

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

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. Le département GEA2 2A accueille 110 étudiants répartis comme dans le tableau ci-dessous

filière \ langue	allemand	espagnol
CG2P	3	27
GEMA	6	50

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

Correction. On va tester l'hypothèse

— H_0 : filière et langue sont indépendantes

versus

— H_1 : filière et langue sont dépendantes.

On va trouver les lois marginales de allemand/espagnol et de CG2P/GEMA :

filière \ langue	allemand	espagnol	Total	Proportion
CG2P	3	27	30	$30/86 \simeq 0.349$
GEMA	6	50	56	$56/86 \simeq 0.651$
Total	9	77	86	1
Proportion	0.105	0.895	1	

Ainsi, les probabilités estimées de chacun est

— $P(\text{CG2P}) \simeq 0.349$ et $P(\text{GEMA}) \simeq 0.651$,

— $P(\text{allemand}) \simeq 0.105$ et $P(\text{espagnol}) \simeq 0.895$.

Ensuite, on calcule les effectifs théoriques

F \ L	allemand	espagnol
CG2P	$0.349 \times 0.105 \times 86 \simeq 3.15$	$0.349 \times 0.895 \times 86 \simeq 26.86$
GEMA	$0.651 \times 0.105 \times 86 \simeq 5.88$	$0.651 \times 0.895 \times 86 \simeq 50.11$

$$n_{\text{theo,all,CG2P}} = n_{\text{total}} \times p_{\text{CG2P}} \times p_{\text{all}} = 86 \times 0.349 \times 0.105 \simeq 3.15.$$

Attention : vérifier dans le tableau que tout est ≥ 5 . Pas le cas ici, mais on va continuer quand même. (En fait il suffit que 75% des champs du tableaux soient supérieurs à 5, c'est le cas ici.)

On peut faire le tableau des χ^2 .

F \ L	allemand	espagnol
CG2P	$\frac{(3.15-3)^2}{3.15} \simeq 0.0071$	0.0007
GEMA	0.0024	0.0002

Rappel : $(n_{\text{theo}} - n_{\text{vrai}})^2 / n_{\text{theo}}$.

Puis on en déduit la statistique $Z = 0.0071 + 0.0007 + 0.0024 + 0.0002 = 0.0104$. Cette statistique suit une loi du χ^2 à $1 = (\text{lignes} - 1) \times (\text{colonnes} - 1) = (2 - 1) \times (2 - 1)$ degrés de liberté.

Par lecture dans la table : ligne 1 et colonne 0.1, on trouve $z = 2.706$. Donc la zone de rejet est $[2.706; +\infty[$. Comme Z n'appartient pas à la zone, on ne rejette pas H_0 , et donc langue et filière semblent indépendants.

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

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

Correction. Hypothèses :

- H_0 : films et arrondissements sont indépendants *versus*
- H_1 : films et arrondissements sont dépendants.

Tableau des effectifs théoriques

	14ème	16ème	19ème
Film d'action	$1000 \times 0.6 \times 0.655 = 393$	147	60
Comédie romantique	262	98	40

Tous les nombres sont ≥ 5 et leur somme est ≥ 30 .

Tableau des χ^2 :

	14ème	16ème	19ème
Film d'action	$(393 - 383)^2 / 393 \simeq 0.254$	0	1.7
Comédie romantique	0.38	0	2.5

La statistique est donc $Z = 2.5 + 1.7 + 0.25 + 0.38 = 4.83$ et suit une loi du χ^2 à $(3 - 1) \times (2 - 1) = 2$ degrés de liberté.

Par lecture dans la table à ligne 2 et colonne 0.01, on trouve $z = 9.210$.

Comme $4.83 < 9.210$, on ne rejette pas H_0 et le type de film semble être indépendant de l'arrondissement.

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 ?

Correction. On teste :

- H_0 : les types d'accidents et des voitures sont indépendants *versus*
- H_1 : les types d'accidents et de voitures sont dépendants.

Tableau des effectifs théoriques

Accident \ Voiture	Citadine	Familiale	Berline
Grave	$346 \times 109/346 \times 195/346 = 61.43$	28.04	19.53
Léger	133.57	60.96	42.47

Tous les nombres sont ≥ 5 et leur somme est ≥ 30 .

Tableau des χ^2 :

Accident \ Voiture	Citadine	Familiale	Berline
Grave	$(61.43 - 67)^2/61.43 = 0.51$	0.15	0.64
Léger	0.23	0.07	0.29

La statistique est donc $Z = 0.51 + 0.15 + 0.64 + 0.23 + 0.07 + 0.29 = 1.89$ et suit une loi du χ^2 a $(3 - 1) \times (2 - 1) = 2$ degrés de liberté.

Par lecture dans la table à ligne 2 et colonne 0.05, on trouve $z = 5.991$.

Comme $1.89 < 5.991$, on ne rejette pas H_0 et le type de voiture semble être indépendant du type d'accident.

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 ?

Correction. On teste :

- H_0 : l'âge et le nombre d'accidents sont indépendants *versus*
- H_1 : l'âge et le nombre d'accidents sont dépendants.

On calcule les totaux :

Accidents \ Âge	21-30	31-40	41-50	51-60	61-70	Total
0	748	821	786	720	672	3747
1	74	60	51	66	50	301
2	31	25	22	16	15	109
3 ou +	9	10	6	5	7	37
Total	862	916	865	807	744	4194

Tableau des effectifs théoriques

Accidents \ Âge	21-30	31-40	41-50	51-60	61-70
0	$4194 \times 3747 / 4194 \times 862 / 4194 = 770.13$	818.37	772.81	720.99	664.70
1	61.87	65.74	62.08	57.92	53.40
2	22.40	23.81	22.48	20.97	19.34
3 ou +	7.60	8.08	7.63	7.12	6.56

Tous les nombres sont ≥ 5 et leur somme est ≥ 30 .

Tableau des χ^2 :

Accidents \ Âge	21-30	31-40	41-50	51-60	61-70
0	$(770.13 - 748)^2 / 770.13 = 0.64$	0.01	0.23	0.001	0.08
1	2.38	0.5	1.98	1.13	0.22
2	3.30	0.06	0.01	1.18	0.97
3 ou +	0.26	0.46	0.35	0.63	0.03

La statistique est donc $Z = 14.42$ et suit une loi du χ^2 à $(5 - 1) \times (4 - 1) = 12$ degrés de liberté.

Par lecture dans la table à ligne 12 et colonne 0.01, on trouve $z = 26.217$.

Comme $14.42 < 26.217$, on ne rejette pas H_0 . L'âge du conducteur semble être indépendant du nombre d'accidents.