CHAPITRE 7

MODÉLISATION et RÉGRESSION

Appliquer la méthode de la droite de régression pour modéliser des données expérimentales à pas variable ou constant.

Les composantes particulières de l'élément de compétence visées par le présent chapitre sont:

- la reconnaissance du lien entre deux variables grâce à l'examen de données observées;
- la modélisation de données expérimentales à pas constant ou à pas variable;
- l'utilisation de papier semi-logarithmique ou logarithmique pour la détermination du lien entre des données expérimentales;
- la transformations de données visant à traduire en un lien affine une relation puissance, une relation exponentielle ou une relation logarithmique.

OBJECTIFS

7.1 Construire un modèle exponentiel ou logarithmique à l'aide de la description verbale d'un phénomène.

7.2 Utiliser la droite de régression pour construire un modèle affine représentant des données numériques.

7.3 Transformer une expression algébrique à l'aide des propriétés des logarithmes.

7.4 Résoudre des équations exponentielles en ayant recours aux propriétés des exposants et des logarithmes.

7.1 Modélisation affine

Même si on peut décrire par un modèle affine un phénomène étudié lors d'une expérience de laboratoire, d'un sondage ou d'une recherche, il faut s'attendre à ce qu'il y ait une différence entre les valeurs observées et les valeurs fournies par le modèle. En effet, aucun modèle n'est une description exacte d'un phénomène expérimental. Une règle de correspondance obtenue à l'aide de données expérimentales est un **modèle empirique** dont la fiabilité dépend de la précision des données expérimentales.

Données à pas constant

Lorsqu'on étudie la relation entre les variables d'un phénomène pour lequel on dispose de données empiriques, la représentation graphique se révèle un moyen efficace pour décider si le phénomène est descriptible par un modèle affine. En pratique, on a plusieurs couples formant un nuage de points et on cherche à déterminer la droite qui décrit le plus fidèlement possible le phénomène. Le cas le plus simple est celui où les données sont à pas constant, c'est-à-dire que les valeurs de la variable indépendante sont à intervalles réguliers.

Critère algébrique

Le taux de variation constant est une caractéristique du modèle affine. Dès que l'on peut déterminer que le taux de variation d'une grandeur donnée est constant, on conclut que le phénomène est modélisable par une fonction affine.

Si on dispose de mesures expérimentales pour des valeurs à intervalles réguliers p de la variable indépendante, et que le phénomène est descriptible par un modèle affine $f(x) = ax + b$. Pour chacun des couples, on a :

$$\begin{aligned} f(x+p) &= a(x+p) + b \\ &= ax + ap + b \\ &= ax + b + ap \\ &= f(x) + ap \end{aligned}$$

$$f(x+p) - f(x) = ap.$$

$$\frac{\Delta f}{p} = \frac{f(x+p) - f(x)}{p} = a.$$

Par conséquent, lorsque les valeurs de la variable indépendante sont à pas constant et que la différence entre les images consécutives de cette variable est proportionnelle au pas, on en conclut que le phénomène est descriptible par un modèle affine.

En bref, l'existence d'un lien affine dans le cas de données à pas constant est confirmée si

$$\Delta f = f(x+p) - f(x) = ap.$$

PROCÉDURE

Modélisation affine, données à pas constant

1. Représenter graphiquement les données. Si le nuage de points semble former une droite, on pose, sous toute réserve, l'hypothèse que le phénomène est descriptible par un modèle affine.
2. Calculer les différences $f(x + p) - f(p)$. Si elles sont relativement constantes, le paramètre a du modèle affine est la moyenne des différences divisée par le pas.
3. Calculer les ordonnées $b_i = f(x_i) - ax_i$ et déterminer le paramètre b en prenant la moyenne des valeurs de ces ordonnées.
4. Utiliser le modèle pour analyser le phénomène et répondre aux questions.

EXEMPLE 7.1.1

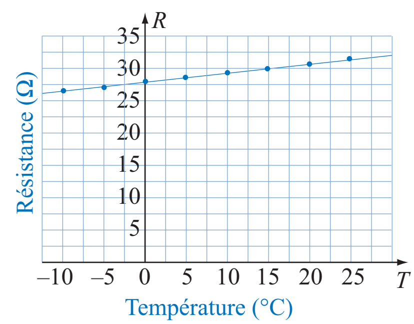
Une équipe étudie en laboratoire la relation entre la température d'un conducteur et sa résistance. On sait qu'une augmentation de la température a pour conséquence une augmentation de l'activité des molécules, ce qui nuit au passage des charges. On a mesuré la résistance du conducteur en modifiant la température à pas constant et on a obtenu les résultats du tableau ci-contre. Déterminer un modèle décrivant la relation entre la température et la résistance.

T (°C)	R (Ω)
–10,0	26,6
–5,0	27,4
0,0	28,2
5,0	28,9
10,0	29,7
15,0	30,3
20,0	31,1
25,0	31,9

Solution

Représentation graphique des données

La représentation graphique des données est un nuage de points qui semblent alignés. On pose donc l'hypothèse qu'il existe un lien affine entre les variables.



Calcul des différences et de la pente

Puisque les mesures ont été prises à des intervalles de 1,2 cm, les données sont à pas constant. En calculant la différence des images, on obtient un critère dont les valeurs sont inscrites dans le deuxième tableau.

Le fait que la différence des images est relativement constante confirme l'hypothèse d'un lien affine entre les variables. La pente est égale à la moyenne des différences divisée par la longueur du pas de la variable indépendante

$$a = \frac{0,741}{5} = 0,1482.$$

T (°C)	R (Ω)	$R(T_i + p) - R(T_i)$
–10,0	26,6	–
–5,0	27,4	0,730
0,0	28,2	0,820
5,0	28,9	0,710
10,0	29,7	0,760
15,0	30,3	0,540
20,0	31,1	0,720
25,0	31,9	0,910
	Moyenne	0,741

Le modèle affine est donc de la forme

$$f(x) = 0{,}148\,2x + b$$

ou, si on utilise les symboles des variables du problème,

$$R(T_i) = 0{,}148\,2T_i + b.$$

T (°C)	R (Ω)	ΔR_i	$R_i - 0{,}1482T_i$
–10,0	26,6	–	28,142
–5,0	27,4	0,730	28,131
0,0	28,2	0,820	28,210
5,0	28,9	0,710	28,179
10,0	29,7	0,760	28,198
15,0	30,3	0,540	27,997
20,0	31,1	0,720	27,976
25,0	31,9	0,910	28,145
	Moyennes	0,741	28,119

Calcul de l'ordonnée

Il reste à calculer la valeur du paramètre b. Si on fait passer par chaque point du graphique une droite de pente 0,148 2, ces droites coupent l'axe vertical en des points distincts dont les ordonnées de ces points sont respectivement :

$$b_i = R(T_i) - 0{,}148\,2T_i.$$

On calcule chacune de ces ordonnées en ajoutant simplement une colonne au tableau. Ce type de traitement est simple et rapide avec le logiciel Excel.

La moyenne des ordonnées calculées est 28,119. On acceptera donc comme modèle :

$$R(T_i) = 0{,}148\,2T_i + 28{,}119\ \Omega.$$

Données à pas variable

Dans ce qui suit, nous présentons trois méthodes applicables autant à des données à pas variable qu'à des données à pas constant, ainsi que des mesures utilisées pour juger de la précision d'un modèle mathématique.

REMARQUE

Le critère algébrique est un moyen rapide de détecter un lien affine entre deux variables, mais il ne constitue pas le meilleur critère pour établir que le lien est bel et bien affine.

Dans une expérience de laboratoire portant sur des variables dont on sait que le lien est affine, le critère algébrique constitue une façon rapide de vérifier que les mesures sont prises correctement.

Méthode graphique

La façon la plus simple de construire un modèle affine consiste à représenter les couples de données sur du papier quadrillé et, à l'aide d'une règle transparente, choisir parmi toutes les droites passant par deux des points obtenus celle qui semble le mieux décrire le phénomène. On utilise les coordonnées $(x_1; y_1)$ et $(x_2; y_2)$ des points choisis pour écrire l'équation de la droite retenue comme modèle, soit en appliquant la procédure de géométrie analytique :

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}.$$

soit en résolvant le système d'équations :

$$y_1 = ax_1 + b$$
$$y_2 = ax_2 + b.$$

EXEMPLE 7.1.2

Une seconde équipe a réalisé la même expérience sur la relation entre la température et la résistance avec un autre conducteur et obtenu les correspondances inscrites dans le tableau ci-contre.

T (°C)	R (Ω)
–12,0	26,51
–8,0	27,22
–4,0	27,62
2,0	28,43
8,0	29,32
15,0	30,24
24,0	31,88
32,0	32,64

a) Construire un modèle mathématique décrivant le lien entre les variables.

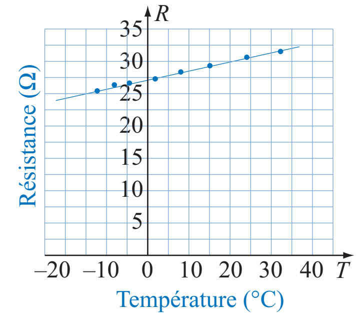
b) À l'aide de ce modèle, estimer la résistance à 50 °C.

c) Effectuer une mesure de la précision de ce modèle.

■ ***Solution***

a) Dans ce cas, le pas de la variable indépendante n'est pas constant et la méthode de l'exemple précédent n'est pas applicable.

On représente graphiquement les données du tableau, on constate que le nuage de points évoque une droite, mais les points ne sont pas parfaitement alignés, ce qui peut s'expliquer par des erreurs de mesure auxquelles s'ajoutent les erreurs lors de la représentation graphique.



Par exemple, supposons que la droite passant par les points (–12; 26,51) et (32; 32,64) nous semble la plus apte à décrire la relation entre les variables, en appliquant la procédure de géométrie analytique, on a

$$\frac{R-26,51}{T-(-12)}=\frac{32,64-26,51}{32-(-12)}=0,139\ 318\ 181.$$

En isolant R, on obtient

$$R = 0,139\,318\,181\,T + 28,181\,818\,18.$$

L'usage est de conserver dans le modèle un chiffre significatif de plus que dans la mesure qui en comporte le moins. On conserve donc cinq chiffres significatifs puisque les mesures de la résistance en comportent quatre et le modèle est

$$R(T) = 0,139\,32\,T + 28,182\ \Omega.$$

b) Pour estimer la résistance à 54 °C, il faut calculer l'image de 54 par le modèle. On obtient :

$$R(54) = 0,139\,32 \times 54 + 28,182 = 35,705\,28.$$

Après avoir effectué les calculs on doit arrondir de telle sorte que le résultat comporte le même nombre de chiffres significatifs que la mesure qui en compte le moins. On estime donc qu'à 54 °C, la résistance du conducteur sera de 35,71 Ω.

c) L'équipe doit mesurer la précision de son modèle. Pour ce faire, il faut calculer, pour chaque valeur de la variable indépendante, la différence entre la valeur observée (C_i) et la valeur donnée par le modèle mathématique (C_m). De telles différences sont appelées **résidus**. La somme des carrés des résidus est une mesure de la précision du modèle mathématique. Les résidus et leurs carrés sont inscrits dans le tableau présenté ci-dessous.

