

Nom :  
 Prénom :

*Le formulaire et la calculatrice sont autorisés. Vous rendrez l'énoncé auquel vous pouvez joindre une copie si vous le souhaitez. Vous pouvez soit répondre directement 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 : Résultats scolaires d'élèves en situation de mal-être

Cet exercice est indépendant de l'exercice 2 (inutile d'avoir réussi l'un pour traiter l'autre). Toutefois, ils portent tous deux sur le dépistage scolaire du mal-être des élèves. Les données sont fictives mais inspirées de l'article *L'accompagnement scolaire des souffrances adolescentes*, de Sidonie Vacher (publié dans *Agora débats/jeunesses*, 2024).

Parmi les élèves de sixième d'un même collège, il y en a 12 dont le mal-être a été détecté par l'institution scolaire.

On désigne par  $X$  la note (la moyenne générale sur leur dernier bulletin de note) de chaque collégien·ne, et on reporte ci-dessous la note des 12 collégien·nes dont le mal-être a été détecté :

12,5	6	11	4,2	11,3	8,8	6,4	9,7	11,7	12,3	15	8
------	---	----	-----	------	-----	-----	-----	------	------	----	---

1. Donner la moyenne des notes de ces 12 collégien·ne.s.

$$m(X) = \frac{\sum x_i}{n} = \frac{12,5+6+11+\dots+8}{12} = \frac{116,9}{12} \simeq 9,74$$

2. Donner aussi leur écart-type.

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{12,5^2+6^2+11^2+\dots+8^2}{12} = \frac{1248,25}{12}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{1248,25}{12} - \left(\frac{116,9}{12}\right)^2 \simeq 9,12$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 3,02$$

3. Donner enfin la médiane.

On ordonne les valeurs par ordre croissant :

4,2    6    6,4    8    8,8    9,7    11    11,3    11,7    12,3    12,5    15

La médiane est la valeur numéro  $\frac{12+1}{2} = 6,5$ , ou plutôt le milieu entre les valeurs numéro 6 et 7. C'est donc le milieu entre 9,7 et 11 c'est-à-dire  $\frac{9,7+11}{2} = 10,35$ .

## Exercice 2 : Dépistage du mal-être des élèves

On s'interroge sur les biais de détection de cette souffrance par l'institution scolaire (qui pourrait ignorer la souffrance de « *bons élèves* » pour se focaliser sur les élèves dont le mal-être limite leur propre possibilité d'apprentissage et/ou nuit à l'ensemble de la classe).

On décide alors d'un diagnostic de l'ensemble des 103 élèves de sixième de ce collège (via un questionnaire inspiré de *l'échelle de Kandel*) et on obtient qu'en tout il y en a 26 qui sont en situation de mal-être (parmi lesquels seuls les 12 de l'exercice 1 avaient attiré l'attention du personnel du collège).

On désigne à nouveau par  $X$  la note (la moyenne sur leur dernier bulletin de note) de chaque collégien·ne, et on reporte ci-dessous la note des 26 collégien·nes en situation de mal-être :

Note $X$	[4 ; 6[	[6 ; 8[	[8 ; 10[	[10 ; 12[	[12 ; 14[	[14 ; 16[
Effectif	1	2	3	5	8	7
Fréquence	0,038	0,077	0,115	0,192	0,308	0,269
Fréq. Cum.	0,038	0,115	0,231	0,423	0,731	1,0

*Vous pouvez utiliser les lignes vides du tableau pour vos propres calculs.*

1. Parmi l'ensemble des collégien·ne·s auxquel·le·s on a fait passer le questionnaire, quelle proportion sont en situation de mal-être ?

Cette proportion est  $\frac{26}{103} \simeq 0,2524$ .

2. Parmi les collégien·ne·s dont le questionnaire a indiqué qu'ils sont en situation de mal-être, quelle proportion a une note moyenne au moins égale à 12 ?

Ceux qui ont une moyenne au moins égale à 12 correspondent au deux dernières colonnes du tableau. Il sont  $8 + 7 = 15$ . Cette proportion est donc  $\frac{15}{26} \simeq 0,5769$ .

