TP

Effectuer une régression linéaire

I Méthode

1 – Côté théorie : quelle est la loi que l'on veut vérifier ?

▶ étape 1 On met la loi à vérifier sous la forme : $\boxed{y = a_{\text{théo}}x + b_{\text{théo}}.}$

> **Exemple :** On souhaite vérifier que la fréquence fondamentale de vibration d'une corde est donnée par $f_1 = c/(2L)$. On réalise plusieurs mesures de cette fréquence pour plusieurs longueurs L de corde différentes. On pose alors $y = f_1$ (que l'on va mesurer), $x = 1/L$ (qu'on mesure aussi).
>
> La loi $f_1 = c/(2L)$ s'écrit alors $y = (c/2) \times x$, ce qui est bien de la forme $y = a_{\text{théo}}x + b_{\text{théo}}$.
>
> On identifie alors $a_{\text{théo}}$ et $b_{\text{théo}}$: ici on a $\boxed{a_{\text{théo}} = \dfrac{c}{2}}$ et $\boxed{b_{\text{théo}} = 0.}$

▶ étape 2 On fait l'application numérique pour $a_{\text{théo}}$ et $b_{\text{théo}}$, avec incertitudes si besoin.

> **Suite de l'exemple :**
>
> ⋆ Ici on a $c = \sqrt{\gamma RT/M} = 345\,\text{m/s}$ pour $T = 25\,°\text{C}$. Donc $a_{\text{théo}} = c/2 = 172{,}5\,\text{m/s}$.
>
> On prend une incertitude de $1\,°\text{C}$ pour T, on utilise la formule de propagation des incertitudes pour obtenir $\Delta a_{\text{théo}} = 1\,\text{m/s}$.
>
> On a donc finalement $\boxed{a_{\text{théo}} = (345 \pm 1)\,\text{m/s}.}$
>
> ⋆ On a de plus $\boxed{b_{\text{théo}} = 0\,\text{Hz},}$ sans incertitude.

2 – Côté expériences : comment vérifier la loi ?

▶ étape 3 On réalise des mesures expérimentales de y (ici la fréquence f_1) pour plusieurs valeurs de x (donc ici de L).

On entre ces valeurs sous le logiciel Regressi (par exemple) dans un tableau à deux entrées.

Remarque : On peut entrer également l'incertitude sur chaque mesure expérimentale (si on la connaît), en cliquant sur "incertitude" en haut. Sous Regressi, dans les options de modélisation, sélectionner "prendre en compte les incertitudes". Il faut alors entrer les *incertitudes types*, et Regressi retournera l'incertitude élargie sur a et b.

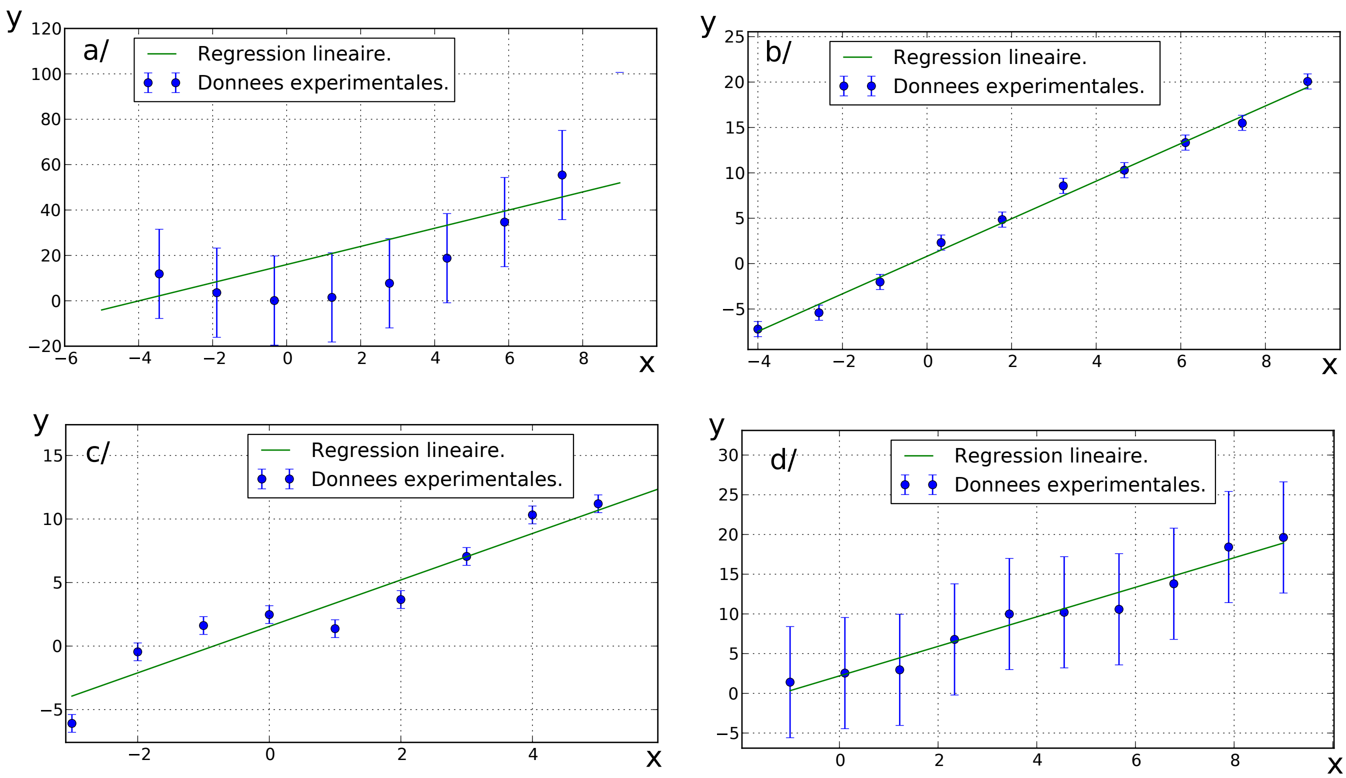
→ On demande ensuite au logiciel de tracer la variable y (donc ici les valeurs de f_1) en fonction de la valeur de x (donc ici les valeurs de $1/L$).

