Partiel - Durée 1 h

Un formulaire d'une page (donc un recto) est autorisé. Les tables des lois sont autorisées. **Tout autre document est prohibé**. Les calculatrices sont autorisées. **Les téléphones sont interdits.**

Problème.

Un chercheur en panne d'inspiration décide de réaliser une expérience pour mesurer l'influence d'un régime huileux sur le poids des souris de laboratoires. Il commande 20 souris blanches (de laboratoire) adultes à un élevage et les nourrit pendant un mois uniquement avec de l'huile de friture usagée (grâce à un partenariat avec une grande chaîne de fast-food). Il mesure le poids de chaque souris avant et après le régime à base d'huile de friture. Sur les 7 souris qui survivent à ce régime, il obtient les mesures suivantes (en grammes)

Souris Numéro	1	2	3	4	5	6	7
Poids sans régime huileux (x_i)	44	41	24	40	32	39	42
Poids avec régime huileux (y_i)	53	40	33	53	43	39	46

Il calcule alors les moyennes empiriques ($\bar{x}$ et $\bar{y}$) et les variances empiriques modifiés (${s^2}'_x$ et ${s^2}'_y$) associés. Il obtient :

– Sans huile : $\bar{x} = 37.43$ et ${s^2}'_x = 57,50$;
– Avec huile : $\bar{y} = 43.86$ et ${s^2}'_y = 63.94$.

Il en déduit que le régime riche en huile usagée fait augmenter le poids des souris et s'apprête à publier ses résultats, mais il se rappelle tout à coup qu'en L2 il avait vaguement entendu son professeur de mathématiques parler de statistique et de tests.

Exercice I) Expliquer pourquoi il ne peut pas conclure directement, juste en comparant $\bar{x}$ à $\bar{y}$, à l'augmentation du poids des souris suite à un régime huileux.

$\hookrightarrow$ ici mettre 1/2 pt si l'étudiant a bien dit que $\bar{x}$ et $\bar{y}$ ne sont ques des estimations et un autre 1/2 pt si l'étudiant a compris que c'était des variables aléatoires (ou disons que leurs valeurs pouvaient changer)

Après une rapide recherche sur internet, il se décide à utiliser un test classique pour comparer ces deux régimes alimentaires, avec un seuil de 5%, **et en supposant les écart-type égaux**.

Exercice II) Dans un premier temps les deux populations de souris (avec et sans huile) sont supposées différentes et indépendantes

1. Quelles sont les populations statistiques ? quels sont les échantillons ?
   $\hookrightarrow$ deux populations de souris : avec et sans huile. Deux échantillons de taille 7 (1/2 pt)
2. Quelles sont les variables étudiées ?
   $\hookrightarrow$ Les variables poids souris sans huile et poids souris avec huile. Elles sont quantitatives continues (1/2 ou 1 pt)

3. Ecrire les hypothèses de test qui permettent de répondre à la question du chercheur.
   $\hookrightarrow$ $H_0 : m_x = m_y$ vs $H_1 : m_x \neq m_y$ test bilatéral (1/2 pt)
4. Quelles sont les conditions nécessaires pour réaliser ce test. Sont-elles raisonnables ?
   $\hookrightarrow$ indépendance des toutes les observations (ie des x, des y, et des x et y entre elles). normalité des x avec même loi (x iid) et normalité des y avec même loi (y iid). On peut accepter la normalité car x et y continues, en supposant que les valeurs négatives sont négligeables. On peut difficilement accepter l'indépendance entre x et y, mais c'est supposé ici. (1 pt)
5. Calculer la valeur de la statistique de test obtenue sur l'échantillon.
   $\hookrightarrow$ on trouve $T = -1.54$ (1pt)
6. Au seuil de 5% peut-on dire que les moyennes des poids sont différentes ?
   Tous les fractiles sont donnés en fin d'énoncé.
   $\hookrightarrow$ T ne dépasse pas (à droite) le fractile -2.18. Donc on ne rejette pas H_0. Les moyenne ne sont pas considérées comme différentes ici. (1pt)
7. Donner un encadrement de la "p-value" (c'est-à-dire la probabilité de se tromper en rejetant H_0).
   $\hookrightarrow$ On peut encadrer la p-value grâce aux valeurs données : elle est comprise entre 10% et 20%. (1pt)
8. Quelles est votre conclusion après cette étude ?
   $\hookrightarrow$ sous nos hypothèses d'indépendance on ne peut pas conclure à une différence de poids due au régime. (1/2 pt)

Satisfait par son étude statistique, félicité par la chaîne de fast-food, il fait écrire un article par son thésard et après une relecture rapide le soumet pour publication dans une revue prestigieuse. Quelques mois plus tard, le rapport du relecteur lui parvient [1] :

> *L'auteur manque clairement de connaissance en statistique. Il n'utilise pas les tests appropriés. Cet article ne mérite pas d'être publié.*

En effet, les 7 souris observées avant et après le régime à base d'huile étaient les mêmes. L'auteur de l'étude décide donc de recommencer son analyse statistique sous de nouvelles conditions.

