Comparaison des distributions de caractères qualitatifs: Le Test du Chi-Deux (χ^2)

TALHI.R
Service de Biostatistique, Faculté de Médecine d'Oran
2020 - 2021

Formulations équivalentes:

Test du chi-deux = Test du chi-carré = Test de Pearson

Le test du χ^2 permet de tester la liaison entre deux ou plusieurs distributions de caractères qualitatifs.

Quelle que soit la nature d’un test, son principe et sa chronologie sont toujours les mêmes.

Selon la situation on distingue:

1-Le test du chi deux de conformité (ou d'ajustement)

2-Le test du chi deux d'homogénéité

3-Le test du chi deux d'indépendance

4-Le test du chi deux corrigé (correction de Yates)

1- Le Chi deux
de conformité (ou d'ajustement)

- Il sert à comparer une distribution observée sur un échantillon à une distribution connue dans la population ou à une distribution théorique.

Exemple n°1

- Dans une maternité , sur 100 naissances, on a observé 44 garçons et 56 filles; cette observation est- elle compatible avec la statistique nationale donnant les proportions de naissances de garçons et de filles de respectivement 53% et 47% ?

Tableau de contingence:

	Garçons	Filles	Total
Statistique nationale	0,53	0,47	1
Effectif théorique (calculé) T_i ou C_i	53	47	100
Effectif observé O_i	44	56	100

Démarche à suivre:

1- **H_0 :**la distribution par sexe observée à la maternité est conforme à la distribution national

2-choix du test → un χ^2 de conformité

3-vérification des conditions d'application :

Tous les Ti ≥ 5 (53 et 47 sont > 5)

$\boxed{\mathbf{Ci = Ti = n \times fi}}$ avec

n= L'effectif total de l'échantillon

fi= Fréquence de chaque classe de la variable dans la population ou dans la distribution théorique.

4-Calcul de la statistique (somme des écarts) :

$$\boxed{\chi^2 = \sum \frac{(Ci - Oi)^2}{Ci}}$$

$\chi^2 = (53\text{-}44)^2 / 53 + (47\text{-}56)^2 / 47 =$ **3, 25**

5-on fixe le seuil de signification au risque **α = 5%**

6-on compare le χ^2 calculé au χ^2 de la table au risque α = 5% et avec un nombre de degré de liberté

$$\boxed{\mathbf{ddl = k - 1}}$$

(k = nombre de modalités de la variable étudiée) $\rightarrow$ ddl = 2-1=1

Table du χ^2

α / d.d.l	0,9	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,015	0,455	1,074	1,642	2,706	**3,841**	5,412	6,635	10,82
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,81
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,34	16,26
4	1,064	3,357	4,878	5,989	7,779	9,488	11,66	13,27	18,46
5	1,610	4,351	6,064	7,289	9,236	11,07	13,38	15,08	20,51
6	2,204	5,348	7,231	8,558	10,64	12,59	15,03	16,81	22,45

Exemple: avec α = 5 % et ddl = 1 $\rightarrow$ χ^2 tabulaire = 3, 84

χ^2 calculé = 3, 25 $\rightarrow$ χ^2 calculé < χ^2 de la table (3, 25 **<** 3, 84)

7-conclusion:

χ^2 calculé < χ^2 de la table (3, 25 **<** 3, 84) $\rightarrow$

H_0 n'est pas rejetée au risque α = 5% $\rightarrow$

la distribution observée est **conforme** à la répartition nationale par sexe des naissances.

2- Le chi- deux d'homogénéité

- Il sert à comparer deux ou plusieurs distributions observées sur des échantillons.

Exemple n°2:

Deux médicaments (A et B) ont été testés sur deux groupes de malades.
A l'issue de l'essai, on a observé les résultats suivants :

	Disparition des symptômes	Persistance des symptômes	Aggravation	Réaction secondaire	Total
A	100	40	20	30	190
B	220	80	70	40	410
Total	320	120	90	70	600

Peut-on dire que ces deux traitements ont les mêmes effets ?

1-H_0 :
les deux traitements ont les mêmes effets →
il n’y a pas de différence entre les deux distributions .

2-tableau de contingence:

	Disparition des symptômes	Persistance des symptômes	Aggravation	Réaction secondaire	Total
A	100 (101,33)	40 (38)	20 (28,5)	30 (22,16)	190
B	220 (218,66)	80 (82)	70 (61,5)	40 (47,83)	410
Total	320	120	90	70	600

3-choix du test → un χ^2 d'homogénéité

4-vérification des conditions d'application : Tous les Ti ≥ 5

calcul des Ti:

$\boxed{\textbf{Ti = (total de la ligne × total de la colonne) / total général}}$

