Chap. 4 – **Tests du Khi-deux**

M 3102 – Estimation et Tests d'hypothèses

IUT Aurillac – STID/CS

UNIVERSITÉ Clermont Auvergne

3 types de **Test du Khi-deux** :

- Test de **conformité** (ou adéquation) :
  $\rightsquigarrow$ test pour voir si les données sont en adéquation avec un modèle théorique.

- Test d'**homogénéité** :
  $\rightsquigarrow$ test pour comparer l'homogénéité de deux échantillons ou plus.

- Test d'**indépendance** :
  $\rightsquigarrow$ test de l'indépendance de deux caractères qualitatifs.

On définit sur une population un système complet d'événements $\{A_1, ...A_I\}$.

♢ **Loi théorique** à tester : on connait les **probabilités** $p_1, ...p_I$ de ces événements : $p_i = P(A_i)$.

♢ Observations sur un échantillon de taille $n \Longrightarrow$ **effectifs** $O_1, ...O_I$ correspondant aux événements.

Idée : comparer ces effectifs observés avec les **effectifs théoriques** $E_1, ..., E_I$,

où $E_i = n \times p_i$, calculés à partir des probabilités théoriques $p_1, ..., p_I$

Événements	A_1	A_2	...	A_I	Total
Effectifs observés	O_1	O_2	...	O_I	n
Effectifs théoriques	E_1	E_2	...	E_I	n

Test de conformité ou d'adéquation

Processus de test comme Chapitre 3 - On rajoute une **étape 0 de construction du tableau**.

1. Deux hypothèses possibles :

   - **L'hypothèse nulle** (H_0) : la population est conforme à la loi. Elle consiste à dire que les différences entres les effectifs observés et les effectifs théoriques ne sont pas assez importantes pour être significatives.

   - **L'hypothèse alternative** (H_1) : la population n'est pas conforme à la loi. Les différences d'effectifs sont trop significatives.

2. α reste quelconque dans un cadre général.

3. La **statistique de test** à utiliser est :

$$\chi^2 = \sum_{i=1}^{I} \frac{(O_i - E_i)^2}{E_i}.$$

   Conditions d'application : sous peine de biaiser les résultats, on demande à ce que tous les effectifs soient supérieurs ou égaux à 5 : $E_i \geq 5$. Si ce n'est pas le cas, il faudra regrouper certains événements ensembles et refaire un tableau comme au dessus avec les nouveaux effectifs.

$$\chi^2 = \sum_{i=1}^{I} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(I - 1 - r),$$

où r est le nombre de paramètres à estimer pour connaître la loi théorique (par exemple, s'il faut estimer une moyenne, on aura $r = 1$).

4. Région critique : $W = [k_{1-\alpha}^{I-1-r}; +\infty[$.

5. Calcul de la statistique.

6. Décision et conclusion :

   - Si $\chi^2 \in W$, c'est-à-dire $\chi^2 \geq k_{1-\alpha}^{I-1-r}$, on décide de rejeter (H_0) au risque α. Cela peut vouloir dire que l'échantillon n'est pas conforme à la loi, ou alors que la loi n'est pas vérifiée (ce n'est pas tout à fait la même chose...)

   - Sinon, on ne rejette pas (H_0), ce qui signifie qu'il est impossible de prouver au risque α que la population n'est pas conforme à cette loi théorique.

Test de conformité ou d'adéquation : exemple

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. Les résultats sont consignés dans le tableau ci-dessous :

Nombre de clients	0	1	2	3	4	≥ 5
Nombres de magasins	35	50	40	20	6	3

On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Test de conformité ou d'adéquation : exemple

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. Les résultats sont consignés dans le tableau ci-dessous :

Nombre de clients	0	1	2	3	4	≥ 5
Nombres de magasins	35	50	40	20	6	3

On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

0. Construction du tableau :

   Les événements sont donnés par : A_0 : "il y a 0 client dans le magasin entre 9h et 10h", A_1 : "il y a 1 client dans le magasin entre 9h et 10h", ... , A_5 : "il y a 5 clients ou plus dans le magasin entre 9h et 10h".

