Test χ^2

0

0

Variables aléatoires – Loi du χ^2

Loi du χ^2 : On dit « loi du khi (lettre grecque) deux »

- **La loi de χ^2** ou loi de Pearson permet de comparer des distributions.
- **Elle permet :**
  - De **calculer la probabilité** d'observer **des écarts** dus au hasard entre des *fréquences observées* et celles *prévues par une loi de probabilité.*
  - de décider si une série d'effectifs observés diffère significativement d'une série d'effectifs théoriques (ou attendus)
- **Sa définition :**
  - Soit X_1, X_2....., X_i..... X_n, n variables normales centrée réduites, on appelle χ^2 la variable aléatoire définie par :

$$\chi^2 = X_1^2 + X_2^2 + ... + X_i^2 + ... + X_n^2 = \sum_{i=1}^{n} X_i^2$$

  - On dit que χ^2 suit loi de **khi deux** à ***n*** **degré de liberté** (ddl)

1

1

Variables aléatoires – Loi du χ^2

❑ Degré de liberté (*ddl*)

- **En générale :** Désigne le nombre de variables aléatoires qui ne peuvent être déterminées ou fixées par une équation : x + y = 6 ➔ *ddl* = n-1 = 2-1

Combien de case puis-je remplir librement ?

Modalités	Effectifs	
a	3	
b	6	3 à remplir!
c	7	
d	8	1 imposée
Total	**24**	

➔ *ddl* = 3

	Variable x		Totaux	
Variable x	**5**	19	24	1 à remplir!
	32	20	52	3 imposées
Totaux	37	39	76	

➔ *ddl* = 1

En règle générale : *ddl* = (C-1) x (L-1)

- **Loi de probabilité**

La variable qui suit cette loi résulte $\sum$ **Plusieurs variables indépendantes**

Suivent chacune une <u>même</u> loi de probabilité

On dit pour cette variable suit une loi de probabilités X Selon la loi suivit par

(Ex: X= χ^2) de n ddl, avec n Somme

Si ces variables sont dépendantes, à ce moment il y a perte de degrés de libertés

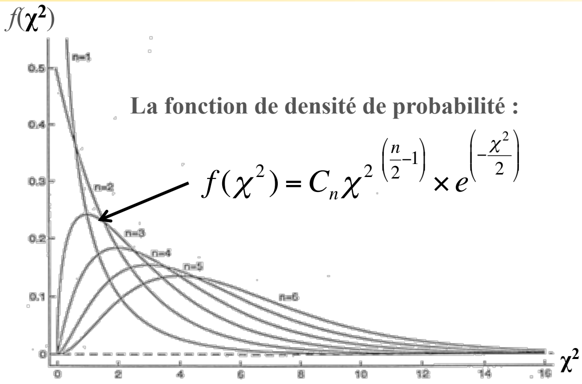
2

2

Variables aléatoires – Loi du χ^2

❑ Distribution

Sa distribution est **asymétrique** et dépend d'un seul paramètre ***n***. Ce dernier dépend du nombre de variables aléatoires normales indépendantes ***X*** intervenant dans la somme.

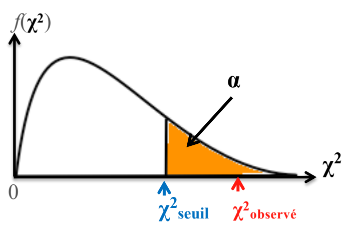


La fonction de densité de probabilité :

$$f(\chi^2) = C_n \chi^{2\left(\frac{n}{2}-1\right)} \times e^{\left(-\frac{\chi^2}{2}\right)}$$

❑ Table du χ^2

Elle donne, en fonction du nombre de degrés de liberté ***n***, les valeurs limites χ^2_α du χ^2 correspondant au coefficient de risque **α**



3

Variables aléatoires – Loi du χ^2

❑ **Exemple** : Pour $n = 8$ et $\alpha = 0{,}05$

En fait, dans les tests statistiques, on utilise souvent comme seuil de risque :

