

RECODIFICATION DU RECENSEMENT 1999 EN NAF RÉV.2

ÉLABORATION D'UN OUTIL POUR LE CALCUL DE PRÉCISION

Benoît BUISSON
Insee, Pôle ingénierie statistique entreprises

Introduction

Le bureau du Comité de Pilotage de l'Action Régionale (COPAR) de l'INSEE, lors de la réunion du 30 septembre 2009, a passé commande au Pôle Ingénierie Statistique Entreprises (PISE) de la recodification de l'activité principale exercée par les établissements (APET) issue du recensement de la population de 1999 en nomenclature naf rév. 2. En effet, le recensement 2007 et les suivants seront uniquement diffusés en naf rév. 2. Pour pouvoir calculer des évolutions d'emploi entre 1999 et les années postérieures à 2006, le bureau du Copar a jugé nécessaire d'appliquer la naf rév. 2 au recensement de 1999. Pour assurer cette codification, il n'a pas été envisagé de retourner au niveau « individuel » des SIRET de 1999. La demande, vis à vis du pôle ISE, portait donc « uniquement » sur le fait de savoir s'il était possible à partir des activités codées en naf rév. 1 d'en déduire une activité en naf rév. 2, en prenant la table de passage utilisée pour double codifier le recensement 2006. Il s'agissait bien de tirer parti de cette table de passage nationale de 2006, et non de proposer une autre table de passage. Cet article se propose de détailler les travaux méthodologiques effectués dans ce cadre notamment en insistant sur la construction d'un outil visant à mesurer la précision statistique des résultats obtenus.

1. Les travaux réalisés pour la recodification

1.1. La table de passage utilisée

Il n'était pas demandé au pôle de réfléchir à **une table de passage** naf rév. 1 naf rév. 2 adéquate mais de reprendre une [table de passage existante](#)¹, utilisée pour codifier le recensement 2006 dans les deux nomenclatures ([note 2107/DG75-F520 du 28 septembre 2009](#)). Après fusion avec les fichiers détails du recensement de 1999, **deux problèmes sont apparus**, en effet des codes du fichier détail ne figuraient pas dans la table de passage. D'une part il s'agissait de codes spécifiques au recensement pour **les activités agricoles**. Le recensement 1999 a en effet été codé spécifiquement pour les activités agricoles, dans une nomenclature plus agrégée que la naf. Le pôle ISE a obtenu du pôle Entrepot de Données Locales une « table de passage » sur le sujet, sans avoir une idée toutefois de la clef de répartition entre les différents sous-postes. Pour approximer cette clef de répartition nous avons utilisé les fichiers SIRENE, sur le champ complémentaire, du premier janvier 2003 en prenant

¹ Les liens proposés pointent sur le site intranet du pôle ISE, accessible uniquement aux utilisateurs internes à l'INSEE. Pour accéder à ces documents de l'extérieur de l'Insee, merci de contacter l'auteur de cet article.

comme variable de répartition le nombre d'entreprises. D'autre part, le recensement 1999 est codé en **naf 1993** et non en naf rév. 1. Il est possible de cerner quatre types de modification entre la naf 1993 et la naf rév. 1 : regroupement de codes, nouveau code sans changement de contour, éclatement, code identique avec changement de contour. Les deux premiers cas ne posent pas de problème, ils ont pu être traités directement à partir de la table de passage. Pour les deux autres il a fallu utiliser des clefs de répartition, à partir de la base de données ALISSE du 31 décembre 2003 avec, comme variable de répartition, le nombre d'entreprises. Après correction de ces deux types de problème, **nous avons de fait reconstitué une [table de passage directement applicable au RP 1999](#)** en complétant la table existante, tous les codes naf utilisés dans le RP s'y retrouvent, ce qui n'était pas le cas auparavant. Il est important de quantifier le nombre et la part des postes concernés par ce traitement. Au niveau France, la part des emplois agricoles au lieu de résidence en 1999 atteint 4,2 %. Cette part atteint son maximum dans le Limousin avec plus de 8 %. Comme déjà signalé, l'ensemble de ces cas n'a pas nécessité de recodification : 2,5 % des emplois ont dû être recodés au titre de la nomenclature spécifique appliquée dans l'agriculture. Cette part dépasse les 5 % dans quatre régions métropolitaines (Aquitaine, Limousin, Languedoc-Roussillon et Corse) ainsi qu'en Martinique. En ce qui concerne le passage naf 93 naf rév. 1 (hors agriculture), cela concerne 4,2 % des emplois avec un maximum de 6 % en Ile de France. Là-aussi tous ces cas n'ont pas nécessité de recodification : 3,2 % des emplois ont été recodés au titre de ce changement de nomenclature, avec un maximum de 5,2 % en Ile-de-France. Globalement **5,8 % des emplois ont dû être recodés**. Plus exactement la table de passage a dû être complétée pour 5,8 % des postes. La Martinique, le Limousin, le Languedoc-Roussillon et la Corse dépassent les 8 % d'emplois à recoder. Dans la suite nous ne considérons pas ces postes comme imputés, donc nous prenons comme hypothèse que ce travail sur la table de passage ne génère pas d'aléa.

1.2. L'imputation aléatoire : le principe de la recodification

Une fois la table de passage complètement définie, il est possible de débiter la recodification de tous les individus de la base du recensement 1999, ayant un code d'activité en naf rév 1. Un certain nombre de cas ne pose absolument aucune difficulté : il s'agit des codes en naf rév. 1 pour lesquels 100 % des emplois se retrouvent dans un seul code de naf rév. 2. La demande d'expertise porte bien sûr sur les cas où **il y a ventilation en plusieurs codes de naf rév. 2**. Dans ce type de situation, le principe général est d'affecter une activité en naf rév. 2 selon le même principe que **le traitement de la non-réponse partielle** dans une enquête, à savoir par un **mécanisme aléatoire**. Ainsi si 80 % des emplois d'un code X en naf rév. 1 se retrouve en A1 en naf rév. 2 et 20 % en A2 ; nous effectuerons un tirage aléatoire qui affectera le code A1 avec une probabilité de 0,8 et le code A2 avec une probabilité de 0,2. Il ne s'agit pas bien sûr d'affecter « de force » le code majoritaire. Le pôle ISE a conçu un outil, une macro SAS, qui permet simplement de réaliser cela. En lien avec cette procédure, nous pouvons définir deux indicateurs utiles pour la suite :

- un indicateur qui distingue simplement les cas où l'imputation est 100 % sûre, cas où le code de naf rév. 1 de l'individu arrive dans 100 % des cas sur un et un seul code de

naf rév 2. Cet indicateur sera appelé imput par la suite, il vaut 0 si l'imputation est 100 % sûre, 1 dans le cas contraire.

