UNIVERSITE DE BATNA_2 **FACULTE DE MEDECINE**

SERIE DE TD N° 10 *en BIOSTATISTIQUE 2022/2023*

Exercices sur les Tests du Khi-Deux

Exercice 1 :

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 α = 5%.

Solution de l'exercice 1 :

Données :

On a deux populations de même maladie et traitées par deux méthodes :

X est une variable catégorielle avec 2 modalités c'est-à-dire les 2 méthodes de traitement (x_1 c'est le radium et x_2 c'est la chirurgie).

Y est une variable catégorielle avec 2 modalités (c'est-à-dire la guérison ou non ; y_1 = guéris et y_2 = non guéris).

La fréquence de guérison dans le 1^{er} groupe est de 25 % et dans le $2^{ème}$ groupe elle est de 36 %.

Alors c'est un problème de comparaison de 2 proportions. Il suffit donc de démontrer l'homogénéité de ces 2 proportions ou leur non homogénéité.

1°) Par la Méthode du Khi-Deux : **Test du Khi-Deux d'homogénéité**

- *Formulation des hypothèses* :

  H₀ : « Les deux traitements sont équivalents » ; **contre**
  H₁ : « Les deux traitements ne sont pas équivalents ».

- *Calcul de la statistique de test* : To = $\chi^2_{(observé)}$
  Les effectifs observés ($E\ O_i$) et attendus ou théoriques ($E\ A_i$ ou $E\ t_i$) sont reportés dans le tableau suivant :

X \ Y	Guéris		Non guéris		Total
Radium	50	59.4	150	140.6	200
Chirurgie	54	44.6	96	105.4	150
Total	104		246		350

Comme tous les effectifs sont supérieurs à 5, alors le test sera validé.

$$\chi^2_{(observé)} = \sum_{i=1}^{4} \frac{(E\ O_i - E\ t_i)^2}{E\ t_i} = \frac{(50-59.4)^2}{59.4} + \frac{(54-44.6)^2}{44.6} + \frac{(150-140.6)^2}{140.6} + \frac{(96-105.4)^2}{105.4} = 4.93548768$$

- *Le seuil critique* : Le tableau de contingence a 2 lignes et 2 colonnes donc le nombre de degré de liberté = (2–1)*(2–1) = 1 ⇒ $\chi^2_{(1)}$ **= 3.84** ;

  Si la statistique de test observé To = $\chi^2_{(observé)} > \chi^2_{(1)}$ alors on ne peut pas accepter H_0.

- *Décision* : comme on a $\chi^2_{(observé)}$ **= 4.94** = To > $\chi^2_{(1)}$ **= 3.84** alors on ne peut pas **Conserver H₀.**

2°) Par la Méthode du test d'homogénéité de 2 proportions :

Traitement de cet exercice avec le test d'homogénéité pour comparaison deux proportions.

- *Formulation des hypothèses :* On effectue un test bilatéral car il y a une équivalence.

  H_0 : {Les deux traitements au radium et à la chirurgie sont équivalents} **contre**

  H_1 : {Les deux traitements au radium et à la chirurgie ne sont pas équivalents}.

- *Le seuil critique :* α = 5 % = P (rejet de $H_0 | H_0$ vraie), comme les effectifs n > 30 (toutes les conditions nécessaires sont remplies) alors on utilise la table de la loi normale n° 2 et on a u_α = 1.96 et si la statistique de test observé $\in IA(H_0) = [-1.96 ; +1.96]$ on conserve H_0.

- *Statistique de test observée :* To = $\frac{f_1 - f_2}{\sigma_c}$ sachant que f_1 = 50/200 = 0.25 et que f_2 = 54/150 = 0.36 et la fréquence commune $f_0 = \frac{50+54}{200+150} = \frac{104}{350}$ = 0.297142857 d'où l''écart type commun : $\sigma_c = \sqrt{f_0(1-f_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{\frac{104}{350} * \frac{246}{350}\left(\frac{1}{200}+\frac{1}{150}\right)}$ = 0.049361639

  Donc To = $\frac{0.25-0.36}{0.049361639}$ = – 2.228451115 ≈ – 2.23

- *Décision :* La statistique de test observé To n'appartient pas à la zone d'acceptation de H_0 {$IA(H_0)$}, on ne peut pas conserver H_0. Donc les deux traitements ne sont pas équivalents.

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

Exercice 2 :

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
Elevée	29	96

La réussite du traitement dépend-elle du niveau de la tension artérielle ?

On prendra un risque α = 5 %.

Solution de l'exercice 2 : **Test du Khi-Deux 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 »

- *Formulation des hypothèses* :

  H_0 : « la réussite est indépendante de la tension artérielle » **contre**

  H_1 : « la réussite est dépendante de la tension artérielle ».

