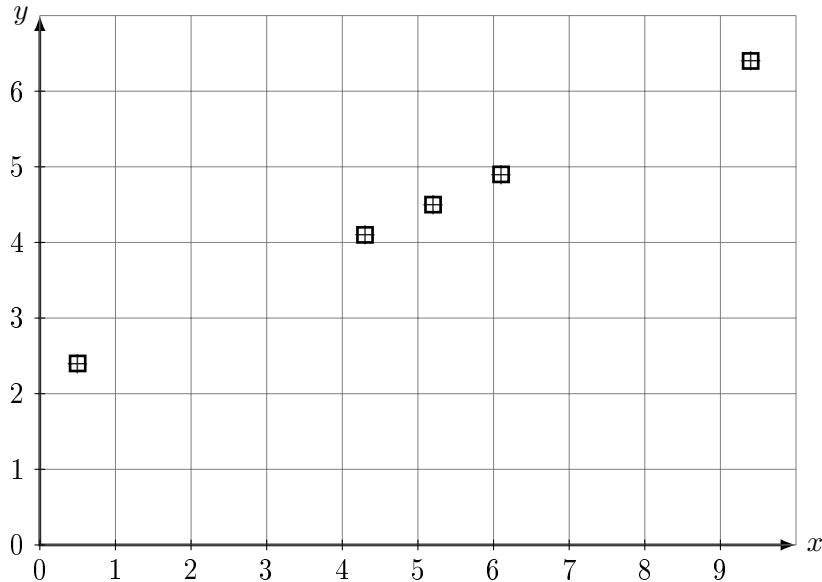


1 Définition du problème avec un exemple

Supposons que l'on fasse le relevé suivant :

x_i	0,5	4,3	5,2	6,1	9,4
y_i	2,4	4,1	4,5	4,9	6,4

Faisons l'hypothèse que les valeurs sont connues à 0,1 près.



On souhaite calculer l'ajustement affine par la méthode des moindres carrés et connaître l'incertitude sur les résultats.

On cherche donc les valeurs de a et b avec leur incertitude dans modèle $y = a \cdot x + b$.

- On notera N le nombre de paires. Ici $N = 5$.
- On sait que $a = \frac{C_{xy}}{V_x}$
- On sait que $b = \bar{y} - a \cdot \bar{x}$

Mais il va falloir préciser les notations pour pouvoir intégrer les erreurs.

2 Variables aléatoires

2.1 Retournement de perspective

Soit une grandeur x réelle (à supposer que l'idée ait un sens). Je la mesure, j'obtiens x_{mes} . x_{mes} est potentiellement différent de x à cause des erreurs de mesures. On pourrait alors dire $x_{mes} = x + \varepsilon$.

Dans cette perspective, c'est l'erreur de mesure qui est aléatoire et x qui est une valeur déterministe fixée.

Mais cela ne nous avance pas à grand chose puisque x est inconnue. De plus, c'est une approche curieuse puisque suite à l'expérience, x_{mes} est connue et plus aléatoire...

Voici comment on peut changer d'approche : On suppose l'existence de x , mais x reste inconnue. Avant toute expérience, on décrit l'erreur comme aléatoire de sorte que la valeur que l'on va mesurer (la proba est faite avant la mesure) est décrite comme aléatoire X_{mes} .

L'étude probabiliste de X_{mes} permet de faire des prédictions sur les valeurs que l'on est susceptible de mesurer. Cette étude revient à faire des probas conditionnelles comme :

$$P_{x=\text{une certaine valeur}}(X_{mes} = x_{mes})$$

On fait une expérience et on obtient x_{mes} . Cette valeur est connue et n'est donc plus sujette à aléa. La probabilité se renverse : maintenant, **sachant** que l'on a obtenu ce résultat, quelle est la probabilité que x valait telle ou telle valeur ?

$$P_{X_{mes}=x_{mes}}(x = \text{une certaine valeur})$$

Cela revient à considérer que c'est la mesure qui est connue et la valeur réelle qui est aléatoire.

2.2 Application au problème

Pour un point i donné, x_i est le résultat de la mesure. Puisque cette mesure est soumise à une incertitude, on suppose que l'écart entre le x_i mesuré et la valeur réelle est aléatoire.

