Annales L2

Biostatistiques S4

Sommaire :

- Sujet 1 , 2017-2018 : Page : 3
  Correction, 2017-2018 : Page : 12

- Sujet 1 , 2018-2019 : Page : 14
  Correction, 2018-2019 : Page : 24

- Sujet 1 , 2019-2020 : Page : 27
  Correction, 2019-2020 : Page : 40

- Sujet 1 , 2021-2022 : Page : 44
  Correction, 2021-2022 : Page : 57

- Sujet 1 , 2022-2023 : Page : 62

Amicale Paris Sciences

Licence Sciences Biomédicales
2017-2018

Session 1 – Semestre 4
Sujet

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

UNIVERSITÉ
PARIS
DESCARTES

Notes
Ex. 1 :
Ex. 2 :
Ex. 3 – A :
Ex. 3 – B :
Ex. 3 – C :
Total :

L2
Année 2017 – 2018
4e semestre
1re session

Faculté des Sciences
Fondamentales et Biomédicales
Faculté de Pharmacie de Paris

UE 4.08 ÉPREUVE DE BIOSTATISTIQUE

Tout signe distinctif porté sur la copie pouvant indiquer sa provenance constitue une fraude.
Aucun document n'est autorisé. La calculette est autorisée.
Il est interdit de désagrafer ce cahier composé de 8 pages numérotées et de 4 pages d'annexes.

Exercice 1 – Dans l'industrie agro-alimentaire, l'étude de la contamination de produits par des micro-organismes est un problème très fréquemment rencontré. En particulier, on étudie le cas de contamination de lait par des spores de *Clostridia*. Cette bactérie, naturellement présente dans les organismes humains et animaux, peut causer chez des individus fragiles des diarrhées voire, dans les cas les plus graves, le botulisme.

1) L'étude de contamination effectuée dans un premier tube de 1 mL de lait indique l'absence (associée au succès) ou la présence (associée à l'échec) de spores. La variable aléatoire X_1 modélise ce résultat. Quelles sont les réalisations possibles et la loi de X_1 ?
2) L'étude est ensuite effectuée sur n tubes indépendants de 1 mL de lait. La variable aléatoire Y modélise le nombre de tubes stériles, c'est-à-dire sans spore, parmi les n tubes. Quelles sont les réalisations de Y ? Comment définir Y à partir de la question 1 ?
3) En supposant que la probabilité qu'un tube soit contaminé est la même quel que soit le tube, quelle est la loi de Y ?
4) On suppose que le nombre de spores dans un tube de 1 mL de lait, noté Z, suit une loi de Poisson de paramètre λ. À quelle(s) réalisation(s) de Z correspond le fait qu'un tube soit stérile ? En fonction de λ, quelle est la probabilité qu'un tube soit stérile ? À quelle(s) réalisation(s) de Z correspond le fait qu'un tube ne soit pas stérile ? Quelle est la probabilité qu'un tube ne soit pas stérile ?
5) Donnez, en fonction de λ, les lois de X_1 et de Y.
6) Pour 10 tubes de 1 mL de lait et une probabilité de 0,6 qu'un tube soit stérile, déterminez le nombre moyen de tubes stériles parmi les 10. Déterminez le nombre moyen de spores dans un tube de 1 mL de lait.

Réponses

1) L'étude de contamination effectuée dans un premier tube de 1 mL de lait indique l'absence (associée au succès) ou la présence (associée à l'échec) de spores. La variable aléatoire X_1 modélise ce résultat. Quelles sont les réalisations possibles et la loi de X_1 ?

2) L'étude est ensuite effectuée sur n tubes indépendants de 1 mL de lait. La variable aléatoire Y modélise le nombre de tubes stériles, c'est-à-dire sans spore, parmi les n tubes. Quelles sont les réalisations de Y ? Comment définir Y à partir de la question 1 ?

3) En supposant que la probabilité qu'un tube soit contaminé est la même quel que soit le tube, quelle est la loi de Y ?

4) On suppose que le nombre de spores dans un tube de 1 mL de lait, noté Z, suit une loi de Poisson de paramètre λ. À quelle(s) réalisation(s) de Z correspond le fait qu'un tube soit stérile ? En fonction de λ, quelle est la probabilité qu'un tube soit stérile ? À quelle(s) réalisation(s) de Z correspond le fait qu'un tube ne soit pas stérile ? Quelle est la probabilité qu'un tube ne soit pas stérile ?

5) Donnez, en fonction de λ, les lois de X_1 et de Y.

6) Pour 10 tubes de 1 mL de lait et une probabilité de 0,6 qu'un tube soit stérile, déterminez le nombre moyen de tubes stériles parmi les 10. Déterminez le nombre moyen de spores dans un tube de 1 mL de lait.

Exercice 2 – Une usine fabrique des comprimés qui doivent être conditionnés dans une plaquette thermoformée afin de les protéger. Les comprimés sont donc alignés puis enfermés entre une feuille d'aluminium et un film de matière plastique transparente constituant, grâce à un cloquage, des petites logettes pour les comprimés. Afin que la procédure se déroule au mieux, le diamètre d'un comprimé, variable aléatoire notée D, se devrait d'être de 8 mm. Cependant, les erreurs d'usinage provoquent des variations de diamètre. D'après une étude antérieure, l'erreur sur le diamètre d'un comprimé, variable aléatoire notée E, suivrait une loi normale centrée d'écart-type σ mm.

1) Quel est le lien entre les variables aléatoires E et D ? Justifiez la loi de D en précisant ses paramètres caractéristiques. Donnez une représentation graphique de E et de D.
2) Un premier contrôleur affirme que σ est égal à 0,02 mm et que tous les comprimés dont le diamètre n'est pas compris entre 7,97 mm et 8,03 mm doivent être rejetés. Donnez une représentation graphique de cette affirmation en utilisant les densités de probabilités de E et D. En supposant cette affirmation exacte, quelle est la proportion de comprimés rejetés ?
3) Un deuxième contrôleur affirme que, pour optimiser la phase de cloquage, la probabilité que le diamètre d'un comprimé soit supérieur à 8,1 mm doit être de 1 %. Dans ce contexte, déterminez σ.
4) À la vue des résultats, un troisième contrôleur confirme la valeur de σ du premier contrôleur ainsi que la probabilité de 1 % du deuxième contrôleur sur la probabilité que le diamètre d'un comprimé soit supérieur à 8,1 mm. Dans ce contexte et sachant que la normalité de la distribution est acquise, quelle hypothèse doit-il remettre en cause dans l'énoncé ? Que vaut-elle ?
5) Un comprimé dont le diamètre est supérieur à 8,1 mm est considéré défectueux avec une probabilité fixée à 1 %. La plaquette thermoformée est constituée de 6 comprimés de diamètres indépendants et identiquement distribués. La présence dans une plaquette d'au moins un comprimé défectueux rend la plaquette défectueuse. Si la variable aléatoire X est le nombre de comprimés défectueux dans une plaquette parmi les 6, quelle est la probabilité qu'une plaquette soit défectueuse ?

Réponses

1) Quel est le lien entre les variables aléatoires E et D ? Justifiez la loi de D en précisant ses paramètres caractéristiques. Donnez une représentation graphique de E et de D.

2) Un premier contrôleur affirme que σ est égal à 0,02 mm et que tous les comprimés dont le diamètre n'est pas compris entre 7,97 mm et 8,03 mm doivent être rejetés. Donnez une représentation graphique de cette affirmation en utilisant les densités de probabilités de E et D. En supposant cette affirmation exacte, quelle est la proportion de comprimés rejetés ?

3) Un deuxième contrôleur affirme que, pour optimiser la phase de cloquage, la probabilité que le diamètre d'un comprimé soit supérieur à 8,1 mm doit être de 1 %. Dans ce contexte, déterminez σ.

4) À la vue des résultats, un troisième contrôleur confirme la valeur de σ du premier contrôleur ainsi que la probabilité de 1 % du deuxième contrôleur sur la probabilité que le diamètre d'un comprimé soit supérieur à 8,1 mm. Dans ce contexte et sachant que la normalité de la distribution est acquise, quelle hypothèse doit-il remettre en cause dans l'énoncé ? Que vaut-elle ?

5) Un comprimé dont le diamètre est supérieur à 8,1 mm est considéré défectueux avec une probabilité fixée à 1 %. La plaquette thermoformée est constituée de 6 comprimés de diamètres indépendants et identiquement distribués. La présence dans une plaquette d'au moins un comprimé défectueux rend la plaquette défectueuse. Si la variable aléatoire X est le nombre de comprimés défectueux dans une plaquette parmi les 6, quelle est la probabilité qu'une plaquette soit défectueuse ?

Exercice 3 – (Les trois parties A, B et C sont indépendantes)
Le diabète gestationnel (ou « diabète de grossesse ») apparaît pour la première fois lors de la grossesse et disparaît généralement après l'accouchement. Ce type de diabète peut entraîner des complications lors de la grossesse avec des conséquences pour la future maman ou pour le bébé. Afin d'étudier le risque de récidive d'apparition d'un diabète gestationnel au cours de la grossesse suivante, la glycémie est mesurée chez des femmes enceintes ayant eu un diabète gestationnel lors d'une précédente grossesse. La glycémie pour la femme n° i est mesurée en début de grossesse ($x_{0,i}$ en g/L) puis à 28 semaines de grossesse ($x_{1,i}$ en g/L). L'étude porte sur les différences des glycémies mesurées, c'est-à-dire $y_i = x_{1,i} - x_{0,i}$.

Partie A
Les différences, y_i, des mesures de la glycémie pour un premier échantillon sont présentées ci-dessous :

0,5 –0,1 0,4 0,9 –0,5 0,6 0,7 0,0

Quelle est la taille de cet échantillon ? Quand cela est envisageable, calculez la valeur modale, la médiane, la moyenne et l'étendue de ces mesures, la variance et l'écart-type de ces résultats. N'oubliez pas de préciser les unités.

Réponses

Partie B – Afin d'affiner les résultats de l'étude de la partie A et en s'appuyant sur le même protocole, une 2e enquête chez 29 femmes est mise en place. L'analyse statistique indique que la moyenne des différences des glycémies est de 0,3 et que la somme des différences des glycémies au carré est de 12,69 sachant que les différences des glycémies sont mesurées en g/L.

1) Calculez la variance des différences de glycémie en précisant l'unité.

2) Après avoir, si besoin est, fait les hypothèses de loi nécessaires, quel est l'intervalle de confiance au risque 5 % de la différence moyenne des glycémies ?

3) Afin de tester l'absence ou pas de diabète gestationnel, vérifiez pour un risque α fixé à 5 % si la moyenne des différences des glycémies est nulle ou pas.

4) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (resp. NS) quand le test est significatif (resp. non significatif). Donnez un encadrement du degré de signification.

α	$\forall \alpha > 20$ %	20 %	10 %	5 %	2 %	1 %	$\forall \alpha < 1$ %
Test S ou NS ?							

Réponses

1) Calculez la variance des différences de glycémie en précisant l'unité.

2) Après avoir, si besoin est, fait les hypothèses de loi nécessaires, quel est l'intervalle de confiance au risque 5 % de la différence moyenne des glycémies ?

3) Afin de tester l'absence ou pas de diabète gestationnel, vérifiez pour un risque α fixé à 5 % si la moyenne des différences des glycémies est nulle ou pas.

4) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (resp. NS) quand le test est significatif (resp. non significatif).

α	$\forall \alpha$ > 20 %	20 %	10 %	5 %	2 %	1 %	$\forall \alpha$ < 1 %
Test S ou NS ?							

Partie C – Afin d'affiner cette étude sur les femmes à risque dans le développement d'un diabète gestationnel, une nouvelle étude est menée sur un échantillon de femmes enceintes ayant des antécédents familiaux de diabète (noté DF) et sur un échantillon de femmes enceintes n'ayant pas d'antécédent familial de diabète (noté PDF). La différence des glycémies est mesurée pour chaque femme des deux échantillons. On se demande si le risque de diabète gestationnel est plus élevé en présence d'antécédents familiaux.
Voici 4 sorties de logiciel R adaptées ou pas à cette étude.

Sortie de logiciel n° 1	Sortie de logiciel n° 2
F test to compare two variances data: PDF and DF F = 0.4719, num df = 13, denom df = ?? p-value = 0.1813 alternative hypothesis: true ratio of variances is not equal to 1	F test to compare two variances data: DF and DF F = 1, num df = 15, denom df = 15, p-value = 1 alternative hypothesis: true ratio of variances is not equal to 1

Sortie de logiciel n° 3	Sortie de logiciel n° 4
Welch Two Sample t-test data: DF and PDF t = 2.7241, df = 26.611, p-value = 0.005623 alternative hypothesis: true difference in means is greater than 0	Two Sample t-test data: PDF and DF t = -2.6575, df = ?? , p-value = 0.00643 alternative hypothesis: true difference in means is less than 0

1) Que vaut le ?? de la sortie n° 1 ? Quelles sont les tailles des deux échantillons ?
2) Parmi les sorties n° 1 et n° 2, laquelle vous semble la plus adaptée à la démarche statistique ? N'oubliez pas de déterminer l'hypothèse nulle, l'alternative, la valeur du critère de test et la conclusion pour un risque fixé à 5 % pour la sortie de logiciel vous semblant la plus adaptée.
3) Que vaut le ?? de la sortie n° 4 ? Parmi les sorties n° 3 et n° 4, laquelle vous semble la plus adaptée à la démarche statistique ? Expliquez la conséquence de votre choix sur les ddl. N'oubliez pas de déterminer l'hypothèse nulle, l'alternative, la valeur du critère de test et la conclusion pour un risque fixé à 5 % pour la sortie de logiciel vous semblant la plus adaptée. Quelle est la conclusion de cette étude ?
4) Si les sorties de logiciel n° 1 et 2 n'étaient pas fournies, auriez-vous favorisé la sortie n° 3 ou la sortie n° 4 ? Pourquoi ?

Réponses

1) Que vaut le ?? de la sortie n° 1 ? Quelles sont les tailles des deux échantillons ? (Attention à vos notations)

2) Parmi les sorties n° 1 et n° 2, laquelle vous semble la plus adaptée à la démarche statistique ? N'oubliez pas de déterminer l'hypothèse nulle, l'alternative, la valeur du critère de test et la conclusion pour un risque fixé à 5 % pour la sortie de logiciel vous semblant la plus adaptée.

3) Que vaut le ?? de la sortie n° 4 ? Parmi les sorties n° 3 et n° 4, laquelle vous semble la plus adaptée à la démarche statistique ? Expliquez la conséquence de votre choix sur les ddl. N'oubliez pas de déterminer l'hypothèse nulle, l'alternative, la valeur du critère de test et la conclusion pour un risque fixé à 5 % pour la sortie de logiciel vous semblant la plus adaptée. Quelle est la conclusion de cette étude ?

4) Si les sorties de logiciel n° 1 et 2 n'étaient pas fournies, auriez-vous favorisé la sortie n° 3 ou la sortie n° 4 ? Pourquoi ?

Amicale Paris Sciences

Licence Sciences Biomédicales
2017-2018

Session 1 – Semestre 4
Correction

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

Pour vérifier les calculs.
L2 Biostats - 2018.

Exercice 1

1 - $X_i = \begin{cases} 1 & \text{si absence de spores} \quad \pi \\ 0 & \text{si présence de spores} \quad 1-\pi \end{cases}$ $X_1 \sim \mathcal{B}(1,\pi)$

2 - $Y = 0, 1, \ldots, n$. $Y = \sum_{i=1}^{n} X_i$ où $X_i = 1$ si absence de spores, 0 sinon.

3 - $(X_i$ iid $+ X_i \sim \mathcal{B}(1,\pi))\ \forall i$ $Y \sim \mathcal{B}(n;\pi)$

4 - $Z \sim \mathcal{P}(\lambda)$ $P(Z=0) = e^{-\lambda}$ (tube stérile) $P(Z>0) = 1 - e^{-\lambda}$ (non stérile)

5 - $X_1 \sim \mathcal{B}(1; e^{-\lambda})$ $Y \sim \mathcal{B}(n; e^{-\lambda})$.