- $\alpha = 5\ \%$ et $n = 8$ $\chi^2_{5\%} = 15.51$
- $\alpha = 1\ \%$ et $n = 8$ $\chi^2_{1\%} = 20.09$

n \ α	0.99	. . .	. . .	0.05	. . .	0.010
1				.		.
.				.		.
8	. . .	. . .	. . .	$\chi^2_{5\%}$	. . .	$\chi^2_{1\%}$
.						
.						
30						

❑ **Remarques :**

- Pour toute la 1ère ligne, les valeurs sont celles du carré de la variable normale centrée réduite
- La table s'arrête pour n=30, au-delà on prend l'approximation de la loi normale et on utilise sa table
- L'espérance $E(\chi^2) = n$ et La variance $V(\chi^2) = 2n$

4

4

Loi χ^2

❑ Types

Test du $\chi 2$ de conformité.

- ✓ Sous H_0, les échantillons observés sont issus de la même population sous-jacente.
- ✓ Donc, les distributions devraient être identiques entre elles.

Test du $\chi 2$ d'indépendance.

- ✓ Soient X_1 et X_2 deux variables qualitatives.
- ✓ Sous H_0, la distribution de X_1 devrait être indépendante de celle de X_2.
- ✓ A contrario, la distribution des variables X_1 et X_2 sont indépendantes (H_1)

Test du $\chi 2$ d'ajustement.

- ✓ Sous H_0, l'échantillon observé provient de la population dont la distribution théorique est connue.
- ✓ Donc, la distribution observée est identique à la distribution théorique

5

5

Loi χ^2

❑ Tableau de contingence

Le test du χ^2 est réalisé sur des données qui ont une représentation bien spéciale. Cette dernière se base sur la notion d'effectifs synthétisés au sein d'un tableau de contingence. La forme de ce dernier dépend du type de test utilisé :

Test du χ2 conformité

Le tableau contient autant de colonnes qu'il y a d'échantillons observés.

Variable X	*Effectif n_i*				
	Échantillon 1	*Échantillon 2*	...	*Échantillon k*	*Total*
Modalité 1	n_{11}	n_{12}	...	n_{1k}	t_1
Modalité 2	n_{21}	n_{22}	...	n_{2k}	t_2
. . .	. . .	. . .	... ...	. . .	. . .
Modalité i	n_{i1}	n_{i2}	...	n_{ik}	t_i
Total	N_1	N_2	...	N_k	N_{total}

6

6

Loi χ^2

Le tableau contient autant de lignes que de modalités de la variable X_1 et autant de colonnes que de modalités de la variable X_2.

Test du χ^2 d'indépendance

Variable X_1	*Variable X_2*				
	Modalité *1*	Modalité *2*	...	Modalité *k*	*Total*
Modalité 1	n_{11}	n_{12}	...	n_{1k}	t_1
Modalité 2	n_{21}	n_{22}	...	n_{2k}	t_2
. . .	. . .	. . .	... ...	. . .	. . .
Modalité i	n_{i1}	n_{i2}	...	n_{ik}	t_i
Total	N_1	N_2	...	N_k	N_{total}

Test du χ^2 d'ajustement

Variable X	*Effectif n_i*
Modalité n_1	n_1
Modalité n_2	n_2
. . .	. . .
Modalité n_k	n_k

Le tableau ne comporte qu'une seule colonne, car la distribution d'une variable est observée sur un seul échantillon. A chaque modalité X_i de la variable est associé à son effectif observée n_i.

7

7

Variables aléatoires – Loi du χ^2

❑ Le test

Exemple : Ce tableau présente le nombre les réponses Oui/Non à une question posée à 179 sujets concernat l'effet positif de deux compléments alimentaires. On cherche s'il existe un lien entre le type du complément (IMPO/LOC) et sont effet sur les sujets ?

	Oui	Non
Complément IMP	68	60
Complément LOC	47	14

A. Identifier si le test χ^2 est applicable :

- Permet de savoir si une distribution est conforme à une distribution qu'on connaît.
- Permet, à partir de l'observation de l'échantillon, de décider si 2 variables nominales sont indépendantes ou non.

