

EPI INFO

- Cross-tabulation și testul χ^2 -

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

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, bolnav-sănătos 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, toxinfecție alimentară etc.).

Pentru efectuarea acestui test în Epi Info, se activează modulul Visual Dashboard și se regsește comanda corespunzătoare în meniul contextual (fig.1a) și activează o fereastră (fig.1b)



Fig.1

Comanda **Crosstabulation** (a) și fereastra corespunzătoare (b)

Fereastra comenzii **Crosstabulation** conține două câmpuri:

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

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

Rezultatul testului afirmă sau infirmă o legătură causală între variabila independentă (**Exposure**) și cea dependentă (**Outcome**) cu ajutorul a câtorva parametri specifici fiecărui tip de studiu - pe cazuri (case studies) sau pe întreaga populație (cohort studies). De obicei 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 precizie deosebită atunci când se analizează un caz-control (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 precizie deosebită 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

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

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

Mantel-Haenszel - pentru t^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 t^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 t^2 se utilizează formula Fisher, rezultatul fiind valabil dacă sunt cel puțin 30 de valori pentru variabile sau dacă în tabel există valori mai mici decât 5

Interpretarea valorilor: rezultatele se interpretează luând în considerare 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 adevărat, dacă **p** are valori $< 0.05 \Rightarrow H_0$ este fals

Exemplu

Utilizând fișierul RiscCardiovascular.xls se determină dacă există o legătură causală 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 2.



Fig.2
Tabelul de incidență pentru Fumat și Hipertensiune

Având în vedere că studiul a fost efectuat pe un e antion (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 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.

2. Tabelul de incidență 2x2xn

Acest tip de tabel se utilizează pentru variabilele stratificate - atunci când este posibil influența a mai multor factori (variabile) asupra variabilei dependente. Se utilizează în special pentru excluderea factorilor "paraziți" - care apar în m surători și au o influență atât asupra variabilei dependente cât și a celei independente. De exemplu se pot enumera: sexul, vârsta, greutatea, etc..

De exemplu, dacă se urmărește influența **Alcoolului** asupra apariției **hipertensiunii** și se presupune că și fumatul are o influență atât asupra hipertensiunii cât și a consumului de alcool, se va folosi un tabel 2x2 stratificat, cu doi factori unul principal (**Alcool**) și unul de influență (Stratify by: **Fumat**), figura 3.



a)



Fig.3

Tabelul de incidență pentru **Alcool**, **Fumat**, **Hipertensiune**
a) set ri; b) tabel 2x2

Summary Results				
Odds Ratio	Estimate	Lower	Upper	
Crude (cross product)	1.3100	0.9363	1.8288	(T)
Crude (MLE)	1.3095	0.9367	1.8288	(M)
Fisher-Exact		0.9242	1.8531	(F)
Adjusted (MH)	1.3684	0.9778	1.9151	(R)
Adjusted (MLE)	1.3700	0.9635	1.9464	(F)
(T=Taylor series, R=RGE, M=Exact mid-P, F=Fisher exact)				
Risk Ratio	Estimate	Lower	Upper	
Crude	1.1686	0.9694	1.4088	
Adjusted	1.2005	0.9522	1.4526	
Chi Square	X ²	2 Tailed P		
Uncorrected (MH)	3.3681	0.0664701082		
Corrected (MH)	3.0572	0.0803813006		
Homogeneity Tests			X ²	2 Tailed P
Breslow-Day-Tarone test for Odds Ratios			2.3168	0.1279852463
Breslow-Day test for Odds Ratios			2.3094	0.1285524129
Breslow-Day test for Risk Ratios			2.7556	0.0969124221

Fig.4

Rezultatele Tabelului de inciden pentru **Fumat, alcool** și **Hipertensiune**

Rezultatele (figura 4) se interpretează astfel:

Odds/Risk ratio:

- valorile din fiecare strat se interpretează ca și la tabelele 2x2
- dacă valorile generale (**Crude** și **Adjusted**) sunt diferite se poate concluziona că factorul secundar este "parazit" și în realitate cauza este factorul secundar și nu cel principal
- dacă valorile generale (**Crude** și **Adjusted**) nu sunt diferite se poate concluziona că factorul secundar nu este "parazit" și în realitate cauza este factorul principal și nu cel secundar.

Dacă există mai mulți factori presupuiți paraziți se utilizează comanda **Add StatCalc - Tables 2x2xn**, din meniul contextual. În fereaștră apărută se completează pe fiecare strat valorile corespunzătoare factorului considerat, pe primul strat fiind plasat factorul principal. Pentru a obține valorile pentru fiecare factor, se rulează tabele 2x2 pentru fiecare factor și se referă la valorile din celule.

Exerciții

1. Să se verifice legătura dintre variabilele **Alcool** și **Diabet** și să se interpreteze rezultatele.
2. Să se verifice legătura dintre variabilele **Sex, Fumat** și **Hipertensiune** și să se interpreteze rezultatele.
3. Să se verifice legătura dintre variabilele, **Sex, Alcool, Fumat** și **Diabet** și să se interpreteze rezultatele.