**RÉGRESSION LINÉAIRE (D2, J)**
(18 / 10 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) De façon générale, une **régression linéaire** est une **régression** dont la propriété principale est la linéarité par rapport au(x) **paramètre**(s) (cf aussi **hypothèse linéaire**, **linéaire**, **modèle de régression linéaire**, **modèle d'interdépendance linéaire**, **modèle linéaire**, **problème linéaire**).

Ainsi, le modèle de régression multuple linéaire standard s'écrit (dans l'**espace des variables**) :

(1) $\eta = \xi' b + \varepsilon$, avec $E\, \varepsilon / \xi = 0$,

où $(\xi, b) \mapsto \xi' b \in \mathbf{R}$ est une forme bilinéaire sur $\mathbf{R}^K$ (donc marginalement une **forme linéaire** pr à $\xi$) (cf **forme multilinéaire**), $\xi$ est à valeurs dans $\mathbf{R}^K$, les **vars** $\eta$ et $\varepsilon$ sont à valeurs dans $\mathbf{R}$, et $b \in \mathbf{R}^K$ représente le **paramètre d'intérêt** de la régression.

Sa forme « observée » s'écrit (dans l'**espace des observations**) :

(2) $y = X b + u$, avec $E\, u / X = 0$, $V\, u / X = \sigma_u^2\, \Omega$,

où l'on a, en outre, spécifié la **dispersion** de u, avec $\sigma_u > 0$ et $\Omega \in S_N(\mathbf{R})$ (**matrice symétrique**). On a ici, de même, une **application linéaire** $b \mapsto X b$ de $\mathbf{R}^K$ dans $\mathbf{R}^N$ (non aléatoire si la **matrice** des exogènes X est supposée fixée).

(ii) La forme analytique du modèle peut être **non linéaire** pr aux (observations des) variables qu'il inclut (cf aussi **non linéarité**).

Ainsi, le modèle élémentaire suivant est-il linéaire pr à b mais non linéaire pr à $\xi$ (modèle logarithmique) :

(1)' $\eta = (\mathrm{Log}\, \xi)' b + \varepsilon$, avec $E\, \varepsilon / \xi = 0$,

où l'on note symboliquement $\mathrm{Log}\, \xi = (\mathrm{Log}\, \xi_1, ..., \mathrm{Log}\, \xi_K)$ et $\xi \in (\mathbf{R}_+^*)^K$.

Des **changements de variables** appropriés permettent de transformer tout modèle non linéaire pr aux observations (mais linéaire pr au paramètre) en un modèle linéaire pr à ces deux données.

(iii) Comme pour toute **procédure statistique** impliquant des relations « linéaires », l'intérêt d'une régression linéaire réside dans la relative facilité des calculs (algèbre linéaire).