

Testarea ipotezelor statistice (1)

Plan de curs

Teste statistice

Etapele testelor statistice

Testul Hi-pătrat: analiza asocierii dintre două variabile calitative

Teste statistice: etape

Parametrul	Este vorba de <u>parametrul populației</u> : μ = media sau p =frecvența caracteristicii în populația de interes
Ipoteze	<p><u>Ipoteza nulă</u> (H_0): $p_F = p_M$</p> <p>Exemplu: Frecvența hematomului post punctie venoasă la pacienții de gen feminin (p_F) nu diferă semnificativ de frecvența hematomului post punctie venoasă bărbați (p_M)</p> <p><u>Ipoteza alternativă</u> (H_A) pentru un test bilateral: $p_F \neq p_M$</p> <p>Frecvența hematomului post punctie venoasă la pacienții de gen feminin (p_F) nu diferă semnificativ hematomului post punctie venoasă bărbați (p_M)</p>
Asumpții/Condiții	Orice test are anumite condiții în care se aplică. De exemplu, un test pe medii se aplică doar dacă datele urmează o distribuție normală (distribuția lui Gauss)
Alegerea testului	Alegerea testului se face în funcție de tipul variabilelor (calitative/cantitative) și condițiile de aplicare ale testului (ex. distribuția normală în cazul mediei)

Teste statistice: Simboluri și definiții utilizate în testarea ipotezelor statistice

Parametrul	Este vorba de <u>parametrul populației</u> : μ în cazul mediei, p în cazul frecvenței
Calcularea statisticii testului	Fiecare test are formula proprie de calculare a statisticii testului
Calcularea probabilității p asociate statisticii testului	Pentru fiecare statistică a testului se calculează probabilitatea (p) de a observa dacă acea statistică s-a obținut doar datorită șansei
Decizia statistică	Valoarea $p < \alpha \rightarrow$ respinge H_0 – rezultat semnificativ statistic Valoarea $p \geq \alpha \rightarrow H_0$ nu se respinge – rezultat nesemnificativ statistic
Concluzia	Semnificativ statistic \neq semnificativ clinic

Nivelul de semnificație alfa (α) = probabilitatea erorii de tip I (probabilitatea de a respinge ipoteza nulă în condițiile în care H_0 este adevărată)

Exemple de articole științifice medicale

> Nurs Rep. 2021 Jul 12;11(3):530-535. doi: 10.3390/nursrep11030050.

Insomnia Prevalence among Italian Night-Shift Nurses

Nicola Ielapi ^{1, 2}, Michele Andreucci ³, Umberto Marcello Bracale ⁴, Davide Costa ^{2, 5}, Egidio Bevacqua ^{2, 6}, Andrea Bitonti ⁷, Sabrina Mellace ⁸, Gianluca Buffone ⁹, Stefano Candido ¹⁰, Michele Provenzano ⁶, Raffaele Serra ^{2, 6}

Affiliations + expand

PMID: 34968328 PMCID: PMC8608071 DOI: 10.3390/nursrep11030050

Free PMC article

FULL TEXT LINKS

FULL TEXT
OPEN ACCESS

FREE
Full text

ACTIONS

Cite

Collections

Table 1. Clinical and demographic characteristics of subjects enrolled in the study. The features are presented for the whole study population (on the left) and according to the presence or absence of insomnia.

