

Régression orthogonale

Denis Vekemans *

Soit (X, Y) une série double représentée dans le plan rapporté à un repère (O, \vec{i}, \vec{j}) par les points $M_k(x_k, y_k)$, $k \in \{1, 2, \dots, n\}$.

On considère le point moyen $G(m(X), m(Y))$ et le changement de variables $x' = x - m(X)$, $y' = y - m(Y)$.

Etant donnée une droite Δ dont une équation dans le repère (G, \vec{i}, \vec{j}) est

$$x' \cos \alpha + y' \sin \alpha - \rho = 0.$$

On note H_k la projection orthogonale de M_k sur la droite Δ .

On se propose de déterminer Δ (i.e. α et ρ) de manière à minimiser

$$S = \sum_{k=1}^n \overline{H_k M_k}^2.$$

1. Montrer que $S = \sum_{k=1}^n (x'_k \cos \alpha + y'_k \sin \alpha - \rho)^2$.
2. Montrer que s'il existe une droite Δ minimisant S , cette droite passe par G .
3. Déterminer le coefficient directeur de Δ qui rend S minimum (préciser dans quel cas, la solution est unique).

La droite Δ ajustée est appelée droite de régression orthogonale.

4. On suppose que la droite de régression orthogonale est unique. Soient D et D' les droites de régression de Y par rapport à X et de X par rapport à Y . Etudier la position relative des droites D , D' et Δ .
5. Soit $T = \sum_{k=1}^n \overline{GM_k}^2$ et $S' = \sum_{k=1}^n \overline{GH_k}^2$.

Montrer que minimiser S revient à maximiser S' .

On pose $q = \frac{S'}{T}$.

Montrer que q est un indice de qualité de l'ajustement compris entre $\frac{1}{2}$ et 1.

Quelle est la signification des valeurs $q = \frac{1}{2}$ et $q = 1$?

6. On donne les notes d'un groupe de 8 étudiants dans deux disciplines : mathématiques (note x sur 20) et informatique (note y sur 20).

x	12	15	8	6	10	12	11	9
y	10	16	11	4	9	14	17	8

*Laboratoire de mathématiques pures et appliquées Joseph Liouville ; 50, rue Ferdinand Buisson BP 699 ; 62 228 Calais cedex ; France

Ajuster sur ces données les deux droites de régression D et D' et calculer le coefficient de corrélation r .

Ajuster sur ces données la droite de régression orthogonale Δ et calculer le coefficient de qualité q de cet ajustement.

Interpréter l'ajustement Δ .

Références

- [1] C. RAFFIN, *Statistiques et Probabilités, DEUG Scientifiques 2^e année*, Flash U, Armand Colin, 1987.