TEST DU COEFFICIENT DE CORRÉLATION (C5, D1, F3, I2)
(14 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Il existe autant de tests du coefficient de corrélation qu'il existe de types différents de **corrélations** ou de **coefficients de corrélation**. On se limite ici au test du **coefficient de corrélation linéaire**.

(i) Soit $(\Omega, \mathcal{T}, \mathcal{P})$ un **modèle statistique** fondamental et $\zeta = (\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **couple aléatoire** réel, supposé de carré intégrable et de **loi** $P^{\zeta} = P^{(\xi, \eta)}$.

Le coefficient de corrélation (théorique) est un **paramètre** de la (2,2)-**matrice des covariances** Σ du couple ζ.

On observe un N-**échantillon** $Z = ((X_1 , Y_1) ,...., (X_N , Y_N))$ de ζ, où le couple (X, Y) désigne une **copie** de (ξ, η), distribuée selon la loi $P^{(\xi, \eta)}$, et l'on note $X = (X_1 ,...., X_N)$ et $Y = (Y_1 ,...., Y_N)$ les sous-échantillons partiels (ie « échantillons marginaux ») extraits de Z. A toute loi P^{ζ}, image d'une probabilité P définie sur $\mathcal{T}$ par ζ, correspond un coefficient de corrélation linéaire (théorique) $\rho\ (\xi, \eta)$ ou $\rho_{\xi\eta}$.

Le **test du coefficient de corrélation** consister à vérifier une **hypothèse de base** tq :

(1) $H_0 : \rho_{\xi\eta} < 0$ (corrélation négative, ou « inverse »),

contre une **alternative** tq (test unilatéral) :

(2) $H_1 : \rho_{\xi\eta} \geq 0$ (corrélation positive, ou « directe »).

Il peut aussi porter sur une hypothèse tq :

(1)' $H_0' : \rho_{\xi\eta} = 0$ (absence de corrélation),

contre une alternative tq (test bilatéral) :

(2)' $H_1' : \rho_{\xi\eta} \neq 0$ (existence d'une corrélation).

La **statistique de test** associée aux tests précédents est généralement le **coefficient de corrélation empirique** entre X et Y, lequel est donc un estimateur « naturel » de $\rho_{\xi\eta}$ (cf **statistique naturelle**), ie :

(3) $r_{XY} = \{(X' P X)^{-1/2} (Y' P Y)^{-1/2} . X' P Y,$

où X et Y sont représentés en vecteurs colonnes et où P est la **matrice de centrage** pr à la moyenne (empirique).

(ii) Dans le cas gaussien, ie lorsque $\zeta \sim \mathcal{N}_2 (\mu, \Sigma)$ (**loi normale** à deux dimensions), et si les coordonnées (X_n , Y_n) de Z sont indépendantes entre elles, la loi de r_{XY} ne dépend que de $\rho_{\xi\eta}$. Par suite :

(a) dans le cas unilatéral {(1),(2)}, il existe un **test uniformément le plus puissant**, qui est aussi un **test sans biais**, défini par une **région critique** de la forme :

(4) $w = [r_{XY} \geq q_{1 - \alpha}]$

où $q_{1 - \alpha}$ est le **quantile** d'ordre $1 - \alpha$ de la loi de r_{XY} et $\alpha \in]0, 1[$ le seuil critique retenu.

Cette région w est déterminée par le fait que, sous l'hypothèse H_0' définie en (1)' (non corrélation), la statistique suivante :

(5) $R_N = (1 - r_{XY}^2)^{-1/2} (N - 2)^{1/2} . r_{XY}$

suit une **loi de STUDENT** $\mathcal{S}_{N-2}$ à N - 2 **degrés de liberté** ;

(b) dans le cas bilatéral H_1' défini en (2)', il n'existe pas de test uniformément le plus puissant, mais l'on peut (sous l'hypothèse H_0) adopter la région critique suivante :

(5)' $w = [|r_{XY}| \geq q_{1 - (\alpha / 2)}]$,

dans laquelle $q_{1 - (\alpha / 2)}$ est le quantile d'ordre $1 - (\alpha / 2)$ de la loi de r_{XY} (pour une région de confiance symétrique). On peut encore utiliser la loi $\mathcal{S}_{N-2}$.

Lorsque N est grand (N >> 0) (eg $N \geq 25$), on utilise l'**approximation** normale de cette loi.

(iii) Dans une **situation statistique** non gaussienne, on peut encore procéder à un test de façon analogue à ce qui précède. Sous l'hypothèse H_0' : $\rho_{\xi\eta} = 0$ (de non corrélation) définie en (1)', on peut en effet établir la **propriété asymptotique** :

(7) $\mathcal{L}(r_{XY}) \rightarrow^{H0'}{}_{N \rightarrow +\infty} \mathcal{N}(0, N^{-1/2})$ (**loi normale** centrée).

ou, plus rigoureusement, $\mathcal{L}(N^{1/2} . r_{XY}) \rightarrow_{N \rightarrow +\infty} \mathcal{N}_1(0, 1)$ (propriété vraie si H_0' l'est).

Si le test concerne une valeur ρ_0 donnée de $\rho_{\xi\eta}$, ie l'hypothèse de base :

(1)'' H_0'' : $\rho_{\xi\eta} < \rho_0$, contre l'alternative :

(2)" H_1" : $\rho_{\xi\eta} \geq \rho_0$,

on montre qu'il existe un test uniformément le plus puissant pour tester H_0".

En pratique, on définit la **statistique** suivante :

(8) $r_N = \text{Arg th } r_{XY} = (1/2) \text{ Log } \{(1 - r_{XY})^{-1} (1 + r_{XY})\}$ (ssi $|r_{XY}| < 1$),

ie $r_N = 2^{-1} \{\text{Log } (1 + r_{XY}) - \text{Log } (1 - r_{XY})\}$, et l'on utilise la propriété asymptotique suivante :

(9) $\mathscr{L}(r_N) \to^{H0"}{}_{N \to +\infty} \mathscr{N}(\rho_{\xi\eta}, (N - 3)^{-1/2})$ (**loi normale**),

ou, plus rigoureusement, $\mathscr{L}\{(N - 3)^{1/2} . (r_{XY} - \rho_{\xi\eta})\} \to_{N \to +\infty} \mathscr{N}_1(0, 1)$ (propriété vraie si H_0" l'est).