

TD7 : test du χ^2 (test d'indépendance)

Rappel :

On teste

— H_0 : les données sont indépendantes *versus*

— H_1 : les données ne sont pas indépendantes.

Le nombre de classes k_1 de la donnée 1.

Le nombre de classes k_2 de la donnée 2.

Statistique : $\hat{z} = \frac{(n_{\text{obs},1,1} - n_{\text{théo},1,1})^2}{n_{\text{théo},1,1}} + \dots + \frac{(n_{\text{obs},k_1,k_2} - n_{\text{théo},k_1,k_2})^2}{n_{\text{théo},k_1,k_2}}$.

Si $n \geq 30$ et tous les $n_{\text{théo},i,j} \geq 5$, alors $\hat{z} \sim \chi^2$ à $(k_1 - 1) \times (k_2 - 1)$ degrés de liberté.

Exercice 1. Trouver la valeur de z telle que

1. $P(\chi^2(4) \leq z) = 0.99$

2. $P(\chi^2(1) \leq z) = 0.9$

1 Exercices

Exercice 2 (Cf slides du 2nd CM). Le département GEA2 2A accueille 86 étudiants répartis comme dans le tableau ci-dessous

filière \ langue	allemand	espagnol
CG2P	3	27
GEMA	6	50

Le choix entre la filière et la langue est-il indépendant au risque 10% ?

Exercice 3 (Examen final 2021-2022). Un gérant parisien de petites salles de cinéma souhaite savoir si, en fonction du quartier, certains genres de films attirent plus de spectateurs. Pour une première étude, il se concentre sur 2 types de films (les films d'action et les comédies romantiques) et sur ces 3 cinémas parisiens situés dans le 14ème, le 16ème et le 19ème arrondissements. Pour cela, il compile le nombre de visiteurs sur une semaine dans ces différentes salles en fonction du type de films. Il obtient

	14ème	16ème	19ème	Total
Film d'action	383	147	70	600
Comédie romantique	272	98	30	400
Total	655	245	100	1000

Est-ce que le type de film et l'arrondissement de la salle sont indépendants au niveau 1% ?

2 Exercices d'entraînement

Exercice 4. Plusieurs compagnies d'assurance se demandent si le montant de la police d'assurance ne doit pas être lié au type de la voiture. Une étude menée sur 346 accidents a donné les résultats suivants :

Accident \ Voiture	Citadine	Familiale	Berline	Total
Grave	67	26	16	109
Léger	128	63	46	237
Total	195	89	62	346

Au risque 5%, les compagnies d'assurance ont-elles raison ?

Exercice 5. Les mêmes compagnies d'assurances se demandent si l'âge du conducteur pourrait avoir un impact sur le nombre d'accidents. Leurs enquêtes amènent aux données suivantes sur un an :

Accidents \ Âge	21-30	31-40	41-50	51-60	61-70
0	748	821	786	720	672
1	74	60	51	66	50
2	31	25	22	16	15
3 ou +	9	10	6	5	7

Au risque 1%, l'âge du conducteur et le nombre d'accidents sont-ils dépendants ou indépendants ?