Examen de statistiques – 6 mai 2016
L1 de psychologie

Réponses

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

Le formulaire et la calculatrice sont autorisés. Le soin de la rédaction entrera en compte dans la correction mais dans les questions où des détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.

Exercice 1 : Parité dans l'enseignement

Il y a en France 357 000 hommes et 685 000 femmes qui sont enseignants.

1. Combien y a-t-il d'enseignants en France, et quelle proportion d'entre eux sont des femmes ?

   En France, il y a 1 042 000 enseignants, parmi lesquels 65,7 % sont des femmes.

---

2. On considère dans cette question un échantillon aléatoire de cinq enseignants, choisi sans remise, et on note X le nombre de femmes au sein de cet échantillon aléatoire.

   (a) Donner la loi de la variable X.

   $X \sim \mathcal{H}(1\,042\,000; 685\,000; 5)$

   (b) Montrer que cette loi peut être approchée par une loi binomiale, et préciser laquelle.

   $\sqrt{\frac{1042000-5}{1042000-1}} \simeq 1$, donc

   $X \approx \mathcal{B}(5 \quad ; 0{,}657 \quad)$.

   On utilisera dans la suite cette approximation par une loi binomiale.

   (c) Calculer $\mathbb{P}[X \leqslant 2]$.

   $$\begin{aligned}\mathbb{P}[X \leqslant 2] &= \mathbb{P}[X=0] + \mathbb{P}[X=1] + \mathbb{P}[X=2] \\ &= \tbinom{5}{0}(1-0{,}657)^5 + \tbinom{5}{1}(0{,}657)(1-0{,}657)^{5-1} + \tbinom{5}{2}(0{,}657)^2(1-0{,}657)^{5-2} \\ &\simeq 0{,}0047 + 0{,}0455 + 0{,}1742 \simeq 0{,}224\end{aligned}$$

   (d) Quelle est la probabilité qu'un échantillon aléatoire de cinq enseignants contienne au moins trois hommes ?
   C'est exactement la probabilité qu'on vient de calculer.

   Cette probabilité vaut 22,4 %.

---

3. On considère dans cette question un échantillon aléatoire de 71 enseignants, choisi sans remise, et on note Y le nombre d'hommes au sein de cet échantillon aléatoire.

   (a) Donner la loi de la variable Y.

   $Y \sim \mathcal{H}(1\,042\,000; 357\,000; 71)$

   (b) Montrer que cette loi peut être approchée par une loi normale, et préciser laquelle.
   On a $n = 71 > 30$, et $n\,p \simeq 24{,}3 > 5$, et $n(1-p) \simeq 46{,}7 > 5$ donc on peut approximer Y par $\mathcal{N}(24{,}3; 4)$
   où $4 \simeq \sqrt{n\,p(1-p)}\sqrt{\frac{N-n}{N-1}}$

   $Y \approx \mathcal{N}(24{,}3\,; 4 \quad)$.

   Dans la suite de cet exercice, on utilisera cette approximation en faisant une correction de continuité.

   (c) Calculer $\mathbb{P}[Y \geqslant 22]$. *On demande ici de détailler les calculs.*

   $$\begin{aligned}\mathbb{P}[Y \geqslant 22] &= \mathbb{P}[Y \geqslant 21{,}5] = \mathbb{P}[\tfrac{21{,}5-24{,}3}{4} < \tfrac{Y-24{,}3}{4}] \simeq \mathbb{P}[-0{,}7 < Z] \\ &\simeq 1 - F(-0{,}7) \simeq 1-(1-0{,}758) \simeq 1 - 0{,}242 \simeq 0{,}76\end{aligned}$$

   (d) Quelle est la probabilité qu'un échantillon aléatoire de 71 enseignants contienne moins de 30% d'hommes ?
   Sur 71 enseignants, 30% correspondent à 21,3 personnes, donc on calcule $\mathbb{P}[Y < 22] = 1 - \mathbb{P}[Y \geqslant 22] = 24\%$.

Exercice 2 : Test de mémoire

On soumet des patients atteints de la maladie de Parkinson à un test de mémoire. Les notes obtenues par les différents patients sont données ci-dessous : *On pourra supposer que ces notes sont issues d'une loi normale.*

Mme Vasseur : 8,1	Mr Begue : 6,9	Mr Coste : 4,1	Mr Dumas : 5,7	Mr Guilbert : 8,4
Mme Andre : 5,7	Mme Boulanger : 8,1	Mme Legros : 7,4	Mme Leger : 7,6	Mr Pelletier : 5,2
Mr Pereira : 9,6	Mr Fernandez : 6,6	Mme Seguin : 5,2	Mr Lelievre : 7,3	Mme Barthelemy : 8,9
Mme Merle : 3,9	Mr Marie : 2,5	Mme Dos Santos : 5,8	Mr Delage : 7,8	

1. Quelle est la taille de l'échantillon ? Combien de femmes l'échantillon comporte-t-il ? Et combien d'hommes ?

   $n = 19$
   Il y a 9 femmes et 10 hommes dans l'échantillon.

