MAP 433 : Introduction aux méthodes statistiques. Cours 8

16 Octobre 2015

Aujourd'hui

1. Le modèle de régression: quelques rappels
   - Régression linéaire multiple
   - Propriétés de l'estimateur des Moindres Carrés: modèle Gaussien
2. Tests d'hypothèses
3. Analyse des résidus

Modèle de régression

Definition

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

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

- $\boldsymbol{x}_i$ *déterministes, donnés (ou choisis) : plan d'expérience.*
- *Hypothèses sur les* ξ_i *: à débattre. Pour simplifier, 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).*
- *Attention ! Les* Y_i *ne sont pas identiquement distribuées.*
- $\theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Modèle de régression

Definition

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

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

- $\boldsymbol{x}_i$ *déterministes, donnés (ou choisis) : plan d'expérience.*
- *Hypothèses sur les* ξ_i *: à débattre. Pour simplifier, 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).*
- *Attention ! Les* Y_i *ne sont pas identiquement distribuées.*
- $\theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Modèle de régression

Definition

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

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

- $\boldsymbol{x}_i$ *déterministes, donnés (ou choisis) : plan d'expérience.*
- *Hypothèses sur les* ξ_i *: à débattre. Pour simplifier, 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).*
- *Attention ! Les* Y_i *ne sont pas identiquement distribuées.*
- $\theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Modèle de régression

Definition

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

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

- $\boldsymbol{x}_i$ *déterministes, donnés (ou choisis) : plan d'expérience.*
- *Hypothèses sur les* ξ_i *: à débattre. Pour simplifier, 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).*
- *Attention ! Les* Y_i *ne sont pas identiquement distribuées.*
- $\theta = (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$.

Régression gaussienne

- Modèle de régression:

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

- Supposons: $\xi_i \sim \mathcal{N}(0,1)$, i.i.d.
- On a alors le modèle de régression gaussienne. Comment estimer θ? On sait expliciter la loi de l'observation $Z = (Y_1, \ldots, Y_n) \Longrightarrow$ appliquer le principe du maximum de vraisemblance.
- La loi de Y_i:

$$\mathbb{P}^{Y_i}(dy) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}\big(y - r(\boldsymbol{\beta}, \boldsymbol{x}_i)\big)^2\Big)\,dy$$
$$\ll dy.$$

EMV pour régression gaussienne

- Le modèle $\{\mathbb{P}_\theta^n = \text{loi de } (Y_1,\ldots,Y_n), \theta = (\boldsymbol{\beta},\sigma^2) \in \mathbb{R}^k \times \mathbb{R}_+^*\}$ est dominé par $\mu^n(dy_1\ldots dy_n) = dy_1\ldots dy_n$.
- D'où

$$\frac{d\,\mathbb{P}_\theta^n}{d\mu^n}(y_1,\ldots,y_n) = \prod_{i=1}^n \tfrac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\tfrac{1}{2\sigma^2}(y_i - r(\boldsymbol{\beta},\boldsymbol{x}_i))^2\big)$$

- La fonction de vraisemblance

$$\boxed{\mathcal{L}_n(\theta, Y_1,\ldots,Y_n) = \frac{1}{(\sqrt{2\pi\sigma^2})^n}\exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n \big(Y_i - r(\boldsymbol{\beta},\boldsymbol{x}_i)\big)^2\Big)}$$

Estimateur des moindres carrés

Maximiser la vraisemblance en régression gaussienne

$$\widehat{\boldsymbol{\beta}}_{\mathrm{n}} \in \mathrm{argmin}_{b\in\mathbb{R}^k} \sum_{i=1}^{n} \big(Y_i - r(b, \boldsymbol{x}_i)\big)^2$$

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \big(Y_i - r(\widehat{\boldsymbol{\beta}}_{\mathrm{n}}, \boldsymbol{x}_i)\big)^2$$

- L'estimateur $\widehat{\boldsymbol{\beta}}_{\mathrm{n}}$ 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 $r(\boldsymbol{\beta}, \boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$. On observe

$$(\boldsymbol{x}_1, Y_1), \ldots, (\boldsymbol{x}_n, Y_n)$$

avec

$$\boxed{Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \sigma\xi_i, \quad i = 1, \ldots, n}$$

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

- Matriciellement

$$\boxed{\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}}$$

avec

  - $\boldsymbol{Y} = (Y_1 \cdots Y_n)^T$,
  - $\boldsymbol{\xi} = (\xi_1 \cdots \xi_n)^T$
  - $\mathbb{X}$ la matrice $(n \times k)$ dont la i-ème ligne est $\mathbb{X}_{i,\cdot} = \boldsymbol{x}_i^T$.

