



Mesures et incertitudes : fiche pour le professeur

Modélisations : nuages de points et graphiques

Comment exploiter un nuage de points ? Une modélisation est-elle toujours pertinente ?

Préambule : que font les tableurs dotés d'une fonction « modélisation » ?

La modélisation d'un nuage de points par une fonction affine est fréquemment mise en œuvre dans les activités expérimentales. Celle-ci peut avoir divers objectifs : tester une loi, faire une prévision, mesurer un paramètre (coefficients directeur ou ordonnée à l'origine) d'une droite supposée, etc. Quel que soit son objectif, si l'on utilise les logiciels usuels, la modélisation d'un nuage de points par une droite repose toujours sur la méthode des moindres carrés. Or non seulement celle-ci est rarement explicitée mais, surtout, elle repose sur des hypothèses à vérifier pour la mettre en œuvre qui, nous le verrons, ne sont pas toujours satisfaites.

La méthode des moindres carrés ordinaires

Que nous utilisions la fonction « courbe de tendance » de Excel ou LibreOffice, la fonction « modélisation » des tableurs à usage pédagogique comme Rgressi ou Latis Pro, la modélisation d'un nuage de points par une fonction affine (ou linéaire) repose, sans que cela ne soit forcément documenté, sur la méthode dite « des moindres carrés ordinaires ».



Focus sur la méthode des moindres carrés ordinaires ou « MCO »

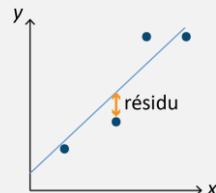
Si l'on dispose d'un échantillon de N valeurs d'une grandeur x et d'un autre échantillon de N valeurs d'une grandeur y , la méthode des moindres carrés ordinaires (MCO) est une des plus simples permettant d'obtenir l'équation d'une droite censée modéliser la relation entre y et x :

$$y = ax + b$$

Le principe de la méthode des MCO

À chaque couple de points $(x_i; y_i)$ on associe un résidu ε_i défini par :

$$\varepsilon_i = y_i - (ax_i + b)$$



Le résidu quantifie l'écart en ordonnée entre le point de l'échantillon et celui ayant la même abscisse mais appartenant à la droite.

La méthode consiste à calculer les valeurs de a (coefficients directeur) et de b (ordonnée à l'origine) qui minimisent la somme des carrés des résidus. On montre que :

$$\sum_{i=1}^N \varepsilon_i^2 \text{ est minimale} \Leftrightarrow \begin{cases} a = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ b = \bar{y} - a\bar{x} \end{cases}$$

C'est ainsi que le coefficient directeur et l'ordonnée à l'origine d'une droite de régression sont calculées par tous les tableurs usuels.

Au plan mathématique, la méthode des MCO peut être appliquée à n'importe quel nuage de points, qu'il soit raisonnable ou non de le modéliser par une droite ; mais cela ne signifie pas que son résultat soit *physiquement* pertinent. Pour que cette méthode donne une modélisation pertinente, trois hypothèses doivent être satisfaites :

- ▶ **Hypothèse 1** : l'incertitude de la grandeur portée en abscisse est nulle (ou de poids négligeable devant celle de la grandeur portée en ordonnée).
- ▶ **Hypothèse 2** : l'incertitude de la grandeur portée en ordonnée est constante.
- ▶ **Hypothèse 3** : les incertitudes de la grandeur portée en ordonnée sont indépendantes entre elles.

Nous le verrons, certains de nos usages courants consistent à exploiter cette méthode en dehors de son champ de validité. Cette fiche a pour objectif d'identifier les cas où on peut le faire et de proposer des solutions alternatives dans les autres cas.



Modéliser pour faire une mesure par prévision (étalonnage)

Le principe

La modélisation nous renvoie deux paramètres (coefficients directeur et ordonné à l'origine) souvent reliés à une grandeur physique. Il est usuel d'exploiter leurs valeurs pour déterminer une propriété d'un objet inconnu.

