

# Software epidemiologic - OpenEpi

## 1. Obiective. Prezentare generală

Cu ajutorul acestui material ne vom familiariza cu Programul *OpenEpi*. În cadrul acestei lucrări practice vă veți însuși principalele tipuri de calcule statistice care se pot face cu acest program în studiile epidemiologice.

Programul *OpenEpi* permite obținerea de statistici pentru determinări cantitative și calitative în studiile descriptive și analitice, analize stratificate cu limite de încredere exacte, analiza perechilor și analize persoană-timp, calcularea mărimii eșantionului și a puterii studiului, furnizarea de numere aleatoare, sensibilitate, specificitate și alte statistici pentru evaluare, tabele R x C, chi-pătrat pentru doză-răspuns, precum și link-uri către alte site-uri utile. *OpenEpi* este gratuit și o sursă deschisă pentru statistici epidemiologice. Acesta poate fi rulat de pe un server web sau poate fi descărcat și folosit fără o conexiune web. URL-ul este [www.openepi.com](http://www.openepi.com). Nu este necesar un server. Programele sunt scrise în JavaScript și HTML și ar trebui să fie compatibile cu recente Linux, Mac și browsere pentru PC, indiferent de sistemul de operare.

Rezultatele testelor sunt furnizate pentru fiecare modul, astfel încât să poată fi analizată fiabilitatea, deși întotdeauna se recomandă să se verifice rezultatele importante cu software-ul de la mai multe surse. Sunt furnizate de asemenea link-uri la sute de programe de calculare pe internet.

*OpenEpi* a fost dezvoltat pentru a efectua analize disponibile în modulele StatCalc și EpiTable din versiunea DOS a Epi Info și pentru a îmbunătăți tipurile de analize furnizate de aceste module și pentru a furniza o serie de instrumente și calcule care nu sunt în prezent disponibile în programul Epi Info. Acesta este primul pas spre un set complet de software epidemiologic disponibil pe web. *OpenEpi* poate fi considerat ca un program de însoțire pentru Epi Info și alte programe, cum ar fi SAS, PSPP, SPSS, Stata, SYSTAT, Minitab, Epidata și R.

Tipurile de calcule, efectuate de către *OpenEpi* în prezent, includ:

- diferite intervale de încredere pentru proporții, rate, rata mortalității standardizate, medie, mediana, percentilele;
- tabele 2x2 brute și stratificate pentru numere și rate;
- analiză pentru studiile caz-control pereche;
- test pentru tendință cu date numerice;
- test t pentru eșantioane independente și one-way ANOVA;
- analiza testelor de diagnostic și de screening cu trasarea curbilor ROC;
- calculul mărimii eșantionului pentru proporții, în anchetele transversale, studiile caz-control nepereche, studiile de cohortă, studiile clinice controlate randomizate și pentru compararea a două valori medii;

- calculul puterii pentru proporții (în studii caz-control nepereche, transversale, de cohortă, studiile clinice controlate randomizate) și pentru compararea a două valori medii;
- generator aleatoriu de numere.



Figura 1. Programul OpenEpi

## 2. Testul t pentru două eșantioane independente

Acest modul compară valorile medii a două eșantioane independente. Este necesară introducerea intervalului de încredere dorit, mărimea eșantionului, media și deviația standard (sau eroarea standard) pentru fiecare din cele două grupuri care se compară. Programul va testa dacă există o diferență semnificativă între mediile celor două eșantioane. Se face și testul Hartley pentru egalitatea varianțelor. Dacă acesta este semnificativ din punct de vedere statistic ( $p < 0.05$ ), așa cum se întâmplă în exemplul de mai jos, se ia în considerare rândul *unequal variance* din tabelul cu rezultate. Dacă testul Hartley pentru egalitatea varianțelor nu este semnificativ din punct de vedere statistic, se ia în considerare rândul *equal variance* din tabelul cu rezultate. Diferența medie, cu intervalul de încredere este de asemenea afișată.



