



# Document décrivant les méthodes de vérification par échantillonnage et extrapolation

Ce descriptif technique de la mise en œuvre de méthodes de vérification par échantillonnage et extrapolation est accompagné de l'arrêté du 11 avril 2007 définissant lesdites méthodes, pris en application de l'article R 243-59-2 du code de la Sécurité Sociale.

Conformément à ces dispositions, ces documents vous sont remis préalablement à la mise en œuvre des méthodes qui ne pourra être effective avant un délai de quinze jours à partir de la date de réception.

Dans ce délai de quinze jours, vous avez la possibilité de vous opposer à leur utilisation, en informant par écrit, l'inspecteur du Recouvrement.

## **1 Le recours aux méthodes de vérification par échantillonnage et extrapolation**

### ***Recours aux méthodes***

Lorsque la vérification doit porter sur des données matérielles (pièces comptables, justificatifs de frais, bulletins de paie...) très nombreuses et dispersées, une investigation et un chiffrage exhaustifs peuvent s'avérer longs et fastidieux ou entraîner de possibles erreurs de mesure avec des contraintes partagées pour l'employeur et l'inspecteur.

En conséquence, dès lors que les premiers constats de l'inspecteur du recouvrement semblent mettre en évidence la détection d'erreurs ou d'anomalies, la vérification et le chiffrage sur la base d'un échantillonnage et d'une extrapolation, permettent l'application d'une méthode fiable et rigoureuse. Cette méthode présente plusieurs avantages :

- Elle est adaptée à la réalité de l'entreprise,
- Elle participe d'une meilleure compréhension des opérations du contrôle,

Elle est également strictement encadrée :

- Les principes du contrôle sont préalablement exposés,
- Ses modalités d'application sont par ailleurs clairement définies,
- L'employeur, enfin, participe à l'ensemble du processus dans le cadre d'une procédure contradictoire.

### ***Analyse des modalités d'accès à l'information***

L'inspecteur s'appuie sur les échanges avec l'employeur pour identifier :

- la nature des informations disponibles,
- les supports contenant les informations recherchées,
- le mode d'archivage de ces supports,
- l'existence éventuelle d'un état récapitulatif, sur lequel sont isolées les informations recherchées.

L'inspecteur procède à l'analyse des modalités d'accès à l'information et recherchera prioritairement avec l'employeur s'il existe une possibilité de traitement automatisé des données disponibles, afin de chiffrer la régularisation sur la base d'une vérification exhaustive.

## **2 Mise en œuvre de la méthode**

Avant d'être mise en application, la démarche de vérification propre à l'entreprise devra être décrite et formalisée. Ce descriptif sera adapté à la situation contrôlée et intégrera toutes les données de l'entreprise et les différents choix effectués relatifs à la population, le tirage de l'échantillon et le mode d'extrapolation.

Ce document est remis à l'employeur qui pourra le cas échéant effectuer des observations. Les éventuelles modifications liées à ces échanges seront également formalisées.

### ***2\_1 Détermination de l'individu statistique***

En fonction de la nature de l'irrégularité détectée, l'inspecteur identifie les catégories de salariés concernées. S'il dénombre plusieurs risques concernant sensiblement la même population de salariés, il ne sera effectué qu'un seul sondage de sorte que tous les chefs de redressements seront étudiés sur le même échantillon.

Suivant le mode de classement des pièces justificatives nécessaires au chiffrage, il détermine ensuite les individus statistiques qui composeront la base de sondage (ou population).

Pour chiffrer les régularisations, l'inspecteur travaille normalement salarié par salarié sur une période déterminée (année, mois...). En fonction du mode de conservation des justificatifs, la méthode, plus détaillée, consistant à retenir l'individu sur une période d'un mois (« salarié-mois ») sera préférée à celle utilisant l'individu sur une période annuelle.

Le recours au sondage, facteur de réduction du temps consacré à l'évaluation du montant des irrégularités, sera alors possible dans toutes les entreprises quelle que soit leur taille (la population étudiée doit néanmoins comporter au minimum 50 individus statistiques).

### ***2\_2 Constitution de la base d'échantillonnage***

L'individu statistique est caractérisé par un certain nombre de variables. Ces dernières sont appelées variables auxiliaires. Elles peuvent être qualitatives ou quantitatives.

Ainsi, à titre d'exemple, la masse de salaires, la masse de frais, la qualification du salarié, le secteur d'activité dans lequel il est affecté, le type de transport qu'il réalise, constituent des variables auxiliaires.

La population étudiée se compose de l'ensemble des individus faisant l'objet du contrôle de points particuliers de législation. Les individus non concernés par les points vérifiés ne seront pas retenus dans la base d'échantillonnage.

