

PROCÉDURE DE TUKEY (F1, F8)

(07 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

On appelle parfois **procédure de J.W. TUKEY** une **méthode d' « écrêtage »**, ou **méthode de « pelage »**, d'un ensemble de **données** qui est basée sur une notion de **censure** étendue à des vecteurs aléatoires ou à des **observations** vectorielles.

(i) Soit $X = (X_1, \dots, X_N)$ un **échantillon aléatoire** constitué de N **vecteurs aléatoires** $X_n : \Omega \mapsto \mathbf{R}^K$. On associe à $X^{(0)} = X$ l'**ensemble** $E^{(0)}$ des N points correspondants dans l'espace \mathbf{R}^K (cf **nuage de points**).

La méthode consiste :

(a) à éliminer les points extrêmes (cf **valeur extrême**) de l'**enveloppe convexe** $\text{Co } E^{(0)}$ de $E^{(0)}$, d'où un ensemble de points $E^{(1)}$;

(b) on élimine, à leur tour, les points extrêmes de l'enveloppe convexe $\text{Co } E^{(1)}$;

(c) etc.

A l'étape p , on aboutit à un ensemble $E^{(p)}$ tq (selon le choix retenu au préalable) :

(a) soit p est un entier fixé d'avance ;

(b) soit la différence $\text{Card } E^{(0)} - \text{Card } E^{(p)} = L$ est fixée d'avance, avec $L < N$ (**censure de type I**) ;

(c) soit :

$$(1) \quad (\text{Card } E^{(0)} - \text{Card } E^{(p)}) / \text{Card } E^{(0)} = [\alpha \cdot N],$$

où $\alpha \in]0, 1[$ désigne la **proportion** de points à éliminer, fixée d'avance (**censure de type II**), et $[.]$ désigne la fonction **partie entière**.

(ii) Si l'on note $X^{(p)} = \{X_{n(1)}, \dots, X_{n(p)}\}$ l'échantillon final obtenu par cette procédure, on peut, notamment en théorie de la **robustesse**, calculer des **statistiques** tq $S_{N,p} = s_{N,p}(X^{(p)})$ au lieu des statistiques « brutes » tq $S_N = s_N(X)$ habituellement utilisées lorsqu'on n'élimine pas les points aberrants (cf **aberration**, **courbe d'influence**).