1re

COURS

Nuage de points — Ajustement affine et point moyen

Programme de mathématiques intégré à l'enseignement scientifique de première générale — arrêté du 26 février 2026, BO n° 14 du 2 avril 2026 — MENE2602916A

On croise maintenant **deux** caractères : qualitatifs, dans un tableau où l'on compare des **proportions** ; quantitatifs, par un nuage de points. Et tout le chapitre tient dans une phrase : **quand le nuage est allongé, on le remplace par une droite qui passe par le point moyen $G(\overline{x}\,;\overline{y})$, et cette droite sert à estimer des valeurs inconnues, avec prudence dès qu'on sort des données.**

1 Croiser deux caractères qualitatifs

DÉFINITION

On observe **deux caractères** sur chaque individu d'une population. Un caractère est **qualitatif** quand ses valeurs sont des catégories (oui ou non, un mode de transport) ; un caractère quantitatif s'y ramène en **regroupant ses valeurs en classes** (moins de 3 h, 3 h ou plus). Le **tableau croisé d'effectifs** compte, dans chaque case, les individus qui ont à la fois la catégorie de la ligne et celle de la colonne. La dernière ligne et la dernière colonne, les **marges**, donnent les totaux.

EXEMPLE

Enquête fictive auprès de 400 lycéens : sur les 250 qui pratiquent un sport en club, 160 passent moins de 3 h par jour devant un écran ; sur les 150 autres, 45. Les sportifs passent-ils moins de temps devant les écrans ?

1. Chaque ligne s'additionne : $250 - 160 = 90$ sportifs et $150 - 45 = 105$ non-sportifs passent 3 h ou plus. Marge du bas : $160 + 45 = 205$ et $90 + 105 = 195$. Contrôle du total : $205 + 195 = 400$
2. Les groupes n'ont pas la même taille : on compare des **proportions**. Parmi les sportifs : $\frac{160}{250} = 0{,}64$ et parmi les non-sportifs, $\frac{45}{150} = 0{,}3$.

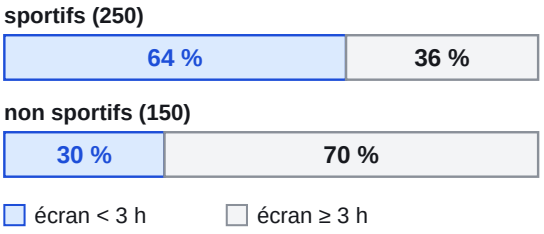
64 % des sportifs passent moins de 3 h par jour devant un écran, contre 30 % des non-sportifs : dans cette enquête, les sportifs passent nettement moins de temps devant les écrans.

le tableau croisé : 400 lycéens (données fictives)

	écran < 3 h	écran ≥ 3 h	total
sportifs	160	90	**250**
non sportifs	45	105	**150**
total	**205**	**195**	**400**

en gris : les marges (totaux)

les proportions dans chaque groupe



À gauche, le tableau complété et ses marges. À droite, un diagramme en barres à 100 % : une barre par groupe, ce qui rend les deux groupes comparables malgré leurs tailles différentes.

COMPARER DEUX GROUPES DANS UN TABLEAU CROISÉ

1. Repérer les **groupes** à comparer : ce sont des lignes, ou des colonnes, du tableau.
2. Pour chaque groupe, diviser l'effectif de la case par la **marge de ce groupe**, jamais par l'effectif total.
   « parmi les sportifs » : on divise par le nombre de sportifs, 250.
3. Comparer les proportions, puis conclure par une phrase sur la situation. Pour les montrer, un diagramme en barres à 100 % ou un diagramme circulaire par groupe ; dans un diagramme circulaire, l'angle d'un secteur vaut **fréquence × 360°**.

PIÈGE

Comparer les effectifs : « 90 sportifs et 105 non-sportifs passent 3 h ou plus devant un écran, la différence est faible. » — les groupes n'ont pas la même taille, 250 et 150. En proportion, c'est 36 % contre 70 % : presque le double.

Ce que ça coûte : la conclusion, qui est l'objet même de la question.

Le réflexe : dès que les groupes ont des tailles différentes, passer aux proportions.

PIÈGE

Pour « la proportion de sportifs parmi les élèves à moins de 3 h », calculer $\frac{160}{250} = 64$ %. — « parmi les élèves à moins de 3 h » désigne la **colonne** : on divise par sa marge, 205, et l'on trouve $\frac{160}{205} \approx 0{,}78$, soit environ 78 %.