- un autre indicateur qui distingue simplement les cas où l'imputation est presque sûre des autres cas. On dira que l'imputation est presque sûre lorsque la probabilité d'affecter le code en naf rév. 2 le plus fréquent est supérieure à 90 %, dans ce cas l'indicateur imput90 vaut 0. Il vaudra 1 dans le cas d'imputation non presque sûre. Par définition imput=0 implique imput90=0.

Bien sûr une fois ces indicateurs définis, il va être possible de calculer des taux moyens d'imputation par zone qui vont permettre de juger de la plus ou moins grande fiabilité, d'un point de vue statistique, des opérations réalisées.

1.3. Le niveau de nomenclature choisi pour recodifier

Comme déjà esquissé, il était primordial de choisir **le niveau de nomenclature** à partir duquel réaliser l'imputation aléatoire. À un niveau fin, les taux d'imputation seront plus élevés mais les études en région pourront se révéler plus riches qu'à un niveau trop agrégé. Il faut donc trouver un bon compromis entre ces deux principes : maintenir des taux d'imputation aléatoires relativement faibles d'une part, permettre des études en région sur ce sujet d'autre part. En plus du niveau de nomenclature, il se pose également la question du **zonage territoriale** sur lequel analyser les taux d'imputation. La commande initiale du bureau du Copar mentionnait de privilégier le niveau zone d'emploi tout en n'excluant pas les possibilités d'études sur les zonages à façon, regroupement de communes en particulier. Pour analyser les taux d'imputation, à partir de la table de passage et de la table à codifier et avant de réaliser l'imputation aléatoire, nous avons conçu une macro SAS qui génère les deux indicateurs imput et imput90 en fonction du zonage retenu (région, département, zone d'emploi, commune) et du niveau de nomenclature (du niveau le plus fin au niveau A10). Cet outil génère également des indicateurs statistiques de synthèse pour chaque croisement zonage * niveau de nomenclature, comme la médiane ou encore le premier décile en ce qui concerne la part d'activité imputée par zone de référence.

Au niveau de nomenclature le plus fin et au lieu de résidence, 45 % des postes naf rev1 (champ : France métropolitaine et DOM) nécessitent une imputation aléatoire en Naf rév.2 alors que l'indicateur imput90 s'élève à 21 %. Plus nous raisonnons à un niveau agrégé de nomenclature en Naf rév.2, plus ces taux ont tendance à diminuer bien sûr. Au niveau A10, ils atteignent respectivement (toujours au lieu de résidence) 22 % et 4 %. Lorsque nous observons les différents niveaux de nomenclature nous repérons un décrochage « intéressant » au niveau A129 de la nomenclature. À ce niveau les taux d'imputation s'élèvent à 32 % (imput) et 10 % (imput90). L'indicateur imput90, sur lequel nous allons fonder l'essentiel de nos conclusions, devient particulièrement faible à ce niveau et chutant de plus de 10 points par rapport au niveau A272. Nous avons privilégié par la suite **le niveau A129** de la nomenclature. Ce niveau de nomenclature n'est pas un niveau international, contrairement au niveau A88 (division au niveau international). Toutefois il propose, par rapport au niveau A88, des

éclatements de postes importants notamment dans l'industrie². Après avoir choisi ce niveau de nomenclature, nous nous sommes intéressés au niveau géographique de pertinence, toujours uniquement à partir d'une analyse descriptive des taux d'imputation et plus particulièrement de l'indicateur imput90. **Les niveaux régions et départements** ne semblent pas poser de problèmes spécifiques, que cela soit au lieu de travail ou au lieu de résidence. De manière générale, la variabilité des taux d'imputation est plus forte au lieu de travail en lien avec la concentration de l'emploi. Au lieu de travail, la région Île-de-France connaît les taux d'imputation les plus élevés, l'indicateur imput90 frôlant les 12 %. Il s'agit toutefois d'un taux qui apparaît faible. Au niveau département, cinq départements³ enregistrent un indicateur imput90 supérieur à 12 %, le maximum étant atteint dans les Hauts de Seine avec 17 % et un taux d'imputation imput de 37 %. Même si le taux d'imputation peut apparaître élevé, moins de 20 % des cas se révèlent être des imputations « non presque sûres ». Pour cette raison nous considérons que le niveau département se révèle fiable au niveau A129 de la nomenclature. Pour **les zones d'emploi**, la fiabilité d'ensemble se révèle là aussi très bonne. Au lieu de travail, plus de la moitié des zones d'emploi ont un indicateur imput90 inférieur à 10 %, près de 95 % des zones ont un indicateur imput90 inférieur à 15 %. Toutefois trois zones apparaissent « extrêmes » avec un indicateur imput90 supérieur à 20 % : Morteau et Nogent le Rotrou (20 %), Le Creusot (25 %). La part des postes imputés atteint 45 % pour cette dernière zone d'emploi. Dans ces cas atypiques, il est facile de présumer que les taux élevés d'imputation proviennent surtout d'un employeur majeur de la zone. Moyennant quelques approximations, il devrait être possible de forcer la codification à la main dans ce cas. Par exemple si nous savons que l'employeur principal d'une zone était codifié XX en naf rev1 et AA en naf rev2, l'idée serait de forcer tous les codes XX de la zone en AA en naf rev2. L'inconvénient serait de basculer à tort un certain nombre d'entreprises et donc d'emplois, toutefois en lien avec le niveau de nomenclature et la concentration de l'emploi cet inconvénient nous apparaît minime. **Le niveau communal** est lui-aussi globalement très bon, toutefois les exceptions se multiplient à ce niveau. Au lieu de travail, plus de 50 % des communes enregistrent un indicateur imput90 inférieur à 5,5 % ; 90 % d'entre-elles ont un indicateur imput90 inférieur à 22 % (pour le lieu de résidence la médiane vaut 8,8% et le 9 ème décile est de 18 %). Au lieu de travail, nous avons sélectionné les communes qui ont un indicateur imput90 supérieur à 20 %. Vingt-sept d'entre-elles ont plus de 8000 postes codés en naf rev1. Les plus importantes sont Courbevoie et Boulogne-Billancourt tandis que l'indicateur imput90 est particulièrement dégradé pour les communes de Grande-Synthe (département 59), Florange (département 57) et Blanquefort (département 33). Au lieu de résidence, nous retrouvons Le Creusot qui est la plus « grande » commune avec un indicateur imput90 supérieur à 20 %. Le fait que quelques communes soient dégradées peut poser des problèmes spécifiques, pas facilement identifiables, lors des regroupements à façon de communes.