6 - $n = 10$ $\pi = 0,6$ $E(Y) = 6$ tubes $E(Z) \simeq 0,51$ nb de spores

Exercice 2

1 - $D = 8 + E$ $D \sim \mathcal{N}(8;\sigma)$

2 - $\sigma = 0,02$ mm. $P(D<7,97) + P(D>8,03) = P(E<-0,03) + P(E>0,03) = 0,1336$

3 - $P(D>8,1) = 1\%$. $\sigma \simeq 0,0429$ mm.

4 - On rejette que le diamètre moyen soit de 8 mm. $\mu = 8,053$ mm.

5 - $X \sim \mathcal{B}(6; 0,01)$ $P(X \geq 1) = 0,0585$.

Exercice 3 - Partie A

$n = 8$ pas de valeur modale $me = 0,45$ g/L $\bar{y} = 0,3125$ g/L $e = 1,4$ g/L

$s^2 \simeq 0,22125$ $(g/L)^2$ $s \simeq 0,4704$ g/L

Exercice 3 - Partie B

1 - $n = 29$ $s^2 = 0,36$ $(g/L)^2$

2 - $\forall i$ $Y_i \sim \mathcal{N}(\mu;\sigma)$ $]0,071 ; 0,528[$ en g/L

3 - $H_0: \mu = 0$ $H_1: \mu \neq 0$. $\forall i$ $Y_i \sim \mathcal{N}(\mu;\sigma)$ $\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim$ Student à $(n-1)$ ddl.

$t_{obs} = 2,6926 > 2,0484$ ⇒ rejet de H_0 pour $\alpha = 5\%$. (S).
⇒ la moyenne des ≠ de glycémie diffère significativement de 0.

4 - $\alpha_S < 5\%$ $1\% < \alpha_S < 2\%$.

Exercice 4 - Partie C

(cf. TD)

Amicale Paris Sciences

Licence Sciences Biomédicales
2018-2019

Session 1 – Semestre 4
Sujet

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

UNIVERSITÉ PARIS DESCARTES

L2

Année 2018-2019
4e semestre
1re session

Faculté des Sciences Fondamentales et Biomédicales
Faculté de Pharmacie de Paris

Ex. 1 :
Ex. 2 :
Ex. 3 – A :
Ex. 3 – B :
Ex. 3 – C :

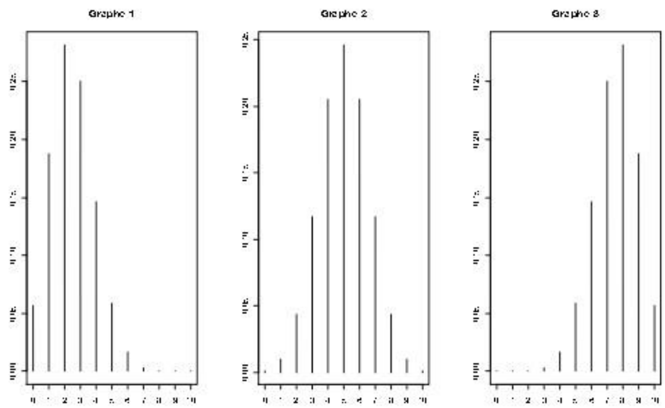
Total :

UE 4.08 — ÉPREUVE DE BIOSTATISTIQUE

Tout signe distinctif porté sur la copie pouvant indiquer sa provenance constitue une fraude. Aucun document n'est autorisé. La calculette **non programmable** est autorisée. **Il est interdit de désagrafer ce cahier (9 pages numérotées et 4 pages d'annexes).**

Exercice 1 – La neuropathie périphérique, atteinte du système nerveux, est l'une des complications du diabète. Lorsque le taux de sucre dans le sang demeure trop élevé sur une longue période de temps, les nerfs des membres inférieurs peuvent être endommagés. Les symptômes aux extrémités du corps peuvent être des engourdissements, des sensations de brûlure ou des pertes de sensibilité à la douleur et à la température. Un service de diabétologie d'un hôpital s'intéresse à la présence de cette pathologie chez ses patients diabétiques.

1) La variable aléatoire Z_1 modélise le résultat du premier patient diabétique. Quelles sont les réalisations de Z_1 ? Quelle est la loi de Z_1 ?
2) Le diabétologue interroge ainsi *n* patients indépendants les uns des autres. La variable aléatoire Y modélise ce décompte de patients diabétiques atteints par cette pathologie parmi les *n*. Quelle est la loi de Y ?
3) Sachant que la probabilité qu'aucun patient diabétique ne soit atteint par cette pathologie parmi les *n* est égale à la probabilité que tous les patients diabétiques soient atteints par cette pathologie parmi les *n*, trouvez la valeur du 2e paramètre de la loi de Y.
4) Un des 3 diagrammes ci-dessous est le diagramme en bâtons de la loi de Y. Lequel ?



5) Sachant que la valeur modale coïncide avec le nombre moyen de patients diabétiques atteints par cette pathologie parmi les *n*, trouvez la valeur du 1er paramètre de la loi de Y.
6) Est-il possible d'approximer la loi de Y ? Si oui, par quelle loi en précisant ses paramètres ?

Réponses (toutes vos réponses doivent être justifiées)

1) La variable aléatoire Z_1 modélise le résultat du premier patient diabétique. Quelles sont les réalisations de Z_1 ? Quelle est la loi de Z_1 ?

2) Le diabétologue interroge ainsi n patients indépendants les uns des autres. La variable aléatoire Y modélise ce décompte de patients diabétiques atteints par cette pathologie parmi les n. Quelle est la loi de Y ?

3) Sachant que la probabilité qu'aucun patient diabétique ne soit atteint par cette pathologie parmi les n est égale à la probabilité que tous les patients diabétiques soient atteints par cette pathologie parmi les n, trouvez la valeur du 2^{e} paramètre de la loi de Y.

4) Un des 3 diagrammes ci-dessus est le diagramme en bâtons de la loi de Y. Lequel ?

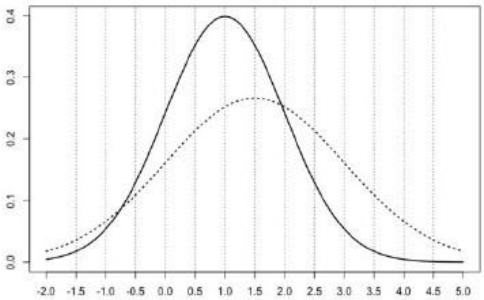
5) Sachant que la valeur modale coïncide avec le nombre moyen de patients diabétiques atteints par cette pathologie parmi les n, trouvez la valeur du 1er paramètre de la loi de Y.

6) Est-il possible d'approximer la loi de Y ? Si oui, par quelle loi en précisant ses paramètres ?

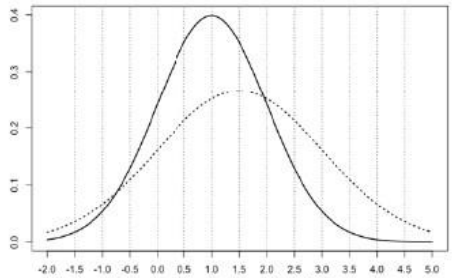
Exercice 2 – La glycémie, taux de sucre dans le sang, varie en fonction des apports nutritionnels et des dépenses corporelles. Elle est contrôlée par l'insuline, hormone déficiente ou absente en cas de diabète. Un lecteur de glycémie permet aux patients diabétiques de contrôler leur maladie et d'adapter eux-mêmes leur régime, leur activité physique et leur traitement. Un fabriquant de lecteur de glycémie affirme que la mesure de la glycémie (taux de glucose dans le sang) en g/L à jeun :

- chez des patients normoglycémiques, variable aléatoire notée G_N, suit une loi normale d'espérance μ_N et d'écart-type σ_N ;
- chez des patients atteints d'hyperglycémie diabétique, variable aléatoire notée G_{H+}, suit une loi normale d'espérance μ_{H+} et d'écart-type σ_{H+} ;
- chez des patients atteints d'hypoglycémie diabétique, variable aléatoire notée G_{H-}, suit une loi normale d'espérance μ_{H-} et d'écart-type σ_{H-}.

1) Les densités de probabilités de G_N et G_{H+} sont représentées ci-dessous. En justifiant vos réponses, légendez le graphique et en déduire μ_N et μ_{H+}. Le fabriquant ne sait plus à qui associer les valeurs des écart-types égales à 1 g/L et l'autre égal à 1,5 g/L. En déduire σ_N et σ_{H+}.



2) Calculez les probabilités que les mesures de la glycémie soient comprises entre leur espérance ± leur écart-type pour les patients normoglycémiques puis pour les patients atteints d'hyperglycémie diabétique. Comparez les deux probabilités puis commentez. Hachurez sur le graphique ci-dessous les zones correspondant aux deux probabilités.



3) 12,1 % des patients atteints d'hypoglycémie diabétique ont une glycémie supérieure à 0,634 g/L. Parmi les deux propositions ci-dessous, l'une est fausse. Pourquoi ?

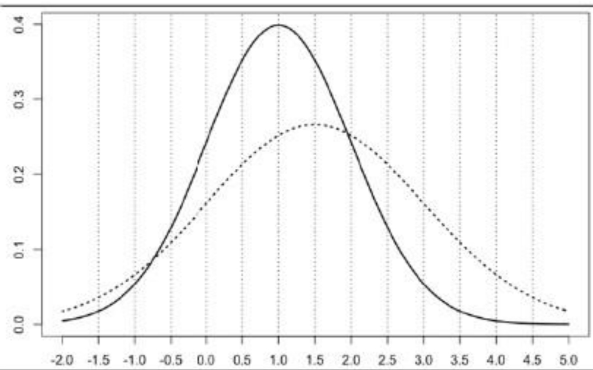
Proposition 1 : « 50 % des patients atteints d'hypoglycémie diabétique ont une glycémie supérieure à 0,4 g/L »

Proposition 2 : « 50 % des patients atteints d'hypoglycémie diabétique ont une glycémie inférieure à 0,7 g/L ».

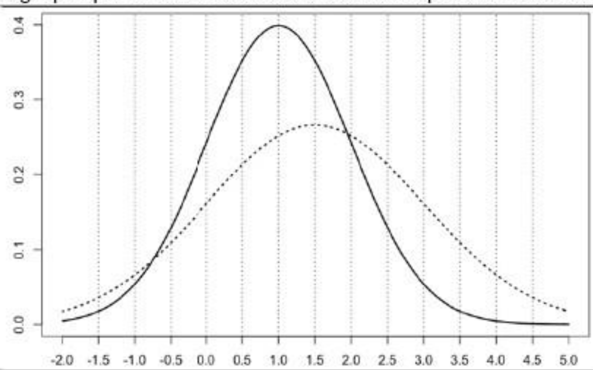
4) On suppose exacte l'autre proposition de la question 3, représentez ces résultats sur la densité de probabilité de G_{H-}.
5) À partir des questions précédentes, déduisez μ_{H-} et σ_{H-}.
6) Que pensez-vous du modèle proposé par le fabriquant ?

Réponses (toutes vos réponses doivent être justifiées)

1) Les densités de probabilités de G_N et G_{H+} sont représentées ci-dessous. En justifiant vos réponses, légendez le graphique et en déduire μ_N et μ_{H+}. Le fabriquant ne sait plus à qui associer les valeurs des écart-types égales à 1 g/L et l'autre égal à 1,5 g/L. En déduire σ_N et σ_{H+}.



2) Calculez les probabilités que les mesures de la glycémie soient comprises entre leur espérance ± leur écart-type pour les patients normoglycémiques puis pour les patients atteints d'hyperglycémie diabétiques. Comparez les deux probabilités puis commentez. Hachurez sur le graphique ci-dessous les zones correspondant aux deux probabilités.



3) 12,1 % des patients atteints d'hypoglycémie diabétique ont une glycémie supérieure à 0,634 g/L. Parmi les deux propositions ci-dessous, l'une est fausse. Pourquoi ?

Proposition 1 : « 50 % des patients atteints d'hypoglycémie diabétique ont une glycémie supérieure à 0,4 g/L »

Proposition 2 : « 50 % des patients atteints d'hypoglycémie diabétique ont une glycémie inférieure à 0,7 g/L ».

4) On suppose exacte l'autre proposition de la question 3, représentez ces résultats sur la densité de probabilité de G_{H-}.

5) À partir des questions précédentes, déduisez μ_{H-} et σ_{H-}.

7) Que pensez-vous du modèle proposé par le fabriquant ?

Exercice 3 – (Les trois parties A, B et C sont indépendantes)
L'hypoglycémie est une complication fréquente du diabète. La glycémie du diabétique est, par définition, variable puisque le diabétique ne peut apporter la quantité d'insuline nécessaire pour contrôler sa glycémie. Le diabétique ayant un taux élevé de sucre dans le sang, l'hypoglycémie est généralement lié aux traitements destinés à faire baisser ces taux de sucre.

Partie A – En cas d'hypoglycémie, une première stratégie thérapeutique est de faire consommer aux patients du sucre (un sucre pour 20 kg de masse corporelle). Les mesures de glycémie en g/L après consommation de sucre, notées x_i, pour un premier échantillon sont les suivantes :

1,5 1,1 1,3 1,4 0,8 1,1 1,2

Quelle est la taille de cet échantillon ? Quand cela est envisageable, calculez la valeur modale, la médiane, la moyenne et l'étendue de ces mesures, la variance et l'écart-type de ces résultats. N'oubliez pas de préciser les unités.

Réponses (toutes vos réponses doivent être justifiées)

Partie B – Un service hospitalier de diabétologie teste la même stratégie thérapeutique auprès de ses patients diabétiques manifestant une hypoglycémie lors des séances d'éducation thérapeutique. Les résultats obtenus sur un échantillon de taille n sont les suivants :

— La somme des mesures de glycémie après consommation de sucre vaut 27,5 g/L ;
— La somme des mesures centrées au carré de glycémie après consommation de sucre vaut 1,5 (g/L)^2 ;
— L'écart-type des mesures de glycémie après consommation de sucre vaut 0,25 g/L.

1) Quelle est la taille de l'échantillon ? Calculez la moyenne des mesures de glycémie après consommation de sucre.
2) Après avoir, si besoin est, fait les hypothèses de loi nécessaires, quel est l'intervalle de confiance au risque α fixé à 5 % de la glycémie moyenne après consommation de sucre ?
3) Le service hospitalier considère qu'une glycémie « normale » est de 1 g/L. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifier pour un risque α fixé à 5 % si la moyenne des glycémies après consommation de sucre reste « anormale ».
4) Le degré de signification du test de la question 3 est-il plus petit ou plus grand que 5 % ?
5) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification.

α	$\forall \alpha > 20$ %	20 %	10 %	5 %	2 %	1 %	$\forall \alpha < 1$ %
Test S ou NS ?							

Réponses (toutes vos réponses doivent être justifiées)

1) Quelle est la taille de l'échantillon ? Calculez la moyenne des mesures de glycémie après consommation de sucre.

2) Après avoir, si besoin est, fait les hypothèses de loi nécessaires, quel est l'intervalle de confiance au risque α fixé à 5 % de la glycémie moyenne après consommation de sucre ?

3) Le service hospitalier considère qu'une glycémie « normale » est de 1 g/L. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifier pour un risque α fixé à 5 % si la moyenne des glycémies après consommation de sucre reste « anormale ».

4) Le degré de signification du test de la question 3 est-il plus petit ou plus grand que 5 % ?

5) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification.

α	$\forall\alpha > 20$ %	20 %	10 %	5 %	2 %	1 %	$\forall\alpha < 1$ %
Test S ou NS ?							

Partie C – En cas d'hypoglycémie, une première stratégie thérapeutique est de faire consommer aux patients du sucre (un sucre pour 20 kg de masse corporelle). Une deuxième stratégie thérapeutique est une injection de glucagon. Afin de comparer ces deux stratégies, une nouvelle étude est menée sur un échantillon de patients ayant consommé du sucre après hypoglycémie (noté S) et sur un échantillon de patients ayant eu une injection de glucagon après hypoglycémie (noté G). Les glycémies des patients sont mesurées pour chacun des patients des deux échantillons. La question est de savoir si les deux stratégies thérapeutiques donnent des résultats différents.
Voici des sorties de logiciel R adaptées, ou pas, à cette étude.

