

MAP 433 : Introduction aux méthodes statistiques. Cours 8

16 Octobre 2015

Aujourd'hui

- 1 Le modèle de régression
- 2 Régression linéaire multiple
- 3 Choix de modèles
- 4 Analyse des résidus

Modèle de régression

Definition

Données : $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)$ avec $Y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^k$, et

$$Y_i = r(\beta, \mathbf{x}_i) + \sigma \xi_i, \quad \mathbb{E}_\theta [\xi_i] = 0,$$

- \mathbf{x}_i *déterministes*
- les variables ξ_i sont centrées, $\mathbb{E}_\theta[\xi_i] = 0$, décorrélées, $\mathbb{E}_\theta[\xi_i \xi_j] = 0$ si $i \neq j$ et de variance unité $\mathbb{E}[\xi_i^2] = 1$ (*homoscédasticité*).
- Paramètres : $\theta = (\beta, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Modèle de régression

Definition

Données : $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)$ avec $Y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^k$, et

$$Y_i = r(\boldsymbol{\beta}, \mathbf{x}_i) + \sigma \xi_i, \quad \mathbb{E}_\theta [\xi_i] = 0,$$

- \mathbf{x}_i déterministes
- les variables ξ_i sont centrées, $\mathbb{E}_\theta[\xi_i] = 0$, décorrélées, $\mathbb{E}_\theta[\xi_i \xi_j] = 0$ si $i \neq j$ et de variance unité $\mathbb{E}[\xi_i^2] = 1$ (homoscédasticité).
- Paramètres : $\theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Modèle de régression

Definition

Données : $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)$ avec $Y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^k$, et

$$Y_i = r(\beta, \mathbf{x}_i) + \sigma \xi_i, \quad \mathbb{E}_\theta [\xi_i] = 0,$$

- \mathbf{x}_i déterministes
- les variables ξ_i sont centrées, $\mathbb{E}_\theta[\xi_i] = 0$, décorrélées, $\mathbb{E}_\theta[\xi_i \xi_j] = 0$ si $i \neq j$ et de variance unité $\mathbb{E}[\xi_i^2] = 1$ (homoscédasticité).
- Paramètres : $\theta = (\beta, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Régression gaussienne

- Modèle de régression Gaussien :

$$Y_i = r(\beta, x_i) + \sigma \xi_i, \quad \theta \in \Theta \subset \mathbb{R}^d \times \mathbb{R}_+.$$

où $\xi_i \sim \mathcal{N}(0, 1)$, i.i.d.

- On sait expliciter la vraisemblance de l'observation $Z = (Y_1, \dots, Y_n) \implies$ appliquer le principe du maximum de vraisemblance.

$$\mathcal{L}_n(\theta, Y_1, \dots, Y_n) = \frac{1}{(\sqrt{2\pi}\sigma^2)^n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - r(\beta, x_i))^2 \right)$$

Estimateur des moindres carrés

Maximiser la **vraisemblance** en régression gaussienne

$$\hat{\beta}_n \in \operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^k} \sum_{i=1}^n (Y_i - r(\mathbf{b}, \mathbf{x}_i))^2$$

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (Y_i - r(\hat{\beta}_n, \mathbf{x}_i))^2$$

- L'estimateur $\hat{\beta}_n$ est appelé l'**estimateur des moindres carrés**. Il peut être appliqué même dans un cas non gaussien.
- **Existence, unicité.**

Régression linéaire multiple (=Modèle linéaire)

- La fonction de régression est $r(\beta, \mathbf{x}_i) = \mathbf{x}_i^T \beta$. On observe

$$(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)$$

avec

$$Y_i = \mathbf{x}_i^T \beta + \sigma \xi_i, \quad i = 1, \dots, n$$

où $\theta \in \Theta = \mathbb{R}^k$, $\mathbf{x}_i \in \mathbb{R}^k$.

- Matriciellement

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} \beta + \sigma \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} = \mathbb{X} \beta + \sigma \boldsymbol{\xi}.$$

EMC en régression linéaire multiple

- Estimateur des **moindres carrés** en régression linéaire multiple : tout estimateur $\hat{\beta}_n$ satisfaisant

$$\sum_{i=1}^n (Y_i - \mathbf{x}_i^T \hat{\beta}_n)^2 = \min_{\mathbf{b} \in \mathbb{R}^k} \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \mathbf{b})^2.$$

- En notation matricielle :

$$\| \mathbf{Y} - \mathbb{X} \hat{\beta}_n \|^2 = \min_{\mathbf{b} \in \mathbb{R}^k} \| \mathbf{Y} - \mathbb{X} \mathbf{b} \|^2$$

Projection orthogonale sur

$$\text{Vect}(\mathbb{X}) = \{ \mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \mathbb{X} \mathbf{b}, \mathbf{b} \in \mathbb{R}^k \}.$$

Géométrie de l'EMC

- L'EMC vérifie

$$\mathbb{X} \hat{\beta}_n = \Pi_{\mathbb{X}} \mathbf{Y}$$

où $\Pi_{\mathbb{X}}$ est le projecteur orthogonal sur $\text{Vect}(\mathbb{X})$.

