Chapitre 4 : correction des exercices théoriques

T.4.1

$$\begin{aligned}
s_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_iy_i - x_i\overline{y} - \overline{x}y_i + \overline{x}\,\overline{y}) \\
&= \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\overline{y} - \overline{x}\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right) + \frac{1}{n}n\overline{x}\,\overline{y} \\
&= \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \overline{x}\,\overline{y} - \overline{x}\,\overline{y} + \overline{x}\,\overline{y} \\
&= \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \overline{x}\,\overline{y}.
\end{aligned}$$

Nous retrouvons bien la relation (4.10).

Etablissons à présent la relation (4.12). Nous partons de :

$$s_{uv} = \frac{1}{n}\sum_{i=1}^{n}(u_i-\overline{u})(v_i-\overline{v}).$$

Or,

$$u_i = \frac{x_i-x_0}{d_x},\; v_i = \frac{y_i-y_0}{d_y},\; \overline{u} = \frac{\overline{x}-x_0}{d_x} \quad \text{et} \quad \overline{v} = \frac{\overline{y}-y_0}{d_y}\;.$$

Dès lors,

$$\begin{aligned}
s_{uv} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-x_0}{d_x} - \frac{\overline{x}-x_0}{d_x}\right)\left(\frac{y_i-y_0}{d_y} - \frac{\overline{y}-y_0}{d_y}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{d_x}\right)\left(\frac{y_i-\overline{y}}{d_y}\right) = \frac{1}{d_xd_y}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})\right] = \frac{s_{xy}}{d_xd_y},
\end{aligned}$$

ce que nous voulions démontrer.

T.4.2

$$\begin{aligned}
s_{xy} &= \frac{1}{n}\sum_j\sum_k n_{jk}(x_j-\overline{x})(y_k-\overline{y}) \\
&= \frac{1}{n}\sum_j\sum_k n_{jk}(x_jy_k - x_j\overline{y} - \overline{x}y_k + \overline{x}\,\overline{y}) \\
&= \frac{1}{n}\sum_j\sum_k n_{jk}x_jy_k - \left[\frac{1}{n}\sum_j\sum_k n_{jk}x_j\right]\overline{y} - \overline{x}\left[\frac{1}{n}\sum_j\sum_k n_{jk}y_k\right] + \left[\frac{1}{n}\sum_j\sum_k n_{jk}\right]\overline{x}\,\overline{y} \\
&= \frac{1}{n}\sum_j\sum_k n_{jk}x_jy_k - \left[\frac{1}{n}\sum_j\left(\sum_k n_{jk}\right)x_j\right]\overline{y} \\
&\qquad -\overline{x}\left[\frac{1}{n}\sum_k\left(\sum_j n_{jk}\right)y_k\right] + \left[\frac{1}{n}\left(\sum_j\sum_k n_{jk}\right)\right]\overline{x}\,\overline{y}.
\end{aligned}$$

Or,

$$\sum_k n_{jk} = n_{j\bullet},\ \sum_j n_{jk} = n_{\bullet k} \quad \text{et} \quad \sum_j\sum_k n_{jk} = n.$$

Dès lors,

$$\begin{aligned}
s_{xy} &= \frac{1}{n}\sum_j\sum_k n_{jk}x_jy_k - \left[\frac{1}{n}\sum_j n_{j\bullet}x_j\right]\overline{y} - \overline{x}\left[\frac{1}{n}\sum_k n_{\bullet k}y_k\right] + \left[\frac{1}{n}n\right]\overline{x}\,\overline{y} \\
&= \frac{1}{n}\sum_j\sum_k n_{jk}x_jy_k - \overline{x}\,\overline{y} - \overline{x}\,\overline{y} + \overline{x}\,\overline{y} \\
&= \frac{1}{n}\sum_j\sum_k n_{jk}x_jy_k - \overline{x}\,\overline{y},
\end{aligned}$$

ce que nous voulions démontrer.

