Contrôle Terminal – 09/05/2025 – 2h00
Statistiques – L1 de psychologie

Réponses

Veuillez rendre ce sujet et votre copie.

Numéro d'anonymat :

Le formulaire et la calculatrice sont autorisés. Merci d'indiquer dans la case Numéro d'anonymat*, ci-dessus, un numéro que vous reporterez aussi sur votre copie. Vous rendrez l'énoncé et votre copie, et pouvez soit répondre sur l'énoncé, soit détailler certaines questions sur la copie si vous avez besoin de plus de place. Le soin de la rédaction entrera en compte dans la notation mais dans les questions où des détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.*

Exercice 1 : Précarité et isolement social

Chaque année, le crédoc (le Centre de recherche pour l'étude et l'observation des conditions de vie) interroge des milliers de français·es, au moyen d'un questionnaire qui comporte des questions sur de nombreux aspects de leurs conditions de vie, parmi lesquels leur niveau de revenu et la solitude qu'ils ressentent. L'analyse des questionnaires indique que 12% des français·es se sentent isolé·e·s.

Dans cet exercice, on s'appuie sur ces questionnaires pour déterminer si les personnes ayant les plus bas revenus sont plus frappées par la solitude que les personnes ayant les revenus les plus élevés.

1. On regarde les réponses de 26 français·es choisi·e·s au hasard parmi celles et ceux qui ont rempli le questionnaire. On constate que 5 d'entre elles/eux souffrent d'isolement. Peut-on en déduire une estimation de la proportion de personnes isolée en France, sous la forme d'un intervalle de confiance correspondant à la confiance 99% ?

Non, on ne peut pas, car la procédure du formulaire nécessite que $n\,p_e\,(1-p_e)$ soit supérieur à 10. Or $n\,p_e\,(1-p_e) = 26 \times \frac{5}{26}\left(1-\frac{5}{26}\right) \simeq 4{,}038 < 10$.

2. On choisit désormais 153 personnes qui ont de hauts revenus (on les choisit au hasard parmi toutes les personnes à haut revenu qui ont été sondées par le crédoc). On constate que 15 d'entre elles se déclarent isolées. En déduire une estimation de la proportion de personnes qui se sentent isolées, parmi l'ensemble des français·es à haut revenu. *On vous demande de réaliser un intervalle de confiance, pour la confiance* 99%.

On a $n\,p_e(1-p_e) = 153 \times \frac{15}{153}\left(1-\frac{15}{153}\right) \simeq 13{,}529 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion de personnes souffrant de solitude, parmi les français·es à haut revenu.
Le formulaire indique que pour la confiance $0{,}99$ on a $z_\alpha \simeq 2{,}576$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 2{,}576 \times \sqrt{\frac{\frac{15}{153}(1-\frac{15}{153})}{153}} \simeq 0{,}0619$ et $p_e - a_\alpha \simeq \frac{15}{153} - 0{,}061\,9 \simeq 0{,}036\,14$ et $p_e + a_\alpha \simeq \frac{15}{153} + 0{,}061\,9 \simeq 0{,}159\,9$.
On estime donc que cette proportion est dans l'intervalle $[0{,}0361; 0{,}1599]$ avec la confiance $c = 0{,}99$.

3. On choisit ensuite 129 personnes qui ont de bas revenus (parmi tous les français·es à bas revenus interrogés par le crédoc), et on constate que 19 d'entre elles se déclarent isolées. En déduire une estimation de la proportion de personnes qui se sentent isolées, parmi l'ensemble des français·es à bas revenu. *On vous demande de réaliser un intervalle de confiance, pour la confiance* 99%.

On a $n\,p_e(1-p_e) = 129 \times \frac{19}{129}\left(1-\frac{19}{129}\right) \simeq 16{,}202 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion de personnes souffrant de solitude, parmi les français·es à bas revenu.
Le formulaire indique que pour la confiance 0,99 on a $z_\alpha \simeq 2{,}576$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 2{,}576 \times \sqrt{\frac{\frac{19}{129}(1-\frac{19}{129})}{129}} \simeq 0{,}0804$ et $p_e - a_\alpha \simeq \frac{19}{129} - 0{,}080\,4 \simeq 0{,}066\,89$ et $p_e + a_\alpha \simeq \frac{19}{129} + 0{,}080\,4 \simeq 0{,}227\,7$.
On estime donc que cette proportion est dans l'intervalle $[0{,}0669; 0{,}2277]$ avec la confiance $c = 0{,}99$.

