

Plaquette 1 – Nuage de points et point moyen

Statistiques à deux variables — Première générale

Corrections détaillées

Boîte à outils

Nuage de points et point moyen.

- **Série à deux variables.** À chaque individu i correspond un couple $(x_i; y_i)$. Le **nuage de points** est l'ensemble de ces points dans un repère.
- **Point moyen.** Le point G a pour coordonnées les deux moyennes :

$$G(\bar{x}; \bar{y}) \quad \text{avec} \quad \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{et} \quad \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

- **Propriété essentielle.** Toute droite d'ajustement obtenue par la méthode des moindres carrés ou par la méthode de Mayer **passse par G** .
- **Méthode de Mayer.** On partage la série en deux groupes de même effectif (les premières valeurs de x , puis les dernières), on calcule les points moyens G_1 et G_2 de chaque groupe, et la droite (G_1G_2) sert d'ajustement. Elle passe toujours par G , qui est le milieu de $[G_1G_2]$ lorsque les deux groupes ont le même effectif.
- **Lire un nuage.** Trois observations à faire, dans cet ordre : le nuage est-il **allongé** (on pourra l'ajuster par une droite) ? la tendance est-elle **croissante** ou **décroissante** ? y a-t-il un **point aberrant** à l'écart des autres ?
- **Vocabulaire.** Une corrélation est dite **positive** quand y augmente avec x , **négative** quand y diminue quand x augmente.
- **⚠ Un point aberrant déplace le point moyen** et fausse l'ajustement : on le signale toujours, et on précise si on le conserve ou non.

Correction de l'exercice 1 – Construire un nuage de points

a) On place les six points de coordonnées (1 ; 12), (2 ; 15), (3 ; 19), (4 ; 22), (5 ; 27) et (6 ; 31). Le nuage obtenu est tracé dans le corrigé graphique ci-dessous.

b) Les six points sont **presque alignés** : le nuage est nettement allongé, dans une direction qui monte de la gauche vers la droite. La tendance est donc **croissante** : le chiffre d'affaires augmente avec les années.

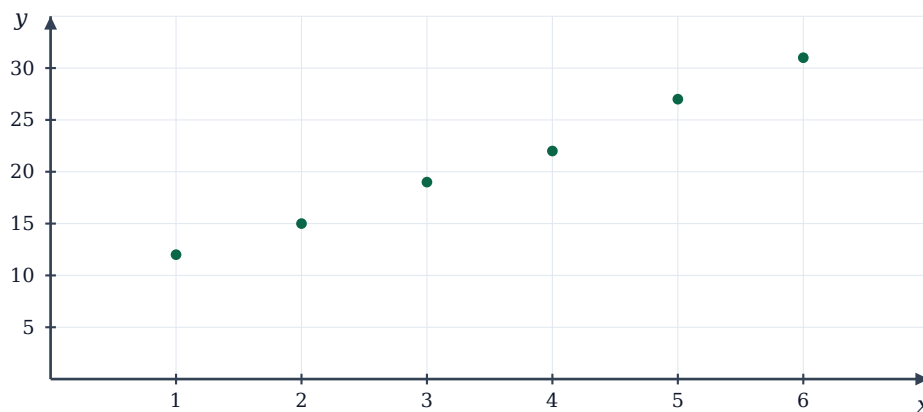
c) Oui. Un ajustement affine est envisageable précisément parce que le nuage est allongé selon une direction et que les points s'écartent peu de cette direction. On peut d'ailleurs le constater sur les écarts annuels :

$$15 - 12 = 3 \quad ; \quad 19 - 15 = 4 \quad ; \quad 22 - 19 = 3 \quad ; \quad 27 - 22 = 5 \quad ; \quad 31 - 27 = 4$$

Les augmentations tournent toutes autour de 4 milliers d'euros par an : la croissance est presque régulière, ce qui est la signature d'un modèle affine.

Ce qu'il ne faut pas conclure. Un nuage allongé autorise un ajustement affine, il ne le **prouve** pas : c'est un choix de modélisation, à valider ensuite par le coefficient de corrélation (plaquette 3).

Réponse. b) nuage allongé, tendance croissante — c) oui, les augmentations annuelles sont toutes voisines de 4.



Le nuage des six points : il est allongé et croissant.