Régression linéaire multiple (=Modèle linéaire)

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

$$(\boldsymbol{x}_1, Y_1), \ldots, (\boldsymbol{x}_n, Y_n)$$

avec

$$\boxed{Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \sigma \xi_i, \quad i = 1, \ldots, n}$$

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

- Matriciellement

$$\boxed{\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}}$$

avec

  - $\boldsymbol{Y} = (Y_1 \cdots Y_n)^T$,
  - $\boldsymbol{\xi} = (\xi_1 \cdots \xi_n)^T$
  - $\mathbb{X}$ la matrice $(n \times k)$ dont la i-ème ligne est $\mathbb{X}_{i,\cdot} = \boldsymbol{x}_i^T$.

Régression linéaire multiple (=Modèle linéaire)

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

$$(\boldsymbol{x}_1, Y_1), \ldots, (\boldsymbol{x}_n, Y_n)$$

avec

$$\boxed{Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \sigma \xi_i, \quad i = 1, \ldots, n}$$

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

- Matriciellement

$$\boxed{\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}}$$

avec

  - $\boldsymbol{Y} = (Y_1 \cdots Y_n)^T$,
  - $\boldsymbol{\xi} = (\xi_1 \cdots \xi_n)^T$
  - $\mathbb{X}$ la matrice $(n \times k)$ dont la i-ème ligne est $\mathbb{X}_{i,\cdot} = \boldsymbol{x}_i^T$.

EMC en régression linéaire multiple

- Estimateur des moindres carrés en régression linéaire multiple : tout estimateur $\widehat{\boldsymbol{\beta}}_{\mathrm{n}}$ satisfaisant

$$\sum_{i=1}^{n}\big(Y_i-\boldsymbol{x}_i^T\,\widehat{\boldsymbol{\beta}}_{\mathrm{n}}\big)^2=\min_{\mathbf{b}\in\mathbb{R}^k}\sum_{i=1}^{n}\big(Y_i-\boldsymbol{x}_i^T\,\mathbf{b}\big)^2.$$

- En notation matricielle :

$$\begin{aligned}\|\boldsymbol{Y}-\mathbb{X}\,\widehat{\boldsymbol{\beta}}_{\mathrm{n}}\|^2&=\min_{\mathbf{b}\in\mathbb{R}^k}\|\boldsymbol{Y}-\mathbb{X}\boldsymbol{\beta}\|^2\\&=\min_{v\in V}\|\boldsymbol{Y}-v\|^2\end{aligned}$$

où $V=\mathrm{Im}(\mathbb{X})=\{v\in\mathbb{R}^n : v=\mathbb{X}\mathbf{b},\ \mathbf{b}\in\mathbb{R}^k\}$. Projection orthogonale sur V.

Géométrie de l'EMC

- L'EMC vérifie

$$\boxed{\mathbb{X}\,\widehat{\boldsymbol{\beta}}_{\mathrm{n}} = P_V\,\boldsymbol{Y}}$$

  où P_V est le projecteur orthogonal sur V.

- Comme $\boldsymbol{Y} - P_V\,\boldsymbol{Y} \perp V$, on en déduit les équations normales des moindres carrés:

$$\boxed{\mathbb{X}^T\mathbb{X}\widehat{\boldsymbol{\beta}}_{\mathrm{n}} = \mathbb{X}^T\,\boldsymbol{Y}.}$$

- Remarques.
  - L'EMC est un Z-estimateur.
  - unicité de $\widehat{\beta}_{\mathrm{n}}$ si la matrice de Gram $\mathbb{X}^T\mathbb{X}$ est inversible (la matrice $\mathbb{X}$ est de rang complet).