- *Calcul de la statistique de test* : To = $\chi^2_{(observé)}$

  $$\chi^2_{(observé)} = \sum_{i=1}^{4} \frac{(E\,O_i - E\,t_i)^2}{E\,t_i} = \frac{(21-25)^2}{25} + \frac{(29-25)^2}{25} + \frac{(104-100)^2}{100} + \frac{(96-100)^2}{100} = 1.6 = \text{To}.$$

- *Le seuil critique* : Le tableau de contingence a 2 lignes et 2 colonnes donc le nombre de degré de liberté = 1 ⇒ $\chi^2_{(1)}$ **= 3.84** ; si To > $\chi^2_{(1)}$ alors on ne peut pas accepter H_0.

T \ R	y_1 = Echec		y_2 = Succès		Total
x_1 = Basse	21	25	104	100	125
x_2 = Elevée	29	25	96	100	125
Total	50		200		250

- *Décision* : comme on a $\chi^2_{(observé)}$ **= 1.6** = To < $\chi^2_{(1)}$ **= 3.84** alors on **Conserve H_0** au risque de 5 %.

Exercice 3 :

De nombreuses observations cliniques ont montré que jusque-là :

- 30% des malades atteints de la maladie 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.

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

Solution de l'exercice 3 : **Test du Khi-Deux d'adéquation ou de conformité**

- *Formulation des hypothèses* :

  H_0 : La distribution de l'échantillon des 80 personnes est conforme à celle de la population (C'est-à-dire cet échantillon est représentatif de la population) ; **contre**

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

- *Calcul de la statistique de test observé* To = $\chi^2_{(observé)}$ *après avoir calculé les effectifs attendus ou théoriques* :

x_i	< 1 an	Entre 1 et 2	Entre 2 et 5	> 5 ans	Total
Proportion %	30	50	10	10	100
Effectifs observés	12	56	8	4	80
Effectifs théoriques	24	40	8	8	80
$\chi^2_{(\text{observé})}$	6	6.4	0	2	14.4

$$\chi^2_{(observé)} = \sum_{i=1}^{4} \frac{(E\,O_i - E\,t_i)^2}{E\,t_i} = \frac{(12-24)^2}{24} + \frac{(56-40)^2}{40} + \frac{(8-8)^2}{8} + \frac{(4-8)^2}{8} = 14.4$$

- *Le seuil critique* : Le nombre de degré de liberté = n – 1 = 3 ⇒ $\chi^2_{(3)}$ **= 7.81** ;

  Si To > $\chi^2_{(3)}$ alors on ne peut pas accepter H_0.

- *Décision* : comme on a $\chi^2_{(observé)}$ **= 14.4** = To > $\chi^2_{(3)}$ **= 7.81** alors on ne peut pas **Conserver H_0.** Donc 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 :

X \ Y	Antibiothérapie	Placébo
Infection	10	29
Pas d'infection	75	27

L'antibiothérapie est-elle efficace dans la prévention des complications infectieuses ?

On prendra un risque α = 5 %.

Solution de l'exercice 4 :

Données : On a 2 variables catégorielles X (présence ou non d'une infection) et Y (méthode de traitement par antibiothérapie et méthode de traitement par placébo) de 2 modalités chacune.

Le groupe ayant reçu l'antibiothérapie a une fréquence de non infection de 75/85 ≈ 88 % ;
Alors que le groupe ayant reçu le placébo a une fréquence de non infection de 27/56 ≈ 48 %.
Les modalités de la première variable X sont {Infection et non-infection}
Les modalités de la deuxième variable Y sont {Antibiothérapie et Placébo}
Par conséquent le problème est de tester s'il y a indépendance de ces 2 variables X et Y ou non.

I) Test du Khi-Deux d'indépendance

- *Formulation des hypothèses* :

  H_0 : « Les deux variables (X et Y) sont indépendantes » ; **contre**

  H_1 : « Les deux variables sont dépendantes ».

- *Calcul de la statistique de test observé* : To = $\chi^2_{(observé)}$

$$\chi^2_{(observé)} = \sum_{i=1}^{4} \frac{(E\,O_i - E\,t_i)^2}{E\,t_i} = \frac{(10-23.5)^2}{23.5} + \frac{(75-61.5)^2}{61.5} + \frac{(29-15.5)^2}{15.5} + \frac{(27-40.5)^2}{40.5} = 26.9768$$

X \ Y	Antibiothérapie		Placébo		Total
Infection	10	23.5	29	15.5	39
Non infection	75	61.5	27	40.5	102
Total	85		56		141

- *Le seuil critique* : Le tableau de contingence a 2 lignes et 2 colonnes d'où le nombre de degré de liberté = 1 ⇒ $\chi^2_{(1)}$ **= 3.84** ; si To > $\chi^2_{(1)}$ alors on ne peut pas accepter H_0.

