UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS
L3 MASS - Statistiques - 2008-2009

TD5 - Corrigé

Ex. 1. On trouve $\widehat{\beta}_2 = 0,39$, $\widehat{\beta}_1 = -12,39$, $R^2 = 0.77$. On ne peut pas en dire grand chose, R^2 est proche de 1, mais vu la courbe, le modèle linéaire n'est pas adapté.

Ex. 2. On utilise que $\mathbb{E}(\epsilon_i) = 0$.

$$\begin{aligned}
\mathbb{E}(\widehat{\beta}_2) &= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_1 + \beta_2 x_i + \mathbb{E}(\epsilon_i))}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_1 + \beta_2 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_2 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \beta_2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \beta_2 \ .
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\widehat{\beta}_1) &= \mathbb{E}(\bar{y}) - \mathbb{E}(\widehat{\beta}_2)\bar{x} \\
&= \beta_1 + \beta_2 \bar{x} - \beta_2 \bar{x} \\
&= \beta_1 \ .
\end{aligned}$$

Ex. 3. On utilise que $\mathbb{E}(\epsilon_i) = 0$, $\mathbb{E}(\epsilon_i^2) = \sigma^2$.

$$\begin{aligned}
\mathbb{E}(\widehat{\beta}_2^2) &= \mathbb{E}\left(\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_1 + \beta_2 x_i + \epsilon_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2\right) \\
&= \frac{\left(\sum_{i=1}^n (x_i - \bar{x})(\beta_1 + \beta_2 x_i)\right)^2 + \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\
&= \frac{\beta_2^2 \left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2 + \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\
&= \beta_2^2 + \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \ .
\end{aligned}$$

Donc, en utilisant le résultat de l'exercice 2, $\mathrm{Var}(\widehat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

Ex. 4. On utilise que $\mathbb{E}(\epsilon_i) = 0$, $\mathbb{E}(\epsilon_i^2) = \sigma^2$.

$$\begin{aligned}
\mathbb{E}(\widehat{\beta}_1 \widehat{\beta}_2) &= \mathbb{E}((\bar{y} - \widehat{\beta}_2 \bar{x})\widehat{\beta}_2) \\
&= \mathbb{E}(\bar{y}\widehat{\beta}_2) - \bar{x}\mathbb{E}(\widehat{\beta}_2^2) \\
&= \mathbb{E}((\beta_1 + \beta_2 \bar{x})\widehat{\beta}_2) + \mathbb{E}(\bar{\epsilon}\widehat{\beta}_2) - \bar{x}\mathbb{E}(\widehat{\beta}_2^2) \ .
\end{aligned}$$

Or

$$\begin{aligned}
\mathbb{E}(\bar{\epsilon}\widehat{\beta}_2) &= \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i-\bar{x})(\beta_1+\beta_2 x_i+\epsilon_i)\bar{\epsilon}}{\sum_{i=1}^n (x_i-\bar{x})^2}\right) \\
&= \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i-\bar{x})\epsilon_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}\right) \\
&= \frac{\sum_{i=1}^n (x_i-\bar{x})}{n\sum_{i=1}^n (x_i-\bar{x})^2}\sigma^2 \\
&= 0\ .
\end{aligned}$$

Donc au total,

$$\mathbb{E}(\widehat{\beta}_1\widehat{\beta}_2) = \beta_1\beta_2 - \bar{x}\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\ .$$

Donc

$$\mathbb{E}(\widehat{\beta}_1\widehat{\beta}_2) - \mathbb{E}(\widehat{\beta}_1)\mathbb{E}(\widehat{\beta}_2) = -\bar{x}\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\ .$$

Ex. 5. (a) Nous avons $\widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2\bar{x}$ et

$$\begin{aligned}
\widehat{\beta}_2 &= \frac{\sum_{i=1}^n (x_i-\bar{x})y_i}{\sum_{i=1}^n (x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2}\ .
\end{aligned}$$

On en déduit $\widehat{\beta}_2 = 6,26/28,09 = 0.22$, $\widehat{\beta}_1 = 18,34 - 39,49 \times 0.22 = 9.65$.