- Comme $\mathbf{Y} - \Pi_{\mathbb{X}} \mathbf{Y} \perp \text{Vect}(\mathbb{X})$, on en déduit **les équations normales** de l'EMC :

$$\mathbb{X}^T \mathbb{X} \hat{\beta}_n = \mathbb{X}^T \mathbf{Y}.$$

- Remarques.

- L'EMC est un Z-estimateur.
- **unicité** de $\hat{\beta}_n$ si la matrice de Gram $\mathbb{X}^T \mathbb{X}$ est inversible (la matrice \mathbb{X} est de rang complet).

Estimateur des moindres carrés

Proposition

Si $\mathbb{X}^T \mathbb{X}$ (matrice $k \times k$) inversible, alors l'EMC $\hat{\beta}_n$ est *unique* et

$$\boxed{\hat{\beta}_n = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbf{Y}} = \mathbb{X}^\# \mathbf{Y}$$

- 1 $\mathbb{X}^\#$ *pseudo-inverse* de \mathbb{X} .
- 2 $\Pi_{\mathbb{X}} = \mathbb{X}(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T = \mathbb{X} \mathbb{X}^\#$ projecteur orthogonal sur $\text{Vect}(\mathbb{X})$.

Propriétés de l'estimateur

Théorème (Régression Gaussienne)

Supposons que $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}$ avec $\boldsymbol{\xi} \sim \mathcal{N}(0, \text{Id}_n)$. Alors,
 $\hat{\boldsymbol{\beta}}_n \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbb{X}^T \mathbb{X})^{-1})$

Démonstration.

Comme $\mathbb{X}^\# = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T$, $\mathbb{X}^\# \mathbb{X} = \text{Id}_k$. Par conséquent :

$$\begin{aligned}\hat{\boldsymbol{\beta}}_n &= \mathbb{X}^\# \mathbf{Y} = \mathbb{X}^\# (\mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}) \\ &= \boldsymbol{\beta} + \sigma \mathbb{X}^\# \boldsymbol{\xi}\end{aligned}$$

Le vecteur $\mathbb{X}^\# \boldsymbol{\xi}$ est Gaussien centré de matrice de covariance

$$\mathbb{E} \left[\mathbb{X}^\# \boldsymbol{\xi} \boldsymbol{\xi}^T (\mathbb{X}^\#)^T \right] = \mathbb{X}^\# \mathbb{E} \left[\boldsymbol{\xi} \boldsymbol{\xi}^T \right] (\mathbb{X}^\#)^T = \mathbb{X}^\# (\mathbb{X}^\#)^T = (\mathbb{X}^T \mathbb{X})^{-1}.$$



Prédiction et Erreur de prédiction

■ Prédiction

$$\hat{Y} = \mathbb{X} \hat{\beta}_n = \Pi_{\mathbb{X}} Y$$

projection des observations sur l'espace de régression.

■ Erreur (résidu) de prédiction :

$$Y - \hat{Y} = (\text{Id}_n - \Pi_{\mathbb{X}}) Y .$$

■ Si $Y = \mathbb{X}\beta + \sigma\xi$, alors

$$\hat{Y} = \mathbb{X}\beta + \sigma\Pi_{\mathbb{X}}\xi$$

$$Y - \hat{Y} = \sigma(\text{Id}_n - \Pi_{\mathbb{X}})\xi$$

car $\Pi_{\mathbb{X}}\mathbb{X} = \mathbb{X}$ ($\Pi_{\mathbb{X}}$ est le projecteur orthogonal sur l'image de \mathbb{X}).

Théorème de Cochran

Théorème

Hypothèses

- $\mathbf{Y} \sim \mathcal{N}(\mu, \sigma^2 \text{Id}_n)$,
- \mathcal{M} un sous espace de \mathbb{R}^n de dimension k ,
- Π la matrice de projection orthogonale sur \mathcal{M} et $\Pi_{\perp} = \text{Id}_n - \Pi$ la matrice de projection orthogonale sur \mathcal{M}^{\perp} .

Alors

- 1 $\Pi \mathbf{Y} \sim \mathcal{N}(\Pi \mu, \sigma^2 \Pi)$, $\Pi_{\perp} \mathbf{Y} \sim \mathcal{N}(\Pi_{\perp} \mu, \sigma^2 \Pi_{\perp})$
- 2 les vecteurs $\Pi \mathbf{Y}$ et $\Pi_{\perp} \mathbf{Y}$ sont indépendants
- 3 $\|\Pi(\mathbf{Y} - \mu)\|^2 / \sigma^2 \sim \chi_k^2$ et $\|\Pi_{\perp}(\mathbf{Y} - \mu)\|^2 / \sigma^2 \sim \chi_{n-k}^2$.

