Part II

Tests du χ^2

5 Test du χ^2 d'indépendance

Dans cette section, on s'intéresse aux relations entre deux variables notées X et Y. Supposons que l'on observe ces deux variables sur n unités statistiques. A chaque individu i, on peut associer un couple d'observations $(x_i; y_i)$. Chaque variable peut-être quantitative ou qualitative. Nous proposons ici un test d'indépendance.

(X, Y) 2 caractères relevés par individu de l'échantillon.

En général, le test du Chi-2 est utilisé pour deux variables qualitatives.

Les données peuvent être représentées dans un tableau à double entrée appelé **Tableau de contingence**.

	m_1^Y	$\ldots$	m_k^Y	$\ldots$	m_K^Y	total
m_1^X	n_{11}	$\ldots$	n_{1k}	$\ldots$	n_{1K}	$n_{1\bullet}$
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
m_j^X	n_{j1}	$\ldots$	n_{jk}	$\ldots$	n_{jK}	$n_{j\bullet}$
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
m_J^X	n_{J1}	$\ldots$	n_{Jk}	$\ldots$	n_{JK}	$n_{J\bullet}$
	$n_{\bullet 1}$	$\ldots$	$n_{\bullet k}$	$\ldots$	$n_{\bullet K}$	n

marges

n_{jk}

ligne

colonne

$\sum_k n_{jk}$

$\Rightarrow$ Taille de l'échantillon

$= \sum_j n_{jk}$

Le **tableau des fréquences** s'obtient en divisant les effectifs par le nombre d'unités statistiques n (effectif total). Comme précédemment on obtient

Notons $m_1^X, \cdots, m_J^X$ les J modalités de X et $m_1^Y, \cdots, m_K^Y$ les K modalités de Y. Si l'une des deux variables (ou les deux) est quantitative continue, les m_j^X ou les m_k^Y sont des classes modales. Introduisons les quantités suivantes :

- n_{jk} est le nombre de fois où le couple (X, Y) prend la modalité (m_j^X, m_k^Y),
- $n_{\bullet k}$ est le nombre de fois où la variable Y prend la valeur m_k^Y,
- $n_{j\bullet}$ est le nombre de fois où la variable X prend la valeur m_j^X.

On a

$$\sum_{j=1}^{J} n_{jk} = n_{\bullet k} \quad \text{et} \quad \sum_{k=1}^{K} n_{jk} = n_{j\bullet}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{J} n_{jk} = \sum_{j=1}^{J} n_{j\bullet} = \sum_{k=1}^{K} n_{\bullet k} = n$$

tableau des fréquences = tableau de contingences/n

$$f_{jk} = \frac{n_{jk}}{n}, \quad f_{\bullet k} = \frac{n_{\bullet k}}{n} \quad f_{j\bullet} = \frac{n_{j\bullet}}{n}$$

fréquences jointes

	m_1^Y	$\dots$	m_k^Y	$\dots$	m_K^Y	total
m_1^X	f_{11}	$\dots$	f_{1k}	$\dots$	f_{1K}	$f_{1\bullet}$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
m_j^X	f_{j1}	$\dots$	f_{jk}	$\dots$	f_{jK}	$f_{j\bullet}$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
m_J^X	f_{J1}	$\dots$	f_{Jk}	$\dots$	f_{JK}	$f_{J\bullet}$
	$f_{\bullet 1}$	$\dots$	$f_{\bullet k}$	$\dots$	$f_{\bullet K}$	1

marges.

fréquences marginales

5.2 Distributions marginales

A partir du tableau de contingence, on peut retrouver la distribution de chacune des variables séparément :

Modalité de Y	m_1^Y	$\dots$	m_k^Y	$\dots$	m_K^Y	total
Fréquence empirique	$f_{\bullet 1}$	$\dots$	$f_{\bullet k}$	$\dots$	$f_{\bullet K}$	1

Modalité de X	m_1^X	$\dots$	m_j^X	$\dots$	m_J^X	total
Fréquence empirique	$f_{1\bullet}$	$\dots$	$f_{j\bullet}$	$\dots$	$f_{J\bullet}$	1

Les distributions de X et de Y sont appelées distributions marginales. Sur chaque variable, on peut calculer les indicateurs habituels (moyenne, variance, écart type si la variable est quantitative...). Ces paramètres sont qualifiés d'indicateurs marginaux.