Exemple : le dosage par étalonnage spectrophotométrique

On suppose admis que les mesures sont effectuées dans le domaine de validité de la loi de Beer-Lambert : le but n'est alors pas de valider cette loi mais de l'exploiter pour faire une mesure.

On obtient les valeurs suivantes à partir de 5 solutions « étalon » (nous revenons plus loin sur le sens de ces guillemets) :

Concentration (mol/L)	0,01	0,005	0,0025	0,00125	0,000625
Absorbance	3,10	2,03	0,89	0,67	0,31

Une solution inconnue est proposée et la mesure de son absorbance donne : $A = 2,5$.

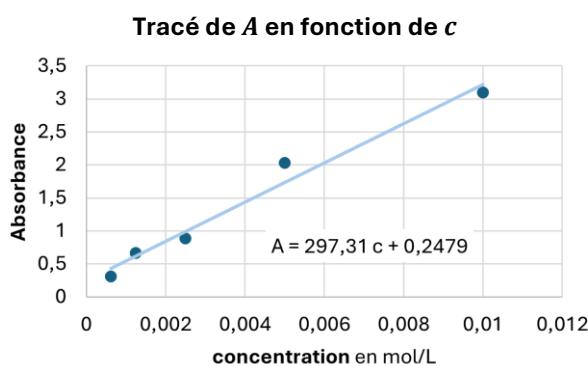
La modélisation permet d'obtenir l'équation d'une droite et d'en déduire la valeur de sa concentration.

Comment faire avec des élèves dans le respect des hypothèses de la méthode des MCO ?

Quelle grandeur porter en abscisse ? en ordonnée ? (ou comment respecter au mieux l'hypothèse 1 ?)

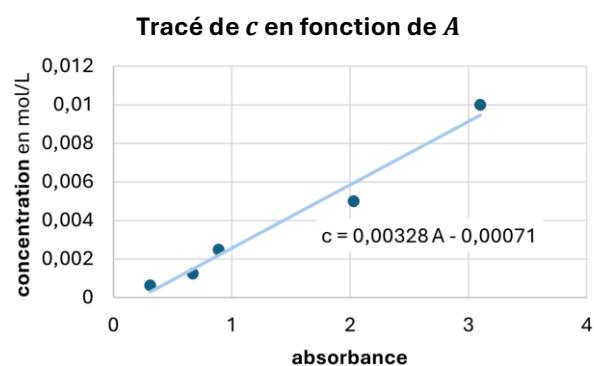
Dans l'exemple précédent, le « réflexe » que nous avons souvent est de porter la concentration en abscisse et l'absorbance en ordonnée : c'est en effet le graphique qui illustre le mieux (pour un élève) la loi de Beer-Lambert. Dans l'objectif de faire une mesure, ce n'est pourtant pas un bon choix. En effet, si les solutions ont été réalisées par des élèves, les incertitudes de leurs concentrations ont certainement un poids beaucoup plus élevé que celles des valeurs d'absorbance. Placer A en abscisse et c en ordonnée permet donc de mieux respecter l'hypothèse 1 de la méthode des moindres carrés.

Il est intéressant de constater que le résultat final est sensible à ce choix. Les deux graphiques ci-dessous ont été tracés à partir des mêmes valeurs de A et c (données dans le tableau précédents).



La concentration de la solution inconnue vaut :

$$c_{\text{inc}} = \frac{2,5 - 0,2479}{297,31} = 7,57 \times 10^{-3} \text{ mol} \cdot \text{L}^{-1}$$



La concentration de la solution inconnue vaut :

$$c_{\text{inc}} = 0,00328 \times 2,5 - 0,00071 = 7,49 \times 10^{-3} \text{ mol} \cdot \text{L}^{-1}$$

On voit que les deux concentrations, bien que déterminées en exploitant le même tableau de valeurs, sont sensiblement différentes. La seconde valeur est plus juste que la première car le graphique tracé respecte mieux l'hypothèse 1 de la méthode des MCO.

