ESTIMATEUR DE HORVITZ - THOMPSON (H, M3, M4)
(16 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**estimateur de HORVITZ - THOMPSON** est une notion importante généralement associée à un **plan de sondage**. La notion est présentée dans le cas d'un **sondage simple**.

(i) Soit Ω une **population** finie $\{\omega_1 ,..., \omega_M\}$ et Π un plan **sondage sans remise** dans Ω. On suppose que la **variable** étudiée est un **vecteur aléatoire** $\eta : \Omega \mapsto \mathbf{R}^G$.

Soit $A \in \mathcal{T}_N$ un **échantillon sans remise** de taille N. On note :

(a) $Y = (Y_1 ,..., Y_M)$, avec $Y_m = \eta\ (\omega_m)$, $\forall\ m \in N_M^*$, les **observations** (ou **mesures**) effectuées sur Ω ;

(b) $y = (y_1 ,..., y_N)$, avec $y_n = \eta\ (a_n)$, $\forall\ n \in N_N^*$ tq $a_n \in A$, les observations analogues effectuées sur A ;

(c) $T = \Sigma_{m=1}^M Y_m = e_M' Y$ le **total** inconnu à estimer, défini sur Ω.

L'**estimateur de D.G. HORVITZ - D.J. THOMPSON** de T est défini comme l'estimateur linéaire (en y) suivant :

(1) $T_N = t_N (y) = \Sigma_{n=1}^N \alpha_n . y_n = \Sigma_{m=1}^M \alpha_m . Y_m . \mathbf{1}_A (\omega_m)$

(où $\mathbf{1}_A$ est une **variable de CORNFIELD**). On suppose aussi que T_N est tq :

(2) $\alpha_m = \pi_m = \Pi\ (A \text{ f } \omega_m)$,

où π_m désigne la **probabilité d'inclusion** de l'unité ω_m dans A (ie la probabilité que l'échantillon aléatoire A contienne l'unité donnée ω_m), ie :

(3) $\pi_m = \Pi\ (A \ni \omega_m)$,

où le symbole $\ni$ est un sybole de contenance : $A \ni \omega_m \Leftrightarrow \omega_m \in A$.

(ii) On appelle **sondage de D.G. HORVITZ - D.J. THOMPSON** le **plan de sondage** défini par le couple (Π, t_N) précédent.

(iii) Par construction, T_N est à la fois un estimateur linéaire et un **estimateur sans biais** de T. Sa **matrice des covariances** s'explicite selon :

(4) $V\ T_N = \Sigma_{m=1}^M\ \pi_m^{-1} (1 - \pi_m)\ Y_m\ Y'_m + \Sigma_{l=1}^M \Sigma_{m=1\ (m \neq l)}^M (\pi_l\ \pi_m)^{-1} (\pi_{lm} - \pi_l\ \pi_m)\ Y_l\ Y'_m$,

où :

(5) $\pi_{lm} = \Pi \{(\omega_l , \omega_m) \in A^2\}$

est la probabilité d'inclusion des unités données ω_l et ω_m ($m \neq l$) dans l'échantillon A.

Cette **dispersion** V T_N est estimée sans biais par :

(6) $v(T_N) = (1/2) \Sigma_{\alpha=1}^{N} \Sigma_{\beta=1}^{N}{}_{(\beta \neq \alpha)} \; \pi_{\alpha\beta}^{-1} (\pi_\alpha \pi_\beta - \pi_{\alpha\beta}) D_{\alpha\beta} D_{\alpha\beta}'$,

expression dans laquelle $D_{\alpha\beta} = (y_\alpha / \pi_\alpha) - (y_\beta / \pi_\beta)$ ($\forall \beta \neq \alpha$).