

LA QUALITE DU CODAGE DE L'ACTIVITE ET DE LA PROFESSION DANS LE RECENSEMENT DE LA POPULATION

Elodie MARTAL - Yves JACQUIN
Insee - Département de la Démographie

INTRODUCTION

Lors de chaque campagne annuelle de recensement, les questionnaires renseignés par les personnes recensées et collectés par les agents recenseurs sont saisis par un dispositif de lecture optique et font l'objet d'un vidéo-codage. L'Insee procède ensuite à une série de traitements statistiques, réalisés en deux volets : l'exploitation « principale » et l'exploitation « complémentaire ».

La première porte sur l'ensemble des questionnaires collectés (notamment pour toutes les questions de type cases à cocher) alors que la seconde n'est réalisée que sur un échantillon de bulletins individuels (BI)¹.

Le volet de l'exploitation statistique, dite « complémentaire », est destiné à produire les variables dont l'élaboration est plus complexe. Il s'agit de celles qui décrivent :

- la structure familiale des ménages : détermination précise de la personne de référence du ménage, identification, le cas échéant, de familles au sein du ménage et composition de ces familles ;
- les secteurs d'activité dans lesquels les emplois sont exercés ;
- la profession et la catégorie socio-professionnelle²

Les données du questionnaire de recensement :

Les questionnaires du recensement de la population comportent deux types de questions : les questions fermées et ouvertes.

Pour les questions fermées, le codage est implicitement proposé dans le questionnaire et ne nécessite donc aucun travail supplémentaire.

Par exemple, pour la question n°8 - Quel est votre état matrimonial légal ? (de type case à cocher), il suffit de coder la case qui a été cochée.

8 Quel est votre état matrimonial légal ? MATR

• Célibataire (<i>jamais légalement marié(e)</i>)	<input type="checkbox"/>	1
• Marié(e) (<i>ou séparé(e) mais non divorcé(e)</i>)	<input type="checkbox"/>	2
• Veuf, veuve	<input type="checkbox"/>	3
• Divorcé(e)	<input type="checkbox"/>	4

¹ L'échantillon destiné à l'exploitation complémentaire est constitué :

- pour les communes de 10 000 habitants ou plus : de l'ensemble des résidences principales recensées et de leurs habitants, de 25% des personnes recensées dans les communautés, ainsi que l'ensemble des personnes recensées en habitations mobiles et des personnes sans abri ;

- pour les communes de moins de 10 000 habitants : de 25 % des résidences principales recensées et de leurs habitants, de 25% des personnes recensées dans les communautés, ainsi que de 25 % des personnes recensées en habitations mobiles et des personnes sans abri ;

² Insee.fr : <http://www.insee.fr/fr/publics/communication/recensement/particuliers/doc/fiche-exploitation.pdf>

Pour les questions ouvertes, la codification est plus complexe : elle vise à traduire un ou plusieurs libellés en clair (récupérés dans les bulletins individuels de la campagne de recensement de la population) en un code dans une nomenclature officielle. Elle concerne :

- les libellés de pays et de nationalités des BI ;
- les libellés de commune des lieux de naissance, de travail, de résidence antérieure, d'étude ;
- l'activité économique de l'établissement employeur déclaré dans le BI (activité);
- les libellés de profession (actuelle pour les actifs et antérieure pour les inactifs).

29 **Quelle est votre profession principale ?**
*Soyez précis. Par exemple : « CAISSIÈRE » (et non « EMPLOYÉE »),
« CHEF DE SERVICE CLIENTÈLE » (et non « CADRE »).
Si vous êtes agent de la fonction publique d'État,
territoriale ou hospitalière, indiquez votre grade (corps,
catégorie, etc.).*

PROFS (profs_c_c)

[Annexe 1](#)
[Bulletin individuel](#)

Parmi les questions ouvertes, celles sur l'activité et la profession sont les plus complexes à coder car le libellé seul ne suffit pas : le codage doit faire appel à de nombreuses autres variables³ dites auxiliaires, comme par exemple le statut et la taille de l'entreprise.

En effet, alors que le taux de codage automatique réussi est proche de 100% pour le pays et la nationalité, il est de 80% pour la profession antérieure, 75% pour la profession actuelle et seulement 45% pour l'activité employeur. On constate que le pourcentage de codage automatique est plus important pour la profession antérieure que pour la profession actuelle : cela tient au fait que le codage repose sur moins d'information disponible, et donc moins de variables auxiliaires, et s'effectue donc souvent par défaut. Cela signifie que la qualité est certainement moindre a priori, ce qui sera confirmé dans notre étude.

Pour le codage des libellés, on utilise dans un premier temps l'outil SICORE⁴ qui repose sur une base de connaissance propre à la variable que l'on souhaite coder (cette base de connaissance est remise à jour tous les ans).

Cependant, SICORE ne permettant pas de coder automatiquement tous les libellés, les non codés sont alors repris par des gestionnaires à l'aide d'un outil de reprise en direction régionale. Ainsi, l'application RECAP (Recherche et codage de l'activité et de la profession) permet le codage de l'activité et de la profession pour plus de 90%⁵ des bulletins individuels n'ayant pu être codés lors du traitement automatique centralisé.

Chaque année, le codage de l'activité et de la profession concerne environ 1 800 000 individus : ce sont tous les actifs occupés d'une exploitation complémentaire annuelle⁶.

Une enquête qualité a été menée sur le codage (automatique et manuel) de l'activité et de la profession pour les campagnes de recensement 2006 à 2009. Elle s'appuie sur une version spécifique de l'application RECAP, dénommée RECAP-Qualité. Les opérations ont été organisées sur un échantillon d'environ 100 000 individus. Elles sont articulées en 2 phases : une phase de recodage, suivie d'une phase d'arbitrage par des experts.

³ La position professionnelle, le département géographique, la fonction, l'activité au niveau détaillé (classe NAF) et au niveau agrégé (division), le nombre de salariés, l'orientation des productions agricoles, la variable Public/Privé (à partir du SIRET et de la catégorie juridique de l'entreprise), le sexe, le statut, la taille de l'entreprise.

