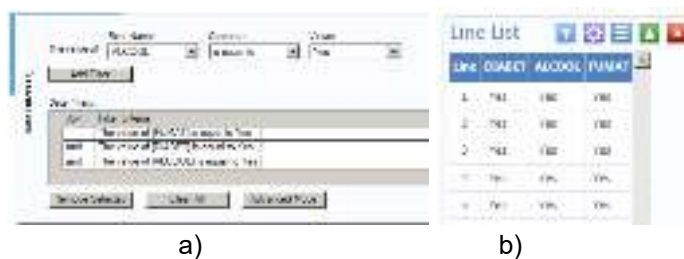


Fig.10
a) fereastra de filtrare; b) rezultat

4.2 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 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 11a) se completează:
 - Assign field: Risc major
 - Assign filed type: text
- Se apasă butonul Create/Edit Condition

- În fereastra nou apărută (similară 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 afișează lista variabilei **Risc major**, folosind comanda Line List
 - Pentru afișarea numărului de pacienți cu risc se calculează 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 calculează frecvența variabilei **Risc major** stratificată după **Hipertensiune** (în zona **Advanced Option - Stratify by**), folosind comanda Frequency, cu Display Option: Frequency
- Rezultat:** Se obțin rezultatele prezentate în figura 11b,c,d.



Fig.11

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

5. 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 variabilă.

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 12a

Rezultat

Se obține graficul din figura 12b.

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.

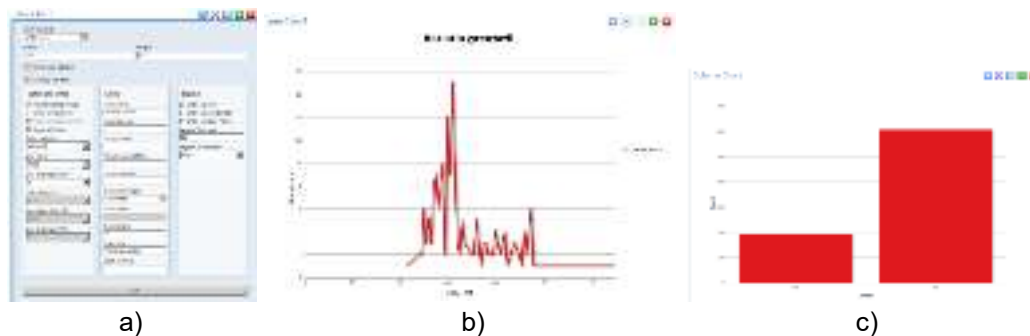


Fig.12

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

Exemplu

Să 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 12c.

5.1 Exportul graficelor

Graficele obținute în Dashboard pot exportate în format html sau în documente word. În acest scop sunt disponibile comenzile: **Send output to word** și **Save output to html**. Prima comanda 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**.

6. Cross-tabulation și testul χ^2 -

Au drept scop verificarea unor ipoteze obținute în urma centralizării datelor unei cercetări statistice și stabilirea posibilelor legături între variabile.

6.1 Tabelul de incidență 2x2

Acest tip de test se bazează pe un tabel numit "tabel de incidență" sau tabel 2x2. În acest tabel se introduc cele două variabile, dintre care una este independentă iar a doua dependentă (este influențată de cea independentă). Testul se utilizează pentru variabile cu valori binare (da-nu, bărbat-femeie, sănătos-bolnav etc.).

Acest tip de test se utilizează în general atunci când se analizează posibilitatea legăturii cauzale între expunerea la un factor de risc - **Exposure** (alcool, fumat, un anumit aliment etc.) și apariția unei afecțiuni - **Outcome** (ciroză, cancer pulmonar, toxi-infecție alimentară etc.).

Pentru efectuarea acestui test în Epi Info, se activează modulul Visual Dashboard și se regăsește comanda corespunzătoare în meniul contextual (fig.13a) și activează o fereastră (fig.13b)

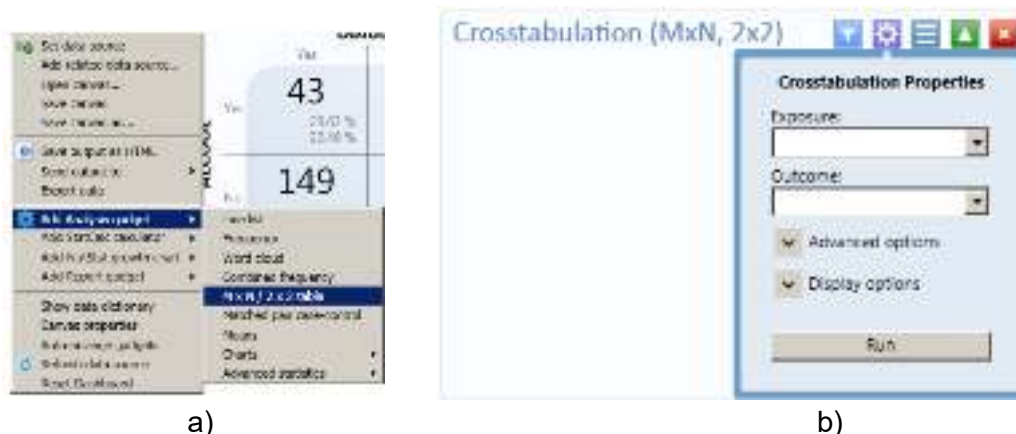


Fig.13
Comanda **Crosstabulation** (a) și fereastra corespunzătoare (b)