T_i	R_i	$R_m(T)$	Rés	Rés^2
–12,0	26,51	26,510 16	0,00016	0,00000
–8,0	27,22	27,067 44	–0,15256	0,02327
–4,0	27,62	27,624 72	0,00472	0,00002
2,0	28,43	28,460 64	0,03064	0,00094
8,0	29,32	29,296 56	–0,02344	0,00055
15,0	30,24	30,271 80	0,03180	0,00101
24,0	31,88	31,525 68	–0,35432	0,12554
32,0	32,64	32,640 24	0,00024	0,00000
			Somme	0,15134

REMARQUE

Le résidu est 0 lorsque la droite passe par le point. Si le résidu est négatif, la valeur calculée à l'aide du modèle est inférieure à celle observée. Si le résidu est positif, la valeur donnée par le modèle est plus grande que celle observée.

Méthode des données groupées

Cette méthode consiste à diviser les points en deux groupes contenant chacun la moitié, ou environ la moitié, des données. On calcule ensuite la valeur moyenne de la variable indépendante et de la variable dépendante dans chaque groupe. Les valeurs moyennes, représentées par $(\bar{x}_1;\ \bar{y}_1)$ et $(\bar{x}_2;\ \bar{y}_2)$, servent à écrire l'équation d'une droite à l'aide de la proportion

$$\frac{y-\bar{y}_1}{x-\bar{x}_1}=\frac{\bar{y}_2-\bar{y}_1}{\bar{x}_2-\bar{x}_1}.$$

On peut également trouver les valeurs de a et b en solutionnant le système d'équations suivant :

$$\bar{y}_1=a\bar{x}_1+b$$
$$\bar{y}_2=a\bar{x}_2+b.$$

EXEMPLE 7.1.3

T (°C)	R (Ω)
–12,0	26,51
–8,0	27,22
–4,0	27,62
2,0	28,43
8,0	29,32
15,0	30,24
24,0	31,88
32,0	32,64

Appliquer la méthode des données groupées aux résultats des tests de la seconde équipe, donnés dans le dernier exemple, pour construire un modèle affine décrivant la relation entre la température T et la résistance R du conducteur. Effectuer le calcul des résidus afin de mesurer la fiabilité du modèle obtenu.

Solution

En regroupant les données et en calculant les moyennes, on obtient

$$\bar{T}_1=\frac{-12,0-8,0-4,0+2,0}{4}=-5,5,$$

$$\bar{R}_1=\frac{26,51+27,22+27,62+28,43}{4}=27,445.$$

$$\bar{T}_2=\frac{8,0+15,0+24,0+32,0}{4}=19,75,$$

$$\bar{R}_2=\frac{29,32+30,24+31,88+32,64}{4}=31,02.$$

En appliquant la procédure de géométrie analytique, on a

$$\frac{R-27,445}{T-(-5,5)}=\frac{31,02-27,445}{19,75-(-5,5)}=0,141\ 584\ 158.$$

En isolant R et en arrondissant les paramètres à cinq chiffres significatifs, on obtient le modèle affine

$$R(T) = 0,141\ 58\,T + 28,224\ \Omega.$$

Le tableau ci-dessous donne la somme des carrés des résidus.

T_i	R_i	$R_m(T)$	Rés	Rés^2
–12,0	26,51	26,52504	0,01504	0,00023
–8,0	27,22	27,09136	–0,12864	0,01655
–4,0	27,62	27,65768	0,03768	0,00142
2,0	28,43	28,50716	0,07716	0,00595
8,0	29,32	29,35664	0,03664	0,00134
15,0	30,24	30,34770	0,10770	0,01160
24,0	31,88	31,62192	–0,25808	0,06661
32,0	32,64	32,75456	0,11456	0,01312
			Somme	0,11682

On constate que ce modèle est plus fiable que celui que l'on obtient avec la méthode graphique puisque la somme des carrés des résidus est plus petite.

Méthode des moindres carrés

La méthode des moindres carrés consiste à calculer :

$\overline{x}$, la moyenne des valeurs de la variable indépendante x;

$\overline{y}$, la moyenne des valeurs de la variable dépendante y;

$\overline{x^2}$, la moyenne des carrés des valeurs de la variable indépendante x;

$\overline{xy}$, la moyenne des produits des valeurs des deux variables.

On obtient ensuite les paramètres a et b de la droite recherchée en solutionnant le système d'équations

$$\begin{aligned} \overline{y} &= a\overline{x} + b \\ \overline{xy} &= a\overline{x^2} + b\overline{x}. \end{aligned}$$

La droite obtenue est appelée **droite de régression**, c'est la droite dont la somme des carrés des différences entre les valeurs observées et les points de la droite est minimale.

Calcul des paramètres d'une droite de régression

Les valeurs moyennes des variables sont définies par

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n} \quad \text{et} \quad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n}.$$

Si on remplace les variables par ces expressions dans les équations

$$\overline{y} = a\overline{x} + b \quad \text{et} \quad \overline{xy} = a\overline{x^2} + b\overline{x}$$

et qu'on isole les paramètres a et b, on obtient

$$a = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n\sum x_i^2 - \left(\sum x_i\right)^2} \quad \text{et} \quad b = \frac{\sum y_i - a\sum x_i}{n}.$$

Nous disposerons les données du problème en tableau et nous effectuerons les calculs dans ce tableau en ajoutant des colonnes pour chacune des sommes nécessaires pour calculer les paramètres. Nous utiliserons directement les expressions pour a et b en substituant les sommes du tableau pour calculer les paramètres de la droite de régression.

PROCÉDURE

Calcul des paramètres d'une droite de régression

1. Représenter graphiquement les données afin de s'assurer que le modèle affine est approprié.
2. Pour simplifier le traitement et la gestion des données, construire un tableau en réservant une colonne à chacune des grandeurs n, x, y, xy et x^2. La dernière ligne du tableau contient les sommations utilisées dans les formules de a et de b.

T (°C)	R (Ω)
–12,0	26,51
–8,0	27,22
–4,0	27,62
2,0	28,43
8,0	29,32
15,0	30,24
24,0	31,88
32,0	32,64

EXEMPLE 7.1.4

Appliquer la méthode de la droite de régression aux résultats des tests de la seconde équipe, rappelés dans le tableau ci-contre, pour construire un modèle affine décrivant la relation entre la température T et la résistance R du conducteur. Effectuer le calcul des résidus afin de mesurer la fiabilité du modèle obtenu.

Solution

On construit un tableau dans lequel on calcule les produits des valeurs correspondantes et les carrés des valeurs de la variable indépendante.

	T_i	R_i	T_iR_i	T_i^2
	–12,0	26,51	–318,12	144,0
	–8,0	27,22	–217,76	64,0
	–4,0	27,62	–110,48	16,0
	2,0	28,43	56,86	4,0
	8,0	29,32	234,56	64,0
	15,0	30,24	453,60	225,0
	24,0	31,88	765,12	576,0
	32,0	32,64	1 044,48	1 024,0
Σ	57,0	233,86	1 908,26	2 117,0

Dans la dernière ligne, on a les sommes à utiliser pour le calcul des paramètres a et b, on obtient :

REMARQUE
Nous utiliserons désormais, sans les démontrer, les expressions permettant de calculer directement les valeurs des paramètres a et b de la droite de régression.

$$a = \frac{n\sum T_iR_i - \left(\sum T_i\right)\left(\sum R_i\right)}{n\sum T_i^2 - \left(\sum T_i\right)^2}$$

$$= \frac{8\times(1\ 908,26) - 57,0\times 233,86}{8\times 2\ 117,0 - (57,0)^2} = 0,141\ 452\ 73$$

$$b = \frac{\sum R_i - a\sum T_i}{n} = \frac{233,86 - 0,141\ 452\ 73\times 57,0}{8}$$

$$= 28,224\ 649\ 3$$

En arrondissant les paramètres à cinq chiffres significatifs, on obtient le modèle affine

$$R(T) = 0,141\ 45\,T + 28,225\ \Omega.$$

Pour effectuer le calcul des résidus, on ajoute simplement des colonnes à notre tableau,

	T_i	R_i	T_iR_i	T_i^2	$R_m(T)$	Rés	Rés²
	–12,0	26,51	–318,12	144,0	26,527	0,01658	0,00027
	–8,0	27,22	–217,76	64,0	27,093	–0,12692	0,01611
	–4,0	27,62	–110,48	16,0	27,659	0,03889	0,00151
	2,0	28,43	56,86	4,0	28,508	0,07760	0,00602
	8,0	29,32	234,56	64,0	29,356	0,03632	0,00132
	15,0	30,24	453,60	225,0	30,346	0,10649	0,01134
	24,0	31,88	765,12	576,0	31,620	–0,26044	0,06783
	32,0	32,64	1 044,48	1 024,0	32,751	0,11118	0,01236
Σ	57,0	233,86	1 908,26	2 117,0			0,11677

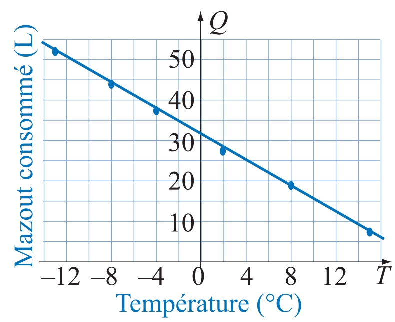
On constate que ce modèle est plus fiable que celui que l'on obtient avec les autres méthodes puisque la somme des carrés des résidus est plus petite.

EXEMPLE 7.1.5

L'entrepreneur en construction pour lequel vous travaillez a décidé d'évaluer les coûts de chauffage des maisons qu'il construit afin de se servir de ce renseignement dans sa publicité. Il a noté, pour des périodes de 24 heures, la consommation moyenne de mazout en fonction de la température extérieure moyenne durant ces 24 heures. Les données qu'il a obtenues sont inscrites dans le tableau présenté ci-contre.

T_i	Q_i
–13	52,0
–8	44,0
–4	36,8
2	28,0
8	18,0
15	6,8

Trouver, par la méthode des moindres carrés, le modèle affine décrivant la relation entre la température et la quantité de mazout consommée.