⁴ Système Informatisé de COdage des Réponses aux Enquêtes

⁵ Sur le champ des actifs occupés, le codage reste à blanc pour 7,2% des codes activité et 10,4% pour la profession actuelle par manque d'information sur le BI. Pour la profession antérieure, 12,5% des codes restent à blanc. Ces codes à blanc sont ensuite imputés lors de la phase de redressement.

⁶ Une enquête annuelle de recensement porte sur environ 9,5 millions d'individus, du fait du sondage à 8% dans les grandes communes (plus de 10 000 habitants) et du recensement d'une petite commune sur 5 chaque année. Sur cette population, la part des actifs occupés est d'environ 40%, et le ratio de l'exploitation complémentaire est d'environ 48%.

Le travail mené a consisté à exploiter les résultats de l'opération RECAP-Qualité, et notamment de déterminer les taux d'erreur de codage pour les variables 'activité', 'profession actuelle' et 'profession antérieure'.

La première partie de ce document a pour objectif de détailler la méthode utilisée, et la seconde présente les résultats des quatre campagnes d'enquêtes RECAP-Qualité.

1. La méthode utilisée

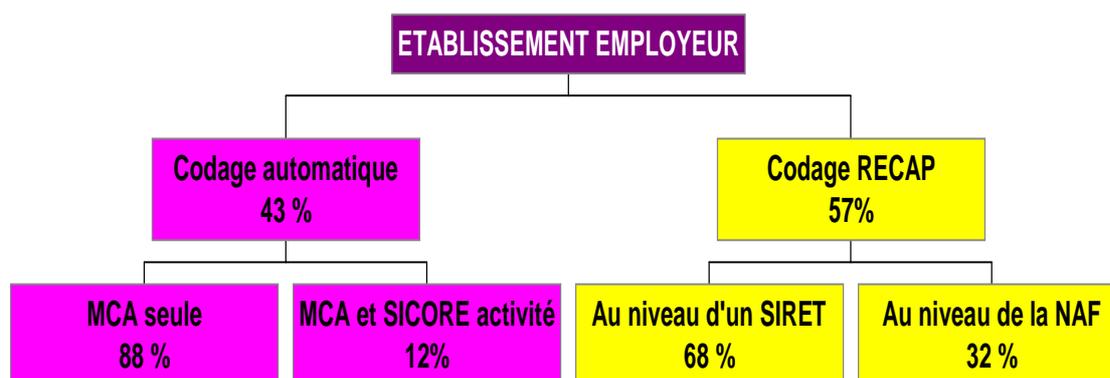
1.1. Processus de codage dans le recensement de la population

Pour l'enquête annuelle de recensement de l'année n, les codages automatiques de l'activité et de la profession commencent en mai (traitement automatique centralisé).

- Le traitement automatique effectué pour le codage de l'établissement employeur s'appuie sur une MCA (Mise en Concordance Automatique) inspirée de SIRENE. Elle compare les déclarations faites sur le BI aux questions 18 (Quel est le nom de l'établissement qui vous emploie ou que vous dirigez ?) et 20 (Quelle est l'adresse de votre lieu de travail ?) à un référentiel d'établissement extrait du répertoire SIRENE⁷ de l'INSEE. Ce traitement utilise aussi SICORE⁸ Activité. En effet, SICORE est utile pour confirmer des échos douteux retournés par la MCA en comparant l'activité de l'établissement écho et celles retournées par SICORE à partir du libellé d'activité (cf. infra).
- Le traitement réalisé pour le codage de la profession est SICORE Profession

En cas d'échec de ces traitements automatiques, les codages manuels prennent la suite de juin à octobre de l'année n : c'est la campagne dite **RECAP (RE**cherche et **C**odage de l'**A**ctivité et de la **P**rofession) qui se déroule dans chaque Direction Régionale⁹ (sauf en DOM). Chaque DR traite en priorité les bulletins individuels de son domaine géographique mais peut être amenée à en traiter d'autres.

Le codage est donc réparti en deux processus : automatique et reprise par RECAP. Pour la variable **activité - employeur**, le taux de codage est réparti ainsi entre les deux processus :



Taux de codage moyen 2005-2009 pour le codage de l'activité

Le processus automatique est divisé en deux sous-processus. Dans le premier, le rapprochement entre les données du BI et le référentiel extrait de SIRENE produit un code jugé de qualité satisfaisante (écho avec un risque d'erreur faible).

⁷ Système informatisé du répertoire national des entreprises et des établissements.

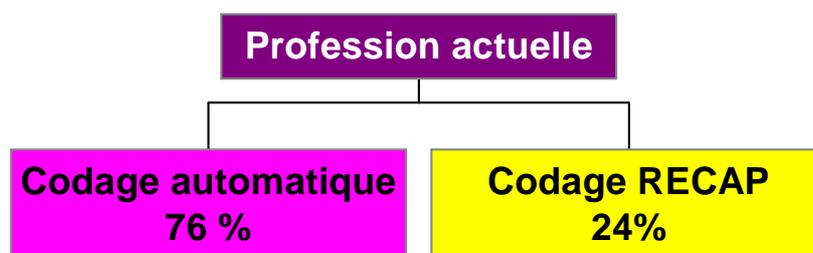
⁸ Système Informatisé de COdage des Réponses aux Enquêtes.

⁹ La campagne RECAP mobilise une centaine de gestionnaires en DR chaque année.

