



UNIVERSITATEA DE MEDICINĂ,
FARMACIE, ȘTIINȚE ȘI TEHNOLOGIE
"GEORGE EMIL PALADE"
DIN TÂRGU MUREȘ

PROBABILITĂȚI ȘI STATISTICĂ ÎN SISTEME MEDICALE

Cursul 12, 22.10.2020

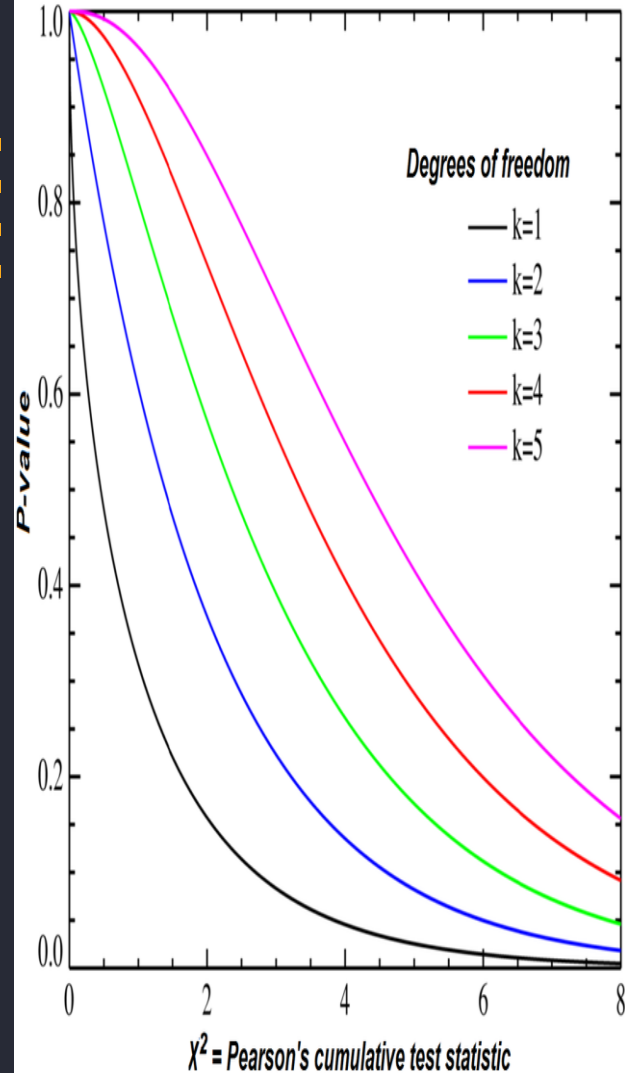
INFERENȚA STATISTICĂ PENTRU DATE DIN SISTEME MEDICALE

prof. univ. dr. habil Manuela Rozalia GABOR

STRUCTURA CURSULUI

1. Testarea ipotezelor statistice – Curs 9
2. Alegerea unui test statistic – Curs 9
3. Normalitatea datelor – Curs 10
4. Compararea variabilelor cantitative – Curs 11
5. **Compararea distribuțiilor variabilelor calitative – Curs 12**

Teste de tip Chi Square (hi pătrat)