Ce que ça coûte : le résultat, et l'interprétation qui suit.

Le réflexe : le mot « parmi » désigne le groupe de référence ; son effectif est le dénominateur.

2 Deux caractères quantitatifs : nuage de points et point moyen

DÉFINITION

Deux caractères **quantitatifs** x et y, mesurés sur n individus, donnent n couples $(x_1\,;y_1)$, ..., $(x_n\,;y_n)$. Dans un repère, les points $M_i(x_i\,;y_i)$ forment le **nuage de points** de la série. Pour une série **chronologique**, le temps se place en abscisse.

DÉFINITION

Le **point moyen** du nuage est le point G dont l'abscisse est la **moyenne des** x_i et l'ordonnée la **moyenne des** y_i. Il résume le nuage en un seul point, mais il n'est en général **pas** l'un de ses points.

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

EXEMPLE

Population d'une ville, en milliers d'habitants (données fictives), avec x le nombre d'années écoulées depuis 2000 :

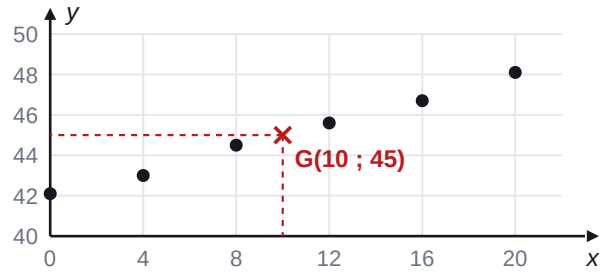
année	2000	2004	2008	2012	2016	2020
x	0	4	8	12	16	20
y	42,1	43,0	44,5	45,6	46,7	48,1

Calculer les coordonnées du point moyen.

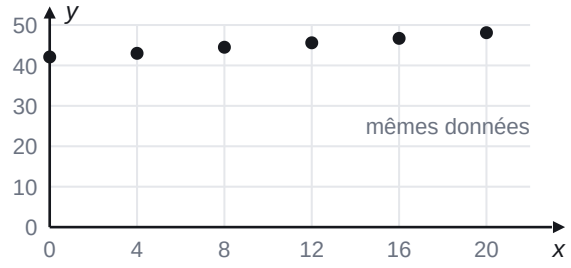
1. Moyenne des abscisses, les six valeurs :
$\overline{x} = \frac{0+4+8+12+16+20}{6} = \frac{60}{6} = 10$
2. Moyenne des ordonnées :
$\overline{y} = \frac{42{,}1+43{,}0+44{,}5+45{,}6+46{,}7+48{,}1}{6} = \frac{270}{6} = 45$

$G(10\,;45)$: l'année 2010, et une population moyenne de 45 milliers d'habitants sur les six relevés. Aucun point du nuage n'est en G.

axe vertical à partir de 40 : la hausse paraît forte



axe vertical à partir de 0 : environ +14 %



À gauche, le nuage de la ville et son point moyen G, axe vertical gradué à partir de 40. À droite, les mêmes points avec un axe qui part de 0 : la hausse réelle est modeste.

AVEC UN TABLEUR

1. Saisir les x_i en A2:A7 et les y_i en B2:B7, sélectionner les deux colonnes et insérer un graphique « nuage de points » (ou « XY »). Vérifier l'origine et les graduations choisies par le logiciel.
2. Dans deux cellules libres, taper =MOYENNE(A2:A7) et =MOYENNE(B2:B7) : ce sont les coordonnées de G.

PIÈGE

Prendre pour G le milieu du premier et du dernier point : $(10\,;45{,}1)$. — ce milieu n'utilise que deux valeurs sur six. Ici, il tombe près de G par chance ; sur un nuage irrégulier, il peut en être très loin.

Ce que ça coûte : le point moyen, et la droite d'ajustement qui doit passer par lui.

Le réflexe : additionner **toutes** les valeurs, et diviser par leur **nombre**.

PIÈGE

Lire sur la figure de gauche que la population a été « multipliée par près de 4 », le dernier point étant presque quatre fois plus haut que le premier. — l'axe commence à 40 : les hauteurs mesurent l'écart à 40 milliers. En réalité, $48{,}1 \div 42{,}1 \approx 1{,}14$: une hausse d'environ 14 %.

Ce que ça coûte : une conclusion fausse, sur une copie comme dans un débat.

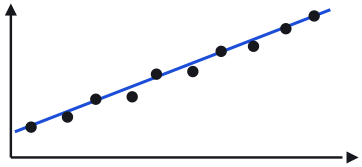
Le réflexe : avant toute lecture, repérer l'**origine** du repère et les **unités** des graduations.