Exercice III) On ne suppose plus l'indépendance des deux populations de souris : le poids de chaque souris est mesuré avant et après le régime huileux

1. Quelle est la population statistique considérée ? quel est l'échantillon ?
   $\hookrightarrow$ souris de labo. Echantillon taille 7. (1/2 pt)
2. Quelle est la variable étudiée ?
   $\hookrightarrow$ la différence des poids : quanti continue (1,5 pt)
3. Ecrire les hypothèses de test qui permettent de répondre à la question du chercheur.
   $\hookrightarrow$ en notant m la moyenne des différence : $H_0 : m = 0$ vs $H_1 : m \neq 0$ (1 pt)
4. Quelles sont les conditions nécessaires pour réaliser ce test. Sont-elles raisonnables ?
   $\hookrightarrow$ indépdendance des $z = x - y$, et normalité des z (et non plus des x ou y) C'est tout a fait raisonnable car souris indépendantes et de même espèce (donc iid car même comportement alimentaire etc...) De plus, la différence est une variable réelle qui peut être négative ce qui renforce l'hypothèse de normalité. (1pt)

1. Les articles scientifiques sont normalement relus par des spécialistes avant d'être publiés

5. Calculer la valeur de la statistique de test obtenue sur l'échantillon.
   $\hookrightarrow$ on trouve $T = -2.87$. (1pt)
6. Au seuil de 5% peut-on dire que les moyennes des poids sont différentes ?
   Tous les fractiles sont donnés en fin d'énoncé.
   $\hookrightarrow$ On rejette H_0 car la stat dépasse (à droite) le fractile -2.45. (1pt)
7. Donner un encadrement de la "p-value" (c'est-à-dire la probabilité de se tromper en rejetant H_0).
   $\hookrightarrow$ la p-value est entre 2% et 5% (1pt)
8. Quelles est votre conclusion après cette étude ?
   $\hookrightarrow$ le régime a une influence sur le poids des souris (1/2pt)
9. Quels sont, dans la situation qui nous intéresse, les avantages de ce second test par rapport au premier ?
   $\hookrightarrow$ le hypothèses sont plus raisonnables : la normalité a plus de sens car la différence peut être négative. D'aute part l'indépendance peut être supposée. Et l'égalité des variances n'est plus à supposer (ou à tester avec risque d'erreur...)... (1pt)

Fort de cette nouvelle analyse statistique, le thésard modifie l'article et le chercheur le soumet à une revue moins prestigieuse. Il attend maintenant avec impatience la réponse du nouveau relecteur. Par contre, la chaîne de fast-food décide d'interrompre son partenariat.

10. Quel serait votre avis scientifique sur cette expérience si vous étiez le relecteur ? (bonus)
    $\hookrightarrow$ peut être un manque d'observations ? l'auteur a-t-il testé le régime sur lui même ? quid des souris mortes ? la recherche a-t-elle un sens ? et donne -t-elle un sens à la vie ? doit-on lutter contre la traite scientifique des thésards

Rappelons que le fractile u d'ordre 0.9 signifie que $P(X < u) = 0.9$, et que pour la loi de Student $P(X > x) = P(X < -x)$. On donne les fractiles suivants :
✓ *Pour une loi de Student à 12 degrés de liberté : le fractile d'ordre 0.90 vaut 1.36, le fractile d'ordre 0.95 vaut 1.78, le fractile d'ordre 0.975 vaut 2.18, et le fractile d'ordre 0.99 vaut 2.68.*
✓ *Pour une loi de Student à 6 degrés de liberté : le fractile d'ordre 0.90 vaut 1.44, le fractile d'ordre 0.95 vaut 1.94, le fractile d'ordre 0.975 vaut 2.45, et le fractile d'ordre 0.99 vaut 3.14.*

Précisons pour finir que ceci est une “caricature”, et que la recherche est (en grande majorité) faite dans des conditions plus sérieuses.