

UE9 – Mathématiques - Biostatistiques

Annales classées corrigées

Introduction à la statistique – Estimation –
Intervalle de Confiance (IC)

SUJET

2023

Question n°10 On souhaite estimer la proportion d'interventions au bloc opératoire qui concernent une opération lourde du rachis, qui nécessite d'occuper le bloc pendant une durée très longue. On tire au sort un échantillon de 1300 interventions dans des blocs opératoires d'hôpitaux similaires à l'hôpital parisien (interventions que l'on considérera comme indépendantes). On trouve 13 opérations lourdes du rachis. On s'intéresse à l'intervalle de confiance à 95 % de la proportion d'interventions qui concernent une opération lourde du rachis. Parmi les propositions suivantes, **laquelle** est exacte ?

- A. Cet intervalle de confiance est [5 %; 15 %] à 1 % près
- B. Cet intervalle de confiance est [0,95 %; 1,05 %] à 0,05 % près
- C. Cet intervalle de confiance est [0,8 %; 1,2 %] à 0,1 % près
- D. Cet intervalle de confiance est [0,5 %; 1,5 %] à 0,1 % près
- E. On ne peut pas le calculer parce que le théorème limite central ne s'applique pas

2022

Question n°10 Un avion en provenance de Singapour (pays vert) atterrit à l'aéroport. La procédure de test aléatoire conduit à tester 36 passagers pour lesquels on observe une moyenne de taux d'antigène de 5 UI avec un écart-type de ce taux de 2,5 UI.

Parmi les propositions suivantes, **laquelle** correspond à l'intervalle de confiance à 95% de la moyenne théorique du taux d'antigène des passagers en provenance de Singapour (à 0,1 UI près)?

- A. [4,9 ; 5,1] UI
- B. [4,9 ; 5,9] UI
- C. [4,2 ; 5,8] UI
- D. [4,5 ; 5,5] UI
- E. [0,0 ; 10,0] UI

2021

Question n°10 On constitue un échantillon représentatif de 49 personnes atteintes d'une nouvelle forme de M. Pour chacun de ces patients, on dose la concentration plasmatique de la protéine utilisée dans le test diagnostic. On observe une moyenne de 4 et un écart-type de 1,4 unités. Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s) ?

- A. L'intervalle de pari à 95% de la moyenne de la concentration plasmatique de la protéine dans cette population est [3,60 ; 4,40] (à 0,01 près)
- B. L'intervalle de confiance à 95% de la moyenne de la concentration plasmatique de la protéine dans cette population est [3,60 ; 4,40] (à 0,01 près)
- C. L'intervalle de confiance à 99% de la moyenne de la concentration plasmatique de la protéine dans cette population est [3,48 ; 4,52] (à 0,01 près)
- D. Il n'est pas nécessaire que la concentration plasmatique de la protéine suive une loi normale pour pouvoir calculer l'intervalle de confiance de la moyenne
- E. Si l'on inclut 98 personnes, la largeur de l'intervalle de confiance à 95% sera divisée par deux

2020

Question n°7 Voyant que le nombre de recours au médiateur dans le service B ne correspond pas aux notations que le responsable reçoit, il mène une enquête auprès de 400 patients sortant du service B et leur demande leurs notes. Il observe alors une moyenne des notes de 6 et une variance des notes de 4.

Cochez la (ou les) réponse(s) exacte(s).

- A. Il peut calculer l'intervalle de confiance à 95% de la moyenne des notes dans le service B car la loi suivie par X_B est normale
- B. L'intervalle de confiance à 95% de la moyenne est $[5,80;6,20]$ à 0,01 près
- C. L'intervalle de confiance à 90% de la moyenne est $[5,84;6,16]$ à 0,01 près
- D. $P(X_B < 2) = 2,5\%$
- E. En incluant 1600 patients, il diminuerait par deux la largeur de l'intervalle de confiance à 95% et de l'intervalle de confiance à 90%

2019

Dans l'hôpital de Champignac, où le logiciel est installé, tous les patients venant aux urgences ont une radio de thorax. Malheureusement, le responsable des urgences ne connaît pas la prévalence de la pneumopathie aux urgences et décide de l'estimer pour évaluer au mieux l'intérêt du logiciel.

