TD8 : Tests du χ^2 d'indépendance et tests d'ajustement à une loi :

Exercice 1. p. 119.

1. On effectue un test d'indépendance entre 1) les caractères "Conventionnel" et "Bio", et 2) les traces de résidus.

On a :

- X = variable aléatoire à $I = 2$ modalités : {"conventionnel", "bio"}.
- Y = variable aléatoire à $J = 3$ modalités : {"pas de résidus", "< LMR", "≥ LMR"}.

Hypothèses du test :

$$\begin{cases} H_0 = \text{" X et Y sont indépendantes ".} \\ H_1 = \text{" X et Y ne sont pas indépendantes ".} \end{cases}$$

Statistique de test :

Soient :
- n_{ij} ($i \in \{1 \ldots I\}$, $j \in \{1 \ldots J\}$) le nombre d'observations de $\{X = i^e \text{ modalité}\}$ et $\{Y = j^e \text{ modalité}\}$.
- d_{ij} ($i \in \{1 \ldots I\}$, $j \in \{1 \ldots J\}$) l'effectif théorique de $\{X = i^e \text{ modalité}\}$ et $\{Y = j^e$

modalité j dans le cas d'indépendance de X et Y.

Alors la statistique de test :

$$\mathcal{D}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - d_{ij})^2}{d_{ij}}$$

Suit sous H_0 une loi du χ^2 à $\underbrace{(I-1)}_{1\times}\underbrace{(J-1)}_{2}$ degrés de liberté, ie :

$$\mathcal{D}^2 \underset{H_0}{\sim} \chi^2_2 .$$

Zone de rejet :

Au risque $\alpha = 0.05$, on note q le quantile d'ordre $1-\alpha$ de la loi du χ^2 à $(I-1)(J-1)$ ddl.

La région de rejet est définie par :

$$R = \{ d^2 > q \}.$$

2. **Application.**

On commence par calculer les effectifs théoriques dans le cas d'indépendance. On a :

Observations :

	$j=1$	$j=2$	$j=3$	marge
$i=1$	$n_{11} = 195$	$n_{12} = 99$	$n_{13} = 6$	$n_{1.} = 300$
$i=2$	$n_{21} = 212$	$n_{22} = 33$	$n_{23} = 5$	$n_{2.} = 250$
marge	$n_{.1} = 407$	$n_{.2} = 132$	$n_{.3} = 11$	$n = 550$

Théorie.

	$j=1$	$j=2$	$j=3$
$i=1$	$d_{11} = \frac{n_{1\cdot}}{n} \times n_{\cdot 1} = \frac{300}{550} \times 407 = 222$	$d_{12} = \frac{n_{1\cdot}}{n} \times n_{\cdot 2} = 72$	$d_{13} = \frac{n_{1\cdot}}{n} \times n_{\cdot 3} = 6$
$i=2$	$d_{21} = 185$	$d_{22} = 60$	$d_{23} = 5$

Sur les données, on obtient :

$$d^2 = 29,499.$$

Par ailleurs, on lit :

$$q = 5,99.$$

Alors : $d^2 > q$: on est dans la zone de rejet : on rejette H_0, et on conclut, au risque $\alpha = 5\%$, que X et Y ne sont pas indépendantes.

Calcul de la p-value.

Pour $Z \sim \chi^2_2$, on cherche :

$$\text{p-value} = \mathbb{P}(Z > d^2)$$
$$= 1 - \mathbb{P}(Z < d^2)$$
$$< 0,001 \text{ (tables)}$$

On peut calculer (logiciel) que p-value = $3,9 \times 10^{-7}$.
On fait alors la même conclusion.

Exercice 2 :

On cherche ici à vérifier si les données peuvent être des observations d'une v.a suivant une loi de Poisson de paramètre λ.

Pour cela, on veut mettre en place un test d'adéquation à une loi.

Étape 1 : trouver la loi.

On connait la famille de loi : Poisson.

Pour connaître la loi, il nous faut λ.

Or, pour $X \sim \mathcal{P}(\lambda)$, on sait que $\mathbb{E}(X) = \lambda$.

Avec nos observations, on a donc comme valeur λ la plus probable $\lambda = \overline{x} = 1,5$.

Étape 2 : Test.

Hypothèses du test :

H_0 : " l'échantillon est la réalisation de va $X_i \sim \mathcal{P}(1,5)$, $i \in \{1 \dots 150\}$, iid "

H_1 : " l'échantillon n'est pas la réalisat° de va $X_i \sim \mathcal{P}(1,5)$ iid, $i \in \{1 \dots 150\}$ ".

Statistique de test.

$$D^2 = \sum_{i=1}^{l} \frac{(n_i - d_i)^2}{d_i} \underset{H_0}{\sim} \chi^2_{l-1-r}$$

avec n_i : nombre d'observations de la modalité i, et d_i : nombre d'observations théoriques si on a une loi $\mathcal{P}(1,5)$.

Le nombre de ddl du χ^2 est :

$$\underbrace{l}_{\text{nombre de modalités}} - 1 - \underbrace{r}_{\text{nombre de paramètres estimés}}$$

Zone de rejet.

Pour le risque $\alpha = 0.05$, on note q le quantile d'ordre $1-\alpha$ de la loi χ^2 à 5 ddl.

Alors : $R = \{d^2 > q\}$.

Application.

$\forall i \in \{1 \dots 7\}$, on calcule :

$$d_i = 450 \times \mathbb{P}(X = i \mid X \sim \mathcal{P}(1,5)).$$

On obtient :

	0	1	2	3	4	5	6
d	33	50	38	19	7	2	1

Pour pouvoir approximer la loi sous H_0 par une loi du χ^2, il faut au moins 5 observations par classes. Alors on considère :

	0	1	2	3	≥ 4
n	32	54	34	21	9
d	33	50	38	19	10.

Ici $\ell = 5$, $r = 1$.

Donc $q = 7,81$. Par ailleurs on calcule $d^2 = 1,03$.

On ne se situe pas dans la zone de rejet.

Au risque $\alpha = 5\%$, on ne rejette pas H_0, et on peut considérer que les piqûres de moustiques sont réparties selon une loi de Poisson de paramètre $\lambda = 1.5$.