

TD 7 – Tests du χ^2 **Exercice 1.** (loi de χ^2)

1. Représentez la densité d'une variable aléatoire distribuée selon χ^2_8 .
2. Avec quelle probabilité cette variable est-elle inférieure à 5 ?
3. Déterminez et interprétez le quantile d'ordre 0,90 de la loi χ^2_8 .
 - (a) La densité de χ^2_8 (Fig. 1, Fig. 2) se représente sous R par : `curve(dchisq(x, 8), xlim=c(0, 15))`.
 - (b) Sous R, `pchisq(5, 8)` donne la probabilité pour une variable distribuée selon χ^2_8 d'être inférieure ou égale à 5 : 0,24 environ (Fig. 1).
 - (c) Sous R, `qchisq(0.90, 8)` donne le quantile d'ordre 0,9 de χ^2_8 : 13,4 environ (Fig. 2).

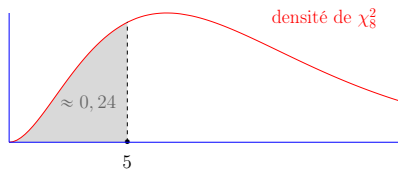


Fig. 1

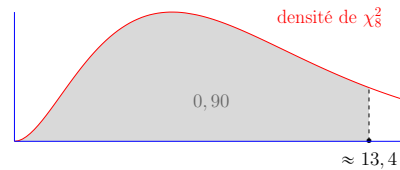


Fig. 2

Exercice 2. (indépendance de variables qualitatives)

Table 1 donne la répartition de soixante consommateurs de produits Danone par niveau de satisfaction (A/B/C/D) et par genre (H/F).

		satisfaction			
		A	B	C	D
genre	H	5	15	5	5
	F	5	10	5	10

Table 1: Répartition observée de soixante sujets par genre et niveau de satisfaction

On souhaite tester au seuil de 5% l'indépendance du genre et du niveau de satisfaction d'un sujet choisi au hasard.

1. Énoncez les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$.
 $\mathcal{H}_0$: la satisfaction est indépendante du genre ; $\mathcal{H}_1$: la satisfaction et le genre sont des variables dépendantes.
2. Sur quelle statistique la décision du test repose-t-elle ?
 La statistique de décision est : $X = \sum_{i=1}^2 \sum_{j=1}^4 (n_{i,j} - n_{i,j}^*)^2 / n_{i,j}^*$ avec :
 - $n_{i,j}$: effectif observé dans la modalité i de la variable *genre* et j de la variable *satisfaction*
 - $n_{i,j}^*$: effectif théorique dans la modalité i de la variable *genre* et j de la variable *satisfaction*.
 Pour rappel :
 - n : taille de l'échantillon.
 - $n_{i,\bullet} = \sum_{j=1}^4 n_{i,j}$ effectif marginal dans la modalité i du genre.
 - $n_{\bullet,j} = \sum_{i=1}^2 n_{i,j}$ effectif marginal dans la modalité j de la satisfaction.
 - $n_{i,j}^* = n_{i,\bullet} \times n_{\bullet,j} / n$.
3. Quelle est la distribution de la statistique de test sous $\mathcal{H}_0$?
 Sous $\mathcal{H}_0$, X est distribuée selon la loi de χ^2 à $(2-1) \times (4-1) = 3$ degrés de liberté.
4. Dressez le tableau des effectifs théoriques sous hypothèse d'indépendance.
 Les effectifs théoriques (Table 2) peuvent être calculés à la main : $n_{1,1}^* = n_{1,\bullet} \times n_{\bullet,1} / n = 30 \times 10 / 60$, etc. On peut aussi les obtenir sous R :
 - effectifs observés : `cont = matrix(c(5, 15, 5, 5, 5, 10, 5, 10), nrow=2, byrow=TRUE)`
 - effectifs théoriques : `chisq.test(cont)$expected`

		satisfaction				
		A	B	C	D	
genre	H	$n_{1,1}^* = 5$	$n_{1,2}^* = 12,5$	$n_{1,3}^* = 5$	$n_{1,4}^* = 7,5$	$n_{1,\bullet} = 30$
	F	$n_{2,1}^* = 5$	$n_{2,2}^* = 12,5$	$n_{2,3}^* = 5$	$n_{2,4}^* = 7,5$	$n_{2,\bullet} = 30$
		$n_{\bullet,1} = 10$	$n_{\bullet,2} = 25$	$n_{\bullet,3} = 10$	$n_{\bullet,4} = 15$	$n = 60$

Table 2: *Tableau des effectifs théoriques sous hypothèse d'indépendance*

5. Quelle est la valeur observée de la statistique de test ?

La valeur observée de statistique de test est : $X_{obs} = (5 - 5)^2/5 + (15 - 12,5)^2/12,5 + \dots + (5 - 5)^2/5 = 8/3 \approx 2,67$. On l'obtient sous R par : `theo = chisq.test(cont)$expected ; (cont-theo)^2/theo` ou : `chisq.test(cont)$statistic`.

