Contrôle terminal – 5 mai 2017 – Durée : 2h00
Statistiques, L1 de psychologie Réponses

Veuillez rendre ce sujet avec votre copie. Numéro d'anonymat :

Le formulaire et la calculatrice sont autorisés. Vous êtes invités à répondre directement sur l'énoncé, mais si vous avez besoin de plus de place, vous pouvez détailler certaines questions sur un copie que vous rendrez avec l'énoncé. 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 : Impact socioprofessionnel de maladies dans l'enfance
Le docteur Guy dispose des dossiers médicaux de ses patients, et cherche à savoir si leurs revenus à l'âge adulte dépendent du fait qu'ils aient été hospitalisés au cours de leur enfance : en effet, il s'attend à ce que les hospitalisations nuisent à la réussite des études, et influent donc directement sur le revenu à l'âge adulte.

À cette fin, il divise sa patientèle en deux groupes : le groupe A est constitué d'adultes qui n'ont jamais été hospitalisé pendant plus d'un mois au cours de leur enfance, et le groupe B est constitué d'adultes qui, au cours de leur enfance, ont été hospitalisé au moins une fois pendant plus d'un mois.

Afin de calculer des intervalles de confiance, on pourra supposer dans cette exercice que les revenus suivent une loi normale.

1. Le Docteur Guy commence par interroger 26 patients choisis au hasard (avec remise) au sein du groupe B. Il obtient les revenus annuels suivants : 0 €, 16 212 €, 0 €, 15 361 €, 31 329 €, 18 833 €, 12 708 €, 7 042 €, 16 058 €, 17 412 €, 48 126 €, 23 782 €, 28 296 €, 28 080 €, 27 131 €, 26 982 €, 28 098 €, 11 279 €, 26 888 €, 36 594 €, 6 705 €, 8 814 €, 22 330 €, 0 €, 22 058 €, et 27 920 €.

   (a) Déterminer la moyenne et l'écart type du revenu annuel au sein de cet échantillon.
   On demande dans cette question de justifier la réponse par un court calcul.

   moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{0+16212+\cdots+27920}{26} = \frac{508038}{26} \simeq 19\,540$
   $m(X^2) = \frac{\sum x_i^2}{n} = \frac{0^2+16212^2+\cdots+27920^2}{26} = \frac{13408215542}{26}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{13408215542}{26} - \left(\frac{508038}{26}\right)^2 \simeq 133\,892\,004$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 11\,571$

   (b) Déterminer aussi le revenu médian au sein de cet échantillon.

   La médiane est la valeur numéro $\frac{26+1}{2} = 13{,}5$ (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 13 et 14. C'est donc le milieu entre 18833 et 22058 c'est-à-dire $\frac{18\,833+22\,058}{2} = 20\,445{,}5$.

   (c) Estimer, avec la confiance 99%, le revenu moyen de l'ensemble des patients du groupe B.
   Répondre en calculant un intervalle de confiance.

   Comme $n = 26 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}005$ et $n-1 = 25$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 2{,}7874$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 6\,450{,}6$.
   On estime donc que μ est dans l'intervalle $[13\,089; 25\,991]$ avec la confiance $c = 0{,}99$

2. Il interroge ensuite 125 patients choisis au hasard (avec remise) au sein du groupe A. Il constate qu'au sein de cet échantillon de 125 patients, le revenu annuel moyen est de 22514€, avec un écart type de 17518€.
   Estimer, avec la confiance 99%, le revenu moyen de l'ensemble des patients du groupe A.
   Répondre en calculant un intervalle de confiance.

   Comme n= 125 >30, on cherche z_α tel que $F(z_\alpha) = \frac{c+1}{2}$
   On a $F(2{,}576) \simeq 0{,}995$ d'où $z_\alpha \simeq 2{,}576$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 4052$
   On estime donc que μ est dans l'intervalle $[22\,514 - 4\,052\,;\, 22\,514 + 4\,052] = [18\,462\,;\, 26\,566]$, avec la confiance $c = 0{,}99$.

