

Biostatistiques – BST1

Chapitre 23

Test de comparaison de plusieurs pourcentages estimés sur des échantillons indépendants

José LABARERE, PU-PH

Arnaud Seigneurin, MCU-PH, Bastien Boussat, MCU-PH, Alexandre Bellier, AHU, Patrice François, PU-PH

Plan

- Objectifs
- Introduction
- Test du chi²
- Généralisation du chi²
- Conditions d'application du chi²

Objectifs

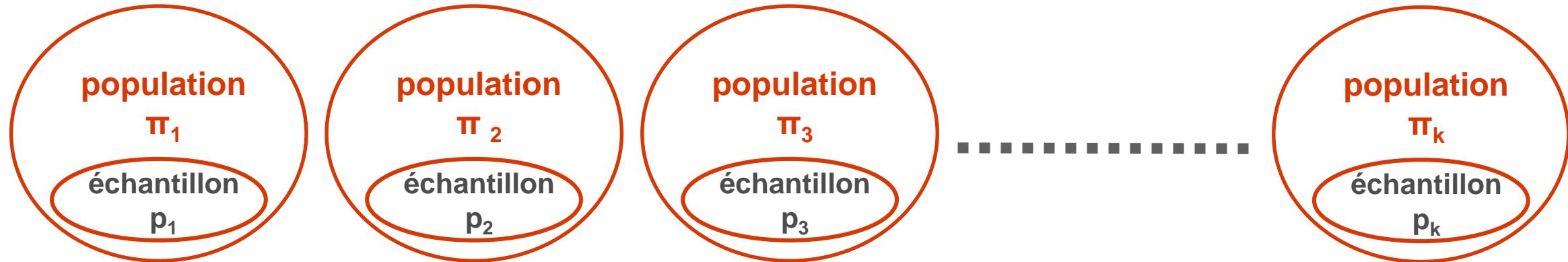
Comparaison de plusieurs pourcentages estimés sur des échantillons indépendants :

- Identifier une situation requérant le test du chi²
- Formuler les hypothèses (nulle et alternative)
- Connaitre et vérifier les conditions d'application du test du chi²
- Interpréter le résultat d'un test du chi²

Plan

- Objectifs
- Introduction
- Test du chi²
- Généralisation du chi²
- Conditions d'application du chi²

Comparaison de plusieurs pourcentages estimés sur des échantillons indépendants



Comparer k pourcentages estimés sur k échantillons indépendants revient à tester l'association entre :

- 1 variable qualitative binaire
- 1 variable qualitative polytomique (nominale ou ordinaire)

Comparaison de plusieurs pourcentages estimés sur des échantillons indépendants

Décès en réanimation

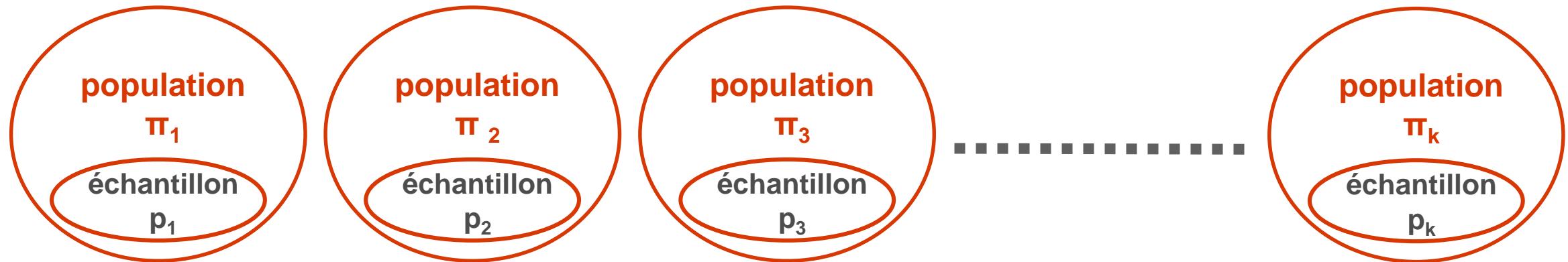
Age	oui	non	total
≤40	4 (7)	54 (93)	58
41 à 50	16 (11)	126 (89)	142
51 à 60	63 (15)	360 (85)	423
61 à 70	174 (29)	422 (71)	596
71 à 80	136 (40)	204 (60)	340
>80	12 (55)	10 (45)	22

Adapté de Grasselli G, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. JAMA 2020.