6. Quelle est la valeur critique du test ?

Puisque le seuil de significativité est 5%, la valeur critique est le quantile d'ordre 0,95 de χ_3^2 : 7,8 environ.

7. Déterminez la p -valeur du test.

La p -valeur est la probabilité que sous $\mathcal{H}_0$, la statistique X prenne une valeur plus atypique encore que la valeur observée : $p\text{-val.} = \mathbb{P}(X > X_{obs}) = \mathbb{P}(\chi_3^2 > 8/3) = 1 - \mathbb{P}(\chi_3^2 \leq 8/3) \approx 0,446$ (* R : `1-pchisq(8/3, 3)`).

8. Enoncez la décision du test de deux façons :

i. en comparant la valeur observée de la statistique de test à la valeur critique.

La valeur observée de X ($X_{obs} = 8/3$) est inférieure à la valeur critique ; au seuil 5% on ne rejette pas l'indépendance des variables *genre* et *satisfaction*.

ii. en comparant la p -valeur au seuil de risque.

La p -valeur est supérieure au seuil de significativité (5%) ; au seuil 5% on ne rejette pas l'indépendance des variables *genre* et *satisfaction*.

Exercice 3. (indépendance de variables qualitatives)

Table 3 consigne les résultats d'une enquête : on demande à deux cents personnes nées en Aquitaine si elles sont favorables à la radiation de la corrida du patrimoine culturel immatériel (PCI).

		Radiation		
		oui	neutre	non
Département	Dordogne	15	14	21
	Gironde	19	14	8
	Landes	10	13	10
	Lot-et-Garonne	14	12	17
	Pyrénées-Atlantiques	11	13	9

Table 3: *Répartition de deux cents personnes selon : (i) leur département d'origine et (ii) leur avis sur la radiation de la corrida du PCI*

On souhaite déterminer dans un test de χ^2 si *Radiation* et *Département* sont des variables indépendantes.

$\mathcal{H}_0$: *Radiation* et *Département* sont des variables indépendantes ; $\mathcal{H}_1$: *Radiation* et *Département* ne sont pas indépendantes.

1. Dressez le tableau des effectifs théoriques sous hypothèse d'indépendance.

Les effectifs théoriques sont dans Table 4.

2. Déterminez la valeur de la statistique du χ^2 .

La valeur observée de la statistique de test : $X_{obs} \approx 8,24$

3. Doit-on rejeter l'indépendance des deux variables au seuil de 5% ?

La valeur critique associée au seuil de significativité 5% est le quantile d'ordre 0,95 de la loi de χ^2 à 8 degrés de liberté : 15,51. Puisque la valeur critique est supérieure à la valeur observée de la statistique de test, on ne rejette pas l'indépendance entre la région d'origine et l'avis d'un sujet sur la radiation de la corrida du PCI.

		Radiation		
		oui	neutre	non
Département	Dordogne	17, 2	16, 5	16, 2
	Gironde	14, 1	13, 5	13, 3
	Landes	11, 4	10, 9	10, 7
	Lot-et-Garonne	14, 8	14, 2	14, 0
	Pyrénées-Atlantiques	11, 4	10, 9	10, 7

Table 4: *Tableau des effectifs théoriques (arrondis) sous hypothèse d'indépendance*

4. Déterminez et interprétez la p -valeur du test.

La p -valeur du test est la probabilité pour une variable distribuée selon χ_8^2 d'être supérieure à $X_{obs} \approx 8,24$; elle vaut environ 0,41. Tant que le seuil de significativité sera inférieur à 41%, on ne rejettera pas l'indépendance.

Exercice 4. (*indépendance de variables qualitatives*)

Table 5 répartit cent téléspectateurs selon le programme préféré et la chaîne favorite.

		programme préféré		
		film	journal	sport
chaîne favorite	Ch. 1	12	13	9
	Ch. 2	10	18	11
	Ch. 3	7	12	8

Table 5: *Répartition de cent téléspectateurs selon le programme préféré et la chaîne favorite*

On souhaite déterminer dans un test de χ^2 si le programme et la chaîne préférés d'un téléspectateur sont indépendants.

