Contrôle Terminal – 11/05/2022 – 2h00
Statistiques – L1 de psychologie

Réponses

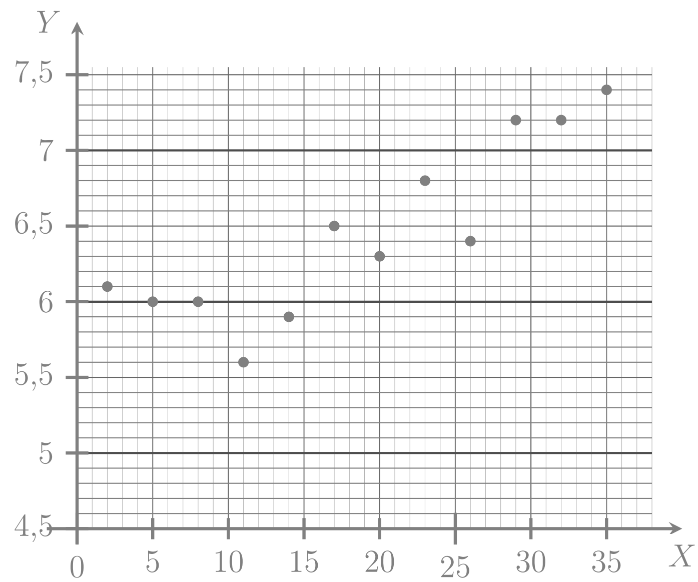
Veuillez rendre ce sujet et votre copie.

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 : Sport et mathématiques

Mme Riou est chercheuse en médecine. Elle cherche à savoir si la pratique du sport, lorsqu'elle met en jeu une plus forte implication des élèves et des activités physiques cognitivement stimulantes, permet d'améliorer les performance dans d'autres disciplines comme les mathématiques. Pour cela, elle considère 12 classes de CM1, où un certain nombre d'heures parmi les cours de sports de l'année sont repensées pour les rendre plus stimulantes (ce nombre d'heure est noté X). Elle fait ensuite passer le même test de mathématiques à toutes ces classes, et regarde la moyenne obtenue par chaque classe (notée Y).
Elle récolte ainsi des données, qu'elle synthétise par le nuage de points ci-contre :



1. Entrer les nombres d'heures X et les moyennes en maths (Y) de chaque classe dans le tableau ci-dessous :

Nombre d'heures (X)	2	5	8	11	14	17	20	23	26	29	32	35
Moyenne en maths (Y)	6,1	6	6	5,6	5,9	6,5	6,3	6,8	6,4	7,2	7,2	7,4

2. Parmi le sous échantillon formé des classes où moins de 18 heures de sports ont été repensées pour les rendre plus stimulantes, quels sont la moyenne et l'écart type de Y ?

Dans ce sous-échantillon, les notes Y sont : 6,1, 6, 6, 5,6, 5,9 et 6,5. On obtient alors
$m(Y) \simeq 6{,}0167$
$s(Y) \simeq 0{,}2672$

3. Parmi le sous échantillon formé des classes où plus de 18 heures de sports ont été repensées pour les rendre plus stimulantes, quels sont la moyenne et l'écart type de Y ?

Dans ce sous-échantillon, les notes Y sont : 6,3, 6,8, 6,4, 7,2, 7,2 et 7,4. On obtient alors
$m(Y) \simeq 6{,}8833$
$s(Y) \simeq 0{,}418$

Exercice 2 : 2$^{\text{ème}}$ tour des élections présidentielles

La famille Mahe compte 6 personnes. Lors du second tour de la présidentielle, trois d'entre eux (qui s'appellent Alexis, Benjamin et Camille) ont voté pour E. Macron ; une d'entre eux (qui s'appelle Delphine) a voté pour M. Le Pen, et deux d'entre eux (qui s'appellent Estelle et Françoise) se sont abstenues (ou ont voté blanc).

