Plaquette 3 – Régression linéaire et corrélation

Statistiques à deux variables — Première générale

Exercices

Méthode

La calculatrice fournit la **meilleure** droite d'ajustement au sens des moindres carrés : celle qui rend minimale la somme des carrés des résidus. Elle donne aussi le **coefficient de corrélation linéaire** r, un nombre compris entre -1 et 1 qui mesure à quel point le nuage est proche d'une droite : $|r|$ voisin de 1 signale un alignement excellent, $|r|$ voisin de 0 un nuage sans direction. Le signe de r est celui du coefficient directeur. Cette plaquette apprend à obtenir ces trois nombres, à les interpréter, et surtout à ne pas leur faire dire plus qu'ils ne disent : **une corrélation forte n'est pas une preuve de causalité**. C'est l'erreur de raisonnement la plus fréquente du chapitre, et la plus systématiquement évaluée.

Exercice 1 – Obtenir la droite de régression

On reprend la série des chiffres d'affaires, en milliers d'euros.

Année x	1	2	3	4	5	6
Chiffre y	12	15	19	22	27	31

À la calculatrice, on obtient $a \approx 3{,}829$ et $b = 7{,}6$.

a) Donner une équation de la droite de régression de y en x, les coefficients arrondis au centième.

b) Vérifier que cette droite passe par le point moyen $G(3{,}5\,;21)$.

c) Comparer avec la droite de Mayer $y = 3{,}78x + 7{,}77$ obtenue dans la plaquette 2.

Exercice 2 – Une régression décroissante

On reprend les relevés de consommation de gaz.

Température x **(°C)**	2	4	6	8	10	12	14
Consommation y (m³/j)	20	18	15	13	10	8	5

La calculatrice donne $a = -1{,}25$, $b \approx 22{,}714$ et $r \approx -0{,}999$.

a) Donner l'équation de la droite de régression, les coefficients arrondis au centième.

b) Interpréter le signe et la valeur de r.

c) Estimer la consommation pour une température de 18 °C. Cette estimation est-elle exploitable ?

Exercice 3 – Comparer deux coefficients de corrélation

Deux séries sont étudiées.

Série A : $(1\,;3)$, $(2\,;5)$, $(3\,;4)$, $(4\,;8)$, $(5\,;9)$, $(6\,;11)$, $(7\,;10)$, $(8\,;14)$, avec $r \approx 0{,}957$.

Série B : $(1\,;6)$, $(2\,;2)$, $(3\,;9)$, $(4\,;3)$, $(5\,;7)$, avec $r \approx 0{,}165$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

a) Pour chaque série, dire si un ajustement affine est pertinent.

b) Quelle est la série la mieux ajustée par une droite ?

c) Pour la série B, la calculatrice donne tout de même $y = 0{,}3x + 4{,}5$. Que penser de cette droite ?

Exercice 4 – Masse et taille de nouveau-nés

On reprend les sept nouveau-nés.

Taille x (cm)	46	47	48	50	51	52	54
Masse y (kg)	2,6	2,8	3,0	3,3	3,5	3,6	4,0

La calculatrice donne $a \approx 0{,}170$, $b \approx -5{,}206$ et $r \approx 0{,}998$.

a) Donner l'équation de la droite de régression et commenter r.

b) Estimer la masse d'un nouveau-né de 49 cm, arrondie au gramme.

c) Interpréter le coefficient a, puis expliquer pourquoi b n'a ici aucun sens concret.

Exercice 5 – Quand la corrélation est parfaite

Le tableau donne la fréquentation d'un musée, en centaines de visiteurs, pour cinq mois (le mois 4 a été écarté, le musée étant fermé).

Mois x	1	2	3	5	6
Visiteurs y	10	13	16	22	25

a) À la calculatrice, on obtient $a = 3$, $b = 7$ et $r = 1$. Que signifie $r = 1$?

b) Vérifier par le calcul que les cinq points appartiennent à la droite $y = 3x + 7$.

c) Cette droite est-elle la même que celle obtenue à la main dans la plaquette 2 ?