$\mathcal{H}_0$: *programme préféré* et *chaîne favorite* sont des variables indépendantes ; $\mathcal{H}_1$: *programme préféré* et *chaîne favorite* ne sont pas indépendantes.

1. Dressez le tableau des effectifs théoriques sous hypothèse d'indépendance.

		programme préféré		
		film	journal	sport
chaîne favorite	Ch. 1	9,86	14,62	9,52
	Ch. 2	11,31	16,77	10,92
	Ch. 3	7,83	11,61	7,56

Table 6: *Effectifs théoriques sous hypothèse d'indépendance*

2. Déterminez la valeur de la statistique du χ^2 .

$$X_{obs} \approx 1,04.$$

3. Doit-on rejeter l'indépendance des deux variables au seuil de 5% ?

$$p\text{-val.} = \mathbb{P}(\chi_4^2 > X_{obs}) = 1 - \mathbb{P}(\chi_4^2 \leq X_{obs}) \approx 0,90. \quad p\text{-val.} \geq \alpha = 0,05 ; \text{ on ne rejette pas } \mathcal{H}_0.$$

Exercice 5. (*adéquation à une loi de probabilité discrète*)

La tableau suivant donne la répartition de deux cents sujets par catégorie professionnelle ainsi que la fréquence de chaque catégorie dans la population.

catégorie	effectif (échantillon)	fréquence en % (population)
agriculteur	14	10
commerçant	61	25
cadres	34	15
prof. intermédiaire	18	10
employé	46	20
ouvrier	27	20

Au seuil 5% doit-on rejeter : l'échantillon est représentatif de la population ?

catégorie	effectifs observés (n_i)	fréquences en % (population)	effectifs théoriques (n_i^*)
agriculteur	14	10	20 = 0,10 × 200
commerçant	61	25	50
cadres	34	15	30
prof. intermédiaire	18	10	20
employé	46	20	40
ouvrier	27	20	40

$\mathcal{H}_0$: l'échantillon est représentatif de la population ; $\mathcal{H}_1$: l'échantillon n'est pas représentatif.

SdT. $X = \sum_{i=1}^6 (n_i - n_i^*)^2 / n_i^*$; $X_{obs} = (14 - 20)^2 / 20 + \dots + (27 - 40)^2 / 40 \approx 10,1$.

ZdR. $\mathcal{H}_0$ est rejetée au seuil 5% si $X_{obs} > \chi_{5;0,95}^2 \approx 11,1$.

Dec. $X_{obs} \leq \chi_{5;0,95}^2$; on ne rejette pas $\mathcal{H}_0$ au seuil 5%.

NB. Sous R, on peut entrer les effectifs observés : `eff=c(14,61,34,18,46,27)` et les fréquences théoriques : `freq=c(0.10,0.25,0.15,0.10,0.20,0.20)` et obtenir la valeur de la statistique de test et la p -valeur par : `mytest=chisq.test(eff,p = freq)` ; `mytest$statistic` ; `mytest$p.value`.

Exercice 6. (adéquation à une loi uniforme discrète)

Le tableau suivant indique le nombre de victoires de huit chevaux ayant participé à cent-quarante-quatre courses.

cheval	1	2	3	4	5	6	7	8
nb victoires	29	19	18	25	17	10	15	11

Au seuil 5% doit-on rejeter : les chevaux ont tous au départ les mêmes chances de gagner ?

cheval	1	2	3	4	5	6	7	8
nb victoires observées (n_i)	29	19	18	25	17	10	15	11
nb victoires théoriques (n_i^*)	18	18	18	18	18	18	18	18

$\mathcal{H}_0$: les huit chevaux ont autant de chances de gagner ; $\mathcal{H}_1$: les chevaux n'ont pas les mêmes chances de gagner.

SdT. $X = \sum_{i=1}^8 (n_i - n_i^*)^2 / n_i^*$; $X_{obs} = (29 - 18)^2 / 18 + \dots + (11 - 18)^2 / 18 \approx 16,3$.

ZdR. $\mathcal{H}_0$ est rejetée au seuil 5% si $X_{obs} > \chi_{7;0,95}^2 \approx 14,1$.

Dec. $X_{obs} > \chi_{7;0,95}^2$; on rejette $\mathcal{H}_0$ au seuil 5%.

