

Devoir surveillé — Information chiffrée : statistique bivariée

1re ens. sci. — 60 min

Durée : 60 min — Barème : 20 points. Calculatrice selon les consignes de l'énoncé. La rédaction et la clarté comptent.

Exercice 1 (5 points) — Sport et licence en club

On a interrogé les 250 élèves d'un lycée pour savoir s'ils sont licenciés dans un club sportif. Le tableau croisé d'effectifs ci-dessous, partiellement rempli, résume les réponses.

	Licenciés	Non licenciés	Total
Filles	60	...	150
Garçons	...	40	...
Total	...	...	250

- (1,5 point) Recopier et compléter le tableau.
- (1 point) Calculer la fréquence marginale des élèves licenciés dans le lycée. Exprimer cette fréquence en pourcentage et l'interpréter par une phrase.
- (1 point) Parmi les filles du lycée, quelle est la fréquence des élèves licenciées dans un club ? Interpréter ce résultat par une phrase.
- (1,5 point) À la lecture du résultat précédent, un rédacteur du journal du lycée écrit : « 40 % des licenciés du lycée sont des filles. » A-t-il raison ? Justifier par un calcul et, si nécessaire, corriger sa phrase.

Corrigé

1. On complète le tableau à l'aide des totaux en ligne et en colonne.

Sur les 150 filles du lycée, 60 sont licenciées, donc le nombre de filles non licenciées est $150 - 60 = 90$.

Le lycée compte 250 élèves dont 150 filles, donc le nombre total de garçons est $250 - 150 = 100$. Parmi ces 100 garçons, 40 ne sont pas licenciés, donc le nombre de garçons licenciés est $100 - 40 = 60$.

On en déduit les totaux des colonnes : $60 + 60 = 120$ élèves licenciés et $90 + 40 = 130$ élèves non licenciés. On vérifie que $120 + 130 = 250$: le compte est bon.

	Licenciés	Non licenciés	Total
Filles	60	90	150
Garçons	60	40	100
Total	120	130	250

2. La fréquence marginale des élèves licenciés s'obtient en divisant l'effectif total des licenciés par l'effectif total du lycée :

$$f = \frac{120}{250} = 0,48$$

soit 48 %.

Interprétation : dans ce lycée, 48 % des élèves sont licenciés dans un club sportif.

3. On cherche ici une fréquence conditionnelle : le groupe de référence n'est plus l'ensemble du lycée, mais uniquement les filles. Parmi les 150 filles, 60 sont licenciées, donc :

$$f = \frac{60}{150} = 0,4$$

soit 40 %.

Interprétation : parmi les filles du lycée, 40 % sont licenciées dans un club sportif.

4. Le rédacteur n'a pas raison. Sa phrase porte sur les licenciés du lycée : le groupe de référence est donc l'ensemble des 120 licenciés, et non l'ensemble des 150 filles. Calculons la fréquence des filles parmi les licenciés :

$$f = \frac{60}{120} = 0,5$$

soit 50 %.

Les deux phrases ne portent pas sur le même groupe de référence : « 40 % des filles sont licenciées » se calcule par rapport aux 150 filles, tandis que « 40 % des licenciés sont des filles » se calculerait par rapport aux 120 licenciés. Le rédacteur a confondu ces deux fréquences conditionnelles.

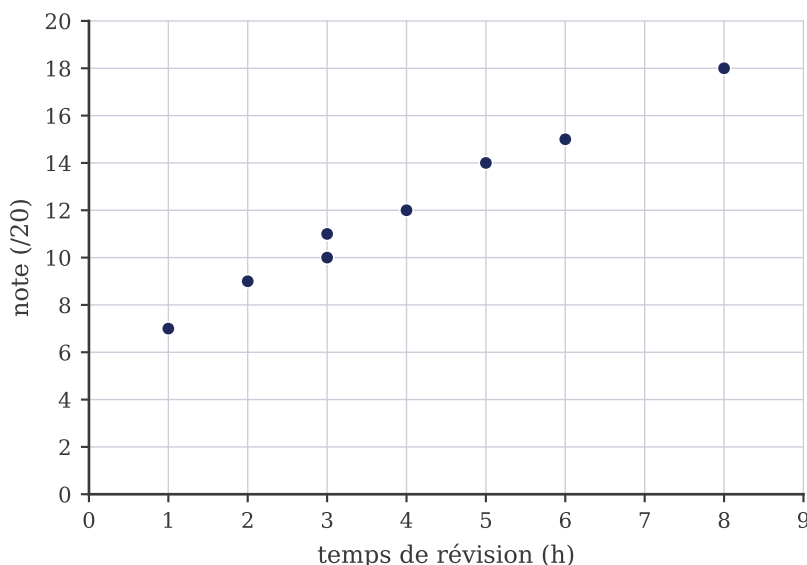
Phrase corrigée : « 50 % des licenciés du lycée sont des filles. »