Quand trois membres de la famille (choisis au hasard) discutent ensemble, on cherche à déterminer la probabilité qu'il y ait précisément un·e d'entre eux qui ait voté pour E. Macron, un·e qui ait voté pour M. Lepen, et un·e qui se soit abstenu·e (ou qui ait voté blanc).

1. Combien y a-t-il de choix possibles de trois personnes parmi la famille Mahe ? *Vous pouvez soit en faire la liste exhaustive, soit donner une formule (et calculer sa valeur) qui indique ce nombre de choix possibles. Écrire la liste de tout les cas peut aider à comprendre, mais attention car cela peut prendre du temps.*

Il y a $\binom{6}{3} = 20$ choix possibles.
Remarque : On peut en donner la liste, ci-dessous
ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF et DEF
Où par exemple ACD signifie "Alexis+Camille+Delphine".

2. Parmi ces choix possibles, combien contiennent une personne qui a voté pour E. Macron, une qui a voté M. Le Pen, et une qui s'est abstenue (ou a voté blanc). *À nouveau, vous pouvez soit en faire la liste exhaustive, soit donner une formule (et calculer sa valeur).*

Il y en a $3 \times 1 \times 2 = 6$. En effet, il y a 3 choix possible d'une personne qui a voté E. Macron, et pour chacun des choix d'une personne qui a voté pour E. Macron il y a 2 possibilités pour la personne qui s'est abstenue (d'où 3×2 possibilités pour les deux personnes ayant voté Macron et abstention/blanc). Pour chaque choix des deux personnes qui ont voté Macron et abstention/blanc, on ajoute finalement Delphine (la seule possibilité pour ajouter une personne qui a voté LePen) et on obtient 6 choix avec "1 vote Macron+1 vote Lepen+1 abstention ou vote blanc".
Remarque :On peut aussi en donner la liste, ci dessous :

ADE	BDE	CDE
ADF	BDF	CDF

On note qu'il y a deux colonnes (le choix de la personne qui a voté Macron), et trois lignes (le choix de la personne qui s'est abstenue ou a voté blanc), et on retrouve donc bien qu'il y a 2×3 possibilités.

3. Conclure : Quand trois membres de la famille (choisis au hasard) discutent ensemble, quelle est la probabilité qu'il y ait précisément un·e d'entre eux qui ait voté pour E. Macron, un·e qui ait voté pour M. Lepen, et un·e qui se soit abstenu·e (ou qui ait voté blanc).

On a vu que cette situation correspond à 6 cas sur les 20 cas possibles, donc cette probabilité vaut $\frac{6}{20} = 0{,}3$.

On cherche désormais à répéter le même calcul parmi l'ensemble des électeurs français :

Lors du second tour de la présidentielle, parmi les 48 752 339 électeurs inscrits, il y en a 18 768 639 qui on voté pour E. Macron, 13 288 686 qui ont voté pour M. Le Pen, et 16 695 014 qui se sont abstenus (ou ont voté blanc ou nul).

4. Lorsque trois électeurs (choisis au hasard) se rencontrent, quelle est la probabilité qu'il s'agisse précisément d'une personne qui a voté pour E. Macron, une qui a voté pour M. Le Pen, et une personne qui s'est abstenue (ou a voté blanc ou nul). *Vous pourrez vous inspirer des questions précédentes en adaptant le calcul.*

Cette fois ci le nombre total de cas est $\binom{48752339}{3} \simeq 1{,}93123 \times 10^{22}$.
Parmi ceux-là, il y a $18768639 \times 13288686 \times 16695014 \simeq 4{,}16391 \times 10^{21}$ cas avec un vote Macron, un vote Le Pen et un blanc ou abstention.
En conséquence, cette probabilité vaut $\frac{4{,}16391 \times 10^{21}}{1{,}93123 \times 10^{22}} \simeq 0{,}215\,6$.

Exercice 3 : Croyances et écologie.