T.4.3

$$s_{uv} = \frac{1}{n}\sum_j\sum_k n_{jk}(u_j-\overline{u})(v_k-\overline{v}).$$

Or,

$$u_j = \frac{x_j-x_0}{d_x},\ v_k = \frac{y_k-y_0}{d_y},\ \overline{u} = \frac{\overline{x}-x_0}{d_x} \quad \text{et} \quad \overline{v} = \frac{\overline{y}-y_0}{d_y}\ .$$

Dès lors,

$$\begin{aligned}
s_{uv} &= \frac{1}{n}\sum_j\sum_k n_{jk}\left(\frac{x_j-x_0}{d_x} - \frac{\overline{x}-x_0}{d_x}\right)\left(\frac{y_k-y_0}{d_y} - \frac{\overline{y}-y_0}{d_y}\right) \\
&= \frac{1}{n}\sum_j\sum_k n_{jk}\left(\frac{x_j-\overline{x}}{d_x}\right)\left(\frac{y_k-\overline{y}}{d_y}\right) \\
&= \frac{1}{d_xd_y}\left[\frac{1}{n}\sum_j\sum_k n_{jk}(x_j-\overline{x})(y_k-\overline{y})\right] = \frac{s_{xy}}{d_xd_y}\ .
\end{aligned}$$

On en déduit que

$$s_{xy} = d_x d_y s_{uv},$$

ce que nous voulions démontrer.

T.4.4

(a) Soient $u_i = \frac{x_i - x_0}{d_x}$ et $v_i = \frac{y_i - y_0}{d_y}$ $(i = 1, \ldots, n)$. Nous devons montrer que $r_{uv} = r_{xy}$, où r_{uv} est le coefficient de corrélation calculé sur la série statistique bivariée $\{(u_i, v_i); i = 1, \ldots, n\}$, tandis que r_{xy} est le coefficient de corrélation calculé sur la série statistique bivariée $\{(x_i, y_i); i = 1, \ldots, n\}$.
Nous avons

$$r_{uv} = \frac{s_{uv}}{s_u s_v} \cdot$$

Or, nous savons que

$$s_{uv} = \frac{s_{xy}}{d_x d_y}, \; s_u = \frac{s_x}{d_x} \quad \text{et} \quad s_v = \frac{s_y}{d_y} \cdot$$

Dès lors,

$$r_{uv} = \frac{\frac{s_{xy}}{d_x d_y}}{\frac{s_x}{d_x} \frac{s_y}{d_y}} = \frac{s_{xy}}{s_x s_y} = r_{xy}.$$

(b) Posons $u_i = \frac{x_i - \overline{x}}{s_x}$ et $v_i = \frac{y_i - \overline{y}}{s_y}$ $(i = 1, \ldots, n)$. Nous avons :

$$\begin{aligned} r_{xy} &= \frac{s_{xy}}{s_x s_y} = \frac{1}{s_x s_y} \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \right] \\ &= \frac{1}{s_x s_y} \left[\frac{1}{n} \sum_{i=1}^{n} (s_x u_i)(s_y v_i) \right] = \frac{s_x s_y}{s_x s_y} \left[\frac{1}{n} \sum_{i=1}^{n} u_i v_i \right] \\ &= \frac{1}{n} \sum_{i=1}^{n} u_i v_i, \end{aligned}$$

ce qui établit la relation (4.18).

T.4.5

$$\begin{aligned} & Q(\alpha, \beta) = \sum_i (y_i - \alpha - \beta x_i)^2 \\ \Rightarrow \quad & \frac{\partial Q}{\partial \alpha} = \sum_i 2(y_i - \alpha - \beta x_i)(-1) = -2 \sum_i (y_i - \alpha - \beta x_i) \\ & \frac{\partial Q}{\partial \beta} = \sum_i 2(y_i - \alpha - \beta x_i)(-x_i) = -2 \sum_i (y_i - \alpha - \beta x_i) x_i \end{aligned}$$

Nous devons donc résoudre le système d'équations normales suivant :