² Au besoin il peut être nécessaire de consulter la [rubrique nomenclature agrégée](#) sur le site internet de l'INSEE.

³ Le Doubs, Le Loir-et-Cher, La Loire, Les Yvelines et Les Hauts de Seines.

1.4. Les redressements déterministes

Après avoir choisi le niveau A129 de la nomenclature pour réaliser l'imputation aléatoire, il apparaissait donc que le taux d'imputation (indicateur imput90) se révélait être supérieur à 20 % pour 27 communes de plus de 8 000 emplois au lieu de travail. Suite à ce premier examen de résultats, nous avons décidé de **recodifier « manuellement »** un certain nombre de cas en cherchant les plus grands employeurs des zones qui posent des difficultés. Pour la recodification manuelle, nous avons considéré le cas des 105 communes qui enregistrent un indicateur imput90 supérieur à 15 % tout en ayant plus de 5 000 emplois au lieu de travail. Pour ces communes, nous avons effectué le rapprochement de la base RP avec la base CLAP. Pour chacune des communes, nous avons vérifié que l'activité concentrant le plus d'emplois correspond au premier établissement employeur selon la base CLAP 2006, codifiée en naf rév. 1. Pour cela nous avons comparé le code APE entre les deux sources et l'ordre de grandeur des effectifs. Dans le cas où l'activité dominante au recensement diverge avec celle du plus grand établissement de la commune selon CLAP, nous avons considéré les établissements qui arrivent aux rangs 2 et 3 selon CLAP. Suite à ce premier travail, les données concordent pour 76 communes qui peuvent potentiellement bénéficier d'un traitement manuel. Pour les autres cas nous n'avons pas fait de recherches complémentaires, il aurait fallu connaître l'historique des plus grands établissements employeurs de la commune entre 1999 et 2006. Pour les 76 communes concernées, il reste à s'assurer que les établissements repérés dans CLAP 2006 sont toujours actifs dans la base CLAP 2008, codifiée en naf rév. 2. C'est le cas pour 70 d'entre elles, pour les 6 autres communes les établissements ont cessé ou subi une restructuration jugée complexe. Pour ces 70 communes, nous avons fait une **imputation déterministe manuelle**. Ainsi, par exemple, pour la commune de Florange dont 46 % de l'emploi au lieu de travail en 1999 est présent dans l'APE 271Z, nous attribuons à ces emplois le code 2410Z de la naf rév. 2 en lien avec la présence du premier établissement employeur de la zone, Arcelor. Dans le cas des redressements déterministes, les indicateurs imput et imput90 valent, après traitement et donc dans le fichier mis à disposition, 0 du fait de l'absence de redressement aléatoire.

Bien sûr ce traitement déterministe, bien balisé grâce aux recherches dans CLAP, permet de faire descendre l'indicateur imput90 dans les zones où il se révélait être particulièrement fort. Ainsi après traitement, 31 communes sur 70 voient leur indicateur imput90 passer sous la barre des 15 % au niveau A129 de la nomenclature. Au niveau zone d'emploi, 12 zones d'emploi (au lieu de 20 initialement) ont un indicateur imput90 supérieur à 15 %, et ce même indicateur demeure supérieur à 20 % dans une seule zone d'emploi, Morteau. La recodification déterministe engendre donc une amélioration sensible des cas les plus dégradés. Vous trouverez en annexe 1 la liste des redressements déterministes effectués.

1.5. Imputation aléatoire et niveaux de nomenclature

Après avoir pratiqué quelques redressements déterministes, l'imputation aléatoire a donc été réalisée au niveau A129 de la nomenclature, comme explicité au paragraphe 1.3.

Pour les niveaux de nomenclature plus agrégés que le A129 (niveaux A88, A64, A38, A21, A17, A10, A5) il figure également le code d'activité correspondant dans la table mise à disposition (par exemple variable C_NA88 pour le niveau A88). Pour faire cela, nous avons juste utilisé les tables de passage correspondantes à partir du niveau A129. L'imputation aléatoire est donc réalisée uniquement au niveau A129 pour éviter les confusions. Pour chaque niveau de nomenclature, l'utilisateur peut trouver dans la table mise à disposition les deux indicateurs correspondants déjà mentionnés, par exemple `imput_A38` et `imput90_A38` pour le niveau en 38 postes. Pour un individu donné du fichier il est possible d'avoir `imput_A129=1` et `imput_A38=0`. L'activité a été imputée au niveau A129 mais si, dans la table de passage, le code d'activité de l'individu en naf rév. 1 est ventilé entre plusieurs codes au niveau A129 qui appartiennent tous au même niveau A38 alors `imput_A38=0`. Plus le raisonnement se fera à un niveau agrégé, plus l'imputation sera fiable et les taux d'imputation faibles. Évidemment le cas inverse n'est pas vrai : nous ne pouvons pas avoir `imput_A129=0` et `imput_A38=1` (ces remarques ne sont pas valides entre les niveaux A21 et A17, ce dernier n'étant pas un regroupement parfait des modalités du A21). Par niveau de nomenclature il y a donc trois variables disponibles dans les fichiers mis à disposition : le code d'activité, les deux indicateurs d'imputation préfixés par `imput` et `imput90`.