Exercice 2 (5 points) — Réviser, ça rapporte ?

Huit élèves d'une classe de 1re ont noté leur temps de révision hebdomadaire x (en heures) avant un contrôle, puis leur note y (sur 20) à ce contrôle.

Temps de révision x (en h)	1	2	3	3	4	5	6	8
Note y (sur 20)	7	9	10	11	12	14	15	18

Le nuage de points correspondant est représenté ci-dessous.



- (1 point) Décrire l'allure du nuage de points : que semble-t-il se passer pour la note lorsque le temps de révision augmente ?
- (1,5 point) Calculer les coordonnées du point moyen G du nuage. Détailler les calculs.
- (1 point) Deux élèves ont révisé exactement 3 heures. Lire sur le graphique les notes qu'ils ont obtenues. Que peut-on en conclure ?
- (1,5 point) Un élève affirme : « Ce graphique prouve que réviser davantage fait forcément augmenter la note. » Expliquer en quelques phrases pourquoi cette conclusion est trop rapide.

Corrigé

1. Les points du nuage sont presque alignés et le nuage « monte » de la gauche vers la droite : lorsque le temps de révision augmente, la note semble augmenter elle aussi. Le nuage suggère donc une tendance croissante : plus un élève a révisé, plus sa note est élevée, en général.

2. Le point moyen G du nuage a pour abscisse la moyenne $\bar{x}$ des temps de révision et pour ordonnée la moyenne $\bar{y}$ des notes.

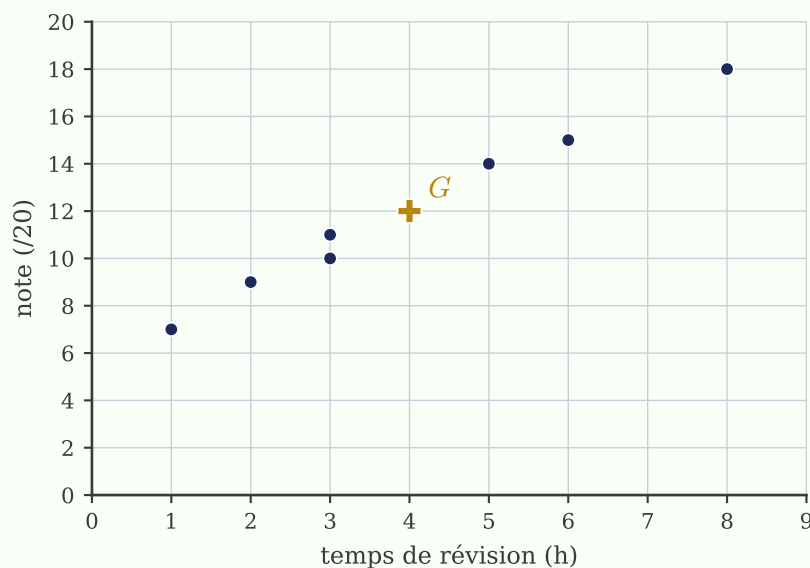
Calcul de l'abscisse :

$$\bar{x} = \frac{1 + 2 + 3 + 3 + 4 + 5 + 6 + 8}{8} = \frac{32}{8} = 4$$

Calcul de l'ordonnée :

$$\bar{y} = \frac{7 + 9 + 10 + 11 + 12 + 14 + 15 + 18}{8} = \frac{96}{8} = 12$$

Le point moyen du nuage est donc $G(4; 12)$: en moyenne, les huit élèves ont révisé 4 heures et obtenu 12 sur 20.