Correction de l'exercice 2 - Calculer les coordonnées du point moyen

a) La moyenne des abscisses est la somme divisée par l'effectif :

$$\bar{x} = \frac{2 + 4 + 6 + 8 + 10}{5} = \frac{30}{5} = 6$$

b) De même pour les ordonnées :

$$\bar{y} = \frac{5 + 9 + 12 + 18 + 21}{5} = \frac{65}{5} = 13$$

c) Le point moyen est donc

$$G(6; 13)$$

Le contrôle de vraisemblance. Le point moyen doit toujours se situer « au milieu » du nuage : $\bar{x} = 6$ est bien entre 2 et 10, et $\bar{y} = 13$ entre 5 et 21 ✓ Si l'une des deux moyennes sortait de l'intervalle des données, c'est qu'il y aurait une erreur d'addition.

L'erreur de division. On divise par l'effectif $n = 5$, c'est-à-dire le nombre de **couples**, et non par 10 (le nombre total de valeurs écrites dans le tableau).

Réponse. a) $\bar{x} = 6$ — b) $\bar{y} = 13$ — c) $G(6; 13)$.

Correction de l'exercice 3 - Un nuage de huit points

a) On calcule les deux moyennes, avec $n = 8$:

$$\bar{x} = \frac{1 + 2 + 3 + 4 + 5 + 6 + 7 + 8}{8} = \frac{36}{8} = 4,5$$

$$\bar{y} = \frac{3 + 5 + 4 + 8 + 9 + 11 + 10 + 14}{8} = \frac{64}{8} = 8$$

Le point moyen est donc $G(4,5 ; 8)$.

b) Non. Si les points étaient alignés, l'augmentation de y serait la même à chaque fois que x augmente de 1. Or :

$$\text{de } x = 2 \text{ à } x = 3: 4 - 5 = -1 \quad \text{de } x = 3 \text{ à } x = 4: 8 - 4 = +4$$

Une fois la note baisse, une fois elle monte de 4 points : le nuage n'est pas aligné, il est seulement **allongé** dans une direction croissante.

c) La corrélation est **positive** : globalement, plus le nombre d'heures de révision est grand, plus la note est élevée. Les deux « accidents » ($x = 3$ et $x = 7$, où la note baisse) n'inversent pas cette tendance d'ensemble.

La prudence à garder. Une corrélation positive ne dit pas que réviser une heure de plus rapporte un nombre fixe de points : elle décrit une tendance moyenne sur l'ensemble des élèves, pas une garantie individuelle.

Réponse. a) $G(4,5 ; 8)$ — b) non : la note baisse de 1 point puis monte de 4 — c) corrélation positive.

Correction de l'exercice 4 - Lire un nuage de points

a) La température la plus basse est 2°C , et le point correspondant a pour ordonnée 20 : c'est le point $(2 ; 20)$. La plus haute est 14°C , avec le point $(14 ; 5)$.

b) Quand x augmente, y diminue : la tendance est **décroissante**, c'est-à-dire que la corrélation est **négative**. Dans le contexte, c'est parfaitement logique : plus il fait chaud dehors, moins on consomme de gaz pour se chauffer.

c) Le point d'abscisse 6 a pour ordonnée 15 : pour une température de 6°C , la consommation est de **15 m^3 par jour**. C'est une lecture directe, puisque cette température figure parmi les relevés.

Ce que le nuage permet de dire. Les sept points sont presque alignés sur une droite descendante : la consommation baisse d'environ $1,25 \text{ m}^3$ par degré gagné. C'est ce coefficient que la droite d'ajustement fournira précisément dans la plaquette suivante.

Réponse. a) $(2 ; 20)$ et $(14 ; 5)$ — b) décroissante : plus il fait chaud, moins on consomme — c) 15 m^3 par jour.

Correction de l'exercice 5 - Reconnaître trois types de nuages

a) Les écarts successifs de y valent tous 2 pour un pas de 1 en x : les points sont **parfaitement alignés**. Un ajustement affine est non seulement pertinent, il est **exact** — la droite est $y = 2x + 1$. La corrélation est **positive**.

b) Les écarts valent $-2, -1, -2, -2$: presque constants et négatifs. Le nuage est allongé et descendant, donc un ajustement affine est **pertinent**, avec une corrélation **négative**.

c) Les ordonnées font 6, 2, 9, 3, 7 : elles montent, descendent, remontent sans régularité. Le nuage est **dispersé**, sans direction privilégiée : un ajustement affine ne serait **pas pertinent**. On peut le voir en calculant le point moyen,

$$\bar{x} = \frac{15}{5} = 3 \quad \bar{y} = \frac{6+2+9+3+7}{5} = \frac{27}{5} = 5,4$$

mais ce point moyen ne résume rien : les points s'en écartent dans toutes les directions.