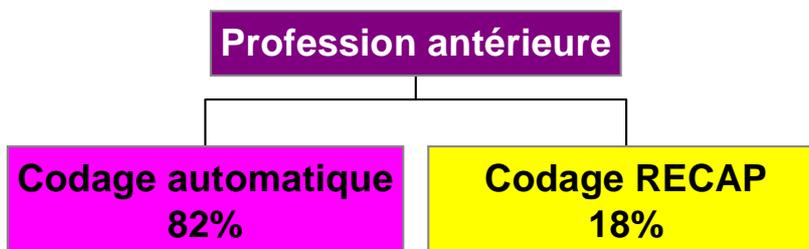
Dans le second, le code proposé par la MCA est incertain. Sachant que la question 19 du BI (Quelle est l'activité de cet établissement ?) est codée par SICORE Activité, on compare le code NAF¹⁰ que SICORE retient avec le code NAF obtenu par l'écho de la MCA.

Le processus reprise dans RECAP est également éclaté en deux sous-processus. Dans le premier, le gestionnaire retrouve l'employeur et donc le SIRET. Dans le second, il code seulement l'activité - employeur au niveau NAF (grâce à la question 19).

Pour le codage de la **profession actuelle et antérieure**, la répartition est la suivante :



Taux de codage moyen 2005-2009 pour le codage de la profession actuelle



Taux de codage moyen 2005-2009 pour le codage de la profession antérieure

1.2. Mesure de la qualité du codage : RECAP Qualité

Comment étudier la qualité du codage ?

L'opération RECAP Qualité est une enquête permettant de vérifier la qualité du codage de l'activité et de la profession réalisé par les gestionnaires des DR mais aussi celui effectué par les processus automatiques (MCA et SICORE). Elle a été mise en place pour l'enquête annuelle de recensement de 2006 et renouvelée chaque année depuis (sauf en 2010 pour raisons budgétaires).

Le principe

1^{ère} étape : Un deuxième codage est réalisé par environ 9 DR « volontaires » entre novembre et décembre de l'année n, il est dénommé "codage qualité".

Dans les conditions habituelles de la campagne RECAP, les gestionnaires de RECAP Qualité recodent un échantillon de bulletins individuels qui l'ont déjà été dans la campagne RECAP. Les BI traités initialement en automatique ou en manuel sont mélangés, ainsi que les provenances géographiques. En outre, le recodeur ne connaît pas le premier codage.

Les échantillons sont représentatifs des types de communes (grandes et petites) et des codes de nomenclatures (NAF pour l'activité et PCS¹¹ pour la profession).

[Annexe 2](#)
[Les échantillons de RECAP Qualité](#)

¹⁰ Nomenclature d'activités françaises : NAF révision 1 (datant de 2003) pour les premières années de RECAP Qualité (2006, 2007 et 2008) puis NAF révision 2 (datant de 2008) pour l'année 2009

¹¹ Nomenclature des Professions et Catégories Socioprofessionnelles (PCS 2003)

2^{ème} étape : Un arbitrage est ensuite réalisé en janvier et février de l'année n+1 par trois DR. C'est un troisième codage. L'arbitre connaît les deux premiers codages attribués et peut choisir entre ces deux derniers ou proposer un troisième codage différent. Le temps consacré à l'examen du BI est plus important que dans la campagne RECAP. On considère que ce code arbitré (troisième codage) est correct.

Ainsi, un codage est considéré comme bon si le codage arbitré est égal au codage initial (car quoi qu'il en soit c'est bien le codage initial qui a été retenu dans la chaîne du RP : il n'y a pas de retour direct de l'opération qualité sur les données finales du RP¹²). Dans le cas contraire, le codage sera considéré comme en erreur. Mais cela dépend bien sûr du niveau de nomenclature auquel on se situe. Un codage pourra en effet être en erreur au niveau détaillé de la NAF (respectivement de la PCS), mais être bon au niveau plus agrégé de la division de la NAF (respectivement de la catégorie socioprofessionnelle - CS - en 42 postes de la PCS)

[Annexe 5](#)
[Les nomenclatures activité et profession](#)

Si le premier et le deuxième codages sont identiques, on parle de convergence.

La convergence est entendue au niveau le plus fin de la nomenclature :

- pour la profession, la convergence s'évalue donc au niveau PCS
- pour l'activité, il s'agit de la convergence au niveau du SIRET¹³

Dans la campagne RECAP Qualité, on expertise tous les codages divergents et 10% des codages convergents (sauf en 2006 où les codages convergents n'étaient pas arbitrés, ce qui revient à les considérer comme bons).

L'objectif de cette étude a été d'exploiter les quatre enquêtes qualité menées de 2006 à 2009 et donc de déterminer :

- des taux d'erreur de codage de l'activité, de la profession antérieure et de la profession actuelle
- l'évolution de ces taux sur la période considérée

1.3. Méthodologie mise en œuvre

1.3.1. Les objectifs

Les premiers objectifs de l'étude ont été très rapidement déterminés, à savoir calculer les taux de mauvais codage selon différentes variables.

Variable	Modalités
Qualité (ou codage 2)	<ul style="list-style-type: none"> - codages divergents (codes initial et arbitré différents) - codages convergents
Variable codée	<ul style="list-style-type: none"> - activité (de l'employeur) - profession actuelle - profession antérieure
Processus de codage	<ul style="list-style-type: none"> - automatique - reprise manuelle

1.3.2. La mise en œuvre

¹² Dans cet ordre d'idée, le codage initial est également désigné sous l'appellation de « codage décidé » dans les graphiques infra et en annexe.

¹³ ou en NAF lorsque les SIRET ne sont renseignés pour aucun des 2 codages

Avant toute chose, il est apparu que les professions salariée et indépendante pouvaient être toutes les deux renseignées (normalement à tort) sur un même BI. En effet, selon que la personne recensée est salariée ou non, elle remplit différentes parties du questionnaire, et la profession apparaît donc 2 fois dans le BI (questions 25 et 29). Dans la campagne RECAP Qualité, une règle a donc été mise en œuvre : le libellé profession salariée prime et on ne teste donc pas la qualité du code profession indépendante lorsque les deux professions sont renseignées. Cette règle a permis de créer une variable unique profession actuelle.

En second lieu, afin de faciliter les calculs, des variables ont été créées à partir des tables SAS, pour tout ce qui concerne la mise en concordance des codages 1 et 2 (convergence / divergence) puis des codages 1 et 3 (bon codage / mauvais codage).