Test de conformité ou d'adéquation : exemple

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. Les résultats sont consignés dans le tableau ci-dessous :

Nombre de clients	0	1	2	3	4	≥ 5
Nombres de magasins	35	50	40	20	6	3

On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

0. Construction du tableau :

   Les événements sont donnés par : A_0 : "il y a 0 client dans le magasin entre 9h et 10h", A_1 : "il y a 1 client dans le magasin entre 9h et 10h", ... , A_5 : "il y a 5 clients ou plus dans le magasin entre 9h et 10h".

   Il nous faut trouver λ, qui coïncide avec l'espérance de la loi. On pourra alors calculer les p_i grâce à la formule : $p_i = P(X = i) = e^{-\lambda} \times \frac{\lambda^i}{i!}$.

Test de conformité ou d'adéquation : exemple

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. Les résultats sont consignés dans le tableau ci-dessous :

Nombre de clients	0	1	2	3	4	≥ 5
Nombres de magasins	35	50	40	20	6	3

On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

0. Construction du tableau :

   Les événements sont donnés par : A_0 : "il y a 0 client dans le magasin entre 9h et 10h", A_1 : "il y a 1 client dans le magasin entre 9h et 10h", ... , A_5 : "il y a 5 clients ou plus dans le magasin entre 9h et 10h".

   Il nous faut trouver λ, qui coïncide avec l'espérance de la loi. On pourra alors calculer les p_i grâce à la formule : $p_i = P(X = i) = e^{-\lambda} \times \frac{\lambda^i}{i!}$.

   On estime λ grâce aux données : $\lambda = \mathbb{E}(X) \simeq 1.49$. On pose alors $\lambda = 1.5$.

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Calculons les p_i et les E_i :

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Calculons les p_i et les E_i :

Événements	A_0	A_1	A_2	A_3	A_4	A_5	Total
Eff observés							
p_i							
Eff théoriques							

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Calculons les p_i et les E_i :

Événements	A_0	A_1	A_2	A_3	A_4	A_5	Total
Eff observés	35	50	40	20	6	3	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0471	0.0186	1
Eff théoriques	34.35	51.54	38.65	19.33	7.25	2.86	154

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Calculons les p_i et les E_i :

Événements	A_0	A_1	A_2	A_3	A_4	A_5	Total
Eff observés	35	50	40	20	6	3	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0471	0.0186	1
Eff théoriques	34.35	51.54	38.65	19.33	7.25	2.86	154

La dernière classe a un effectif inférieur à 5, il faut donc regrouper les deux derniers événements. Le tableau devient :

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Calculons les p_i et les E_i :

Événements	A_0	A_1	A_2	A_3	A_4	A_5	Total
Eff observés	35	50	40	20	6	3	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0471	0.0186	1
Eff théoriques	34.35	51.54	38.65	19.33	7.25	2.86	154

La dernière classe a un effectif inférieur à 5, il faut donc regrouper les deux derniers événements. Le tableau devient :

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

1. On pose (H_0) : la population suit une loi de Poisson (avec $\lambda = 1.5$) et (H_1) : la population ne suit pas une loi de Poisson.

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

1. On pose (H_0) : la population suit une loi de Poisson (avec $\lambda = 1.5$) et (H_1) : la population ne suit pas une loi de Poisson.

2. $\alpha = 5\%$.

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

1. On pose (H_0) : la population suit une loi de Poisson (avec $\lambda = 1.5$) et (H_1) : la population ne suit pas une loi de Poisson.
2. $\alpha = 5\%$.
3. Sous l'hypothèse (H_0), la statistique de test à utiliser est :

$$\chi^2 = \sum_{i=1}^{I} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(I - 1 - r).$$

Les conditions d'effectifs sont bien vérifiées.