3. On lit sur le graphique les ordonnées des deux points d'abscisse 3 : les deux élèves qui ont révisé exactement 3 heures ont obtenu 10 et 11.

Conclusion : à temps de révision égal, deux élèves peuvent obtenir des notes différentes. La note ne dépend donc pas uniquement du temps de révision : d'autres facteurs entrent en jeu.

4. Cette conclusion est trop rapide car le graphique montre seulement un **lien apparent** entre le temps de révision et la note : il ne prouve pas que l'un est la cause de l'autre.

D'abord, d'autres explications sont possibles : les élèves déjà à l'aise avec le chapitre révisent peut-être davantage par goût pour la matière, l'attention en classe ou le sommeil jouent aussi sur la note. Ensuite, la question 3 a montré qu'à temps de révision égal, les notes peuvent différer : le temps de révision n'explique pas tout. Enfin, l'échantillon est très réduit : huit élèves d'une seule classe ne permettent pas de généraliser.

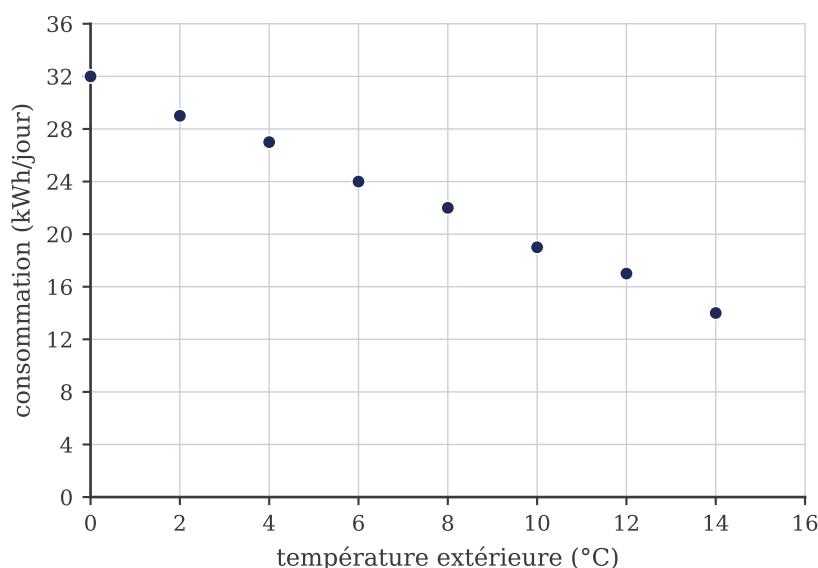
On peut donc dire que réviser davantage **semble** associé à de meilleures notes dans cette classe, mais pas que cela fait « forcément » augmenter la note.

Exercice 3 (6 points) — Chauffage et température

Une famille relève, pendant 8 journées d'hiver, la température extérieure moyenne du jour x (en °C) et la consommation électrique de sa maison y (en kWh par jour), chauffage électrique compris. Les journées sont classées par température croissante.

Température x (en °C)	0	2	4	6	8	10	12	14
Consommation y (en kWh/jour)	32	29	27	24	22	19	17	14

Le nuage de points correspondant est représenté ci-dessous.



- (1,5 point) On partage la série en deux moitiés : les 4 premiers points du tableau d'une part, les 4 derniers d'autre part. Calculer les coordonnées du point moyen G_1 de la première moitié et du point moyen G_2 de la seconde.
- (1,5 point) Déterminer l'équation réduite de la droite (G_1G_2) , appelée droite de Mayer de la série.
- (1 point) Interpréter le coefficient directeur de cette droite dans le contexte de l'exercice.
- (1 point) À l'aide de l'équation de la droite, estimer la consommation de la maison pour une journée où la température extérieure moyenne est de 7 °C. Comment s'appelle cette démarche ?
- (1 point) Un membre de la famille utilise la même équation pour prévoir la consommation lors d'une journée de printemps à 20 °C. Calculer la valeur obtenue, puis expliquer pourquoi cette prévision n'est pas fiable.