Théorème de Cochran

Théorème

Hypothèses

- $\mathbf{Y} \sim \mathcal{N}(\mu, \sigma^2 \text{Id}_n)$,
- \mathcal{M} un sous espace de \mathbb{R}^n de dimension k ,
- Π la matrice de projection orthogonale sur \mathcal{M} et $\Pi_{\perp} = \text{Id}_n - \Pi$ la matrice de projection orthogonale sur \mathcal{M}^{\perp} .

Alors

- 1 $\Pi \mathbf{Y} \sim \mathcal{N}(\Pi \mu, \sigma^2 \Pi)$, $\Pi_{\perp} \mathbf{Y} \sim \mathcal{N}(\Pi_{\perp} \mu, \sigma^2 \Pi_{\perp})$
- 2 les vecteurs $\Pi \mathbf{Y}$ et $\Pi_{\perp} \mathbf{Y}$ sont indépendants
- 3 $\|\Pi(\mathbf{Y} - \mu)\|^2 / \sigma^2 \sim \chi_k^2$ et $\|\Pi_{\perp}(\mathbf{Y} - \mu)\|^2 / \sigma^2 \sim \chi_{n-k}^2$.

Théorème de Cochran

Théorème

Hypothèses

- $\mathbf{Y} \sim \mathcal{N}(\mu, \sigma^2 \text{Id}_n)$,
- \mathcal{M} un sous espace de \mathbb{R}^n de dimension k ,
- Π la matrice de projection orthogonale sur \mathcal{M} et $\Pi_{\perp} = \text{Id}_n - \Pi$ la matrice de projection orthogonale sur \mathcal{M}^{\perp} .

Alors

- 1 $\Pi \mathbf{Y} \sim \mathcal{N}(\Pi \mu, \sigma^2 \Pi)$, $\Pi_{\perp} \mathbf{Y} \sim \mathcal{N}(\Pi_{\perp} \mu, \sigma^2 \Pi_{\perp})$
- 2 les vecteurs $\Pi \mathbf{Y}$ et $\Pi_{\perp} \mathbf{Y}$ sont indépendants
- 3 $\|\Pi(\mathbf{Y} - \mu)\|^2 / \sigma^2 \sim \chi_k^2$ et $\|\Pi_{\perp}(\mathbf{Y} - \mu)\|^2 / \sigma^2 \sim \chi_{n-k}^2$.

Résidus et variance résiduelle

Hypothèse : $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\xi$, $\xi \sim \mathcal{N}(0, \text{Id}_n)$.

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \sigma\mathbb{X}^\# \xi$$

$$\mathbf{Y} - \hat{\mathbf{Y}} = (\text{Id}_n - \Pi_{\mathbb{X}}) \mathbf{Y} = \sigma(\text{Id}_n - \Pi_{\mathbb{X}})\xi \quad \text{car } (\text{Id}_n - \Pi_{\mathbb{X}})\Pi_{\mathbb{X}} = 0$$

$$\hat{\sigma}_n^2 = (n - k)^{-1} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Theorem

- 1 $(n - k)\hat{\sigma}_n^2/\sigma^2$ suit une loi du χ^2 à $(n - k)$ degrés de liberté.
- 2 $\hat{\boldsymbol{\beta}}_n$ et $\hat{\sigma}_n^2$ sont indépendants.

Démonstration.

$(\text{Id}_n - \Pi_{\mathbb{X}})\xi \sim \mathcal{N}(0, (\text{Id}_n - \Pi_{\mathbb{X}}))$ et donc $\|(\text{Id}_n - \Pi_{\mathbb{X}})\xi\|^2$ suit une loi du χ^2 à $(n - k) = \text{Tr}(\text{Id}_n - \Pi_{\mathbb{X}})$ d.l. □

Résidus et variance résiduelle

Hypothèse : $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\xi$, $\xi \sim \mathcal{N}(0, \text{Id}_n)$.

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \sigma\mathbb{X}^\# \xi$$

$$\mathbf{Y} - \hat{\mathbf{Y}} = (\text{Id}_n - \Pi_{\mathbb{X}}) \mathbf{Y} = \sigma(\text{Id}_n - \Pi_{\mathbb{X}})\xi \quad \text{car } (\text{Id}_n - \Pi_{\mathbb{X}})\Pi_{\mathbb{X}} = 0$$

$$\hat{\sigma}_n^2 = (n - k)^{-1} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Theorem

- 1 $(n - k)\hat{\sigma}_n^2/\sigma^2$ suit une loi du χ^2 à $(n - k)$ degrés de liberté.
- 2 $\hat{\boldsymbol{\beta}}_n$ et $\hat{\sigma}_n^2$ sont indépendants.

Démonstration.