Question n°2 Combien faut-il inclure de patients au minimum pour s'assurer que la largeur de l'intervalle de confiance à 95% de l'estimation soit inférieure à 0,01 si on considère que la prévalence de la pneumopathie aux urgences vaut entre 10% et 20%. Cochez la (ou les) réponse(s) exacte(s).

- A. On applique le théorème central limite pour faire ce calcul
- B. Il faut environ 25 500 patients (à 1 000 patients près)
- C. Il faut environ 14 500 patients (à 1 000 patients près)
- D. On suppose que les patients sont indépendants les uns des autres concernant la pneumopathie pour faire ce calcul
- E. L'erreur β ou erreur de type II est de 5%

2018**Question n°12**

La PMI de Champignac décide de planifier une étude pour estimer, dans 2 ans, la proportion d'enfants ayant effectivement été complètement vaccinés. Elle prévoit de tirer au sort un échantillon d'enfants de taille n pour estimer le pourcentage d'enfants ayant reçu les 11 vaccins et son intervalle de confiance à 95%. Elle souhaite que la largeur de cet intervalle soit inférieure à 5%, suppose une proportion attendue de 0,90 et cherche à calculer une valeur minimale de n . Que peut-on dire?

Cochez LA bonne réponse.

- A. Il faudrait inclure au moins $n=144$ patients.
- B. Si finalement la proportion observée est de 0,95, la largeur de l'intervalle de confiance à 95% sera plus petite à n égal.
- C. Si finalement la proportion observée est de 0,85, la largeur de l'intervalle de confiance à 95% sera la même à n égal.
- D. Si finalement 100% des enfants reçoivent les 11 vaccins, l'intervalle de confiance à 95% sera symétrique.
- E. Il n'est pas nécessaire d'appliquer le théorème limite central pour effectuer le calcul de la valeur minimale de n .

2016

Question n°5 Le médecin aimerait estimer la prévalence de la résistance. Il constitue un échantillon de prélèvements bactériologiques de méningocoques et vérifie la résistance à l'antibiothérapie in vitro. Il obtient une estimation de la proportion de résistance de 0,2 et avec une largeur de l'intervalle de confiance à 95% de 0,2. Cochez LA réponse exacte.

- A. Il a étudié la résistance sur 16 prélèvements.
- B. Il a étudié la résistance sur 32 prélèvements.
- C. Il a étudié la résistance sur 50 prélèvements.
- D. Il a étudié la résistance sur 64 prélèvements.
- E. Il a étudié la résistance sur 128 prélèvements.

2015

QCM 8 – Sur les 10 communes wallonnes choisies, il en observe 6 sans médecin. Cochez la (ou les) réponse(s) exacte(s).

- A - Il peut estimer que 60% des communes de Wallonie manquent de médecins
- B - Il peut affirmer que 60% des communes de Wallonie manquent de médecins
- C - Il peut dire que la proportion de communes wallonnes sans médecin est vraisemblablement comprise entre 30% et 90% (au risque 5%)
- D - Il peut parier que la proportion de communes wallonnes sans médecin sur son échantillon de 10 est comprise entre 30% et 90% (au risque 5%)
- E - Le théorème de la limite centrale ne s'applique pas

2014

Une étude publiée récemment a été réalisée en Grande Bretagne pour apporter des informations épidémiologiques sur la violence. Au total, le groupe interrogé comprenait 8000 individus (4000 hommes et 4000 femmes), dont 1000 (750 hommes et 250 femmes) ont rapporté un comportement violent. Deux tiers des sujets violents avaient des troubles de la personnalité contre un tiers des non violents. *Les données de l'étude ont été légèrement modifiées pour simplifier les calculs.*

QCM 12 On souhaite alors étudier si la proportion de sujets violents est ou non identique en France à celle notée dans cette étude anglaise. Sur les 100 individus français interrogés, on observe 10 individus « violents ». Cochez la ou les propositions exactes.

