*Université de Batna*

*Faculté de médecine*

# *Corrigé-type de TD10*

## *Test de KHI-2*

## *Rappel de cour :*

Les tests du Khi-deux sont basés sur la statistique de $\chi^2$ proposée par Karl Pearson. L'objectif de ces tests est principalement de comparer des distributions entre elles. Ces tests peuvent être appliqués à des variables de nature qualitative ou quantitative.

### *Trois types de test de khi-2 :*

***Test de conformité :***

l'objectif est de comparer une distribution observée sur un échantillon à une distribution théorique.

1. **Choix des hypothèses :**

   $H_0$ : La distribution de l'échantillon est conforme à celle de la population (échantillon représentatif de la population).

   $H_1$ : L'échantillon n'est pas représentatif de la population.

2. **Calcul des effectifs théoriques :**

   $$\boldsymbol{i = \overline{1,k} \qquad A_i = a_i N}$$

   **Condition de validité de test** $\boldsymbol{A_i \geq 5}$

3. **La statistique de test :**

   $$\boldsymbol{T_0 = \sum_{i=1}^{k} \frac{(O_i - A_i)^2}{A_i}}$$

4. **Décision :** Le seuil critique se lit dans la **table 4 (loi de Khi deux)**

   En ligne la valeur de $\alpha$ et en colonne $v = l - 1$

   Si $T_0 \leq \chi^2_{v;1-\alpha}$ On accepte $H_0$

   Si $T_0 > \chi^2_{v;1-\alpha}$ On rejette $H_0$

***Test d'homogénéité :***

l'objectif est de comparer deux ou plusieurs distributions observées sur des échantillons

1. **Choix des hypothèses :**

   $H_0$ : Les distributions observées sont identiques.

   $H_1$ : Les distributions observées ne sont pas identiques.

2. **Calcul des effectifs théoriques :**

$$i = \overline{1,k} \qquad j = \overline{1,l}$$

$$N_i = \sum_{j=1}^{l} O_{ij} \qquad N_j = \sum_{i=1}^{k} O_{ij} \qquad N = \sum_{i=1}^{k}\sum_{j=1}^{l} O_{ij}$$

$$a_j = \frac{N_j}{N} \qquad A_{ij} = N_i a_j = \frac{N_i N_j}{N}$$

**Condition de validité de test** $A_{ij} \geq 5$

3. **La statistique de test :**

$$T_0 = \sum_{j=1}^{l}\sum_{i=1}^{k} \frac{(O_{ij} - A_{ij})^2}{A_{ij}}$$

4. **Décision :** Le seuil critique se lit dans la **table 4 (loi de Khi deux)**

   En ligne la valeur de $\alpha$ et en colonne $v = (k-1)(l-1)$

   Si $T_0 \leq \chi^2_{v;1-\alpha}$ On accepte $H_0$

   Si $T_0 > \chi^2_{v;1-\alpha}$ On rejette $H_0$

***Test d'indépendance :*** Qui est utilisé pour étudier la liaison entre deux variables

1. **Choix des hypothèses :**

   $H_0$ : Les variables sont indépendantes.

   $H_1$ : Il existe une liaison entre les variables

2. **Calcul des effectifs théoriques :**

$$\mathrm{i} = \overline{1,k} \qquad \mathrm{j} = \overline{1,l}$$

$$n_{i.} = \sum_{j=1}^{l} O_{ij} \qquad n_{.j} = \sum_{i=1}^{k} O_{ij} \qquad N = \sum_{i=1}^{k}\sum_{j=1}^{l} O_{ij}$$