Exercice 7. (adéquation à une loi de Poisson)

Le tableau suivant indique la répartition de cinquante familles selon le nombre de voitures par foyer.

nb voitures	0	1	2	3	4
nb familles	21	18	7	3	1

Au seuil 10%, doit-on rejeter l'hypothèse : le nombre de voitures par foyer est distribué selon une loi de Poisson ?

nb voitures	0	1	2	3	4
effectif observé n_i	$n_1 = 21$	$n_2 = 18$	$n_3 = 7$	$n_4 = 3$	$n_5 = 1$
fréquence théorique f_i^*	$f_1^* \approx 0,407$	$f_2^* \approx 0,366$	$f_3^* \approx 0,165$	$f_4^* \approx 0,049$	$f_5^* \approx 0,013$
effectif théorique n_i^*	$n_1^* \approx 20,35$	$n_2^* \approx 18,30$	$n_3^* \approx 8,25$	$n_4^* \approx 2,45$	$n_5^* \approx 0,65$

Le nombre moyen de voitures par foyer est : $(0 \times 21 + \dots + 4 \times 1) / 50 = 0,9$. Supposons que le nombre X de voitures d'un foyer choisi au hasard est distribué selon $\mathcal{P}(0,9)$. Alors, pour $i \in \{1, 2, 3, 4\}$: $f_i^* = \mathbb{P}(X = i - 1) = \mathbb{P}(\mathcal{P}(0,9) = i - 1) = e^{-0,9} 0,9^{i-1} / (i - 1)!$ et $n_i^* = 50 \times f_i^*$. Par ailleurs : $f_5^* = \mathbb{P}(X \geq 4) = \mathbb{P}(\mathcal{P}(0,9) \geq 4) = 1 - \sum_{i=1}^4 f_i^* \approx 0,013$.

$\mathcal{H}_0$: le nombre de voitures par foyer est distribué selon $\mathcal{P}(0,9)$; $\mathcal{H}_1$: le nombre de voitures par foyer n'est pas distribué selon $\mathcal{P}(0,9)$.

SdT. $X = \sum_{i=1}^5 (n_i - n_i^*)^2 / n_i^*$; $X_{obs} = (21 - 20,35)^2 / 20,35 + \dots + (1 - 0,65)^2 / 0,65 \approx 0,527$.

ZdR. $\mathcal{H}_0$ est rejetée au seuil 10% si $X_{obs} > \chi_{4;0,90}^2 \approx 7,78$.

Dec. $X_{obs} \leq \chi_{4;0,90}^2$; on ne rejette pas $\mathcal{H}_0$ au seuil 10%.

Exercice 8. (adéquation à une loi binômiale)

Le tableau suivant indique la répartition de trois-cent-vingt familles de cinq enfants selon le nombre de garçons parmi les enfants.

nb garçons	0	1	2	3	4	5
nb familles	18	56	110	88	40	8

Au seuil 10%, doit-on rejeter l'hypothèse : la naissance d'une fille ou d'un garçon sont équiprobables ?

nb garçons	0	1	2	3	4	5
nb familles observé	18	56	110	88	40	8
fréquence théorique	$f_1^* = 0,03125$	$f_2^* = 0,15625$	$f_3^* = 0,3125$	$f_4^* = 0,3125$	$f_5^* = 0,15625$	$f_6^* = 0,03125$
effectif théorique	$n_1^* = 10$	$n_2^* = 50$	$n_3^* = 100$	$n_4^* = 100$	$n_5^* = 50$	$n_6^* = 10$

S'il est équiprobable d'avoir un garçon ou une fille, le nombre X de garçons d'une famille de cinq enfants est distribué selon la loi binomiale de paramètres : $n = 5$ et $p = 1/2$. Ainsi, la fréquence théorique de familles de cinq

enfants comportant i garçons ($i \in \{0, 1, \dots, 5\}$) est : $f_{i+1}^* = \mathbb{P}(b(5; 0, 5) = i) = \binom{5}{i} \times 0,5^i \times (1 - 0,5)^{5-i} = \binom{5}{i}/32$ et l'effectif théorique sur trois-cent-vingt familles est : $n_{i+1}^* = 320 \times f_{i+1}^*$.

$\mathcal{H}_0$: il est équiprobable d'avoir un garçon ou une fille (X est distribuée selon $b(5; 0, 5)$) ; $\mathcal{H}_1$: il n'est pas équiprobable d'avoir un garçon ou une fille.

SdT. $X = \sum_{i=1}^6 (n_i - n_i^*)^2 / n_i^*$; $X_{obs} = (18 - 10)^2/10 + \dots + (8 - 10)^2/10 \approx 11,96$.

ZdR. $\mathcal{H}_0$ est rejetée au seuil 10% si $X_{obs} > \chi_{5;0,90}^2 \approx 9,2$.

Dec. $X_{obs} > \chi_{5;0,90}^2$; on rejette $\mathcal{H}_0$ au seuil 10%.