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \sigma\mathbb{X}^\# \xi = \boldsymbol{\beta} + \sigma\mathbb{X}^\# \Pi_{\mathbb{X}} \xi \quad \mathbf{Y} - \hat{\mathbf{Y}} = \sigma(\text{Id}_n - \Pi_{\mathbb{X}})\xi$$

Par le théorème de Cochran, $\Pi_{\mathbb{X}}\xi$ et $(\text{Id}_n - \Pi_{\mathbb{X}})\xi$ sont indépendants. □

Loi de Student

Definition

Soit Z une variable aléatoire de loi normale centrée et réduite et soit U une variable indépendante de Z et distribuée suivant la loi du χ^2 à p degrés de liberté. La variable

$$T = \frac{Z}{\sqrt{U/p}}$$

suit une loi de Student à p degrés de liberté.

Loi de Fisher

Definition

Soient U_1 et U_2 deux variables aléatoires indépendantes distribuées selon des lois du χ^2 à d_1 et d_2 degrés de liberté. La variable

$$\frac{U_1/d_1}{U_2/d_2}$$

est distribuée suivant une loi de Fisher à (d_1, d_2) d.l.

Intervalle de confiance

Théorème

Hypothèse : $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}$, $\boldsymbol{\xi} \sim \mathcal{N}(0, \text{Id}_n)$.

Pour $j = 1, \dots, k$,

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_n \sqrt{[(\mathbb{X}^T \mathbb{X})^{-1}]_{j,j}}} \sim \mathcal{T}_{n-k}$$

\mathcal{T}_{n-k} : Loi de **Student** à $n - k$ d.d.l.

Démonstration.

- $\hat{\beta}_j - \beta_j \sim \mathcal{N}(0, \sigma^2[(\mathbb{X}^T \mathbb{X})^{-1}]_{j,j})$
- $(n - k)\hat{\sigma}_n^2 / \sigma^2 \sim \chi^2(n - k)$ d.l.
- $\hat{\beta}_j - \beta_j$ et $(n - k)\hat{\sigma}_n^2$ sont indépendants.



Régions de confiance

Théorème

Hypothèse : $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}$, $\boldsymbol{\xi} \sim \mathcal{N}(0, \text{Id}_n)$.

Soit R une matrice $(q \times k)$ de rang q ($q \leq k$) alors

$$\frac{1}{q\hat{\sigma}_n^2} (R\{\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\})^T \left[R(\mathbb{X}^T \mathbb{X})^{-1} R^T \right]^{-1} R\{\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\} \sim \mathcal{F}_{q, n-k}$$

$\mathcal{F}_{q, n-k}$: Loi de Fisher à $(q, n - k)$ -d.d.l.

Démonstration.

- $R\{\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\} \sim \mathcal{N}(0, \sigma^2 R(\mathbb{X}^T \mathbb{X})^{-1} R^T)$ car $\hat{\boldsymbol{\beta}}_n \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (\mathbb{X}^T \mathbb{X})^{-1})$
- $(n - k)\hat{\sigma}_n^2/\sigma^2$ suit une loi du χ^2 à $(n - k)$ -d.l.
- $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ et $\hat{\sigma}_n^2$ sont indépendants.



Intervalles de confiance

Théorème

- 1 *Un intervalle de confiance bilatéral de niveau $1 - \alpha$, pour un β_j , $j = 1, \dots, p$, est donné par*

$$[\hat{\beta}_j - t_{n-p}(1 - \alpha/2)\hat{\sigma}_n\sqrt{[(X^T X)^{-1}]_{jj}}, \hat{\beta}_j + t_{n-p}(1 - \alpha/2)\hat{\sigma}_n\sqrt{[(X^T X)^{-1}]_{jj}}]$$

- 2 *Un intervalle de confiance bilatéral de niveau $1 - \alpha$, pour σ^2 est donné par*

$$\left[\frac{(n-p)\hat{\sigma}_n^2}{c_2}, \frac{(n-p)\hat{\sigma}_n^2}{c_1} \right] \quad \text{où} \quad \mathbb{P}(c_1 \leq \chi_{n-p}^2 \leq c_2) = 1 - \alpha.$$

Régions de confiance

Théorème

Une région de confiance pour q ($q \leq k$) paramètres β_j notés $(\beta_{j_1}, \dots, \beta_{j_q})$ de niveau $1 - \alpha$ est donnée par

$$\left\{ (\beta_{j_1}, \dots, \beta_{j_q}) \mathbb{R}^q, \right. \\ \left. \frac{1}{q\hat{\sigma}_n^2} [E(\hat{\beta}_n - \beta)]^T [E(X^T X)^{-1} E^T]^{-1} [E(\hat{\beta}_n - \beta)] \leq f_{q, n-p}(1 - \alpha) \right\},$$

où

- E est la matrice de taille $q \times k$ telle que

$$\begin{bmatrix} \beta_{j_1} \\ \vdots \\ \beta_{j_q} \end{bmatrix} = E\beta$$

(tous les éléments sont nuls sauf les $[E]_{ij_i}$, $i = 1, \dots, q$ qui valent 1)

- $f_{q, n-k}(1 - \alpha)$ est le quantile de niveau $(1 - \alpha)$ d'une loi de Fisher admettant $(q, n - k)$ d.d.l.