Les hypothèses 1 et 2 sont-elles satisfaites ?

Dans l'exemple précédent, la réponse dépend de la méthode employée pour préparer les solutions. Si l'élève les prépare par dilutions successives, alors non : l'incertitude de la concentration augmente après chaque dilution et chaque concentration a une incertitude qui dépend de la solution mère qui a été diluée. Les incertitudes des concentrations ne sont donc ni égales ni indépendantes entre elles. Si l'on procède différemment il peut en être autrement.



Alors que faire ?

Au lycée la régression par la méthode des MCO est la seule accessible ; or les mesures par étalonnage sont au programme et sont un support intéressant dont il ne faut pas se priver, quitte à faire quelques entorses aux hypothèses de la méthode. Le professeur devra cependant veiller à :

- ▶ porter en abscisse la grandeur qui a l'incertitude la plus faible ;
- ▶ *si possible*, fournir un protocole qui assure que les incertitudes des valeurs portées en ordonnées soient voisines et indépendantes entre elles.

Modéliser pour déterminer un paramètre inconnu

■ Le principe

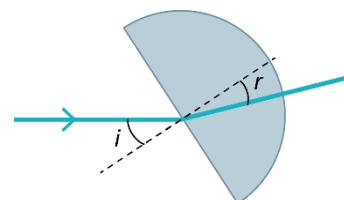
Dans de nombreux cas, la grandeur à mesurer est un paramètre de la supposée droite de régression : soit son ordonnée à l'origine, soit (plus fréquemment) son coefficient directeur.

Exemple classique n°1 : mesure d'une accélération en mécanique

Le mouvement d'un objet soumis à une force constante (par exemple une chute) est étudié par pointage vidéo, les valeurs de sa vitesse sont calculées. Le but est de déterminer la valeur de son accélération.

Exemple classique n°2 : mesure d'un indice de réfraction en optique

À l'aide d'un dispositif type « discoptic » les élèves mesurent, pour plusieurs angles d'incidence, les angles de réfraction correspondants. Le but est de déterminer l'indice de réfraction du matériau réfractant en exploitant la relation de Snell-Descartes, qui donne dans ce cas : $\sin(i) = n \sin(r)$.



■ Comment exploiter de telles séries de mesures avec les élèves

Si l'on veut exploiter un nuage de points avec la méthode des MCO, il faut veiller à ce que les hypothèses de la méthode soient satisfaites au mieux.

Un (bon) exemple : retour sur la mesure d'une accélération (exemple 1 précédent)

Une solution pour déterminer l'accélération peut consister à tracer l'évolution de la vitesse en fonction du temps et à exploiter la méthode des MCO pour déterminer le coefficient directeur de la droite obtenue, qui est l'accélération recherchée (à condition d'avoir validé le fait que l'objet soit bien soumis à des forces constantes). C'est une méthode pertinente car les hypothèses de la méthode des MCO sont respectées :

- ▶ le temps (qui dépend de la fréquence de la caméra utilisée) a une incertitude très faible : le placer en abscisse est donc pertinent ;
- ▶ les valeurs de vitesse résultent du pointage, chaque point ayant été obtenu indépendamment des autres en employant la même méthode : les incertitudes des vitesses sont donc voisines et indépendantes entre elles.

Un contre-exemple : la mesure d'un indice de réfraction (retour sur l'exemple 2 précédent)

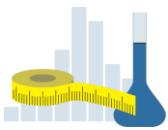
Une exploitation classique du dispositif « discoptic » consiste à représenter graphiquement $\sin(i)$ en fonction de $\sin(r)$, de modéliser comme une droite la courbe obtenue et de faire calculer son coefficient directeur au tableur utilisé. Mais, dans ce cas-là, on est en dehors du champ de validité de la méthode des MCO. En effet :

- ▶ l'incertitude de $\sin(r)$, grandeur portée en abscisse, est du même ordre que celle de $\sin(i)$;
- ▶ les incertitudes de $\sin(i)$ et $\sin(r)$ ne sont pas constantes car plus les angles augmentent plus le faisceau est élargi et atténué par le dispositif.

