TRANSFORMATION DE FISHER (C2, C5, F3)
(09 / 10 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **transformation de FISHER** possède deux significations :

(a) transformation du coefficient de corrélation linéaire empirique ;

(b) transformation d'un rapport de variances.

(i) **Première signification**. Soit $(\Omega , \mathcal{T}, P)$ un **espace probabilisé**, $(\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **couple aléatoire** réel et $((X_1 , Y_1) ,..., (X_N , Y_N))$ un **échantillon aléatoire** indépendant constitué de **copies** de (ξ, η), ie un **échantillon indépendant équidistribué** selon la loi $P^{(\xi, \eta)}$ du couple (ξ, η).

On note r_{XY} le **coefficient de corrélation linéaire** empirique entre les échantillons partiels $X = (X_1 ,..., X_N)$ et $Y = (Y_1 ,..., Y_N)$.

On appelle alors **transformation de R.A. FISHER** la fonction suivante, définie par l'argument de la tangente hyperbolique :

(1) $r_{XY} \mapsto z_{XY} = \text{Arg th } r_{XY} = (1/2) \text{ Log } \{(1 - r_{XY})^{-1} . (1 + r_{XY})\}$,

avec Arg th : $]0, 1[\mapsto \mathbf{R}$. La **variable aléatoire** (ou **statistique**) z_{XY} est donc à valeurs dans $\mathbf{R}$. On l'utilise souvent à la place de r_{XY} car, dans le cas gaussien où $P^{(\xi, \eta)} = \mathcal{N}_2 (\mu, \Sigma)$ (**loi normale multidimensionnelle** à deux dimensions), la **convergence en loi** suivante :

(2) $\mathcal{L}\{(\sigma (z_{XY}))^{-1} . (z_{XY} - E z_{XY})\} \to_{N \to +\infty} \mathcal{N}_1 (0, 1)$ (**loi normale réduite**)

est plus « rapide » que celle correspondant à r_{XY} , où l'on note $E z_{XY}$ l'**espérance** et $\sigma (z_{XY})$ l'**écart-type** de z_{XY} .

(ii) **Seconde signification**. On appelle aussi **transformation de R.A. FISHER** une transformation du **rapport des variances** (empiriques) de deux échantillons. Si X^i est un échantillon de taille N_i dont la **variance empirique** est notée S_i^2 (i = 1, 2), cette transformation de FISHER est de la forme :

(3) $(X^1, X^2) \mapsto Z_{N(1)N(2)} = (1/2) \text{ Log } (S_1^2 / S_2^2)$,

où N(i) désigne, par commodité, N_i (i = 1, 2).

La statistique $Z_{N(1)N(2)}$ est souvent simplement notée Z ou z.