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

4. $W = [7.81; +\infty[$ d'après la table de la loi du Khi-deux à $5 - 1 - 1 = 3$ degrés de liberté (5 événements après regroupement, 1 paramètre à estimer λ).

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

4. $W = [7.81; +\infty[$ d'après la table de la loi du Khi-deux à $5-1-1=3$ degrés de liberté (5 événements après regroupement, 1 paramètre à estimer λ).

5. On a : $\chi^2 = \frac{(35-34.35)^2}{34.35} + \frac{(50-51.54)^2}{51.54} + ... + \frac{(9-10.11)^2}{10.11} \simeq 0.254$.

Test de conformité ou d'adéquation : exemple (suite)

Une enquête auprès d'un groupe de commerçant du centre-ville d'Aurillac a permis de relever le nombre de clients entrés dans le magasin pendant une heure de 9h à 10h. On cherche à savoir si, au risque 5%, on peut admettre que la population suit une loi de Poisson.

Événements	A_0	A_1	A_2	A_3	A_4	Total
Eff observés	35	50	40	20	9	154
p_i	0.2231	0.3347	0.2510	0.1255	0.0657	1
Eff théoriques	34.35	51.54	38.65	19.33	10.11	154

4. $W = [7.81; +\infty[$ d'après la table de la loi du Khi-deux à $5 - 1 - 1 = 3$ degrés de liberté (5 événements après regroupement, 1 paramètre à estimer λ).

5. On a : $\chi^2 = \frac{(35-34.35)^2}{34.35} + \frac{(50-51.54)^2}{51.54} + ... + \frac{(9-10.11)^2}{10.11} \simeq 0.254$.

6. Décision et conclusion : Comme $\chi^2 \notin W$, on ne rejette pas (H_0) au risque 5%. Cela revient à dire (à notre niveau) que la population suit une loi de Poisson.

Test d'homogénéité

$\rightsquigarrow$ **comparer deux échantillons** ou plus.

On dispose d'une population P, et d'un caractère pouvant prendre un ensemble de valeurs ou modalités (pas forcément des valeurs numériques).

Ces valeurs ou **modalités** sont notées $A_1, \dots, A_I$ (par exemple : employé en CDD, en CDI, en Intérim...)

On dispose de J **échantillons** notés $L_1, \dots, L_J$.

On connaît enfin l'**effectif observé** de la valeur A_i dans l'échantillon L_j qui sera noté $O_{i,j}$.

On a alors : $N = \sum_{i=1}^{I} \sum_{j=1}^{J} O_{i,j}$, l'**effectif total** des échantillons.

On aura besoin de présenter ces résultats sous forme d'un tableau :

Événts	A_1	A_2	...	A_i	...	A_I	Totaux
Ech. L_1	$O_{1,1}/E_{1,1}$	$O_{2,1}/E_{2,1}$	...	$O_{i,1}/E_{i,1}$	...	$O_{I,1}/E_{I,1}$	N_1
Ech. L_2	$O_{1,2}/E_{1,2}$	$O_{2,2}/E_{2,2}$	...	$O_{i,2}/E_{i,2}$	...	$O_{I,2}/E_{I,2}$	N_2
...	...	...	...	...	...	...	...
Ech. L_j	$O_{1,j}/E_{1,j}$	$O_{2,j}/E_{2,j}$	...	$O_{i,j}/E_{i,j}$	...	$O_{I,j}/E_{I,j}$	N_j
...	...	...	...	...	...	...	...
Ech. L_J	$O_{1,J}/E_{1,J}$	$O_{2,J}/E_{2,J}$	...	$O_{i,J}/E_{i,J}$	...	$O_{I,J}/E_{I,J}$	N_J
Totaux	T_1	T_2	...	T_i	...	T_I	$N = \sum_{i=1}^{I} T_i = \sum_{j=1}^{J} N_j$

On calcule les **valeurs** $E_{i,j} = \frac{T_i \times N_j}{N}$ pour remplir le tableau précédent.