Solution

Identification des variables

La quantité de mazout consommé Q (L) dépend de la température extérieure T (°C). La représentation graphique des données est un nuage de points (présenté ci-contre) qui évoque une droite, même si les points ne sont pas parfaitement alignés.

Définition du lien entre les variables

Pour déterminer la valeur des paramètres de la droite, il faut calculer les produits des valeurs correspondantes et le carré des valeurs de la variable indépendante, puis faire la somme des données et de ces résultats. On peut présenter tous les calculs dans un même tableau, dont la dernière ligne est réservée aux sommes des valeurs inscrites dans les colonnes. En utilisant les formules des paramètres, on obtient :

Valeurs observées			
T	Q	TQ	T^2
–13	52,0	–676,0	169
–8	44,0	–352,0	64
–4	36,8	–147,2	16
2	28,0	56,0	4
8	18,0	144,0	64
15	6,8	102,0	225
0	185,6	–873,2	542

$$a = \frac{n\sum T_iQ_i - \left(\sum T_i\right)\left(\sum Q_i\right)}{n\sum T_i^2 - \left(\sum T_i\right)^2} = \frac{6\times(-873,2) - 0\times 185,6}{6\times 542 - (0)^2} = -1,611...$$

$$b = \frac{\sum Q_i - a\sum T_i}{n} = \frac{185,6 - (-1,611...)\times 0}{6} = 30,93...$$

Le modèle est donc $Q(T) = -1,611T + 30,93$.

Mesures de la précision du modèle

Jusqu'à quel point le modèle mathématique construit est-il fiable? Il existe des mesures qui permettent de répondre partiellement à cette question. Nous avons déjà présenté la somme des carrés des résidus, mais on utilise également le coefficient de corrélation et le coefficient de détermination.

Calcul des résidus

Pour mesurer le calcul de la somme des carrés des résidus, on utilise le tableau à l'aide duquel on a déterminé les paramètres du modèle affine. On ajoute des colonnes à ce tableau pour calculer les images à l'aide du modèle ainsi que les résidus et les carrés des résidus dont on fait la somme. Ainsi, pour le dernier exemple on obtient le tableau complémentaire suivant.

Valeurs observées				Valeurs du modèle	Résidus	Carrés des résidus
T_i	Q_i	T_iQ_i	T_i^2	$Q(T)$	r	r^2
–13	52,0	–676,0	169	51,87	0,123	0,015069
–8	44,0	–352,0	64	43,822	0,178	0,031722
–4	36,8	–147,2	16	37,378	–0,578	0,333638
2	28,0	56,0	4	27,711	0,289	0,083409
8	18,0	144,0	64	18,045	–0,045	0,002005
15	6,8	102,0	225	6,767	0,033	0,001070
0	185,6	–873,2	542			0,466913

Coefficient de corrélation

Le coefficient de corrélation est une mesure de l'intensité du lien entre deux variables. Il indique le degré de regroupement des points dans le voisinage de la droite. Il est défini par

$$r=\frac{n\sum x_iy_i-\left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\sum x_i^{\,2}-\left(\sum x_i\right)^2}\;\sqrt{n\sum y_i^{\,2}-\left(\sum y_i\right)^2}}.$$

Ainsi, pour le dernier exemple, on a

$$r=\frac{n\sum T_iQ_i-\left(\sum T_i\right)\left(\sum Q_i\right)}{\sqrt{n\sum T_i^{\,2}-\left(\sum T_i\right)^2}\;\sqrt{n\sum Q_i^{\,2}-\left(\sum Q_i\right)^2}}.$$

Le tableau du dernier exemple donne quatre des sommes apparaissant dans la formule de r. Il manque seulement $\sum Q_i^2$. On peut donc facilement calculer le coefficient de corrélation :

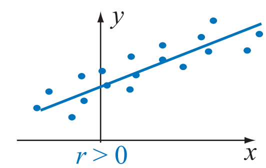
$$r=\frac{6\times(-873,2)-0\times 185,6}{\sqrt{6\times 542-(0)^2}\;\sqrt{6\times 7\ 148,48-(185,6)^2}}=-0,999\ 8.$$

Le coefficient de corrélation linéaire r est un nombre compris entre –1 et 1 ($-1 \leq r \leq 1$). Lorsque $r = 0$ (corrélation nulle), le modèle affine n'est pas du tout approprié au phénomène. Lorsque r est proche de 1 ou de –1, le regroupement des points dans le voisinage de la droite est important. Si la valeur de r est positive, les variables varient dans un même sens, c'est-à-dire que la valeur de la variable dépendante augmente lorsque la valeur de la variable indépendante augmente. Si la valeur de r est négative, les valeurs des variables varient en sens inverse, c'est-à-dire que la valeur de la variable dépendante diminue lorsque la valeur de la variable indépendante augmente. L'exemple 7.1.6 illustre de dernier cas : la quantité de mazout consommée diminue lorsque la température augmente. De plus, le coefficient r est –0,999 8, ce qui est très proche de –1. La corrélation est donc très forte.

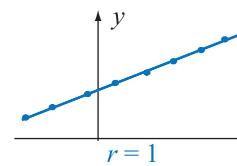
Le coefficient de détermination est le carré du coefficient de corrélation. Il est une mesure de la pertinence d'utiliser un modèle affine en faisant abstraction du fait que la corrélation peut être positive ou négative. C'est une mesure de l'adéquation entre le modèle et les données observées.

Droite de tendance

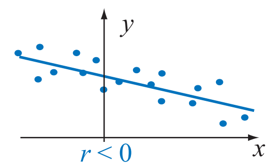
La droite de régression permet de construire un modèle simple, utilisé pour analyser des phénomènes ou décrire une tendance. On l'appelle alors **droite de tendance**. On distingue deux cas dans l'analyse de tendance, selon que les valeurs estimées sont à l'intérieur ou à l'extérieur de l'ensemble des données observées.



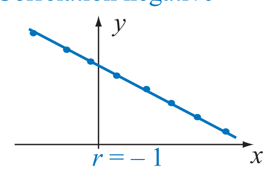
Corrélation positive



Corrélation positive parfaite



Corrélation négative



Corrélation négative parfaite

Interpolation

Lorsque les prévisions portent sur des valeurs à l'intérieur de l'intervalle des données, le processus est appelé **interpolation**. Généralement, les estimations par interpolation sont plutôt fiables.

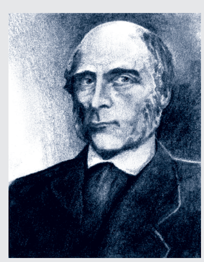
Extrapolation

Si les prévisions portent sur des valeurs à l'extérieur de l'ensemble des données, le processus est appelé **extrapolation**. Il est à noter que la fiabilité est plus grande lorsqu'on fait des prédictions pour des valeurs proches de l'ensemble des données observées. Une prédiction portant sur une valeur éloignée de cet intervalle donne une estimation qui, sans être à rejeter, doit être utilisée avec circonspection. Dans les deux cas, il ne faut pas s'attendre à ce que le soit modèle plus précis que les données qu'il décrit.

Un peu *d'histoire*

FRANCIS GALTON

1822-1911



Influencés par les travaux de Charles Darwin (1809-1882), les statisticiens anglais de la fin du XIX^{e} siècle utilisèrent les statistiques dans des contextes plus proches de la biologie que de la sociologie, comme le faisaient les statisticiens du continent européen. Francis Galton, cousin de Darwin, s'intéressa à des questions statistiques liées à la génétique, l'hérédité et le comportement humain. Alors qu'Adolphe Quételet (1796-1874) avait réalisé des travaux sur des données biométriques de l'homme, comme le poids, la taille et le périmètre thoracique, et avait montré que ces données se répartissaient selon une courbe normale, Galton mena des recherches sur la **variabilité des caractères**, les différences entre individus et les moyens pour conserver et favoriser les meilleurs d'entre eux. Sa contribution majeure est la notion de **corrélation** et la mesure de celle-ci, le **coefficient de corrélation**.

Lors d'études sur l'hérédité, réalisées en 1877, Galton se rendit compte que des parents de petite taille avaient des enfants plus petits que la moyenne, mais plus grands que leurs parents. De même, des parents plus grands que la moyenne avaient des enfants plus grands que la moyenne, mais plus petits que leurs parents. Ce phénomène indique qu'il y a **corrélation** entre la taille des parents et celle des enfants, mais qu'il y a également une **régression** par rapport à la moyenne, d'où l'appellation **droite de régression**. La régression vers la moyenne est en inversement proportionnelle à la corrélation. Dans ses travaux sur l'eugénisme, Galton étudia la dispersion des résultats et élabora les notions de médiane et de quartile. À l'époque, les travaux de Galton étaient perçus comme une contribution importante dans la lutte de la science contre l'obscurantisme religieux. Ils furent malheureusement utilisés comme justification pour les exactions commises dans l'Allemagne nazie.

À partir de 1865, Galton se consacra à la statistique dans le but de quantifier les caractéristiques physiques, psychiques et comportementales de l'être humain, ainsi que leur évolution.

Darwin avait énoncé ses lois de l'évolution dans aucune considération du calcul des probabilités, mais ses théories ont assuré le triomphe d'une description probabiliste du monde. Galton fit le lien entre la théorie de la sélection naturelle et la recherche mathématique, consacrant une large partie de son activité à la défense de la théorie de l'évolution et cherchant à montrer qu'elle fournit des prévisions susceptibles d'être vérifiées.

Un peu *d'histoire*

CARL FRIEDRICH GAUSS

1777-1855

Carl Friedrich Gauss, astronome, mathématicien et physicien, naquit le 23 avril 1777 à Brunswick, en Allemagne, dans une famille très modeste. Au cours de ses études élémentaires, deux de ses professeurs, Büttner et Bartels, remarquèrent son talent pour les mathématiques et l'aidèrent à entrer à l'école secondaire. En 1791, Bartels le présenta au duc de Brunswick. La recommandation de Bartels et les réalisations de Gauss incitèrent le duc à accorder une bourse à Gauss, en 1792, ce qui lui permit d'entrer au Collegium Carolinum de Brunswick en 1872, puis à l'Université de Göttingen en 1795. Gauss quitta l'université en 1798 sans avoir obtenu de diplôme, mais il avait déjà fait une importante découverte, soit la construction à la règle et au compas du polygone régulier à dix-sept côtés. Il retourna à Brunswick où il reçut un diplôme. Le duc de Brunswick décida de continuer à subvenir aux besoins de Gauss mais exigea que celui-ci soutienne une thèse de doctorat à l'Université de Helmstedt. Cette thèse portait sur le théorème fondamental de l'algèbre, qui stipule que tout polynôme est le produit de binômes de degré 1 et de trinômes de degré 2, tous irréductibles.

