Chapitre 5

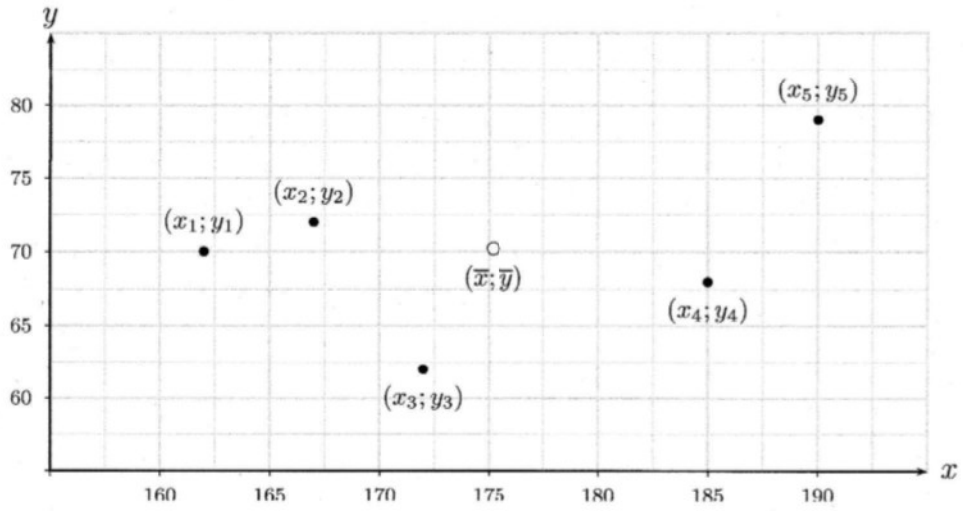
Régression linéaire

Supposons que l'on se donne deux caractéristiques X et Y sur une même population. Par exemple, on pourrait mesurer la taille en centimètres pour X et le poids en kilogrammes pour Y des lycéens de première année. Pour cela, on mesure les caractéristiques X et Y sur un échantillon aléatoire de taille n. On obtient ainsi des observations à deux coordonnées (x_i, y_i) pour $i \in \{1, 2, \ldots, n\}$. Pour cet exemple, prenons $n = 5$.

élève i	1	2	3	4	5
taille en cm x_i	162	167	172	185	190
poids en kg y_i	70	72	62	68	79

moyennes
$\overline{x} = 175.2$
$\overline{y} = 70.2$

La méthode la plus simple pour observer la relation entre X et Y est de représenter ces points dans le plan où l'axe horizontal représente la caractéristique X et l'axe vertical la caractéristique Y. Une telle représentation est appelée un *diagramme de dispersion.*



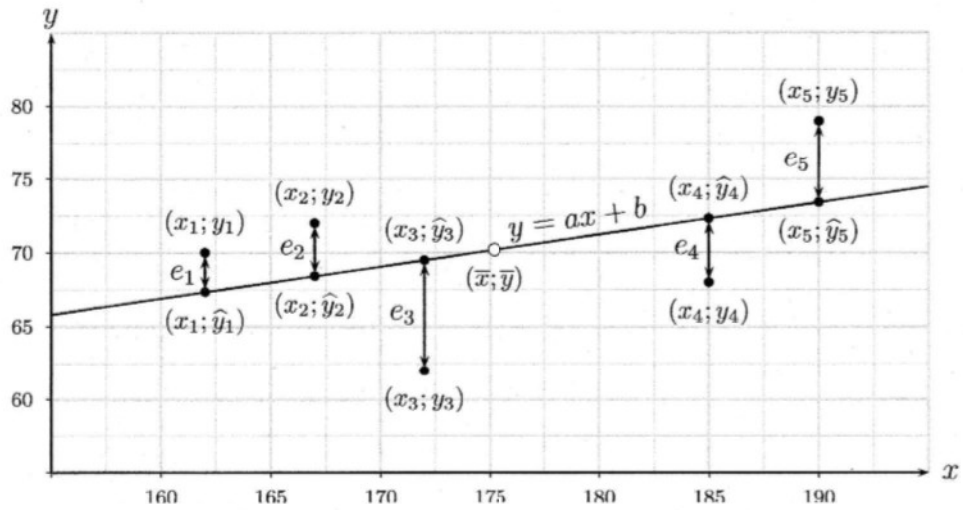
Si la relation entre X et Y est exacte, alors on devrait pouvoir trouver, pour une mesure x_i donnée, l'UNIQUE valeur pour y_i. Ainsi, Y serait une FONCTION de X $(y = f(x))$.

Malheureusement (ou heureusement), il se trouve que dans la plupart des cas, la relation n'est pas exacte (par exemple deux individus de même taille n'ont pas exactement le même poids). Néanmoins, même s'il n'y a pas de relation exacte, il se pourrait qu'il y ait une relation théorique et que, dans chaque mesure, il y ait une part aléatoire.

Dans un tel contexte, on dit que Y est la *variable dépendante*, et X est la *variable indépendante.*

Pour commencer, on va regarder s'il y a une chance pour que la relation (exacte ou non) entre X et Y soit affine (le graphe d'une fonction affine est une droite).

On va donc essayer de faire passer une droite "au mieux" parmi les points (x_i, y_i).



Les notations de la régression linéaire

y_i	y_i est la mesure effective associée à x_i.
$y = ax + b$	Il s'agit du modèle affine théorique entre les caractéristiques X et Y. Il faut déterminer les valeurs des bons paramètres a et b.
$\widehat{y}_i = ax_i + b$	$\widehat{y}_i$ est l'approximation théorique par le modèle affine associée à x_i.
$e_i = y_i - \widehat{y}_i$	e_i est l'erreur entre la mesure effective y_i et son approximation théorique $\widehat{y}_i$ associée à la i-ième mesure. Les e_i sont appelés les *résidus* associés au modèle.