$$A_{ij} = \frac{n_{i.} n_{.j}}{N}$$

**Condition de validité de test** $A_{ij} \geq 5$

3. **La statistique de test :**

$$T_0 = \sum_{j=1}^{l}\sum_{i=1}^{k} \frac{(O_{ij} - A_{ij})^2}{A_{ij}}$$

4. **Décision :** Le seuil critique se lit dans la **table 4 (loi de Khi deux)**

   En ligne la valeur de $\alpha$ et en colonne $v = (k-1)(l-1)$

   Si $T_0 \leq \chi^2_{v;1-\alpha}$ On accepte $H_0$

   Si $T_0 > \chi^2_{v;1-\alpha}$ On rejette $H_0$

| Test de conformité Ajustement Adéquation | Test d'homogénéité | Test d'indépendance |
|---|---|---|
| ➢ Une population mère avec un échantillon.<br>➢ L'objectif : comparer ou ajuster une distribution observée sur un échantillon à une distribution théorique (Normale, poisson, binomiale ou quelconque).<br>➢ **Choix des hypothèses :**<br>$H_0$ : la distribution observée est conforme à la distribution théorique<br>$H_1$ : la distribution observée est différente de la distribution théorique | ➢ Une variable mesurée sur Deux échantillons ou plusieurs.<br>➢ L'objectif : comparer deux ou plusieurs distributions observées sur des échantillons.<br>➢ **Choix des hypothèses :**<br>$H_0$ : les distributions sont identiques entre elles.<br>$H_1$ : les distributions sont différentes entre elles | ➢ Deux variables étudiés sur un même échantillon.<br>➢ L'objectif : étudier sur un même échantillon la liaison entre deux variables<br>➢ **Choix des hypothèses :**<br>$H_0$ : les deux variables sont indépendantes<br>$H_1$ : il existe une liaison entre les deux variables. |

***<u>Exercice1 :</u>***

Dans un groupe de 200 malades atteints du cancer du col de l'utérus, un traitement par application locale du radium a donné 50 guérisons.

Un autre groupe de 150 sujets atteints de la même maladie a été traité par chirurgie, on a trouvé 54 guérisons.

Que peut-on dire sur l'équivalence des deux traitements ? On prendra un risque $\alpha = 5\%$.

***<u>Solution :</u>*** ***Test d'homogénéité :***

Une variable qualitative mesurée sur deux populations :

$$\begin{cases} P_1 malades\ atteints\ du\ cancer\ du\ col\ de\ l'utérus\ traités\ par\ application\ du\ radium \\ P_2 malades\ atteints\ du\ cancer\ du\ col\ de\ l'utérus\ traités\ par\ chirurgie \end{cases}$$

**La variable : «** la réussite d'un traitement »

On dispose de deux échantillons indépendants prélevés de $P_1$ $et$ $P_2$ de Tailles respectives $n_1 = 200 \ \ et \ n_2 = 150$.

Au total, On a tiré 350 malades.

**Tableau des effectifs observés :**

| Traitement\ Résultat | Guéri | Non guéri | Total |
|---|---|---|---|
| **Radium** | 50 | 150 | 200 |
| **Chirurgie** | 54 | 96 | 150 |
| **Total** | 104 | 246 | 350 |

1. **Choix des hypothèses :**
   $H_0$ : Les deux traitements sont équivalents
   $H_1$ : Les deux traitements ne sont pas équivalents
2. **Calcul des effectifs théoriques :**

$$\boldsymbol{k = 2 \quad i = \overline{1,2}}$$
$$\boldsymbol{l = 2 \quad j = \overline{1,2}}$$
$$\boldsymbol{N_i = \sum_{j=1}^{2} O_{ij} \qquad N_j = \sum_{i=1}^{2} O_{ij} \qquad N = \sum_{i=1}^{2}\sum_{j=1}^{2} O_{ij}}$$
$$\boldsymbol{a_j = \frac{N_j}{N} \qquad A_{ij} = N_i a_j = \frac{N_i N_j}{N}}$$

| Traitement \ Résultat | Guéri | | Non guéri | | Total $N_i$ |
|---|---|---|---|---|---|
| | $O_{ij}$ | $A_{ij}$ | $O_{ij}$ | $A_{ij}$ | |
| **Radium** | **50** | **59** | **150** | **141** | **200** |
| **Chirurgie** | **54** | **45** | **96** | **105** | **150** |
| **Total** $N_j$ | **104** | | **246** | | **350** |
| $a_j$ | **0.297** | | **0.703** | | **1** |

$$\boldsymbol{les\ A_{ij} \geq 5 \qquad test\ valide}$$

3. **La statistique de test :**

$$\boldsymbol{T_0 = \sum_{j=1}^{l}\sum_{i=1}^{k}\frac{(O_{ij}-A_{ij})^2}{A_{ij}} = 4.94}$$

4. **Le seuil critique :**

$$\boldsymbol{1-\alpha = 0.95 \qquad v = (k-1)(l-1) = 1 \qquad alors \qquad \chi^2_{1;0.95} = 3.841}$$

5. **Décision :** $\boldsymbol{T_0 > \chi^2_{1;0.95}}$

Décision statistique : On rejette$H_0$.