2. Comment juger de la pertinence statistique ?

2.1. Simulations et règle de décision

Vis à vis des utilisateurs, chargés d'études en directions régionales notamment, il nous semblait important d'édicter des précautions d'usage de la table mise à disposition, d'établir une **règle de décision** pour juger de la fiabilité statistique de l'imputation ainsi réalisée. Pour établir cette règle de décision, nous avons souhaité calculer des variances, sur des estimateurs du type « nombre d'emplois par zone selon le code d'activité XXX en NAF rév. 2 »⁴ et donc calculer des intervalles de confiance ainsi que des coefficients de variation. Pour cela nous avons généré le mécanisme d'imputation aléatoire non pas une fois mais 500. Pour des raisons de temps de calcul⁵ mais aussi de facilité d'analyse, nous avons limité le nombre de cas à regarder et à analyser. Tout d'abord nous avons limité le nombre de zones d'emploi sur lesquelles réaliser ces **imputations multiples**, en essayant de prendre des cas « type » en termes de taille de zone d'emploi et de taux d'imputation. Pour les zones d'emploi sélectionnées, nous n'avons pas retenu tous les postes de la nomenclature en 129 postes pour les estimations, nous avons sélectionné 4-5 postes par zone, en mixant des postes fréquents dans la zone et des postes à fréquence plus réduite. La liste des zones d'emploi

⁴ Au niveau 129 de la nomenclature

⁵ Générer 500 imputations sur un fichier de 6 millions d'observations peut s'avérer long.

étudiées ainsi que les postes de la nomenclature retenus figurent en annexe 2. L'idée générale était de considérer comme fiable les estimations avec un **coefficient de variation** inférieur à 5 %. Nous avons considéré également une zone de « fiabilité moyenne » avec des coefficients de variation compris entre 5 % et 10 %. Notre objectif était de relier la valeur du coefficient de variation avec deux paramètres : le taux d'imputation imput90 et la moyenne des postes à estimer (notion de taille) pour édicter une règle de décision. En effet l'utilisateur disposera en premier lieu des taux d'imputation, il pourra estimer un nombre d'emplois par poste mais ne pourra pas « en fonctionnement standard » avoir recours à l'imputation multiple pour juger de la fiabilité des résultats. Cette analyse n'a pas été des plus simples, mais nous sommes arrivés dans un première temps à la règle de décision suivante :

- si l'indicateur imput90 est inférieur à 5 % l'imputation sera jugée fiable quelle que soit la valeur estimée ;
- si l'indicateur imput90 est inférieur à 20 % avec une estimation supérieure à 200 et inférieure à 600 l'imputation sera jugée également fiable ;
- Pour les estimations supérieures à 600, l'imputation sera jugée fiable quel que soit le taux d'imputation.

De fait, cette règle était par nature « prudente » pour éviter de considérer comme fiables des situations où le coefficient de variation se révélait être supérieur à 5 %. Il est naturel de voir que plus le taux d'imputation est élevé et plus le niveau d'emploi estimé est faible plus la fiabilité peut être mise en cause. C'est donc l'étude des simulations qui nous a permis de définir une règle empirique associant les deux critères. Il faut bien avoir en tête ce que nous entendons ici par le terme « fiabilité ». Il s'agit bien d'une fiabilité « uniquement » statistique au sens **d'indicateur de précision** de type variance d'un estimateur. À partir du fichier mis à disposition, l'utilisateur peut estimer des nombres d'emplois par activité et par zone géographique. Ce fichier est issu d'une imputation aléatoire. Si nous faisons retourner cette imputation aléatoire, le fichier en sortie changerait. L'utilisateur obtiendra donc une nouvelle estimation pour l'activité en question. Quand le résultat est fiable, cela signifie ici que l'estimation varie peu potentiellement en fonction de l'imputation aléatoire.

Nous avons voulu compléter ces tests, revenir au besoin sur la règle de décision, à partir des travaux demandés par le PSAR (pôle de service à l'action régionale) Offre de Données Locales. Depuis le début de l'opération, le calendrier de celle-ci a été défini par le bureau du Copar en lien avec une mise à disposition des données pour le PSAR offre de données locales. La demande du PSAR portait sur une diffusion de l'emploi, au lieu de travail et en nomenclature agrégée en 5 postes (NA5), pour les communes de plus de 2000 habitants pour le millésime 2008 du recensement diffusé au 1^{er} juillet 2011. Il s'agissait au travers du site Web-RP de renouer avec la comparaison des données du dernier millésime du recensement avec celles du recensement 1999⁶. Après avoir réalisé l'imputation aléatoire sur l'ensemble du fichier et effectué quelques « corrections » déterministes décrits précédemment, nous avons regardé si les données diffusées correspondaient à la règle énoncée plus haut. En effectuant de nouveau une analyse de type **imputation multiple**, nous nous sommes

⁶ Cette comparaison était présente pour les recensements 2006 et 1999 mais en naf rév.1 ; elle avait été abandonnée pour le recensement 2007 diffusé en Naf rév.2

aperçus que nous avons été trop prudents dans la règle énoncée dans un premier temps. L'analyse nous a permis de modifier **la règle de décision** de la façon suivante :

- Si l'indicateur imput90 est inférieur à 10 % l'imputation sera jugée fiable quelle que soit la valeur estimée ;
- Si l'indicateur imput90 est inférieur à 20 % avec une estimation supérieure à 80 et inférieure à 160 l'imputation sera jugée également fiable ;
- Si l'indicateur imput90 est inférieur à 40 % avec une estimation supérieure à 160 et inférieure à 400 l'imputation sera jugée également fiable ;
- Pour les estimations supérieures à 400, l'imputation sera jugée fiable quel que soit le taux d'imputation.

En appliquant cette règle, il reste des cas jugés problématiques mais quasi-exclusivement pour des **estimations sur des très petits effectifs**, inférieures à 100 en données pondérées. Bien que l'imputation ajoute des difficultés, la diffusion sur ces petits effectifs est déjà source d'imprécision avec un sondage au quart. Il n'est pas inutile de rappeler qu'au niveau NA5 le taux d'imputation sur l'ensemble du fichier est inférieur à 2 %, ce qui est très faible. Nous avons mis à disposition du pôle entrepôt de données locales une base imputée, ce qui a permis la diffusion de ces tableaux par le PSAR Offre de Données Locales en juillet 2011. De fait une des premières conséquences des travaux du pôle ISE concerne donc la mise à disposition via le service universel, des tableaux de la ventilation de l'emploi du recensement de 1999 en NA5, nouvelle nomenclature donc, pour les communes de plus de 2000 habitants. Une précision : compte tenu de la finesse des tableaux détaillés et pour être en conformité avec les règles du secret statistique, les données des tableaux détaillés ne sont effectivement proposées que pour des zones de plus de 2000 habitants. Néanmoins, les bases de données communales associées fournissent l'ensemble des données, quelle que soit la taille de la commune donc y compris les communes de moins de 2000 habitants, car elles offrent la possibilité de regroupements à façon.