Grâce au soutien du duc de Brunswick, Gauss n'eut pas à chercher d'emploi et il put se consacrer entièrement à la recherche. Il apporta des contributions originales en théorie des nombres, en astronomie, en géodésie, en cartographie et dans toutes les branches des mathématiques. Il s'intéressa beaucoup aux géométries euclidiennes et non euclidiennes et élabora la méthode d'approximation par les moindres carrés. Il s'en servit pour résoudre de façon brillante un problème de son époque. L'Italien Giuseppe Piazzi (1746-1826) venait de découvrir, soit le 1er janvier 1801, le plus gros astéroïde entre Mars et Jupiter, nommé Cérès. Il n'avait pu observer qu'une petite partie de son orbite, soit 9°, avant que l'astéroïde ne disparaisse derrière le Soleil. Plusieurs savants tentèrent de décrire la trajectoire de Cérès à l'aide des données recueillies par Piazzi afin de déterminer à quel endroit l'astéroïde serait à nouveau visible. La prédiction la plus précise fut celle de Gauss.

Ayant perdu son protecteur, tué dans une bataille avec l'armée prussienne, Gauss quitta Brunswick, en 1807 pour occuper le poste de directeur de l'observatoire de Göttingen. Il s'intéressa à l'astéroïde Pallas, découvert par l'astronome Olbers en mars 1802. Ses travaux sur cet astéroïde l'amenèrent à résoudre un système de six équations linéaires à six inconnues. Gauss a laissé son nom à la méthode qu'il employa, soit la construction d'un système d'équations équivalent à celui à résoudre, dans lequel la première équation contient les six inconnues, la seconde seulement cinq, la troisième quatre, et ainsi de suite. Un tel système se résout par substitution en commençant par la sixième équation.

Au début du XIXe siècle, s'imposa une loi fondamentale de la statistique, soit la **loi normale** (d'abord appelée **loi des possibilités**, puis renommée par Karl Pearson), issue de la méthode des moindres carrés. Employée comme critère d'optimisation, en **théorie des erreurs**, cette loi fut élaborée indépendamment par Adrien Marie Legendre et Carl Friedrich Gauss.

Méthode des moindres carrés

Il est toujours délicat de tirer des conclusions de mesures. Le problème, c'est que toute mesure comporte une erreur. L'effet de la température sur les instruments, l'imprécision des lectures et des visées ne sont que quelques sources d'erreurs, et celles-ci sont d'autant plus nombreuses que le nombre de mesures est grand. Le fait de ne prendre que quelques mesures ne règle pas le problème, car il est alors difficile d'estimer l'ordre de grandeur des erreurs. Par ailleurs, lorsque le nombre de mesures est élevé, la manipulation de celles-ci pose un autre problème : « comment utiliser toutes ces mesures de façon à minimiser l'effet des erreurs sur le résultat final ». Le problème fut étudié par Legendre (Adrien Marie, 1752-1833) au début du XIXe siècle. Legendre voulait déterminer l'orbite d'une comète à l'aide de mesures de sa position. Son objectif était de combiner plusieurs mesures pour calculer la meilleure estimation de l'orbite. Ce type de problème est également relié à la découverte d'éventuelles planètes inconnues grâce aux perturbations de la trajectoire d'une planète ou d'une comète qu'elle provoquent. À l'époque, le problème des mesures s'est aussi posé dans les opérations de triangulation visant à mesurer un méridien (définition du mètre) ou à déterminer la forme de la Terre (aplatissement aux pôles). C'est en 1805 que Legendre publia sa méthode de minimalisation de la somme des carrés des écarts. Gauss avait déjà élaboré cette méthode en 1794 et l'utilisa, en 1801, pour calculer l'orbite de l'astéroïde Cérès, mais il ne la publia qu'en 1809. Gauss établit de plus des liens entre cette méthode et les lois de probabilité, ce que Legendre ne fit pas. Plus précisément, Gauss montra, en utilisant la méthode des moindres carrés, que les facteurs aléatoires indépendants entraînent des erreurs de mesure qui se répartissent selon la loi normale, et que la moyenne arithmétique des mesures donne l'estimation qui minimise la somme des carrés des erreurs.

Gauss fut l'un des savants éminents de l'histoire. Il est considéré par plusieurs historiens des sciences comme l'un des trois plus grands mathématiciens de tous les temps, avec Archimède et Newton.

7.2 Exercices

1. Lors d'une expérience consistant à étudier la relation entre la température et la résistance d'un conducteur, on a déterminé les données du tableau suivant.

T (°C)	R (Ω)
70	26,4
80	26,7
90	26,9
100	27,2
110	27,4
120	27,7

a) Représenter graphiquement les données. Quelle hypothèse peut-on faire sur le lien entre ces variables ?

b) Appliquer la méthode des données à pas constant pour déterminer un modèle affine décrivant la relation entre les variables. Effectuer le calcul des résidus.

c) Appliquer la méthode des moindres carrés pour déterminer un modèle affine décrivant la relation entre les variables. Effectuer le calcul des résidus.

d) Indiquer lequel est le meilleur modèle e justifiant votre conclusion.

2. Un entrepreneur en construction a décidé d'évaluer les coûts de chauffage des maisons qu'il bâtit afin de se servir de ce renseignement dans sa publicité. Il a noté la consommation moyenne de mazout durant des périodes de 24 heures en fonction de la température extérieure moyenne durant ces 24 heures. Les données qu'il a obtenues sont présentées dans le tableau suivant.

T (°F)	Q (L)
–11	48,0
–7	41,0
–1	32,0
2	27,0
6	20,0
12	11,0

a) Représenter graphiquement les données. Que suggère cette représentation ?

b) Construire un modèle affine décrivant la relation entre la température et la quantité de mazout consommée.

c) Évaluer la quantité de mazout consommée en une journée lorsque la température extérieure moyenne est de 9 °F.

d) Dans le cas où la température moyenne en janvier est de –12 °F, estimer la consommation de mazout durant ce mois.

e) Évaluer la quantité de mazout consommée en une journée lorsque la température extérieure moyenne est de –20 °F.

3. Une association d'automobilistes a demandé à ses membres de noter la distance qu'ils ont parcourue et le coût d'utilisation de leur véhicule au cours de la dernière année, en incluant les coûts de l'immatriculation, des assurances, de l'essence et de l'entretien. L'association a dressé le tableau suivant à l'aide des informations reçues pour la voiture la plus populaire auprès de ses membres.

Distance (km)	Coût ($)
5 000	3 950
10 000	4 860
15 000	5 740
20 000	6 600
25 000	7 520
30 000	8 460

a) Construire un modèle mathématique décrivant la correspondance entre les deux variables.

b) Donner une mesure de la précision du modèle en calculant les résidus.

c) Prévoir, à l'aide du modèle, le coût d'utilisation de la voiture en question dans le cas où la distance parcourue en une année est de 45 000 km.

4. Une entreprise veut mettre sur le marché un nouveau jeu vidéo. Elle a effectué une étude de marché afin de fixer le prix de ce produit. Les résultats de l'étude sur le prix et le volume estimé des ventes annuelles, sont présentés dans le tableau suivant.

Prix ($)	Volume des ventes
35	540
40	492
45	458
50	406
55	336
60	294

a) Déterminer la règle de correspondance entre le prix de l'article et le nombre de clients potentiels.

b) Estimer la précision du modèle à l'aide des résidus et du coefficient de corrélation.

5. Lors d'une expérience consistant à étudier la relation entre la température et la résistance d'un conducteur d'un nouvel alliage, on a déterminé les données du tableau suivant.

T (°C)	R (Ω)
–25,5	11,62
–10,4	12,96
2,5	14,15
6,3	14,51
12,7	15,03
23,4	16,03
55,4	18,88
72,8	20,50

a) Représenter graphiquement les données. Quelle hypothèse peut-on faire sur le lien entre ces variables ?
b) Si l'hypothèse d'un modèle affine est plausible, appliquer la méthode des moindres carrés pour déterminer un modèle décrivant la relation entre les variables et effectuer le calcul des résidus.
c) Utiliser le modèle pour estimer la résistance si le conducteur est utilisé à une température de 45 °C.

6. Dans une expérience de laboratoire, on a obtenu les données du tableau suivant.

x	y
2,02	53,15
5,04	49,79
8,12	46,32
12,14	41,92
15,37	38,12
19,22	33,75
22,51	30,08
27,42	24,62

a) Représenter graphiquement les données. Quelle hypothèse peut-on faire sur le lien entre ces variables ?
b) Si l'hypothèse d'un modèle affine est plausible, appliquer la méthode des moindres carrés pour déterminer un modèle décrivant la relation entre les variables, et effectuer le calcul des résidus.
c) Utiliser le modèle pour estimer la valeur de la variable dépendante lorsque la variable indépendante est égale à 24,65.
d) Utiliser le modèle pour estimer la valeur de la variable indépendante lorsque la variable dépendante est égale à 32,15.

7. Dans une expérience de laboratoire, on a obtenu les données du tableau suivant.

x	y
–5,28	5,15
1,14	18,65
6,15	29,84
8,17	33,67
19,31	57,85
24,67	68,04
29,32	79,22
35,43	92,12

a) Représenter graphiquement les données. Quelle hypothèse peut-on faire sur le lien entre ces variables ?
b) Si l'hypothèse d'un modèle affine est plausible, appliquer la méthode des moindres carrés pour déterminer un modèle décrivant la relation entre les variables, et effectuer le calcul des résidus.
c) Utiliser le modèle pour estimer la valeur de la variable dépendante lorsque la variable indépendante est égale à 12,15.
d) Utiliser le modèle pour estimer la valeur de la variable indépendante lorsque la variable dépendante est égale à 21,15.

8. Dans une expérience de laboratoire, on a obtenu les données du tableau suivant.

x	y
2,5	21,41
3,8	15,47
4,5	12,99
6,1	8,70
7,3	6,45
8,1	5,28
10,2	3,12
11,3	2,37
13,5	1,37
15,8	0,77

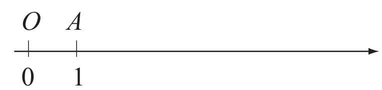
a) Représenter graphiquement les données. Quelle hypothèse peut-on faire sur le lien entre ces variables ?
b) Si l'hypothèse d'un modèle affine est plausible, appliquer la méthode des moindres carrés pour déterminer un modèle décrivant la relation entre les variables, et effectuer le calcul des résidus.
c) Utiliser le modèle pour estimer la valeur de la variable dépendante lorsque la variable indépendante est égale à 18,65.

7.3 Échelles graphiques

Dans la présente section, nous appliquons la régression affine à l'élaboration d'une procédure d'analyse permettant de décider quel est le modèle le plus approprié et de définir ce dernier. Cette méthode repose sur la représentation graphique à échelle logarithmique. Nous allons donc d'abord, présenter les caractéristiques d'une telle échelle en la comparant à une échelle linéaire.