Un exemple

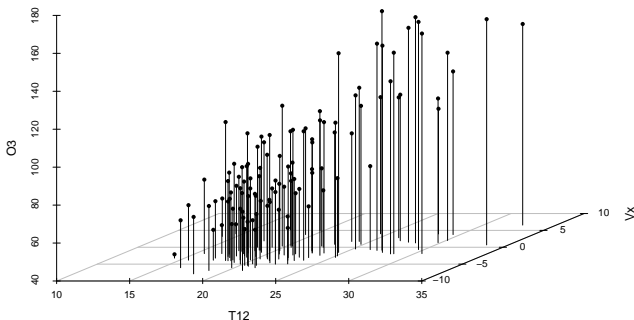


FIGURE – Représentation brute des données : modèle d'explication de l'ozone (O3) par la température à 12h (T12) et le Vent à 12h (Vx12)

Régression multiple

Modèle de régression

$$\text{maxO3} = \beta_1 + \beta_2 \text{T12} + \beta_3 \text{Vx12} + \beta_4 \text{Ne12} + \sigma \xi$$

Intervalles de confiance

	2.5 %	97.5 %
(Intercept)	-25.4886483	33.280203
T12	3.4819098	5.544563
Vx12	0.3264694	2.931560
Ne12	-3.6368523	0.399082

Régions de confiance

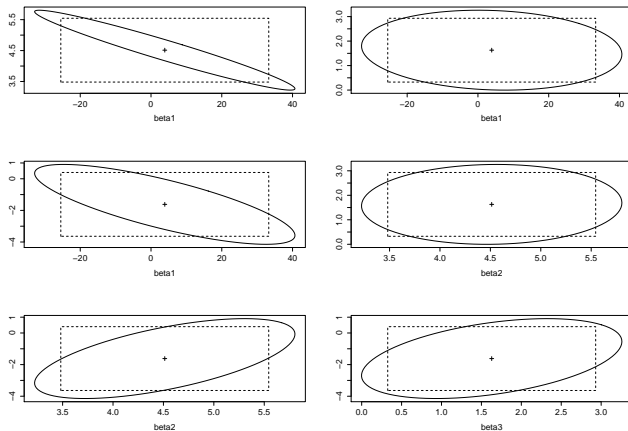


FIGURE – Régions de confiance et rectangle des couples de paramètres

Problème

Nous avons modélisé les pics d'ozone par T12, Vx12 et Ne12. Il paraît raisonnable de se poser les questions suivantes :

- 1 Est-ce que la valeur de O3 est influencée par Vx ?
- 2 Y a-t-il un effet nébulosité ?
- 3 Est-ce que la valeur de O3 est influencée par Vx ou T12 ?

Rappelons que le modèle utilisé est le suivant :

$$O3 = \beta_1 + \beta_2 T12 + \beta_3 Vx12 + \beta_4 Ne12 + \sigma \xi$$

Nous pouvons expliciter les trois questions précédentes en terme de test d'hypo- thèse :

- 1 correspond à $H_0 : \beta_3 = 0$, contre $H_1 : \beta_3 \neq 0$;
- 2 correspond à $H_0 : \beta_4 = 0$, contre $H_1 : \beta_4 \neq 0$;
- 3 correspond à $H_0 : \beta_2 = \beta_3 = 0$, contre $H_1 : \beta_2 \neq 0$ ou $\beta_3 \neq 0$.

Test entre modèles emboîtés

- **Modèle :**

$$Y = X\beta + \sigma\xi \text{ où } \xi \sim \mathcal{N}(0, \sigma^2 \text{Id}_n) ,$$

ce qui implique

$$\mathbb{E}_{\beta, \sigma^2}[Y] = X\beta \in \text{Vect}(X) .$$

- On cherche à tester si

$$\mathbb{E}_{\beta, \sigma^2}[Y] \in \text{Vect}(X_0)$$

où $\text{Vect}(X_0) \subset \text{Vect}(X)$ est un sous espace linéaire (strict) de $\text{Vect}(X)$

- **Exemple typique :** $H_0 : \beta_{j_1} = \dots = \beta_{j_q} = 0$. Dans ce cas, X_0 sont les colonnes de la matrice X qui correspondent aux indices $\{1, \dots, k\} \setminus \{j_1, \dots, j_q\}$

Test du rapport de vraisemblance généralisé

$$\Lambda_n = \frac{\sup_{(\beta_0, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}_0\beta_0 \|^2)}{\sup_{(\beta, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}\beta \|^2)}$$

Test du rapport de vraisemblance généralisé

$$\Lambda_n = \frac{\sup_{(\beta_0, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}_0 \beta_0 \|^2)}{\sup_{(\beta, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X} \beta \|^2)}$$

