

ÉCHANTILLON REPRÉSENTATIF (M3)

(18 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

La notion de **représentativité d'un échantillon**, relativement à une population qu'il sert à « connaître », constitue un problème important de la **théorie des sondages**.

(i) Soit $A = \{a_1, \dots, a_N\}$ un **échantillon** extrait d'une **population** finie $\Omega = \{\omega_1, \dots, \omega_M\}$, partitionnée en **strates** selon $\Pi_\Omega = \{\Omega_1, \dots, \Omega_H\}$, avec $H \leq M$.

On dit que A est un **échantillon représentatif** ssi :

$$(1) \quad N_h / N = M_h / M, \quad \forall h \in N_H^*,$$

où $N_h = \text{Card } A_h$, $\forall h$, $\Pi_A = \{A_1, \dots, A_H\}$ est la **partition « trace »** de Π_Ω sur A , et $M_h = \text{Card } \Omega_h$.

Autrement dit, les **taux de sondage partiels** $F_h = N_h / M_h$ sont tous égaux au **taux de sondage global** $F = N / M$.

(ii) En pratique, la partition Π_Ω est déterminée par référence à (au moins) une **variable statistique** (eg un **caractère**) $\xi : \Omega \mapsto \mathcal{X}$ définie sur Ω , parfois appelée **variable de contrôle**, ou **variable de stratification**. Cette variable permet, en effet, de définir les strates Ω_h dans lesquelles seront tirées les échantillons A_h .

Dans certains cas, l'échantillon représentatif est aussi un **échantillon équilibré**.