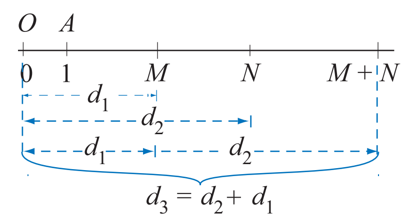
Échelle linéaire

Une échelle est dite **linéaire** si son pas est constant, c'est-à-dire si chaque nombre est situé à une distance de l'origine qui est proportionnelle à sa valeur. La droite représentée ci-contre comporte un point origine O et un point A qui détermine la valeur unitaire ou la longueur du pas de l'échelle.

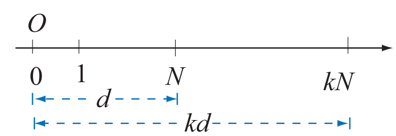


Pas unitaire d'une échelle linéaire

Si la droite est graduée selon la longueur unitaire $\overline{OA}$ et si M et N sont deux nombres positifs *situés* à des distances respectives d_1 et d_2 de l'origine, en respectant la proportionnalité, alors leur somme est un nombre $V = M + N$ représenté par un point situé à une distance $d_1 + d_2$ de l'origine. De plus, si un nombre $N > 0$ est situé à une distance d de l'origine, pour tout nombre $k > 0$, le nombre kN est situé à une distance kd de l'origine.

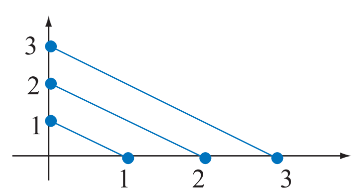


Somme sur une échelle linéaire



Produit sur une échelle linéaire

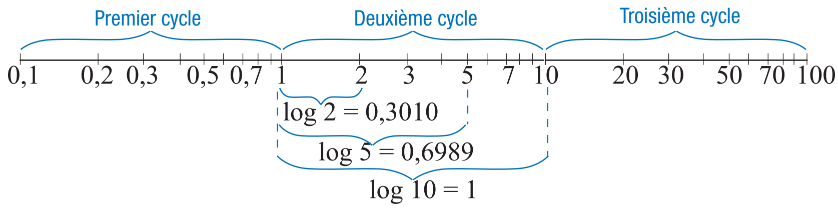
Dans un système d'axes gradués selon des échelles linéaires de pas différents, les segments de droite joignant deux points des axes de même valeur déterminent des triangles semblables, car la distance à l'origine de n'importe quel nombre est proportionnelle à sa valeur.



Proportionnalité de deux échelles linéaires

Échelle logarithmique

Nous avons déjà souligné que la droite est la représentation graphique la plus facile à reconnaître. Pour déceler un lien non affine entre deux variables, il est d'usage d'utiliser du papier quadrillé dont au moins l'une des deux échelles est graduée à l'aide du logarithme de base 10. Sur une échelle logarithmique, l'origine correspond au nombre 1, car (0; 0) = (0; log 1). La position des autres nombres est déterminée de telle sorte que leur distance à l'origine soit proportionnelle au logarithme du nombre. Ainsi, puisque le logarithme de base 10 de 5 est 0,698 9 et que le logarithme de 10 est 1, la distance de 1 à 5 correspond à 69,89 % de la distance de 1 à 10. Puisque le logarithme de 100 est 2, la distance de 1 à 100 est égale à deux fois la distance de 1 à 10; la distance entre 0,1 et 1 est égale à la distance entre 1 et 10 puisque le logarithme de 0,1 est égal à –1. Chacun des intervalles représentant une unité logarithmique est appelé **cycle**. Ainsi, l'intervalle de 0,1 à 1 est un cycle, tout comme les intervalles de 1 à 10 et de 10 à 100.



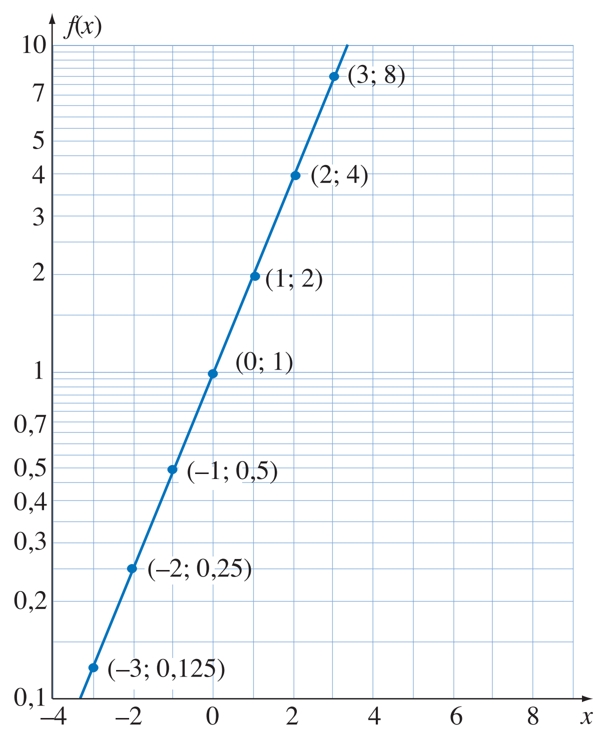
Graduations sur une échelle logarithmique

Du papier quadrillé suivant une échelle linéaire et une échelle logarithmique est appelé **papier semi-logarithmique** et un papier quadrillé suivant deux échelles logarithmiques est appelé **papier logarithmique**. Sur ces deux types de papiers, il n'y a pas de nombres indiquant les graduations; l'échelle commence à n'importe quel nombre suivant les besoins du problème. Dans les premiers exercices, nous indiquons les graduations pour permettre au lecteur de se familiariser avec ce genre de représentations graphiques. La caractéristique la plus intéressante est le fait que le graphique d'une fonction exponentielle sur du papier semi-logarithmique est une droite, comme l'illustre la représentation de la fonction $f(x) = 2^x$ sur un papier semi-logarithmique à deux cycles.

REMARQUE
Dans la démarche d'analyse menant à la modélisation d'une liste de données, nous représentons ces données dans un système de référence bilinéaire, semi-logarithmique ou bi-logarithmique.

Dans le graphique, le point désigné par (2; 4) correspond en réalité au point (2; log 4) puisque sa distance à l'axe des x est proportionnelle au logarithme de la valeur de la variable dépendante.

x	$f(x)$
–3	0,125
–2	0,25
–1	0,5
0	1
1	2
2	4
3	8



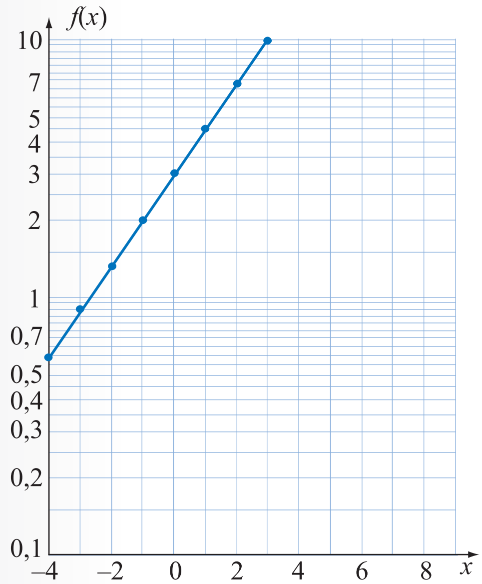
EXEMPLE 6.3.1

Représenter $f(x) = 3 \times 1{,}5^x$ sur du papier semi-logarithmique.

■ ***Solution***

On calcule d'abord quelques correspondances que l'on inscrit dans un tableau comme le suivant.

x	$f(x)$
–4	0,59
–3	0,89
–2	1,33
–1	2
0	3
1	4,5
2	6,75
3	10,1
4	15,2
5	22,8



Même si la base de la fonction exponentielle est différente de 10, la représentation graphique est une droite. La raison en est fort simple. Soit, par exemple, une fonction exponentielle de la forme

$$y = ab^x.$$

On a

$\log y = \log(ab^x)$. Logarithme de chaque membre de l'équation.

$\log y = \log a + \log b^x$. Propriété du logarithme d'un produit.

$\log y = \log a + x \log b$. Propriété du logarithme d'une puissance.

$\log y = x \log b + \log a$. Commutativité de l'addition.

$Y = Ax + B$, où $Y = \log y$, $A = \log b$ et $B = \log a$.

Puisque $\log b$ et $\log a$ sont des constantes, il existe une relation affine entre x et $\log y$. C'est pourquoi la représentation graphique sur du papier semi-logarithmique donne une droite.

En prenant plutôt le logarithme naturel de chaque membre de l'équation $y = ab^x$, on obtient la relation

$$\ln y = x \ln b + \ln a.$$

Ainsi, qu'on effectue les calculs dans **l'une ou l'autre des bases, on obtient le même modèle.**

REMARQUE
Le texte ci-contre est particulièrement important. Lorsqu'on dit, par exemple, que :

$$\log[\alpha] = 21{,}7\times10^{-3}T + 36{,}8\times10^{-4}$$

définit une relation affine, il faut comprendre que la relation dont on parle est entre T et $\log[\alpha]$ alors que $\log[\alpha]$ est une fonction de T :

$$\log[\alpha] = f(T)..$$

Échelle logarithmique et modélisation

Les caractéristiques des échelles logarithmiques indiquent comment utiliser du papier semi-logarithmique pour reconnaître un lien exponentiel entre des variables et déterminer la règle de correspondance décrivant ce lien. L'exemple suivant illustre la marche à suivre.

EXEMPLE 6.3.2

Au cours d'une expérience de laboratoire, on a obtenu les grandeurs physiques inscrites dans le tableau présenté ci-contre.

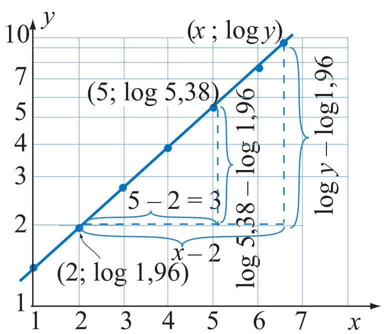
x	y
1	1,40
2	1,96
3	2,74
4	3,84
5	5,38
6	7,53

a) Vérifier l'hypothèse d'un lien exponentiel entre les variables x et y.

b) Déterminer la règle de correspondance décrivant le lien entre les deux variables.

