1 Equation de la droite des moindres carrés

A retenir

La droite des moindres carrés d'une série statistique double (x ;y) a pour équation : $y = a(x - \overline{x}) + \overline{y}$ avec $a = \dfrac{cov(x;y)}{V(x)}$

Le principe

On va chercher à minimiser la somme des carrés des distances entre les points du nuage et ceux de la droite d'ajustement .

On va considérer cette somme d'abord comme un polynôme du second degré en b et calculer sa dérivée

Puis comme un polynôme du second degré en a et calculer sa dérivée

La démonstration

Soient $M_i(x_i;y_i)$ les points du nuage pour i allant de à n .

Soit D la droite d'ajustement affine d'équation $y = ax + b$

Soient $P_i(x_i;ax_i + b$ les points de la droite .

Par définition , la droite des moindre carrés vérifie a et b minimisent la somme $S = M_1P_1^2 + M_2P_2^2 + ... + M_nP_n^2$

Or : $S = \displaystyle\sum_{k=1}^{n}(y_k - ax_k - b)^2$

$= \displaystyle\sum_{k=1}^{n}((y_k - ax_k) - b)^2$

$= \displaystyle\sum_{k=1}^{n}(y_k - ax_k)^2 - 2b\sum_{k=1}^{n}(y_k - ax_k) + \sum_{k=1}^{n} b^2$

$= \displaystyle\sum_{k=1}^{n}(y_k - ax_k)^2 - 2b\sum_{k=1}^{n}(y_k - ax_k) + nb^2$

S peut donc être considéré comme un polynôme du second degré en b . On sait que n est positif donc S admet un minimum lorsque sa dérivée est nulle .

$$S' = 0 \iff -2\sum_{k=1}^{n}(y_k - ax_k) + 2bn = 0 \iff b = \frac{\displaystyle\sum_{k=1}^{n}(y_k - ax_k)}{n}$$

Or $\dfrac{\displaystyle\sum_{k=1}^{n}(y_k - ax_k)}{n} = \dfrac{\displaystyle\sum_{k=1}^{n} y_k}{n} - a\dfrac{\displaystyle\sum_{k=1}^{n} x_k}{n} = \overline{y} - a\overline{x}$

On a donc en remplaçant dans l'expression du départ :

$$S = \sum_{k=1}^{n}(y_k - ax_k - b)^2 = \sum_{k=1}^{n}(y_k - ax_k - \overline{y} + a\overline{x})^2 = \sum_{k=1}^{n}((y_k - \overline{y}) - a(x_k - \overline{x}))^2$$

$= \sum_{k=1}^{n}(y_k - \overline{y})^2 - 2a\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y}) + a^2\sum_{k=1}^{n}(x_k - \overline{x})^2$

$= nV(y) - 2ancov(x;y) + a^2nV(x)$

On peut donc considérer maintenant S comme un polynôme du second degré en a , minimum quand sa dérivée s'annule

$S' = 0 \iff -2ncov(x;y) + 2anV(x) = 0 \iff a = \dfrac{cov(x;y)}{V(x)}$