La valeur réelle sera donc pour nous comme une variable aléatoire $X_i = x_i + \varepsilon_{X_i}$. L'écart entre la valeur réelle et la valeur mesurée est aussi une variable aléatoire. On la supposera toujours d'**espérance nulle** (si ce n'était pas le cas, il suffirait d'ajouter $\overline{\varepsilon_{X_i}}$ à x_i)

Dans l'exemple, ε_{X_i} suit une loi uniforme d'espérance 0 et d'écart-type $\frac{0,1}{\sqrt{3}}$.

On fait l'hypothèse que toutes les mesures subissent la même incertitude. Donc tous les ε_{X_i} suivent la même loi de probabilité, ils ont tous une espérance et un même écart-type que l'on notera σ_X .

La variable X_i suit donc une loi de même nature que ε_{X_i} , avec l'écart-type σ_X mais une espérance x_i .

Naturellement, tout ce qui vient d'être dit s'applique à y .

Important : x et y ne sont pas indépendants puisque justement on cherche à établir un modèle $y = a \cdot x + b$. Mais les erreurs de mesures peuvent être considérées comme indépendantes. Donc les ε_{X_i} et ε_{Y_i} sont indépendants.

Par conséquent, les X_i et Y_i **en tant que variables aléatoires** sont indépendants.

Cela entraîne par exemple que $E(X_i \cdot Y_i) = E(X_i) \cdot E(Y_i)$.

3 Grandeurs composées

3.1 Calculer a et b

Il faut bien définir le cadre. Sur la base de la série de N points (x_i, y_i) que l'on a obtenu lors d'une expérience, tenant compte de l'incertitude on veut établir une loi de probabilité pour les valeurs de a et b .

Cela peut sembler un peu étrange : c'est comme si j'envisageai de répéter un grand nombre de fois la mesure du nuage, qu'à chaque fois j'obtiendrais le même nuage (x_i, y_i) mais que les valeurs aléatoires (X_i, Y_i) (qui je le rappelle sont censées représenter les valeurs réelles) se réalisaient différemment, selon leurs lois de probabilités.

Ces valeurs différentes des N paires (X_i, Y_i) constituant le nuage à chaque expérience, donneraient chaque fois une réalisation différente de A et B .

A et B doivent donc être vues comme des variables aléatoires.

On considère que l'espérance $E(A)$ est la valeur de référence et l'écart-type $\sigma(A)$ est une mesure de l'incertitude sur A . Idem pour B .

Il nous faut donc déterminer les fonctions

$$(X_i, Y_i) \mapsto A, B$$

Et faire une propagation d'incertitude pour déterminer les incertitudes sur A et B tenant compte des incertitudes sur les points du nuage.

Le calcul de a et b nécessite les grandeurs C_{xy} , V_x , \bar{x} et \bar{y} , nous allons les détailler.

4 Les moyennes

Notations

- Je note $\bar{x} = \frac{1}{N} \sum_i x_i$, c'est à dire la moyenne des valeurs mesurées.
- σ_x est l'écart-type pour la série de valeurs du nuage.
- Je note $\overline{X_i} = E(X_i)$ l'espérance (donc moyenne probabiliste) pour la variable aléatoire X_i . On a déjà dit $\overline{X_i} = x_i$.
- σ_X est l'écart-type relatif à l'incertitude sur chaque mesure x_i .
- Je note $\overline{X_N}$ la moyenne des réalisations des N valeurs X_i lors d'une expérience (potentielle puisque ce sont des probas)

Ne pas confondre σ_X et σ_x qui n'ont rien à voir. σ_X correspond aux incertitudes, σ_x correspond à la dispersion des valeurs dans un même nuage. Exemples extrêmes :

- S'il n'y avait qu'un seul point dans le nuage, on aurait $\sigma_x = 0$. Mais ce point pourrait être mesuré avec une erreur, alors $\sigma_X \neq 0$.
- Si les mesures étaient parfaites, alors $\sigma_X = 0$. Mais cela n'empêche pas d'avoir des valeurs x_i différentes dans le nuage, alors $\sigma_x \neq 0$.