Sortie de logiciel n° 1	Sortie de logiciel n° 2
F test to compare two variances data: S and G F = 8.1921, num df = 10, denom df = 11, p-value = 0.9991 alternative hypothesis: true ratio of variances is less than 1	F test to compare two variances data: S and G F = ?? , num df = 10, denom df = 11, p-value = 0.001723 alternative hypothesis: true ratio of variances is not equal to 1

Sortie de logiciel n° 3	Sortie de logiciel n° 4
F test to compare two variances data: G and S F = ?? , num df = ??, denom df = ??, p-value = ?? alternative hypothesis: true ratio of variances is not equal to 1	Welch Two Sample t-test data: S and G t = 0.79732, df = 12.224, p-value = 0.44 alternative hypothesis: true difference in means is not equal to 0

Sortie de logiciel n° 5	Sortie de logiciel n° 6
Welch Two Sample t-test data: G and S t = -0.79732, df = ?? , p-value = ?? alternative hypothesis: true difference in means is not equal to 0	Two Sample t-test data: G and S t = -0.82631, df = 21, p-value = 0.4179 alternative hypothesis: true difference in means is not equal to 0

1) Que vaut le « ?? » de la sortie n° 2 ?
2) Que valent les « ?? » de la sortie n° 3 ?
3) Quelles sont les tailles des deux échantillons ?
4) Pour la sortie n° 2, quelles sont les hypothèses nulle et alternative et la conclusion de ce test ?
5) Que valent les « ?? » de la sortie n° 5 ?
6) Parmi les sorties n° 4, 5 ou 6, laquelle (ou lesquelles) vous semble(nt) la (ou les) plus adaptée(s) à la démarche statistique pour répondre à la question sur les deux stratégies thérapeutiques ? N'oubliez pas de déterminer l'hypothèse nulle, l'alternative, la valeur du critère de test et la conclusion pour un risque fixé à 5 %.
7) Quelle est la conclusion de cette étude ?

Réponses (toutes vos réponses doivent être justifiées)

1) Que vaut le « ?? » de la sortie n° 2 ?

2) Que valent les « ?? » de la sortie n° 3 ?

3) Quelles sont les tailles des deux échantillons ?

4) Pour la sortie n° 2, quelles sont les hypothèses nulle et alternative et la conclusion de ce test ?

5) Que valent les « ?? » de la sortie n° 5 ?

6) Parmi les sorties n° 4, 5 ou 6, laquelle (ou lesquelles) vous semble(nt) la (ou les) plus adaptée(s) à la démarche statistique pour répondre à la question sur les deux stratégies thérapeutiques ? N'oubliez pas de déterminer l'hypothèse nulle, l'alternative, la valeur du critère de test et la conclusion pour un risque fixé à 5 %.

7) Quelle est la conclusion de cette étude ?

Amicale Paris Sciences

Licence Sciences Biomédicales
2018-2019

Session 1 – Semestre 4
Correction

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

Pour vérifier vos calculs
L2 - Biostat. 2015

Exercice 1

1 - $Z_i = \begin{cases} 1 & \text{si neuropathie} \quad \pi \\ 0 & \text{sinon} \quad 1-\pi \end{cases}$ $\qquad Z_i \sim \mathcal{B}(1,\pi)$

2 - $Y = \sum_{i=1}^{n} Z_i \sim \mathcal{B}(n ; \pi)$ car les Z_i sont iid + $Z_i \sim \mathcal{B}(1,\pi)$ $\forall i$

3 - $\pi = 1/2$

4 - Graphe 2

5 - $n = 10$

6 - limite de l'approximation gaussienne : $Y \sim \mathcal{N}(5 ; \sqrt{2{,}5})$

Exercice 2

1 - trait plein : ddp de G_N $\quad \mu_N = 1$ g/L $\quad \sigma_N = 1$ g/L

trait pointillé : ddp de G_{H+} $\quad \mu_{H+} = 1{,}5$ g/L $\quad \sigma_{H+} = 1{,}5$ g/L

2 - $P(0 \leq G_N \leq 2) = 0{,}6826$
$P(0 \leq G_{H+} \leq 3) = 0{,}6826$ $\Big\}$ égalité

3 - Proposition 2 impossible.

5 - $\mu_{H-} = 0{,}4$ g/L $\qquad \sigma_{H-} \simeq 0{,}2$ g/L

6 - Modélisation via des valeurs négatives possibles $\Rightarrow$ oups !

Exercice 3 - Partie A

$n = 7 \qquad m_0 = 1{,}1$ g/L $\qquad me = 1{,}2$ g/L $\qquad \bar{x} = 1{,}2$ g/L

$e = 0{,}7$ g/L $\qquad s^2 \simeq 0{,}053$ (g/L)^2 $\qquad s = 0{,}23$ g/L.

Exercice 3 - Partie B

1 - $n = 25 \qquad \bar{x} = 1{,}1$ g/L

2 - $\forall i \; X_i \sim \mathcal{N}(\mu ; \sigma)$ $\qquad] \, 0{,}9968 ; 1{,}2032 \, [$ (en g/L)

3 - $H_0 : \mu = 1 \qquad H_1 : \mu \neq 1$ $\qquad$ conditions : $\forall i \; X_i \sim \mathcal{N}(\mu ; \sigma)$

$t_{obs} = 2 \qquad t_{1-\alpha/2} = 2{,}0639$ pour $\alpha = 5\%$ à 24 ddl.

$\Rightarrow$ Non rejet de H_0 pour $\alpha = 5\%$ $\Rightarrow$ la consommation ne reste pas "anormale"

4 - $ds > 5\%$

5 - $5\% < ds < 10\%$.

Exercice 3 - Partie C

1 - $f = 8,1921$

2 - $F = 0,1221$ numdf = 11 denomdf = 10 pvalue = 0,001723

3 - $n_G = 12$ $n_S = 11$

4 - $H_0 : \sigma_S^2 = \sigma_G^2$ $H_1 : \sigma_S^2 \neq \sigma_G^2$

pvalue = 0,1723 % < 5% ⇒ Rejet de H_0 pour $\alpha = 5\%$

⇒ "Pas d'égalité des variances"

5 - $df = 12,224$ pvalue = 0,44

6 - $H_0 : \mu_G = \mu_S$ $H_1 : \mu_G \neq \mu_S$.

Sorties 4 et 5 $|t_{obs}| = 0,79732$.

pvalue = 44% > 5% ⇒ Non rejet de H_0 pour $\alpha = 5\%$.

⇒ "pas de mise en évidence de différence"

7 - Pas de mise en évidence de différence entre les 2 stratégies thérapeutiques

Amicale Paris Sciences

Licence Sciences Biomédicales
2020-2021

Session 1 – Semestre 4
Sujet

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

PHARMACIE
Santé
Université de Paris

Notes
Ex. 1 :
Ex. 2 :
Ex. 3 – A :
Ex. 3 – B :
Ex. 3 – C :
Total :

L2

Année 2020 – 2021
Premier semestre
Première session

UFR des Sciences Fondamentales et Biomédicales
Faculté de Pharmacie de Paris

SA04M020

ÉPREUVE DE BIOSTATISTIQUE

Tout signe distinctif porté sur la copie pouvant indiquer sa provenance constitue une fraude.
Aucun document n'est autorisé. La calculette **non programmable** est autorisée.
Il est interdit de désagrafer ce cahier (12 pages numérotées).

Exercice 1 – Une étude menée en 2020 à très grande échelle et publiée en janvier dans « *Journal of internal medicine* » portait sur la perte de l'odorat dans le cadre d'une forme légère de Covid-19 et la non-récupération de la pleine capacité olfactive après soixante jours. Dans l'ensemble de cet exercice, les patients sont tous atteints d'une forme légère de Covid-19 avec une perte d'odorat. On note π la probabilité de non-récupération de leur pleine capacité olfactive après 60 jours. On interroge un premier patient sur la non-récupération de sa pleine capacité olfactive.

1) Si la variable aléatoire modélisant le résultat de ce premier patient est notée Y_1, quelles sont ses réalisations ? Quelle est la loi de Y_1 ?
2) Afin d'interroger $n-1$ autres patients, on s'assure qu'ils n'ont ni lien de parenté, ni lien social. On considère que la probabilité de non-récupération de leur pleine capacité olfactive après 60 jours est la même pour tous les individus. Pourquoi ces deux informations sont-elles importantes ?
3) On interroge les $n-1$ autres patients sur la non-récupération de leur pleine capacité olfactive. On note Y_i le résultat du patient i. Soit la variable aléatoire X modélisant le décompte total de patients n'ayant pas récupéré leur pleine capacité olfactive après 60 jours parmi les n patients. Quelle est la loi de X ? Quels sont son moment d'ordre 1, son moment d'ordre 2 et sa variance ?
4) Soit la variable aléatoire $Z = \sum_{i=1}^{n}(1-Y_i)$, que modélise-t-elle ? Quelle est la loi de Z ? Quel est le lien entre les deux variables aléatoires X et Z ?
5) Si l'espérance de la somme de X et Z est égale à 20 et $E(Z) = 3E(X)$, quels sont les paramètres des lois de X et Z ?
6) Que pensez-vous de l'affirmation « la somme de X et Z n'est pas aléatoire » ?

Réponses (toutes vos réponses doivent être justifiées)

1) Si la variable aléatoire modélisant le résultat de ce premier patient est notée Y_1, quelles sont ses réalisations ? Quelle est la loi de Y_1 ?

2) Afin d'interroger $n-1$ autres patients, on s'assure qu'ils n'ont ni lien de parenté, ni lien social. On considère que la probabilité de non-récupération de leur pleine capacité olfactive après 60 jours est la même pour tous les individus. Pourquoi ces deux informations sont-elles importantes ?

3) On interroge les $n-1$ autres patients sur la non-récupération de leur pleine capacité olfactive. On note Y_i le résultat du patient i. Soit la variable aléatoire X modélisant le décompte total de patients n'ayant pas récupéré leur pleine capacité olfactive après 60 jours parmi les n patients. Quelle est la loi de X ? Quels sont son moment d'ordre 1, son moment d'ordre 2 et sa variance ?

4) Soit la variable aléatoire $Z = \sum_{i=1}^{n}(1 - Y_i)$, que modélise-t-elle ? Quelle est la loi de Z ? Quel est le lien entre les deux variables aléatoires X et Z ?

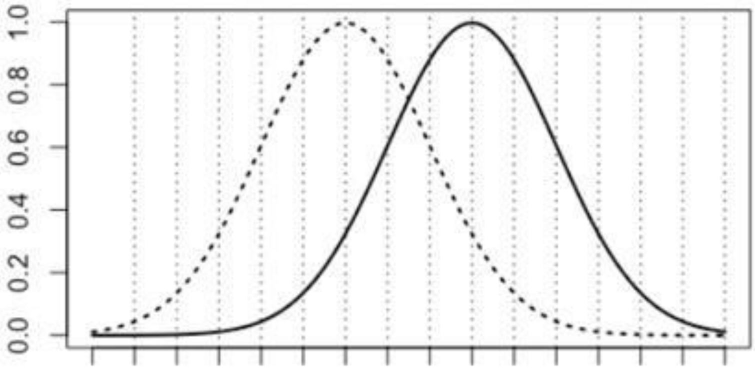
5) Si l'espérance de la somme de X et Z est égale à 20 et $E(Z) = 3E(X)$, quels sont les paramètres des lois de X et Z ?

6) Que pensez-vous de l'affirmation « la somme de X et Z n'est pas aléatoire » ?

Exercice 2 – De nombreux mois après avoir été atteints par la Covid-19, certains individus conservent une capacité respiratoire limitée. La capacité respiratoire mesurée par la quantité d'air (en litres, noté L) rejetée lors d'une expiration forcée est modélisée par :

- une loi normale notée C_S d'espérance μ_S L et d'écart-type σ_S L pour un individu ayant retrouvé une capacité respiratoire normale ;
- une loi normale notée C_{NS} d'espérance μ_{NS} L et d'écart-type σ_{NS} L pour un individu ayant une capacité respiratoire limitée quelques mois après avoir été atteint par la Covid-19.

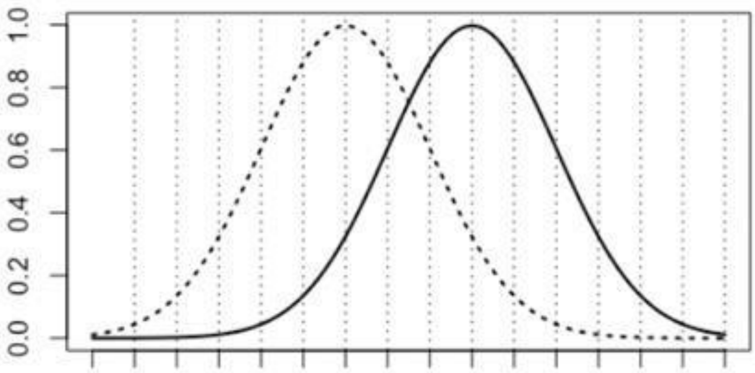
Les densités de probabilité de C_S et de C_{NS} sont représentées ci-dessous.



1) En justifiant vos réponses, légendez le graphique. Positionnez μ_S et μ_{NS}.
2) Que pensez-vous de l'affirmation « les écarts-types σ_S et σ_{NS} sont égaux » ?
3) Si le maximum de la densité de probabilité d'une loi normale réduite est de 0,4, pensez-vous que les écarts-types σ_S et σ_{NS} sont plus petits ou plus grands que 1 ?
4) En utilisant le résultat de la question 2, uniquement l'annexe B et sachant les trois affirmations suivantes :
   - 50 % des individus ayant retrouvé une capacité respiratoire normale ont une capacité respiratoire inférieure à 1,8 L ;
   - 38,3 % des individus ayant retrouvé une capacité respiratoire normale ont une capacité respiratoire comprise entre 1,6 L et 2 L ;
   - 6,68 % des individus ayant une capacité respiratoire limitée quelques mois après avoir été atteints par la Covid-19 ont une capacité respiratoire supérieure à 1,8 L,

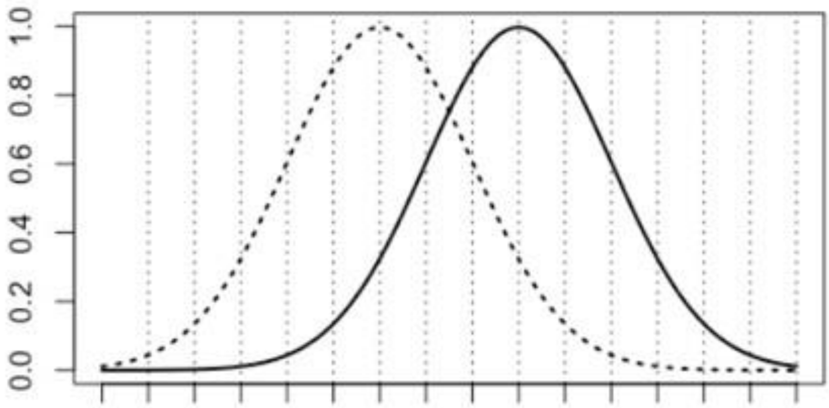
déduisez les espérances et les écarts-types de ces deux lois normales en précisant les unités.

5) Représentez sur le graphique ci-dessous les trois affirmations de la question 4.



Réponses (toutes vos réponses doivent être justifiées)

1) En justifiant vos réponses, légendez le graphique. Positionnez μ_S et μ_{NS}.



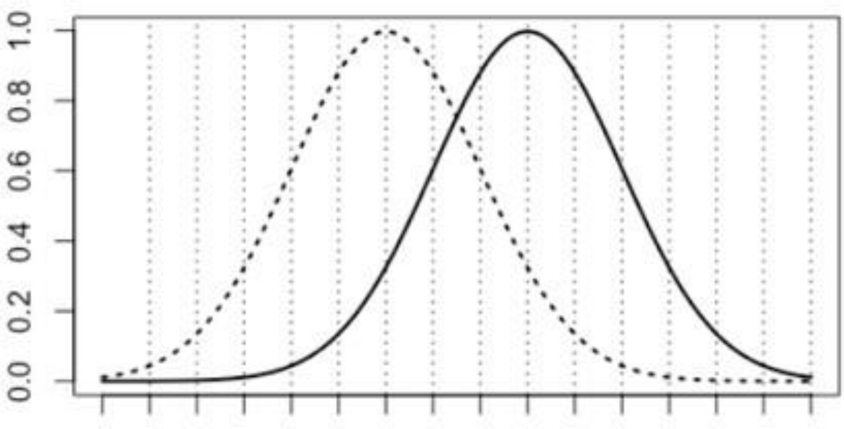
Pour la justification :

2) Que pensez-vous de l'affirmation « les écarts-types σ_S et σ_{NS} sont égaux » ?