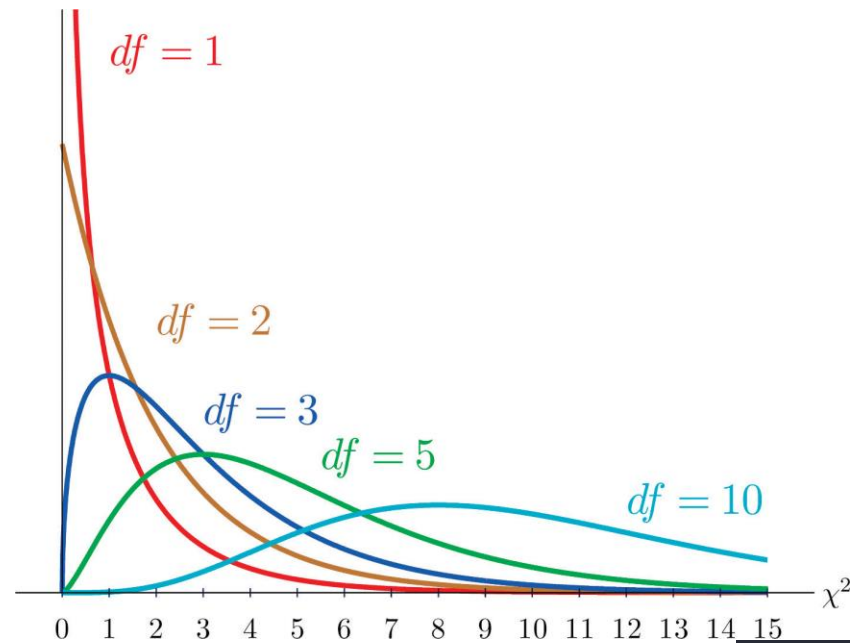
1. Testul chi square de independență
2. Testarea normalității unei distribuții (chi square goodness of fit/univariat)
3. Testul chi square de semnificație **2** x 2 (bivariat)
4. Testul chi square de semnificație **k** x 2

TESTE DE TIP CHI SQUARE (HI PĂTRAT)

Așa cum variabilele cantitative au teste specifice care sunt folosite în funcție de condițiile care sunt îndeplinite, și pentru variabilele calitative avem teste care se bazează pe compararea a două distribuții (frecvențe).

Testul Chi-square (Hi-pătrat), unul dintre cele mai des folosite și citate teste din literatura biomedicală, evaluează relațiile între două sau mai multe variabile calitative.

Pentru a aplica acest test, este necesar să se calculeze statistica asociată testului, χ^2 , care urmează o distribuție chi square



1. Testul chi square de independență

Cum se poate testa dacă două variabile calitative sunt independente?

Presupunem ca dorim să verificăm dacă există o asocierie între fumat (F) și apariția unei anumite boli (B). Considerăm ca avem un eșantion de 400 de indivizi împărțiți astfel: 160 prezintă boala B (240 nu prezintă boala) și 130 sunt fumători F (270 nu sunt fumători).

Ambele variabile sunt calitative dihotomiale, deoarece o persoană fie fumează, fie nu fumează, iar o boala este prezentă sau absentă. Repartiția indivizilor în funcție de cele două criterii se prezintă în următorul tabel de contingență (2 x 2):

Fumat	Boala		Total
	B bolnavi	\bar{B} sănătoși	
F (fumători)	80	50	130
\bar{F} (nefumători)	80	190	270
Total	160	240	400

EXEMPLU

Se caută să se stabilească dacă fumatul influențează apariția bolii B sau dacă apariția acesteia este independentă de fumat.

“

Ipotezele testului de independență

Ipoteza nulă (H_0):
Fumatul (F) și diagnosticul de boală (B) sunt independente

Ipoteza alternativă (H_1):
Fumatul (F) și diagnosticul de boală (B) sunt dependente.

Testul chi square de independență – etapele aplicării testului

	Cazul general	Ilustrarea printr-un exemplu
Problema	Se încearcă să se determine, cu ajutorul unui eșantion de n subiecți, dacă două caractere A și B având L și respectiv C modalități de realizare sunt sau nu independente.	Fumatul (F) și o boală (B) sunt independente? În acest caz, L=C=2, iar eșantionul observat are n=400 subiecți repartizați în tabelul de contingență prezentat mai sus.
Etapa 1. <u>Definirea ipotezei nule H₀</u>	H ₀ : caracterele A și B sunt independente.	H ₀ : fumatul nu are influență asupra apariției bolii B.
Etapa 2. <u>Definirea unui parametru</u>	$\chi^2 = \sum_{i=1}^{L \cdot C} \frac{(O_i - T_i)^2}{T_i}$ urmează o lege χ^2 cu (L-1) x (C-1) grade de libertate	$\chi^2 = \sum_{i=1}^{L \cdot C} \frac{(O_i - T_i)^2}{T_i}$ urmează o lege χ^2 cu 1 grad de libertate.
Etapa 3. <u>Pragul de semnificație</u>	Fie α pragul de semnificație al testului.	S-a ales pragul de semnificație $\alpha = 0.05$