3. Peut-on conclure que, comme s'y attendait le docteur Guy, les patients du groupe B aient en moyenne un revenu plus faible que ceux du groupe A ? *(on répondra avec la confiance 99%)*

Les intervalles calculés se chevauchent, ils ne permettent donc pas d'affirmer, avec la confiance 99%, quel groupe de patient à le revenu moyen le plus élevée. Ceci ne suffit donc pas à confirmer l'hypothèse du docteur Guy.

Exercice 2 : Étude de la satisfaction au travail
Des psychologues ont établi un questionnaire évaluant la satisfaction au travail des employés. Ils l'ont construit de telle sorte que parmi l'ensemble des salariés français, la distribution des notes obtenues est la loi normale $\mathcal{N}(80; 24)$ (c'est-à-dire que si l'on note X la note de satisfaction au travail d'un salarié choisi au hasard, on a $X \sim \mathcal{N}(80; 24)$).

1. Calculer la probabilité $\mathbb{P}[99{,}2 \leqslant X \leqslant 111{,}2]$.
   Dans cette question on demande une justification par un court calcul utilisant la table du formulaire.

$$\begin{aligned}\mathbb{P}[99{,}2 \leqslant X \leqslant 111{,}2] &= \mathbb{P}[\tfrac{99{,}2-80}{24} < \tfrac{X-80}{24} < \tfrac{111{,}2-80}{24}] \simeq \mathbb{P}[0{,}8 < Z < 1{,}3] \\ &\simeq F(1{,}3) - F(0{,}8) \simeq 0{,}903\,2 - 0{,}788\,1 \simeq 0{,}1151\end{aligned}$$

2. Donner aussi la valeur des probabilités $\mathbb{P}[41{,}6 \leqslant X \leqslant 111{,}2]$ et $\mathbb{P}[X \geqslant 68]$.

$$\begin{aligned}\mathbb{P}[41{,}6 \leqslant X \leqslant 111{,}2] &= \mathbb{P}[\tfrac{41{,}6-80}{24} < \tfrac{X-80}{24} < \tfrac{111{,}2-80}{24}] \simeq \mathbb{P}[-1{,}6 < Z < 1{,}3] \\ &\simeq F(1{,}3) - F(-1{,}6) \simeq 0{,}9032 - (1 - 0{,}945\,2) \simeq 0{,}903\,2 - 0{,}054\,8 \simeq 0{,}8484\end{aligned}$$

$$\begin{aligned}\mathbb{P}[X \geqslant 68] &= \mathbb{P}[\tfrac{68-80}{24} < \tfrac{X-80}{24}] \simeq \mathbb{P}[-0{,}5 < Z] \\ &\simeq 1 - F(-0{,}5) \simeq 1 - (1 - 0{,}691\,5) \simeq 1 - 0{,}308\,5 \simeq 0{,}6915\end{aligned}$$

3. Quelle est la proportion théorique d'individus dont le score de satisfaction au travail est inférieur à 40 ?
   Dans cette question on demande à nouveau une justification par un court calcul utilisant la table du formulaire. Vous effectuerez une interpolation linéaire pour augmenter la précision du résultat.

Cette proportion théorique n'est autre que la probabilité $\mathbb{P}[X \leqslant 40]$, calculée ci-dessous :

$$\begin{aligned}\mathbb{P}[X \leqslant 40] &= \mathbb{P}[\tfrac{X-80}{24} < \tfrac{40-80}{24}] \simeq \mathbb{P}[Z < -1{,}6667] \\ &\simeq F(-1{,}6667) \simeq 1 - 0{,}952\,2 \simeq 0{,}0478\end{aligned}$$

Où l'on a utilisé une interpolation linéaire :

$$F(1{,}6667) \simeq 0{,}9515 + \tfrac{0{,}9525-0{,}9515}{1{,}67-1{,}66}(1{,}6667 - 1{,}66) \simeq 0{,}9522.$$

4. Les chercheurs ont remarqué qu'en dessous de la note 40, les salariés ont des risques particulièrement élevés de développer des troubles psychosociaux en rapport avec le travail. Ils qualifient donc d'"*insatisfaction préoccupante*" la situation des salariés dont la note de satisfaction est plus faible que 40. Ils constatent que 5% des salariés français ont une *insatisfaction préoccupante* au travail.

