STATISTIQUE CONCOMITANTE (F1, F6)
(08 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **statistique concomitante** est un **vecteur aléatoire** dont les coordonnées sont disposées comme celles de la statistique d'ordre associée à un échantillon donné.

Soit $(\Omega, \mathcal{T}, \mathcal{P})$ un **modèle statistique**, $\zeta = (\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **couple aléatoire** réel de loi P^ζ et $Z = ((X_1 , Y_1) ,..., (X_N , Y_N))$ un N-**échantillon** issu de ζ.

On pose $X = (X_1 ,..., X_N)$ et $Y = (Y_1 ,..., Y_N)$, et l'on note $X^{(.)} = (X^{(1)} ,..., X^{(N)})$ la **statistique ordonnée** associée à X et $R = (R_1 ,..., R_N)$ sa **statistique de rang**, avec :

(1) $\quad X^{(R(n))} = X_n , \qquad \forall\ n \in \{1 ,..., N\}$ (en notant R(n) pour désigner R_n).

(i) Soit alors $\sigma \in \boldsymbol{\sigma}_N$ la **permutation** des coordonnées de X correspondant à $X^{(.)}$, ie :

(2) $\quad X^{(n)} = X_{\sigma (n)} , \qquad \forall\ n \in \{1 ,..., N\}.$

On appelle **statistique concomitante**, ou **échantillon concomitant**, de $X^{(.)}$ (donc de X) la statistique $Y^{\sim}$ définie par :

(3) $\quad Y^{\sim} = Y_{\sigma (n)} , \qquad \forall\ n \in \{1 ,..., N\}.$

(ii) Si l'on pose $\sigma (z) = (z_{\sigma(1)} ,..., z_{\sigma(N)})$, $\forall\ z \in \mathbf{R}^N$, on a :

(4)
$$X^{(.)} = (X_{\sigma (1)} ,..., X_{\sigma (N)}) = \sigma (X),$$
$$Y^{\sim} = (Y_{\sigma (1)} ,..., Y_{\sigma (N)}) = \sigma (Y).$$

(iii) Lorsque X comporte des valeurs (ou « coordonnées ») multiples (ie égales entre elles), on identifie généralement les permutations correspondantes (cf aussi **rang multiple**).

(iv) La notion de statistique concomitante intervient souvent en **Statistique non paramétrique**. La loi P^ζ du couple (ξ, η) étant donnée, on peut définir et étudier celle du couple $(X^{(.)}, Y^{\sim})$, ou encore rechercher l'existence d'une **régression négative** (resp positive) entre $X^{(.)}$ et $Y^{\sim}$.

Elle intervient aussi dans d'autres circonstances, telle l'étude de la **courbe de LORENZ** généralisée (empirique). En effet, si $(X_n , Y_n)_{n=1,...,N}$ est une **suite** d'**observations** d'un **couple aléatoire** (à coordonnées non négatives) $\zeta = (\xi, \eta) : \Omega \mapsto \mathbf{R}_+^2$, la **courbe de GINI empirique** est définie par les couples $(u_n , v_n)_{n=1,...,N}$ selon les relations :

$$(5)\qquad \begin{cases} u_n = \Sigma_{\alpha=1}^{n} X^{(\alpha)} / \Sigma_{\alpha=1}^{N} X^{(\alpha)} \\ v_n = \Sigma_{\alpha=1}^{n} Y^{\sim}_{\alpha} / \Sigma_{\alpha=1}^{N} Y^{\sim}_{\alpha} \end{cases} \qquad \forall\ n = 1 ,..., N,$$

dans lesquelles $Y^{\sim}$ désigne la concomitante de $X^{(.)}$.