4. Pour chacune de ces affirmations, peut-on conclure, avec la confiance 99% qu'elle soit vraie, ou fausse ? *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse à partir des estimations des questions précédentes.*

Parmi les Français·es à haut revenu, il y en a moins que 12% qui sont isolé·e·s.
☐ Vraie ☐ Fausse ☒ On ne peut rien affirmer
En effet, la valeur 12% est à l'intérieur de l'intervalle de la question 2, cet intervalle ne dit pas si la proportion est plus grande ou plus petite que 12%.

Parmi les Français·es à bas revenu, il y en a moins que 12% qui sont isolé·e·s.
☐ Vraie ☐ Fausse ☒ On ne peut rien affirmer
En effet, la valeur 12% est à l'intérieur de l'intervalle de la question 3, cet intervalle ne dit pas si la proportion est plus grande ou plus petite que 12%.

Il y avait une plus grande proportion de personnes isolées parmi les français·es à bas revenu que parmi les français·es à haut revenu.
☐ Vraie ☐ Fausse ☒ On ne peut rien affirmer
En effet, les intervalles calculés en 2 et 3 se chevauchent : ils ne permettent pas de dire (avec la confiance 99%) laquelle des deux proportions est la plus grande.

5. Si on avait voulu estimer à 2% près la proportion de personnes isolées parmi les français·es à bas revenu, quelle taille d'échantillon aurait-il fallu prendre ?

Pour avoir une précision 0,02 avec la confiance 0,99, il faut $n > z_\alpha^2\frac{p_e(1-p_e)}{h^2} \simeq 2{,}576^2\,\frac{\frac{19}{129}(1-\frac{19}{129})}{0{,}02^2} \simeq 2083$

Exercice 2 : Manuels scolaire et éco-anxiété

Des chercheur·e·s travaillent sur l'*écoanxiété* et ont analysé des manuels scolaires, pour voir si leur traitement du changement climatique influençait la perception par les étudiant·e·s de la situation climatique et des perspectives.

Dans cet exercice on se focalise plus spécifiquement sur les manuels édités pour les étudiant·e·s en licence de biologie. Depuis les années 1990 il est systématique que ces manuels contiennent une partie sur le changement climatique, et on regarde dans cet exercice si cette partie dédiée au changement climatique présente plutôt un constat ou une perspective avec des solutions.

Pour chaque manuel de biologie, on calcule un score que l'on note X et qui indique si le manuel se concentre sur des solutions au changement climatique (si X est proche de zéro c'est que le livre n'en parle presque pas, s'il est proche de cent c'est que le livre se focalise vraiment sur ces solutions). *On supposera que X suit une loi normale.*

1. On considère tout d'abord un échantillon de 100 livres édités entre 2010 et 2019. Les scores X sont indiqués ci-dessous

Score X	[0 ; 1[	[1 ; 2[	[2 ; 3[	[3 ; 4[	[4 ; 5[	[5 ; 6[	[6 ; 7[
Effectif	1	7	14	28	27	15	8

On a laissé des lignes vides au cas où vous en ayez besoin pour des calculs intermédiaires.

(a) Au sein de cet échantillon, quelle est la proportion de manuels pour lesquels X est inférieur à 3.

Au sein de cet échantillon, il y a $1 + 7 + 14 = 22$ manuels dont le score est inférieur à 3. Cela correspond à une proportion $\frac{22}{100} = 0{,}22$.