Questions	Overall (n = 2'355)	Insomnia (n = 1'542)	NO-Insomnia (n = 813)	p
Age, years	40.4 ± 10.3	39.9 ± 10.5	41.4 ± 10.0	0.001
Male gender, %	15.0	14.6	15.7	0.456
Smoking habit, %	34.6	37.2	29.5	<0.001
Months of work, median [IQR]	168 [72–300]	144 [72–300]	180 [84–312]	<0.001
Night shifts, %	99.7	99.7	99.6	0.642
Months spent in night shifts, median [IQR]	144 [60–264]	132 [60–252]	156 [60–276]	0.007
Night shifts per month, number	6.3 ± 1.4	6.3 ± 1.4	6.2 ± 1.4	0.163
Time to reach workplace, minutes	45 [45–65]	45 [44–65]	45 [45–65]	0.256
Rest time, minutes	180 [4–240]	180 [4–240]	180 [4–240]	0.821
Rest period after night shift, %	50.8	49.3	53.5	0.055
Rest in the afternoon, minutes	30 [0–120]	30 [0–120]	60 [0–120]	0.016
Number of coffees, mean	2.5 ± 1.5	2.5 ± 1.5	2.5 ± 1.4	0.698
Consumption of coffee during work, %	71.8	72.6	70.2	0.232
Number of coffees during night shift, mean	1.4 ± 1.1	1.4 ± 1.1	1.2 ± 1.5	0.178
Work in emergency department, %	38.4	37.6	39.9	0.250
History of insomnia previous to start working, %	25.7	29.4	18.7	<0.001
Referral to a Medical Specialist, %	9.0	10.2	6.8	0.006
Impairment (or new onset) of insomnia symptoms during COVID-19 period, %	75.2	82.5	61.4	<0.001
Insomnia symptoms impairment (or new onset) attributed to COVID-19 period, %	62.8	69.7	49.6	<0.001
Changing bedtime times during COVID-19 period, %	27.7	22.4	37.9	<0.001

Link articol medical:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8608071/pdf/nursrep-11-00050.pdf>

Testul Hi-pătrat (de independență)

- **Utilitate**

- Testarea asocierii / relației / legăturii dintre două variabile calitative
- Fiecare dintre ele cu 2 sau mai multe categorii
 - Presupuse mutual exclusive.

- **Exemple**

- Relatia dintre:
 - *un factor (posibil) de risc și o boală (sau un eveniment)*
 - Hematom post puncție venoasă (da/nu) și dificultatea prelevării probei de sange (da/nu)
 - Consumul excesiv de alcool (da/nu) și ciroza hepatică (prezentă/absentă)

Testul Hi-pătrat

- **Formularea ipotezei nule** (exemple de formulări):
 - Există asociere dintre un factor de risc (prezent/absent) și boala (prezentă/absentă)?
- **Ipoteza nulă** (formulari echivalente):
 - H_0 : Nu **există** relație statistic semnificativă între factorul de studiu și boală
 - H_0 : Nu **există** asociere semnificativa statistic între factorul de studiu și boală
 - H_0 : Nu **există** diferență semnificativă statistic între frecvența bolii la pacienții expuși factorului de risc și frecvența bolii la cei neexpuși factorului.

Testul Hi-patrat

- **Formularea ipotezei alternativă** (exemple de formulari):
 - Există asociere dintre un factor de risc (prezent/absent) și boală (prezentă/absentă)?
 - **Ipoteza alternativă** (formulari echivalente):
 - H_1 : **există** relație statistic semnificativă între factorul de studiu și boală pe populația de interes
 - H_1 : **există** asociere semnificativă statistic între factorul de studiu și boală pe populația de interes
 - H_1 : **există** diferență semnificativă statistic între frecvența bolii la pacienții expuși factorului și frecvența bolii la cei neexpuși factorului pe populația de interes

Testul Hi-pătrat (χ^2)

- **Testul Hi-pătrat** se aplică dacă următoarele **asumpții** sunt îndeplinite:
- DOUĂ variabile CALITATIVE
- observații independente în eșantioane
- eșantioane independente (putem considera că categoriile unei variabile indică eșantioanele care sunt comparate – ex.: fără factor de risc, tratament 1 sau 2, ...)
- **Mai mult (>) 80% din frecvențele teoretice trebuie să fie mai mari decât 5**
- Dacă mai mult de 20% dintre frecvențele teoretice sunt <5 ȘI toate frecvențele teoretice sunt mai mari decât 1 → **testul exact Hi-patrat cu corecția lui Yates**
- dacă cel puțin 20% din frecvențele teoretice sunt mai mici de 5, sau există cel puțin o frecvență mai mică de 1 → **testul exact al lui Fisher**

Testarea asocierii: Hi-pătrat

Exemplu:

- S-a realizat un studiu pentru a investiga asocierea dintre apariția hematomului după puncția venoasă și genul pacientului.
- Eșantionul de studiu a cuprins un număr de 100 pacienți adulți aleși aleatoriu, care s-au prezentat în perioada Decembrie 2015-Decembrie 2017 la secția ATI din cadrul unei clinici universitare din Cluj Napoca și cărora li s-a efectuat puncție venoasă.