- A - On peut dire que la proportion de sujets violents en France est de 10%
- B - L'intervalle de pari à 95% de la proportion de sujets violents en France est compris entre 4 et 16%
- C - L'intervalle de confiance à 95% de la proportion de sujets violents en France est compris entre 4 et 16%
- D - Un intervalle de pari à 90% est toujours plus étroit qu'un intervalle de pari à 95% de la même variable pour une taille d'échantillon donnée
- E - On peut réaliser un test de l'écart-réduit

2013

QCM3- Pour éviter les problèmes liés aux naissances par le siège, il décide de proposer un traitement aux femmes enceintes dont le fœtus est en siège à la 33^{ème} semaine de gestation, afin qu'il se retourne avant l'accouchement. Il apprend que les médecins chinois chauffent une herbe à proximité de la peau du point d'acupuncture BL67 (« Zhiyin ») près du 5^{ème} orteil, ce qui conduit au retournement spontané du fœtus en moins de deux semaines dans 75% des cas. Il décide d'évaluer ce traitement sur un échantillon de $n = 27$ femmes. En supposant que ces données sont valides, cochez la ou les propositions exactes.

- A. Il peut parier à 95% que le nombre de fœtus qui vont se retourner sera compris entre 7 et 11 (à 1 près)
- B. Il peut parier à 95% que le nombre de fœtus qui vont se retourner sera compris entre 16 et 25 (à 1 près)
- C. Il s'agit d'un intervalle de pari
- D. Il s'agit d'un intervalle de confiance
- E. Le théorème de la limite centrale ne peut s'appliquer car $n = 27$

Le score d'Apgar a été développé en 1952 par une médecin américaine Virginia Apgar pour mesurer l'état de santé du nouveau-né.

QCM7- Le médecin mesure les valeurs des scores Apgar à 5 minutes de vie chez les 27 nouveaux-nés de son échantillon. Il observe une moyenne de 9 et un écart-type de 0,5. Cochez LA proposition exacte.

- A. Il peut affirmer que l'APGAR d'un nouveau-né est de 9
- B. Il peut affirmer que l'APGAR moyen de tous les nouveaux-nés est de 9
- C. Il peut dire qu'il y a 95 chances sur 100 pour que l'APGAR moyen de tous les nouveaux-nés soit dans l'intervalle 8-10 (à 0,1 près)
- D. Il peut dire qu'il y a 95 chances sur 100 pour que l'APGAR moyen de tous les nouveaux-nés soit dans l'intervalle 8,8-9,2 (à 0,1 près)
- E. Il ne peut rien dire sur cet intervalle car le théorème de la limite centrale ne s'applique pas

2012

QCM 7 - Il effectue un échantillon de 36 coureurs, effectue son test et en déclare 13 dopés (36%).
Cochez la ou les propositions exactes.

- A. Il peut affirmer que la proportion de dopés dans le tour est de 36%
- B. Il peut estimer que la proportion de dopés dans le tour est de $36\% \pm 8\%$ avec un niveau de confiance de 95%
- C. Il peut estimer que la proportion de dopés dans le tour est de $36\% \pm 16\%$ avec un niveau de confiance de 95%
- D. Il a pu appliquer le théorème de la limite centrale car $n \geq 30$
- E. Il a pu appliquer le théorème de la limite centrale car $0,2 \times 36 \geq 5$

2011

QCM 7 Le médecin veut tester si son hypothèse sur le temps moyen d'une souris quelconque à trouver le fromage est correcte en posant $H_0 : \mu = 5$. Il observe sur les 100 souris un temps moyen de 6,5 secondes à trouver le fromage, avec un écart-type de 2,5. Parmi les réponses suivantes, dire la(les) proposition(s) vraie(s).

- a - Il peut affirmer qu'en moyenne les 100 souris ont mis 6,5 secondes à trouver le fromage
- b - Il peut affirmer qu'en moyenne une souris quelconque met entre 1,5 et 11,5 secondes à trouver le fromage avec un niveau de confiance de 95%
- c - Il peut affirmer qu'en moyenne une souris quelconque met entre 6 et 7 secondes à trouver le fromage avec un niveau de confiance de 95%
- d - Pour faire le test, il peut construire sous H_0 un intervalle de pari de la moyenne observée chez les 100 souris et regarder si 6,5 appartient à celui-ci
- e - Pour faire le test, il peut construire un intervalle de confiance de la moyenne μ et regarder si 6,5, appartient à celui-ci

2009

La résistance aux traitements du paludisme est un problème grave en Afrique. Une large étude faite en 2007 chez des patients infectés par le paludisme rapporte que 50% des patients étaient résistants à la chloroquine et 10% étaient multi-résistants (c'est à dire résistants à au moins 3 molécules). On effectue à nouveau une étude en 2008 sur un échantillon de n patients atteints de paludisme. Pour chaque patient, on mesure sur un prélèvement parasitaire la résistance à la chloroquine (oui/non), ainsi que la multi-résistance (oui/non).