(b) Quelles sont la moyenne et l'écart type de X ?

moyenne : $m(X) = \frac{\sum c_i n_i}{n} = \frac{0{,}5\times1+1{,}5\times7+2{,}5\times14+\cdots+6{,}5\times8}{100} = \frac{400}{100} = 4$

$m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{0{,}5^2\times1+1{,}5^2\times7+2{,}5^2\times14+\cdots+6{,}5^2\times8}{100} = \frac{1\,785}{100}$

$Var(X) = m(X^2) - m(X)^2 = \frac{1\,785}{100} - \left(\frac{400}{100}\right)^2 = 1{,}85$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}36$

(c) À partir de cet échantillon, estimer la moyenne du score X parmi les manuels de biologie édités entre 2010 et 2019. *On demande de déterminer un intervalle de confiance, pour la confiance 95%.*

Comme $n = 100 > 30$, le formulaire indique que pour la confiance 0,95 on a $z_\alpha \simeq 1{,}96$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}96 \times \frac{1{,}36}{\sqrt{100-1}} \simeq 0{,}2679$

On estime donc que la moyenne de ce score est dans l'intervalle $[4 - 0{,}267\,9\,;\, 4 + 0{,}267\,9] \simeq [3{,}73\,;\, 4{,}27]$, avec la confiance $c = 0{,}95$.

2. On considère ensuite un échantillon de 8 livres édités entre 1990 et 1999. Les scores X sont indiqués ci contre :

15	12	9	22	18	19	15	6

   (a) Quelles sont la moyenne et l'écart type de X au sein de cet échantillon ?

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{15+12+9+\cdots+6}{8} = \frac{116}{8} = 14{,}5$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{15^2+12^2+9^2+\cdots+6^2}{8} = \frac{1\,880}{8}$
$Var(X) = m(X^2) - m(X)^2 = \frac{1\,880}{8} - \left(\frac{116}{8}\right)^2 = 24{,}75$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 4{,}97$

   (b) À partir de cet échantillon, estimer la moyenne du score X parmi les manuels de biologie édités entre 1990 et 1999. *On demande de déterminer un intervalle de confiance, pour la confiance 95%.*

Comme $n = 8 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n-1 = 7$ degrés de liberté (ddl)
On lit $t_\alpha \simeq 2{,}3646$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2{,}3646 \times \frac{4{,}97}{\sqrt{8-1}} \simeq 4{,}441\,9$.
On estime donc que la moyenne de ce score est dans l'intervalle $[14{,}5 - 4{,}441\,9\,;\,14{,}5 + 4{,}441\,9] \simeq [10{,}06\,;\,18{,}94]$ avec la confiance $c = 0{,}95$

   (c) Estimer de même l'écart type de ce score X parmi les manuels de biologie édités entre 1990 et 1999. *On demande de déterminer un intervalle de confiance, pour la confiance* 95%.

On lit dans la table du χ^2 à 7 degrés de liberté la valeur $x_1 = 1{,}69$ correspondant à $q = 0{,}025$ et la valeur $x_2 = 16{,}01$ correspondant à $p = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 4{,}97\sqrt{\frac{8}{16{,}01}} \simeq 3{,}51$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 4{,}97\sqrt{\frac{8}{1{,}69}} \simeq 10{,}8$
On estime donc que σ est dans l'intervalle $[3{,}51; 10{,}8]$ avec la confiance $c = 0{,}95$.

3. Pour chacune des affirmations ci-dessous, peut-on conclure, avec la confiance 95% qu'elle soit vraie, ou fausse ? Justifiez votre réponse en vous appuyant sur les intervalles de confiance calculés ci-avant. *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse à partir des estimations des questions précédentes.*

Sur la période 1990-1999, l'écart type du score X (parmi les manuels de biologie de licence) était plus grand que 1.
☒ Vraie ☐ Fausse ☐ On ne peut rien affirmer
En effet, la valeur 1 est en dehors de l'intervalle de la question 2c, dont toutes les valeurs sont plus grandes que 1.

La moyenne du score X est plus élevée pour les manuels de la décennie 2010-2019 que pour les manuels de la décennie 1990-1999.
☐ Vraie ☒ Fausse ☐ On ne peut rien affirmer
En effet, les intervalles calculés en 1c et 2b ne se chevauchent pas, et permettent d'affirmer que le score moyen en 2010-2019 (question 1c) est plus petit que le score moyen en 1990-1999 (question 2b).