(101,33 - 218,66 - 38 - 82 - 28,5 - 61,5 - 22,16 et 47,83) sont tous > 5

5-Calcul de la statistique :

$$\boxed{\chi^2 = \sum \frac{(Ci - Oi)^2}{Ci}}$$

$\chi^2 = (100 - 101{,}33)^2 / 101{,}33 + (220 - 218{,}66)^2 / 218{,}66 + (40 - 38)^2 / 38$
$+ (80 - 82)^2 / 82 + (20 - 28{,}5)^2 / 28{,}5 + (70 - 61{,}5)^2 / 61{,}5 +$
$(30 - 22{,}16)^2 / 22{,}16 + (40 - 47{,}83)^2 / 47{,}83$

$$\boldsymbol{\chi^2 = 7{,}94}$$

6-on fixe le seuil de signification au risque **α = 5%**

7-on compare le χ^2 calculé au χ^2 de la table au risque α = 5% et avec un nombre de degré de liberté

$$\boxed{\mathbf{ddl = (C-1) \times (L - 1)}}$$

C= nombre de colonnes
L= nombre de lignes

ddl = (4-1) × (2-1) = 3

Table du χ^2

α / d.d.l	0,9	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,015	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,82
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,81
3	0,584	2,366	3,665	4,642	6,251	**7,815**	9,837	11,34	16,26
4	1,064	3,357	4,878	5,989	7,779	9,488	11,66	13,27	18,46
5	1,610	4,351	6,064	7,289	9,236	11,07	13,38	15,08	20,51
6	2,204	5,348	7,231	8,558	10,64	12,59	15,03	16,81	22,45

Exemple: avec α = 5 % et ddl = 3 $\rightarrow$ χ^2 tabulaire = 7,815

χ^2 calculé = 7,94 $\rightarrow$ χ^2 calculé > χ^2 de la table

8-conclusion:

χ^2 calculé > χ^2 de la table (7,94 > 7,81) →

$\mathbf{H_0}$ est rejetée →

La différence est significative au risque α = 5% →

les deux distributions **ne sont pas homogènes.**

3- Le chi- deux d'indépendance

- Il sert à étudier sur un même échantillon la liaison entre les distributions de deux variables.

Exemple n°3

- Un épisode d'intoxication alimentaire collective (TIAC) est survenu parmi les élèves d'une école primaire à Oran.

Le docteur X.Y fut chargé de l'enquête. Il a dressé le tableau suivant croisant l'information sur la consommation de glace au chocolat, l'un des desserts proposés au cours du dernier repas pris en commun par les écoliers, et le statut malade / non malade.

Tableau de contingence:

	Malades	Sains	Total
Glace au chocolat	69 (76)	83 (76)	152
Pas de glace	31 (24)	17 (24)	48
Total	100	100	200

H_0: il n’y a pas de liaison entre la consommation de glace au chocolat et la survenue de la gastro- entérite.

Le principe et le calcul du test sont identiques a ceux d'un χ^2 d'homogénéité

Le test utilisé : χ^2 d'indépendance
χ^2 calculé = **5,4**
ddl = **1**
χ^2 calculé > χ^2 de la table (5,4 > 3,84) →
$\mathbf{H_0}$ est rejetée au risque $\alpha = 5\%$ →
Il existe une liaison entre la consommation de glace au chocolat et la survenue de la gastro- entérite .

4- L a correction de Yates

- S’applique aux tableaux de contingence **2×2** (ou tableau a quatre cases).
- On utilise un **χ^2 corrigé**
- Les conditions d’application:

Au moins un des effectifs théoriques est **< 5** et tous sont **≥ 3**

Exemple n°4:

On compare les résultats de deux traitements A et B :

	Traitement A	Traitement B	Total
Échec	17 (18,33)	18 (16,66)	35
Succès	5 **(3,66)**	2 **(3,33)**	7
Total	22	20	42

Peut-on dire que ces deux traitements sont équivalents ?

H_0: il n'existe pas de différence entre les deux traitements ou les deux traitements sont équivalents.

- Deux effectifs théoriques sont < 5 (**3,66** et **3,33**) et tous les effectifs sont >3

- Calcul de la statistique:

$$\boxed{\chi^2 \text{ corrigé} = \sum \frac{(\lvert C_i - O_i \rvert - 0{,}5)^2}{C_i}}$$

χ^2 calculé = **1,35** et ddl = 1

χ^2 calculé **<** χ^2 tabulaire   (1,31 < 3,84) →

$\mathbf{H_0}$ est non rejetée au risque α = 5% →

la différence n'est pas significative → les deux traitements sont équivalents.

Remarque:

Si au moins un des effectifs théoriques est **< 3** →

il existe une façon de tester l’homogénéité de deux variables binaires: le test exact de Fisher .

Les références:

- Dabis F, Drucker J, Moren A. Épidémiologie d’intervention. Édition 1992
- Schwartz D. Méthodes statistiques.1992
- Ancelle T. Statistique Épidémiologie. Édition 2002
- Mesli MF, Mokhtari A. Biostatistique. Édition mai 2007