### Exercițiu

Se compară două valori medii obținute pe două grupuri: grupul 1, 7 subiecți, media=11,57, deviația standard=8,81; grupul 2, 18 subiecți, media=7,44, deviația standard=3,698. Se introduc datele în tabelul de mai jos și se apasă butonul calculate.

**Two-Sample Independent t Test**

Calculate Clear

Confidence Interval (%) (two-sided)  Enter a value between 0 and 100, usually 95%

	Sample Size	Mean	Std. Dev.	(or)	Std. Error
Group 1	7	11.57	8.81		
Group 2	18	7.44	3.698		

Figura 2. Tabelul pentru introducerea de date

**Two-Sample Independent t Test**

Input Data

Two-sided confidence interval 95%

	Sample size	Mean	Std. Dev.	Std. Error
Group-1	7	11.57	8.81	
Group-2	18	7.44	3.698	

Result	z statistics	df	p-value <sup>1</sup>	Mean Difference	Lower Limit	Upper Limit
Equal variance	1.68286	23	0.1059	4.13	-0.946799	9.2068
Unequal variance	1.19986	7	0.2692	4.13	-4.00922	12.2692

Test for equality of variance<sup>2</sup>

	F statistics	df(numerator,denominator)	p-value <sup>1</sup>
	5.67568	6,17	0.004296

<sup>1</sup> p-value (two-tailed)  
<sup>2</sup> Hartley's test for equality of variance

Figura 3. Tabelul pentru introducerea de date;  
testul pentru egalitatea varianțelor

### Soluția

Deoarece testul pentru egalitatea varianțelor este semnificativ din punct de vedere statistic ( $p=0.004296$ ) se va alege rezultatul de pe rândul varianțe inegale (unequal variance) care dă o valoare a lui  $p=0,2692$ , adică diferența este nesemnificativă statistic.

## 3. Analiza varianței (ANOVA)

Testul ANOVA compară valorile medii pentru două sau mai multe eșantioane independente. Este necesară introducerea mărimii eșantionului, a mediei și a deviației standard (sau erorii standard)

pentru fiecare grup care se compară. Programul va testa, dacă există, o diferență semnificativă între mediile eșantioanelor. Intervale de încredere pentru fiecare valoare medie individuală, de asemenea, sunt afișate.

#### 4. Tabelele 2x2

Tabelele 2x2 sunt folosite pentru a evalua asocierea dintre un posibil factor de risc ("expunere") și un rezultat ("boala"). Cifrele corespunzătoare tuturor celor patru combinații posibile de evenimente în populația studiată sunt introduse în celulele corespunzătoare. Tabelul poate fi modificat (rotit sau "în oglindă") astfel încât rândurile sau coloanele care reprezintă expunerea, precum și denumirile coloanelor (+) și (-) pot fi în orice ordine pentru a se potrivi cu teoria din principalele manuale de epidemiologie. Poate fi folosit un singur tabel sau se pot introduce mai multe straturi.

Statisticile produse includ testele Fisher exact și mid p, testul chi pătrat, odds ratio, estimarea probabilității maxime pentru odds ratio, raportul risc/prevalența (riscul relativ), diferența de risc și fracțiile etiologice cu limite de încredere, produse prin mai multe metode, cu analiză stratificată.



#### Exercițiu

Un studiu de cohortă desfășurat pe 162 de subiecți examinează relația dintre expunere și boală. Din totalul celor 102 subiecți expuși, 66 au făcut boala, comparativ cu 28 din totalul celor 60 de subiecți neexpuși. Se introduc datele în tabelul de mai jos și se apasă butonul calculate.