	Cazul general	Ilustrarea printr-un exemplu
Etapa 4. <u>Definirea regiunii critice</u>	Ținând seama de faptul că χ^2 urmează legea χ^2 cu $(L-1) \times (C-1)$ grade de libertate se determină valoarea χ_{α}^2 încât $P(\chi^2 \geq \chi_{\alpha}^2) = \alpha$. Regiunea critică este $[\chi_{\alpha}^2, \infty)$.	Pentru pragul $\alpha = 0.05$ și χ^2 cu 1 grad de libertate valoarea $\chi_{\alpha}^2 = 3,84$, astfel că în acest caz regiunea critică este intervalul $[3,84, \infty)$.
Etapa 5. <u>Calcularea valorii observate a parametrului</u>	Se calculează frecvențele teoretice $T_i = \frac{\text{total linie} \times \text{total coloana}}{n}$ Se calculează $\chi^2 = \sum_{i=1}^{L \cdot C} \frac{(O_i - T_i)^2}{T_i}$	Se calculează $\chi^2 = \frac{(80 - 52)^2}{52} + \frac{(50 - 78)^2}{78} + \frac{(80 - 108)^2}{108} + \frac{(190 - 162)^2}{162}$ $= 37,2$
Etapa 6. <u>Decizia</u>	Dacă $\chi^2 \in [3,84, \infty)$ se respinge H_0 cu un risc de eroare de prima speță $\leq \alpha$. Dacă $\chi^2 \notin [3,84, \infty)$ atunci H_0 nu se respinge, neputându-se respinge H_0 cu un risc de eroare de speță a doua β	$\chi^2 \gg 3,84$ deci ipoteza nulă H_0 se respinge cu un risc inferior lui 5%. În concluzie, <u>fumatul se asociază cu boala B.</u> <u>Atenție direcția asociației dintre fumat și boală trebuie demonstrată prin parametri medicali.</u>

Condiții de aplicare

Testele pentru variabilele calitative, la fel ca orice test statistic, trebuie să îndeplinească anumite

cerinte/condiții:

Eșantioanele sunt extrase aleator din populație și păstrează toate caracteristicile acesteia

Fiecare subiect face parte dintr-un singur grup sau clasă, acestea fiind bine definite

Testul χ^2

Frecvențele din tabelul de contingență teoretic teoretice trebuie să fie minim 5 pentru cel puțin 80% din celule (Atenție la tabelele 2x2 toate celulele trebuie să aibă valorile mai mari sau egale cu 5).

Nicio frecvență nu trebuie să fie nulă (0).

2. Testarea normalității unei distribuții (univariat)

Pentru testarea normalității se poate utiliza și testul Chi-pătrat (Chi Square goodness of fit), prin care frecvențele (de clasă sau cumulate) observate sunt comparate cu frecvențele teoretice (de clasă sau cumulate) determinate utilizând legea normală.

Ipotezele statistice care se formulează pentru acest test sunt:

- | | | |
|----|---|--|
| H0 | { | • Ipoteza nulă (H0): Eșantionul este normal distribuit |
| H1 | { | • Ipoteza alternativă (H1): Eșantionul nu este normal distribuit |

Valori teoretice pentru statistica χ^2 (χ^2_{teoretic})

Critical values of the Chi-square distribution with d degrees of freedom							
Probability of exceeding the critical value							
d	0.05	0.01	0.001	d	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

INTRODUCTION TO POPULATION GENETICS, Table D.1
© 2013 Sinauer Associates, Inc.