3. Calculer la note moyenne au sein de l'échantillon des collégiens en situation de mal-être. *Dans cette question, on vous demande de présenter le calcul effectué.*

$$m(X) = \frac{\sum c_i n_i}{n} = \frac{5 \times 1 + 7 \times 2 + 9 \times 3 + \dots + 15 \times 7}{26} = \frac{310}{26} \simeq 11,92$$

4. Donner aussi l'écart-type.

$$m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{5^2 \times 1 + 7^2 \times 2 + 9^2 \times 3 + \dots + 15^2 \times 7}{26} = \frac{3898}{26}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{3898}{26} - \left(\frac{310}{26}\right)^2 \simeq 7,76$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 2,79$$

5. À partir de la formule du formulaire, trouver la médiane approchée de cette note. *Dans cette question, on vous demande de présenter le calcul effectué.*

On commence par calculer les fréquences et fréquences cumulées  
(on les a insérées dans le tableau en début d'exercice).

Classe de la médiane : [12 ; 14[

$$\text{Méd} \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,5 - F_X(a_i)) \simeq 12 + \frac{14 - 12}{0,731 - 0,423} (0,5 - 0,423) \simeq 12,5$$

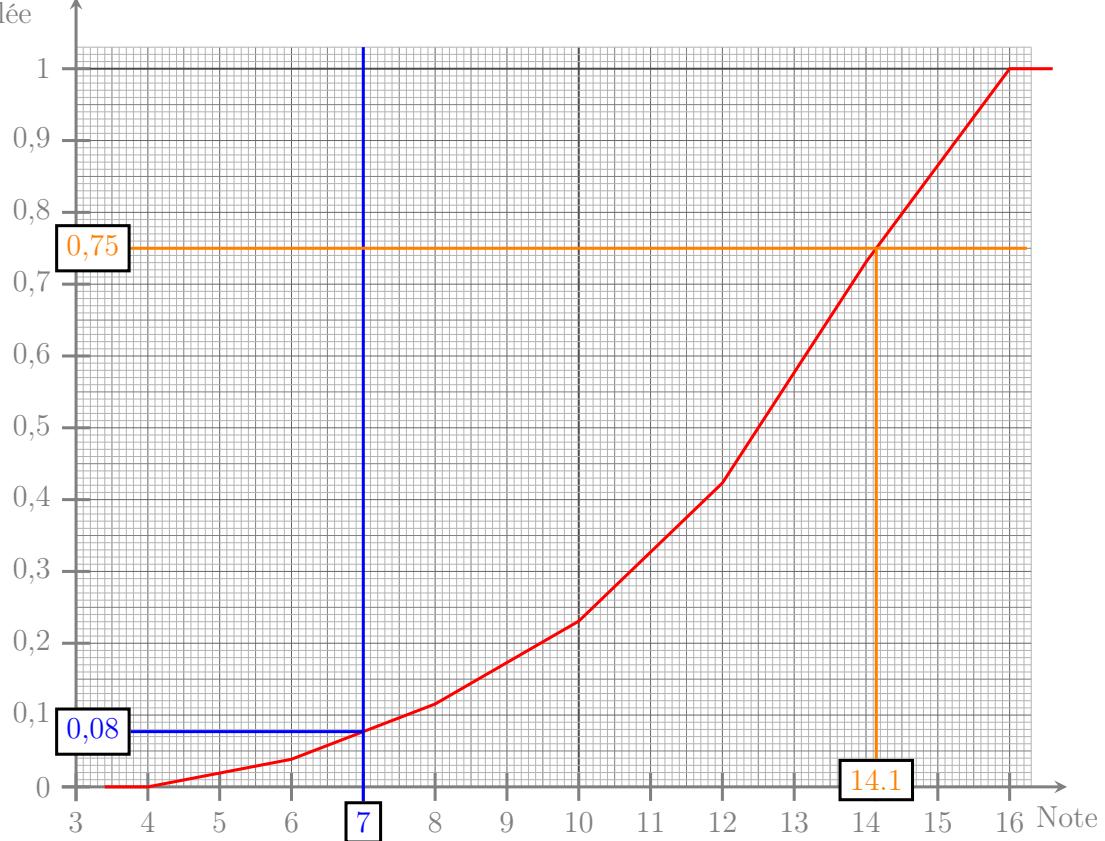
6. De même, combien vaut (environ) le premier quartile ?

Classe du premier quartile :  $[10 ; 12[$

$$Q_1 \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)}(0,25 - F_X(a_i)) \simeq 10 + \frac{12 - 10}{0,423 - 0,231}(0,25 - 0,231) \simeq 10,198$$

7. On trace ci-dessous le polygone des fréquences cumulées des notes au sein de cet échantillon.

Fréquence cumulée



(a) Déterminer graphiquement une valeur approchée du troisième quartile.