- Plusieurs particularités sont à noter pour l'anquête qualité **2006** :
 - tous les codages convergents étaient considérés comme bons et donc non arbitrés
 - aucun arbitrage n'a été réalisé pour la profession antérieure

- Pour l'année **2007**, tous les cas n'ont pas pu être arbitrés par manque de temps : un seul cas n'a pas pu être arbitré pour la profession actuelle. Pour l'activité, c'est le cas de 6 920 codes et de 2 051 codes pour la profession antérieure : soit respectivement 14% et 28% de l'échantillon qualité. La répartition de ces cas par la variable qualité distinguant les codes convergents des codes divergents a été comparée à la répartition des cas arbitrés, ce qui donne :

ACTIVITE en 2007		
En %	Codes convergents	Codes divergents
Cas non arbitrés	11,65	88,35
Cas arbitrés	11,38	88,61

PROFESSION ANTERIEURE en 2007		
En %	Codes convergents	Codes divergents
Cas non arbitrés	3,46	96,54
Cas arbitrés	3,14	96,86

Au vu de ces résultats, on peut considérer qu'il n'y a pas de biais, c'est à dire que l'on considère que les taux d'erreur de codage des cas non arbitrés sont « similaires » à ceux des cas arbitrés.

- En 2006, 2007, et 2008, la nomenclature d'activité utilisée est la NAF, révision 1 (révision de 2003). En **2009**, c'est la NAF révision2 (révision de 2008).

Ces particularités notées, trois grandes étapes ont été nécessaires afin de calculer les taux d'erreur de codage :

- calculer le taux d'erreur de codage dans l'échantillon arbitré, et ceci par variable, par caractéristique convergent/divergent, par année et par processus (et sous-processus pour l'activité).
- extrapoler ces taux à tous les individus de l'échantillon, en veillant à rassembler convergents et divergents
- extrapoler cette deuxième série de taux à la population totale, dans laquelle on connaît la part des codages par processus automatiques et manuels.

2. Les résultats

2.1. Les taux d'erreur de codage

Dans cette partie, les résultats présentés sont ceux de l'enquête qualité 2009 - Les résultats des autres années sont présentés en annexes.

2.1.1. L'activité

Cette étape de l'étude a consisté à calculer les taux d'erreur de codage, selon les trois étapes énoncées dans la partie méthodologie.

ACTIVITE (NIVEAU DIVISION) - Enquête qualité 2009

ECHANTILLON

Processus de codage		Taille échantillon	Taille échantillon arbitré		Taille échantillon	Taille échantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de l'activité %	
Codage automatique	Automatique	6 000	1 442	Convergents	5 064	506	1	0,2	8,1
				Divergents	936	936	476	50,9	
	Automatique + SICORE	6 000	1 520	Convergents	4 977	497	3	0,6	2,9
				Divergents	1 023	1 023	144	14,1	
Codage manuel	Recap avec choix SIRET	6 000	2 461	Convergents	3 932	393	1	0,3	5,5
				Divergents	2 068	2 068	320	15,5	
	Recap avec choix d'un code NAF	33 079	19 481	Convergents	15 095	1 497	62	4,1	17,8
				Divergents	17 984	17 984	5 274	29,3	
Total	des convergents				29 068	2 893	67		
	des divergents				22 011	22 011	6 214		
TOTAL :					51 079	24 904	6 281		

POPULATION TOTALE

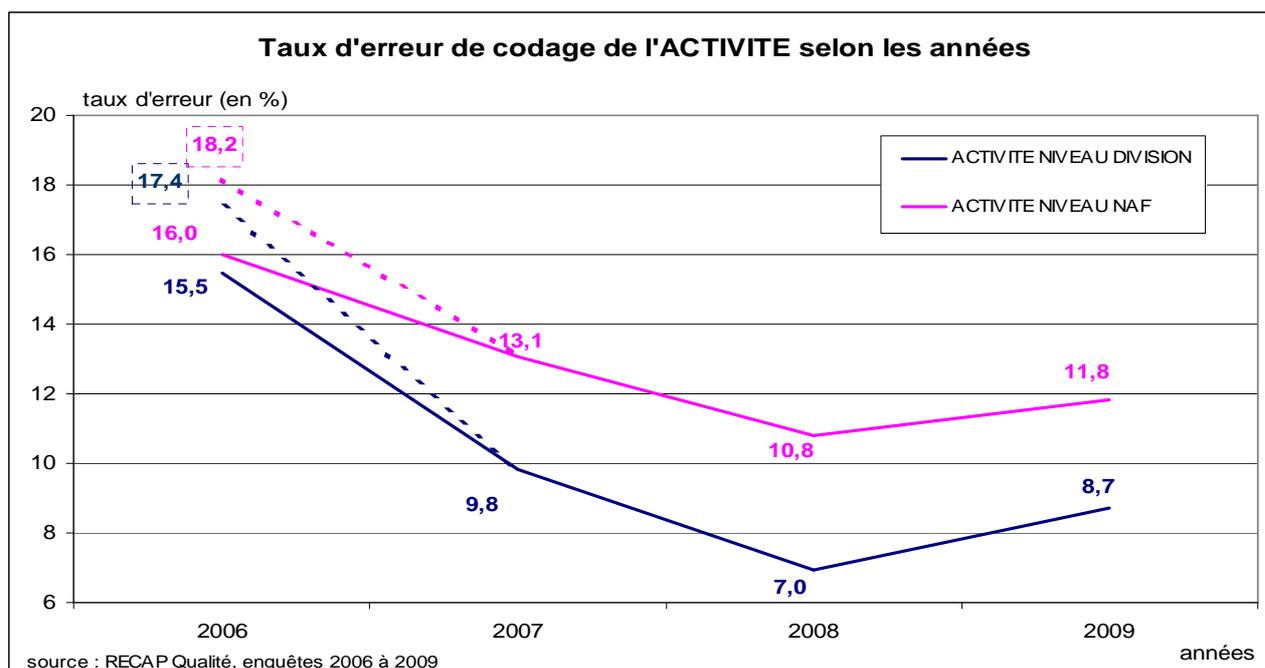
Processus de codage		Population		Taux d'erreur de codage de l'activité par sous-processus%	Taux d'erreur de codage de l'activité par processus%	Taux d'erreur global %
Codage automatique	Automatique	746 035	41,2%	8,1	7,8	8,7
	Automatique + SICORE	51 849	2,9%	2,9		
Codage manuel	Recap avec choix SIRET	687 395	37,9%	5,5	9,5	
	Recap avec choix d'un code NAF	326 088	18,0%	17,8		

1 811 367

La même démarche a été réalisée pour les autres années, ce qui a permis de construire le graphique ci-dessous.