Géométrie de l'EMC

Proposition

Si $\mathbb{X}^T\mathbb{X}$ (matrice $k \times k$) inversible, alors $\widehat{\beta}_n$ est unique et

$$\boxed{\widehat{\beta}_n = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T \boldsymbol{Y}} = \mathbb{X}^{\#}\, \boldsymbol{Y}$$

$$\Pi_{\mathbb{X}} = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T = \mathbb{X}\mathbb{X}^{\#}$$

est dite matrice chapeau (`hat matrix`).

Proposition

Si $\mathbb{X}^T\mathbb{X} > 0$, alors $\Pi_{\mathbb{X}}$ est le projecteur sur V: $\Pi_{\mathbb{X}} = P_V$ et $\mathrm{rang}(\Pi_{\mathbb{X}}) = k$.

Hypothèses

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

1. $\mathbb{X}$ est de rang complet.
2. $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{Id}_n)$

Hypothèses

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

1. $\mathbb{X}$ est de rang complet.
2. $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{Id}_n)$

Propriétés de l'estimateur

Théorème

Pour tout $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$, *sous* $\mathbb{P}_{\boldsymbol{\beta},\sigma^2}$, *l'estimateur* $\widehat{\boldsymbol{\beta}}_{\mathrm{n}}$ *est un vecteur Gaussien de moyenne* $\boldsymbol{\beta}$ *et de variance* $\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$

Proof.

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\mathrm{n}} &= \mathbb{X}^{\#}\,\boldsymbol{Y} = \mathbb{X}^{\#}(\mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}) \\ &= \boldsymbol{\beta} + \sigma\mathbb{X}^{\#}\boldsymbol{\xi}\end{aligned}$$

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

$$\mathbb{X}^{\#}\left(\mathbb{X}^{\#}\right)^T = (\mathbb{X}^T\mathbb{X})^{-1}.$$

$\square$

Prédiction et Erreur de prédiction

- Prédiction
$$\hat{\boldsymbol{Y}} = \mathbb{X}\,\widehat{\boldsymbol{\beta}}_{\mathrm{n}} = \Pi_{\mathbb{X}}\,\boldsymbol{Y}$$
projection des observations sur l'espace de régression.
- Erreur de prédiction:
$$\hat{\boldsymbol{\xi}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\mathrm{Id}_n - \Pi_{\mathbb{X}})\,\boldsymbol{Y}\ .$$
- Sous $\mathbb{P}_\theta$, $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}$. Donc,
$$\hat{\boldsymbol{Y}} = \mathbb{X}\boldsymbol{\beta} + \sigma\Pi_{\mathbb{X}}\boldsymbol{\xi}$$
$$\hat{\boldsymbol{\xi}} = \sigma(\mathrm{Id}_n - \Pi_{\mathbb{X}})\boldsymbol{\xi}$$
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

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

1. $\Pi\, \boldsymbol{Y} \sim \mathcal{N}(\Pi\mu, \sigma^2\Pi)$, $\Pi_\perp\, \boldsymbol{Y} \sim \mathcal{N}(\Pi_\perp\mu, \sigma^2\Pi_\perp)$
2. *les vecteurs* $\Pi\, \boldsymbol{Y}$ *et* $\Pi_\perp\, \boldsymbol{Y}$ *sont indépendants*
3. $\|\Pi(\boldsymbol{Y} - \mu)\|^2/\sigma^2 \sim \chi^2_k$ *et* $\Pi_\perp(\boldsymbol{Y} - \mu)\|^2/\sigma^2 \sim \chi^2_{n-k}$.

Théorème de Cochran

Théorème

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

1. $\Pi \boldsymbol{Y} \sim \mathcal{N}(\Pi\mu, \sigma^2 \Pi)$, $\Pi_\perp \boldsymbol{Y} \sim \mathcal{N}(\Pi_\perp \mu, \sigma^2 \Pi_\perp)$
2. *les vecteurs* $\Pi \boldsymbol{Y}$ *et* $\Pi_\perp \boldsymbol{Y}$ *sont indépendants*
3. $\|\Pi(\boldsymbol{Y} - \mu)\|^2/\sigma^2 \sim \chi^2_k$ et $\Pi_\perp(\boldsymbol{Y} - \mu)\|^2/\sigma^2 \sim \chi^2_{n-k}$.