On a ainsi.

$$e_i = y_i - \widehat{y}_i \iff y_i = \widehat{y}_i + e_i \iff y_i = \overbrace{ax_i + b}^{\text{modèle linéaire}} + e_i$$

5.1 La droite des moindres carrés

Dans ce cas le modèle théorique $y = ax + b$ est construit de manière à ce que la somme des carrés des résidus soit la plus petite possible.

Autrement dit, on veut a et b tels que la somme $\sum_{i=1}^{n} e_i^2$ soit minimale.

Pourquoi les carrés

1. L'élévation au carré néglige les signes. Ainsi une erreur négative ne sera pas compensée par une erreur positive.
2. L'élévation au carré réduit les petits écarts (car $(\frac{1}{2})^2 = \frac{1}{4}$; $(\frac{1}{4})^2 = \frac{1}{16}$) et amplifie les grands écarts (car $2^2 = 4$; $4^2 = 16$).

Bien sûr, il existe d'autres méthodes, comme la méthode des moindres valeurs absolues où l'on cherche les paramètres a et b qui minimisent $\sum_{i=1}^{n} |e_i|$. Mais les calculs associés à cette méthode sont plus compliqués.

Un résultat pratique

Les deux sommes suivantes sont égales.

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \sum_{i=1}^{n} x_i y_i \;-\; n\,\overline{x}\,\overline{y}$$

Preuve

On commence par développer la somme de gauche (les sommes ayant toujours les mêmes indices, on les notera juste $\sum$). À la fin du calcul, on utilise $n\overline{x} = \sum x_i$ et $n\overline{y} = \sum y_i$.

$$\begin{aligned}
\sum(x_i-\overline{x})(y_i-\overline{y}) &= \sum(x_iy_i - \overline{x}y_i - x_i\overline{y} + \overline{x}\,\overline{y}) \\
&\overset{\Sigma_1}{=} \sum x_iy_i \;-\; \sum \overline{x}y_i \;-\; \sum x_i\overline{y} \;+\; \sum \overline{x}\,\overline{y} \\
&\overset{\Sigma_2}{=} \sum x_iy_i \;-\; \overline{x}\sum y_i \;-\; \overline{y}\sum x_i \;+\; \overline{x}\,\overline{y}\sum 1 \\
&\overset{\Sigma_3}{=} \sum x_iy_i \;-\; \overline{x}\,n\,\overline{y} \;-\; \overline{y}\,n\,\overline{x} \;+\; n\,\overline{x}\,\overline{y} \\
&= \sum x_iy_i \;-\; n\,\overline{x}\,\overline{y}
\end{aligned}$$

Notations

On va ainsi noter[1] chacune de ces deux sommes σ_{XY}.

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \sigma_{XY} = \sum_{i=1}^{n} x_i y_i \;-\; n\,\overline{x}\,\overline{y}$$

En remplaçant Y par X ou X par Y, on a les formules suivantes.

$$\sum_{i=1}^{n}(x_i-\overline{x})^2 = \sigma_{XX} = \sum_{i=1}^{n} x_i^2 \;-\; n\,\overline{x}^{\,2} \quad \text{et} \quad \sum_{i=1}^{n}(y_i-\overline{y})^2 = \sigma_{YY} = \sum_{i=1}^{n} y_i^2 \;-\; n\,\overline{y}^{\,2}$$

Théorème des moindres carrés

Les valeurs des paramètres a et b pour la droite des moindres carrés $y = ax + b$ sont :

$$\boxed{a = \frac{\sigma_{XY}}{\sigma_{XX}}} \quad \text{et} \quad \boxed{b = \overline{y} - a\,\overline{x}}$$

Ces formules ne sont valables que s'il existe au moins deux x_i qui ont des valeurs différentes (sinon, il y a une division par zéro dans la formule pour a).

Conséquences graphiques

Dans la section 5.6, on montre deux propriétés graphiquement intéressantes.

1. la droite des moindres carrés $y = ax + b$ passe par $(\overline{x};\overline{y})$, qui est le *centre de gravité* des points $(x_i; y_i)$. En effet, on a la relation $\overline{y} = a\overline{x} + b$.
2. la droite des moindres carrés $y = ax + b$ est telle que la somme des résidus est nulle. En effet, la relation $\sum \widehat{y}_i = \sum y_i$ est équivalente à $\sum e_i = 0$.

Ainsi, la droite de régression est agréable à regarder.

---

1. Par rapport aux notations en probabilités, on a $\sigma_{XX} = n\sigma^2(X)$ et $\sigma_{YY} = n\sigma^2(Y)$ où $\sigma(X)$ et $\sigma(Y)$ représentent les écarts types respectifs de X et de Y. De même, $\sigma_{XY} = n\mathrm{Cov}(X,Y)$ où $\mathrm{Cov}(X,Y)$ est la covariance de X et de Y. De plus, si on travaille avec des échantillons, les estimateurs de la variance et de la covariance sont sans biais lorsqu'on remplace n par $(n-1)$.

5.2 Le coefficient de corrélation

On est maintenant capable de faire passer "au mieux" une droite parmi un nuage de points selon la méthode des moindres carrés. Cela ne nous dit toujours pas s'il y a une relation (ne serait-ce que linéaire) entre les caractères X et Y. Pour cela, les mathématiciens ont inventé un outil : il s'agit du *coefficient de corrélation* défini par

$$\rho = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$$

Propriétés de ce coefficient

Le coefficient de corrélation est toujours compris entre -1 et 1.