Les tableaux synoptiques des années 2006 à 2008 sont regroupés en annexe 3.

[Annexe 3](#)
[Tableaux synoptiques](#)



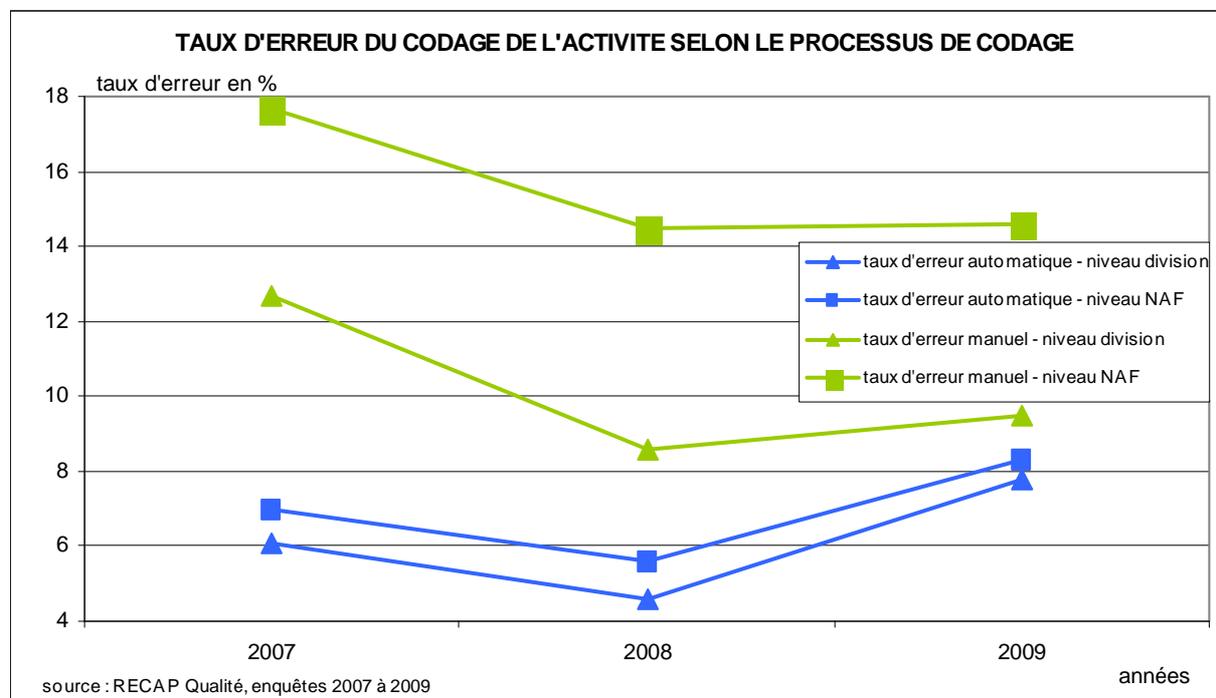
En 2006, les taux d'erreur de codage des codes convergents étaient considérés comme nuls. Les calculs effectués sur les années suivantes prouvant que ce n'est pas le cas, les taux d'erreur de 2006 ont été réévalués afin qu'ils soient comparables aux autres années. Les taux d'erreur des codes convergents de 2007 ont ainsi été appliqués en 2006, d'où les courbes en pointillés.

La tendance de ces deux courbes est nette : le taux d'erreur de codage de l'activité diminue (-6,4 points au niveau NAF et -8,7 points au niveau division). La différence de niveau des deux courbes est due au caractère plus ou moins agrégé de la nomenclature NAF.

En 2009, une légère hausse du taux d'erreur intervient. On peut penser que la nomenclature NAF révision 2, utilisée pour la première fois, a pu entraîner des difficultés d'appropriation de la part des gestionnaires. Mais ce sont surtout les processus automatiques qui sont moins bons qu'en 2008 (mais

meilleurs qu'en 2007), ce qui peut être considéré comme une contrepartie de taux de codages automatiques supérieurs (on est passé grosso modo de 42% à 44% en taux de codage automatique pour l'activité), liée à la refonte de l'algorithme de recherche par la MCA qui a accompagné le changement de NAF. Il peut donc s'agir d'une petite baisse de la qualité au profit de la quantité.

Les taux d'erreur de codage de l'activité par processus (automatique ou reprise) sont les suivants :



En 2007, la différence entre taux d'erreur automatique et manuel est très nette : on note une différence de 6,6 points au niveau division. En 2009, l'écart se resserre nettement et la différence n'est plus que de 1,7 points au niveau division. Les taux d'erreurs manuels impactent donc de moins en moins le niveau des taux d'erreurs globaux.