Décision pratique : Les deux traitements ne sont pas équivalents.

**2ème méthode :**

On applique un test d'homogénéité de comparaison de deux proportions :

Il s'agit d'un test bilatéral (égale conte inégale)

1) **Choix des hypothèses :**

$H_0$ : Les deux méthodes de traitement sont équivalentes

$H_1$ : Les deux méthodes de traitement ne sont pas équivalentes

$$H_0 : P_A = P_B$$
$$H_1 : P_A \neq P_B$$

2) **Calcul de la statistique de test observée :**

$$\widetilde{P_A} = {}^{50}/_{200} = 0.25 \qquad \widetilde{\boldsymbol{P}}_{\boldsymbol{B}} = {}^{\boldsymbol{54}}/_{\boldsymbol{150}} = \boldsymbol{0.36}$$

$$\boldsymbol{T_0} = \frac{\widetilde{\boldsymbol{P}}_{\boldsymbol{A}} - \widetilde{\boldsymbol{P}}_{\boldsymbol{B}}}{\sqrt{\widetilde{\boldsymbol{P}}(\boldsymbol{1} - \widetilde{\boldsymbol{P}})\left(\frac{\boldsymbol{1}}{\boldsymbol{n_A}} + \frac{\boldsymbol{1}}{\boldsymbol{n_B}}\right)}} = -\boldsymbol{2.23}$$

**Telle que la proportion commune** $\widetilde{\boldsymbol{P}} = \frac{\boldsymbol{X+Y}}{\boldsymbol{n_A+n_B}}$ **=0.297**

3) Identification de seuil critique **:** le test est **valable car les conditions sont bien remplies :**

$$\boldsymbol{n_A = 200 > 30 \qquad n_A\widetilde{P}_A = 50 > 5 \qquad n_A(1 - \widetilde{P}_A) = 150 > 5}$$

$$\boldsymbol{n_B = 150 > 30 \qquad n_B\widetilde{P}_B = 54 > 5 \qquad n_B(1 - \widetilde{P}_B) = 96 > 5}$$

Alors on cherche la valeur critique dans la table 2 (table de variable normale réduite)

$$\boldsymbol{\alpha = 0.05 \qquad \overline{Z}_\alpha = \overline{Z}_{0.05} = 1.96}$$

4) Décision :

$$\boldsymbol{T_0 < -\overline{Z}_\alpha}$$
$$\boldsymbol{-2.23 < -1.96}$$

Décision statistique : La statistique de test se trouve dans la zone de rejet de $H_0$ , donc on accepte $H_1$ .

Décision pratique : Les deux méthodes de traitements ne sont pas équivalentes

**Remarque :**

Pour l'exercice 1, on utilise deux méthodes de résolution, soit par un test de khi deux d'homogénéité, soit par un test d'hypothèse de comparaison de deux proportions car il s'agit d'un test bilatéral (équivalentes contre non équivalentes).

| **A RETENIR** |
|---|

Mme Medouer nawel

Pour les tests d'homogénéités bilatéraux de comparaison de deux proportions, un test de Khi-deux d'homogénéité est valable.

***Exercice2 :***

On veut savoir si la réussite (R) d'un traitement est indépendante du niveau de la tension artérielle du malade (T).

On dispose pour cela de 250 observations réparties comme suit :

| T \ R | échec | succès |
|---|---|---|
| basse | 21 | 104 |
| élevée | 29 | 96 |

La réussite de traitement dépend-elle du niveau de la tension artérielle ? On prendra un risque $\alpha = 5\%$.

***Solution :***

***Test d'indépendance :***

Deux variables qualitatives mesurées simultanément sur un échantillon de taille n=250.

Variable X : « Tension » à deux modalités « Basse, Elevée ».

Variable Y : « Réussite » à deux modalités « Echec=Non, Succès=Oui »

**Tableau des effectifs observés :**

| Tension \ Résultat | Echec | Succès | Total |
|---|---|---|---|
| Basse | 21 | 104 | 125 |
| Elevée | 29 | 96 | 125 |
| Total | 50 | 200 | 250 |

1. **Choix des hypothèses :**
   $H_0$ : La réussite de traitement est indépendante du niveau de la tension artérielle
   $H_1$ : La réussite de traitement est liée au niveau de la tension artérielle.