Exploiter la méthode des moindres carrés avec une telle série de mesures n'est donc pas recommandé.

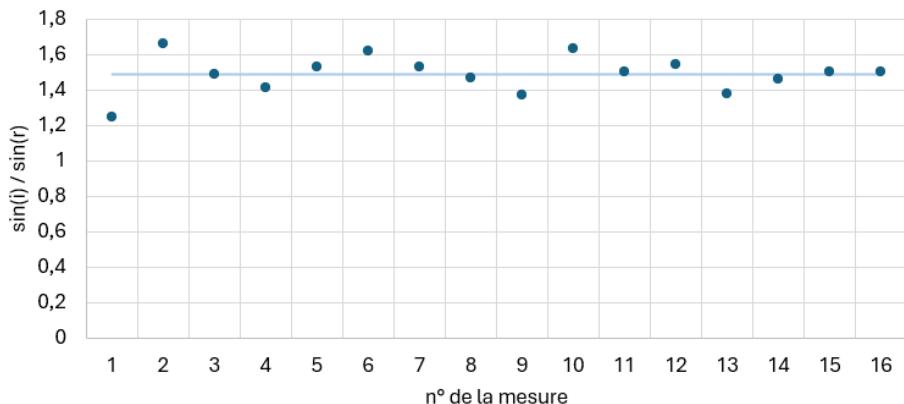
Alors comment faire sans utiliser la méthode des MCO ?

Une solution alternative au tracé de la droite et à la mesure de son coefficient directeur consiste à obtenir un échantillon de valeurs de la grandeur mesurée et à l'exploiter **avec des outils statistiques**.



Dans l'exemple 2 précédent (mesure de l'indice de réfraction), cela consiste à :

- ▶ calculer la valeur de $n = \sin(i)/\sin(r)$ correspondant à chaque couple $(i; r)$;
- ▶ calculer la valeur moyenne des indices obtenus ;
- ▶ si l'on veut faire tracer un graphique aux élèves : attribuer à chaque couple d'angles mesurés un numéro et représenter l'indice calculé en fonction de ce numéro :



les points bleu foncé représentent les mesures individuelles, la droite claire la valeur moyenne

Cette approche a plusieurs avantages :

- ▶ la constance de $\sin(i)/\sin(r)$ est visuellement illustrée (ce qui équivaut à constater que le tracé de $\sin(i)$ en fonction de $\sin(r)$ ressemble à un droite passant par l'origine) ;
- ▶ l'incertitude-type de l'indice de réfraction mesuré devient facile d'accès, par une méthode de type A : il suffit pour l'évaluer de calculer l'écart-type de l'échantillon des valeurs de n et de calculer s/\sqrt{N} .

Modéliser pour « tester une loi »

■ Que signifie « tester une loi » ?

Par « tester une loi » (expression présente dans les programmes officiels) nous entendons : comparer une série de mesures expérimentales à la prévision théorique issue d'une loi. Cela regroupe des activités diverses.

▶ Vérifier que des résultats expérimentaux sont compatibles avec ce qu'exprime une loi

L'exemple précédent de la mesure de l'indice de réfraction s'inscrit dans cette démarche : il ne s'agit pas de « valider » la loi de Snell-Descartes, celle-ci n'est pas en cause ; la question est plutôt : les résultats de mesures sont-ils suffisamment justes et fidèles pour être compatibles avec la loi de Snell-Descartes ?

▶ Tester si une loi est valide

Dans ce cas c'est la loi qui est mise à l'épreuve et les résultats de mesure qui font foi. Ce cas se présente pour les lois dont le champ de validité est restreint. C'est le cas, par exemple, de la relation que les élèves considèrent comme la définition du pH : $pH = -\log [H_3O^+]$. Il s'agit en réalité d'une approximation et on est fondé à tester si les solutions dont on dispose ont des concentrations qui appartiennent au **champ de validité** de cette approximation.

