

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.

Un groupe de chercheurs est spécialisé dans l'étude de la dépression chez l'adolescent. Ils étudient plus précisément les symptômes dépressifs chez les lycéens français, au moyen d'une nouvelle méthode de mesure de l'état dépressif. Cette méthode s'appuie à la fois sur un questionnaire psychologique et sur des dosages neuro-endocriniens, et elle attribue à chaque individu un score (noté S), qui peut prendre n'importe quelle valeur entre 0 et 60. Elle a été construite de façon à ce que les propriétés de ce score (moyenne, écart type, etc) soient assez proches de ce que donne une échelle pré-existante : l'échelle de Montgomery-Åsberg.

Exercice 1 : Sévérité des épisodes dépressifs

Comme premier test de la pertinence de leur méthode de mesure, ils l'appliquent à un échantillon de 65 lycéens atteints d'épisodes dépressifs majeurs. Ils obtiennent les score suivants :

Score (état dépressif)	[30 ; 35[[35 ; 40[[40 ; 45[[45 ; 50[[50 ; 55[[55 ; 60[
effectif	19	18	20	5	1	2
Fréquence	0,292	0,277	0,308	0,077	0,015	0,031
Fréquence cumulée	0,292	0,569	0,877	0,954	0,969	1

- Indiquez la population et la variable statistique considérées. Quel est le type de cette variable ?

La variable statistique est le score S , elle est quantitative continue. La population est les lycéens français, et plus précisément (dans cet exercice) les lycéens français atteints d'épisodes dépressifs majeurs.

- Calculer les fréquences et les fréquences cumulées. *Vous les remplirez dans les lignes disponibles du tableau*
- Calculer la moyenne et l'écart type de cette variable au sein de l'échantillon. *Dans cette question on vous demande de justifier votre réponse en présentant le calcul effectué.*

$$\begin{aligned} \text{moyenne : } m(S) &= \frac{\sum_i c_i n_i}{n} = \frac{32,5 \times 19 + 37,5 \times 18 + 42,5 \times 20 + 47,5 \times 5 + 52,5 \times 1 + 57,5 \times 2}{65} = \frac{2547,5}{65} \simeq 39,19 \\ m(S^2) &= \frac{\sum_i c_i^2 n_i}{n} = \frac{32,5^2 \times 19 + 37,5^2 \times 18 + 42,5^2 \times 20 + 47,5^2 \times 5 + 52,5^2 \times 1 + 57,5^2 \times 2}{65} = \frac{102156,25}{65} \\ \text{Var}(S) &= m(S^2) - m(S)^2 = \frac{102156,25}{65} - \left(\frac{2547,5}{65}\right)^2 \simeq 35,6 \\ \text{Écart-type : } s(S) &= \sqrt{\text{Var}(S)} \simeq 5,97 \end{aligned}$$

4. Quelle est, au sein de l'échantillon, la proportion d'individus dont le score est supérieur à 35 ?

$\mathbb{P}_r[S > 35] = 1 - \mathbb{P}_r[S \leq 35] \simeq 1 - 0,292 = 0,708$ (où on a utilisé la fréquence cumulée de la 1^{ère} colonne).

5. Calculer la médiane de S sur cet échantillon. *Pour cette médiane, on vous demande de justifier votre réponse en présentant le calcul effectué*

