ESTIMATEUR PAR RÉGRESSION (H, J1, M)

(i) En **théorie des sondages**, dans le même cadre que celui définissant l'**estimateur par différence**, on suppose que Y vérifie une hypothèse bayésienne de la forme (cf **superpopulation**) :

(1) $E\,Y = b_0 + b_1\,X$, avec $V\,Y = \sigma^2 \,.\, I_M$.

Pour estimer eg le total $T = e_M'\,Y$, la méthode de prédiction bayésienne revient ici à estimer $(b_0\,,\, b_1)$ à l'aide de l'**estimateur des moindres carrés ordinaires** $(b_0^{\wedge}, b_1^{\wedge})$ lié au modèle (1), ie à l'aide du modèle :

(2) $y = b_0 + b_1\,x + u$, avec $E\,u = 0$, $V\,u = \sigma^2 \,.\, I_N$.

On appelle alors **estimateur par régression** de T l'estimateur :

(3) $T_N' = M \,.\, \overline{y}_N + b_1 \,.\, (e_M'\,X + M\,\overline{x}_N)$,

où $\overline{x}_N$ (resp $\overline{y}_N$) est la **moyenne empirique** de ξ (resp de η) calculée à l'aide de l'**échantillon** (x, y). La formule (3) suppose donc X connu.

L'estimateur par régression de la **moyenne** $\overline{Y}_M = T / M$ (moyenne théorique de η dans la population Ω) se déduit de (3) selon :

(4) $T_N'' = M^{-1} \,.\, T_N' = \overline{y}_N + b_1 \,.\, (\,\overline{X}_M - \overline{x}_N\,)$,

où $\overline{X}_M = M^{-1} \,.\, e_M'\,X$ est la moyenne théorique de ξ dans la population Ω.

(ii) On montre que l'estimateur par régression est asymptotiquement sans biais (cf **biais asymptotique**).