Association pour l'Accès Santé Université Paris Cité
A2SUP

Annales Corrigées 2025-2026

Tome **Tutorat**
2022 - 2025

UE 9 : Mathématiques - Biostatistiques
Sujets et Corrections

ASSOCIATION POUR L'ACCÈS SANTÉ – UNIVERSITÉ PARIS CITÉ

A2SUP

Bureau T203 – 45 rue des Saints-Pères, 75006 Paris
Bureau TP15 – 4 avenue de l'Observatoire, 75006 Paris
Bureau 101 – 15 rue de l'École de Médecine, 75006 Paris

01 42 86 40 59

www.a2sup.fr

Table des matières

La version numérique de ce tome, disponible sur le site `a2sup.fr`, contient également :

- UE9 – 2020-2021 Tuto 2
- UE9 – 2021-2022 Tuto 2
- UE9 – 2023-2024 Tuto 1
- UE9 – 2024-2025 Tuto 1

Pour la confection de ce tome, un grand merci à vos RM de Maths de cette année Gulliver Smith (DFGSM2) et Yazid RAZIK(DFGSM2), à vos VP Tuto PASS 25-26 Souad BENABDI (DFGSM2), Lùca BLONDEL-JORAND (DFGSM2), aux LaTeXiseurs, à la team MAghreb United chargée de la respo annales : Adam EL ABSI (DFGSM2, MA 2025-2026), Amira KHORSI (DFGSM2, MA 2025-2026) et Reza ISKER (DFGSP2, MA 2025-2026) et enfin à Semih YACIZI 🕸 (DFGSM2, RMEB Chimie 2025-2026), Abdussamed YAZICI (Césure médecine, 2e année de Thèse de Neurosciences, RMGT Maths 2021-2022) et Salma DRIS (Césure médecine, M2 BiP PPH, RM HE 2023-2024).

Bonjouuuuuur (et oui les tuto c'est fini #sad)

Tout d'abord, si tu es arrivé jusqu'ici, on tient à te dire qu'on est extrêmement **FIER DE TOI** ! Tu as travaillé pendant plus de 6 mois, réalisé des tonnes de QCM, GT, tuto et autres... Cette année n'est pas la plus facile, donc bravo de ne pas avoir lâché et de maintenant arriver à une des dernières grandes étapes : celles des ANNALES ✨. Tes VPs favoris vont alors te donner quelques conseils pour que cette période se déroule au mieux pour toi, et que tu sois prêt à briller pendant les exams.

Pour commencer : dès que tu reçois tes annales, prends un moment pour commencer à organiser tes deux semaines de révisions. Tu peux faire un **rétroplanning** pour être sûr d'avoir assez de temps pour alterner entre annales et révisions ! Commence par les tutos puis les sujets d'examen en laissant les plus récents (et donc les plus représentatifs) pour la fin. Laisse un ou deux jours libres avant les examens, au cas où tu aurais du retard dans ton planning, mais aussi pour ralentir le rythme et te reposer.

Surtout, n'oublie jamais que pour être efficace, il faut garder une **bonne hygiène de vie**. Et ça commence par le sommeil, aie une routine stable et un sommeil de qualité pendant cette période, ça aide vraiment. (petite anecdote de Souad : j'ai super mal dormi la veille des exams du S2 au profit de quelques dernières révisions, et j'ai regretté parce qu'à partir des dernières épreuves, c'était hyper dur de se concentrer, donc svp ne refaites pas mon erreur 😔)
Ensuite, on n'oublie pas de faire les annales en **conditions d'examen** : aucun bruit, montre en main, stylo noir et grilles de QCM. (dispo sur le site ✨)

Dernier point, prévois **beaucoup de temps pour les corrections**. C'est le point le plus important de l'annale, bien comprendre la correction pour ne plus refaire les mêmes erreurs. Privilégiez la qualité plutôt que la quantité : il est préférable de faire moins d'annales, mais en ayant bien compris plutôt que de toutes les bâcler. Écrivez vos erreurs et relisez les régulièrement !

On vous remets le barème pour les QCM (encore une fois, on ne se décourage pas pour une note à une annale, c'est hyper normal et vous ne pouvez que progresser durant ces revisions)

Nombre de différences par rapport à la correction	Points
0 différence	1 point
1 différence	0,7 point
2 différences	0,1 point
3 différences	0 point

Après celles du S1, les annales n'ont maintenant plus aucun secret pour toi, donc fais toi confiance et donne tout : the end is near 🙌.
Courage, donne ton maximum pour ce dernier semestre, on croit en toi ❤️

Bisous de LùSo,
Vos VPs Tuto PASS

BRAVO !

Simplement BRAVO d'avoir tenu jusqu'à maintenant ! Vous êtes les meilleurs, n'en doutez jamais. L'année arrive bientôt à sa fin et vous serez enfin ✨ libres ✨ ne lâchez rien. Aussi, un immense MERCI à tous les LaTeXiseurs de folies qui ont permis de vous présenter ces bébés et de vous fournir des supports de qualité.

Signé AmiRezAdam
-MAghreb Fitna : la team préca United

Salut les P1wans !!!

La période des annales du S2 commence 🔥, c'est maintenant que votre niveau va s'améliorer de façon e^x.

Vous pouvez être fier de ce que vous avez fait jusqu'ici et de ce que vous allez faire sur cette période, si tu ouvres ces tomes en pensant ne pas être prêt, tu dois savoir que c'est le cas de tout le monde. C'est maintenant que ça se joue et vos efforts auront des résultats. On te conseille de faire les sujets dans l'ordre chronologique si tu as le temps commence par les tutos, mais si c'est compliqué niveau timing fait les examens en priorité ❄️.

Le format de l'épreuve se compose de 13 à 15 questions à faire en 1 h 00 avec l'aide du formulaire et de la calculatrice, familiarisez vous +++ avec le formulaire pendant cette période pour être à l'aise le jour j ❤️.

Prenez votre temps sur la correction, c'est ce qui vous permettra de vous améliorer de sujet en sujet, posez vos questions sur le forum ou directement aux professeurs 👓.

Pour mathématiser, il faut bien manger 🍕 et essayer de dormir suffisamment. Ne négligez pas vos pauses sinon je viens dans vos rêves vous parler de la loi normale.

Bref, rappelez-vous de vos objectifs, chaque rencontre d'un obstacle est une occasion de grandir et si on tente sa chance, on a déjà gagné quoi qu'il se passe après 🏅. Le PASS ce n'est pas fastoche, mais vous êtes arrivés jusqu'ici et ça personne ne pourra jamais vous l'enlever. On vous dit que l'année n'est pas un sprint, mais un marathon, cependant la fin du marathon approche, finissez-le avec brio que ce soit en maths ou ailleurs. Faite le photo-finish !!!📷.

L'A2SUP te pousse, on te pousse, on sait que c'est difficile, mais avance, et quand tu auras grimpé le sommet P1, tu seras content du chemin accompli 🐦.

Association pour l'Accès Santé – Université Paris Cité
Année Universitaire 2022-2023

SUJET

Examen Blanc n°2 PASS

UE 9 : Mathématiques - Biostatistiques

Durée de l'épreuve : **1h**

A LIRE AVANT DE COMMENCER L'ÉPREUVE

Vérifiez que les informations saisies sur votre grille QCM sont correctes : nom, prénom et numéro étudiant.
Les correcteurs liquides ou en ruban de type Blanco, Tipp-Ex, et autres sont interdits car chaque question comporte une ligne de droit au remords.
Seule l'utilisation du stylo à bille noir est autorisée pour cocher les grilles.

INFORMATIONS RÉGLEMENTAIRES

- Les questions sans réponse seront considérées comme nulles.
- Une grille QCM est à remplir pour l'ensemble de l'épreuve.
- Veiller à remplir complètement toute la surface des cases choisies.
- Ne pas gratter, ne pas raturer, ne pas mettre de croix ni aucun autre signe.
- Toute fraude ou tentative de fraude fera l'objet de poursuites disciplinaires (Décret n°92-657 du 13 juillet 1992). Tout signe distinctif porté sur la grille QCM pouvant indiquer sa provenance constitue une fraude.
- Les calculatrices sont **autorisées**
- Aucun candidat n'est admis à quitter la salle d'examen avant la fin de l'épreuve.

RECOMMANDATIONS SPÉCIFIQUES À L'ÉPREUVE

INFORMATIONS SUR L'ÉPREUVE

Le sujet contient 4 pages numérotées de 1 à 4 et comporte 15 questions.
Merci de vérifier au début de l'épreuve que le sujet est complet.

Exercice 1 (Questions 1 à 5)

L'anesthésie générale permet de priver un patient des sensations douloureuses et de l'endormir lors d'une intervention chirurgicale. Celle-ci peut être obtenue par injection ou par inhalation. La répartition dans le corps de l'anesthésiant est suivie par le modèle (1) suivant :

$$(1) : \begin{cases} \frac{dx_{a_1}}{dt} &= -\alpha x_{a_1}(t) + \beta x_{a_2}(t) \\ \frac{dx_{a_2}}{dt} &= \gamma x_{a_1}(t) - \beta x_{a_2}(t) \end{cases}$$

Où $x_{a_1}(t)$ et $x_{a_2}(t)$ sont respectivement les doses d'anesthésiant dans le sang et dans les muscles et α, β, γ des constantes réelles positives.

Question 1 Soient C_1 et C_2 deux constantes réelles. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'équation caractéristique associée à $x_{a_1}(t)$ est $r^2 - (\alpha+\beta)r + \beta(\gamma - \alpha) = 0$
B. L'équation caractéristique associée à $x_{a_1}(t)$ est $r^2 + \alpha r - \beta = 0$
C. Pour $\alpha = \gamma$, $x_{a_1}(t) = e^{-(\alpha+\beta)t}(C_1 + C_2 t)$
D. Pour $\alpha = \gamma$, $x_{a_1}(t) = C_1 + C_2 e^{-(\alpha+\beta)t}$
E. Pour $\alpha > \gamma$, $x_{a_1}(t)$ admet deux solutions imaginaires

On s'intéresse à présent à l'écart entre la dose d'anesthésiant administrée x_a et la dose optimale x_e. On notera cet écart x tel que $x = x_a - x_e$ vérifiant le système (2) suivant :

$$(2) : \begin{cases} x'(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{cases}$$

Où $u(t)$ est le débit d'administration supposée constant tel que $u(0) = u_0$ $mg.min^{-1}$ et y l'état d'endormissement du patient. A, B et C sont des constantes réelles. On donne $x(0) = 0$ $mg.kg^{-1}$. Le temps est ici exprimé en minutes.

Question 2 **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Les équations du système (2) vérifient deux équations différentielles du second ordre
B. $x'(t) = Ax(t) + Bu(t)$ est une équation différentielle linéaire
C. $y(t)$ vérifie : $\frac{Bu_0}{A}(Ce^{At} - C)$
D. $x'(t)$ s'exprime en $mg.kg^{-1}$
E. Au voisinage de l'origine, à l'ordre 2, on peut approcher Ce^{At} par $C + CAt + \frac{C^2A^2t^2}{2}$

Question 3 L'état d'endormissement du patient peut être suivi par la variation de la surface pupillaire (VSP) qui s'exprime en pourcentage en fonction du temps en secondes. On obtient les données suivantes :

t (en s)	0	50	100	150	200	250	300
VSP (en %)	0	0	42	20	14	8	22

On souhaite approcher $V = \int_0^{300} \text{VSP}(t)dt$ par la méthode des trapèzes. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. On peut approcher V par $\frac{1}{50} \times (42 + 20 + 14 + 8 + 22)$
B. On peut approcher V par $\frac{1}{50} \times (42 + 20 + 14 + 8 + 11)$
C. On peut approcher V par $50 \times (42 + 20 + 14 + 8 + 11)$
D. $V \approx 4750$ %.s
E. $V \approx 1{,}9$ %.s

Le flux sanguin peut être suivi durant l'opération chirurgicale par échodoppler pulsé. Pour ce faire, les chirurgiens placent la sonde à un angle θ avec le vaisseau sanguin d'intérêt dans lequel les globules rouges circulent à une vitesse v. Un faisceau d'ultrasons de fréquence f et de vitesse C non nulle permet de calculer le flux d'écoulement sanguin. Cette méthode satisfait l'équation Doppler suivante :

$$\Delta f(f, v, \theta, C) = \frac{2fv\cos(\theta)}{C}$$

Des hémorragies peuvent survenir durant l'intervention, on modélise alors la perte de globules rouges par la fonction p suivante qui dépend du nombre de globules rouges n :

Pour tout $n > 1$,

$$p(n) = \frac{6n^2 - 11n + 8}{1 - 3n}$$

Question 4 Soit $P(n) = \int p(n)dn$. Soit k une constante. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. $\frac{\partial(\Delta f)}{\partial \theta} = \frac{2fv}{C}$
B. $\frac{\partial(\Delta f)}{\partial C} = \frac{2fv\cos(\theta)}{C^2}$
C. Par division euclidienne, on peut écrire $p(n) = -3n + 1 + \frac{5}{1-3n}$
D. $P(n) = -\frac{3n^2}{2} + n - 5\ln|1 - 3n| + k$
E. $P(n) = -n^2 + 3n - \frac{5}{3}\ln|1 - 3n| + k$

Les patients opérés présentent un risque de développer une infection virale ou bactérienne à la suite de leur intervention. Lors d'une infection virale, on peut modéliser l'évolution de la concentration d'antigène [Ag] au cours du temps par l'équation différentielle suivante :

Pour tout $t \geq 0$,

$$[Ag]'(t) = \varepsilon V(t) - \delta[Ag](t)$$

Où $V(t) = e^{2t}$ représente la quantité de virus, ε correspond au taux de mortalité des virus et δ la constante de dégradation des antigènes ($\delta > 0$). On précise qu'au temps initial, la quantité de virus est nulle. On pourra retirer la dépendance au temps ($[Ag]$ au lieu de $[Ag](t)$) dans le but de simplifier les écritures.

Question 5 Soit S une constante. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. La solution générale de l'équation différentielle sans second membre est $Se^{-\delta t}$
B. La solution générale de l'équation différentielle sans second membre est $\varepsilon e^{\delta t}$
C. Une solution particulière de l'équation différentielle avec second membre est $\frac{\varepsilon}{\delta}e^{\delta t}$
D. La solution générale de l'équation avec second membre est $\frac{\varepsilon}{2+\delta}\left(-e^{-\delta t} + e^{2t}\right)$
E. La solution générale de l'équation avec second membre est $\frac{\varepsilon}{\delta}\left(e^{\delta t} - e^{2t}\right)$

Exercice 2 (Questions 6 à 10)

Dans un hôpital comptant 450 patients, 150 patients sont âgés de moins de 30 ans. Parmi eux, 2% présentent une atteinte de l'appareil locomoteur. Chez les patients âgés de plus de 30 ans, 100 ont un déficit en vitamine D, parmi eux 4% ont une atteinte de l'appareil locomoteur ; contre 3% pour ceux n'ayant pas de déficit en vitamine D.

Question 6 Soit n le nombre de patients de plus de 30 ans n'ayant pas de déficit de vitamine D qui présente une atteinte de l'appareil locomoteur. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. $n = 3$
B. $n = 4$
C. $n = 5$
D. $n = 6$
E. $n = 7$

Question 7 **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. La probabilité d'avoir plus de 30 ans et d'avoir un déficit en vitamine D sachant qu'on a une atteinte de l'appareil locomoteur est de 0,31 (à 0,01 près)
B. La probabilité d'avoir plus de 30 ans et d'avoir un déficit en vitamine D sachant qu'on a une atteinte de l'appareil locomoteur est de 0,35 (à 0,01 près)
C. La probabilité d'avoir plus de 30 ans et d'avoir un déficit en vitamine D sachant qu'on a une atteinte de l'appareil locomoteur est de 0,40 (à 0,01 près)
D. La probabilité de ne pas avoir d'atteinte de l'appareil locomoteur est de 0,90 (à 0,01 près)
E. La probabilité de ne pas avoir d'atteinte de l'appareil locomoteur est de 0,97 (à 0,01 près)

Question 8 Si on considère maintenant que chacun des $n = 450$ patients a la même probabilité $p = 0{,}01$ d'avoir une atteinte de l'appareil locomoteur. X est la variable aléatoire qui compte le nombre d'atteintes de l'appareil locomoteur (on considère que les survenues des atteintes sont indépendantes entre elles). **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. X est une variable aléatoire discrète
B. X suit rigoureusement une loi binomiale $\mathcal{B}in(n = 450, p = 0{,}01)$
C. X suit rigoureusement une loi normale $N(n = 450, p = 0{,}01)$
D. X suit rigoureusement une loi de Poisson $\mathcal{P}ois(\lambda)$ avec $\lambda = np$
E. Lorsque $n \geq 20$ et $p < 0{,}01$, on peut approximer une loi binomiale de paramètres n et p par une loi de Poisson de paramètre $\lambda = np$

Question 9 Une blessure fréquente est l'entorse légère. La durée du rétablissement d'une entorse légère suit une loi exponentielle de paramètre $\lambda = 0{,}25\ j^{-1}$. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Au bout de 2 jours, la probabilité d'être guéri est de 0,4 à 0,1 près
B. Au bout de 2 jours, la probabilité d'être guéri est de 0,6 à 0,1 près
C. La moyenne de la durée de rétablissement est de 3 jours
D. La moyenne de la durée de rétablissement est de 4 jours
E. Si au bout de deux jours l'entorse n'a toujours pas guérie, la probabilité qu'elle soit guérie deux jours plus tard est de 0,6 (à 0,1 près)

Question 10 Parmi les patients, 10% ont des ecchymoses. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Sur un groupe de 30 patients, l'intervalle de pari à 95% des patients présentant des ecchymoses est $[0; 6]$
B. Sur un groupe de 60 patients, l'intervalle de pari à 95% des patients présentant des ecchymoses est $[1; 11]$
C. Sur un groupe de 30 patients, l'intervalle de confiance à 95% des patients présentant des ecchymoses est $[0; 6]$
D. Sur un groupe de 60 patients, l'intervalle de confiance à 95% des patients présentant des ecchymoses est $[1; 11]$
E. Lorsque le risque α diminue, la largeur de l'intervalle de pari à $1-\alpha$ du nombre de patients présentant des ecchymoses augmente

Exercice 3 (Questions 11 à 15)

Le raton laveur est un animal qui peut transmettre un certain nombre de maladies. On s'intéresse à deux populations de ratons laveurs : les ratons américains et les ratons européens. Ces populations sont indépendantes.

On étudie la longueur de la queue d'un échantillon de ratons américains d'effectif $n_a = 81$. On obtient une longueur de queue moyenne de $m_a = 26\ cm$ avec une variance de $s_a^2 = 9\ cm^2$.

Question 11 Concernant l'intervalle de confiance à 99% de la moyenne théorique μ_a de la longueur de la queue des ratons laveurs américains, **parmi les propositions suivantes, laquelle est exacte ?**

A. $IC_{99\%}(\mu_a) = [20; 30] \pm 0{,}1$ cm

B. $IC_{99\%}(\mu_a) = [25{,}1; 26{,}9] \pm 0{,}1$ cm

C. $IC_{99\%}(\mu_a) = [25{,}3; 26{,}7] \pm 0{,}1$ cm

D. $IC_{99\%}(\mu_a) = [25; 27{,}1] \pm 0{,}1$ cm

E. Les conditions ne sont pas réunies pour appliquer le théorème central limite

Question 12 On compare la longueur de la queue de la population de ratons laveurs américaine avec une population d'origine européenne, d'effectif $n_e = 162$, de moyenne $m_e = 29\ cm$, et de variance $s_e^2 = 10\ cm^2$. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'hypothèse nulle est $H_0 : \mu_a \neq \mu_e$

B. Les longueurs de queues moyennes sont significativement différentes entre les deux populations au risque 5%

C. On ne met pas en évidence de différence de longueur de queue entre les deux populations au risque 5%

D. Si on rejette l'hypothèse nulle H_0, on peut commettre une erreur de type 1

E. Il faut que la moyenne des effectifs des deux groupes soit supérieure à 30 pour respecter les conditions de validité

Question 13 On compare à présent la longueur de la queue des ratons laveurs américains à une population de référence, les ratons laveurs captifs dont la longueur de la queue moyenne vaut $\mu_c = 25\ cm$. On a déjà réalisé un test de comparaison de la population de ratons captifs avec l'échantillon de ratons laveur américains, et on a trouvé la statistique de test $Z = 3$. On suppose alors la longueur de la queue dans la population de ratons laveurs américains est 1 cm supérieure à celle des ratons laveurs captifs. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Il y a une différence significative entre les ratons américains et les ratons captifs au risque 5%

B. On n'oberve pas de différence significative entre les deux populations au risque 5%

C. L'hypothèse alternative est $H_1 : m_a = \mu_c + 1$

D. Le risque β (ou erreur de type II) est de 83% (à 1% près)

E. Le degré de signification est compris entre 10^{-3} et 10^{-2}

Question 14 Certains ratons laveurs sont atteints de la rage. Parmi les ratons laveurs sauvages américains, on observe une proportion $p_a = 10\%$ d'individus atteints de la rage. Dans la population des ratons laveurs captifs, prise comme référence, cette proportion vaut $\pi_c = 4\%$. On compare ces populations. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'hypothèse nulle est $H_0 : \pi_a = \pi_c$

B. Pour comparer la proportion dans un échantillon à une proportion théorique, on peut faire un test du χ^2 à 1 degré de liberté

C. Il y a une différence significative dans la prévalence de la rage entre les deux populations au risque 5%

D. On ne rejette pas l'hypothèse nulle au risque 5%

E. Il n'y a pas de différence significative entre les deux populations au risque 5%

Question 15 On cherche à déterminer maintenant s'il y a un lien entre le nombre de bandes sur la queue des ratons laveurs et la population d'origine. Les données recueillies sont présentées dans le tableau suivant :

Nombre de bandes	5	6	7	Total
Ratons américains	13	20	48	81
Ratons européens	36	29	97	162
Total	49	49	145	243

On réalise un test du χ^2, et on trouve $K = 2{,}3$. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'hypothèse alternative H_1 suppose une corrélation entre l'origine géographique et le nombre de bandes sur la queue

B. On fait un test du χ^2 à 6 degrés de liberté

C. On fait un test de comparaison de moyennes

D. Les conditions de validité sont vérifiées car tous les effectifs observés sont supérieurs à 5

E. On peut affirmer au risque 5% qu'il n'y a pas de corrélation entre les origines géographiques et le nombre de bandes sur la queue

FIN DU SUJET

Association pour l'Accès Santé – Université Paris Cité
Année Universitaire 2022-2023

CORRECTION

Examen Blanc n°2 PASS

UE 9 : Mathématiques - Biostatistiques

Durée de l'épreuve : **1h**

A LIRE AVANT DE COMMENCER L'ÉPREUVE

Vérifiez que les informations saisies sur votre grille QCM sont correctes : nom, prénom et numéro étudiant.
Les correcteurs liquides ou en ruban de type Blanco, Tipp-Ex, et autres sont interdits car chaque question comporte une ligne de droit au remords.
Seule l'utilisation du stylo à bille noir est autorisée pour cocher les grilles.

INFORMATIONS RÉGLEMENTAIRES

- Les questions sans réponse seront considérées comme nulles.
- Une grille QCM est à remplir pour l'ensemble de l'épreuve.
- Veiller à remplir complètement toute la surface des cases choisies.
- Ne pas gratter, ne pas raturer, ne pas mettre de croix ni aucun autre signe.
- Toute fraude ou tentative de fraude fera l'objet de poursuites disciplinaires (Décret n°92-657 du 13 juillet 1992). Tout signe distinctif porté sur la grille QCM pouvant indiquer sa provenance constitue une fraude.
- Les calculatrices sont **autorisées**
- Aucun candidat n'est admis à quitter la salle d'examen avant la fin de l'épreuve.

RECOMMANDATIONS SPÉCIFIQUES À L'ÉPREUVE

INFORMATIONS SUR L'ÉPREUVE

Le sujet contient 13 pages numérotées de 1 à 13 et comporte 15 questions.
Merci de vérifier au début de l'épreuve que le sujet est complet.

Université Paris Cité
A2SUP - Tutorat

Tuto n° : 2　　UE (spé) : 9

Nom :
Prénom :
Numéro A2SUP :

IDENTIFICATION

Numéro A2SUP : Diz. Mil., Mil., Cent., Diz., Unit. (0 1 2 3 4 5 6 7 8 9)

Contrôle : M Unit., M Diz., J Unit., J Diz. (0 1 2 3 4 5 6 7 8 9)
Date de naissance (JJ/MM)

	A	B	C	D	E
1	□	□	□	■	□
2	□	■	■	□	□
3	□	□	■	■	□
4	□	□	□	□	■
5	■	□	□	■	□
6	□	□	□	■	□
7	■	□	□	□	■
8	■	■	□	□	■
9	■	□	□	■	□
10	□	■	□	□	■
11	□	■	□	□	□
12	□	■	□	■	□
13	■	□	□	□	■
14	■	■	□	□	□
15	■	□	□	□	□
16	□	□	□	□	□
17	□	□	□	□	□
18	□	□	□	□	□
19	□	□	□	□	□
20	□	□	□	□	□
21	□	□	□	□	□
22	□	□	□	□	□
23	□	□	□	□	□
24	□	□	□	□	□
25	□	□	□	□	□
26	□	□	□	□	□
27	□	□	□	□	□
28	□	□	□	□	□
29	□	□	□	□	□
30	□	□	□	□	□
31	□	□	□	□	□
32	□	□	□	□	□
33	□	□	□	□	□
34	□	□	□	□	□
35	□	□	□	□	□
36	□	□	□	□	□
37	□	□	□	□	□
38	□	□	□	□	□
39	□	□	□	□	□
40	□	□	□	□	□
41	□	□	□	□	□
42	□	□	□	□	□
43	□	□	□	□	□
44	□	□	□	□	□
45	□	□	□	□	□

Exercice 1 (Questions 1 à 5)

L'anesthésie générale permet de priver un patient des sensations douloureuses et de l'endormir lors d'une intervention chirurgicale. Celle-ci peut être obtenue par injection ou par inhalation. La répartition dans le corps de l'anesthésiant est suivie par le modèle (1) suivant :

$$(1) : \begin{cases} \frac{dx_{a_1}}{dt} &= -\alpha x_{a_1}(t) + \beta x_{a_2}(t) \\ \frac{dx_{a_2}}{dt} &= \gamma x_{a_1}(t) - \beta x_{a_2}(t) \end{cases}$$

Où $x_{a_1}(t)$ et $x_{a_2}(t)$ sont respectivement les doses d'anesthésiant dans le sang et dans les muscles et α, β, γ des constantes réelles positives.

Question 1 Soient C_1 et C_2 deux constantes réelles. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'équation caractéristique associée à $x_{a_1}(t)$ est $r^2 - (\alpha+\beta)r + \beta(\gamma-\alpha) = 0$
B. L'équation caractéristique associée à $x_{a_1}(t)$ est $r^2 + \alpha r - \beta = 0$
C. Pour $\alpha = \gamma$, $x_{a_1}(t) = e^{-(\alpha+\beta)t}(C_1 + C_2 t)$
D. Pour $\boldsymbol{\alpha = \gamma}$, $\boldsymbol{x_{a_1}(t) = C_1 + C_2 e^{-(\alpha+\beta)t}}$
E. Pour $\alpha > \gamma$, $x_{a_1}(t)$ admet deux solutions imaginaires

Question 1

Items A et B

Pour simplifier l'écriture du système, on l'écrira sans les t :

$$\begin{cases} {x'}_{a_1} &= -\alpha x_{a_1} + \beta x_{a_2} \\ {x'}_{a_2} &= \gamma x_{a_1} - \beta x_{a_2} \end{cases}$$

On cherche l'équation caractéristique associée à $x_{a_1}(t)$. On commence par dériver la première ligne :

$$({x'}_{a_1})' = (-\alpha x_{a_1} + \beta x_{a_2})'$$

Soit :

$$x''_{a_1} = -\alpha x'_{a_1} + \beta x'_{a_2}$$

On injecte la deuxième ligne pour remplacer x'_{a_2} par son expression :

$$x''_{a_1} = -\alpha x'_{a_1} + \beta(\gamma x_{a_1} - \beta x_{a_2}) = -\alpha x'_{a_1} + \beta\gamma x_{a_1} - \beta^2 x_{a_2}$$

Or,

$$x'_{a_1} = -\alpha x_{a_1} + \beta x_{a_2}$$

Donc :

$$\beta x_{a_2} = x'_{a_1} + \alpha x_{a_1}$$

Ce qui nous donne :

$$\begin{aligned} x''_{a_1} &= -\alpha x'_{a_1} + \beta\gamma x_{a_1} - \beta \times (x'_{a_1} + \alpha x_{a_1}) \\ &= -(\alpha+\beta)x'_{a_1} + \beta(\gamma-\alpha)x_{a_1} \end{aligned}$$

On réarrange un peu notre équation :

$$x''_{a_1} + (\alpha+\beta)x'_{a_1} - \beta(\gamma-\alpha)x_{a_1} = 0$$

On remplace ainsi x_{a_1} par $r^0 = 1$, x'_{a_1} par $r^1 = r$ et x''_{a_1} par r^2. Donc :

$$r^2 + (\alpha+\beta)r - \beta(\gamma-\alpha) = 0$$

(A FAUX) (B FAUX)

Items C et D

On calcule le discriminant du système :

$$\Delta = (\alpha+\beta)^2 + 4\beta(\gamma-\alpha)$$

Si $\alpha = \gamma$

$$\Delta = (\alpha+\beta)^2$$

On nous précisait que les constantes α et β sont réelles positives donc le discriminant ici est positif. On aura donc deux racines réelles :

$$r_1 = \frac{-(\alpha+\beta) + (\alpha+\beta)}{2} = 0$$

Et :

$$r_2 = \frac{-(\alpha+\beta) - (\alpha+\beta)}{2} = -(\alpha+\beta)$$

Là vous pouviez utiliser le formulaire pour retrouver l'expression de la solution :

$$x_{a_1}(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t}$$

Or $r_1 = 0$:

$$x_{a_1}(t) = C_1 e^{0 \times t} + C_2 e^{r_2 t}$$

Soit :

$$x_{a_1}(t) = C_1 + C_2 e^{-(\alpha+\beta)t}$$

(C FAUX) **(D VRAI)**

✘ Item E → *Pour $\alpha > \gamma$, $x_{a_1}(t)$ admet deux solutions imaginaires*

C'était un item doublement faux ! Si $\alpha > \gamma$, on ne peut rien conclure car on aura :

$$\begin{cases} \Delta = (\alpha+\beta)^2 + 4\beta(\gamma-\alpha) \\ 4\beta(\gamma-\alpha) < 0 \end{cases}$$

Mais on ne pourra rien dire sur le signe de Δ. Et quand bien même Δ était négatif, on aurait eu deux RACINES imaginaires conjuguées mais une seule solution pour $x_{a_1}(t)$!! *(E FAUX)*

Réponse vraie : D

On s'intéresse à présent à l'écart entre la dose d'anesthésiant administrée x_a et la dose optimale x_e.

2

On notera cet écart x tel que $x = x_a - x_e$ vérifiant le système (2) suivant :

$$(2) : \begin{cases} x'(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{cases}$$

Où $u(t)$ est le débit d'administration supposée constant tel que $u(0) = u_0$ $mg.min^{-1}$ et y l'état d'endormissement du patient. A, B et C sont des constantes réelles. On donne $x(0) = 0$ $mg.kg^{-1}$. Le temps est ici exprimé en minutes.

Question 2 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les équations du système (2) vérifient deux équations différentielles du second ordre

B. $x'(t) = Ax(t) + Bu(t)$ est une équation différentielle linéaire

C. $y(t)$ vérifie : $\frac{Bu_0}{A}(Ce^{At} - C)$

D. $x'(t)$ s'exprime en $mg.kg^{-1}$

E. Au voisinage de l'origine, à l'ordre 2, on peut approcher Ce^{At} par $C + CAt + \frac{C^2A^2t^2}{2}$

Question 2

✘ **Item A** → *Les équations du système (2) vérifient deux équations différentielles du second ordre*

Revoici le système :

$$(2) : \begin{cases} x'(t) = Ax(t) + Bu(t) & (*) \\ y(t) = Cx(t) & (**) \end{cases}$$

C'est un système comportant une seule équation différentielle du premier ordre. On peut réécrire la première ligne du système $(*)$:

$$x'(t) = Ax(t) + Bu_0$$

Car $u(t) = u_0$ (u étant une fonction constante).

