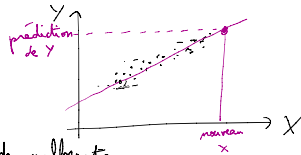


Exercice 1

- 1) On a toujours une variable à expliquer (Y) → But: si on donne de nouveaux X , comment prédire les Y correspondants.



Ici, à expliquer = Y , taux de malformation
explicative = X , pollution.

2) $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ pour $i \in \{1, \dots, 8\}$
 dépendance affine
 incertitude de mesure / aléa ...

$$Y = X\beta + \varepsilon$$

$$X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_8 \end{pmatrix} \quad \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_8 \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} X_1 \\ \vdots \\ X_8 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_8 \end{pmatrix}$$

3) Astuce: si $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est inversible alors $A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Donc $Y_i = \beta_1(C_1)_i + \beta_2(C_2)_i + \varepsilon_i$

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

$n=8$
 $d=2$
 Ici β a 2 lignes.

$X^T X = d \begin{pmatrix} 1 & X \\ X^T & 1 \end{pmatrix}$
 $(X^T X)^{-1} = \frac{1}{d} \begin{pmatrix} 1 & X \\ X^T & 1 \end{pmatrix}^{-1}$

$X^T X = \begin{pmatrix} 1 & -1 \\ X_1 & -X_8 \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ 1 & X_8 \end{pmatrix} = \begin{pmatrix} 2 & 2X_1 \\ 2X_1 & 2X_1^2 \end{pmatrix} = \begin{pmatrix} 2 & 16 \\ 16 & 352 \end{pmatrix}$
 $\det(X^T X) = 167.99$

donc $(X^T X)^{-1} = \frac{1}{167.99} \begin{pmatrix} 352.53 & -5.5 \\ -5.5 & 2 \end{pmatrix}$ et $X^T Y = \begin{pmatrix} 2Y_1 \\ 2X_1 Y_1 \end{pmatrix} = \begin{pmatrix} 83.5 \\ 569.91 \end{pmatrix}$

Donc $\hat{\beta} = \begin{pmatrix} 3.39 \\ 1.09 \end{pmatrix}$ On a estimé $Y_i = 3.39 + 1.09 X_i + \varepsilon_i$.

4) On prédit $Y^{\text{préd}} = 3.39 + 1.09 \times 8 = 12.11$.

Exercice 2

- 1) Variable à expliquer: S_i
Variable explicative: I_i

2) $S = X\beta + \varepsilon$ où $X = \begin{pmatrix} 1 & I_1 \\ \vdots & \vdots \\ 1 & I_n \end{pmatrix}$, $S = \begin{pmatrix} S_1 \\ \vdots \\ S_n \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

3) $\det(X^T X) = 289.626$

$\hat{\beta} = \begin{pmatrix} 5.337 \\ -3.29 \end{pmatrix}$

$S_i = \beta_1 + \beta_2 I_i + \varepsilon_i$
Est-ce que S_i dépend de I_i ? → Est-ce que $\beta_2 \neq 0$?

- 4) Modèle gaussien homosédatique:
 les composantes de ε sont i.i.d. $\mathcal{N}(0, \sigma^2)$.
 ε vecteur gaussien $(\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n))$

5) σ^2 est inconnue il faut l'estimer.

$$\hat{\sigma}_m^2 = \frac{1}{n-d} \|Y - X\hat{\beta}\|^2$$

6) Théorème: $\hat{\beta}_m \sim \mathcal{N}(\beta, \sigma^2 (X^T X)^{-1})$
 $(n-d) \frac{\hat{\sigma}_m^2}{\sigma^2} \sim \chi^2(n-d)$
 $\hat{\beta}_m$ et $\hat{\sigma}_m^2$ sont indépendants.

Comment faire un test sur β_2 ?

