ESTIMATEUR PAR DIFFÉRENCE (H, M)
(27 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) En **théorie des sondages**, on considère une **population** finie Ω (Card Ω = M) sur les éléments ω_m desquels on peut observer des « **caractères** » η et ξ. Ces derniers définissent resp les vecteurs $Y = (Y_1, ..., Y_M)'$ et $X = (X_1, ..., X_M)'$ dans la population, avec $Y_m = \eta(\omega_m)$ et $X_m = \xi(\omega_m)$, pour tout $\omega_m \in \Omega$, m = 1 ,..., M.

Soit $A = \{a_1, ..., a_N\}$ un N-**échantillon** aléatoire constitué d'éléments (ou **événements** élémentaires) a_n extraits de Ω selon un **plan de sondage** Π donné. On peut alors observer des **mesures** ou **observations**, aléatoires puisque A l'est, de η et de ξ qui définissent resp les vecteurs $y = (y_1, ..., y_N)'$ et $x = (x_1, ..., x_N)'$ dans l'échantillon, avec $y_n = \eta(a_n)$ et $x_n = \xi(a_n)$, pour tout $a_n \in A$, n = 1 ,..., N.

On suppose que Y vérifie une hypothèse bayésienne de la forme (cf **superpopulation**) :

(1) $E\,Y = X + b_0\, e_M$, avec $V\,Y = \sigma^2 . I_M$.

Pour estimer eg le total :

(2) $T = e_M'\,Y = \Sigma_{m=1}^{M} Y_m$,

la méthode de prédiction bayésienne (cf **loi de prédiction bayésienne**) consiste ici à estimer b_0 par la **méthode des moindres carrés ordinaires**, ie selon :

(3) $b_0^\wedge = e_N'\,(y - x) / N$.

On appelle **estimateur par différence** de T l'estimateur :

(4) $T_N' = e_M'\,X + (M / N)\,(e_N'\,y - e_N'\,x)$.

L'estimateur par différence de la **moyenne** théorique $\overline{Y}_M = T / M$ s'en déduit selon :

(5) $T_N'' = T_N' / M = \overline{y}_N + (\overline{X}_M - \overline{x}_N)$,

où $\overline{y}_N$ (resp $\overline{x}_N$) est la **moyenne empirique** de η (resp de ξ) calculée avec l'échantillon y (resp x) et $\overline{X}_M$ est la moyenne théorique (supposée connue) du caractère ξ dans la population Ω.

(ii) L'**estimateur par régression** est généralement préféré au précédent.