Il s'agit donc d'une équation différentielle du premier ordre (car ne dépend que de la dérivée première) à second membre constant :

$$x'(t) - Ax(t) = Bu_0 \quad (A\ FAUX)$$

✔ **Item B** → *$x'(t) = Ax(t) + Bu(t)$ est une équation différentielle linéaire*

Une équation différentielle linéaire s'écrit sous la forme :

$$a_0y + a_1y' + a_2y'' + \ldots + a_ny^{(n)} = f(t)$$

Ici, l'équation différentielle s'écrit :

$$-Ax(t) + x'(t) = Bu_0$$

On retrouve bien la même forme, avec $a_0 = A$, $a_1 = 1$, et $f(t) = Bu_0$ **(B VRAI)**

✔ **Item C** → *$y(t)$ vérifie : $\frac{Bu_0}{A}(Ce^{At} - C)$*

On résout l'équation différentielle en passant bien par les 4 étapes à savoir :

Etapes de la résolution d'une équation différentielle

1. La solution générale de l'équation sans second membre : **SGESSM**
2. La solution particulière de l'équation avec second membre : **SPEASM**
3. La solution générale de l'équation avec second membre : **SGEASM**
4. Détermination de la constante d'intégration (apparue lors de la première étape)

SGESSM : x_h

On résout :

$$x_h' - Ax_h = 0$$

Soit :

$$\begin{aligned} & \frac{dx_h}{dt} = Ax_h \\ \iff & \frac{dx_h}{x_h} = Adt \\ \implies & \ln|x_h| = At + cste \\ \iff & e^{\ln|x_h|} = e^{At+cste} \\ \iff & x_h = Ke^{At} \quad (K \in \mathbb{R}) \end{aligned}$$

SPEASM : X

Ici notre second membre est une constante : Bu_0. Donc la solution particulière sera également une constante X :

$$X' - AX = Bu_0$$

Comme X = cste, on a $X' = 0$:

$$-AX = Bu_0$$

Ce qui nous donne :

$$X = -\frac{Bu_0}{A}$$

SGEASM : x

3

$$x(t) = x_h + X = Ke^{At} - \frac{Bu_0}{A}$$

La constante K une fois que la SGEASM a été déterminée ! On trouve cette constante généralement à partir des conditions initiales :

$$x(0) = 0 \iff Ke^{A\times 0} - \frac{Bu_0}{A} = 0 \iff K = \frac{Bu_0}{A}$$

Donc :

$$x(t) = \frac{Bu_0}{A}(e^{At} - 1)$$

Comme $y(t) = Cx(t)$ ((**) de (2)), on a :

$$y(t) = \frac{CBu_0}{A}(e^{At} - 1)$$

Soit :

$$y(t) = \frac{Bu_0}{A}(Ce^{At} - C) \quad \textbf{(C VRAI)}$$

✘ Item D → *$x'(t)$ s'exprime en $mg.kg^{-1}$*

On nous demande ici l'unité de $x'(t)$. On peut écrire comme en fysik (oh le comeback incroyable la team fysik vous êtes les plus beaux) :

$$x'(t) = \frac{dx}{dt}$$

On connaît l'unité de x qui sera la même que celle de dx : $mg.kg^{-1}$ (dx n'est qu'une variation de x, donc les deux auront la même unité). Et dt représente un temps en min. Donc $x'(t)$ aura pour unité des $mg.kg^{-1}.min^{-1}$ *(D FAUX)*.

✘ Item E → *Au voisinage de l'origine, à l'ordre 2, on peut approcher Ce^{At} par $C + CAt + \frac{C^2A^2t^2}{2}$*

Vous pouviez ici reprendre votre formulaire et chercher le développement limité (DL) de e^x au voisinage de l'origine à l'ordre 2. Votre formulaire nous donne :

$$e^x = 1 + x + \frac{x^2}{2} + \text{reste}$$

Ici, on cherchait le DL de Ce^{At} on utilise la formule en remplaçant x par At (ce qu'on a bien le droit de faire car, lorsque t tend vers 0, At tend vers 0, ainsi l'argument de l'exponentielle est bien au voisinage de 0).

$$Ce^{At} \approx C\left(1 + At + \frac{At^2}{2}\right)$$

Ce qui nous donnait :

$$Ce^{At} \approx C + CAt + C \times \frac{(At)^2}{2} \approx C + CAt + \frac{CA^2t^2}{2} \quad \textit{(E FAUX)}$$

Réponses vraies : B et C

Question 3 L'état d'endormissement du patient peut être suivi par la variation de la surface pupillaire (VSP) qui s'exprime en pourcentage en fonction du temps en secondes. On obtient les données suivantes :

t (en s)	0	50	100	150	200	250	300
VSP (en %)	0	0	42	20	14	8	22

On souhaite approcher $V = \int_0^{300} \text{VSP}(t)dt$ par la méthode des trapèzes. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. On peut approcher V par $\frac{1}{50} \times (42 + 20 + 14 + 8 + 22)$
B. On peut approcher V par $\frac{1}{50} \times (42 + 20 + 14 + 8 + 11)$
C. On peut approcher V par $50 \times (42 + 20 + 14 + 8 + 11)$
D. $V \approx 4750$ %.s
E. $V \approx 1{,}9$ %.s

Question 3

? Items A, B et C

On a une petite méthode des trapèzes à appliquer ici. On se rappelle que c'est une méthode qui permet d'approcher l'aire sous une courbe en approchant cette aire par une succession de trapèzes. La formule qui s'applique est :

$$V \approx \text{pas} \times \left(\frac{\text{VSP}(t_1)}{2} + \text{VSP}(t_2) + \ldots + \frac{\text{VSP}(t_n)}{2}\right)$$

On identifie le pas ici, c'est-à-dire l'espace entre deux valeurs. On remarque qu'on avance de 50 en 50 donc notre pas vaudra 50 :

$$V \approx 50 \times \left(\frac{0}{2} + 0 + 42 + 20 + 14 + 8 + \frac{22}{2}\right) \approx 50 \times (42 + 20 + 14 + 8 + \frac{22}{2})$$

(A FAUX) (B FAUX) **(C VRAI)**

4

? Items D et E

On pouvait faire le calcul sans calculatrice en partant de l'idée que $50 = \frac{100}{2}$. Ainsi :

$$\begin{aligned} V &\approx \frac{100}{2} \times (42 + 20 + 14 + 8 + \frac{22}{2}) \\ &\approx 100 \times \frac{(42 + 20 + 14 + 8 + \frac{22}{2})}{2} \\ &\approx 100 \times (21 + 10 + 7 + 4 + 5{,}5) \\ &\approx 4750\ s \quad \textbf{(D VRAI)} \quad \textit{(E FAUX)} \end{aligned}$$

Réponses vraies : C et D

Le flux sanguin peut être suivi durant l'opération chirurgicale par échodoppler pulsé. Pour ce faire, les chirurgiens placent la sonde à un angle θ avec le vaisseau sanguin d'intérêt dans lequel les globules rouges circulent à une vitesse v. Un faisceau d'ultrasons de fréquence f et de vitesse C non nulle permet de calculer le flux d'écoulement sanguin. Cette méthode satisfait l'équation Doppler suivante :

$$\Delta f(f, v, \theta, C) = \frac{2fv\cos(\theta)}{C}$$

Des hémorragies peuvent survenir durant l'intervention, on modélise alors la perte de globules rouges par la fonction p suivante qui dépend du nombre de globules rouges n :

Pour tout $n > 1$,

$$p(n) = \frac{6n^2 - 11n + 8}{1 - 3n}$$

Question 4 Soit $P(n) = \int p(n)dn$. Soit k une constante. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. $\frac{\partial(\Delta f)}{\partial\theta} = \frac{2fv}{C}$
B. $\frac{\partial(\Delta f)}{\partial C} = \frac{2fv\cos(\theta)}{C^2}$
C. Par division euclidienne, on peut écrire $p(n) = -3n + 1 + \frac{5}{1-3n}$
D. $P(n) = -\frac{3n^2}{2} + n - 5\ln|1 - 3n| + k$
E. $P(n) = -n^2 + 3n - \frac{5}{3}\ln|1 - 3n| + k$

Question 4

✗ Item A → $\frac{\partial(\Delta f)}{\partial\theta} = \frac{2fv}{C}$

On calcule la dérivée partielle de la fonction Δf par rapport à θ. On va donc considérer que tous les autres termes sont constants et on aura :

$$\frac{\partial(\Delta f)}{\partial\theta} = \frac{2fv}{C} \times (\cos(\theta))'$$

Or $(\cos(\theta))' = -\sin(\theta)$, donc :

$$\frac{\partial(\Delta f)}{\partial\theta} = -\frac{2fv}{C} \times \sin(\theta) \quad \textit{(A FAUX)}$$

✗ Item B → $\frac{\partial(\Delta f)}{\partial C} = \frac{2fv\cos(\theta)}{C^2}$

On fait la même chose en dérivant cette fois-ci par rapport à C en considérant le reste comme constant :

$$\frac{\partial(\Delta f)}{\partial C} = 2fv\cos(\theta) \times \left(\frac{1}{C}\right)'$$

Or, $\left(\frac{1}{C}\right)' = -\frac{1}{C^2}$, donc :

$$\begin{aligned} \frac{\partial(\Delta f)}{\partial C} &= 2fv\cos(\theta) \times \left(-\frac{1}{C^2}\right) \\ &= \frac{-2fv\cos(\theta)}{C^2} \quad \textit{(B FAUX)} \end{aligned}$$

✗ Item C → *Par division euclidienne, on peut écrire* $p(n) = -3n + 1 + \frac{5}{1-3n}$

On a ici un quotient de polynômes où le numérateur est de degré supérieur à celui du dénominateur. On va donc devoir faire une division euclidienne :

$$\begin{array}{rrr|rr} +6n^2 & -11n & +8 & -3n & +1 \\ \cline{4-5} -(6n^2 & -2n) & & -2n & +3 \\ \cline{1-2} 0n^2 & -9n & +8 & & \\ & -(-9n & +3) & & \\ \cline{2-3} & 0n & +5 & & \end{array}$$

On peut donc écrire la fonction p sous la forme suivante :

$$p(n) = -2n + 3 + \frac{5}{1 - 3n} \quad \textit{(C FAUX)}$$

? Items D et E

On cherche à intégrer $p(n)$:

$$\begin{aligned} P(n) &= \int \left(-2n + 3 + \frac{5}{1-3n}\right) dn \\ &= \int -2n\ dn + \int 3\ dn + \int \frac{5}{1-3n}\ dn \\ &= -n^2 + 3n + 5\int \frac{1}{1-3n}\ dn \end{aligned}$$

Calculons $\int \frac{1}{1-3n}\ dn$:

$$\begin{aligned} \int \frac{1}{1-3n}\ dn &= \int \left(-\frac{1}{3} \times \frac{-3}{1-3n}\right) dn \\ &= -\frac{1}{3} \times \ln|1 - 3n| + k \quad (k \in \mathbb{R}) \end{aligned}$$

Ainsi :

$$P(n) = -n^2 + 3n + 5 \times (-\frac{1}{3})\ln|1 - 3n| + k$$

(D FAUX) **(E VRAI)**

Réponse vraie : E

Les patients opérés présentent un risque de développer une infection virale ou bactérienne à la suite de leur intervention. Lors d'une infection virale, on peut modéliser l'évolution de la concentration d'antigène [Ag] au cours du temps par l'équation différentielle suivante :

Pour tout $t \geq 0$,

$$[Ag]'(t) = \varepsilon V(t) - \delta [Ag](t)$$

Où $V(t) = e^{2t}$ représente la quantité de virus, ε correspond au taux de mortalité des virus et δ la constante de dégradation des antigènes ($\delta > 0$). On précise qu'au temps initial, la quantité de virus est nulle. On pourra retirer la dépendance au temps ($[Ag]$ au lieu de $[Ag](t)$) dans le but de simplifier les écritures.

Question 5 Soit S une constante. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. La solution générale de l'équation différentielle sans second membre est $Se^{-\delta t}$

B. La solution générale de l'équation différentielle sans second membre est $\varepsilon e^{\delta t}$

C. Une solution particulière de l'équation différentielle avec second membre est $\frac{\varepsilon}{\delta} e^{\delta t}$

D. La solution générale de l'équation avec second membre est $\frac{\varepsilon}{2+\delta}\left(-e^{-\delta t} + e^{2t}\right)$

E. La solution générale de l'équation avec second membre est $\frac{\varepsilon}{\delta}\left(e^{\delta t} - e^{2t}\right)$

Question 5

? Items A et B

On a ici une équation différentielle du premier ordre avec second membre. On commence par résoudre l'équation sans second membre (pour trouver la solution générale de l'équation sans second membre qu'on notera $[Ag]_0$) :

$$[Ag]_0' = -\delta [Ag]_0$$

On peut écrire cette équation sous la forme suivante :

$$\frac{d[Ag]_0}{dt} = -\delta [Ag]_0$$

Soit :

$$\frac{d[Ag]_0}{[Ag]_0} = -\delta dt$$

En intégrant de chaque côté, on obtient :

$$\ln [Ag]_0 = -\delta t + cste$$

On applique la fonction exponentielle sur les deux membres :

$$[Ag]_0 = Se^{-\delta t} \qquad \textit{(B FAUX)} \;\; \textbf{(A VRAI)}$$

✘ Item C → *Une solution particulière de l'équation différentielle avec second membre est $\frac{\varepsilon}{\delta} e^{\delta t}$*

On cherche à présent la solution particulière de l'équation avec second membre, notée $[Ag]_p$. On nous donnait $\varepsilon V(t) = \varepsilon e^{2t}$.

Donc, d'après le formulaire, la solution particulière sera de la forme Te^{2t} ($T \in \mathbb{R}$) :

$$\overbrace{2Te^{2t}}^{(Te^{2t})'} = \varepsilon e^{2t} - \delta T e^{2t}$$

Soit :

$$\begin{aligned} & 2Te^{2t} + \delta T e^{2t} = \varepsilon e^{2t} \\ \iff & 2T + \delta T = \varepsilon \\ \iff & T(2+\delta) = \varepsilon \\ \iff & T = \frac{\varepsilon}{2+\delta} \end{aligned}$$

Donc la solution particulière de l'équation avec second membre est :

$$[Ag]_p = \frac{\varepsilon}{2+\delta} e^{2t} \qquad \textit{(C FAUX)}$$

? Item D et E

On aura donc pour la solution générale de l'équation avec second membre :

$$\begin{aligned} [Ag] &= [Ag]_0 + [Ag]_p \\ &= Se^{-\delta t} + \frac{\varepsilon}{2+\delta} e^{2t} \end{aligned}$$

Pour $t = 0$,

$$[Ag](t=0) = Se^{-\delta \times 0} + \frac{\varepsilon}{2+\delta} e^{2\times 0} = 0$$

Soit :

$$\begin{aligned} & S + \frac{\varepsilon}{2+\delta} = 0 \\ \iff & S = -\frac{\varepsilon}{2+\delta} \end{aligned}$$

Donc :

$$[Ag] = \frac{\varepsilon}{2+\delta}\left(-e^{-\delta t} + e^{2t}\right) \qquad \textit{(E FAUX)} \;\; \textbf{(D VRAI)}$$

Réponses vraies : A et D

Exercice 2 (Questions 6 à 10)

Dans un hôpital comptant 450 patients, 150 patients sont âgés de moins de 30 ans. Parmi eux, 2% présentent une atteinte de l'appareil locomoteur. Chez les patients âgés de plus de 30 ans, 100 ont un déficit en vitamine D, parmi eux 4% ont une atteinte de l'appareil locomoteur ; contre 3% pour ceux n'ayant pas de déficit en vitamine D.

Question 6 Soit n le nombre de patients de plus de 30 ans n'ayant pas de déficit de vitamine D qui présente une atteinte de l'appareil locomoteur. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

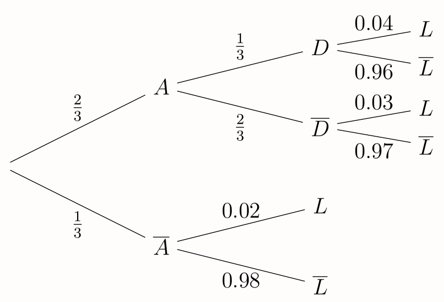
A. $n = 3$
B. $n = 4$
C. $n = 5$
D. $n = 6$
E. $n = 7$

Question 6

? Items A, B, C, D et E

Il vaut mieux tracer l'arbre de probabilité. Soit :

- A l'évènement "avoir plus de 30 ans" et $\overline{A}$ sont complémentaires
- D l'évènement "avoir un déficit en vitamine D" et $\overline{D}$ sont complémentaires
- L l'évènement "avoir une atteinte de l'appareil locomoteur" et $\overline{L}$ sont complémentaires



Sur les 450 patients de départ, 150 ont moins de 30 ans, donc 300 ont plus de 30 ans. Parmi ces 300 patients ayant plus de 30 ans, 100 ont un déficit de vitamine D. Il y a en tout 200 patients qui n'ont pas de déficit en vitamine D. Chez ces derniers le taux d'atteintes de l'appareil locomoteur est de 3% :

$$200 \times 0{,}03 = 6$$

Réponse vraie : D

Question 7 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La probabilité d'avoir plus de 30 ans et d'avoir un déficit en vitamine D sachant qu'on a une atteinte de l'appareil locomoteur est de 0,31 (à 0,01 près)
B. La probabilité d'avoir plus de 30 ans et d'avoir un déficit en vitamine D sachant qu'on a une atteinte de l'appareil locomoteur est de 0,35 (à 0,01 près)
C. La probabilité d'avoir plus de 30 ans et d'avoir un déficit en vitamine D sachant qu'on a une atteinte de l'appareil locomoteur est de 0,40 (à 0,01 près)
D. La probabilité de ne pas avoir d'atteinte de l'appareil locomoteur est de 0,90 (à 0,01 près)
E. La probabilité de ne pas avoir d'atteinte de l'appareil locomoteur est de 0,97 (à 0,01 près)

Question 7

? Items A, B, et C

On notera, pour rendre les choses plus lisibles :

- A l'événement "avoir plus de 30 ans" et $\overline{A}$ sont complémentaires
- D l'événement "avoir un déficit en vitamine D" et $\overline{D}$ sont complémentaires
- L l'événement "avoir une atteinte de l'appareil locomoteur" et $\overline{L}$ sont complémentaires

On applique la formule de Bayes :

$$\begin{aligned}\mathbb{P}(D|L) &= \frac{\mathbb{P}(D \cap L)}{\mathbb{P}(L)} \\ &= \frac{\mathbb{P}(L|D) \times \mathbb{P}(D)}{\mathbb{P}(L)}\end{aligned}$$

On recherche alors $\mathbb{P}(L)$. Avec le schéma de la question précédente, on a :

$$\begin{aligned}\mathbb{P}(L) &= \frac{4+6+3}{450} \\ &= \frac{13}{450}\end{aligned}$$

Autre moyen possible

On peut utiliser la loi de probabilité totale mais c'est moins rapide :

$$\begin{aligned}\mathbb{P}(L) &= \mathbb{P}(L \cap D) + \mathbb{P}(L \cap \overline{D}) + \mathbb{P}(L \cap \overline{A}) \\ &= \frac{2}{3} \times \frac{1}{3} \times 0{,}04 + \frac{2}{3} \times \frac{2}{3} \times 0{,}03 + \frac{1}{3} \times 0{,}02 \\ &= \frac{13}{450}\end{aligned}$$

Passons à l'application numérique :

$$\begin{aligned}\mathbb{P}(D|L) &= \frac{0{,}04 \times \frac{100}{450}}{\frac{13}{450}} \\ &\approx 0{,}31\end{aligned}$$

(A VRAI) *(B FAUX) (C FAUX)*

7

Items D et E

On cherche $\mathbb{P}\left(\overline{L}\right)$. A l'aide de la question précédente :

$$\begin{aligned}\mathbb{P}\left(\overline{L}\right) &= 1 - \mathbb{P}(L)\\ &= 1 - \frac{13}{450}\\ &\approx 0{,}97 \quad \textit{(D FAUX)}\ \textbf{(E VRAI)}\end{aligned}$$

Réponses vraies : A et E

Question 8 Si on considère maintenant que chacun des $n = 450$ patients a la même probabilité $p = 0{,}01$ d'avoir une atteinte de l'appareil locomoteur. X est la variable aléatoire qui compte le nombre d'atteintes de l'appareil locomoteur (on considère que les survenues des atteintes sont indépendantes entre elles). **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. X est une variable aléatoire discrète
B. X suit rigoureusement une loi binomiale $\mathcal{B}in(n = 450, p = 0{,}01)$
C. X suit rigoureusement une loi normale $N(n = 450, p = 0{,}01)$
D. X suit rigoureusement une loi de Poisson $\mathcal{P}ois(\lambda)$ avec $\lambda = np$
E. Lorsque $n \geq 20$ et $p < 0{,}01$, on peut approximer une loi binomiale de paramètres n et p par une loi de Poisson de paramètre $\lambda = np$

Question 8

Items A, B, C et D

On répète $n = 450$ fois la même expérience aléatoire de Bernoulli de manière identique et indépendante. La probabilité du succès « avoir une atteinte de l'appareil locomoteur » est p. La variable aléatoire X suit donc une loi binomiale **(A VRAI) (B VRAI)** *(C FAUX) (D FAUX)*.

Item E → *Lorsque $n \geq 20$ et $p < 0{,}01$, on peut approximer une loi binomiale de paramètres n ...*

Approximation $\mathcal{B}in \rightsquigarrow \mathcal{P}ois$

On peut approximer une loi binomiale par une loi de Poisson si : $n \geq 20$ et $p \leq 0{,}01$. **(E VRAI)**

Réponses vraies : A, B et E

Question 9 Une blessure fréquente est l'entorse légère. La durée du rétablissement d'une entorse légère suit une loi exponentielle de paramètre $\lambda = 0{,}25\ j^{-1}$. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Au bout de 2 jours, la probabilité d'être guéri est de 0,4 à 0,1 près
B. Au bout de 2 jours, la probabilité d'être guéri est de 0,6 à 0,1 près
C. La moyenne de la durée de rétablissement est de 3 jours
D. La moyenne de la durée de rétablissement est de 4 jours
E. Si au bout de deux jours l'entorse n'a toujours pas guérie, la probabilité qu'elle soit guérie deux jours plus tard est de 0,6 (à 0,1 près)

Question 9

Items A et B

On sait que :

$$X \sim Exp(\lambda = 0{,}25)$$

D'après les propriétés de la loi exponentielle :

$$\begin{aligned}\mathbb{P}(X \leq 2) &= F(2)\\ &= 1 - e^{-2\times 0{,}25}\\ &= 1 - e^{-0{,}5}\\ &\approx 1 - 0{,}6\\ &\approx 0{,}4 \quad \textbf{(A VRAI)}\ \textit{(B FAUX)}\end{aligned}$$

Items C et D

Toujours d'après les propriétés de la loi exponentielle, on a :

$$\begin{aligned}\mathbb{E}(X) &= \frac{1}{\lambda}\\ &= \frac{1}{0{,}25}\\ &= 4 \quad \textbf{(D VRAI)}\ \textit{(C FAUX)}\end{aligned}$$

Médiane et moyenne

Ne pas confondre médiane et moyenne ! Elles commencent toutes les deux par un "m", mais ne désignent pas la même chose 😄

8

✗ **Item E** → *Si au bout de deux jours l'entorse n'a toujours pas guérie, la probabilité qu'elle soit ...*

La loi exponentielle est une loi sans mémoire :

$$\begin{aligned}\mathbb{P}(X \leq 4 | X \geq 2) &= \mathbb{P}(X \leq 4-2)\\ &= \mathbb{P}(X \leq 2)\\ &= F(2)\\ &= 1 - e^{-2\times 0{,}25}\\ &= 1 - e^{-0{,}5}\\ &\approx 1 - 0{,}6\\ &\approx 0{,}4 \quad \textit{(E FAUX)}\end{aligned}$$

Réponses vraies : A et D

Question 10 Parmi les patients, 10% ont des ecchymoses. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Sur un groupe de 30 patients, l'intervalle de pari à 95% des patients présentant des ecchymoses est $[0;6]$

B. Sur un groupe de 60 patients, l'intervalle de pari à 95% des patients présentant des ecchymoses est $[1;11]$

C. Sur un groupe de 30 patients, l'intervalle de confiance à 95% des patients présentant des ecchymoses est $[0;6]$

D. Sur un groupe de 60 patients, l'intervalle de confiance à 95% des patients présentant des ecchymoses est $[1;11]$

E. Lorsque le risque α diminue, la largeur de l'intervalle de pari à $1-\alpha$ du nombre de patients présentant des ecchymoses augmente

Question 10

? **Items C et D**

Nous connaissons les paramètres de la loi, il s'agit donc d'un intervalle de pari (et non de Marseille) d'une proportion *(C FAUX) (D FAUX)*.

✗ **Item A** → *Sur un groupe de 30 patients, l'intervalle de pari à 95% des patients présentant des ...*

Comme il s'agit d'un IP d'une proportion, on vérifie les conditions de validité. Quelles sont-elles vaillant.e P1 ?

♥ **Conditions de validité d'un IP d'une proportion**

$$np \geq 5$$

ET :

$$n(1-p) \geq 5$$

Ici, on a $\overbrace{np = 0{,}1 \times 30 = 3}^{\times}$ et $\overbrace{n(1-p) = 0{,}9 \times 30 = 18}^{\checkmark}$ Les conditions ne sont pas respectées *(A FAUX)*.

✓ **Item B** → *Sur un groupe de 60 patients, l'intervalle de pari à 95% des patients présentant des ...*

On vérifie les conditions de validité :

$$\begin{cases} np = 0{,}1 \times 60 = 6 > 5\\ n(1-p) = 0{,}9 \times 60 = 54 > 5\end{cases}$$

Les conditions sont respectées !

En s'aidant du formulaire pour trouver l'expression de l'IP d'une proportion (en sachant que $\varepsilon_{\alpha=0.05} = 1.96$) :

$$\begin{aligned} IP_{95\%} &= 60 \times \left[0{,}1 \pm 1{,}96 \times \sqrt{\frac{0.1 \times 0.9}{60}}\right]\\ &\approx [1;11] \quad \textbf{(B VRAI)}\end{aligned}$$

✓ **Item E** → *Lorsque le risque α diminue, la largeur de l'intervalle de pari à $1-\alpha$ du nombre de ...*

La largeur de l'intervalle de pari d'une proportion vaut :

$$l = 2\varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}}$$

Lorsque α diminue, ε_α augmente. En effet, quantitativement, on voit que $\varepsilon_{0{,}05} = 1{,}96$ et $\varepsilon_{0{,}01} = 2{,}576$; qualitativement, si α diminue, c'est le risque d'erreur qui diminue. Or, si ce risque diminue, c'est qu'on réduit notre zone d'erreur, ce qui implique que l'intervalle qu'on étudie (l'IP) devient plus grand : l augmente donc. Je donne un exemple concret : je te demande ma taille. Tu prendras plus de risque en me donnant l'intervalle $[170;180]cm$ que l'intervalle $[150;200]cm$.

Réponses vraies : B et E

Exercice 3 (Questions 11 à 15)

Le raton laveur est un animal qui peut transmettre un certain nombre de maladies. On s'intéresse à deux populations de ratons laveurs : les ratons américains et les ratons européens. Ces populations sont indépendantes.

On étudie la longueur de la queue d'un échantillon de ratons américains d'effectif $n_a = 81$. On obtient une longueur de queue moyenne de $m_a = 26\ cm$ avec une variance de $s_a^2 = 9\ cm^2$.

Question 11 Concernant l'intervalle de confiance à 99% de la moyenne théorique μ_a de la longueur de la queue des ratons laveurs américains, **parmi les propositions suivantes, laquelle est exacte ?**

A. $IC_{99\%}(\mu_a) = [20; 30] \pm 0{,}1$ cm

B. $IC_{99\%}(\mu_a) = [25{,}1; 26{,}9] \pm 0{,}1$ cm

C. $IC_{99\%}(\mu_a) = [25{,}3; 26{,}7] \pm 0{,}1$ cm

D. $IC_{99\%}(\mu_a) = [25; 27{,}1] \pm 0{,}1$ cm

E. Les conditions ne sont pas réunies pour appliquer le théorème central limite

Question 11

Items A, B, C, D et E

On calcule l'intervalle de confiance d'une moyenne pour les valeurs données (on peut se servir du formulaire si besoin).
On commence par vérifier les conditions de validité, pour vérifier qu'on peut appliquer le théorème central limite (le TCL). Quelle(s) est (sont) la (les) condition(s) brave P1 ?

Conditions d'application du TCL

Le TCL est applicable si et seulement si $n > 30$ (pour toute loi autre que la loi binomiale)

C'est le cas ici car on a $n_a = 81$ individus *(E FAUX)*. Exprimons alors notre intervalle de confiance (Quelle est le comble pour un mathématicien ? De se faire berner par son intervalle de confian... ok ok c'était nul) :

$$IC_{1-\alpha}(\mu_a) = m_a \pm \varepsilon_\alpha \sqrt{\frac{s_a^2}{n_a}}$$

Appliquons avec les valeurs adéquates :

$$\begin{aligned} IC_{1-\alpha=0{,}99}(\mu_a) &\approx 26 \pm 2{,}576 \times \sqrt{\frac{9}{81}} \\ &\approx 26 \pm 2{,}576 \times \sqrt{\frac{1}{9}} \\ &\approx 26 \pm 2{,}576 \times \frac{1}{3} \\ &\approx 26 \pm 0{,}86 \\ &\approx [25{,}14; 26{,}86] \approx [25{,}1; 26{,}9] \quad \textbf{(B VRAI)} \end{aligned}$$

Il fallait bien voir que l'on cherchait un intervalle de confiance à 99% : $\alpha = 0{,}01$ donc $\varepsilon_\alpha = 2{,}576$

Réponse vraie : B

Question 12 On compare la longueur de la queue de la population de ratons laveurs américaine avec une population d'origine européenne, d'effectif $n_e = 162$, de moyenne $m_e = 29\ cm$, et de variance $s_e^2 = 10\ cm^2$. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'hypothèse nulle est $H_0 : \mu_a \neq \mu_e$

B. Les longueurs de queues moyennes sont significativement différentes entre les deux populations au risque 5%

C. On ne met pas en évidence de différence de longueur de queue entre les deux populations au risque 5%

D. Si on rejette l'hypothèse nulle H_0, on peut commettre une erreur de type 1

E. Il faut que la moyenne des effectifs des deux groupes soit supérieure à 30 pour respecter les conditions de validité

Question 12

Items A et D

L'hypothèse nulle $H_0 : \mu_e = \mu_a$ est celle qui suppose qu'il n'y a pas de différence entre les deux populations. On parle bien d'une **égalité** dans le cas de l'hypothèse nulle *(A FAUX)*. C'est pour l'hypothèse alternative $H_1 : \mu_e \neq \mu_a$ qu'on suppose qu'il y a une différence.

L'erreur de type 1 est la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie, on la fait lorsque l'on rejette H_0 à tort **(D VRAI)**.

✘ Item E → *Il faut que la moyenne des effectifs des deux groupes soit supérieure à 30 pour ...*

La condition pour appliquer le TCL est que tous les effectifs de toutes les populations soient supérieurs à 30. Ici $n_a = 81 > 30$ et $n_e = 162 > 30$ donc les conditions de validité sont bien vérifiées *(E FAUX)*.

Items B et C

Ici, on réalise un test statistique de l'écart réduit de comparaison de moyennes. Construisons alors le test statistique adéquat :

$$\begin{aligned} Z &= \frac{m_e - m_a}{\sqrt{\frac{s_e^2}{n_e} + \frac{s_a^2}{n_a}}} \\ &= \frac{29 - 26}{\sqrt{\frac{10}{162} + \frac{9}{81}}} \\ &= \frac{3}{\sqrt{\frac{5}{81} + \frac{9}{81}}} \\ &= \frac{3}{\sqrt{\frac{14}{81}}} \\ &= \frac{3 \times 9}{\sqrt{14}} \\ &\approx 7{,}22 > 1{,}96 \end{aligned}$$

On rejette donc l'hypothèse nulle au risque $\alpha = 5\%$. On peut donc dire (au risque 5%) qu'il y a une différence significative entre les deux populations de ratons laveurs **(B VRAI)** *(E FAUX)*.

Réponses vraies : B et D

Question 13 On compare à présent la longueur de la queue des ratons laveurs américains à une population de référence, les ratons laveurs captifs dont la longueur de la queue moyenne vaut $\mu_c = 25\ cm$. On a déjà réalisé un test de comparaison de la population de ratons captifs avec l'échantillon de ratons laveur américains, et on a trouvé la statistique de test $Z = 3$. On suppose alors la longueur de la queue dans la population de ratons laveurs américains est 1 *cm* supérieure à celle des ratons laveurs captifs. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. Il y a une différence significative entre les ratons américains et les ratons captifs au risque 5%
B. On n'oberve pas de différence significative entre les deux populations au risque 5%
C. L'hypothèse alternative est $H_1 : m_a = \mu_c + 1$
D. Le risque β (ou erreur de type II) est de 83% (à 1% près)
E. Le degré de signification est compris entre 10^{-3} et 10^{-2}

Question 13

? Items A, B et E

On prend la valeur de Z que l'on nous donne ; on regarde dans la table 2a (ou on se souvient) : $\varepsilon_{0,05} = 1{,}96$ et on voit que $Z > 1{,}96$.

Donc il y a une différence significative entre les deux populations de ratons laveurs au risque 5% **(A VRAI)** *(B FAUX)*.

Pendant que l'on y est, on regarde le degré de signification dans notre tableau. On voit que $\varepsilon_{0,01} = 2{,}576 < 3$ et que $\varepsilon_{0,001} \approx 3{,}29 > 3$ Ainsi, le degré de signification p est bien compris entre 0,001 et 0,01, soit entre 10^{-3} et 10^{-2} **(E VRAI)**.

✘ Item C → *L'hypothèse alternative est $H_1 : m_a = \mu_c + 1$*

On pose l'hypothèse nulle $H_1 : \mu_a = \mu_c + 1$. Souvenez-vous qu'on n'utilise JAMAIS les valeurs observées (m) dans l'hypothèse nulle mais bien les valeurs théoriques (μ) *(C FAUX)*.

✘ Item D → *Le risque β (ou erreur de type II) est de 83% (à 1% près)*

Pour calculer l'erreur de type II β, on va commencer par reprendre la définition de la statistique de test. On note X la variable aléatoire qui prend pour valeur la moyenne de la longueur de la queue des ratons laveurs américains. Cette variable aléatoire X suit une loi qu'on peut approcher à une loi normale car $n_a > 30$ (vive le TCL !). L'écart-type de X vaut donc $\sqrt{\frac{s_a^2}{n_a}}$ (écart-type de la moyenne).