Une chercheuse en géopolitique s'intéresse au lien entre religions et écologie. En effet, de nombreuses personnalités emblématiques dans l'histoire de l'écologie entretenaient des liens étroits avec une religion.

Elle interroge des militant·e·s écologistes français·es, et compte combien d'entre eux se déclarent athées.

Elle souhaite le comparer à la proportion d'athées parmi l'ensemble de la population française, qui compte 40% d'athées. Pour cela on peut calculer les probabilités pour des échantillons de Français·es au hasard, et comparer le résultat avec des échantillons de militant·e·s écologistes.

1. On considère un échantillon aléatoire de 29 français·es, et on désigne par S le nombre d'athées au sein de l'échantillon.
   (a) Justifiez que S suit une loi binomiale, indiquez laquelle et donnez sa moyenne et son écart type.

   S suit une loi binomiale car la taille de l'échantillon (29) est beaucoup plus petite que la population (des dizaines de millions de français) parmi laquelle est choisi l'échantillon.
   $S \sim \mathcal{B}(29; 0{,}4)$.
   $m(S) = 29 \times 0{,}4 = 11{,}6$ $\qquad$ $s(S) = \sqrt{29 \times 0{,}4\,(1 - 0{,}4)} \simeq 2{,}638$

   (b) Calculer la probabilité $\mathbb{P}[S \geqslant 17]$. *En plus du résultat, on vous demande d'indiquer quel calcul vous avez effectué (et/ou quel calcul vous avez demandé à la calculette).*

   La Calculette indique que $\mathbb{P}[S \leqslant 16] \simeq 0{,}967\,12$
   Donc $\mathbb{P}[S \geqslant 17] = 1 - \mathbb{P}[S \leqslant 16] \simeq 1 - 0{,}967\,12 = 0{,}032\,88$

2. On considère désormais un échantillon aléatoire de 12 000 français·es, et on désigne par X le nombre d'athées au sein de l'échantillon.
   (a) Justifier que X peut être approchée par une loi normale, et préciser laquelle. Faut il faire une correction de continuité quand on utilise cette approximation par une loi normale ?

   L'échantillon est beaucoup plus gros qu'en question 1, mais reste beaucoup plus petit que le population totale (des dizaines de millions de français). Donc $X \sim \mathcal{B}(12000; 0{,}4)$.
   $n\,p(1-p) = 12\,000 \times 0{,}4\ (1-0{,}4) = 2\,880 > 1\,000$, donc on peut approcher $\mathcal{B}(12000\ ;\ 0{,}4) \approx \mathcal{N}(12\,000 \times 0{,}4\ ;\ \sqrt{12\,000 \times 0{,}4\ (1-0{,}4)}) \simeq \mathcal{N}(4800\ ;\ 53{,}666)$, et il n'y a pas besoin de faire de correction de continuité.

   (b) Calculer $\mathbb{P}[4\,703 \leqslant X \leqslant 4\,899]$

   Comme $X \sim \mathcal{N}(4800\ ;\ 53{,}666)$, la calculatrice indique que $\mathbb{P}\,[4\,703 \leqslant X \leqslant 4\,899] \simeq 0{,}932\,12$

   (c) Déterminer k tel que $\mathbb{P}[X \leqslant k] \simeq 0{,}05$. *On vous demande dans cette question de présenter un calcul qui utilise une table du formulaire, mais n'utilise pas les fonctions qui manipulent la loi normale sur la calculatrice.*

   Pour résoudre l'équation $\mathbb{P}[X \leqslant k] \simeq 0{,}05$, on note tout d'abord que
   $\mathbb{P}[X \leqslant k] = \mathbb{P}[Z \leqslant \frac{k-4\,800}{53{,}666}] = F\left(\frac{k-4\,800}{53{,}666}\right)$
   d'où l'équation $F\left(\frac{k-4\,800}{53{,}666}\right) = 0{,}05$, c'est à dire $F\,(z) = 0{,}05$, où on note $z = \frac{k-4\,800}{53{,}666}$.
   Comme $F(z) \simeq 0{,}05 < 0{,}5$, on sait que $z < 0$.
   On commence donc par résoudre l'équation $F\,(|z|) \simeq 0{,}95$:
   $|z| \simeq 1{,}645$ d'après la table inverse du formulaire.
   D'où $z \simeq -1{,}645$.
   D'où $\frac{k-4\,800}{53{,}666} = z = (-1{,}645)$, qui donne $k \simeq z \times 53{,}666 + 4\,800 \simeq -1{,}645 \times 53{,}666 + 4\,800 \simeq 4712$.