2. **Calcul des effectifs théoriques :**

$$\boldsymbol{k = 2 \quad i = \overline{1,2}}$$
$$\boldsymbol{l = 2 \quad j = \overline{1,2}}$$
$$\boldsymbol{n_{i.} = \sum_{j=1}^{2} O_{ij} \qquad n_{.j} = \sum_{i=1}^{2} O_{ij} \qquad N = \sum_{i=1}^{2}\sum_{j=1}^{2} O_{ij}}$$
$$\boldsymbol{A_{ij} = \frac{n_{i.}n_{.j}}{N}}$$

| Tension\ Résultat | Echec | | Succès | | Total $n_{i.}$ |
|---|---|---|---|---|---|
| | $\boldsymbol{O_{ij}}$ | $\boldsymbol{A_{ij}}$ | $\boldsymbol{O_{ij}}$ | $\boldsymbol{A_{ij}}$ | |
| Basse | 21 | 25 | 104 | 100 | 125 |
| Elevée | 29 | 25 | 96 | 100 | 125 |
| Total $n_{.j}$ | 50 | | 200 | | 250 |

$$\boldsymbol{les\ A_{ij} \geq 5 \qquad test\ valide}$$

3. **La statistique de test :**

$$T_0 = \sum_{j=1}^{l} \sum_{i=1}^{k} \frac{(O_{ij} - A_{ij})^2}{A_{ij}} = 1.6$$

4. **Le seuil critique :**

$$1 - \alpha = 0.95 \qquad v = (k-1)(l-1) = 1 \qquad alors \qquad \chi^2_{1;0.95} = 3.841$$

5. **Décision :** $T_0 \leq \chi^2_{1;0.95}$

Décision statistique : On accepte $H_0$.

Décision pratique : La réussite de traitement est indépendante du niveau de la tension artérielle.

**Exercice 3**

De nombreuses observations cliniques ont montré que -là :

- 30% des malades atteints de M ont une survie inférieure à un an
- 50% ont une survie entre un an et deux ans.
- 10% ont une survie entre deux ans et cinq ans.
- 10% ont une survie supérieure à cinq ans.

On applique un nouveau traitement à 80 malades atteints de la maladie M et on constate :

- 12 ont une survie inférieure à un an
- 56 ont une survie entre un an et deux ans.
- 8 ont une survie entre deux ans et cinq ans.
- 4 ont une survie supérieure à cinq ans.

  Que peut-on conclure ? On prendra un risque $\alpha = 5\%$.

***Solution :***

***Test de conformité :***

1. **Choix des hypothèses :**

   $H_0$ : La distribution de l'échantillon est conforme à celle de la population (échantillon représentatif de la population).

   $H_1$ : L'échantillon n'est pas représentatif de la population.

2. **Calcul des effectifs théoriques :**

| Survie | Effectifs observées $O_i$ | Probabilités théoriques $a_i$ | Effectifs Théoriques $A_i$=$Na_i$ |
|---|---|---|---|
| $survie \leq 1$ans | 12 | 0.3 | 24 |
| 1 an$<$ $survie \leq 2\ ans$ | 56 | 0.5 | 40 |
| 2 ans$<$ $survie \leq 5\ ans$ | 8 | 0.1 | 8 |
| $survie > 5ans$ | 4 | 0.1 | 8 |
| Total N | 80 | 1 | 80 |

$$les\ A_i \geq 5 \qquad test\ valide$$

3. **La statistique de test :**

$$T_0 = \sum_{i=1}^{k} \frac{(O_i - A_i)^2}{A_i} = 14.4$$

4. **Le seuil critique :**

$$1 - \alpha = 0.95 \qquad v = (k-1) = 3 \qquad alors \qquad \chi^2_{3;0.95} = 7.815$$

5. **Décision :** $T_0 > \chi^2_{3;0.95}$

Décision statistique : On rejette$H_0$.

Décision pratique : L'échantillon n'est pas représentatif de la population.