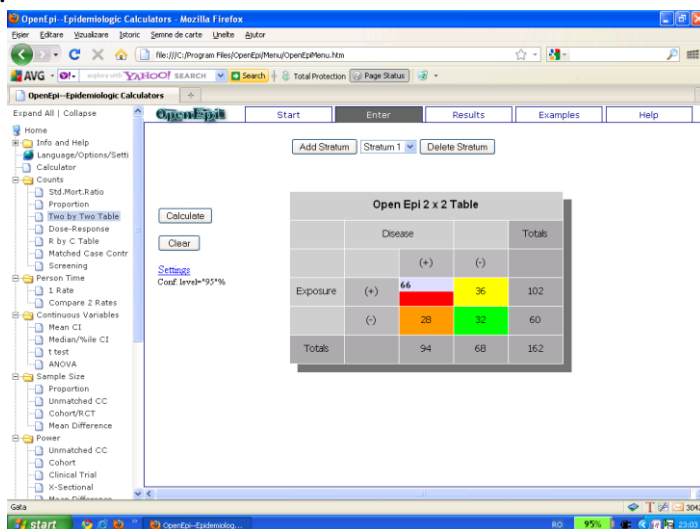


Figura 4. Tabelul 2x2

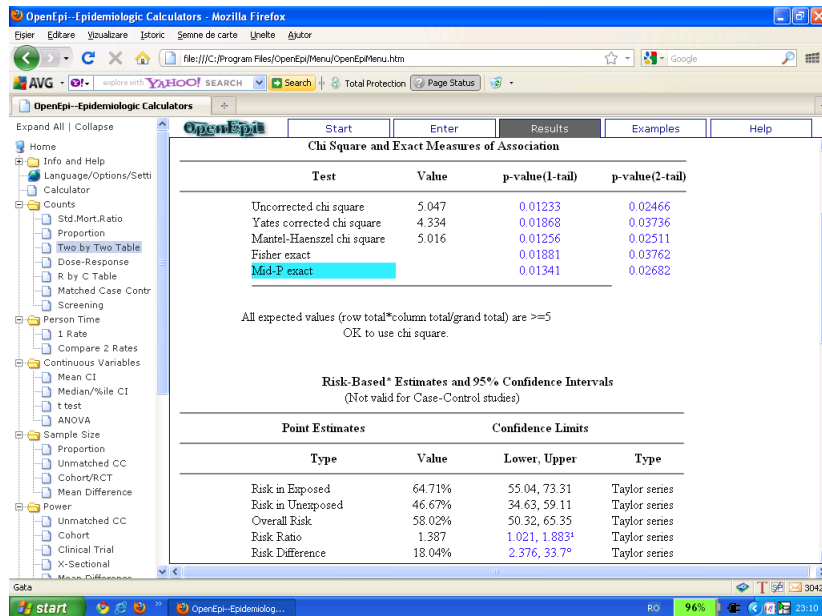


Figura 5. Test de referință

## 5. Evaluarea testelor de diagnostic și de screening

Acest modul ajută la evaluarea performanțelor unui test de diagnostic sau a unei proceduri de screening față de un *test de referință* sau *standard de aur*. Se pleacă de la presupunerea că există un standard de referință (care permite diagnosticul pozitiv). Modulul vă permite să alegeți tipul de test de diagnosticare pe baza rezultatelor pe care le raportează. Apoi, modulul folosește statistici relevante pentru a evalua performanța testului de diagnostic. Pentru două sau mai multe niveluri de expunere, programul calculează *sensibilitatea*, *specificitatea*, *valoarea predictivă pozitivă și negativă*, *acuratețea diagnosticului*, *raporturile probabilităților* (likelihood ratios), *șansa diagnosticului* (diagnostic odds), *indicele Kappa* a lui Cohen, *reducerea entropiei* și un *indice de eroare* (bias index).



### Exercițiu

Vi se propune să comparați rezultatele unui nou test pentru depistarea cancerului mamar cu *testul de referință*. În acest scop, 200 de femei cu vârsta peste 50 de ani au participat la studiu. Dintre cele 200 de femei, 20 aveau valori crescute față de normal și doar 8 dintre acestea 20 aveau un cancer mamar confirmat histopatologic. Se introduc datele în tabelul de mai jos și se apasă butonul calculate.