Plan

- Objectifs
- Introduction
- Test du chi²
- Conditions d'application du chi²

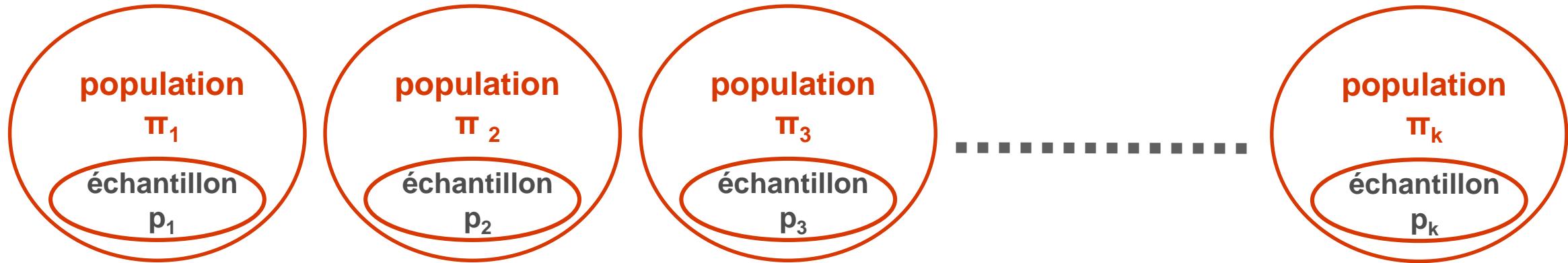
Hypothèse nulle (H_0)



$$\Pi_1 = \Pi_2 = \Pi_3 = \dots = \Pi_k = \Pi$$

**Absence d'association entre les 2 variables qualitatives
(les 2 variables qualitatives sont indépendantes)**

Hypothèse alternative (H1)



Au moins 1 des pourcentages Π_i diffère des $(k - 1)$ autres pourcentages

Il existe une association entre les 2 variables qualitatives

Effectifs observés

Age	Décès en réanimation			total
	oui	non		
≤40	O_{11}	O_{12}	L_1	
41 à 50	O_{21}	O_{22}	L_2	
51 à 60	O_{31}	O_{32}	L_3	
61 à 70	O_{41}	O_{42}	L_4	
71 à 80	O_{51}	O_{52}	L_5	
>80	O_{61}	O_{62}	L_6	
	C_1	C_2	n	

Effectifs théoriques attendus sous H0*

Age	Décès en réanimation			total
	oui	non		
≤40	T_{11}	T_{12}	L_1	
41 à 50	T_{21}	T_{22}	L_2	
51 à 60	T_{31}	T_{32}	L_3	
61 à 70	T_{41}	T_{42}	L_4	
71 à 80	T_{51}	T_{52}	L_5	
>80	T_{61}	T_{62}	L_6	
	C_1	C_2	n	

$$\sum_{i,j} \frac{(O_{ij} - T_{ij})^2}{T_{ij}} \rightarrow \chi^2_{(L-1)(C-1)ddl}$$

$$* T_{ij} = \frac{L_i \times C_j}{n}$$

Effectifs observés

Décès en réanimation

Age	Décès en réanimation			total
	oui	non		
≤40	4	54	58	
41 à 50	16	126	142	
51 à 60	63	360	423	
61 à 70	174	422	596	
71 à 80	136	204	340	
>80	12	10	22	