- On calcule d'abord l'EMV sous le modèle contraint

$$\mathbf{Y} = \mathbb{X}_0 \beta_0 + \sigma \boldsymbol{\xi}.$$

- **régresseur** $\hat{\beta}_{n0} = \mathbb{X}_0^\# \mathbf{Y}$, **variance** $\hat{\sigma}_n^2 = n^{-1} \| \mathbf{Y} - \mathbb{X}_0 \hat{\beta}_{n0} \|^2$
- **vraisemblance**

$$\sup_{(\beta_0, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}_0 \beta_0 \|^2) = \frac{\exp(-n)}{(2\pi n^{-1} \| \mathbf{Y} - \mathbb{X}_0 \hat{\beta}_{n0} \|^2)^{n/2}}$$

Test du rapport de vraisemblance généralisé

$$\Lambda_n = \frac{\sup_{(\beta_0, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}_0 \beta_0 \|^2)}{\sup_{(\beta, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X} \beta \|^2)}$$

- On calcule ensuite l'EMV sous le modèle non contraint
- **régresseur** $\hat{\beta}_n = \mathbb{X}^\# \mathbf{Y}$, **variance** $\hat{\sigma}_n^2 = n^{-1} \| \mathbf{Y} - \mathbb{X} \hat{\beta}_n \|^2$
- **vraisemblance**

$$\sup_{(\beta, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X} \beta \|^2) = \frac{\exp(-n)}{(2\pi n^{-1} \| \mathbf{Y} - \mathbb{X} \hat{\beta}_n \|^2)^{n/2}}$$

Test du rapport de vraisemblance généralisé

$$\Lambda_n = \frac{\sup_{(\beta_0, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}_0 \beta_0 \|^2)}{\sup_{(\beta, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X} \beta \|^2)}$$

- En posant $\hat{\mathbf{Y}} = \mathbb{X} \hat{\beta}_n$ et $\hat{\mathbf{Y}}_0 = \mathbb{X}_0 \hat{\beta}_{n0}$, le RVG est donné par

$$\Lambda_n = \frac{\| \mathbf{Y} - \hat{\mathbf{Y}} \|^n}{\| \mathbf{Y} - \hat{\mathbf{Y}}_0 \|^n}$$

- $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 \in \text{Vect}(\mathbb{X})$ car $\hat{\mathbf{Y}} \in \text{Vect}(\mathbb{X})$ et $\hat{\mathbf{Y}}_0 \in \text{Vect}(\mathbb{X}_0) \subset \text{Vect}(\mathbb{X})$
- $\mathbf{Y} - \hat{\mathbf{Y}} \perp \text{Vect}(\mathbb{X})$ car $\hat{\mathbf{Y}} = \Pi_{\mathbb{X}} \mathbf{Y}$ est la projection orthogonale de \mathbf{Y} sur $\text{Vect}(\mathbb{X})$,
- **Conclusion**

$$\| \mathbf{Y} - \hat{\mathbf{Y}}_0 \|^2 = \| \mathbf{Y} - \hat{\mathbf{Y}} \|^2 + \| \hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 \|^2$$

Test du rapport de vraisemblance généralisé

$$\Lambda_n = \frac{\sup_{(\beta_0, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X}_0 \beta_0 \|^2)}{\sup_{(\beta, \sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \mathbf{Y} - \mathbb{X} \beta \|^2)}$$

- On considère la statistique de test

$$F_n = \frac{\| \hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 \|^2 / q}{\| \mathbf{Y} - \hat{\mathbf{Y}} \|^2 / (n - k)}$$

- Le test du RVG s'écrit donc

$$\Lambda_n = (1 + \{q/(n - k)\} F_n)^{-n/2}$$

- On rejette H_0 si Λ_n est inférieur à un seuil ce qui revient à tester que $F_n > d$.

Distribution du test

Modèle général (sans contrainte) :

$$Y = X\beta + \sigma\xi$$

- Comme $\hat{Y} = \Pi_X Y$ et $\hat{Y}_0 = \Pi_{X_0} Y$ et $\Pi_X \Pi_{X_0} = \Pi_{X_0} \Pi_X = \Pi_{X_0}$,

$$\begin{aligned}\hat{Y} - \hat{Y}_0 &= X\beta + \sigma\Pi_X\xi - \Pi_{X_0}X\beta + \sigma\Pi_{X_0}\xi \\ &= (X\beta - \Pi_{X_0}X\beta) + \sigma\Pi_X(\text{Id}_n - \Pi_{X_0})\xi,\end{aligned}$$

- D'autre part, comme $Y - \hat{Y}$

$$Y - \hat{Y} = \sigma(\text{Id}_n - \Pi_X)\xi.$$

- Par le théorème de Cochran, $\Pi_X(\text{Id}_n - \Pi_{X_0})\xi$ et $(\text{Id}_n - \Pi_X)\xi$ sont **indépendants**
- **Conclusion** : Le numérateur et le dénominateur de la statistique de test sont **indépendants**

$$F_n = \frac{\|\hat{Y} - \hat{Y}_0\|^2/q}{\|Y - \hat{Y}\|^2/(n-k)}$$

Distribution du test sous l'hypothèse nulle

- **Hypothèse nulle** : $\mathbb{X}\beta = \Pi_{\mathbb{X}_0}\mathbb{X}\beta$ car $\mathbb{X}\beta \in \text{Vect}(\mathbb{X}_0)$.
- **Conséquence** : sous H_0 ,

$$\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0 = \sigma \Pi_{\mathbb{X}}(\text{Id}_n - \Pi_{\mathbb{X}_0})\xi$$

- **Conclusion** : par le Théorème de Cochran, sous H_0 ,

$$\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2 / \sigma^2$$

est distribué suivant une variable de χ^2 à q d.d.l., qui est le nombre de coefficients nuls.