Corrigé

- La première moitié de la série est formée des 4 premiers points du tableau : (0 ; 32), (2 ; 29), (4 ; 27) et (6 ; 24). Le point moyen partiel G_1 a pour coordonnées les moyennes des abscisses et des ordonnées de ces points :

$$\bar{x}_1 = \frac{0 + 2 + 4 + 6}{4} = \frac{12}{4} = 3 \quad \text{et} \quad \bar{y}_1 = \frac{32 + 29 + 27 + 24}{4} = \frac{112}{4} = 28$$

Donc $G_1(3 ; 28)$.

La seconde moitié est formée des 4 derniers points : (8 ; 22), (10 ; 19), (12 ; 17) et (14 ; 14). De même :

$$\bar{x}_2 = \frac{8 + 10 + 12 + 14}{4} = \frac{44}{4} = 11 \quad \text{et} \quad \bar{y}_2 = \frac{22 + 19 + 17 + 14}{4} = \frac{72}{4} = 18$$

Donc $G_2(11; 18)$.

2. La droite de Mayer est la droite (G_1G_2) . Son équation réduite est de la forme $y = mx + p$.
Calcul du coefficient directeur :

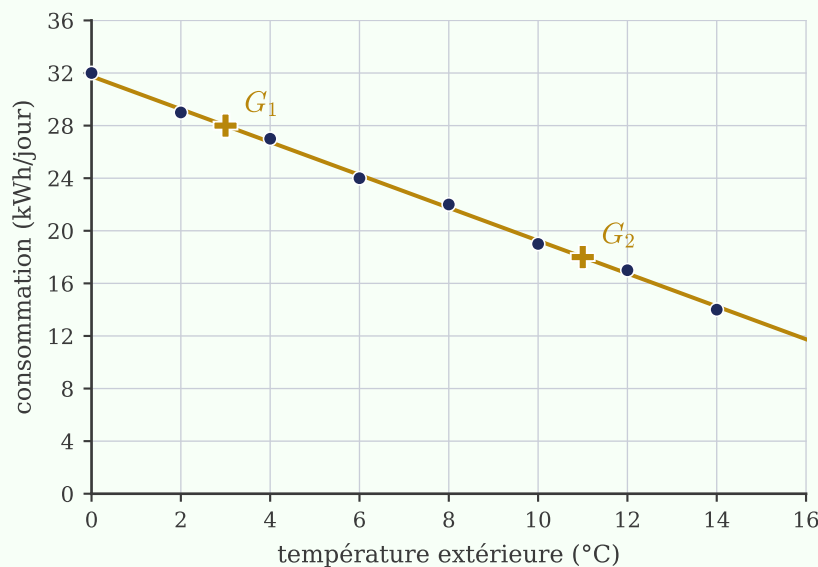
$$m = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{18 - 28}{11 - 3} = \frac{-10}{8} = -1,25$$

Calcul de l'ordonnée à l'origine : la droite passe par $G_1(3; 28)$, donc $28 = -1,25 \times 3 + p$, c'est-à-dire $28 = -3,75 + p$, d'où $p = 28 + 3,75 = 31,75$.

La droite de Mayer de la série a donc pour équation réduite :

$$y = -1,25x + 31,75$$

Vérification avec G_2 : $-1,25 \times 11 + 31,75 = -13,75 + 31,75 = 18$, qui est bien l'ordonnée de G_2 .



3. Le coefficient directeur de la droite vaut $-1,25$. Il est négatif, ce qui traduit une tendance décroissante : quand la température augmente, la consommation diminue.

Interprétation dans le contexte : d'après ce modèle, lorsque la température extérieure moyenne augmente de $1\text{ }^{\circ}\text{C}$, la consommation électrique de la maison diminue d'environ $1,25\text{ kWh}$ par jour.

4. La température de $7\text{ }^{\circ}\text{C}$ se situe à l'intérieur de la plage des températures observées (de $0\text{ }^{\circ}\text{C}$ à $14\text{ }^{\circ}\text{C}$). On remplace x par 7 dans l'équation de la droite :

$$y = -1,25 \times 7 + 31,75 = -8,75 + 31,75 = 23$$

On estime donc la consommation à 23 kWh par jour pour une journée à $7\text{ }^{\circ}\text{C}$.

Cette démarche, qui consiste à estimer une valeur **à l'intérieur** de la plage des données observées, s'appelle une **interpolation**.

