

TP2

Régression linéaire simple (2) : données simulées

Objectifs du TP

- Simuler plusieurs jeux de données selon un même modèle.
- Observer la précision et la variabilité des estimateurs par moindres carrés en régression linéaire.
- Illustrer les résultats du cours sur la loi de ces estimateurs.

Données simulées

On considère le modèle

$$y = \beta_1 + \beta_2 x + \varepsilon \quad (1)$$

avec x, ε indépendantes, x de loi uniforme sur $[0, 1]$ et ε de loi gaussienne de moyenne 0 et de variance σ^2 .

On fixe $\beta_1 = -1$, $\beta_2 = 3$ et $\sigma^2 = 2$.

1. Simuler $n = 100$ réalisations indépendantes $(x_i, y_i)_{1 \leq i \leq n}$ du couple (x, y) défini par (1) (on pourra utiliser les fonctions `runif` et `rnorm`).
2. À l'aide du seul échantillon $(x_i, y_i)_{1 \leq i \leq n}$, estimer par moindres carrés les paramètres β_1 et β_2 du modèle (1), ainsi que la variance résiduelle σ^2 .
Comparer les valeurs estimées $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\sigma}^2$ aux valeurs exactes de ces paramètres.
3. Répéter les questions précédentes en changeant la valeur de la graine du générateur de nombres aléatoires (utiliser la fonction `set.seed`). Que constate-t-on ?
4. À l'aide d'un script `R`, stocker les valeurs estimées $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\sigma}^2$ obtenues pour $N = 200$ échantillons indépendants.
 - (a) Estimer la moyenne, la médiane, la variance de $\hat{\beta}_1$. Commenter.
 - (b) Représenter la loi de $\hat{\beta}_1$ sous forme d'histogramme. Commenter.
 - (c) Reprendre les deux questions ci-dessus en remplaçant $\hat{\beta}_1$ par $\hat{\beta}_2$, puis $\hat{\sigma}^2$.
5. Même question en générant une fois pour toutes les x_i (ici, seuls les y_i changent d'un échantillon à l'autre).
6. (Question subsidiaire) Changer les valeurs de $\beta_1, \beta_2, \sigma^2$ pour rendre le problème de régression linéaire plus difficile / plus facile.
Même chose en changeant uniquement la loi des x_i .