On construit maintenant notre statistique de test :

$$Z = \frac{X - \mu_c}{\sqrt{\frac{s_a^2}{n_a}}}$$

Sous H_0, $\mathbb{E}(X) = \mu_a = \mu_c$ donc $Z \sim \mathcal{N}(0{,}1)$

On fait le test au risque $\alpha = 5\%$. La condition pour rejeter H_0 est $|Z| > 1{,}96$

Sous H_1, on a $\mu_a - \mu_c = 1$

Ainsi, la nouvelle espérance de la statistique de test Z est :

$$\begin{aligned}\mathbb{E}(Z) &= \mathbb{E}\left(\frac{X-\mu_c}{\sqrt{\frac{s^2}{n}}}\right)\\ &= \frac{\mu_a - \mu - c}{\sqrt{\frac{9}{81}}}\\ &= \frac{1}{\frac{1}{3}}\\ &= 3\end{aligned}$$

Ainsi, sous H_1, $Z \sim \mathcal{N}(3{,}1)$. Donc la puissance $\mathcal{P}$ vaut :

$$\begin{aligned}\mathcal{P} &= \mathbb{P}(|Z| > 1{,}96 | Z \sim \mathcal{N}(3;1))\\ &= \mathbb{P}(Z > 1{,}96 | Z \sim \mathcal{N}(3;1)) + \mathbb{P}(Z < -1{,}96 | Z \sim \mathcal{N}(3;1))\end{aligned}$$

Donc, en posant $Z' = \frac{Z-3}{1}$ la loi normale centrée réduite, on obtient :

$$\begin{aligned}\mathcal{P} &= \mathbb{P}(Z' > \frac{1{,}96-3}{1}) + \mathbb{P}(Z' < \frac{-1{,}96-3}{1})\\ &\approx \mathbb{P}(Z' > -1{,}05) + \mathbb{P}(Z' < -5)\\ &\approx 1 - 0{,}147 + 0{,}00000029\\ &\approx 0{,}853\end{aligned}$$

Ainsi,

$$\beta = 1 - \mathcal{P} \approx 0{,}147$$

Réponses vraies : A et E

Question 14 Certains ratons laveurs sont atteints de la rage. Parmi les ratons laveurs sauvages américains, on observe une proportion $p_a = 10\%$ d'individus atteints de la rage. Dans la population des ratons laveurs captifs, prise comme référence, cette proportion vaut $\pi_c = 4\%$. On compare ces populations. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'hypothèse nulle est $H_0 : \pi_a = \pi_c$
B. Pour comparer la proportion dans un échantillon à une proportion théorique, on peut faire un test du χ^2 à 1 degré de liberté
C. Il y a une différence significative dans la prévalence de la rage entre les deux populations au risque 5%
D. On ne rejette pas l'hypothèse nulle au risque 5%
E. Il n'y a pas de différence significative entre les deux populations au risque 5%

Question 14

✔ **Item A** → *L'hypothèse nulle est $H_0 : \pi_a = \pi_c$*

L'hypothèse nulle suppose que les proportions théoriques sont les mêmes entre les deux populations :

$$H_0 : \pi_a = \pi_c \quad \textbf{(A VRAI)}$$

✔ **Item B** → *Pour comparer la proportion dans un échantillon à une proportion théorique, on ...*

Un test du χ^2 à 1 ddl est strictement identique à un test z de comparaison d'une proportion à une norme **(B VRAI)**.

? **Items C, D et E**

Avant de faire tout test, on vérifie les conditions de validité. Il faut que :

$$n(1-\pi_0) \geq 5 \quad \text{et} \quad n\pi_0 \geq 5$$

(On n'oublie pas que la vérification se fait avec la proportion théorique et non la proportion observée !)

On vérifie les conditions :

$$0{,}04 \times 81 = 3{,}24 < 5 \quad \text{et} \quad 0{,}96 \times 81 = 77{,}76 > 5$$

Les conditions de validité ne sont pas réunies pour réaliser le test *(C FAUX) (D FAUX) (E FAUX)*.

Réponses vraies : A et B

Question 15 On cherche à déterminer maintenant s'il y a un lien entre le nombre de bandes sur la queue des ratons laveurs et la population d'origine. Les données recueillies sont présentées dans le tableau suivant :

Nombre de bandes	5	6	7	Total
Ratons américains	13	20	48	81
Ratons européens	36	29	97	162
Total	49	49	145	243

On réalise un test du χ^2, et on trouve $K = 2{,}3$. **Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?**

A. L'hypothèse alternative H_1 suppose une corrélation entre l'origine géographique et le nombre de bandes sur la queue

B. On fait un test du χ^2 à 6 degrés de liberté
C. On fait un test de comparaison de moyennes
D. Les conditions de validité sont vérifiées car tous les effectifs observés sont supérieurs à 5
E. On peut affirmer au risque 5% qu'il n'y a pas de corrélation entre les origines géographiques et le nombre de bandes sur la queue

Question 15

✔ **Item A** → *L'hypothèse alternative H_1 suppose une corrélation entre l'origine géographique et ...*

On pose bien l'hypothèse nulle H_0 comme l'absence de corrélation et l'hypothèse alternative H_1 comme une corrélation **(A VRAI)**.

✘ **Item D** → *Les conditions de validité sont vérifiées car tous les effectifs observés sont supérieurs à 5*

Les conditions de validité sont respectées car tout les effectifs **calculés** (et non pas observés) sont ≥ 5 *(D FAUX)*. Il suffit de vérifier que le plus petit effectif calculé est supérieur ou égal à 5. On rappelle l'expression de l'effectif calculé :

$$\text{Effectif calculé} = \frac{(\text{Total ligne}) \times (\text{Total colonne})}{\text{Total}}$$

La plus petite ligne correspond aux ratons américains ($n_a = 81$), et la plus petite colonne correspond aux ratons à 5 ou 6 bandes (qui ont le même effectif total). Calculons l'effectif théorique des ratons américains à 5 bandes :

$$C_{1,1} = \frac{81 \times 49}{243} = \frac{49}{3} \approx 16{,}3 \geq 5$$

Ainsi, le plus petit effectif calculé est bien supérieur à 5, donc tous les effectifs calculés le sont.

✘ **Item B** → *On fait un test du χ^2 à 6 degrés de liberté*

On calcule le nombre de degrés de liberté :

$$\text{ddl} = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1)$$

Attention, on ne prend pas en compte les totaux ! On a 2 lignes (américains et européens), et 3 colonnes (5, 6 et 7 bandes), donc :

$$\text{ddl} = (2-1) \times (3-1) = 2$$

C'est donc un test du χ^2 à 2 degrés de liberté *(B FAUX)*.

✘ **Item C** → *On fait un test de comparaison de moyennes*

On fait ici un test du χ^2 pour déterminer si le nombre de bandes est indépendant de l'origine géographique. C'est un test du χ^2 d'indépendance. Un test de comparaison de moyennes sert (*comme son nom l'indique*) à comparer des moyennes, et ne peut pas être remplacé par un test du χ^2.
En revanche, un test du χ^2 à **1 ddl** équivaut à un test de comparaison de **proportions** (et non de moyennes). Ce n'est pas le cas ici puisqu'on a 2 ddl *(C FAUX)*.

✘ **Item E** → *On peut affirmer au risque 5% qu'il n'y a pas de corrélation entre les origines ...*

On veut savoir si on peut rejeter ou non H_0. On cherche dans la table 3, à la ligne 2 ddl, la valeur seuil pour $\alpha = 0{,}05$ (puisqu'on est au risque 5%), et on trouve **5,991**. On a :

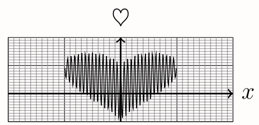
$$K = 2{,}3 < 5{,}991$$

Donc on ne peut pas rejeter l'hypothèse nulle : on ne peut pas affirmer qu'il y a une corrélation entre l'origine et le nombre de bandes. En revanche, on n'accepte **JAMAIS** l'hypothèse nulle, donc <u>on ne peut pas affirmer qu'il n'y a pas de corrélation</u> ! La formulation de

12

l'item E revient à accepter H_0, ce qui est formellement interdit *(oui, on finit sur un piège méchant mais au moins vous ne tomberez plus dedans) (E FAUX).*

C'est la fin de cette correction ! Si tu as tout compris, c'est parfait, donne-toi à fond, ça finira par payer ! Si certaines notions, certains passages ou développements n'ont pas été clairs, clique <ICI> pour nous poser toutes tes questions ! Utilise ce forum sans modération, on est là pour ça !
Et n'oublie pas que tu as aussi le forum du moodle, les enseignants te répondront avec graannd plaisir également !



Réponse vraie : A

Association pour l'Accès Santé – Université Paris Cité
Année Universitaire 2023-2024

SUJET

Examen Blanc n°2 PASS

UE 9 : Mathématiques - Biostatistiques

Durée de l'épreuve : **1h**

A LIRE AVANT DE COMMENCER L'ÉPREUVE

Vérifiez que les informations saisies sur votre grille QCM sont correctes : nom, prénom et numéro étudiant.
Les correcteurs liquides ou en ruban de type Blanco, Tipp-Ex, et autres sont interdits car chaque question comporte une ligne de droit au remords.
Seule l'utilisation du stylo à bille noir est autorisée pour cocher les grilles.

INFORMATIONS RÉGLEMENTAIRES

- Les questions sans réponse seront considérées comme nulles.
- Une grille QCM est à remplir pour l'ensemble de l'épreuve.
- Veiller à remplir complètement toute la surface des cases choisies.
- Ne pas gratter, ne pas raturer, ne pas mettre de croix ni aucun autre signe.
- Toute fraude ou tentative de fraude fera l'objet de poursuites disciplinaires (Décret n°92-657 du 13 juillet 1992). Tout signe distinctif porté sur la grille QCM pouvant indiquer sa provenance constitue une fraude.
- Les calculatrices sont **autorisées**
- Aucun candidat n'est admis à quitter la salle d'examen avant la fin de l'épreuve.

RECOMMANDATIONS SPÉCIFIQUES À L'ÉPREUVE

INFORMATIONS SUR L'ÉPREUVE

Le sujet contient 4 pages numérotées de 1 à 4 et comporte 15 questions.
Merci de vérifier au début de l'épreuve que le sujet est complet.

Exercice 1 (questions 1 à 5)

Une agence de musique décide d'étudier le marché de la pop au Japon et en Corée en s'intéressant à l'évolution du nombre de groupe J-pop et de K-pop au cours du temps. Le nombre de groupes de J-pop est représenté par la fonction $j(t)$ et celui de K-pop par la fonction $k(t)$. Les données de l'étude nous renseignent sur le nombre de groupes au début des mesures : $k(t=0)=10$ et $j(t=0)=30$

On pose le système suivant pour décrire l'évolution des deux groupes de populations :
$\forall t \in [0;+\infty[$,

$$\begin{cases} k'(t) = 2k(t) + j(t) \\ j'(t) = -3k(t) + 2j(t) \end{cases}$$

Question 1 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Il s'agit d'une relation de compétition
B. Il s'agit d'une relation proie/prédateur
C. L'équation caractéristique de k(t) est : $r^2 - 4r + 7 = 0$
D. $=k(t) = e^{2t}\left(30\cos\left(\sqrt{3}t\right) - 10\sqrt{3}\sin\left(\sqrt{3}t\right)\right)=$
E. $j(t) = e^{2t}\left(30\cos\left(\sqrt{3}t\right) - 10\sqrt{3}\sin\left(\sqrt{3}t\right)\right)$

Un nouveau modèle est utilisé pour étudier le nombre de groupes de K-pop $k(t)$ sous la forme :

$$k''(t) + 2k'(t) - 3k(t) = t^2 + 2t + 3e^{-2t}$$

Avec ce modèle, on retrouve à l'origine $k(0) = 1$ et un point d'inflexion à cette même valeur.

Question 2 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La dérivée seconde $k''(t)$ s'annule en 0
B. Il s'agit d'une équation différentielle du second ordre
C. k(t)=$\frac{11}{4}e^t + \frac{23}{108}e^{-3t} - \frac{1}{3}t^2 - \frac{10}{9}t - \frac{26}{27} - e^{-2t}$
D. k(t)=$\frac{11}{4}e^t - \frac{23}{108}e^{-3t} + \frac{1}{3}t^2 - \frac{10}{9}t + \frac{26}{27} + e^{-2t}$
E. k est définie sur $]-\infty;+\infty[$

On étudie un troisième modèle étudiant l'évolution du nombre de groupes de K-pop. Elle est définie pour tout réel positif t d'expression :

$$k(t) = -e^{-t}(t^2 - 2t - 4) + \cos(t) + 1$$

Question 3 On veut étudier l'aire sous la courbe (AUC) de cette fonction à 0,01 près. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. $AUC_0^2 \approx 4{,}00$
B. $AUC_0^2 \approx 6{,}91$
C. $AUC_0^2 \approx -2{,}80$
D. $AUC_0^4 \approx 10{,}00$
E. $AUC_0^4 \approx 20{,}00$

On étudie un modèle qui traduit l'évolution des groupes de J-pop $j(t)$ selon deux paramètres a et b avec a supérieur à b. Le modèle définie sur $\mathbb{R}^2$ est le suivant :

$$j(a,b) = (a-b)e^{\frac{a}{b}}$$

Question 4 On donne les valeurs des paramètres suivants pour une valeur de $j = 593{,}65$, $a = 5{,}00 \pm 0{,}05$ et $b = 1{,}00 \pm 0{,}01$. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La plus grosse incertitude vient de la variable b
B. $\Delta j \leq 44{,}52$
C. $\Delta j \leq 68{,}27$
D. $j \approx 593{,}65 \pm 44{,}52$
E. $j \approx 593{,}65 \pm 68{,}27$

On pose un ultime modèle, pour tout t positif :

$$k(t) = \frac{3t^3 - 2t^2 - 6t + 12}{3t^2 - 2t + 2}$$

On donne aussi :

$$\int_0^1 \frac{2t - 10}{3t^2 - 2t + 2} dt \approx -4{,}67$$

Question 5 Concernant cette fonction, laquelle (lesquelles) des propositions est (sont) exacte(s) ?

A. $k'(t) = \frac{-9t^2+4t+6}{(3t^2-2t+2)^2}$
B. $k'(t) = \frac{-9t^4+12t^3-40t^2+80t-12}{3t^2-2t+2}$
C. k est définie sur $]-\infty, +\infty[$
D. $\int_0^1 k(t) = [t]_0^1 - [\ln(3t^2 - 2t + 2)]_0^1 + \int_0^1 \frac{2t-10}{3t^2-2t+2} dt$
E. $\int_0^1 k(t)\, dt \approx 4{,}76$

Exercice 2 (questions 6 à 10)

Au service des urgences, 50% des patients viennent consulter pour des difficultés respiratoires. Parmi ces patients, 30% ont une pneumonie. On note également que 10% des patients se présentant aux urgences sans difficultés respiratoires ont tout de même une pneumonie.

Question 6 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La sensibilité du signe "difficulté respiratoire" pour la maladie "pneumonie" est de 30%
B. La spécificité du signe "difficulté respiratoire" pour la maladie "pneumonie" est de 10%
C. Un patient qui se présente aux urgences avec des difficultés respiratoires à 30% de chances d'avoir une pneumonie
D. Un patient qui se présente aux urgences avec difficultés respiratoires à 30% de chances de ne pas avoir une pneumonie
E. Si la prévalence de la pneumonie aux urgences augmente, la sensibilité du signe "difficultés respiratoires" augmente

Parmi les 50% des patients se présentant pour une difficulté respiratoire, on en choisit au hasard 20. La probabilité pour ces patients d'avoir une pneumonie est la même que précédemment. On note X la variable aléatoire qui donne le nombre de patients ayant une pneumonie parmi ces 20 personnes.

Question 7 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La variable X suit une loi binomiale de paramètres $n = 20$ et $\pi = 0.5$
B. La variable X suit une loi de Poisson de paramètre $\lambda = 0.6$
C. La probabilité qu'aucun des patients examinés ne soit malade est nulle
D. La probabilité que la moitié des patients soit malade est de 50%
E. L'écart-type de X est d'environ 2 patients

On prend désormais un échantillon de 30 patients, qui ont toujours la même probabilité d'être malade. X est toujours la variable aléatoire qui donne le nombre de patients malade parmi l'échantillon.

Question 8 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. On peut approximer X par une loi normale de paramètres $\mu = 9$ et $\sigma^2 = 6.3$
B. La variable $Z = \frac{X-\mu}{\sigma}$ suit une loi normale centrée non-réduite
C. La probabilité que le nombre de patients malades parmi les 30 interrogés soit supérieur à 15 est d'environ 0.02
D. La probabilité que le nombre de patients malades parmi les 30 interrogés soit inférieur à 10 est d'environ 0.31
E. Si $Z \sim \mathcal{N}(0{,}1)$, $\mathbb{P}(Z > 2) = \mathbb{P}(Z < 2)$

La durée de l'hospitalisation des patients atteints de pneumonie (notée D) suit une loi exponentielle de paramètre λ. On sait que la probabilité qu'un patient déjà hospitalisé depuis 3 jours reste encore au moins 5 jours de plus est de 0.37.

Question 9 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La variable D suit une loi exponentielle de paramètre $\lambda = 0.4$
B. La probabilité qu'un patient reste au total moins de 2 semaines alors qu'il est déjà là depuis 6 jours est égale à 0.8
C. 50% des patients restent moins de 5 jours
D. En moyenne, les patients sont hospitalisés pendant 5 jours
E. On peut approcher la variable D par une loi de Poisson de paramètre λ

On s'intéresse à la moyenne théorique du nombre de patients atteints de pneumonie se présentant aux urgences sur 1 mois (M1) et sur 1 an (M12).

On tire au sort un échantillon de 150 patients aux urgences d'un hôpital parisien, et on trouve parmi les patients de cet échantillon 60 cas de pneumonie.

Question 10 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La probabilité que le patient de l'échantillon soit malade est notée $p = 0.4$
B. L'intervalle de pari à 95% de M1 est plus large que l'intervalle de pari à 95% de M12
C. L'intervalle de pari à 99% de M1 est plus large que l'intervalle de pari à 95% de M1
D. L'intervalle de confiance à 95% de la proportion de patients atteints de pneumonie est $[32.3; 47.8]$% à 0.05% près
E. L'intervalle de pari à 95% de la proportion de patients atteints de pneumonie est $[32.3; 47.8]$% à 0.1% près

Exercice 3 (questions 11 à 15)

Importée d'Ethiopie, le café s'est vite répandu dans l'Empire Ottoman, devenant ce qu'on connaît aujourd'hui comme étant le café turc. Il est considéré comme stimulant la mémoire et réduisant certaines maladies. Une étude parue en 2015 dans le Journal of the International Society of Sports Nutrition s'intéresse de près à l'effet du café turc sur la performance et le temps de réaction. Plusieurs aspects de ces effets sont étudiés dans cet exercice. L'étude a été réalisée sur $n = 60$ participants. $n_{TC} = 30$ participants ont ingéré du café turc et $n_{DC} = 30$ autres du café turc décaféiné. Les données utilisées sont tirées ou simulées à partir de cet article. On prend $\alpha = 5\%$.

La pression artérielle diastolique a été mesurée chez les participants juste avant l'exercice. Le groupe des participants ayant ingéré du café turc (TC) présente une pression artérielle diastolique moyenne de 122 $mmHg$. Le groupe des participants ayant ingéré du café turc décaféiné (DC) présente une pression diastolique moyenne de 115 $mmHg$. Un test a été conduit afin de savoir si la pression artérielle diastolique varie avec le type de café. L'intervalle de la différence entre la pression artérielle diastolique TC et DC est $[6.7; 7.3]$ au risque $\alpha = 5\%$.

Question 11 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. L'hypothèse nulle est l'indépendance entre le type de café et la pression artérielle diastolique dans la population
B. Les pressions artérielles diastoliques entre les participants TC et DC est significativement identique au risque 5%
C. La pression artérielle diastolique est significativement différente entre la population TC et la population DC au risque 5%
D. La conclusion du test peut reposer sur un intervalle de pari ou un critère de test
E. Réaliser un test statistique revient à tester l'appartenance du critère de test z à la distribution de la variable aléatoire du critère de test Z sous H_0

La performance physique a ensuite été évaluée sur une course. Le nombre de participants ayant parcouru plus de 5 km a été dénombré. Dans le groupe TC, 60% des participants ont couru plus de $5km$ alors que dans le groupe DC, 20% des participants ont couru plus de $5km$. Un test a été réalisé afin de comparer la proportion de personnes ayant couru plus de $5km$ dans les deux cas. Le degré de signification est entre 1% et 2%.

Question 12 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La statistique de test est entre 2,57 et 2,32
B. La statistique de test est entre 1,64 et 1,28
C. La proportion de proportion de personnes ayant couru plus de $5km$ est significativement différente entre le groupe TC et DC au risque 5%
D. La proportion de proportion de personnes ayant couru plus de $5km$ est significativement identique entre une population TC et DC au risque 5%
E. Ici, le critère de test z est dans les queues de distribution de la variable aléatoire du critère de test Z

Des tests de concentration ont été réalisés en mesurant le temps de réaction des participants. Parmi les participants TC, 3 ont présenté un temps de réaction supérieur (>) à 0,1s et 27 ont présenté un temps de réaction inférieur (<) à 0,1s. Le temps de réaction dans la population d'individus DC se distribue de la manière suivante : $40\% > 0{,}1s$ et $60\% < 0{,}1s$.
Un test du χ^2 est réalisé.

Question 13 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Le test réalisé peut être un test d'indépendance
B. Le degré de signification de ce test est supérieur à 1%
C. Le degré de signification de ce test est inférieur à 0,1%
D. Si le degré de signification aurait été plus grand, cela signifie que la différence entre les temps de réaction TC et DC aurait été plus grande
E. Si la loi du temps de réaction n'était pas une loi normale, on aurait pu quand même faire ce test

La concentration de lactate après l'exercice a été mesurée dans chacun des groupes TC et DC. L'écart-type de la concentration de lactate dans les groupes TC et DC s_{TC} et s_{DC} est de 0,5 $mmol/L$. On prend $H_1 : \mu_{TC} - \mu_{DC} = 2$.

Question 14 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La puissance de ce test est proche de 100%
B. La puissance de ce test est proche de 50%
C. La puissance de ce test est proche de 80%
D. Un plus grand nombre de participants aurait permis de mettre en évidence une différence plus petite
E. La puissance augmente lorsque le risque α augmente

Question 15 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les estimateurs ponctuels ne sont pas utilisés dans les tests statistiques, ce sont les estimateurs par intervalle qui sont utilisés.
B. Les tests statistiques utilisent les paramètres de la population et pas les estimateurs.
C. Les conditions de validité des tests permettent de vérifier que les hypothèses sont raisonnables pour réaliser un test dit paramétrique.
D. Si $ds < \alpha$, on ne rejette pas H_0
E. Si $ds > \alpha$, on rejette H_0

FIN DU SUJET

Association pour l'Accès Santé – Université Paris Cité
Année Universitaire 2023-2024

CORRECTION

Examen Blanc n°2 PASS

UE 9 : Mathématiques - Biostatistiques

Durée de l'épreuve : **1h**

A LIRE AVANT DE COMMENCER L'ÉPREUVE

Vérifiez que les informations saisies sur votre grille QCM sont correctes : nom, prénom et numéro étudiant.
Les correcteurs liquides ou en ruban de type Blanco, Tipp-Ex, et autres sont interdits car chaque question comporte une ligne de droit au remords.
Seule l'utilisation du stylo à bille noir est autorisée pour cocher les grilles.

INFORMATIONS RÉGLEMENTAIRES

- Les questions sans réponse seront considérées comme nulles.
- Une grille QCM est à remplir pour l'ensemble de l'épreuve.
- Veiller à remplir complètement toute la surface des cases choisies.
- Ne pas gratter, ne pas raturer, ne pas mettre de croix ni aucun autre signe.
- Toute fraude ou tentative de fraude fera l'objet de poursuites disciplinaires (Décret n°92-657 du 13 juillet 1992). Tout signe distinctif porté sur la grille QCM pouvant indiquer sa provenance constitue une fraude.
- Les calculatrices sont **autorisées**
- Aucun candidat n'est admis à quitter la salle d'examen avant la fin de l'épreuve.

RECOMMANDATIONS SPÉCIFIQUES À L'ÉPREUVE

INFORMATIONS SUR L'ÉPREUVE

Le sujet contient 15 pages numérotées de 1 à 15 et comporte 15 questions.
Merci de vérifier au début de l'épreuve que le sujet est complet.

Université Paris Cité
A2SUP - Tutorat

Tuto n° : 2 UE (spé) : 9

Nom :
Prénom :
Numéro A2SUP :

IDENTIFICATION

Numéro A2SUP : Diz. Mil. / Mil. / Cent. / Diz. / Unit. (0 1 2 3 4 5 6 7 8 9)

Contrôle : M Unit. / M Diz. / J Unit. / J Diz. (0 1 2 3 4 5 6 7 8 9)
Date de naissance (JJ/MM)

1	B C E
2	A B C E
3	B
4	C E
5	C E
6	C
7	E
8	A
9	B D
10	A B C
11	A C E
12	A E
13	C
14	A D E
15	C
16–45	

Exercice 1 (questions 1 à 5)

Une agence de musique décide d'étudier le marché de la pop au Japon et en Corée en s'intéressant à l'évolution du nombre de groupe J-pop et de K-pop au cours du temps. Le nombre de groupes de J-pop est représenté par la fonction $j(t)$ et celui de K-pop par la fonction $k(t)$. Les données de l'étude nous renseignent sur le nombre de groupes au début des mesures : $k(t=0)=10$ et $j(t=0)=30$

On pose le système suivant pour décrire l'évolution des deux groupes de populations :
$\forall t \in [0;+\infty[$,

$$\begin{cases} k'(t) = 2k(t) + j(t) \\ j'(t) = -3k(t) + 2j(t) \end{cases}$$

Question 1 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Il s'agit d'une relation de compétition

B. Il s'agit d'une relation proie/prédateur

C. L'équation caractéristique de k(t) est : $r^2 - 4r + 7 = 0$

D. $=k(t) = e^{2t}\left(30\cos\left(\sqrt{3}t\right) - 10\sqrt{3}\sin\left(\sqrt{3}t\right)\right) =$

E. $j(t) = e^{2t}\left(30\cos\left(\sqrt{3}t\right) - 10\sqrt{3}\sin\left(\sqrt{3}t\right)\right)$

Question 1

Items A et B

Petite question de cours pour bien commencer. Pour connaître le type de relation il suffit de regarder les coefficients de notre système. Ici les variations de $k(t)$ dépendent positivement de $j(t)$, en effet pour une valeur de cette dernière croissante, la valeur de $k'(t)$ va également augmenter. Au contraire, les variations de $j(t)$ dépendent négativement de $k(t)$, puisque pour un $k(t)$ croissant la valeur de $j(t)$ diminue.

Ainsi la K-pop croît avec la J-pop et la J-pop diminue avec la K-pop, ce qui traduit une relation de type proie/prédateur (où la K-pop est le « prédateur » et la J-pop la « proie »). *(A FAUX)* **(B VRAI)**

Items C, D et E

On notera j et k pour éviter d'alourdir les équations. Il faut ici résoudre le système pour pouvoir répondre. Pour cela on va d'abord déterminer une expression de k puis en déduire une de j (mais faire l'inverse est tout à fait possible et revient au même).

1 – Arriver à l'équation caractéristique

Pour cela on va dériver la première ligne de notre système :

$$k'' = 2k' + j'$$

On injecte ensuite l'expression de j' donnée dans la deuxième ligne de notre système :

$$k'' = 2k' - 3k + 2j \qquad \text{(a)}$$

On reprend ensuite la première ligne du système pour isoler $j(t)$:

$$j = k' - 2k \qquad \text{(b)}$$

On réinjecte maintenant cette nouvelle expression (b) dans la principale (a) :

$$\begin{aligned} k'' &= 2k' - 3k + 2(k' - 2k) \\ &= 2k' - 3k + 2k' - 4k \\ &= 4k' - 7k \end{aligned}$$

$$\iff k'' - 4k' + 7k = 0$$

On se retrouve avec une équation différentielle de degré 2 linéaire à coefficients constants sans second membre, il n'y aura donc pas besoin de faire une recherche de solution particulière (puisque celle-ci vaut $k_0(t) = 0$) et on peut se tourner directement vers la solution homogène.

On a presque terminé, il suffit de remplacer les k par des r *(protip : chaque symbole ′ correspond au numéro en exposant)*, on a ainsi pour équation caractéristique :

$$r^2 - 4r + 7$$

2 – Déterminer la solution homogène

On résout l'équation caractéristique : $r^2 - 4r + 7$

$$\Delta = (-4)^2 - 4 \times 1 \times 7 = -12$$

Vu que le discriminant est négatif les deux racines de cette équation seront des nombres complexes (car $\sqrt{-12}$ n'existe pas dans les nombres réels on va utiliser le complexe i qui donne $i^2 = -1$, avec $i\sqrt{12}$ au carré on retrouve bien -12). On a donc

$$\begin{aligned} r_1 &= \frac{4 + i\sqrt{12}}{2} \\ &= 2 + i\sqrt{3} \end{aligned} \qquad \begin{aligned} r_2 &= \frac{4 - i\sqrt{12}}{2} \\ &= 2 - i\sqrt{3} \end{aligned}$$

Les deux solutions s'expriment sous la forme $r = \alpha \pm i\beta$ avec $\alpha = 2$ et $\beta = \sqrt{3}$. La solution homogène s'écrit donc :

$$\begin{aligned} k_h(t) &= e^{\alpha t}\left(A\cos(\beta t) + B\sin(\beta t)\right) \\ &= e^{2t}\left(A\cos(\sqrt{3}t) + B\sin(\sqrt{3}t)\right) \end{aligned}$$

3 – Obtenir la solution générale

La solution générale s'écrit :

$$\begin{aligned} k(t) &= k_h(t) + k_0(t) \\ &= e^{2t}\left(A\cos\left(\sqrt{3}t\right) + B\sin\left(\sqrt{3}t\right)\right) \end{aligned}$$

4 – Calculer $j(t)$

On va d'abord calculer la dérivée $k'(t)$. k est de la forme $u \times v$, on aura alors :

$$k'(t) = 2e^{2t}\left[A\cos\left(\sqrt{3}t\right) + B\sin\left(\sqrt{3}t\right)\right] + e^{2t}\left[-\sqrt{3}A\sin\left(\sqrt{3}t\right) + \sqrt{3}B\cos\left(\sqrt{3}t\right)\right]$$
$$= 2k(t) + e^{2t}\left[-\sqrt{3}A\sin\left(\sqrt{3}t\right) + \sqrt{3}B\cos\left(\sqrt{3}t\right)\right]$$

On a plus qu'à reprendre l'expression (b) avec cette nouvelle expression de $k'(t)$ et $k(t)$:

$$\begin{aligned} j(t) &= k'(t) - 2k(t) \\ &= 2k(t) + e^{2t}\left[-\sqrt{3}A\sin\left(\sqrt{3}t\right) + \sqrt{3}B\cos\left(\sqrt{3}t\right)\right] - 2k(t) \\ &= e^{2t}\left[-\sqrt{3}A\sin\left(\sqrt{3}t\right) + \sqrt{3}B\cos\left(\sqrt{3}t\right)\right] \end{aligned}$$

5 – Déterminer A et B

D'après l'énoncé on sait que $k(0) = 10$, mais on sait également :

$$\begin{aligned} k(0) &= e^{0\times t}\left[-\sqrt{3}A\cos\left(\sqrt{3}\times 0\right) + \sqrt{3}B\sin\left(\sqrt{3}\times 0\right)\right] \\ &= 1 \times (1 \times A + 0 \times B) \\ &= A \end{aligned}$$

et idem pour $j(0)$, on a :

$$\begin{aligned} j(0) &= 1 \times \left(0 \times A + \sqrt{3}B\right) \\ &= \sqrt{3}B \end{aligned}$$

On a donc $A = 10$ et $B = \frac{30}{\sqrt{3}} = 10\sqrt{3}$

Pour résumer on a donc :

$$\begin{aligned} k(t) &= e^{2t}\left(10\cos(\sqrt{3}t) + 10\sqrt{3}\sin(\sqrt{3}t)\right) \\ j(t) &= e^{2t}\left(-10\sqrt{3}\sin(\sqrt{3}t) + 30\cos(\sqrt{3}t)\right) \end{aligned}$$

(C VRAI)*(D FAUX)***(E VRAI)**

Réponses vraies : B, C et E

Un nouveau modèle est utilisé pour étudier le nombre de groupes de K-pop $k(t)$ sous la forme :

$$k''(t) + 2k'(t) - 3k(t) = t^2 + 2t + 3e^{-2t}$$

Avec ce modèle, on retrouve à l'origine $k(0) = 1$ et un point d'inflexion à cette même valeur.

Question 2 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La dérivée seconde $k''(t)$ s'annule en 0
B. Il s'agit d'une équation différentielle du second ordre
C. k(t)=$\frac{11}{4}e^t + \frac{23}{108}e^{-3t} - \frac{1}{3}t^2 - \frac{10}{9}t - \frac{26}{27} - e^{-2t}$
D. k(t)=$\frac{11}{4}e^t - \frac{23}{108}e^{-3t} + \frac{1}{3}t^2 - \frac{10}{9}t + \frac{26}{27} + e^{-2t}$
E. k est définie sur $]-\infty;+\infty[$

Question 2

Item A → *La dérivée seconde $k''(t)$ s'annule en 0*

Puisqu'on a un point d'inflexion en 0, cela se traduit par un changement de la variation (= dérivée) de la fonction. Mathématiquement, cela se caractérise par un dérivée seconde nulle changeant de signe. **(A VRAI)**

Item B → *Il s'agit d'une équation différentielle du second ordre*

Encore une petite question de cours : ici on retrouve jusqu'à la dérivée seconde de la fonction, il s'agit donc d'une équation différentielle du second ordre. **(B VRAI)**

Items C, D et E

Comme avant on va procéder par étapes :

1 – Déterminer l'équation caractéristique

Dans notre situation, il suffit de supprimer le second membre. On se retrouve avec une équation de la forme : $k_0'' + 2k_0' - 3k_0$. On en déduit l'équation caractéristique :

$$r^2 + 2r - 3$$

2 – Déterminer la solution générale sous second membre

On reprendre l'équation caractéristique afin de la résoudre, en trouvant les racines :

$$\begin{aligned} \Delta &= 2^2 - 4 \times (-3) = 16 \\ \Longrightarrow r_1 &= 1 \quad \text{et} \quad r_2 = -3 \end{aligned}$$

Puisque le discriminant est positif, la solution homogène s'écrit :

$$k_0(t) = Ae^t + Be^{-3t}$$

3 – Déterminer la solution particulière

Notre second membre est une somme de deux fonctions usuelles *(fonction carré et exponentielle)*. La solution particulière K est donc la somme de la solution particulière de la fonction exponentielle qu'on va noter k_1 et la solution particulière d'un polynôme du second degré qu'on va noter k_2. On a donc :

$$\begin{aligned} K(t) &= k_1(t) + k_2(t) \\ &= at^2 + bt + c + de^{-2t} \end{aligned}$$

On va ensuite la dériver 2 fois pour la réinjecter dans l'équation de base :

$$\begin{aligned} K'(t) &= k_1'(t) + k_2'(t) \\ &= 2at + b - 2de^{-2t} \end{aligned}$$

$$\Longrightarrow K''(t) = 2a + 4de^{-2t}$$

On les réinjecte ensuite dans l'équation de base :

3

$$K'' + 2K' - 3K = t^2 + 2t + 3e^{-2t}$$

$$\iff \left[2a + 4de^{-2t}\right] + 2\left[2at + b - 2de^{-2t}\right] - 3\left[at^2 + bt + c + de^{-2t}\right] = t^2 + 2t + 3e^{-2t}$$

$$\iff (-3a)t^2 + (4a - 3b)t + (2a + 2b - 3c) + (-3d)e^{-2t} = t^2 + 2t + 3e^{-2t}$$

Par identification, on obtient alors un système à résoudre pour connaître les valeurs de a, b, c et d. Ils doivent êtres égaux aux coefficients du second membre (ici 1, 2, 0 et 3).