Effectifs théoriques attendus sous H0

Décès en réanimation

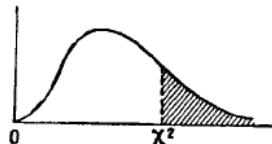
Age	Décès en réanimation			total
	oui	non		
≤40	14,9	43,1	58	
41 à 50	36,4	105,6	142	
51 à 60	108,4	314,6	423	
61 à 70	152,7	443,3	596	
71 à 80	87,1	252,9	340	
>80	5,6	16,4	22	

$$\chi_o^2 = \frac{(4-14,9)^2}{14,9} + \dots + \frac{(10-16,4)^2}{16,4} = 102,1$$

Détermination de la valeur de X_{α}^2 correspondant à un risque $\alpha = 0,05$ (5%)

Table de χ^2 (*).

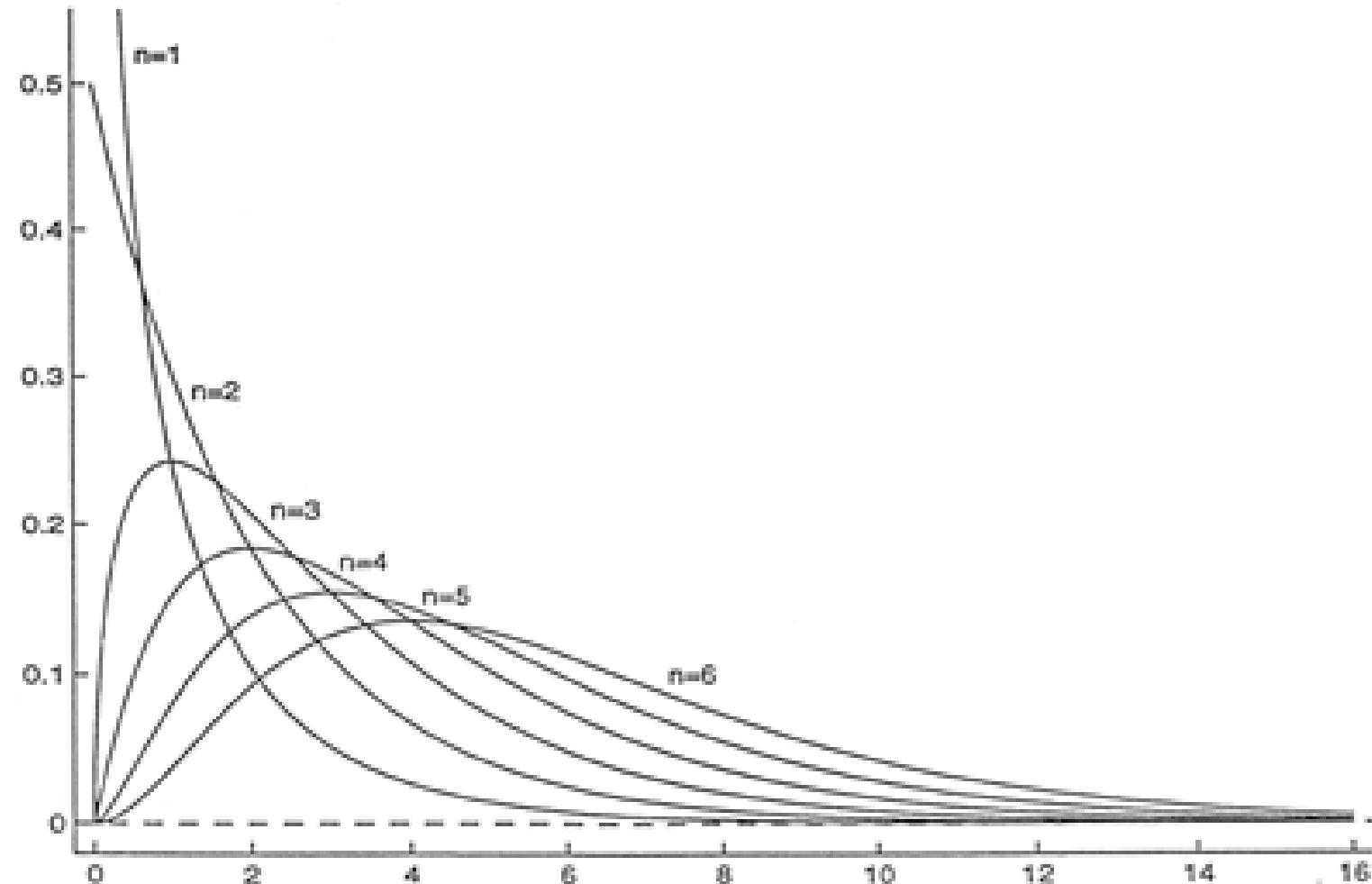
La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).