Distribution du test sous l'hypothèse nulle

- $(\mathbf{Y} - \hat{\mathbf{Y}})$ est indépendant de $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0$ et $\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2/\sigma^2$ est distribué suivant une loi du χ^2 à $(n - k)$ d.d.l.
- Sous l'hypothèse H_0 , $\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2/\sigma^2$ et $\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$ sont **indépendants** et distribués suivant des lois du χ^2 à q et $n - k$ d.d.l.
- **Conclusion** Sous l'hypothèse H_0 , la statistique de test est donc distribuée suivant **un loi de Fisher** à $(q, n - k)$ d.d.l

Synthèse Test entre modèles emboîtés

- Considérons l'hypothèse nulle : $H_0 : \mathbb{X}\beta \in \text{Vect}(\mathbb{X}_0)$ (où $\text{Vect}(\mathbb{X}_0)$ est un sous-espace vectoriel de $\text{Vect}(\mathbb{X})$).
- **Test** : On rejette H_0 si

$$\frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2/q}{\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2/(n-k)} \geq f_{1-\alpha}(q, n-k)$$

où $f_{1-\alpha}(q, n-k)$ est le quantile $1 - \alpha$ d'une loi de Fisher à $(q, n-k)$ -d.d.l.

Test de Student

- Dans le cas où $H_0 : \beta_j = 0$, pour $j \in \{1, \dots, k\}$, le test est équivalent au test de Student

$$T_j = \frac{\hat{\beta}_j}{\hat{\sigma}_n \sqrt{[(\mathbb{X}^T \mathbb{X})^{-1}]_{i,i}}}$$

- Sous H_0 , T_j est distribué suivant une loi de Student à $(n - p)$ -d.d.l
- Le test rejette H_0 si

$$|T_j| \geq t_{n-k}(1 - \alpha/2)$$

où $t_{n-k}(1 - \alpha/2)$ est le quantile d'ordre $1 - \alpha/2$ de la loi de Student à $(n - p)$ -d.d.l.

Les différents résidus

- Les résidus théoriques $\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}$ sont estimés par

$$\mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \Pi_{\mathbb{X}} \mathbf{Y} = \sigma \Pi_{\mathbb{X}} \boldsymbol{\xi}.$$

- Absence de bias $\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}] = 0$ et $\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[(\mathbf{Y} - \hat{\mathbf{Y}})] = 0$.

- Covariance $\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^T] = \sigma^2 \text{Id}_n$ mais

$$\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})^T] = \sigma^2 (\text{Id}_n - \Pi_{\mathbb{X}}).$$

- Résidus standardisés : Pour $i \in \{1, \dots, n\}$, nous avons :

$$t_i = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_n \sqrt{1 - \pi_{i,i}}}, \quad \pi_{i,i} = [\Pi_{\mathbb{X}}]_{i,i}$$

La loi de ces résidus est difficile à calculer car le **numérateur** et le **dénominateur** sont dépendants.

Les différents résidus

- Les résidus théoriques $\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}$ sont estimés par

$$\mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \Pi_{\mathbb{X}} \mathbf{Y} = \sigma \Pi_{\mathbb{X}} \boldsymbol{\xi}.$$

- Absence de bias $\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}] = \mathbf{0}$ et $\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[(\mathbf{Y} - \hat{\mathbf{Y}})] = \mathbf{0}$.
- Covariance $\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^T] = \sigma^2 \text{Id}_n$ mais

$$\mathbb{E}_{\boldsymbol{\beta}, \sigma^2}[(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})^T] = \sigma^2 (\text{Id}_n - \Pi_{\mathbb{X}}).$$

- Résidus studentisés :

$$t_{n,i}^* = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,(i)} \sqrt{1 - \pi_{i,i}}}, \quad \pi_{i,i} = [\Pi_{\mathbb{X}}]_{i,i}.$$

où $\hat{\sigma}_{n,(i)}$ est l'estimateur de la variance du modèle linéaire, privé de l'observation i .