Rappel : en proba : indépendance $\Longleftrightarrow P_A(B) = P(B)$

$\Longleftrightarrow P(A)P_A(B) = P(A)P(B)$

$\Longleftrightarrow P(A \cap B) = P(A)P(B)$

5.3 Statistique du χ^2

Dans notre cadre : f_{ij} fréquence jointe = produit des marges $f_{i\bullet} f_{\bullet j}$

En présence de deux variables, l'un des enjeux principaux est d'étudier (c'est à dire quantifier voire expliquer) la dépendance entre les deux caractères.

Si on était dans le cadre des probabilités, ce qui n'est pas le cas, alors deux caractères sont indépendants si la valeur de l'un n'a aucune influence sur la distribution de l'autre. Si tel était le cas, alors les distributions conditionnelles

$$f_{j|k} = \frac{f_{jk}}{f_{\bullet k}} \qquad \text{et} \qquad f_{k|j} = \frac{f_{jk}}{f_{j\bullet}}$$

seraient toutes semblables à la distribution marginale. Pour tout (j, k), on devrait avoir

$$f_{j|k} = f_{j\bullet} \qquad \text{et} \qquad f_{k|j} = f_{\bullet k}.$$

Ainsi, on aurait :

$$f_{kj} = f_{j|k} f_{\bullet k} = f_{j\bullet} f_{\bullet k}.$$

D'où, si les deux variables étaient indépendantes, on aurait

$$n_{jk} = \frac{n_{j\bullet} n_{\bullet k}}{n}.$$

En stat, il n'y a pas l'indépendance "pure" des proba, mais on regarde la distance au cadre d'indépendance probabiliste.

En statistiques, on ne peut que "quantifier la distance à l'indépendance" par la statistique du χ^2,

$$D_{\chi^2} = n \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(f_{jk} - f_{j\bullet} f_{\bullet k})^2}{f_{j\bullet} f_{\bullet k}}.$$

différence entre f_{ij} et les marges.

présence technique

on normalise par le produit des marges pour éviter qu'une différence ne prennent un rapport de force injustifié (si f_{ij} est fort par exemple).

On peut remarquer que

$$D_{\chi^2} = n \left(\sum_{j=1}^{J} \sum_{k=1}^{K} \frac{n_{jk}^2}{n_{j\bullet} n_{\bullet k}} - 1 \right),$$

ou de façon équivalente

$$D_{\chi^2} = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(n_{jk} - \frac{n_{j\bullet} n_{\bullet k}}{n} \right)^2}{\frac{n_{j\bullet} n_{\bullet k}}{n}},$$

où J et K sont le nombre de modalités de chacune des deux variables considérées.

Le cas d'indépendance probabiliste serait alors équivalent à $D_{\chi^2} = 0$.

Test du Chi-2 :

H_0 = " X et Y sont indépendantes "

$$D_{\chi^2} = n \sum_j \sum_k \frac{(f_{jk} - f_{j\circ} f_{\circ k})^2}{f_{j\circ} f_{\circ k}}$$

Theorem 5.1 *La variable du test D_{χ^2} suit une loi du χ^2 à $(K-1)(L-1)$ degrés de liberté.*

modalités de X ↗ ↑ modalités de Y.

Démonstration : f_n estimateur d'une fréquence f sur un échantillon de taille n.

$$f_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

les $X_i \sim \mathcal{B}(p)$ i.i.d.