QCM 8 Sur le n -échantillon de patients étudiés en 2008, on appelle N_n la variable aléatoire "Nombre de patients résistants à la chloroquine". On définit les variables suivantes :

$$P_n = N_n/n \text{ et } Q_n = N_n/(n-1)$$

On notera $E(P_n)$ et $E(Q_n)$ l'espérance respectivement de P_n et Q_n . Parmi les propositions suivantes, dire **LA** proposition vraie.

- a - P_n est une estimation de la proportion de résistance à la chloroquine en 2008
- b - Q_n est une estimation de la proportion de résistance à la chloroquine en 2008
- c - P_n est un estimateur sans biais de la proportion de résistance à la chloroquine en 2008
- d - $E(Q_n) = E(P_n)$
- e - $E(Q_n) < E(P_n)$

2007

On s'intéresse aux septicémies (infections bactériennes sanguines) à *E. Coli*, à leur sévérité et à leur traitement. Les bactéries *E. Coli* appartiennent à quatre groupes phylogénétiques A, B1, D, B2. Ce sont usuellement des bactéries commensales de l'homme (hôtes usuels de l'intestin) qui peuvent devenir pathogènes, provoquant des septicémies parfois mortelles.

QCM 10 - Dans un échantillon de 100 patients ayant une septicémie non traitée par une antibiothérapie adaptée, on observe 50 décès.

Parmi les questions suivantes, **laquelle est vraie ?**

- a - L'intervalle de pari à 95% de la proportion observée de décès pour des septicémies non traitées par une antibiothérapie adaptée dans un échantillon de taille 100 est de $[0,4 ; 0,6]$
- b - L'intervalle de confiance à 95% de la proportion de décès pour des septicémies non traitées par une antibiothérapie adaptée est de $[0,5 ; 0,7]$
- c - Les conditions de validité du test pour le calcul de l'intervalle de confiance sont vérifiées car $0,5 \times 100$ est supérieur à 5
- d - L'intervalle de confiance à 99% de la proportion de décès pour des septicémies non traitées par une antibiothérapie adaptée est de $[0,4 ; 0,6]$
- e - L'intervalle de confiance à 95% de la proportion de décès pour des septicémies non traitées par une antibiothérapie adaptée est de $[0,4 ; 0,6]$

Vous décidez d'utiliser la température corporelle, mesurée en degrés Celsius ($^{\circ}\text{C}$), des patients atteints de grippe sévère, 3 jours après le début de leur maladie, pour détecter les malades qui s'aggravent. **On supposera par la suite que cette mesure de la température est une variable Gaussienne.** On sait que, chez les malades qui ne s'aggravent pas (neuf sur 10), la température fluctue autour de sa moyenne μ_1 dans 95% des cas entre 35,5 et 37,5 et que chez les malades aggravés, la température fluctue autour de sa moyenne $\mu_2=39$ avec un écart-type $\sigma_2=1$.

L'Oseltamivir, un antiviral efficace contre les virus de la grippe, est commercialisé sous le nom de Tamiflu. **On considère que, sous Tamiflu, la température des malades atteints de grippe sévère qui s'aggravent suit toujours une loi Normale, d'écart-type 1°C .** On voudrait tester l'hypothèse nulle H_0 que le Tamiflu ne modifie pas la température moyenne contre l'hypothèse alternative H_1 qu'elle est diminuée en moyenne de 1°C .

QCM 12 On appelle X la température des patients qui s'aggravent. On s'intéresse au carré de la température, c'est à dire à la variable aléatoire $Y = X^2$. On mesure la température sur un échantillon de n patients. On appelle M_x la moyenne expérimentale des températures et M_y la moyenne

expérimentale des carrés des températures. On définit $N = \frac{1}{n} \sum x_i^2$ et $T = \left(\frac{1}{n} \sum x_i \right)^2$ Parmi les

propositions suivantes, **laquelle** est vraie ?