Par exemple, dans le cadre d'une entreprise de travail temporaire, cette base d'échantillonnage pourra se composer des salariés intérimaires ayant perçu des sommes exclues de l'assiette des cotisations. Seront alors exclus, les salariés administratifs, ainsi que les salariés intérimaires n'ayant pas perçu, sur la période contrôlée, des sommes non soumises à cotisations.

#### ***La stratification***

L'objectif premier de la stratification est d'améliorer la précision des estimations en renforçant l'homogénéité des ensembles à partir desquels l'échantillon est tiré. Elle est notamment nécessaire lorsque la population est vaste et hétérogène au regard des variables analysées.

En pratique, la population est scindée en sous-groupes homogènes par rapport à une ou plusieurs variables connues.

Les sous-groupes obtenus sont définis en tant que « strates ». C'est au sein de ces strates qu'un tirage aléatoire des individus qui composeront l'échantillon est effectué.

Il s'agit de constituer des groupes d'individus les plus semblables possible. Le recueil des données qualitatives et quantitatives établi dans les paragraphes précédents doit permettre de définir les variables les plus discriminantes, c'est-à-dire, celles qui permettent de réaliser au mieux des groupes homogènes d'individus, au regard du risque de redressement. Dans le cas d'une variable qualitative (exemple : le service), chaque strate sera constituée d'une des modalités de la variable (Direction, production, commercial, SAV...).

Dans le cas d'une variable comportant un trop grand nombre de modalités (exemple : qualifications professionnelles), il sera peut-être indispensable d'effectuer des regroupements logiques de celles-ci (exemple : personnel de bureau, soudeurs, maçons...).

La stratification peut être effectuée à plusieurs niveaux en utilisant différentes variables. Si une variable qualitative a été retenue, elle doit être utilisée en premier niveau de stratification. Le nombre de groupes sera déterminé en fonction du nombre de modalités de la variable. Ces groupes seront ensuite subdivisés en fonction de la deuxième variable retenue et ainsi de suite.

Exemple de stratification :

Type d'entreprise : ETT

1<sup>ère</sup> variable de stratification : le secteur d'activité (Industrie, Bâtiment...)

On connaît pour chaque salarié le secteur d'activité dans lequel il est affecté. On peut donc ranger ces individus dans des sous-groupes par type d'activité.

2<sup>ème</sup> variable de stratification : le ratio frais/ salaire

On a calculé pour chaque salarié son ratio frais / salaire \*. Dans chaque sous-groupe d'activité, on range les salariés dans de nouveaux sous-groupes en fonction de leur ratio frais / salaire.

*\* Précision : le ratio frais/salaire est une variable créée à partir des données existantes.*

*A titre d'exemple le ratio frais/salaire permet :*

*- d'apprécier la diversité de situations des salariés*

*- de neutraliser, en partie, l'hétérogénéité de la population relative au temps de travail passé dans l'entreprise ou au salaire (salaires et temps de travail étant assez fortement corrélés).*

### *Analyse et traitement des atypismes*

Tout au long du processus d'élaboration et d'observation de la base d'échantillonnage, notamment lors des échanges avec l'employeur, l'inspecteur peut mettre à jour des situations atypiques d'individus ou de groupes d'individus. C'est le cas de salariés qui, du fait de leur statut, de leurs conditions de travail ou de tout autre motif, présentent des caractéristiques très différentes des autres au regard du risque, de sorte que leur intégration à l'échantillon pourrait nuire à la qualité des résultats.

Si ces individus sont peu nombreux, ils devront être sortis de la base de sondage, et seront analysés séparément. De plus, ils ne devront être utilisés dans le cadre de l'extrapolation.

Dans l'hypothèse où le nombre des individus présentant des caractéristiques particulières n'est pas marginal, ces individus seront regroupés dans une strate à part entière qui viendra s'ajouter aux autres.

### *Nombre de strates*

En théorie, le nombre de strates doit être le plus grand possible. Plus ce nombre est élevé, plus il est facile de constituer des groupes homogènes.

En pratique, il peut être difficile d'en retenir un grand nombre, c'est donc en fonction de sa connaissance de l'entreprise et des échanges avec l'employeur que l'inspecteur décidera du nombre de strates qu'il devra constituer. Le nombre de strates sera donc fonction de la complexité de l'entreprise.

### *Bornage des strates*

Dans le cadre d'une stratification à partir d'une variable quantitative (montant des frais, ratio frais/salaire...), l'étude préalable des tendances et des distributions de la variable choisie est nécessaire pour réaliser le bornage des strates. Ces tendances, présentées sous forme de graphiques simples ou cumulés, permettent d'identifier des coupures et aident à décider du nombre de strates à effectuer.

