Preuve

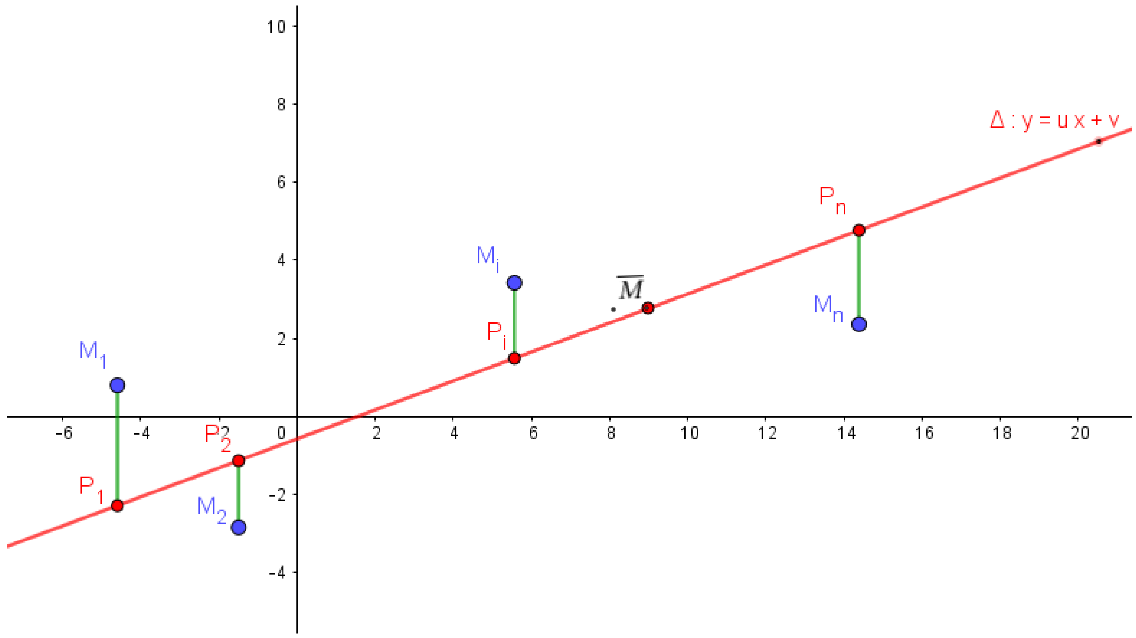
Soit $n \in \mathbb{N}^*$ et soient $x = (x_1; \ldots; x_n)$ et $y = (y_1; \ldots; y_n)$ deux séries statistiques à n termes réels.

Soient $M_1(x_1; y_1); \ldots; M_n(x_n; y_n)$ les points de coordonnées respectives $(x_1; y_1); \ldots; M(x_n; y_n)$ dans un repère orthogonal du plan.

Soit $\overline{M}(\bar{x}; \bar{y})$ le point de coordonnées les moyennes $\bar{x}$ et $\bar{y}$ séries statistiques x et y.

Soit $(u; v)$ un couple de nombres réels et soit Δ la droite d'équation $y = ux + v$.

Pour tout entier i entre 1 et n, soit P_i le point d'abscisse x_i de la droite Δ.



La droite Δ d"équation $y = ux + v$ passe par le point moyen $\overline{M}(\bar{x}; \bar{y})$ si et seulement si

$$\bar{y} = u\bar{x} + v$$

$$v = \bar{y} - u\bar{x} \tag{1}$$

Par définition, le point M_i a pour coordonnées $(x_i; y_i)$, et le point P_i étant le point d'abscisse x_i de la droite Δ d"équation $y = ux + v$, les coordonnées du point P_i sont $(x_i; ux_i + v)$.

Les points P_i et M_i ayant la même abscisse x_i, la distance P_iM_i est la valeur absolue de la différence de leurs ordonnées, de telle sorte que

$$\begin{aligned} P_iM_i^2 &= \left(y_i - (ux_i + v)\right)^2 \\ &= \left(y_i - ux_i - v\right)^2 \end{aligned}$$

et puisque $v = \bar{y} - u\bar{x}$ d'après (1), on obtient

$$\begin{aligned} P_iM_i^2 &= \left(y_i - ux_i - \bar{y} + u\bar{x}\right)^2 \\ &= \left((y_i - \bar{y}) - u(x_i - \bar{x})\right)^2 \\ &= \left(y_i - \bar{y}\right)^2 - 2\left(y_i - \bar{y}\right) u\,(x_i - \bar{x}) + (u(x_i - \bar{x}))^2 \\ &= \left(y_i - \bar{y}\right)^2 - 2u\,(x_i - \bar{x})\left(y_i - \bar{y}\right) + u^2\,(x_i - \bar{x})^2 \\ &= (x_i - \bar{x})^2\,u^2 - 2\,(x_i - \bar{x})\left(y_i - \bar{y}\right)u + \left(y_i - \bar{y}\right)^2 \end{aligned}$$