■ *Solution*

a) On représente les données dans un système d'axes semi-logarithmique. Puisque les valeurs de la variable dépendante s'échelonnent de 1,40 à 7,53, on a besoin d'un seul cycle. Le graphique sur du papier semi-logarithmique dont l'échelle logarithmique est verticale est une droite, ce qui indique l'existence d'un lien exponentiel entre les variables.



b) Pour trouver la description algébrique du lien entre les variables, on applique la méthode des moindres carrés en utilisant le logarithme des valeurs de la variable dépendante. On peut effectuer les calculs dans la base naturelle ou la base 10. Si on choisit la base e, on obtient le tableau présenté ci-contre. Pour déterminer les paramètres du modèle par la méthode des moindres carrés, on prend le logarithme des valeurs de la variable dépendante. En effectuant les calculs en base e, on obtient :

x	y	$\ln y$	$x \ln y$	x^2
1	1,40	0,336 5	0,336 5	1
2	1,96	0,672 9	1,345 9	4
3	2,74	1,008 0	3,023 9	9
4	3,84	1,345 5	5,381 9	16
5	5,38	1,682 7	8,413 4	25
6	7,53	2,018 9	12,113 4	36
21	22,85	7,064 4	30,614 9	91

$$A = \frac{n\sum x_i \ln y_i - \left(\sum x_i\right)\left(\sum \ln y_i\right)}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$= \frac{6\times 30{,}614\ 9 - 21\times 7{,}064\ 4}{6\times 91 - 21\times 21} = 0{,}336\ 5.$$

$$B = \frac{\sum \ln y_i - A\ \left(\sum x_i\right)}{n}$$

$$= \frac{7{,}064\ 4 - 0{,}336\ 5\times 21}{6} = -0{,}000\ 48.$$

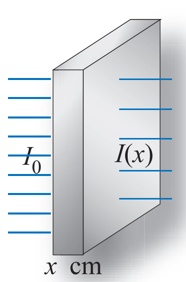
La valeur de B est négligeable, compte tenu de la précision des données de départ. Le lien entre les variables est donc :

$$\ln y = 0{,}3365x.$$

On trouve le lien exponentiel en exprimant cette équation sous forme exponentielle

$$y = e^{0{,}3365x} = 1{,}4^x.$$

EXEMPLE 6.3.3



x	$I(x)$
0,00	1
2,00	0,84
5,00	0,65
6,50	0,57
9,00	0,46
12,00	0,35
16,00	0,25
22,00	0,15

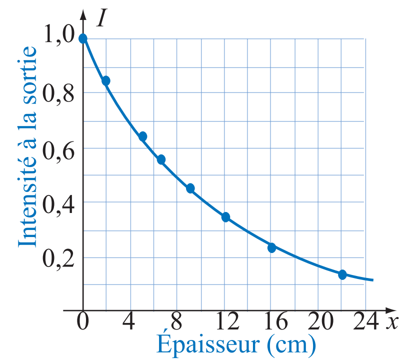
On désire analyser la capacité d'absorption des rayons X d'un matériau. Pour ce faire, on bombarde des plaques de différentes épaisseurs et on mesure l'intensité $I(x)$ des radiations à la sortie des plaques. En prenant que $I_0 = 1$ unité comme intensité des radiations à l'entrée, on a obtenu les mesures inscrites dans le tableau pour différentes épaisseurs x, en centimètres.

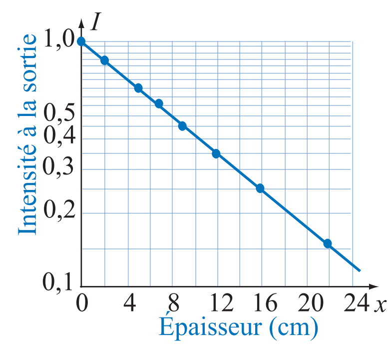
a) Quel type de correspondance relie les variables.

b) Déterminer la règle de correspondance.

■ ***Solution***

a) On représente les données sur du papier bilinéaire, ce qui donne la représentation ci-contre. Le graphique représenté ci-contre étant une courbe, on conclut qu'il ne s'agit pas d'une correspondance affine. La représentation graphique sur du papier semi-logarithmique dont l'échelle logarithmique est verticale est une droite, ce qui confirme l'existence d'un lien exponentiel entre les variables.





b) Pour trouver la description algébrique du lien exponentiel entre les variables, on calcule le logarithme des valeurs de la variable dépendante, ce qui donne le tableau présenté ci-contre. On obtient par régression

$$A = \frac{n\sum x_i \ln I_i - \left(\sum x_i\right)\left(\sum \ln I_i\right)}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$= \frac{8\times(-89,660\ 4)-(72,50\times -6,277\ 0)}{8\times 1036,25-72,50\times 72,50} \approx -0,086\ 427\ 919...$$

$$B = \frac{\sum \ln I_i - A\left(\sum x_i\right)}{n}$$

$$= \frac{-6,2770-(-0,086\ 427\times 72,50)}{8} \approx -0,000\ 138\ 03.$$

La valeur de B est négligeable compte tenu de la précision des données. Elle est théoriquement nulle, car la valeur initiale de I est 1 et une plaque d'épaisseur nulle n'absorbe pas de rayons X. Donc,

$$\ln I = -0,086\ 428x.$$

Ce qui s'écrit sous forme exponentielle,

$$I = e^{-0,086\ 428x} = 0,917^x.$$

Le modèle est $I(x) = 0,917^x$. Il est à noter que la plupart des calculatrices sont munies de fonctions permettant de calculer directement les paramètres A et B.

x	I	$\ln I$	$x \ln I$	x^2
0,00	1,00	0,0000	0,0000	0,00
2,00	0,84	–0,1744	–0,3487	4,00
5,00	0,65	–0,4308	–2,1539	25,00
6,50	0,57	–0,5621	–3,6538	42,25
9,00	0,46	–0,7765	–6,9888	81,00
12,00	0,35	–1,0498	–12,5979	144,00
16,00	0,25	–1,3863	–22,1807	256,00
22,00	0,15	–1,8971	–41,7366	484,00
72,50	4,27	–6,2770	–89,6604	1 036,25

Fonction puissance

La représentation graphique sur du papier logarithmique permet également de reconnaître une fonction puissance ou une fonction logarithmique. Une fonction puissance est de la forme $y = ax^b$. En prenant le logarithme de chaque membre de l'équation, on obtient :

$\log y = \log(ax^b)$

$\log y = \log a + \log x^b$. Propriété du logarithme d'un produit.

$\log y = \log a + b \log x$. Propriété du logarithme d'une puissance.

$\log y = b \log x + \log a$. Commutativité de l'addition.

En posant $Y = \log y$, $A = b$, $X = \log x$ et $B = \log a$, on a $Y = AX + B$.

Il existe donc entre $\log x$ et $\log y$, une correspondance affine que la représentation des données sur du papier logarithmique met en évidence.

Fonction logarithmique

En ce qui concerne la fonction logarithmique, l'équation $y = a \log x + b$, indique clairement qu'il existe une relation affine entre y et $\log x$, représentée symboliquement par $y = AX + B$ où $A = a$, $X = \log x$ et $B = b$. Cette relation est mise en évidence par sa représentation sur du papier semi-logarithmique, la variable indépendante étant portée sur l'échelle logarithmique. Si le nuage de points évoque une droite, le modèle est logarithmique.

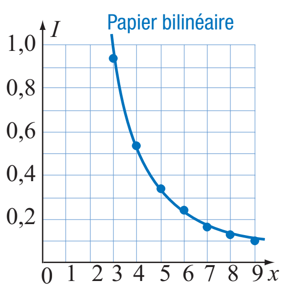
EXEMPLE 6.3.4

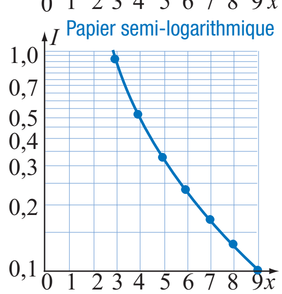
On a obtenu en laboratoire les données présentées dans le tableau ci-contre.

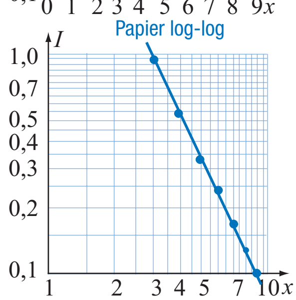
a) Quel type de correspondance relie les deux variables.

b) Déterminer la règle de correspondance.

x	$I(x)$
3,00	0,94
4,00	0,53
5,00	0,34
6,00	0,24
7,00	0,17
8,00	0,13
9,00	0,10







Solution

a) La représentation graphique sur du papier bilinéaire ou du papier semi-logarithmique est une courbe.

La représentation sur du papier log-log laisse penser qu'une fonction puissance pourrait décrire le lien entre les deux variables. On établit donc une relation affine entre $\ln x$ et $\ln I$. Les résultats des calculs préliminaires sont rassemblés dans le tableau suivant.

	x	I	$\ln x$	$\ln I$	$\ln x \ln I$	$(\ln x)^2$
	3,00	0,94	1,098 6	–0,061 9	–0,068 0	1,206 9
	4,00	0,53	1,386 3	–0,634 9	–0,880 1	1,921 8
	5,00	0,34	1,609 4	–1,078 8	–1,736 3	2,590 3
	6,00	0,24	1,791 8	–1,427 1	–2,557 0	3,210 4
	7,00	0,17	1,945 9	–1,772 0	–3,448 1	3,786 6
	8,00	0,13	2,079 4	–2,040 2	–4,242 5	4,324 1
	9,00	0,10	2,197 2	–2,302 6	–5,059 3	4,827 8
Σ	42,00	2,45	12,108 7	–9,317 4	–17,991 3	21,867 9

$$A = \frac{n\sum \ln x_i \ln I_i - \left(\sum \ln x_i\right)\left(\sum \ln I_i\right)}{n\sum (\ln x_i)^2 - \left(\sum \ln x_i\right)^2}$$

$$= \frac{7\times(-17{,}991\ 3) - (12{,}108\ 7\times -9{,}317\ 4)}{7\times 21{,}867\ 9 - (12{,}108\ 7)^2} \approx -2{,}032\ 24...$$

$$B = \frac{\sum \ln I_i - A\left(\sum \ln x_i\right)}{n}$$

$$= \frac{-9{,}317\ 4 - (-2{,}032\ 244...\times 12{,}108\ 7)}{7} \approx 2{,}184\ 34...$$

b) La relation affine est $\ln I = -2{,}032\,2 \ln x + 2{,}184\,3$; donc,

$$\ln I = \ln x^{-2{,}032\,2} + 2{,}184\,3$$

$$I = e^{2{,}184\,3} x^{-2{,}032\,2} = 8{,}8844 x^{-2{,}032\,2}.$$

Compte tenu de la précision des données, le modèle est :

$$I = \frac{8{,}88}{x^{2{,}03}}.$$

Les valeurs des paramètres A et B varient selon le nombre de chiffres significatifs retenus pour les calculs, particulièrement si y interviennent des logarithmes. Idéalement, on devrait retenir tous les chiffres et arrondir seulement à la fin des calculs en tenant compte de la précision des mesures à modéliser.

