

Corrigé de la série de TD n1 :

Exercice 0.1. Les estimateurs MCO sont :

$$\tilde{a} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ et } \hat{b} = \bar{y} - \hat{a}\bar{x}.$$

1)– On suppose que b est connu, trouvons $\tilde{a}$:

L'estimateur $\tilde{a}$ correspond à l'argmin de la quantité

$$\begin{aligned} S(a) &= \sum_{i=1}^n (y_i - (ax_i + b))^2 \\ \Rightarrow \frac{\partial S}{\partial a} &= 0 \Leftrightarrow \sum_{i=1}^n (-2x_i)(y_i - \tilde{a}x_i - b) = 0 \\ \Rightarrow -2 \sum_{i=1}^n x_i y_i + 2\tilde{a} \sum_{i=1}^n x_i^2 + 2b \sum_{i=1}^n x_i &= 0 \\ \Rightarrow \tilde{a} &= \frac{\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2} \\ \tilde{a} &= \frac{\sum_{i=1}^n x_i (y_i - b)}{\sum_{i=1}^n x_i^2} \end{aligned}$$

2)– $\text{var}(\tilde{a}) = ?$

On a : $y_i = ax_i + b + \xi_i$

$$\begin{aligned} \tilde{a} &= \frac{\sum_{i=1}^n x_i (y_i - b)}{\sum_{i=1}^n x_i^2} \\ &= \frac{\sum_{i=1}^n x_i (ax_i + b + \xi_i - b)}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i (ax_i + \xi_i)}{\sum_{i=1}^n x_i^2} \\ &= \frac{a \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i \xi_i}{\sum_{i=1}^n x_i^2} = a + \frac{\sum_{i=1}^n x_i \xi_i}{\sum_{i=1}^n x_i^2} \\ \Rightarrow \text{var}(\tilde{a}) &= \frac{\sum_{i=1}^n x_i^2 \text{var}(\xi_i)}{(\sum_{i=1}^n x_i^2)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} = \text{var}(\tilde{a}) \end{aligned}$$

* On a $\text{var}(\hat{a}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

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

$$\begin{aligned}
&= \sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \preceq \sum_{i=1}^n x_i^2 \\
&\Rightarrow \text{var}(\tilde{a}) \preceq \text{var}(\hat{a})
\end{aligned}$$

2)– Supposons que a est connu, mais pas b :

(a)– Trouvons $\tilde{b}$:

L'estimateur $\tilde{b}$ correspond à l'argmin de

$$\begin{aligned}
S(b) &= \sum_{i=1}^n (y_i - (ax_i + b))^2 \\
\Rightarrow \frac{\partial S}{\partial a} = 0 &\Rightarrow \frac{\partial S}{\partial a} = -2 \sum_{i=1}^n (y_i - ax_i - \tilde{b}) = 0 \\
&\Rightarrow \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - n\tilde{b} = 0 \\
&\Rightarrow \tilde{a} = \frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} \\
&\quad \tilde{b} = \bar{y} - a\bar{x}
\end{aligned}$$

(b)– $\text{var}(\tilde{b}) = ?$

$$\begin{aligned}
y_i &= ax_i + b + \xi_i \\
\bar{y} &= a\bar{x} + b + \frac{1}{n} \sum_{i=1}^n \xi_i \\
\tilde{b} &= a\bar{x} + b + \frac{1}{n} \sum_{i=1}^n \xi_i - a\bar{x} \\
\tilde{b} &= b + \frac{1}{n} \sum_{i=1}^n \xi_i \\
\text{var}(\tilde{b}) &= \text{var}(b) + \frac{1}{n^2} \sum_{i=1}^n \text{var}(\xi_i) \\
\text{var}(\tilde{b}) &= \frac{\sigma^2}{n}
\end{aligned}$$

On a :

$$\begin{aligned}
\text{var}(\hat{b}) &= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \succeq \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2} = \frac{\sigma^2}{n} \\
&\Rightarrow \text{var}(\hat{b}) \succeq \text{var}(\tilde{b}).
\end{aligned}$$

Exercice 0.2. 1)– Calculons la droite de régression des MCO du poids au fonction du poids des pères

$$f_i = \beta_1 + \beta_2 p_i + \xi_i$$

Soient $\hat{\beta}_1$ et $\hat{\beta}_2$ les estimateurs des MCO de β_1 et β_2 tel que :

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (p_i - \bar{p}) f_i}{\sum_{i=1}^n (p_i - \bar{p})^2} = \frac{\sum_{i=1}^n p_i f_i - \bar{p} \sum_{i=1}^n f_i}{\sum_{i=1}^n p_i^2 - n \bar{p}^2}$$

$$\hat{\beta}_2 = \frac{54107 - (800/12)811}{(53418) - 12(800/12)^2} = 0.48$$

$$\hat{\beta}_1 = \bar{f} - \hat{\beta}_2 \bar{p} = (811/12) - 0.48(800/12)$$

$$\hat{f}_i = 35.8 + 0.48 p_i + \hat{\xi}_i$$

2)– Calculons la droite de régression du poids des pères en fonction des fils :

$$p_i = \beta_1 + \beta_2 f_i + \xi_i$$

$\hat{\beta}_1$ et $\hat{\beta}_2$ sont les estimateurs de β_1 et β_2 .