Cu un risc de eroare de 5%, putem afirma că există o asociere semnificativă între hematomul post puncție venoasă și genul pacientului la adulți?

(sau altfel spus: Există diferență semnificativă statistic între frecvența hematomului la femei și frecvența hematomului la bărbați??)

Testarea asocierii: Hi-pătrat

- Repartiția pacienților în funcție de cele 2 variabile de interes în eșantionul de studiu este dată în tabelul de contingență observat alăturat:

	Hematom (da)	Hematom (nu)	Total
F	31	22	53
M	17	30	47
Total	48	52	100

Testul Hi-pătrat

- **Etapa 1:** Formularea ipotezelor statistice H_0 (ipoteza nula) si H_1 (ipoteza alternativa)

H_0 : **Nu există o asociere semnificativ statistică** între hematomul post puncție venoasă și genul pacientului la **pacienții adulți**.

SAU

H_0 : Hematomul post punctie venoasa și genul pacientului sunt **independente** din punct de vedere statistic **pacienții adulți**.

SAU

H_0 : **Nu există o diferență semnificativă statistic** între **frecvența** hematomului post punctie venoasă la pacienții adulți de gen feminine (populația de gen feminin) și **frecvența** hematomului la pacienții de gen masculin (populația de gen masculin).

Testul Hi-pătrat

H_1 : **Există o asociere semnificativ statistică** între hematomul post puncție venoasă și genul pacientului.

SAU

H_1 : Hematomul post puncție venoasă și genul pacientului sunt **dependente** din punct de vedere statistic

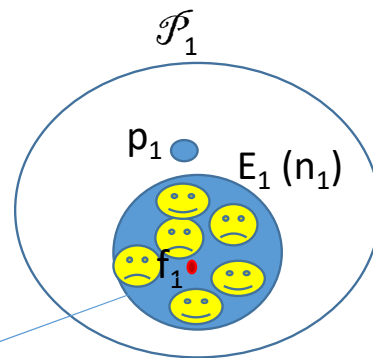
SAU

H_1 : **Există o diferență semnificativă statistic** între frecvența hematomului post puncție venoasă la pacienții de gen feminin (populația de gen feminin) și frecvența hematomului la pacienții de gen masculin (populația de gen masculin)..

TESTUL HI-PĂTRAT – ETAPELE

1. Formularea ipotezelor:

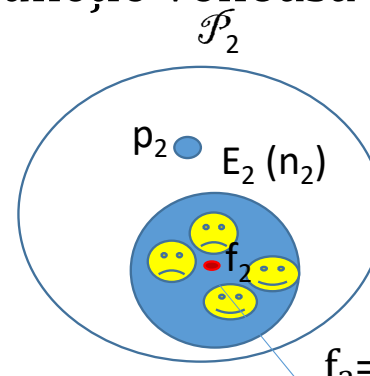
- H_0 : Nu există o **asociere semnificativ statistică** între hematomul post puncție venoasă și genul pacientului la adulți.
- H_1 : Există o **asociere semnificativ statistică** între hematomul post puncție venoasă și genul pacientului la adulți.
- Variabila de interes: Hematom post-puncție venoasă (😊 , 😞)



E_1 = esantion de femei

f_1 = Frecvența relativă a hematomului pe esantion

p_1 = Frecvența relativă a hematomului pe esantion



f_2 = Frecvența relativă a hematomului la bărbați

Testul Hi-pătrat – ETAPELE

2. Alegerea nivelului de semnificație $\alpha = 0,05$

3. Statistica Hi-pătrat:

- Se va calcula tabelul de contingenta teoretic (asteptat) presupunand ca H_0 este adevarata
- Se va calcula **diferența** dintre tabelul de contingență **observat** și tabelul de contingență **teoretic**

4. Determinarea valorii p si a regiunii de respingere a ipotezei nule

5. Decizia testului:

- Daca diferența este:
 - Mică \Rightarrow
 - NU se respinge ipoteza nulă
 - diferențele obținute se datorează hazardului
 - Mare \Rightarrow
 - Se va respinge ipoteza nulă
 - diferența se datorează mai degrabă existenței unei asocieri între variabile decât hazardului