On passe au processus de test ...

Test d'homogénéité

1. On a alors deux hypothèses possibles :
   - **L'hypothèse nulle** (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié.

Test d'homogénéité

1. On a alors deux hypothèses possibles :
   - **L'hypothèse nulle** (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié.

   $\rightsquigarrow$ Différences entres les observations réalisées sur les échantillons : pas assez importantes.
   $\rightsquigarrow$ Soit les échantillons proviennent bien d'une même population, soit la population est homogène sur ce caractère.

Test d'homogénéité

1. On a alors deux hypothèses possibles :
   - **L'hypothèse nulle** (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié.

   $\rightsquigarrow$ Différences entres les observations réalisées sur les échantillons : pas assez importantes.
   $\rightsquigarrow$ Soit les échantillons proviennent bien d'une même population, soit la population est homogène sur ce caractère.

   - **L'hypothèse alternative** (H_1) : les populations ne sont pas homogènes.

   $\rightsquigarrow$ Différences d'effectifs trop significatives.

Test d'homogénéité

1. On a alors deux hypothèses possibles :
   - **L'hypothèse nulle** (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié.

   $\rightsquigarrow$ Différences entres les observations réalisées sur les échantillons : pas assez importantes.
   $\rightsquigarrow$ Soit les échantillons proviennent bien d'une même population, soit la population est homogène sur ce caractère.

   - **L'hypothèse alternative** (H_1) : les populations ne sont pas homogènes.

   $\rightsquigarrow$ Différences d'effectifs trop significatives.

2. α reste quelconque dans un cadre général.

Test d'homogénéité

1. On a alors deux hypothèses possibles :
   - **L'hypothèse nulle** (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié.

   $\rightsquigarrow$ Différences entres les observations réalisées sur les échantillons : pas assez importantes.
   $\rightsquigarrow$ Soit les échantillons proviennent bien d'une même population, soit la population est homogène sur ce caractère.

   - **L'hypothèse alternative** (H_1) : les populations ne sont pas homogènes.

   $\rightsquigarrow$ Différences d'effectifs trop significatives.

2. α reste quelconque dans un cadre général.

3. La **statistique de test** est :

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((I-1)(J-1)).$$

<u>**Conditions d'application**</u> : il faut que tous les **effectifs soient supérieurs ou égaux à** 5 : $E_{i,j} \geq 5$. Sinon on regroupe certains événements ensembles, et on refait le tableau.

Test d'homogénéité

4. Région critique : $W = [k_{1-\alpha}^{(I-1)(J-1)}; +\infty[$.

5. Calcul de la statistique.

6. Décision et conclusion :

   Si $\chi^2 \in W$, c'est-à-dire $\chi^2 \geq k_{1-\alpha}^{(I-1)(J-1)}$, on décide de rejeter (H_0) au risque α. Cela veut dire qu'avec $\alpha\%$ de chance de se tromper, on peut affirmer que les populations ne sont pas homogènes par rapport au caractère considéré.

   Sinon, on ne rejette pas (H_0), cela revient à conclure (à notre niveau) que la population est homogène. Cela veut aussi dire qu'on ne peut distinguer les populations en utilisant ce caractère comme justificatif.

Test d'homogénéité : exemple

On cherche à savoir si les élèves de deux collèges sont inscrits en proportion égale à leur club sportif respectif.
On note P_1 la population d'élève du collège A et P_2 la population d'élève du collège B.
On va réaliser un test d'homogénéité des deux populations sur le caractère "inscription au club sportif" dont les modalités sont "oui" et "non".

