FICHE MÉTHODE N°04
Regression linéaire

Objectif

On dispose de données expérimentales (x_i, y_i) et des incertitudes-types correspondantes u(xi) et u(yi) (voir fiche incertitudes) qui ne sont pas forcément les mêmes pour chaque point. On cherche à ajuster à ces données un modèle affine d'équation :

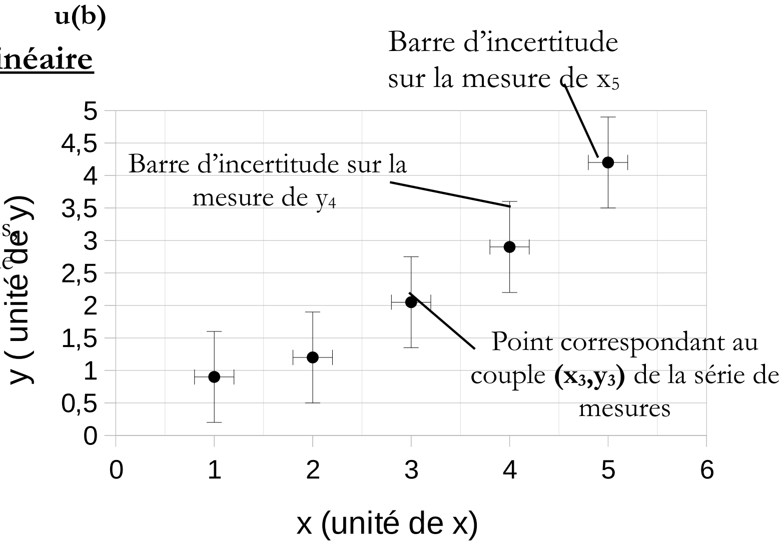
$$y = a\,x + b$$

a et b sont les paramètres du modèle que l'on souhaite déterminer avec leurs incertitudes correspondantes u(a) et u(b)

I Interprétation graphique de la régression linéaire

Nuage de points

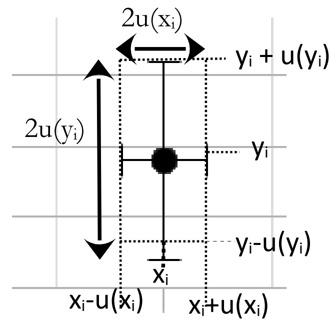
Si on place dans un plan Oxy les points correspondants aux couples formés par les deux série de données fournies, on obtient ce qu'on appelle un nuage de points (graphique ci-contre)



Barres d'incertitude

On reporte les incertitudes-type sur les valeurs représentées. Pour ce faire, on peut tracer des **"croix"** dont la taille représente les incertitudes $u(x_i)$ et $u(y_i)$.

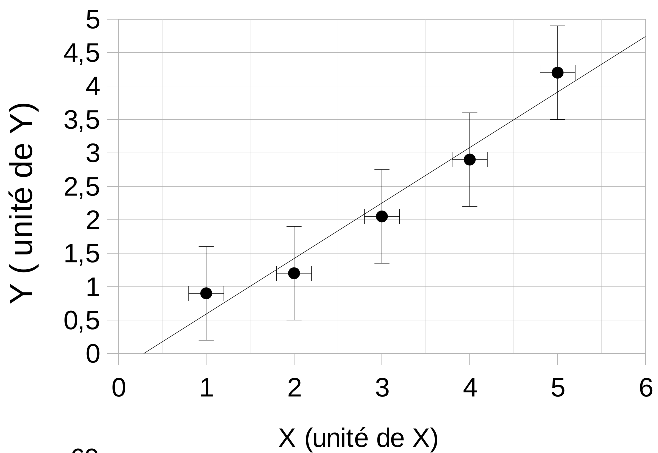
Interprétation des barres d'incertitude



Si on suppose que les erreurs aléatoires suivent une statistique gaussienne alors :
- La valeur vraie $x_{i\,vraie}$ associée à la mesure x_i a 68% de chance de se trouver entre $x_i - u(x_i)$ et $x_i + u(x_i)$
- La valeur vraie $y_{i\,vraie}$ associée à la mesure y_i a 68% de chance de se trouver entre $y_i - u(y_i)$ et $y_i + u(y_i)$

Remarque : Souvent en physique l'incertitude sur x est négligeable devant celle sur y. Les barres d'incertitudes sont alors purement verticales.

Effectuer une régression linéaire entre les deux série consiste à trouver la droite qui passe au plus près de l'ensemble de ces points.



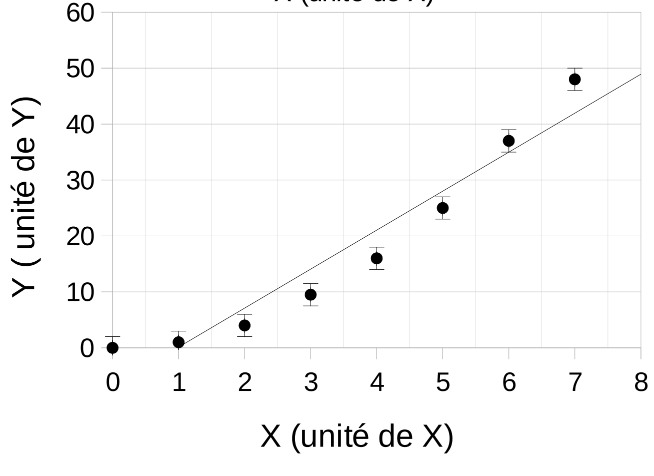
C'est le cas de la droite noir sur le graphique ci-contre.