Exercice 3 : Apprentissage du calcul en primaire

On cherche à comparer deux méthodes pour enseigner le calcul (addition et soustractions) en CP, que l'on désigne par "*méthode A*" et "*méthode B*" : la méthode A consiste à apprendre à faire des opérations « *avec les doigts* » avant de manipuler symboliquement des chiffres, alors que la méthode B procède directement aux manipulation symboliques avec des chiffres.

1. Un école de primaire compte 66 élèves en CP, dont la moitié (c'est à dire 33 élèves) ont appris à compter selon la méthode A et l'autre moitié (c'est à dire 33 élèves) selon la méthode B. On choisit au hasard un échantillon de 65 élèves de CP de cette école (on n'autorise pas à piocher deux fois le même élève), et on désigne par X le nombre d'élèves, au sein de l'échantillon qui ont appris à compter selon la méthode A.

   La variable aléatoire X suit elle une loi binomiale ? *Justifier pourquoi, et si vous répondez qu'elle suit une loi binomiale, précisez laquelle en indiquant les valeurs de n et p*

   ☐ X suit la loi $\mathcal{B}(\ ;\ )$ ☐ X suit presque la loi $\mathcal{B}(\ ;\ )$ ☒ X ne suit pas une loi binomiale.
   En effet, le tirage est sans remise (on n'autorise pas à piocher deux fois le même élève). Pour approcher par une loi binomiale il aurait fallu que la taille de l'échantillon soit beaucoup plus petite (au moins dix fois plus petite) que la taille de la population parmi laquelle on choisit les individus au hasard. Ici la taille de l'échantillon est 65 et la taille de la population est 66 donc cette condition n'est pas satisfaite.
   Remarque : on pouvait calculer explicitement les probabilités : parmi les 66 élèves de primaire on en choisit 65, donc il n'y en a qu'un qui n'est pas d'en l'échantillon. Il y a une chance sur deux que cet élèves ait appris avec la méthode A (et dans ce cas $X = 32$), et une chance sur deux qu'il ait appris avec la méthode B (et dans ce cas $X = 33$), donc la loi de X est donnée par $\mathbb{P}[X = 32] = 0{,}5$ et $\mathbb{P}[X = 33] = 0{,}5$.

2. On considère désormais les 3 192 élèves de CP qui habitent dans une même agglomération. Parmi eux, il y en a la moitié (c'est à dire 1 596 élèves) qui ont appris à compter selon la méthode A et l'autre moitié (c'est à dire 1 596 élèves) selon la méthode B. On choisit au hasard un échantillon de 65 élèves de CP de cette agglomération (on n'autorise pas à piocher deux fois le même élève), et on désigne par X le nombre d'élèves, au sein de l'échantillon qui ont appris à compter selon la méthode A.

   (a) La variable aléatoire X suit elle une loi binomiale ? *Justifier pourquoi, et si vous répondez qu'elle suit une loi binomiale, précisez laquelle en indiquant les valeurs de n et p*

   ☐ X suit la loi $\mathcal{B}(\ ;\ )$ ☒ X suit presque la loi $\mathcal{B}(65; 0{,}5)$ ☐ X ne suit pas une loi binomiale
   En effet, le tirage est sans remise (on n'autorise pas à piocher deux fois le même élève), mais la taille de l'échantillon (à savoir 65) est plus de 10 fois plus petite que la taille de la population parmi laquelle on pioche (à savoir 3 192). Donc le cours indique que X suit presque une loi binomiale : la loi $\mathcal{B}(65; 0{,}5)$ car l'échantillon est de taille 65 et qu'on a pioché dans une population parmi laquelle 50% des élèves ayant suivi la méthode A.

   (b) Combien vaut la probabilité $\mathbb{P}[X \geqslant 43]$?

   La calculatrice indique (pour cette loi binomiale) que $\mathbb{P}[X \geqslant 43] \simeq 0{,}00625$.

