

CORRIGÉ 13

**Exercice 1.** (a) La droite est  $y(x) = \hat{a} + \hat{\beta}x$ . En mettant  $x = \bar{x}$  on trouve

$$y(\bar{x}) = \hat{a} + \hat{\beta}\bar{x} = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta}\bar{x} = \bar{y}.$$

(b) On a

$$\sum_{j=1}^n r_j = \sum_{j=1}^n (y_j - \hat{y}_j) = n\bar{y} - n\hat{a} - \hat{\beta} \sum_{j=1}^n x_j = n\bar{y} - n(\bar{y} - \hat{\beta}\bar{x}) - \hat{\beta}n\bar{x} = 0.$$

(c) On a

$$\begin{aligned} \sum_{j=1}^n x_j r_j &= \sum_{j=1}^n x_j (y_j - \hat{a} - \hat{\beta}x_j) = \sum_{j=1}^n x_j y_j - n\bar{x}(\bar{y} - \hat{\beta}\bar{x}) - \hat{\beta} \sum_{j=1}^n x_j^2 \\ &= \sum_{j=1}^n x_j y_j - n\bar{x}\bar{y} - \hat{\beta}(\sum_{j=1}^n x_j^2 - n\bar{x}^2) = 0, \end{aligned}$$

par la formule de  $\hat{\beta}$ .

(d) On a

$$\sum_{j=1}^n \hat{y}_j r_j = \sum_{j=1}^n r_j (\hat{a} + \hat{\beta}x_j) = \hat{a} \sum_{j=1}^n r_j + \hat{\beta} \sum_{j=1}^n r_j x_j = 0 + 0,$$

d'après les parties 2. et 3.

**Exercice 2.** La corrélation entre une variable aléatoire  $X$  et une variable aléatoire  $Y$  est définie par  $r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \times \text{Var}(Y)}}$ . Son équivalent empirique (ou encore la réalisation de son estimateur) est donc

$$\hat{r} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (1)$$

D'après le cours, on sait que la réalisation de l'estimateur de la pente de la droite de régression s'écrit

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Ainsi, en utilisant le fait que  $\sum_{i=1}^n (x_i - \bar{x}) = 0$ , on obtient

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2)$$

La combinaison de (1) et (2) donne

$$\hat{\beta} = \hat{r} \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

On trouve ainsi que  $\hat{\beta} = 0.6\sqrt{20/10} \approx 0.85$ . Le cours nous donne également que la réalisation de l'estimateur de l'ordonnée à l'origine de la droite de régression est  $\hat{a} = \bar{y} - \hat{\beta}\bar{x}$ . On a donc  $\hat{a} = -4.5$ .

**Exercice 3.** (a) D'après l'énoncé, on doit minimiser  $\sum_{i=1}^n [y_i - y_0 - \beta(x_i - x_0)]^2$ . On a donc

$$\begin{aligned} \frac{\partial \sum_{i=1}^n [y_i - y_0 - \beta(x_i - x_0)]^2}{\partial \beta} &= 0 \\ \Leftrightarrow -2 \sum_{i=1}^n [y_i - y_0 - \beta(x_i - x_0)](x_i - x_0) &= 0 \\ \Leftrightarrow \beta &= \frac{\sum_{i=1}^n (x_i - x_0)(y_i - y_0)}{\sum_{i=1}^n (x_i - x_0)^2}. \end{aligned}$$

Par ailleurs on a

$$\frac{\partial^2 \sum_{i=1}^n [y_i - y_0 - \beta(x_i - x_0)]^2}{\partial \beta^2} = 2 \sum_{i=1}^n (x_i - x_0)^2 > 0.$$

On obtient donc que le minimum est atteint pour

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - x_0)(y_i - y_0)}{\sum_{i=1}^n (x_i - x_0)^2}.$$

(b) Si l'on pose  $(x_0, y_0) = (\bar{x}, \bar{y})$ , on obtient

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

On retombe alors sur l'estimateur de la pente de la droite de régression classique. En d'autres termes, l'estimateur de la pente de régression classique correspond à la pente qui minimise l'erreur de la droite de régression forcée à passer par la moyenne  $(\bar{x}, \bar{y})$ .