## Exercice 4 :

On désire tester l'effet d'une antibiothérapie systématique sur l'apparition d'une infection post-opératoire. Une expérience randomisée est conduite. Un premier groupe de patients reçoit une antibiothérapie. Un deuxième groupe reçoit un placebo. Les résultats sont les suivants :

| | Antibiothérapie | placebo |
|---|---|---|
| infection | 10 | 29 |
| Pas d'infection | 75 | 27 |

L'antibiothérapie est-elle efficace dans la prévention des applications infectieuses ? On prendra un risque$\alpha = 5\%$.

**Solution :**

***Test d'indépendance :***

Deux variables qualitatives mesurées simultanément sur un échantillon de taille n=141.

Variable X : « Produit consommé » à deux modalités « Antibiothérapie, Placebo ».

Variable Y : « Infection » à deux modalités « Infection, Pas d'infection »

**On premier lieu**, on applique un test d'indépendance de khi-deux pour discuter **la liaison** entre **l'antibiothérapie** et **l'apparition des infections post-opératoires.**

**Tableau des effectifs observés :**

| | Antibiothérapie | Placebo | Total |
|---|---|---|---|
| Infection | 10 | 29 | 39 |
| Pas d'infection | 75 | 27 | 102 |
| Total | 85 | 56 | 141 |

1. **Choix des hypothèses :**
   $H_0$ : L'antibiothérapie n'influe pas sur l'apparition d'une infection post-opératoire. (Elles sont indépendantes)
   $H_1$ : L'antibiothérapie influe sur l'apparition d'une infection post-opératoire (liées).

2. **Calcul des effectifs théoriques :**

$$\boldsymbol{k = 2 \quad i = \overline{1,2}}$$
$$\boldsymbol{l = 2 \quad j = \overline{1,2}}$$

$$\boldsymbol{n_{i.} = \sum_{j=1}^{2} O_{ij} \qquad n_{.j} = \sum_{i=1}^{2} O_{ij} \qquad N = \sum_{i=1}^{2}\sum_{j=1}^{2} O_{ij} \quad A_{ij} = \frac{n_{i.}n_{.j}}{N}}$$

| | Antibiothérapie | | Placébo | | Total $n_{i.}$ |
|---|---|---|---|---|---|
| | $\boldsymbol{O_{ij}}$ | $\boldsymbol{A_{ij}}$ | $\boldsymbol{O_{ij}}$ | $\boldsymbol{A_{ij}}$ | |
| Infection | 10 | 23.5 | 29 | 15.5 | 39 |
| Pas d'infection | 75 | 61.5 | 27 | 40.5 | 102 |
| Total $n_{.j}$ | 85 | | 56 | | 141 |

**Condition de validité de test :**

$$\boldsymbol{A_{ij} \geq 5}$$

3. **La statistique de test :**

$$\boldsymbol{T_0 = \sum_{j=1}^{l}\sum_{i=1}^{k}\frac{(O_{ij}-A_{ij})^2}{A_{ij}}} = \mathbf{26.97}$$

4. **Le seuil critique :**

$$\boldsymbol{1-\alpha = 0.95 \qquad v = (k-1)(l-1) = 1 \qquad alors \qquad \chi^2_{1;0.95} = 3.841}$$

5. **Décision :** $\boldsymbol{T_0 > \chi^2_{1;0.95}}$

Décision statistique : On rejette$H_0$.

Décision pratique : L'antibiothérapie influe sur l'apparition d'une infection post-opératoire.

(Efficace dans la prévention des applications infectieuses).

**<u>En 2ème lieu :</u>**

Il s'agit d'un test unilatéral à droite (égale contre supérieure)

1) **Choix des hypothèses :**
   $H_0$ : L'antibiothérapie et le placebo sont équivalents dans la prévention des applications infectieuses
   $H_1$ : L'antibiothérapie est plus efficace dans la prévention des applications infectieuses.

$$H_0 : P_A = P_B$$
$$H_1 : P_A > P_B$$

2) **Calcul de la statistique de test :**
   Proportion des non infectés en utilisant l'antibiothérapie :
   $$\widetilde{P_A} = {}^{75}/_{85} = 0.88$$
   Proportion des non infectés en utilisant le placebo :
   $$\widetilde{P_B} = {}^{\mathbf{27}}/_{\mathbf{56}} = \mathbf{0.48}$$

$$\boldsymbol{T_0} = \frac{\widetilde{\boldsymbol{P}}_A - \widetilde{\boldsymbol{P}}_B}{\sqrt{\widetilde{\boldsymbol{P}}(\mathbf{1} - \widetilde{\boldsymbol{P}})\left(\frac{\mathbf{1}}{\boldsymbol{n_A}} + \frac{\mathbf{1}}{\boldsymbol{n_B}}\right)}} = \mathbf{5.198}$$