B. Poser l'Hypothèse :

- **Hypothèse nulle (H_0)** : Hypothèse selon laquelle il n'y a pas de lien entre les variables. C'est à partir de cette hypothèse qu'est calculée la probabilité p.
- **Hypothèse alternative (H_1)** : Hypothèse selon laquelle il y a un lien entre les variables

H_0 : « il n'y a pas de lien statistique significatif entre le type du complément et sont effet sur les sujets ".

8

8

Variables aléatoires – Loi χ^2

D. Calculer les totaux :

	Oui	Non	Total
Complément IMP	68	60	**128**
Complément LOC	47	14	**51**
Total	**105**	**74**	**179**

- La ligne et la colonne de totaux sont appelées marges du tableau.

58.66% $= \frac{105}{179} \times 100$ **41.34%** $= \frac{74}{179} \times 100$

E. Calculer des effectifs théoriques (tableau d'indépendance):

- Ce calcul s'effectue grâce aux marges du tableau.

	Oui	Non	Total
Complément IMP	75.1	52.9	**128**
Complément LOC	29.9	21.1	**51**
Total	**105**	**74**	**179**

$128 \times 0.5866 = 75.1$
$128 \times 0.4134 = 52.9$

$51 \times 0.5866 = 29.9$
$51 \times 0.4134 = 21.1$

9

9

Variables aléatoires – Loi χ^2

E. Calculer les écarts à l'indépendance :

- Calcul des écarts quadratique entre les fréquences observées et théoriques.
- Puis standardiser par la fréquence théorique

	Oui	Non	Total
Complément IMP	0.668	0.948	-
Complément LOC	1.670	2.380	-
Total	-	-	$\boxed{\textbf{5.66}}$

$$\frac{\left(n_{i\,k-observé} - n_{i\,k-théorique}\right)^2}{n_{i\,k-théorique}} = \frac{(68-75{,}1)^2}{75{,}1}$$

$$\chi^2_{obs} = \sum_{i=1}^{n}\sum_{k=1}^{j} \frac{\left(n_{i\,k-observé} - n_{i\,k-théorique}\right)^2}{n_{i\,k-théorique}} = 5{,}66$$

F. Déterminer $\chi^2_{n,\,\alpha}$ depuis la table :

- Degré de liberté $\boldsymbol{n}$ = (nbre de lignes – 1) x (nbre de colonnes – 1) ➔ (2 – 1) x (2 – 1) =1
- Le seuil de risque $\boldsymbol{\alpha}$ = 5 %
- Lire la table : déterminons $\chi^2_{n,\,\alpha}$ théorique qui à 5 % de chance d'être dépassée avec une loi du χ^2 observée à 1d.d.l : $\boldsymbol{n}$ = 1 et $\boldsymbol{\alpha}$ = 0.05 ➔ $\chi^2_{1,\,0.05} = \boxed{\mathbf{3.84}}$

Seuil de rejet

10

10

Variables aléatoires – Loi χ^2

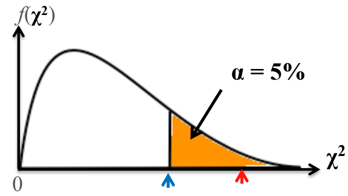
❑ La table inverse du χ^2 pour le calcul du χ^2 théorique (p = 1- α)

n\p	0,005	0,010	0,025	0,050	0,100	0,250	0,500	0,750	0,900	0,950	0,975	0,990	0,995
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	32,0	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,56	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0

11

Variables aléatoires – Loi χ^2

5.66 > 3.88 ➔ $\chi^2_{calculé} > \chi^2_{seuil\ critique}$



Donc on rejet l'hypothèse H_0

« il existe un de lien statistique significatif entre le type du complément et sont effet sur les sujets ".

En règle générale

✓ **Si $\chi^2_{calculé} > \chi^2_{seuil\ critique}$ ➔ on rejette l'hypothèse H_0 on garde H_1**

✓ **Si $\chi^2_{calculé} < \chi^2_{seuil\ critique}$ ➔ on garde l'hypothèse H_0**

12