→ 1 si cas favorable

→ 0 sinon.

$$\sum_{i=1}^{n} X_i \sim \mathcal{B}(n,f) \xrightarrow[\mathcal{L}]{n \text{ grand}} \mathcal{N}. \quad (TCL)$$

On peut approcher $\sum_{i=1}^{n} X_i$ par $\mathcal{N}\left(nf, \sqrt{nf(1-f)}\right)$

$Var(aX+b) = a^2 Var(X)$

$Var\left(\frac{Z}{n}\right) = \frac{1}{n^2} Var(Z)$

$\sigma_{Z/n} = \frac{1}{n}\sigma_Z$

Donc $f_n = \frac{\sum_{i=1}^{n} X_i}{n}$ s'approche par $\mathcal{N}\left(f; \sqrt{\frac{f(1-f)}{n}}\right)$

Si f_n, f sont petits, $1-f \simeq 1$.

$$\frac{f_n - f}{\sqrt{f/n}} \sim \mathcal{N}(0,1).$$

$$\Leftrightarrow \sqrt{n}\,\frac{f_n - f}{\sqrt{f}} \sim \mathcal{N}(0,1)$$ pour f petit et n grand.

Retour au contexte :

indépendance Sous H_0, $\forall i\,j$ $\quad f_{ij} = f_{i\bullet} f_{\bullet j}$.

$\iff$ Sous H_0, les estimateurs des fréquences jointes estiment le produit des marges.

$$f_{ij} \qquad\qquad \sqrt{n}\, \frac{f_n - f}{\sqrt{f}} \qquad\qquad f_{i\bullet} f_{\bullet j}.$$

Les f_{ij} étant nombreuses, on peut les supposer petites. ainsi que les $f_{i\bullet} f_{\bullet j}$. On considère donc que chaque quotient :

$$\sqrt{n}\, \frac{f_{ij} - f_{i\bullet} f_{\bullet j}}{\sqrt{f_{i\bullet} f_{\bullet j}}} \sim \mathcal{N}(0,1).$$

(Sous H_0).

En sommant les carrés, il vient que

$$D_{\chi^2} = \sum_i \sum_j \left(\frac{\sqrt{n}\,(f_{ij} - f_{i\bullet} f_{\bullet j})}{\sqrt{f_{i\bullet} f_{\bullet j}}} \right)^2 \sim \chi^2(d)$$

en tant que somme de carrés de lois normales.

d, degrés de liberté, correspond au nombre d'aléas indépendants.

fréquences liées aux J-1 premières variables.

	m_1^Y	$\dots$	m_k^Y	$\dots$	m_K^Y	total
m_1^X	f_{11}	$\dots$	f_{1k}	$\dots$	$\boxed{f_{1K}}$	$f_{1\bullet}$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
m_j^X	f_{j1}	$\dots$	f_{jk}	$\dots$	f_{jK}	$f_{j\bullet}$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	
m_J^X	f_{J1}	$\dots$	f_{Jk}	$\dots$	f_{JK}	$f_{J\bullet}$
	$f_{\bullet 1}$	$\dots$	$f_{\bullet k}$	$\dots$	$f_{\bullet K}$	1

$f_{1\bullet} = \sum_k f_{1k}$

f_{1K} se déduit de $f_{1\bullet}$ et des K-1 premières f_{ij}.

fréquences liées aux autres

pas d'aléa sur les marges sous H_0.

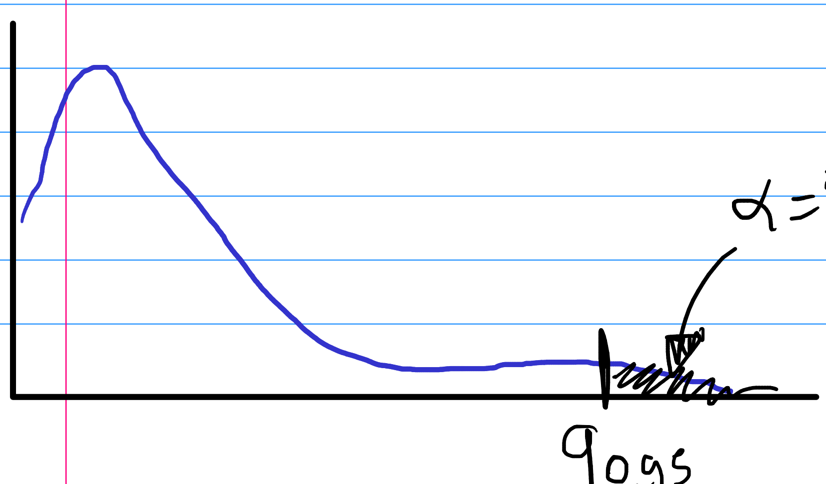
	m_1^Y	$\dots$	m_k^Y	$\dots$	m_K^Y	total
m_1^X	f_{11}	$\dots$	f_{1k}	$\dots$	f_{1K}	$f_{1\bullet}$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
m_j^X	f_{j1}	$\dots$	f_{jk}	$\dots$	f_{jK}	$f_{j\bullet}$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
m_J^X	f_{J1}	$\dots$	f_{Jk}	$\dots$	f_{JK}	$f_{J\bullet}$
	$f_{\bullet 1}$	$\dots$	$f_{\bullet k}$	$\dots$	$f_{\bullet K}$	1