(Plus précisément $H_0: \beta_2 = 0$ v.s. $H_1: \beta_2 \neq 0$)

• $(\hat{\beta}_m)_2 \sim \mathcal{N}(\beta_2, \sigma^2 (X^T X)^{-1}_{22}) = \mathcal{N}(\beta_2, \frac{1}{\det(X)} (X^T X)_{44})$

• Peut-on construire une statistique de test fondée sur $(\hat{\beta}_m)_2$ dont la loi ne dépend pas de σ^2 ?

$$(A^{-1})_{22} = \frac{1}{\det(A)} A_{11}$$

Si $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ alors $\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

Si $Z \sim \mathcal{N}(\mu, \Sigma)$, alors $Z_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$
 $(\text{Var}(Z_i) = \Sigma_{ii}, \text{Cov}(Z_i, Z_j) = \Sigma_{ij})$

$(\hat{\beta}_m)_2 \sim \mathcal{N}(\beta_2, \frac{\sigma^2}{\det(X)})$ donc $\frac{(\hat{\beta}_m)_2 - \beta_2}{\frac{\sigma}{\sqrt{\det(X)}}} \sim \mathcal{N}(0, 1)$

Remplacez σ par $\hat{\sigma}_m$

puis cherchez à vous ramener à une loi de Student.

(Si $Z \sim \mathcal{N}(0,1)$ et $U \sim \chi^2(d)$ et Z et U sont indépendants $\frac{Z}{\sqrt{U/d}} \sim \mathcal{U}(d)$)

On cherche $Z \sim \mathcal{N}(0,1)$ indépendante
 $U \sim \chi^2(d)$

$$Z = \frac{(\hat{\beta}_m)_2 - \mathbb{E}[(\hat{\beta}_m)_2]}{\sqrt{\text{Var}[(\hat{\beta}_m)_2]}} \sim \mathcal{N}(0,1)$$

$$\text{et } U = \frac{(n-d) \hat{\sigma}_m^2}{\sigma^2}$$

Donc

$$T = \frac{(\hat{\beta}_m)_2 - \beta_2}{\frac{\hat{\sigma}_m}{\sqrt{\det(X)}}} \times \frac{1}{\sqrt{\frac{(n-d) \hat{\sigma}_m^2}{\sigma^2} / (n-d)}} \sim \mathcal{U}(n-d)$$

$$T = \frac{(\hat{\beta}_m)_2 - \beta_2}{\frac{\hat{\sigma}_m}{\sqrt{\det(X)}}} \times \sqrt{\frac{\det(X)}{n}}$$

- On veut tester H_0 : pas de dépendance H_1 : dépendance entre S et I.
- Modèle linéaire gaussien homosédastique à variance inconnue: $S_i = \beta_1 + \beta_2 I_i + \varepsilon_i$ où les ε_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, σ^2 inconnue.

• Stat de test: $T = \frac{(\hat{\beta}_m)_2}{\frac{\hat{\sigma}_m}{\sqrt{\det(X)}}} \sim \mathcal{U}(n-d) = \mathcal{U}(9)$ sous H_0 .

• Région de rejet: $R =]-\infty, -t] \cup [t, +\infty[$

• Niveau: on cherche t tel que $P_{H_0}(T \in R) = P_{T \sim \mathcal{U}(9)}(T \leq -t) + P_{T \sim \mathcal{U}(9)}(T \geq t) = 2 P_{T \sim \mathcal{U}(9)}(T \geq t) \leq 5\%$
 On choisit donc $t =$ la quantile d'ordre 97.5% de $\mathcal{U}(9)$: $t = 2.252$

• Calcul: $T = \frac{-3.29}{\sqrt{262.124}} \sqrt{\frac{288.626}{11}} = -3.29 \times \sqrt{0.1001} \approx -3 \sqrt{\frac{1}{9}} = -1 \notin R$

Donc on conserve H_0 : on ne peut pas conclure que la solidité dépend de l'intensité.
 de la posture $\uparrow$ corrélation $\uparrow$ par causalité $\uparrow$ de la résonance

Exercice 3

On sait

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad \text{lorsque } Y = X\beta + \varepsilon.$$

Question: si on fait une petite erreur en mesurant les Y , quelle erreur sur $\hat{\beta}$ en découle?