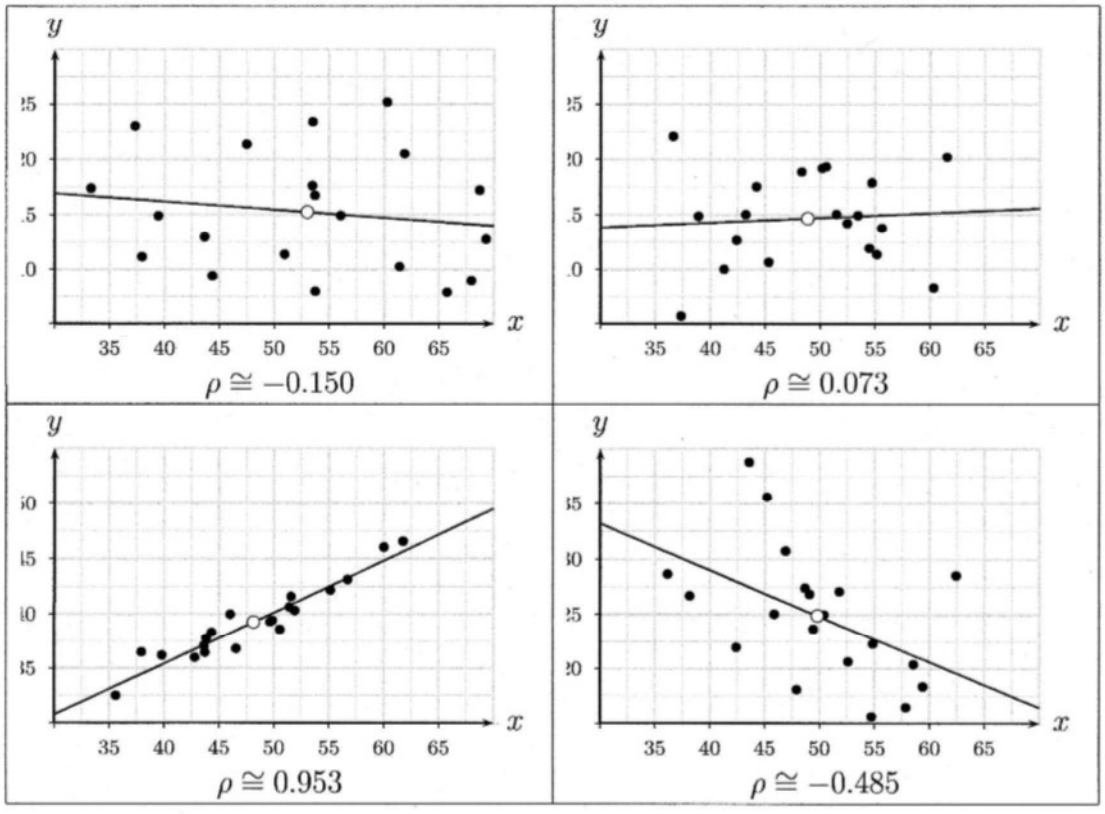
$$-1 \leqslant \rho \leqslant 1$$

C'est un outil qui permet de mesurer si la relation est linéaire, presque linéaire ou très peu linéaire.

- Lorsque ρ est proche de 0, alors la relation n'est pas linéaire. Soit les variables sont indépendantes, soit il y a un autre type de relation (voir page 72).
- Plus ρ est proche de 1, plus les points sont proches de la droite des moindres carrés (qui sera de pente positive).
- Plus ρ est proche de -1, plus les points sont proches de la droite des moindres carrés (qui sera de pente négative).

Moralité Plus $|\rho|$ est proche de 1, meilleure est l'approximation par la droite des moindres carrés. On dit alors que X et Y sont *corrélés* (ou *linéairement dépendants*).

Exemples



5.3 Le coefficient de détermination

Relation évidente

Les valeurs y_i et $\overline{y}$ proviennent directement des données. Les valeurs $\widehat{y}_i$ sont obtenues à partir du modèle. Ces trois valeurs sont liées par la relation suivante.

$$\overbrace{\underbrace{(y_i - \widehat{y}_i)}_{\text{résidu}} + (\widehat{y}_i - \overline{y})}^{\text{dépend aussi du modèle}} = \overbrace{(y_i - \overline{y})}^{\text{ne dépend que des données}}$$

Relation «miraculeuse»

Cette relation est vraie lorsqu'on utilise le modèle de la droite des moindres carrés.

$$\underbrace{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}_{\substack{\text{variation due} \\ \text{aux résidus}}} + \underbrace{\sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2}_{\substack{\text{variation due} \\ \text{au modèle}}} = \underbrace{\sum_{i=1}^{n} (y_i - \overline{y})^2}_{\substack{\text{variation totale} \\ \text{(car somme des} \\ \text{deux autres)}}}$$

Définition du coefficient de détermination

Le coefficient de détermination, noté R^2, est défini par

$$R^2 = \frac{\sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

Propriétés de ce coefficient

Lorsque la relation «miraculeuse» est vraie, le coefficient de détermination détermine le rapport entre la variation due au modèle et la variation totale. C'est donc le pourcentage de la variation due au modèle dans la variation totale.

Pour cette raison, le coefficient de détermination est toujours compris entre 0 et 1.

$$0 \leqslant R^2 \leqslant 1$$

- Lorsque R^2 est proche de 0, près du 100% de la variation totale est expliquée par la variation due aux résidus. Cela signifie que le modèle n'est pas adapté aux données.
- Lorsque R^2 est proche de 1, près du 100% de la variation totale est expliquée par la variation due au modèle. Cela signifie que le modèle est bien adapté aux données.