TESTUL HI-PĂTRAT

Realizarea tabelului de contingenta teoretic (asteptat)

- Frecvența teoretică dintr-o celulă = (totalul de pe linia celulei × totalul de pe coloana celulei) / talia esantionului de pacienti

Tabelul de
contingență
observat

	Hematom (da)	Hematom (nu)	Total
F	31	22	53
M	17	30	47
Total	48	52	100

	Hematom (da)	Hemato m (nu)	Total
F	25	28	53
M	23	24	47
Total	48	52	100

Tabelul de
contingență
teoretic

Se recopiază tabelul cu
totalurile dar fără conținut!

	Hematom (da)	Hematom (nu)	Total
F	$= \frac{53 \times 48}{100}$	$= \frac{53 \times 52}{100}$	53
M	$= \frac{47 \times 48}{100}$	$= \frac{47 \times 52}{100}$	47
Total	48	52	100

TESTUL HI-PĂTRAT

Tabelul de contingenta teoretic fiind calculat sub ipoteza de independență între Hematom și Gen, va avea același procent (proporție) de pacienți cu hematom de gen feminin și masculin

Tabel de contingență observat

	Hematom (da)	Hematom (nu)	Total
F	31	22	53
M	17	30	47
Total	48	52	100

T. observat: procentajele pe LINII

	Hematom (da)	Hematom (nu)	Total
F	58%	42%	53
M	36%	64%	47
Total	48	52	100

T. observat: procentajele pe COLOANE

	Hematom (da)	Hematom (nu)	Total
F	65%	42%	53
M	35%	58%	47
Total	48	52	100

Tabel de contingenta teoretic

	Hematom (da)	Hematom (nu)	Total
F	25	28	53
M	23	24	47
Total	48	52	100

T. teoretic: procentajele pe LINII

	Hematom (da)	Hematom (nu)	Total
F	48%	52%	53
M	48%	52%	47
Total	48	52	100

T. teoretic: procentajele pe COLOANE

	Hematom (da)	Hematom (nu)	Total
F	53%	53%	53
M	47%	47%	47
Total	48	52	100

TESTUL HI-PĂTRAT

Calcul diferenței dintre tabelul de contingenta observat și cel teoretic

Statistica test Hi-pătrat este

$$\chi^2 = \frac{(f_1^0 - f_1^t)^2}{f_1^t} + \frac{(f_2^0 - f_2^t)^2}{f_2^t} + \frac{(f_3^0 - f_3^t)^2}{f_3^t} + \frac{(f_4^0 - f_4^t)^2}{f_4^t}$$

f^0 – frecvența absolută **observată**, f^t – frecvența absolută **teoretică**

Tabel de contingență observat

	Hematom (da)	Hematom (nu)	Total
F	31	22	53
M	17	30	47
Total	48	52	100

Tabel de contingență teoretic

	Hematom (da)	Hematom (nu)	Total
F	25	28	53
M	23	24	47
Total	48	52	100

Testul Hi-pătrat: etape

2. Nivel de semnificație: $\alpha = 0,05$

3. Valoarea calculată a statisticii testului:

$$\chi^2 = \frac{(f_1^0 - f_1^t)^2}{f_1^t} + \frac{(f_2^0 - f_2^t)^2}{f_2^t} + \frac{(f_3^0 - f_3^t)^2}{f_3^t} + \frac{(f_4^0 - f_4^t)^2}{f_4^t}$$

Tabel de contingenta observat

	Hematom (da)	Hematom (nu)	Total
F	31	22	53
M	17	30	47
Total	48	52	100

Tabel de contingenta teoretic

	Hematom (da)	Hematom (nu)	Total
F	25	28	53
M	23	24	47
Total	48	52	100

$$\chi^2 = \frac{(31-25)^2}{25} + \frac{(22-28)^2}{28} + \frac{(17-23)^2}{23} + \frac{(30-24)^2}{24} = 20,39$$