2.1.2. La profession actuelle

PROFESSION ACTUELLE (NIVEAU CS en 42 postes) - Enquête qualité 2009

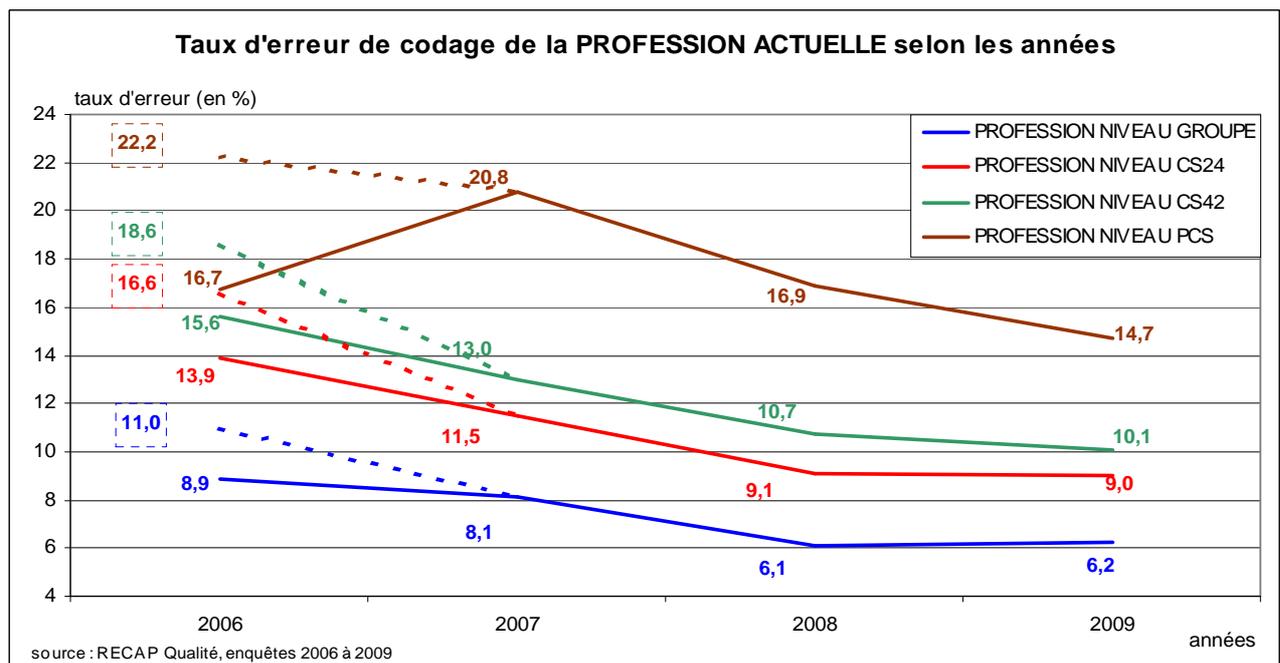
ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession %	
Codage automatique	6 000	1 642	Convergen	4 842	484	8	1,7	7,2
			Divergen	1 158	1 158	354	30,6	
Codage manuel	33 593	17 579	Convergen	17 791	1 777	59	3,3	19,1
			Divergen	15 802	15 802	5 837	36,9	
Total :	des convergen			22 633	2 261	67		
	des divergen			16 960	16 960	6 191		
TOTAL :				39 593	19 221	6 258		

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
	Codage automatique	1 329 555	75,6%	
Codage manuel	429 801	24,4%	19,1	

TOTAL 1 759 356



Au niveau groupe de la PCS, le taux d'erreur passe de 9% en 2006 à 6,2% en 2009, soit une baisse de pratiquement 3 points.

Plus le niveau de la nomenclature est détaillé, plus la baisse est forte entre 2006 et 2009 : -7,6 points pour le niveau PCS.

Cette baisse peut être interprétée comme l'amélioration du niveau d'expertise des gestionnaires de reprise en région. Aux niveaux plus agrégés, le taux d'erreur se stabilise en 2008.

Il convient de noter que les erreurs se compensent en grande partie au niveau agrégé car les erreurs de codages sont peu visibles sur les distributions (cf partie II/B)

2.1.3. La profession antérieure

Le codage de la profession antérieure est nécessairement plus fragile car on possède, contrairement à la profession actuelle, de nettement moins de variables auxiliaires. De plus, la taille de l'échantillon utilisé pour l'arbitrage de la profession antérieure est plus de 5 fois inférieure à celle de l'échantillon pour la profession actuelle, les résultats sont donc forcément moins fiables.

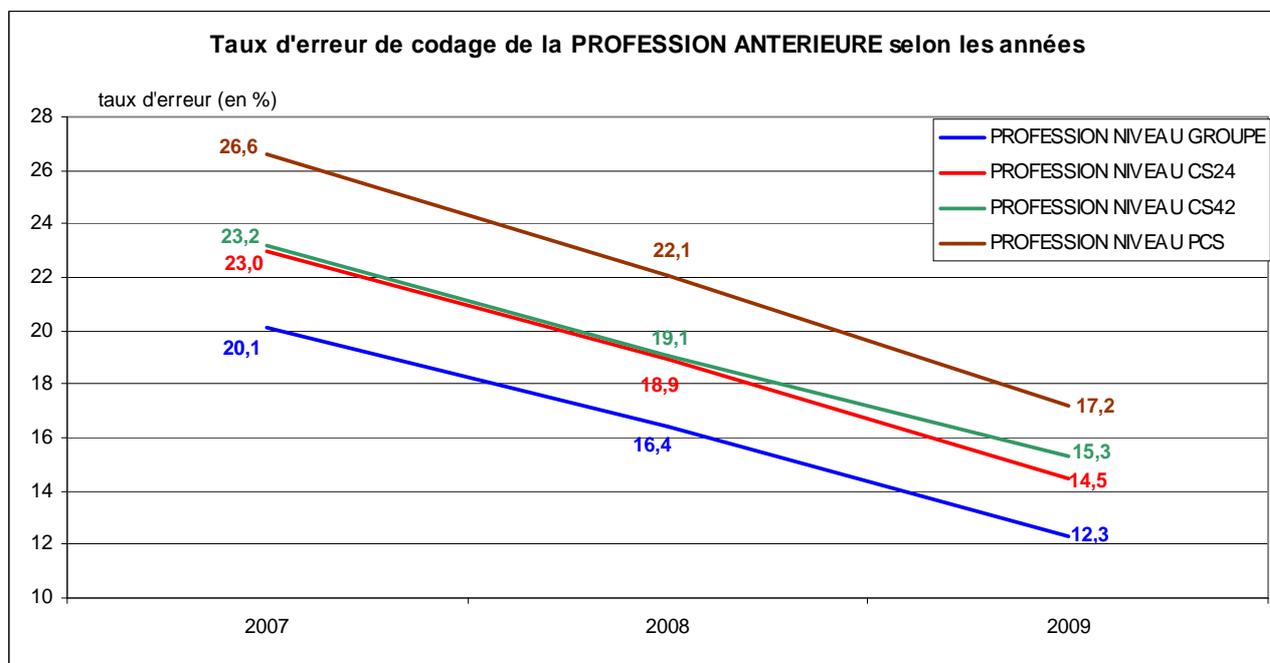
ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession au niveau groupe %	
Codage automatique	1 000	269	Convergen	812	81	2	2,5	9,6
			Divergen	188	188	76	40,4	
Codage manuel	6 333	3 829	Convergen	2 574	70	7	10,0	41,8
			Divergen	3 759	3 759	2 392	63,6	
Total :	des convergen			3 386	151	9		
	des divergen			3 947	3 947	2 468		
TOTAL :				7 333	4 098	2 477		