Étudions ici la variable aléatoire $\overline{X_N}$.

$$\begin{aligned} E(\overline{X_N}) &= \frac{1}{N} \sum_i x_i + \varepsilon_{X_i} \\ &= \bar{x} + \frac{1}{N} \sum_i E(\varepsilon_{X_i}) \\ &= \bar{x} + \frac{1}{N} \sum_i 0 \\ &= \bar{x} \end{aligned}$$

Pour la propagation d'incertitude, il faut les dérivées partielles.

$$\frac{\partial \overline{X_N}}{\partial X_i} = \frac{\partial}{\partial X_i} \frac{1}{N} \sum_i X_i = \frac{1}{N}$$

Partant de là on peut calculer :

$$\sigma_{\overline{X_N}} = \sqrt{\sum_i \left(\frac{\partial \overline{X_N}}{\partial X_i} \sigma_X \right)^2} = \sqrt{\sum_i \left(\frac{\sigma_X}{N} \right)^2} = \frac{\sigma_X}{\sqrt{N}}$$

On en déduit donc que $\overline{X_N}$ suit une loi de probabilité d'espérance \bar{x} et d'écart-type $\frac{\sigma_X}{\sqrt{N}}$.

Même si mes ε des incertitudes ne sont pas gaussiens, la moyenne $\overline{X_N}$ tend à être gaussienne.

Même chose pour y .

4.1 Variance

Notations

- V_x est la variance calculée sur les valeurs de la série mesurée

$$V_x = \frac{1}{N} \sum_i (x_i - \bar{x})^2 = \left(\frac{1}{N} \sum_i x_i^2 \right) - \bar{x}^2$$

- V_{X_N} est la variance calculée sur les valeurs potentielles du nuage (X_i, Y_i) .

$$V_{X_N} = \frac{1}{N} \sum_i (X_i - \overline{X_N})^2$$

On commence par l'espérance. On rappelle qu'en général : $V_A = E(A^2) - E(A)^2$.

$$\begin{aligned} E(V_{X_N}) &= \frac{1}{N} E \left(\sum_i (X_i - \overline{X_N})^2 \right) \\ &= \frac{1}{N} E \left(\sum_i X_i^2 \right) - \frac{2}{N} E \left(\overline{X_N} \sum_i X_i \right) + \frac{1}{N} \times N \times E \left(\overline{X_N}^2 \right) \\ &= \frac{1}{N} \sum_i E(X_i^2) - 2 E \left(\overline{X_N}^2 \right) + E \left(\overline{X_N}^2 \right) \\ &= \frac{1}{N} \sum_i \left(E(X_i)^2 + \sigma_X^2 \right) - \left(E \left(\overline{X_N} \right)^2 + \frac{\sigma_X^2}{N} \right) \\ &= \frac{1}{N} \sum_i (x_i^2 + \sigma_X^2) - \left(\bar{x}^2 + \frac{\sigma_X^2}{N} \right) \\ &= \frac{1}{N} \sum_i x_i^2 - \bar{x}^2 + \frac{N-1}{N} \sigma_X^2 \\ &= V_x + \frac{N-1}{N} \sigma_X^2 \end{aligned}$$

Cela rappelle le résultat bien connu : quand on obtient un écart-type sur un échantillon, il faut majorer cet écart-type par un facteur $\sqrt{\frac{n}{n-1}}$ pour correctement estimer l'écart-type de la population.

Maintenant les dérivées partielles :

$$\begin{aligned} \frac{\partial V_{X_N}}{\partial X_\alpha} &= \frac{2}{N} (X_\alpha - \overline{X_N}) - \frac{2}{N} \frac{\partial \overline{X_N}}{\partial X_\alpha} \sum_i (X_i - \overline{X_N}) \\ &= \frac{2}{N} (X_\alpha - \overline{X_N}) - \frac{2}{N} \frac{\partial \overline{X_N}}{\partial X_\alpha} \times 0 \\ &= \frac{2}{N} (X_\alpha - \overline{X_N}) \end{aligned}$$

Cette expression devant être prise pour les erreurs nulles :

$$\frac{\partial V_{X_N}}{\partial X_\alpha} = \frac{2}{N} (x_\alpha - \bar{x})$$

Donc l'écart-type :

$$\begin{aligned}
\sigma(V_{X_N}) &= \sqrt{\sum_{\alpha} \left(\frac{\partial V_{X_N}}{\partial X_{\alpha}} \sigma_X \right)^2} \\
&= 2 \frac{\sigma_X}{N} \sqrt{\sum_{\alpha} (x_{\alpha} - \bar{x})^2} \\
&= 2\sigma_X \sqrt{\frac{V_x}{N}}
\end{aligned}$$

5 Corrélation

On s'intéresse maintenant à la corrélation.