TESTUL HI-PĂTRAT: ETAPE

4. Valoare critică și regiunea de respingere (RR):

Din tabla distribuției Hi-patrat: valoare critică $\chi^2_{critic} = 3,84 \Rightarrow \mathbf{RR} = [3,84, +\infty)$

regiune de respingere (intervalul de valori unde respingem ipoteza nula a testului)

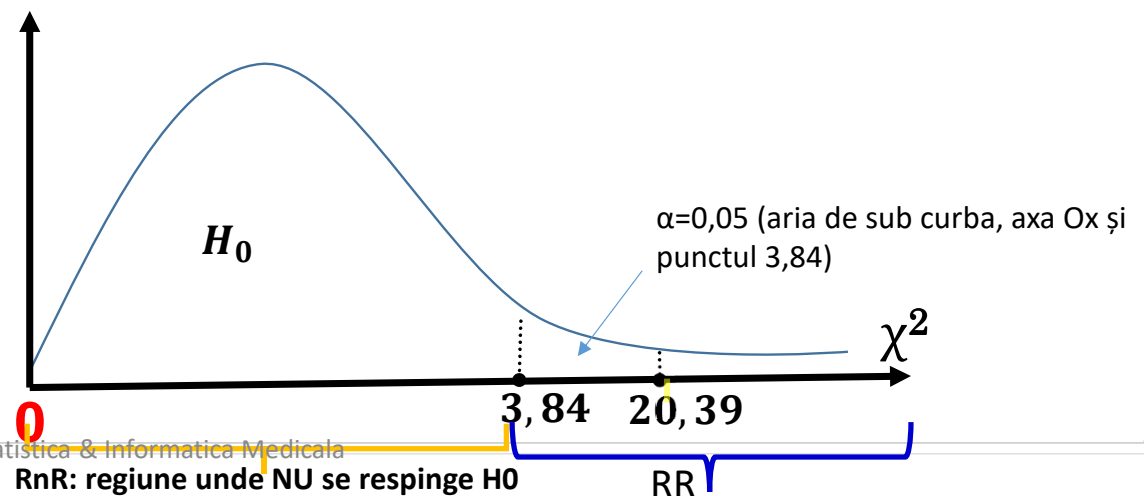
TESTUL HI-PĂTRAT: ETAPE

5. a. Decizia testului în funcție de regiunea de respingere (RR):

- Dacă χ^2 aparține lui RR → se respinge H_0 , deci suntem în favoarea lui H_1
- Dacă χ^2 NU aparține lui RR → NU se respinge H_0 , deci **nu am găsit** asociere între Hematom și Gen pe populația de adulți
- în cazul nostru: $\chi^2 = 20,39$ aparține intervalului $[3,84, +\infty)$, deci respingem $H_0 \Rightarrow H_1$: există asociere între hematom și genul pacientului (frecvența hematomului este diferită la femei și bărbați)