3) Si le maximum de la densité de probabilité d'une loi normale réduite est de 0,4, pensez-vous que les écarts-types σ_S et σ_{NS} sont plus petits ou plus grands que 1 ?

4) En utilisant le résultat de la question 2, uniquement l'annexe B et sachant les trois affirmations suivantes : « 50 % des individus ayant retrouvé une capacité respiratoire normale ont une capacité respiratoire inférieure à 1,8 L », « 38,3 % des individus ayant retrouvé une capacité respiratoire normale ont une capacité respiratoire comprise entre 1,6 L et 2 L », « 6,68 % des individus ayant une capacité respiratoire limitée quelques mois après avoir été atteints par la Covid-19 ont une capacité respiratoire supérieure à 1,8 L », déduisez les espérances et les écarts-types de ces deux lois normales en précisant les unités.

5) Représentez sur le graphique ci-dessous les trois affirmations de la question 4.



Exercice 3 – (Les trois parties A, B et C sont indépendantes)
Ces trois parties portent sur le temps d'hospitalisation en soins intensifs de patients diagnostiqués Covid-19. Ces temps d'hospitalisation en soins intensifs (notés t_i) calculés en soustrayant la date et l'heure de sortie en soins intensifs avec la date et l'heure d'entrée en soins intensifs, ont été arrondis à une décimale et exprimés en jours.

Partie A – Dans un premier hôpital, les temps d'hospitalisation (en jours) d'un premier échantillon ont été recueillis :

12,2 9,8 14,1 11,9 14,1 12,9

Quelle est la taille de cet échantillon ? Quand cela est envisageable, calculez la valeur modale, la médiane, la moyenne, l'étendue, la variance et l'écart-type de ces résultats en précisant les unités.

Réponses (toutes vos réponses doivent être justifiées)

Partie B – Un deuxième hôpital fournit deux intervalles de confiance du temps moyen d'hospitalisation en soins intensifs de patients diagnostiqués Covid-19 en oubliant de préciser le risque de première espèce α :

IC1 : [10,56172 ; 11,43828]
IC2 : [10,46715 ; 11,53285]

1) Quelle hypothèse de loi a été nécessaire pour établir ces intervalles de confiance ?
2) Dans ce contexte, donnez la formule de l'intervalle de confiance du temps moyen d'hospitalisation.
3) Ce 2e hôpital indique que l'un des intervalles de confiance a été calculé pour un risque de première espèce de 5 % et l'autre pour un risque de 10 %. Associez ces deux risques aux deux intervalles de confiance IC1 et IC2 en justifiant vos choix.
4) En utilisant la propriété de symétrie de cet intervalle de confiance, quelle est la moyenne des temps d'hospitalisation ?
5) En utilisant les demi-largeurs de ces intervalles de confiance à 15 ddl, quels sont la taille de cet échantillon et l'écart-type des temps d'hospitalisation ?
6) Au niveau national et dans ce contexte, il est considéré que le temps d'hospitalisation standard est de 11,5 jours. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifiez pour un risque α fixé à 5 % si la moyenne des temps d'hospitalisation est standard.
7) Le degré de signification du test de la question 6 est-il plus petit ou plus grand que 5 % ?
8) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification. N'hésitez pas à regarder vos intervalles de confiance.

α	$\forall\alpha > 20$ %	20 %	10 %	5 %	2 %	1 %	$\forall\alpha < 1$ %
Test S ou NS ?							

Réponses (toutes vos réponses doivent être justifiées)

1) Quelle hypothèse de loi a été nécessaire pour établir ces intervalles de confiance ?

2) Dans ce contexte, donnez la formule de l'intervalle de confiance du temps moyen d'hospitalisation.

3) Ce 2[e] hôpital indique que l'un des intervalles de confiance a été calculé pour un risque de première espèce de 5 % et l'autre pour un risque de 10 %. Associez ces deux risques aux deux intervalles de confiance IC1 et IC2 en justifiant vos choix.

4) En utilisant la propriété de symétrie de cet intervalle de confiance, quelle est la moyenne des temps d'hospitalisation ?

5) En utilisant les demi-largeurs de ces intervalles de confiance à 15 ddl, quels sont la taille de cet échantillon et l'écart-type des temps d'hospitalisation ?

6) Au niveau national et dans ce contexte, il est considéré que le temps d'hospitalisation standard est de 11,5 jours. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifiez pour un risque α fixé à 5 % si la moyenne des temps d'hospitalisation est standard.

7) Le degré de signification du test de la question 6 est-il plus petit ou plus grand que 5 % ?

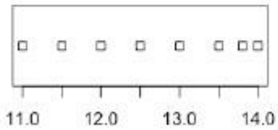
8) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification. N'hésitez pas à regarder vos intervalles de confiance.

α	$\forall\alpha > 20$ %	20 %	10 %	5 %	2 %	1 %	$\forall\alpha < 1$ %
Test S ou NS ?							

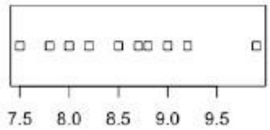
Partie C – En plus du temps d'hospitalisation, un troisième hôpital a recueilli le groupe sanguin des patients Covid-19. Leur affirmation est que les patients Covid-19 ayant le groupe sanguin A passent plus de jours en soins intensifs que les patients Covid-19 ayant le groupe sanguin B. Pour vérifier cette hypothèse, une analyse est menée sur le temps d'hospitalisation en soins intensifs de patients Covid-19 ayant le groupe sanguin A (noté A) ainsi que sur le temps d'hospitalisation en soins intensifs de patients Covid-19 ayant le groupe sanguin B (noté B).

1) À partir de la représentation graphique ci-dessous, que pensez-vous de l'affirmation de ce troisième hôpital ? Sachant qu'il n'y aucun ex-aequo dans les échantillons, quelles sont les tailles des deux échantillons ?

Pour le groupe sanguin A



Pour le groupe sanguin B



2) Afin de tester les temps moyens d'hospitalisation, quelles sont les hypothèses de loi nécessaires ?
3) Voici des sorties du logiciel R :

Sortie de logiciel n° 1	Sortie de logiciel n° 2
F test to compare two variances data: B and A F = 0.42417, num df = ??, denom df = ??, p-value = 0.2302 alternative hypothesis: true ratio of variances is not equal to 1	F test to compare two variances data: A and B F = ??, num df = ??, denom df = ??, p-value = 0.2302 alternative hypothesis: true ratio of variances is not equal to 1

Sortie de logiciel n° 3	Sortie de logiciel n° 4
F test to compare two variances data: A and B F = ??, num df = ??, denom df = ??, p-value = 0.8849 alternative hypothesis: true ratio of variances is less than 1	F test to compare two variances data: B and A F = 0.42417, num df = ??, denom df = ??, p-value = ?? alternative hypothesis: true ratio of variances is greater than 1

3.1) Pour la sortie n° 1, que valent les deux « ?? » ? Quelles sont les hypothèses nulle et alternative ?
3.2) Pour la sortie n° 2, que valent les trois « ?? » ?
3.3) Pour la sortie n° 3, que valent les trois « ?? » ? Quelles sont les hypothèses nulle et alternative ?
3.4) Pour la sortie n° 4, que valent les trois « ?? » ?
3.5) Parmi les sorties numérotées de 1 à 4, quelles sont celles qui vous semblent adaptées à la démarche statistique ? Quelle est votre conclusion pour un risque de première espèce α fixé à 5 % ?

4) Voici des sorties du logiciel R :

Sortie de logiciel n° 5	Sortie de logiciel n° 6
Two Sample t-test data: A and B t = 9.5573, df = ??, p-value = 5.138e-08 alternative hypothesis: true difference in means is not equal to 0	Two Sample t-test data: A and B t = ??, df = ??, p-value = 2.569e-08 alternative hypothesis: true difference in means is greater than 0

Sortie de logiciel n° 7	Sortie de logiciel n° 8
Two Sample t-test data: B and A t = ??, df = ??, p-value = 1 alternative hypothesis: true difference in means is greater than 0	Welch Two Sample t-test data: A and B t = 9.1103, df = 11.525, p-value = 1.316e-06 alternative hypothesis: true difference in means is not equal to 0

Sortie de logiciel n° 9	Sortie de logiciel n° 10
Welch Two Sample t-test data: A and B t = 9.1103, df = ??, p-value = 6.582e-07 alternative hypothesis: true difference in means is greater than 0	Welch Two Sample t-test data: B and A t = -9.1103, df = ??, p-value = 1 alternative hypothesis: true difference in means is greater than 0

4.1) Pour les sorties n° 5, 6 et 7, que vaut le « ?? » associé au df ?
4.2) Pour la sortie 6, que vaut le « ?? » associé au t ?
4.3) Pour la sortie 7, que vaut le « ?? » associé au t ?
4.4) Pour les sorties 9 et 10, que vaut le « ?? » associé au df ?
4.5) Quelles sont les sorties de logiciel qui ne supposent pas l'égalité des variances ?

5) Quelle est la sortie de logiciel adaptée à la question de ce troisième hôpital ? Précisez les hypothèses nulle et alternative. Quelle est votre conclusion pour un risque α fixé à 5 % ?

Réponses (toutes vos réponses doivent être justifiées)

1) À partir de la représentation graphique ci-dessus, que pensez-vous de l'affirmation de ce troisième hôpital ? Sachant qu'il n'y aucun ex-aequo dans les échantillons, quelles sont les tailles des deux échantillons ?

2) Afin de tester les temps moyens d'hospitalisation, quelles sont les hypothèses de loi nécessaires ?

3) À partir des sorties numérotées de 1 à 4 du logiciel R ci-dessus :

3.1) Pour la sortie n° 1, que valent les deux « ?? » ? Quelles sont les hypothèses nulle et alternative ?

3.2) Pour la sortie n° 2, que valent les trois « ?? » ?

3.3)	Pour la sortie n° 3, que valent les trois « ?? » ? Quelles sont les hypothèses nulle et alternative ?

3.4)	Pour la sortie n° 4, que valent les trois « ?? » ?

3.5)	Parmi les sorties numérotées de 1 à 4, quelles sont celles qui vous semblent adaptées à la démarche statistique ? Quelle est votre conclusion pour un risque de première espèce α fixé à 5 % ?

) À partir des sorties numérotées de 5 à 10 du logiciel R :

4.1)	Pour les sorties n° 5, 6 et 7, que vaut le « ?? » associé au df ?

4.2)	Pour la sortie 6, que vaut le « ?? » associé au t ?

4.3)	Pour la sortie 7, que vaut le « ?? » associé au t ?

4.4)	Pour les sorties 9 et 10, que vaut le « ?? » associé au df ?

4.5)	Quelles sont les sorties de logiciel qui ne supposent pas l'égalité des variances ?

5) Quelle est la sortie de logiciel adaptée à la question de ce troisième hôpital ? Précisez les hypothèses nulle et alternative. Quelle est votre conclusion pour un risque α fixé à 5 % ?

Si manque de place pour une question de cet examen en précisant le numéro de l'exercice et de la question.

Amicale Paris Sciences

Licence Sciences Biomédicales
2020-2021

Session 1 – Semestre 4
Correction

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

Exercice 1

1 - $Y_1 = \begin{cases} 1 & \text{Pas de récupération} \quad \pi \\ 0 & \text{Récupération} \quad 1-\pi \end{cases}$

$Y_1 \sim \mathcal{B}(1;\pi)$

2 - Pour avoir les Y_i iid.

3 - $X = \sum_{i=1}^{n} Y_i \sim \mathcal{B}(n;\pi)$ car les Y_i sont iid + $\forall i \; Y_i \sim \mathcal{B}(1;\pi)$

$E(X) = n\pi \qquad V(X) = n\pi(1-\pi) \qquad E(X^2) = n\pi(1-\pi+n\pi)$.

4 - $Z = \sum_{i=1}^{n}(1-Y_i) = n - \sum_{i=1}^{n} Y_i$ ⇒ nb de patients ayant retrouvé la capacité olfactive après 60 j

$Z \sim \mathcal{B}(n;1-\pi)$

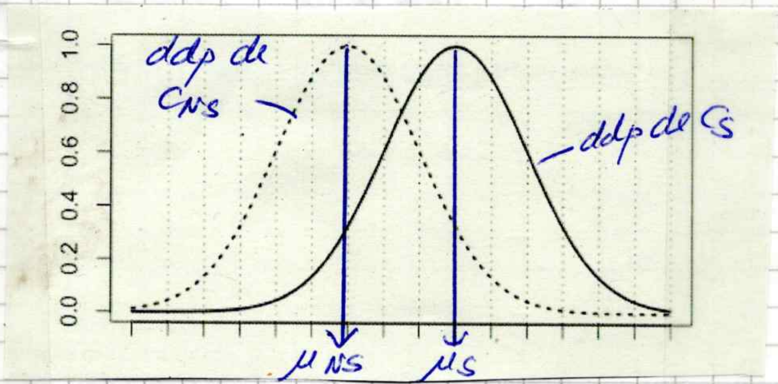
$Z = n - X$

5 - $n = 20 \quad \pi = 1/4 \quad X \sim \mathcal{B}(20;1/4) \quad E(X) = 5$
$Z \sim \mathcal{B}(20;3/4) \quad E(Z) = 15$.

6 - Exacte car $X + Z$ est une constante $(=20)$

Exercice 2

$C_S \sim \mathcal{N}(\mu_S;\sigma_S)$ en L
↳ Normal



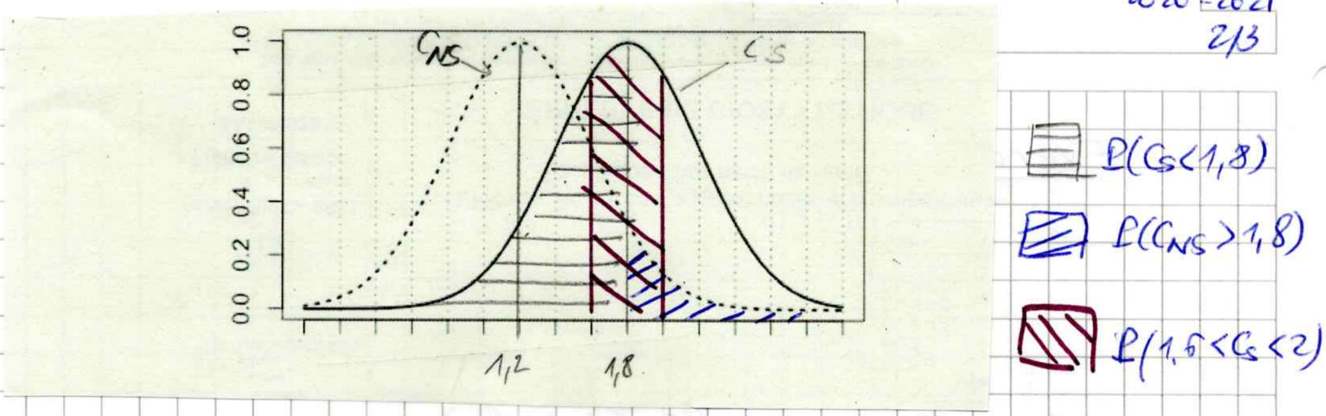
$C_{NS} \sim \mathcal{N}(\mu_{NS};\sigma_S)$ en L
↳ Capacité limitée

donc $\mu_{NS} < \mu_S$

2 - 2 courbes identiques à une translation près (même hauteur, même dispersion) ⇒ affirmation OK

3 - max des ddp de C_S et $C_{NS} \simeq 1$
le max ↑ quand σ ↓ donc σ_S et $\sigma_{NS} < 1$

4 - $\mu_S = 1{,}8$ L $\qquad \sigma_S = \sigma_{CS} = 0{,}4$ L $\qquad \mu_{NS} = 1{,}2$ L



Exercice 3 t_i = temps d'hospitalisation du patient i

(A) $n=6$ $mo = 14,1j$ $me = 12,55j$ $\bar{t} = \frac{\sum t_i}{n} = 12,5j$

$e = 4,3j$ $s^2 = \frac{1}{n-1}\left(\sum t_i^2 - \frac{(\sum t_i)^2}{n}\right) = 2,604j^2$ $s = \sqrt{s^2} = 1,614j$

(B) T_i = v.a = temps d'hospitalisation du patient i

1 – $\forall i$ $T_i \sim \mathcal{N}(\mu;\sigma)$ + iid.

