

Exercice 1 Pour le site $i \in \{1, \dots, 8\}$, on note X_i le taux de pollution et Y_i le taux d'arbres malformés.

1) Dans un modèle $Y = X\beta + \varepsilon$ on veut être capable de prédire de nouveaux Y_i si une nouvelle valeur de X_i se présente.

Ici, on veut expliquer le taux de malformations à partir du taux de pollution donc X_i = variable explicative.
 Y_i = variable à expliquer

2) $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$

Interprétation:
 β_1 : dépendance affine
 ε_i : incertitude de mesure / aléa / incertitude liée à d'autres variables non observées ...

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_8 \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} X_1 \\ \vdots \\ X_8 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_8 \end{pmatrix}$$

$$Y = X\beta + \varepsilon \quad \text{où} \quad X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_8 \end{pmatrix}$$

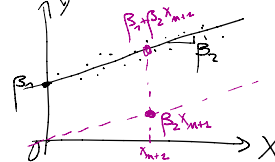
$$\text{et } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

3) Quels sont les paramètres du modèle? $\rightarrow \beta$

Comment les estimer? $\hat{\beta} = (X^T X)^{-1} X^T Y$

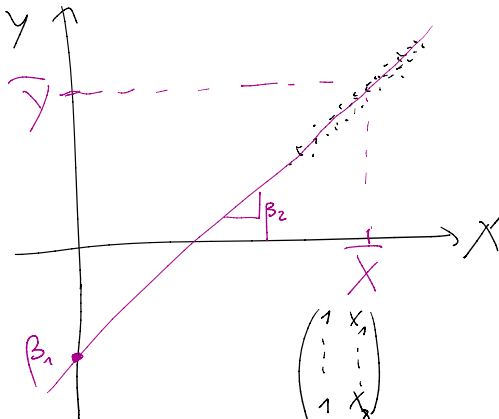
Application numérique: $\hat{\beta} = \begin{pmatrix} \dots \\ -17 \\ \dots \\ 1,59 \end{pmatrix}$?

Astuce si $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ inversible alors $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$



On donne (X_i, Y_i) pour $i = 1, \dots, n$ (entraînement)

On donne X_{n+1} : que peut-on dire de Y_{n+1} ? (prédiction)



$$X^T X = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_8 \end{pmatrix} \begin{pmatrix} 8 \\ \sum X_i \\ \sum X_i^2 \end{pmatrix} = \begin{pmatrix} 8 & 51,5 \\ 51,5 & 352,53 \end{pmatrix}$$

$$(X^T X)^{-1} = \frac{1}{8 \cdot 352,53 - 51,5^2} \begin{pmatrix} 352,53 & -51,5 \\ -51,5 & 8 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & \dots & 1 \\ X_1 & \dots & X_8 \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_8 \end{pmatrix} = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} = \begin{pmatrix} 83,5 \\ 560,51 \end{pmatrix}$$

Donc

$$\hat{\beta} = \frac{1}{8 \cdot 352,53 - 51,5^2} \begin{pmatrix} 352,53 & -51,5 \\ -51,5 & 8 \end{pmatrix} \begin{pmatrix} 83,5 \\ 560,51 \end{pmatrix}$$

$$= \frac{1}{167,99} \begin{pmatrix} 569,63 \\ 183,83 \end{pmatrix} \approx \begin{pmatrix} 3,39 \\ 1,09 \end{pmatrix}$$

$$Y_i = 3,39 + 1,09 X_i + \varepsilon_i$$

$$Y_i = 3,39 + 1,09 X_i + \varepsilon_i$$

$$4) \quad Y_{\text{nouveau}} = 3,39 + 1,09 \cdot 8 = 12,11$$

Exercice 2 $\ll \text{test } \beta_2 = 0 \text{ v.s. } \beta_2 \neq 0 \gg$ (y a-t-il un lien entre x et y?)