Ce que la question c) enseigne. On peut **toujours** calculer une droite d'ajustement, même sur un nuage informe. C'est le nuage, puis le coefficient de corrélation, qui disent si cette droite a un sens.

Réponse. a) alignement parfait, corrélation positive — b) ajustement pertinent, corrélation négative — c) nuage dispersé, ajustement non pertinent.

Correction de l'exercice 6 - La méthode de Mayer

a) Pour les quatre premiers points, avec $n = 4$:

$$\bar{x}_1 = \frac{1+2+3+4}{4} = 2,5 \quad \bar{y}_1 = \frac{3+5+4+8}{4} = \frac{20}{4} = 5$$

donc $G_1(2,5; 5)$. Pour les quatre derniers :

$$\bar{x}_2 = \frac{5+6+7+8}{4} = 6,5 \quad \bar{y}_2 = \frac{9+11+10+14}{4} = \frac{44}{4} = 11$$

donc $G_2(6,5; 11)$.

b) Le coefficient directeur de la droite (G_1G_2) est

$$m = \frac{11-5}{6,5-2,5} = \frac{6}{4} = 1,5$$

Son équation s'écrit alors $y = 1,5(x - 2,5) + 5$, soit

$$y = 1,5x - 3,75 + 5 = 1,5x + 1,25$$

c) On remplace x par 4,5 :

$$y = 1,5 \times 4,5 + 1,25 = 6,75 + 1,25 = 8$$

On obtient bien 8, l'ordonnée de G : la droite de Mayer passe par le point moyen de la série ✓

Pourquoi c'est toujours vrai. Les deux sous-groupes ont le même effectif, donc G est le **milieu** du segment $[G_1G_2]$: $\frac{2,5+6,5}{2} = 4,5$ et $\frac{5+11}{2} = 8$ ✓ Un milieu appartient toujours au segment, donc à la droite.

Réponse. a) $G_1(2,5; 5)$ et $G_2(6,5; 11)$ — b) $y = 1,5x + 1,25$ — c) pour $x = 4,5$, on trouve $y = 8$ ✓

Correction de l'exercice 7 - L'effet d'un point aberrant

a) L'abscisse moyenne ne change pas, puisque les x sont inchangés : $\bar{x} = 4,5$. Pour les ordonnées :

$$\bar{y} = \frac{3+5+4+8+9+11+10+2}{8} = \frac{52}{8} = 6,5$$

Le nouveau point moyen est $G'(4,5; 6,5)$.

b) L'ordonnée est passée de 8 à 6,5 : elle a **baissé de 1,5 point**. Une seule valeur modifiée a donc déplacé le point moyen de tout le nuage — c'est logique, puisque l'écart de 12 points sur une note se répartit sur les 8 élèves : $\frac{12}{8} = 1,5$.

c) Ce point est **aberrant** : l'élève a beaucoup révisé (8 heures) et obtenu la note la plus basse, ce qui contredit toute la tendance du nuage. Comme l'énoncé en donne la raison (la maladie), cet individu n'est pas représentatif du lien étudié : il est légitime de l'**écarter**, à condition de le dire explicitement.

Sans ce point, la série des sept premiers élèves donne $\bar{x} = \frac{28}{7} = 4$ et $\bar{y} = \frac{50}{7} \approx 7,14$, et la tendance croissante est nette.

La règle de rédaction. On n'écarte jamais une donnée « parce qu'elle gêne ». On l'écarte quand une raison extérieure aux chiffres montre qu'elle ne relève pas du phénomène étudié — et on le mentionne dans la conclusion.

Réponse. a) $G(4,5; 6,5)$ — b) l'ordonnée baisse de 1,5 — c) oui, on peut l'écarter : c'est un point aberrant dont la cause (maladie) est extérieure au phénomène.

