Université d'Evry-Val d'Essonne 2018-2019
L2 Informatique
DL2 Biologie Informatique
Cours de Statistiques

Feuille de TD 5 : Tests statistiques

Exercice 1
On suppose que le poids (en kg) d'une boîte de conserve vendue par un certain fabricant est une variable aléatoire X de loi $\mathcal{N}(\mu; 0,04)$. Jusqu'à présent le poids moyen d'une boîte était de $1kg$. Le fabricant prétend avoir augmenté le contenu de chaque boîte de 25% en gardant le même prix. On prélève un échantillon de 10 boîtes et on obtient $\overline{x}_n = 1.13kg$ avec $\overline{x}_n = n^{-1}\sum_{i=1}^{n} x_i$ et x_i le poids observé de la i$^{\text{ème}}$ boîte.
On note

$$\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i.$$

On pose $(H_0)\ : \mu = 1$ et $(H_1)\ : \mu = 1.25$. On va procéder au test de (H_0) contre (H_1).

1. Quelles sont les deux erreurs que l'on peut commettre lors de ce test ?
2. Donner la définition du risque de première espèce.
3. Donner la définition du risque de seconde espèce.
4. Donner la définition du niveau du test.
5. Quelle est la statistique de test ? Justifier.
6. Construire le test au niveau α en précisant la zone de rejet et la règle de décision du test.
7. Application numérique : que concluez-vous au niveau $\alpha = 1\%$, $\alpha = 5\%$ et $\alpha = 10\%$. Commenter.
8. Calculer la p-value du test. Conclusion ?
9. Donner la définition de la puissance du test. Calculer la puissance du test Π_α pour $\alpha = 1\%$, $\alpha = 5\%$ et $\alpha = 10\%$.
10. En prenant $\alpha = 5\%$, représenter (en les précisant) sur un même graphique les lois de la statistique de test, sous l'hypothèse (H_0) et sous l'hypothèse (H_1). On indiquera également où se lisent sur le graphique la zone de rejet, le niveau, la puissance et la p-value.

    On considère à présent les hypothèses $(H_0)\ : \mu = 1$ et $(H_1)\ : \mu > 1$. On teste (H_0) contre (H_1).
11. Construire le test et comparer avec ce qui a été fait précédemment.
12. Donner la définition de la puissance du test. Calculer la puissance du test de niveau $\alpha = 5\%$ lorsque sous l'alternative (H_1), $\mu = 1.15$. Commenter.

13. Que devient cette puissance si on observe $\overline{x} = 1.13$ sur un échantillon de 100 boîtes au lieu de 10. Commenter.

Exercice 2

Un casino propose un nouveau jeu aléatoire où, à chaque partie du jeu, ou bien on gagne un montant x euros donné ou bien on en perd l'équivalent. Le casino affirme que les parties successives du jeu sont indépendantes et que le jeu est équilibré, autrement dit, à chaque partie du jeu on a autant de chance de gagner que de perdre. Avant de me mettre à jouer, je laisse se dérouler soixante parties du jeu et j'observe les résultats suivants (où G signifie que la partie est remportée par le joueur et P signifie qu'elle est remportée par le casino) :

$$\begin{array}{cccccccccc}
P & G & G & P & P & G & P & P & P & G \\
G & P & G & P & P & P & P & G & P & G \\
P & G & P & P & P & G & G & P & P & G \\
P & P & G & P & P & P & P & P & G & P \\
P & P & P & G & G & P & P & P & G & G \\
P & P & P & G & P & P & P & G & G & G
\end{array}$$

1. Donner une estimation de la probabilité p que je perde une partie du jeu.
2. Je doute que le jeu soit équilibré et je souhaite vérifier cette intuition à travers un test statistique. Je veux tester l'hypothèse nulle
   $H_0 : p = 0.5$ contre l'alternative $H_1 : p = 0.6$.
   (a) Quelle est la statistique de test ?
   (b) Quelle est la région de rejet du test de niveau asymptotique α ? Justifier.
   (c) Quel est le risque de première et de seconde espèce associé au test ? Sait-on les calculer de façon exacte ?
   (d) Quelle est la puissance asymptotique du test de niveau asymptotique $\alpha = 0.05$?
   (e) Rejette-t-on ou non l'hypothèse que le jeu est équilibré, au niveau asymptotique $\alpha = 0.03$? $\alpha = 0.05$? $\alpha = 0.01$?
   (f) Quelle est la p-value (ou niveau de signification) asymptotique du test ?

