Université Paris 7 — **Économétrie**
Licence 3, 2009-2010 — **Pierre Alquier - Joseph Salmon**

TD N° 2 : Modèle linéaire simple

On s'intéresse dans tout ce TD au modèle suivant :

$$\forall i \in \{1, ..., n\}, \quad y_i = b_0 + b_1 x_i + \varepsilon_i \tag{1}$$

avec les hypothèses suivantes
- **(H0)** les valeurs des x_i sont supposées déterministes (non-aléatoires) ;
- **(H1)** les ε_i sont aléatoires avec $\mathbb{E}(\varepsilon_i) = 0$;
- **(H2)** les ε_i sont indépendants les uns des autres ;
- **(H3)** pour tout i, $\mathrm{Var}(\varepsilon_i) = \sigma^2$;
- **(H4)** les ε_i sont gaussiens.

Le premier exercice est une étude théorique de ce modèle. Le deuxième exercice traite un exemple (modélisation de la relation entre rendement agricole et utilisation d'engrais). Enfin, les deux derniers exercices s'interrogent sur ce qui se passe si on modifie certaines des hypothèses **(H0)**-**(H4)**.

Exercice 1. Dans le contexte donné par l'Equation (1), on observe (x_1, y_1), ..., (x_n, y_n).

1) Ecrire le modèle sous forme matricielle.
2) On souhaite calculer l'estimateur des moindres carrés de b_0 et de b_1 "à la main", sans utiliser la formule du cours. On rappelle que $(\hat{b}_0, \hat{b}_1)$ minimise la fonction

$$\mathrm{sc}(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[y_i - \Big(\beta_0 + \beta_1 x_i \Big) \right]^2 .$$

   Quelle est la condition sur les x_i sous laquelle ce minimum est unique ? On appellera **(H5)** cette hypothèse et on la supposera vérifiée dans la suite.
3) Déterminer $(\hat{b}_0, \hat{b}_1)$ en fonction de (x_1, y_1), ..., (x_n, y_n).
4) A partir de la formule obtenue à la question 3), calculer $\mathbb{E}(\hat{b}_0)$ et $\mathbb{E}(\hat{b}_1)$.
5) A partir de la question 3), calculer $\mathrm{Var}(\hat{b}_1)$ et donner la loi de $\hat{b}_1$.

Exercice 2. On va maintenant traiter un exemple. On souhaite modéliser y le rendement de maïs (en quintal) d'une parcelle de terrain en fonction de x la quantité d'engrais utilisée (en kilos). On a les données complètes dans la Table 1.

De façon à éviter de faire trop de calculs, on donne les quantités suivantes :

$$\sum x_i = 261 \ , \ \sum y_i = 304 \text{ et } \sum x_i y_i = 8286.$$

1) Donner la valeur des estimateurs $\hat{b}_0$ et $\hat{b}_1$.
2) Quel rendement prévoit le modèle pour une parcelle ayant reçu 20 kilos d'engrais ?
3) De façon générale, pour une augmentation $\Delta x = 1$ kg d'engrais, quelle est l'augmentation estimée du rendement Δy ?
4) Rappeler la formule de l'estimateur de la variance $\hat{\sigma^2}$ vue en cours et calculer cette valeur.

TAB. 1 – Données pour l'exercice 2.

i	x_i	y_i
1	16	20
2	18	24
3	23	28
4	24	22
5	28	32
6	29	28
7	26	32
8	31	36
9	32	41
10	34	41

EXERCICE 3. On revient au modèle théorique donné par l'Equation (1). On suppose que les hypothèses **(H0)**, **(H2)**, **(H3)**, **(H4)** et **(H5)** sont toujours vraies, mais que **(H1)** n'est plus vraie :

$$\mathbb{E}(\varepsilon_i) \neq 0.$$

Montrer que l'on peut toujours estimer le paramètre b_1. Quelles en sont les conséquences (cf. par exemple la Question 3 de l'Exercice 2) ?

EXERCICE 4. On étudie toujours le modèle théorique donné par l'Equation (1), dans le cas particulier où y_i est la consommation du ménage i et x_i son revenu. On suppose que l'hypothèse **(H0)** est vraie, et que

$$\varepsilon_i = e_i \sqrt{x_i}$$

où les e_i, pour $1 \leq i \leq n$, sont iid de loi $\mathcal{N}(0, \sigma^2)$.

1) Vérifier que les hypothèses **(H1)**, **(H2)** et **(H4)** sont toujours vraies.
2) Qu'en est-il de l'hypothèse **(H3)** ?
3) Discuter la pertinence de cette modélisation.

Remarque : on verra plus tard dans le cours comment estimer les paramètres dans le cas de l'Exercice 4.