Seules ces variables forment des variables indépendantes.
Il y en a $(L-1)(K-1)$

Donc $D_{\chi^2} \sim \chi^2((L-1)(K-1))$.

Décision du test : $\alpha = 5\%$.

$H_0 \Leftrightarrow$ indépendance $\Leftrightarrow D_{\chi^2} = 0$



→ Si $D_{\chi^2} > q_{0,95}$, rejet de H_0 au seuil 5%.

→ Si $D_{\chi^2} < q_{0,95}$, non rejet.

5.4 interprétation

Au seuil $\alpha\%$ (le plus souvent $\alpha = 5$), il faut comparer D_{χ^2} au quantile d'ordre $1 - \alpha\%$ à savoir $q_{1-\alpha}$ ($q_{0,95}$ le plus souvent) d'une loi du χ^2_d, où

$$d = (J - 1)(K - 1)$$

est le degré de liberté de la loi (c'est à dire le paramètre de la loi du χ^2).

L'interprétation est la suivante :

- si $D_{\chi^2} \geq q_{1-\alpha}$, on conclut que les deux variables sont dépendantes, (Rejet)
- sinon, on conclut qu'elles sont indépendantes. (Non rejet).

Les logiciels de statistiques (type R, Excel ...) calculent la $p-$value (ou valeur p). On retiendra qu'on rejettera l'hypothèse d'indépendance si $p \leq \alpha/100$ (le plus souvent si $p \leq 0,05$.)

→ $p < 5\%$ on rejette

→ $p > 5\%$ on ne rejette pas.

En pratique, on évite d'utiliser le test du χ^2 si un effectif du tableau est inférieur ou égal à 5 car l'approximation par le Théorème Central Limit est alors trop grossière.

5.5 Exemple

À partir de 200 dossiers d'une agence immobilière, on recense les réponses positives et négatives selon la situation maritale du demandeur (célibataire ou en couple). On obtient les résultats suivants :

	Célibataire	En couple
Dossier accepté	34	58
Dossier refusé	66	42

(i) On Donne le tableau des fréquences.

Pour calculer les fréquences, on divise chaque effectif par l'effectif total (ici 200) :

÷ 200

	Célibataire	En couple	Total
Dossier accepté	0.17	0.29	0.46
Dossier refusé	0.33	0.21	0.54
Total	0.5	0.5	1

marges.

	Célibataire	En couple	Total
Dossier accepté	0.17	0.29	0.46
Dossier refusé	0.33	0.21	0.54
Total	0.5	0.5	1

(ii) On Calcule la statistique du Chi-deux.
La statistique du Chi-deux est donnée par :

$$D_{\chi^2} = n \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(f_{jk} - f_{\bullet j} f_{k\bullet})^2}{f_{\bullet j} f_{k\bullet}}$$

Ici on a donc :

$$\begin{aligned} D_{\chi^2} =& \underbrace{200}_{\text{effectif total}}\Bigg(\frac{(0.17 - 0.46 \times 0.5)^2}{(0.46 \times 0.5)} + \frac{(0.29 - 0.46 \times 0.5)^2}{0.46 \times 0.5} + \frac{(0.33 - 0.54 \times 0.5)^2}{0.54 \times 0.5} \\ &+ \frac{(0.21 - 0.54 \times 0.5)^2}{0.54 \times 0.5}\Bigg) \\ =& 200\Big(0.016 + 0.016 + 0.013 + 0.013\Big) \\ =& 11.6 \end{aligned}$$

À comparer au quantile $q_{0,95}$ . . .