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
Codage automatique	1 036 415	82,3%	9,6	15,3
Codage manuel	222 934	17,7%	41,8	

TOTAL 1 259 349



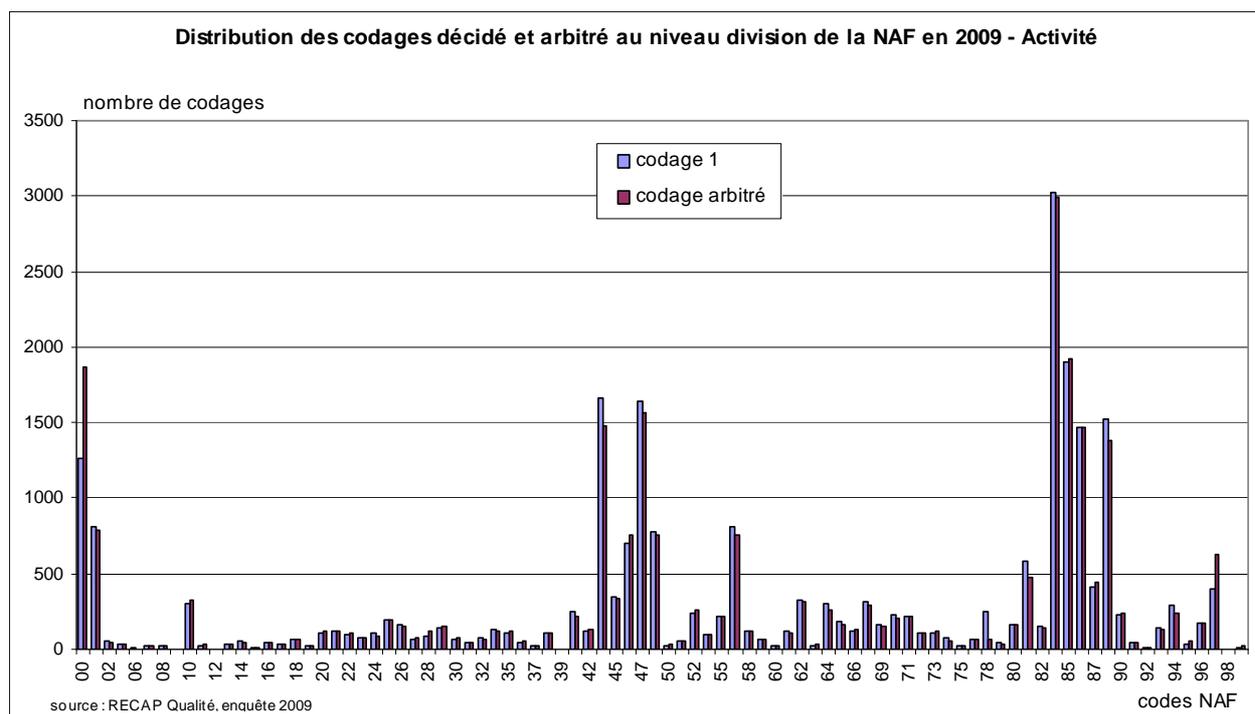
Les taux d'erreur diminuent très fortement entre 2006 à 2009 : de 9,4 points au niveau PCS et de près de 8 points au niveau groupe. Le niveau du taux d'erreur reste néanmoins plus important que pour la profession actuelle.

2.2. Distribution des codages initial et arbitré

Un autre contrôle a consisté à comparer les deux distributions brutes du codage initial et du troisième codage (codage arbitré) afin d'identifier les éventuels « postes fragiles ». Ce contrôle a été réalisé pour chacune des variables activité, profession actuelle, et profession antérieure et par année. Il complète la partie précédente en donnant une analyse sur l'ensemble du spectre des activités et professions.

2.2.1. L'activité

Il convient de noter la similitude de niveaux entre les deux distributions pour toutes les années (voir annexe 4 pour 2006, 2007 et 2008). Par exemple, en 2009 :



La modalité 00 correspond aux codages à blanc¹⁴.

[Annexe 4](#)
[Distribution des codages initial et arbitré](#)

¹⁴ Dans l'échantillon 'activité' arbitré, le taux de codes initiaux restés à blanc est de 17,6% en 2006, 6,1% en 2007, 5,8% en 2008 et 5,1% en 2009.