3 Ajuster le nuage par une droite

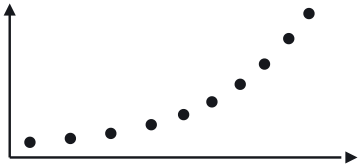
DÉFINITION

Quand les points du nuage sont **à peu près alignés**, on dit que le nuage est **allongé**, et on peut le remplacer par une **droite d'ajustement** d'équation $y = ax + b$, qui passe au plus près des points. C'est un **modèle** : il décrit une **tendance**, sans en donner la cause, et il ne passe en général exactement par aucun point.

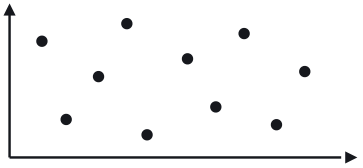
nuage allongé
un ajustement affine convient



nuage courbé
pas d'ajustement affine



nuage sans forme
aucune tendance à ajuster



Avant tout calcul, regarder la forme du nuage : seule la première justifie une droite.

PROPRIÉTÉ

Plusieurs droites d'ajustement existent, qui ne coïncident pas tout à fait ; aucune théorie n'est attendue. **Au jugé** : à la règle, une droite qui suit le nuage en laissant à peu près autant de points de chaque côté. **Droite de Mayer** : on range les points par abscisses croissantes, on les partage en deux groupes de même effectif (à un près), et la droite passe par les points moyens G_1 et G_2 des deux groupes. **Moindres carrés** : la droite que trace le tableur quand on ajoute au nuage une « courbe de tendance » linéaire, équation affichée. Les droites de Mayer et des moindres carrés **passent par le point moyen** G (admis) : on fait passer la droite au jugé par G, elle aussi. Pour la ville, le tableur affiche $y = 0{,}3014x + 41{,}986$, et $0{,}3014 \times 10 + 41{,}986 = 45$: elle passe bien par $G(10\,;45)$, tout près de la droite de Mayer calculée ci-dessous.

ÉQUATION DE LA DROITE PASSANT PAR DEUX POINTS A ET B

1. **Coefficient directeur** : la variation de y divisée par la variation de x.
   $a = \frac{y_B - y_A}{x_B - x_A}$
2. **Ordonnée à l'origine** : A est sur la droite, donc $y_A = a\,x_A + b$, d'où $b = y_A - a\,x_A$.
3. **Contrôler** avec le second point, ou avec G : ses coordonnées doivent vérifier l'équation.

EXEMPLE

Déterminer la droite de Mayer du nuage de la ville.

1. Groupe 1, les trois premiers points : abscisse moyenne $\frac{0+4+8}{3} = 4$, ordonnée moyenne : $\frac{42{,}1+43{,}0+44{,}5}{3} = \frac{129{,}6}{3} = 43{,}2$
2. Groupe 2, les trois derniers : abscisse moyenne $\frac{12+16+20}{3} = 16$, ordonnée moyenne : $\frac{45{,}6+46{,}7+48{,}1}{3} = \frac{140{,}4}{3} = 46{,}8$
3. Coefficient directeur de (G_1G_2), avec $G_1(4\,;43{,}2)$ et $G_2(16\,;46{,}8)$: $a = \frac{46{,}8-43{,}2}{16-4} = \frac{3{,}6}{12} = 0{,}3$
4. Ordonnée à l'origine, avec G_1 : $b = 43{,}2 - 0{,}3 \times 4 = 42$
   contrôle avec $G(10\,;45)$: $0{,}3 \times 10 + 42 = 45$, G est bien sur la droite.

La droite de Mayer a pour équation $y = 0{,}3x + 42$: selon ce modèle, la ville gagne environ 0,3 millier d'habitants, soit 300 habitants, par an.

PIÈGE

Calculer le coefficient directeur à l'envers : $a = \dfrac{16-4}{46{,}8-43{,}2} \approx 3{,}33$. — c'est la variation de x divisée par celle de y : la droite monte beaucoup trop vite et ne passe plus par G.

Ce que ça coûte : l'équation, et toutes les estimations qui en découlent.

Le réflexe : « y en haut » : a est la variation de y divisée par celle de x. Puis contrôler que G vérifie l'équation.