▶étape 4 On confirme d'abord *visuellement* que les points forment bien approximativement une droite.

Critères de validation d'un modèle linéaire

Pour cela, il faut 1/que les points ne suivent pas une tendance nettement non linéaire (une courbure, une oscillation, etc.), et 2/ que la droite de régression passe à l'intérieur des barres d'erreur.

Examinons quelques exemples :



- Cas **a** : la droite de régression passe par les barres d'erreur, mais cependant on note une nette courbure : le modèle linéaire n'est donc certainement pas bon.
- Cas **b** : pas de tendance particulière, et la droite de régression passe par les barres d'erreur : on valide le modèle linéaire.
- Cas **c** : la droite de régression ne passe pas par les barres d'erreur : on rejette le modèle. (il faudra vérifier qu'on n'a pas pris des incertitudes trop petites)
- Cas **d** : pas de tendance nette et la droite passe par les barres d'erreur : on valide le modèle. Remarquons toutefois que l'incertitude est grande et que d'autres modèles, non linéaires, pourraient aussi convenir : la conclusion n'est donc pas très forte.

→ Si on valide le modèle linéaire, on passe à l'étape suivante.

→ Si ce n'est pas le cas on arrête ici et on dit que le modèle linéaire ou affine n'est pas en accord avec les mesures.

▶étape 5 On demande au logiciel d'effectuer une modélisation, affine dans notre cas ($y = ax + b$).

Le logiciel trouve alors les paramètres a et b qui sont tels que l'écart entre la droite $y = ax + b$ et les points expérimentaux soit minimal (cet écart étant défini comme $C(a,b) = \sum_i [y_i - (ax_i + b)]^2$, avec (x_i,y_i) les points mesurés).

Régressi retourne les valeurs de a et de b, assorties d'une incertitude élargie (même si on n'a pas entré d'incertitude sur les valeurs expérimentales).

Il s'agit donc des **valeurs expérimentales** de a et b, que l'on va noter $a_{\text{exp}} \pm \Delta a_{\text{exp}}$ et $b_{\text{exp}} \pm \Delta b_{\text{exp}}$.

Suite de l'exemple : On écrira donc $\boxed{a_{\text{exp}} = (.. \pm ..)\,\text{m/s}}$ et $\boxed{b_{\text{exp}} = (.. \pm ..)\,\text{Hz}}$ avec les valeurs obtenues.