DDL :

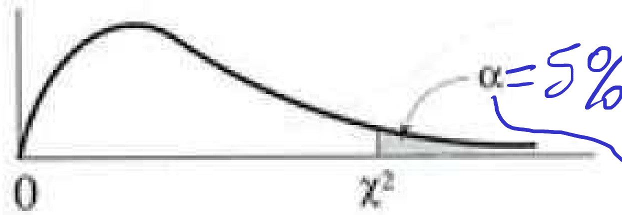
	Célibataire	En couple
Dossier accepté	34	58
Dossier refusé	66	42

$(2-1)(2-1) = 1$

DDL.

(iii) On fait le test du Chi-deux pour conclure.
On compare cette statistique à la valeur de la table du Chi-deux à 1 degré de liberté (2 modalités pour chaque variable).

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2



$q_{0,95} = 3,84$.

dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15

$D\chi^2 = 11,6$ $q_{0,95} = 3,84$. Au seuil 5%, on rejette l'indépendance.

On trouve 3.84.
On a donc $D_{\chi^2} > q_{0.95}$ et on conclue que les variables sont dépendantes : la situation maritale influence l'acceptation ou le refus du dossier.

6 Test du χ^2 pour l'ajustement d'une série à une loi de probabilité

Lorsqu'une loi statistique a une distribution qui ressemble celle d'une loi de probabilité connue, on peut se poser la question de leur adéquation. Des méthodes descriptives (qq-plot, droite de Henry...à suivre...) peuvent permettre une première approche. Les statistiques inférentielles, et en particulier le test du χ^2, peut donner un outil de choix supplémentaire.

On considère la série suivante détaillant le poids de 500 sacs de ciment.

Poids (kg)	effectif
[0,45]	35
]45,47]	53
]47,49]	76
]49,51]	100
]51,53]	88
]53,55]	78
]55,57]	42
]57,∞]	28

On souhaite savoir si cette série peut être ajustée par une loi Normale (m, σ). On peut faire une estimation ponctuelle des paramètres m et σ sur la série,

$$m = \frac{44 \times 35 + \cdots + 46 \times 53 + \cdots 58 \times 28}{500} \approx 50,78, \qquad \sigma \approx 3,74.$$

H_0 = "adéquation à $\mathcal{N}(50,78 ; 3,74)$".

⚠ on teste bien l'adéquation à la loi normale (pas moyenne ou écart-type).

Effectifs théoriques : $T_i = \underset{\text{effectif total}}{500}\, P\big(a_i \leq \mathcal{N}(50{,}78; 3{,}74) \leq b_i\big)$

pour l'effectif théorique de la classe $]a_i, b_i]$

On souhaite donc comparer la série observée à une loi normale $\mathcal{N}(50,78;3,74)$.

On note O_i les effectifs observés. On pose t_i les extrémités des classes. On complète le tableau avec les effectifs trouvés avec la loi gaussienne,

$$T_i = 500 \times (F(t_i) - F(t_{i-1})),$$

où F est la fonction de répartition de le loi normale $\mathcal{N}(50,78;3,74)$. On obtient

poids	O_i	T_i
[0,45]	35	30,5
]45,47]	53	47,5
]47,49]	76	80
]49,51]	100	104
]51,53]	88	99
]53,55]	78	74,5
]55,57]	42	40,5
]57,∞]	28	24

6.1 Hypothèse $\mathcal{H}_0$

On considère l'hypothèse

$$\mathcal{H}_0 = \text{“La série observée est distribuée selon une loi normale } \mathcal{N}(50,78;3,74)''.$$

6.2 Variable du test

On étudie la distance à l'adéquation des effectifs de la série observée à la série théorique

$$D_{\chi^2} = \sum_{i=1}^{l} \frac{(O_i - T_i)^2}{T_i}.$$

Theorem 6.1 *La variable du test D_{χ^2} suit une loi du χ^2 à $l-s-1$ degrés de liberté, où l est le nombre de modalités observées, s est le nombre de paramètres estimés (m,σ,...).*

Sous $\mathcal{H}_0$, proba d'être dans la classe O_i.

