
ESSENTIEL RÉGRESSION LINÉAIRE

La représentation graphique d'une acquisition expérimentale est un nuage de points, issu d'un tableau de valeurs. Le plus souvent on contrôle expérimentalement un paramètre, qui sera l'abscisse du nuage de points, et on mesure une autre grandeur qui varie au cours de l'expérience, qui permettra d'obtenir l'ordonnée du graphique.

La première question qui se pose est : « existe-t-il une relation linéaire entre le paramètre contrôlé et le paramètre mesuré » ? Autrement dit : si on dispose d'une valeur pour la grandeur en ordonnée, peut-on prévoir la valeur de l'abscisse correspondante, ou le nuage de points semble-t-il totalement aléatoire ?

L'outil pour quantifier la dépendance entre abscisse et ordonnée est le coefficient de corrélation. Lorsque $R^2 = 1$, les deux grandeurs sont parfaitement corrélées : il existe un lien fort entre leurs valeurs.

Attention ! Corrélation ne signifie pas causalité ! D'ailleurs on retrouve la même valeur de coefficient de corrélation entre permutant abscisse et ordonnée : quelle serait alors la cause, et quelle serait la conséquence ?

Par exemple, la valeur du coefficient de corrélation permet de tester des hypothèses en cinétique chimique, lorsqu'on souhaite déterminer l'ordre d'une réaction mettant en jeu un unique réactif A. Les modèles cinétiques (loi de vitesse) nous permettent de montrer que :

- dans le cas d'une réaction d'ordre 0, le nuage de points représentant la concentration $[A]$ en fonction du temps doit présenter un coefficient de corrélation égal à 1 ;
- dans le cas d'une réaction d'ordre 1, le nuage de points représentant $\ln[A]$ en fonction du temps doit présenter un coefficient de corrélation égal à 1 ;
- dans le cas d'une réaction d'ordre 2, le nuage de points représentant $\frac{1}{[A]}$ en fonction du temps doit présenter un coefficient de corrélation égal à 1.

En pratique il est impossible d'obtenir exactement $R^2 = 1$ du fait des incertitudes expérimentales. D'ailleurs en toute rigueur il ne s'agit pas de nuage de points, mais d'un nuage de rectangles du fait des incertitudes sur les valeurs des grandeurs mesurées. Ce point ne sera pas développé ici.

Une fois la corrélation montrée, on souhaite généralement disposer d'une équation de droite. Le logiciel utilisé détermine par calcul la droite qui décrit le mieux le nuage de points. L'équation de droite permet ensuite d'exploiter la courbe pour n'importe quelle valeur, et pas seulement pour les valeurs correspondant à des points expérimentaux. Autrement dit, on extrapole les données expérimentales dont on dispose, afin de ne pas avoir à acquérir un trop grand nombre de points. On transforme un jeu discret de données en un ensemble continu, ce qui ouvre la voie à la modélisation du système : modèle empirique qui décrit directement les données, puis éventuellement théorie plus générale.