- *Décision* : comme on a $\chi^2_{(observé)}$ **≈ 26.98** = To > $\chi^2_{(1)}$ **= 3.84** alors on ne peut pas **Conserver H_0.** Par conséquent c'est H_1 qu'il faut conserver (Liaison des variables).

II) Pour démontrer l'efficacité de l'antibiothérapie, il faut faire un test unilatéral droit de comparaison de deux fréquences f_1 et f_2 (égal contre supérieur), car la question posée est : L'antibiothérapie est-elle efficace dans la prévention des complications infectieuses ?

La fréquence des non infectés en utilisant l'antibiothérapie est : $f_1 = \frac{75}{85} = 0.882352941$

La fréquence des non infectés en utilisant le placébo est : $f_2 = \frac{27}{56} = 0.482142857$

La fréquence commune est $f_0 = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \frac{75+27}{85+56} = \frac{102}{141} = 0.723404255$

Formulation des hypothèses :

H_0 : $\{p_1 - p_2 = 0\}$ ou $\{p_1 = p_2\}$; contre **H_1** : $\{p_1 - p_2 > 0\}$ ou $\{p_1 > p_2\}$ ⇒ (= contre >).

Le seuil critique : $u_{0.05} = +1.645$ ⇒ si To > Seuil Critique = $u_{0.05} = +1.645$ ⇒ on ne peut pas accepter H_0.

Calcul de la statistique de test observé :

$$To = \frac{f_1 - f_2}{\sigma_c} = \frac{(f_1 - f_2)}{\sqrt{f_0\,(1-f_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{(0.88235 - 0.48214)}{\sqrt{0.7234(1-0.7234)(85^{-1}+56^{-1})}} = 5.198387748 \approx 5.2$$

Décision / comme To = 5.2 > $u_{0.05}$ = 1.645. On ne peut pas accepter H_0 : les observations disponibles montrent une différence significative entre les proportions p_1 et p_2 (c'est-à-dire on a effectivement $p_1 > p_2$) au risque de 5 %.

III) Conclusion finale : L'antibiothérapie est la plus efficace dans ce traitement.

Remarque : La solution est faite du point de vue mathématique, deux distributions « antibiothérapie » et « placebo », donc la deuxième étape est obligatoire (comme il y a dépendance des deux variables, on voudrait savoir quel est le traitement le plus efficace, et en médecine ce ne sera pas le placébo).

Mais du point de vue de la médecine le placebo est sans principe actif et 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 sociale. On sait que dans une population particulière de patients atteints de la phobie sociale, on a la distribution suivante (les patients sont classés en trois niveaux de phobie sociale, le niveau 1 étant le plus faible et le niveau 3 le plus marqué).

Niveau de la phobie	Niveau-1	Niveau-2	Niveau-3
Proportion	30%	40%	30%

200 patients tirés au sort dans cette population ont suivi une thérapie cognitive-comportementale. Les résultats post-traitement sont les suivants :

Niveau de la phobie	Niveau-1	Niveau-2	Niveau-3
effectif	120	50	30

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

Solution de l'exercice 5 : Test du Khi-Deux d'adéquation ou de conformité

- *Formulation des hypothèses* :

***H*₀** : « La distribution des niveaux de la phobie sociale de l'échantillon des 200 personnes est conforme à la distribution théorique de la population » ; **contre**

***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 il y a un effet sur la phobie sociale) ».

- *Calcul de la statistique de test* : To = $\chi^2_{(observé)}$

$$\chi^2_{(observé)} = \sum_{i=1}^{4} \frac{(E\,O_i - E\,t_i)^2}{E\,t_i} = \frac{(120-60)^2}{60} + \frac{(50-80)^2}{80} + \frac{(30-60)^2}{60} = 86.25$$

Phobie	Niveau-1	Niveau-2	Niveau-3	Total
Proportion %	30	40	30	100
Effectif observé	120	50	30	200
Effectif théorique	60	80	60	200

- *Le seuil critique* : Le nombre de degré de liberté = n – 1 = 2 ⇒ $\chi^2_{(2)}$ **= 5.99** ; si To > $\chi^2_{(2)}$ alors on ne peut pas accepter H_0.
- *Décision* : comme on a $\chi^2_{(observé)}$ **= 86.25** = To > $\chi^2_{(2)}$ **= 5.99** alors on ne peut pas **Conserver H_0.**

Exercice 6 :

Dans une université où les initiatives pédagogiques différenciées sont vivement encouragées, trois groupes de professeurs ont mis au 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 :

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

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

Solution de l'exercice 6 **: Test du Khi-Deux d'homogénéité**

- *Formulation des hypothèses* :

  H0 : Les trois méthodes sont équivalentes ; **contre**

  H1 : L'une au moins des trois méthodes est plus efficace.