Correction de l'exercice 8 – Placer le point moyen sur un nuage

a) Les deux moyennes, avec $n = 6$:

$$\bar{x} = \frac{10 + 20 + 30 + 40 + 50 + 60}{6} = \frac{210}{6} = 35$$

$$\bar{y} = \frac{4 + 7 + 11 + 14 + 18 + 22}{6} = \frac{76}{6} = 12,67 \text{ (arrondi au centième)}$$

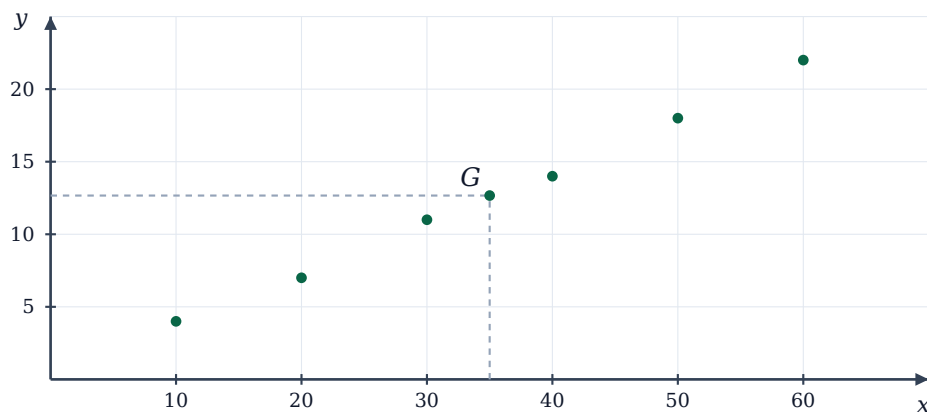
Le point moyen est donc $G(35; \frac{38}{3})$, soit environ $G(35; 12,67)$.

b) On place les six points du tableau, puis G à l'abscisse 35 (entre 30 et 40) et à l'ordonnée 12,67. Le tracé figure ci-dessous : G tombe bien au cœur du nuage.

c) Non, G n'est **pas** un point du nuage : aucun colis n'a un volume de 35 dm^3 . Et ce n'est pas un cas particulier — le point moyen est une **construction**, obtenue par un calcul de moyennes, et il ne coïncide avec un point du nuage que par hasard.

Ce que G représente. C'est le « centre de gravité » du nuage : si l'on posait des masses égales sur chaque point, la plaque s'équilibrerait exactement en G . C'est pourquoi toute droite d'ajustement doit passer par lui.

Réponse. a) $G(35; 12,67)$ environ — c) non, G n'appartient généralement pas au nuage.



Le nuage des six colis et son point moyen G .

Correction de l'exercice 9 – Point moyen avec des valeurs décimales

a) On additionne d'abord les quatre abscisses, en alignant les décimales :

$$1,2 + 2,5 + 3,1 + 4,2 = 11 \quad \text{donc} \quad \bar{x} = \frac{11}{4} = 2,75$$

b) De même pour les ordonnées :

$$8,4 + 10,2 + 11,6 + 13,8 = 44 \quad \text{donc} \quad \bar{y} = \frac{44}{4} = 11$$

c) Le point moyen est donc

$$G(2,75; 11)$$

Cohérence : 2,75 est bien compris entre la plus petite et la plus grande abscisse (1,2 et 4,2), et 11 entre 8,4 et 13,8 ✓ De plus, $\bar{y}$ tombe entre la deuxième et la troisième valeur, ce qui est attendu pour une série croissante de quatre termes.

L'erreur d'addition à éviter. Additionner les décimaux sans aligner les virgules donne presque toujours un résultat faux. Le contrôle « la moyenne est entre le minimum et le maximum » détecte immédiatement le problème.

Réponse. a) $\bar{x} = 2,75$ — b) $\bar{y} = 11$ — c) $G(2,75; 11)$, coordonnées bien encadrées par les données.

Correction de l'exercice 10 - Corrélation positive, négative ou nulle

a) Corrélation **positive** : à localisation comparable, plus un appartement est grand, plus il se vend cher. Les deux variables augmentent ensemble.

b) Corrélation **négative** : plus une voiture est ancienne, moins elle se revend cher. Quand x augmente, y diminue.

c) Corrélation **quasi nulle** : rien ne lie la taille des pieds au salaire d'un adulte. Le nuage serait dispersé, sans direction privilégiée, et un ajustement affine n'aurait aucun sens.

d) Corrélation **négative** : la température moyenne baisse d'environ 0,6 °C tous les 100 m d'altitude gagnés. C'est une relation physique bien établie.

La mise en garde indispensable. Une corrélation, même forte, ne prouve **pas** un lien de cause à effet. On observe par exemple une corrélation positive entre la pointure et le salaire **chez les enfants et les adultes confondus** : ce n'est pas le pied qui fait le salaire, c'est l'âge qui agit sur les deux. On appelle cela une variable cachée.

