Exploration Numérique 1

30 août 2023

Soient $t_1, \dots, t_n$ des réels ; on suppose qu'il existe au moins deux indices $i \neq j$ tels que $t_i \neq t_j$. On considère le modèle statistique

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{ p_\theta \cdot \mathrm{dLeb}^{\otimes n} := \bigotimes_{i=1}^{n} \mathrm{N}(\beta_1 + \beta_2 t_i, \sigma^2) \ : \ \theta := (\beta_1, \beta_2, \sigma^2) \in \Theta = \mathbb{R}^2 \times \mathbb{R}_+^* \right\} \right) .$$

On note $\mathbf{1}$ et $\mathbf{t}$ les vecteurs de $\mathbb{R}^n$ définis par $\mathbf{1} := (1, \dots, 1)^T$ et $\mathbf{t} := (t_1, \dots, t_n)^T$. Dans la suite, on pose

$$s := \frac{1}{n} \|\mathbf{t}\|^2 - (\bar{t})^2 , \quad \text{avec} \quad \bar{t} := n^{-1} \sum_{i=1}^{n} t_i .$$

Nous notons X_i, $i \in \{1, \dots, n\}$ les observations canoniques et posons $\mathbf{X} = [X_1, \dots, X_n]^T$ et $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$.

On appelle estimateurs des *moindres carrés* de β_1 et β_2 les estimateurs $\hat{\beta}_1$ et $\hat{\beta}_2$ obtenus en minimisant

$$S(\beta_1, \beta_2) = \sum_{i=1}^{n} (X_i - \beta_1 - \beta_2 t_i)^2 = \|\mathbf{X} - \beta_1 \mathbf{1} - \beta_2 \mathbf{t}\|^2 .$$

Les questions théoriques ont été traitées en PC. Une rédaction succincte est simplement demandée [il vous suffit de rappeler les résultats pour référence, mais soyez sûr de savoir les établir !]

1. Montrer que

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} (t_i - \bar{t}) X_i}{\sum_{i=1}^{n} (t_i - \bar{t})^2} \quad \text{et} \quad \hat{\beta}_1 = \bar{X} - \hat{\beta}_2 \bar{t} . \tag{1}$$

Nous notons $\hat{X}_i = \hat{\beta}_1 + \hat{\beta}_2 t_i$, $\hat{\varepsilon}_i = X_i - \hat{X}_i$ le résidu de prédiction et

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 / (n-2) \tag{2}$$

On rappelle que

— $\frac{(n-2)}{\sigma^2} \hat{\sigma}^2$ suit une loi du χ^2 à $(n-2)$ degrés de liberté.

— Pour $j = 1, 2$, $\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_j}$ suit une loi de Student à $(n-2)$ degrés de liberté où

$$\hat{\sigma}_1^2 = \hat{\sigma}^2 \left(\frac{\sum_{i=1}^{n} t_i^2}{n \sum_{i=1}^{n} (t_i - \bar{t})^2} \right) \quad \text{et} \quad \hat{\sigma}_2^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (t_i - \bar{t})^2}$$

Nous considérons les données des anomalies de températures annuelles à la surface du globe décrites sur `https://data.giss.nasa.gov/gistemp/`. On utilisera les données fournies dans `https://data.giss.nasa.gov/gistemp/tabledata_v4/GLB.Ts+dSST.txt` (txt) ou `https://data.giss.nasa.gov/gistemp/tabledata_v4/GLB.Ts+dSST.csv` (csv). Dans les applications numériques t_i est l'année.

2. Visualiser les données d'anomalies de températures sur la période 1880-2020.
3. Estimer les paramètres $\hat{\beta}_1$, $\hat{\beta}_2$ et $\hat{\sigma}^2$ en utilisant les données d'anomalies sur des intervalles de temps de 40 ans (1880-1920), (1890-1930), etc en vous décalant de 10 ans.
4. Pour chaque paramètre et chaque intervalle de temps déterminer les intervalles de confiance de niveau de couverture 0.95.
5. Visualiser en superposition du graphe des observations les différentes droites de régression $\hat{X}_i = \hat{\beta}_1 + \hat{\beta}_2 t_i$ et les intervalles de confiance de prédiction de niveau de couverture 0.95.
6. Visualiser pour chaque intervalle les résidus de standardisés - voir Regression Linéaire avec R, chapitre 4.
7. Conclure.