- *Calcul de la statistique de test* : To = $\chi^2_{(observé)}$

$$\sum_{i=1}^{6}\frac{(E\,O_i-E\,t_i)^2}{E\,t_i}=\frac{(51-56.7)^2}{56.7}+\frac{(38-35.4)^2}{35.4}+\frac{(86-82.9)^2}{82.9}+\frac{(29-23.3)^2}{23.3}+\frac{(12-14.6)^2}{14.6}+\frac{(31-34.1)^2}{34.1}=3.019$$

	Admis		ajournés		Total
Méthode-1	51	56.7	29	23.3	80
Méthode-2	38	35.4	12	14.6	50
Méthode-3	86	82.9	31	34.1	117
total	175		72		247

- *Le seuil critique* : Le tableau de contingence a 3 lignes et 2 colonnes donc le nombre de degré de liberté = 2 ∗ 1 = 2 ⇒ $\chi^2_{(2)}$ **= 5.99** ; si To > $\chi^2_{(2)}$ alors on ne peut pas accepter H_0.

- *Décision* : Comme on a $\chi^2_{(observé)}$ **= 3.019** = To < $\chi^2_{(2)}$ **= 5.99** alors on ne peut pas **Rejeter H0** car il n'y a pas de différence significative au risque de 5 %. Donc on peut dire pratiquement les trois méthodes sont équivalentes.

Exercice supplémentaire

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

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)

- **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 aléatoire 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 ; la variable X ~ P (λ) avec P (X = k) = $e^{\lambda}\,\frac{\lambda^k}{k!}$

Nombre de doses vendues x_i	0	1	2	3	4	5	6	Σ
Nombre de jours n_i	14	27	26	18	9	4	2	100
$n_i * x_i$	0	27	52	54	36	20	12	201

La moyenne $\lambda = \frac{\sum n_i \ x_i}{n} = \frac{201}{100} = 2.01$ s = 1.43896 $s^2 = 2.07061$

- Calcul des effectifs théoriques :

Doses vendues	**Effectifs observés**	**Probabilités théoriques p_i**	**Eff. Attendus = Np_i**
0	14	$P(X = 0) = e^{-2.01}$	13.39886747 ≅ 13.399
1	27	$P(X = 1) = e^{-2.01} (2.01/1\,!)$	26.93172361 ≅ 26.932
2	26	$P(X = 2) = e^{-2.01} (2.01^2/2\,!)$	27.06638223 ≅ 27.066
3	18	$P(X = 3) = e^{-2.01} (2.01^3/3\,!)$	18.13447609 ≅ 18.134
4	9	$P(X = 4) = e^{-2.01} (2.01^4/4\,!)$	9.112574236 ≅ 9.113
5	4	$P(X = 5) = e^{-2.01} (2.01^5/5\,!)$	3.663254843 ≅ 3.663
6	2	$P(X = 6) = e^{-2.01} (2.01^6/6\,!)$	1.227190372 ≅ 1.227
> 6	0	$P(X > 6) = 0.00465531155$	0.465531155 ≅ 0.466
Total	100	1	100

La condition sur les effectifs théoriques ($\boldsymbol{Ai} \geq \mathbf{5}$) n'étant pas respectée pour les trois dernières classes, alors leur regroupement est obligatoire. On obtient donc le tableau suivant :

	X = 0	X = 1	X = 2	X = 3	X= 4	X ≥ 5	Σ
O_i	14	27	26	18	9	6	100
A_i	13.399	26.932	27.066	18.134	9.113	5.356	100

- **Calcul de la statistique de test :**

$$\mathbf{To} = \sum_{i=1}^{3} \frac{(O_i - A_i)^2}{A_i} = \frac{(14-13.399)^2}{13.399} + \frac{(27-26.932)^2}{26.932} + \frac{(26-27.066)^2}{27.066} + \frac{(18-18.134)^2}{18.134} +$$

$$\frac{(9-9.113)^2}{9.113} + \frac{(6-5.356)^2}{5.356} = 0.148938907 \cong \mathbf{0.149}$$

- **Seuil critique :** Le risque à prendre est α = 5 %, le nombre de degrés de liberté (ddl) est : $\nu = (6 - 1) = 5$, le critère du Khi-Deux tiré de la table 4 est $\chi^{\mathbf{2}}_{\mathbf{5\,;\,0.05}} = 11.07$
- **Décision :** Comme To < $\chi^2_{5\,;\,0.05}$ alors on doit accepter H_0.
  L'échantillon est représentatif de la population, c'est-à-dire les ventes suivent une loi de Poisson.