On a :

$$\begin{cases} -3a = 1 \\ 4a - 3b = 2 \\ 2a + 2b - 3c = 0 \\ -3d = 3 \end{cases} \iff \begin{cases} a = -\frac{1}{3} \\ b = -\frac{10}{9} \\ c = -\frac{26}{27} \\ d = -1 \end{cases}$$

Soit,

$$K(t) = -\frac{1}{3}t^2 - \frac{10}{9}t - \frac{26}{27} - e^{-2t}$$

📌 4 – Déterminer la solution générale

On n'a plus qu'à faire la somme des deux solutions pour trouver la générale *(pour rappel, cette somme nous permet d'obtenir la solution à l'équation puisque les SGESSM et SPEASM sont des solutions qui sont combinées linéairement ! La combinaison linéaire de solutions est aussi solution ! Merci thomas j'ai appris un truc aujourd'hui 👍)* :

$$\begin{aligned} k(t) &= k_0(t) + K(t) \\ &= Ae^t + Be^{-3t} - \frac{1}{3}t^2 - \frac{10}{9}t - \frac{26}{27} - e^{-2t} \end{aligned}$$

📌 5 – Déterminer A et B

On sait qu'il y a un point d'inflexion à l'origine, ce qui se caractérise par $k''(0) = 0$. On va alors d'abord dériver 2 fois la fonction $k(t)$:

$$\begin{aligned} k(t) &= Ae^t + Be^{-3t} - \frac{1}{3}t^2 - \frac{10}{9}t - \frac{26}{27} - e^{-2t} \\ \Longrightarrow k'(t) &= Ae^t - 3Be^{-3t} - \frac{2}{3}t - \frac{10}{9} + 2e^{-2t} \\ \Longrightarrow k''(t) &= Ae^t + 9Be^{-3t} - \frac{2}{3} - 4e^{-2t} \end{aligned}$$

De plus on a d'après l'énoncé $k(0) = 1$. On a donc :

$$k(0) = Ae^0 + Be^{-3\times 0} - \frac{1}{3} \times 0^2 - \frac{10}{9} \times 0 - \frac{26}{27} - e^{-2\times 0}$$

$$\iff 1 = A + B - \frac{26}{27} - 1$$

Et

$$k''(0) = Ae^0 + 9Be^{-3\times 0} - \frac{2}{3} - 4e^{-2\times 0}$$

$$\iff 0 = A + 9B - \frac{2}{3} - 4$$

On obtient alors le système suivant :

$$\begin{cases} A + B - \frac{26}{27} - 1 = 1 \\ A + 9B - \frac{2}{3} - 4 = 0 \end{cases} \iff \begin{cases} A = \frac{11}{4} \\ B = \frac{23}{108} \end{cases}$$

Finalement,

$$k(t) = \frac{11}{4}e^t + \frac{23}{108}e^{-3t} - \frac{1}{3}t^2 - \frac{10}{9}t - \frac{26}{27} - e^{-2t} \qquad \textbf{(C VRAI)}\,\textit{(D FAUX)}$$

On a une fonction qui est la somme d'un polynôme et d'exponentielles qui sont des fonctions définies sur $\mathbb{R}$. On a donc k définie sur $]-\infty; +\infty[$. **(E VRAI)**

Réponses vraies : A, B, C et E

On étudie un troisième modèle étudiant l'évolution du nombre de groupes de K-pop. Elle est définie pour tout réel positif t d'expression :

$$k(t) = -e^{-t}(t^2 - 2t - 4) + \cos(t) + 1$$

Question 3 On veut étudier l'aire sous la courbe (AUC) de cette fonction à 0,01 près. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. $AUC_0^2 \approx 4{,}00$

B. $AUC_0^2 \approx 6{,}91$

C. $AUC_0^2 \approx -2{,}80$

D. $AUC_0^4 \approx 10{,}00$

E. $AUC_0^4 \approx 20{,}00$

Question 3

❓ Items A, B et C

Pour ces deux premiers items on doit d'abord intégrer la fonction entre 0 et 2 puisque le tableau ne fournit pas les valeurs dans cet intervalle.

$$\int_0^2 k(t)\,dt = \int_0^2 \left[-e^{-t}(t^2 - 2t - 4) + cos(t) + 1\right] dt$$

On va linéariser l'intégrale pour y voir plus clair :

$$\int_0^2 \left[-e^{-t}(t^2 - 2t - 4) + cos(t) + 1\right] dt = \underbrace{\int_0^2 -e^{-t}t^2\,dt}_{1} + \underbrace{\int_0^2 2te^{-t}\,dt}_{2} + \underbrace{\int_0^2 4e^{-t}\,dt}_{3} + \underbrace{\int_0^2 \cos(t)\,dt}_{4} + \underbrace{\int_0^2 1\,dt}_{5}$$

1️⃣ Pour la première intégrale, on réalise une IPP. On remarque qu'on a le produit d'une fonction (t^2) et de la dérivée d'une autre ($-e^{-t}$, dérivée de e^t). On a donc :

$$\int_0^2 -e^{-t}t^2\,dt = \left[e^{-t}t^2\right]_0^2 - \int_0^2 2te^{-t}\,dt$$

4

3️⃣ On s'intéresse ensuite à la 3ème intégrale, on peut ici simplement primitiver la fonction $-e^{-t}$ après avoir sorti la constante de l'intégrale :

$$\int_0^2 4e^{-t}\,dt = -4\int_0^2 -e^{-t}\,dt = -4\left[e^{-t}\right]_0^2$$

4️⃣ De même pour la quatrième intégrale :

$$\int_0^2 \cos(t)dt = [\sin(t)]_0^2$$

5️⃣ De même pour la dernière intégrale :

$$\int_0^2 1dt = [t]_0^2$$

Si on remet tout bout à bout on a :

$$\begin{aligned}\int_0^2 k(t)dt &= \left[e^{-t}t^2\right]_0^2 - \int_0^2 2te^{-t}dt + \int_0^2 2te^{-t}dt - 4\left[e^{-t}\right]_0^2 + [\sin(t)]_0^2 + [t]_0^2\\ &= \left[e^{-t}t^2\right]_0^2 - 4\left[e^{-t}\right]_0^2 + [\sin(t)]_0^2 + [t]_0^2\\ &= e^{-2}2^2 - 0 - 4\left(e^{-2}-1\right) + \sin(2) - 0 + 2 - 0\\ &= 4 + \sin(2) + 2\\ &\approx 6{,}91 \qquad \textit{(A FAUX)(C FAUX)}\textbf{(B VRAI)}\end{aligned}$$

? Items D et E

On va utiliser ici la même méthode mais entre 0 et 4 :

$$\begin{aligned}\int_0^4 k(t)dt &= \left[e^{-t}t^2\right]_0^4 - \int_0^4 2te^{-t}dt + \int_0^4 2te^{-t}dt - 4\left[e^{-t}\right]_0^4 + [\sin(t)]_0^4 + [t]_0^4\\ &= \left[e^{-t}t^2\right]_0^4 - 4\left[e^{-t}\right]_0^4 + [\sin(t)]_0^4 + [t]_0^4\\ &= e^{-4}4^2 - 0 - 4\left(e^{-4}-1\right) + \sin(4) - 0 + 4 - 0\\ &= 12e^{-4} + \sin(4) + 8\\ &\approx 8{,}20 \qquad \textit{(D FAUX)}\textbf{(E FAUX)}\end{aligned}$$

Réponse vraie : B

On étudie un modèle qui traduit l'évolution des groupes de J-pop $j(t)$ selon deux paramètres a et b avec a supérieur à b. Le modèle définie sur $\mathbb{R}^2$ est le suivant :

$$j(a,b) = (a-b)e^{\frac{a}{b}}$$

Question 4 On donne les valeurs des paramètres suivants pour une valeur de $j = 593{,}65$, $a = 5{,}00 \pm 0{,}05$ et $b = 1{,}00 \pm 0{,}01$. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La plus grosse incertitude vient de la variable b

B. $\Delta j \leq 44{,}52$

C. $\Delta j \leq 68{,}27$

D. $j \approx 593{,}65 \pm 44{,}52$

E. $j \approx 593{,}65 \pm 68{,}27$

Question 4

? Items A, B, C, D et E

Alors là il faut se rappeler de la formule du cours qui nous dit :

$$\Delta f(x,y) \leq \left|\frac{\partial f}{\partial x}\right|\Delta x + \left|\frac{\partial f}{\partial y}\right|\Delta y$$

Concrètement cette formule nous dit que la variation possible d'une fonction à plusieurs variables est forcément plus petite ou égale à la somme des valeurs absolues des différentielles selon chaque variable. Appliqué à notre fonction, on a :

$$\Delta j(a,b) \leq \left|\frac{\partial j}{\partial a}\right|\Delta a + \left|\frac{\partial j}{\partial b}\right|\Delta b$$

On va ensuite faire la dérivée partielle de la fonction selon chaque variable (on se rappelle que faire la dérivée partielle selon une variable c'est simplement dériver selon cette dernière en considérant les autres variables comme des constantes dans le calcul) :

Calcul de $\frac{\partial j}{\partial a}$

$$\begin{aligned}\frac{\partial j}{\partial a} &= \frac{\partial (a-b)e^{\frac{a}{b}}}{\partial a}\\ &= \frac{\partial a e^{\frac{a}{b}}}{\partial a} - b\frac{\partial e^{\frac{a}{b}}}{\partial a}\\ &= \frac{\partial a}{\partial a}\times e^{\frac{a}{b}} + a\times\frac{\partial e^{\frac{a}{b}}}{\partial a} - e^{\frac{a}{b}}\\ &= e^{\frac{a}{b}} + a\times\frac{1}{b}\times e^{\frac{a}{b}} - e^{\frac{a}{b}}\\ &= \frac{a}{b}e^{\frac{a}{b}}\end{aligned}$$

5

Calcul de $\frac{\partial j}{\partial b}$

$$\begin{aligned}\frac{\partial j}{\partial b} &= \frac{\partial (a-b)e^{\frac{a}{b}}}{\partial b}\\ &= a\frac{\partial e^{\frac{a}{b}}}{\partial b} - \frac{\partial b e^{\frac{a}{b}}}{\partial b}\\ &= -\frac{a^2}{b^2}e^{\frac{a}{b}} - \left(\frac{\partial b}{\partial b}\times e^{\frac{a}{b}} + b\times\frac{\partial e^{\frac{a}{b}}}{\partial b}\right)\\ &= -\frac{a^2}{b^2}e^{\frac{a}{b}} - e^{\frac{a}{b}} + \frac{a}{b}e^{\frac{a}{b}}\\ &= e^{\frac{a}{b}}\left(-\frac{a^2}{b^2}+\frac{a}{b}-1\right)\end{aligned}$$

Si on met tout ça bout à bout on se retrouve avec :

$$\begin{aligned}\Delta j(a,b) &\leq \left|\frac{a}{b}e^{\frac{a}{b}}\right|\Delta a + \left|e^{\frac{a}{b}}\left(-\frac{a^2}{b^2}+\frac{a}{b}-1\right)\right|\Delta b\\ &\leq \left|\frac{5}{1}e^{\frac{5}{1}}\right|\times 0{,}05 + \left|e^{\frac{5}{1}}\left(-\frac{5^2}{1^2}+\frac{5}{1}-1\right)\right|\times 0{,}01\\ &\leq 37{,}1033 + 31{,}1667\\ &\leq 68{,}2700\end{aligned}$$

On constate qu'avec la dérivée partielle en a a la valeur la plus grande, c'est donc l'incertitude en a qui donne le plus de poids à l'incertitude totale. *(A FAUX)(B FAUX)***(C VRAI)***(D FAUX)***(E VRAI)**

Réponses vraies : C et E

On pose un ultime modèle, pour tout t positif :

$$k(t) = \frac{3t^3 - 2t^2 - 6t + 12}{3t^2 - 2t + 2}$$

On donne aussi :

$$\int_0^1 \frac{2t-10}{3t^2-2t+2}dt \approx -4{,}67$$

Question 5 Concernant cette fonction, laquelle (lesquelles) des propositions est (sont) exacte(s) ?

A. $k'(t) = \frac{-9t^2+4t+6}{(3t^2-2t+2)^2}$

B. $k'(t) = \frac{-9t^4+12t^3-40t^2+80t-12}{3t^2-2t+2}$

C. k est définie sur $]-\infty, +\infty[$

D. $\int_0^1 k(t) = [t]_0^1 - [\ln(3t^2-2t+2)]_0^1 + \int_0^1 \frac{2t-10}{3t^2-2t+2}dt$

E. $\int_0^1 k(t)\,dt \approx 4{,}76$

Question 5

Items A et B

Comment faire pour calculer la petite (grande mdr) dérivée ?? Pas de panique, on va décomposer tout ça.
On va commencer par poser :

$$u = 3t^3 - 2t^2 - 6t + 12 \quad et \quad v = 3t^2 - 2t + 2$$

On va ensuite dériver ces petites fonctions : $u' = 9t^2 - 4t - 6$ et $v' = 6t - 2$.

Maintenant, on sait que $(u/v)' = \frac{(u'v - uv')}{v^2}$, on a plus qu'à remplacer tout ça et voilà !

$$\begin{aligned}k'(t) &= \frac{-(-9t^2+4t+6)(3t^2-2t+2) - (3t^3-2t^2-6t+12)(6t-2)}{(3t^2-2t+2)^2}\\ &= \frac{(27t^4-30t^3+8t^2+4t-12)-(18t^4-18t^3-32t^2+84t-24)}{(3t^2-2t+2)^2}\\ &= \frac{9t^4-12t^3+40t^2-80t+12}{(3t^2-2t+2)^2}\end{aligned}$$

Item C → *k est définie sur $]-\infty, +\infty[$*

Comment trouver le domaine de définition ?

Ici, on a un rapport de polynômes comme fonction, le seul interdit serait de diviser par 0, c'est-à-dire que le polynôme du bas $3t^2 - 2t + 2$ s'annule. On doit donc déterminer ses racines pour les exclure de l'ensemble de définition car **ON NE DIVISE JAMAIS PAR 0**

Ainsi, le discriminant $\Delta = (-2)^2 - 4\times 3\times 2 = -20 < 0$

Puisque le discriminant est négatif, le polynôme n'admet aucune racine réelle, et donc ne s'annule pas. Autrement dit, pour toute valeur de t, le polynôme sera différent de 0. La fonction k est donc définie pour tout réel, soit sur $]-\infty, +\infty[$.

Items D et E

Ici, on vous demande d'intégrer la fonction, qui est un rapport de polynômes : pas évident à faire comme ça. Il faut d'abord essayer de simplifier au maximum l'expression avant de pouvoir décomposer l'intégrale $\int_0^1 k(t)\,dt$ en plusieurs intégrales plus simples. Pour cela, on va utiliser la division euclidienne de polynômes avec ici en numérateur $3t^3 - 2t^2 - 6t + 12$ et en dénominateur $3t^2 - 2t + 2$:

$$\begin{array}{rrrrr|l} & 3t^3 & -\ 2t^2 & -\ 6t & +\ 12 & 3t^2-2t+2\\ - & (3t^3 & -\ 2t^2 & +\ 2t) & & t\\ & & & -\ 8t & +\ 12 & \end{array}$$

(Avec : $-(t)(3t^2-2t+2) = -3t^3 + 2t^2 - 2t$ pour pouvoir faire la première soustraction)

6

Donc,

$$\frac{3t^3-2t^2-6t+12}{3t^2-2t+2} = t + \frac{-8t+12}{3t^2-2t+2} = t + \frac{-6t+2}{3t^2-2t+2} + \frac{-2t+10}{3t^2-2t+2}$$

On va déjà séparer au maximum et primitiver ce que l'on peut pour y voir plus clair :

- $\int_0^1 (t)\,dt = \left[\frac{1}{2}t^2\right]_0^1$
- $-\int_0^1 \frac{6t-2}{3t^2-2t+2}\,dt = -\left[\ln(3t^2-2t+2)\right]_0^1$

On se retrouve alors avec :

$$\int_0^1 k(t)\,dt = \left[\frac{1}{2}t^2\right]_0^1 - \left[\ln(3t^2-2t+2)\right]_0^1 - \int_0^1 \frac{2t-10}{3t^2-2t+2}\,dt$$

On a alors plus qu'à remplacer la dernière intégrale par sa valeur donnée dans l'énoncé et calculer le reste !

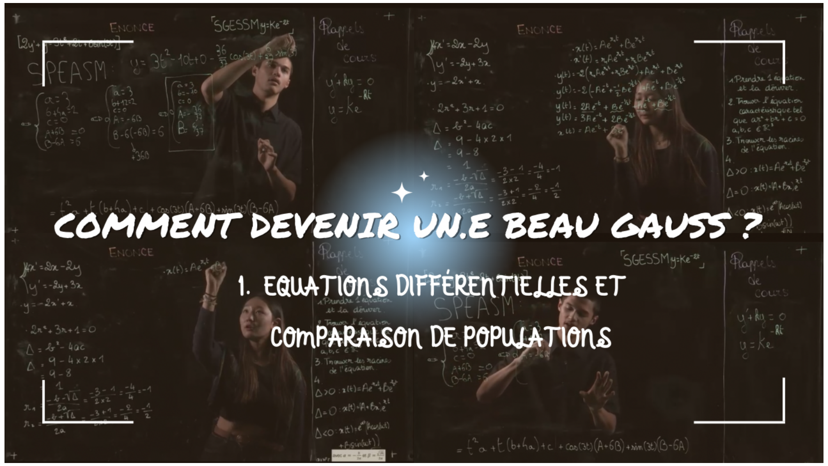
$$\int_0^1 k(t)\,dt = \frac{1}{2} - (\ln(3)-\ln(2)) + 4{,}67 \approx 4{,}76$$

(D FAUX) **(E VRAI)**

Bien joué à toi !

Tu peux être fier et peut être triste d'avoir terminé l'exo de bio-maths, si t'as galéré voici des vidéos qu'on ta préparé ! BON COURAGE ♡

Réponses vraies : C et E



Exercice 2 (questions 6 à 10)

Au service des urgences, 50% des patients viennent consulter pour des difficultés respiratoires. Parmi ces patients, 30% ont une pneumonie. On note également que 10% des patients se présentant aux urgences sans difficultés respiratoires ont tout de même une pneumonie.

Question 6 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La sensibilité du signe "difficulté respiratoire" pour la maladie "pneumonie" est de 30%
B. La spécificité du signe "difficulté respiratoire" pour la maladie "pneumonie" est de 10%
C. Un patient qui se présente aux urgences avec des difficultés respiratoires à 30% de chances d'avoir une pneumonie
D. Un patient qui se présente aux urgences avec difficultés respiratoires à 30% de chances de ne pas avoir une pneumonie
E. Si la prévalence de la pneumonie aux urgences augmente, la sensibilité du signe "difficultés respiratoires" augmente

Question 6

✘ **Item A** → *La sensibilité du signe "difficulté respiratoire" pour la maladie "pneumonie" est de 30%*

La sensibilité est la probabilité d'avoir le signe si on est malade. On note M l'évènement "le patient a une pneumonie" et S l'évènement "le patient présente le signe "difficulté respiratoire"". Mathématiquement parlant,

$$Se = \mathbb{P}(S|M)$$

D'après la formule de Bayes :

$$\mathbb{P}(S|M) = \frac{\mathbb{P}(M|S)\times\mathbb{P}(S)}{\mathbb{P}(M)}$$

Or d'après l'énoncé, $\mathbb{P}(S) = 0.5$, et $\mathbb{P}(M|S) = 0.3$. On peut calculer, d'après la formule des probabilités totales :

$$\mathbb{P}(M) = \mathbb{P}(M\cap S) + \mathbb{P}(M\cap\overline{S}) = 0.5\times 0.3 + 0.5\times 0.1 = 0.2$$

La probabilité d'être malade est de 0.2, on peut donc reprendre le calcul de la sensibilité :

$$Se = \mathbb{P}(S|M) = \frac{\mathbb{P}(M|S)\times\mathbb{P}(S)}{\mathbb{P}(M)} = \frac{0.3\times 0.5}{0.2} = 0.75$$

La sensibilité est ici de 75%.

7

Item B → *La spécificité du signe "difficulté respiratoire" pour la maladie "pneumonie" est de 10%*

La spécificité est la probabilité de ne pas avoir le signe lorsqu'on n'est pas malade. C'est-à-dire :

$$Sp = P\left(\overline{S}|\overline{M}\right)$$

Toujours d'après la formule de Bayes (présente dans le formulaire) :

$$\mathbb{P}\left(\overline{S}|\overline{M}\right) = \frac{\mathbb{P}(\overline{S}) \times \mathbb{P}(\overline{M}|\overline{S})}{\mathbb{P}(\overline{M})}$$

Donc,

$$\begin{aligned} Sp &= P(\overline{S}|\overline{M}) \\ &= \frac{\mathbb{P}(\overline{S}) \times \mathbb{P}(\overline{M}|\overline{S})}{\mathbb{P}(\overline{M})} \\ &= \frac{\mathbb{P}(\overline{S}) \times \mathbb{P}(\overline{M}|\overline{S})}{1 - \mathbb{P}(M)} \\ &= \frac{0.5 \times 0.9}{1 - 0.2} \\ &= 0.56 \end{aligned}$$

La spécificité est donc de 56%.

? Items C et D

L'énoncé dit que 30% des patients qui se présentent aux urgences pour des difficultés respiratoires ont une pneumonie. Donc un patient qui se présente aux urgences avec des difficultés respiratoires à effectivement 30% de chances d'avoir une pneumonie. **(C VRAI)**

Cela veut aussi dire que parmi ces patients qui se sont présentés pour des difficultés respiratoires, 70% n'ont pas de pneumonie ! Un patient qui se présente aux urgences avec difficultés respiratoires à 70% de chances de ne pas avoir une pneumonie. *(D FAUX)*

Item E → *Si la prévalence de la pneumonie aux urgences augmente, la sensibilité du signe ...*

Ici question de cours !
↪ la spécificité et la sensibilité sont indépendantes de la prévalence de la maladie. A l'inverse, la VPP augmente quand la prévalence augmente et la VPN diminue quand la prévalence augmente.

Réponse vraie : C

Parmi les 50% des patients se présentant pour une difficulté respiratoire, on en choisit au hasard 20. La probabilité pour ces patients d'avoir une pneumonie est la même que précédemment. On note X la variable aléatoire qui donne le nombre de patients ayant une pneumonie parmi ces 20 personnes.

Question 7 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La variable X suit une loi binomiale de paramètres $n = 20$ et $\pi = 0.5$
B. La variable X suit une loi de Poisson de paramètre $\lambda = 0.6$
C. La probabilité qu'aucun des patients examinés ne soit malade est nulle
D. La probabilité que la moitié des patients soit malade est de 50%
E. L'écart-type de X est d'environ 2 patients

Question 7

Item A → *La variable X suit une loi binomiale de paramètres $n = 20$ et $\pi = 0.5$*

On répète 20 fois une expérience aléatoire de manière indépendante et identique, qui possède deux issues : le patient est malade ou le patient n'est pas malade. La probabilité que le patient interrogé soit malade est de 0.3.

La variable X, qui compte le nombre de patients malades parmi les 20 personnes interrogées, suit donc une loi binomiale de paramètres $n = 20$ et $\pi = 0.3$.

Item B → *La variable X suit une loi de Poisson de paramètre $\lambda = 0.6$*

La loi de Poisson est applicable si $\pi << 1$ et n très grand devant π.

On voit déjà que la condition $\pi << 1$ n'est pas validée car $\pi = 0.3$, la loi de Poisson est dite loi des événements rares. Aucune condition n'est validée ici. mince notre évènement n'est pas assez rare !

Item C → *La probabilité qu'aucun des patients examinés ne soit malade est nulle*

On peut ici calculer $\mathbb{P}(X = 0)$ avec la formule de la loi binomiale, qui est :

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

On obtient,

$$\begin{aligned} \mathbb{P}(X = 0) &= \binom{20}{0} \times 0.3^0 \times (1 - 0.3)^{20-0} \\ &= 1 \times 1 \times 0.7^{20} \\ &\approx 0.0008 \end{aligned}$$

Donc la valeur est bien proche de zéro mais pas nulle ! C'est un évènement très peu probable mais pas impossible :)

Sinon, plus rapide, on peut s'aider du tableau présent dans le formulaire. On prend $n = 20$ et $k = 0$. On se rend compte que la valeur $p = 0.3$ n'est pas présente, donc on va encadrer notre résultat entre les valeurs pour $p = 0.25$ et $p = 0.4$.

On lit que $\mathbb{P}(X = 0)$, lorsque $p = 0.25$, est égale à 0.0032 et $\mathbb{P}(X = 0)$, pour $p = 0.4$, est égal à 0.

Donc on peut supposer $\mathbb{P}(X = 0)$ pour $p = 0.3$ est comprise entre 0.032 et 0.

✘ **Item D** → *La probabilité que la moitié des patients soit malade est de 50%*

On cherche $\mathbb{P}(X = 10)$. On prend la formule de la loi binomiale

$$\begin{aligned}\mathbb{P}(X = 10) &= \binom{20}{10} \times 0.3^{10} \times (1 - 0.3)^{20-10} \\ &\approx 0.3\end{aligned}$$

La probabilité qu'exactement la moitié des patients soit malade est de 3%.

✔ **Item E** → *L'écart-type de X est d'environ 2 patients*

Pour la loi binomiale,

$$\begin{aligned}\sigma &= \sqrt{\mathbb{V}ar(X)} \\ &= \sqrt{np(1-p)}\end{aligned}$$

Donc ici

$$\begin{aligned}\sigma &= \sqrt{20 \times 0.3 \times 0.7} \\ &\approx 2 \text{ patients}\end{aligned}$$

Réponse vraie : E

On prend désormais un échantillon de 30 patients, qui ont toujours la même probabilité d'être malade. X est toujours la variable aléatoire qui donne le nombre de patients malade parmi l'échantillon.

Question 8 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. On peut approximer X par une loi normale de paramètres $\mu = 9$ et $\sigma^2 = 6.3$

B. La variable $Z = \frac{X-\mu}{\sigma}$ suit une loi normale centrée non-réduite

C. La probabilité que le nombre de patients malades parmi les 30 interrogés soit supérieur à 15 est d'environ 0.02

D. La probabilité que le nombre de patients malades parmi les 30 interrogés soit inférieur à 10 est d'environ 0.31

E. Si $Z \sim \mathcal{N}(0,1)$, $\mathbb{P}(Z > 2) = \mathbb{P}(Z < 2)$

Question 8

✔ **Item A** → *On peut approximer X par une loi normale de paramètres $\mu = 9$ et $\sigma^2 = 6.3$*

X suit une loi binomiale de paramètre $n = 30$ et $\pi = 0.3$. Pour pouvoir approximer une loi binomiale par une loi normale, il y a deux conditions :

$$\begin{cases} n(1-\pi) \geq 5 \\ n\pi \geq 5 \end{cases}$$

Or ici :

$$\begin{cases} n(1-\pi) = 30 \times 0.7 = 21 \\ n\pi = 30 \times 0.3 = 9 \end{cases}$$

Les trois conditions sont validées, on peut donc approximer la loi suivie par X par une loi normale de paramètres :

$$\begin{cases} \mu = n\pi = 9 \\ \sigma^2 = n\pi(1-\pi) = 9 \times 0.7 = 6.3 \end{cases}$$

✘ **Item B** → *La variable $Z = \frac{X-\mu}{\sigma}$ suit une loi normale centrée non-réduite*

La variable $Z = \frac{X-\mu}{\sigma}$ suit une loi normale dite "standard" ou "centrée-réduite" car sa variance est de 1 et son espérance est de 0. Les démonstrations sont dans les cours si besoin.

✘ **Item C** → *La probabilité que le nombre de patients malades parmi les 30 interrogés soit ...*

On va se servir ici de la fameuse variable $Z = \frac{X-\mu}{\sigma}$, pour pouvoir utiliser la loi normale centrée réduite. On cherche $\mathbb{P}(X > 15)$, on va donc calculer :

$$\begin{aligned}\mathbb{P}(X > 15) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} > \frac{15-9}{\sqrt{6.3}}\right) \\ &= \mathbb{P}(Z > 2{,}39)\end{aligned}$$

Attention, le dénominateur est σ donc on prend bien $\sqrt{\sigma^2}$, c'est-à-dire $\sqrt{6.3}$.

On prend la table statistique 2b du formulaire, qui indique que $\mathbb{P}(Z > x_\beta) = \beta$. D'après nos calculs, ici $x_\beta \approx 2.39$. Dans la table la valeur la plus proche de 2.39 est 2.4.

Pour $x_\beta = 2.36$, $\beta = 0.0082$.

Donc $\mathbb{P}(Z > 2.36) = 0.0082$, ce qui signifie que la probabilité qu'il y ait plus de 15 malades dans notre échantillon est de 0.0082 soit 0.82%

✘ **Item D** → *La probabilité que le nombre de patients malades parmi les 30 interrogés soit ...*

On cherche $\mathbb{P}(X < 10)$. On reprend notre variable Z, et on calcule comme suit :

$$\begin{aligned}\mathbb{P}(X > 10) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} > \frac{10-9}{\sqrt{6.3}}\right) \\ &= \mathbb{P}(Z > 0{,}399)\end{aligned}$$

$$Z = \frac{10-9}{\sqrt{6.3}} \approx 0.399$$

On reporte dans la table 2b et pour $x_\beta = 0.399$ (valeur la plus proche 0.4) on trouve $\beta = 0.345$.

β est la probabilité que Z soit supérieur à x_β, c'est à dire la probabilité que X soit supérieur à 10 dans notre échantillon. Donc ici la probabilité cherchée est $1 - \beta = 0.655$.

La probabilité qu'il y ait moins de 10 malades dans notre échantillon est d'environ 0.655.

✘ **Item E** → *Si $Z \sim \mathcal{N}(0,1)$, $\mathbb{P}(Z > 2) = \mathbb{P}(Z < 2)$*

La loi normale centrée réduite est symétrique, cela signifie que que $\mathbb{P}(Z \geq a) = P(Z \leq -a)$

C'est déjà la deuxième moitié du sujet 😢

J'espère que tu comprends bien la correction, bon courage à toi pour la suite, peu importe que tu aies réussi ou pas le tuto, continue bien la correction, l'important dans les défaites c'est que ça ne devienne pas des échecs dont l'on en apprend rien. Donc lezz goo bien faire la correction et si question gogo fofo

Réponse vraie : A

La durée de l'hospitalisation des patients atteints de pneumonie (notée D) suit une loi exponentielle de paramètre λ. On sait que la probabilité qu'un patient déjà hospitalisé depuis 3 jours reste encore au moins 5 jours de plus est de 0.37.

Question 9 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La variable D suit une loi exponentielle de paramètre $\lambda = 0.4$

B. La probabilité qu'un patient reste au total moins de 2 semaines alors qu'il est déjà là depuis 6 jours est égale à 0.8

C. 50% des patients restent moins de 5 jours

D. En moyenne, les patients sont hospitalisés pendant 5 jours

E. On peut approcher la variable D par une loi de Poisson de paramètre λ

Question 9

✘ **Item A** → *La variable D suit une loi exponentielle de paramètre $\lambda = 0.4$*

La loi exponentielle est une loi sans mémoire, ce qui veut dire que :

$$\mathbb{P}(D > a + c | D > a) = \mathbb{P}(D > c)$$

Si on traduit mathématiquement l'énoncé, on nous dit en fait que :

$$\begin{aligned} P(D > 3 + 5 | D > 3) &= P(D > 5) \\ &= 0.37 \end{aligned}$$

Or, la loi exponentielle nous dit que $\mathbb{P}(D > X) = e^{-\lambda X}$
Donc ici

$$\begin{aligned} &\mathbb{P}(D > 5) = 0.37 \\ \iff &e^{-5\lambda} = 0.37 \\ \iff &-5\lambda = \ln(0.37) \\ \iff &\lambda = -\frac{\ln(0.37)}{5} \\ &\approx 0.2 \end{aligned}$$

Donc $D \sim \exp(0.2)$

✔ **Item B** → *La probabilité qu'un patient reste au total moins de 2 semaines alors qu'il est déjà ...*

Ici on cherche $\mathbb{P}(D < 14 | D > 6)$, équivalente à $\mathbb{P}(D < 14 - 6)$. On n'oublie pas de tout mettre en jour !! Si on prend 2 semaines à la place de 14 jours, ça va cafouiller.

$$\begin{aligned} \mathbb{P}(D < 14 | D > 6) &= \mathbb{P}(D < 8) \\ &= \left(1 - e^{-8\lambda}\right) \\ &= 0.8 \end{aligned}$$

✘ **Item C** → *50% des patients restent moins de 5 jours*

Pour cet item on cherche la médiane. Pour la loi exponentielle,

$$\begin{aligned} \text{med} &= \ln(2) \times \mathbb{E}(D) \\ &= \ln(2) \times \frac{1}{\lambda} \\ &= \ln(2) \times \frac{1}{0.2} \\ &= 3.5 \text{ jours} \end{aligned}$$

✔ **Item D** → *En moyenne, les patients sont hospitalisés pendant 5 jours*

Ici on cherche la moyenne, c'est à dire l'espérance.

$$\mathbb{E}(X) = \frac{1}{\lambda} = \frac{1}{0.2} = 5 \text{ jours}$$

Attention, au delà des formules à connaître, le piège de cette question est que la moyenne et la médiane sont deux grandeurs différentes, à ne pas confondre !!!