Exercice 3

Dans un service de pédiatrie, on traite les enfants asthmatiques au moyen d'un corticoïde C et l'on cherche à savoir si ce traitement modifie la durée moyenne de sommeil par nuit. Pour cela, on constitue par tirage au sort un groupe de 40 enfants traités avec le corticoïde C et on relève sur ces 40 enfants, la durée de sommeil au cours d'une nuit. On obtient sur cet échantillon une durée moyenne (empirique) de sommeil de 9,5h. Chez les enfants non traités et de cette tranche d'âge, on sait que la durée moyenne théorique de sommeil est de 10h par nuit avec un écart-type de 1h. On suppose que l'écart-type de la durée de sommeil est la même chez les enfants traités et non traités. Construire un test permettant de savoir s'il y a une modification de la durée moyenne de sommeil chez les enfants traités par rapport à la durée moyenne attendue chez un enfant non traité. La construction du test sera rédigée de façon claire et détaillée.

Exercice 4
On suppose que le nombre X d'agressions par jour dans une agglomération suit une loi de Poisson de paramètre λ. La mairie annonce qu'il y a en moyenne $\lambda_0 = 1,3$ agressions par jour dans cette agglomération. Ce chiffre est contesté par une association qui pense qu'il est sous-estimé. L'association veut vérifier cela à travers un test statistique. Pour cela, elle a relevé sur 30 jours, les nombres d'agressions chaque jour, qui sont donnés par le tableau suivant :

3	4	2	5	3	4	2	1	7	0
2	1	4	2	1	1	2	3	3	1
1	1	4	1	1	0	2	1	1	3

1. Donner une estimation de λ.
2. Quel test peut-on construire pour prouver que la mairie sous estime le nombre moyen d'agression par jour ? Donner les hypothèses testées.
3. Donner la statistique de test et la forme de la zone de rejet en justifiant votre réponse.
4. Décide-t-on du rejet ou non de l'hypothèse H_0 lorsque le niveau asymptotique α vaut 0.03 ? lorsque $\alpha = 0.05$? lorsque $\alpha = 0.01$?
5. L'association peut-elle remettre en cause le chiffre avancé par la mairie et conclure à une sous estimation ?

Exercice 5
On envisage d'ajouter un adjuvant au traitement usuel contre un certain type de rhumatisme. Sans adjuvant, la durée (exprimée en jours) séparant deux crises de récurrence rhumatismale peut être modélisée par une variable aléatoire de loi gaussienne d'espérance $\mu = 560$. On administre le traitement avec adjuvant à 10 sujets. Les durées de récurrence observées sont les suivantes : 646, 573, 485, 752, 742, 636, 607, 665, 506, 575. Au niveau $\alpha = 5\%$, peut-on conclure que l'adjuvant modifie la durée moyenne de récurrence ?

Exercice 6
Un homme politique a remporté les dernières élections avec 52% de voix. Un sondage effectué un an après son élection auprès de 1000 personnes indiquent que 508 d'entre elles voteront pour lui si de nouvelles élections ont lieu. Ces données permettent-elles de conclure que :
1. si de nouvelles élections avaient lieu, son score serait différent de celui obtenu l'année précédente ?
2. si de nouvelles élections avaient lieu, son score serait inférieur à celui obtenu l'année précédente ?
Répondre à ces questions à l'aide de tests dont on détaillera la construction.