3 – Comparaison entre théorie et expérience

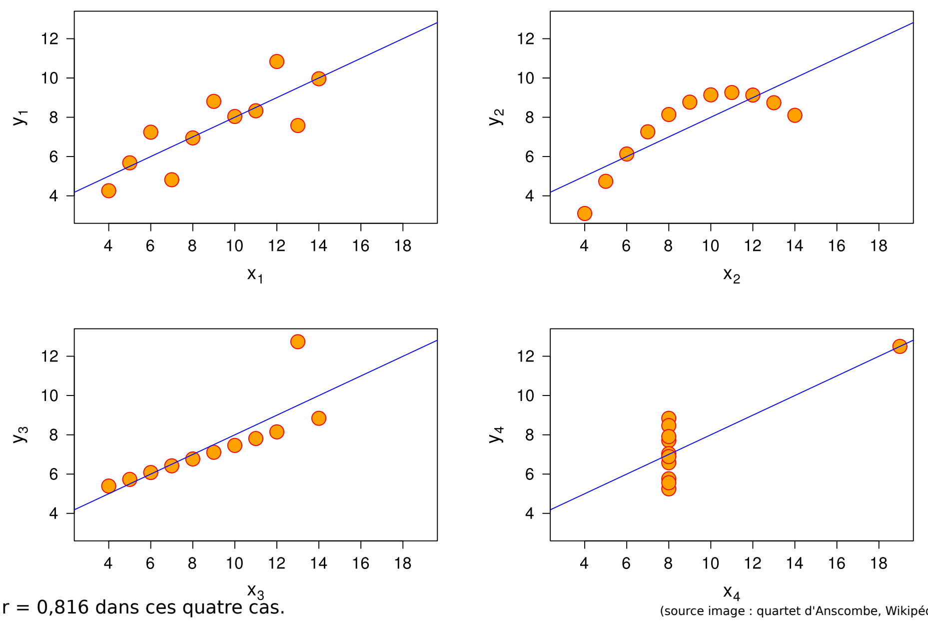
►$_{\text{étape 6}}$ On conclut en comparant théorie et expérience, donc en regardant si les intervalles $a_{\text{exp}} \pm \Delta a_{\text{exp}}$ et $a_{\text{théo}} \pm \Delta a_{\text{théo}}$ ont des valeurs en commun, et de même pour $b_{\text{théo}} \pm \Delta b_{\text{théo}}$ et $b_{\text{exp}} \pm \Delta b_{\text{exp}}$.

Remarque : Si $b_{\text{théo}} = 0$ et que cette valeur nulle est comprise dans l'intervalle $b_{\text{exp}} \pm b_{\text{exp}}$, alors il est possible de faire une régression avec une loi linéaire (du type $y = ax$). Il faut alors vérifier visuellement qu'il n'y a pas d'ordonnée à l'origine. Puis voir si la nouvelle valeur de a est compatible avec la valeur théorique.

II Remarque sur le coefficient de corrélation r^2

La calculatrice ou les tableurs donnent parfois un coefficient r^2, appelé coefficient de corrélation linéaire. Ce coefficient permet de savoir si les données x et y dépendent linéairement l'une de l'autre : il est proche de ± 1 quand c'est le cas.
Cependant, il n'est pas adapté à la vérification d'une loi physique, pour essentiellement deux raisons :

- Il ne prend pas en compte les incertitudes. Il peut donc être très proche de 1, tout en ayant une droite de régression qui ne passe pas du tout par les barres d'erreur. Il ne remplace donc pas le critère de validation numéro 2/ ci-dessus.
- Il peut être proche de 1 dans des cas où la dépendance n'est pas linéaire. Sur les quatre exemples ci-dessous, il vaut 0,816, alors qu'il y a clairement des tendances non linéaires. Il ne remplace donc pas le critère de validation 1/ ci-dessus.



r = 0,816 dans ces quatre cas.

(source image : quartet d'Anscombe, Wikipédia)

III Exemples

1 - On souhaite vérifier que la vitesse du son dans l'air est bien proportionnelle à la racine carrée de la température comme prédit par la théorie : $c_s = \alpha\sqrt{T}$, avec α une constante. On effectue des mesures de la vitesse du son c_s pour différentes températures T :

c_s (m/s)	310	325	331	345	355
T (K)	240	260	280	300	320

Que faut-il poser pour mettre la loi à vérifier sous la forme $y = a_{\text{théo}}x + b_{\text{théo}}$?

À l'aide de votre calculatrice, vérifier si ces données expérimentales sont en accord avec la loi théorique $c_s = 20.05\sqrt{T}$ (T en K, c_s en m/s).

2 - On souhaite vérifier la loi de Descartes de la réfraction : $n_1 \sin i_1 = n_2 \sin i_2$.

Expérimentalement, pour plusieurs angles d'incidence i_1, on mesure l'angle réfracté i_2.

Que faut-il poser pour mettre la loi à vérifier sous la forme $y = a_{\text{théo}}x + b_{\text{théo}}$?

Remarque : On essaie toujours de se ramener à une loi linéaire ou affine. La raison est que le cerveau humain est doué pour voir si des points sont alignés, mais pas du tout pour estimer s'ils suivent une loi en racine carrée, en $1/x$, en exponentiel ou autre. Toutefois, il se peut dans certains cas qu'on ne puisse pas se ramener à une loi affine ou linéaire. Dans ce cas on utilise les autres modèles disponibles dans le logiciel, tout en regardant à l'œil si le modèle est vraisemblable.