$$\left\{\begin{array}{l} \left.\frac{\partial Q}{\partial \alpha}\right|_{\alpha=a,\beta=b} = 0 \\ \left.\frac{\partial Q}{\partial \beta}\right|_{\alpha=a,\beta=b} = 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} \sum_i (y_i - a - bx_i) = 0 \\ \sum_i (y_i - a - bx_i)x_i = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} \sum_i y_i - na - b\sum_i x_i = 0 \\ \sum_i x_i y_i - a\sum_i x_i - b\sum_i x_i^2 = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} \frac{1}{n}\sum_i y_i - a - b\left(\frac{1}{n}\sum_i x_i\right) = 0 \\ \frac{1}{n}\sum_i x_i y_i - a\left(\frac{1}{n}\sum_i x_i\right) - b\left(\frac{1}{n}\sum_i x_i^2\right) = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} a = \overline{y} - b\overline{x} \\ \frac{1}{n}\sum_i x_i y_i - a\overline{x} - b\left(\frac{1}{n}\sum_i x_i^2\right) = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} a = \overline{y} - b\overline{x} \\ \frac{1}{n}\sum_i x_i y_i - \left(\overline{y} - b\overline{x}\right)\overline{x} - b\left(\frac{1}{n}\sum_i x_i^2\right) = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} a = \overline{y} - b\overline{x} \\ \frac{1}{n}\sum_i x_i y_i - \overline{x}\,\overline{y} + b\overline{x}^2 - b\left(\frac{1}{n}\sum_i x_i^2\right) = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} a = \overline{y} - b\overline{x} \\ s_{xy} - b\left(\frac{1}{n}\sum_i x_i^2 - \overline{x}^2\right) = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} a = \overline{y} - b\overline{x} \\ s_{xy} - bs_x^2 = 0 \end{array}\right.$$

$$\Leftrightarrow \left\{\begin{array}{l} a = \overline{y} - b\overline{x} \\ b = \frac{s_{xy}}{s_x^2} \end{array}\right.$$

T.4.6

(a)

$$\begin{aligned} \overline{e} &= \frac{1}{n}\sum_i e_i = \frac{1}{n}\sum_i (y_i - \widehat{y}_i) \\ &= \frac{1}{n}\sum_i y_i - \frac{1}{n}\sum_i \widehat{y}_i \\ &= \overline{y} - \frac{1}{n}\sum_i \widehat{y}_i. \end{aligned}$$

Or,

$$\begin{aligned}\frac{1}{n}\sum_i \widehat{y}_i &= \frac{1}{n}\sum_i (a+bx_i) = a + b\left(\frac{1}{n}\sum_i x_i\right)\\ &= a + b\overline{x} = \overline{y} - b\overline{x} + b\overline{x} = \overline{y}.\end{aligned}$$

Dès lors,

$$\overline{e} = \overline{y} - \overline{y} = 0.$$

(b)

$$\begin{aligned}s^2_{y.x} &= \frac{1}{n}\sum_i e_i^2 = \frac{1}{n}\sum_i (y_i - \widehat{y}_i)^2\\ &= \frac{1}{n}\sum_i (y_i - a - bx_i)^2\\ &= \frac{1}{n}\sum_i (y_i - (\overline{y} - b\overline{x}) - bx_i)^2\\ &= \frac{1}{n}\sum_i \left[(y_i - \overline{y}) - b(x_i - \overline{x})\right]^2\\ &= \frac{1}{n}\sum_i \left[(y_i - \overline{y})^2 - 2b(x_i - \overline{x})(y_i - \overline{y}) + b^2(x_i - \overline{x})^2\right]\\ &= \frac{1}{n}\sum_i (y_i - \overline{y})^2 - 2b\frac{1}{n}\sum_i (x_i - \overline{x})(y_i - \overline{y}) + b^2\frac{1}{n}\sum_i (x_i - \overline{x})^2\\ &= s_y^2 - 2bs_{xy} + b^2 s_x^2 = s_y^2 - 2\frac{s_{xy}}{s_x^2}s_{xy} + \left(\frac{s_{xy}}{s_x^2}\right)^2 s_x^2\\ &= s_y^2 - 2\frac{s_{xy}^2}{s_x^2} + \frac{s_{xy}^2}{s_x^2} = s_y^2 - \frac{s_{xy}^2}{s_x^2} = s_y^2 - \frac{s_{xy}^2}{s_x^2 s_y^2}s_y^2\\ &= s_y^2 - r^2 s_y^2 = (1 - r^2)s_y^2\end{aligned}$$

(c)

$$\begin{aligned}s^2_{\text{reg}} &= \frac{1}{n}\sum_i (\widehat{y}_i - \overline{y})^2 = \frac{1}{n}\sum_i (a + bx_i - \overline{y})^2\\ &= \frac{1}{n}\sum_i (\overline{y} - b\overline{x} + bx_i - \overline{y})^2 = \frac{1}{n}\sum_i \left[b(x_i - \overline{x})\right]^2\\ &= b^2\frac{1}{n}\sum_i (x_i - \overline{x})^2 = b^2 s_x^2\\ &= \left(\frac{s_{xy}}{s_x^2}\right)^2 s_x^2 = \frac{s_{xy}^2}{s_x^2} = \frac{s_{xy}^2}{s_x^2 s_y^2}s_y^2 = r^2 s_y^2\end{aligned}$$