Exercice 6 – Surface et loyer

Le tableau donne le loyer mensuel y (en euros) de six appartements en fonction de leur surface x (en m²).

x	20	30	40	50	60	70
y	480	620	760	880	1 020	1 150

La calculatrice donne $a \approx 13{,}343$, $b \approx 217{,}905$ et $r \approx 0{,}9998$.

a) Donner l'équation de la droite de régression, les coefficients arrondis au centième, et commenter r.

b) Estimer le loyer d'un appartement de 45 m².

c) Interpréter les deux coefficients dans le contexte.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

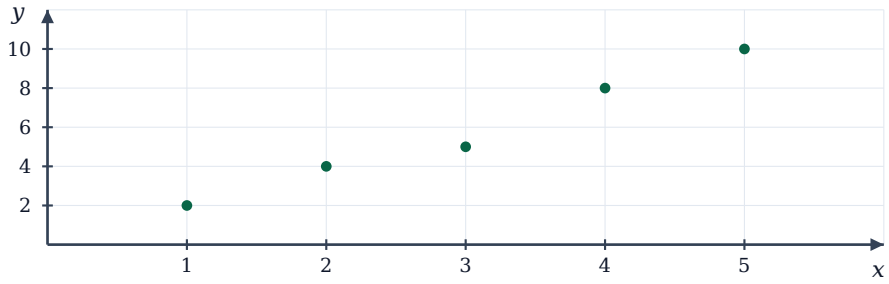
Exercice 7 – Associer trois nuages à trois corrélations

Trois nuages de cinq points sont tracés ci-dessous. Les coefficients de corrélation sont, dans un ordre inconnu, $-0{,}99$, $0{,}15$ et $0{,}98$.

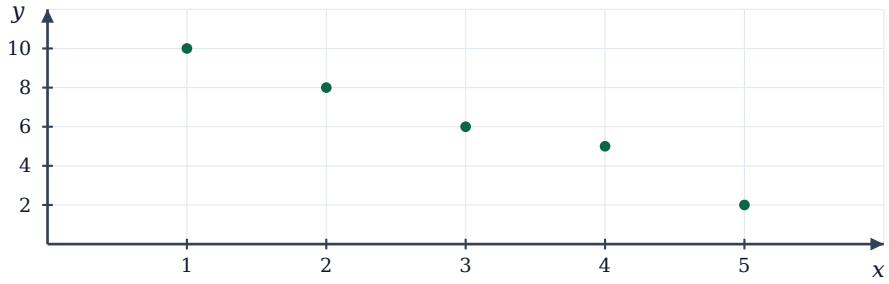
a) Attribuer son coefficient à chaque nuage, en justifiant.

b) Pour lequel des trois un ajustement affine est-il à écarter ?

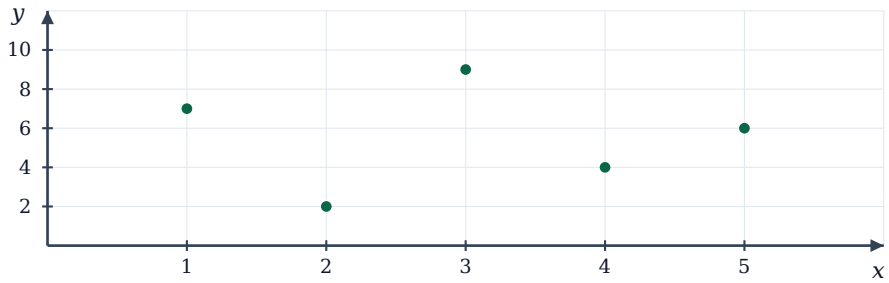
c) Que vaudrait r pour un nuage dont les cinq points seraient parfaitement alignés sur une droite descendante ?



Nuage n° 1.



Nuage n° 2.



Nuage n° 3.