Decizia de respingere sau acceptare a ipotezei nule H_0



Dacă:

$$\chi^2_{\text{calculat}} > \chi^2_{\text{teoretic}}$$

se **RESPINGE** H_0

Dacă:

$$\chi^2_{\text{calculat}} < \chi^2_{\text{teoretic}}$$

se **ACCEPTĂ** H_0

3. Testul chi square de semnificație 2 x 2 (bivariat)

Scopul:

Să determine dacă există diferențe semnificative între frecvențele observate pentru două variabile dicotomiale

CONDIȚII DE APLICARE

Eșantioanele sunt extrase aleator din populație și păstrează toate caracteristicile acesteia

Fiecare subiect face parte dintr-un singur grup sau clasă, acestea fiind bine definite

Testul χ^2 bivariat

Este necesar ca talia celor două eșantioane să fie suficient de mare. Această condiție este satisfăcută dacă $n > 20$ și frecvențele din tabelul de contingență teoretic teoretice trebuie să fie minim 5, pentru cel puțin 80% din celule și nicio celulă să nu fie 0

Observațiile în cadrul eșantioanelor să fie independente.

Algoritmul testului χ^2 bivariat

1. Atunci când două eșantioane sunt divizate fiecare în două clase, se poate construi următorul tabel de contingență 2 x 2

Tabel de contingență 2x2

	C1	C2	Total
Eșantionul 1	a	b	a + b
Eșantionul 2	c	d	c + d
Total	a + c	b + d	n = a + b + c + d

3. Statistica testului este comparată cu valoarea obținută din tabelul χ^2 cu 1 grad de libertate.

Dacă χ^2 este mai mare decât valoarea teoretică, vom respinge ipoteza nulă de independență între eșantioane și clase.

Altfel spus, cele două eșantioane nu au fost extrase din aceeași populație.

4. Testul chi square de semnificație $k \times 2$

Scopul testului este investigarea semnificației diferențelor dintre k distribuții de frecvențe observate, cu clasificare dihotomică.

EXEMPLU

Sunt studiate trei eșantioane formate din **trei** grupe de vârstă a femeilor gravide. Se dorește să se verifice dacă există o posibilă asocierie cu apariția preeclampsiei. În total este investigat un număr de 80 de persoane, obținându-se următorul tabel de contingență **3 X 2**:

Ipoteza nulă H_0 : Prezența preeclampsiei nu depinde de grupa de vârstă din care face parte gravida.

Tabel observat	Preeclampsie prezentă	Preeclampsie absentă	Total
Vârsta <20 de ani	14	22	36
Vârsta 20 - 40 de ani	18	16	34
Vârsta > 40 de ani	8	2	10
Total	40	40	80

Valoarea teoretică $\chi^2_{(2; 0,05)} = 5,991$

Valoarea calculată $\chi^2_{\text{calculat}} = 5,495$

Deoarece $\chi^2_{\text{calculat}} < \chi^2_{\text{teoretic}}$,

ipoteza nulă H_0 se acceptă și deci preeclampsia nu depinde de grupa de vârstă din care face parte gravida.

Testul McNemar (eșantioane perechi)

Testul McNemar este asemănător testului Chi square, folosește aceeași distribuție cu acesta (de aici și parametrul statistic identic), utilizându-se în situația în care avem **eșantioane perechi**.

1. Eșantioanele au fost extrase aleator din populație și păstrează toate caracteristicile acesteia.

2. Fiecare subiect face parte dintr-un singur grup (clasă), care este bine definit

3. Observațiile în cadrul eșantioanelor să fie independente

4. Suma valorilor din celulele $b+c > 25$

Condiții de aplicare





EXEMPLU

Testul McNemar pentru eșantioane perechi

Pacienților diabetici li s-a recomandat o dietă care are ca efect scăderea glicemiei. Prin studiul nostru dorim să urmărim dacă această dietă a avut și un rol în scăderea colesterolului. Avem un eșantion de 70 de diabetici:

Ipoteză nulă (H_0):

Nu există o legătură între hipercolesterolemie și dieta urmată

Tabel observat	<i>Înainte de dietă</i>	<i>După dietă</i>	Total
Hipercolesterolemie	15	25	40
Normocolesterolemie	10	20	30
Total	25	45	70