RAPPORT DE CORRÉLATION (C5, D2)
(20 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

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

On appelle **rapport de corrélation (théorique)** de η pr (ou relatif) à ξ, ou de η conditionnel(lement) à ξ (ou de η sachant ξ, ou encore de η en ξ), **simple** le nombre réel signé $\rho_{\eta\xi}$, dont le carré s'écrit :

(1) $\rho_{\eta / \xi}^2 = \{V_2\, \eta - E_1\, V(\eta / \xi)\} / V_2\, \eta = V_1\, E(\eta / \xi) / V_2\, \eta$,

rapport dans lequel $V_2\, \eta = V\, \eta$ désigne la **variance** propre (ie la variance marginale) de η, $V(\eta / \xi)$ la variance conditionnelle de η sachant ξ, E_1 l'opérateur espérance mathématique pr à ξ, V_1 la variance calculée pr à ξ et $E(\eta / \xi)$ l'espérance conditionnelle de η sachant ξ (cf **caractéristique marginale**, **caractéristique conditionnelle**).

Le rapport $\rho_{\eta / \xi}^2$ exprime un « **degré** » **de dépendance** de η pr à ξ, car il exprime la réduction relative de la variance de η lorsqu'on centre η pr à sa moyenne conditionnelle $E(\eta / \xi)$ au lieu de sa propre moyenne $E\, \eta = E_2\, \eta$ (espérance marginale).

Par permutation des rôles de ξ et η, on définit un rapport de corrélation « dual » $\rho_{\xi / \eta}^2 \leq 1$, défini (par permutation de rôle) selon $\rho_{\xi / \eta}^2 = (V\, \eta - E_2\, V_1) / V\, \eta$.

(ii) Le rapport de corrélation (simple) possède diverses propriétés, dont les suivantes :

(a) si $V(\eta / \xi) = 0$, alors $V(\eta / \xi = x) = 0$ (P-p.s.), x e R. Autrement dit, à toute valeur de x ne correspond qu'une valeur $E^{\xi}\, \eta$ de η (P-p.s.). Par suite, $E_1\, (V^{\xi}\, \eta) = 0$ et $\rho_{\eta / \xi}^2 = 1$. Si ξ et η sont indépendantes, $E(\eta / \xi) = E\, \eta$ (et ne dépend donc pas de ξ) et $V_1\, (E(\eta / \xi)) = 0$. Par suite, $\rho_{\eta / \xi}^2 = 0$.

Cette situation correspond au cas où il existe une liaison « fonctionnelle » φ entre η et ξ, ie une fonction mesurable $\varphi : \mathbf{R} \mapsto \mathbf{R}$ tq $\eta = \varphi(\xi)$ (P-p.s.) (cf **application mesurable**). Dans ce cas, $\rho_{\eta / \xi}^2 = 1$;

(b) en général, $\rho_{\xi / \eta}^2 \neq \rho_{\eta / \xi}^2$;

(c) si $\rho_{\xi\eta}^2$ désigne le carré du **coefficient de corrélation linéaire** entre ξ et η, on utilise parfois le rapport $\rho_{\eta / \xi}^2 / \rho_{\xi\eta}^2$, ou toute fonction adéquate de ce rapport, comme **indice** de **non linéarité**. Un exemple de telle fonction est $(\rho_{\eta / \xi}^2 - \rho_{\xi / \eta}^2) / (1 - \rho_{\xi / \eta}^2)$.

(iii) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $(\xi, \eta) : \Omega \mapsto \mathbf{R}^K \times \mathbf{R}$ un **couple aléatoire** réel dans lequel ξ = est un **vecteur aléatoire** réel, à valeurs dans $\mathbf{R}^K$. On considère une sous-suite $(k_1 ,..., k_m)$ de $(1 ,..., K) = \mathbf{N}_K^*$ composée d'indices distincts deux à deux (ie $k_{j''} \neq k_{j'}$ si $j'' \neq j'$) ainsi que la sous-suite $\xi(m) = (\xi_{k(1)}, ..., \xi_{k(m)})$ correspondante, où k(i) désigne, par commodité, k_i (tout i = 1 ,..., m).

On appelle **rapport de corrélation (théorique)** de η pr (ou relatif) à $\xi(m)$, ou de η conditionnel(lement) à $\xi(m)$ (ou de η sachant $\xi(m)$, ou encore de η en $\xi(m)$), **multiple d'ordre (k_1 ,..., k_m)** le nombre réel signé $\rho_{\eta / \xi(m)}$, dont le carré s'écrit :

$$(2) \qquad \rho_{\eta / \xi(m)}{}^2 = \begin{cases} \{V_{k(1)...k(m)} (E (\eta / \xi(m))\} / V_2\, \eta \;=\; V_1\, E (\eta / \xi(m)) / V_2\, \eta, \\ 1 - \{E_{k(1)...k(m)} (V (\eta / \xi(m))\} / V_2\, \eta, \end{cases}$$

pour toute partie $\{k_1, ..., k_m\} \subset \{1, ..., K\}$ tq $j'' \neq j' \Rightarrow k_{j''} \neq k_{j'}$.

Il existe donc K (2^{K-1} - 1) rapports de corrélation multiples (théoriques) possibles de η pr à $\xi(m) = (\xi_{k(1)}, ..., \xi_{k(m)})$, avec $\{k_1, ..., k_m\} \subset \{1, ..., K\}$.

(v) On montre alors que (extensions des propriétés précédentes) :

(a) $0 \leq \rho_{\eta / \xi(m)}{}^2 \leq 1$, pour toute suite ($k_1$,..., k_m) définie comme précédemment et tout $m \in N_K{}^*$;

(b) η est non corrélée avec $\xi(m)$ ssi $\rho_{\eta / \xi(m)}{}^2 = 0$;

(c) η est liée fonctionnellement à $\xi(m)$ ssi $\rho_{\eta / \xi(m)}{}^2 = 1$ (auquel cas $V (\eta / \xi(m)) = 0$).

(v) Le **rapport de corrélation empirique (simple)** se définit, de façon analogue, à l'aide d'un **échantillon aléatoire** $((X_1, Y_1), ..., (X_N, Y_N))$ dont la **variable parente** est le couple (ξ, η), ie à l'aide de la **fonction de répartition empirique** fondée sur cet échantillon (cf **statistique naturelle**).

C'est une **statistique** (non paramétrique) qui permet de tester le caractère non linéaire d'une régression (cf **régression non linéaire**).

De même, le **rapport de corrélation empirique (multiple)** se définit, de façon analogue, à l'aide d'un **échantillon aléatoire** $((X_1(m), Y_1), ..., (X_N(m), Y_N))$ issu du couple $(\xi(m), \eta)$, ie à l'aide de la fr empirique fondée sur cet échantillon.