Exercice 8 – La droite passe toujours par G

Pour une série de neuf relevés, on sait que $\overline{x} = 12$ et $\overline{y} = 45$. La calculatrice donne un coefficient directeur $a = 2{,}8$ pour la droite de régression.

a) Déterminer b sans connaître les données.

b) Donner l’équation de la droite de régression.

c) Estimer y pour $x = 20$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Exercice 9 – Budget publicitaire et ventes

Une entreprise relève son budget publicitaire x (en milliers d'euros) et ses ventes y (en milliers d'euros).

x	5	8	12	15	20	25
y	120	150	190	215	260	300

La calculatrice donne $a \approx 8{,}999$, $b \approx 78{,}350$ et $r \approx 0{,}999$.

a) Donner l'équation de la droite de régression, les coefficients arrondis au centième.

b) Interpréter le coefficient a. Augmenter le budget publicitaire est-il rentable, d'après ce modèle ?

c) Estimer les ventes pour un budget de 30 milliers d'euros.

Exercice 10 – Corrélation n'est pas causalité

Dans chacune des situations suivantes, une corrélation forte est observée. Dire si l'on peut en déduire un lien de cause à effet, et proposer une explication.

a) Le nombre de pompiers mobilisés sur un incendie et les dégâts constatés.

b) La taille d'un enfant et son niveau de lecture.

c) Le nombre d'heures d'entraînement d'un sportif et sa performance.

d) Les ventes de crème solaire et celles de glaces.

Exercice 11 – L'effet d'un arrondi trop brutal

Pour la série des chiffres d'affaires, la droite de régression a pour coefficients $a \approx 3{,}828\,6$ et $b = 7{,}6$.

a) Estimer le chiffre d'affaires de l'année 10 avec ces valeurs.

b) Refaire l'estimation en arrondissant a à 4 (l'unité).

c) Comparer les deux résultats et conclure sur la règle d'arrondi.

Exercice 12 – Régression sur les notes de huit élèves

On reprend la série des huit élèves (heures de révision x, note y).

x	1	2	3	4	5	6	7	8
y	3	5	4	8	9	11	10	14

La calculatrice donne $a \approx 1{,}476$, $b \approx 1{,}357$ et $r \approx 0{,}957$.

a) Donner l'équation de la droite de régression et commenter r.

b) Vérifier que la droite passe par $G(4{,}5\,;\,8)$.

c) Estimer la note d'un élève ayant révisé 10 heures. Quelles réserves formuler ?

Exercice 13 – Ventes de vélos : régression et Mayer

On reprend les ventes trimestrielles de vélos électriques.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Trimestre x	1	2	3	4	5	6
Ventes y	34	41	52	58	69	76

La calculatrice donne $a \approx 8{,}571$, $b = 25$ et $r \approx 0{,}998$.

a) Donner l'équation de la droite de régression et estimer les ventes du huitième trimestre.

b) Comparer avec l'estimation obtenue par la droite de Mayer, $y = 8{,}45x + 25{,}43$, dans la plaquette 2.

c) Laquelle des deux estimations retenir ? Justifier.

Exercice 14 – Lire les sorties d'une calculatrice

Un élève relève l'affichage suivant de sa calculatrice après avoir saisi une série de dix couples.

Grandeur	a	b	r	r^2
Valeur	$-0{,}845$	53,17	$-0{,}872$	0,760

a) Donner l'équation de la droite de régression.

b) L'ajustement affine est-il satisfaisant ? Justifier avec les repères du cours.

c) Le signe de a et celui de r sont-ils cohérents ? Que se passerait-il s'ils étaient différents ?

Exercice 15 – Synthèse : consommation d'un véhicule

On relève, pour six trajets, la vitesse moyenne x (en km/h) et la consommation y (en L/100 km).

x	50	70	90	100	110	130
y	5,2	5,0	5,4	5,9	6,5	7,8

a) Calculer les coordonnées du point moyen, arrondies au centième.

b) La calculatrice donne $a \approx 0{,}032\,65$, $b \approx 2{,}973$ et $r \approx 0{,}890$. Donner l'équation de la droite de régression et commenter r.

c) Vérifier que la droite passe par le point moyen.

d) Estimer la consommation à 120 km/h.

e) Le nuage est-il vraiment bien décrit par une droite ? Observer les trois premières valeurs et conclure.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).