5. On remplace x par 20 dans l'équation de la droite :

$$y = -1,25 \times 20 + 31,75 = -25 + 31,75 = 6,75$$

Le modèle prévoit une consommation de $6,75\text{ kWh}$ par jour.

Mais $20\text{ }^{\circ}\text{C}$ est **en dehors** de la plage des températures observées ($0\text{ }^{\circ}\text{C}$ à $14\text{ }^{\circ}\text{C}$) : cette démarche est une **extrapolation**, et rien ne garantit que la tendance observée en hiver se poursuive au printemps. En effet, à $20\text{ }^{\circ}\text{C}$ le chauffage est éteint : la consommation ne peut plus baisser au

même rythme et se stabilise autour d'une consommation de base (électroménager, eau chaude, éclairage...). La baisse linéaire de 1,25 kWh par degré n'a donc plus de sens à cette température : le modèle n'est fiable que sur la plage des températures pour lesquelles il a été construit.

Exercice 4 (4 points) — Esprit critique

Pour chaque affirmation, dire si elle est **vraie** ou **fausse**, puis justifier soigneusement la réponse. Une réponse non justifiée ne rapporte aucun point.

1. (1,5 point) Le lycée A compte 400 élèves et le lycée B en compte 1 200. Dans le diagramme circulaire des régimes du lycée A, le secteur des demi-pensionnaires occupe la moitié du disque (50 %); dans celui du lycée B, il occupe le quart du disque (25 %). Affirmation : « Comme le secteur est deux fois plus grand au lycée A, il y a plus de demi-pensionnaires au lycée A qu'au lycée B. »
2. (1 point) Affirmation : « Une extrapolation est toujours aussi fiable qu'une interpolation. »
3. (1,5 point) Affirmation : « Les coordonnées du point moyen d'un nuage sont la moyenne des abscisses et la moyenne des ordonnées des points du nuage, mais le point moyen n'est pas forcément l'un des points du nuage. »

Corrigé

1. L'affirmation est **fausse**.

Un diagramme circulaire représente des **proportions** (des fréquences), pas des effectifs. Pour comparer les effectifs, il faut tenir compte du nombre total d'élèves de chaque lycée. Calculons les effectifs de demi-pensionnaires :

- Lycée A : 50 % de 400 élèves, soit $0,5 \times 400 = 200$ demi-pensionnaires ;
- Lycée B : 25 % de 1 200 élèves, soit $0,25 \times 1\,200 = 300$ demi-pensionnaires.

Or $200 < 300$: il y a donc **plus** de demi-pensionnaires au lycée B qu'au lycée A, alors que son secteur est deux fois plus petit. Un secteur plus grand signifie une proportion plus grande, pas un effectif plus grand, car les totaux des deux lycées sont différents.

2. L'affirmation est **fausse**.

Une interpolation consiste à estimer une valeur **à l'intérieur** de la plage des données observées : le modèle y a été construit à partir de valeurs réellement mesurées, l'estimation est donc raisonnablement fiable. Une extrapolation consiste au contraire à estimer une valeur **en dehors** de cette plage : rien ne garantit que la tendance observée se poursuive au-delà des données. Une extrapolation est donc en général **moins** fiable qu'une interpolation, et elle peut même donner des résultats absurdes. L'exercice 3 en fournit un exemple : à 20 °C, le chauffage est éteint et le modèle linéaire construit sur des journées d'hiver ne s'applique plus.

3. L'affirmation est **vraie**.

Par définition, le point moyen d'un nuage de points est le point $G(\bar{x}; \bar{y})$, où $\bar{x}$ est la moyenne des abscisses des points du nuage et $\bar{y}$ la moyenne de leurs ordonnées : la première partie de l'affirmation est exacte.

De plus, le point moyen ne coïncide pas forcément avec l'un des points du nuage. Justifions-le par un exemple : pour le nuage formé des deux points (0 ; 10) et (2 ; 20), on a $\bar{x} = \frac{0+2}{2} = 1$ et $\bar{y} = \frac{10+20}{2} = 15$, donc $G(1; 15)$, qui n'est aucun des deux points du nuage. La seconde partie de l'affirmation est donc exacte elle aussi.