Réponse. a) positive — b) négative — c) quasi nulle — d) négative.

Correction de l'exercice 11 - Vérifier un point moyen annoncé

a) On calcule la moyenne des abscisses, avec $n = 4$:

$$\bar{x} = \frac{1 + 3 + 5 + 7}{4} = \frac{16}{4} = 4$$

L'abscisse annoncée est donc **correcte**.

b) Pour les ordonnées :

$$\bar{y} = \frac{6 + 8 + 13 + 17}{4} = \frac{44}{4} = 11$$

L'ordonnée annoncée, 10, est donc **fausse** : elle vaut 11.

c) L'élève a raison sur l'abscisse mais se trompe sur l'ordonnée : le point moyen est $G(4; 11)$ et non $G(4; 10)$. Son erreur vient probablement d'une addition inexacte ($6 + 8 + 13 + 17 = 44$, et non 40).

Le contrôle rapide à faire. La moyenne de quatre nombres dont le plus petit est 6 et le plus grand 17 doit se trouver entre les deux, plutôt vers le milieu : 11 est plausible, 10 l'était aussi — le contrôle par encadrement ne suffit donc pas ici, il faut refaire l'addition.

Réponse. a) $\bar{x} = 4$, correct — b) $\bar{y} = 11$, et non 10 — c) non, le point moyen est $G(4; 11)$.

Correction de l'exercice 12 - Retrouver une valeur manquante

a) La moyenne des abscisses :

$$\bar{x} = \frac{1 + 2 + 3 + 4 + 5}{5} = \frac{15}{5} = 3$$

C'est bien l'abscisse de G ✓

b) Notons v la valeur manquante. L'ordonnée du point moyen s'écrit

$$\bar{y} = \frac{4 + 7 + v + 13 + 16}{5} = 10$$

On multiplie les deux membres par 5 :

$$4 + 7 + v + 13 + 16 = 50 \Leftrightarrow 40 + v = 50 \Leftrightarrow v = 10$$

La valeur manquante est donc 10.

c) Contrôle : $\frac{4 + 7 + 10 + 13 + 16}{5} = \frac{50}{5} = 10$ ✓ On retrouve bien l'ordonnée annoncée.

La méthode générale. « Connaître la moyenne » revient à connaître la **somme** : c'est toujours en passant par la somme (somme = $n \times$ moyenne) que l'on retrouve une donnée manquante. On remarque au passage que les écarts 4, 7, 10, 13, 16 sont constants : la série est parfaitement alignée sur la droite $y = 3x + 1$.

Réponse. a) $\bar{x} = 3$ ✓ — b) la valeur manquante est 10 — c) $\bar{y} = \frac{50}{5} = 10$ ✓

Correction de l'exercice 13 - Taille et masse de nouveau-nés

a) Somme des abscisses : $46 + 47 + 48 + 50 + 51 + 52 + 54 = 348$, donc

$$\bar{x} = \frac{348}{7} = 49,71 \text{ (arrondi au centième)}$$

Somme des ordonnées : $2,6 + 2,8 + 3,0 + 3,3 + 3,5 + 3,6 + 4,0 = 22,8$, donc

$$\bar{y} = \frac{22,8}{7} = 3,26 \text{ (arrondi au centième)}$$

Le point moyen est donc environ $G(49,71 ; 3,26)$.

b) La corrélation est **positive** : les masses augmentent régulièrement avec les tailles, sans aucune exception dans ce tableau.

c) On compare les deux individus extrêmes :

$$\frac{4,0 - 2,6}{54 - 46} = \frac{1,4}{8} = 0,175$$

La masse augmente donc d'environ 0,175 kg, soit 175 **g par centimètre** de taille supplémentaire.

Le statut de cette estimation. Ce calcul n'utilise que deux points sur sept : c'est une première approximation, sensible aux valeurs extrêmes. La droite de régression de la plaquette 3 utilisera les **sept** points et donnera un coefficient plus fiable.

Réponse. a) $G(49,71 ; 3,26)$ environ — b) positive — c) environ 0,175 kg par centimètre, soit 175 g.

