La méthode des moindres carrés

Le principe des moindres carrés
Ce principe a été annoncé par Legendre:
« La valeur la plus probable de n'importe quelle quantité observée est telle que la somme des carrés des différences entre les observations et cette valeur est minimum. »

La moyenne
Supposons d'avoir fait N mesures $x_1, x_2, x_3, \ldots , x_N$ d'une quantité dont la valeur vraie soit x. D'après le principe des moindres carrés, pour déterminer x on doit chercher la valeur qui minimise la quantité $E = \sum_{i=1}^{N} (x_i - x)^2$, donc qui satisfait $\frac{dE}{dx} = 0 \quad \Rightarrow \quad -2\sum_{i=1}^{N} (x_i - x) = 0$. On obtient donc $(x_1 - x) + (x_2 - x) + \ldots + (x_N - x) = 0$, c'est-à-dire $x_1 + x_2 + \ldots + x_N = Nx$, du quel on tire $x = \frac{x_1 + x_2 + \ldots + x_N}{N}$. Ca signifie donc que le meilleur choix pour x est la moyenne arithmétique.

La meilleure droite
Souvent on sait qu'une grandeur physique y est reliée à une autre grandeur x selon une fonction $y = f(x)$ qui décrit le processus physique. La fonction la plus simple est la droite:

$$y = ax + b \tag{1}$$

Normalement dans une expérience on fait plusieurs mesures de y pour différentes x et on cherche la relation entre les deux, c'est-à-dire on cherche les paramètres a et b. Comme chaque mesure a une erreur, les points (x_i, y_i) ne sont sur une droite que approximativement; le problème est donc de trouver la meilleure droite.
Si on connaisse déjà les paramètres a et b, on pourrait calculer pour chaque x_i la correspondante valeur 'théorique' y_i*. Envisageons donc la somme des différences carrées entre la valeur théorique et la valeur mesurée:

$$E = \sum_i (y_i - y_i^*)^2 = \sum_i (y_i - ax_i - b)^2 \tag{2}$$

D'après le principe des moindres carrés on doit chercher les valeurs a et b qui minimisent cette quantité, c'est-à-dire qui satisfont:

$$\frac{\partial E}{\partial a} = 0 \qquad et \qquad \frac{\partial E}{\partial b} = 0 \tag{3}$$

Ca correspond à trouver la solution du système:

$$\begin{cases} \frac{\partial E}{\partial a} = -2\sum_i (y_i - ax_i - b)x_i = 0 \\ \frac{\partial E}{\partial b} = -2\sum_i (y_i - ax_i - b) = 0 \end{cases} \tag{4}$$

Supposons qu'on est ici dans la condition que les erreurs sur les x sont négligeables par rapport aux erreurs sur les y (ça signifie dans le cas de la droite que $\sigma_y >> a\sigma_x$). Comme chaque mesure y_i a son erreur, nous souhaitons que dans le calcul les mesures avec une grande erreur comptent moins que celles avec une erreur petite, donc c'est mieux envisager la quantité

$$E = \sum_i \frac{(y_i - ax_i - b)^2}{\sigma_i^2} \tag{5}$$

et donc résoudre le système:
$$\begin{cases} \dfrac{\partial E}{\partial a} = -2\sum_i \dfrac{(y_i - ax_i - b)x_i}{\sigma_i^2} = 0 \\ \dfrac{\partial E}{\partial b} = -2\sum_i \dfrac{(y_i - ax_i - b)}{\sigma_i^2} = 0 \end{cases} \quad (6)$$

qui peut être écrit :
$$\begin{cases} \sum_i \dfrac{y_i x_i}{\sigma_i^2} - a\sum_i \dfrac{x_i^2}{\sigma_i^2} - b\sum_i \dfrac{x_i}{\sigma_i^2} = 0 \\ \sum_i \dfrac{y_i}{\sigma_i^2} - a\sum_i \dfrac{x_i}{\sigma_i^2} - b\sum_i \dfrac{1}{\sigma_i^2} = 0 \end{cases} \quad (7)$$

On définit
$$\sum_i \frac{y_i x_i}{\sigma_i^2} = [yx]_N \quad ; \quad \sum_i \frac{x_i}{\sigma_i^2} = [x]_N \quad ; \quad \sum_i \frac{y_i}{\sigma_i^2} = [y]_N \quad ; \quad \sum_i \frac{x_i^2}{\sigma_i^2} = [x^2]_N \quad ; \quad \sum_i \frac{1}{\sigma_i^2} = [1]_N$$

et on écrit donc:
$$\begin{cases} [yx]_N - a[x^2]_N - b[x]_N = 0 \\ [y]_N - a[x]_N - b[1]_N = 0 \end{cases} \quad (8)$$