Notations

- C_{xy} est la corrélation calculée sur les valeurs de la série mesurée

$$C_{xy} = \frac{1}{N} \sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \left(\frac{1}{N} \sum_i x_i \cdot y_i \right) - \bar{x} \cdot \bar{y}$$

- C_{XY} est la corrélation calculée sur les valeurs potentielles du nuage (X_i, Y_i) .

$$C_{XY_N} = \frac{1}{N} \sum_i (X_i - \bar{X}_N) \cdot (Y_i - \bar{Y}_N)$$

Calculons d'abord l'espérance en tenant compte de l'indépendance de X et Y .

$$\begin{aligned}
E(C_{XY_N}) &= E\left(\frac{1}{N} \sum_i (X_i - \bar{X}_N) \cdot (Y_i - \bar{Y}_N)\right) \\
&= \frac{1}{N} \sum_i E(X_i - \bar{X}_N) \cdot E(Y_i - \bar{Y}_N) \\
&= \frac{1}{N} \sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y}) \\
&= C_{xy}
\end{aligned}$$

Et les dérivées partielles :

$$\begin{aligned}
\frac{\partial C_{XY_N}}{\partial X_{\alpha}} &= \frac{1}{N} \frac{\partial}{\partial X_{\alpha}} \sum_i X_i (Y_i - \bar{Y}_N) - \frac{1}{N} \frac{\partial}{\partial X_{\alpha}} \bar{X}_N \sum_i (Y_i - \bar{Y}_N) \\
&= \frac{1}{N} (Y_{\alpha} - \bar{Y}_N) - \frac{1}{N} \cdot \frac{1}{N} \sum_i (Y_i - \bar{Y}_N) \\
&= \frac{1}{N} (Y_{\alpha} - \bar{Y}_N) - 0
\end{aligned}$$

Cette dérivée devant être calculée pour les erreurs à 0, donc :

$$\frac{\partial C_{XY}}{\partial X_{\alpha}} = \frac{1}{N} (y_{\alpha} - \bar{y})$$

De la même façon (c'est symétrique) :

$$\frac{\partial C_{XY_N}}{\partial Y_{\alpha}} = \frac{1}{N} (x_{\alpha} - \bar{x})$$

On peut en déduire l'écart-type pour C_{XY_N} :

$$\begin{aligned}
 \sigma(C_{XY_N}) &= \sqrt{\sum_{\alpha} \left(\frac{\partial C_{XY}}{\partial X_{\alpha}} \sigma_X \right)^2 + \left(\frac{\partial C_{XY}}{\partial Y_{\alpha}} \sigma_Y \right)^2} \\
 &= \frac{1}{N} \sqrt{\sum_{\alpha} (y_{\alpha} - \bar{y})^2 \sigma_X^2 + (x_{\alpha} - \bar{x})^2 \sigma_Y^2} \\
 &= \frac{1}{N} \sqrt{N V_y \sigma_X^2 + N V_x \sigma_Y^2} \\
 &= \sqrt{\frac{V_y \sigma_X^2 + V_x \sigma_Y^2}{N}}
 \end{aligned}$$

6 Coefficient a

On sait que $a = \frac{C_{xy}}{V_x}$. Donc on définit $A = \frac{C_{XY_N}}{V_{X_N}}$ et on cherche l'espérance de A et son écart-type.