Théorème de Cochran

Théorème

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

1. $\Pi\,\boldsymbol{Y} \sim \mathcal{N}(\Pi\mu, \sigma^2\Pi)$, $\Pi_\perp\,\boldsymbol{Y} \sim \mathcal{N}(\Pi_\perp\mu, \sigma^2\Pi_\perp)$
2. *les vecteurs* $\Pi\,\boldsymbol{Y}$ *et* $\Pi_\perp\,\boldsymbol{Y}$ *sont indépendants*
3. $\|\Pi(\boldsymbol{Y} - \mu)\|^2/\sigma^2 \sim \chi^2_k$ *et* $\Pi_\perp(\boldsymbol{Y} - \mu)\|^2/\sigma^2 \sim \chi^2_{n-k}$.

Résidus et variance résiduelle

Sous $\mathbb{P}_{\boldsymbol{\beta},\sigma^2}$, $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}$,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{n}} = \boldsymbol{\beta} + \sigma\mathbb{X}^{\#}\boldsymbol{\xi}$$
$$\hat{\boldsymbol{\xi}} = \sigma(\mathrm{Id}_n - \Pi_{\mathbb{X}})\boldsymbol{\xi}$$
$$\hat{\sigma}^2 = (n-k)^{-1}\|\hat{\boldsymbol{\xi}}\|^2$$

Theorem

Pour tout $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}_+$,

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

Proof.

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

Résidus et variance résiduelle

Sous $\mathbb{P}_{\boldsymbol{\beta},\sigma^2}$, $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi}$,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathrm{n}} &= \boldsymbol{\beta} + \sigma\mathbb{X}^{\#}\boldsymbol{\xi} \\
\hat{\boldsymbol{\xi}} &= \sigma(\mathrm{Id}_n - \Pi_{\mathbb{X}})\boldsymbol{\xi} \\
\hat{\sigma}^2 &= (n-k)^{-1}\|\hat{\boldsymbol{\xi}}\|^2
\end{aligned}$$

Theorem

Pour tout $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}_+$,

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

Proof.

$$\widehat{\boldsymbol{\beta}}_{\mathrm{n}} = \boldsymbol{\beta} + \sigma\mathbb{X}^{\#}\boldsymbol{\xi} = \boldsymbol{\beta} + \sigma\mathbb{X}^{\#}\Pi_{\mathbb{X}}\boldsymbol{\xi}$$

et $\Pi_{\mathbb{X}}\boldsymbol{\xi}$ et $(\mathrm{Id}_n - \Pi_{\mathbb{X}})\boldsymbol{\xi}$ sont indépendants. $\square$

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. Par définition 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 une Loi du χ^2 à d_1 et d_2 degrés de liberté. Par définition, 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.

Lois des estimateurs

Théorème

Pour tout $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}_+$,

1. *Pour* $j = 1, \ldots, k$,

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

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

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

Proof.

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

$\square$

Lois des estimateurs

Théorème

Pour tout $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^k \times \mathbb{R}_+$,

1. *Pour* $j = 1, \ldots, k$,

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

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

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

Proof.

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

$\square$

Intervalles de confiance

Théorème

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

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

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

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

Régions de confiance

Théorème

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

$$\left\{ R\beta \in \mathbb{R}^q, \frac{1}{q\hat{\sigma}^2}[R(\hat{\beta}-\beta)]'[R(X'X)^{-1}R']^{-1}[R(\hat{\beta}-\beta)] \leq f_{q,n-p}(1-\alpha) \right\},$$

où R est la matrice de taille $q \times p$ dont tous les éléments sont nuls sauf les $[R]_{ij_i}$, $i = 1, \ldots, q$ qui valent 1 et $f_{q,n-p}(1-\alpha)$ est le quantile de niveau $(1-\alpha)$ d'une loi de Fisher admettant $(q, \; n-p)$ d.l.