regiunea de respingere a lui $H_0 = RR$

Regiunea unde nu respingem $H_0 = RnR$



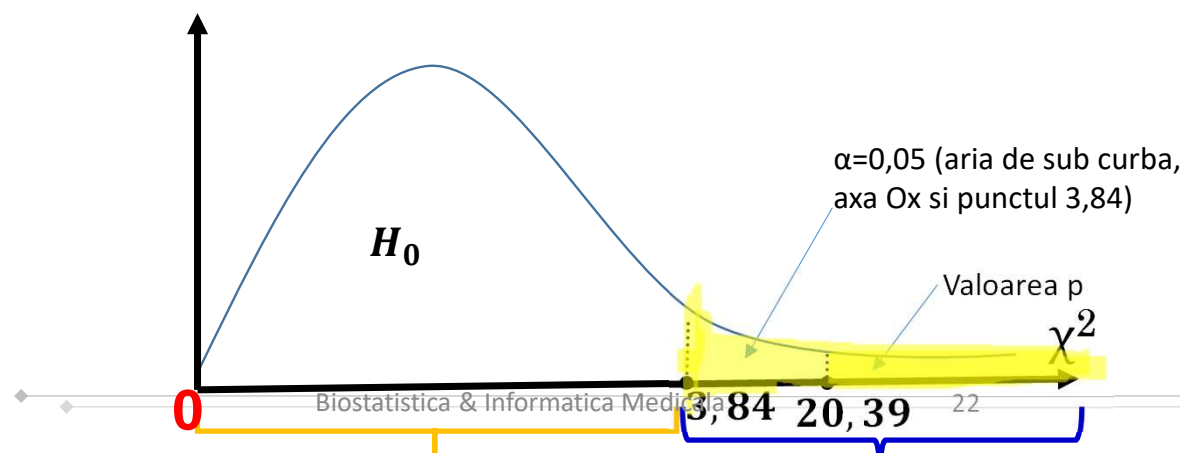
TESTUL HI-PĂTRAT: ETAPE

5. b. Decizia testului în funcție de valoarea p:

Daca valoarea $p < 0,05 \rightarrow$ se respinge H_0 , deci suntem în favoarea lui H_1

Daca valoarea $p \geq 0,05 \rightarrow$ NU se respinge $H_0 \rightarrow$ NU am gasit asociere semnificativă între cele doua variabile (formulare echivalentă: nu avem suficiente dovezi statistice pentru a afirma că există o asociere semnificativă între **hematomul post punctie venoasă și genul pacientului la adulți)**

- în cazul exemplului dat: $p = 0,025 < 0,05 \rightarrow$ se respinge H_0 , deci suntem în favoarea lui H_1



Indicatori medicali pentru cuantificarea asocierii

1. RIE =riscul individual pentru cei expusi

- $RIE = a / (a + b)$: procentul % de bolnavi dintre pacientii expusi la factor

2. RIN = riscul individual pentru cei expusi

- $RIN = c / (c + d)$: procentul % de bolnavi dintre pacientii neexpusi

3. RR = riscul relativ

- $RR = RIE / RIN$: de cate ori este mai mare riscul de boala la cei expusi fata de cei ne-expusi

	B=da	B=nu	Total
F=prezent	a	b	a+b
F=absent	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

B=patologia de interes; F=factorul de studiu (gen; fumat; etc.)

Indicatori medicali pentru cuantificarea asocierii

	Hematom (da)	Hematom (nu)	Total
Sex F	31	22	53
Sex M	17	30	47
Total	48	52	100

$$RIE = \Pr(\text{Hematom}/F) = a / (a + b) = 31/53 = 0,58$$

$$RIN = \Pr(\text{Hematom}/M) = c / (c + d) = 0,36$$

$$RR = \Pr(\text{Hematom}/F) / \Pr(\text{Hematom}/M) = RIE / RIN = 1,62$$

Interpretare RR: în eșantionul de studiu, sexul feminin a fost un factor de risc pentru apariția hematomului (frecvența hematomului a fost de 1,62 ori mai mare la femei decât la bărbați).

De reținut:

- testul Hi-pătrat se utilizează pentru testarea asocierii dintre două variabile calitative (nominale, dihotomiale, ordinale cu număr mic de clase)
- aplicarea testului Hi-pătrat implică realizarea a două tabele de contingență (observat și teoretic)
- decizia testului se va lua în raport cu statistica testului calculată și apartenența acesteia la regiunea de respingere (**RR**) SAU calculul **valorii p** (engl. p-value)
- ipoteza nulă (H_0) asociată testului se respinge (dacă $p < 0,05$) sau nu (dacă $p \geq 0,05$)
- dacă nu se respinge $H_0 \Rightarrow$ **NU am găsit asociere semnificativă** între cele două variabile (formulare echivalentă: nu avem suficiente dovezi statistice pentru a afirma că există o asociere semnificativă între cele două variabile pe populație)

Exemple de probleme

Q1. Se caută să se studieze dacă stresul se poate asocia cu apariția hipertensiunii arteriale. În urma observațiilor unui eșantion de 500 de persoane, 220 prezentau hipertensiune arterială, dintre aceștia 100 prezentând și stresul. Au fost identificați 210 pacienți fără tensiunea arterială și fără stres. Care dintre următoarele afirmații sunt adevărate:

