

## Echantillonnage et Equilibrage

Guillaume Chauvet, Ensai/Crest – Vendredi 21 novembre de 9h à 12h30

Lorsque l'on souhaite produire des estimations sur une population, l'échantillonnage permet de réduire le volume de données à traiter. Utilisé dans les enquêtes, il permet de sélectionner aléatoirement un échantillon d'unités dans une base de sondage, unités qui seront ensuite enquêtées pour inférer sur la population d'étude. La méthode d'échantillonnage utilisée est généralement un compromis entre la recherche d'estimations précises, et la nécessité de respecter un budget imposé. L'échantillonnage peut également être utilisé en situation de données volumineuses, afin de se restreindre à un volume de données à traiter compatible avec ses ressources informatiques. Dans ce cas, le choix de la méthode d'échantillonnage repose également sur des contraintes techniques (algorithme séquentiel, possibilité de paralléliser).

Dans cet atelier, nous reviendrons sur les méthodes d'échantillonnage en population finie. Après avoir rappelé le principe d'inférence, nous ferons quelques rappels sur des méthodes classiques d'échantillonnage. Nous présenterons des algorithmes de tirage pour le sondage aléatoire simple, en discutant de leurs avantages/inconvénients dans le cas d'un volume de données important. Nous rappellerons ensuite les principes d'un échantillonnage équilibré, et le lien entre l'équilibrage et la précision des estimateurs. Nous présenterons la méthode du Cube, ainsi que des exemples d'application de l'échantillonnage équilibré.

### Plan de l'atelier

- 1) Rappels sur les méthodes d'échantillonnage
  - a. Principes généraux
  - b. Sondage aléatoire simple : principes et algorithmes d'échantillonnage
  - c. Sondage stratifié
  - d. Sondage à probabilités inégales
- 2) Échantillonnage équilibré
  - a. Principe
  - b. La méthode du Cube
  - c. La méthode du pivot
  - d. Application à l'imputation équilibrée

### Bibliographie

- Ardilly, P. (2006), Les techniques de Sondage, Technip.
- Chauvet, G., Deville, J.-C., and Haziza, D. (2011). On balanced random imputation in surveys, *Biometrika*.
- Chauvet, G., Tillé, Y. (2006). A fast algorithm for balanced sampling. *Computational Statistics*, 21, 53-62.

**Ateliers de formation – 17 et 21 novembre 2014**



**8ème colloque francophone sur les sondages - Dijon**

- Deville, J-C. , Tillé, Y (2004), Efficient balanced sampling : the cube method, *Biometrika*, 91, 893-912.
- Tillé, Y. (2006), *Sampling Algorithms*, Springer-Verlag.