Contrôle Terminal – 28/06/2018 – Durée : 2h
Statistiques ($2^{\text{ème}}$ session) – 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 : Entraînement de chiens en prison

On étudie l'impact d'animaux domestiques (en l'occurrence des chiens) sur le niveau dépressif de femmes détenues.

1. On considère un groupe de 15 femmes détenues (volontaires pour participer à l'expérience), dont on note X le niveau dépressif (mesuré par le "test IPAT") et Y le niveau d'estime de soi (mesuré par le "test de Coopersmith"). On obtient les résultats suivants :

Individu	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Niveau dépressif (X)	12	14	45	11	11	13	11	12	14	36	14	12	11	11	45
Estime de soi (Y)	92	93	22	92	92	94	90	98	96	44	94	86	92	90	20
x_i'	7	11	14,5	3	3	9	3	7	11	13	11	7	3	3	14,5
y_i'	8,5	11	2	8,5	8,5	12,5	5,5	15	14	3	12,5	4	8,5	5,5	1
$(x_i'-y_i')^2$	2,25	0,0	156,25	30,25	30,25	12,25	6,25	64,0	9,0	100,0	2,25	9,0	30,25	6,25	182,25

(a) Calculer le coefficient de corrélation des rangs de spearmann. Que peut-on déduire de la valeur de ce coefficient ?

On calcule tout d'abord les rangs x_i' et y_i', entrés dans la table ci-dessus.
le coefficient de corrélation des rangs de Spearman est donc
$1-\left(6\,\frac{2{,}25+0+156{,}25+30{,}25+30{,}25+\cdots+182{,}25}{15\left(15^2-1\right)}\right)\simeq -0{,}144$
Cela ne met pas en évidence de lien significatif entre les deux variables.

(b) Calculer la moyenne et l'écart type du niveau dépressif au sein de cet échantillon.

moyenne : $m(X)=\frac{\sum x_i}{n}=\frac{12+14+\cdots+45}{15}=\frac{272}{15}\simeq 18{,}13$
$m(X^2)=\frac{\sum x_i^2}{n}=\frac{12^2+14^2+\cdots+45^2}{15}=\frac{7140}{15}$
$Var(X)=m(X^2)-m(X)^2=\frac{7140}{15}-\left(\frac{272}{15}\right)^2\simeq 147{,}18$
Écart-type : $s(X)=\sqrt{Var(X)}\simeq 12{,}13$

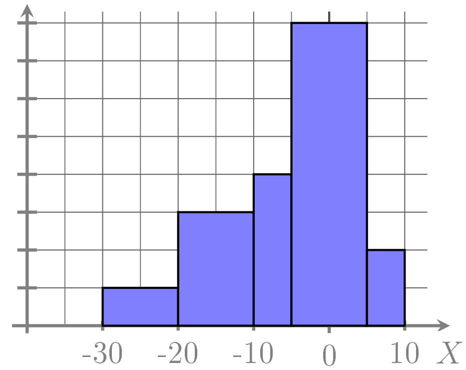
(c) Calculer la médiane du niveau d'estime de soi au sein de cet échantillon.

La médiane est la valeur numéro $\frac{15+1}{2}=8$ (en ordonnant par ordre croissant). C'est donc 92.

(d) Au sein de cet échantillon, quelle est la proportion de détenues dont le niveau dépressif dépasse 25 ?

> Cela correspond à 15 individus sur 3.
> La proportion demandée est donc $\frac{3}{15} = 0{,}2$.

2. On demande à chacune des 15 femmes de ce groupe d'entraîner un chien, destiné à aider des personnes âgées ou souffrant de troubles de la mobilité. On mesure la variation de leur niveau dépressif, c'est à dire la différence entre le niveau dépressif de chaque individu après avoir entraîné les chiens et le niveau du même individu avant l'entraînement. On admettra que cette différence est assez proche d'une loi normale. On obtient l'histogramme ci contre, réalisé sur un papier quadrillé :



(a) Déterminer les fréquences des différentes classes, et les reporter dans le tableau ci-dessous.