Je commence par l'écart-type et les dérivées partielles qui sont plus simples. Je rappelle que l'on prend la dérivées en 0, c'est à dire pour les erreurs nulles.

$$\begin{aligned}
 \frac{\partial A}{\partial X_{\alpha}} &= \frac{1}{V_{X_N}^2} \left(V_{X_N} \frac{\partial C_{XY_N}}{\partial X_{\alpha}} - C_{XY_N} \frac{\partial V_{X_N}}{\partial X_{\alpha}} \right) \\
 &= \frac{1}{V_x^2} \left(V_x \frac{1}{N} (y_{\alpha} - \bar{y}) - C_{xy} \frac{2}{N} (x_{\alpha} - \bar{x}) \right)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial A}{\partial Y_{\alpha}} &= \frac{1}{V_{X_N}} \frac{\partial C_{XY_N}}{\partial Y_{\alpha}} \\
 &= \frac{1}{V_x} \frac{1}{N} (x_{\alpha} - \bar{x})
 \end{aligned}$$

Dans le calcul on va exploiter que $C_{xy}^2 = r^2 V_x V_y$.

$$\begin{aligned}
 \sigma(A) &= \sqrt{\sum_{\alpha} \left(\frac{\partial A}{\partial X_{\alpha}} \sigma_X \right)^2 + \left(\frac{\partial A}{\partial Y_{\alpha}} \sigma_Y \right)^2} \\
 &= \sqrt{\frac{\sigma_X^2}{N^2 V_x^4} \sum_{\alpha} (V_x (y_{\alpha} - \bar{y}) - 2C_{xy} (x_{\alpha} - \bar{x}))^2 + \frac{\sigma_Y^2}{N^2 V_x^2} \sum_{\alpha} (x_{\alpha} - \bar{x})^2} \\
 &= \sqrt{\frac{\sigma_X^2}{N^2 V_x^4} \sum_{\alpha} [V_x^2 (y_{\alpha} - \bar{y})^2 - 4V_x C_{xy} (x_{\alpha} - \bar{x}) (y_{\alpha} - \bar{y}) + 4C_{xy}^2 (x_{\alpha} - \bar{x})^2] + \frac{\sigma_Y^2}{N V_x^2} V_x} \\
 &= \sqrt{\frac{\sigma_X^2}{N V_x^4} [V_x^2 V_y - 4V_x C_{xy}^2 + 4C_{xy}^2 V_x] + \frac{\sigma_Y^2}{N V_x^2} V_x} \\
 &= \sqrt{\frac{\sigma_X^2}{N V_x^4} [V_x^2 V_y - 4r^2 V_x^2 V_y + 4r^2 V_x^2 V_y] + \frac{\sigma_Y^2}{N V_x^2} V_x} \\
 &= \frac{1}{V_x} \sqrt{\frac{\sigma_X^2 V_y + \sigma_Y^2 V_x}{N}}
 \end{aligned}$$

Pour l'espérance je vais supposer $V_{X_N} - V_x$ assez petit.

$$\begin{aligned}
 E(A) &= E\left(\frac{C_{XY_N}}{V_{X_N}}\right) \\
 &= E\left(\frac{C_{XY_N}}{V_x + V_{X_N} - V_x}\right) \\
 &= \frac{1}{V_x} E\left(\frac{C_{XY_N}}{1 + \frac{V_{X_N} - V_x}{V_x}}\right) \\
 &\approx \frac{1}{V_x} E\left(C_{XY_N} \left(1 - \frac{V_{X_N} - V_x}{V_x}\right)\right) \\
 &\approx \frac{1}{V_x} E\left(C_{XY_N} \left(2 - \frac{V_{X_N}}{V_x}\right)\right) \\
 &\approx 2 \frac{C_{xy}}{V_x} - \frac{1}{V_x^2} E(C_{XY_N} V_{X_N})
 \end{aligned}$$

Et en faisant l'hypothèse audacieuse que C_{XY_N} et V_{X_N} sont à peu près indépendants :

$$\begin{aligned}
 E(A) &= 2 \frac{C_{xy}}{V_x} - \frac{1}{V_x^2} E(C_{XY_N}) \times E(V_{X_N}) \\
 &= 2 \frac{C_{xy}}{V_x} - \frac{1}{V_x^2} C_{xy} \times \left(V_x + \frac{\sigma_X^2}{N}\right) \\
 &= \frac{C_{xy}}{V_x} - \frac{C_{xy} \sigma_X^2}{NV_x^2}
 \end{aligned}$$

