

LA VÉRIFICATION DES DONNÉES ("DATA EDITING") PRINCIPES, MISE EN OEUVRE

Philippe Brion

La vérification des données ("data editing") recouvre les activités de détection et de correction des erreurs qui sont menées dans le cadre de la production de statistiques. Bien que ces activités nécessitent beaucoup de ressources pour un institut statistique, il semble qu'elles aient fait l'objet d'assez peu d'articles en français (en particulier si l'on compare aux problèmes posés par le traitement des non-réponses), alors qu'elles ont connu un certain nombre d'évolutions fortes durant les trente dernières années.

L'objectif de cet atelier est donc de poser, d'abord, le problème, et dresser ensuite un panorama des différentes méthodes utilisées à l'heure actuelle.

Plan de l'atelier

1. Principes:

- (a) la nécessité de contrôler les données collectées: types de contrôles (logiques, fourchettes, évolutions, ...), variables liées
- (b) le lien entre contrôles et redressements
- (c) l'impossibilité de "tout nettoyer", les contraintes budgétaires
- (d) micro-contrôles et macro-contrôles.

2. Les différentes méthodes:

- (a) les méthodes de contrôle automatique ("automatic editing"), en particulier présentation rapide de la méthode de Fellegi Holt
- (b) la vérification sélective ("selective editing"): l'utilisation de scores relatifs à chaque unité pour mesurer l'impact d'une donnée jugée suspecte sur une statistique cible, les scores locaux et globaux
- (c) la vérification en sortie ("output editing")
- (d) la mise en place d'une stratégie globale articulant les différentes méthodes
- (e) l'importance des métadonnées, en particulier dans le cadre d'un processus d'amélioration en continu du dispositif de contrôle redressements.

Bibliographie

- [1] ISTAT, CBS and SFSO (2007). Recommended practices for editing and imputation in cross-sectional business surveys, manuel mis au point dans le cadre d'un projet d'Eurostat.
- [2] De Waal, T., Pannekoek, J., Scholtus, S. (2011). Handbook of statistical data editing and imputation, Wiley.
- [3] Documentation relative aux "work sessions" des Nations unies consacrées au data editing, disponible à l'adresse
<http://www1.unece.org/stat/platform/display/kbase/UNECE+Work+Sessions+on+Statistical+Data+Editing>