Feuille N° 8
Statistiques Inférentielles

Exercice 1

Dans un étude sur les retards de train aux Pays-Bas, on se demande si les retards montrent plus de variabilité dans les heures de pointe que dans les heures creuses. Les retards observés pendant les heures de pointe sont modélisés par un échantillon distribué selon une loi de variance σ_1^2, et ceux observés pendant les heures creuses par un échantillon distribué selon une loi de variance σ_2^2. On souhaite tester l'hypothèse nulle $H_0 : \sigma_1^2 = \sigma_2^2$, quelle hypothèse alternative choisiriez-vous ?

Exercice 2

En 1986, deux chercheurs ont étudié le nombre de cycles menstruels nécessaires aux femmes, fumeuses et non-fumeuses, pour tomber enceinte, à partir du moment où elles ont décidé d'avoir un enfant. On suppose que les nombres de cycles observés sont modélisés comme les réalisations d'échantillons de loi géométrique. On note p_1 la probabilité qu'une femme fumeuse tombe enceinte pendant un cycle donné, indépendamment des cycles précédents et p_2 la probabilité qu'une femme non-fumeuse tombe enceinte pendant un cycle donné, indépendamment des cycles précédents. On se demande si p_1 est différent de p_2, et on se propose donc d'étudier le test $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.

1. On suppose que l'on a les données suivantes :

Cycles	1	2	3	4	5	6	7	8	9	10	11	12
Fumeuses	29	16	17	4	3	9	4	5	1	1	1	3
Non-fumeuses	198	107	55	38	18	22	7	9	5	3	6	6

   Quelle statistique de test proposeriez-vous ? La calculer avec les données.

2. On se demande si les femmes fumeuses ont plus de difficultés à tomber enceinte que les femmes non-fumeuses. Quelle serait l'hypothèse alternative dans ce cas ?

Exercice 3

On suppose que l'on dispose d'un jeu de données, réalisation d'un échantillon $X_1, \dots, X_n$ de loi uniforme sur $[0, \theta]$, pour un certain $\theta > 0$ inconnu. On fait le test $H_0 : \theta = 5$ contre $H_1 : \theta \neq 5$.

1. On prend $T_1 = \max\{X_1, \dots, X_n\}$ comme statistique de test. Quelles valeurs prend T_1 ? Lesquelles sont en faveur de H_0 ? de H_1 ? On pourra faire un dessin.
2. Même question avec $T_2 = \left|2\overline{X}_n - 5\right|$.

Exercice 4

Pour tester une certaine hypothèse nulle H_0, on utilise une statistique de test T construite à partir d'un échantillon de loi continue. On rejette H_0 si l'on observe une valeur t de la statistique de test pour laquelle sous H_0, $\mathbb{P}\{T \geqslant t\} \leqslant 0.05$. On donne différentes valeurs de t et une queue de probabilité à droite ou à gauche correspondantes (sous H_0). Dire dans chaque cas que vaut la p-valeur et si on devrait rejeter H_0.

1. $t = 2.34$ et $\mathbb{P}\{T \geqslant 2.34\} = 0.23$.
2. $t = 2.34$ et $\mathbb{P}\{T \leqslant 2.34\} = 0.23$.
3. $t = 0.03$ et $\mathbb{P}\{T \geqslant 0.03\} = 0.968$.
4. $t = 1.07$ et $\mathbb{P}\{T \leqslant 1.07\} = 0.981$.
5. $t = 1.07$ et $\mathbb{P}\{T \leqslant 2.34\} = 0.01$.

6. $t = 2.34$ et $\mathbb{P}\{T \leqslant 1.07\} = 0.981$.
7. $t = 2.34$ et $\mathbb{P}\{T \leqslant 1.07\} = 0.800$.

Exercice 5

En moyenne, le nombre de bébés nés à Cleveland au mois de Septembre est 1472. Le 26 janvier 1977, la ville a été immobilisée par un blizzard. Neuf mois plus tard, en Septembre 1977, le nombre enregistré de naissances était 1718. Peut-on attribuer la hausse de 246 à la chance ? Pour répondre à cette question, on modélise le nombre de naissances par une variable aléatoire T de loi de Poisson de paramètre μ, qui représente le nombre moyen de naissances. On souhaite tester l'hypothèse H_0 : $\mu = 1472$.

1. Quelle hypothèse alternative H_1 choisiriez-vous ?
2. On décide de rejeter H_0 pour $t = 1718$. Quelles valeurs de T sont en faveur de H_0 ? En faveur de H_1 ?
3. Calculer la p-valeur pour $t = 1718$, en approchant la loi de T par la loi $\mathcal{N}(\mu, \mu)$.

Exercice 6

On se donne un réel t, qui est la réalisation d'une variable aléatoire T de loi $\mathcal{N}(\mu, 1)$. Pour tester $H_0 \colon \mu = 0$ contre $H_1 \colon \mu \neq 0$, on utilise T comme statistique de test, et on décide de rejeter H_0 si $|t| \geqslant 2$. Calculer la probabilité de faire une erreur de type I.