Théorème de retrouvailles

Lorsque le modèle est celui de la droite des moindres carrés, le coefficient de détermination est égal au carré du coefficient de corrélation. Autrement dit

$$R^2 = \rho^2$$

5.4 La droite des moindres carrés forcée à l'origine

Dans ce cas le modèle théorique $y = ax + b$ est construit de manière à ce que

1. la droite passe par l'origine du plan ;
2. la somme des carrés des résidus soit la plus petite possible.

Autrement dit, on veut a et b tels que

1. $b = 0$, ainsi le modèle est $y = ax$.
2. la somme $\sum_{i=1}^{n} e_i^2$ soit minimale.

Théorème des moindres carrés pour la version forcée à l'origine

La valeur du paramètre a pour la droite des moindres carrés $y = ax$ est :

$$\boxed{a = \frac{\sigma_{XY}^{(0)}}{\sigma_{XX}^{(0)}}} \qquad \text{où} \quad \sigma_{XY}^{(0)} = \sum_{i=1}^{n} x_i y_i \quad \text{et} \quad \sigma_{XX}^{(0)} = \sum_{i=1}^{n} x_i^2$$

Cette formule n'est valable que s'il existe au moins deux x_i qui ont des valeurs différentes (sinon, il y a une division par zéro dans la formule pour a).

Ennuis

1. la droite des moindres carrés $y = ax$ ne passe pas forcément par $(\overline{x}; \overline{y})$, qui est le *centre de gravité* des points $(x_i; y_i)$.
2. la droite des moindres carrés $y = ax$ ne vérifie pas forcément la relation $\sum \widehat{y}_i = \sum y_i$, et ainsi la somme des résidus $\sum e_i$ n'est pas forcément nulle.

Ainsi, à l'œil, cette droite peut paraître un peu bizarre.

Exemple

Lors d'une expérience d'osmose en biologie, un arbre chimique grandit dans une solution à 30% de saccharose. On cherche à déterminer la vitesse moyenne de croissance de l'arbre chimique durant les 15 premières minutes qui sera la pente de la droite des moindres carrés. Lors des mesures, des élèves ont obtenus les nombres suivants.

temps (min)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
hauteur (cm)	0.0	1.5	2.9	4.3	5.5	6.6	7.8	9.0	10.3	11.4	12.5	13.7	14.7	15.8	16.8	17.7

On trouve

$$a \cong 1.23 \text{ cm} \cdot \text{min}^{-1}$$

Les relations des pages 68 et 69 ne sont pas toujours vraies dans le cas du modèle $y = ax$. L'exemple précédent infirme chacune de ces formules.

Néanmoins, on peut les retrouver si, dans ces relations, on remplace $\overline{x}$ et $\overline{y}$ par 0, comme on le voit à la page suivante. Ce qui explique la notation avec les exposants $^{(0)}$.

Bien évidemment, si le centre de gravité $(\overline{x}; \overline{y})$ est l'origine du plan, les droites des moindres carrés $y = ax + b$ et le modèle qui force la droite à l'origine sont les mêmes.

5.4.1 Les coefficients de détermination et de corrélation

Relation évidente

$$\overbrace{\underbrace{(y_i - \widehat{y}_i)}_{\text{résidu}} + \widehat{y}_i}^{\text{dépend aussi du modèle}} = \overbrace{y_i}^{\text{ne dépend que des données}}$$

Relation «miraculeuse»

$$\underbrace{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}_{\text{variation due aux résidus}} + \underbrace{\sum_{i=1}^{n} \widehat{y}_i^2}_{\text{variation due au modèle par rapport à 0}} = \underbrace{\sum_{i=1}^{n} y_i^2}_{\text{variation totale (car somme des deux autres)}}$$

Les coefficients de détermination et de corrélations

$$R^2 = \frac{\sum_{i=1}^{n} \widehat{y}_i^2}{\sum_{i=1}^{n} y_i^2} \quad \text{et} \quad \rho = \frac{\sigma_{XY}^{(0)}}{\sqrt{\sigma_{XX}^{(0)}}\sqrt{\sigma_{YY}^{(0)}}} \quad \text{où} \quad \begin{aligned} \sigma_{XY}^{(0)} &= \sum_{i=1}^{n} x_i y_i \\ \sigma_{XX}^{(0)} &= \sum_{i=1}^{n} x_i^2 \\ \sigma_{YY}^{(0)} &= \sum_{i=1}^{n} y_i^2 \end{aligned}$$

Théorème de retrouvailles

Même dans ce modèle où la droite des moindres carrés est forcée à l'origine, on a

$$R^2 = \rho^2$$

5.4.2 Preuves

Même si on a perdu l'ingrédient $\widehat{y}_i = \sum y_i$. On conserve l'ingrédient $\sum \widehat{y}_i^2 = \sum \widehat{y}_i y_i$. En effet

$$\sum_{i=1}^{n} \widehat{y}_i^2 = \sum_i (a x_i)^2 = a^2 \sum_i x_i^2 = \left(\frac{\sum_i x_i y_i}{\sum_i x_i^2}\right)^2 \sum_i x_i^2 = \frac{\left(\sum_k x_k y_k\right)\left(\sum_i x_i y_i\right)}{\sum_k x_k^2}$$

$$= \sum_i \left(\frac{\sum_k x_k y_k}{\sum_k x_k^2} x_i y_i\right) = \sum_i \big(\underbrace{a x_i}_{\widehat{y}_i} y_i\big) = \sum_{i=1}^{n} \widehat{y}_i y_i$$

Preuve de la relation «miraculeuse»

On peut maintenant prouver la relation «miraculeuse».