0. Construction du tableau :

Échantillon Participation	Oui	Non	Total
Collège A	12	38	50
Collège B	26	34	60
Total	38	72	110

Test d'homogénéité : exemple

On cherche à savoir si les élèves de deux collèges sont inscrits en proportion égale à leur club sportif respectif.
On note P_1 la population d'élève du collège A et P_2 la population d'élève du collège B.
On va réaliser un test d'homogénéité des deux populations sur le caractère "inscription au club sportif" dont les modalités sont "oui" et "non".

0. Construction du tableau :

Échantillon Participation	Oui	Non	Total
Collège A	12	38	50
Collège B	26	34	60
Total	38	72	110

On complète avec les valeurs des effectifs théoriques $E_{i,j}$:

Test d'homogénéité : exemple

On cherche à savoir si les élèves de deux collèges sont inscrits en proportion égale à leur club sportif respectif.
On note P_1 la population d'élève du collège A et P_2 la population d'élève du collège B.
On va réaliser un test d'homogénéité des deux populations sur le caractère "inscription au club sportif" dont les modalités sont "oui" et "non".

0. Construction du tableau :

Échantillon Participation	Oui	Non	Total
Collège A	12	38	50
Collège B	26	34	60
Total	38	72	110

On complète avec les valeurs des effectifs théoriques $E_{i,j}$:

Échantillon Participation	Oui	Non	Total
Collège A	12 / 17.27	38 / 32.73	50
Collège B	26 / 20.73	34 / 39.27	60
Total	38	72	110

Test d'homogénéité : exemple (suite)

On cherche à savoir si les élèves de deux collèges sont inscrits en proportion égale à leur club sportif respectif.
On note P_1 la population d'élève du collège A et P_2 la population d'élève du collège B.
On va réaliser un test d'homogénéité des deux populations sur le caractère "inscription au club sportif" dont les modalités sont "oui" et "non".

Échantillon Participation	Oui	Non	Total
Collège A	12 / 17.27	38 / 32.73	50
Collège B	26 / 20.73	34 / 39.27	60
Total	38	72	110

1. Hypothèses : (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié et (H_1) : les populations ne sont pas homogènes.

2. $\alpha = 5\%$.

Test d'homogénéité : exemple (suite)

On cherche à savoir si les élèves de deux collèges sont inscrits en proportion égale à leur club sportif respectif.
On note P_1 la population d'élève du collège A et P_2 la population d'élève du collège B.
On va réaliser un test d'homogénéité des deux populations sur le caractère "inscription au club sportif" dont les modalités sont "oui" et "non".

Échantillon Participation	Oui	Non	Total
Collège A	12 / 17.27	38 / 32.73	50
Collège B	26 / 20.73	34 / 39.27	60
Total	38	72	110

1. Hypothèses : (H_0) : les populations d'échantillon sont homogènes par rapport au caractère étudié et (H_1) : les populations ne sont pas homogènes.
2. $\alpha = 5\%$.
3. La **statistique de test** est : $\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((2-1)(2-1))$.
   Conditions d'application : OK, $E_{i,j} \geq 5$.

Test d'homogénéité : exemple (suite)

Échantillon Participation	Oui	Non	Total
Collège A	12 / 17.27	38 / 32.73	50
Collège B	26 / 20.73	34 / 39.27	60
Total	38	72	110

4. Région critique : $W = [k_{1-\alpha}^{(2-1)(2-1)}; +\infty[= [k_{0.95}^{1}; +\infty[= [3.841; +\infty[$.

Test d'homogénéité : exemple (suite)

Échantillon Participation	Oui	Non	Total
Collège A	12 / 17.27	38 / 32.73	50
Collège B	26 / 20.73	34 / 39.27	60
Total	38	72	110

4. Région critique : $W = [k_{1-\alpha}^{(2-1)(2-1)}; +\infty[= [k_{0.95}^{1}; +\infty[= [3.841; +\infty[$.

5. Calcul : $\chi^2 = \frac{(12-17.27)^2}{17.27} + ... + \frac{(34-39.27)^2}{39.27} \simeq 4.5$.

Test d'homogénéité : exemple (suite)