Correction de l'exercice 14 - Deux sous-nuages de tailles différentes

a) Avec $n = 6$:

$$\bar{x} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3,5 \quad \bar{y} = \frac{2 + 5 + 6 + 9 + 11 + 13}{6} = \frac{46}{6} \approx 7,67$$

Donc $G\left(3,5 ; \frac{23}{3}\right)$, soit environ $G(3,5 ; 7,67)$.

b) Pour les trois premiers couples :

$$\bar{x}_1 = \frac{1 + 2 + 3}{3} = 2 \quad \bar{y}_1 = \frac{2 + 5 + 6}{3} = \frac{13}{3} \approx 4,33$$

Pour les trois derniers :

$$\bar{x}_2 = \frac{4+5+6}{3} = 5 \quad \bar{y}_2 = \frac{9+11+13}{3} = \frac{33}{3} = 11$$

Donc $G_1\left(2; \frac{13}{3}\right)$ et $G_2(5; 11)$.

c) Le milieu de $[G_1G_2]$ a pour coordonnées les moyennes des coordonnées :

$$\frac{2+5}{2} = 3,5 \quad \frac{\frac{13}{3} + 11}{2} = \frac{\frac{13}{3} + \frac{33}{3}}{2} = \frac{\frac{46}{3}}{2} = \frac{23}{3}$$

Ce sont exactement les coordonnées de G ✓

Le coefficient directeur de la droite (G_1G_2) est

$$m = \frac{11 - \frac{13}{3}}{5 - 2} = \frac{\frac{20}{3}}{3} = \frac{20}{9} \approx 2,22$$

et son équation est $y = \frac{20}{9}(x - 5) + 11$, soit environ $y \approx 2,22x - 0,11$.

Pourquoi ça marche. Les deux sous-groupes ont le **même effectif** (3 et 3) : la moyenne générale est alors la moyenne des deux moyennes, donc G est le milieu. Avec des effectifs différents, ce serait faux — il faudrait pondérer.

Réponse. a) $G(3,5; 7,67)$ environ — b) $G_1(2; 4,33)$ et $G_2(5; 11)$ — c) le milieu de $[G_1G_2]$ est bien G , et $y \approx 2,22x - 0,11$.

Correction de l'exercice 15 - Synthèse : abonnements d'une salle de sport

a) Les deux moyennes, avec $n = 6$:

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

$$\bar{y} = \frac{220+260+310+340+400+430}{6} = \frac{1960}{6} \approx 326,67$$

Donc $G(3,5; 326,67)$ environ.

b) Pour les trois premières années :

$$\bar{x}_1 = \frac{1+2+3}{3} = 2 \quad \bar{y}_1 = \frac{220+260+310}{3} = \frac{790}{3} \approx 263,33$$

Pour les trois dernières :

$$\bar{x}_2 = \frac{4+5+6}{3} = 5 \quad \bar{y}_2 = \frac{340+400+430}{3} = \frac{1170}{3} = 390$$

Donc $G_1(2; 263,33)$ et $G_2(5; 390)$.

c) Le coefficient directeur :

$$m = \frac{390 - 263,33}{5 - 2} = \frac{126,67}{3} \approx 42,2$$

L'équation s'écrit $y = 42,2(x - 5) + 390$, soit

$$y = 42,2x - 211 + 390 = 42,2x + 179$$

d) Pour $x = 3,5$:

$$y = 42,2 \times 3,5 + 179 = 147,7 + 179 = 326,7$$

C'est bien l'ordonnée de G , aux arrondis près ✓ Pour l'année 8 :

$$y = 42,2 \times 8 + 179 = 337,6 + 179 = 516,6 \approx 517 \text{ abonnés}$$

e) L'estimation est **raisonnable mais pas garantie** : elle suppose que la croissance de

42 abonnés par an se poursuive deux années au-delà des données observées, alors que la capacité de la salle, la concurrence ou la conjoncture peuvent l'infléchir. Toute prévision hors de l'intervalle des données est une **extrapolation**, à annoncer comme telle — ce sera le sujet de la plaquette 4.

Le bilan de la plaquette. Un nuage se lit (allongé ? croissant ? point aberrant ?), puis se résume par son point moyen. La méthode de Mayer fournit un premier ajustement, simple et entièrement calculable à la main, qui passe nécessairement par G .

Réponse. a) $G(3,5 ; 326,67)$ — b) $G_1(2 ; 263,33)$ et $G_2(5 ; 390)$ — c) $y = 42,2x + 179$ — d) la droite passe par G ; environ 517 abonnés l'année 8 — e) c'est une extrapolation, à prendre avec prudence.