2. Calculer les moyennes et écarts types des notes obtenues par les hommes et par les femmes.
   (a) Notes des patients hommes :
   $m_e = 6{,}41$
   $s_e = 1{,}993$
   (b) Notes des patients femmes :
   $m_e = 6{,}744$
   $s_e = 1{,}561$

3. Peut-on affirmer, avec la confiance $c = 95\%$, que parmi l'ensemble des malades atteints de la maladie de Parkinson, les femmes ont en moyenne de meilleures notes que les hommes à ces tests de mémoire ?
   On attend un raisonnement construit à partir des outils vous semblant pertinents au sein du programme de statistiques. On détermine les intervalles de confiances suivant pour les moyennes des hommes et des femmes :

   Estimation de la moyenne de l'ensemble des femmes : dans l'intervalle $[5{,}471; 8{,}017]$.
   Estimation de la moyenne de l'ensemble des hommes : dans l'intervalle $[4{,}907; 7{,}913]$.
   Les intervalles se chevauchent, donc on ne peut pas conclure avec la confiance 95%

Exercice 3 : Épreuve de comptage

Des enfants sont soumis à une épreuve de comptage : on leur présente 9 objets, on leur demande de les compter (aussi vite que possible), et on enregistre le temps qu'il leur a fallu pour les compter (on note X ce temps, exprimé en secondes). *Note : Les questions 2a, 2b, et 2c sont complètement indépendantes de la question 1.*

1. Des études antérieures ont montré que sur l'ensemble des enfants de 12 ans, la moyenne et l'écart type de X sont $m = 12{,}09$ et $\sigma = 1{,}35$. On étudie dans cette question l'hypothèse selon laquelle $X \sim \mathcal{N}(12{,}09; 1{,}35)$.
   (a) Calculer le troisième quartile de X (sous l'hypothèse que $X \sim \mathcal{N}(12{,}09; 1{,}35)$). *Détailler les calculs et donner un résultat arrondi à 0,1 près.*
   $\mathbb{P}[X \geqslant Q_3] = \mathbb{P}[Z \geqslant \frac{Q_3 - 12{,}09}{1{,}35}] = 1 - F\left(\frac{Q_3 - 12{,}09}{1{,}35}\right)$ d'où l'équation $1 - F\left(\frac{Q_3 - 12{,}09}{1{,}35}\right) = 0{,}25$.
   On déduit $\frac{Q_3 - 12{,}09}{1{,}35} = z = 0{,}675$, qui donne
   $Q_3 \simeq z \times 1{,}35 + 12{,}09 \simeq 0{,}675 \times 1{,}35 + 12{,}09 \simeq 13{,}0$.

   (b) Au sein d'un échantillon aléatoire de 31 enfants de 12 ans (choisi avec remise), quelle serait la loi du nombre d'enfants mettant moins de 13 secondes à compter les 9 objets ?

   En notant Y ce nombre d'enfants, on aurait $Y \sim \mathcal{B}(31; 0{,}75)$, car on vient de calculer que $\mathbb{P}[X \leqslant 13] \simeq 0{,}75$.

2. On s'intéresse dans cette question à un échantillon d'enfants de 12 ans, dont les temps de réponse à l'épreuve de comptage sont donnés ci-dessous : *(On considérera que cet échantillon est choisi avec remise.)*

temps X (en secondes)	[9 ; 10[	[10 ; 11[	[11 ; 12[	[12 ; 13[	[13 ; 14[	[14 ; 15[
effectif	2	5	12	6	5	3

   (a) Au sein de cet échantillon, quelle proportion d'enfants met au moins 13 secondes à compter les 9 objets ?
   Parmi l'échantillon, 8 enfants mettent au moins 13 secondes à compter les 9 objets. Cela correspond à une proportion $\frac{8}{33} \simeq 0{,}242$. Cette proportion vaut 24,2 %.

   (b) En utilisant uniquement cet échantillon, estimer parmi l'ensemble des enfants de 12 ans la proportion de ceux qui mettent au moins 13 secondes à compter les 9 objets.
   On déterminera un intervalle de confiance avec la confiance 98%.