2.2. Imputation multiple et calcul de précision

Comme nous l'avons vu dans la partie précédente, un bon moment pour juger de la pertinence statistique consiste à réaliser des imputations multiples. Par contre, il est très coûteux de réaliser ces imputations multiples sur l'ensemble du fichier. De plus, cela peut se révéler difficile à utiliser pour les chargés d'études en région. Pour estimer des emplois par code NAF et par code géographique, en prenant la moyenne des estimations obtenues sur chaque imputation, cela aurait l'avantage de fiabiliser les résultats. Toutefois la mise à disposition d'un fichier détail, avec la réalisation de 500 imputations, peut entraîner des difficultés. De ce fait, pour tester de manière plus approfondie la fiabilité statistique de la codification du RP 1999 en naf rév. 2, le pôle a conçu et mis à disposition sur [son site intranet](#) un **kit sas** qui permet de calculer des indicateurs de fiabilité de l'imputation. Ceci peut s'avérer particulièrement utile pour les cas qui posent des difficultés par rapport à la règle de décision précédente, sur les **zonages à façon** par exemple. L'idée générale de cet outil est de réaliser l'imputation un grand nombre de fois sur une zone et pour un niveau de nomenclature défini par l'utilisateur. La macro SAS permet de réaliser l'imputation 500 fois et fournit en sortie **des indicateurs**

de fiabilité et de dispersion : moyenne de l'estimation, sur les 500 imputations réalisées, valeurs minimale et maximale, médiane et surtout **coefficient de variation**. C'est notamment la lecture du coefficient de variation qui peut fournir une aide précieuse. Selon les critères Eurostat pour les enquêtes européennes, un coefficient de variation inférieur à 5 % donne des gages suffisants de fiabilité, au-delà de 10 % l'incertitude s'avère trop grande pour diffuser des informations robustes. Entre 5 et 10 %, la prudence demeure. Pour les cas qui posent des difficultés au regard de la règle de décision énoncée dans la partie précédente, il peut être intéressant de faire tourner la macro SAS et de regarder le coefficient de variation associée. Si celui-ci est inférieur à 5 %, l'analyse et la diffusion pourront se faire. Dans la plupart des cas la règle de décision et la lecture du coefficient de variation s'avèrent liées : un cas qui sort de la règle de décision a deux grandes chances d'entraîner un coefficient de variation supérieur à 5 %. Outre le principe de décision lui-même, l'utilisation de cet outil SAS donne des indicateurs de précision que ne permettent pas d'obtenir, directement, les indicateurs d'imputation.

Pour illustrer cela nous pouvons prendre un exemple. L'idée est ici d'estimer l'emploi dans l'aire urbaine⁷ de Saint-Nazaire au niveau A38 de la nomenclature, naf rév. 2, en 1999. A ce niveau de nomenclature, trois postes posent problème au regard de la règle de décision précédemment énoncée. Ainsi pour les activités informatiques et les services d'information (poste JC de la nomenclature), l'estimation du nombre d'emplois au lieu de travail est comprise entre 80 et 160 pour un taux imputé de près de 32 %, supérieur à la borne de 20 % édictée dans la règle de décision. Utiliser la macro SAS permet de mieux comprendre la situation. En réalisant 500 imputations sur cette zone et à ce niveau de nomenclature, nous constatons que la valeur moyenne estimée est de 105 emplois dans les activités informatiques, avec un minimum de 84 et un maximum de 128. Le coefficient de variation est de 6,6 % : il indique bien que la diffusion sur ce poste de nomenclature peut s'avérer problématique, car trop incertaine. La règle de décision est confirmée ici. Le coefficient de variation bien que n'étant pas très « élevé » est supérieur à 5.

Le contenu du kit pour calculer la précision est indiqué en annexe 3. Il figure également dans cette même annexe 3 le mode d'emploi de la macro SAS, à lire attentivement avant utilisation. Générer 500 imputations aléatoires peut prendre du temps de calcul, plusieurs heures si les zones s'avèrent grandes. Il est préférable de faire des tests dans un premier temps, sur des zones réduites, pour estimer le temps de calcul nécessaire.

Enfin dernier point : même sans recodification en naf rév. 2, il ne faut pas oublier que l'utilisation du **recensement complémentaire** ne peut fournir que des estimations, et non des grandeurs réelles, en lien avec le sondage au 1/4. Ces estimations peuvent elles aussi se révéler problématiques dans le cas de très faibles effectifs.

⁷ Travail effectué avant la refonte des aires urbaines en 2011

2.3. Pertinences statistique et économique

Les éléments précédents, règle de décision et macro SAS de calcul de précision, permettent de juger de la fiabilité statistique des éléments que l'utilisateur souhaite analyser ou diffuser. Il est nécessaire toutefois d'aller plus loin. En effet, la fiabilité statistique n'est pas tout en la matière. Il s'avère indispensable de vérifier les estimations fournies avec des éléments de terrain.