$$\sum (y_i - \widehat{y}_i)^2 + \sum \widehat{y}_i^2 = \sum y_i^2 - 2\sum y_i \widehat{y}_i + \sum \widehat{y}_i^2 + \sum \widehat{y}_i^2$$

$$\overset{\text{ingrédient}}{=} \sum y_i^2 - 2\sum \widehat{y}_i^2 + \sum \widehat{y}_i^2 + \sum \widehat{y}_i^2 = \sum y_i^2$$

Preuve du théorème de retrouvailles

$$R^2 = \frac{\sum \widehat{y}_i^2}{\sum y_i^2} = \frac{\sum (a x_i)^2}{\sum y_i^2} = a^2 \cdot \frac{\sum x_i^2}{\sum y_i^2} = \left(\frac{\sigma_{XY}^{(0)}}{\sigma_{XX}^{(0)}}\right)^2 \cdot \frac{\sigma_{XX}^{(0)}}{\sigma_{YY}^{(0)}} = \frac{{\sigma_{XY}^{(0)}}^2}{\sigma_{XX}^{(0)} \sigma_{YY}^{(0)}} = \rho^2$$

5.5 Autres types de régression

5.5.1 Préambule : une autre vision de la régression linéaire

On suppose que les données sont liées par une relation linéaire, non exacte, de la façon suivante.

$$y_i = ax_i + b + e_i$$

En suivant la méthode des moindres carrés, on trouve le minimum de $\sum_{i=1}^{n} e_i^2$ en annulant le gradient [2] : cela revient à résoudre le système suivant d'inconnues a et b.

$$\begin{cases} \sum_{i=1}^{n} y_i x_i &= a \sum_{i=1}^{n} x_i^2 &+ b \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i &= a \sum_{i=1}^{n} x_i &+ b\, n \end{cases}$$

5.5.2 Régression quadratique

On suppose que les données sont liées par une relation quadratique, non exacte, de la façon suivante.

$$y_i = ax_i^2 + bx_i + c + e_i$$

En suivant la méthode des moindres carrés, on trouve le minimum de $\sum_{i=1}^{n} e_i^2$ en annulant le gradient[2] : cela revient à résoudre le système suivant d'inconnues a, b et c.

$$\begin{cases} \sum_{i=1}^{n} y_i x_i^2 &= a \sum_{i=1}^{n} x_i^4 &+ b \sum_{i=1}^{n} x_i^3 &+ c \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} y_i x_i &= a \sum_{i=1}^{n} x_i^3 &+ b \sum_{i=1}^{n} x_i^2 &+ c \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i &= a \sum_{i=1}^{n} x_i^2 &+ b \sum_{i=1}^{n} x_i &+ c\, n \end{cases}$$

5.5.3 Régression hyperbolique

On suppose que les données sont liées par une relation hyperbolique, non exacte, de la façon suivante.

$$y_i = \frac{1}{ax_i + b + e_i} \iff \frac{1}{y_i} = ax_i + b + e_i$$

La méthode des moindres carrés peut s'appliquer [3] en utilisant le changement de variable suivant :

$$z = \frac{1}{y} \iff y = \frac{1}{z} \quad \text{et} \quad z_i = \frac{1}{y_i} \iff y_i = \frac{1}{z_i}$$

On trouve que :

$$y = \frac{1}{ax + b} \quad \left(\iff z = ax + b \right) \qquad \text{où} \quad \boxed{a = \frac{\sigma_{XZ}}{\sigma_{XX}}} \quad \text{et} \quad \boxed{b = \overline{z} - a\,\overline{x}}$$

2. Le gradient est une notion vue à l'université qui ne peut être expliquée qu'en deuxième année de lycée (il faut savoir dériver).

3. La *vraie* méthode des moindres carrés consisterait à minimiser la somme des carrés des ε_i donnés par la relation $y_i = \frac{1}{ax_i+b} + \varepsilon_i$. Néanmoins les calculs sont ici très complexes.

5.5.4 Régression exponentielle

On suppose que les données sont liées par une relation exponentielle, non exacte, de la façon suivante.

$$y_i = b \cdot a^{x_i} \cdot 10^{e_i} \iff \log(y_i) = \log(a)x_i + \log(b) + e_i$$

La méthode des moindres carrés peut s'appliquer[4] en utilisant le changement de variable suivant :

$$w = \log(y) \iff y = 10^w \qquad \text{et} \qquad w_i = \log(y_i) \iff y_i = 10^{w_i}$$

On trouve que :

$$\Big(\begin{array}{c} y = b \cdot a^x \\ \iff w = \log(a)x + \log(b) \end{array} \Big) \text{ où } \begin{array}{l} \log(a) = \dfrac{\sigma_{XW}}{\sigma_{XX}} \\ \Leftrightarrow \boxed{a = \exp_{10}\left(\dfrac{\sigma_{XW}}{\sigma_{XX}}\right)} \end{array} \text{ et } \begin{array}{l} \log(b) = \overline{w} - a\,\overline{x} \\ \Leftrightarrow \boxed{b = 10^{\overline{w} - a\,\overline{x}}} \end{array}$$

5.5.5 Régression d'une puissance

On suppose que les données sont liées par une puissance, non exacte, de la façon suivante.

$$y_i = b \cdot x_i^a \cdot 10^{e_i} \iff \log(y_i) = a\log(x_i) + \log(b) + e_i$$

La méthode des moindres carrés peut s'appliquer[5] en utilisant le changement de variable suivant :

$$\begin{array}{lll} w = \log(y) \iff y = 10^w & \text{et} & w_i = \log(y_i) \iff y_i = 10^{w_i} \\ & \text{et} & \\ v = \log(x) \iff x = 10^v & \text{et} & v_i = \log(x_i) \iff x_i = 10^{v_i} \end{array}$$

