ÉCHANTILLON DE NEYMAN (M1, M3)
(18 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

Dans un **sondage stratifié**, l'utilisation de **taux de sondage** variables d'une **strate** à l'autre permet une meilleure **précision** qu'avec un sondage où ces taux sont égaux (échantillon stratifié « représentatif »).

(i) Soit A un N-**échantillon** stratifié selon une **partition** $\Pi_A = \{A_1 ,..., A_H\}$.

On cherche à déterminer les tailles N_h = Card A_h des **échantillons de strate** A_h tq eg la **variance** de l'**estimateur** (moyenne pondérée d'ensemble) :

(1) $T_N'' = \Sigma_{h=1}^{H} (M_h / M) . \bar{y}_h$

soit minimum, la taille d'ensemble N = Card A étant fixée a priori.

On appelle alors **échantillon de J. NEYMAN** un N-échantillon A solution du problème de **programmation mathématique** contraint suivant :

(2) $\min V T_N''$ sous $\Sigma_h N_h = N$.

(ii) La méthode des **multiplicateurs de LAGRANGE** conduit à des conditions du premier ordre dont on déduit :

(3) $f_h = N S_h / \Sigma_h M_h S_h , \quad \forall h \in N_H^*$.

Autrement dit, les taux de sondage par strate $f_h = N_h / M_h$ doivent être proportionnels aux **écarts-types** S_h des strates h : une allocation de A par strates vérifiant cette propriété est appelée **allocation de NEYMAN**.

(iii) En pratique, les S_h ne sont pas connus (sauf si le sondage est un recensement) et doivent être estimés (cf aussi **sondage post-stratifié**).