COEFFICIENT DE HERFINDAHL - HIRSCHMANN (F3)
(23 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **coefficient de HERFINDAHL - HIRSCHMANN** est un **coefficient de concentration** ou, alternativement, un **coefficient** de **dispersion**, d'une **variable numérique** positive (ou non négative) observable dans un **ensemble** d'**unités statistiques**.

Il permet donc d'apprécier le **degré de regroupement**, ou le **degré de dissémination**, de ces unités relativement au **critère** représenté par la variable considérée, eg magnitude d'un type donné répartie dans un ensemble :

(a) de corps célestes : masse ou dimension (physique) ;

(b) de bactéries : volume (biologie) ;

(c) d'animaux : niveau de santé (écologie) ;

(d) de personnes physiques : revenu (sociologie : économie).

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}_+$ une **vars** positive. On observe un N-**échantillon** $X = (X_1, ..., X_N)$ dont la **variable parente** est ξ et l'on pose $f = (f_1, ..., f_N)'$ (vecteur des **fréquences relatives**), avec $f_n = X_n / e_N' X$, où $e_N = (1, ..., 1)'$ est le premier vecteur bissecteur de $\mathbf{R}^N$. Les va f_n s'interprètent souvent comme des pourcentages.

On appelle **coefficient de O.C. HERFINDAHL - A.O. HIRSCHMANN** la **statistique** :

(1) $H_N^2(f) = \Sigma_{n=1}^N f_n^2.$

$H_N^2(f)$ se note aussi H_N^2 et la va $1 - H_N^2$ est aussi appelée **coefficient de A.O. HIRSHMANN**.

On montre que :

(2) $H_N^2 = N^{-1} . (C_N^2 + 1),$

où C_N^2 est le carré du **coefficient de variation** (empirique) $C_N = S_N / \overline{X}_N$, avec $\overline{X}_N = e_N' X / e_N' e_N$ (**moyenne empirique**) et $S_N^2 = X' P X / e_N' e_N$ (**variance empirique**), P étant la **matrice de centrage par rapport à la moyenne** $\overline{X}_N$.

Par suite :

(a) si N = 1, alors $f_1 = 1$ et $H_N^2 = 1$ (concentration absolue) ;

(b) si $f_n = 1 / N$, $\forall n \in \mathbf{N}_N^*$ (distribution uniforme) (cf **loi uniforme discrète**), alors $H_N^2 = 1 / N$;

(c) $N^{-1} \leq H_N^2 \leq 1$, d'où (lorsque $N \to +\infty$) $0 \leq H_\infty^2 \leq 1$.

(ii) La formule (1) se généralise selon :

(3) $\quad H_N^\alpha (f)$ ou $H_N^\alpha = \Sigma_{n=1}^N f_n^\alpha, \qquad \forall \alpha \in [1, +\infty[$,

avec $H_N^\infty (f) = \max_{n=1}^N f_n$ lorsque $\alpha = +\infty$.

(iii) Si $X^{(.)} = (X^{(1)} ,..., X^{(N)})$ désigne la la **statistique ordonnée** associée à X en ordre croissant (ie tq $X^{(n)} \leq X^{(n+1)}, \forall n \in N_{N-1}^*$) et si l'on pose :

(4) $\quad f^{(n)} = X^{(n)} / e_N' X, \quad \forall n \in N_N^* \quad$ et $\quad f^{(.)} = (f^{(1)} ,..., f^{(N)})$,

on peut définir, $\forall n \in N_N^*$, le poids des n plus grandes observations selon :

(5) $\quad C_n = \Sigma_{\alpha=1}^n f^{(\alpha)}$.

En particulier, $C_1 = H_N^\infty (f)$.