▶ Comparer des résultats de mesures à un modèle

C'est ce que nous faisons lorsque nous comparons une situation expérimentale réelle à un modèle dont les hypothèses *ne peuvent pas* être exactement satisfaites. Le cas d'école est la chute libre. On dispose de moyens fiables pour étudier expérimentalement la chute d'un objet, on a une confiance totale dans les relations théoriques, issues des lois de Newton, qui décrivent la chute libre et, enfin, on est totalement certain que le mouvement étudié *ne peut pas* être une chute libre, puisqu'il a lieu dans l'air.

La question qui se pose est, dès lors : avec quel degré d'approximation le modèle de la chute libre décrit-il la situation expérimentale ?

Nous revenons dans la suite de cette fiche sur chacune de ces trois manières de « tester une loi » et, nous le verrons, le tracé d'une droite et sa modélisation par la méthode des MCO ne sont pas toujours pertinents.



■ Comment vérifier que des résultats expérimentaux sont compatibles avec une loi ?

Les limites de l'approche « classique »

Reprenons l'exemple de la mesure d'un indice de réfraction. Si le but est de « tester la loi de Snell-Descartes » (ce qui ne peut qu'avoir lieu *avant de vouloir l'exploiter pour mesurer l'indice de réfraction*), une approche courante consiste à :

- ▶ représenter graphiquement $\sin(i)$ en fonction de $\sin(r)$;
- ▶ modéliser le nuage de points obtenu par une fonction affine ou linéaire ;
- ▶ constater que le coefficient de détermination est « proche de 1 ».

Nous l'avons vu, un tel procédé consiste à utiliser la méthode des MCO en-dehors de son champ de validité. De plus, cela pose le problème de la validation : puisque les points ne sont pas réellement alignés, quel critère utiliser pour affirmer que le nuage de points est « presque une droite » ou non ? La plupart des logiciels fournissent un coefficient de corrélation et/ou un coefficient de détermination (voire confondent les deux : c'est le cas de Latis Pro qui fournit un coefficient de détermination en l'appelant « corrélation »). Nous ne recommandons pas de l'exploiter, pour au moins deux raisons :

- les élèves n'ont pas accès à la signification de ce coefficient ;
- cela ne règle pas la question du critère de validation : que signifie « proche de 1 » ? Un élève qui obtient un coefficient de détermination de 0,98 a-t-il « bien » ou « mal » travaillé ?

Alors comment faire ?

Au lycée et *a fortiori* en classe de 2^{nde}, si l'on admet que la loi de Snell-Descartes est juste et que ce sont les résultats de mesures qui sont à valider, une approche qualitative peut suffire. Si l'on trace le quotient $\sin(i)/\sin(r)$ en fonction du numéro de la mesure (voir graphique de la page précédente), constater que le nuage de points obtenu fluctue peu autour de la valeur moyenne est d'ores et déjà un critère de compatibilité entre la loi admise et les résultats de mesure.

Et si l'on veut un argument quantitatif ?

Avertissement : ce qui suit décrit une démarche possible, pratiquée en laboratoire de métrologie, mais n'est pas préconisé devant des lycéens.

Une approche quantitative peut consister, pour chaque valeur d'indice, à :

- ▶ calculer l'écart $n_i - \bar{n}$ entre chaque indice mesuré et la moyenne de l'échantillon ;
- ▶ comparer cet écart à l'écart-type de l'échantillon (qui est l'incertitude-type de chaque valeur individuelle) et vérifier que les deux ont le même ordre de grandeur : si c'est le cas les résultats expérimentaux, compte-tenu de leurs incertitudes, sont compatibles avec la loi.

On peut par exemple faire calculer pour chaque valeur d'indice le quotient :

$$z = \frac{|n_i - \bar{n}|}{s}$$

s étant l'écart-type de l'échantillon et se fixer comme critère : une valeur est jugée aberrante si $z \geq 2$.