   On considère dans cette question un échantillon de 150 salariés français, choisis avec remise, et on note S_{150} le nombre de salariés qui, au sein de l'échantillon, ont une insatisfaction au travail préoccupante.

   (a) Quelle est la loi de la variable S_{150} ? Précisez de plus sa moyenne et son écart type.

   Comme le tirage est avec remise, on a :
   $S_{150} \sim \mathcal{B}(150; 0{,}05)$
   $m(S_{150}) = 150 \times 0{,}05 = 7{,}5$
   $s(S_{150}) = \sqrt{150 \times 0{,}05\ (1 - 0{,}05)} \simeq 2{,}669$

   (b) Justifier que cette loi peut être approchée par une autre loi et préciser laquelle.

   On a : $n = 150 > 30$, $n\,p = 7{,}5 > 5$, $n\,(1-p) = 142{,}5 > 5$, donc on peut approcher cette loi par $\mathcal{N}(7{,}5; 2{,}669)$.

   (c) En déduire une valeur approchée du troisième quartile de S_{150}.

   Utiliser le résultat de la question précédente et justifier la réponse par un court calcul.

   Pour résoudre l'équation $\mathbb{P}[S_{150} \leqslant Q_3] \simeq 0{,}75$, on note tout d'abord que
   $\mathbb{P}[S_{150} \leqslant Q_3] = \mathbb{P}[Z \leqslant \frac{Q_3 - 7{,}5}{2{,}669}] = F\left(\frac{Q_3-7{,}5}{2{,}669}\right)$
   d'où l'équation $F\left(\frac{Q_3-7{,}5}{2{,}669}\right) = 0{,}75$, c'est à dire $F(z) = 0{,}75$, où on note $z = \frac{Q_3-7{,}5}{2{,}669}$.
   On effectue une interpolation linéaire en posant
   $x_1 = 0{,}67$ et $y_1 = F(0{,}67) \simeq 0{,}7486$.
   $x_2 = 0{,}68$ et $y_2 = F(0{,}68) \simeq 0{,}7517$.
   $x = z$ et $y = F(z) = 0{,}75$. La formule d'interpolation linéaire donne alors :
   $z = x \simeq 0{,}67 + \frac{0{,}68-0{,}67}{0{,}7517-0{,}7486}(0{,}75 - 0{,}7486) \simeq 0{,}675$.
   D'où $\frac{Q_3-7{,}5}{2{,}669} = z = 0{,}675$, qui donne $Q_3 \simeq z \times 2{,}669 + 7{,}5 \simeq 0{,}675 \times 2{,}669 + 7{,}5 \simeq 9{,}3$.

Exercice 3 : Télétravail

On s'intéresse dans cet exercice aux salariés français : ils sont 21,5 millions, parmi lesquels 5% ont une insatisfaction au travail préoccupante. De plus, 12,4% des salariés français travaillent (au moins une partie du temps) à distance (on parle de *télétravail*).

1. Combien y a-t-il en France de salariés faisant du télétravail ?

   Il y en a environ $21\,500\,000 \times 0{,}124 = 2\,666\,000$.

2. On considère un échantillon aléatoire (avec remise) de 65 salariés en télétravail, parmi lequel on constate que 7 ont une insatisfaction au travail préoccupante.

   Estimer, avec la confiance 99%, le taux d'insatisfaction au travail préoccupante parmi les salariés en télétravail.

   Répondre par le calcul d'un intervalle de confiance.

   On a $n = 65 > 30$, et $p_e = \frac{7}{65} \simeq 0{,}108$, d'où $n\,p_e = 7 > 5$, et $n(1-p_e) = 58 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
   On a $F(2{,}576) \simeq 0{,}995$ d'où $z_\alpha \simeq 2{,}576$
   d'où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}099$.
   On estime donc que p est dans l'intervalle $[0{,}0087; 0{,}2067]$ avec la confiance $c = 0{,}99$.

3. Peut-on en conclure (avec la confiance 99%) que les salariés en télétravail sont plus souvent en situation d'insatisfaction au travail préoccupante que les autres salariés ? Ou peut-on conclure au contraire, qu'ils le sont moins souvent ?

