

Feuille de TD : Tests non-paramétriques et tests du khi-deux

Exercice 1. On observe les 10 valeurs suivantes pour X_1, \dots, X_{10} :

0.39 0.22 0.99 0.62 0.18 0.53 0.15 0.96 0.7 0.26

Proposez un test au niveau 5% pour tester l'adéquation à la loi uniforme $\mathcal{U}([0, 1])$.

Exercice 2. Huit individus ont été traités avec le soporifique S et huit autres individus avec un produit inactif I . Pour chacun des 16 sujets, le temps de sommeil moyen après traitement a été enregistré : X_i représente le temps de sommeil moyen pour l'individu i traité avec le soporifique (S) et Y_j représente le temps de sommeil moyen pour l'individu j traité avec le produit (I). On a observé (en minutes)

i, j	1	2	3	4	5	6	7	8
X_i (S)	578	568	548	478	458	538	618	548
Y_j (I)	430	360	430	570	490	480	380	400

1. Peut-on considérer que les deux échantillons sont indépendants ?

2. Peut-on conclure à l'efficacité du soporifique ?

Proposez deux tests pour répondre à cette question. Vous préciserez bien le modèle utilisé, les hypothèses testées, les statistiques de test utilisées, la zone de rejet ainsi que les conclusions des deux tests proposés pour un niveau $\alpha = 5\%$.

Exercice 3. Dans une usine, on a du mal à fixer le taux d'acidité des yaourts. En comparant ce taux pour 10 pots après 5 heures de fabrication au taux pour ces mêmes 10 pots juste au moment de la fabrication, on obtient :

0h (X_i)	12.51	12.48	12.91	12.56	12.58	12.82	12.53	12.50	12.51	12.42
5h (Y_i)	12.82	12.79	12.74	12.88	12.82	12.40	12.84	12.81	12.91	12.39

1. Peut-on considérer que les échantillons (X_1, \dots, X_n) et (Y_1, \dots, Y_n) sont indépendants?

2. On considère les variables $Z_i = Y_i - X_i$ pour $i = 1, \dots, n$. On supposera dans un premier temps que les Z_i sont i.i.d. de loi normale. Proposez un test, basé sur (Z_1, \dots, Z_n) pour tester s'il y a eu une variation significative du taux d'acidité.

3. On ne suppose plus que les Z_i sont de loi normale. On introduit la variable :

$$N = \sum_{i=1}^n \mathbf{1}_{Z_i \geq 0}.$$

Quelle est la loi de N sous l'hypothèse que la médiane des variables Z_i est nulle ?

4. En utilisant la variable N , testez s'il y a eu une variation significative du taux d'acidité.

Exercice 4. On pèse 20 plaquettes de beurre pris au hasard dans une production normande. Les résultats, en grammes, sont :

247.0 247.8 250.2 251.3 251.9 249.4 248.8 247.1 255.0 247.0
254.8 244.8 250.7 250.7 252.6 251.1 254.1 249.2 252.0 254.0

On suppose que le poids en grammes d'une plaquette de beurre de cette production peut être modélisé par une v.a.r X . Peut-on affirmer que X suit une loi normale ?

Exercice 5. Une urne contient des boules de quatre couleurs différentes avec des proportions inconnues. On tire au hasard 160 boules avec remise et l'on obtient les proportions suivantes sur l'échantillon observé :

Couleur	Noir	Rouge	Jaune	Vert	Total
Effectifs observés	100	18	24	18	160

1. Accepteriez-vous, au seuil $\alpha = 5\%$, l'hypothèse que les proportions de boules noires, rouges, jaunes et vertes soient respectivement $9/16$, $3/16$, $3/16$ et $1/16$? Soyez précis dans votre réponse : vous donnerez la statistique de test ainsi que sa loi asymptotique sous l'hypothèse nulle et la zone de rejet. Donnez un encadrement de la p -valeur du test réalisé.
2. On note N_1, N_2, N_3, N_4 le nombre de boules noires, rouges, jaunes et vertes. On souhaite préciser l'inhomogénéité de la répartition, et tester s'il y a autant de boules rouges et jaunes que de vertes, i.e. le taux de boules de chaque couleur est de la forme $(p_1, p_2, p_3, p_4) = (1 - 3\theta, \theta, \theta, \theta)$. En supposant que (N_1, N_2, N_3, N_4) suit une loi multinomiale de paramètre $(n, (p_1, p_2, p_3, p_4))$ où n est le nombre total de boules dans l'urne, déterminez l'estimateur du maximum de vraisemblance de θ .
3. Effectuez un test d'adéquation à la forme postulée pour la loi. Qu'en concluez-vous ?

Exercice 6. Pour comparer la proportion de personnes atteintes par la grippe en ville et à la campagne, deux échantillons ont été étudiés :

- sur 100 personnes habitant une grande agglomération, on a observé une proportion $\hat{p}_0 = 0.24$ de sujets ayant eu la grippe.
- sur 100 personnes habitant à la campagne, on a observé une proportion $\hat{p}_1 = 0.20$ de sujets ayant eu la grippe.

Selon vous, la proportion de sujets atteints par la grippe est-elle différente en ville et à la campagne ?

Exercice 7. A une élection, deux candidats se présentent : N.S. et S.R.

500 électeurs sont sondés. On leur demande non seulement leur opinion politique (i.e. leur choix à la future élection) mais aussi leur appartenance sociale (Rentiers/Actifs/Retraités). On obtient les résultats suivants :

	N.S.	S.R.
Rentiers	40	10
Actifs	140	210
Retraités	77	23

L'opinion politique dépend-elle de l'appartenance sociale ?