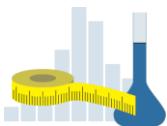
Avec les valeurs ayant servi à tracer le graphique de la page précédente cela donne :

mesure n°	i (°)	r (°)	$n = \frac{\sin(i)}{\sin(r)}$	\bar{n}	$z = \frac{ n_i - \bar{n} }{s}$
1	5	4	1,24942873	1,491761	2,31917776
2	10	6	1,66125256	1,491761	1,62207476
⋮	⋮	⋮	⋮	⋮	⋮
15	75	40	1,50271382	1,491761	0,10482109

La conclusion est, ici :

- la première mesure est aberrante (on pourra donc l'éliminer lorsque l'on calculera l'indice moyen) ;
- les autres mesures, selon notre critère, sont compatibles avec la loi de Snell-Descartes.

En suivant cette démarche on confronte des résultats expérimentaux **compte-tenu de leur incertitude**, à une loi **que l'on suppose valide a priori**.



■ Comment tester si une loi est valide

De nombreuses lois que nous utilisons en physique ou en chimie sont des relations fondées sur des approximations : la loi de Kohlrausch en conductimétrie, la loi de Beer-Lambert en spectrophotométrie, la loi d'Ohm en électricité et la définition du pH enseignée au lycée par exemple. Toutes ces relations ont un **champ de validité** et il peut être intéressant de caractériser celui-ci.

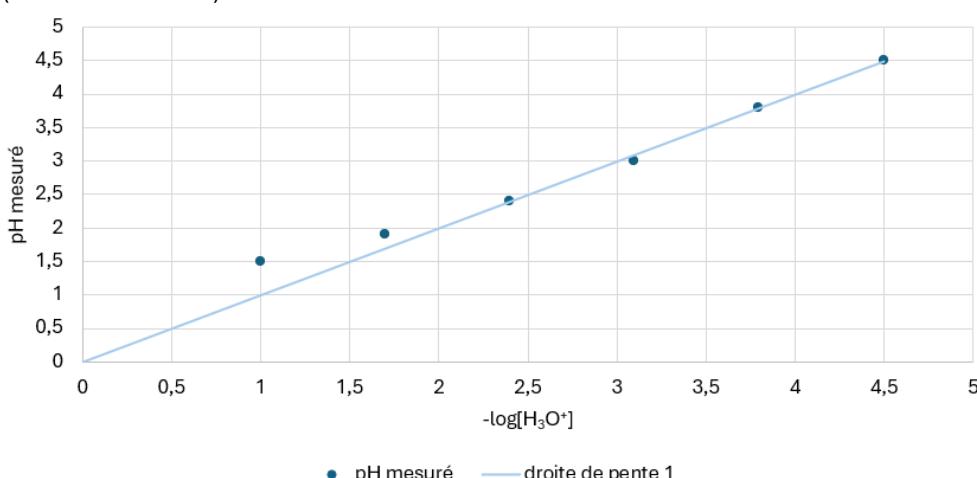
Nous choisissons comme exemple la relation qui, au lycée, fait office de définition du pH : $pH = -\log [H_3O^+]$. Une expérience classique consiste à faire mesurer les pH de solutions d'acide fort dont les concentrations sont connues et de représenter graphiquement le pH mesuré en fonction de $-\log [H_3O^+]$. La question est : que fait-on du nuage de points obtenu ?

Faut-il modéliser le nuage de points ?

Une modélisation par la méthode des MCO est, ici, hors de question : aucune des hypothèses de cette méthode n'est vérifiée, surtout si les élèves ont étalonné eux-mêmes leur pH-mètre et préparé leurs solutions par dilutions successives d'une solution mère (il serait d'ailleurs très dommage de les dispenser de ces tâches). De plus, quel sens aurait une droite de pente forcément différente de 1 et d'ordonnée à l'origine forcément non-nulle, puisque l'on sait que cette relation n'est pas valide pour tous les pH, le but étant justement de le vérifier ?