Par ailleurs, afin d'optimiser le bornage des strates, il est également possible d'avoir recours à une application statistique dénommée « **nuées dynamiques** ».

Il s'agit d'une méthode automatique de partition d'une population aboutissant à minimiser la dispersion des classes ainsi formées. (voir formules jointes en annexe)

## **2\_3\_Tirage de l'échantillon**

Le tirage est effectué en fonction de la taille globale de l'échantillon et de la répartition des individus entre les strates.

### *La taille globale de l'échantillon*

Le nombre de strates est un indicateur du degré de complexité de l'entreprise, et donc de la dispersion de la variable d'intérêt. Il convient donc de déterminer un nombre total d'individus proportionnel au nombre de strates constitué dans l'entreprise.

Lorsque la base d'échantillonnage n'est pas stratifiée, on procède à un tirage aléatoire simple avec un nombre minimal d'individus de 50 (loi faible des grands nombres).

Selon cette loi faible des grands nombres, la taille globale minimale de l'échantillon serait équivalente à : 50 que l'on multiplie par le nombre de strates (en fonction de l'importance de la population ou de l'hétérogénéité de la variable étudiée, le nombre d'individus composant l'échantillon pourra être augmenté).

Pour certaines vérifications s'appuyant sur des bases d'échantillonnage pauvres en variables auxiliaires et rendant difficile, a priori, l'analyse de l'homogénéité des individus, l'inspecteur peut, dans un premier temps, procéder au tirage d'un pré-échantillon d'une taille assez importante duquel seront extraits les x (minimum 50) premiers individus pour constituer l'échantillon. En fonction des résultats obtenus après analyse de ces premiers individus, il pourra augmenter la taille de l'échantillon en puisant les individus suivants du pré-échantillon.

### *L'effectif des sous populations et la répartition des individus dans les strates*

Une fois la taille globale de l'échantillon déterminée, il importe de répartir de manière optimale les individus que l'on va sélectionner au sein de chaque strate. A ce titre, deux modes opératoires sont envisageables :

- Soit, on procède au calcul des proportions de chaque strate en nombre d'individus par rapport à la population de la base et on applique ces ratios au nombre d'individus de l'échantillon, il s'agit alors d'une répartition proportionnelle,
- Soit, on utilise la méthode de l'allocation de Neyman.

Cette dernière consiste à répartir les individus entre les différentes strates au prorata de la dispersion (écart type) de la variable auxiliaire utilisée (ex : montant des frais) et de la taille de chaque strate. (Voir formules jointe en annexe)

En effet la stratification permet de créer des groupes homogènes au regard de la variable fixée. Cela ne veut pas dire que les strates auront toutes la même taille et la même dispersion de leurs variables. A partir de ce constat, il serait contradictoire de prendre autant d'individus dans une strate très homogène et dans une autre qui le serait beaucoup moins.

Cette allocation aura donc pour effet de sur-représenter les strates à forte dispersion et/ou à effectif important. Cette variable ne sera pas obligatoirement la même que celle ayant servi à la stratification, mais sera en revanche une variable plus fortement corrélée avec la variable d'intérêt (montant du redressement).

Exemple : Si la variable frais/salaire a servi à effectuer la stratification, il est préférable d'utiliser la variable frais pour déterminer l'allocation de Neyman, car à l'intérieur d'une strate donnée les frais sont plus fortement corrélés avec le montant du redressement.

En tout état de cause, en prévision d'une éventuelle extrapolation par le ratio, chaque strate devra être représentée dans l'échantillon par au moins quinze individus.

### *Le tirage des individus*

Deux méthodes peuvent être retenues

#### - Tirage aléatoire simple

Il consiste à attribuer à chaque individu un nombre généré aléatoirement, de trier ensuite la population selon ce nombre, et de sélectionner les n premiers individus pour constituer un échantillon de taille n.

#### - Tirage aléatoire systématique ordonné

Dans chaque strate les individus sont classés dans un ordre croissant en fonction de la variable auxiliaire retenue (par exemple en fonction de leur montant de frais ou des sommes exclues).

Ils sont ensuite sélectionnés selon un tirage systématique de la façon suivante :

- un nombre appelé pas de tirage «p» est calculé en fonction des tailles de la population et de l'échantillon,
- puis, un individu est sélectionné aléatoirement parmi les «p» premiers individus du fichier.