Deux pistes d'approfondissement peuvent être envisagées sur ce sujet. La première concerne bien sûr **la connaissance locale du système productif** de la zone d'étude. Le chargé d'étude dispose pour cela de la ventilation de l'emploi en 1999 en naf rév. 1, une estimation de cette ventilation en naf rév. 2 et des éléments plus récents à partir du recensement. La connaissance de l'activité fine des grands établissements de la zone d'étude reste bien sûr primordiale en la matière. Associée à la lecture de la table de passage, cela peut permettre de déceler des cas qui peuvent se révéler problématiques. Prenons un exemple « fictif » : en 1999 dans une zone d'étude beaucoup d'emplois relevaient du poste en naf rév. 1 « autres imprimeries » (poste 22.2C). Dans la table de passage, nous pouvons voir que ce poste est ventilé en trois postes de la nomenclature A129 de la naf rév. 2 : le poste « imprimerie et reproduction d'enregistrements » (poste C18Z avec une probabilité de 88 %), le poste « fabrication d'articles en papier et en carton » (poste C17B avec une probabilité de 11,8 %) et enfin le poste « édition » (poste J58Z avec une probabilité très faible de 0,2%). En lien avec les probabilités d'affectation, près de neuf emplois sur dix de la zone d'étude risquent donc de se voir affecter l'activité « imprimerie et reproduction d'enregistrement ». Si nous savons par ailleurs, ici c'est la connaissance locale qui intervient, que l'établissement en question est une papeterie cela pose évidemment problème. La deuxième piste d'approfondissement concerne **la double recodification du recensement 2006**. Contrairement à ce qui a été fait ici, cette double recodification s'est opérée majoritairement par établissement et non à partir d'une table de passage globale. Même s'il reste de l'incertitude, cette double codification peut donc être jugée très fiable. Ainsi sur le zonage d'étude, l'utilisateur dispose des éléments qui suivent par activité : niveau de l'emploi en 1999 naf rev 1, niveau de l'emploi en 2006 naf rév. 1, niveau de l'emploi naf rév. 2 en 2006 pour établir un diagnostic sur l'estimation de 1999 en naf rév. 2. Ici aussi, prenons un exemple fictif. Supposons que dans un code X donné en naf rév. 1 le nombre d'emplois soit de 125 en 1999 et 150 en 2006 sur une zone à façon donnée. En 2006 et en naf rév. 2, ces 150 emplois sont ventilés en 50 emplois pour le code A1 et 100 emplois pour le code A2. De ce fait, une ventilation des 125 emplois en 1999 en 100 emplois en A1 et 25 emplois en A2 peut paraître suspecte. Des approfondissements sont sans doute nécessaires.

Annexe 1 : les redressements déterministes

A) liste des 70 communes pour lesquelles une imputation déterministe manuelle a été effectuée

10081	La Chapelle-Saint-Luc	76057	Barentin
12176	Onet-le-Château	77118	Claye-Souilly
13039	Fos-sur-Mer	77243	Lagny-sur-Marne
14437	Mondeville	77305	Montereau-Fault-Yonne
17028	Aytré	77350	Ozoir-la-Ferrière
21355	Longvic	78117	Buc
24520	Sarlat-la-Canéda	78297	Guyancourt
25031	Audincourt	78362	Mantes-la-Ville
26235	Pierrelatte	78423	Montigny-le-Bretonneux
26281	Romans-sur-Isère	78640	Vélizy-Villacoublay
28134	Dreux	88209	Golbey
28280	Nogent-le-Rotrou	91086	Bondoufle
29046	Douarnenez	91161	Chilly-Mazarin
33056	Blanquefort	91223	Étampes
35115	Fougères	91661	Villebon-sur-Yvette
36088	Issoudun	92009	Bois-Colombes
38449	Saint-Quentin-Fallavier	92025	Colombes
41018	Blois	92026	Courbevoie
41269	Vendôme	92040	Issy-les-Moulineaux
42005	Andrézieux-Bouthéon	92062	Puteaux
42071	Le Coteau	92073	Suresnes
42187	Roanne	92078	Villeneuve-la-Garenne
42207	Saint-Chamond	94015	Bry-sur-Marne
44026	Carquefou		
44035	La Chapelle-sur-Erdre		
45284	Saint-Jean-de-Braye		
46102	Figeac		
49267	Saint-Barthélemy-d'Anjou		
57221	Florange		
59009	Villeneuve-d'Ascq		
59271	Grande-Synthe		
59286	Haubourdin		
59527	Saint-André-lez-Lille		
60395	Méru		
60414	Montataire		
61169	Flers		
62239	Coquelles		
62427	Hénin-Beaumont		
63075	Chamalières		
64260	Hendaye		
67348	Obernai		
68278	Rixheim		
68287	Rouffach		
69081	Écully		
69282	Meyzieu		
71153	Le Creusot		
74011	Annecy-le-Vieux		

B) Liste des redressements déterministes effectués par commune

```
if commune_lt = "10081" and apen = "251A" then apen2 = "2211Z" ;
if commune_lt = "12176" and apen = "343Z" then apen2 = "2932Z" ;
if commune_lt = "13039" and apen = "271Z" then apen2 = "2410Z" ;
if commune_lt = "14437" and apen = "316A" then apen2 = "2931Z" ;
if commune_lt = "17028" and apen = "352Z" then apen2 = "3020Z" ;
if commune_lt = "21355" and apen = "641A" then apen2 = "5310Z" ;
if commune_lt = "24520" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "25031" and apen = "343Z" then apen2 = "2932Z" ;
if commune_lt = "26235" and apen = "233Z" then apen2 = "2013A" ;
if commune_lt = "26281" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "28134" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "28280" and apen = "343Z" then apen2 = "2811Z" ;
if commune_lt = "29046" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "33056" and apen = "343Z" then apen2 = "2932Z" ;
if commune_lt = "35115" and apen = "322B" then apen2 = "2630Z" ;
if commune_lt = "36088" and apen = "361A" then apen2 = "3030Z" ;
if commune_lt = "38449" and apen = "316A" then apen2 = "2931Z" ;
if commune_lt = "41018" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "41269" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "42005" and apen = "343Z" then apen2 = "2932Z" ;
if commune_lt = "42071" and apen = "182E" then apen2 = "1413Z" ;
if commune_lt = "42187" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "42207" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "44026" and apen = "251E" then apen2 = "2219Z" ;
if commune_lt = "44035" and apen = "722Z" then apen2 = "6202B" ;
if commune_lt = "45284" and apen = "245C" then apen2 = "2042Z" ;
if commune_lt = "46102" and apen = "353A" then apen2 = "3030Z" ;
if commune_lt = "49267" and apen = "343Z" then apen2 = "2811Z" ;
if commune_lt = "57221" and apen = "271Z" then apen2 = "2410Z" ;
if commune_lt = "59009" and apen = "803Z" then apen2 = "8542Z" ;
if commune_lt = "59271" and apen = "271Z" then apen2 = "2410Z" ;
if commune_lt = "59286" and apen = "156D" then apen2 = "1062Z" ;
if commune_lt = "59527" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "60395" and apen = "252H" then apen2 = "2229A" ;
if commune_lt = "60414" and apen = "271Z" then apen2 = "2410Z" ;
if commune_lt = "61169" and apen = "361A" then apen2 = "2932Z" ;
if commune_lt = "62239" and apen = "748K" then apen2 = "8299Z" ;
if commune_lt = "62427" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "63075" and apen = "222C" then apen2 = "1812Z" ;
if commune_lt = "64260" and apen = "601Z" then apen2 = "4910Z" ;
if commune_lt = "67348" and apen = "312A" then apen2 = "2712Z" ;
if commune_lt = "68278" and apen = "151E" then apen2 = "1013A" ;
if commune_lt = "68287" and apen = "343Z" then apen2 = "2931Z" ;
if commune_lt = "69081" and apen = "521F" then apen2 = "4711F" ;
if commune_lt = "69282" and apen = "331B" then apen2 = "3250A" ;
if commune_lt = "71153" and apen = "271Z" then apen2 = "2410Z" ;
if commune_lt = "74011" and apen = "722Z" then apen2 = "5829C" ;
if commune_lt = "76057" and apen = "312A" then apen2 = "2712Z" ;
if commune_lt = "77118" and apen = "521F" then apen2 = "4711F" ;
if commune_lt = "77243" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "77305" and apen = "313Z" then apen2 = "2732Z" ;
if commune_lt = "77350" and apen = "751A" then apen2 = "8411Z" ;
if commune_lt = "78117" and apen = "331A" then apen2 = "2660Z" ;
if commune_lt = "78297" and apen = "341Z" then apen2 = "2910Z" ;
if commune_lt = "78362" and apen = "363Z" then apen2 = "3220Z" ;
if commune_lt = "78423" and apen = "742C" then apen2 = "7112B" ;
if commune_lt = "78640" and apen = "341Z" then apen2 = "2910Z" ;
if commune_lt = "88209" and apen = "292F" then apen2 = "2825Z" ;
if commune_lt = "91086" and apen = "641A" then apen2 = "5310Z" ;
if commune_lt = "91161" and apen = "602M" then apen2 = "4941A" ;
if commune_lt = "91223" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "91661" and apen = "721Z" then apen2 = "6202A" ;
if commune_lt = "92009" and apen = "751A" then apen2 = "8411Z" ;
if commune_lt = "92025" and apen = "751A" then apen2 = "8411Z" ;
if commune_lt = "92026" and apen = "741J" then apen2 = "7010Z" ;
if commune_lt = "92040" and apen = "642A" then apen2 = "6110Z" ;
if commune_lt = "92062" and apen = "651C" then apen2 = "6419Z" ;
if commune_lt = "92073" and apen = "851A" then apen2 = "8610Z" ;
if commune_lt = "92078" and apen = "751A" then apen2 = "8411Z" ;
if commune_lt = "94015" and apen = "922C" then apen2 = "5911A" ;
```