Alors que faire de ce nuage de points ?

Une proposition peut être de comparer ce nuage de points à une droite d'ordonnée à l'origine nulle et de coefficient directeur 1 : elle représente le résultat que l'on aurait si la loi testée était exactement satisfaite. Cela donne (résultats d'élèves) :



Ici il n'y a pas d'ambiguïté : les points bleu foncé représentent les mesures réalisées et la droite bleu clair représente la relation testée. Visuellement, on voit que la loi est respectée pour des pH suffisamment élevés (ce qui est cohérent : l'activité des ions H_3O^+ tend vers leur concentration molaire lorsque celle-ci tend vers 0, l'approximation sous-jacente à la définition du pH niveau lycée est donc d'autant plus satisfaite que le pH est élevé).

Si l'on veut affiner, il est possible de fixer un critère quantitatif : « on estime que la relation est compatible avec les mesures si l'écart entre les deux est inférieur à... ».

En suivant une telle démarche, **on fait confiance à nos résultats de mesures et c'est la loi qui est mise à l'épreuve**. C'est la situation inverse de celle précédemment évoquée.

Cela suppose que les valeurs mesurées soient fiables. Dans l'exemple ci-dessus : des solutions soigneusement préparées et un pH-mètre étalonné sont indispensables si l'on veut conclure concernant le champ de validité de la loi testée.



Comment comparer des résultats de mesures à un modèle

Dans certains cas, on dispose de résultats de mesure fiables et d'un modèle théorique reposant sur diverses hypothèses. Le modèle n'est pas en cause, il s'agit en revanche de vérifier dans quelle mesure ses hypothèses sont satisfaites par la situation réelle. C'est exactement ce à quoi nous sommes confrontés lorsque nous faisons étudier par pointage la chute d'un objet, dans le but de comparer les résultats au modèle de la chute libre. On sait d'avance que la chute libre ne peut pas exister sur Terre ; il existe cependant des situations dans lesquelles, *compte-tenu de nos moyens de mesure*, on ne voit pas de différence entre la chute réelle et la chute libre. S'il n'y a pas compatibilité, cela n'indique donc pas que les mesures sont mauvaises, ni que le modèle est « faux », c'est juste que l'une de ses hypothèses (le plus souvent l'absence de force de frottement) n'est pas réalisée par la chute étudiée (cela peut aussi révéler un mauvais pointage ou, plus souvent, un mauvais étalonnage de l'écran, mais il est alors possible d'y remédier).

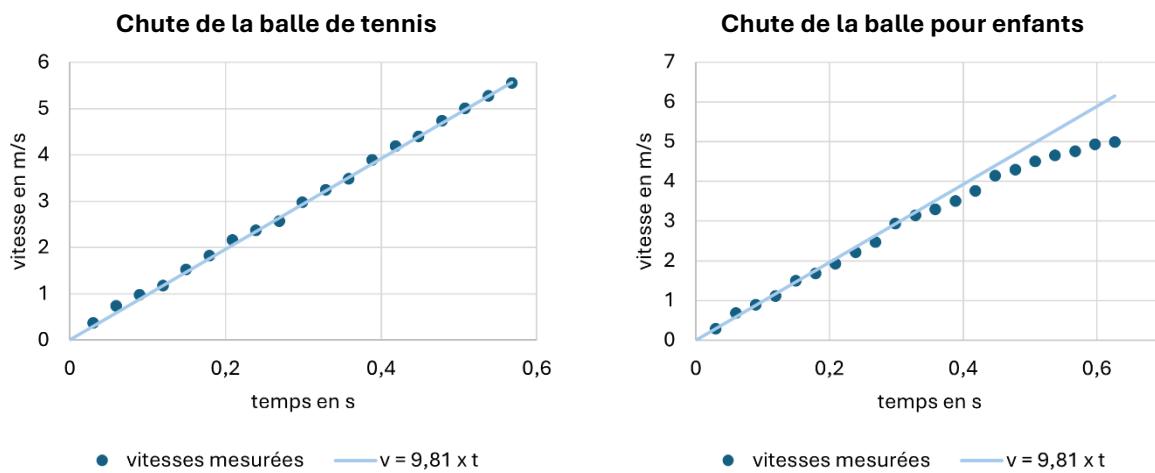
Comment exploiter le nuage de points ?