Un exemple

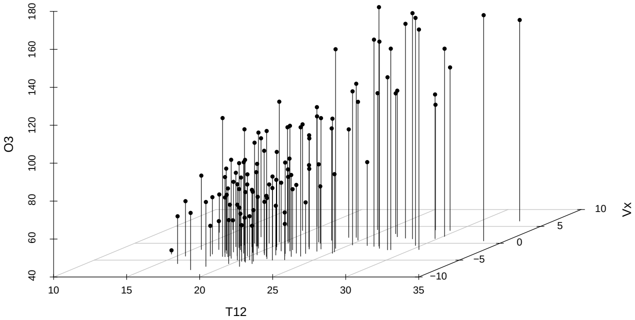

Figure: Représentation brute des données: modèle d'explication de l'ozone (O3) par la température à 12h (T12) et le Vent à 12h (Vx12)

Régression multiple

Modèle de régression

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

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

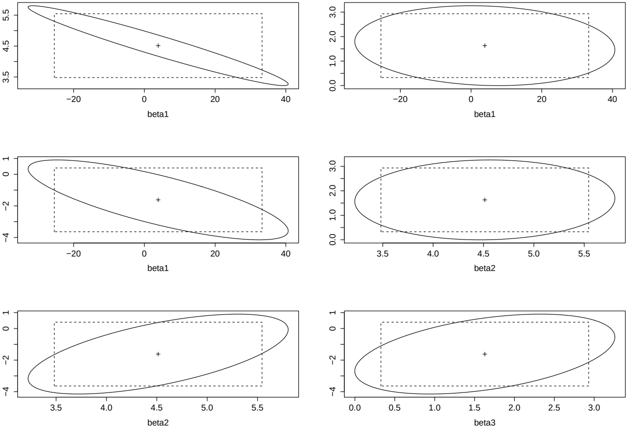

Figure: Régions de confiance et rectangle des couples de paramètres

Problème

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

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

Rappelons que le modèle utilisé est le suivant:

$$\mathrm{O3} = \beta_1 + \beta_2 \mathrm{T12} + \beta_3 \mathrm{Vx12} + \beta_4 \mathrm{Ne12} + \sigma \boldsymbol{\xi}$$

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

1. correspond à $\mathrm{H}_0 : \beta_3 = 0$, contre $\mathrm{H}_1 : \beta_3 \neq 0$;
2. correspond à $\mathrm{H}_0 : \beta_4 = 0$, contre $\mathrm{H}_1 : \beta_4 \neq 0$;
3. correspond à $\mathrm{H}_0 : \beta_2 = \beta_3 = 0$, contre $\mathrm{H}_1 : \beta_2 \neq 0$ ou $\beta_3 \neq 0$.

Test entre modèles emboîtés

- Modèle:
$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\xi} \text{ où } \boldsymbol{\xi} \sim \mathcal{N}(0,\ \sigma^2 \mathrm{Id}_n)\ ,$$
ce qui implique
$$\mathbb{E}_{\boldsymbol{\beta},\sigma^2}[\boldsymbol{Y}] = \mathbb{X}\boldsymbol{\beta} \in \mathrm{Vect}(\mathbb{X})\,.$$
- On cherche à tester si
$$\mathbb{E}_{\boldsymbol{\beta},\sigma^2}[\boldsymbol{Y}] \in \mathrm{Vect}(\mathbb{X}_0)$$
où $\mathrm{Vect}(\mathbb{X}_0) \subset \mathrm{Vect}(\mathbb{X})$ est un sous espace linéaire (strict) de $\mathrm{Vect}(\mathbb{X})$
- Exemple typique: H_0: $\beta_{j_1} = \cdots = \beta_{j_q} = 0$. Dans ce cas, $\mathbb{X}_0$ sont les colonnes de la matrice $\mathbb{X}$ qui correspondent aux indices $\{1, \ldots, k\} \setminus \{j_1, \ldots, j_q\}$

