ÉQUATIONS NORMALES (H3)
(26 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

Les **équations normales** sont des **équations estimantes** particulières, qui interviennent principalement dans l'étude du **modèle de régression** ou d'autres modèles analogues.

L'écriture et la terminologie adoptent souvent la forme plurielle lorsqu'elles se réfèrent à un « ensemble » d'**équations normales** « scalaires », ie à une **suite** (finie) de conditions nécessaires du premier ordre.

Lorsque la notion est définie dans un contexte « vectoriel », l'**espace d'observation** disponible est un espace vectoriel (eg $\mathbf{R}^N$) : on parle alors d'**équation normale** (vectorielle) au singulier.

(i) Soit $y = z + u$, avec $z = X\, b$, $E\, u = 0$ et $V\, u = \Sigma > 0$, un **modèle de régression linéaire**, exprimé dans un **espace d'observation** vectoriel (fini), de dimension N (généralement $\mathbf{R}^N$).

On appelle (ensemble des) **équations « normales »** du **modèle linéaire** précédent la suite des équations exprimant les conditions du premier ordre nécessaires pour minimiser $\|y - z\|_\Sigma$ (**distance** de y à z dans la **métrique** définie par la **dispersion** Σ), ie tq :

(0) $\|y - z\|_\Sigma^2 = (y - z)' \, \Sigma^{-1} \, (y - z).$

Comme $\|u\|_\Sigma^2 = \|y - z\|_\Sigma^2 = \|y - X\, b\|_\Sigma^2 = (y - X\, b)' \, \Sigma^{-1} \, (y - X\, b)$, l'annulation de la dérivée première pr à b définit l'**équation normale** vectorielle :

(1) $X' \, \Sigma^{-1} \, X\, b = X' \, \Sigma^{-1} \, y.$

Réécrite sous la forme :

(1)' $X' \, \Sigma^{-1} \, (y - X\, b) = X' \, \Sigma^{-1} \, u = 0,$

cette équation s'interprète ainsi : lorsque b est solution de (1), les vecteurs colonnes de X sont orthogonaux (dans la métrique Σ ainsi définie dans $\mathbf{R}^N$) au vecteur $u = y - z$ représentant la **perturbation aléatoire**.

(ii) Dans le cas d'un **modèle non linéaire** de régression $y = F(b) + u$, avec $E\, u = 0$ et $V\, u = \Sigma > 0$, la notion se transpose directement.

L'équation normale est alors l'équation (vectorielle) traduisant la condition du premier ordre nécessaire pour minimiser la distance $\|y - z\|_\Sigma$ de y à z selon la métrique Σ. Ceci implique que la **dérivée** première de l'expression :

(2) $\|u\|_\Sigma^2 = \|y - z\|_\Sigma^2 = \|y - F(b)\|_\Sigma^2 = (y - F(b))' \, \Sigma^{-1} \, (y - F(b))$

pr à b est nulle, ce qui conduit à l'**équation normale** :

(3) $(D\ F(b))'\ \Sigma^{-1}\ F(b) = (D\ F(b))'\ \Sigma^{-1}\ y$,

où D F(b) est la **matrice** de l'**application dérivée** D F $\in \mathscr{L}(\mathbf{R}^Q, \mathbf{R}^N)$ de F au point b $\in \mathbf{R}^Q$.

Comme précédemment, l'équation (3), réécrite selon :

(3)' $(D\ F(b))'\ \Sigma^{-1}\ (y - F(b)) = (D\ F(b))'\ \Sigma^{-1}\ u = 0$,

s'interprète ainsi : lorsque b est solution de (3), les vecteurs colonnes de D F(b) sont Σ-orthogonaux (dans $\mathbf{R}^N$) à u = y - z.

En particulier, si Q = K et F(b) = X b, (3) se réduit à (1).

(iii) Dans l'expression **équations normales**, le qualificatif « normal » dérive de l'**hypothèse** de **normalité** $u \sim \mathcal{N}_N(0, \Sigma)$ (**loi normale multidimensionnelle**), souvent supposée dans la **méthode des moindres carrés**.

Ainsi (eg en matière de **test** ou d'**estimation** par **vraisemblance**), on suppose que $y / z \sim \mathcal{N}_N(z, \Sigma)$, ie de façon explicite que $\mathscr{L}(y / X) = \mathcal{N}_N(X\ b, \Sigma)$ dans le cas linéaire, ou que $\mathscr{L}(y / X) = \mathcal{N}_N(F(b), \Sigma)$ dans le cas non linéaire.

Cependant, cette hypothèse de normalité n'est pas nécessaire à la définition de la méthode des mco elle-même.