Il s'agit bien d'une forme de tirage aléatoire, le premier individu étant sélectionné aléatoirement. (voir formule jointe en annexe)

Le tirage systématique ordonné est préconisé lorsque la variable auxiliaire est connue pour tous les individus de la population et qu'elle n'obéit pas à une forme de classement non maîtrisée par l'inspecteur. Son utilisation offre des garanties supplémentaires quant à la qualité de l'échantillon : tous les ordres de grandeurs de la variable utilisée seront ainsi balayés.

Les tirages sont effectués en présence de l'employeur qui peut donc, s'il le souhaite, participer à l'opération. Une copie de la liste des individus de l'échantillon est remise à l'employeur, un autre exemplaire est conservé par l'inspecteur. Ces deux listes sont paraphées pour attester de la validité de l'échantillon et de sa prise de connaissance par les deux parties.

## **2\_4\_Exploitation et optimisation de l'échantillon**

Sur la base de la liste d'individus composant l'échantillon, l'entreprise est invitée à produire les différents éléments et justificatifs nécessaires à la vérification de chacun des individus sur le ou les points de législation préalablement identifiés.

Une maquette «idéale» de dossier individuel aura préalablement été définie avec l'entreprise quant aux documents indispensables au contrôle. Les individus composant l'échantillon feront l'objet d'une vérification exhaustive.

L'inspecteur pourra, en accord avec l'entreprise, juger opportun d'effectuer ce contrôle notamment sur des dossiers réputés sensibles afin d'éviter des contestations ultérieures du chiffrage.

Dans le cadre d'un sondage aléatoire avec constitution d'un pré-échantillon, l'inspecteur, à l'issue de la vérification de l'échantillon, peut s'assurer de la précision des résultats obtenus en effectuant à l'aide de l'assistant sondages une simulation du chiffrage global du redressement. Il visualisera alors l'intervalle de confiance obtenu. Si cet intervalle de confiance est trop important, il pourra augmenter la taille de son échantillon en utilisant les individus en réserve dans le pré-échantillon.

## **2\_5 Présentation du détail des régularisations de l'échantillon à l'employeur.**

L'employeur est informé :

- des régularisations opérées par individu sur l'échantillon,
- des motifs de réintégration,
- des références des textes justifiant cette réintégration.

L'inspecteur, après avoir pris en compte les arguments développés par l'employeur, procédera, le cas échéant, à l'ajustement du chiffrage de son échantillon.

## **2\_6 Extrapolation à la population totale (voir les formules jointes en annexe)**

L'extrapolation des résultats déterminés lors de l'analyse des individus de l'échantillon peut s'effectuer :

- soit par le ratio
- soit par les nombres

### *Extrapolation par le ratio*

Le pourcentage de réintégration obtenu sur l'échantillon (montant de la variable d'intérêt / montant de la variable auxiliaire) est appliqué au montant total de la variable auxiliaire pour l'ensemble de la population dont est issu l'échantillon.

Exemple de calcul :

Population totale : 1 000 individus                      Echantillon : 50 individus  
Montant total des frais (population) : 5 000 000 euros (Variable auxiliaire)  
Montant du redressement sur l'échantillon : 2 500 euros (Variable d'intérêt)  
Montant des frais sur l'échantillon : 25 000 euros (Variable auxiliaire)

**Ratio de réintégration** = 2 500 / 25 000 = 10%

**Extrapolation à la population totale** : 5 000 000 \* 10 % = 500 000 euros

Lorsque la population a été préalablement stratifiée, la connaissance du montant global de la variable auxiliaire pour chaque strate avec une représentation dans l'échantillon d'un nombre d'individus supérieur ou égal à 15 de chaque strate, permettent à l'inspecteur d'utiliser l'extrapolation par le ratio séparé par strate (méthode permettant d'améliorer encore la précision du résultat).

Exemple de calcul avec le ratio séparé par strate :

**Strate 1**

Population : 1 000 individus  
Montant des frais : 2 000 000 euros  
Echantillon : 30 individus  
Montant frais moyen  
sur l'échantillon : 116,67 euros  
Montant de redressement  
moyen sur échantillon : 16,67 euros

**Ratio strate 1** :  $16,67/166,67 = 10\%$

**Extrapolation**

$2\,000\,000 * 10\% = 200\,000$  +  
soit au total : 650 000 euros

**Strate 2**

Population : 2 000 individus  
Montant des frais : 3 000 000 euros  
Echantillon : 70 individus  
Montant frais moyen  
sur l'échantillon : 190,48 euros  
Montant de redressement  
moyen sur échantillon : 28,57 euros

**Ratio strate 2** :  $28,57/190,48 = 15\%$

A défaut, lorsque l'inspecteur ne dispose que du montant global alloué à la population totale, il appliquera un ratio de redressement calculé sur la base des ratios déterminés sur chaque strate pondérés par la taille des strates, au montant global. Dans ce cas d'échantillonnage stratifié, il s'agit de la méthode d'extrapolation par le ratio combiné.