Annexe 2 : les tests pour établir la règle de décision

Tableau des zones d'emploi⁸ retenues pour le test

	ZE avec moins de 20 000 emplois	ZE entre 20 000 et 40 000 emplois	ZE entre 40 000 et 80000 emplois	ZE avec plus de 80 000 emplois
Taux imput2 inférieur < 10 %	Sartène (9417)	Bergerac (7205)	Avranches (2506)	Saint-Nazaire (5202)
Taux imput2 compris entre 10 % et 15 %	Mont de Marsan (7281)	Romorantin (2453)	Blois (2451)	Saint-Etienne (8213)
Taux imput 2 Supérieur > 15 %	Pays de Bray (2307)	Le Creusot (2664)	Evreux (2325)	Boulogne Billancourt (1133) Dunkerque (312)

Liste des codes NAF (nomenclature A129) retenus pour le test

C16Z : Travail du bois et fabrication d'articles en bois, à l'exception des meubles ; fabrication d'articles en vannerie et sparterie ;
 C28A : Fabrication de machines et d'équipements d'usage général ;
 C29B : Fabrication d'équipements automobiles ;
 C33Z : Réparation et installation de machines et d'équipements ;
 F42Z : Génie Civil ;
 F43Z : Travaux de construction spécialisés ;
 M71Z : Activités d'architecture et d'ingénierie ; activités de contrôles et analyses techniques ;
 P85Z : Enseignement ;
 Q88Z : Action sociale sans hébergement

Exemple de résultats pour la zone d'emploi de Romorantin

	Poste C33Z	Poste F42Z	Poste F43Z	Poste P85Z
Valeur moyenne	167	95	1671	1321
Valeur minimum	120	84	1656	1300
Valeur maximum	216	104	1684	1336
Coefficient de variation	10,3 %	3,2 %	5,2 %	0,45 %

⁸ Travail effectué avant la refonte des zones d'emploi en 2011

Annexe 3 : mode d'emploi de la macro SAS de calcul de précision

Le kit precision_RP, mis à disposition [le site intranet du pôle ISE](#) comporte les éléments suivants :

- Le programme SAS "precision_RP", qui contient la **macro** du même nom, ainsi qu'un exemple de lancement ; la macro, une fois convenablement paramétrée, fonctionne en faisant appel à deux autres programmes, ainsi qu'à huit tables de passage, outils également fournis dans ce kit.
- Une courte **notice** d'utilisation à consulter avant utilisation (voir ci-dessous) ;
- Les programmes SAS "**_m1_imputations_multiples.sas**" et "**_m2_cumul_et_stats.sas**", qui seront appelés par la macro ;
- Huit **tables de passage** de la NAF Rév 1 adaptée au RP vers huit niveaux de la NAF rév 2.

Objet de la macro

Cet outil accompagne la mise à disposition de la recodification en NAF rév 2 du RP 1999.

Il a pour vocation de mettre à disposition des statistiques (coefficient de variation, quartiles, ...) permettant d'apprécier la précision de l'estimation du nombre d'emplois, pour un niveau géographique et un niveau de nomenclature donnés.

Il procède en deux temps :

D'abord il impute 500 fois l'activité en NAF rév 2 à partir de la variable en NAF rév 1, en se basant sur les tables de passage par niveau de nomenclature fournies par le pôle ISE avec ce programme. (Attention, les observations comportant un code NAF rév 1 égal à 'ZZZZ' -inactifs ou actifs inoccupés- ne sont pas conservées dans la table en sortie).

Ensuite il propose en sortie une table SAS donnant pour chaque modalité de NAF rév 2 les statistiques obtenues suite à ces 500 imputations :

- nombre d'imputations (donc 500),
- nombre moyen d'emplois, écart-type, et **coefficient de variation**,
- nombre d'emplois minimum, et maximum,
- médiane, quartiles, déciles et centiles extrêmes,
- taux d'imputation, et **taux d'imputation « non presque sûr »**.

