

ÉCHANTILLONNAGE OPTIMUM SOUS CONTRAINTES FORTES

Antoine DUPONT (*), Alii (**)

(*) Insee, Direction de la méthodologie et de la coordination statistique et internationale

(**) Insee, Département de la Démographie

antoine.dupont@insee.fr

Mots-clés (6 maximum) : Échantillonnage, calage, collecte, équilibrage, multimode, séries temporelles.

Domaines : Voir nomenclature du site (1 ou 2 domaines ou sous-domaines).

Résumé (entre 1 et 2 pages)

Cet article reprend et enrichit une problématique déjà présentée aux JMS 2001¹. Il s'agit de constituer, au sein d'une population de référence, des classes présentant des conditions d'homogénéité ou d'hétérogénéité maximales vis-à-vis de certaines caractéristiques quantitatives.

Le projet Nautila va se substituer au projet Octopusse : mais la filiation est claire.

Il est possible en fait de généraliser cette notion d'inertie en introduisant une pseudo-distance d entre les unités de la population de référence, non nécessairement euclidienne, et non nécessairement nulle lorsqu'on mesure la distance d'une unité à elle-même. Sous réserve de donner des poids α_i à chaque unité i de la population, l'inertie d'une classe K sera donnée par :

$$I(K) = \frac{1}{2} \frac{\sum_{i \in K} \sum_{j \in K} \alpha_i \alpha_j d_{i,j}^2}{\sum_{i \in K} \alpha_i}$$

Cette formule redonne l'expression de la variance de la classe lorsque la distance considérée $d_{i,j}$ est euclidienne.

Dans le second groupe d'exemples, on cherchera à appliquer ces techniques à des cas non euclidiens, par exemple en comparant département de naissance et département de résidence.

1. « Un algorithme de regroupement d'unités statistiques selon certains critères de similitude », Marc Christine et Michel Isnard, VIIèmes Journées de Méthodologie statistique, 4-5 décembre 2000.

Bibliographie

[1] Elston Lafata J., Koch G.G., Weissert W.G., Estimating activity limitation in the non-institutionalized population: a method for small areas, American Journal of Public Health, vol 84, n° 11, pp 1813-1817, november 1994.

[2] Ératosthène, un algorithme de détermination du PGCD par le crible éponyme, IIIe siècle avant J.-C.

[3] Molkogorov A., un théorème central limite universel.