(c) La réalisation de l'estimateur de la pente de la droite de régression calculée sur notre jeu de données est  $\hat{\beta} = \frac{24.75}{35} = 0.71$ . La droite de régression est donc  $y = \hat{\beta}x - \hat{\beta}x_0 + y_0 = 3.87 + 0.71x$ .

**Exercice 4.** (a) On a

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2.$$

Ainsi, en utilisant les données de l'énoncé, on obtient  $\sum_{i=1}^n (x_i - \bar{x})^2 = 76.9$ . De même, on a  $\sum_{i=1}^n (y_i - \bar{y})^2 = 108.76$ . Enfin, on obtient

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n\bar{x}\bar{y} = 72.17.$$

Maintenant, on rappelle que

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

Ainsi, on obtient les estimations  $\hat{b} = 0.94$  et  $\hat{a} = -4.56$ . Finalement, on sait que l'estimateur de la variance du bruit Gaussien  $\eta$  est

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{a} - \hat{b}x_i)^2.$$

Sa réalisation est donc  $\hat{\sigma}^2 = 5.13$ .

- (b) On teste l'hypothèse  $H_0 : b = 0$  contre  $H_1 : b \neq 0$  au niveau de signification de 1%. On sait d'après le cours que sous  $H_0$

$$\frac{\hat{b}}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2} = t_8,$$

Ainsi on rejette  $H_0$  si  $\left| \frac{\hat{b}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \right| > t_{8,0.995} = 3.355$ . On a  $\left| \frac{\hat{b}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \right| = 3.64$ . On rejette donc  $H_0$  en faveur de  $H_1 : b \neq 0$ .

**Exercice 5.** (a) Puisque  $PV^\gamma = C$ , on a

$$\log(P) + \gamma \log(V) = \log(C) \quad \text{et donc} \quad \log(P) = \log(C) - \gamma \log(V).$$

En posant  $X = \log(V)$  et  $Y = \log(P)$ , l'équation de la droite du modèle linéaire s'écrit

$$Y = \alpha + \beta X,$$

où  $\alpha = \log(C)$  et  $\beta = -\gamma$ . Nous souhaitons estimer les paramètres  $\alpha$  et  $\beta$ .

- (b) On sait d'après le cours que les estimateurs des paramètres de la droite de régression sont donnés par

$$\hat{\beta} = \frac{\sum_{i=1}^n Y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x},$$

où  $x_i = \log(v_i)$  et  $Y_i = \log(P_i)$ ,  $i = 1, \dots, 6$ . On trouve  $\hat{\beta} = -1.4$  et  $\hat{\alpha} = 9.66$ . Ainsi, on a  $\hat{C} = \exp(\hat{\alpha}) = 15677.78$  et  $\hat{\gamma} = -\hat{\beta} = 1.4$ .

- (c) On a  $\hat{y} = \log(\hat{p}) = \hat{\alpha} + \hat{\beta} \log(v)$ . Ainsi, pour  $v = 100$ , on a  $\hat{p} = \exp(\hat{y}) = 24.85 \text{ kg/cm}^2$ .

- (d) Soit

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

On sait d'après le cours que

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2}.$$

L'intervalle de confiance à 95% est donc donné par les bornes

$$\hat{\beta} \pm \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} t_{n-2,0.975},$$

où

$$\hat{\sigma} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2}.$$

On a  $\hat{\sigma} \approx 0.04$ ,  $\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \approx 1.05$  et  $t_{n-2,0.975} = 2.776$ . L'intervalle de confiance recherché est donc approximativement  $[-1.51, -1.29]$ .