2 – $\frac{\bar{T}-\mu}{S/\sqrt{n}}$ ~ Student à $(n-1)$ddl

$\mu \in I\ \bar{t} - t_{1-\alpha/2}\frac{s}{\sqrt{n}}\ ;\ \bar{t} + t_{1-\alpha/2}\frac{s}{\sqrt{n}}\ I$

3 – Plus $\alpha \uparrow$, plus $t_{1-\alpha/2} \downarrow$, plus IC sera petit.

IC1 → $\alpha = 10\%$ IC2 → $\alpha = 5\%$.

4 – $\bar{t} = 11j$

5 – $n=16$ $s = 1j$

6 – $H_0: \mu = 11,5$ $H_1: \mu \neq 11,5$

$\forall i$ $T_i \sim \mathcal{N}(\mu;\sigma)$ + iid $\frac{\bar{T}-\mu}{S/\sqrt{n}}$ ~ Student à $(n-1)$ddl

$t_{obs} = \frac{\bar{t} - 11,5}{s/\sqrt{n}} = -2$

ZR de H_0. ZR de H_0

$-2,131$ 0 $2,131$ $\left\{\begin{array}{l}\alpha = 5\% \\ 15 ddl\end{array}\right.$

↳ Non rejet de H_0 pour $\alpha = 5\%$.
↳ Temps d'hospitalisation "standard"
ou pas d'argument contre...

7 – $ds > 5\%$

8 – $5\% < ds < 10\%$

C

1 - 2 aux échelles   temps de A » temps de B
$n_A = 8$ $n_B = 10$

2 - $\left.\begin{array}{l} T_{i,A} \\ T_{i,B} \end{array}\right\}$ = temps d'hospitalisation du patient i $\left\{\begin{array}{l} \text{pour A} \\ \text{pour B} \end{array}\right.$

$\forall i$ $T_{i,A} \sim \mathcal{N}(\mu_A ; \sigma_A)$ $T_{i,B} \sim \mathcal{N}(\mu_B ; \sigma_B)$ + iid

3 - ⊗ numdf = $n_B - 1 = 9$ dnomdf = $n_A - 1 = 7$

$H_0 : \sigma_B^2 = \sigma_A^2$ $H_1 : \sigma_B^2 \neq \sigma_A^2$

⊗ numdf = $n_A - 1 = 7$ dnomdf = $n_B - 1 = 9$.

$f = 1/0,42417 \simeq 2,3575$

⊗ $f = 2,3575$ numdf = 7   dnomdf = 9.

$H_0 : \sigma_B^2 = \sigma_A^2$ $H_1 : \sigma_A^2 < \sigma_B^2$.

⊗ numdf = 7   dnomdf = 7   même test que la sortie 3
⇒ pvalue = 0,8849.

⊗ Sorties 1 et 2.
- × $\alpha = 5\%$
- × pvalue = 23,02 % > 5 %
- × Non rejet de H_0 pour $\alpha = 5\%$.
  ⇒ égalité des variances (Pas d'argument contre ...)

4 - ⊗ ?? = 16

⊗ $t = 9,5573$

⊗ $t = -9,5573$

⊗ df = 11,585

⊗ Sorties 8, 9 et 10.

5 - $H_0 : \mu_A = \mu_B$ $H_1 : \mu_A > \mu_B$

Sortie 6
- × $\alpha = 5\%$
- × pvalue $\simeq 10^{-8}$
- × pvalue < 5 % ⇒ rejet de H_0 pour $\alpha = 5\%$

Le temps d'hospitalisation pour les patients du groupe sanguin A est en moyenne significativement plus grand que celui des patients du groupe sanguin B.

Amicale Paris Sciences

Licence Sciences Biomédicales
2021-2022

Session 1 – Semestre 4
Sujet

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

PHARMACIE
Santé
Université Paris Cité

Notes
Ex. 1 :
Ex. 2 :
Ex. 3 – A :
Ex. 3 – B :
Ex. 3 – C :
Total :

L2 Sciences Biomédicales

Année 2021 – 2022
Deuxième semestre
Première session

UFR des Sciences Fondamentales et Biomédicales
Faculté de Pharmacie de Paris

SA04M020 **ÉPREUVE DE BIOSTATISTIQUE**

Tout signe distinctif porté sur la copie pouvant indiquer sa provenance constitue une fraude.
Aucun document n'est autorisé. La calculette **non programmable** est autorisée.
Il est interdit de désagrafer ce cahier (12 pages numérotées).

La cachexie (« mauvais état » en grec) désigne une perte progressive de masse musculaire squelettique survenant au cours d'une maladie chronique. Cette perte conduit par la suite à une détérioration fonctionnelle progressive. L'anorexie (« absence d'appétit » en grec) correspond à un manque d'appétit ou à une impression d'un besoin réduit en apports alimentaires. Le syndrome d'anorexie-cachexie (SAC) bénéficie de la plus grande attention dans le cadre des cancers et fait l'objet des recherches les plus intensives ces dernières années. En dépit de connaissances considérables, la médecine en est encore relativement à ses débuts dans la compréhension des principes de base diagnostiques et thérapeutiques du SAC alors qu'il revêt une importance à la fois pour les patients et pour le système de santé, dans la mesure où il est associé à une morbidité et une mortalité accrues, à une qualité de vie réduite et à des séjours prolongés dans des hôpitaux de soins aigus. Ce syndrome sera étudié dans le cadre du mésothéliome pleural malin (MPM).

Exercice 1 – Le mésothéliome pleural malin (MPM) est un cancer rare dont la principale cause est l'exposition prolongée à l'amiante. Cette maladie touche principalement les hommes et l'âge moyen au moment du diagnostic est d'environ 60 ans. Une équipe de chercheurs australiens avait trouvé qu'environ 50 % des patients étaient déjà atteints du syndrome d'anorexie-cachexie (SAC) lors du diagnostic de leur MPM. Pour confirmer ou pas ce résultat, un service d'oncologie recrute n hommes diagnostiqués MPM afin de décompter le nombre de SAC lors de ce diagnostic.

1) En notant π la probabilité d'être atteint du SAC lors du diagnostic du MPM et en supposant que cette probabilité est la même pour tous ces hommes, quelle condition obtient-on ?
2) La sélection de ces n hommes est telle qu'ils soient sans lien de parenté et sans lien sur leur passé professionnel. Quelle condition obtient-on ?
3) Si l'on note Y la variable aléatoire modélisant le décompte du nombre de SAC lors du diagnostic de MPM parmi ces n hommes, quelles sont ses réalisations ? Quelle est la loi de Y ?
4) La fonction de répartition de Y est notée $F_Y(y)$. Que vaut la fonction de répartition si $y < 0$? Que vaut la fonction de répartition si $y > n$?
5) Sachant que $F_Y(y) = 0{,}00034$ si $0 \leq y < 1$, que pouvez-vous en déduire sur la probabilité qu'aucun homme parmi ces n ne soit atteint du SAC ? Exprimez cette probabilité en fonction de n et de π.
6) Sachant que $F_Y(y) = 0{,}00450$ si $1 \leq y < 2$, que pouvez-vous en déduire sur la probabilité qu'un seul homme parmi ces n soit atteint du SAC ? Exprimez cette probabilité en fonction de n et de π.
7) Le nombre moyen de SAC lors du diagnostic de MPM parmi ces n hommes est de 5,5. Exprimez ce résultat en fonction de n et de π.
8) En utilisant le rapport entre la probabilité qu'aucun homme parmi ces n ne soit atteint du SAC et la probabilité qu'un seul homme parmi ces n soit atteint du SAC, trouvez la probabilité π d'être atteint du SAC lors du diagnostic du MPM (résultat à arrondir à 2 chiffres après la virgule). Que vaut n ?

Réponses (toutes vos réponses doivent être justifiées)

1) En notant π la probabilité d'être atteint du SAC lors du diagnostic du MPM et en supposant que cette probabilité est la même pour tous ces hommes, quelle condition obtient-on ?

2) La sélection de ces n hommes est telle qu'ils soient sans lien de parenté et sans lien sur leur passé professionnel. Quelle condition obtient-on ?

3) Si l'on note Y la variable aléatoire modélisant le décompte du nombre de SAC lors du diagnostic de MPM parmi ces n hommes, quelles sont ses réalisations ? Quelle est la loi de Y ?

4) La fonction de répartition de Y est notée $F_Y(y)$. Que vaut la fonction de répartition si $y < 0$? Que vaut la fonction de répartition si $y \geq n$?

5) Sachant que $F_Y(y) = 0{,}00034$ si $0 \leq y < 1$, que pouvez-vous en déduire sur la probabilité qu'aucun homme parmi ces n ne soit atteint du SAC ? Exprimez cette probabilité en fonction de n et de π.

6) Sachant que $F_Y(y) = 0{,}00450$ si $1 \leq y < 2$, que pouvez-vous en déduire sur la probabilité qu'un seul homme parmi ces n soit atteint du SAC ? Exprimez cette probabilité en fonction de n et de π.

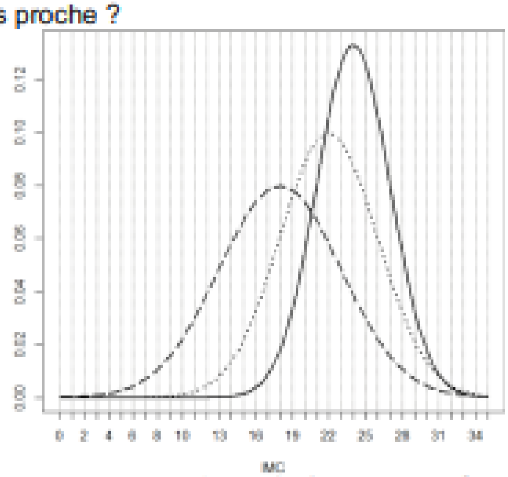
7) Le nombre moyen de SAC lors du diagnostic de MPM parmi ces n hommes est de 5,5. Exprimez ce résultat en fonction de n et de π.

8) En utilisant le rapport entre la probabilité qu'aucun homme parmi ces n ne soit atteint du SAC et la probabilité qu'un seul homme parmi ces n soit atteint du SAC, trouvez la probabilité π d'être atteint du SAC lors du diagnostic du MPM (résultat à arrondir à 2 chiffres après la virgule) ? Que vaut n ?

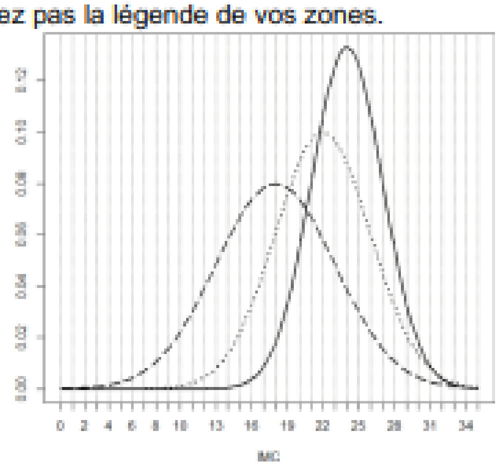
Exercice 2 – L'indice de masse corporelle, appelé aussi IMC, permet d'estimer la corpulence d'une personne en fonction de sa taille et de sa masse corporelle. Pour des hommes atteints d'un cancer du mésothéliome pleural malin (MPM), l'IMC exprimé en kg/m^2 est un indicateur qui diminue dans le temps. En cas de syndrome d'anorexie-cachexie (SAC), cette diminution s'accroît. Cet IMC est modélisé par :

- une loi normale notée I d'espérance μ et d'écart-type σ pour un homme sain,
- une loi normale notée I_{MSC} d'espérance μ_{MSC} et d'écart-type σ_{MSC} pour un homme non atteint par le SAC lors du diagnostic de son MPM,
- une loi normale notée I_{MAC} d'espérance μ_{MAC} et d'écart-type σ_{MAC} pour un homme atteint par le SAC lors du diagnostic de son MPM.

1) Quelles sont les unités des espérances et des écarts-types ci-dessus ?
2) Parmi ces deux propositions, laquelle retenez-vous : « $\mu > \mu_{MSC}$ » ou « $\mu < \mu_{MSC}$ » ?
3) Parmi ces deux propositions, laquelle retenez-vous : « $\mu_{MSC} > \mu_{MAC}$ » ou « $\mu_{MSC} < \mu_{MAC}$ » ?
4) Sachant que les trois densités de probabilités sont représentées ci-dessous, positionnez les trois espérances sur le graphique. Que valent ces trois espérances en précisant leurs unités et en arrondissant à l'entier le plus proche ?



5) Quelle inégalité pouvez-vous proposer entre les trois écarts-types ?
6) Si l'écart-type d'un homme sain vaut 3, quelle est la probabilité que l'IMC d'un homme sain soit supérieur strictement à 27 kg/m^2 ?
7) Si la probabilité que l'IMC d'un homme non atteint par le SAC lors du diagnostic de son MPM soit inférieur strictement à 20 kg/m^2 est de 30,9 %, déduisez-en σ_{MSC} (résultat à arrondir à une décimale).
8) Si la probabilité que l'IMC d'un homme atteint par le SAC lors du diagnostic de son MPM soit compris entre 13 et 23 kg/m^2 (bornes incluses) est de 68,4 %, déduisez-en σ_{MAC} (résultat à arrondir à une décimale).
9) Représentez sur le graphique ci-dessous par des zones hachurées les probabilités des questions 6, 7 et 8. N'oubliez pas la légende de vos zones.



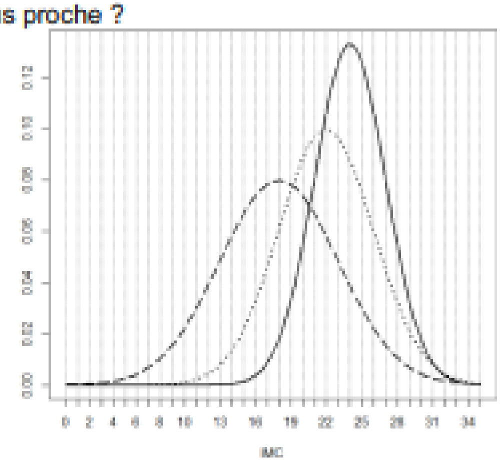
Réponses (<u>toutes vos réponses doivent être justifiées</u>)

1) Quelles sont les unités des espérances et des écarts-types ci-dessus ?

2) Parmi ces deux propositions, laquelle retenez-vous : « $\mu > \mu_{MSC}$ » ou « $\mu < \mu_{MSC}$ » ?

3) Parmi ces deux propositions, laquelle retenez-vous : « $\mu_{MSC} > \mu_{MAC}$ » ou « $\mu_{MSC} < \mu_{MAC}$ » ?

4) Sachant que les trois densités de probabilités sont représentées ci-dessous, positionnez les trois espérances sur le graphique. Que valent ces trois espérances en précisant leurs unités et en arrondissant à l'entier le plus proche ?



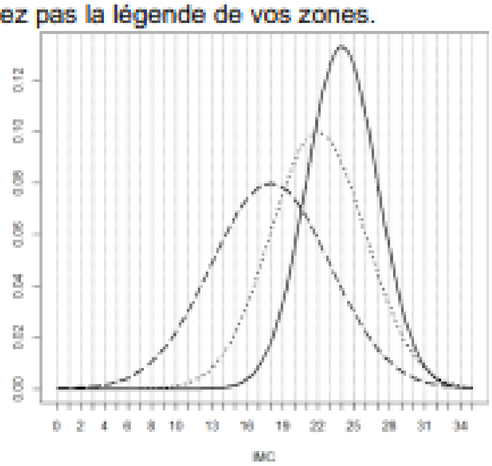
5) Quelle inégalité pouvez-vous proposer entre les trois écarts-types ?

6) Si l'écart-type d'un homme sain vaut 3, quelle est la probabilité que l'IMC d'un homme sain soit supérieur strictement à 27 kg/m^2 ?

7) Si la probabilité que l'IMC d'un homme non atteint par le SAC lors du diagnostic de son MPM soit inférieur strictement à 20 kg/m^2 est de 30,9 %, déduisez-en σ_{MSC} (résultat à arrondir à une décimale).