Niveau dépressif	$[-30\,;\,-20[$	$[-20\,;\,-10[$	$[-10\,;\,-5[$	$[-5\,;\,5[$	$[5\,;\,10[$
Fréquence	0,067	0,2	0,133	0,533	0,067

Remarque : Pour les calculer, on a compté la surface (le nombre de cases) de chaque rectangle (2, 6, 4, 16 et 2) et la surface totale ($2 + 6 + 4 + 16 + 2 = 30$). La fréquence de chaque classe est alors la surface du rectangle correspondant, divisé par la surface totale.

On ne demande pas de le démontrer, mais on pourrait calculer à partir de ces fréquences que le niveau dépressif au sein de l'échantillon a diminué en moyenne de 5,17 points à l'issue de l'entraînement du chien, avec un écart type de 8,44.

(b) Estimez, pour l'ensemble des femmes détenues, quelle serait la variation moyenne de leur niveau dépressif si on leur faisait entraîner ainsi un chien.

Vous déterminerez un intervalle de confiance avec la confiance $c = 95\%$.

> Comme $n = 15 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n - 1 = 14$ degrés de liberté (ddl)
> On lit $t_\alpha \simeq 2{,}1448$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 4{,}838$.
> On estime donc que μ est dans l'intervalle $[-10{,}01; -0{,}332]$ avec la confiance $c = 0{,}95$
> **Remarque :** La réponse $[0{,}332; 10{,}01]$ était accépté aussi même si elle ne met pas en évidence que la variation est négative

Exercice 2 : Taille et salaire

En France, le revenu moyen est d'environ 22500€ par an. On considère dans cet exercice un échantillon de 100 personnes "de grande taille", c'est à dire mesurant plus de 1m88. Leur revenu est donné ci-dessous :

Revenu (en milliers d'euros)	$[0\,;\,10[$	$[10\,;\,20[$	$[20\,;\,30[$	$[30\,;\,40[$	$[40\,;\,60[$	$[60\,;\,100[$
Effectif	8	41	33	9	3	6
Fréquence	0,08	0,41	0,33	0,09	0,03	0,06
Fréquence cumulée	0,08	0,49	0,82	0,91	0,94	1

1. Calculer les fréquences et fréquences cumulées, et le reporter dans le tableau.

2. Calculer la moyenne et l'écart type du revenu des personnes de cet échantillon.

Pour cette question, on vous demande d'indiquer les calculs effectués.

moyenne : $m(X) = \frac{\sum_i c_i n_i}{n} = \frac{5\times8+15\times41+25\times33+35\times9+50\times3+80\times6}{100} = \frac{2425}{100} = 24{,}25$
$m(X^2) = \frac{\sum_i c_i^2 n_i}{n} = \frac{5^2\times8+15^2\times41+25^2\times33+35^2\times9+50^2\times3+80^2\times6}{100} = \frac{86975}{100}$
$Var(X) = m(X^2) - m(X)^2 = \frac{86975}{100} - \left(\frac{2425}{100}\right)^2 \simeq 281{,}69$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 16{,}78$
Remarque : Ces valeurs sont en milliers d'euros, c'est à dire que le revenu moyen de l'échantillon est environ 24 250€ avec un écart type de 16 780€.

3. Déterminer de même la médiane.

Pour cette question, on vous demande à nouveau d'indiquer les calculs effectués.

Classe de la médiane : $[20\,;\,30[$
Méd $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}5 - F_X(a_i)) \simeq 20 + \frac{30-20}{0{,}82-0{,}49}(0{,}5-0{,}49) \simeq 20{,}3$

4. Estimer le revenu moyen de l'ensemble des français de plus de 1m88.

Vous déterminerez un intervalle de confiance, avec la confiance $c = 95\%$.

Comme n= 100 >30, on cherche z_α tel que $F(z_\alpha) = \frac{0{,}95+1}{2} = 0{,}975$
On lit sur la table inverse que pour la confiance 0,95 on a $z_\alpha \simeq 1{,}96$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 3{,}305$
On estime donc que μ est dans l'intervalle $[24{,}25 - 3{,}305\,;\, 24{,}25 + 3{,}305] \simeq [20{,}95\,;\, 27{,}55]$, avec la confiance $c = 0{,}95$.

5. Peut-on conclure qu'en moyenne, les français "de grande taille" ont un revenu plus élevé que le reste de la population française ?