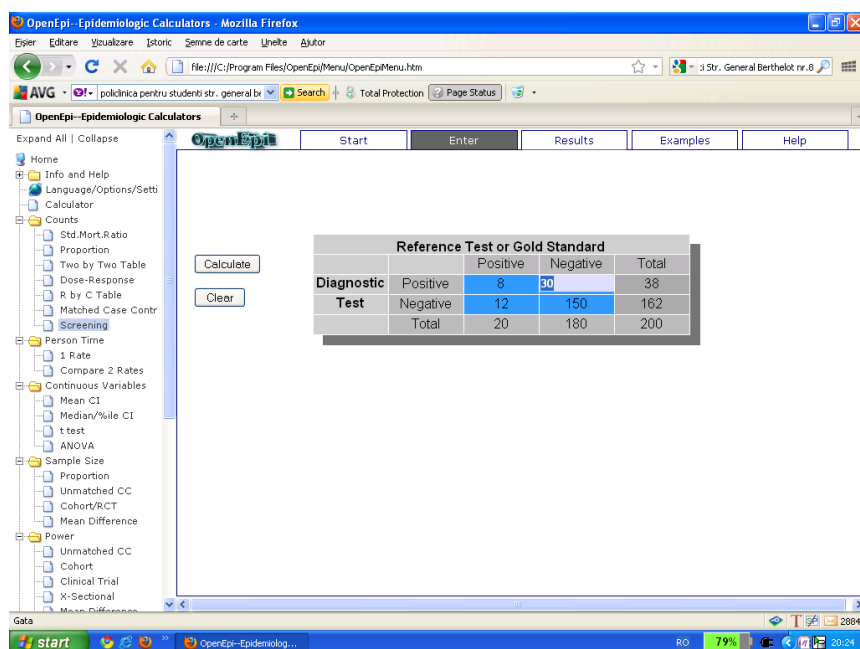


Figura 6. Introducerea de date

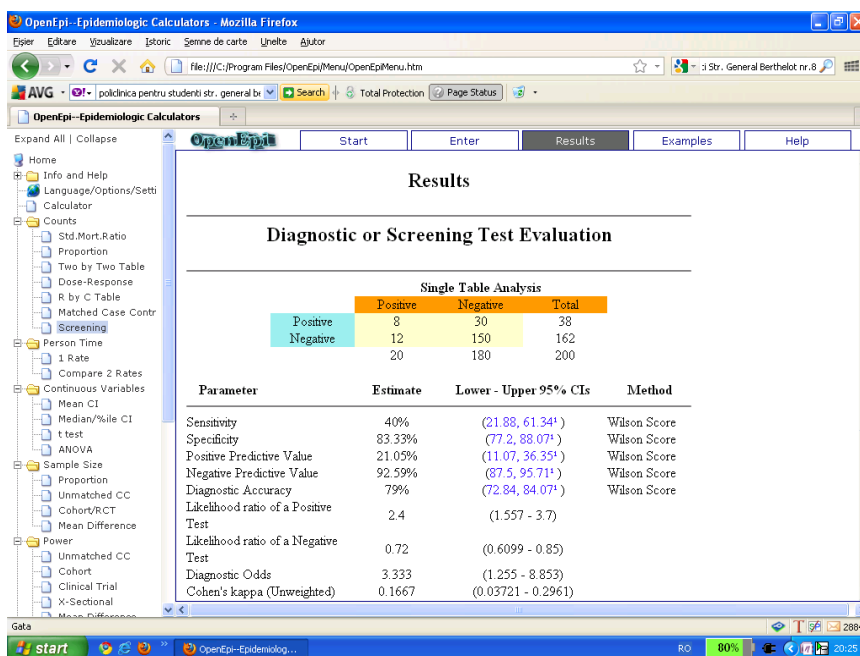


Figura 7. Testul de referință

Dacă sunt specificate mai mult de două niveluri ale rezultatelor, se trasează un grafic (*curba ROC*), care prezintă proporțiile rezultatelor adevărat pozitive față de rezultatele fals pozitive, cu calcularea ariei de sub curbă, așa cum se vede în exemplul de mai jos.