**Telle que la proportion commune** $\widetilde{\boldsymbol{P}} = \frac{\boldsymbol{X+Y}}{\boldsymbol{n_A+n_B}} = \mathbf{0,72}$

3) Identification de seuil critique **:** le test est **valable car les conditions sont bien remplies :**

$$\boldsymbol{n_A = 85 > 30} \qquad \boldsymbol{n_A\widetilde{P}_A = 75 > 5} \qquad \boldsymbol{n_A(1 - \widetilde{P_A}) = 10 > 5}$$

$$\boldsymbol{n_B = 56 > 30} \qquad \boldsymbol{n_B\widetilde{P}_B = 27 > 5} \qquad \boldsymbol{n_B(1 - \widetilde{P_B}) = 29 > 5}$$

   Alors on cherche la valeur critique dans la table Normale
   $$\boldsymbol{\alpha = 0.05} \qquad \boldsymbol{\overline{Z}_{2\alpha} = \overline{Z}_{0.10} = 1.645}$$

4) Décision :

$$\boldsymbol{T_0 > \overline{Z}_{2\alpha}}$$
$$\mathbf{5.198 > 1.645}$$

Décision statistique : La statistique de test se trouve dans la zone de rejet de $H_0$ , donc on accepte $H_1$ .

La décision pratique : On conclure que l'antibiothérapie est plus efficace que le placebo dans la prévention des applications infectieuses.

**Remarque :** La solution est faite en point de vue mathématique, deux distributions « antibiothérapie » et « placebo », donc la deuxième étape est obligatoire.

Mais en point de vue médicale le placebo est sans principe actif donc on prévoit la décision sans faire un test d'homogénéité entre l'antibiothérapie et le placebo.

**A RETENIR**
Pour les tests unilatéraux, un test d'homogénéité de khi-deux pour la comparaison entre deux proportions observées est **non valide**.

***Exercice 5 :***

On souhaite évaluer l'effet éventuel de différentes psychothérapies sur la phobie social. On sait que dans une population particulière de patients atteints de la phobie social, on a la distribution suivante (les patients sont classés en trois niveaux de phobie sociale, le niveau 1 étant le plus faible, le niveau 3 le plus marqué :

| Niveau de la phobie | Niveau1 | Niveau 2 | Niveau 3 |
|---|---|---|---|
| proportion | 30% | 40% | 30% |

200 patients tirés au sort dans cette population ont suivi une thérapie cognitive-comportementales.

Les résultats post-traitement sont les suivants :

| Niveau de la phobie | Niveau1 | Niveau 2 | Niveau 3 |
|---|---|---|---|
| effectif | 120 | 50 | 30 |

Peut-on dire que la thérapie a un effet ? On prendra un risque$\alpha = 5\%$.

***Solution :***

***Test de conformité :***

1. **Choix des hypothèses :**
   $H_0$ : La distribution des niveaux de la phobie de l'échantillon est conforme à celle de la population (c'est-à-dire la thérapie cognitive-comportementales n'a pas d'effet sur la phobie).
   $H_1$ : La distribution des niveaux de la phobie de l'échantillon n'est pas conforme à celle de la population (c'est-à-dire la thérapie cognitive-comportementales a un effet sur la phobie).

2. **Calcul des effectifs théoriques :**

| Niveau de la phobie | Effectifs observées $O_i$ | Probabilités théoriques $a_i$ | Effectifs Théoriques $A_i$=$Na_i$ |
|---|---|---|---|
| **Niveau 1** | **120** | **0.3** | **60** |
| **Niveau 2** | **50** | **0.4** | **80** |
| **Niveau 3** | **30** | **0.3** | **60** |
| **Total N** | **200** | **1** | **200** |

$$\boldsymbol{les\ A_i \geq 5} \qquad \textbf{test valide}$$

3. **La statistique de test :**

$$T_0 = \sum_{i=1}^{k} \frac{(O_i - A_i)^2}{A_i} = 86.25$$

4. **Le seuil critique :**

$$1 - \alpha = 0.95 \qquad v = (k-1) = 2 \qquad alors \qquad \chi^2_{2;0.95} = 5.991$$

5. **Décision :** $T_0 > \chi^2_{2;0.95}$

Décision statistique : On rejette$H_0$.

