

RANG MOYEN (F6)

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

La notion de **rang moyen** se rapporte au « rangement » d'un **ensemble d'observations** dont plusieurs d'entre elles sont égales (ou identiques, ou encore non distinguables).

(i) Soit $(\Omega, \mathcal{F}, P)$ un **espace probabilisé**, $(\mathcal{X}, \leq)$ un ensemble ordonné (cf **relation d'ordre**), $\xi : \Omega \mapsto \mathcal{X}$ une **va** et $X = (X_1, \dots, X_N)$ un **échantillon aléatoire** constitué de $N \geq 2$ copies, indépendantes entre elles, de la **variable parente** ξ . Soit $X^{(.)} = (X^{(1)}, \dots, X^{(N)})$ l'échantillon ordonné associé à X (cf **statistique ordonnée**).

On suppose qu'il existe deux entiers L et M (avec $1 \leq L < L + M \leq N$) tq X comporte $M = (L + M) - L$ valeurs (ie coordonnées) multiples (ie égales) :

$$(1) \quad X^{(L)} < X^{(L+1)} = \dots = X^{(L+M)} < X^{(L+M+1)}.$$

On appelle **rang moyen**, associé aux valeurs multiples précédentes, la valeur (non nécessairement entière) r définie par la moyenne des rangs des coordonnées identiques concernées :

$$(2) \quad r = M^{-1} \cdot \{(L+1) + \dots + (L+M)\} = M \cdot \{L + (1/2)(M + 1)\}.$$

(ii) En **Statistique non paramétrique**, on attribue parfois aux M valeurs multiples $X^{(n)}$ (avec $L+1 \leq n \leq L+M$) le rang moyen r défini en (2) (**méthode des rangs moyens**). Cette méthode intervient dans la **correction** de certaines **statistiques**, ainsi « modifiées », afin de traiter des observations comportent des valeurs multiples (« multiplets »).

(iii) Dans une suite tq X peuvent exister plusieurs « paquets » de coordonnées identiques. Le traitement statistique est analogue au précédent.

(iv) Plus généralement, la suite X peut être quelconque, ie ni équidistribuée, ni indépendante.