La solution du système est:
$$\begin{cases} a = \dfrac{[1]_N [yx]_N - [x]_N [y]_N}{[1]_N [x^2]_N - [x]_N^2} \\ b = \dfrac{[x^2]_N [y]_N - [x]_N [yx]_N}{[1]_N [x^2]_N - [x]_N^2} \end{cases} \quad (9)$$

Les erreurs:
Les erreurs sur les paramètres a et b peuvent être calculées avec la formule de propagation des erreurs. Si on est dans la condition que les erreurs sur les *x* sont négligeables par rapport aux erreurs sur les *y* ce calcul donne:
$$\begin{cases} \sigma_a^2 = \dfrac{[1]_N}{[1]_N [x^2]_N - [x]_N^2} \\ \sigma_b^2 = \dfrac{[x^2]_N}{[1]_N [x^2]_N - [x]_N^2} \end{cases} \quad (10)$$

Généralisation
La méthode qu'on a discutée est valable aussi si la relation entre x et y n'est pas linéaire. Si par exemple, on sait que la relation est parabolique ($f(x) = ax^2 + bx + c$) on devrait minimiser la quantité
$$E = \sum_i \frac{(y_i - ax_i^2 - bx_i - c)^2}{\sigma_i^2}, \quad (11)$$

donc résoudre le système:

$$\begin{cases} \dfrac{\partial E}{\partial a} \propto \sum_i \dfrac{\left(y_i - ax_i^2 - bx_i - c\right)x_i^2}{\sigma_i^2} = 0 \\ \dfrac{\partial E}{\partial b} \propto \sum_i \dfrac{\left(y_i - ax_i^2 - bx_i - c\right)x_i}{\sigma_i^2} = 0 \\ \dfrac{\partial E}{\partial c} \propto \sum_i \dfrac{\left(y_i - ax_i^2 - bx_i - c\right)}{\sigma_i^2} = 0 \end{cases} \qquad (12)$$

Le calcul commence à devenir lourde, et on va de mal en pis si on a des fonctions transcendantes (log, sin, cos...). Si possible on doit chercher de changer la relation en linéaire. Supposons par exemple d'avoir la fonction $y = a\, e^{-\gamma x}$: il est convenable d'envisager les logarithmes et d'écrire $\ln y = \ln a - \gamma\, x$. Dans ce cas la relation entre ln y et x est linéaire et on peut calculer aisément les paramètres $\ln a$ et γ, donc a aussi.

N.B. Le cas discuté dans le premier paragraphe, 'la moyenne', correspond donc à prendre *f(x)=cost.*

Observations sur les erreurs

a) Les erreurs sur les *y* ont toutes la même valeur σ_y .

Dans ce cas on peut faire les calculs avec le système (4), ce qui simplifie le résultat: les paramètres sont donnés toujours par les formules (9), mais on utilise

$$[yx]_N = \sum_i y_i x_i \;\; ; \;\; [x]_N = \sum_i x_i \;\; ; \;\; [y]_N = \sum_i y_i \;\; ; \;\; \left[x^2\right]_N = \sum_i x_i^2 \;\; ; \;\; [1]_N = 1$$

Les erreurs sur les parametres sont aussi simplifiés:

$$\sigma_a^2 = \sigma_y^2 \frac{1}{\left[x^2\right]_N - [x]_N^2} \qquad \text{et} \qquad \sigma_b^2 = \sigma_y^2 \frac{\left[x^2\right]_N}{\left[x^2\right]_N - [x]_N^2} \qquad (13)$$

b) Jusqu'ici on a considéré que les erreurs sur les *x* sont négligeables par rapport aux erreurs sur les *y*. S'il arrive le contraire, il est facile de trouver les formules correspondantes. On doit minimiser la quantité suivante (σ_i sont maintenant les erreurs sur les *x*):

$$E = \sum_i \frac{(x_i - x_i *)^2}{\sigma_i^2} = \sum_i \frac{(x_i - y_i/a - b/a)^2}{\sigma_i^2} \qquad (14)$$

c) Si les variables *x* et *y* ont erreurs similaires et on ne peut négliger ni les unes ni les autres, les calculs à faire deviennent vite compliqués. Envisageons les figures suivantes:

Dans la première figure on suppose que les mesures ont seulement les erreurs sur les *y*, dans la deuxième seulement sur les *x*. Si on a les deux, il faut envisager la somme des différences soit horizontales soit verticales, donc

$$E = \sum_i \left\{ \frac{(y_i - y_i *)^2}{\sigma_y^2} + \frac{(x_i - x_i *)^2}{\sigma_x^2} \right\} \qquad (15)$$

et calculer les paramètres avec la procédure habituelle de minimisation.