Pré-requis

L'utilisateur de l'outil doit disposer :

- du logiciel SAS
- d'une table SAS extraite du RP 1999 exploitation complémentaire, et comprenant au minimum les variables suivantes :
 - o commune du lieu de travail : cette variable devra être nommée : **COMMUNE_LT** ;
 - o poids de l'individu dans le RP99 exploitation complémentaire : cette variable devra être nommée : **IPONDIC** ;
 - o activité selon le niveau 700 de la NAF adaptée au RP : le nom de cette variable est libre, mais devra être indiqué en paramètre en entrée de la macro (par défaut : **NAF**) ;
 - o Ainsi que les variables utiles pour filtrer les données (telles REGION_LT, etc...)

Utilisation

ETAPE 1 - Renseigner les paramètres de la macro, dans le paragraphe « paramétrage par l'utilisateur de la macro "precision_RP" ». Les informations requises sont :

Le dossier (indiquer le chemin complet) où sont stockés les programmes sas appelés par le programme m0_précision.sas :

`l_chemin = D:\Mes Documents\RP99\programmes ,`

Le dossier où se trouve la table à imputer (indiquer son chemin complet) :

`l_in = D:\Mes Documents\RP99\tables ,`

Le nom de la table à imputer :

`l_tabin = RP99nafrev1 ,`

Le nom de la variable indiquant l'activité en naf rev1 dans la table à recodifier ('APEN', par défaut) :

`l_nafrev1 = naf ,`

Le dossier où devront être stockées les sorties (indiquer son chemin complet) . Ces sorties sont d'une part la table imputée 500 fois en NAF rev 2, et d'autre part la table comprenant les statistiques calculées à partir de la table imputée. Attention, la table imputée est volumineuse : exemple pour 10000 observations et 519 variables - 17 en entrée + 502 générées par la macro - elle occupe déjà 36 Mo).

`l_out = D:\Mes Documents\RP99\mes_sorties ,`

Le nom que devra porter la table en sortie, après imputation :

`l_tabout = RP99_recodifie_a5 ,`

Le dossier où se trouvent les huit tables de passage en NAF rév 2 (indiquer le chemin complet) :

`l_pass = D:\Mes Documents\RP99\tablespassage ,`

Le niveau de nomenclature retenu pour l'imputation (8 choix possibles : 5, 10, 17, 21, 38, 64, 88 ou 129) :

`l_nivo = 5 ,`

Le filtrage géographique, selon le lieu de travail ou le lieu de résidence. Le texte saisi sera repris tel quel dans une clause « WHERE » (ne pas oublier les parenthèses) :

`l_filtre = (region_lt in ("53","52","54","72")) ,`

Le nom (associé au filtrage géographique) à donner au fichier SAS de statistiques en sortie. Ce nom doit donc respecter les normes associées au nommage des tables SAS (pas d'espace, ni de caractères spéciaux) :

`l_nomzone = regions_cote_ouest ,`

Le choix d'affichage (ou non) des commentaires SAS dans la log. Indiquer « oui » ou « non » selon ses préférences. En raison des 500 itérations, les commentaires SAS peuvent être d'un volume conséquent. Par défaut, le programme les désactive. (Ce paramètre étant le dernier, il n'a pas à être suivi d'une virgule) .

`messages_sas = oui`

L'utilisateur veillera à ne modifier que le texte surligné.

ETAPE 2 - Lancer l'ensemble du programme. Le résultat du traitement se trouve dans la "WORK", dans la table SAS suffixée en "_stats", le nom complet de la table reprenant également le nom indiqué pour le paramètre l_nomzone, ainsi que le niveau de nomenclature (*Soit dans l'exemple ci-dessus : "regions_cote_ouest_A5_stats"*).

Exemple de lancement et d'utilisation

Voici un exemple d'appel de la macro :

```
%precision_RP(  
L_chemin=D:\Mes Documents\RP99\recodification\programmes,  
L_in= D:\Mes Documents\RP99\recodification\tables,  
L_tabin=RP99,  
L_nafrev1=naf,  
L_out= D:\Mes Documents\RP99\recodification\mes_sorties,  
L_tabout=Chateaubriant_a5,  
L_pass= D:\Mes Documents\RP99\recodification\t_passage,  
L_nivo=5,  
L_filtre=(commune_It in ("44036")),  
L_nomzone=Chateaubriant,  
messages_sas=non  
);
```

Voici la table Chateaubriant_a5_stats obtenue en sortie :

stat	_AZ	_BE	_FZ	_GU	_OQ
N	500	500	500	500	500
MIN	115	2329	288	2756	2454
MAX	118	2341	294	2777	2469
MEAN	116.974	2336.374	291.83	2764.956	2463.866
STD	0.8090941024	2.2503894119	0.9267241419	3.5506729724	2.7244980785
VAR	0.6546332665	5.064252505	0.8588176353	12.607278557	7.4228897796
CV	0.6916871291	0.0963197421	0.3175561601	0.1284169792	0.1105781759
P1	115	2331	289	2758	2456.5
P10	116	2333	291	2761	2460
Q1	116	2335	291	2763	2462
MEDIAN	117	2337	292	2765	2464
Q3	118	2338	292	2767	2466
P90	118	2339	293	2770	2467
P99	118	2340.5	294	2774	2468
IMPUT	4.2522269906	4.0393361679	15.019017921	2.2769259258	0.5221874891
IMPUT90	1.0446765948	0.6525496346	0.6784771956	0.2793534508	0.5206451974

Lecture :

Pour AZ (Agriculture, sylviculture et pêche), on lit que :

- le **coefficient de variation** (ligne « CV ») est de 0,69 %, et indique donc que l'estimation obtenue est tout à fait fiable (on considère qu'un coefficient de variation inférieur à 5% est correct).
- Le **taux d'imputation « non presque sûre »** (ligne « IMPUT90 »), également directement calculable à partir du fichier recodifié, est de 1,04%, ce qui signifie qu'à peine plus de 1% des emplois de l'agriculture, de la sylviculture et de la pêche, sont obtenus par une imputation affectée d'un aléa conséquent. Par ailleurs la ligne « IMPUT » indique quant à elle que, au total, 4,25% des emplois de cette activité sont obtenus par imputation (presque sûre, ou non presque sûre). (se référer également aux préconisations relatives au taux imput90 et au nombre d'emplois obtenus).