1) Variable explicative: I_i

Variable à expliquer: S_i

$$2) \quad S_i = \beta_1 + \beta_2 I_i + \varepsilon_i \quad \text{donc} \quad S = X\beta + \varepsilon \quad \text{ou} \quad X = \begin{pmatrix} 1 & I_1 \\ \vdots & \vdots \\ 1 & I_n \end{pmatrix}$$

$$3) \quad \text{Formule du cours: } \hat{\beta} = \frac{1}{\sum_{i=1}^n (X_i^2 - \bar{X}^2)} \begin{pmatrix} \sum_{i=1}^n X_i^2 Y_i - \bar{X} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i - \bar{X} \sum_{i=1}^n Y_i \end{pmatrix}$$

$$= \frac{1}{11 \times 1213,602 - 3614^2} \begin{pmatrix} 1213,602 \times 46,799 - 3614 \times 15,283 \\ 15,283 \times 46,799 - 3614 \times 3,64 \end{pmatrix}$$

$$= \frac{1}{288,626} \begin{pmatrix} 56,795,359,398 - 55,254,852,698 \\ 168,180,177 - 169,131,586 \end{pmatrix}$$

$$= \begin{pmatrix} 5,337 \\ -3,296 \end{pmatrix}$$

4) Un modèle linéaire $Y = X\beta + \varepsilon$ est gaussien homoscedastique si $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, c'est-à-dire si les ε_i sont i.i.d $\mathcal{N}(0, \sigma^2)$.

5) L'estimateur $\hat{\beta} = \frac{1}{n} \|S - X\hat{\beta}\|^2$ est biaisé: $E[\hat{\beta}] = \frac{n-2}{n} \sigma^2$

L'estimation non biaisée est $\hat{\sigma}^2 = \frac{1}{n-2} \|S - X\hat{\beta}\|^2$.
 dans le cours, m.d. 1 où d = dim de X
 parce que le terme constant n'est pas
 inclus dans X: $Y = \beta_0 + \beta_1 X + \dots + \beta_k X^k + \varepsilon$
 d+1 termes

6) Rappel: $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X^T X)^{-1})$
 Théorème: $\frac{n \hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$
 sont indépendantes

On veut tester $H_0: \beta_2 = 0$ v.s. $H_1: \beta_2 \neq 0$.

Sous H_0 , quelle est la loi de $\hat{\beta}_2$? $\mathcal{N}(\beta_2, \sigma^2 ((X^T X)^{-1})_{22})$

$Z \sim \mathcal{N}(\mu, \Sigma)$, $E[Z] = \mu$, $\text{Var}(Z_i) = \Sigma_{ii}$, $\text{Cov}(Z_i, Z_j) = \Sigma_{ij}$.
 (notion de $(X^T X)^{-1}$)

Def: Si $Z \sim \mathcal{N}(q, 1)$, $U \sim \chi^2(d)$ et $Z \perp U$
 alors $\frac{Z}{\sqrt{U/d}} \sim \mathcal{T}(d)$

Def: Si $Z \sim \mathcal{N}(0,1)$, $U \sim \chi^2(d)$ et $Z \perp U$
 alors $\frac{Z}{\sqrt{U/d}} \sim \mathcal{Z}(d)$

Que prend-on pour Z et pour U ? Que devient la formule $\frac{Z}{\sqrt{U/d}}$ avec ce choix?

Ici $\hat{\beta}_2 \sim \mathcal{N}(\beta_2, \sigma^2 (X^T X)^{-1}) = \mathcal{N}(\beta_2, \sigma^2 \frac{(X^T X)^{-1}_{11}}{\det(X^T X)})$ donc
 et $X^T X = \begin{pmatrix} 11 & \sum I_i \\ \sum I_i & \sum I_i^2 \end{pmatrix}$

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\sigma^2 \frac{11}{\det(X^T X)}}} \sim \mathcal{N}(0,1).$$

Par ailleurs $U = \frac{11 \hat{\sigma}_n^2}{\sigma^2} = (n-2) \frac{\hat{\sigma}^2}{\sigma^2} = 9 \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(9)$ est indépendant de Z

donc $\frac{Z}{\sqrt{U/9}} \sim \mathcal{Z}(9)$, c-à-d $T = \frac{(\hat{\beta}_2 - \beta_2) / \sqrt{\sigma^2 \frac{11}{\det(X^T X)}}}{\sqrt{9 \hat{\sigma}^2 / \sigma^2} / 3} = \frac{(\hat{\beta}_2 - \beta_2) \sqrt{\frac{\det(X^T X)}{11 \hat{\sigma}^2}}}{\sqrt{9 \hat{\sigma}^2 / \sigma^2} / 3} \sim \mathcal{Z}(9)$

- On veut tester si I a une influence sur S : H_0 : pas d'influence, H_1 : influence
- On suppose $S = X\beta + \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ (modèle gaussien homosédastique comme dans Z et 4)
 et σ^2 inconnue.

