Statistiques: Série 5
Corrigé

Exercice 1.

a) On construit le tableau des données

X	Y	$X \cdot Y$	X^2	Y^2
-4	16	-64	16	256
-2	10	-20	4	100
1	1	1	1	1
4	-8	-32	16	64
-1	19	-115	37	421

On en tire

— $\mathrm{Var}(X) = \overline{X^2} - \overline{X}^2 = \frac{37}{4} - \left(-\frac{1}{4}\right)^2 = \frac{147}{16}$;

— $\mathrm{Var}(Y) = \overline{Y^2} - \overline{Y}^2 = \frac{421}{4} - \left(\frac{19}{4}\right)^2 = \frac{1323}{16}$;

— $\mathrm{Cov}(X;Y) = \overline{X \cdot Y} - \overline{X} \cdot \overline{Y} = -\frac{115}{4} - \frac{-1}{4} \cdot \frac{19}{4} = -\frac{441}{16}$.

Il s'ensuit que

$$p = \frac{\mathrm{Cov}(X;Y)}{\mathrm{Var}(X)} = \frac{-\frac{441}{16}}{\frac{147}{16}} = -\frac{441}{147} = -3$$

et

$$h = \overline{Y} - p\overline{X} = \frac{19}{4} - (-3) \cdot \frac{-1}{4} = \frac{19}{4} - \frac{3}{4} = 4.$$

Ainsi, la droite des moindres carrés admet l'équation

$$y = -3x + 4.$$

Comme

$$r = \frac{\mathrm{Cov}(X;Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{-\frac{441}{16}}{\sqrt{\frac{147}{16}} \cdot \sqrt{\frac{1323}{16}}} = -1,$$

La corrélation est totale ; les quatre points étant alignés.

b) On construit le tableau des données

X	Y	$X \cdot Y$	X^2	Y^2
-2	4	-8	4	16
-1	1	-1	1	1
1	1	1	1	1
2	4	8	4	16
0	10	0	10	34

On en tire

— $\mathrm{Var}(X) = \overline{X^2} - \overline{X}^2 = \dfrac{10}{4} - 0^2 = \dfrac{5}{2}$;

— $\mathrm{Var}(Y) = \overline{Y^2} - \overline{Y}^2 = \dfrac{34}{4} - \left(\dfrac{10}{4}\right)^2 = \dfrac{9}{4}$;

— $\mathrm{Cov}(X;Y) = \overline{XY} - \overline{X} \cdot \overline{Y} = 0 - 0 \cdot \dfrac{10}{4} = 0$.

Il s'ensuit que

$$p = \frac{\mathrm{Cov}(X;Y)}{\mathrm{Var}(X)} = \frac{0}{\frac{5}{2}} = 0$$

et

$$h = \overline{Y} - p\overline{X} = \frac{5}{2} - 0 \cdot \frac{0}{4} = \frac{5}{2}.$$

Ainsi, la droite des moindres carrés admet l'équation

$$y = \frac{5}{2}.$$

Comme

$$r = \frac{\mathrm{Cov}(X;Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{0}{\sqrt{\frac{5}{2}} \cdot \sqrt{\frac{9}{4}}} = 0,$$

La corrélation est nulle ; les quatre points étant situés sur une parabole.

Exercice 2.

a) On construit le tableau des données

X	Y	XY	X^2	Y^2
20	80	1'600	400	6'400
28	142	3'976	784	20'164
32	176	5'632	1'024	30'976
34	191	6'494	1'156	3'6481
36	210	7'560	1'296	44'100
150	799	25'252	4'660	138'121

On en tire

— $\text{Var}(X) = \overline{X^2} - \overline{X}^2 = \dfrac{4060}{5} - \left(\dfrac{150}{5}\right)^2 = 32\,;$

— $\text{Var}(Y) = \overline{Y^2} - \overline{Y}^2 = \dfrac{138'121}{5} - \left(\dfrac{799}{5}\right)^2 = \dfrac{52'204}{25}\,;$

— $\text{Cov}(X;Y) = \overline{XY} - \overline{X} \cdot \overline{Y} = \dfrac{25'262}{5} - \dfrac{150}{5} \cdot \dfrac{799}{5} = \dfrac{1'292}{5}.$

On a

$$p = \frac{\text{Cov}(X;Y)}{\text{Var}(X)} = \frac{\frac{1'292}{5}}{32} = \frac{323}{40}$$

et

$$h = \overline{Y} - p\overline{X} = \frac{799}{5} - \frac{323}{40} \cdot \frac{150}{5} = \frac{1649}{20}.$$

D'où $y = \dfrac{323}{40}x - \dfrac{1649}{20}.$

b) Comme

$$r = \frac{\text{Cov}(X;Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{\frac{1292}{5}}{\sqrt{32} \cdot \sqrt{\frac{52'204}{25}}} \cong 0,9996,$$

la corrélation linéaire entre X et Y est très forte.

c) Pour $x = 15^{\text{C}}$, le taux d'ozone y est donné par

$$y = \frac{323}{40} \cdot 15 - \frac{1649}{20} \cong 38,675\ \mu\text{g/m}^3.$$

d) On pose $y = 240\ \mu\text{g/m}^3$:

$$\begin{aligned} \frac{323}{40}x - \frac{1649}{20} &= 240 \\ 323x - 3298 &= 9600 \\ 323x &= 12'898 \\ x &= \frac{12'898}{323} \cong 39,94°\text{C} \end{aligned}$$

Exercice 3.

a) On construit le tableau des données

X	Y	XY	X^2	Y^2
56	147	8232	3136	21609
42	125	5250	1764	15625
72	160	11520	5184	25600
36	118	4248	1296	13924
63	149	9387	3969	22201
47	128	6016	2209	16384
55	150	8250	3025	22500
49	145	7105	2401	21025
420	1122	60008	22984	158868

On en tire

— $\mathrm{Var}(X) = \overline{X^2} - \overline{X}^2 = \dfrac{22'984}{8} - \left(\dfrac{420}{8}\right)^2 = \dfrac{467}{4}$;

— $\mathrm{Var}(Y) = \overline{Y^2} - \overline{Y}^2 = \dfrac{158'868}{8} - \left(\dfrac{1'122}{8}\right)^2 = \dfrac{3015}{16}$;

— $\mathrm{Cov}(X;Y) = \overline{XY} - \overline{X} \cdot \overline{Y} = \dfrac{60'008}{8} - \dfrac{420}{8} \cdot \dfrac{1'122}{8} = \dfrac{1103}{8}$.

b) On a

$$p = \frac{\mathrm{Cov}(X;Y)}{\mathrm{Var}(X)} = \frac{\frac{1103}{8}}{\frac{467}{4}} = \frac{1103}{934}$$

et

$$h = \overline{Y} - p\overline{X} = \frac{1122}{8} - \frac{1103}{934} \cdot \frac{420}{8} = \frac{36543}{467}.$$

D'où

$$y = \frac{1103}{934}x + \frac{36543}{467}.$$

c) Comme

$$r = \frac{\mathrm{Cov}(X;Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{\frac{1103}{468}}{\sqrt{\frac{467}{4}} \cdot \sqrt{\frac{3015}{16}}} \cong 0,9295,$$

la corrélation linéaire entre X et Y est très forte.

d) La tension d'une femme âgée d'environ 43 ans est de

$$y = \frac{1103}{934} \cdot 43 + \frac{36543}{467} \cong 129,031 \text{ mmHg}.$$