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (f_i - \bar{f}) p_i}{\sum_{i=1}^n f_i^2 - n \bar{f}^2} = \frac{\sum_{i=1}^n f_i p_i - \bar{f} \sum_{i=1}^n p_i}{\sum_{i=1}^n f_i^2 - n \bar{f}^2}$$

$$= \frac{54107 - (811/12)(800)}{54849 - 12(811/12)^2} = 1.04$$

$$\hat{\beta}_1 = \bar{p} - \hat{\beta}_2 \bar{f} = (800/12) - 1.04(811/12) = -3.62$$

$$\hat{p}_i = -3.62 + 1.04 f_i + \hat{\xi}_i$$

Exercice 0.3. I)–

1)– La matrice $X'X$ est symétrique, alors :

$$X'X = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 9.3 & 5.4 \\ 0 & 5.4 & 12.7 \end{pmatrix}$$

2)– $n = 25$. 3)– Le coefficient de corrélation entre x et z

$$r_{x,z} = \frac{\sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (z_i - \bar{z})^2}}$$

$$\begin{aligned}
 r_{x,z} &= \frac{\sum_{i=1}^n x_i z_i - n\bar{x}\bar{z}}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \sqrt{\sum_{i=1}^n z_i^2 - n\bar{z}^2}} \\
 &= \frac{5.4 - 25(0)(0)}{\sqrt{9.3}\sqrt{12.7}} = 0.49
 \end{aligned}$$

II)–

1)–

$$\hat{y} = -1.6 + 0.61x + 0.46z + \hat{\xi}$$

$$\Rightarrow \bar{y} = -1.6 + 0.61\bar{x} + 0.46\bar{z} + \hat{\xi}$$

$$= -1.6$$

2)– $SCE = ?$

$$\begin{aligned}
 SCE &= \|\hat{y} - \bar{y}\|^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\
 &= \sum_{i=1}^n (-1.6 + 0.61x_i + 0.46z_i + 1.6)^2 \\
 &= \sum_{i=1}^n (0.61x_i + 0.46z_i)^2 \\
 &= (0.61)^2 \sum_{i=1}^n x_i^2 + (0.46)^2 \sum_{i=1}^n z_i^2 + 2(0.61)(0.46) \sum_{i=1}^n x_i z_i \\
 &= 0.37(9.3) + 0.21(12.7) + 0.56(5.4) \\
 &= 9.13
 \end{aligned}$$

$$= SCT = SCE + SCR = 9.13 + 0.3 = 9.43$$

et

$$R^2 = \frac{SCE}{SCT} = \frac{9.13}{9.43} = 0.96$$

On remarque 96 pour cent de la variation des données est expliquée par le modèle. 3)–

$$R_a^2 = 1 - \frac{n-1}{n-p} \frac{SCR}{SCT} = 1 - \frac{24}{22} \frac{0.3}{9.43} = 0.96 \simeq R^2$$

$$\hat{\sigma}^2 \frac{SCR}{n-p} = \frac{\|\hat{\xi}\|^2}{n-p} = \frac{0.3}{25-3} = 0.01$$

$$\Rightarrow \text{var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = (0.01)(X'X)^{-1}$$

* Calculons $(X'X)^{-1}$:

$$X'X = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 9.3 & 5.4 \\ 0 & 5.4 & 12.7 \end{pmatrix}$$

$$dzt(X'X) = 25(9.3 * 12.7 - 5.4 * 5.4) = 2223.75$$

$$X'X = \begin{pmatrix} 88.95 & 0 & 0 \\ 0 & 317.5 & -135 \\ 0 & -135 & 232.5 \end{pmatrix}$$

$$X'X = \frac{1}{2223.76} C^t = \begin{pmatrix} 0.04 & 0 & 0 \\ 0 & 0.143 & -0.0607 \\ 0 & -0.0607 & -0.1045 \end{pmatrix}$$

L'estimateur de la $var(\hat{\beta})$ est :

$$var(\hat{\beta}) = 0.01 \begin{pmatrix} 0.04 & 0 & 0 \\ 0 & 0.143 & -0.0607 \\ 0 & -0.0607 & 0.1045 \end{pmatrix}$$