Décision pratique : La distribution des niveaux de la phobie de l'échantillon n'est pas conforme à celle de la population (c'est-à-dire la thérapie cognitive-comportementales a un effet sur la phobie).

***Exercice6 :***

Dans une université ou les initiatives pédagogiques différenciées sont vivement encouragées, trois groupes de professeurs ont mis aux point trois méthodes différentes d'apprentissage de bio-statistique qu'on a appliqué à trois échantillons d'étudiants ayant le même niveau initial. A l'examen les résultats furent les suivants :

| Observée | Admis | Ajournés |
|---|---|---|
| Méthode 1 | 51 | 29 |
| Méthode 2 | 38 | 12 |
| Méthode 3 | 86 | 31 |

Peut-on affirme que l'une des trois méthodes est efficace que les autres en terme de réussite à l'examen ? On prendra un risque $\alpha = 5\%$.

***Solution :***

***Test d'homogénéité :***

Une variable qualitative mesurée sur trois populations :

$$\begin{cases} P_1 : Etudiants\ apprentés\ par\ la\ première\ méthode \\ P_2 : Etudiants\ apprentés\ par\ la\ deuxième\ méthode \\ P_3 : Etudiants\ apprentés\ par\ la\ troisième\ méthode \end{cases}$$

**La variable :** « Résultat de l'examen »

On dispose de trois échantillons indépendants prélevés de $P_1, P_2 \ et\ P_3$ de Tailles respectives $n_1 = 80,\ n_2 = 50\ et\ n_3 = 117.$

Au total, On a tiré 247 étudiants.

| **Méthode \Résultat** | **Admis** | **Ajournés** | Total |
|---|---|---|---|
| **Méthode 1** | 51 | 29 | 80 |
| **Méthode 2** | 38 | 12 | 50 |
| **Méthode 3** | 86 | 31 | 117 |
| **Total** | 175 | 72 | 247 |

1. **Choix des hypothèses :**
   $H_0$ : Les trois méthodes sont équivalentes.
   $H_1$ : L'une des trois méthodes est efficace que les autres.
2. **Calcul des effectifs théoriques :**

$$\boldsymbol{k = 3 \quad i = \overline{1,3}}$$
$$\boldsymbol{l = 2 \quad j = \overline{1,2}}$$
$$\boldsymbol{N_i = \sum_{j=1}^{2} O_{ij} \qquad N_j = \sum_{i=1}^{3} O_{ij} \qquad N = \sum_{i=1}^{3}\sum_{j=1}^{2} O_{ij}}$$
$$\boldsymbol{a_j = \frac{N_j}{N} \qquad A_{ij} = N_i a_j = \frac{N_i N_j}{N}}$$

| **Méthode \Résultat** | **Admis** | | **Ajournés** | | Total $\boldsymbol{N_i}$ |
|---|---|---|---|---|---|
| | $\boldsymbol{O_{ij}}$ | $\boldsymbol{A_{ij}}$ | $\boldsymbol{O_{ij}}$ | $\boldsymbol{A_{ij}}$ | |
| **Méthode 1** | 51 | 57 | 29 | 23 | 80 |
| **Méthode 2** | 38 | 35 | 12 | 15 | 50 |
| **Méthode 3** | 86 | 83 | 31 | 34 | 117 |
| **Total** $\boldsymbol{N_j}$ | 175 | | 72 | | 247 |
| $\boldsymbol{a_j}$ | 0.71 | | 0.29 | | 1 |

$\boldsymbol{les\ A_{ij} \geq 5}$ **test valide**

3. **La statistique de test :**

$$\boldsymbol{T_0 = \sum_{j=1}^{l}\sum_{i=1}^{k}\frac{(O_{ij}-A_{ij})^2}{A_{ij}}} = \mathbf{3.27}$$

4. **Le seuil critique :**

$$\boldsymbol{1-\alpha = 0.95 \qquad v = (k-1)(l-1) = 2 \qquad alors \qquad \chi^2_{2;0.95} = 5.991}$$

5. **Décision :** $\boldsymbol{T_0 \leq \chi^2_{2;0.95}}$

Décision statistique : On accepte $H_0$.