8) Si la probabilité que l'IMC d'un homme atteint par le SAC lors du diagnostic de son MPM soit compris entre 13 et 23 kg/m^2 (bornes incluses) est de 68,4 %, déduisez-en σ_{MAC} (résultat à arrondir à une décimale).

9) Représentez sur le graphique ci-dessous par des zones hachurées les probabilités des questions 6, 7 et 8. N'oubliez pas la légende de vos zones.



Exercice 3 – (Les 3 parties A, B et C sont indépendantes)

Les trois parties portent sur l'indice de masse corporelle d'hommes atteints d'un cancer du mésothéliome pleural malin (MPM). Cet indice, appelé aussi IMC, permet d'estimer la corpulence d'une personne en fonction de sa taille et de sa masse corporelle. L'IMC exprimé en kg/m^2 peut devenir un indicateur de cachexie en cas de diminution de l'IMC dans le temps.
Une étude est menée afin d'étudier si la prise en compte de facteurs liés au mode de vie, comme l'activité physique ou l'alimentation, pourrait offrir des avantages aux patients atteints de MPM en limitant la cachexie.

Partie A – Lors de l'inclusion des patients atteints de MPM, une étude intermédiaire est menée sur les premiers patients dont les IMC (notés y_i) sont donnés ci-dessous :

25	28	23	26	25	22	19

Quelle est la taille de cet échantillon ? Quand cela est envisageable, calculez la valeur modale, la médiane, la moyenne, l'étendue, la variance et l'écart-type de ces IMC en précisant les unités.

Réponses (<u>toutes vos réponses doivent être justifiées</u>)

Partie B – À la fin de l'inclusion de tous les patients atteints de MPM dans cette étude, les calculs conduisent à une estimation de l'IMC moyen des patients de 22 kg/m^2 avec un écart-type estimé de 4 kg/m^2.

1) Quelle hypothèse de loi est-il nécessaire de faire pour établir l'intervalle de confiance de l'IMC moyen de patients atteints de MPM ?
2) Dans ce contexte, donnez la formule de l'intervalle de confiance de l'IMC moyen.
3) Si, pour un risque α fixé à 10 %, le quantile adapté au calcul de l'intervalle de la question 2 est de 1,7459 en utilisant l'annexe appropriée, quelle est la taille de l'échantillon ?
4) Calculez l'intervalle de confiance de l'IMC moyen pour un risque α fixé à 5 %.
5) L'oncologue suivant cette étude affirme que l'IMC moyen attendu serait de 24,5 kg/m^2 d'après la bibliographie. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifiez pour un risque α fixé à 5 % si la moyenne des IMC est en adéquation avec l'affirmation de l'oncologue.
6) Le degré de signification du test statistique de la question précédente est-il plus petit ou plus grand que 5 % ?
7) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification.

α	$\alpha > 20\ \%$	$20\ \%$	$10\ \%$	$5\ \%$	$2\ \%$	$1\ \%$	$\alpha < 1\ \%$
Test S ou NS ?							

Réponses (toutes vos réponses doivent être justifiées)

1) Quelle hypothèse de loi est-il nécessaire de faire pour établir l'intervalle de confiance de l'IMC moyen de patients atteints de MPM ?

2) Dans ce contexte, donnez la formule de l'intervalle de confiance de l'IMC moyen.

3) Si, pour un risque α fixé à 10 %, le quantile adapté au calcul de l'intervalle de la question 2 est de 1,7459 en utilisant l'annexe appropriée, quelle est la taille de l'échantillon ?

4) Calculez l'intervalle de confiance de l'IMC moyen pour un risque α fixé à 5 %.

5) L'oncologue suivant cette étude affirme que l'IMC moyen attendu serait de 24,5 kg/m^2 d'après la bibliographie. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifiez pour un risque α fixé à 5 % si la moyenne des IMC est en adéquation avec l'affirmation de l'oncologue.

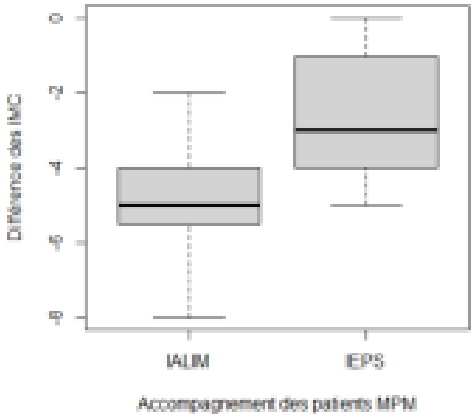
6) Le degré de signification du test statistique de la question précédente est-il plus petit ou plus grand que 5 % ?

7) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification.

α	$\alpha > 20\ \%$	$20\ \%$	$10\ \%$	$5\ \%$	$2\ \%$	$1\ \%$	$\alpha < 1\ \%$
Test S ou NS ?							

Partie C – Les patients de MPM inclus dans cette étude ont été aléatoirement mis dans deux groupes. Le premier groupe (noté IALIM) a reçu des compléments alimentaires lors du suivi. Le deuxième groupe (noté IEPS) a été accompagné pour maintenir une activité physique en adéquation avec leur pathologie. À la fin de la période de suivi et pour les deux groupes, les mesures obtenues sont les différences pour chaque patient i de son IMC après suivi (et donc après l'un des deux accompagnements) avec son IMC du début de l'étude : $d_i = imc_{après,i} - imc_{début,i}$. Sachant qu'aucune amélioration de l'état de santé ne peut être espérée, l'objectif de cette étude est de savoir si l'activité physique freine plus la détérioration de l'IMC qu'un apport en complément alimentaire pour ces patients atteints de MPM.

1) Quel est le nom de la représentation graphique ci-dessous ? Pourquoi les résultats sont-ils tous négatifs ? Quelle réponse intuitive vous suggère cette représentation graphique ?



2) Afin de tester les différences moyennes des IMC, quelles sont les hypothèses nulle et alternative ? Quelles sont les hypothèses de loi nécessaires ? (Attention à vos notations).
3) Voici des sorties du logiciel R :

Sortie de logiciel n° 1	Sortie de logiciel n° 2
F test to compare two variances data: IALIM and IEPS F = 0.99405, num df = 7, denom df = ??, p-value = 0.9931 alternative hypothesis: true ratio of variances is not equal to 1	F test to compare two variances data: IEPS and IALIM F = ?? , num df = 8, denom df = ??, p-value = 0.9931 alternative hypothesis: true ratio of variances is not equal to 1

3.1) Quelles sont les hypothèses nulle et alternative de ces deux sorties ?
3.2) Quelles sont les tailles des deux échantillons ?
3.3) Pour la sortie n° 1, que vaut le « ?? » ?
3.4) Pour la sortie n° 2, que valent les deux « ?? » ?
3.5) Avec une hypothèse alternative unilatérale, quelle serait la valeur du critère de test ?
3.6) Quelle est la conclusion de ce test pour un risque α fixé à 5 % ?

4) Voici des sorties du logiciel R :

Sortie de logiciel n° 3	Sortie de logiciel n° 4
Welch Two Sample t-test data: IALIM and IEPS t = -2.628, df = 14.777, p-value = 0.01919 alternative hypothesis: true difference in means is not equal to 0	Welch Two Sample t-test data: IEPS and IALIM t = ??, df = ??, p-value = 0.9904 alternative hypothesis: true difference in means is less than 0

Sortie de logiciel n° 5	Sortie de logiciel n° 6
Two Sample t-test data: IALIM and IEPS t = -2.6275, df = ??, p-value = 0.01903 alternative hypothesis: true difference in means is not equal to 0	Two Sample t-test data: IALIM and IEPS t = -2.6275, df = ??, p-value = 0.9905 alternative hypothesis: true difference in means is greater than 0

Sortie de logiciel n° 7	Sortie de logiciel n° 8
Two Sample t-test data: IALIM and IEPS t = -2.6275, df = ??, p-value = 0.009513 alternative hypothesis: true difference in means is less than 0	Two Sample t-test data: IEPS and IALIM t = 2.6275, df = ??, p-value = 0.009513 alternative hypothesis: true difference in means is greater than 0

4.1) Quelle est l'hypothèse nulle de ces sorties ?

4.2) Pour la sortie n° 4, que valent les deux « ?? » ?

4.3) Pour les sorties n° 5, 6, 7 et 8, que vaut le « ?? » associé à df ?

4.4) En fonction de votre conclusion à la question 3.6, quelles sont les sorties qu'il faudrait favoriser ?

5) Afin d'atteindre l'objectif de cette étude, quelle(s) sortie(s) retenez-vous ? Précisez les hypothèses nulle et alternative. Quelle est votre conclusion pour un risque α fixé à 1 % ?

Réponses (toutes vos réponses doivent être justifiées)

1) Quel est le nom de la représentation graphique ci-dessus ? Pourquoi les résultats sont-ils tous négatifs ? Quelle réponse intuitive vous suggère cette représentation graphique ?

2) Afin de tester les différences moyennes des IMC, quelles sont les hypothèses nulle et alternative ? Quelles sont les hypothèses de loi nécessaires ? (Attention à vos notations).

3) À partir des sorties du logiciel R numérotées de 1 à 2 :

3.1) Quelles sont les hypothèses nulle et alternative de ces deux sorties ?

3.2) Quelles sont les tailles des deux échantillons ?

3.3) Pour la sortie n° 1, que vaut le « ?? » ?

3.4) Pour la sortie n° 2, que valent les deux « ?? » ?

3.5) Avec une hypothèse alternative unilatérale, quelle serait la valeur du critère de test ?

3.6) Quelle est la conclusion de ce test pour un risque α fixé à 5 % ?

4) À partir des sorties du logiciel R numérotées de 3 à 8 :

4.1) Quelle est l'hypothèse nulle de ces sorties ?

4.2) Pour la sortie n° 4, que valent les deux « ?? » ?

4.3) Pour les sorties n° 5, 6, 7 et 8, que vaut le « ?? » associé à df ?

4.4) En fonction de votre conclusion à la question 3.6, quelles sont les sorties qu'il faudrait favoriser ?

5) Afin d'atteindre l'objectif de cette étude, quelle(s) sortie(s) retenez-vous ? Précisez les hypothèses nulle et alternative. Quelle est votre conclusion pour un risque α fixé à 1 % ?

Amicale Paris Sciences

Licence Sciences Biomédicales
2021-2022

Session 1 – Semestre 4
Correction

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

Pour vérifier vos calculs
L2 - Biostat - (2021 - 2022)

Exercice 1 n hommes MPM pour décompter le nb de SAC.

1 - π = proba d'être atteint du SAC lors du diagnostic du MPM.
↳ la même pour tous les hommes ⇒ identiquement distribué.

2 - Sélection sans lien de parenté et sans lien sur passé pro ⇒ indépendance.

3 - Y = v.a. = nb de SAC lors du diagnostic parmi n.
$Y = 0, 1, \ldots, n$. (réalisations)

loi de Y ? $X_i = \begin{cases} 1 & SAC \quad \pi \\ 0 & \overline{SAC} \quad 1-\pi \end{cases}$ $X_i \sim \mathcal{B}(1, \pi)$ + iid $\forall i$

$$Y = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n; \pi)$$

4 - $F_Y(y) = P(Y \leq y)$ Par définition $\begin{cases} F_Y(y) = 0 & \text{si } y < 0 \\ F_Y(y) = 1 & \text{si } y > n \end{cases}$

5 - $F_Y(y) = 0{,}00034$ si $0 \leq y < 1$.

si $0 \leq y < 1$ $F_Y(y) = P(Y=0) = \boxed{(1-\pi)^n = 0{,}00034}$

6 - $F_Y(y) = 0{,}00450$ si $1 \leq y < 2$

si $1 \leq y < 2$ $F_Y(y) = P(Y=0) + P(Y=1)$.

$P(Y=1) = 0{,}0045 - 0{,}00034 = 0{,}00416$

$= \binom{n}{1} \pi^1 (1-\pi)^{n-1} \Leftrightarrow \boxed{n\pi(1-\pi)^{n-1} = 0{,}00416}$

7 - $E(Y) = n\pi = 5{,}5$

8 - $\dfrac{P(Y=0)}{P(Y=1)} = \dfrac{(1-\pi)^n}{n\pi(1-\pi)^{n-1}} = \dfrac{1-\pi}{5{,}5} = \dfrac{0{,}00034}{0{,}00416} \Rightarrow \boxed{\pi = 0{,}55}$

$\boxed{n = 10}$

Exercice 2 IMC en kg/m^2 :

Pour un homme sain. $I \sim \mathcal{N}(\mu ; \sigma)$

Pour un homme non atteint par le SAC $I_{MSC} \sim \mathcal{N}(\mu_{MSC} ; \sigma_{MSC})$

Pour un homme atteint du SAC $I_{MAC} \sim \mathcal{N}(\mu_{MAC} ; \sigma_{MAC})$

1 – μ et σ en kg/m^2.

2 – $\mu > \mu_{MSC}$ car l'IMC ↓ pour les hommes atteints de NAF

3 – $\mu_{MAC} < \mu_{MSC}$ car le SAC diminue plus l'IMC.

4 – $\mu_{MAC} = 18\ kg/m^2$ $\mu_{MSC} = 22\ kg/m^2$ $\mu = 24\ kg/m^2$

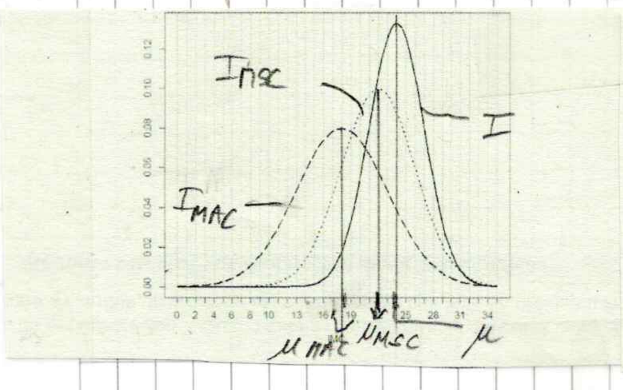
5 – $\sigma_{MAC} > \sigma_{MSC} > \sigma$.

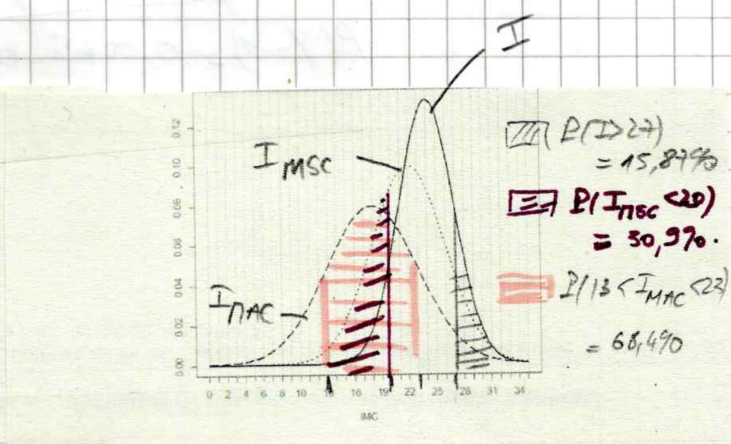
6 – $\sigma = 3\ kg/m^2$ $P(I > 27) = 1 - F_\gamma(1) = 0,1587$

7 – $P(I_{MSC} < 20) = 0,309 \Rightarrow \sigma_{MSC} = 4\ kg/m^2$.

8 – $P(13 < I_{MAC} < 23) = 0,684$

$F_\gamma\left(\frac{5}{\sigma_{MAC}}\right) = 0,842 \Leftrightarrow \sigma_{MAC} = 5\ kg/m^2$





Exercice 3

(A) $n = 7$ $\quad m_0 = 25\ kg/m^2$ $\quad me = 25\ kg/m^2$

$\bar{y} = \frac{\sum y_i}{n} = \frac{168}{7} = 24\ kg/m^2$

$e = 28 - 19 = 9\ kg/m^2$

$s^2 = \frac{1}{n-1}\left\{ \sum y_i^2 - \frac{(\sum y_i)^2}{n} \right\} = 8{,}66\ (kg/m^2)^2$

$s = 2{,}94\ kg/m^2$

(B) $\bar{y} = 22\ kg/m^2$ $\quad s = 4\ kg/m^2$

Y_i = v.a. = IMC du i^e patient

(1) $\forall i\ Y_i \sim \mathcal{N}(\mu, \sigma)$ + iid.

