

Correction du sur-échantillonnage

Fatou SALL, CESP

Mars 2019

Considérons un sondage stratifié par zones et villes. Pour les besoins de l'étude, nous décidons de ne pas respecter toutes les contraintes de l'enquête, on choisit par exemple de distordre la proportion des individus de certaines zones. Une correction doit être effectuée sur les poids des individus afin d'annuler cette distorsion.

Soient

K le nombre de zones et H le nombre de villes d'un pays donné,
 n_{hk} : le nombre d'individus de la ville h dans la zone k dans l'échantillon,
 $n_{.k}$: le nombre d'individus de la zone k dans l'échantillon,
 $d_{hk} = \frac{N_{hk}}{n_{hk}}$: le poids de la ville h dans la zone k .
Le poids de la zone k est défini par l'enquête comme suit:

$$d_k = \sum_h d_{hk} = \sum_h \frac{N_{hk}}{n_{hk}} \quad (1)$$

Pour calculer le coefficient de correction de la distorsion, on suppose qu'un sondage stratifié proportionnel aurait été fait dans chaque zone et le poids aurait été calculé comme suit:

$$w_k = \frac{N_{.k}}{n_{.k}} = \frac{\sum_h N_{hk}}{\sum_h n_{hk}} \quad (2)$$

Le coefficient de correction de la zone k est défini comme suit:

$$Correction = \frac{d_k}{w_k} = \frac{\sum_h \frac{N_{hk}}{n_{hk}}}{\sum_h \frac{N_{hk}}{n_{hk}}} \quad (3)$$

En appliquant ce coefficient de correction par zone au jeu de poids, on neutralise l'effet cette distorsion liée au sur-échantillonnage.