4 Interpoler, extrapoler

DÉFINITION

Estimer une valeur inconnue à l'aide de la droite d'ajustement, c'est **interpoler** si la valeur de x est **à l'intérieur** de l'intervalle des données, et **extrapoler** si elle est **à l'extérieur**. Extrapoler suppose que la tendance observée **se poursuit** : c'est une hypothèse, d'autant plus fragile qu'on s'éloigne des données. Loin d'elles, une droite peut même donner des valeurs absurdes, comme une quantité négative.

ESTIMER UNE VALEUR AVEC LA DROITE D'AJUSTEMENT

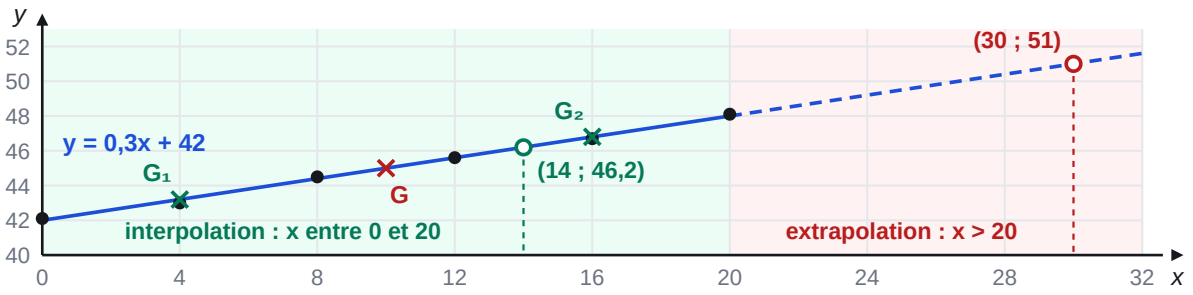
1. Traduire la question en valeur de x, en relisant ce que x représente : ici, le nombre d'années **depuis 2000**.
2. Remplacer x dans l'équation ; pour une valeur de y donnée, résoudre $ax + b = y$.
3. Conclure dans l'unité de la situation, en précisant interpolation ou extrapolation ; une extrapolation s'énonce au conditionnel.

EXEMPLE

Avec la droite $y = 0{,}3x + 42$, estimer la population de la ville en 2014, puis en 2030.

1. 2014 donne $x = 14$, dans l'intervalle $[0\,;20]$ des données : interpolation.
   $0{,}3 \times 14 + 42 = 46{,}2$
2. 2030 donne $x = 30$, hors de l'intervalle : extrapolation.
   $0{,}3 \times 30 + 42 = 51$

Environ 46 200 habitants en 2014 ; environ 51 000 en 2030, mais c'est une extrapolation : seulement si la tendance se poursuit.



La droite de Mayer passe par G_1, G et G_2. En trait plein, là où sont les données ; en pointillés, son prolongement, qui n'est plus qu'une hypothèse.

PIÈGE

Remplacer x par l'année elle-même : $0{,}3 \times 2030 + 42 = 651$ milliers d'habitants en 2030. — x compte les années **depuis 2000**. Une ville dont la population serait multipliée par plus de treize en dix ans : le résultat aurait dû alerter.

Ce que ça coûte : toute la question, et souvent les suivantes.

Le réflexe : relire la définition de x avant de remplacer, et confronter le résultat à l'ordre de grandeur des données.

À VÉRIFIER

- Dans un tableau croisé, ai-je comparé des **proportions**, en divisant par la **marge du groupe** désigné par « parmi » ?
- Ai-je repéré l'**origine** et les **graduations** des axes avant de lire un graphique ?
- Mon point moyen utilise-t-il **toutes** les valeurs ? Mon a est-il la variation de y sur celle de x, et G vérifie-t-il l'équation ?
- Ai-je remplacé x par la bonne valeur et dit si j'interpole ou si j'extrapole ?

À RETENIR

- Tableau croisé : « parmi » désigne le groupe de référence, dont la **marge** est le dénominateur. Entre groupes de tailles différentes, on compare des **proportions**.
- Point moyen : $G(\overline{x}\,;\overline{y})$, avec la moyenne de **toutes** les abscisses et de **toutes** les ordonnées.
- Ajustement affine : une droite $y = ax + b$ proche des points d'un nuage **allongé** ; au jugé, de Mayer ou des moindres carrés, elle passe par G.
- Mayer : deux groupes rangés par x croissants, puis la droite (G_1G_2). Deux points : $a = \dfrac{y_B - y_A}{x_B - x_A}$, puis $b = y_A - a\,x_A$.
- Interpoler : x dans l'intervalle des données. Extrapoler : en dehors, et au conditionnel.