Preuve : $O_i = \sum_{k=1}^{n} X_k$ avec $X_k \sim \mathcal{B}(T_i/n)$

↳ 1 si le sac est dans la classe liée à O_i.

$O_i \sim \mathcal{B}(n, T_i/n)$

Pour n grand, O_i s'approche par une loi normale (TCL) de paramètres :

$$E\,O_i = n\frac{T_i}{n} = T_i$$

$$Var(O_i) = n\left(\frac{T_i}{n}\right)\underbrace{\left(1-\frac{T_i}{n}\right)}_{\text{petit}} \simeq n\left(\frac{T_i}{n}\right) = T_i.$$

O_i s'approche par $\mathcal{N}(T_i, \sqrt{T_i})$.

D'où $$\frac{O_i - T_i}{\sqrt{T_i}} \sim \mathcal{N}(0,1)$$

et $$D_{\chi^2} = \sum_i \frac{(O_i - T_i)^2}{T_i} \sim \chi^2(d).$$

d est le nombre d'aléa indépendants.

Les DDL :

poids	O_i	T_i
[0,45]	35	30,5
]45,47]	53	47,5
]47,49]	76	80
]49,51]	100	104
]51,53]	88	99
]53,55]	78	74,5
]55,57]	42	40,5
]57,∞]	28	24

$\sum_i O_i = n$.

imposé par H_0.

↳ l'un des effectifs se déduit des autres, il est lié.

On a fixé d'autres paramètres : $m = \sum_i \frac{T_i c_i}{n}$ fixé par H_0.

Pour chaque nouveau paramètre, on lie un nouveau T_i aux autres.

Donc, si l est le nombres de classes, s le nombre de paramètres estimés,

$$d = l - 1 - s$$

ddl

effectifs total connu

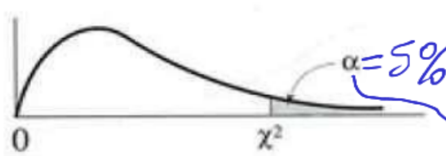
$$n = \sum_i T_i$$

poids	O_i	T_i
[0,45]	35	30,5
]45,47]	53	47,5
]47,49]	76	80
]49,51]	100	104
]51,53]	88	99
]53,55]	78	74,5
]55,57]	42	40,5
]57,∞]	28	24

$$D_{\chi^2} = \frac{(35-30,5)^2}{30,5} + \frac{(53-47,5)^2}{47,5} + \dots + \frac{(28-24)^2}{24}$$

$$\underline{DDL} = \underset{\text{classes}}{8} - \underset{n=500}{1} - \underset{\text{2 paramètres estimés } m,\ \sigma}{2} = 5.$$

Dans notre cas, la loi du χ^2 a 5 d.d.l., et $D_{\chi^2} = 3,76$.



$q_{0,95} = 3,84$

dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15

ddl 5

$q_{0,95} = 11,07$

$D_{\chi^2} = 3,76 < q_{0,95}$

Au seuil 5%, on ne rejette pas H_0.

On peut considérer que les observations sont en adéquation avec $\mathcal{N}(50,78\,;3,74)$.

6.3 Interprétation

Au seuil $\alpha\%$ (le plus souvent $\alpha = 5$), il faut comparer D_{χ^2} au quantile d'ordre $1 - \alpha\%$ à savoir $q_{1-\alpha}$ ($q_{0,95}$ le plus souvent) de la loi du χ^2_d

L'interprétation est la suivante :

- si $D_{\chi^2} \geq q_{1-\alpha}$, on conclut que les deux distributions ne peuvent pas être identiques,
- sinon, on ne rejette pas cette hypothèse.

Les logiciels de statistiques (type R, Excel ...) calculent la $p-$value (ou valeur p). On retiendra qu'on rejettera l'hypothèse d'adéquation si $p \leq \alpha/100$ (le plus souvent si $p \leq 0,05$.)

En pratique, dans ce cas également, on évite d'utiliser le test du χ^2 si un effectif du tableau est inférieur ou égal à 5 à cause de l'approximation avec le Théorème Central Limit.

Dans notre exemple, $q_{0,95} = 11,07$. Puisque $D_{\chi^2} \leq q_{0,95}$, on ne rejette pas l'adéquation des lois.