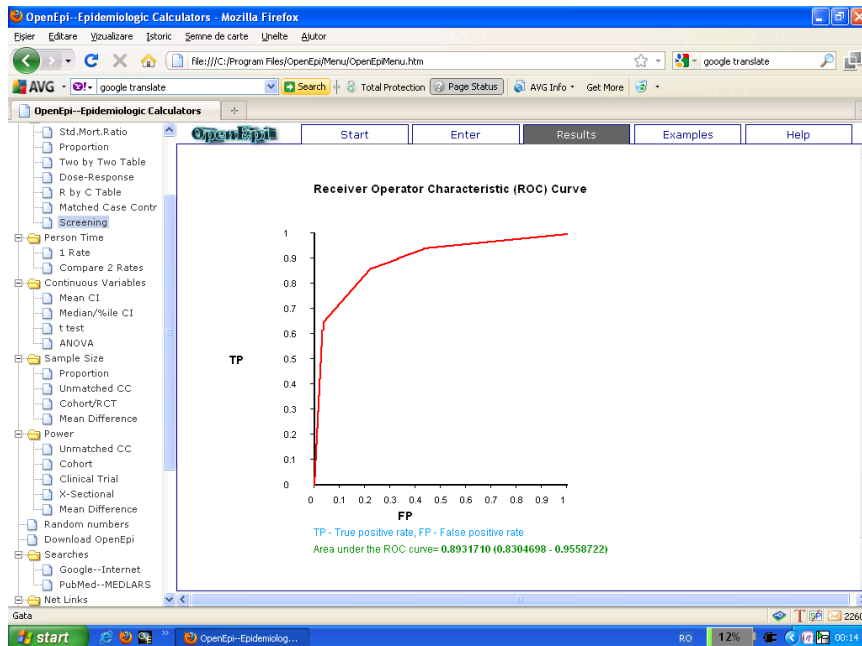


Figura 8. Obținerea curbei ROC

## 6. Calculul puterii unui studiu atunci când se compară două valori medii

Acest modul estimează puterea pentru compararea a două valori medii. În urma introducerii valorilor medii (sau a diferenței dintre ele), a deviației standard (sau a varianței) și a mărimii fiecărui grup care se compară în tabelul prezentat mai jos, se determină puterea pentru intervalul de încredere dorit.



### Exercițiu

Dorim să determinăm care este puterea unui studiu de a demonstra că există o diferență semnificativă statistic pentru un interval de încredere de 95%. Cele două grupuri sunt formate din câte 100 de subiecți. Pentru grupul 1, media=125 și deviația standard=15,34; pentru grupul 2, media=120 și deviația standard=18,23. Se introduc datele în tabelul de mai jos și se apasă butonul calculate.