$$\hat{\beta}^{\text{perturb}} = (X^T X)^{-1} X^T Y^{\text{perturb}}$$

1) a) Par définition de Y^{perturb} , $Y - Y^{\text{perturb}} = -\delta U e_1$ donc $\|Y - Y^{\text{perturb}}\| = \delta \|U e_1\|$
 $= \delta \|e_1\|$ car $U \in O_n(\mathbb{R})$
 $= \delta$ car $\|e_1\| = 1$.

$$b) \quad \|\hat{\beta} - \hat{\beta}^{\text{perturb}}\| = \|(X^T X)^{-1} X^T (Y - Y^{\text{perturb}})\| = \delta \|(X^T X)^{-1} X^T U e_d\|.$$

On utilise $X = U \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_d \\ & & & 0 \end{pmatrix} V^T$ donc $X^T X = V \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \\ & & & 0 \end{pmatrix} U^T U \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_d \\ & & & 0 \end{pmatrix} V^T$
 $\stackrel{=I_d}{=} \underbrace{U^T U}_{\text{car } U \in O_n(\mathbb{R})} V^T$
d'où $X^T X = V \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \\ & & & 0 \end{pmatrix} V^T$

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

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

$$\text{Donc pour tout } z \in \mathbb{R}^d, \quad (X^T X)^{-1} X^T U z = V \begin{pmatrix} \sigma_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \sigma_d^{-1} \\ & & & 0 \end{pmatrix} z = V \begin{pmatrix} z_1/\sigma_1 \\ \vdots \\ z_d/\sigma_d \\ 0 \end{pmatrix}$$

$$\text{D'où } \|(X^T X)^{-1} X^T U z\| = \sqrt{\sum_{i=1}^d \frac{z_i^2}{\sigma_i^2}}, \text{ en particulier } \|(X^T X)^{-1} X^T U e_d\| = \frac{1}{\sigma_d}.$$

$$\text{D'où } \|\hat{\beta} - \hat{\beta}^{\text{perturb}}\| = \frac{\delta}{\sigma_d}.$$

$$c) \quad \frac{\|\hat{\beta} - \hat{\beta}^{\text{perturb}}\|}{\|Y - Y^{\text{perturb}}\|} = \frac{1}{\sigma_d} \xrightarrow{\sigma_d \rightarrow 0} +\infty : \beta \text{ est instable, la moindre erreur sur } Y \text{ cause une énorme erreur sur } \hat{\beta}.$$

2) en exo.

$$3) \quad M = X^T X + \lambda I_d = V \begin{pmatrix} \sigma_1^2 + \lambda & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 + \lambda \end{pmatrix} V^T.$$

$$4) \quad \forall z \in \mathbb{R}^m, \quad M^{-1} X^T U z = (\dots) = V^T \begin{pmatrix} \frac{\sigma_1}{\sigma_1^2 + \lambda} & & 0 \\ & \ddots & \\ 0 & & \frac{\sigma_d}{\sigma_d^2 + \lambda} \\ & & & 0 \end{pmatrix} z$$

d'où

$$\|M^{-1} X^T U z\| = \sqrt{\sum_{i=1}^d \left(\frac{\sigma_i}{\sigma_i^2 + \lambda}\right)^2 z_i^2} \leq \sqrt{\sum_{i=1}^d \left(\sup_{\sigma \geq 0} \frac{\sigma}{\sigma^2 + \lambda}\right)^2 z_i^2} \leq \left(\sup_{\sigma \geq 0} \frac{\sigma}{\sigma^2 + \lambda}\right) \|z\|.$$

En prenant

$$z = U^{-1} (Y - Y^{\text{perturb}}), \text{ alors } Y - Y^{\text{perturb}} = U z \text{ donc } \|z\| = \|Y - Y^{\text{perturb}}\|$$

d'où

$$\frac{\|\hat{\beta} - \hat{\beta}^{\text{perturb}}\|}{\|Y - Y^{\text{perturb}}\|} \leq \sup_{\sigma \geq 0} \frac{\sigma}{\sigma^2 + \lambda} = \frac{1}{2\sqrt{\lambda}}$$