On trouve que :

$$\Big(\begin{array}{c} y = b \cdot x^a \\ \iff w = av + \log(b) \end{array} \Big) \text{ où } \boxed{a = \frac{\sigma_{VW}}{\sigma_{VV}}} \text{ et } \begin{array}{l} \log(b) = \overline{w} - a\,\overline{v} \\ \iff \boxed{b = 10^{\overline{w} - a\,\overline{v}}} \end{array}$$

5.5.6 Régression logarithmique

On suppose que les données sont liées par une relation logarithmique, non exacte, de la façon suivante.

$$y_i = a\log(x_i) + b + e_i$$

Cette fois la *vraie* méthode des moindres carrés peut s'appliquer en utilisant le changement de variable suivant :

$$v = \log(x) \iff x = 10^v \qquad \text{et} \qquad v_i = \log(x_i) \iff x_i = 10^{v_i}$$

On trouve que : $\Big(\begin{array}{c} y = a\log(x_i) + b \\ \iff y = av + b \end{array} \Big)$ où $\boxed{a = \dfrac{\sigma_{VY}}{\sigma_{VV}}}$ et $\boxed{b = \overline{y} - a\,\overline{v}}$

4. La *vraie* méthode des moindres carrés consisterait à minimiser la somme des carrés des ε_i donnés par la relation $y_i = b \cdot a^{x_i} + \varepsilon_i$. Néanmoins les calculs sont ici très complexes.

5. La *vraie* méthode des moindres carrés consisterait à minimiser la somme des carrés des ε_i donnés par la relation $y_i = b \cdot x_i^a + \varepsilon_i$. Néanmoins les calculs sont ici très complexes.

5.6 Preuves des théorèmes

5.6.1 Preuve des théorèmes des moindres carrés

Rappel sur les paraboles

Une parabole d'expression fonctionnelle $p(x) = \alpha x^2 + \beta x + \gamma$ avec $\alpha > 0$ a un minimum pour $x = -\frac{\beta}{2\alpha}$.

Preuve du théorème sur les moindres carrés

Trouvons une valeur de b tel que la somme des résidus au carré soit minimale. En d'autres termes, on veut trouver le minimum de l'expression suivante.

$$\sum_{i=1}^{n} e_i^2$$

En utilisant le fait que $\widehat{y}_i = ax_i + b$, on peut rendre cette somme dépendante du paramètre b. C'est pourquoi, cette somme est momentanément appelée $S(b)$.

$$\begin{aligned}
S(b) &= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2 \\
&= \sum_{i=1}^{n} \big((y_i - ax_i) - b\big)^2 = \sum_{i=1}^{n} \big((y_i - ax_i)^2 - 2(y_i - ax_i)b + b^2\big) \\
&= \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + \sum_{i=1}^{n} b^2 \\
&= \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \left(\sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i \right) + \sum_{i=1}^{n} b^2
\end{aligned}$$

Or $\sum_{i=1}^{n} y_i = n\,\overline{y}$ et $\sum_{i=1}^{n} x_i = n\,\overline{x}$, ainsi on a

$$\begin{aligned}
S(b) &= \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b\big(n\,\overline{y} - a\,n\,\overline{x}\big) + nb^2 \\
&= \underbrace{n}_{\alpha>0} b^2 \; \underbrace{-2n(\overline{y} - a\,\overline{x})}_{\beta} b \; + \; \underbrace{\sum_{i=1}^{n} (y_i - ax_i)^2}_{\gamma}
\end{aligned}$$

Par le rappel ci-dessus, la valeur de b qui minimise $S(b)$ est donnée par

$$b = \frac{2n(\overline{y} - a\,\overline{x})}{2n} = \overline{y} - a\,\overline{x}$$

Maintenant qu'on a établi la relation $b = \overline{y} - a\,\overline{x}$, l'expression de la droite des moindres carrés est ainsi devenue

$$y = ax + b = ax + \overline{y} - a\,\overline{x} = a(x - \overline{x}) + \overline{y}$$

Il faut maintenant trouver a tel que la somme des résidus au carré soit minimale. En d'autres termes, on veut trouver le minimum de l'expression suivante.

$$\sum_{i=1}^{n} e_i^2$$

En utilisant le fait que $\widehat{y}_i = ax_i + b = a(x_i - \overline{x}) + \overline{y}$, on peut rendre cette somme uniquement dépendante du paramètre a. C'est pourquoi, on a décidé d'appeler la somme $S(a)$.

$$S(a) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} \Big(y_i - \big(a(x_i - \overline{x}) + \overline{y}\big)\Big)^2 = \sum_{i=1}^{n} \Big(y_i - a(x_i - \overline{x}) - \overline{y}\Big)^2$$

On cherche à trouver a tel que $S(a)$ est le plus petit possible. On sait que $S(a) \geqslant 0$ (car une somme de nombres positifs (ou nuls) ne peut être que positive (ou nulle)).
On a donc

$$\begin{aligned}
S(a) &= \sum_{i=1}^{n} \Big(y_i - a(x_i - \overline{x}) - \overline{y}\Big)^2 = \sum_{i=1}^{n} \Big((y_i - \overline{y}) - a(x_i - \overline{x})\Big)^2 \\
&= \sum_{i=1}^{n} \Big((y_i - \overline{y})^2 - 2a(x_i - \overline{x})(y_i - \overline{y}) + a^2(x_i - \overline{x})^2\Big) \\
&= \underbrace{\sum_{i=1}^{n} (y_i - \overline{y})^2}_{\sigma_{YY}} - 2a \underbrace{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}_{\sigma_{XY}} + a^2 \underbrace{\sum_{i=1}^{n} (x_i - \overline{x})^2}_{\sigma_{XX}} \\
&= \underbrace{\sigma_{XX}}_{>0} a^2 - 2\sigma_{XY} a + \sigma_{YY}
\end{aligned}$$