(2) $\bar{Y} \sim \mathcal{N}(\mu; \sigma/\sqrt{n})$ $\quad T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim$ Student à $(n-1)$ ddl.

$$\mu \in \left] \bar{y} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}\ ;\ \bar{y} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right[$$

où $t_{1-\alpha/2}$ lu dans la table de Student à $(n-1)$ ddl.

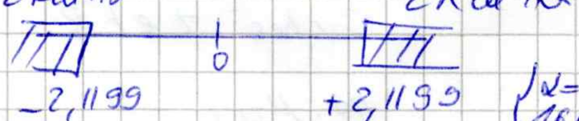
(3) $\alpha = 10\%$ $\quad t_{1-\alpha/2} = 1{,}7459 \Rightarrow n - 1 = 16 \Rightarrow n = 17$

(4) $]19{,}943\ ;\ 24{,}057[$ (en kg/m^2)

(5) $H_0 : \mu = 24{,}5$ $\quad H_1 : \mu \neq 24{,}5$

$\forall i\ Y_i \sim \mathcal{N}(\mu, \sigma)$ + iid $\quad T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim$ Student à $(n-1)$ ddl.

$t_{obs} = \frac{\bar{y} - 24{,}5}{s/\sqrt{n}} = -2{,}57$



$\left\{ \begin{array}{l} \alpha = 5\% \\ 16\ ddl. \end{array} \right.$

$\Rightarrow$ Rejet de Ho pour $\alpha = 5\%$.
$\Rightarrow$ Pas en adéquation avec l'affirmation de l'oncologue.

(6) $ds < 5\%$

(7) $2\% < ds < 5\%$

Ⓒ

① Box. Plot
Les résultats sont négatifs car aucune amélioration ne peut être espérée.
L'activité physique semblerait ralentir la dégradation de l'IMC

② $X_{i, IALIN}$ = v.a. : ≠ IMC avec complément alimentaire

$X_{i, IEPS}$ = " : ≠ " avec activité physique.

H: $X_{i, IALIN} \sim \mathcal{N}(\mu_{IALIM} ; \sigma_{IALIN})$ + iid.
$X_{i, IEPS} \sim \mathcal{N}(\mu_{IEPS} ; \sigma_{IEPS})$

③ $H_0 : \sigma^2_{IALIN} = \sigma^2_{IEPS}$ $H_1 : \sigma^2_{IALIN} \neq \sigma^2_{IEPS}$

numdf = $n_{IALIN} - 1 = 7$ $n_{IALIN} = 8$ (sortie 1)
numdf = $n_{IEPS} - 1 = 8$ $n_{IEPS} = 9$ (sortie 2)

?? = 8

$F = \frac{1}{0,99405} = 1,005$ denomdf = 7

même valeur du critère que pour le bilatéral (car sous H_0)

ds = 99,31 % > 5 % ⇒ non rejet de H_0
⇒ "Égalité des variances"
"pas d'argument contre l'égalité des variances"

④ $H_0 : \mu_{IALIN} = \mu_{IEPS}$

$t = + 2,620$ df = 14,777

df = 15

Sorties 7 et 8

$H_0 : \mu_{IEPS} = \mu_{IALIN}$ $H_1 : \mu_{IEPS} > \mu_{IALIN}$
7 et 8

$\alpha = 1\%$ ds < 1% ⇒ Rejet de H_0 pour $\alpha = 1\%$.

⇒ L'activité physique freine plus la détérioration de l'IMC qu'un apport en complément alimentaire

Amicale Paris Sciences

Licence Sciences Biomédicales
2022-2023

Session 1 – Semestre 4

Sujet

Biostatistiques

Siège administratif : Amical Paris Sciences – 45 Rue des Saints-Pères – 75006 Paris
http://www.aps-paris5.fr - Email : assosaps@gmail.com
Association régit par la loi 1901 enregistrée à la préfecture de Paris

PHARMACIE
Santé
Université Paris Cité

Notes
Ex. 1 :
Ex. 2 :
Ex. 3 – A :
Ex. 3 – B :
Ex. 3 – C :
Total :

L2 Sciences Biomédicales

Année 2022 – 2023
Deuxième semestre
Première session

UFR des Sciences Fondamentales et Biomédicales
Faculté de Pharmacie de Paris

SA04M020

ÉPREUVE DE BIOSTATISTIQUE
Durée de l'épreuve : 1 h

Tout signe distinctif porté sur la copie pouvant indiquer sa provenance constitue une fraude.
Aucun document n'est autorisé. La calculette **non programmable** est autorisée.
Il est interdit de désagrafer ce cahier (12 pages numérotées).

L'hirudothérapie est la thérapie par les sangsues médicinales. L'effet thérapeutique recherché est la combinaison d'un effet de saignée décongestionnant, d'un effet anticoagulant et d'un effet analgésique lié directement à la morsure et à la succion. La sangsue est installée sur la zone concernée, par période de 30 minutes à chaque fois. Quand la sangsue est gorgée de sang, elle se détache du site. Les indications pour une application de sangsues s'effectuant sur prescription médicale sont la stagnation veineuse, la greffe tissulaire et la réimplantation chirurgicale.

Exercice 1 – L'un des effets indésirables de l'hirudothérapie est la complication infectieuse. En effet, le tube digestif des sangsues est colonisé par une bactérie du genre *Aeromonas* qui peut entraîner une infection chez le patient. Afin d'étudier la nécessité de mettre en place une antibioprophylaxie lors de la pose de sangsues médicinales, on étudie l'absence de cette infection parmi n patients.

1) Le fait qu'un patient développe ou pas une infection n'a aucun impact sur la possible infection d'un autre patient. Quelle condition obtient-on ?
2) Que doit-on supposer sur la probabilité, notée π, de ne pas développer une infection à la bactérie du genre *Aeromonas* ?
3) Tous les patients ont eu la même microchirurgie à un doigt nécessitant la pose du même nombre de sangsues pour revasculariser leur doigt et leur état immunitaire était semblable. Est-ce en accord avec l'une des conditions précédentes ?
4) Si la variable aléatoire modélisant l'absence ou pas d'une infection chez l'individu n° i est notée Z_i, quelles sont ses réalisations ? Quelle est la loi de Z_i ? Que valent son espérance et sa variance ?
5) Si l'on note X la variable aléatoire modélisant le décompte du nombre de non-infections parmi ces n patients, quelles sont ses réalisations ? Quelle est la loi de X en faisant le lien avec les Z_i ? Que valent son espérance et sa variance ?
6) Quelle est la probabilité (à exprimer en fonction de n et π) que les n patients soient non infectés après la pose de sangsues ?
7) Quelle est la probabilité (à exprimer en fonction de n et π) que $n-1$ patients parmi les n soient non infectés après la pose de sangsues ?
8) Si la probabilité demandée à la question 7 vaut 0,2684 et si la variance de X vaut 1,6, quelle est la valeur de π^{n-2} ?
9) Que vaut la fonction de répartition, notée $F_X(x)$, de X si $x \geq n$?
10) Que vaut la fonction de répartition de X (à exprimer en fonction de n et π) si $n-1 \leq x < n$?
11) Si $F_X(x) = 0{,}8926$ pour $n-1 \leq x < n$, quelle est la valeur de la probabilité de la question 6 ?
12) En utilisant le rapport entre les résultats des questions 11 et 8, trouvez la probabilité π de ne pas développer une infection après la pose de sangsues (résultat à arrondir à 1 chiffre après la virgule).
13) Que vaut n en arrondissant à l'entier le plus proche ?
14) Quel est le nombre moyen de patients développant une infection après la pose de sangsues ?

Réponses <u>(toutes vos réponses doivent être justifiées)</u>

1) Le fait qu'un patient développe ou pas une infection n'a aucun impact sur la possible infection d'un autre patient. Quelle condition obtient-on ?

2) Que doit-on supposer sur la probabilité, notée π, de ne pas développer une infection à la bactérie du genre *Aeromonas* ?

3) Tous les patients ont eu la même microchirurgie à un doigt nécessitant la pose du même nombre de sangsues pour revasculariser leur doigt et leur état immunitaire était semblable. Est-ce en accord avec l'une des conditions précédentes ?

4) Si la variable aléatoire modélisant l'absence ou pas d'une infection chez l'individu n° i est notée Z_i, quelles sont ses réalisations ? Quelle est la loi de Z_i ? Que valent son espérance et sa variance ?

5) Si l'on note X la variable aléatoire modélisant le décompte du nombre de non-infections parmi ces n patients, quelles sont ses réalisations ? Quelle est la loi de X en faisant le lien avec les Z_i ? Que valent son espérance et sa variance ?

6) Quelle est la probabilité (à exprimer en fonction de n et π) que les n patients soient non infectés après la pose de sangsues ?

7) Quelle est la probabilité (à exprimer en fonction de n et π) que $n-1$ patients parmi les n soient non infectés après la pose de sangsues ?

8) Si la probabilité demandée à la question 7 vaut 0,2684 et si la variance de X vaut 1,6, quelle est la valeur de π^{n-2} ?

9) Que vaut la fonction de répartition, notée $F_X(x)$, de X si $x \geq n$?

10) Que vaut la fonction de répartition de X (à exprimer en fonction de n et π) si $n-1 \leq x < n$?

11) Si $F_X(x) = 0{,}8926$ pour $n-1 \leq x < n$, quelle est la valeur de la probabilité de la question 6 ?

12) En utilisant le rapport entre les résultats des questions 11 et 8, trouvez la probabilité π de ne pas développer une infection après la pose de sangsues (résultat à arrondir à 1 chiffre après la virgule).

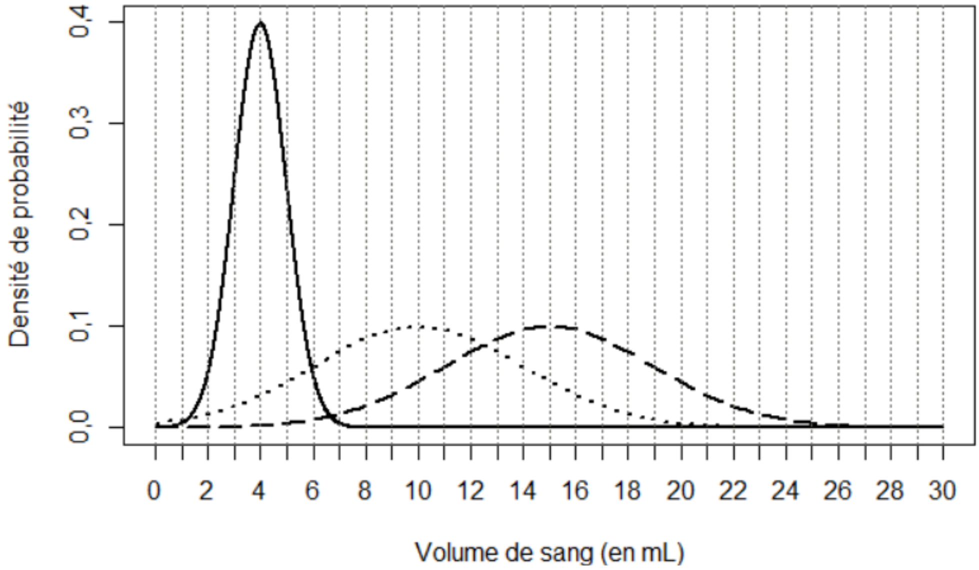
13) Que vaut n en arrondissant à l'entier le plus proche ?

14) Quel est le nombre moyen de patients développant une infection après la pose de sangsues ?

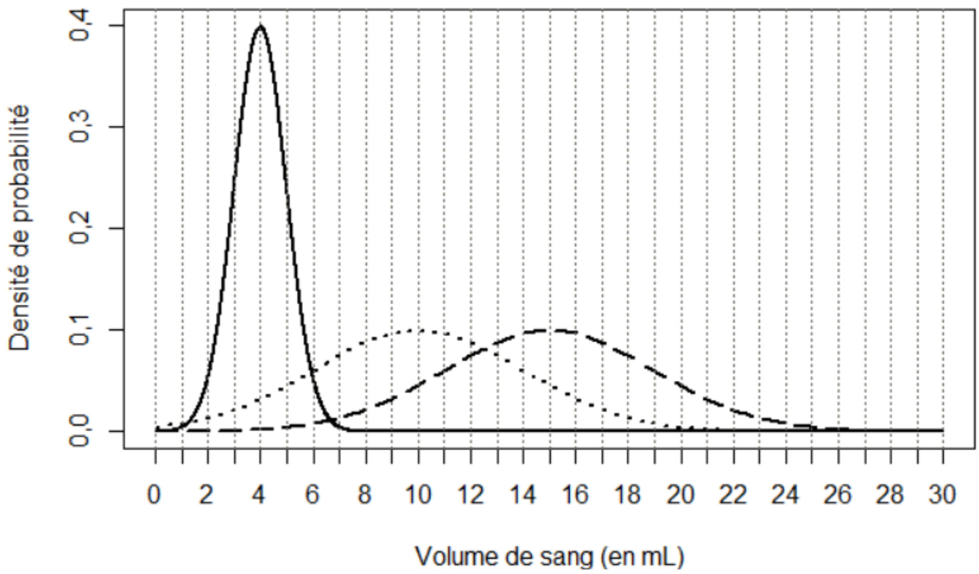
Exercice 2 – L'application et la morsure d'une sangsue sont indolores. La difficulté de cette thérapie est de bien faire mordre la sangsue en la faisant jeûner pour qu'elle puisse se gaver de sang puis se détacher d'elle-même, rassasiée, après avoir absorbé quelques mL de sang sachant qu'une sangsue peut survivre plus de 2 ans après un seul repas. On supposera qu'une sangsue n'ayant pas jeûné se gave de très peu de sang et que plus le temps de jeûne est important, plus la quantité moyenne de sang absorbée est importante.
Le volume de sang absorbé (en mL) est modélisé par une loi normale :

– notée V_0 d'espérance μ_0 et d'écart-type σ_0 pour une sangsue n'ayant pas jeûné ;
– notée V_3 d'espérance μ_3 et d'écart-type σ pour une sangsue ayant jeûné 3 mois ;
– notée V_6 d'espérance μ_6 et d'écart-type σ pour une sangsue ayant jeûné 6 mois.