✘ **Item E** → *On peut approcher la variable D par une loi de Poisson de paramètre λ*

FAUX ! On ne peut faire aucune approximation entre la loi exponentielle et la loi de Poisson.

Réponses vraies : B et D

On s'intéresse à la moyenne théorique du nombre de patients atteints de pneumonie se présentant aux urgences sur 1 mois (M1) et sur 1 an (M12).

On tire au sort un échantillon de 150 patients aux urgences d'un hôpital parisien, et on trouve parmi les patients de cet échantillon 60 cas de pneumonie.

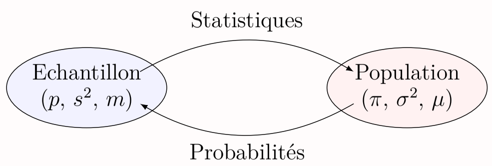
Question 10 Parmi les propositions suivantes, laquelle (ou lesquelles) est (sont) exacte(s) ?

A. La probabilité que le patient de l'échantillon soit malade est notée $p = 0.4$
B. L'intervalle de pari à 95% de M1 est plus large que l'intervalle de pari à 95% de M12
C. L'intervalle de pari à 99% de M1 est plus large que l'intervalle de pari à 95% de M1
D. L'intervalle de confiance à 95% de la proportion de patients atteints de pneumonie est $[32.3; 47.8]\%$ à 0.05% près
E. L'intervalle de pari à 95% de la proportion de patients atteints de pneumonie est $[32.3; 47.8]\%$ à 0.1% près

Question 10

✔ **Item A** → *La probabilité que le patient de l'échantillon soit malade est notée $p = 0.4$*

Ici on part d'un échantillon, donc la probabilité est noté p, la variance s^2 et la moyenne m. Quand on part de données théoriques, à l'inverse, on va utiliser des lettres grecques : la probabilité est notée π, la variance σ^2 et la moyenne μ.



✔ **Item B** → *L'intervalle de pari à 95% de M1 est plus large que l'intervalle de pari à 95% de M12*

Ce qui change entre l'intervalle de pari à 95% de M1 et l'intervalle de pari à 95% de M12, c'est uniquement n le nombre de mois.

Or selon la formule, plus on augmente n, plus la largeur de l'intervalle est basse (il y a plus de semaines à comparer entre elles, donc la moyenne variera moins). Plus on a des données, donc sur un an, plus nos résultats vont être précis et l'intervalle va être aussi précis - c'est à dire petit.

Donc, ici, l'intervalle de pari à 95% de M1 est plus large que l'intervalle de pari à 95% de M12 car n est plus petit.

✔ **Item C** → *L'intervalle de pari à 99% de M1 est plus large que l'intervalle de pari à 95% de M1*

Ici on compare l'intervalle de pari à 99%, qui est lié à $\varepsilon_{0,01}$, de M1 contre l'intervalle de pari à 95%, qui est lié à $\varepsilon_{0,05}$, de M1.

Selon le formulaire, on a

$$\varepsilon_{0,05} = 1{,}960 \qquad \text{et} \qquad \varepsilon_{0,01} = 2{,}576$$

donc

$$\varepsilon_{0,01} > \varepsilon_{0,05}$$

Or, selon la formule, si on augmente ε_α, alors on augmente la largeur de l'intervalle de pari. Donc l'intervalle de pari à 99% de M1 est plus large que l'intervalle de pari à 95% de M1.

? **Items D et E**

On peut rayer d'office un des items : on essaie de calculer la proportion théorique π, à partir d'une proportion observée p ; on effectue donc ici un intervalle de confiance. L'item E est donc faux.

Place au calcul :

$$IC(\pi) = p \pm \varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}}$$

On trouve cette formule dans le formulaire.

On nous demande ici un intervalle de confiance à 95%, donc notre $\varepsilon_{\alpha=0.05}$, vaut 1.96. On calcule ensuite notre probabilité p, qui vaut, $p = \frac{60}{150} = 0.4$

On y retourne :

$$\begin{aligned} IC(\pi) &= 0.4 \pm 1.96\sqrt{\frac{0.4(1-0.4)}{150}} \\ &= 0.4 \pm 0.078 \\ &= 40\% \pm 7.8\%, \text{ à } 0.1\% \text{ près} \end{aligned}$$

$$IC(\pi) \in [32.2\%; 47.8\%], \text{ à } 0.1\% \text{ près}$$

L'item D est faux car l'IC est à 0.1% près et non 0.05%. Et si l'item D indiquait à 0.1% près, on aurait pu cocher ?? Peut-être pas !!!

On doit d'abord vérifier les conditions du TCL, sinon TOUT est faux !!!

⚠ **Attention !!!**

Pour valider le TCL et pouvoir faire un IC, il faut :

$$\begin{cases} n \times l_{inf} \geq 5 \\ n \times (1 - l_{inf}) \geq 5 \\ n \times l_{sup} \geq 5 \\ n \times (1 - l_{sup}) \geq 5 \end{cases}$$

Ça c'est vraiment à **connaître** par coeur ❤️ et à **vérifier systématiquement**, ce piège de validité du TCL tombe tout le temps !!

Réponses vraies : A, B et C

Exercice 3 (questions 11 à 15)

Importée d'Ethiopie, le café s'est vite répandu dans l'Empire Ottoman, devenant ce qu'on connaît aujourd'hui comme étant le café turc. Il est considéré comme stimulant la mémoire et réduisant certaines maladies. Une étude parue en 2015 dans le Journal of the International Society of Sports Nutrition s'intéresse de près à l'effet du café turc sur la performance et le temps de réaction. Plusieurs aspects de ces effets sont étudiés dans cet exercice. L'étude a été réalisée sur $n = 60$ participants. $n_{TC} = 30$ participants ont ingéré du café turc et $n_{DC} = 30$ autres du café turc décaféiné. Les données utilisées sont tirées ou simulées à partir de cet article. On prend $\alpha = 5\%$.

La pression artérielle diastolique a été mesurée chez les participants juste avant l'exercice. Le groupe des participants ayant ingéré du café turc (TC) présente une pression artérielle diastolique moyenne de 122 $mmHg$. Le groupe des participants ayant ingéré du café turc décaféiné (DC) présente une pression diastolique moyenne de 115 $mmHg$. Un test a été conduit afin de savoir si la pression artérielle diastolique varie avec le type de café. L'intervalle de la différence entre la pression artérielle diastolique TC et DC est $[6.7; 7.3]$ au risque $\alpha = 5\%$.

Question 11 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. L'hypothèse nulle est l'indépendance entre le type de café et la pression artérielle diastolique dans la population

B. Les pressions artérielles diastoliques entre les participants TC et DC est significativement identique au risque 5%

C. La pression artérielle diastolique est significativement différente entre la population TC et la population DC au risque 5%

D. La conclusion du test peut reposer sur un intervalle de pari ou un critère de test

E. Réaliser un test statistique revient à tester l'appartenance du critère de test z à la distribution de la variable aléatoire du critère de test Z sous H_0

Question 11

✔ Item A → *L'hypothèse nulle est l'indépendance entre le type de café et la pression artérielle ...*

Avant de réponse directement à l'item, on doit comprendre le contexte ici. On cherche à savoir si la pression artérielle diastolique varie avec le type de café, c'est à dire si la pression artérielle diastolique (PAD) moyenne chez une **population** TC et DC est la même.

On s'intéresse aux paramètres, donc la PAD moyenne, dans la population TC et DC et pas dans les groupes à disposition. Il est facile de conclure à partir d'un échantillon mais cela n'a pas un grand intérêt. La volonté d'un test statistique est de conclure sur la population, de savoir ce qu'il se passe dans la population. Comme on n'a pas les données sur toute la population, on **se base sur les données d'échantillon**.

Notre but premier est de conclure sur la population, c'est à dire sur ce qui se passe en général. Les échantillons sont des cas particuliers de la population et on ne peut pas conclure dessus. Vous devez bien comprendre cela.

Ici notre H_0 est l'égalité entre les PAD dans la population TC et la population DC. Cela signifie aussi que le type de café est indépendant de la PAD. L'item est bien vraie.

? Items B, C et D

Maintenant on doit réaliser ce test de comparaison de moyennes. Cela peut être fait en réalité de deux manières : soit en calculant le critère de test z et en le comparant à ε_α, ou en utilisant un intervalle de confiance.

Pour calculer le critère de test, il faut avoir les moyennes m et les écart-types s. Ici s n'est pas donné donc on ne peut pas utiliser cette méthode.

En revanche on nous donne un intervalle et on peut l'utiliser. Pour comprendre pourquoi on peut l'utiliser, il faut reprendre le principe du test :

Le test veut conclure sur la population. Mais seules les données observées sont disponibles. Donc on va prendre ces données disponibles en prenant en compte la variabilité de la variable aléatoire. **On définit ainsi un champ de possibilité de la valeur théorique, en utilisant les données de l'échantillon.**

On compare ensuite une valeur fixée, en fonction de H_0, à ce champ de possibilité qui est en réalité un intervalle. Par exemple si on compare une moyenne à une valeur fixée μ_0, on va chercher à savoir si μ_0 est dans cet intervalle. Si on compare deux moyennes, on va chercher à savoir si les deux intervalles se superposent.

L'intervalle construit repose sur les données de l'échantillon : c'est un intervalle de **confiance** et non pas de pari qui est lui construit à partir des données de la population. Dans notre cas ici, on a fait un intervalle de confiance de la **différence** des deux moyennes.

On se rappelle aussi que :

$$H_0 : \mu_{TC} = \mu_{DC} \iff H_0 : \mu_{TC} - \mu_{DC} = 0$$

avec μ_{TC} la PAD moyenne dans la population TC et μ_{DC} la PAD moyenne dans la population DC. Sous H_0 la différence des PAD dans les populations TC et DC est nulle. On compare alors ce 0 à l'intervalle de confiance de la différence.

Cet intervalle ne contient pas la valeur de H_0 ce qui signifie que l'on rejette H_0. Rejeter H_0 ici revient à conclure à une différence entre μ_{TC} et μ_{DC} au risque 5% car c'est le risque avec lequel est construit l'intervalle de confiance.

Les conditions de validité sont vérifiées car :

$$n_{TC} = 30 \quad \text{et} \quad n_{DC} = 30$$

✔ Item E → *Réaliser un test statistique revient à tester l'appartenance du critère de test z à la ...*

Pourquoi cet item est vrai ? On va revenir encore une fois au principe des tests mathématiquement parlant :

On cherche à tester deux hypothèses H_0 et H_1. Pour cela, on va construire une variable aléatoire critère de test avec une distribution sous H_0, c'est à dire en partant du principe que H_0 est vraie.

Ensuite on va calculer le critère de test à partir de l'échantillon. Si cette estimation du critère de test est une valeur réaliste de la variable aléatoire sous H_0 alors cela signifie que l'échantillon en soi est sous H_0, donc on ne rejettera pas H_0. Si ce n'est pas une valeur réaliste, cela signifie que l'échantillon n'est pas conforme à H_0 donc on rejettera H_0.

Pour savoir si le critère de test est réaliste, on le compare à la distribution de la variable aléatoire. S'il est dans les queues de distribution, c'est à dire $> \varepsilon_\alpha$, il n'est pas très réaliste.

Réponses vraies : A, C et E

La performance physique a ensuite été évaluée sur une course. Le nombre de participants ayant parcouru plus de 5 km a été dénombré. Dans le groupe TC, 60% des participants ont couru plus de 5km alors que dans le groupe DC, 20% des participants ont couru plus de 5km. Un test a été réalisé afin de comparer la proportion de personnes ayant couru plus de 5km dans les deux cas. Le degré de signification est entre 1% et 2%.

Question 12 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La statistique de test est entre 2,57 et 2,32

B. La statistique de test est entre 1,64 et 1,28

C. La proportion de proportion de personnes ayant couru plus de $5km$ est significativement différente entre le groupe TC et DC au risque 5%

D. La proportion de proportion de personnes ayant couru plus de $5km$ est significativement identique entre une population TC et DC au risque 5%

E. Ici, le critère de test z est dans les queues de distribution de la variable aléatoire du critère de test Z

Question 12

? Items A et B

Dédicace à ma professeure de Biostatistiques pour ces deux items ingénieux.

Le degré de signification ds est directement donné dans l'énoncé. Or si on reprend la définition de ce pourcentage : c'est le plus petit risque α que l'on peut prendre en continuant à rejeter H_0 pour cet échantillon.

ds est donc le risque α si $\varepsilon_\alpha = |z|$.

Pour retrouver le critère de test z, on doit chercher ε_α correspondant à ce ds. En regardant dans la table $2a$, on regarde les ε_α correspondant à un risque α entre 1% et 2%. On trouve deux ε_α : 2,57 et 2,32.

$$2{,}32 < z < 2{,}57$$

? Items C et D

Pour répondre à ces items, on peut utiliser deux méthodes :

1. La méthode traditionnelle consiste à comparer z à $\varepsilon_{5\%}$.

$$\varepsilon_{5\%} = 1{,}96$$
$$z > 1{,}96$$

Les conditions de validité sont vérifiées ici aussi.
On rejette alors H_0 et on conclut à une différence au risque 5%.

2. La deuxième méthode utilisée en réalité est la comparaison du ds au risque α qui est ici de 5%. En effet, si le ds est inférieur à 5%, cela signifie que $z > 1{,}96$, donc on rejette H_0.

On conclut sur la population et pas sur l'échantillon. Ce dernier n'est qu'un outil pour avoir des informations générales. L'échantillon n'est qu'un cas particulier.

✔ Item E → *Ici, le critère de test z est dans les queues de distribution de la variable aléatoire du ...*

On se rappelle qu'on a rejeté H_0 et que $z > 1{,}96$. Cela signifie donc que z est une valeur extrême de Z, elle est dans la queue de distribution de Z.

Le principe du test c'est de rejeter H_0 lorsque z n'est pas une valeur réaliste de Z justement, comme expliqué dans la correction de la question 1.

Réponses vraies : A et E

Des tests de concentration ont été réalisés en mesurant le temps de réaction des participants. Parmi les participants TC, 3 ont présenté un temps de réaction supérieur ($>$) à $0{,}1s$ et 27 ont présenté un temps de réaction inférieur ($<$) à $0{,}1s$. Le temps de réaction dans la population d'individus DC se distribue de la manière suivante : $40\% > 0{,}1s$ et $60\% < 0{,}1s$.
Un test du χ^2 est réalisé.

Question 13 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Le test réalisé peut être un test d'indépendance

B. Le degré de signification de ce test est supérieur à 1%

C. Le degré de signification de ce test est inférieur à 0,1%

D. Si le degré de signification aurait été plus grand, cela signifie que la différence entre les temps de réaction TC et DC aurait été plus grande

E. Si la loi du temps de réaction n'était pas une loi normale, on aurait pu quand même faire ce test

Question 13

✘ Item A → *Le test réalisé peut être un test d'indépendance*

On réalise un test du χ^2 en comparant des données observées à des données standards, théoriques. Ce type de test est un test d'adéquation et non d'indépendance.

Si on comparait deux échantillons, ç'aurait été effectivement un test d'indépendance.

? Items B et C

On réalise donc ce test d'ajustement en construisant le tableau de contingence suivant :

	$> 0.1s$	$< 0.1s$
Effectifs observés	3	27
Effectifs théoriques	12	18

La deuxième ligne correspond aux effectifs théoriques c. Ce sont les effectifs dans l'échantillon lorsqu'on applique rigoureusement les proportions de la population dans cet échantillon.

13

On calcule ensuite le critère de test K :

$$\begin{aligned} K &= \sum \frac{(o-c)^2}{c} \\ &= \frac{(3-12)^2}{12} + \frac{(27-18)^2}{18} \\ &= \frac{81}{12} + \frac{81}{18} \\ &= 11{,}25 \end{aligned}$$

On a deux modalités ici, donc $k = 2$.

$$\begin{aligned} ddl &= k - 1 \\ &= 1 \end{aligned}$$

On compare alors K à $c_{0{,}05,\ 1ddl}$ qui vaut 3,84.

$$K > 3{,}84$$

On rejette H_0 et on conclut que la distribution du temps de réaction dans la population TC et dans la population DC est différente, au risque 5%.

Pour trouver ds, on encadre d'abord K puis on regarde à quel risque α cela correspond.

A partir de la table 3, on trouve :

$$\begin{aligned} & K > c_{0{,}1\%,\,1ddl} \\ \iff & ds < 0{,}1\% \end{aligned}$$

✘ **Item D** → *Si le degré de signification aurait été plus grand, cela signifie que la différence entre ...*

Il faut comprendre que le degré de signification ne dépend pas **uniquement** de la différence observée. Il dépend aussi et surtout du nombre de participants n.

Retenez que ds reflète surtout la conviction que l'on a du résultat présenté et non pas la différence constatée.

✘ **Item E** → *Si la loi du temps de réaction n'était pas une loi normale, on aurait pu quand même ...*

Les tests présentés cette année sont des tests paramétriques, c'est à dire des tests où la variable aléatoire étudiée suit une loi normale. La loi du χ^2 est dérivée d'une loi normale donc le principe est le même.

Si le temps de réaction ne suivait pas une loi normale, on n'aurait pas pu faire un test du χ^2, on aurait pu faire uniquement des tests non paramétriques.

Réponse vraie : C

La concentration de lactate après l'exercice a été mesurée dans chacun des groupes TC et DC. L'écart-type de la concentration de lactate dans les groupes TC et DC s_{TC} et s_{DC} est de 0,5 $mmol/L$. On prend $H_1 : \mu_{TC} - \mu_{DC} = 2$.

Question 14 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La puissance de ce test est proche de 100%

B. La puissance de ce test est proche de 50%

C. La puissance de ce test est proche de 80%

D. Un plus grand nombre de participants aurait permis de mettre en évidence une différence plus petite

E. La puissance augmente lorsque le risque α augmente

Question 14

? **Items A, B et C**

Comment on calcule la puissance ?

On définit d'abord la nouvelle loi que suit la variable aléatoire du critère de test Z' sous H_1. Elle suit une loi normale d'espérance $\frac{\Delta}{s_D}$ et d'écart-type 1.

La différence Δ dont il est question est celle précisée par H_1. C'est pour cela qu'on peut calculer la puissance uniquement si H_1 est précisée.

s_D est ce qu'on peut appeler l'écart-type commun.

$$\begin{aligned} s_D^2 &= \frac{s_{TC}^2}{n_{TC}} + \frac{s_{DC}^2}{n_{DC}} \\ &= \frac{0{,}5^2}{30} + \frac{0{,}5^2}{30} \\ &= \frac{1}{60} \\ \Longrightarrow s_D &= \sqrt{s_D^2} \\ &= 0{,}13 \end{aligned}$$

Donc dans le cas de l'énoncé :

$$\begin{aligned} Z' \sim \mathcal{N}\left(\frac{\Delta}{s_D},1\right) &\iff Z' \sim \mathcal{N}\left(\frac{2}{0.13},1\right) \\ &\iff Z' \sim \mathcal{N}(15.38,1) \end{aligned}$$

La puissance est la probabilité de rejeter H_0, donc que $z > 1{,}96$ si H_0 est fausse, soit H_1 est vraie.

$$\mathbb{P}(Z' > 1{,}96 | Z' \sim \mathcal{N}(15.38,1))$$

On calcule maintenant cette probabilité en centrant et réduisant Z' :

$$\begin{aligned}\mathbb{P}(Z' > 1{,}96) &= \mathbb{P}\left(\frac{Z'-\mu}{\sigma} > \frac{1{,}96-\mu}{\sigma}\right)\\ &= \mathbb{P}\left(U > \frac{1{,}96-15{,}38}{1}\right)\\ &= \mathbb{P}(U > -13{,}42)\\ &= 1-\mathbb{P}(U < -13{,}42)\\ &= 1-\mathbb{P}(U > 13{,}42)\end{aligned}$$

avec U la variable aléatoire centrée et réduite, appelée aussi Z dans votre cours.

Dans la table 2b, $\mathbb{P}(U > 13{,}42)$ est très proche de 0. Cela signifie que :

$$\mathbb{P}(Z' > 1{,}96) \approx 100\%$$

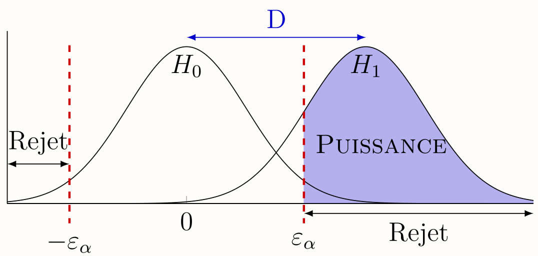
✔ Item D → *Un plus grand nombre de participants aurait permis de mettre en évidence une ...*

C'est bien cela. Plus le nombre de participants n est grand, plus on peut mettre en évidence une petite différence.

✔ Item E → *La puissance augmente lorsque le risque α augmente*

La puissance augmente lorsque la différence Δ augmente et lorsque le risque de première espèce, soit le risque α, augmente.

On peut voir cela sur le graphique suivant. La puissance est la partie en bleue, soit le rejet de H_0 sous la distribution de H_1. Lorsque l'erreur de type I est grande, cela signifie que ε_α se décale à gauche.



Alors la partie en bleu – *la puissance* – augmente.

Réponses vraies : A, D et E

Question 15 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les estimateurs ponctuels ne sont pas utilisés dans les tests statistiques, ce sont les estimateurs par intervalle qui sont utilisés.
B. Les tests statistiques utilisent les paramètres de la population et pas les estimateurs.
C. Les conditions de validité des tests permettent de vérifier que les hypothèses sont raisonnables pour réaliser un test dit paramétrique.
D. Si $ds < \alpha$, on ne rejette pas H_0
E. Si $ds > \alpha$, on rejette H_0

Question 15

? Items A et B

Ces deux items sont faux car les tests statistiques peuvent utiliser les estimateurs ponctuels, des paramètres observés car ceux de la population ne sont pas accessibles, et les estimateurs par intervalle c'est à dire les intervalles de confiance.

Pour plus de détails je vous invite à relire la correction de la question 1.

✔ Item C → *Les conditions de validité des tests permettent de vérifier que les hypothèses sont ...*

La condition pour réaliser des tests paramétriques est la normalité de la loi que suit la variable aléatoire en question. Pour vérifier cela, on pense aux conditions de validité.
S'ils ne sont pas validés, tout le raisonnement mathématique construit ne tient plus et le test devient alors bancal.
Donc oui, les conditions de validité sont importantes et ne sont pas là uniquement pour vous embêter 😉.

? Items D et E

Si le ds est inférieur à 5%, cela signifie que $z > 1{,}96$, donc on rejette H_0.
Ainsi, pour conclure à un test, il suffit en réalité de comparer ds et α.

Lorsque :

$$\begin{cases} ds < \alpha : \text{on rejette alors } H_0 \\ ds > \alpha : \text{on ne rejette pas } H_0 \end{cases}$$

Réponse vraie : C

15

Association pour l'Accès Santé – Université Paris Cité
Année Universitaire 2024-2025

Examen Blanc n°2 PASS

UE 9 : Mathématiques - Biostatistiques

Durée de l'épreuve : **1h**

A LIRE AVANT DE COMMENCER L'ÉPREUVE

Vérifiez que les informations saisies sur votre grille QCM sont correctes : nom, prénom et numéro étudiant.
Les correcteurs liquides ou en ruban de type Blanco, Tipp-Ex, et autres sont interdits car chaque question comporte une ligne de droit au remords.
Seule l'utilisation du stylo à bille noir est autorisée pour cocher les grilles.

INFORMATIONS RÉGLEMENTAIRES

- Les questions sans réponse seront considérées comme nulles.
- Une grille QCM est à remplir pour l'ensemble de l'épreuve.
- Veiller à remplir complètement toute la surface des cases choisies.
- Ne pas gratter, ne pas raturer, ne pas mettre de croix ni aucun autre signe.
- Toute fraude ou tentative de fraude fera l'objet de poursuites disciplinaires (Décret n°92-657 du 13 juillet 1992). Tout signe distinctif porté sur la grille QCM pouvant indiquer sa provenance constitue une fraude.
- Les calculatrices sont **autorisées**
- Aucun candidat n'est admis à quitter la salle d'examen avant la fin de l'épreuve.

RECOMMANDATIONS SPÉCIFIQUES À L'ÉPREUVE

INFORMATIONS SUR L'ÉPREUVE

Le sujet contient 5 pages numérotées de 1 à 5 et comporte 15 questions.
Merci de vérifier au début de l'épreuve que le sujet est complet.

Exercice 1 (questions 1 à 5)

On s'intéresse à l'infection par Mycobacterium tuberculosis, aussi appelé bacille de Koch, causant la tuberculose. Lorsqu'elle infecte les poumons, M. tuberculosis connaît une croissance rapide initiale. La réponse immunitaire adaptative limite la progression de l'infection, souvent en encapsulant les bactéries dans des granulomes, où elles survivent mais cessent de croître activement.

Pour décrire l'évolution de la population bactérienne dans l'organisme au cours du temps t, on propose le modèle suivant :

$$N'(t) = rN(t)\left(1 - \frac{N(t)}{K}\right)$$

Avec :

— $N(t)$: Taille de la population à l'instant t
— r : Taux intrinsèque de croissance de la population
— K : Capacité de charge (la taille maximale de la population que l'environnement peut soutenir)

r et K sont des constantes réelles strictement positives.

Avec N_0 la taille initiale de la population M. tuberculosis, la résolution de l'équation différentielle permet d'obtenir l'expression du nombre de bactéries au cours du temps :

$$N(t) = \frac{K}{1 + \left(\frac{K-N_0}{N_0}\right)e^{-rt}}$$

Question 1 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Le modèle proposé est un modèle exponentielle de Malthus
B. Il s'agit d'une équation différentielle non linéaire du premier ordre à coefficients constants et sans second membre
C. Si $N > K$, la population décroît
D. Au voisinage de l'origine, à l'ordre 2, on peut approximer e^{-rt} par $1 + rt + \frac{r^2t^2}{2}$
E. Au voisinage de l'origine, à l'ordre 3, on peut approximer e^{-rt} par $1 - rt + \frac{r^2t^2}{2} - \frac{r^3t^3}{3}$

Le traitement de la tuberculose se base principalement sur les antibiotiques. Après administration, ils peuvent être distribués dans deux compartiments : le compartiment central ou plasmatique (sang) et le compartiment périphérique (tissus).

Ce système d'équations décrit comment la concentration du médicament évolue dans chaque compartiment au cours du temps :

$$\begin{cases} C_1'(t) = -(\alpha+\gamma)C_1(t) + \beta C_2(t) \\ C_2'(t) = \alpha C_1(t) - \beta C_2(t) \end{cases}$$

Avec :

— $C_1(t)$: la concentration du médicament dans le compartiment central
— $C_2(t)$: la concentration du médicament dans le compartiment périphérique
— α, β, γ : des constantes réelles

Question 2 $\forall \alpha \in \mathbb{R}$, $\forall \beta \in \mathbb{R}$, $\forall \gamma \in \mathbb{R}$ et soient A et B deux constantes réelles. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Il s'agit d'un système d'équations différentielles d'ordre 2
B. On peut affirmer que les concentrations C_1 et C_2 sont dans une relation de proie/prédateur
C. Si C_1 augmente, alors C_2 diminue
D. L'équation caractéristique associée à $C_1(t)$ est $r^2 + (\alpha + \beta + \gamma)r + \beta\gamma = 0$
E. Pour un couple $(\alpha, \beta, \gamma) = (2{,}2{,}3)$: $C_1(t) = Ae^{-t} + Be^{-6t}$

Une nouvelle étude a permis de déterminer l'évolution au cours du temps t en heures h de la concentration plasmatique du médicament chez un individu, exprimée en mg/L :

$$C_1(t) = C_0\left(e^{-0{,}1t} - e^{-0{,}5t}\right)$$

Avec $C_0 > 0$ la dose initiale administrée. On injecte initialement une dose de 10 mg/L.

On s'intéresse à la biodisponibilité du médicament, afin de s'assurer que la dose injectée est appropriée. Pour cela, on étudie l'aire sous la courbe AUC :

$$AUC = \int_0^{+\infty} C_1(t) \cdot dt$$

On définit la zone thérapeutique pour une AUC comprise entre 100 UA et 200 UA. En dessous, il existe un risque d'inefficacité. Au-dessus, il existe un risque de toxicité.

Question 3 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. $\lim\limits_{t \to +\infty} C_1(t) = C_0$
B. Au voisinage de l'origine, à l'ordre 2, on peut approximer $C_1(t)$ par $C_0\,(0{,}4t - 0{,}12t^2)$
C. La dose administrée est adéquate
D. $AUC = 80\ mg \cdot L^{-1} \cdot h$
E. La valeur moyenne de $C_1(t)$ sur l'intervalle de temps $[0;4]h$ est de 20 mg/L

Enoncé commun aux question 4 et 5 :

En réalité, la concentration plasmatique d'un médicament dépend de caractéristiques physiopathologiques du patient. On utilise un nouveau modèle permettant de prendre en compte ces caractéristiques :

$$C_1(t) = C_0 \ln\left(\frac{80t - 7}{(t+1)^2}\right)$$

Avec :

- $C_1(t)$: la concentration plasmatique du médicament
- $C_0 > 0$: la concentration initiale du médicament
- $t \geq 0$: le temps en h

La valeur 80 dans ce modèle dépend de la morphologie et du sexe du patient selon cette formule :

$$f(x,y,z) = e^{-x+4\sqrt{y}+\frac{1}{z}}$$

Avec :

- x une constante qui dépend du sexe : $F = 0{,}5$; $H = 1$
- y la taille en m
- z le poids en kg

Question 4 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. $C_1(t)$ est définie sur $]0;+\infty[$
B. $C_1'(t) = \frac{-80t+94}{(80t-7)(t+1)}$
C. La concentration est maximale pour $t = 1{,}175\ h$
D. La concentration augmente puis diminue
E. Plus on attend, plus $C_1(t)$ tend vers C_0

Question 5 On étudie la concentration plasmatique d'un homme de $1{,}80 \pm 0{,}10\ m$ qui pèse $65{,}0 \pm 0{,}1\ kg$. Parmi les propositions suivantes, laquelle est exacte ?

A. $f \approx 80 \pm 10$
B. $f \approx 80 \pm 11$
C. $f \approx 80 \pm 12$
D. $f \approx 80 \pm 13$
E. $f \approx 80 \pm 14$

Exercice 2 (questions 6 à 10)

Dans le cadre d'un dépistage du cancer du sein, on dispose d'un test de dépistage T dont les deux issues sont : positif ou négatif. Deux échantillons distincts de 1000 femmes sont constitués (en « Population générale » et en « Population à risque »). On connaît le vrai statut de ces femmes face à la maladie (malade M ou saine $\overline{M}$). Les résultats du test de dépistage T sont les suivants dans les deux échantillons.

Au sein de la « Population générale » :

- Parmi les femmes saines, environ 5% sont positives au test.
- La probabilité d'avoir un test négatif est de 94,5%.
- 7 patientes sont malades.

Au sein de la « Population à risque » :

- Parmi les femmes malades, 25% des femmes sont négatives au test.
- La probabilité d'avoir un test positif est de 20%.
- 800 patientes sont saines.

Question 6 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Dans la « Population à risque », la prévalence est de 20%
B. La sensibilité dans la « Population générale » est de 80%
C. Dans la « Population à risque », la valeurs prédictives positives est de 75%
D. Dans le cas général, la spécificité augmente si la prévalence augmente
E. Si on tire au hasard une patiente dans la « population générale », il y a 0,5% de chance qu'elle soit malade et positive au test T

À la suite d'études cliniques, des statisticiens ont constaté que ce sont plus souvent les femmes exposées à des facteurs de risques qui se font dépister. Nous nous intéresserons ainsi à l'échantillon « Population à risque » de $n = 1000$ patientes et considérerons que chaque test T est indépendant et identiquement distribué.

Soit X la variable aléatoire qui prend la valeur 1 lorsque le test est positif, et la valeur 0 lorsque le test est négatif.

Question 7 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. X est une variable aléatoire qui suit une loi de Bernoulli de paramètres $n = 1000$ et $\pi = 20\%$
B. X est une variable aléatoire discrète
C. En considérant Y une variable aléatoire comptant le nombre de tests positifs, $\mathbb{E}(Y) = 200$ et $\mathbb{V}ar(Y) = 160$
D. Y est approximable par une loi gaussienne car $n \geq 30$
E. X est approximable par une loi gaussienne car $n\pi \geq 5$ et $n(1-\pi) \geq 5$

On effectue un nouvel essai clinique sur un échantillon cette fois-ci de $n = 20$ personnes, de même probabilité d'avoir un test positif que la variable X.