$H_0: \beta_2 = 0$ $H_1: \beta_2 \neq 0$ (car $S_i = \beta_1 + \beta_2 I_i + \varepsilon_i$)

- Statistique de test: $T = \frac{\hat{\beta}_2}{\sqrt{\frac{\det(X^T X)}{11 \hat{\sigma}^2}}} \sim \mathcal{Z}(9)$ sous H_0 .

- Région de rejet: $R =]-\infty, -t] \cup [t, +\infty[$ (symétrique car $\mathcal{Z}(9)$ est symétrique)

On veut $P_{H_0}(T \in R) \leq \alpha = 5\%$, c-à-d $P_{T \sim \mathcal{Z}(9)}(T \leq -t) + P_{T \sim \mathcal{Z}(9)}(T \geq t) \leq 5\%$.

On prend $t =$ la quantile d'ordre 97,5% de $\mathcal{Z}(9)$,

c-à-d $t = 3,262$.

- Calcul de T :

$X^T X = \begin{pmatrix} 11 & \sum I_i \\ \sum I_i & \sum I_i^2 \end{pmatrix}$

$T^{obs} = -33 \cdot \sqrt{\frac{11 \cdot 123602 - 3614^2}{11 \cdot 262124}} = -33 \cdot \sqrt{\frac{288624}{2883344}} = -33 \cdot \sqrt{0,1}$

$T^{obs} = -1,04$

$T^{obs} \notin R$ donc on conserve H_0 : l'intensité de la résonance n'a pas d'influence sur la solidité de la poutre.

Exercice 3 $X = U \Sigma V^T$

$$X = U \Sigma V^T$$

$(A^T A)_{ij} = \sum_k (A^T)_{ik} (A)_{kj} = \sum_k A_{ki} A_{kj}$

$(A^2)_{ij} = \sum_k A_{ik} A_{kj}$

$X^T X = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$

$\sim \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

$\Sigma = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

$\Sigma = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

$$X^T X = V \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix} V^T$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix} \quad \text{donc } \Sigma^T \Sigma = \text{diag}(\sigma_i^2)$$

$$(X^T X)^{-1} = V \begin{pmatrix} \sigma_1^{-2} & & 0 \\ & \ddots & \\ 0 & & \sigma_d^{-2} \end{pmatrix} V^T$$

$$(X^T X)^{-1} X^T = V \begin{pmatrix} \sigma_1^{-2} & & 0 \\ & \ddots & \\ 0 & & \sigma_d^{-2} \end{pmatrix} V^T \underbrace{\Sigma^T U^T}_{I_d} = V \begin{pmatrix} \sigma_1^{-2} & & 0 \\ & \ddots & \\ 0 & & \sigma_d^{-2} \end{pmatrix} U^T$$

Donc si $z \in \mathbb{R}^r$,

$$(X^T X)^{-1} X^T U z = V \begin{pmatrix} \sigma_1^{-2} & & 0 \\ & \ddots & \\ 0 & & \sigma_d^{-2} \end{pmatrix} z$$

d'où $(X^T X)^{-1} X^T U e_d = V \sigma_d^{-2} e_d$
 donc $\|(X^T X)^{-1} X^T U e_d\| = \frac{1}{\sigma_d^2} \|e_d\| = \frac{1}{\sigma_d}$

2) Utilisez que si $X \in \mathbb{R}^d$ vecteur colonne, $\|X\|^2 = X^T X$.