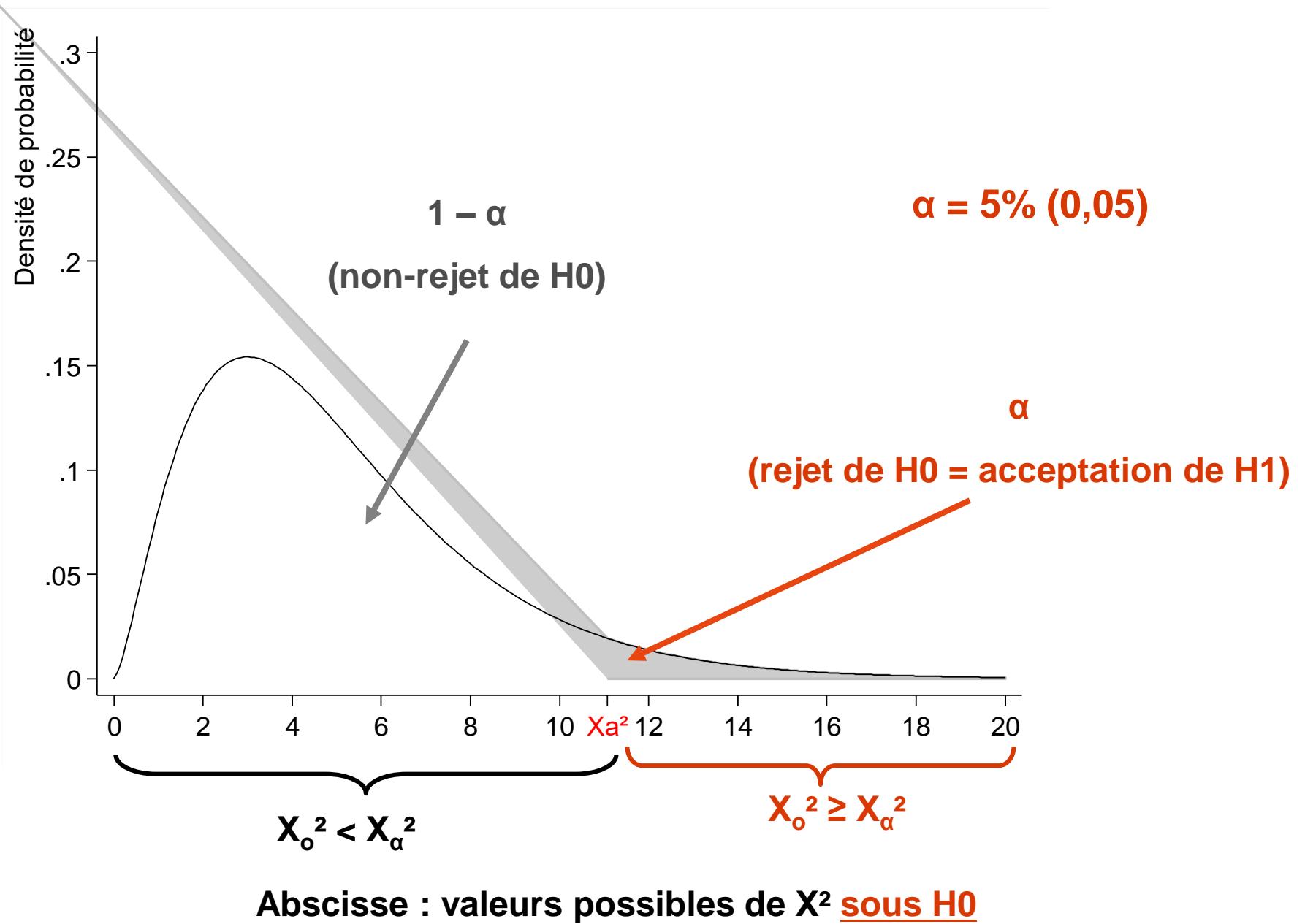


d.d.l. \ α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,888	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,239	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	13,240	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	14,041	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	14,848	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	15,659	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	16,473	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	17,292	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	18,114	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	18,939	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	19,768	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	20,599	29,336	33,530	36,250	40,256	43,773	47,962	50,892	59,703

$X_{\alpha}^2 = 11,070$ pour $(L - 1) \times (C - 1) = 5$ degrés de liberté

Densité de probabilité de loi du Chi^2





X_{α}^2 = valeur de X^2 pour le risque α

X_o^2 = valeur observée de X^2 pour l'échantillon

Détermination du degré de signification associé à X_o^2 (P -value)

$$\chi_o^2 = \sum_{i,j} \frac{(O_{ij} - T_{ij})^2}{T_{ij}} = 102,1$$

$P < 0.001$

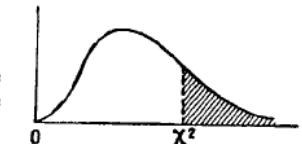
$P < \alpha \rightarrow$ rejet de H_0 (acceptation de H_1)

Le pourcentage de décès en réanimation
différait significativement en fonction de l'âge

Rappel : Probabilité conditionnelle d'observer une valeur de X^2
au moins aussi grande que X_o^2 si l'hypothèse nulle H_0 est vraie

Table de χ^2 (*).

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



$\alpha \backslash d.d.l.$	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	13,240	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	14,041	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	14,848	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	15,659	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	16,473	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	