TP n°3
Tests

N.B. : en fin d'énoncé se trouve un rappel de quelques tests statistiques classiques.

1 Régime alimentaire

Dans cet exercice, on se demande quels facteurs peuvent influencer le type de beurre consommé, ce qui permettrait de répondre à des questions du type "la consommation de beurre salé est-elle liée au fait d'être breton ?". Pour cela, on a fait une étude du type de beurre consommé parmi un échantillon de population de 80 personnes, présenté dans le tableau suivant :

Origine	Sexe	Effectif consommant du beurre salé	Effectif consommant du beurre doux
Parisienne	Homme	12	8
Parisienne	Femme	10	10
Bretonne	Homme	16	4
Bretonne	Femme	15	5

1. Dans un premier temps, on se demande s'il y a une dépendance entre le sexe et la consommation de beurre salé.
   a. Proposer un test ? Quelle hypothèse veut-on tester ?
   b. On a le tableau suivant :

	Beurre salé	Beurre doux	Total
Homme	28	12	40
Femme	25	15	40
Total	53	27	80

   Calculer la statistique du χ^2 de ces données.
   c. Avec un niveau de confiance 95%, doit-on considérer que le sexe a une influence sur la consommation de beurre salé ?
2. Dans la suite, on cherche à créer une fonction `Chi2Indep`(M, α) retournant le résultat d'un test du χ^2 d'indépendance avec un niveau de confiance $1 - \alpha$, où M est la matrice compilant les données du problème. Par exemple, dans le cas de la question 1,
$$M = \begin{bmatrix} 28 & 12 \\ 25 & 15 \end{bmatrix}.$$
3. a. Ecrire une fonction `Chi2Indep`(M, α) retournant un booléen *True* ou *False*, répondant à la question "Ces deux caractères sont-ils indépendants". On pourra utiliser à bon escient la fonction Scilab `cdfchi`.
   b. La consommation de beurre salé est-elle liée au fait d'être breton ?

2 Qualité du générateur de nombres aléatoires

Dans cet exercice, on se demande si le générateur de nombres aléatoires de Scilab fournit bien des nombres aléatoires. Plus précisément, on se demande si un échantillon obtenu à l'aide de la fonction Scilab `rand` suit bien une loi $\mathscr{U}([0,1])$.

1. Quelle hypothèse veut-on tester ? Quelle est l'hypothèse alternative ?
2. Générer un échantillon de taille 20, qu'on appellera X, et tracer son histogramme. Qu'en pensez-vous ?
3. Pour résoudre ce problème, on peut utiliser un test de Kolmogorov-Smirnov.
   a. Tracer sur l'intervalle $[0,1]$ (sur un même graphe) la fonction de répartition F de $\mathscr{U}([0,1])$ et la fonction de répartition empirique de $X = (X_1, \ldots, X_N)$:
   $$F_N(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{\{X_k \leq x\}}$$
   b. Créer une fonction $\texttt{KolSmi}(X)$ retournant *True* ou *False* à la question "L'échantillon suit-il la loi $\mathscr{U}([0,1])$?" avec un niveau de confiance 95%. Attention, on tirera l'échantillon X en amont (et pas dans la fonction).
   c. Conclure, pour un échantillon de taille 100.
4. a. Si l'on regarde 1000 échantillons générés à l'aide de `rand`, quelle proportion d'entre eux est validée par le test de Kolmogorov-Smirnov ?
   b. Même question avec la commande $\texttt{grand}(\cdot,\cdot,\texttt{'unf'},0,1)$.

3 Baccalauréat

On a demandé à tous les élèves d'une classe de terminal de choisir un menu pour l'épreuve d'EPS au baccalauréat :

Menu	Nombre d'élèves
Menu n°1 : Endurance - Natation - Badminton	7
Menu n°2 : Endurance - Judo - Volley	5
Menu n°3 : Endurance - Saut en hauteur - Handball	5
Menu n°4 : Natation - Saut de haie - Football	4
Menu n°5 : Natation - Ping-pong - Basket-ball	9

1. On se demande si chaque menu est choisi par la même proportion d'élèves.
   a. Quelle hypothèse veut-on tester ? Quelle est l'hypothèse alternative ?
   b. Pour résoudre ce problème, on peut utiliser un test du χ^2 d'ajustement à une distribution. Quelle est la loi de cette distribution ?

c. Créer une fonction `Chi2Adeq`$(M^{\text{obs}}, p^{\text{theo}}, \alpha)$ retournant *True* ou *False* à la question "L'échantillon suit-il la distribution théorique ?" avec un niveau de confiance $1-\alpha$, les vecteurs M^{obs} et p^{theo} représentant respectivement les effectifs observés et les proportions théoriques de l'échantillon.

d. Conclure.

2. On étend l'étude à toutes les classes de terminale du lycée :

Menu	Nombre d'élèves
Menu n°1	72
Menu n°2	49
Menu n°3	48
Menu n°4	46
Menu n°5	85

Que peut-on en conclure ?

Rappels : Tests classiques

Test du χ^2 d'adéquation

Statistique du χ^2 d'adéquation pour un N-échantillon :

$$\xi_N = N \sum_{i \le k} \frac{(p_i^{\text{obs}} - p_i^{\text{theo}})^2}{p_i^{\text{theo}}}, \quad p_i^{\text{obs}} : \text{prop. observée}, \quad p_i^{\text{theo}} : \text{prop. théorique}.$$

On ne rejette pas H_0 si $\xi_N < q_\alpha$, où q_α est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(k-1)$.

Test du χ^2 d'indépendance

Statistique du χ^2 d'indépendance pour un N-échantillon :

$$\xi_N = N \sum_{i \le k, j \le l} \frac{(p_{ij}^{\text{obs}} - p_{ij}^{\text{theo}})^2}{p_{ij}^{\text{theo}}}, \quad p_{ij}^{\text{obs}} : \text{prop. observée}, \quad p_{ij}^{\text{theo}} : \text{prop. théorique}.$$

On ne rejette pas H_0 si $\xi_N < q_\alpha$, où q_α est le quantile d'ordre $1-\alpha$ de la loi $\chi^2((k-1)(l-1))$.

Test de Kolmogorov-Smirnov

Statistique de Kolmogorov-Smirnov :

$$\xi_N = \sup_x |F_n(x) - F(x)|, \quad F_n : \text{f.d.r. empirique}, \quad F : \text{f.d.r. théorique}.$$

On ne rejette pas H_0 si $\xi_N < q_\alpha N^{-1/2}$, où q_α est le quantile d'ordre $1-\alpha$ de la loi de Kolmogorov ($q_\alpha = 1.36$ pour $\alpha = 0.05$).