Par suite

$$\begin{aligned} P_1M_1^2 + \ldots + P_nM_n^2 &= \sum_{i=1}^{n} P_iM_i^2 \\ &= \sum_{i=1}^{n}\left[(x_i - \bar{x})^2\,u^2 - 2\,(x_i - \bar{x})\left(y_i - \bar{y}\right)u + \left(y_i - \bar{y}\right)^2\right] \\ &= \sum_{i=1}^{n}\left[(x_i - \bar{x})^2\,u^2\right] - \sum_{i=1}^{n}\left[2\,(x_i - \bar{x})\left(y_i - \bar{y}\right)u\right] + \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \\ &= \left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]u^2 - 2\left[\sum_{i=1}^{n}(x_i - \bar{x})\left(y_i - \bar{y}\right)\right]u + \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \\ &= n\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]u^2 - 2n\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\left(y_i - \bar{y}\right)\right]u + n\left[\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right] \\ &= n\,\mathrm{var}(x)\,u^2 - 2n\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\left(y_i - \bar{y}\right)\right]u + n\,\mathrm{var}(y) \\ &= n\,\mathrm{var}(x)\,u^2 - 2n\,\mathrm{cov}(x;y)\,u + n\,\mathrm{var}(y) \quad \text{où} \quad \mathrm{cov}(x;y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\left(y_i - \bar{y}\right) \\ &= au^2 + bu + c \end{aligned}$$

avec

$$\left\{\begin{array}{rcr} a &=& n\,\mathrm{var}(x) \\ b &=& -2n\,\mathrm{cov}(x;y) \\ c &=& n\,\mathrm{var}(y) \end{array}\right.$$

La tableau de variation de la fonction $au^2 + bu + c$ avec $a > 0$ est

u	$-\infty$	$-\frac{b}{2a}$	$+\infty$
$au^2 + bu + c$	$\searrow$	$-\frac{\Delta}{4a}$	$\nearrow$

avec $\Delta = b^2 - 4ac$.

Par suite, $P_1M_1^2 + \ldots + P_nM_n^2 = au^2 + bu + c$ est minimal lorsque

$$u = -\frac{b}{2a} = -\frac{-2n\,\mathrm{cov}(x;y)}{2n\,\mathrm{var}(x)} = \frac{\mathrm{cov}(x;y)}{\mathrm{var}(x)}$$

et le minimum de $P_1M_1^2 + \ldots + P_nM_n^2 = au^2 + bu + c$ est

$$\begin{aligned}
-\frac{\Delta}{4a} &= -\frac{1}{4a} \times \left(b^2 - 4ac\right) \\
&= -\frac{1}{4n\,\mathrm{var}(x)} \times \left[\left(-2n\,\mathrm{cov}(x;y)\right)^2 - 4 \times n\,\mathrm{var}(x) \times n\,\mathrm{var}(y)\right] \\
&= -\frac{1}{4n\,\mathrm{var}(x)} \times \left[4n^2\,\mathrm{cov}(x;y)^2 - 4n^2\,\mathrm{var}(x)\,\mathrm{var}(y)\right] \\
&= -\frac{4n^2}{4n\,\mathrm{var}(x)} \times \left[\mathrm{cov}(x;y)^2 - \mathrm{var}(x)\,\mathrm{var}(y)\right] \\
&= -\frac{n}{\mathrm{var}(x)} \times \left[\mathrm{cov}(x;y)^2 - \sigma_x^2\sigma_y^2\right] \\
&= \frac{n}{\mathrm{var}(x)} \times \left[\sigma_x^2\sigma_y^2 - \mathrm{cov}(x;y)^2\right] \\
&= \frac{n\sigma_x^2\sigma_y^2}{\mathrm{var}(x)} \times \left[1 - \frac{\mathrm{cov}(x;y)^2}{\sigma_x^2\sigma_y^2}\right] \\
&= \frac{n\sigma_x^2\sigma_y^2}{\mathrm{var}(x)} \times \left[1 - \left(\frac{\mathrm{cov}(x;y)}{\sigma_x\sigma_y}\right)^2\right] \\
&= k\left[1 - r^2\right] \quad \text{avec} \quad r = \frac{\mathrm{cov}(x;y)}{\sigma_x\sigma_y} \quad \text{et} \quad k = \frac{n\sigma_x^2\sigma_y^2}{\mathrm{var}(x)}
\end{aligned}$$

Puisque $k\left(1 - r^2\right)$ est le minimum de $P_1M_1^2 + \ldots + P_nM_n^2 \geq 0$, on a

$$k\left(1 - r^2\right) \geq 0$$

et compte tenu que $k \geq 0$ puisque $k = \frac{n\sigma_x^2\sigma_y^2}{\mathrm{var}(x)}$, on a $1 - r^2 \geq$, soit $r^2 \leq 1$, ou encore

$$-1 \leq r \leq 1$$

Notons également que, lorsque $r = \pm 1$, on a $k\left(1 - r^2\right) = 0$, c'est-à-dire

$$P_1M_1^2 + \ldots + P_nM_n^2 = 0$$

$$\forall\, i \in \{1; \ldots; n\} \quad P_iM_i^2 = 0$$

$$\forall\, i \in \{1; \ldots; n\} \quad M_i = P_i$$

autrement dit tous les points $P_1, \ldots, P_n$ sont alignés sur la droite Δ.

Pour conclure, notons que si on considère

$$P_1M_1^2 + \ldots + P_nM_n^2 = k\left(1 - r^2\right) \quad \text{avec} \quad r = \frac{\mathrm{cov}(x;y)}{\sigma_x\sigma_y} \quad \text{et} \quad k = \frac{n\sigma_x^2\sigma_y^2}{\mathrm{var}(x)}$$

comme une fonction $f(r) = k\left(1 - r^2\right)$ de $r \in [-1;1]$ avec $k \geq 0$, on a $f'(r) = -2kr$ et

r	-1		0		1
$f'(r)$		$+$	0	$-$	
$f(r)$	0	$\nearrow$		$\searrow$	0

de telle sorte que le nuage de points $M_1\left(x_1; y_1\right); \ldots; M_n\left(x_n; y_n\right)$ a une forme d'autant plus allongée que $f(r)$ est proche de 0 soit que r est proche de ± 1.