Décision pratique : Les trois méthodes d'apprentissages peuvent être considérées comme équivalentes, c'est-à-dire malgré les différences résultats, les méthodes se valent.

***Remarque :***

La condition d'application ou de validité du test de Khi-deux c'est que les effectifs théoriques doivent être supérieurs ou égale à 5. Si ce n'est pas le cas, on procède à un regroupement de modalités.

**Exercice supplémentaire :**

**Exercice1 :**

Le responsable des stocks d'un laboratoire de produits pharmaceutiques souhaite savoir combien de doses de vaccin il doit tenir en stock.

1) Il relève donc les ventes de ce vaccin sur les 100 derniers jours, supposés représentatifs, à savoir :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Nombre de doses vendues | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Nombre de jours | 14 | 27 | 26 | 18 | 9 | 4 | 2 |

Peut-on dire que les ventes de vaccin sont distribuées selon une loi de Poisson ?

**Solution :**

***Test de conformité (Ajustement par la loi de Poisson) :***

1. **Choix des hypothèses :**
   $H_0$ : La distribution observée suit la loi de Poisson.
   (Echantillon représentatif de la population).
   $H_1$ : La distribution observée ne suit pas la loi de Poisson.
   (L'échantillon n'est pas représentatif de la population)

**Détermination de la distribution théorique :**

La variable X étudiée ici est le nombre de doses de vaccins vendus par jour.

Cette variable peut prendre les valeurs 0, 1, 2, 3, 4, 5,6,…

On étudie la distribution de cette variable sur un échantillon de 100 jours.

$$X \sim P(\lambda)$$

Tel que $\lambda = \frac{0\times14+1\times27+2\times26+3\times18+4\times9+5\times4+6\times2}{100} = 2.01$ ($\lambda$ c'est la moyenne)

$$P(X = i) = \frac{e^{-2.01}2.01^i}{i!}$$

2. **Calcul des effectifs théoriques :**

| Nombre de doses vendues | Effectifs observées $O_i$ | Probabilités théoriques $a_i$ | Effectifs Théoriques $A_i$=$Na_i$ |
|---|---|---|---|
| 0 | 14 | $P(X=0)=e^{-2.01}$ | 13.4 |
| 1 | 27 | $P(X=1)=\frac{e^{-2.01}2.01}{1!}$ | 26.93 |
| 2 | 26 | $P(X=2)=\frac{e^{-2.01}2.01^2}{2!}$ | 27.07 |
| 3 | 18 | $P(X=3)=\frac{e^{-2.01}2.01^3}{3!}$ | 18.13 |
| 4 | 9 | $P(X=4)=\frac{e^{-2.01}2.01^4}{4!}$ | 9.11 |
| 5 | 4 | $P(X=5)=\frac{e^{-2.01}2.01^5}{5!}$ | 3.66 |
| 6 | 2 | $P(X=6)=\frac{e^{-2.01}2.01^6}{6!}$ | 1.70 |
| Total | 100 | 1 | 100 |

La condition sur les effectifs théoriques (**$A_i \geq 5$**) n'étant pas respectée pour les deux dernières classes, un regroupement des deux dernières classes est obligatoire. On obtient donc le tableau suivant :

| | X=0 | X=1 | X=2 | X=3 | X=4 | X≥ 5 | T |
|---|---|---|---|---|---|---|---|
| Observées | 14 | 27 | 26 | 19 | 9 | 6 | 100 |
| Théoriques | 13.4 | 26.93 | 27.27 | 18.13 | 9.11 | 5.36 | 100 |

$$\boldsymbol{les\ A_i \geq 5 \qquad test\ valide}$$

3. **La statistique de test :**

$$\boldsymbol{T_0 = \sum_{i=1}^{l} \frac{(O_i - A_i)^2}{A_i} = 0.15}$$

4. **Le seuil critique :**

$$\boldsymbol{1-\alpha = 0.95 \qquad v = (k-1) = 5 \qquad alors \qquad \chi^2_{5;0.95} = 11.07}$$

*5.* **Décision :** $\boldsymbol{T_0 < \chi^2_{5;0.95}}$

Décision statistique : On accepte$H_0$.

Décision pratique : L'échantillon est représentatif de la population, c'est-à-dire les ventes suivent une loi de Poisson.