- A. Frecvențele absolute observate asociate tabelului de contingenta al problemei sunt: 100–70–120–210
- B. Pentru testarea asocierii dintre stres si aparitia hipertensiunii arteriale s-ar putea utiliza testul Hi-patrat
- C. Ipoteza nula a testului Hi-patrat afirma ca: Nu există asociere semnificativ statistică între stres și hipertensiunea arterială
- D. Ipoteza nula a testului Hi-patrat afirma ca: Exista diferenta semnificativa statistic intre frecventa hipertensiunii arteriale la pacientii expusi la stres și frecvența hipertensiunii arteriale la cei neexpuși la stres
- E. Frecvențele absolute observate asociate tabelului de contingență al problemei sunt:170–330-220–280

R: A, B, C

Exemple de probleme

Q2. Se caută să se studieze dacă stresul se poate asocia cu apariția hipertensiunii arteriale. În urma observațiilor unui eșantion de 500 de persoane, 220 prezentau hipertensiune arterială, dintre aceștia 100 prezentând și stresul. Au fost identificați 210 pacienți fără tensiunea arterială și fără stres. Care dintre următoarele afirmații sunt adevarate:

- A. Frecvențele absolute teoretice asociate tabelului de contingenta teoretic al problemei sunt: 75–95–145–185
- B. Pentru testarea asocierii dintre stres si aparitia hipertensiunii arteriale s-ar putea utiliza testul Hi-patrat
- C. Ipoteza alternativă a testului Hi-patrat afirma ca: Nu exista asociere semnificativ statistică între stres si hipertensiunea arteriala
- D. Ipoteza alternativă a testului Hi-patrat afirma ca: Există diferență semnificativă statistic între frecvența hipertensiunii arteriale la pacientii expuși la stres și frecvența hipertensiunii arteriale la cei neexpuși la stres
- E. Frecvențele absolute observate asociate tabelului de contingență al problemei sunt: 100–70–120–210

R: A, B, D, E

Exemple de probleme

Q3. Se caută să se studieze dacă stresul se poate asocia cu apariția hipertensiunii arteriale. În urma observațiilor unui eșantion de 500 de persoane, 220 prezentau hipertensiune arterială, dintre aceștia 100 prezentând și stresul. Au fost identificați 210 pacienți fără tensiunea arterială și fără stres. Valoarea testului statistic aplicat a fost de 22,97 iar probabilitatea acestuia a fost egală cu 0,0000016. Care din următoarele este interpretarea corectă a acestui rezultat pentru testul aplicat?

- A. Deoarece $p < 0,05$ acceptăm ipoteza nulă
- B. Deoarece $p < 0,05$ respingem ipoteza nulă
- C. Pe baza rezultatelor obținute nu se poate trage nici o concluzie statistică
- D. Deoarece $p < 0,05$ exista asociere semnificativa intre stres si apartitia hipertensiunii arteriale
- E. Deoarece $p < 0,05$ frecventa hipertensiunii arteriale este semnificativ diferita la persoanele expuse la stres fata de persoanele neexpuse la stres.

R: B, D, E

Exemple de probleme

Q4. Se caută să se studieze dacă stresul se poate asocia cu apariția hipertensiunii arteriale. În urma observațiilor unui eșantion de 500 de persoane, 220 prezentau hipertensiune arterială, dintre aceștia 100 prezentând și stresul. Au fost identificați 210 pacienți fără tensiunea arterială și fără stres. Valoarea testului statistic aplicat a fost de 22,97 iar probabilitatea acestuia a fost egală cu 0,0000016.

Care din următoarele este interpretarea corectă a acestui rezultat pentru testul aplicat?

- A. Deoarece $p < 0,05$ respingem ipoteza nulă deci exista asociere semnificativa între stres și apariția hipertensiunii arteriale
- B. Testul statistic folosit pentru testarea asocierii dintre hipertensiunea arterială și stres este un test de tip Hi-pătrat
- C. Testul statistic folosit pentru testarea asocierii dintre hipertensiunea arterială și stres este un test de tip Student
- D. Pe baza rezultatelor obținute riscul relativ este egal cu $RR=1,62$
- E. Pe esantionul de studiu, riscul de a face hipertensiune arteriala a fost de 1,62 ori mai mare la persoanele expuse la stres versus persoanele neexpuse la stres

R: A, B, D, E

MULȚUMESC PENTRU ATENȚIE!