EXEMPLE 6.3.5

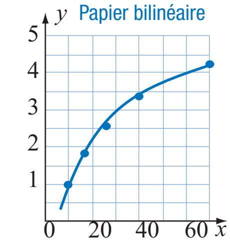
On a obtenu les données du tableau présenté ci-contre en laboratoire.

a) Quel type de correspondance relie les deux variables.

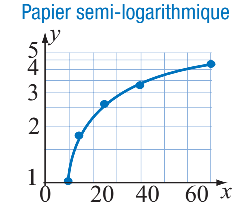
b) Déterminer la règle de correspondance.

x	y
10	1,00
15	1,70
25	2,59
40	3,41
70	4,38

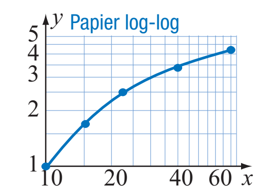
Papier bilinéaire



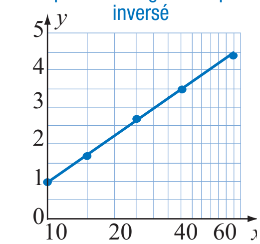
Papier semi-logarithmique



Papier log-log



Papier semi-logarithmique inversé



Solution

a) La représentation graphique sur du papier bilinéaire, sur papier semi-logarithmique ou log-log est une courbe. Cependant, en portant les valeurs de la variable indépendante sur l'axe logarithmique d'un papier semi-logarithmique, on obtient une droite. Le modèle logarithmique est donc approprié.

b) Les résultats des calculs préliminaires sont rassemblées dans le tableau suivant.

	x	y	$\log x$	$y \log x$	$(\log x)^2$
	10	1,00	1,000	1,000	1,000
	15	1,70	1,176	1,999	1,383
	25	2,59	1,398	3,621	1,954
	40	3,41	1,602	5,463	2,567
	70	4,38	1,845	8,081	3,404
Σ	160	13,08	7,021	20,165	10,308

$$A = \frac{n\sum(\log x_i)\, y_i - \left(\sum \log x_i\right)\left(\sum y_i\right)}{n\sum(\log x_i)^2 - \left(\sum \log x_i\right)^2}$$

$$= \frac{5\times(20{,}165) - (7{,}021\times 13{,}08)}{5\times 10{,}308 - (7{,}021)^2} \approx 4{,}003\ 59\ldots$$

$$B = \frac{\sum y_i - A\left(\sum \log x_i\right)}{n}$$

$$= \frac{13{,}08 - (4{,}003\ 59\ldots\times 7{,}021)}{5} \approx -3{,}005\ 855\ldots$$

Le modèle est donc : $y = 4{,}00 \log x - 3{,}01$.

Il n'est pas toujours nécessaire de représenter les données sur autant d'échelles différentes. Si on sait quel type de relation lie les variables est connue, on choisit tout de suite une échelle appropriée. On peut cependant utiliser du papier linéaire tout en sachant que la relation est exponentielle, logarithmique ou de puissance. Il faut alors effectuer des calculs sur les données expérimentales pour déterminer les valeurs à porter dans le graphique. La démarche élaborée dans le présent chapitre est très utile pour analyser des données expérimentales. .

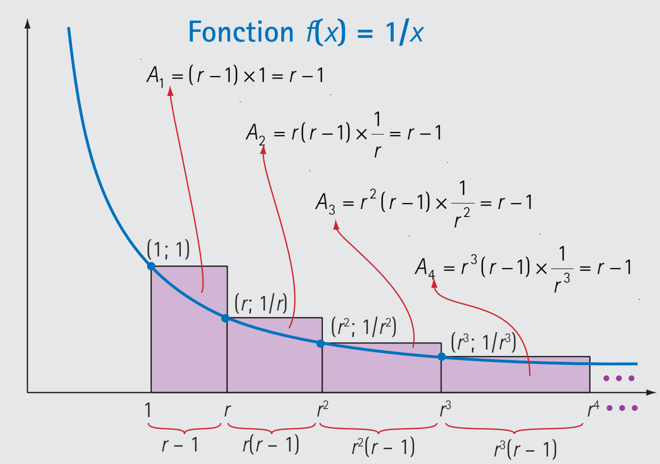
Un peu d'histoire

AVÈNEMENT DES LOGARITHMES

En développant les logarithmes, Napier a fait beaucoup plus que développer des méthodes facilitant les calculs. Il a permis la découverte d'une nouvelle fonction et ouvert la voie à de nouvelles techniques de modélisation.

Grégoire de Sant-Vincent 1584-1667

En 1647, Grégoire de Saint-Vincent (1584-1687) travaille sur la quadrature de l'hyperbole[1] et met en évidence une nouvelle fonction (NH St-Vincent). En 1661, Christiaan Huygens (1629-1695) remarque que cette fonction se trouve être une fonction logarithme particulière : le logarithme naturel. La notion de fonction et la correspondance entre les fonctions exponentielles et les fonctions logarithmes apparaissent après le travail de Leibniz sur la notion de fonction, en 1697.



La somme des bases des rectangles est une progression géométrique, et la somme des aires est une progression arithmétique, puisque les rectangles ont la même aire.

1. La quadrature d'une surface est la recherche d'un carré ayant même aire que la surface en question. La quadrature de l'hyperbole porte sur l'aire entre la courbe et l'axe horizontal.

Le mathématicien anglais d'origine galloise Edmund Gunter (1581-1626) développe l'échelle logarithmique (NH ÉchelleLog) et l'utilise dans une première version de règle à calcul. Jumelée à la régression linéaire, l'échelle logarithmique constitue un outil de modélisation important du lien entre deux variables (NH Log-Modélisation).

Henrietta Leavitt 1868-1921

C'est en utilisant une représentation de mesures à l'aide d'une échelle logarithmique qu'Henrietta Leavitt (NH Leavitt) a pu déterminer la relation entre la luminosité et la période des céphéïdes. La relation période-luminosité découverte par Leavitt est à la base d'une méthode d'évaluation des distances des amas stellaires et des galaxies dans l'Univers. C'est la méthode utilisée par Edwin Hubble pour conclure à l'expansion de l'univers.

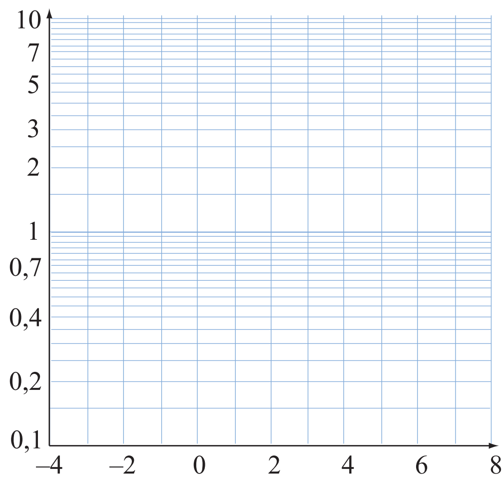
C'est également en utilisant une échelle logarithmique que Vilfredo Pareto a établi la relation entre le revenu familial et le nombre de familles ayant ce revenu (NH Pareto) . En fait, on retrouve les exponentielles et les logarithmes dans plusieurs domaines, la croissance des capitaux en gestion, la croissance des populations en biologie et la radioactivité en physique. On les retrouve également en chimie, notamment dans les équations d'Arrhenius (NH Arrhenius).

La découverte de la radioactivité et la description de la quantité de matière radioactive en fonction du temps par une fonction exponentielle a mis à la disposition des savants un moyen de datation très efficace. La fonction inverse, qui est une fonction logarithmique, permet de calculer le temps écoulé depuis le début de la désintégration de la matière radioactive. Cette découverte a facilité les démarches de datation en archéologie et en sciences de la nature (NH Log-Datation01-02).

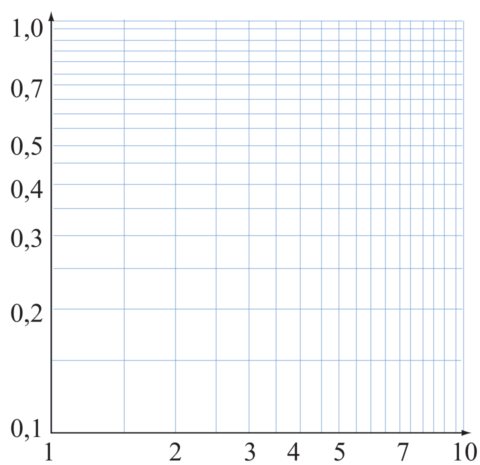
Vilfredo Pareto 1848-1923

7.4 Exercices

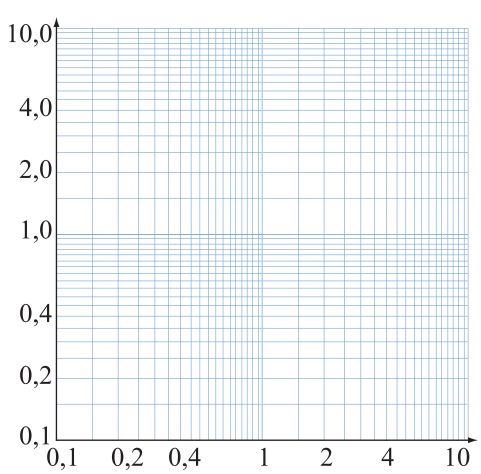
1. Représenter $f(x) = 1{,}8^x$ sur du papier semi-logarithmique.



2. Représenter $f(x) = 1/x$ sur du papier log-log.



3. Représenter $f(x) = 3x^2$ sur du papier log-log.



4. On a mesuré la vitesse angulaire N (en tours par minute) d'une roue d'entraînement à différents instants t (en minutes) après avoir coupé le courant.

Temps (min)	Vitesse angulaire, N (r/min)
0,5	22,75
1,0	12,80
1,5	7,24
2,0	4,10
2,5	2,32
3,0	1,31

a) Quel type de correspondance relie les deux variables ?
b) Déterminer la règle de correspondance entre les deux variables.

5. La pression atmosphérique (p, en kilopascals) dépend de l'altitude (h, en kilomètres) au-dessus du niveau de la mer. On a pris les mesures regroupées dans le tableau suivant.