Les probabilités calculées précédemment pourraient être comparées avec des résultats expérimentaux sur des échantillons réels. Mais à la place la chercheuse choisi simplement d'interroger un échantillon de 85 militant·e·s écologistes, choisis au hasard parmi l'ensemble des militant·e·s écologistes français·es : elle constate qu'il y a 36 athées au sein de l'échantillon.

3. En déduire une estimation de la proportion d'athées parmi l'ensemble des militant·e·s écologistes. *Vous déterminerez un intervalle de confiance, pour la confiance 95%.*

On a $n\,p_e(1-p_e) = 85 \times \frac{36}{85}\left(1-\frac{36}{85}\right) \simeq 20{,}753 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
On a $F(1{,}96) \simeq 0{,}975$ d'où $z_\alpha \simeq 1{,}96$
d'où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}105$ et $p_e - a_\alpha \simeq \frac{36}{85} - 0{,}105 \simeq 0{,}318\,5$ et $p_e + a_\alpha \simeq \frac{36}{85} + 0{,}105 \simeq 0{,}528\,5$.
On estime donc que p est dans l'intervalle $[0{,}3185; 0{,}5285]$ avec la confiance $c = 0{,}95$.

4. Quelle taille d'échantillon aurait-il fallu pour estimer cette proportion à 2% près avec la confiance 95% ?

Pour avoir une précision 0,02 avec la confiance 0,95, il faut $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2} \simeq 1{,}96^2 \frac{\frac{36}{85}(1-\frac{36}{85})}{0{,}02^2} \simeq 2344$

5. Pour chacune des phrases suivantes, indiquer si on peut affirmer, avec la confiance 95% que la phrase est vraie, ou si on peut affirmer qu'elle est fausse, ou bien si on ne peut rien affirmer avec cette confiance 95%. *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse principalement à partir des estimations des questions précédentes.*

La proportion d'athées est plus grande parmi les militant·e·s écologistes que parmi le reste de la population.

☐Vraie ☐Fausse ☒ On ne peut rien affirmer

En effet, l'intervalle calculé contient à la fois des valeurs plus grandes que 40% et des valeurs plus petites.

Prendre conscience de la situation environnementale rend souvent les gens athées

☐Vraie ☐Fausse ☒ On ne peut rien affirmer

En effet, on n'a pas mis en évidence de lien entre athéisme et militantisme écologique.

Exercice 4 : Influence des écrans sur l'apparition de troubles autistiques

On cherche à savoir si l'exposition des très jeunes enfants aux écrans a un impact sur l'apparition de troubles du spectre de l'autisme (qui peuvent être diagnostiqués à partir de 2 ou 3 ans).

On commence tout d'abord par considérer un échantillon d'enfants bourguignons de un an : on désigne par X la durée quotidienne qu'ils passent devant les écrans, et on récolte les données ci-dessous

Temps X en minutes	[0 ; 27[	[27 ; 54[	[54 ; 81[	[81 ; 108[	[108 ; 135[	[135 ; 162[
Effectif	161	123	79	29	6	2

1. Combien y a-t-il d'enfants dans l'échantillon ?

Il y en a $161 + 123 + 79 + 29 + 6 + 2 = 400$

2. Au sein de cet échantillon, quels sont la moyenne et l'écart type du temps quotidien passé par les enfants devant les écrans ?