Cet intervalle ne permet pas d'affirmer, avec la confiance 99%, si ce taux d'insatisfaction au travail est plus grand ou plus petit que 5%, c'est-à-dire si les salariés en télétravail sont plus souvent ou moins souvent en situation d'insatisfaction au travail préoccupante que les autres salariés.

4. Si l'on voulait estimer à 2% près, et avec la confiance 99%, la proportion d'insatisfaction au travail parmi les salariés en télétravail, quelle taille d'échantillon serait nécessaire ?

Pour avoir une précision 0,02 avec la confiance 0,99, il faut $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2} \simeq 2{,}576^2 \frac{0{,}108(1-0{,}108)}{0{,}02^2} \simeq 1598$

Exercice 4 : Entreprise ACME
On étudie dans cet exercice la satisfaction au travail des employés de l'entreprise ACME. Cette entreprise compte 160 salariés dont 8 souffrent d'une insatisfaction au travail préoccupante.

Au cours de cet exercice, on considère des échantillons aléatoires de 6 salariés, choisis **sans** remise parmi les salariés de l'entreprise ACME.

1. Combien y a-t-il d'échantillons possibles de 6 salariés choisis parmi les salariés de l'entreprise ACME ?
On justifiera par un court calcul ou en indiquant quelle fonction de la calculette est utilisée.

Il y a $\binom{160}{6}$ échantillons possibles, car c'est le nombre de façons de choisir 6 personnes parmi 160.
Or $\binom{160}{6} = \frac{160!}{6! \times 160-6!} = \frac{1}{6\times5\times\ldots\times1} \times \frac{160\times159\times\ldots\times1}{154\times153\times\ldots\times1} = \frac{1}{6\times5\times\ldots\times1} \times 160 \times 159 \times \ldots \times 155 = \frac{160\times159\times158\times157\times156\times155}{6\times5\times4\times3\times2\times1} =$ 21 193 254 160. Donc cela fait 21 193 254 160 échantillons, c'est-à-dire environ $2{,}12 \times 10^{10}$.

2. On note X le nombre de salariés souffrant d'une insatisfaction au travail préoccupante, au sein d'un échantillon aléatoire (sans remise) de 6 salariés de cette entreprise.
Quelle est la loi de X ?

$X \sim \mathcal{H}(160; 8; 6)$

3. Calculer la probabilité $\mathbb{P}[X \geqslant 3]$. *On demande de justifier par une court calcul.*

$$\begin{aligned}\mathbb{P}[X \geqslant 3] &= \mathbb{P}[X = 3] + \mathbb{P}[X = 4] + \mathbb{P}[X = 5] + \mathbb{P}[X = 6] \\ &= \frac{\binom{8}{3}\binom{160-8}{6-3}}{\binom{160}{6}} + \frac{\binom{8}{4}\binom{160-8}{6-4}}{\binom{160}{6}} + \frac{\binom{8}{5}\binom{160-8}{6-5}}{\binom{160}{6}} + \frac{\binom{8}{6}\binom{160-8}{6-6}}{\binom{160}{6}} \\ &\simeq 0{,}00152 + 4 \times 10^{-05} + 0 + 0 \simeq 0{,}0016\end{aligned}$$

4. Le service des relations clients compte 6 salariés, parmi lesquels 5 souffrent d'une insatisfaction au travail préoccupante. Peut-on conclure qu'au sein de cette entreprise, il y ait un lien entre l'insatisfaction au travail et le fait d'être affecté au service des relations clients ?

S'il n'y avait pas de lien entre l'insatisfaction et le fait d'être dans ce service, alors l'insatisfaction au travail dans ce service serait la même qu'au sein d'un échantillon aléatoire de 6 salariés de l'entreprise et (d'après la question précédente) il serait très peu probable d'avoir au moins 3 salariés du service qui aient une insatisfaction au travail préoccupante. Or on nous indique qu'il y en a 5 (donc plus que 3). On peut donc penser qu'il y ait un lien entre l'insatisfaction au travail et le fait d'être affecté au service client.