   On a $n = 33 > 30$, et $n\,p_e = 8 > 5$, et $n(1 - p_e) = 25 > 5$ donc on peut utiliser la pocédure du formulaire pour estimer la proportion p.
   On a $F(2{,}326) \simeq 0{,}99$ d'où $z_\alpha \simeq 2{,}326$ et $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}1735$.
   On estime donc que p est dans l'intervalle $[0{,}0689; 0{,}4159]$ avec la confiance $c = 0{,}98$.

(c) Comparer ce résultat à la question 1 : avec la confiance 98%, la proportion d'enfants mettant au moins 13 secondes à compter les 9 objets est-elle compatible avec l'hypothèse selon laquelle $X \sim \mathcal{N}(12{,}09; 1{,}35)$?

D'après la question 1, la proportion de ceux qui mettent au moins 13 secondes à compter les 9 objets (parmi l'ensemble des enfants de 12 ans) devrait être de 25%. Comme 0,25 est bien dans l'intervalle calculé en question 2,(b), il n'y a aucune contradiction : l'échantillon est compatible avec « $X \sim \mathcal{N}(12{,}09; 1{,}35)$ » .

Exercice 4 : Stress et orientation

On étudie le temps nécessaire à des rats de laboratoire pour sortir d'un labyrinthe, et l'influence du stress sur ce temps de parcours. Les rats sont soumis à un niveau de stress noté X, et le temps de parcours mesuré (et exprimé en secondes) est noté Y et reporté dans le tableau ci-dessous :

Sujet	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Niveau de stress (X)	0	0	0	0	1	1	1	4	4	4	4	9	9	9	9
Temps de parcours (Y)	8	7	9	11	11	10	16	16	12	18	20	17	16	15	18

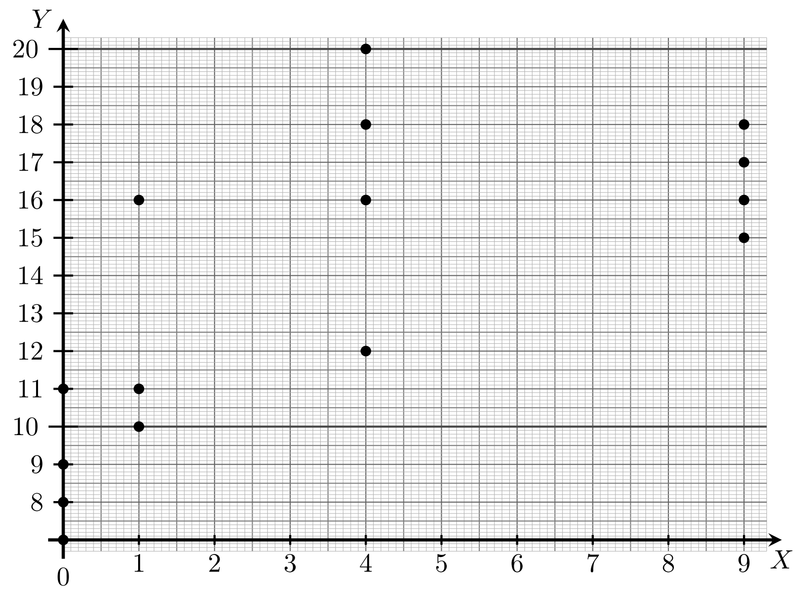
1. Calculer la moyenne et l'écart type de chacune des variables X et Y.
   $m(X) = 3{,}7$

   $s(X) = 3{,}6$

   $m(Y) = 13{,}6$

   $s(Y) = 4$

2. Représenter ces données par un nuage de points, sur le quadrillage ci-dessous :



3. **Droite de régression**

   (a) En détaillant les calculs, déterminer le coefficient de corrélation linéaire des variables X et Y.

   On calcule tout d'abord la covariance :
   $Cov(X,Y) = m(X\,Y) - m(X)m(Y) = \frac{895}{15} - \frac{55}{15}\frac{204}{15} = 9{,}8$

   $r(X;Y) = \frac{\mathrm{Cov}(X;Y)}{\mathrm{s}(X)\mathrm{s}(Y)} \simeq \frac{9{,}8}{3{,}6\times 4} \simeq 0{,}681.$

   (b) Déterminer la droite de régression qui exprime le temps de parcours en fonction du niveau de stress.

   on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{9{,}8}{12{,}62} \simeq 0{,}777$ et $b = m(Y) - a\,m(X) \simeq 13{,}6 - 0{,}777 \times 3{,}667 \simeq 10{,}751$

   $D_{Y|X} : Y = 0{,}777\,X + 10{,}751$

   (c) Pour un rat soumis à un niveau de stress $x = 5$, à quelle valeur y du temps de parcours s'attend-on ?
   On s'attend à $y \simeq 0{,}777 \times 5 + 10{,}751 \simeq 14{,}6$.