moyenne : $m(X) = \frac{\sum c_i n_i}{n} = \frac{13{,}5\times161+40{,}5\times123+67{,}5\times79+\cdots+148{,}5\times2}{400} = \frac{16\,254}{400} \simeq 40{,}64\,\text{min}$

$m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{13{,}5^2\times161+40{,}5^2\times123+67{,}5^2\times79+\cdots+148{,}5^2\times2}{400} = \frac{982\,692}{400}$

$Var(X) = m(X^2) - m(X)^2 = \frac{982\,692}{400} - \left(\frac{16\,254}{400}\right)^2 \simeq 805{,}53$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 28{,}38\,\text{min}$

3. En supposant que cet échantillon est représentatif des enfants bourguignons, estimer le temps moyen que les enfants bourguignons de un an passent quotidiennement devant les écrans. *Vous déterminerez un intervalle de confiance, pour la confiance 96%.*

> Comme n= 400 >30, on cherche z_α tel que $F(z_\alpha) = \frac{0{,}96+1}{2} = 0{,}98$
> On a $F(2{,}054) \simeq 0{,}98$ d'où $z_\alpha \simeq 2{,}054$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2{,}918$
> On estime donc que μ est dans l'intervalle $[40{,}64 - 2{,}918\,;\, 40{,}64 + 2{,}918] \simeq [37{,}72\,;\, 43{,}56]$, avec la confiance $c = 0{,}96$.

On considère désormais un échantillon de 23 enfants (bourguignons) de 3 ans qui ont été diagnostiqués autistes. En interrogeant leurs parents, assistant·e·s maternelles, etc on détermine combien de temps environ ils passaient quotidiennement devant les écrans à l'âge de 1 an, et on obtient une moyenne de 100,87 minutes et un écart-type de 57,17 minutes.

4. En supposant que cet échantillon est représentatif des enfants bourguignons autistes, estimer le temps moyen que l'ensemble des enfants bourguignons autistes passaient quotidiennement devant les écrans à l'âge de un an (donc avant d'être diagnostiqués autistes). *Vous déterminerez un intervalle de confiance, pour la confiance 96%.*

> Comme $n = 23 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}02$ et $n - 1 = 22$ degrés de liberté (ddl)
> On lit $t_\alpha \simeq 2{,}1829$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 26{,}607$.
> On estime donc que μ est dans l'intervalle $[100{,}87 - 26{,}607\,;\, 100{,}87 + 26{,}607] \simeq [74{,}26\,;\, 127{,}48]$ avec la confiance $c = 0{,}96$

5. Pour chacune des phrases suivantes, indiquer si on peut affirmer, avec la confiance 96% que la phrase est vraie, ou si on peut affirmer qu'elle est fausse, ou bien si on ne peut rien affirmer avec cette confiance 96%. *En plus de cocher Vraie/Fausse/On-ne-peut-rien-affirmer, on vous demande de justifier votre réponse principalement à partir des estimations des questions précédentes.*

> En moyenne, les enfants diagnostiqués autistes sont des enfants qui, à l'âge de un an, passaient plus de temps devant les écrans que les autres enfants.
> ☒ Vraie ☐Fausse ☐On ne peut rien affirmer
> En effet, les intervalles calculés en questions 3 et 5 ne se chevauchent pas et indiquent que la moyenne est plus élevées chez les enfants qui sont ensuite diagnostiqués autistes.
>
> Regarder beaucoup les écrans à l'âge de 1 an est un des facteurs qui provoquent plus tard des troubles autistiques.
> ☐Vraie ☐Fausse ☒ On ne peut rien affirmer
> En effet, les données recueillies ne permettent pas de déterminer si les écrans sont la cause et l'autisme la conséquence, ou si c'est par exemple l'inverse (que même s'ils n'ont pas encore été diagnostiqués, les enfants qui vont développer un trouble autistiques ont dès l'âge de 1 an des traits de personnalités différents qui les attirent plus vers les écrans).