La représentation graphique de ces trois densités de probabilités est proposée ci-dessous :

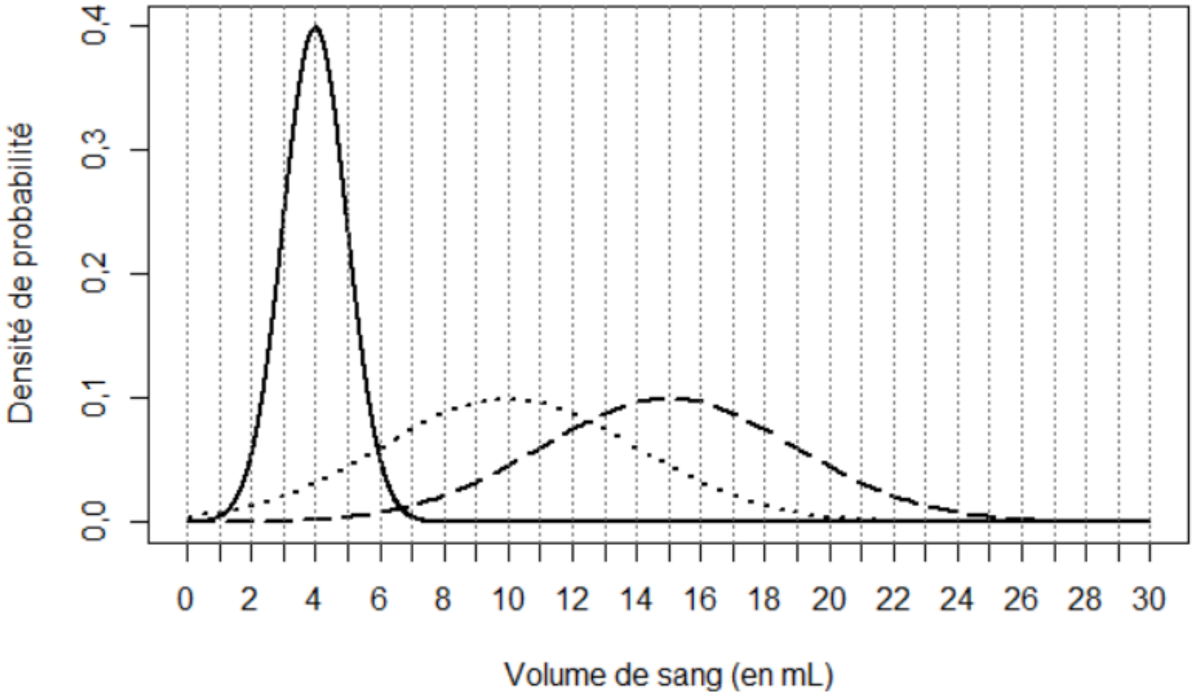


1) Proposez une inégalité entre les trois espérances.
2) Positionnez les espérances sur le graphique ci-dessus. Que valent ces trois espérances en précisant leurs unités et en arrondissant à l'entier le plus proche ?
3) Quelle inégalité pouvez-vous proposer entre les deux écarts-types ?
4) La probabilité que le volume de sang absorbé par une sangsue n'ayant pas jeûné soit compris entre 3 et 5 mL vaut 68,26 %, déduisez-en σ_0 (résultat à arrondir à l'entier le plus proche).
5) Si la probabilité que le volume de sang absorbé par une sangsue ayant jeûné soit strictement supérieur à 14 mL est de 59,87 %, le jeûne de la sangsue est-il de 3 ou 6 mois ?
6) Que vaut σ (résultat à arrondir à l'entier le plus proche) ?
7) Quelle est la probabilité que le volume de sang absorbé par une sangsue ayant jeûné 3 mois soit compris entre $\mu_3 - \sigma$ et $\mu_3 + \sigma$ (bornes comprises) ?
8) Représentez sur le graphique ci-dessous par des zones hachurées les probabilités des questions 4, 5 et 7. N'oubliez pas la légende de vos zones.



Réponses (toutes vos réponses doivent être justifiées)

La représentation graphique de ces trois densités de probabilités est proposée ci-dessous :



1) Proposez une inégalité entre les trois espérances.

2) Positionnez les espérances sur le graphique ci-dessus. Que valent ces trois espérances en précisant leurs unités et en arrondissant à l'entier le plus proche ?

3) Quelle inégalité pouvez-vous proposer entre les deux écarts-types ?

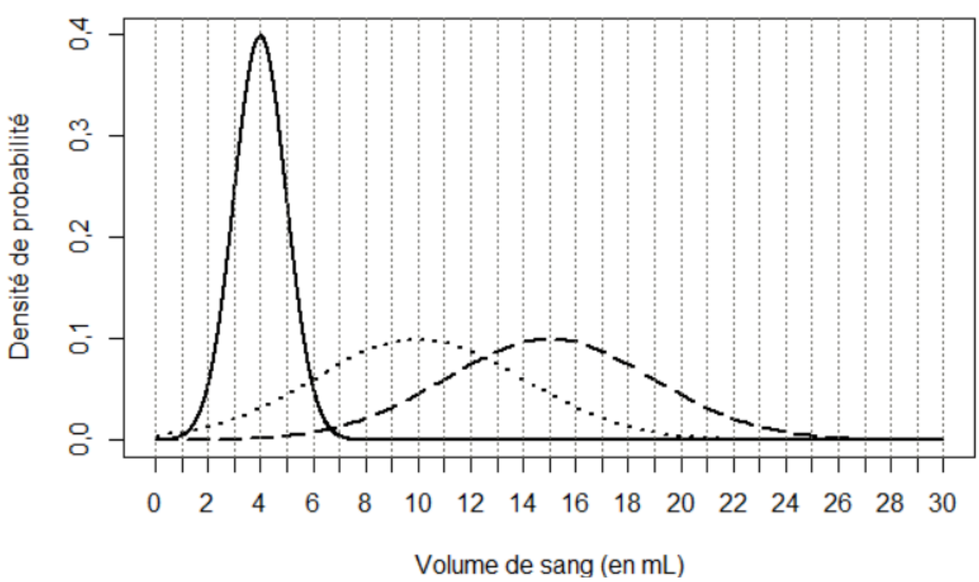
4) La probabilité que le volume de sang absorbé par une sangsue n'ayant pas jeûné soit compris entre 3 et 5 mL vaut 68,26 %, déduisez-en σ_0 (résultat à arrondir à l'entier le plus proche).

5) Si la probabilité que le volume de sang absorbé par une sangsue ayant jeûné soit strictement supérieur à 14 mL est de 59,87 %, le jeûne de la sangsue est-il de 3 ou 6 mois ?

6) Que vaut σ (résultat à arrondir à l'entier le plus proche) ?

7) Quelle est la probabilité que le volume de sang absorbé par une sangsue ayant jeûné 3 mois soit compris entre $\mu_3 - \sigma$ et $\mu_3 + \sigma$ (bornes comprises) ?

8) Représentez sur le graphique ci-dessous par des zones hachurées les probabilités des questions 4, 5 et 7. N'oubliez pas la légende de vos zones.



Exercice 3 – **(Les 3 parties A, B et C sont indépendantes)**

Les sangsues médicinales sont considérées comme un dispositif médical et, à ce titre, elles sont prises en charge par la pharmacie de l'hôpital. Afin d'utiliser des sangsues pour revasculariser, par exemple, un doigt après une microchirurgie, un pharmacien du centre « SOS MAIN » étudie le volume de sang absorbé par une sangsue afin d'évaluer les risques d'anémie en cas de mise en place de plusieurs sangsues. Dans ce contexte, chaque sangsue préalablement pesée est apposée à un volontaire. Lorsque cette dernière se détache d'elle-même, elle est de nouveau pesée pour en déduire le volume de sang absorbé par la sangsue (noté VS). Chaque sangsue est à usage unique.

Partie A – Une première série de mesures du VS (en mL) est présentée ci-dessous :

10,7 9,8 12,6 8,6 12,4 8,9

Sur combien de volontaires a été réalisée cette première série ? Quand cela est envisageable, calculez la valeur modale, la médiane, la moyenne, l'étendue, la variance et l'écart-type de ces volumes de sang en précisant les unités.

Réponses (<u>toutes vos réponses doivent être justifiées</u>)

Partie B – Lors d'une deuxième série de mesures des VS (en mL) sur 22 volontaires, le pharmacien publie l'intervalle de confiance du VS moyen pour un risque α fixé à 5 % :

[10,20193 ; 11,79807] (en mL)

1) Quelle hypothèse de loi est nécessaire pour établir l'intervalle de confiance du VS moyen ?
2) Dans ce contexte, donnez la formule de l'intervalle de confiance du VS moyen.
3) Quelle est la moyenne des VS pour cet échantillon de volontaires ?
4) Quel est l'écart-type des VS pour cet échantillon de volontaires ?
5) Le pharmacien affirme que le VS moyen attendu serait de 10,4 mL d'après la bibliographie. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifiez pour un risque α fixé à 10 % si la moyenne des VS est en adéquation avec l'affirmation du pharmacien.
6) Le degré de signification du test statistique de la question précédente est-il plus petit ou plus grand que 10 % ?
7) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification.

α	$\alpha < 1\ \%$	$1\ \%$	$2\ \%$	$5\ \%$	$10\ \%$	$20\ \%$	$40\ \%$	$\alpha > 40\ \%$
Test S ou NS ?								

Réponses (<u>toutes vos réponses doivent être justifiées</u>)

1) Quelle hypothèse de loi est nécessaire pour établir l'intervalle de confiance du VS moyen ?

2) Dans ce contexte, donnez la formule de l'intervalle de confiance du VS moyen.

3) Quelle est la moyenne des VS pour cet échantillon de volontaires ?

4) Quel est l'écart-type des VS pour cet échantillon de volontaires ?

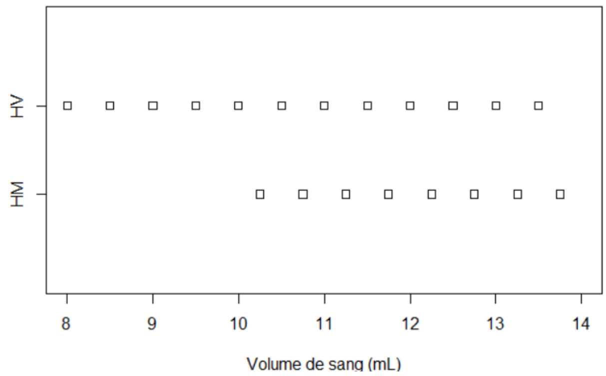
5) Le pharmacien affirme que le VS moyen attendu serait de 10,4 mL d'après la bibliographie. Uniquement avec un test statistique dont les hypothèses nulle et alternative seront précisées ainsi que les conditions d'utilisation, vérifiez pour un risque α fixé à 10 % si la moyenne des VS est en adéquation avec l'affirmation du pharmacien.

6) Le degré de signification du test statistique de la question précédente est-il plus petit ou plus grand que 10 % ?

7) En fonction des différentes valeurs du risque α, remplissez le tableau suivant avec un S (respectivement NS) quand le test est significatif (respectivement non significatif). Donnez un encadrement du degré de signification.

α	$\alpha < 1\ \%$	$1\ \%$	$2\ \%$	$5\ \%$	$10\ \%$	$20\ \%$	$40\ \%$	$\alpha > 40\ \%$
Test S ou NS ?								

Partie C – Il existe différentes espèces de sangsues médicinales, en particulier *Hirudo medicinalis* (notée HM) et *Hirudo verbana* (notée HV). Afin de vérifier si ces deux espèces de sangsues absorbent le même volume de sang, le pharmacien continue son étude en traitant des volontaires par *Hirudo medicinalis* et d'autres volontaires par *Hirudo verbana*.



1) Quel est le nom de la représentation graphique précédente ? Quel est le nombre de mesures pour HV et pour HM sachant qu'il n'y aucun *exæquo* (attention à vos notations) ?
2) Représentez synthétiquement les deux boîtes à moustaches de ces deux séries de mesures sur la représentation graphique précédente. Quelle réponse intuitive à la question du pharmacien vous suggère cette représentation graphique ?
3) Afin de vérifier si ces deux espèces de sangsues absorbent en moyenne le même volume de sang, quelles sont les hypothèses nulle et alternative ? Quelles sont les hypothèses de loi nécessaires ? (Attention à vos notations).
4) Voici des sorties du logiciel R :

Sortie de logiciel n° 1	Sortie de logiciel n° 2
F test to compare two variances data: HV and HM F = ??, num df = ??, denom df = ??, p-value = 0.3141 alternative hypothesis: true ratio of variances is not equal to 1	F test to compare two variances data: HM and HV F = ??, num df = ??, denom df = ??, p-value = 0.3141 alternative hypothesis: true ratio of variances is not equal to 1

4.1) Quelles sont les hypothèses nulle et alternative de ces deux sorties ?
4.2) Pour les sorties n° 1 et 2, que valent les « ?? » associés à « num df » et à « denom df » ?
4.3) En vous aidant de la représentation graphique précédente, quelle proposition retenez-vous parmi les deux suivantes : $s_{HV} > s_{HM}$ ou $s_{HV} < s_{HM}$?
4.4) À quelle sortie, n° 1 ou n° 2, associez-vous la valeur 2,166 pour « F » ? Déduisez-en la valeur de « F » pour l'autre sortie.
4.5) Quelle est la conclusion de ce test pour un risque α fixé à 5 % ?

5) Voici des sorties du logiciel R :

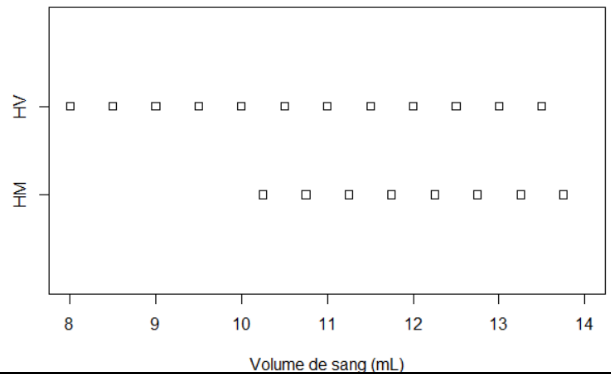
Sortie de logiciel n° 3	Sortie de logiciel n° 4
Two Sample t-test data: HV and HM t = -1.7085, df = ??, p-value = 0.1047 alternative hypothesis: true difference in means is not equal to 0	Two Sample t-test data: HM and HV t = ??, df = ??, p-value = 0.1047 alternative hypothesis: true difference in means is not equal to 0

Sortie de logiciel n° 5	Sortie de logiciel n° 6
Welch Two Sample t-test data: HM and HV t = ??, df = 17.969, p-value = 0.08138 alternative hypothesis: true difference in means is not equal to 0	Welch Two Sample t-test data: HV and HM t = ??, df = ??, p-value = 0.08138 alternative hypothesis: true difference in means is not equal to 0

5.1) Quelle est l'hypothèse nulle de ces sorties ?
5.2) Pour les sorties n° 3 et 4, que valent les « ?? » associés à « df » ?
5.3) Pour la sortie n° 6, que vaut le « ?? » associé à « df » ?
5.4) Pour la sortie n° 4, que vaut le « ?? » associé à « t » ?
5.5) À quelle sortie, n° 5 ou n° 6, associez-vous la valeur 1,8464 pour « t » ? Déduisez-en la valeur de « t » pour l'autre sortie.
5.6) Quelle est l'hypothèse alternative de ces sorties ? Avec une hypothèse alternative formulée différemment, quelle serait la valeur du critère de test ?
5.7) En fonction de votre conclusion à la question 4.5, quelles sont les sorties qu'il faudrait favoriser ?

6) Ces deux espèces de sangsues absorbent-elles en moyenne le même volume de sang pour un risque α fixé à 5 % ?

Réponses (<u>toutes vos réponses doivent être justifiées</u>)



1) Quel est le nom de la représentation graphique précédente ? Quel est le nombre de mesures pour HV et pour HM sachant qu'il n'y aucun *exæquo* (attention à vos notations) ?

2) Représentez synthétiquement les deux boîtes à moustaches de ces deux séries de mesures sur la représentation graphique précédente. Quelle réponse intuitive à la question du pharmacien vous suggère cette représentation graphique ?

3) Afin de vérifier si ces deux espèces de sangsues absorbent en moyenne le même volume de sang, quelles sont les hypothèses nulle et alternative ? Quelles sont les hypothèses de loi nécessaires ? (Attention à vos notations).

4) À partir des sorties n° 1 et 2 du logiciel R :

4.1) Quelles sont les hypothèses nulle et alternative de ces deux sorties ?

4.2) Pour les sorties n° 1 et 2, que valent les « ?? » associés à « num df » et à « denom df » ?

4.3) En vous aidant de la représentation graphique précédente, quelle proposition retenez-vous parmi les deux suivantes : $s_{HV} > s_{HM}$ ou $s_{HV} < s_{HM}$?

4.4) À quelle sortie, n° 1 ou n° 2, associez-vous la valeur 2,166 pour « F » ? Déduisez-en la valeur de « F » pour l'autre sortie.

4.5) Quelle est la conclusion de ce test pour un risque α fixé à 5 % ?

5) À partir des sorties n° 3, 4, 5 et 6 du logiciel R :

5.1) Quelle est l'hypothèse nulle de ces sorties ?

5.2) Pour les sorties n° 3 et 4, que valent les « ?? » associés à « df » ?

5.3) Pour la sortie n° 6, que vaut le « ?? » associé à « df » ?

5.4) Pour la sortie n° 4, que vaut le « ?? » associé à « t » ?

5.5) À quelle sortie, n° 5 ou n° 6, associez-vous la valeur 1,8464 pour « t » ? Déduisez-en la valeur de « t » pour l'autre sortie.

5.6) Quelle est l'hypothèse alternative de ces sorties ? Avec une hypothèse alternative formulée différemment, quelle serait la valeur du critère de test ?

5.7) En fonction de votre conclusion à la question 4.5, quelles sont les sorties qu'il faudrait favoriser ?

6) Ces deux espèces de sangsues absorbent-elles en moyenne le même volume de sang pour un risque α fixé à 5 % ?