Autour du modèle linéaire simple

(Devoir à la maison)

4 postulats du modèle

Pour tous $i = 1, \ldots, n$, $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, ε_i sont i.i.d. de loi gaussienne et X_i sont déterministes.

Les estimateurs

$$\hat{\beta} = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}, \quad \hat{\mu} = \bar{Y} - \hat{\beta}\bar{X}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-2}$$

où $\mathrm{Cov}(Y, X) = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})$, $\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$, $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ et $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$.

Les propriétés des estimateurs

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{\sigma^2 \sum X_i^2}{n^2 \mathrm{Var}(X)}\right), \hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{n\mathrm{Var}(X)}\right), \mathrm{Cov}(\hat{\mu}, \hat{\beta}) = -\frac{\sigma^2 \bar{X}}{n\mathrm{Var}(X)}, \frac{\hat{\sigma}^2(n-2)}{\sigma^2} \sim \chi^2(n-2),$$

$(\hat{\mu}, \hat{\beta})$ et $\hat{\sigma}^2$ sont indépendantes.

Loi du χ^2 à n degrés de liberté

Soient $X_1, \ldots, X_n$ n v. a. indép. de loi $\mathcal{N}(0, 1)$, alors

$$S = X_1^2 + \cdots + X_n^2$$

suit une loi du χ^2 à n degrés de liberté, notée $\chi^2(n)$.

Loi de Student à n degrés de liberté

La loi de Student à n degrés de liberté, notée $T(n)$, est la loi du quotient

$$T = \frac{N}{\sqrt{S/n}}$$

où N suit une loi $\mathcal{N}(0, 1)$ et S suit une loi $\chi^2(n)$, N et S étant deux v. a. indép..

Loi de Fisher à n_1 et n_2 degrés de liberté

Soient S_1 et S_2 deux v. a. indép. de loi respectives $\chi^2(n_1)$ et $\chi^2(n_2)$. Alors le quotient

$$F = \frac{S_1/n_1}{S_2/n_2}$$

suit une loi de Fisher à n_1 et n_2 degrés de liberté, notée $F(n_1, n_2)$.

Théorème de Cochran

Soient E_1 et E_2 deux sous-espaces vectoriels orthogonaux de $E = \mathbb{R}^d$ de dimensions respectives k_1 et k_2 et soit Y un vecteur aléatoire de $\mathbb{R}^d$ de loi normale centrée isotrope de variance σ^2. Alors $P_{E_1}(Y)$ et $P_{E_2}(Y)$ sont deux v. a. gaussienne centrées indépendantes et $\|P_{E_1}(Y)\|^2$ (resp. $\|P_{E_2}(Y)\|^2$) est une loi $\sigma^2\chi^2(k_1)$ (resp. $\sigma^2\chi^2(k_2)$).

Exercice 1

Montrer les propriétés des estimateurs suivantes.

1. $\mathbb{E}\,(\hat{\beta}) = \beta$
2. $\mathbb{E}\,(\hat{\mu}) = \mu$
3. $\mathrm{Var}\,(\hat{\beta}) = \dfrac{\sigma^2}{n\mathrm{Var}\,X}$

(Indication : $\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X}) = \sum_{i=1}^n (X_iY_i - \bar{X}\bar{Y})$)

4. Admettons les propriétés des estimateurs résumées dans la page précédentes, montrer que $\mathrm{Var}\,(\hat{Y}) = \dfrac{\sigma^2}{n}\left(1 + \dfrac{(X-\bar{X})^2}{\mathrm{Var}\,X}\right)$.

Exercice 2

Montrer que SCR $= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ est une v. a. de loi $\sigma^2\chi^2(n-2)$.

Exercice 3

Montrer que sous l'hypothèse nulle : $\beta = 0$, la statistique $\hat{t} = \dfrac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$ suit la loi $T(n-2)$, où $\hat{\sigma}^2_{\hat{\beta}} = \dfrac{\hat{\sigma}^2}{n\mathrm{Var}\,X}$.

Exercice 4

Montrer que sous l'hypothèse nulle : $\beta = 0$, la statistique $\hat{F} = (n-2)\dfrac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$ suit la loi $F(1, n-2)$.

Exercice 5

Notons Y la nouvelle observation qu'on veut prédire. Montrer que l'intervalle de confiance au niveau $1-\alpha$ pour $\mathbb{E}\,(Y)$ est

$$\left[\hat{Y} - t_{n-2}(\alpha/2)\sqrt{\widehat{\mathrm{Var}}\,(\hat{Y})}, \hat{Y} + t_{n-2}(\alpha/2)\sqrt{\widehat{\mathrm{Var}}\,(\hat{Y})}\right]$$

et l'intervalle de confiance au niveau $1-\alpha$ pour Y est

$$\left[\hat{Y} - t_{n-2}(\alpha/2)\sqrt{\widehat{\mathrm{Var}}\,(\hat{Y}) + \hat{\sigma}^2}, \hat{Y} + t_{n-2}(\alpha/2)\sqrt{\widehat{\mathrm{Var}}\,(\hat{Y}) + \hat{\sigma}^2}\right]$$

où $\widehat{\mathrm{Var}}\,(\hat{Y}) = \dfrac{\hat{\sigma}^2}{n}\left(1 + \dfrac{(X-\bar{X})^2}{\mathrm{Var}\,X}\right)$ et $t_{n-2}(\alpha/2)$ est le $\alpha/2$-quantile d'une loi de Student à $n-2$ degrés de liberté.