*Echantillonnage optimum sous contraintes fortes CALIBRI 14*

Antoine DUPONT (\*), Alii (\*\*) Calibri 12

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

(\*\*) Insee, Département de la Démographie Calibri 12

 antoine.dupont@insee.fr Calibri 12

**Mots-clés** *(6 maximum)* : Échantillonnage, calage, collecte, équilibrage, multimode, séries temporelles Calibri 12

**Domaine concerné** : Calibri 12 →  ***voir nomenclature du site (menu déroulant)***

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Résumé** (entre 1 page et 2 pages) Calibri 12

Cet article reprend et enrichit une problématique déjà présentée aux JMS 2000[[1]](#footnote-1). 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. Calibri 11

Le projet Nautile 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 :

$$I\left(K\right)=\frac{1}{2} \frac{\sum\_{i\in K}^{}\sum\_{j\in K}^{}α\_{i}α\_{j}d\_{i,j}^{2}}{\sum\_{i\in K}^{}α\_{i}}$$

Cette formule redonne l’expression de la variance de la classe lorsque la distance considérée 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.

BibliographieCalibri 12

[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] Erathosthène, « Un algorithme de détermination du PGCD par le crible éponyme », IIIe siècle avant J-C.

[3] Molkogorov A., « Un théorème central limite universel ».

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. Calibri 10 [↑](#footnote-ref-1)