École d'actuariat
Université Laval

Modèles linéaires en actuariat
ACT-2003
Exercices Supplémentaires Chapitre 2

Chargée de cours : Marie-Pier Côté
Automne 2014

Question 1. On s'intéresse à l'impact du sexe sur l'espérance de vie. On connaît les durées de vie de $n_F = 300$ femmes et $n_H = 200$ hommes. On choisit d'utiliser la variable indicatrice

$$x_i = \begin{cases} 0 & \text{, si } \texttt{SEXE}_i = \texttt{H} \\ 1 & \text{, si } \texttt{SEXE}_i = \texttt{F} \end{cases} .$$

On note $\bar{Y}_F$ la moyenne des durées de vie des femmes et $\bar{Y}_H$ la moyenne des durées de vie des hommes.

a) Montrer que l'estimateur des moindres carrés $\hat{\beta}_1$ (lié à la variable explicative x) est égal à $\bar{Y}_F - \bar{Y}_H$.
Indice : On peut exprimer $\bar{Y}$ en termes de $\bar{Y}_F$ et $\bar{Y}_H$.

b) Ce résultat permet-il d'interpréter le coefficient relié à une variable catégorique binaire ? Expliquer.

c) Que représente $\hat{\beta}_0$ dans ce cas ?

Question 2. On s'intéresse à la covariance entre deux résidus.

(i) D'abord, trouver $\mathbb{C}\text{ov}(Y_i, \hat{Y}_j)$.
(ii) Puis, calculer $\mathbb{C}\text{ov}(\hat{Y}_i, \hat{Y}_j)$.
(iii) Déduire de (i) et (ii) que

$$\mathbb{C}\text{ov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = -\sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \right) .$$

SOLUTIONS

Question 1.

a) On a

$$\bar{Y} = \frac{\sum_{i=1}^{500} Y_i}{500} = \frac{300\bar{Y}_F + 200\bar{Y}_H}{500}.$$

Aussi,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{500} x_i Y_i - 500\bar{x}\bar{Y}}{\sum_{i=1}^{500} x_i^2 - 500\bar{x}^2}.$$

Or,

$$\bar{x} = \frac{\sum_{i=1}^{500} x_i}{500} = \frac{300}{500},$$
$$\sum_{i=1}^{500} x_i^2 = 300,$$
$$\sum_{i=1}^{500} x_i Y_i = 300\bar{Y}_F$$

Donc,

$$\begin{aligned}\hat{\beta}_1 &= \frac{300\bar{Y}_F - 500 \times \frac{300}{500} \times \frac{300\bar{Y}_F + 200\bar{Y}_H}{500}}{300 - 500\left(\frac{300}{500}\right)^2} \\ &= \frac{500\bar{Y}_F - 300\bar{Y}_F - 200\bar{Y}_H}{500 - 300} \\ &= \bar{Y}_F - \bar{Y}_H.\end{aligned}$$

b) Oui, le coefficient relié à la variable indicatrice qui vaut 1 si le sexe est F représente la différence etre la moyenne de l'espérance de vie pour les femmes et la moyenne de l'espérace de vie pour les hommes.

c)

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x} = \bar{Y} - (\bar{Y}_F - \bar{Y}_H)\frac{300}{500} = \bar{Y}_H.$$

$\Rightarrow \hat{\beta}_0$ est la moyenne de l'espérance de vie pour les hommes.

Question 2.

(i)

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(Y_i, \hat{Y}_j) &= \mathbb{C}\mathrm{ov}(Y_i, \hat{\beta}_0 + \hat{\beta}_1 x_j) \\
&= \mathbb{C}\mathrm{ov}(Y_i, \bar{Y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_j) \\
&= \mathbb{C}\mathrm{ov}(Y_i, \bar{Y}) + (x_j - \bar{x})\mathbb{C}\mathrm{ov}(Y_i, \hat{\beta}_1) \text{ par indépendance des observations} \\
&= \frac{\sigma^2}{n} + \frac{(x_j - \bar{x})}{S_{xx}} \sum_{l=1}^{n} (x_l - \bar{x})\mathbb{C}\mathrm{ov}(Y_i, Y_l) \\
&= \frac{\sigma^2}{n} + \frac{(x_j - \bar{x})(x_i - \bar{x})}{S_{xx}}\sigma^2 \text{ par indépendance des observations.}
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(\hat{Y}_i, \hat{Y}_j) &= \mathbb{C}\mathrm{ov}(\hat{\beta}_0 + \hat{\beta}_1 x_i, \hat{\beta}_0 + \hat{\beta}_1 x_j) \\
&= \mathbb{V}\mathrm{ar}(\hat{\beta}_0) + (x_i + x_j)\mathbb{C}\mathrm{ov}(\hat{\beta}_0, \hat{\beta}_1) + x_i x_j \mathbb{V}\mathrm{ar}(\hat{\beta}_1) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) - (x_i + x_j)\frac{\bar{x}\sigma^2}{S_{xx}} + x_i x_j \frac{\sigma^2}{S_{xx}} \\
&= \ldots \\
&= \sigma^2 \left(\frac{1}{n} + \frac{(x_j - \bar{x})(x_i - \bar{x})}{S_{xx}} \right).
\end{aligned}$$

(iii)

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) &= \mathbb{C}\mathrm{ov}(Y_i - \hat{Y}_i, Y_j - \hat{Y}_j) \\
&= \mathbb{C}\mathrm{ov}(Y_i, Y_j) - \mathbb{C}\mathrm{ov}(Y_i, \hat{Y}_j) - \mathbb{C}\mathrm{ov}(\hat{Y}_i, Y_j) + \mathbb{C}\mathrm{ov}(\hat{Y}_i, \hat{Y}_j) \\
&= 0 - 2\sigma^2 \left(\frac{1}{n} + \frac{(x_j - \bar{x})(x_i - \bar{x})}{S_{xx}} \right) + \left(\frac{1}{n} + \frac{(x_j - \bar{x})(x_i - \bar{x})}{S_{xx}} \right) \\
&= -\sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \right).
\end{aligned}$$