COVARIANCE (C5, F3)
(20 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

La **covariance** entre deux **va** indique comment celles-ci varient l'une pr à l'autre. Alors que l'**espérance mathématique** correspond à une opération linéaire (cf **forme linéaire**), la covariance correspond à une opération bilinéaire (cf **forme multilinéaire**).

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\zeta = (\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **couple aléatoire** réel.

On appelle **covariance (théorique)** entre ξ et η le nombre réel :

(1) $C(\xi, \eta) = E(\xi - E\,\xi) . (\eta - E\,\eta) = E(\xi . \eta) - (E\,\xi) . (E\,\eta)$.

Si le couple ζ est de carré intégrable, alors (**inégalité de CAUCHY-SCHWARZ**) la covariance existe (ie est finie). Elle s'interprète comme le **produit scalaire** des va $\xi - E\,\xi$ et $\eta - E\,\eta$ dans l'espace $L_{\mathbf{R}}^2(\Omega, \mathcal{T}, P)$. L'application C définie en (1), linéaire pr à chacun de ses arguments, est une application bilinéaire sur cet espace (cf **application multilinéaire**).

(ii) La notion de covariance permet de définir celle de **variance**, donc celle d'**écart-type** : en effet, une variance est une covariance particulière (la covariance de ξ avec elle-même.

(iii) Une généralisation de la notion de covariance réside dans celle de **matrice de covariance**, définie pour un **vecteur aléatoire** (cf aussi **opérateur de covariance**).

(iv) On peut étendre la définition (1) en supposant que ξ est à valeurs dans $\mathbf{R}^K$ et η à valeurs dans $\mathbf{R}^G$.

On définit alors la **matrice des inter-covariances**, ou **matrice des covariances croisées** (ou parfois simplement des covariances), ou **matrice d'inter-covariance**, **(théorique)** selon :

(2) $C(\xi , \eta) = E(\xi - E\,\xi)(\eta - E\,\eta)' = E(\xi\,\eta') - (E\,\xi)(E\,\xi)'$.

En particulier, si G = K et $\eta = \xi$, on obtient la définition d’une matrice de covariance précédente.

(v) Tout ce qui précède conduit directement à des équivalents empiriques.

Si (X, Y) est un N-**échantillon aléatoire** issu du couple (ξ, η), il suffit de remplacer, dans les calculs de définition, la probabilité théorique par la **probabilité empirique** P_N déduite de (X, Y).