Modèle de la droite de regression

Cadre pratique On pense qu'une quantité X dépend de manière affine d'une quantité Y, on mesure donc n fois la valeur de X et la valeur de Y associée, on obtient un nuage de n points (Y_i, X_i).

24.1 **Modèle** $\boxed{\tilde{X} = [\mathbb{1}\ Y]\,\Theta \quad \text{où } Y = (Y_i)_{i \in [1..n]} \text{ et } \Theta \in \mathbb{R}^2}$

24.2 **Régularité** $\tilde{X} = [\mathbb{1}\, Y]\,\Theta$ est régulier $\Leftrightarrow \exists (i,j) \in [1..n]^2\ Y_i \neq Y_j$ (a)

$\Leftrightarrow \mathrm{Var}(Y) > 0$ (b)

(a) Si $\forall (i,j) \in [1..n]^2\ Y_i = Y_j$ alors $Y = Y_i \mathbb{1}$ donc $rg([\mathbb{1}, \Theta]) = 1 < 2$ dc non rég.

Si $\exists (i,j) \in [1..n]^2\ Y_i \neq Y_j$. Soit $(\alpha, \beta) \in \mathbb{R}^2$ tq $\alpha Y + \beta \mathbb{1} = 0$. $\begin{matrix} \alpha Y_i = -\beta \\ \alpha Y_j = -\beta \end{matrix}$ donc $\alpha(Y_i - Y_j) = 0$

donc $\alpha = 0$ donc $0 + \beta \mathbb{1} = 0$ donc $\beta = 0$. Aut. dit $(\mathbb{1}, Y)$ est libre et $rg([\mathbb{1}, Y]) = 2$.

(b) $\mathrm{Var}(Y) = 0 \Leftrightarrow \forall i \in [1..n]\ Y_i = Y_j \Leftrightarrow$ Modèle non rég. De plus $\mathrm{Var}(Y) \geq 0$ donc

$\mathrm{Var}(Y) > 0 \Leftrightarrow \mathrm{Var}(Y) \neq 0 \Leftrightarrow$ modèle régulier.

Dans la suite on note $A = [\mathbb{1}\, Y]$ et on suppose que le modèle est régulier.

24.3 **Projection** $\mathrm{Im}(A) = \mathrm{Vect}\{\mathbb{1}, Y\} = \mathrm{Vect}\{\mathbb{1}, Y - P_{\mathbb{1}}^{\perp}(Y)\} = \mathrm{Vect}\{\mathbb{1}, Y - m(Y)\mathbb{1}\}$.

En fait on se ramène à une base orthogonale de $\mathrm{Im}\, A$.

Posons $B = [\mathbb{1}, Y - m(Y)\mathbb{1}]$. $P^{\perp}_{\mathrm{Im}(A)} = P^{\perp}_{\mathrm{Im}(B)}$ et puisque les colonnes de B sont orthogonales libres, d'après 22

$$P^{\perp}_{\mathrm{Im}A} = \frac{\mathbb{1}\mathbb{1}'}{\mathbb{1}'\mathbb{1}} + \frac{(Y - m(Y)\mathbb{1})(Y - m(Y)\mathbb{1})'}{(Y - m(Y)\mathbb{1})'(Y - m(Y)\mathbb{1})} = \underline{\underline{\frac{\mathbb{1}\mathbb{1}'}{n} + \frac{(Y - m(Y)\mathbb{1})(Y - m(Y)\mathbb{1})'}{n\,\mathrm{Var}(Y)}}}$$

24.4 **Moindres carrés** La solution au problème des moindres carrés pour ce modèle est donnée par $\hat{\Theta} = \begin{pmatrix} \hat{b} \\ \hat{a} \end{pmatrix}$ où $\boxed{\hat{a} = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)} \qquad \hat{b} = \bar{X} - \hat{a}\bar{Y}}$

$\hat{X} = \hat{a} Y + \hat{b}\mathbb{1}$ induit l'équation $\mathcal{D}: y = \hat{a}x + \hat{b}$ de la droite de regression du nuage de points.

Puisque $rg(A) = 2$ (modèle rég), d'après 21.1 $P^{\perp}_{\operatorname{Im} A} = A(A'A)^{-1}A'$. Avec 24.3

on a donc $A(A'A)^{-1}A' = \frac{\mathbb{1}\mathbb{1}'}{n} + \frac{(Y-m(Y)\mathbb{1})(Y-m(Y)\mathbb{1})'}{n\, var(Y)}$

$$= \mathbb{1}\left[\frac{\mathbb{1}'}{n} - \frac{m(Y)}{n\,var(Y)}(Y-m(Y)\mathbb{1})'\right] + Y\left[\frac{(Y-m(Y)\mathbb{1})'}{n\,var(Y)}\right]$$

$$= \underbrace{[\mathbb{1}, Y]}_{A}\ \underbrace{\begin{bmatrix} \frac{\mathbb{1}'}{n} - \frac{m(Y)}{n\,var(Y)}(Y-m(Y)\mathbb{1})' \\ \frac{(Y-m(Y)\mathbb{1})'}{n\,var(Y)} \end{bmatrix}}_{C}$$

$\forall U \in \mathbb{R}^2\ A((A'A)^{-1}A')U = ACU$ or A est injective dc $\forall U \in \mathbb{R}^2\ ((A'A)^{-1}A')U = CU$

d'où $(A'A)^{-1}A' = C$. D'après 23.3 on a alors $\hat\Theta = BX$ soit

$\hat a = \frac{(Y-m(Y)\mathbb{1})'}{n\,var(Y)} X$ or $c(X,Y) = \frac{1}{n}(Y-m(Y)\mathbb{1})'(X-m(X)\mathbb{1})$

$$= \frac{1}{n}\left((Y-m(Y)\mathbb{1})'X - m(X)\underbrace{(Y-m(Y)\mathbb{1})'\mathbb{1}}_{=0 \text{ car } Y-m(Y)\mathbb{1} \text{ centré}}\right)$$

$$= \hat a\, var(Y)$$

$\hat b = \frac{\mathbb{1}'X}{n} - m(Y)\left(\frac{(Y-m(Y)\mathbb{1})'X}{n\,var(Y)}\right) = m(X) - m(Y)\hat a$

<u>Remarque</u> La droite de régression passe par le barycentre du nuage de points.

G = barycentre du nuage a pour coordonnées $(\underbrace{m(Y)}_{x_G}, \underbrace{m(X)}_{y_G})$.

donc $\hat a x_G + \hat b = \hat a m(Y) + (m(X) - \cancel{m(Y)\hat a}) = m(X) = x_G$

d'où $G \in \mathcal{D}$

<u>Coefficient de détermination</u>

$\boxed{R^2 = \rho^2(X,Y)}$

$$R^2 = \frac{var(\hat X)}{var(X)} = \frac{var(\hat a Y + \hat b \mathbb{1})}{var(X)} = \frac{\hat a^2\, var(Y)}{var(X)}$$

$$= \frac{cov(X,Y)^2}{var(Y)^{\cancel{2}}}\ \frac{\cancel{v(Y)}}{v(X)} = \frac{cov(X,Y)^2}{var(Y)\,var(X)} = \rho^2(X,Y).$$