CC – 15/03/2018 – Durée : 1h30
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 une 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.

On s'intéresse au niveau d'anglais des Français, au travers des trois exercices ci-dessous, qui sont indépendants les uns des autres.

Afin d'éviter les biais liés à l'année de naissance (qui influe notamment sur le point auquel les individus ont été confrontés à une culture anglophone dans leur enfance), on considère uniquement des personnes nées en 1976.

Exercice 1 : Âge d'apprentissage de l'anglais

On s'intéresse tout d'abord à l'âge (noté X) auquel des français ont commencé à apprendre l'anglais. Pour un échantillon de 60 français nés en 1976, on obtient les données suivantes :

Âge	[1 ; 4[	[4 ; 7[	[7 ; 10[	[10 ; 13[	[13 ; 16[	[16 ; 19[
Effectif	5	12	5	21	15	2
Fréquence	0,083	0,2	0,083	0,35	0,25	0,033
Fréquence cumulée	0,083	0,283	0,366	0,716	0,966	0,999

1. Calculer les fréquences et les fréquences cumulées (inscrivez les dans les lignes vides de la table).
2. Calculer la moyenne et l'écart type de l'âge auquel les individus de l'échantillon ont commencé à apprendre l'anglais. *Dans cette question, on vous demande de justifier votre réponse en présentant le calcul effectué.*

moyenne : $m(X) = \frac{\sum_i c_i n_i}{n} = \frac{2{,}5\times5+5{,}5\times12+8{,}5\times5+11{,}5\times21+14{,}5\times15+17{,}5\times2}{60} = \frac{615}{60} = 10{,}25$

$m(X^2) = \frac{\sum_i c_i^2 n_i}{n} = \frac{2{,}5^2\times5+5{,}5^2\times12+8{,}5^2\times5+11{,}5^2\times21+14{,}5^2\times15+17{,}5^2\times2}{60} = \frac{7299}{60}$

$Var(X) = m(X^2) - m(X)^2 = \frac{7299}{60} - \left(\frac{615}{60}\right)^2 \simeq 16{,}59$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 4{,}07$

3. Quel est, eu sain de cet échantillon, l'âge médian de début d'apprentissage de l'anglais ? *Dans cette question, on vous demande de justifier votre réponse en présentant le calcul effectué.*

Classe de la médiane : [10 ; 13[

Méd $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}5 - F_X(a_i)) \simeq 10 + \frac{13-10}{0{,}716-0{,}366}(0{,}5 - 0{,}366) \simeq 11{,}15$

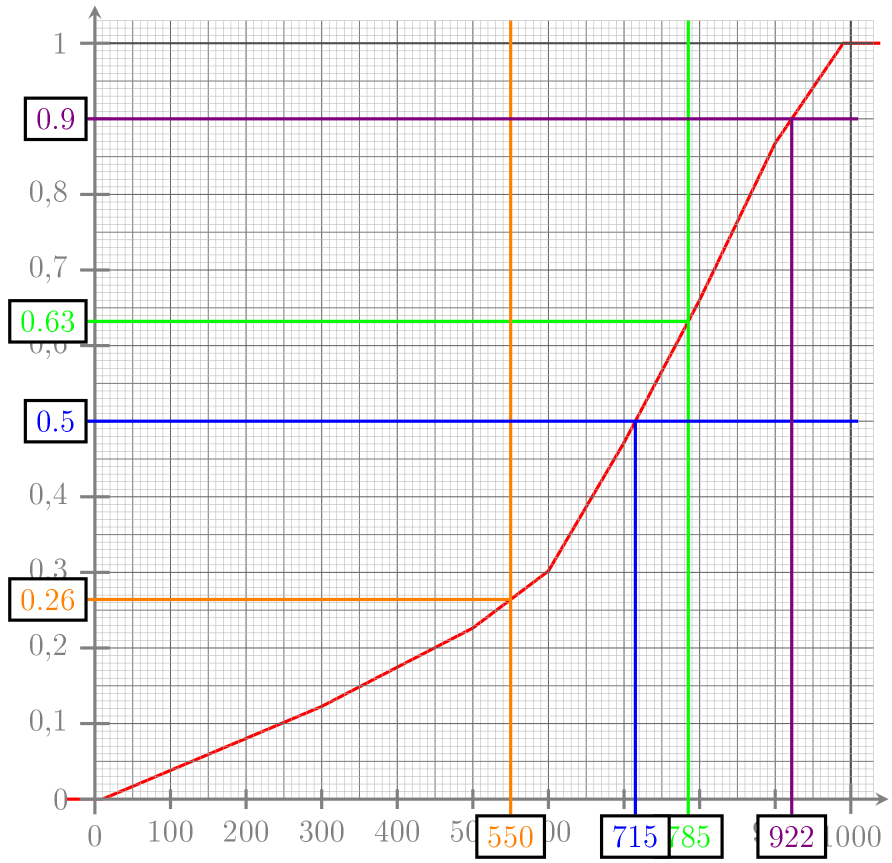
4. Calculer aussi les quartiles.

Classe du premier quartile : [4 ; 7[
$Q_1 \simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}25 - F_X(a_i)) \simeq 4 + \frac{7-4}{0{,}283-0{,}083}(0{,}25 - 0{,}083) \simeq 6{,}5$

Classe du troisième quartile : [13 ; 16[
$Q_3 \simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}75 - F_X(a_i)) \simeq 13 + \frac{16-13}{0{,}966-0{,}716}(0{,}75 - 0{,}716) \simeq 13{,}41$

Exercice 2 : Niveau d'anglais en France

On considère aussi le niveau d'anglais des français nés en 1976, en se basant sur un test très répandu : le TOEIC®, qui attribue à chaque individu une note entre 10 et 990. Parmi les nombreuses études effectuées sur le niveau des Français, l'une d'entre elle indique des données pour 106 français nés en 1976. À partir des données de cette étude, on réalise le graphique ci-contre (reproduit sur papier "millimétré"), qui est le polygone des fréquences cumulées.