Fereastra comenzii **Crosstabulation** conține două câmpuri:

Exposure - pentru variabila independentă, va fi afișată pe liniile tabelului

Outcome - pentru variabila dependentă, va fi afișată pe coloanele tabelului

Rezultatul testului afirmă sau infirmă o legătură cauzală între variabila independentă (**Exposure**) și cea dependentă (**Outcome**) cu ajutorul a câtorva parametri specifici fiecărui tip de studiu - pe eșantion (case studies) sau pe întreaga populație (cohort studies). Deși pentru fiecare tip de analiză sunt semnificativi doar anumiți parametri, Epi Info afișează toate valorile, selecția urmând a fi efectuată de utilizator.

Odds ratio - este raportul dintre cota de îmbolnăviri la cei expuși și cota de îmbolnăviri la cei ne-expuși (cota reprezentând raportul pe linie - expuși --> bolnavi/sănătoși, respectiv ne-expuși --> bolnavi/sănătoși).

Acest parametru se utilizează cu precădere atunci când se analizează un eșantion și nu întreaga populație (studii de caz - case control studies)

Risk ratio - este raportul dintre riscul ca indivizii care au fost expuși la factorul de risc să dezvolte afecțiunea și riscul ca indivizii care nu au fost expuși la factorul de risc să dezvolte afecțiunea.

Acest parametru se utilizează cu precădere atunci când se analizează întreaga populație (studii complete - cohort studies).

Interpretarea valorilor:

Odds ratio sau **Risk ratio** = 1 expunerea la factorul de risc nu are efect asupra îmbolnăvirii

Odds ratio sau **Risk ratio** > 1 expunerea la factorul de risc are efect pozitiv asupra îmbolnăvirii

Odds ratio sau **Risk ratio** < 1 expunerea la factorul de risc are efect negativ asupra îmbolnăvirii

χ^2 - reprezintă rezultatul testului chi² (valabil dacă există cel puțin 30 valori pentru variabile), calculat în mai multe variante:

uncorrected - pentru χ^2 se utilizează formula Pearson, rezultatul fiind valabil dacă sunt cel puțin 100 de valori pentru variabile.

Mantel-Haenszel - pentru pentru χ^2 se utilizează formula Mantel-Haenszel, rezultatul fiind valabil dacă sunt cel puțin 30 de valori pentru variabile acestea fiind de tip ordinal.

corrected - pentru χ^2 se utilizează formula Yates, rezultatul fiind valabil dacă sunt cel puțin 30 de valori pentru variabile sau dacă toate valorile din tabel sunt mai mari decât 5

Fisher's exact - pentru χ^2 se utilizează formula Fisher, rezultatul fiind valabil dacă sunt cel mult 30 de valori pentru variabile sau dacă în tabel există valori mai mici decât 5

Interpretarea valorilor: rezultatele se interpretează luând în considerație parametrul **p**:

p - probabilitatea ca ipoteza H_0 să fie adevărată ("Între cele două variabile nu există nicio legătură"). Dacă p are valori $> 0.05 \Rightarrow H_0$ este adevărată, dacă p are valori $< 0.05 \Rightarrow H_0$ este falsă

Exemplu

Utilizând fișierul RiscCardiovascular.xls să se determine dacă există o legătură cauzală între fumat și hipertensiune

Rezolvare

- Se deschide modulul Visual Dashboard și se importă fișierul RiscCardiovascular.xls
- Se lansează comanda **MxN / 2x2 table**, din meniul **Add Analysis Gadget**
- Se alege variabila **Fumat** în câmpul **Exposure** și variabila **Hipertensiune** în câmpul **Outcome**
- Se apasă butonul **Run**

Rezultat

Ca urmare a aplicării comenzii se obține fereastra din figura 14.



Fig.14

Tabelul de incidență pentru **Fumat** și **Hipertensiune**

Având în vedere că studiul a fost efectuat pe un eșantion (Case Study) și nu pe întreaga populație, se va observa valoarea **Odds ratio = 0.7075 < 1** deci expunerea la fumat are efect negativ asupra hipertensiunii.

Deoarece sunt mai mult de 100 de înregistrări și în nicio celulă din tabelul de incidență nu există valori < 5 se urmărește valoarea testului χ^2 necorectat (Pearson) care conduce la probabilitatea $p = 0.029 < 0.05$ deci ipoteza H_0 este falsă și între cele două variabile există o legătură, adică fumatul generează hipertensiune.

7. Exerciții

1. Să se afișeze lista pacienților fumători cu glicemia între 90 și 100 și înălțimea între 165 cm și 185 cm.
2. Să se creeze o nouă variabilă (**Bărbați 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: **Bărbați cu risc cardiac**.
3. Să se verifice legătura dintre variabilele **Alcool** și **Diabet** și să se interpreteze rezultatele.