17,292	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	18,114	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	18,939	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	19,768	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	20,599	29,336	33,530	36,250	40,256	43,773	47,962	50,892	59,703

X

Plan

- Objectifs
- Introduction
- Test du chi²
- Généralisation du chi²
- Conditions d'application du chi²

Test du chi² d'association

Tester la liaison entre :

- 1 variable qualitative (polytomique à L modalités)
- 1 variable qualitative (polytomique à C modalités)

$$\sum_{i,j} \frac{(O_{ij} - T_{ij})^2}{T_{ij}} \rightarrow \chi^2_{(L-1)(C-1) \text{ ddl}}$$

Plan

- Objectifs
- Introduction
- Test du chi²
- Généralisation du chi²
- Conditions d'application du chi²

Conditions d'application du test du Chi²

- Indépendance des observations
- Effectif théorique T_{ij} attendu sous l'hypothèse nulle (H_0) supérieur ou égal à 5, pour chaque cellule du tableau de contingence

Messages clés

p_1	p_k	test	conditions
estimé	estimé	Chi ²	Effectifs théoriques ≥ 5
(échantillons indépendants)	(L-1) (C-1) ddl		Indépendance des observations

Mentions légales

L'ensemble de ce document relève des législations française et internationale sur le droit d'auteur et la propriété intellectuelle. Tous les droits de reproduction de tout ou partie sont réservés pour les textes ainsi que pour l'ensemble des documents iconographiques, photographiques, vidéos et sonores.

Ce document est interdit à la vente ou à la location. Sa diffusion, duplication, mise à disposition du public (sous quelque forme ou support que ce soit), mise en réseau, partielles ou totales, sont strictement réservées à l'Université Grenoble Alpes (UGA).

L'utilisation de ce document est strictement réservée à l'usage privé des étudiants inscrits à l'Université Grenoble Alpes (UGA), et non destinée à une utilisation collective, gratuite ou payante.