1. Quel est le score médian au sein de cet échantillon ?

On lit sur le graphique (ajouté en bleu) le score 715 pour la fréquence cumulée 0,5.
La médiane est donc environ 715.

2. On considère qu'un niveau avancé (correspondant à au moins B2 dans le "Cadre européen commun de référence pour les langues") est atteint lorsque l'on a au moins la note 785 au TOEIC®.

Quelle est, au sein de cet échantillon, la proportion d'individus qui ont un niveau avancé en anglais ?

On lit sur le graphique (ajouté en vert) la fréquence 0,63 pour le score 785. Donc (en utilisant la lettre Y pour le score au TOEIC) $\mathbb{P}_r[Y \leqslant 785] \simeq 0{,}63$, d'où l'on déduit que $\mathbb{P}_r[Y \geqslant 785] \simeq 1 - 0{,}63 = 0{,}37$.
La proportion demandée vaut donc environ 0,37.

3. Le niveau B1 est un niveau où l'on commence à être indépendant et correspond aux scores entre 550 et 785 points.

   Déterminer quel est, au sein de cet échantillon, la proportion d'individus qui ont le niveau B1.

On vient de voir que $\mathbb{P}_r[Y \leqslant 785] \simeq 0{,}63$. De plus on lit sur le graphique (ajouté en orange) que $\mathbb{P}_r[Y \leqslant 550] \simeq 0{,}26$.
La proportion demandée est donc $\mathbb{P}_r[550 \leqslant Y \leqslant 785] \simeq 0{,}63 - 0{,}26 = 0{,}37$.

4. Combien d'individus, parmi ceux de l'échantillon, ont un niveau inférieur à B1 ?

On a déjà vu que $\mathbb{P}_r[Y \leqslant 550] \simeq 0{,}26$. Cette proportion correspond donc environ à $106 \times 0{,}26 \simeq 28$ individus

5. On souhaite caractériser les 10% d'individus ayant les meilleures notes. Compléter la phrase ci-dessous pour la rendre correcte :

On lit sur la graphique (en violet) le score 922 pour la fréquence cumulée 0,9 (c'est à dire que 90% de l'échantillon a moins que ce score, tandis que 10% a plus que ce score). Ainsi,

Dans cet échantillon, les 10% d'individus ayant la meilleure note sont ceux qui ont au moins la note 922 .

Exercice 3 : Lien entre le niveau d'anglais et l'âge de début d'apprentissage
On souhaite désormais déterminer s'il y a un fort lien entre le niveau en anglais et l'âge auquel on a commencé à apprendre l'anglais. On décide donc de faire passer le TOEIC® à un échantillon de 16 français nés en 1976. On note par X l'âge auquel ils ont commencé à apprendre l'anglais et par Y le score qu'ils obtiennent, et on obtient les données ci-dessous :

X	5	11	2	11	13	14	3	13	6	7	11	4	11	7	5	13
Y	70	80	205	390	395	655	720	755	615	750	225	645	65	920	680	175
x'_i	4,5	10,5	1	10,5	14	16	2	14	6	7,5	10,5	3	10,5	7,5	4,5	14
y'_i	2	3	5	7	8	11	13	15	9	14	6	10	1	16	12	4

1. Calculer la moyenne et l'écart type de la variable X au sein de l'échantillon.

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{5+11+\cdots+13}{16} = \frac{136}{16} = 8{,}5$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{5^2+11^2+\cdots+13^2}{16} = \frac{1400}{16}$
$Var(X) = m(X^2) - m(X)^2 = \frac{1400}{16} - \left(\frac{136}{16}\right)^2 = 15{,}25$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 3{,}91$

2. Calculer le coefficient de corrélation linéaire des variables X et Y. Que peut-on conclure quant au lien entre ces deux variables ?

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{70+80+\cdots+175}{16} = \frac{7345}{16} \simeq 459{,}06$
$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{70^2+80^2+\cdots+175^2}{16} = \frac{4629925}{16}$
$Var(Y) = m(Y^2) - m(Y)^2 = \frac{4629925}{16} - \left(\frac{7345}{16}\right)^2 \simeq 78\,631{,}93$
Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 280{,}41$

$m(X\,Y) = \frac{\sum x_i\, y_i}{n} = \frac{5\times70+11\times80+...+13\times175}{16} = \frac{59035}{16} \simeq 3\,689{,}688$
$Cov(X{,}Y) = m(X\,Y) - m(X)\,m(Y) = \frac{59035}{16} - \frac{136}{16}\frac{7345}{16} \simeq -212{,}344$
$r(X{,}Y) = \frac{Cov(X{,}Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-212{,}344}{\sqrt{15{,}25\times78\,631{,}934}} \simeq -0{,}194$
Cela ne met pas en évidence de lien linéaire significatif entre les deux deux variables.

3. Calculer le coefficient de corrélation des rangs ("de Spearman") des variables X et Y. Que peut-on conclure quant au lien entre ces deux variables ?

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
$1 - \left(6\,\frac{6{,}25+56{,}25+16+12{,}25+36+\cdots+100}{16\left(16^2-1\right)}\right) \simeq -0{,}049$
Cela ne met pas en évidence de lien significatif entre les deux variables.