STATISTIQUE SYSTÉMATIQUE (G1)
(23 / 10 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) Soit ($\mathbf{R}^N$, $\mathscr{B}(\mathbf{R}^N)$, $\mathscr{P}^X$) le **modèle image** d'un modèle fondamental (Ω, $\mathscr{T}$, $\mathscr{P}$) par un **vecteur aléatoire** (**échantillon**) $X = (X_1 ,..., X_N)$, $X^{(.)} = (X^{(1)} ,..., X^{(N)})$ l'échantillon ordonné associé à X (cf **statistique ordonnée**) et ($\mathscr{Y}$, $\mathscr{C}$) un **espace mesurable** auxiliaire (en général, $\mathscr{Y} = \mathbf{R}^Q$).

On appelle **statistique systématique**, ou même **statistique ordonnée**, ou parfois **statistique méthodique**, toute **statistique** S définie par une **application mesurable** s : $\mathbf{R}^N_{\leq} \mapsto \mathscr{Y}$, dans laquelle $\mathbf{R}^N_{\leq} = \{x \in \mathbf{R}^N : x_1 \leq ... \leq x_N\}$.

Autrement dit, la statistique :

(1) $\quad S = s\,(X^{(.)}) \ : \ \Omega \mapsto \mathscr{Y}$

définit une statistique systématique.

Si l'on note $\sigma\,(X) = X^{(.)}$ la statistique ordonnée précédente (où $\sigma \in \boldsymbol{\sigma}_N$ est la **permutation** des coordonnées de X correspondante), on peut aussi définir la **statistique systématique** s (resp S) selon l'expression symbolique :

(1)' $\quad s \ = \ s\,(\sigma\,(X)) \ = \ s \circ \sigma\,(X)$.

(ii) Lorsque $\mathscr{Y} = \mathbf{R}^Q$ et que s est une **application linéaire**, dont la **matrice** est encore notée s, on dit que la **statistique vectorielle** (produit matriciel) :

(2) $\quad S \ = \ s\,X^{(.)}$,

dans laquelle $X^{(.)}$ est écrit sous forme de matrice colonne, est une **statistique systématique linéaire** (F. MOSTELLER).

(iii) A titre d'exemples :

(a) une statistique systématique élémentaire est celle dans laquelle $\mathscr{Y} = \mathbf{R}^N$ (ie Q = N) et s est la **restriction** de l'**application identique** id ($\mathbf{R}^N$) de $\mathbf{R}^N$ à $\mathbf{R}^N_{\leq}$. Alors s : $X^{(.)} \mapsto X^{(.)}$ est une statistique systématique ;

(b) si Q = 1 dans (2) (s est alors une **forme linéaire** sur $\mathbf{R}^N$) et si certains coefficients de s sont nuls, on peut définir diverses statistiques systématiques : eg moyenne tronquée (cf **troncature**), **moyenne censurée** (cf **censure**), **moyenne équilibrée** ou transformée de WINSOR (cf **transformation de WINSOR**).

Ainsi, lorsque $2 \leq L < M \leq N - 1$ et que $s_1 = ... = s_{L-1} = s_{M+1} = ... = s_N = 0$, on obtient la **moyenne empirique** censurée (cf **moyenne censurée**) :

(3) $\overline{X}_N (L, M) = (M - L + 1)^{-1} (X^{(L)} + ... + X^{(M)})$,

avec $s_L = ... = s_M = (M - L + 1)^{-1}$;

(c) enfin, les statistiques qui dépendent des « extrêmes » d'un échantillon (cf **valeur extrême**, **statistique des extrêmes**) sont aussi des statistiques systématiques.

(iv) Les statistiques systématiques interviennent souvent en **Statistique non paramétrique** (tests ou estimations diverses). Leur qualificatif de « systématique » n'est guère approprié : il vaudrait mieux les appeler eg **statistiques de F. MOSTELLER**.