Échantillon	Participation	Oui	Non	Total
Collège A		12 / 17.27	38 / 32.73	50
Collège B		26 / 20.73	34 / 39.27	60
Total		38	72	110

4. Région critique : $W = [k_{1-\alpha}^{(2-1)(2-1)}; +\infty[= [k_{0.95}^{1}; +\infty[= [3.841; +\infty[$.

5. Calcul : $\chi^2 = \frac{(12-17.27)^2}{17.27} + \ldots + \frac{(34-39.27)^2}{39.27} \simeq 4.5$.

6. Décision et conclusion : $4.5 \in W$, on décide donc de rejeter (H_0) au risque 5%. Cela veut dire que les deux collèges ne sont pas homogènes par rapport au nombre d'adhérent de leur club sportif.

Test d'indépendance

On dispose d'**une population** où chaque individu est repéré par **deux caractères qualificatifs** A et B qui prennent les modalités $A_1, ..., A_I$ et $B_1, ..., B_J$.

Exemple : $A \rightarrow$ couleur des yeux dont les modalités seraient : bleus, marrons, verts, noirs...

On connaît l'**effectif observé** $O_{i,j}$ de la modalité $A_i \cap B_j$.

On a alors $N = \sum_{i,j} O_{i,j}$ l'**effectif total** de l'échantillon.

On cherche à savoir si les **caractères** A **et** B **sont indépendants ou non**.

Le calculs des effectifs théoriques se présentent de la même façon que dans le test précédent d'homogénéité bien que le problème soit différent.

Test d'indépendance

On aura besoin de présenter ces résultats sous forme d'un tableau :

Caractères A et B	A_1	A_2	...	A_i	...	A_I	Totaux
B_1	$O_{1,1}/E_{1,1}$	$O_{2,1}/E_{2,1}$	...	$O_{i,1}/E_{i,1}$	...	$O_{I,1}/E_{I,1}$	N_1
B_2	$O_{1,2}/E_{1,2}$	$O_{2,2}/E_{2,2}$	...	$O_{i,2}/E_{i,2}$	...	$O_{I,2}/E_{I,2}$	N_2
...	...	...	...	...	...	...	...
B_j	$O_{1,j}/E_{1,j}$	$O_{2,j}/E_{2,j}$	...	$O_{i,j}/E_{i,j}$	...	$O_{I,j}/E_{I,j}$	N_j
...	...	...	...	...	...	...	...
B_J	$O_{1,J}/E_{1,J}$	$O_{2,J}/E_{2,J}$	...	$O_{i,J}/E_{i,J}$	...	$O_{I,J}/E_{I,J}$	N_J
Totaux	T_1	T_2	...	T_i	...	T_I	$N = \sum_{i=1}^{I} T_i = \sum_{j=1}^{J} N_j$

On calcule les valeurs $E_{i,j} = \frac{T_i \times N_j}{N}$ pour remplir le tableau précédent.

Puis on déroule le processus de test ...

Test d'indépendance

1. On a alors deux hypothèses possibles :
   - **L'hypothèse nulle** (H_0) : les caractères étudiés sont indépendants.

   Elle consiste à dire que les différences entres les observations réalisées sur les échantillons et les effectifs théoriques ne sont pas assez importantes pour être significatives.

   - **L'hypothèse alternative** (H_1) : les deux caractères ne sont pas indépendants.

2. α reste quelconque dans un cadre général.

3. La **statistique de test** est :

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((I-1)(J-1)).$$

   <u>**Conditions d'application**</u> : il faut que tous les **effectifs soient supérieurs ou égaux à** 5 : $E_{i,j} \geq 5$. Sinon on regroupe certains événements ensembles et on refait un tableau.