Par le rappel ci-dessus, la valeur de a qui minimise $S(a)$ est donnée par

$$a = \frac{2\sigma_{XY}}{2\sigma_{XX}} = \frac{\sigma_{XY}}{\sigma_{XX}}$$

Mais, il faut que $\sigma_{XX} > 0$ afin d'avoir un minimum. C'est le cas s'il y a au moins deux x_i qui sont différents. □

Preuve pour la version forcée à l'origine

Trouvons une valeur de a tel que la somme des résidus au carré soit minimale. En d'autres termes, on veut trouver le minimum de l'expression suivante.

$$\begin{aligned}
S(a) &= \sum_{i=1}^{n} \Big(y_i - ax_i\Big)^2 = \sum_{i=1}^{n} \Big(y_i^2 - 2ax_iy_i + a^2x_i^2\Big) = \underbrace{\sum_{i=1}^{n} y_i^2}_{\sigma_{YY}^{(0)}} - 2a \underbrace{\sum_{i=1}^{n} x_iy_i}_{\sigma_{XY}^{(0)}} + a^2 \underbrace{\sum_{i=1}^{n} x_i^2}_{\sigma_{XX}^{(0)}} \\
&= \underbrace{\sigma_{XX}^{(0)}}_{>0} a^2 - 2\sigma_{XY}^{(0)} a + \sigma_{YY}^{(0)}
\end{aligned}$$

Par le rappel ci-dessus, la valeur de a qui minimise $S(a)$ est donnée par

$$a = \frac{2\sigma_{XY}^{(0)}}{2\sigma_{XX}^{(0)}} = \frac{\sigma_{XY}^{(0)}}{\sigma_{XX}^{(0)}}$$

Mais, il faut que $\sigma_{XX}^{(0)} > 0$ afin d'avoir un minimum. C'est le cas s'il y a au moins deux x_i qui sont différents. □

5.6.2 Preuves de la relation «miraculeuse»

On a besoin de deux **ingrédients**.

$$\boxed{\sum \widehat{y}_i = \sum y_i \quad \text{et} \quad \sum \widehat{y}_i^2 = \sum \widehat{y}_i y_i} \ \bigstar$$

Si le modèle vérifie ces deux ingrédients, alors la relation «miraculeuse» est vraie.

En effet, on a

$$\begin{aligned}
& \sum (y_i - \widehat{y}_i)^2 \ + \ \sum (\widehat{y}_i - \overline{y})^2 \\
= \ & \sum (y_i^2 - 2\widehat{y}_i y_i + \widehat{y}_i^2) \ + \ \sum (\widehat{y}_i^2 - 2\widehat{y}_i \overline{y} + \overline{y}^2) \\
\overset{\Sigma_1,\Sigma_2,\Sigma_3}{=} \ & \sum y_i^2 - 2\underbrace{\sum \widehat{y}_i y_i} + \sum \widehat{y}_i^2 \ + \ \sum \widehat{y}_i^2 - 2\overline{y}\underbrace{\sum \widehat{y}_i} + n\overline{y}^2 \\
\overset{\bigstar}{=} \ & \sum y_i^2 - 2\sum \widehat{y}_i^2 + \sum \widehat{y}_i^2 \ + \ \sum \widehat{y}_i^2 - 2\overline{y}\sum y_i + n\overline{y}^2 \\
\overset{\sum y_i = n\overline{y}}{=} \ & \sum y_i^2 - 2\sum \widehat{y}_i^2 + \sum \widehat{y}_i^2 \ + \ \sum \widehat{y}_i^2 - 2n\overline{y}^2 + n\overline{y}^2 \\
= \ & \sum y_i^2 - \sum \widehat{y}_i^2 \ + \ \sum \widehat{y}_i^2 - n\overline{y}^2 \\
= \ & \sum y_i^2 \ - \ n\overline{y}^2 \\
\overset{\text{notation } \sigma_{YY}}{\underset{\text{page 67}}{=}} \ & \sum (y_i - \overline{y})^2
\end{aligned}$$

5.6.3 Les deux visions pour la droite de régression

À la page 72, on affirme que a et b satisfont le système suivant.

$$\left\{ \begin{array}{rcl} \sum y_i x_i & = & a\sum x_i^2 \ + \ b\sum x_i \\ \sum y_i & = & a\sum x_i \ + \ b\,n \end{array} \right. \iff \left\{ \begin{array}{rcl} \sum y_i x_i & = & a\sum x_i^2 \ + \ b\,n\,\overline{x} \\ n\overline{y} & = & a\,n\,\overline{x} \ + \ b\,n \end{array} \right.$$

À la page 67, on a donné les valeurs suivantes de a et b.

$$a = \frac{\sigma_{XY}}{\sigma_{XX}} \quad \text{et} \quad b = \overline{y} - a\,\overline{x}$$

On retrouve ces coefficients en résolvant le système d'équation. En effet, la deuxième ligne est équivalente à

$$\overline{y} = a\overline{x} + b \iff b = \overline{y} - a\overline{x}$$

De plus si, à la première ligne, on soustrait $\overline{x}$ fois la deuxième, cette première ligne devient

$$\begin{aligned}
& \sum y_i x_i - n\,\overline{x}\,\overline{y} = a\sum x_i^2 - a\,n\,\overline{x}^2 \iff \sum y_i x_i - n\,\overline{x}\,\overline{y} = a\Big(\sum x_i^2 - n\,\overline{x}^2\Big) \\
& \iff \sigma_{XY} = a\,\sigma_{XX} \iff a = \frac{\sigma_{XY}}{\sigma_{XX}}
\end{aligned}$$