- $\nabla (\beta \mapsto A\beta) = A^T$
- $\nabla (\beta \mapsto \beta^T C \beta) = (C + C^T)\beta$

$A(B+C) = AB + AC$
 $(A+B)^T = A^T + B^T$
 $(AB)^T = B^T A^T$
 Si X, Y sont des vecteurs colonne,
 alors $X^T Y = Y^T X$
 (car c'est un réel donc $=$ sa transposée)

$$\|\beta\|^2 = \beta^T \beta = \beta^T \text{Id} \beta \quad \text{donc}$$

$$\nabla (\|\beta\|^2) = (\text{Id} + \text{Id}^T) \beta = 2\beta$$

3) Le critère est $\mathcal{C}^1$ donc son minimum annule son gradient, c-à-d annule $-2X^T(Y - X\beta) + 2\beta$.
 On cherche β t.q. $-X^T Y + X^T X \beta + \beta = 0$, c-à-d $(X^T X + \text{Id}) \beta = X^T Y$
 donc $\beta = (X^T X + \text{Id})^{-1} X^T Y$

$$\beta = \underbrace{(X^T X + \text{Id})^{-1} X^T Y}_{= M}$$

4) Si $X = U \Sigma V^T$, $X^T X = V \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix} V^T$

d'où $M^{-1} = X^T X + \text{Id} = V \begin{pmatrix} \sigma_1^2 + 1 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 + 1 \end{pmatrix} V^T$ donc

$$M = V \begin{pmatrix} \frac{1}{\sigma_1^2 + 1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_d^2 + 1} \end{pmatrix} V^T$$

D'où pour tout $z \in \mathbb{R}^m$, si on regarde $\hat{\beta}^{\text{perturb}} - \hat{\beta}$ avec $y^{\text{perturb}} = Y + U z$,

$$\hat{\beta}^{\text{perturb}} - \hat{\beta} = M^{-1} X^T (U z) = V \begin{pmatrix} \frac{1}{\sigma_1^2 + 1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_d^2 + 1} \end{pmatrix} V^T \underbrace{U^T U}_{= \text{Id}} z$$

$$= V \begin{pmatrix} \frac{\sigma_1^2}{\sigma_1^2 + 1} & & 0 \\ & \ddots & \\ 0 & & \frac{\sigma_d^2}{\sigma_d^2 + 1} \end{pmatrix} z$$

$$= \sqrt{\sum_{i=1}^d \left(\frac{\sigma_i^2}{\sigma_i^2 + 1} \right)^2} \|z\|$$

$$\leq \left(\sup_{i=1, \dots, d} \left(\frac{\sigma_i}{\sigma_i^2 + 1} \right) \right) \cdot \sum_{i=1}^d \sigma_i^2$$

Donc $\|\hat{\beta}^{\text{perturb}} - \hat{\beta}\| = \left\| \begin{pmatrix} \frac{\sigma_1}{\sigma_1^2 + \lambda} & 0 \\ 0 & \frac{\sigma_d}{\sigma_d^2 + \lambda} \end{pmatrix} z \right\| \leq \sup_{i=1, \dots, d} \left(\frac{\sigma_i}{\sigma_i^2 + \lambda} \right) \|z\|$

$\leq \sup_{\sigma \geq 0} \left(\frac{1}{\sigma + \frac{\lambda}{\sigma}} \right) \|z\| = \frac{1}{2\sqrt{\lambda}} \|z\|$
 (atteint en $\sigma = \sqrt{\lambda}$)

$\left| \frac{\sigma_0}{\sigma_0^2 + \lambda} \right| \rightarrow \sqrt{\sigma_0^2 + \lambda} \cdot 0$

$\leq \left(\sup_{i=1, \dots, d} \left(\frac{\sigma_i}{\sigma_i^2 + \lambda} \right) \right)^2 \cdot \sum_{i=1}^d z_i^2 \leq \|z\|^2$

D'où

$$\frac{\|\hat{\beta}^{\text{perturb}} - \hat{\beta}\|}{\|y^{\text{perturb}} - y\|} \leq \frac{1}{2\sqrt{\lambda}}$$

$$\sum_i a_i b_i \leq (\max_i a_i) \sum_i b_i$$

$$a_i, b_i \geq 0$$