Exemple de calcul avec le ratio combiné :

Montant des frais pour l'ensemble de la population : 500 000 euros

**Strate 1**

Population : 1 000 individus  
Echantillon : 30 individus  
Montant frais moyen  
sur l'échantillon : 116,67 euros  
Montant de redressement  
moyen sur échantillon : 16,67 euros

**Strate 2**

Population : 2 000 individus  
Echantillon : 70 individus  
Montant frais moyen  
sur l'échantillon : 190,48 euros  
Montant de redressement  
moyen sur échantillon : 28,57 euros

**Ratio combiné** :  $(16,67 * 1000 + 28,57 * 2000) / 3000 = 13,48\%$   
 $(166,67 * 1000 + 190,48 * 2000) / 3000$

**Extrapolation par le ratio combiné à la totalité des frais :**

$500\,000 * 13,48\% = 67391,30$  euros

*Extrapolation par le nombre d'individus de la population*

Il s'agit d'appliquer l'assiette moyenne du redressement (par individu statistique) calculée dans l'échantillon à l'ensemble des individus de la base de sondage. On l'utilise lorsqu'il n'existe aucune information sur les sommes versées. La méthode est inchangée dans le cas où l'exploitation de l'échantillon révèle que certains individus tirés de la base de sondage ne sont pas en réalité concernés par le chef de redressement (individus « absents »). Ce peut être notamment le cas lors du tirage de couples (salarié, mois) si l'on tire un mois correspondant aux congés d'un salarié. Les formules de chiffrage de l'extrapolation par le nombre d'individus tiennent compte de cette situation dès lors que le montant de l'assiette redressée pour ces individus est saisi à zéro.

Exemple de calcul :

Population totale : 1 000 individus  
Echantillon : 50 individus  
Montant du redressement sur l'échantillon : 2 500 euros  
Montant du redressement moyen sur l'échantillon :  $2\,500 / 50 = 50$  euros

**Extrapolation à la population totale** :  $1\,000 \text{ ind.} * 50 = 50\,000$  euros

Lorsque la population est stratifiée, c'est le redressement moyen obtenu sur chaque strate représentée dans l'échantillon qui est extrapolé aux nombres d'individus de ces strates.

Exemple de calcul pour une population stratifiée :

<b>Strate 1</b>		<b>Strate 2</b>	
Population :	1 000 individus	Population :	2 000 individus
Echantillon :	30 individus	Echantillon :	70 individus
Redressement :	500 euros	Redressement :	2 000 euros
Redressement moyen :	16,67 euros	Redressement moyen :	28,57 euros

**Extrapolation dans chaque strate :**

$1000 * 16,67 = 16\ 670$  euros

$2000 * 28,57 = 57\ 140$  euros

**Redressement total :**  $16\ 670 + 57\ 140 = 73\ 810$  euros

### 3 Communication des résultats à l'entreprise

Les résultats des chiffrages déterminés au moyen des techniques d'échantillonnage et d'extrapolation sont communiqués à l'entreprise dans le cadre de la lettre d'observations prévue par l'article R 243-59 du code de la Sécurité sociale.

Cette communication prévoit la reprise complète de tous les éléments de l'analyse avec notamment la description des méthodes utilisées pour définir la population, pour effectuer le tirage du ou des échantillons ainsi que le mode d'extrapolation.

Dans le délai de 30 jours ouvert par la réception de la lettre d'observations, l'employeur peut décider de procéder lui-même au calcul des rectifications formalisées par l'inspecteur. Il doit alors, dans ce même délai, en informer l'inspecteur par courrier recommandé avec accusé de réception.

L'employeur doit, dans un délai de 30 jours suivant la réception, par l'inspecteur, de la décision de régularisation, transmettre le résultat de sa proposition de régularisation. Cette proposition doit porter sur la totalité des individus composant la population dont est issu l'échantillon, et doit également être accompagnée des éléments permettant de s'assurer de son exactitude.

La proposition de régularisation établie par l'employeur, est également susceptible d'être contrôlée par l'inspecteur du recouvrement, qui pourra, le cas échéant, utiliser la base des données de l'employeur où figureront les régularisations opérées par ce dernier afin de procéder au tirage d'un échantillon. L'analyse de cet échantillon permettra de vérifier la validité des régularisations faites par l'employeur pour chaque individu de la population.