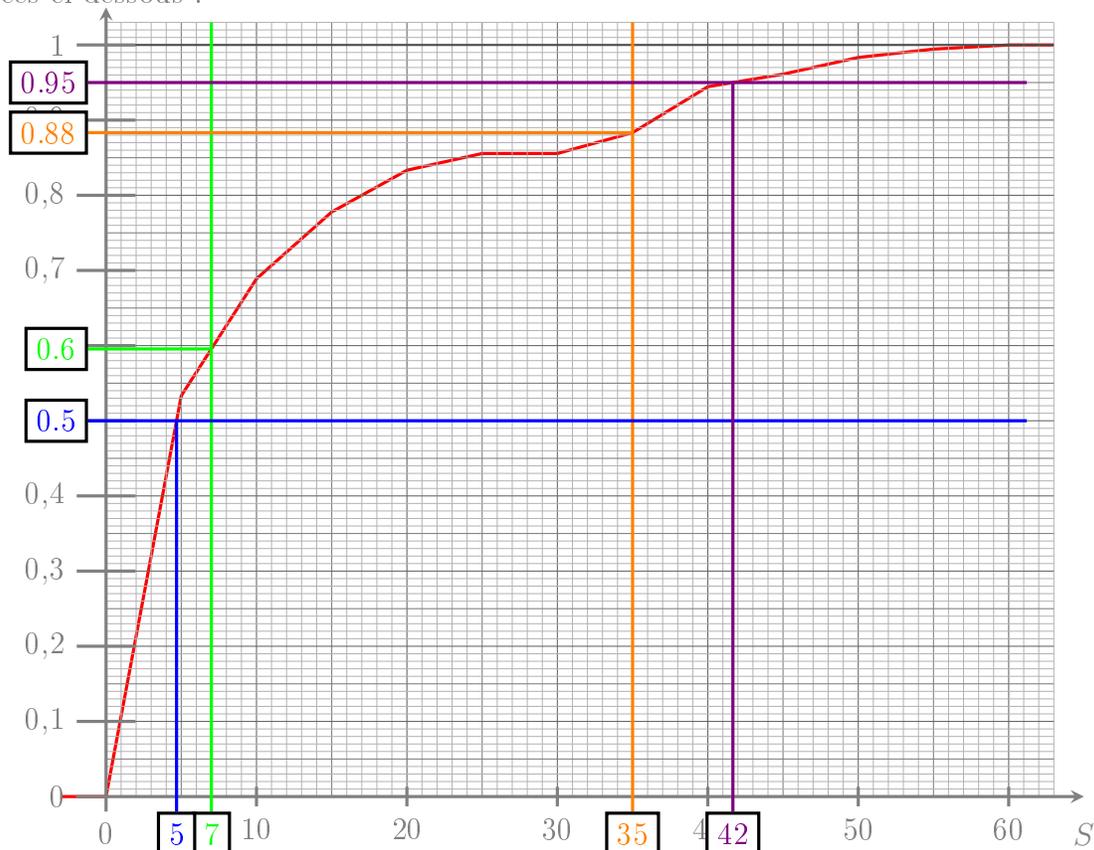
Classe de la médiane : $[35; 40[$
 $\text{Méd} \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,5 - F_X(a_i)) \simeq 35 + \frac{40 - 35}{0,569 - 0,292} (0,5 - 0,292) \simeq 38,75$

6. Calculer enfin son premier quartile.

Classe du premier quartile : $[30; 35[$
 $Q_1 \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,25 - F_X(a_i)) \simeq 30 + \frac{35 - 30}{0,292 - 0} (0,25 - 0) \simeq 34,28$

Exercice 2 : Prévalence de la dépression

Les chercheurs étudient ensuite l'ensemble des lycéens, pour déterminer la prévalence de la dépression : ils considèrent cette fois-ci les scores (notés S) attribués par leur méthode à un échantillon de 180 lycéens français choisis au hasard. Ils obtiennent le polygone des fréquences cumulées ci-dessous :



1. À partir du graphique, déterminer le score médian.

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

2. On s'intéresse aux 5% de lycéens qui ont le score le plus élevé. Compléter la phrase ci-dessous pour les décrire :

On lit sur la graphique (en violet) le score 42 pour la fréquence cumulée 0,95 (c'est à dire que 95% de l'échantillon a moins que ce score, tandis que 5% a plus que ce score). Ainsi,
Dans cet échantillon, les 5% d'individus ayant le plus grand score sont ceux qui ont au moins le score 42 .

3. On interprète les scores de la façon suivante :

Score	$0 \leq S < 7$	$7 \leq S < 20$	$20 \leq S < 35$	$S \geq 35$
État du patient	Normal (sans symptômes)	dépression légère	dépression modérée	dépression sévère

- (a) Quelle proportion des lycéens de l'échantillon souffrent d'un état dépressif léger ou modéré ?

On lit sur le graphique (ajouté en vert) la fréquence 0,6 pour le score 7, et (en orange) la fréquence 0,88 pour le score 35.
Donc $\mathbb{P}_r[7 \leq S < 35] = \mathbb{P}_r[S < 35] - \mathbb{P}_r[S < 7] \simeq 0,88 - 0,6 = 0,28$.