(d)

$$s_y^2 = (1 - r^2)s_y^2 + r^2 s_y^2 = s^2_{y.x} + s^2_{\text{reg}}$$

T.4.7

L'expression (4.39) nous indique que :

$$r_S = \frac{\frac{1}{n}\sum_i [R(x_i) - \overline{R}_x][R(y_i) - \overline{R}_y]}{\sqrt{\left\{\frac{1}{n}\sum_i [R(x_i) - \overline{R}_x]^2\right\}\left\{\frac{1}{n}\sum_i [R(y_i) - \overline{R}_y]^2\right\}}} \cdot$$

Or,

$$\begin{aligned}
\frac{1}{n}\sum_i [R(x_i) - \overline{R}_x][R(y_i) - \overline{R}_y] &= \frac{1}{n}\sum_i R(x_i)R(y_i) - \overline{R}_x\overline{R}_y \\
&= \frac{1}{n}\sum_i R(x_i)R(y_i) - \left(\frac{n+1}{2}\right)^2 .
\end{aligned}$$

Par ailleurs,

$$\begin{aligned}
\frac{1}{n}\sum_i [R(x_i) - \overline{R}_x]^2 &= \frac{1}{n}\sum_i [R(x_i)]^2 - \overline{R}_x^2 \\
&= \frac{1}{n}\sum_{j=1}^{n} j^2 - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 ,
\end{aligned}$$

puisque $\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$, comme énoncé au point (b) de l'exercice E.A.3.5 de l'annexe 3 consultable à l'adresse `http://elements.ulb.ac.be`.

Enfin,

$$\frac{1}{n}\sum_i [R(y_i) - \overline{R}_y]^2 = \frac{1}{n}\sum_i [R(x_i) - \overline{R}_x]^2.$$

On déduit de ces différents résultats que :

$$r_S = \frac{\frac{1}{n}\sum_i R(x_i)R(y_i) - \left(\frac{n+1}{2}\right)^2}{\frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2} \cdot$$

Mais

$$\begin{aligned}
\frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{2(n+1)(2n+1) - 3(n+1)^2}{12} \\
&= \frac{(n+1)(4n+2-3n-3)}{12} = \frac{(n+1)(n-1)}{12} \\
&= \frac{n^2-1}{12} \cdot
\end{aligned}$$

Nous avons donc

$$
\begin{aligned}
r_S &= \frac{\frac{1}{n}\sum_i R(x_i)R(y_i) - \left(\frac{n+1}{2}\right)^2}{\frac{n^2-1}{12}} \\
&= \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) - \frac{12}{n^2-1}\ \frac{(n+1)^2}{4} \\
&= \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) - \frac{3(n+1)}{n-1}\ . \qquad (1)
\end{aligned}
$$

Par ailleurs,

$$
\begin{aligned}
& 1 - \frac{6\sum_i [R(x_i) - R(y_i)]^2}{n(n^2-1)} \\
= \ & 1 - \frac{6}{n(n^2-1)}\left\{\sum_i [R(x_i)]^2 - 2\sum_i R(x_i)R(y_i) + \sum_i [R(y_i)]^2\right\} \\
= \ & 1 - \frac{6}{n^2-1}\left\{\frac{(n+1)(2n+1)}{6} - \frac{2}{n}\sum_i R(x_i)R(y_i) + \frac{(n+1)(2n+1)}{6}\right\} \\
= \ & 1 - \frac{6}{n^2-1}\left\{\frac{2(n+1)(2n+1)}{6} - \frac{2}{n}\sum_i R(x_i)R(y_i)\right\} \\
= \ & 1 - \frac{2(n+1)(2n+1)}{(n+1)(n-1)} + \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) \\
= \ & \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) + \frac{(n-1) - 2(2n+1)}{n-1} \\
= \ & \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) + \frac{n-1-4n-2}{n-1} \\
= \ & \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) - \frac{3n+3}{n-1} \\
= \ & \left(\frac{12}{n^2-1}\right)\frac{1}{n}\sum_i R(x_i)R(y_i) - \frac{3(n+1)}{n-1} \\
= \ & r_S \qquad \text{par (1)},
\end{aligned}
$$

ce que nous voulions démontrer.