COEFFICIENT DE CORRÉLATION DES RANGS DE SPEARMAN
(14 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit (ξ, η) un **couple aléatoire**, dont la **loi** est notée $P^{(\xi, \eta)}$, (X, Y) ou $((X_1, Y_1), ..., (X_N, Y_N))$ un N-**échantillon** (non nécessairement indépendant) généré par ce couple, $R = (R_1, ..., R_N)$ et $S = (S_1, ..., S_N)$ les **statistiques de rang** resp associées à $X = (X_1, ..., X_N)$ et $Y = (Y_1, ..., Y_N)$.

Le **coefficient de corrélation des rangs de SPEARMAN** (empirique) est défini comme le **coefficient de corrélation linéaire** (empirique) entre les vecteur de rangs R et S (cf **coefficient de corrélation des rangs**), ie comme **statistique** :

(1) $\quad r_{XY}' = r_{RS} = ||P R||^{-1} ||P S||^{-1} R' P S,$

dans laquelle $P \in M_N(\mathbf{R})$ désigne **matrice de centrage** pr à chaque **moyenne empirique** (celle de R et celle de S).

Ce coefficient, souvent aussi noté $R_{S,N}$, ou ρ_{XY} ou simplement ρ (« rho » de SPEARMAN), vaut :

(2) $\quad r_{XY}' = 1 - 6 N^{-1} (N^2 - 1)^{-1} \Sigma_{n=1}^{N} D_n^2,$

avec $D_n = R_n - S_n$ (différence entre rangs), formule qui s'écrit aussi :

(2)' $\quad r_{XY}' = 1 - 6 N^{-1} (N^2 - 1)^{-1} ||R - S||^2.$

(ii) Dans le cadre du **problème de l'indépendance**, ce coefficient sert de **statistique de test** pour l'hypothèse H_0 d'**indépendance** (plus exactement, de non **corrélation**) entre ξ et η, ie :

$H_0 : \ P^{(\xi, \eta)} = P^{\xi} \otimes P^{\eta},$

où P^{ξ} (resp P^{η}) est la loi propre (ie la **loi marginale**) de ξ (resp η), contre des hypothèses alternatives H_1 de corrélation (positive ou négative).

Si l'on note T_N la statistique de test associée au **test de SPEARMAN**, le coefficient de SPEARMAN s'écrit :

(3) $\quad r_{XY}' = 12 . N^{-1} (N^2 - 1)^{-1} \{T_N - (1/4) N (N+1)^2\}.$

On montre que, quelles que soient les lois P^{ξ} et P^{η}, la loi de r_{XY}' est une **loi symétrique** pr à 0. Par suite :

$$(4) \quad \begin{aligned} & E_0 (r_{XY}')^{j} = 0, \qquad \forall j \in 2\mathbf{N} + 1 \text{ (impairs)}, \\ & V_0 (r_{XY}') = (N-1)^{-1}, \ \forall N \in \mathbf{N}^* \setminus \{1\}. \end{aligned}$$

La **distribution** de r_{XY}' sous H_0 est tabulée.

De plus, on établit les **propriétés asymptotiques** suivantes, vraies sous H_0 :

(a) convergence vers la loi normale (cf **normalité asymptotique**) :

(5) $\mathscr{L}\{(N-1)^{1/2}\, r_{XY}'\} \to_{N \to +\infty} \mathscr{N}(0, 1)$ (**loi normale réduite**) ;

(b) **convergence en loi** vers la **loi de STUDENT** (à N - 2 **degrés de liberté**). La statistique :

(4) $T_N = (N - 2)^{1/2} . \{1 - (r_{XY}')^2\}^{-1/2} . r_{XY}'$

vérifie :

(5) $\mathscr{L}(T_N) \to_{N \to +\infty} \mathscr{S}_{N-2}$.

Ces propriétés permettent de tester la non corrélation. Ainsi, lorsque l'**alternative** H_1 suppose une corrélation positive, la **région critique** du test (asymptotique) correspondant est de la forme :

(3) $w = [r_{XY}' < q_{1-\alpha}]$,

où q_α est le **quantile** d'ordre $\alpha \in]0, 1[$ ($\alpha << 1$) de la loi ($\mathscr{N}(0, 1)$ ou $\mathscr{S}_{N-2}$) considérée.