- (b) Au sein de cet échantillon, combien de lycéens souffrent d'une dépression sévère ?

On calcule d'abord la proportion $\mathbb{P}_r[S \geq 35] = 1 - \mathbb{P}_r[S < 35] \simeq 1 - 0,88 = 0,12$.
On en déduit ensuite que le nombre de lycéens souffrant d'une dépression sévère est d'environ $180 \times 0,12 \simeq 22$.

Exercice 3 : Observance et efficacité des traitements

Parmi les lycéens souffrant d'épisodes dépressifs de sévérité modérée, certains se sont vus proposer un traitement médicamenteux. Pour 10 d'entre eux, les chercheurs ont mesuré leur état dépressif avant et après deux mois de traitement : ils notent Y la diminution du score à l'issue du traitement.

De plus ils demandent à chacun s'il a respecté précisément la prescription. La réponse est résumée par un nombre X entre 0 et 3, selon la règle ci dessous :

"prescription suivie très précisément"	$\rightsquigarrow X = 3$
"prescription plutôt bien suivie"	$\rightsquigarrow X = 2$
"prescription peu suivie"	$\rightsquigarrow X = 1$
"aucune prise de médicament"	$\rightsquigarrow X = 0$

Ils obtiennent alors les données suivantes :

Individu	A	B	C	D	E	F	G	H	I	J
Observance (X)	1	1	0	3	0	3	2	2	0	0
Variation (Y)	7	6	2	14	0	12	3	4	10	-5
x'_i	5,5	5,5	2,5	9,5	2,5	9,5	7,5	7,5	2,5	2,5
y'_i	7	6	3	10	2	9	4	5	8	1

On notera que comme Y est la diminution du score à l'issue du traitement, lorsqu' Y est négatif cela traduit une agmentation du score.

- Calculer la moyenne et l'écart type de la variable Y .

$$\begin{aligned} \text{moyenne : } m(Y) &= \frac{\sum x_i}{n} = \frac{7+6+\dots+(-5)}{10} = \frac{53}{10} = 5,3 \\ m(Y^2) &= \frac{\sum x_i^2}{n} = \frac{7^2+6^2+\dots+(-5)^2}{10} = \frac{579}{10} \\ \text{Var}(Y) &= m(Y^2) - m(Y)^2 = \frac{579}{10} - \left(\frac{53}{10}\right)^2 = 29,81 \\ \text{Écart-type : } s(Y) &= \sqrt{\text{Var}(Y)} \simeq 5,46 \end{aligned}$$

- (a) Quelle est la moyenne de Y pour les individus qui n'ont pris aucun médicament ?

Les individus qui n'ont pris aucun médicament (c'est à dire ceux qui ont " $X = 0$ ") sont les individus C, E, I et J.
La moyenne de Y sur ces individus vaut $\frac{2+0+10-5}{4} = 1,75$.

- (b) Quelle est la moyenne de Y pour les individus qui ont au moins "plutôt bien" suivi la prescription ?

Les individus qui ont " $X \geq 2$ " sont les individus D, F, G et H.
La moyenne de Y sur ces individus vaut $\frac{14+12+3+4}{4} = 8,25$.

Bien qu'on l'ait résumée par un nombre, la variable X est plutôt une variable qualitative ordinale, de sorte qu'il n'est pas très pertinent d'en calculer la moyenne et l'écart type.

- Calculer la médiane de la variable X .

La médiane est la valeur numéro $\frac{10+1}{2} = 5,5$ (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 5 et 6. C'est donc le milieu entre 1 et 1 c'est-à-dire $\frac{1+1}{2} = 1$.
En d'autres termes la médiane est « prescription peu suivie »

- Afin de déterminer s'il y a un lien entre les variable X et Y , calculez le coefficient de corrélation des rangs de Spearman. Que peut-on en conclure ?

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
$$1 - \left(6 \frac{2,25+0,25+0,25+0,25+0,25+\dots+2,25}{10(10^2-1)}\right) \simeq 0,67$$

Cela suggère un lien entre les deux variables, où Y augmente quand X augmente.