Question 8 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La probabilité que le nombre de tests positifs soit strictement supérieur à 15 dans la population à risque est nulle
B. La probabilité que le nombre de tests négatifs soit au moins égal à 12 dans la population à risque est d'environ 0,99
C. La probabilité que le nombre de tests négatifs soit égal à 17 est d'environ 20,5%
D. Si le nombre de personnes interrogées valait 30, la probabilité que le nombre de tests positifs soit égal à 1 serait inférieure à 1%
E. Une variable aléatoire est dite centrée réduite si $\mu = 0$ et $\sigma = 1$

Les dépistages sont efficaces pour éviter que les symptômes ne s'aggravent, mais nombreuses sont les patientes qui n'y ont pas recours. Ainsi, celles ayant des symptômes graves se font hospitaliser. On choisit un nouvel échantillon de 200 patientes. La durée d'hospitalisation H suit une loi exponentielle de paramètre λ. Une patiente se remet en moyenne au bout de 4 mois de chimiothérapie.

Question 9 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Il faudrait moins de 3 mois pour que la moitié des femmes soient déshospitalisées
B. La probabilité qu'une patiente soit hospitalisée plus d'un mois est de 0,78 au centième près
C. On peut approcher la variable M_{200} de la moyenne de 200 échantillons de la variable H par une loi normale de moyenne 20 et de variance 5
D. La probabilité d'être hospitalisée pendant plus d'un an et demi est nulle
E. Si, au bout de 2 mois, les cellules cancéreuses prolifèrent toujours, la probabilité qu'elle soit guérie au bout de 2 mois et demi est de 0,12 au centième près

On prend un échantillon de 200 patientes hospitalisées à la même date. On cherche à estimer la proportion au risque α de patientes toujours malades au bout de 4 et 12 mois d'hospitalisation.

Question 10 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les proportions théoriques de malades à 4 et 12 mois sont respectivement d'environ $\pi_4 = 37\%$ et $\pi_{12} = 5\%$
B. L'intervalle de pari au risque 15% de la proportion de malades à 12 mois est [0,03 ; 0,07] à 1% près
C. L'intervalle de pari au risque 1% de la proportion de malade au bout de 4 mois est plus étroit que l'intervalle de pari au risque 5% de cette même proportion
D. L'intervalle de pari au risque 5% de la proportion de malades à 12 mois est plus large que l'intervalle de pari au risque 1% de la proportion de malades à 4 mois
E. Le risque alpha est l'erreur de première espèce et il correspond à la probabilité de se tromper sous l'hypothèse nulle

Exercice 3 (questions 11 à 15)

Les tardigrades sont des Eumétazoaires capables de résister pendant des périodes prolongées à des conditions environnementales extrêmes. Un laboratoire en biologie animale souhaite étudier les mécanismes qui permettent à ces organismes de survivre face à ces milieux. Ils s'intéressent tout particulièrement aux tardigrades vivant dans l'océan et dans l'Himalaya. Pour cela, des biologistes étudient un échantillon A de tardigrades venant de l'océan et un échantillon B de tardigrades issus de l'Himalaya. Ils décident de comparer les tailles moyennes dans les différents populations.
Dans le groupe A, d'effectif $n_A = 120$, les adultes ont une taille de moyenne $m_A = 1{,}54$ mm et d'écart-type $s_A = 0{,}12$ mm. Dans le groupe B, d'effectif $n_B = 160$, les adultes ont une taille de moyenne $m_B = 1{,}56$ mm et d'écart-type $s_B = 0{,}04$ mm.

Question 11 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Sous l'hypothèse nulle H_0, $m_A = m_B$
B. On réalise un test de l'écart-réduit de statistique $|Z| > 2$
C. Les tailles moyennes des deux populations sont égales
D. Au risque $\alpha = 5\%$, on ne rejette pas H_0
E. Le degré de significativité du test est supérieur à 7%

Le laboratoire décide alors d'exposer deux échantillons de tardigrades vivant dans l'océan à différentes conditions environnementales. Le groupe C d'effectif $n_C = 120$ est exposé à une très basse température tandis que le groupe D d'effectif n_D est exposé à une pression très élevée. Au bout de deux heures, on note la proportion p de tardigrades ayant survécu. On note $p_C = 79\%$ pour le groupe C et $p_D = 86\%$ pour le groupe D et on décide de comparer les proportions théoriques entre les deux échantillons par un test de l'écart réduit.

Question 12 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. On peut réaliser ce test car $n_C p_C \geq 5$, $n_C(1 - p_C) \geq 5$, $n_D p_D \geq 5$ et $n_D(1 - p_D) \geq 5$
B. L'intervalle de pari à 95% de la proportion théorique de survivants dans le groupe C est $[0{,}71; 0{,}87] \pm 0{,}01$
C. L'intervalle de confiance à 96% de la proportion théorique de survivants dans le groupe D est $[0{,}80; 0{,}92] \pm 0{,}01$
D. En acceptant H_1 à tort, le laboratoire risque de commettre une erreur de type II ou risque de deuxième espèce
E. La statistique du test est $|Z| \simeq 1{,}543$

Deux autres échantillons de tardigrades (E et F) sont ensuite tirés au sort parmis les tardigrades vivant dans les hautes montagnes et on expose l'échantillon E à une dose plus faible de rayons UV que l'échantillon F. Au bout de trois heures, on observe le statut des différents tardigrades (vivants actifs, morts ou en dormance) et on construit le tableau de contingence suivant :

	Échantillon E	Échantillon F	Total
Tardigrades vivants actifs	80	70	150
Tardigrades morts	70	80	150
Tardigrades en dormance	50	150	200
Total	200	300	500

Question 13 Le laboratoire cherche alors à savoir s'il existe un lien entre la dose de rayonnements UV absorbée et le statut de survie des tardigrades. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. On peut réaliser un test du χ^2 d'indépendance à 2 degrés de liberté
B. On peut réaliser un test du χ^2 d'homogénéité à 6 degrés de liberté
C. La statistique de test est $K = 32{,}6$ à 0,1 près
D. Le degré de significativité du test est inférieur à 1%
E. Au risque $\alpha = 0{,}1\%$, il n'existe pas de lien entre la dose de rayonnements UV absorbée et le statut de survie

Satisfaits de leurs résultats, les biologistes décident de se pencher sur le mécanisme permettant aux tardigrades de résister à la dessication. Pour cela, on réalise un dosage de TDP (protéines intrinsèquement désordonnées spécifiques des tardigrades synthétisées en cas de dessication) dans chacun des tardigrades afin de savoir s'il existe un lien entre résistance à la dessication et taux de TDP.

On note T le taux de TDP dans les tardigrades. On pose H_0, sous laquelle aucun tardigrade ne vient d'un milieu sec, que l'on décide de rejeter lorsque $T > 180\ UI \cdot L^{-1}$. On considère que sous l'hypothèse alternative, $T \sim \mathcal{U}([150; 500]\ UI \cdot L^{-1})$ et que sous l'hypothèse nulle, $T \sim \mathcal{U}([25; 200]\ UI \cdot L^{-1})$.

Question 14 Les biologistes décident alors de sélectionner $n = 200$ tardigrades de façon aléatoire afin d'y doser la TDP et déterminer s'ils viennent d'un milieu sec ou humide. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La puissance du test serait plus grande si l'on décidait de rejeter H_0 lorsque $T > 160\ UI \cdot L^{-1}$
B. La puissance du test est de 91,4%
C. Le risque de première espèce de ce test est de 11,4%
D. On calcule la puissance grâce à la loi de T sous H_0
E. En affirmant qu'aucun tardigrade ne vient d'un milieu sec, les biologistes peuvent commettre une erreur de type II

4

Enfin, la capacité des tardigrades à survivre à des sécheresses extrêmes est également permise par la synthèse de la tréhalose, sucre remplaçant l'eau afin de protéger les cellules.

Le laboratoire regroupe tous les tardigrades qu'il possède les expose à un environnement très sec. Le laboratoire décide de classer les familles de tardigrades selon la proportion de tréhalose dans le corps en période de sécheresse, puis il note les effectifs de chaque catégorie dans ses échantillons afin de les comparer à la population générale. Il dresse le tableau suivant :

	Classe L	Classe M	Classe N	Classe O
Effectifs observés en laboratoire	$O_{1,1} = 140$	$O_{2,1} = 320$	$O_{3,1} = 280$	$O_{4,1} = 260$
Effectifs calculés (théoriques)	$C_{1,2} = 160$	$C_{2,2} = 300$	$C_{3,2} = 260$	$C_{4,2} = 280$

Question 15 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les conditions de validité du théorème central limite sont vérifiées car $\forall\ i = 1,...,k\ C_i \geq 5$
B. Au risque $\alpha = 2\%$, on ne rejette pas l'hypothèse nulle
C. On réalise un test du χ^2 d'adéquation à un modèle théorique
D. On réalise un test z de comparaison de proportions
E. L'hypothèse alternative est $\pi_O \neq \pi_C$ pour au moins une classe

Association pour l'Accès Santé – Université Paris Cité
Année Universitaire 2024-2025

CORRECTION

Examen Blanc n°2 PASS

UE 9 : Mathématiques - Biostatistiques

Durée de l'épreuve : **1h**

A LIRE AVANT DE COMMENCER L'ÉPREUVE

Vérifiez que les informations saisies sur votre grille QCM sont correctes : nom, prénom et numéro étudiant.
Les correcteurs liquides ou en ruban de type Blanco, Tipp-Ex, et autres sont interdits car chaque question comporte une ligne de droit au remords.
Seule l'utilisation du stylo à bille noir est autorisée pour cocher les grilles.

INFORMATIONS RÉGLEMENTAIRES

- Les questions sans réponse seront considérées comme nulles.
- Une grille QCM est à remplir pour l'ensemble de l'épreuve.
- Veiller à remplir complètement toute la surface des cases choisies.
- Ne pas gratter, ne pas raturer, ne pas mettre de croix ni aucun autre signe.
- Toute fraude ou tentative de fraude fera l'objet de poursuites disciplinaires (Décret n°92-657 du 13 juillet 1992). Tout signe distinctif porté sur la grille QCM pouvant indiquer sa provenance constitue une fraude.
- Les calculatrices sont **autorisées**
- Aucun candidat n'est admis à quitter la salle d'examen avant la fin de l'épreuve.

RECOMMANDATIONS SPÉCIFIQUES À L'ÉPREUVE

INFORMATIONS SUR L'ÉPREUVE

Le sujet contient 20 pages numérotées de 1 à 20 et comporte 15 questions.
Merci de vérifier au début de l'épreuve que le sujet est complet.

Université Paris Cité
A2SUP - Tutorat

Tuto n° : 2 UE (spé) : 9

Nom :
Prénom :
Numéro A2SUP :

IDENTIFICATION

Numéro A2SUP: Diz. Mil. / Mil. / Cent. / Diz. / Unit. (0 1 2 3 4 5 6 7 8 9)

Contrôle: M Unit. / M Diz. / J Unit. / J Diz. (0 1 2 3 4 5 6 7 8 9)
Date de naissance (JJ/MM)

	A	B	C	D	E
1		■	■		
2				■	■
3		■		■	
4			■	■	
5			■		
6	■		■		■
7		■	■		
8		■	■	■	■
9	■	■			■
10	■	■			■
11				■	■
12			■		■
13	■		■	■	
14	■	■	■		
15	■	■	■		■
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					

Exercice 1 (questions 1 à 5)

On s'intéresse à l'infection par Mycobacterium tuberculosis, aussi appelé bacille de Koch, causant la tuberculose. Lorsqu'elle infecte les poumons, M. tuberculosis connaît une croissance rapide initiale. La réponse immunitaire adaptative limite la progression de l'infection, souvent en encapsulant les bactéries dans des granulomes, où elles survivent mais cessent de croître activement.

Pour décrire l'évolution de la population bactérienne dans l'organisme au cours du temps t, on propose le modèle suivant :

$$N'(t) = rN(t)\left(1 - \frac{N(t)}{K}\right)$$

Avec :

— $N(t)$: Taille de la population à l'instant t

— r : Taux intrinsèque de croissance de la population

— K : Capacité de charge (la taille maximale de la population que l'environnement peut soutenir)

r et K sont des constantes réelles strictement positives.

Avec N_0 la taille initiale de la population M. tuberculosis, la résolution de l'équation différentielle permet d'obtenir l'expression du nombre de bactéries au cours du temps :

$$N(t) = \frac{K}{1 + \left(\frac{K-N_0}{N_0}\right)e^{-rt}}$$

Question 1 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Le modèle proposé est un modèle exponentielle de Malthus

B. Il s'agit d'une équation différentielle non linéaire du premier ordre à coefficients constants et sans second membre

C. Si $N > K$, la population décroît

D. Au voisinage de l'origine, à l'ordre 2, on peut approximer e^{-rt} par $1 + rt + \frac{r^2t^2}{2}$

E. Au voisinage de l'origine, à l'ordre 3, on peut approximer e^{-rt} par $1 - rt + \frac{r^2t^2}{2} - \frac{r^3t^3}{3}$

Question 1

✘ **Item A** → *Le modèle proposé est un modèle exponentielle de Malthus*

On retrouve ici le modèle logistique de Verhulst. En effet dans votre cours/formulaire l'équation différentielle qui décrit ce modèle est de la forme :

$$N' = \lambda(a - N)N \text{ ou } dN = \lambda(a - N)N dt$$

L'équation différentielle de l'énoncé est bien sous cette forme (retirons les t pour y voir plus clair) :

$$N' = rN\left(1 - \frac{N}{K}\right)$$
$$\iff N' = \frac{r}{K}N(K - N)$$
$$\iff N' = \frac{r}{K}(K - N)N$$

Avec $\lambda = \frac{r}{K}$ et $a = K$ dans notre situation.

✔ **Item B** → *Il s'agit d'une équation différentielle non linéaire du premier ordre à coefficients ...*

Revenons sur quelques définitions :

- Une équation différentielle est dite *non linéaire* si on applique une quelconque fonction à la fonction inconnue (fonctions trigonométriques, puissance, ln, exp,...) et *linéaire* si aucune fonction n'y est appliquée
- Une équation différentielle est dite *de premier ordre* lorsqu'elle met en jeu la fonction inconnue et sa dérivée première. Elle est dite *de second ordre* lorsqu'elle contient y a la fonction inconnue et sa dérivée seconde.
- *Avec coefficients constants* ou *variables* désignent si les coefficients devant la fonction inconnue et ses dérivées sont constants ou dépendent d'un paramètre (temps/espace/...)
- *Avec ou sans second membre* désignent si il existe un terme libre, c'est-à-dire non lié à la fonction inconnue et ses dérivées

Dans notre cas :

$$N' = rN\left(1 - \frac{N}{K}\right) \iff N' = -\frac{r}{K}N^2 + rN$$

- La fonction carré est appliquée à notre fonction inconnue (N^2) ⇒ équation différentielle non linéaire
- On observe la fonction inconnue N et sa dérivée première N' ⇒ premier ordre
- Les coefficients r et K sont des constantes réelles positives ⇒ coefficients constants
- On observe aucun terme libre ⇒ sans second membre

✔ **Item C** → *Si $N > K$, la population décroît*

Pour répondre à cette question, jetons un coup d'œil à notre équation différentielle :

$$N'(t) = rN(t)\left(1 - \frac{N(t)}{K}\right)$$

$N'(t)$ désigne l'évolution de la population au cours du temps, c'est-à-dire sa croissance !

Si $N'(t) > 0$, la population croît

Si $N'(t) < 0$, la population décroît

Le signe de $N'(t)$ dépend de $1 - \frac{N}{K}$ car $r > 0$ et $N \geq 0$. Ainsi,

$$N > K \Rightarrow \frac{N}{K} > 1 \Rightarrow 1 - \frac{N}{K} < 0 \Rightarrow N'(t) < 0$$

Donc la population décroît selon cette condition !

On voit ici bien la définition de la capacité de charge K : si la population dépasse la taille maximale de la population que l'environnement peut soutenir, alors elle décroît.

? Items D et E

Ah voilà les fameux développements limités !

La manière la plus rapide de procéder, consiste à reprendre les développements limités usuels au voisinage de 0 (disponible dans votre formulaire également), en l'occurrence celui de la fonction exponentielle :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

Ici, on va considérer que $x = -rt$ pour pouvoir réutiliser cette formule dans notre situation. On a alors :

$$\begin{aligned} e^{-rt} &= 1 + (-rt) + \frac{(-rt)^2}{2!} + \frac{(-rt)^3}{3!} \\ &= 1 - rt + \frac{r^2t^2}{2} - \frac{r^3t^3}{6} \end{aligned}$$

On choisit ensuite l'ordre qui nous intéresse comme précédemment.

Réponses vraies : B et C

Le traitement de la tuberculose se base principalement sur les antibiotiques. Après administration, ils peuvent être distribués dans deux compartiments : le compartiment central ou plasmatique (sang) et le compartiment périphérique (tissus).

Ce système d'équations décrit comment la concentration du médicament évolue dans chaque compartiment au cours du temps :

$$\begin{cases} C_1'(t) = -(\alpha+\gamma)C_1(t) + \beta C_2(t) \\ C_2'(t) = \alpha C_1(t) - \beta C_2(t) \end{cases}$$

Avec :

— $C_1(t)$: la concentration du médicament dans le compartiment central

— $C_2(t)$: la concentration du médicament dans le compartiment périphérique

— α, β, γ : des constantes réelles

Question 2 $\forall \alpha \in \mathbb{R}$, $\forall \beta \in \mathbb{R}$, $\forall \gamma \in \mathbb{R}$ et soient A et B deux constantes réelles. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Il s'agit d'un système d'équations différentielles d'ordre 2

B. On peut affirmer que les concentrations C_1 et C_2 sont dans une relation de proie/prédateur

C. Si C_1 augmente, alors C_2 diminue

D. L'équation caractéristique associée à $C_1(t)$ est $r^2 + (\alpha+\beta+\gamma)r + \beta\gamma = 0$

E. Pour un couple $(\alpha, \beta, \gamma) = (2,2,3)$: $C_1(t) = Ae^{-t} + Be^{-6t}$

Question 2

✘ Item A → *Il s'agit d'un système d'équations différentielles d'ordre 2*

On commence doucement mais sûrement. Qu'avons-nous là ? Un système d'équations différentielles, ici du **premier ordre**.

En effet, on a les dérivées premières des concentrations, donc ordre 1. Si on avait les dérivées secondes, alors ce serait d'ordre 2. Dérivées troisièmes, ordre 3, ...

✘ Item B → *On peut affirmer que les concentrations C_1 et C_2 sont dans une relation de ...*

Pour savoir dans quelle relation se trouvent C_1 et C_2, reprenons le cours :

$$\begin{cases} x'(t) = a_{11}x(t) + a_{12}y(t) \\ y'(t) = a_{21}x(t) + a_{22}y(t) \end{cases}$$

- a_{12} et a_{21} positifs $\Longrightarrow$ relation de symbiose
- a_{12} et a_{21} négatifs $\Longrightarrow$ relation de compétition
- a_{12} et a_{21} de signes opposés $\Longrightarrow$ relation de proie/prédateur

Appliqué à notre cas, on a $a_{12} = \beta$ et $a_{21} = \alpha$.

Or l'énoncé ne nous dit uniquement que α, β et γ sont des constantes réelles. On a aucune information sur leurs signes. Donc on ne peut cocher l'item !

✘ Item C → *Si C_1 augmente, alors C_2 diminue*

Le système permet d'étudier la dynamique d'évolution des 2 concentrations, c'est-à-dire comment évolue une concentration si l'autre évolue également.

Cependant, on a aucune information sur les valeurs et les signes de α, β et γ. Donc on ne peut pas déterminer ici comment évolue une concentration par rapport à l'autre.

✔ Item D → *L'équation caractéristique associée à $C_1(t)$ est $r^2 + (\alpha+\beta+\gamma)r + \beta\gamma = 0$*

Pour simplifier l'écriture du système, on l'écrira sans les t :

$$\begin{cases} C_1' = -(\alpha+\gamma)C_1 + \beta C_2 \\ C_2' = \alpha C_1 - \beta C_2 \end{cases}$$

Il suffit de suivre une méthode pour ne pas se perdre :

✍ Déterminer l'équation caractéristique

1. Sélectionner l'équation qui nous intéresse et la dériver
2. Remplacer la deuxième inconnue dérivée grâce à l'autre équation
3. Remplacer la deuxième inconnue non dérivée grâce à l'équation initialement choisie
4. Remplacer les dérivées par les puissances de r correspondantes

Appliquons cela pas à pas :

3

1. On s'intéresse à C_1. Dérivons :

$$C_1'' = -(\alpha+\gamma)C_1' + \beta C_2'$$

2. Remplaçons C_2' grâce à l'autre équation :

$$\begin{aligned} C_1'' &= -(\alpha+\gamma)C_1' + \beta(\alpha C_1 - \beta C_2) \\ &= -(\alpha+\gamma)C_1' + \alpha\beta C_1 - \beta^2 C_2 \end{aligned}$$

3. Remplaçons C_2 grâce à la première équation :

$$\begin{aligned} C_1' = -(\alpha+\gamma)C_1 + \beta C_2 &\iff \beta C_2 = C_1' + (\alpha+\gamma)C_1 \\ &\iff C_2 = \frac{C_1' + (\alpha+\gamma)C_1}{\beta} \end{aligned}$$

On a l'expression de C_2, injections-la :

$$\begin{aligned} C_1'' &= -(\alpha+\gamma)C_1' + \alpha\beta C_1 - \beta^2\left(\frac{C_1' + (\alpha+\gamma)C_1}{\beta}\right) \\ &= -(\alpha+\gamma)C_1' + \alpha\beta C_1 - \beta C_1' - \beta(\alpha+\gamma)C_1 \\ &= -(\alpha+\beta+\gamma)C_1' + (\alpha-\alpha-\gamma)\beta C_1 \\ &= -(\alpha+\beta+\gamma)C_1' - \beta\gamma C_1 \end{aligned}$$

On réarrange un peu notre équation :

$$C_1'' + (\alpha+\beta+\gamma)C_1' + \beta\gamma C_1 = 0$$

4. Remplaçons C_1'' par r^2, C_1' par $r^1 = r$ et C_1 par $r^0 = 1$:

$$r^2 + (\alpha+\beta+\gamma)r + \beta\gamma = 0$$

And voilà !

✔ **Item E** → *Pour un couple $(\alpha,\beta,\gamma) = (2{,}2{,}3)$: $C_1(t) = Ae^{-t} + Be^{-6t}$*

Pour répondre à ces questions, il faut s'aider du formulaire :

- $D > 0 \Rightarrow y_0 = Ae^{r_1x} + Be^{r_2x}$
- $D = 0 \Rightarrow y_0 = e^{rx}(A + Bx)$
- $D < 0 \Rightarrow y_0 = e^{\mathcal{A}x}\left(Acos(\mathcal{B}x) + Bsin(\mathcal{B}x)\right)$

Pour savoir le signe du discriminant D, repartons de l'équation caractéristique associée à $C_1(t)$ trouvée à l'tem A :

$$r^2 + (\alpha+\beta+\gamma)r + \beta\gamma = 0$$

Réécrivons cette formule sous une autre forme pour faciliter les choses :

$$ar^2 + br + c = 0$$

Avec $a = 1$, $b = \alpha+\beta+\gamma$ et $c = \beta\gamma$
La formule du discriminant est : $D = b^2 - 4ac$

Comme on a $\alpha = 2$, $\beta = 2$ et $\gamma = 3$, on a alors :

$$\begin{aligned} a &= 1 \\ b &= 2+2+3 = 7 \\ c &= 2 \times 3 = 6 \end{aligned}$$

En ce qui concerne le discriminant :

$$\begin{aligned} D &= 7^2 - 4 \times 1 \times 6 \\ &= 49 - 24 = 25 \end{aligned}$$

On a $D > 0$!

Calculons les solutions de l'équation caractéristique :

$$\begin{aligned} r &= \frac{-b \pm \sqrt{D}}{2a} \\ &= \frac{-7 \pm \sqrt{25}}{2 \times 1} \\ &= \frac{-7 \pm 5}{2} \end{aligned}$$

Ainsi les solutions sont :

$$r_1 = \frac{-7+5}{2} = -1$$

$$r_2 = \frac{-7-5}{2} = -6$$

Implémentons cela à la solution grâce au formulaire :

$$C_1(t) = Ae^{-1t} + Be^{-6t} = Ae^{-t} + Be^{-6t}$$

Remarque : on aurait pu décider d'interchanger $r_1 = -6$ et $r_2 = -1$. Cela ne change rien à la réponse finale car A et B sont des constantes quelconques ! On peut également interchanger A et B pour retomber sur l'item.

Pour aller plus vite lors de l'épreuve on peut vérifier que -6 et -1 sont bien des solutions de l'équation caractéristique de C_1. C'est moins beaucoup moins long que de les trouver en utilisant le calcul du discriminant.

Réponses vraies : D et E

Une nouvelle étude a permis de déterminer l'évolution au cours du temps t en heures h de la concentration plasmatique du médicament chez un individu, exprimée en mg/L :

$$C_1(t) = C_0\left(e^{-0{,}1t} - e^{-0{,}5t}\right)$$

Avec $C_0 > 0$ la dose initiale administrée. On injecte initialement une dose de 10 mg/L.

On s'intéresse à la biodisponibilité du médicament, afin de s'assurer que la dose injectée est appropriée. Pour cela, on étudie l'aire sous la courbe AUC :

$$AUC = \int_0^{+\infty} C_1(t) \cdot dt$$

On définit la zone thérapeutique pour une AUC comprise entre 100 UA et 200 UA. En dessous, il existe un risque d'inefficacité. Au-dessus, il existe un risque de toxicité.

Question 3 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. $\lim_{t \to +\infty} C_1(t) = C_0$

B. Au voisinage de l'origine, à l'ordre 2, on peut approximer $C_1(t)$ par $C_0\left(0{,}4t - 0{,}12t^2\right)$

C. La dose administrée est adéquate

D. $AUC = 80\ mg \cdot L^{-1} \cdot h$

E. La valeur moyenne de $C_1(t)$ sur l'intervalle de temps $[0;4]h$ est de 20 mg/L

Question 3

✘ **Item A** → $\lim_{t \to +\infty} C_1(t) = C_0$

C'est un simple calcul de limite, allons-y !

$$\begin{aligned}\lim_{t \to +\infty} C_1(t) &= \lim_{t \to +\infty} C_0\left(e^{-0{,}1t} - e^{-0{,}5t}\right)\\ &= C_0\left(\lim_{t \to +\infty} e^{-0{,}1t} - \lim_{t \to +\infty} e^{-0{,}5t}\right)\\ &= C_0(0-0) = 0\end{aligned}$$

✔ **Item B** → *Au voisinage de l'origine, à l'ordre 2, on peut approximer $C_1(t)$ par $C_0\left(0{,}4t - 0{,}12t^2\right)$*

Comme pour la question 1, on a deux méthodes : soit les formules de Taylor, soit les DLs usuels. Ici on va prendre la deuxième méthode beaucoup plus rapide, qu'on applique à chaque fonction exponentielle de la formule :

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

- Pour $e^{-0{,}1t}$, on prend $x = -0{,}1t$. À l'ordre 2, on a :

$$\begin{aligned}e^{-0{,}1t} &= 1 + (-0{,}1t) + \frac{(-0{,}1t)^2}{2!}\\ &= 1 - 0{,}1t + \frac{0{,}01t^2}{2}\end{aligned}$$

- Pour $e^{-0{,}5t}$, on prend $x = -0{,}5t$. À l'ordre 2, on a :

$$\begin{aligned}e^{-0{,}5t} &= 1 + (-0{,}5t) + \frac{(-0{,}5t)^2}{2!}\\ &= 1 - 0{,}5t + \frac{0{,}25t^2}{2}\end{aligned}$$

On injecte, on a alors :

$$\begin{aligned}C_1(t) &\approx C_0\left(1 - 0{,}1t + \frac{0{,}01t^2}{2} - \left(1 - 0{,}5t + \frac{0{,}25t^2}{2}\right)\right)\\ &\approx C_0\left(1 - 0{,}1t + \frac{0{,}01t^2}{2} - 1 + 0{,}5t - \frac{0{,}25t^2}{2}\right)\\ &\approx C_0\left(0{,}4t - 0{,}12t^2\right)\end{aligned}$$

? **Items C et D**

Déterminons l'AUC :

$$\begin{aligned}AUC &= \int_0^{+\infty} 10\left(e^{-0{,}1t} - e^{-0{,}5t}\right)\cdot dt\\ &= 10\int_0^{+\infty}\left(e^{-0{,}1t} - e^{-0{,}5t}\right)\cdot dt\\ &= 10\left(\int_0^{+\infty} e^{-0{,}1t}\cdot dt - \int_0^{+\infty} e^{-0{,}5t}\cdot dt\right)\end{aligned}$$

Calculons la première intégrale :

$$\begin{aligned}\int_0^{+\infty} e^{-0{,}1t}\cdot dt &= \left[-\frac{1}{0{,}1}\times e^{-0{,}1t}\right]_0^{+\infty} = \left[-10\times e^{-0{,}1t}\right]_0^{+\infty}\\ &= -10\left(\lim_{t\to+\infty} e^{-0{,}1t} - e^{-0{,}1\times 0}\right) = -10\times(0-1) = 10\end{aligned}$$

Au tour de la deuxième :

$$\begin{aligned}\int_0^{+\infty} e^{-0{,}5t}\cdot dt &= \left[-\frac{1}{0{,}5}\times e^{-0{,}5t}\right]_0^{+\infty} = \left[-2\times e^{-0{,}5t}\right]_0^{+\infty}\\ &= -2\left(\lim_{t\to+\infty} e^{-0{,}5t} - e^{-0{,}5\times 0}\right) = -2\times(0-1) = 2\end{aligned}$$

Ainsi :

$$AUC = 10\times(10-2) = 80\ mg\cdot L^{-1}\cdot h$$

Par rapport à l'unité, on multiplie une concentration en mg/L par dt qui s'exprime en h. Donc on retrouve bien des $mg\cdot L^{-1}\cdot h$

Ainsi $AUC \notin [100; 200]$, plus précisément $AUC < 100$. Donc la dose administrée présente un risque d'inefficacité.

✘ **Item E** → *La valeur moyenne de $C_1(t)$ sur l'intervalle de temps $[0;4]h$ est de 20 mg/L*

On appelle valeur moyenne m de la fonction sur l'intervalle $[a,b]$ une valeur de $f(x)$ telle que :

$$m = \frac{1}{b-a}\int_a^b f(x)\cdot dx$$

Lorsqu'on applique à notre cas, on a :

$$m = \frac{1}{4-0}\int_0^4 C_1(t)\cdot dt$$

5

Attention aux bornes ! Pour répondre correctement à la question, il ne fallait pas simplement diviser l'AUC précédemment trouvée par 4, ce qui donnerait 20 mg/L. Il faut refaire le calcul de l'intégrale avec les nouvelles bornes 0 et 4 :

$$\begin{aligned} m &= \frac{1}{4-0} \times \int_0^4 10\left(e^{-0,1t} - e^{-0,5t}\right) \cdot dt \\ &= \frac{1}{4} \times 10\left(\left[-10 \times e^{-0,1t}\right]_0^4 - \left[-2 \times e^{-0,5t}\right]_0^4\right) \\ &= 2,5 \times \left(-10 \times \left(e^{-0,4} - e^{-0}\right) + 2 \times \left(e^{-2} - e^{-0}\right)\right) \\ &= -25 \times \left(e^{-0.4} - 1\right) + 5 \times \left(e^{-2} - 1\right) \end{aligned}$$

À vos calculatrices ! On trouve alors :

$$m \approx 4\ mg/L$$

Réponses vraies : B et D

Enoncé commun aux question 4 et 5 :

En réalité, la concentration plasmatique d'un médicament dépend de caractéristiques physiopathologiques du patient. On utilise un nouveau modèle permettant de prendre en compte ces caractéristiques :

$$C_1(t) = C_0 \ln\left(\frac{80t-7}{(t+1)^2}\right)$$

Avec :

- $C_1(t)$: la concentration plasmatique du médicament
- $C_0 > 0$: la concentration initiale du médicament
- $t \geq 0$: le temps en h

La valeur 80 dans ce modèle dépend de la morphologie et du sexe du patient selon cette formule :

$$f(x,y,z) = e^{-x+4\sqrt{y}+\frac{1}{z}}$$

Avec :

- x une constante qui dépend du sexe : $F = 0,5$; $H = 1$
- y la taille en m
- z le poids en kg

Question 4 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. $C_1(t)$ est définie sur $]0;+\infty[$
B. $C_1'(t) = \frac{-80t+94}{(80t-7)(t+1)}$
C. La concentration est maximale pour $t = 1,175\ h$
D. La concentration augmente puis diminue
E. Plus on attend, plus $C_1(t)$ tend vers C_0

Question 4

✘ **Item A** → *$C_1(t)$ est définie sur $]0;+\infty[$*

Pour déterminer le domaine de définition, la première chose à faire c'est d'analyser les termes qui peuvent poser problème dans notre fonction :

$$C_1(t) = C_0 \ln\left(\frac{80t-7}{(t+1)^2}\right)$$

Tout d'abord, nous avons une fraction. Qu'est ce qui peut poser problème ? La division par 0 !
Or $t > 0$ donc, $\forall t, (t+1)^2 > 0$

Ensuite, nous avons la fonction ln. Son argument doit être strictement positif, c'est-à-dire :

$$\frac{80t-7}{(t+1)^2} > 0$$

On avait vu que $(t+1)^2 > 0$. Intéressons-nous au numérateur :

$$80t - 7 > 0 \iff 80t > 7 \iff t > \frac{7}{80} = 0,0875$$

Le domaine de définition de notre fonction est donc,

$$t \in]0,0875;+\infty[$$

✘ **Item B** → $C_1'(t) = \frac{-80t+94}{(80t-7)(t+1)}$

Hop un ptit calcul de dérivée, allons-y :

Il faut absolument connaître la dérivée de $\ln(w)$ avec w une fonction de t :

$$\ln'(w) = \frac{w'}{w}$$

Posons $w = \frac{80t-7}{(t+1)^2}$. On cherche à déterminer w'. Nous avons ici une fraction. Il faut maintenant se rappeler de la dérivée d'une fraction :

$$\left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$$

Posons :

✦ $u = 80t - 7 \Rightarrow u' = 80$
✦ $v = (t+1)^2 \Rightarrow v' = 2(t+1)$

Ainsi, on a :

$$\begin{aligned} w' &= \frac{80 \times (t+1)^2 - 2 \times (t+1) \times (80t-7)}{(t+1)^4} \\ &= \frac{80 \times (t+1) - 2 \times (80t-7)}{(t+1)^3} \\ &= \frac{80t + 80 - 160t + 14}{(t+1)^3} \\ &= \frac{-80t+94}{(t+1)^3} \end{aligned}$$

Et enfin :

$$\begin{aligned}\ln'(w) &= \frac{w'}{w} = w' \times \frac{1}{w} \\ &= \frac{-80t+94}{(t+1)^3} \times \frac{(t+1)^2}{80t-7} \\ &= \frac{-80t+94}{(80t-7)(t+1)}\end{aligned}$$

Mais ce qu'on a fait, c'est la dérivée du terme en ln. N'oublions pas la constante C_0 pour obtenir $C_1'(t)$!!!

$$C_1'(t) = C_0 \times \frac{-80t+94}{(80t-7)(t+1)}$$

? Items C et D

Des petits souvenirs du lycée ? Et oui place au tableau de variations !