Altitude (km)	Pression (kPa)
0,5	95,15
1,0	89,36
1,5	83,93
2,0	78,82
2,5	74,02
3,0	69,52
3,5	65,29
4,0	61,32

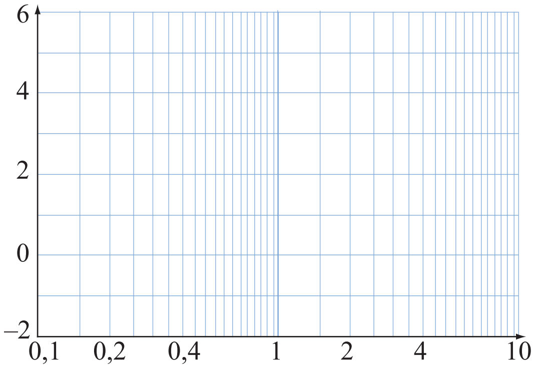
a) Quel type de correspondance relie les deux variables ?
b) Déterminer la règle de correspondance entre les deux variables.

6. On a obtenu les données suivantes en laboratoire.

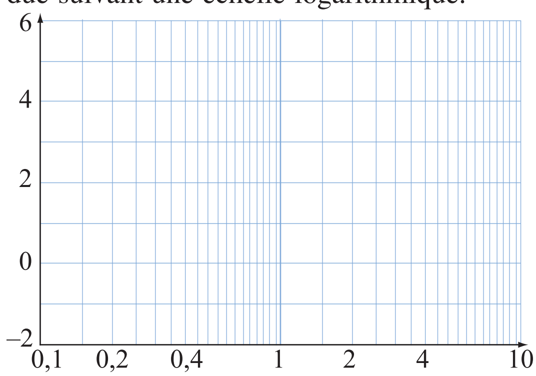
E	I
2,00	37,73
10,00	97,73
20,00	148,95
30,00	190,59
40,00	227,02
50,00	260,01
60,00	290,49

a) Quel type de correspondance relie les deux variables ?
b) Déterminer la règle de correspondance entre les deux variables.

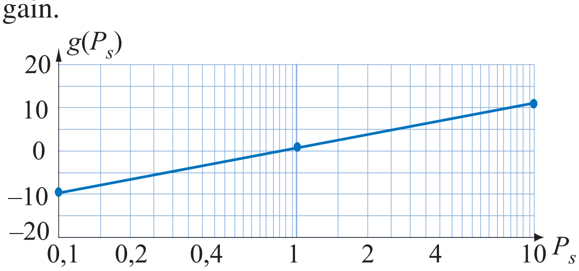
7. Représenter la fonction $f(x) = \ln x$ dans le système suivant dont l'axe horizontal est gradué suivant une échelle logarithmique.



8. Représenter la fonction $f(x) = 2 \log 2x$ dans le système suivant dont l'axe horizontal est gradué suivant une échelle logarithmique.



9. Le gain d'un composant électrique est illustré par le graphique suivant. Décrire algébriquement ce gain.



10. On a obtenu les données suivantes en laboratoire.

c_x	E
20	11,49
50	14,24
90	16,00
130	17,10
150	17,53
210	18,54

Déterminer, à l'aide d'une droite de régression le modèle qui décrit le mieux la relation entre les deux variables.

11. Un réservoir cylindrique de 12 m de hauteur et de 2,5 m de diamètre contient un liquide. On laisse s'écouler le liquide par une valve située à la base du cylindre. On note la vitesse d'écoulement du liquide et la hauteur de celui-ci dans le réservoir. Les données sont rassemblées dans le tableau suivant.

h (m)	v (m^3/min)
1	0,50
2	0,71
3	0,87
4	1,00
5	1,12
6	1,22
7	1,32

Déterminer à l'aide d'une droite de régression le modèle qui décrit le mieux la relation entre les deux variables et en déduire par extrapolation la vitesse d'écoulement du liquide quand le niveau est de 10 m.

12. On a mesuré le courant dans un circuit comprenant une source de tension constante et une résistance variable en faisant varier la résistance. On a obtenu les valeurs inscrites dans le tableau suivant :

Résistance (Ω)	Courant (A)
1,2	6,1
1,7	4,3
2,2	3,3
2,7	2,7
3,2	2,3
3,7	2,0

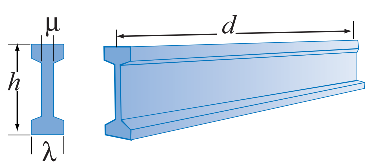
a) Quelle est la variable indépendante dans ce problème ?
b) Représenter graphiquement les données obtenues expérimentalement. Quel modèle mathématique la représentation graphique suggère-t-elle d'employer pour décrire la relation entre les deux variables ?
c) Décrire mathématiquement la relation en procédant par régression.

13. On a mesuré le volume occupé par 32 g d'ammoniac à 0 °C en faisant varier la pression exercée sur e gaz. On a obtenu les valeurs regroupées dans le tableau suivant.

Pression (kPa)	Volume (L)
10	420,0
20	210,0
40	105,0
60	70,0
80	52,5
100	42,0

a) Quelle est la variable indépendante dans ce problème ?
b) Représenter graphiquement les données obtenues expérimentalement. Quel modèle mathématique la représentation graphique suggère-t-elle d'employer pour décrire la relation entre les deux variables ?
c) Décrire mathématiquement la correspondance en procédant par régression.

14. On a fabriqué des poutres en I avec un nouveau matériau. La largeur μ de la bande centrale est le tiers de la largeur total λ. On veut déterminer la charge que peuvent supporter des poutres de même longueur et de même largeur mais de différentes épaisseur h, sans déformation.

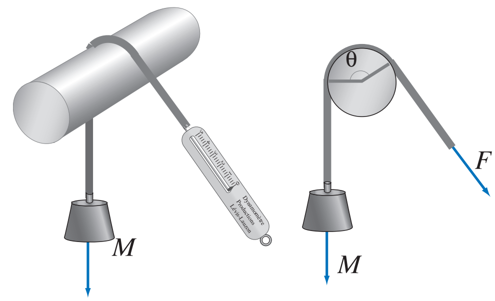


Le tableau suivant donne les mesures prises; la charge C est en kilogrammes (kg) et l'épaisseur h de la poutre en centimètres (cm).

Épaisseur (cm)	Charge (kg)
4	2 190
6	4 930
8	8 770
10	13 700
11	16 580
12	19 730
13	23 150

a) Montrer qu'il existe un lien de puissance entre la charge et le carré de l'épaisseur d'une poutre.
b) Utiliser le modèle pour déterminer la charge que peut supporter une poutre de 9 cm d'épaisseur.

15. On a tenté d'établir la relation entre la force nécessaire pour équilibrer une masse M en utilisant une corde enroulée sur une poutre ronde, et l'angle d'enroulement θ de la corde.

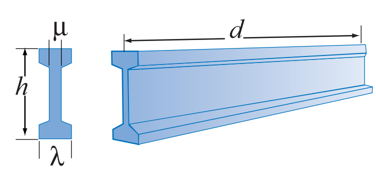


À l'aide d'un dynamomètre, on a mesuré la force minimale nécessaire pour équilibrer la masse M, qui exerce une force de 500 N selon l'angle d'enroulement de la corde sur la poutre. Les valeurs obtenues sont rassemblées dans le tableau suivant.

Angle (nombre de tours)	Force équilibrante
0,4	390
0,8	304
1,0	269
1,5	197
2,1	136
2,7	94
3,2	70
4,1	41
5,6	19

a) Montrer qu'il existe un lien exponentiel entre les deux variables.
b) Quelle force permettra d'équilibrer 800 N en faisant deux tours complets avec la corde?
c) Combien de tours faut-il faire pour que 50 N équilibrent 400 N ?

16. On a fabriqué des poutres en I ont été fabriquées avec un nouveau matériau. La largeur μ de la bande centrale est le tiers de la largeur totale λ. On veut déterminer la charge que peuvent supporter des poutres de même épaisseur et de même largeur, mais de différentes longueur d, sans déformation.

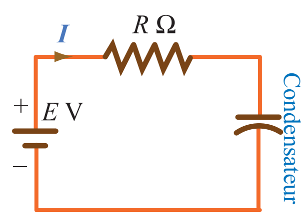


Le tableau suivant donne les mesures prises; la charge C est en kilogrammes (kg) et l'épaisseur h de la poutre en centimètres (cm).

Longueur (m)	Charge (kg)
2	12 490
3	8 340
4	6 250
5	5 000
6	4 170
7	3 570
8	3 130
9	2 780
10	2 500

Montrer qu'il existe un lien inversement proportionnel entre la charge et la longueur d d'une poutre.

17. On ferme un circuit comportant un condensateur monté en série avec une résistance et on enregistre le courant après la fermeture du circuit.

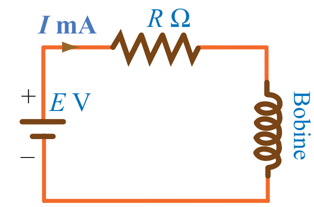


Les mesures obtenues sont données dans le tableau suivant.

Temps (s)	Courant (mA)
0,05	32,75
0,10	26,81
0,15	21,95
0,20	17,97
0,25	14,72
0,30	12,05
0,35	9,86
0,40	8,08
0,45	6,61
0,50	5,41

a) Représenter graphiquement les données sur un papier bilinéaire. Cette représentation graphique permet-elle de faire l'hypothèse d'un lien affine ? Justifier votre réponse.
b) Représenter graphiquement les données sur un papier semi-log. Cette représentation graphique permet-elle de faire l'hypothèse d'un lien affine ? Justifier votre réponse.
c) Déterminer le modèle décrivant le lien entre les variables su problème.

18. On ferme un circuit comportant une bobine montée en série avec une résistance et on enregistre la tension aux bornes de la bobine après la fermeture du circuit.



Les mesures obtenues sont données dans le tableau suivant.

t (s)	v (V)
0,05	86,07
0,10	74,08
0,15	63,76
0,20	54,88
0,25	47,24
0,30	40,66
0,35	34,99
0,40	30,12
0,45	25,92
0,50	22,31

a) Représenter graphiquement les données sur un papier bilinéaire. Cette représentation graphique permet-elle de faire l'hypothèse d'un lien affine ? Justifier votre réponse.
b) Représenter graphiquement les données sur un papier semi-log. Cette représentation graphique permet-elle de faire l'hypothèse d'un lien affine ? Justifier votre réponse.
c) Déterminer le modèle décrivant le lien entre les variables su problème.

19. Le tableau suivant donne le gain de puissance S d'un amplificateur en fonction de la puissance de sortie. Déterminer le modèle décrivant cette relation.

P_s (W)	S (dB)
10,00	18,00
20,00	21,01
30,00	22,77
40,00	24,02
50,00	24,99
60,00	25,78
70,00	26,45
80,00	27,03
90,00	27,54
100,00	28,00