Test du rapport de vraisemblance généralisé

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

Test du rapport de vraisemblance généralisé

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

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

$$\boldsymbol{Y} = \mathbb{X}_0 \boldsymbol{\beta}_0 + \sigma \boldsymbol{\xi}\,.$$

- régresseur $\widehat{\boldsymbol{\beta}}_{\mathrm{n}0} = \mathbb{X}_0^{\#} \boldsymbol{Y}$, variance $\hat{\sigma}^2 = n^{-1} \| \boldsymbol{Y} - \mathbb{X}_0 \widehat{\boldsymbol{\beta}}_{\mathrm{n}0} \|^2$
- vraisemblance

$$\sup_{(\boldsymbol{\beta}_0,\sigma^2)} (2\pi\sigma^2)^{-n/2} \exp(-1/(2\sigma^2) \| \boldsymbol{Y} - \mathbb{X}_0 \boldsymbol{\beta}_0 \|^2) = \frac{exp(-n)}{(2\pi n^{-1} \| \boldsymbol{Y} - \mathbb{X}_0 \widehat{\boldsymbol{\beta}}_{\mathrm{n}0} \|^2)^{n/2}}$$

Test du rapport de vraisemblance généralisé

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

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

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

Test du rapport de vraisemblance généralisé

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

- En posant $\hat{\boldsymbol{Y}} = \mathbb{X}\,\widehat{\boldsymbol{\beta}}_{\mathrm{n}}$ et $\hat{\boldsymbol{Y}}_0 = \mathbb{X}_0\,\widehat{\boldsymbol{\beta}}_{\mathrm{n}0}$, le RVG est donné par

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

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

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

Test du rapport de vraisemblance généralisé

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

- On considère la statistique de test

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

- Le test du RVG s'écrit donc

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

- On rejette H0 si Λ_n est inférieur à un seuil ce qui revient à tester que $F_n > d$.

Distribution du test

Modèle général (sans contrainte):

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

- Comme $\hat{\boldsymbol{Y}} = \Pi_{\mathbb{X}}\,\boldsymbol{Y}$ et $\hat{\boldsymbol{Y}}_0 = \Pi_{\mathbb{X}_0}\,\boldsymbol{Y}$ et $\Pi_{\mathbb{X}}\Pi_{\mathbb{X}_0} = \Pi_{\mathbb{X}_0}\Pi_{\mathbb{X}} = \Pi_{\mathbb{X}_0}$,

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

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

$$\boldsymbol{Y} - \hat{\boldsymbol{Y}} = \sigma(\mathrm{Id}_n - \Pi_{\mathbb{X}})\boldsymbol{\xi}\,.$$

- Par le théorème de Cochran, $\Pi_{\mathbb{X}}(\mathrm{Id}_n - \Pi_{\mathbb{X}_0})\boldsymbol{\xi}$ et $(\mathrm{Id}_n - \Pi_{\mathbb{X}})\boldsymbol{\xi}$ sont indépendants
- Conclusion: Le numérateur et le dénominateur de la statistique de test sont indépendants

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

Distribution du test sous l'hypothèse nulle

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

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

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

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

  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

- $(\boldsymbol{Y} - \hat{\boldsymbol{Y}})$ est indépendant de $\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_0$ et $\| \boldsymbol{Y} - \hat{\boldsymbol{Y}} \|^2 / \sigma^2$ est distribué suivant une loi du χ^2 à $(n-k)$ d.d.l.
- Sous l'hypothèse H_0, $\| \hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_0 \|^2 / \sigma^2$ et $\| \boldsymbol{Y} - \hat{\boldsymbol{Y}} \|^2$ 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: $\mathbb{X}\boldsymbol{\beta} \in \mathrm{Vect}(\mathbb{X}_0)$ (où $\mathrm{Vect}(\mathbb{X}_0)$ est un sous-espace vectoriel de $\mathrm{Vect}(\mathbb{X})$).
- Test: On rejette H_0 si

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

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 $j \in \{1, \ldots, k\}$, le test est équivalent au test de Student
$$T_j = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{[(\mathbb{X}^T\mathbb{X})^{-1}]_{i,i}}}$$

- 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 $\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}$ sont estimés par