Pour connaître les variations de la concentration, il faut étudier le signe de sa dérivée.

Décortiquons chaque terme de $C_1'(t)$:

- ✦ C_0 est une constante strictement positive. Donc $C_0 > 0\ \forall t$
- ✦ $-80t+94$
  Il s'agit d'une fonction linéaire. Elle coupe l'axe des abscisses en $t = \frac{94}{80} = 1{,}175$. Ainsi :

  $$-80t+94 < 0 \text{ pour } t > 1{,}175$$

  $$-80t+94 > 0 \text{ pour } t < 1{,}175$$
- ✦ $80t-7$
  Il s'agit toujours d'une fonction linéaire. Elle coupe l'axe des abscisses en $t = 0{,}0875$ (vu à l'item A). Ainsi :

  $$80t-7 < 0 \text{ pour } t < 0{,}0875$$

  $$80t-7 > 0 \text{ pour } t > 0{,}0875$$
- ✦ $t+1$
  Il s'agit encore et toujours d'une fonction linéaire. Elle coupe l'axe des abscisses en $t = -1$. Ainsi :

  $$t+1 < 0 \text{ pour } t < -1$$

  $$t+1 > 0 \text{ pour } t > -1$$

Et hop on fout tout ça dans notre tableau de signe et de variations, en prenant les bonnes bornes du domaine de définition :

t	0.0875		1.175		$+\infty$
$-80t+94$		+	0	−	
$80t-7$	0	+		+	
$t+1$		+		+	
signe de $C_1'(t)$	‖	+	0	−	
variation de $C_1(t)$		↗	max	↘	

Grâce à l'étude de fonction, nous pouvons voir que la concentration augmente puis diminue, en passant par un maximum pour $t = 1{,}175\ h$.

✘ Item E → *Plus on attend, plus $C_1(t)$ tend vers C_0*

Pour finir avec cette fonction, une ptite étude de limite. Plus on attend signifie que $t \longrightarrow +\infty$. On a alors :

$$\lim_{t\to+\infty} C_1(t) = \lim_{t\to+\infty} C_0 \ln\left(\frac{80t-7}{(t+1)^2}\right)$$

Pour cela étudions la limite de ce qui se trouve dans le ln :

$$\lim_{t\to+\infty} \frac{80t-7}{(t+1)^2} = \lim_{t\to+\infty} \frac{80t-7}{t^2+2t+1} = \lim_{t\to+\infty} \frac{80-\frac{7}{t}}{t+2+\frac{1}{t}} = 0^+$$

On peut alors remplacer dans notre limite initiale :

$$\lim_{X\to0^+} C_0 \ln(X) = -\infty$$

On peut d'ailleurs voir que ce modèle se comporte bizarrement à l'infini : on ne peut pas avoir de concentration négative. Il faudrait alors choisir un domaine d'étude.

Réponses vraies : C et D

Question 5 On étudie la concentration plasmatique d'un homme de $1{,}80 \pm 0{,}10\ m$ qui pèse $65{,}0 \pm 0{,}1\ kg$. Parmi les propositions suivantes, laquelle est exacte ?

A. $f \approx 80 \pm 10$
B. $f \approx 80 \pm 11$
C. $f \approx 80 \pm 12$
D. $f \approx 80 \pm 13$
E. $f \approx 80 \pm 14$

7

Question 5

Items A, B, C, D et E

Tout d'abord, il faut déterminer l'incertitude de f, qui dépend de x,y,z. Ainsi on utilise la formule de la propagation d'incertitude :

$$\Delta f \leq \left|\frac{\partial f}{\partial x}\right| \Delta x + \left|\frac{\partial f}{\partial y}\right| \Delta y + \left|\frac{\partial f}{\partial z}\right| \Delta z$$

Identifions qui est qui :

- $x = 1$ car c'est un homme et $\Delta x = 0$ car elle n'est pas donnée, x semble être théorique et ne pas avoir d'incertitude associée
- $y = 1{,}80$ et $\Delta y = 0{,}10$
- $z = 65{,}0$ et $\Delta z = 0{,}1$

Calculons les dérivées partielles :

- $\left|\frac{\partial f}{\partial x}\right|$: Nul besoin de la calculer, car $\Delta x = 0$. Elle s'annulera dans la formule
- $\left|\frac{\partial f}{\partial y}\right| = \left|\frac{\partial}{\partial y}\left(e^{-x+4\sqrt{y}+\frac{1}{z}}\right)\right| = \left|\frac{4}{2\sqrt{y}} \times e^{-x+4\sqrt{y}+\frac{1}{z}}\right| = \frac{2}{\sqrt{y}} \times e^{-x+4\sqrt{y}+\frac{1}{z}}$
- $\left|\frac{\partial f}{\partial z}\right| = \left|\frac{\partial}{\partial z}\left(e^{-x+4\sqrt{y}+\frac{1}{z}}\right)\right| = \left|-\frac{1}{z^2} \times e^{-x+4\sqrt{y}+\frac{1}{z}}\right| = \frac{1}{z^2} \times e^{-x+4\sqrt{y}+\frac{1}{z}}$

Remplaçons dans la formule :

$$\Delta f \leq \frac{2}{\sqrt{y}} \times e^{-x+4\sqrt{y}+\frac{1}{z}} \times \Delta y + \frac{1}{z^2} \times e^{-x+4\sqrt{y}+\frac{1}{z}} \times \Delta z$$

$$\Rightarrow \Delta f \leq e^{-x+4\sqrt{y}+\frac{1}{z}} \left(\frac{2\Delta y}{\sqrt{y}} + \frac{\Delta z}{z^2}\right)$$

$$\Rightarrow \Delta f \leq e^{-1+4\sqrt{1{,}80}+\frac{1}{65{,}0}} \left(\frac{2 \times 0{,}10}{\sqrt{1{,}80}} + \frac{0{,}1}{65{,}0^2}\right)$$

$$\Rightarrow \Delta f \leq 11{,}927$$

Ici les items imposent que l'incertitude soit sans chiffre après la virgule. Donc il faut arrondir l'incertitude de f à l'unité près. Mais comment ? Dans votre cours, on vous dit :

On arrondit l'incertitude vers le haut et pas au plus proche.

Bon dans notre cas on arrondit vers le haut et également au plus proche, mais la règle c'est toujours vers le haut $\Rightarrow \Delta f \approx 12$

De plus, $f = e^{-1+4\sqrt{1{,}80}+\frac{1}{65{,}0}} \approx 79{,}99 \approx 80$

Ici, aucune restriction sur l'approximation de f est imposée dans votre cours. On l'arrondit alors à l'unité le plus proche.

Finalement :

$$f \approx 80 \pm 12$$

Réponse vraie : C

Exercice 2 (questions 6 à 10)

Dans le cadre d'un dépistage du cancer du sein, on dispose d'un test de dépistage T dont les deux issues sont : positif ou négatif. Deux échantillons distincts de 1000 femmes sont constitués (en « Population générale » et en « Population à risque »). On connaît le vrai statut de ces femmes face à la maladie (malade M ou saine $\overline{M}$). Les résultats du test de dépistage T sont les suivants dans les deux échantillons.

Au sein de la « Population générale » :

- Parmi les femmes saines, environ 5% sont positives au test.
- La probabilité d'avoir un test négatif est de 94,5%.
- 7 patientes sont malades.

Au sein de la « Population à risque » :

- Parmi les femmes malades, 25% des femmes sont négatives au test.
- La probabilité d'avoir un test positif est de 20%.
- 800 patientes sont saines.

Question 6 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Dans la « Population à risque », la prévalence est de 20%

B. La sensibilité dans la « Population générale » est de 80%

C. Dans la « Population à risque », la valeurs prédictives positives est de 75%

D. Dans le cas général, la spécificité augmente si la prévalence augmente

E. Si on tire au hasard une patiente dans la « population générale », il y a 0,5% de chance qu'elle soit malade et positive au test T

Question 6

Au premier abord, ce long énoncé peut te faire peur, mais pas de panique ! Allons-y pas à pas.

Étape n°1 : retranscrire mathématiquement l'énoncé littéral et faire un schéma propre de la situation à laquelle on fait face. Les maths, c'est comme la guerre : il faut assembler les munitions nécessaires avant d'aller à l'assaut. On a deux échantillons distincts : « Population générale » et « Population à risque ».

On définit les évènements suivants :

- M : être malade
- $\overline{M}$: être saine / pas malade
- T^+ : le test T est positif
- T^- : le test T est négatif

À présent, on peut attaquer l'énoncé : il faut assembler toutes les informations dans l'échantillon de la **population générale**. En retranscrivant tiret par tiret l'énoncé, on a :

- $P(T^+|\overline{M}) = 5\% = 0{,}05$
- $P(T^-) = 94{,}5\% = 0{,}945$

- $P(M) = \frac{7}{1000}$

À première vue, ça peut piquer des yeux : on va alors placer toutes ces informations sur un tableau à double entrée (c'est le même raisonnement qu'un arbre de probabilités) pour que les informations soient plus visuelles. Il faut alors calculer le nombre de personnes dans chacune des situations possibles dans le cas d'un tableau à double-entrée (et les probabilités dans un arbre de probabilités).

Si on s'intéresse à la population générale (même raisonnement pour la population à risque) :

- Il y a 7 personnes malades, donc 1000 - 7 = 993 personnes saines.
- 94,5% des femmes testées sont négatives donc 945 femmes sont négatives au test.
- Il y a $1000 - 945 = 55$ femmes testées positives.

On peut alors remplir le tableau de cette manière :

En population générale

	Statut du patient		
Test	Malade	Pas malade	*Total*
+			55
-			945
Total	7	993	1000

- On complète ensuite l'intérieur du tableau. Le croisement entre une ligne et une colonne correspond à l'intersection (« ET » ou $\cap$) de ces deux évènements.
- On peut calculer $P(T^+ \cap \overline{M})$ à l'aide de la relation :

$$P(T^+ \cap \overline{M}) = P(T^+|\overline{M}) \times P(\overline{M}) = 0{,}05 \times \frac{993}{1000} \approx 0{,}05$$

- Pour calculer le nombre de patientes positives et saines, on fait $0{,}05 \times 1000 = 50$. On peut en déduire que $card(T^+ \cap M) = 55 - 50 = 5$.

Finalement, on obtient ce tableau :

En population générale

	Statut du patient		
Test	Malade	Pas malade	*Total*
+	5	50	55
-	2	943	945
Total	7	993	1000

Dans la **population à risque**, en retranscrivant l'énoncé en probabilité, on a :

- $P(T^-|M) = 0{,}25$
- $P(T^+) = 0{,}2$
- $P(\overline{M}) = \frac{800}{1000}$

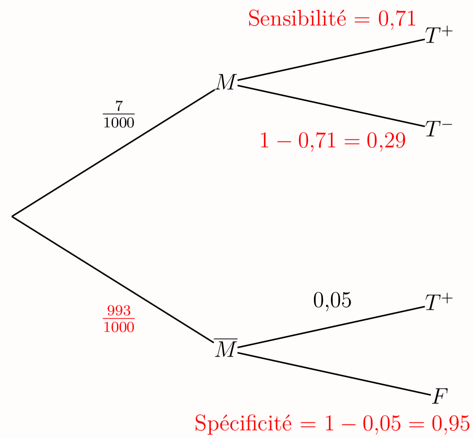
En suivant la même méthode, on obtient :

En population à risque

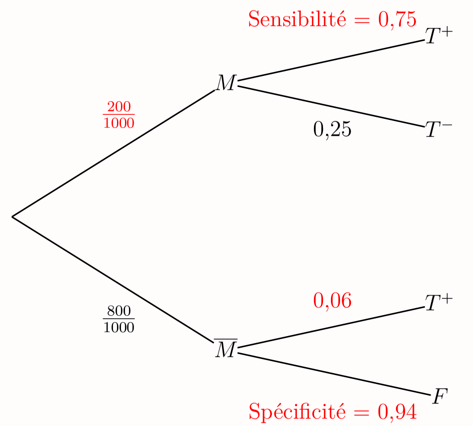
	Statut du patient		
Test	Malade	Pas malade	*Total*
+	150	50	200
-	50	750	800
Total	200	800	1000

En raisonnant avec un arbre des probabilités (les valeurs en noires sont celles qui sont données dans l'énoncé, les valeurs en rouge sont à calculer par vos soins), on obtiendrait :

🌲 Pour la population générale :



🌲 Pour la population à risque :



À présent, on peut s'attaquer aux items 😉 !

✔ **Item A** → *Dans la « Population à risque », la prévalence est de 20%*

La prévalence est le nombre de cas d'une maladie dans une population donnée. Ici, il s'agit donc de $P(M)$. Dans l'échantillon « Population à risque », $P(M) = 1 - \frac{800}{1000} = 20\%$.

✘ **Item B** → *La sensibilité dans la « Population générale » est de 80%*

La sensibilité est la probabilité de la présence d'un signe S sachant que le sujet est malade, autrement dit la probabilité d'avoir un test positif sachant que le sujet est malade.

D'après le formulaire :

$$Se = P(T^+|M) = \frac{P(T^+ \cap M)}{P(M)}$$

On fait l'application numérique :

$$Se_G = \frac{0{,}005}{0{,}007} = 0{,}71$$

✔ **Item C** → *Dans la « Population à risque », la valeurs prédictives positives est de 75%*

Le formulaire nous indique que

$$VPP = P(M|T^+)$$
$$\iff VPP = \frac{P(M \cap T^+)}{P(T^+)}$$

Avec l'aide du tableau que nous avons dressé, on trouve

$$VPP = \frac{150}{200} = 75\%$$

✘ **Item D** → *Dans le cas général, la spécificité augmente si la prévalence augmente*

Notez-le dans votre cahier d'erreur +++ :

📝 La spécificité et la sensibilité sont INDÉPENDANTES de la prévalence de la maladie.

📝 La VPP et la VPN DÉPENDENT de la prévalence : VPP augmente quand la prévalence augmente et VPN diminue quand la prévalence augmente.

✔ **Item E** → *Si on tire au hasard une patiente dans la « population générale », il y a 0,5% de ...*

Si on retranscrit l'item en mathématique, cela revient à calculer :

$$P(M \cap T^+) = \frac{5}{1000} = 0{,}005 = 0{,}5\%$$

Réponses vraies : A, C et E

À la suite d'études cliniques, des statisticiens ont constaté que ce sont plus souvent les femmes exposées à des facteurs de risques qui se font dépister. Nous nous intéresserons ainsi à l'échantillon « Population à risque » de $n = 1000$ patientes et considérerons que chaque test T est indépendant et identiquement distribué.

Soit X la variable aléatoire qui prend la valeur 1 lorsque le test est positif, et la valeur 0 lorsque le test est négatif.

Question 7 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. X est une variable aléatoire qui suit une loi de Bernoulli de paramètres $n = 1000$ et $\pi = 20\%$

B. X est une variable aléatoire discrète

C. En considérant Y une variable aléatoire comptant le nombre de tests positifs, $\mathbb{E}(Y) = 200$ et $\mathbb{V}ar(Y) = 160$

D. Y est approximable par une loi gaussienne car $n \geq 30$

E. X est approximable par une loi gaussienne car $n\pi \geq 5$ et $n(1-\pi) \geq 5$

Question 7

✘ **Item A** → *X est une variable aléatoire qui suit une loi de Bernoulli de paramètres $n = 1000$ ...*

D'une part : rappelons les différences entre loi de Bernoulli et loi Binomiale :

🎲 La loi de Bernoulli est

- Une loi de probabilité d'une variable aléatoire discrète qui prend la valeur 1 lors du « succès » (avec une probabilité π) ou 0 lors d'un « échec » (avec une probabilité $1 - \pi$).
- Elle ne dépend que d'un seul paramètre, la probabilité d'un succès π.

🎲 La loi Binomiale est

- Une répétition de n épreuves de Bernoulli indépendantes et identiquement distribuées (iid). Ainsi, on **compte** le **nombre** de succès en répétant les épreuves de Bernoulli.
- Elle dépend de deux paramètres : n (nombre d'épreuves de Bernoulli) et π (la probabilité d'un succès).

« X est la variable aléatoire qui prend la valeur 1 lorsque le test est positif, et la valeur 0 lorsque le test est négatif. ». Elle suit une loi de Bernoulli. Ici, on a essayé de vous embrouiller entre le nom de la loi et le nombre de paramètres. Faites bien attention à ce qu'on calcule avec cette variable aléatoire !

D'autre part :

$$\pi = \frac{200}{1000} = 0{,}2 \Rightarrow X \sim \mathcal{B}(\pi = 0{,}2)$$

Petit tips de la maison 🏠 : on peut dire qu'une loi de Bernoulli est une loi binomiale de paramètres $n = 1$ et de probabilité π, car on a qu'une seule épreuve de Bernoulli à considérer.

✔ **Item B** → *X est une variable aléatoire discrète*

$\mathcal{B}(\pi)$, X est bien une loi discrète, pas de piège !

✔ **Item C** → *En considérant Y une variable aléatoire comptant le nombre de tests positifs, ...*

« Y une variable aléatoire comptant le nombre de tests positifs » → On a ici $Y \sim \mathcal{B}in(1000; 0{,}2)$.

D'après le formulaire, pour une loi binomiale :

$$\mathbb{E}(X) = n\pi \text{ et } \mathbb{V}ar(X) = n\pi(1-\pi)$$

Application numérique :

$$\mathbb{E}(X) = 1000 \times 0{,}2 = 200 \text{ et } \mathbb{V}ar(X) = 1000 \times 0{,}2 \times 0{,}8 = 160$$

RAS, tout est bon, on coche !

✘ **Item D** → *Y est approximable par une loi gaussienne car $n \geq 30$*

Rappel du Théorème Central Limite (TCL) pour une **loi Binomiale** : il y a deux conditions de validité à satisfaire : $n\pi \geq 5$ et $n(1-\pi) \geq 5$.

L'item est donc faux.

✘ **Item E** → *X est approximable par une loi gaussienne car $n\pi \geq 5$ et $n(1-\pi) \geq 5$*

Attention, on n'approxime pas une loi de Bernoulli par le TCL car une loi de Bernoulli est une loi binomiale de paramètres $n = 1$ et de probabilité π. Ainsi, on ne satisfait pas la première condition ($n \geq 30$).

Réponses vraies : B et C

On effectue un nouvel essai clinique sur un échantillon cette fois-ci de $n = 20$ personnes, de même probabilité d'avoir un test positif que la variable X.

Question 8 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La probabilité que le nombre de tests positifs soit strictement supérieur à 15 dans la population à risque est nulle

B. La probabilité que le nombre de tests négatifs soit au moins égal à 12 dans la population à risque est d'environ 0,99

C. La probabilité que le nombre de tests négatifs soit égal à 17 est d'environ 20,5%

D. Si le nombre de personnes interrogées valait 30, la probabilité que le nombre de tests positifs soit égal à 1 serait inférieure à 1%

E. Une variable aléatoire est dite centrée réduite si $\mu = 0$ et $\sigma = 1$

Question 8

✘ **Item A** → *La probabilité que le nombre de tests positifs soit strictement supérieur à 15 dans ...*

On a changé nos paramètres de Y : $Y \sim \mathcal{B}\rangle\backslash(20; 0{,}2)$.

On s'intéresse au nombre de tests positifs/négatifs, ce qui correspond à un nombre de succès et d'échecs. Ainsi, on se sert de la variable aléatoire Y qui compte le nombre de tests + et de tests -.

Pour l'item A, on retranscrit mathématiquement l'énoncé :

$$P(Y > 15) = P(Y = 16) + P(Y = 17) + P(Y = 18) + P(Y = 19) + P(Y = 20)$$

HOP, petit rendez-vous à la page 10 de votre 💎 PRÉCIEUX 💎 formulaire :

20	0	0,1216	0,0115	0,0032	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,2702	0,0576	0,0211	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000
	2	0,2852	0,1369	0,0669	0,0031	0,0002	0,0000	0,0000	0,0000	0,0000
	3	0,1901	0,2054	0,1339	0,0123	0,0011	0,0000	0,0000	0,0000	0,0000
	4	0,0898	0,2182	0,1897	0,0350	0,0046	0,0003	0,0000	0,0000	0,0000
	5	0,0319	0,1746	0,2023	0,0746	0,0148	0,0013	0,0000	0,0000	0,0000
	6	0,0089	0,1091	0,1686	0,1244	0,0370	0,0049	0,0000	0,0000	0,0000
	7	0,0020	0,0545	0,1124	0,1659	0,0739	0,0146	0,0002	0,0000	0,0000
	8	0,0004	0,0222	0,0609	0,1797	0,1201	0,0355	0,0008	0,0001	0,0000
	9	0,0001	0,0074	0,0271	0,1597	0,1602	0,0710	0,0030	0,0005	0,0000
	10	0,0000	0,0020	0,0099	0,1171	0,1762	0,1171	0,0099	0,0020	0,0000
	11	0,0000	0,0005	0,0030	0,0710	0,1602	0,1597	0,0271	0,0074	0,0001
	12	0,0000	0,0001	0,0008	0,0355	0,1201	0,1797	0,0609	0,0222	0,0004
	13	0,0000	0,0000	0,0002	0,0146	0,0739	0,1659	0,1124	0,0545	0,0020
	14	0,0000	0,0000	0,0000	0,0049	0,0370	0,1244	0,1686	0,1091	0,0089
	15	0,0000	0,0000	0,0000	0,0013	0,0148	0,0746	0,2023	0,1746	0,0319
	16	0,0000	0,0000	0,0000	0,0003	0,0046	0,0350	0,1897	0,2182	0,0898
	17	0,0000	0,0000	0,0000	0,0000	0,0011	0,0123	0,1339	0,2054	0,1901
	18	0,0000	0,0000	0,0000	0,0000	0,0002	0,0031	0,0669	0,1369	0,2852
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0005	0,0211	0,0576	0,2702
	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0032	0,0115	0,1216

On a donc $P(Y > 15) \approx 0$ mais on ne peut pas cocher cet item car :

$$\text{Si } Y \sim \mathcal{B}\rangle\backslash(n, \pi),\ P(Y = k) \neq 0,\ \text{avec}(0 \leq k \leq n)$$

Autrement dit, lorsqu'on est dans le cas d'une loi Binomiale une probabilité n'est jamais nulle. En revanche, elle peut être TRÈS PROCHE (vreeeeuuument très proche) de 0.

✔ **Item B** → *La probabilité que le nombre de tests négatifs soit au moins égal à 12 dans la ...*

On cherche à calculer la probabilité d'échecs (nombre de tests NÉGATIFS). Si on veut au moins 12 tests négatifs ($\geq$ 12 tests négatifs), ça veut dire qu'on veut au maximum 8 tests positifs.

Cela revient donc à calculer :

$$P(Y \leq 8) = P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) \\ + P(Y = 5) + P(Y = 6) + P(Y = 7) + P(Y = 8)$$

Je vous vois paniquer là, mais calmez-vous (Tu veux pleurer ? Arrête). Ça revient à sommer cette partie du formulaire :

20	0	0,1216	0,0115	0,0032	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,2702	0,0576	0,0211	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000
	2	0,2852	0,1369	0,0669	0,0031	0,0002	0,0000	0,0000	0,0000	0,0000
	3	0,1901	0,2054	0,1339	0,0123	0,0011	0,0000	0,0000	0,0000	0,0000
	4	0,0898	0,2182	0,1897	0,0350	0,0046	0,0003	0,0000	0,0000	0,0000
	5	0,0319	0,1746	0,2023	0,0746	0,0148	0,0013	0,0000	0,0000	0,0000
	6	0,0089	0,1091	0,1686	0,1244	0,0370	0,0049	0,0000	0,0000	0,0000
	7	0,0020	0,0545	0,1124	0,1659	0,0739	0,0146	0,0002	0,0000	0,0000
	8	0,0004	0,0222	0,0609	0,1797	0,1201	0,0355	0,0008	0,0001	0,0000
	9	0,0001	0,0074	0,0271	0,1597	0,1602	0,0710	0,0030	0,0005	0,0000
	10	0,0000	0,0020	0,0099	0,1171	0,1762	0,1171	0,0099	0,0020	0,0000
	11	0,0000	0,0005	0,0030	0,0710	0,1602	0,1597	0,0271	0,0074	0,0001
	12	0,0000	0,0001	0,0008	0,0355	0,1201	0,1797	0,0609	0,0222	0,0004
	13	0,0000	0,0000	0,0002	0,0146	0,0739	0,1659	0,1124	0,0545	0,0020
	14	0,0000	0,0000	0,0000	0,0049	0,0370	0,1244	0,1686	0,1091	0,0089
	15	0,0000	0,0000	0,0000	0,0013	0,0148	0,0746	0,2023	0,1746	0,0319
	16	0,0000	0,0000	0,0000	0,0003	0,0046	0,0350	0,1897	0,2182	0,0898
	17	0,0000	0,0000	0,0000	0,0000	0,0011	0,0123	0,1339	0,2054	0,1901
	18	0,0000	0,0000	0,0000	0,0000	0,0002	0,0031	0,0669	0,1369	0,2852
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0005	0,0211	0,0576	0,2702
	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0032	0,0115	0,1216

On a beaucoup de chiffres assez longs, c'est plus pratique de calculer dans l'autre sens :

$$\begin{aligned} P(Y \leq 8) &= 1 - P(Y > 8) \\ &= 1 - P(Y \geq 9) \\ &= 1 - (P(Y = 9) + P(Y = 10) + P(Y = 11) + ... + P(Y = 19) + P(Y = 20)) \\ &= 1 - (0 + 0{,}0001 + 0{,}0005 + 0,0020 + 0,0074) \\ &= 1 - 0,01 \\ &= 0{,}99 \end{aligned}$$

✔ **Item C** → *La probabilité que le nombre de tests négatifs soit égal à 17 est d'environ 20,5%*

La probabilité que le nombre de tests négatifs soit égal à 17 (donc la probabilité que le nombre de tests positifs soit égal à 3) est retranscrit mathématiquement par $P(Y = 3)$. Un petit date encore une fois au formulaire :

20	0	0,1216	0,0115	0,0032	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,2702	0,0576	0,0211	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000
	2	0,2852	0,1369	0,0669	0,0031	0,0002	0,0000	0,0000	0,0000	0,0000
	3	0,1901	0,2054	0,1339	0,0123	0,0011	0,0000	0,0000	0,0000	0,0000
	4	0,0898	0,2182	0,1897	0,0350	0,0046	0,0003	0,0000	0,0000	0,0000
	5	0,0319	0,1746	0,2023	0,0746	0,0148	0,0013	0,0000	0,0000	0,0000
	6	0,0089	0,1091	0,1686	0,1244	0,0370	0,0049	0,0000	0,0000	0,0000
	7	0,0020	0,0545	0,1124	0,1659	0,0739	0,0146	0,0002	0,0000	0,0000
	8	0,0004	0,0222	0,0609	0,1797	0,1201	0,0355	0,0008	0,0001	0,0000
	9	0,0001	0,0074	0,0271	0,1597	0,1602	0,0710	0,0030	0,0005	0,0000
	10	0,0000	0,0020	0,0099	0,1171	0,1762	0,1171	0,0099	0,0020	0,0000
	11	0,0000	0,0005	0,0030	0,0710	0,1602	0,1597	0,0271	0,0074	0,0001
	12	0,0000	0,0001	0,0008	0,0355	0,1201	0,1797	0,0609	0,0222	0,0004
	13	0,0000	0,0000	0,0002	0,0146	0,0739	0,1659	0,1124	0,0545	0,0020
	14	0,0000	0,0000	0,0000	0,0049	0,0370	0,1244	0,1686	0,1091	0,0089
	15	0,0000	0,0000	0,0000	0,0013	0,0148	0,0746	0,2023	0,1746	0,0319
	16	0,0000	0,0000	0,0000	0,0003	0,0046	0,0350	0,1897	0,2182	0,0898
	17	0,0000	0,0000	0,0000	0,0000	0,0011	0,0123	0,1339	0,2054	0,1901
	18	0,0000	0,0000	0,0000	0,0000	0,0002	0,0031	0,0669	0,1369	0,2852
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0005	0,0211	0,0576	0,2702
	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0032	0,0115	0,1216

$0{,}2054 \approx 20{,}5\%$ L'item est donc juste !

✔ **Item D** → *Si le nombre de personnes interrogées valait 30, la probabilité que le nombre de ...*

On a changé le nombre de femmes interrogées, donc il faut créer une nouvelle variable aléatoire W telle que $W \sim \mathcal{B}in(30; 0{,}2)$.

On veut calculer cette fois... $P(W = 2)$. Le formulaire se limite à $n = 20$. Il va donc falloir calculer à la main.

On rappelle que pour une loi Binomiale :

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

Ainsi, on calcule (à la calculatrice) :

$$\begin{aligned} \binom{n}{k} = \binom{30}{1} &= \frac{n!}{k! \times (n-k)!} \\ &= \frac{30!}{1! \times 29!} \\ &= \frac{30 \times 29!}{1 \times 29!} \\ &= \frac{30}{1} \\ &= 30 \end{aligned}$$

$$\begin{aligned} \Rightarrow P(W = 1) &= 30 \times 0{,}2^1 \times 0{,}8^{29} \\ &\approx 0{,}009 \\ &< 0{,}01 \end{aligned}$$

$\binom{30}{1}$ correspond au différents façon d'avoir 1 test positif parmi 30 tests. Il y a 30 combinaisons possibles (soit le test de la première patiente est positif, soit celui de la seconde, soit celui de la troisième, ..., soit celui de la 30-ème).
L'item est donc correct.

✔ **Item E** → *Une variable aléatoire est dite centrée réduite si $\mu = 0$ et $\sigma = 1$*

Pas de piège, c'est la définition d'une variable aléatoire centrée réduite ! On coche à fond !

Réponses vraies : B, C, D et E

Les dépistages sont efficaces pour éviter que les symptômes ne s'aggravent, mais nombreuses sont les patientes qui n'y ont pas recours. Ainsi, celles ayant des symptômes graves se font hospitaliser. On choisit un nouvel échantillon de 200 patientes. La durée d'hospitalisation H suit une loi exponentielle de paramètre λ. Une patiente se remet en moyenne au bout de 4 mois de chimiothérapie.

Question 9 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Il faudrait moins de 3 mois pour que la moitié des femmes soient déshospitalisées
B. La probabilité qu'une patiente soit hospitalisée plus d'un mois est de 0,78 au centième près

C. On peut approcher la variable M_{200} de la moyenne de 200 échantillons de la variable H par une loi normale de moyenne 20 et de variance 5
D. La probabilité d'être hospitalisée pendant plus d'un an et demi est nulle

E. Si, au bout de 2 mois, les cellules cancéreuses prolifèrent toujours, la probabilité qu'elle soit guérie au bout de 2 mois et demi est de 0,12 au centième près

Question 9

Avant répondre aux questions, on écrit mathématiquement notre variable aléatoire H qui représente la durée d'hospitalisation d'une patiente :

$$H \sim Exp(\lambda)$$

Pour déterminer la valeur de λ, on regarde les données qu'on a :

- 📈 $n = 200$
- 📈 « Une patiente se remet en moyenne en 4 mois de chimiothérapie. ». Ainsi, $\mathbb{E}(H) = 4$.

Or, pour une loi exponentielle, $\lambda = \frac{1}{\mathbb{E}(H)} = \frac{1}{4} = 0{,}25$. On a donc la valeur de notre paramètre, on peut à présent démarrer nos questions en paix !

✔ **Item A** → *Il faudrait moins de 3 mois pour que la moitié des femmes soient déshospitalisées*

On cherche le temps t telle que la moitié des femmes soient déshospitalisées. Mathématiquement, cela revient à chercher la date t qui vérifie :

$$P(H < t) = 1 - e^{-0{,}25\times t} = 0{,}5$$

$$\iff e^{-0{,}25\times t} = 0{,}5$$

$$\begin{aligned}\iff -0{,}25 \times t &= \ln 0{,}5\\ &= \ln \frac{1}{2}\\ &= -\ln 2\end{aligned}$$

$$\iff t = \frac{\ln 2}{0{,}25} \approx 2{,}77 \text{ mois}$$

Il faut donc environ 2,77 mois (moins de 3 mois) pour que la moitié des femmes soient déshospitalisées.

✔ **Item B** → *La probabilité qu'une patiente soit hospitalisée plus d'un mois est de 0,78 au ...*

Mathématiquement, on cherche à calculer $P(H > 1)$.

On fait simplement l'application numérique :

$$P(H > 1) = 1 - P(H < 1) = 1 - (1 - e^{-\lambda x}) = e^{-0{,}25\times 1} \approx 0{,}78$$

✘ **Item C** → *On peut approcher la variable M_{200} de la moyenne de 200 échantillons de la variable ...*

Condition de validité du TCL pour une loi continue :

- $n \geq 30$ ce qui est bien le cas !

Ainsi on peut approximer M_{200} par une loi normale telle que $M_{200} \sim \mathcal{N}(\mu; \sigma^2)$.