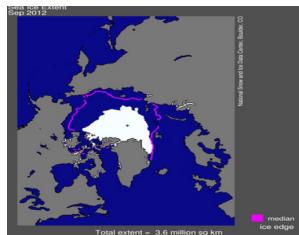
Distribution des résidus studentisés

Théorème

Si la matrice \mathbb{X} est de rang k et $\boldsymbol{\xi} \sim \mathcal{N}(0, \text{Id}_n)$ alors le résidu studentisé

$$t_{n,i}^* = \frac{Y_i - \hat{Y}_i}{\hat{\sigma}_{n,(i)} \sqrt{1 - \pi_{i,i}}}, \quad \pi_{i,i} = [\Pi_{\mathbb{X}}]_{i,i},$$

est distribué suivant une loi de Student à $(n - k - 1)$ d.d.l



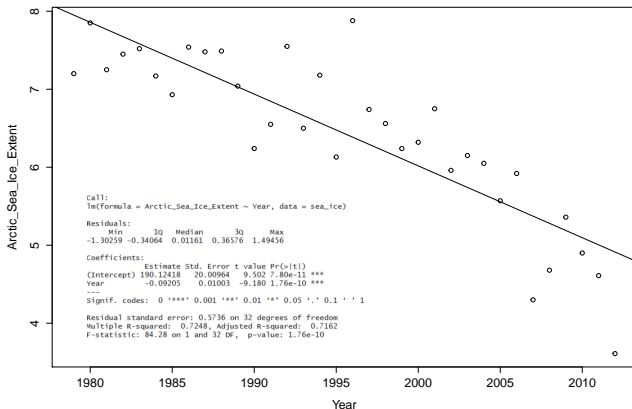


FIGURE – Une pente de -0.0921 montre que la surface de la banquise de 1979-2012 a perdu 92,100 km². C'est à peu près 1/5 de la surface de la France qui disparaît chaque année

la régression est elle-constante ?

Début	Fin	cste	pente
1979	2001	98.3	0.0459
1979	2002	108.5	0.0510
1979	2003	112.0	0.0528
1979	2004	115.6	0.0546
1979	2005	125.2	0.0594
1979	2006	126.8	0.0602
1979	2007	149.5	0.0716
1979	2008	162.2	0.0780
1979	2009	163.4	0.0786
1979	2010	168.8	0.0813
1979	2011	175.4	0.0847
1979	2012	190.1	0.0921

TABLE – La pente semble croître lorsque l'on augmente l'horizon temporel

Résidus standardisés

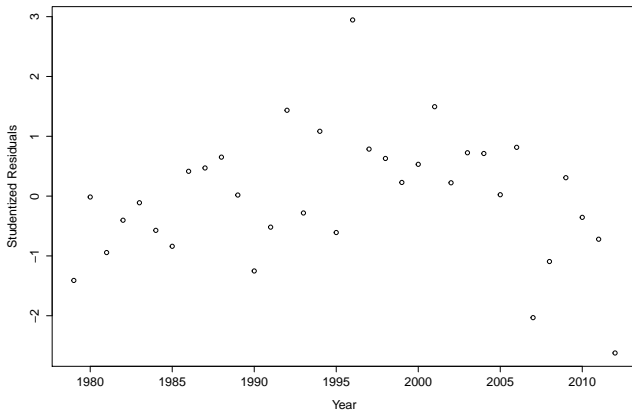


FIGURE – Les résidus studentisés sont négatifs en début et en fin de périodes : suspicion d'un effet non-linéaire

Modèle quadratique

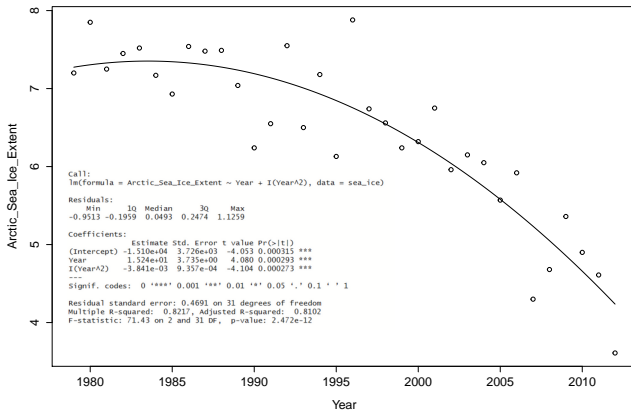


FIGURE – Ajustement quadratique

Choix de modèles

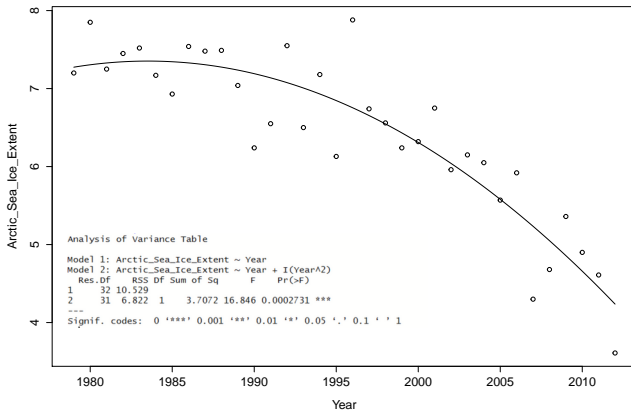


FIGURE – Ajustement quadratique