(c) Justifier que l'on peut approcher X par la loi $\mathcal{N}(32{,}5\ ;\ 4{,}03)$. Y aurait-il besoin d'effectuer une correction de continuité ?

$n\,p(1-p) = 65 \times 0{,}5\,(1-0{,}5) = 16{,}25 > 10$, donc on peut approcher $\mathcal{B}(65\ ;\ 0{,}5) \approx \mathcal{N}(65 \times 0{,}5\ ;\ \sqrt{65 \times 0{,}5\,(1-0{,}5)}) \simeq \mathcal{N}(32{,}5\ ;\ 4{,}03)$, mais il est nécessaire de faire une correction de continuité.

(d) Pour donner une idée de ce qu'est la loi $\mathcal{N}(32{,}5\ ;\ 4{,}03)$, calculer les effectifs théoriques d'une variable suivant la loi $\mathcal{N}(32{,}5\ ;\ 4{,}03)$, si l'effectif total est de 100. *Dans cette question, on n'essaiera pas de faire de correction de continuité.*

Avec la calculatrice, on obtient $100 \times \mathbb{P}[X < 23] \simeq 100 \times 0{,}009\,203\,69 \simeq 0{,}92$,
$100 \times \mathbb{P}[23 \leqslant X < 27] \simeq 100 \times 0{,}076\,959\,9 \simeq 7{,}7$,
$100 \times \mathbb{P}[27 \leqslant X < 31] \simeq 100 \times 0{,}268\,705 \simeq 26{,}87$,
$100 \times \mathbb{P}[31 \leqslant X < 35] \simeq 100 \times 0{,}377\,617 \simeq 37{,}76$,
$100 \times \mathbb{P}[35 \leqslant X < 39] \simeq 100 \times 0{,}214\,132 \simeq 21{,}41$
et $100 \times \mathbb{P}[X \geqslant 39] \simeq 100 \times 0{,}053\,382\,8 \simeq 5{,}34$. D'où les effectifs théoriques ci-dessous :

Score	[19 ; 23[	[23 ; 27[	[27 ; 31[	[31 ; 35[	[35 ; 39[	[39 ; 43[
Effectif théorique	0,92	7,7	26,87	37,76	21,41	5,34

3. On réalise l'expérience dans une agglomération où il y a 3 192 élèves de CP : on apprend à compter selon la méthode A à la moitié d'entre eux (1 596 élèves) et selon la méthode B à l'autre moitié d'entre eux. En fin de CP, les élèves passent tous le même test de mathématiques, et on se focalise sur ceux qui ont les meilleurs notes à ce test.

   Si on observe que parmi les 65 élèves qui ont eu les meilleurs notes au test, il y en a au moins 43 qui ont appris à compter selon la méthode A, cela porterait-il à penser que les élèves ayant appris à compter selon la méthode A et ceux ayant appris selon la méthode B aient les mêmes chances de bien réussir ce test ? *Votre réponse ne doit pas nécessairement être très longue, mais on vous demande de vous appuyer sur les résultat d'une (ou plusieurs) question·s précédente·s en expliquant en quoi elle·s permet·tent de conclure.*

On désigne par X le nombre d'élèves ayant appris à compter selon la méthode A, parmi les 65 élèves qui réussissent le mieux le test.
Si les élèves avaient les mêmes chances de bien réussir le test selon qu'ils aient appris selon la méthode A ou B, alors X suivrait la loi $\mathcal{B}(65; 0{,}5)$ (car sous cette hypothèse ça revient juste à choisir 65 élèves au hasard comme en question 2a), donc la probabilité que $X \geqslant 43$ serait de 0,00625 (comme en question 2b), c'est à dire qu'il aurait été très improbable d'avoir $X \geqslant 43$.
En conséquence si on observe $X \geqslant 43$, cela porterait à remettre en cause ce calcul de probabilité, donc à remettre en cause l'hypothèse sur laquelle se fonde le calcul, donc à penser que les élèves n'avaient pas les mêmes chances de bien réussir le test selon qu'ils aient appris avec la méthode A ou avec la méthode B.