🤔 **μ est la moyenne de la durée des hospitalisations.**

$$\mathbb{E}(M_{200}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(H_i)$$

On divise par $\frac{1}{n}$ parce qu'on calcule une moyenne.

Or, $\sum_{i=1}^{n} \mathbb{E}(H_i) = n \times \mathbb{E}(H)$ car tous les $E(H_i)$ sont égaux. Ainsi :

$$\begin{aligned}\mu &= \frac{n \times \mathbb{E}(H)}{n}\\ &= \mathbb{E}(H) = 4 \text{ mois}\end{aligned}$$

🤔 **σ^2 est la variance de la durée des hospitalisations.**

On raisonne de la même manière pour la variance. On rappelle que :

$$\mathbb{V}ar(H) = \frac{1}{\lambda^2} = \mathbb{E}(H)^2 = 16$$

On obtient donc :

$$\begin{aligned}\sigma^2 &= \frac{1}{n^2} \times \sum_{i=1}^{n} \mathbb{V}ar(H_i)\\ &= \frac{1}{n^2} \times n \times \mathbb{V}ar(H)\\ &= \frac{\mathbb{V}ar(H)}{n}\\ &= \frac{16}{200} = 0{,}08 \text{ mois}^2\end{aligned}$$

L'item est donc FAUX, $M_{200} \sim \mathcal{N}(4\,;\,0{,}08)$.

✘ **Item D** → *La probabilité d'être hospitalisée pendant plus d'un an et demi est nulle*

Sans même calculer, on peut rejeter cet item car dans le cas d'une loi continue, une probabilité n'est jamais nulle !

En faisant l'application numérique, on a $P(H > 12 + 6) = P(H > 18) \approx 0{,}011109$, et non 0 tout rond (comme la tête à toto *ok faut que j'aille me coucher*, et toi aussi ne dors pas trop tard et prends soin de toi ❤️).

13

Item E → *Si, au bout de 2 mois, les cellules cancéreuses prolifèrent toujours, la probabilité ...*

$$\begin{aligned} P(H < 2{,}5 | H > 2) &= P(H < 2{,}5 - 2) \text{ (puisque l'exponentielle est une **loi sans mémoire**)} \\ &= P(H < 0{,}5) \\ &= 1 - e^{-\frac{1}{4}\times 0{,}5} \\ &\approx 0{,}12 \end{aligned}$$

L'item est bien correct !

Réponses vraies : A, B et E

On prend un échantillon de 200 patientes hospitalisées à la même date. On cherche à estimer la proportion au risque α de patientes toujours malades au bout de 4 et 12 mois d'hospitalisation.

Question 10 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les proportions théoriques de malades à 4 et 12 mois sont respectivement d'environ $\pi_4 = 37\%$ et $\pi_{12} = 5\%$

B. L'intervalle de pari au risque 15% de la proportion de malades à 12 mois est [0,03 ; 0,07] à 1% près

C. L'intervalle de pari au risque 1% de la proportion de malade au bout de 4 mois est plus étroit que l'intervalle de pari au risque 5% de cette même proportion

D. L'intervalle de pari au risque 5% de la proportion de malades à 12 mois est plus large que l'intervalle de pari au risque 1% de la proportion de malades à 4 mois

E. Le risque alpha est l'erreur de première espèce et il correspond à la probabilité de se tromper sous l'hypothèse nulle

Question 10

Item A → *Les proportions théoriques de malades à 4 et 12 mois sont respectivement d'environ ...*

Avant de commencer, on va chercher à déterminer les paramètre π_4 et π_{12} associés à la ...
On trouve, grâce à la question précédente que $\pi_4 = P(H > 4)$ et $\pi_{12} = P(H > 12)$. On a

$$\begin{aligned} P(H > 4) &= e^{-\frac{1}{4}\times 4} \\ &= e^{-1} \\ &\approx 0{,}37 \end{aligned}$$

$$\begin{aligned} P(H > 12) &= e^{-\frac{1}{4}\times 12} \\ &= e^{-3} \\ &\approx 0{,}05 \end{aligned}$$

Ainsi, $\pi_4 = 0{,}37\%$ et $\pi_{12} = 5\%$!

Item B → *L'intervalle de pari au risque 15% de la proportion de malades à 12 mois est ...*

Premièrement, on vérifie si on parle d'intervalle de pari ou d'intervalle de confiance.

- On s'intéresse à la proportion **observée** $P_{12,n}$ de malades après 12 mois dans un n-échantillon. On dispose des paramètres théoriques en population générale. Ainsi, cherchera un intervalle de PARI d'une proportion.
- Petit moyen mnémotechnique. Probabilités ⇒ Intervalle de Pari. (Ça commence par la même lettre).
  On s'intéresse à un intervalle de PARI d'une proportion.

Ensuite, **avant de calculer tout intervalle de pari**, il faut vérifier si on peut appliquer le TCL :

- $n\pi_{12} = 200 \times 0{,}05 = 10 \geq 5$
- $n(1 - \pi_{12}) = 200 \times (1 - 0{,}05) = 190 \geq 5$

Les conditions d'application du TCL sont vérifiées, on peut calculer $IP_{1-\alpha}(P_{12,n})$

Il faut encore utiliser le **précieux** formulaire. Il nous indique que :

$$IP_{1-\alpha\%}(P_{12,n}) = \pi_{12} \pm \varepsilon_\alpha \sqrt{\frac{\pi_{12}(1-\pi_{12})}{n}}$$

Pour trouver ε_α on regarde la table 2a. On sait que $\alpha = 0{,}15 = 0{,}1 + 0{,}05$.

α	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,1	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

On a $\varepsilon_{0,15} = 1{,}44$

On fait l'application numérique et on trouve :

$$\begin{aligned} \pi_{12} \pm \varepsilon_\alpha \sqrt{\frac{\pi_{12}(1-\pi_{12})}{n}} &= 0{,}05 \pm 1{,}44\sqrt{\frac{0{,}05\times 0{,}95}{200}} \\ &\approx [0{,}028\ ;0{,}072] \\ &= [0{,}03\ ;0{,}07] \quad \text{(à 1\% près)} \end{aligned}$$

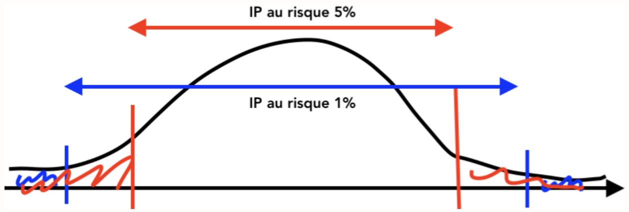
Item C → *L'intervalle de pari au risque 1% de la proportion de malade au bout de 4 mois est ...*

Deux manières d'aborder le sujet :

- Méthode 1 : méthode calculatoire :
  On s'intéresse au quantile $\varepsilon_{0,01} = 2{,}576$ et $\varepsilon_{0,05} = 1{,}960$. En s'appuyant sur cette formule : $P_n : \pi_4 \pm \sqrt{\frac{\pi_4(1-\pi_4)}{n}}$, comme $2{,}576 > 1{,}960$, l'intervalle de pari sera plus grand au risque 1 % qu'au risque 5%. (en effet, plus $\varepsilon_\alpha \uparrow$, plus l'IP $\uparrow$).

14

Méthode 2 : méthode schématique :
On s'intéresse aux zones de rejet :



Plus α augmente, plus la zone de rejet est grande, donc plus l'IP est petite. Inversement, plus l'alpha diminue, plus la zone de rejet est petite, plus l'IP est grande.

Ainsi, l'item est FAUX.

✘ **Item D** → *L'intervalle de pari au risque 5% de la proportion de malades à 12 mois est plus ...*

On est toujours dans le cas d'un intervalle de pari.
On sait qu'on peut appliquer le TCL à la variable $P_{12,n}$. Qu'en est-il de $P_{4,n}$?

- $n\pi_4 = 200 \times 0{,}37 = 74 \geq 5$
- $n(1-\pi_4) = 200 \times (1-0{,}37) = 126 \geq 5$

On peut donc calculer les intervalles de pari.

On se rappelle de la formule de la largeur d'un intervalle de pari.

$$l = 2\varepsilon_\alpha\sqrt{\frac{\pi(1-\pi)}{n}}$$

De plus

$$\alpha = 5\% \Rightarrow \varepsilon_\alpha = 1{,}96$$

et

$$\alpha = 1\% \Rightarrow \varepsilon_\alpha = 2{,}576$$

Soient $l_{4,\ 5\%}$ la largeur de l'intervalle de paris au risque 5% de la proportion de malades après 4 mois,

$$\begin{aligned} l_{4,\ 5\%} &= 2\varepsilon_\alpha\sqrt{\frac{\pi_4(1-\pi_4)}{n}} \\ &= 2 \times 1{,}96 \times \sqrt{\frac{0{,}37 \times 0{,}63}{200}} \\ &= 0{,}13 \end{aligned}$$

De la même manière, on calcule $l_{12,\ 1\%}$, la largeur de l'intervalle de paris au risque 1% de la proportion de malades après 12 mois,

$$\begin{aligned} l_{12,\ 1\%} &= 2\varepsilon_\alpha\sqrt{\frac{\pi_{12}(1-\pi_{12})}{n}} \\ &= 2 \times 2{,}576 \times \sqrt{\frac{0{,}05 \times 0{,}95}{200}} \\ &= 0{,}08 \end{aligned}$$

On a $0{,}08 < 0{,}13$, donc $l_{12,\ 1\%} < l_{4,\ 5\%}$. Ainsi l'item est FAUX

✔ **Item E** → *Le risque alpha est l'erreur de première espèce et il correspond à la probabilité de ...*

Retenez bien que :

	Sous H_0	Sous H_1
Rejet de H_0	α	$1-\beta$
Non rejet de H_0	1 - α	β

- α : erreur de première espèce = risque alpha = seuil de signification.
  Il s'agit du rejet de H_0 en supposant que H_0 est vraie. Autrement dit, on « se trompe » d'un risque alpha de rejeter H_0 par erreur.
- $1-\beta$: puissance.
  C'est la probabilité d'avoir encore plus de raisons de rejeter H_0 sachant qu'on a supposé que H_0 est faux (donc H_1 vraie).

L'item est donc VRAI !

Réponses vraies : A, B et E

Exercice 3 (questions 11 à 15)

Les tardigrades sont des Eumétazoaires capables de résister pendant des périodes prolongées à des conditions environnementales extrêmes. Un laboratoire en biologie animale souhaite étudier les mécanismes qui permettent à ces organismes de survivre face à ces milieux. Ils s'intéressent tout particulièrement aux tardigrades vivant dans l'océan et dans l'Himalaya. Pour cela, des biologistes étudient un échantillon A de tardigrades venant de l'océan et un échantillon B de tardigrades issus de l'Himalaya. Ils décident de comparer les tailles moyennes dans les différents populations.
Dans le groupe A, d'effectif $n_A = 120$, les adultes ont une taille de moyenne $m_A = 1{,}54$ *mm* et d'écart-type $s_A = 0{,}12$ *mm*. Dans le groupe B, d'effectif $n_B = 160$, les adultes ont une taille de moyenne $m_B = 1{,}56$ *mm* et d'écart-type $s_B = 0{,}04$ *mm*.

Question 11 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Sous l'hypothèse nulle H_0, $m_A = m_B$
B. On réalise un test de l'écart-réduit de statistique $|Z| > 2$
C. Les tailles moyennes des deux populations sont égales
D. Au risque $\alpha = 5\%$, on ne rejette pas H_0
E. Le degré de significativité du test est supérieur à 7%

Question 11

✘ **Item A** → *Sous l'hypothèse nulle H_0, $m_A = m_B$*
On réalise ici un test de comparaisons de moyennes entre deux populations de tardigrades vivant à différents endroits à partir des moyennes observées dans les échantillons A et B . Ici, on dispose déjà des paramètres observés dans chacun des échantillons (m_A et m_B) et l'on souhaite déterminer s'il existe une différence significative ou pas pour les paramètres des échantillons théoriques. On compare donc μ_A et μ_B, moyennes théoriques des populations desquelles sont issus les échantillons A et B.

On pose alors deux hypothèses devant s'exclure mutuellement :
— Sous l'hypothèse nulle H_0 : les paramètres théoriques sont égaux entre les deux échantillons donc $\mu_A = \mu_B$
— Sous l'hypothèse alternative H_1 : il y a une différence entre les paramètres théoriques donc $\mu_A \neq \mu_B$

✘ **Item B** → *On réalise un test de l'écart-réduit de statistique $|Z| > 2$*
Il faut tout d'abord vérifier les conditions de validité du théorème central limite : les effectifs n_A et n_B sont supérieurs à 30 donc ces conditions sont validées.
D'après le formulaire :

$$Z = \frac{m_A - m_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{1{,}54 - 1{,}56}{\sqrt{\frac{0{,}12^2}{120} + \frac{0{,}04^2}{160}}} = \frac{-0{,}02}{\sqrt{\frac{0{,}0144}{120} + \frac{0{,}0016}{160}}} \simeq -1{,}754$$
$$|Z| \simeq 1{,}754 < 2$$

✘ **Item C** → *Les tailles moyennes des deux populations sont égales*
Dire qu'il existe une égalité entre les paramètres théoriques des deux échantillons revient à accepter H_0. Or on n'accepte jamais H_0, on ne peut que "ne pas la rejeter" !

✔ **Item D** → *Au risque $\alpha = 5\%$, on ne rejette pas H_0*
Grâce à l'item B, on sait que $|Z| \simeq 1{,}754$.
D'après la table 2a, le risque α correspond à la probabilité que $|Z| > \varepsilon_\alpha$.
Au risque $\alpha = 5\%$, $\varepsilon_\alpha = 1{,}96 > |Z|$. $|Z|$ ne se situe pas dans la zone de rejet de H_0 donc on ne rejette pas H_0 à ce risque.

✔ **Item E** → *Le degré de significativité du test est supérieur à 7%*
De la même manière, le degré de significativité du test correspond au risque α à partir duquel on ne peut plus rejeter H_0. On cherche donc la valeur de α pour laquelle $|Z| > \varepsilon_\alpha$. On sait également que $|Z| = 1{,}754$.
D'après la table 2a, $\varepsilon_{8\%} = 1{,}751$ et $\varepsilon_{7\%} = 1{,}812$.
On a donc $\varepsilon_{8\%} = 1{,}751 < 1{,}754 < 1{,}821 = \varepsilon_{7\%}$. On sait que α diminue lorsque ε_α augmente, donc le degré de significativité est bien supérieur à 7%.

Réponses vraies : D et E

Le laboratoire décide alors d'exposer deux échantillons de tardigrades vivant dans l'océan à différentes conditions environnementales. Le groupe C d'effectif $n_C = 120$ est exposé à une très basse température tandis que le groupe D d'effectif n_D est exposé à une pression très élevée. Au bout de deux heures, on note la proportion p de tardigrades ayant survécu. On note $p_C = 79\%$ pour le groupe C et $p_D = 86\%$ pour le groupe D et on décide de comparer les proportions théoriques entre les deux échantillons par un test de l'écart réduit.

Question 12 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. On peut réaliser ce test car $n_C p_C \geq 5$, $n_C(1-p_C) \geq 5$, $n_D p_D \geq 5$ et $n_D(1-p_D) \geq 5$
B. L'intervalle de pari à 95% de la proportion théorique de survivants dans le groupe C est $[0{,}71; 0{,}87] \pm 0{,}01$
C. L'intervalle de confiance à 96% de la proportion théorique de survivants dans le groupe D est $[0{,}80; 0{,}92] \pm 0{,}01$
D. En acceptant H_1 à tort, le laboratoire risque de commettre une erreur de type II ou risque de deuxième espèce
E. La statistique du test est $|Z| \simeq 1{,}543$

Question 12

✘ **Item A** → *On peut réaliser ce test car $n_C p_C \geq 5$, $n_C(1-p_C) \geq 5$, $n_D p_D \geq 5$ et $n_D(1-p_D) \geq 5$*
On sait que l'on réalise un test de comparaison entre les proportions de deux échantillons. Afin de vérifier les conditions de validité de ce test, on doit calculer la proportion commune p. D'après le formulaire :

$$p = \frac{n_C p_C + n_D p_D}{n_C + n_D} = \frac{120 \times 0{,}79 + 160 \times 0{,}86}{120 + 160} = 0{,}83$$

Les conditions de validité de ce test sont :
— $n_C p = 120 \times 0{,}83 = 99{,}6 \geq 5$ et $n_C(1-p) = 120 \times 0{,}17 = 20{,}4 \geq 5$
— $n_D p = 160 \times 0{,}83 = 132{,}8 \geq 5$ et $n_D(1-p) = 160 \times 0{,}17 = 27{,}2 \geq 5$

On invalide tout de même l'item car les conditions de validité proposées utilisent les proportions observées p_C et p_D dans chaque échantillon au lieu de p.

Items B et C

Attention à ne pas confondre intervalle de pari et intervalle de confiance !

- Un intervalle de confiance peut se calculer lorsque l'on dispose des paramètres observés d'une loi dans un échantillon : on cherche alors à prédire l'intervalle dans lequel les paramètres théoriques de la loi en population générale peuvent se trouver (statistiques).
- Un intervalle de pari peut se calculer au contraire lorsque l'on dispose des paramètres théoriques d'une loi dans une population générale : on cherche alors à parier l'intervalle dans lequel les paramètres de la loi peuvent se trouver dans un échantillon observé (probabilités).

On dispose ici des proportions observées dans chacun des groupes de tardigrades et on cherche à déterminer les proportions théoriques : on calcule donc bien un intervalle de confiance.
On cherche donc un intervalle de confiance à 96%. D'après la table 2a, $\varepsilon_{4\%} = 2{,}054$. D'après le formulaire :

$$IC_{1-\alpha}(\pi) = p \pm \varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}}$$
$$IC_{96\%}(\pi_D) = p_B \pm 2{,}054\sqrt{\frac{p_D(1-p_D)}{n_D}}$$
$$= 0{,}86 \pm 2{,}054\sqrt{\frac{0{,}86\times 0{,}14}{160}}$$
$$\simeq [0{,}80; 0{,}92] \pm 0{,}01$$

✗ Item D → *En acceptant H_1 à tort, le laboratoire risque de commettre une erreur de type II ou ...*

Le risque de deuxième espèce est aussi appelé erreur de type II, mais il correspond à la probabilité de ne pas rejeter H_0 alors que H_1 est la bonne hypothèse. À l'inverse, le risque de première espèce correspond à la probabilité de rejeter H_0 alors qu'elle est vraie (donc accepter H_1 à tort).

	H_0 vraie	H_1 vraie
H_0 non rejetée Test non significatif	Hypothèse correcte $1-\alpha$	Risque de deuxième espèce β Erreur de type II Manque de puissance (faux négatif)
H_0 rejetée Test significatif	Risque de première espèce α Erreur de type I Faux positif	Hypothèse correcte $1-\beta$

✔ Item E → *La statistique du test est $|Z| \simeq 1{,}543$*

On a calculé la proportion commune aux deux échantillons p à item A.
D'après le formulaire :

$$p = \frac{n_C p_C + n_D p_D}{n_C + n_D} = \frac{120\times 0{,}79 + 160 \times 0{,}86}{120+160} = 0{,}83$$
$$Z = \frac{p_C - p_D}{\sqrt{p(1-p)(\frac{1}{n_C}+\frac{1}{n_D})}} = \frac{0{,}79-0{,}86}{\sqrt{0{,}83(1-0{,}83)(\frac{1}{120}+\frac{1}{160})}} \simeq -1{,}543$$

Réponses vraies : C et E

Deux autres échantillons de tardigrades (E et F) sont ensuite tirés au sort parmis les tardigrades vivant dans les hautes montagnes et on expose l'échantillon E à une dose plus faible de rayons UV que l'échantillon F. Au bout de trois heures, on observe le statut des différents tardigrades (vivants actifs, morts ou en dormance) et on construit le tableau de contingence suivant :

	Échantillon E	Échantillon F	Total
Tardigrades vivants actifs	80	70	150
Tardigrades morts	70	80	150
Tardigrades en dormance	50	150	200
Total	200	300	500

Question 13 Le laboratoire cherche alors à savoir s'il existe un lien entre la dose de rayonnements UV absorbée et le statut de survie des tardigrades. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. On peut réaliser un test du χ^2 d'indépendance à 2 degrés de liberté

B. On peut réaliser un test du χ^2 d'homogénéité à 6 degrés de liberté

C. La statistique de test est $K = 32{,}6$ à 0,1 près

D. Le degré de significativité du test est inférieur à 1%

E. Au risque $\alpha = 0{,}1\%$, il n'existe pas de lien entre la dose de rayonnements UV absorbée et le statut de survie

Question 13

Items A et B

Attention à ne pas lire les items trop vite ! Ici, on cherche bien à comparer des proportions entre deux échantillons par un test du χ^2.

- Le test d'homogénéité de comparaison de proportions se fait lorsque l'on dispose d'une seule variable que l'on étudie dans différents groupes (par exemple si on étudiait seulement la répartition de survie dans différents groupes).
- Le test d'indépendance de comparaison de proportions se fait lorsque l'on dispose de deux variables dont on souhaite étudier le lien entre elles.

On cherche ici à savoir s'il existe un lien entre dose absorbée (représentée par les échantillons) et statut de survie : on a donc bien affaire à deux variables dont on veut étudier le lien. On réalise donc un test du χ^2 d'indépendance (et non d'homogénéité) entre ces deux variables.
D'après le formulaire, $ddl = (l-1)(c-1)$ avec l et c le nombre de modalités respectif de chaque variable.
Ici, il existe $c = 2$ modalités pour la dose absorbée (2 échantillons donc 2 colonnes) et $l = 3$ modalités pour le statut (3 statuts possibles donc 3 lignes). Il y a donc $1 \times 2 = 2$ degrés de liberté au test (et non 6). Il faut également faire attention à ne pas prendre en compte les lignes et colonnes des totaux.

✔ Item C → *La statistique de test est $K = 32{,}6$ à 0,1 près*

Avant de réaliser le test, il faut veiller à déterminer les effectifs calculés (ou théoriques). Dans un test du χ^2, on calcule les effectifs par la formule $C_{ij} = \frac{m_j n_i}{n}$ avec m_j et n_i les effectifs totaux des lignes et colonnes respectives et n l'effectif total. On peut alors compléter le tableau suivant :

	Échantillon E	Échantillon F	Total
Tardigrades vivants actifs	$O_{1,1} = 80$ $C_{1,1} = 60$	$O_{2,1} = 70$ $C_{2,1} = 90$	150
Tardigrades morts	$O_{1,2} = 70$ $C_{1,2} = 60$	$O_{2,2} = 80$ $C_{2,2} = 90$	150
Tardigrades en dormance	$O_{1,3} = 50$ $C_{1,3} = 80$	$O_{2,3} = 150$ $C_{2,3} = 120$	200
Total	200	300	500

Tous les effectifs calculés sont supérieurs à 5 donc les conditions de validité du test sont vérifiées.
On peut désormais calculer la statistique de test. D'après le formulaire :

$$K = \sum_{i=1}^{l} \sum_{j=1}^{c} \frac{(C_{ij} - O_{ij})^2}{C_{ij}}$$
$$K = \frac{(60-80)^2}{60} + \frac{(90-70)^2}{90} + \frac{(60-70)^2}{60} + \frac{(90-80)^2}{90} + \frac{(80-50)^2}{80} + \frac{(120-150)^2}{120}$$
$$K \simeq 32{,}64 = 32{,}6 \pm 0{,}1$$

✔ Item D → *Le degré de significativité du test est inférieur à 1%*

On utilise la table 3 afin de déterminer le degré de significativité du test. On a calculé à l'item précédent que $K = 32{,}6$. On sait que le test a $\nu = 2$ degrés de libertés et on cherche à connaître la valeur de α à partir de laquelle $K < \chi^2_{\alpha,\nu}$.
D'après la table 3 : $\chi^2_{\alpha=0{,}01}(\nu = 2) = 9{,}210 < K$. Donc le degré de significativité du test est bien inférieur à 0,01, soit 1%.

✘ Item E → *Au risque $\alpha = 0{,}1\%$, il n'existe pas de lien entre la dose de rayonnements UV...*

On utilise à nouveau la table 3 afin de conclure sur les hypothèses du test. On sait que $K = 32{,}6$ et que $\chi^2_{\alpha=0{,}001}(\nu = 2) = 13{,}815 < K$.
On rejette H_0 lorsque $K > \chi^2_\alpha \nu$, ce qui est le cas ici. Dans un test d'indépendance entre deux variables, le rejet de l'hypothèse nulle correspond au fait que les deux variables ne sont pas indépendantes. Autrement dit, au risque $\alpha = 0{,}1\%$, il peut exister un lien entre la dose de rayonnements UV absorbée et le statut de survie.

Réponses vraies : A, C et D

Satisfaits de leurs résultats, les biologistes décident de se pencher sur le mécanisme permettant aux tardigrades de résister à la dessication. Pour cela, on réalise un dosage de TDP (protéines intrinsèquement désordonnées spécifiques des tardigrades synthétisées en cas de dessication) dans chacun des tardigrades afin de savoir s'il existe un lien entre résistance à la dessication et taux de TDP.

On note T le taux de TDP dans les tardigrades. On pose H_0, sous laquelle aucun tardigrade ne vient d'un milieu sec, que l'on décide de rejeter lorsque $T > 180\ UI \cdot L^{-1}$. On considère que sous l'hypothèse alternative, $T \sim \mathcal{U}([150; 500]\ UI \cdot L^{-1})$ et que sous l'hypothèse nulle, $T \sim \mathcal{U}([25; 200]\ UI \cdot L^{-1})$.

Question 14 Les biologistes décident alors de sélectionner $n = 200$ tardigrades de façon aléatoire afin d'y doser la TDP et déterminer s'ils viennent d'un milieu sec ou humide. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. La puissance du test serait plus grande si l'on décidait de rejeter H_0 lorsque $T > 160\ UI \cdot L^{-1}$
B. La puissance du test est de 91,4%
C. Le risque de première espèce de ce test est de 11,4%
D. On calcule la puissance grâce à la loi de T sous H_0
E. En affirmant qu'aucun tardigrade ne vient d'un milieu sec, les biologistes peuvent commettre une erreur de type II

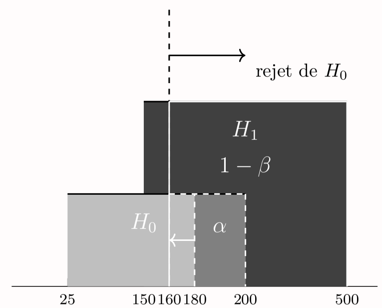
Question 14

✔ Item A → *La puissance du test serait plus grande si l'on décidait de rejeter H_0 lorsque $T >$ ...*

Avant de commencer, il faut bien noter les hypothèses du test afin de ne pas se tromper :

— L'hypothèse nulle H_0 correspond au fait qu'aucun tardigrade ne vient d'un milieu sec
— Par opposition (car si une hypothèse est vraie, l'autre est forcément fausse), l'hypothèse alternative H_1 correspond au fait qu'au moins un tardigrade vient d'un milieu sec. L'hypothèse alternative est souvent moins précise car on cherche à rejeter une hypothèse nulle bien définie.

On décide de diminuer le seuil à partir duquel on décide de rejeter H_0, ce qui a pour conséquence d'augmenter cette zone. La puissance d'un test se présentant comme la partie de la loi de distribution sous H_1 située dans la zone de rejet, celle-ci augmente donc lorsque le risque de première espèce α augmente. Sachez que la puissance augmente aussi lorsque l'on éloigne les zones de distribution des deux hypothèses.



? Items B et C

On utilise la fonction de densité de la loi uniforme. D'après le formulaire, on note a et b les bornes inférieure et supérieure de la loi uniforme de T sous H_0 et x le seuil de la zone de rejet de H_0. Le risque de première espèce se présente comme la partie de la loi de distribution de T sous H_0 située dans la zone de rejet. Celle-ci se situe dans la partie supérieure de la loi uniforme de T sous H_0 (on rejette H_0 si $T \in [180; 200]$), donc il faut modifier légèrement la formule afin de prendre cela en compte.

$$F(a \le x \le b) = \frac{b-x}{b-a}$$
$$\alpha = \mathrm{P}(T_{H_0} > 180) = \frac{180-200}{200-25} \simeq 0{,}114 \neq 0{,}886$$

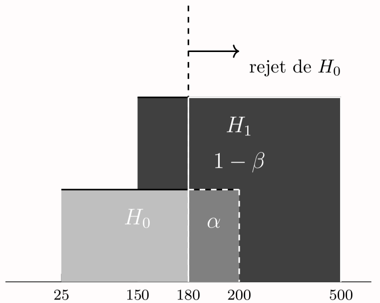
18

De la même manière, la puissance est égale à $1-\beta$ et correspond à la partie de la loi de distribution sous H_1 (et non H_0) située dans la zone de rejet. D'après le formulaire :

$$F(a \leq x \leq b) = \frac{x-a}{b-a}$$

$$\beta = \mathrm{P}(_{H_1} < 180) = \frac{180-150}{500-150}$$

$$1-\beta = 1 - \frac{30}{500-350} \simeq 0{,}914$$



Item D → *On calcule la puissance grâce à la loi de T sous H_0*

La puissance se calcule avec la formule $\mathcal{P} = 1-\beta$ et on sait que le risque de deuxième espèce β correspond à la probabilité de ne pas rejeter H_0 sachant que H_1 est vraie. La puissance se calcule donc avec la distribution suivie par la variable T sous H_1. Sur le diagramme représentant les différentes hypothèses, $1-\beta$ se situe du côté du seuil où l'on rejette H_0 puisque β correspond à la probabilité de non-rejet.

Item E → *En affirmant qu'aucun tardigrade ne vient d'un milieu sec, les biologistes peuvent ...*

Affirmer qu'aucun tardigrade ne vient d'un milieu sec revient à accepter l'hypothèse nulle : il s'agit d'une erreur de type I et non de type II.

Réponses vraies : A, B et C

Enfin, la capacité des tardigrades à survivre à des sécheresses extrêmes est également permise par la synthèse de la tréhalose, sucre remplaçant l'eau afin de protéger les cellules.

Le laboratoire regroupe tous les tardigrades qu'il possède les expose à un environnement très sec. Le laboratoire décide de classer les familles de tardigrades selon la proportion de tréhalose dans le corps en période de sécheresse, puis il note les effectifs de chaque catégorie dans ses échantillons afin de les comparer à la population générale. Il dresse le tableau suivant :

	Classe L	Classe M	Classe N	Classe O
Effectifs observés en laboratoire	$O_{1,1} = 140$	$O_{2,1} = 320$	$O_{3,1} = 280$	$O_{4,1} = 260$
Effectifs calculés (théoriques)	$C_{1,2} = 160$	$C_{2,2} = 300$	$C_{3,2} = 260$	$C_{4,2} = 280$

Question 15 Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

A. Les conditions de validité du théorème central limite sont vérifiées car $\forall\ i = 1,...,k\ C_i \geq 5$
B. Au risque $\alpha = 2\%$, on ne rejette pas l'hypothèse nulle
C. On réalise un test du χ^2 d'adéquation à un modèle théorique
D. On réalise un test z de comparaison de proportions
E. L'hypothèse alternative est $\pi_O \neq \pi_C$ pour au moins une classe

Question 15

Item A → *Les conditions de validité du théorème central limite sont vérifiées car $\forall\ i = ...$*

On cherche ici à comparer des proportions observées à un modèle théorique (population générale) par un test du χ^2. Pour pouvoir faire un test de comparaison de proportions à un modèle théorique, il faut que les effectifs calculés C_{ij} soient tous supérieurs à 5.

Item B → *Au risque $\alpha = 2\%$, on ne rejette pas l'hypothèse nulle*

On dispose ici des effectifs observés et calculés et on sait que les conditions de validité du test sont vérifiées : on doit donc réaliser un test du χ^2 à $(k-1)$ degrés de libertés, avec k le nombre de modalités de la variable *classe*.
D'après le formulaire :

$$K = \sum_{i=1}^{k} \frac{(C_{ij} - O_i)^2}{C_i}$$

$$K = \frac{(160-140)^2}{160} + \frac{(300-320)^2}{300} + \frac{(260-280)^2}{260} + \frac{(280-260)^2}{280} \simeq 6{,}80$$

De plus, comme on est dans le cas d'un test d'homogénéité , on a $\nu = (c-1) = 4-1 = 3$ D'après la table 3, $\chi^2_{\alpha=2\%}(\nu = 3) = 9{,}837 > K$
On rejette l'hypothèse nulle lorsque $K > \chi^2_\alpha(\nu)$ donc au risque $\alpha = 2\%$, elle n'est pas rejetée.

✔ **Item C** → *On réalise un test du χ^2 d'adéquation à un modèle théorique*

Exactement !
En effet, on a calculé l'effectif théorique en multipliant les proportions connues en population générale par la taille de notre échantillon (1000). On cherche à savoir si les tardigrades du laboratoire se comportent de la même façon que de la même façon que l'ensemble de la population des tardigrades.

✘ **Item D** → *On réalise un test z de comparaison de proportions*

On peut réaliser un test z ou test de l'écart-réduit lorsque l'on souhaite comparer des proportions entre deux échantillons entre lesquels une variable à deux modalités change de paramètres. Ici, on a réparti les effectifs selon quatre classes : la variable associée à la classe prend donc quatre modalités et il n'est pas possible de réaliser un test z.

✔ **Item E** → *L'hypothèse alternative est $\pi_O \neq \pi_C$ pour au moins une classe*

C'est tout à fait ça ! On cherche à démontrer par l'hypothèse nulle que pour toutes les modalités observées, les proportions théoriques sont identiques aux proportions observés. L'hypothèse alternative est donc le contraire de l'hypothèse nulle : il y a donc au moins une modalité pour laquelle l'effectif théorique diffère de l'effectif observé.
Bravo à toi d'être arrivé au bout de ce long exercice ! Essaie de refaire le sujet à la maison et de t'entraîner le plus possible pour ne pas retomber dans les mêmes pièges le jour de l'exam. Courage !

Réponses vraies : A, B, C et E