Comme indiqué en orange sur le graphique ci-avant, on lit la note 14,1 pour la fréquence cumulée 75%. Donc le troisième quartile est 14,1.

(b) Au sein de cet échantillon, déterminer les deux proportions  $\mathbb{P}_r[X \leqslant 7]$  et  $\mathbb{P}_r[X \geqslant 7]$ .

Comme indiqué en bleu sur le graphique ci-avant, on lit la proportion 0,08 pour la note 7. Donc  $\mathbb{P}_r[X \leqslant 7] \simeq 0,08$ .

En conséquence,  $\mathbb{P}_r[X \geqslant 7] = 1 - \mathbb{P}_r[X < 7] \simeq 1 - 0,08 = 0,92$  (on a considéré que  $\mathbb{P}_r[X \leqslant 7] = \mathbb{P}_r[X < 7]$  car les données sont regroupées en classes, on les manipule comme une variable continue).

### Exercice 3 : Santé mentale de jeunes adultes selon l'âge d'obtention du premier smartphone

Une ONG nommée "Sapien Labs" cherche à savoir s'il existe un lien entre la santé mentale de jeunes adultes et l'âge auquel ils ont eu leur premier *terminal mobile* (terme qui regroupe les tablettes et les *smartphones*). Cette ONG utilise un score appelé "Quotient de Santé mentale" (que l'on notera  $Y$ ), qui constitue une auto-évaluation par chaque individu de sa propre santé mentale (un questionnaire à choix multiples donne un score allant de -100 à 200 ; les notes négatives traduisant des troubles mentaux significatifs). On désignera par  $X$  l'âge (en années) auquel un individu a eu son premier terminal mobile.

Pour se prémunir contre les différences de santé mentales liées au genre, on choisit un échantillon qui ne comporte que des femmes : au sein d'un échantillon de 10 jeunes femmes (âgées de 18 à 24 ans), on obtient les données ci-dessous.

$X$	13	12	11	14	10	16	13	11	16	8
$Y$	12	-30	34	-12	14	36	11	4	40	-23
$x'_i$	6,5	5	3,5	8	2	9,5	6,5	3,5	9,5	1
$y'_i$	6	1	8	3	7	9	5	4	10	2
$(x'_i - y'_i)^2$	0,25	16	20,25	25	25	0,25	2,25	0,25	0,25	1

On a laissé des lignes vides au cas où vous ayez besoin d'y écrire des données pour vos calculs.

1. Quelles sont les variables statistiques étudiées ? Donner la nature de chacune d'entre elles. Sont elles appariées ?

Les variables statistiques sont

- L'âge  $X$  d'obtention du premier terminal mobile. C'est une variable quantitative. En principe elle est continue (l'âge n'a pas à être pile un nombre entier d'années) même si en pratique on a considéré des valeurs entières.
- Le quotient de Santé mentale  $Y$ . C'est une variable quantitative discrète (car issue d'un questionnaire à choix multiples).

Ces variables sont toutes les deux définies pour les même individus (les 10 jeunes femmes de l'échantillon), donc elles sont appariées.

2. Calculer le coefficient de corrélation des rangs entre  $X$  et  $Y$ . Quelles conclusions peut-on tirer de ce coefficient ? (présenter les calculs effectués)

On calcule tout d'abord les rangs, que l'on a ajouté dans le tableau au début de l'exercice.

Le coefficient de corrélation des rangs vaut donc  $1 - \left( 6 \times \frac{0,25+16+20,25+25+25+\dots+1}{10(10^2-1)} \right) \simeq 0,4515$

Ce coefficient n'est pas assez proche de 1 pour parler de fort lien entre les variables. Il suggère qu'il y ait quand même un lien, mais qui n'est pas très fort. Il indique que la santé mentale  $Y$  a tendance à être meilleure quand l'âge  $X$  auquel on a eu son premier smartphone est plus élevé.

3. Calculer le coefficient de corrélation linéaire entre  $X$  et  $Y$ . *Dans cette question on demande de présenter les calculs effectués.* Quelles conclusions peut-on tirer de ce coefficient ?

