

## 5.1 - Présentation générale des redressements

La phase de redressements permet d'analyser les données brutes et de les corriger lorsqu'elles sont incorrectes. En particulier, elle doit permettre de redresser les non-réponses totales ou partielles, et de corriger les incohérences entre variables. Par exemple, un logement d'une seule pièce mais de 150 m<sup>2</sup>. Ces traitements sont entièrement automatiques, et sont réalisés en batch, lot par lot.

Cette phase comprend plusieurs étapes :

- 1 – Batch R1 : Redressement des premières variables de l'exploitation principales + sélection du donneur pour chaque FLNE.
- 2 – Batch R2A : Redressement des autres variables de l'exploitation principale (hors Analyse Ménage-Famille).
- 3 – Batch R2B : Redressement des variables de l'AMF
- 4 – Batch R3 : Redressement de l'exploitation complémentaire (en lien avec activité, profession).

Les étapes 1, 2 et 3 s'enchaînent pour chaque lot de traitement (il y a environ 30 LT chaque année), juste après la phase de codification automatique (Cf fiche 4.1) et en parallèle à la reprise des chiffrements (Cf fiche 4.3). La partie « sélection du donneur pour les FLNE » de l'étape 1 est détaillée dans la fiche 5.3.

Suite au passage du batch R1, plus spécifiquement du choix des donneurs pour les FLNE, le nombre d'individus d'un lot est fixé. Donc lorsque l'étape 1 est passée sur tous les lots de traitement de l'année, on dispose des données suffisantes pour commencer le calcul des populations légales.

Pour l'étape 4, on a besoin des données issues de la reprise manuelle (faite dans l'application Recap), des variables non codées automatiquement. Il faut donc attendre que tous les BI d'un lot de traitement soient traités pour lancer le batch R3 sur ce lot.

### La méthode de redressement utilisée :

La méthodologie de redressements retenue est de nature séquentielle. On traite les logements et les individus les uns après les autres et on traite les variables les unes après les autres dans un ordre pré-établi (pour l'ordre des variables voir la fiche 5.2). L'ordre de passage permet de s'assurer qu'une variable utilisée pour le redressement d'une autre variable a été redressée auparavant.

Pour l'ensemble des redressements de la non-réponse, la méthode la plus utilisée est celle du **Hot Deck**.

Dans un premier temps, les fichiers du lot de traitement sont triés selon la clé primaire (dep\_code, com\_code, iris\_ilot, adr\_ang, log\_rang pour le fichier logement, et on ajoute le tri sur ind\_rang pour le fichier individu). Pour redresser une valeur manquante, on impute par le plus proche voisin (au sens du tri de fichier), déjà redressé, ayant des caractéristiques proches et pour lequel la variable à redresser est remplie.

Lorsque le donneur est trouvé, la variable est alors affectée de la valeur trouvée.

Pour le **redressement des incohérences** ou lorsque qu'aucun donneur n'a pu être trouvé, on utilise alors une imputation déterministe, c'est-à-dire pré-définie en fonction des valeurs rencontrées.

### La présentation des redressements :

Le détail du redressement des variables sont présentés dans les fiches 5.4 par ordre alphabétique des variables. Ces fiches présentent successivement :

- 1 - Le champ des unités statistiques à traiter pour ce redressement.
- 2 - La variable à redresser (signification, modalités, ..), ainsi que les variables auxiliaires utilisées dans ce redressement.
- 3 – La définition de toutes les incohérences associées à la variable, présentées sous forme de régions d'incohérences.
- 4 - Les traitements à réaliser pour chaque région d'incohérence.