Non : l'intervalle que l'on vient de calculer ne permet pas de savoir si le revenu moyen des personnes de grande taille est supérieur (ou inférieur) à la moyenne française (22,5)

Exercice 3 : Test de personnalité

Un test de personnalité est composé de 90 questions. Il est conçu de telle sorte qu'à chaque question, une personne "altruiste" ait 97% de chances de répondre *oui*, alors qu'une personne "égoïste" a 23% de chances de répondre *oui*.

1. On considère tout d'abord les scores d'un individu altruiste, dont on suppose qu'à chaque question il a 97% de chances de répondre *oui*. On note X le nombre de questions auxquelles il répond *oui*.
   (a) Argumenter que la loi de X est une loi binomiale (et précisez laquelle).

   On répète 90 fois l'expérience "répondre à une question du test" avec à chaque fois la probabilité 97% de répondre *oui*. Donc (en supposant que la réponse à une question n'influe pas sur les autres questions), le nombre de réponse *oui* suit la loi $\mathcal{B}(90; 0{,}97)$.

(b) Quelle est la moyenne de X ?

$m(X) = 90 \times 0{,}97 = 87{,}3$

(c) Calculer la probabilité $\mathbb{P}[X \geqslant 87]$.

$$\begin{aligned}\mathbb{P}[X \geqslant 87] &= \mathbb{P}[X = 87] + \mathbb{P}[X = 88] + \mathbb{P}[X = 89] + \mathbb{P}[X = 90] \\ &= \binom{90}{87}(0{,}97)^{87}(1-0{,}97)^{90-87} + \binom{90}{88}(0{,}97)^{88}(1-0{,}97)^{90-88} + \cdots + \binom{90}{90}(0{,}97)^{90}(1-(\\ &\simeq 0{,}2241 + 0{,}247 + 0{,}1795 + 0{,}0645 \simeq 0{,}715\end{aligned}$$

2. On considère ensuite les scores d'un individu égoïste, dont on suppose qu'à chaque question il a 23% de chances de répondre *oui*. On note Y le nombre de questions auxquelles il répond *oui*.

(a) Justifier que la loi de Y peut être approchée par une loi normale (et précisez laquelle).

$Y \sim \mathcal{B}(90\,;\,0{,}23)$.
On a $n = 90 > 30$, et $n\,p \simeq 20{,}7 > 5$, et $n(1-p) = 69{,}3 > 5$ donc on peut approximer Y par $\mathcal{N}(20{,}7; 3{,}99)$ où $3{,}99 \simeq \sqrt{n\,p(1-p)}$

(b) Calculer la probabilité $\mathbb{P}[Y \geqslant 45]$, en utilisant l'approximation par une loi normale et une "correction de continuité".

Pour cette question, on vous demande d'indiquer les calculs effectués, et d'utiliser la table du formulaire.

$$\begin{aligned}\mathbb{P}[Y \geqslant 45] &= \mathbb{P}[Y \geqslant 44{,}5] = \mathbb{P}[\tfrac{44{,}5-20{,}7}{3{,}99} < \tfrac{Y-20{,}7}{3{,}99}] \\ &\simeq \mathbb{P}[5{,}96 < Z] \simeq 1 - F(5{,}96) \\ &\simeq 1 - 1 \simeq 0\end{aligned}$$

3. Au vu de ces résultats, on considérera pour simplifier qu'avoir au moins 50% de réponses *oui* signifie que l'on est plutôt altruiste.

Dans cette question, on considère un échantillon de 200 personnes se qualifiant d'"épicuriens". On leur fait passer ce test, et on constate qu'au sein de cet échantillon, 72 individus ont au moins 50% de réponses *oui*. Estimer, parmi l'ensemble des personnes se qualifiant d'"épicuriens", la proportion d'individus plutôt altruiste (c'est à dire la proportion qui ont au moins 50% de réponses *oui*).

Vous déterminerez un intervalle de confiance, avec la confiance 95%.

On a $n = 200 > 30$, et $p_e = \frac{72}{200} = 0{,}36$, d'où $n\,p_e = 72 > 5$, et $n(1-p_e) = 128 > 5$ 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 $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}0665$.
On estime donc que p est dans l'intervalle $[0{,}2935; 0{,}4265]$ avec la confiance $c = 0{,}95$.