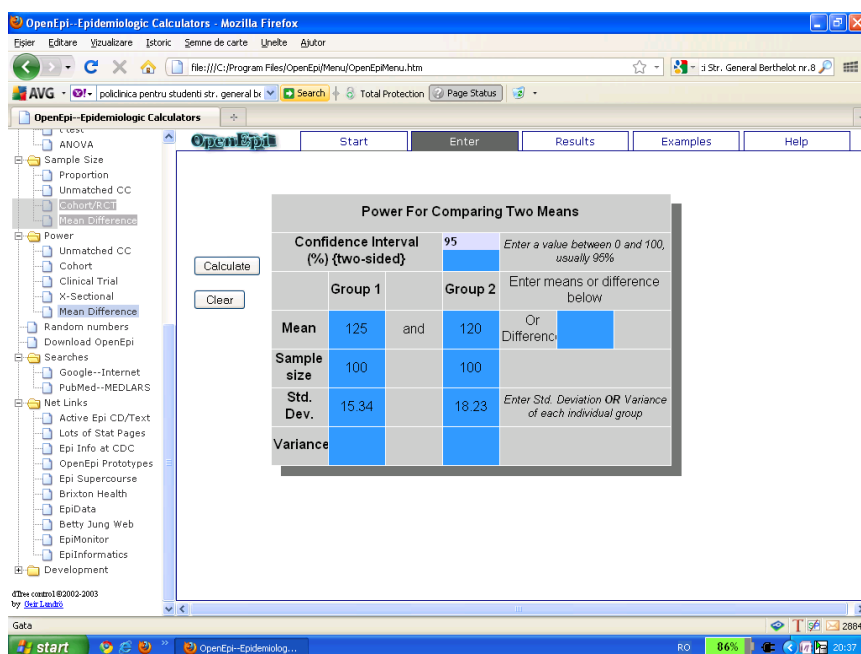


Figura 9. Introducerea de date

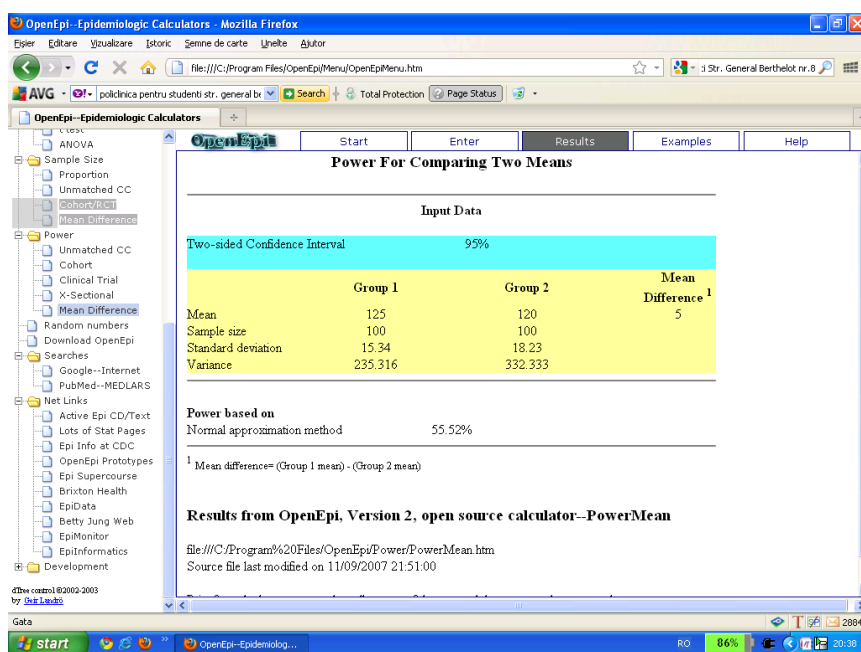


Figura 10. Tabel de comparare pentru două grupuri

## 7. Concluzii

În decursul acestei ore v-ați familiarizat cu aplicațiile programului OpenEpi:

- ✚ ați învățat să folosiți testul t pentru două eșantioane independente atunci când comparați două valori medii, respectiv testul ANOVA pentru mai multe eșantioane independente;
- ✚ ați învățat să folosiți tabelele de contingență 2x2 pentru a evalua asocierea dintre un posibil factor de risc ("expunere") și un rezultat ("boala");
- ✚ ați învățat să evaluați performanțele unui test de diagnostic sau a unei proceduri de screening;
- ✚ ați învățat să estimați puterea unui studiu atunci când se compară două valori medii.

### Informații suplimentare

Vernic CV, Mihalaș GI, Apostol SA, Cătu CO, Mada L, Tudor A. Informatică Medicală Aplicată în Balneo-fiziokinetoterapie și Recuperare. Editura Victor Babeș Timișoara, 2009.

<http://en.wikipedia.org/wiki/OpenEpi>

<http://www.openepi.com>