Pour réaliser les courbes ci-dessous nous avons filmé les chutes, sans vitesse initiale, de deux objets : une balle de tennis et une balle très légère, de diamètre voisin de la précédente, issue d'une piscine à balles pour enfants.

Avec un logiciel de pointage, nous obtenons les valeurs de vitesse à chaque position.

Le modèle de la chute libre et les lois de Newton permettent d'établir que la vitesse d'un objet en chute libre évolue selon la loi : $v(t) = g \times t$.

Nous choisissons donc de représenter graphiquement la vitesse de l'objet en chute en fonction du temps et de superposer le nuage de points obtenu à la droite d'équation $v(t) = 9,81 \times t$ (et non pas de modéliser le nuage de points par une fonction linéaire ou affine), puisque le but est de confronter nos mesures au modèle de la chute libre. Cela donne :

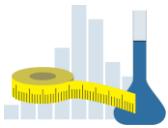


Pour que le modèle de la chute libre soit respecté, deux conditions sont à satisfaire :

- ▶ **1** : l'évolution de la vitesse en fonction du temps doit être linéaire (ou affine selon ce que l'on considère comme l'origine des dates et selon le lieu de l'origine du repère) ;
- ▶ **2** : le coefficient directeur doit être égal au champ de pesanteur terrestre.

Pour tester la condition 1, nous ne préconisons pas une modélisation par la méthode des MCO (laquelle suppose d'examiner un coefficient de détermination à propos duquel nous avons émis plus haut quelques réserves). En revanche il ne faut pas se priver, selon nous, de l'analyse visuelle des graphiques : ceux-ci illustrent que la balle de tennis semble évoluer de manière conforme aux prévisions de la chute libre, tandis que la balle pour enfants, elle, s'éloigne du modèle au fur et à mesure que sa vitesse augmente. Modéliser le second graphique par la méthode des MCO n'aurait aucun sens puisque l'on voit à l'œil nu que ce n'est pas une droite mais une courbe concave.

NB : À bien y réfléchir, l'affirmation « *les points expérimentaux semblent former une droite* » n'est pas beaucoup moins rigoureuse que « *mon coefficient de détermination, dont j'ignore la définition, ne vaut pas 1 mais presque* ».



Ce constat est d'ores et déjà riche en enseignements et l'on peut en déduire de la belle physique, en expliquant pourquoi la force de frottement exercée par l'air a davantage de conséquences sur l'accélération de la balle la plus légère, pourquoi cela s'accentue à la fin du mouvement, etc.

Si l'on veut un critère quantitatif pour évaluer la concordance entre les mesures et le modèle de la chute libre, on peut tester la condition 2 avec le graphique donné par la balle de tennis uniquement.

Puisque les hypothèses de la méthode des MCO sont ici satisfaites (incertitudes très faibles des valeurs de temps, incertitudes voisines et indépendantes entre elles des valeurs de vitesses), on peut modéliser le nuage de points par une fonction affine ou linéaire et en calculer le coefficient directeur : c'est l'accélération de la balle de tennis.

La comparaison entre cette accélération et le champ de pesanteur terrestre, compte-tenu éventuellement de son incertitude, peut fournir un argument quantitatif pour conclure.

En conclusion : la méthode des MCO est utilisée dans ce cas pour mesurer un coefficient directeur, une fois admis qu'il était raisonnable de considérer comme linéaire l'évolution de la grandeur portée en ordonnée. Elle n'est en revanche pas exploitée pour valider la linéarité.