En utilisant $a = \frac{C_{xy}}{V_x}$ calculé sur le nuage de points mesurés :

$$\begin{aligned}
 E(A) &= a - a \frac{\sigma_X^2}{NV_x} \\
 &= a \left(1 - \frac{\sigma_X^2}{NV_x}\right) \approx \frac{C_{xy}}{V_x + \frac{\sigma_X^2}{N}}
 \end{aligned}$$

7 Coefficient b

$$\begin{aligned}
 E(B) &= E(\overline{Y_N} - A \times \overline{X_N}) \\
 &= \bar{y} - E(A) \cdot \bar{x}
 \end{aligned}$$

J'ai fait une hypothèse un peu foireuse que A et $\overline{X_N}$ sont indépendants. Ce n'est sûrement pas le cas. On peut dire aussi que $A = E(A) + \varepsilon_A$ et $\overline{X_N} = \bar{x} + \varepsilon_{\overline{X_N}}$ et en supposant que les ε sont petits, le terme d'ordre 2 peut être négligé ce qui revient au même.

$$E(B) = \bar{y} - \frac{C_{xy}}{V_x + \frac{\sigma_X^2}{N}} \bar{x}$$

On passe aux dérivées :

$$\begin{aligned}
 \frac{\partial B}{\partial X_\alpha} &= \frac{\partial \overline{Y_N}}{\partial X_\alpha} - A \times \frac{\partial \overline{X_N}}{\partial X_\alpha} - \frac{\partial A}{\partial X_\alpha} \times \overline{X_N} \\
 &= 0 - a \times \frac{1}{N} - \bar{x} \times \frac{1}{V_x^2} \left(V_x \frac{1}{N} (y_\alpha - \bar{y}) - C_{xy} \frac{2}{N} (x_\alpha - \bar{x}) \right)
 \end{aligned}$$

$$\begin{aligned}\frac{\partial B}{\partial Y_\alpha} &= \frac{\partial \bar{Y}_N}{\partial Y_\alpha} - A \times \frac{\partial \bar{X}_N}{\partial Y_\alpha} - \frac{\partial A}{\partial Y_\alpha} \times \bar{X}_N \\ &= \frac{1}{N} - a \times 0 - \bar{x} \times \frac{1}{V_x} \frac{1}{N} (x_\alpha - \bar{x})\end{aligned}$$

$$\begin{aligned}\sigma(B) &= \sqrt{\sum_\alpha \left(\frac{\partial B}{\partial X_\alpha} \sigma_X \right)^2 + \left(\frac{\partial B}{\partial Y_\alpha} \sigma_Y \right)^2} \\ &= \frac{1}{\sqrt{N}} \sqrt{\sigma_X^2 \left[a^2 + \frac{\bar{x}^2}{V_x^4} (V_x^2 V_y + 4C_{xy}^2 V_x - 4V_x C_{xy}^2) \right] + \sigma_Y^2 \left[1 + \frac{\bar{x}^2}{V_x^2} V_x \right]} \\ &= \frac{1}{\sqrt{N}} \sqrt{\sigma_X^2 \left[a^2 + \frac{\bar{x}^2}{V_x^2} V_y \right] + \sigma_Y^2 \left[1 + \frac{\bar{x}^2}{V_x^2} V_x \right]} \\ &= \sqrt{\frac{\sigma_Y^2}{N} + a^2 \frac{\sigma_X^2}{N} + \bar{x}^2 \frac{V_y \sigma_X^2 + V_x \sigma_Y^2}{NV_x^2}}\end{aligned}$$

8 Exemple

Reprenons l'exemple donné au début :

x_i	0,5	4,3	5,2	6,1	9,4
y_i	2,4	4,1	4,5	4,9	6,4

Faisons l'hypothèses que les valeurs sont connues à 0,1 près.