5.6.4 Preuve des ingrédients pour le modèle linéaire

On se rappelle que a et b sont solutions du système

$$(\bigstar): \begin{cases} \sum y_i x_i &= a\sum x_i^2 &+ b\sum x_i \\ \sum y_i &= a\sum x_i &+ b\,n \end{cases}$$

Preuve du premier ingrédient

On utilise le fait que $\widehat{y}_i = ax_i + b$, on développe et on observe le système.

$$\sum \widehat{y}_i = \sum \big(ax_i + b\big) = a\sum x_i + b\,n \overset{(\bigstar)}{=} \sum y_i$$

Preuve du deuxième ingrédient

On utilise le fait que $\widehat{y}_i = ax_i + b$, on développe et on observe le système.

$$\begin{aligned} \sum \widehat{y}_i^2 &= \sum \big(ax_i + b\big)^2 = \sum \big(a^2x_i^2 + 2abx_i + b^2\big) = a^2\sum x_i^2 + 2ab\sum x_i + b^2 n \\ &= a\big(a\sum x_i^2 + b\sum x_i\big) + ab\sum x_i + b^2 n \\ &= a\big(a\sum x_i^2 + b\sum x_i\big) + b\big(a\sum x_i + b\,n\big) \\ &\overset{(\bigstar)}{=} a\sum y_i x_i + b\sum y_i \\ &= \sum y_i(ax_i + b) \\ &= \sum y_i\widehat{y}_i = \sum \widehat{y}_i y_i \end{aligned}$$

5.6.5 Preuve du théorème de retrouvailles

On se rappelle que et $a = \dfrac{\sigma_{XY}}{\sigma_{XX}}$ et $b = \overline{y} - a\,\overline{x}$.

$$\begin{aligned} R^2 &= \frac{\sum \big(\widehat{y}_i - \overline{y}\big)^2}{\sum \big(y_i - \overline{y}\big)^2} = \frac{\sum \big(ax_i + b - \overline{y}\big)^2}{\sum \big(y_i - \overline{y}\big)^2} = \frac{\sum \big(ax_i + \overline{y} - a\,\overline{x} - \overline{y}\big)^2}{\sum \big(y_i - \overline{y}\big)^2} \\ &= \frac{\sum \big(ax_i - a\,\overline{x}\big)^2}{\sum \big(y_i - \overline{y}\big)^2} = a^2 \cdot \frac{\sum \big(x_i - \overline{x}\big)^2}{\sum \big(y_i - \overline{y}\big)^2} = \left(\frac{\sigma_{XY}}{\sigma_{XX}}\right)^2 \cdot \frac{\sigma_{XX}}{\sigma_{YY}} = \frac{\sigma_{XY}^2}{\sigma_{XX}^2} \cdot \frac{\sigma_{XX}}{\sigma_{YY}} \\ &= \frac{\sigma_{XY}^2}{\sigma_{XX}\sigma_{YY}} = \left(\frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}\right)^2 = \rho^2 \end{aligned}$$

5.6.6 Preuve des ingrédients pour le modèle quadratique

On se rappelle que a, b et c sont solutions du système

$$(\bigstar): \begin{cases} \sum y_i x_i^2 &= a\sum x_i^4 &+ b\sum x_i^3 &+ c\sum x_i^2 \\ \sum y_i x_i &= a\sum x_i^3 &+ b\sum x_i^2 &+ c\sum x_i \\ \sum y_i &= a\sum x_i^2 &+ b\sum x_i &+ c\,n \end{cases}$$

Preuve du premier ingrédient

On utilise le fait que $\widehat{y}_i = ax_i^2 + bx_i + c$, on développe et on observe le système.

$$\sum \widehat{y}_i = \sum \left(ax_i^2 + bx_i + c\right) = a\sum x_i^2 + b\sum x_i + c\,n \overset{(\bigstar)}{=} \sum y_i$$

Preuve du deuxième ingrédient

On utilise le fait que $\widehat{y}_i = ax_i^2 + bx_i + c$, on développe et on observe le système.

$$\begin{aligned}
\sum \widehat{y}_i^2 &= \sum \left(ax_i^2 + bx_i + c\right)^2 \\
&= \sum \left(a^2x_i^4 + 2abx_i^3 + b^2x_i^2 + 2acx_i^2 + 2bcx_i + c^2\right) \\
&= a^2\sum x_i^4 + 2ab\sum x_i^3 + b^2\sum x_i^2 + 2ac\sum x_i^2 + 2bc\sum x_i + c^2 n \\
&= a\big(a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2\big) \\
&\quad + ab\sum x_i^3 + b^2\sum x_i^2 + ac\sum x_i^2 + 2bc\sum x_i + c^2 n \\
&= a\big(a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2\big) \\
&\quad + b\big(a\sum x_i^3 + b\sum x_i^2 + c\sum x_i\big) \\
&\quad + ac\sum x_i^2 + bc\sum x_i + c^2 n \\
&= a\big(a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2\big) \\
&\quad + b\big(a\sum x_i^3 + b\sum x_i^2 + c\sum x_i\big) \\
&\quad + c\big(c\sum x_i^2 + b\sum x_i + c\,n\big) \\
&\overset{(\bigstar)}{=} a\sum y_i x_i^2 + b\sum y_i x_i + c\sum y_i \\
&= \sum y_i(ax_i^2 + bx_i + c) \\
&= \sum y_i \widehat{y}_i \;=\; \sum \widehat{y}_i y_i
\end{aligned}$$