Un logiciel comme Régressi calcule alors :
- a (la pente de la droite)
- b (l'ordonnée à l'origine de la droite)
- u(a) et u(b) (Si on indique les barres d'incertitude) *

Remarque sur la régression linéaire

La régression permet de tracer la droite la plus proche de tous les points même si le modèle de relation affine entre y et x est faux !

C'est le cas sur la courbe ci-contre. Une relation de puissance (d'équation $y = a\,x^2$) serait plus adaptée.



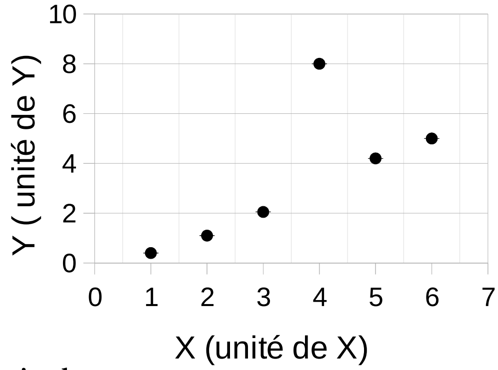
** On verra dans l'année en TP comment déterminer u(a) et u(b)*

II Vérifier la qualité d'une régression linéaire en TP

Étape 1 : On représente les données expérimentales sous la forme d'un nuage de point.

Première analyse : Les points semblent-ils alignés ?
Si la réponse est non, deux possibilités :

- **cas 1 :** Tous les points semblent globalement alignés **sauf un seul.**
(voir graphique ci-contre)
Dans ce cas on vérifie qu'on n'a pas fait une erreur en rentrant les données dans le tableau.



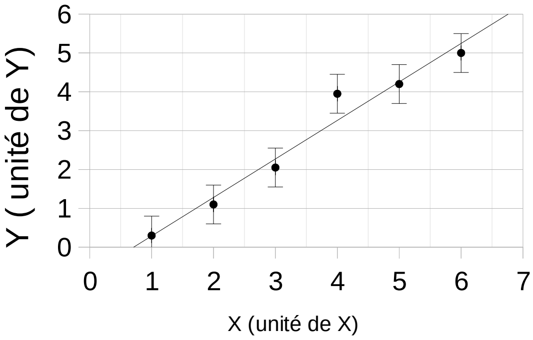
- cas 2 : La relation qui relie y et x ne semble pas être linéaire.
(voir dernier graphique de la page précédente)
Dans ce cas le modèle choisi n'est pas approprié, il faut linéariser la relation entre y et x * .

Étape 2 : On ajoute la courbe d'ajustement sur le graphique et **les barres d'incertitudes.**

Deuxième analyse : l'ajustement est il cohérent avec ce que j'attends ?

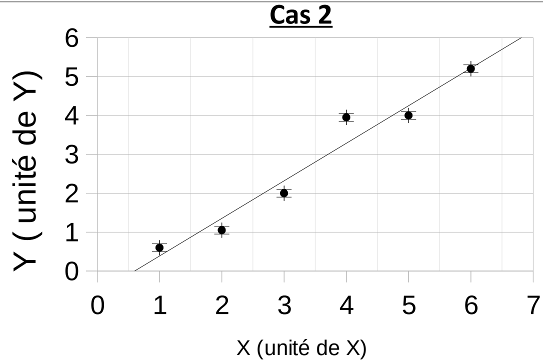
Si la droite issue de l'ajustement coupe la majorité des barres d'incertitude, on peut en déduire que la modélisation est cohérente avec les données expérimentales (compte tenu des incertitudes de mesure)

Cas 1



Dans mon compte rendu de TP j'écris :
Comme la droite issue de la modélisation passe par la majorité des barres d'incertitude, une relation affine entre y et x est cohérente avec les données expérimentales.

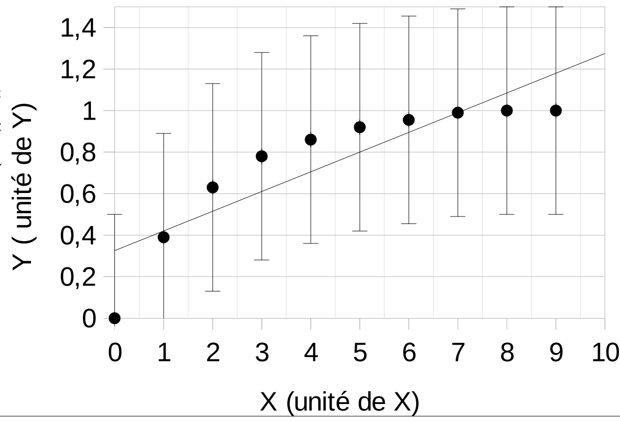
Cas 2



Dans mon compte rendu de TP j'écris :
Compte tenu des incertitudes, une relation affine entre y et x ne semble pas cohérente avec les données expérimentales.
Mais d'après l'allure du nuage de points, les incertitudes ont probablement été sous-estimées.

⚠ il faut avoir un regard critique sur le résultat ! ⚠
Par exemple, si l'incertitude type est largement surestimée, la droite d'ajustement peut passer par toutes les barres d'incertitude alors que les points expérimentaux ne sont pas alignés (d'où l'importance de la première analyse).

Sur le graphique ci-contre les points ne sont clairement pas alignés mais la droite d'ajustement coupe toutes les barres d'incertitude.



Cas particulier d'une relation de proportionnalité

Même si on soupçonne une relation de proportionnalité entre x et y, on commence toujours par réaliser un modélisation affine d'équation y = ax + b

Si 0 appartient à l'intervalle [b-2u(b) , b + 2u(b)] on peut en déduire que la relation de proportionnalité peut être acceptée (après les étapes d'analyse). On peut ensuite réaliser la modélisation d'équation y =ax.

* *On verra comment faire plusieurs fois dans l'année en TP*