4. Région critique : $W = [k_{1-\alpha}^{(I-1)(J-1)}; +\infty[$.

5. Calcul de la statistique.

6. Décision et conclusion :

   Si $\chi^2 \in W$, c'est-à-dire $\chi^2 \geq k_{1-\alpha}^{(I-1)(J-1)}$, on décide de rejeter (H_0) au risque α. Cela veut dire qu'avec α% de chance de se tromper, on peut affirmer que les caractères ne sont pas indépendants.

   Sinon, on ne rejette pas (H_0), cela revient à conclure (à notre niveau) que les deux caractères étudiés sont indépendants ou du moins que l'on a pas su prouver par ce test qu'ils ne l'étaient pas.

Test d'indépendance : exemple

On cherche à savoir si le type d'alimentation peut jouer sur le développement d'un cancer.

Pour cela, on considère le caractère A : type d'alimentation ayant pour modalités "rapide", "végétarienne", ou "traditionnelle".

On considère aussi le caractère B ayant pour modalités "présence d'un cancer", "absence d'un cancer".

On donne le tableau suivant :

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31	93	124
Végétarienne	12	37	49
Traditionnelle	81	243	324
Total	123	373	496

On calcule à présent les effectifs théoriques sous l'hypothèse d'indépendance : $E_{i,j}$.

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

1. (H_0) : les caractères étudiés sont indépendants, (H_1) : les caractères étudiés ne sont pas indépendants.

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

1. (H_0) : les caractères étudiés sont indépendants, (H_1) : les caractères étudiés ne sont pas indépendants.

2. On fixe $\alpha = 5\%$.

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

1. (H_0) : les caractères étudiés sont indépendants, (H_1) : les caractères étudiés ne sont pas indépendants.

2. On fixe $\alpha = 5\%$.

3. La **statistique de test** est :$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((2-1)(3-1))$.

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

1. (H_0) : les caractères étudiés sont indépendants, (H_1) : les caractères étudiés ne sont pas indépendants.
2. On fixe $\alpha = 5\%$.
3. La **statistique de test** est :$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((2-1)(3-1))$.
4. Région critique : $W = [k_{0.95}^{(2-1)(3-1)}; +\infty[= [k_{0.95}^{2}; +\infty[= [5.99; +\infty[$.

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

1. (H_0) : les caractères étudiés sont indépendants, (H_1) : les caractères étudiés ne sont pas indépendants.
2. On fixe $\alpha = 5\%$.
3. La **statistique de test** est :$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((2-1)(3-1))$.
4. Région critique : $W = [k_{0.95}^{(2-1)(3-1)}; +\infty[= [k_{0.95}^{2}; +\infty[= [5.99; +\infty[$.
5. Calcul de la statistique : $\chi^2 = \frac{(31-30.75)^2}{30.75} + ... + \frac{(243-243.65)^2}{243.65} \simeq 0.0121$.

Test d'indépendance : exemple (suite)

Alimentation / Cancer	Cancer	Pas de cancer	Total
Rapide	31 / 30.75	93 / 93.25	124
Végétarienne	12 / 12.15	37 / 36.85	49
Traditionnelle	81 / 80.35	243 / 243.65	324
Total	123	373	496

1. (H_0) : les caractères étudiés sont indépendants, (H_1) : les caractères étudiés ne sont pas indépendants.
2. On fixe $\alpha = 5\%$.
3. La **statistique de test** est :$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi^2((2-1)(3-1))$.
4. Région critique : $W = [k_{0.95}^{(2-1)(3-1)}; +\infty[= [k_{0.95}^{2}; +\infty[= [5.99; +\infty[$.
5. Calcul de la statistique : $\chi^2 = \frac{(31-30.75)^2}{30.75} + ... + \frac{(243-243.65)^2}{243.65} \simeq 0.0121$.
6. Décision et conclusion :

   Comme $\chi^2 \notin W$, on décide de ne pas rejeter (H_0), cela revient à conclure (à notre niveau) que les deux caractères étudiés sont indépendants.