- a - $N = M_x^2$
- b - $T = M_y$
- c - $E(T) = E(X)^2$
- d - N est un estimateur sans biais de $E(Y)$
- e - T est un estimateur sans biais de $E(Y)$

2023 Session 2

Exercice 3

L'activité physique contribue à la prévention des maladies cardiovasculaires.

Question n°9 On cherche à estimer la durée hebdomadaire d'activité physique dans la population. On constitue par tirage au sort un échantillon de $n = 200$ personnes dans la population générale. Si on note x_i la durée hebdomadaire d'activité physique en heures de l'individu i dans l'échantillon, on obtient les résultats suivants : $\sum_{i=1}^n x_i = 1500$ h et $\sum_{i=1}^n x_i^2 = 26175$ h². Parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s)?

- A. On estime la variance de la durée hebdomadaire d'activité physique dans la population à 75 h² (à 0,1 h² près)
- B. On estime la moyenne de la durée hebdomadaire d'activité physique dans la population à 7,5 h (à 0,1 h près)
- C. On estime l'écart-type de la durée hebdomadaire d'activité physique dans la population à 0,6 h (à 0,1 h près)
- D. L'intervalle de pari à 95% de la durée hebdomadaire d'activité physique moyenne dans l'échantillon est [6,3; 8,7] heures (à 0,1 h près)
- E. L'intervalle de confiance à 95 % de la durée hebdomadaire d'activité physique moyenne dans la population est [6,3 ; 8,7] heures (à 0,1 h près)

Question n°8 Dans un échantillon de taille $n = 64$, représentatif d'une autre population, on observe $\sum x_i = 960$ U/L et $\sum x_i^2 = 15408$ (U/L)². Parmi les propositions suivantes, laquelle correspond à l'intervalle de confiance à 95% de la moyenne théorique de concentration de la protéine P dans cette population (à 0,1 près)?

- A. [14 ; 18]
- B. [12 ; 17]
- C. [11 ; 19]
- D. [14; 16]
- E. [15; 17]

Exercice 3

Une des possibles complications médicales suite à l'infection d'un adulte par le virus de la rougeole est l'apparition d'une pneumonie rougeoleuse qui impose une hospitalisation en réanimation.

On a interrogé l'ensemble des services de réanimation de France et tiré au sort 200 patients adultes hospitalisés pour pneumonie suite à une rougeole (cas) et 300 patients adultes, hospitalisés en réanimation pour une pneumonie d'une autre cause (témoins).

On souhaite dans un premier temps évaluer si la proportion de patients vaccinés diffère entre les cas et les témoins. On interroge pour cela les 500 patients sur leur éventuelle vaccination, que l'on catégorise en trois classes :

- correctement vacciné (2 doses reçues),
- insuffisamment vacciné (1 seule dose reçue)
- et non vacciné (aucune vaccination).

Les résultats obtenus sont présentés dans le tableau suivant :

	Non vaccinés	Insuffisamment vaccinés	Correctement vaccinés
Cas	50	50	100
Témoins	50	0	250

Question n°12 Parmi les propositions suivantes, **laquelle** correspond à l'intervalle de confiance à 95% de la proportion théorique de non vaccinés dans la population des cas (à 0,02 près)?

- A. [0,10 ; 0,40]
- B. [0,19 ; 0,31]
- C. [0,25 ; 0,30]
- D. [0,32 ; 0,35]
- E. [0,38 ; 0,43]

2021 Session 2

On considère une population potentiellement plus susceptible d'être atteinte de I . Dans un échantillon représentatif de cette population comprenant 60 personnes, on en trouve 12 atteintes de I . On

pourra approximer $\sqrt{\frac{0,2 \times 0,8}{60}}$ par 0,05.

Question n°10 Concernant l'intervalle de confiance de la proportion théorique de patients atteints de I dans cette nouvelle population, parmi les propositions suivantes, laquelle est (ou lesquelles sont) exacte(s)?

- A. L'intervalle de confiance à 95% vaut $[0,10; 0,30]$ à 0,02 près
- B. L'intervalle de confiance à 99% vaut $[0,07; 0,33]$ à 0,02 près
- C. L'intervalle de confiance à 99% est plus large que l'intervalle de confiance à 95%
- D. En augmentant la taille d'échantillon, on augmente la largeur de l'intervalle de confiance (en supposant la proportion observée identique)
- E. La largeur de l'intervalle de confiance à 90% est 0,16 à 0,02 près