- $N = 5$
- $\sigma_X = \sigma_Y = \frac{0,1}{\sqrt{12}} \approx 0,0289$
- $\bar{x} = 5,1$, $V_x = 8,26$
- $\bar{y} = 4,46$, $V_x = 1,6664$
- $C_{xy} = 3,71$
- $a \approx 0,44915$ et $b \approx 2,1693$ (calculs sur le nuage)

On obtient :

- $E(A) \approx 0,44914$ et $\sigma(A) \approx 0,00492$
- $E(B) \approx 2,16937$ et $\sigma(B) \approx 0,02953$

9 Méthode OLS

9.1 De quelles erreurs on parle ?

D'après ce que j'ai compris, OLS = Ordinary Least Square, donc moindre carré. Donc exactement ce que j'ai fait avant.

Dans cette méthode, on dispose d'une série de paires x_i, y_i et on aimerait trouver un modèle $y = a \cdot x + b$. Il s'agit de trouver a et b qui minimisent une erreur.

On pourrait vouloir chercher la droite la plus proche (qui minimise la somme des distance, ou même la somme des distance au carré par rapport au nuage de points) mais cela n'a pas grand sens. En effet, la grandeur en x et la grandeur en y n'ont pas forcément la même unité. Quel serait le sens de faire une distance avec des secondes en x et des Volts en y par exemple ? En géométrie analytique, on calcule des distance $\sqrt{x^2 + y^2}$ quand le repère est orthonormé.

Donc on casse la symétrie entre x et y en décidant que ce que l'on minimisera, ce sera la somme des carrés des erreurs selon y .

On pourrait se dire : Problème, il y a des erreurs sur x et y , pourquoi ne prendre en compte que les erreurs sur y ?

Ce n'est pas de cela qu'il s'agit. On ne parle pas d'erreur de mesures. Quand on fait le calcul OLS, on suppose que les x_i et y_i sont justes et on ne prend pas d'erreur de mesure en compte. L'erreur dont il est question correspond à l'erreur de prédiction du modèle. En effet, même en choisissant bien a et b , pour un point du nuage, je n'aurais pas forcément $y_i = a \cdot x_i + b$. Les points ne sont pas forcément alignés donc nécessairement, il y aura un écart entre le $y = a \cdot x + b$ prévu par le modèle et le y_i effectivement mesuré.

Pourquoi les points ne sont-ils pas alignés ?

- On peut supposer qu'ils devraient être alignés, que le modèle est parfait et que le défaut d'alignement vient d'erreurs de mesures. Ce n'est pas très réaliste.
- On peut considérer que le modèle n'est qu'une approximation d'une réalité plus complexe. y n'est pas parfaitement lié à x par une loi affine.
- Peut-être (et cela pourrait être un mélange des précédents points) que y dépend d'autres paramètres que x non pris en compte dans le modèle. Cette dépendance à d'autres paramètres peut être vue comme une imperfection du modèle lui même et/ou comme un bruit interprété comme une erreur aléatoire.

En tout cas, on peut choisir d'interpréter l'erreur faite par le modèle comme une erreur aléatoire.

Prenons un exemple :

x_i	1	2	3	4
y_i	2	2,5	3,1	3,4

Disons que je décide d'adopter le modèle $y = 0,5x + 1,5$ (pas la droite des moindres carrés, mais après tout, je choisis le modèle que je veux)

x_i	1	2	3	4
y_i	2	2,5	3,1	3,4
$\hat{y}_i = a \cdot x_i + b$	2	2,5	3	3,5
$\varepsilon_i = \hat{y}_i - y_i$	0	0	-0,1	0,1

Donc le modèle prédit des valeurs \hat{y} qui ne correspondent pas toujours aux valeurs mesurées. Il y a des erreurs ε_i .

La droite des moindres carrés passe par le point moyen $G(\bar{x}; \bar{y})$ ce qui garantit que l'erreur moyenne est nulle : $\bar{\varepsilon} = 0$.