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{13+12+11+\dots+8}{10} = \frac{124}{10} = 12,4$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{13^2+12^2+11^2+\dots+8^2}{10} = \frac{1596}{10}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{1596}{10} - \left(\frac{124}{10}\right)^2 = 5,84$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 2,42$$

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{12+(-30)+34+\dots+(-23)}{10} = \frac{86}{10} = 8,6$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{12^2+(-30)^2+34^2+\dots+(-23)^2}{10} = \frac{6102}{10}$$

$$Var(Y) = m(Y^2) - m(Y)^2 = \frac{6102}{10} - \left(\frac{86}{10}\right)^2 = 536,24$$

$$\text{Écart-type : } s(Y) = \sqrt{Var(Y)} \simeq 23,16$$

$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{13 \times 12 + 12 \times -30 + \dots + 8 \times -23}{10} = \frac{1361}{10} = 136,1$$

$$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{1361}{10} - \frac{124}{10} \frac{86}{10} = 29,46$$

$$r(X,Y) = \frac{Cov(X,Y)}{s(X)s(Y)} = \frac{29,46}{2,42 \times 23,16} \simeq 0,525\,6$$

Ce coefficient n'est pas assez proche de 1 pour parler de fort lien linéaire entre les variables. Il suggère qu'il y ait quand même un lien, mais qui n'est pas très fort ou qui n'est pas linéaire (ne correspond pas à un droite), et où la santé mentale  $Y$  a tendance à être meilleure quand l'âge  $X$  auquel on eu a son premier smartphone est plus élevé.

4. Cette ONG, "Sapien Labs", a publié en mai 2023 une étude sur le lien entre santé mentale et âge d'obtention du premier terminal mobile. L'échantillon est beaucoup plus grand que dans les questions précédentes : il porte sur 27 969 jeunes adultes (de 18 à 24 ans) dont 15 995 femmes. Pour chaque âge d'obtentention du premier terminal mobile (6ans, 7ans, 8ans, etc), la moyenne des quotients de santé mentale de plusieurs dizaines ou centaines de jeunes femmes est calculé.

À partir de leurs données, on peut calculer les coefficients de corrélation entre l'âge d'obtentention du premier terminal mobile et le quotient de santé mentale moyen. Pour les habitantes de pays occidentaux, on obtient un coefficient de corrélation linéaire de 0,9631 et un coefficient de corrélation des rangs de 0,989.

À partir de ces coefficients, peut-on conclure qu'il y a une relation de cause à effet, où un quotient de santé mentale moyen plus faible serait la conséquence d'une obtention plus précoce de son premier terminal mobile ?

Non, en aucun cas des données ne peuvent, à elles seules montrer un lien de cause à effet.

**Remarque sur la dernière question :** Il n'était pas demandé de beaucoup argumenter sur la dernière question, mais pour les personnes qui liraient ce corrigé, quelques commentaires peuvent aider à mieux comprendre cette question :

- Les données montrent qu'il y a un lien particulièrement fort entre l'âge d'obtention du premier terminal numérique et la santé mentale moyenne une fois (jeune) adulte. Mais cela ne peut pas, en soit, montrer qu'il y a un lien de cause à effet. En effet, il y a plein d'hypothèses que l'on pourrait formuler pour expliquer qu'il y ait un lien entre les deux variables :
  - On pourrait d'une part, comme le fait Sapien Labs, penser que les smartphones sont directement la cause d'une mauvaise santé mentale, par la façon dont ils modifient les interactions sociales.
  - On pourrait aussi imaginer que leur détresse psychique ait été déjà présente dans l'enfance et que la seule façon dont les parents aient alors su s'en sortir ait été de leur donner un smartphone plus rapidement (auquel cas  $Y$  serait plutôt la cause et  $X$  la conséquence).
  - Il se pourrait aussi qu'il y ait un troisième facteur (par exemple le milieu social) qui influence à la fois l'âge d'obtention du premier smartphone et la santé mentale, et dans cette troisième hypothèse  $X$  et  $Y$  seraient toutes deux les conséquences d'une même propriété (le milieu social).
- Les données permettent juste de constater qu'il y a un fort lien entre les variables (les points sont presque alignés), mais elle ne permettent absolument pas de distinguer entre ces trois hypothèses, donc la réponse est « **Non, ces données ne permettent pas de conclure qu'il y ait une relation de cause à effet, où  $X$  serait la cause et  $Y$  la conséquence.** »
- Cela sort du cadre de ce cours, mais dans les données qui aboutissent à ces coefficients de corrélation de 0,9631 et 0,989, le mode de calcul est légèrement différent de celui qu'on a vu en cours : en calculant la moyenne des quotients de santé mentale des personnes qui avaient eu leur premier smartphone à un âge donné, on aboutit à des coefficients de corrélation qui sont artificiellement beaucoup plus proches de 1 ou de -1 que ce qui a été fait en cours.