(b)

$$\begin{aligned}
R^2 &= \frac{\sum_{i=1}^n (\widehat{y}_i-\bar{y})^2}{\sum_{i=1}^n (y_i-\bar{y})^2} \\
&= \frac{\sum_{i=1}^n (\widehat{\beta}_1+\widehat{\beta}_2 y_i-\bar{y})^2}{\sum_{i=1}^n (y_i-\bar{y})^2}\ .
\end{aligned}$$

On calcule d'abord

$$\begin{aligned}
\sum_{i=1}^n x_i^2 &= \sum_{i=1}^n [(x_i-\bar{x})^2 - \bar{x}^2 + 2x_i\bar{x}] \\
&= \sum_{i=1}^n (x_i-\bar{x})^2 + n\bar{x}^2\ .
\end{aligned}$$

Nous avons donc

$$\begin{aligned}
\sum_{i=1}^n (\widehat{\beta}_1+\widehat{\beta}_2 x_i-\bar{y})^2 &= \sum_{i=1}^n [\widehat{\beta}_1^2 + \widehat{\beta}_2^2 x_i^2 + \bar{y}^2 + 2\widehat{\beta}_1\widehat{\beta}_2 x_i - 2\widehat{\beta}_2 x_i\bar{y} - 2\widehat{\beta}_1\bar{y}] \\
&= n\widehat{\beta}_1^2 + \widehat{\beta}_2^2\left(\sum_{i=1}^n (x_i-\bar{x})^2 + n\bar{x}\right) + n\bar{y}^2 - 2\widehat{\beta}_1\widehat{\beta}_2\bar{x} + 2\widehat{\beta}_2\bar{y}\bar{x} - 2\widehat{\beta}_1\bar{y}\ .
\end{aligned}$$

Donc, si on pose

$$\begin{aligned}
A := {} & 20\times 9,65^2 + 0,22^2(20\times 28,29 + 20\times 39,49) + 20\times 18,34^2 \\
& +2\times 9,65\times 0,22\times 39,49 \\
& +2\times 0,22\times 18,34\times 39,49 - 2\times 9,65\times 18,34\ ,
\end{aligned}$$

on a alors

$$R^2 = \frac{A}{20\times 2,85}\ .$$

Ex. 6. (a) On veut

$$\begin{aligned}\beta_2 = \mathbb{E}(\widetilde{\beta}_2) &= \sum_{i=1}^n \widetilde{\lambda}_i(\beta_1 + \beta_2 x_i)\\ &= \beta_1 \sum_{i=1}^n \widetilde{\lambda}_i + \beta_2 \sum_{i=1}^n \widetilde{\lambda}_i x_i\end{aligned}$$

et ce pour n'importe quels β_1, β_2. Donc

$$\begin{aligned}\sum_{i=1}^n \widetilde{\lambda}_i &= 0\\ \sum_{i=1}^n \widetilde{\lambda}_i x_i &= 1\ .\end{aligned}$$

(b) Notons $\lambda_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}$. Les λ_i vérifient les mêmes relations que les $\widetilde{\lambda}_i$.

$$\begin{aligned}\mathbb{E}((\widetilde{\beta}_2 - \widehat{\beta}_2)(\widehat{\beta}_2 - \beta_2)) &= \mathbb{E}((\widetilde{\beta}_2 - \widehat{\beta}_2)\widehat{\beta}_2)\\ &= \mathbb{E}((\sum_{i=1}^n \widetilde{\lambda}_i(\beta_1 + \beta_2 x_i + \epsilon_i))(\sum_{i=1}^n \lambda_i(\beta_1 + \beta_2 x_i + \epsilon_i)))\\ &= \mathbb{E}((\beta_2 + \sum_{i=1}^n \widetilde{\lambda}_i \epsilon_i)(\beta_2 + \sum_{i=1}^n \lambda_i \epsilon_i))\\ &= \mathbb{E}((\sum_{i=1}^n \widetilde{\lambda}_i \epsilon_i)(\sum_{i=1}^n \lambda_i \epsilon_i))\\ &= \sum_{i=1}^n \widetilde{\lambda}_i \lambda_i \sigma^2\\ &= \sum_{i=1}^n \left[\widetilde{\lambda}_i \frac{(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}\sigma^2\right]\\ &= 0.\end{aligned}$$

Donc

$$\begin{aligned}\mathrm{Var}(\widetilde{\beta}_2) &= \mathbb{E}((\widetilde{\beta}_2 - \beta_2)^2)\\ &= \mathbb{E}((\widetilde{\beta}_2 - \widehat{\beta}_2)^2 + (\widehat{\beta}_2 - \beta_2)^2 + 2(\widetilde{\beta}_2 - \widehat{\beta}_2)(\widehat{\beta}_2 - \beta_2))\\ &= \mathbb{E}((\widetilde{\beta}_2 - \widehat{\beta}_2)^2) + \mathrm{Var}(\widehat{\beta}_2)\\ &\geq \mathrm{Var}(\widehat{\beta}_2)\ .\end{aligned}$$