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

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

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

- Résidus standardisés: Pour $i \in \{1, \ldots, n\}$, 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 $\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}$ sont estimés par

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

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

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

- 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* $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{Id}_n)$ *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*

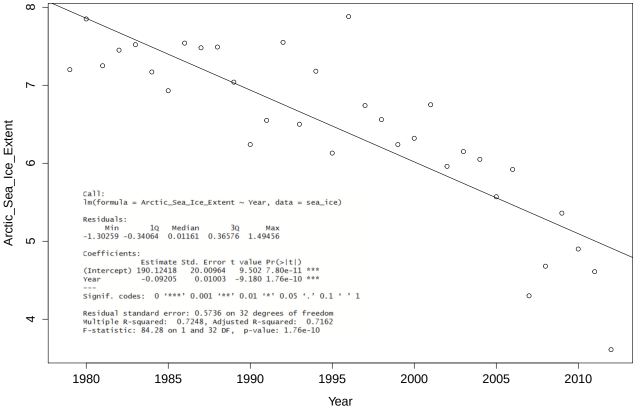

Figure: Une pente de -0.0921 montre que la surface de la banquise de 1979-2012 a perdu 92,100 km^2. C'est à peu près $1/5$ de la surface de la France qui disparaît chaque année

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

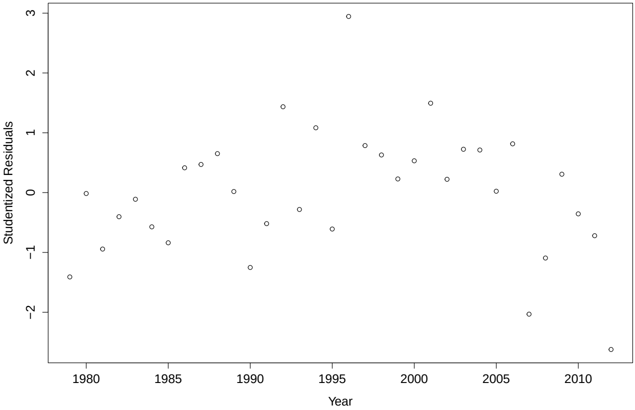

Figure: Les résidus studentisés sont négatifs en début et en fin de périodes: suspicion d'un effet non-linéaire

Modèle quadratique

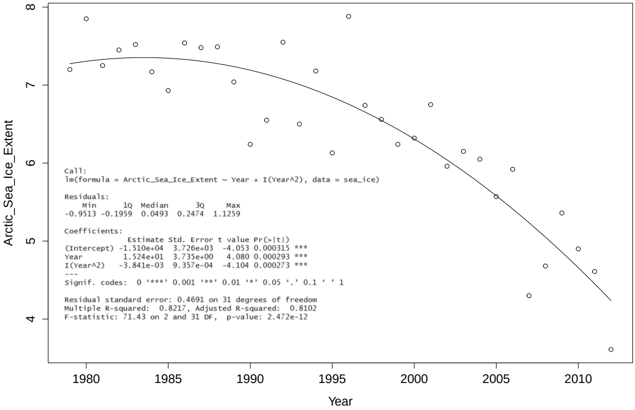

Figure: Ajustement quadratique

Choix de modèles

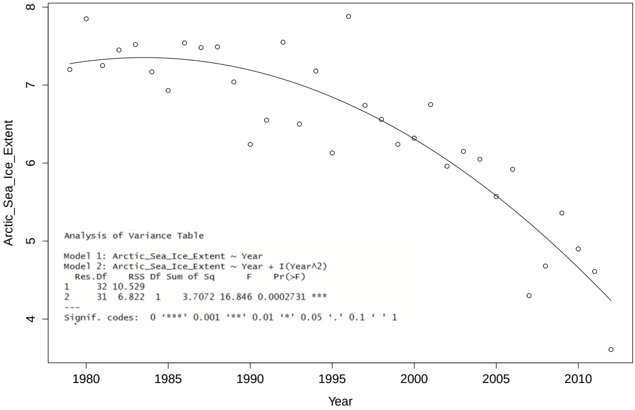

Figure: Ajustement quadratique