On peut donc calculer la variance :

$$V_\varepsilon = \frac{1}{N} \sum_i (\varepsilon_i - \bar{\varepsilon})^2 = \sum_i \varepsilon_i^2$$

C'est justement cette quantité qui est minimisée quand on calcule la droite des moindres carrés. Ici on obtient :

$$V_\varepsilon = \frac{1}{4} (0^2 + 0^2 + 0,1^2 + 0,1^2) = 0,005 \Rightarrow \sigma_\varepsilon = \sqrt{V_\varepsilon} \approx 0,071$$

Voilà comment on arrive à dire que mon modèle $y = 0,5x + 1,5$ fait une erreur sur y , moyenne de l'erreur : 0 (toujours avec une droite des moindres carrés) et σ de l'erreur : 0,071 (cette valeur est minimale avec la droite des moindres carrés)

D'ailleurs, si dans cet exemple, au lieu de prendre $y = 0,5x + 1,5$, je calcule la droite des moindres carrés, j'obtiens $y = 0,48x + 1,55$. Je peux refaire le tableau :

x_i	1	2	3	4
y_i	2	2,5	3,1	3,4
$\hat{y}_i = a \cdot x_i + b$	2,03	2,51	2,99	3,47
$\varepsilon_i = \hat{y}_i - y_i$	0,03	0,01	-0,11	0,07

La moyenne des erreurs est bien nulle. La variance est : $V_\varepsilon = 0,0045 \Rightarrow \sigma_\varepsilon \approx 0,067$. C'est mieux que ce que l'on avait avec la droite précédente. On ne pourra pas faire mieux, la droite des moindres carrés est optimale de ce point de vue.

9.2 Avec des matrices

J'ai plutôt l'habitude du calcul $a = \frac{C_{xy}}{V_x}$ toutefois on peut poser le calcul autrement. On peut même généraliser au cas où il y aurait plus d'une variable.

On dispose de variables x_1, x_2, \dots, x_p en entrée. Ces variables sont censées expliquer y . On voudrait trouver une formule de forme :

$$y = \beta_0 + \sum_j \beta_j x_j$$

Mais une seule mesure ne suffit pas alors on en fait n . Les n mesures de la grandeur x_1 seront notées $x_{11}, x_{21}, \dots, x_{n1}$. De façon générale, on note x_{ij} la mesure de la variable x_1 pour le point i . De même, y_i est la grandeur de sortie pour le point i .

On note \hat{y} la valeur de y prédite par le modèle. On prévoit donc :

$$\hat{y}_i = \beta_0 + \sum_j x_{ij} \beta_j$$

Cela ressemble beaucoup à du matriciel mais le β_0 est à part...

On fabrique alors les matrices Y , X et β de la façon suivante :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} ; \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

De sorte que $\hat{Y} = X \cdot \beta$. Je vous épargne les détails, on obtient que le vecteur de coefficients β optimal (selon le critère des moindres carrés) sera obtenu par :

$$\beta = (X^t X)^{-1} X^t Y$$

On peut le faire avec l'exemple donné dans la section précédente :

$$Y = \begin{pmatrix} 2 \\ 2,5 \\ 3,1 \\ 3,4 \end{pmatrix} ; \quad X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} b \\ a \end{pmatrix}$$

Le calcul :

$$X^t X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix} \Rightarrow (X^t X)^{-1} = \frac{1}{4 \times 30 - 10 \times 10} \begin{pmatrix} 30 & -10 \\ -10 & 4 \end{pmatrix}$$

$$\begin{aligned}
\beta = \begin{pmatrix} b \\ a \end{pmatrix} &= \frac{1}{20} \begin{pmatrix} 30 & -10 \\ -10 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 2,5 \\ 3,1 \\ 3,4 \end{pmatrix} \\
&= \frac{1}{20} \begin{pmatrix} 30 & -10 \\ -10 & 4 \end{pmatrix} \cdot \begin{pmatrix} 11 \\ 29,9 \end{pmatrix} \\
&= \frac{1}{20} \begin{pmatrix} 31 \\ 9,6 \end{pmatrix} \\
&= \begin{pmatrix} 1,55 \\ 0,48 \end{pmatrix}
\end{aligned}$$

Et on retrouve les valeurs de a et b précédentes.

On peut obtenir $\sum_i \varepsilon_i^2$ par le calcul matriciel :

$$\sum_i \varepsilon_i^2 = (\hat{Y} - Y)^t (\hat{Y} - Y) = n \cdot V_\varepsilon = n \cdot \sigma_\varepsilon^2$$

À la main, ça n'a pas grand intérêt mais certains langages comme Matlab ou Numpy (Python) sont optimisés pour le calcul matriciel. On peut alors tout écrire de façon très courte et efficace.