On observe toutefois des différences plus importantes entre les deux distributions pour les postes de la nomenclature suivants (niveau division) :

2006 (NAF révision 1)	2007 NAF révision 1)	2008 (NAF révision 1)	2009 (NAF révision 2)
<ul style="list-style-type: none"> - Sylviculture, exploitation forestière, services annexes (division 02) - Pêche, aquaculture, services annexes (division 05) - Extraction d'hydrocarbures, services annexes (division 11) - Autres industries extractives (division 14) - Industries textile (division 17) - Industries du cuir et de la chaussure (division 19) - Travail du bois et fabrication d'articles en bois (division 20) - Industries du caoutchouc et des plastiques (division 25) - Production et distribution d'électricité, de gaz et de chaleur (division 40) - Commerce et réparation automobile (division 50) - Commerce de gros et intermédiaires du commerce (division 51) - Transports par eau (division 61) - Intermédiation financière (division 65) - Assurance (division 66) - Administration publique (division 75) - Activités indifférenciées des ménages en tant que producteurs de Biens pour usage propre (division 96) - Activités indifférenciées des ménages en tant que producteurs de Services pour usage propre (division 97) 	<ul style="list-style-type: none"> - Sylviculture, exploitation forestière, services annexes (division 02) - Pêche, aquaculture, services annexes (division 05) - Autres industries extractives (division 14) - Industries de l'habillement et des fourrures (division 18) - Travail du bois et fabrication d'articles en bois (division 20) - Fabrication de machines de bureau et de matériel informatique (division 30) - Assurance (division 66) - Activités indifférenciées des ménages en tant que producteurs de Services pour usage propre (division 97) 	<ul style="list-style-type: none"> - Extraction d'hydrocarbures, services annexes (division 11) - Extraction de minerais d'uranium (division 12) - Industries du cuir et de la chaussure (division 19) 	<ul style="list-style-type: none"> - Extraction d'hydrocarbures (division 06) - Collecte et traitement des eaux usées (division 37) - Services d'information (division 63) - Activités liées à l'emploi (division 78) - Réparation d'ordinateurs et de biens personnels et domestiques (division 95) - Activités des ménages en tant qu'employeurs de personnel domestique (division 97)

A noter : les postes identifiés ici comme fragiles correspondent aux postes de la nomenclature :

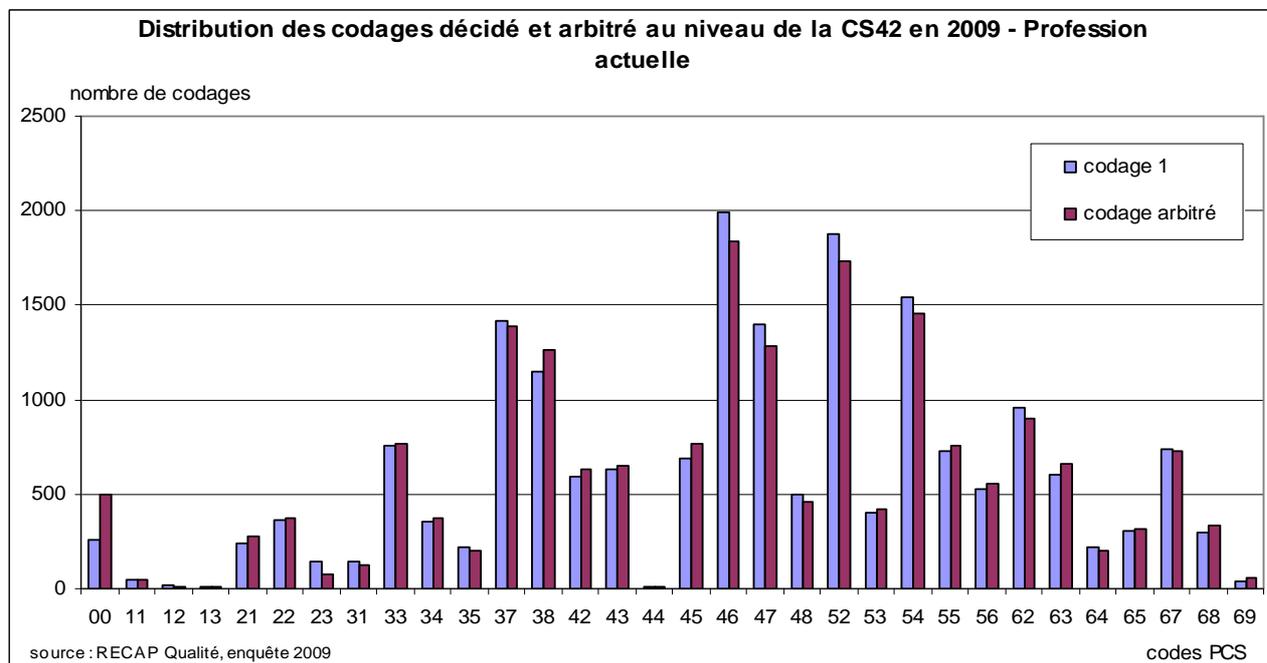
- dont la différence relative entre l'effectif initial et l'effectif arbitré est supérieure à **30%** ;
- dont l'effectif dans l'échantillon arbitré est au moins de 5 individus.

Certains postes fragiles sont les mêmes d'une année sur l'autre (extraction d'hydrocarbures, sylviculture, pêche,...), avec toutefois une diminution du nombre de postes fragiles entre 2006 et 2008, ce qui va dans le sens de l'amélioration du codage observée par ailleurs. Pour 2009, c'est différent avec l'introduction de la nouvelle nomenclature NAF révision 2 utilisée pour le codage du recensement, et de nouveaux postes fragiles apparaissent (fabrication de boissons, activités liées à l'emploi, ...). On note que tous les postes fragiles de 2009 (sauf l'extraction d'hydrocarbures) sont des postes créés avec le changement de nomenclature.

2.2.2. La profession actuelle

Là encore, les deux distributions au niveau de la CS en 42 postes sont assez semblables et ceci pour toutes les années.

Par exemple, en 2009 :



Les postes ressortant comme les plus fragiles sont les suivants :

2006	2007	2008	2009
<ul style="list-style-type: none"> - Agriculteurs sur petite exploitation (CS 11) - Agriculteurs sur moyenne exploitation (CS 12) - Artisans (CS 21) - Chefs d'entreprise de 10 salariés ou plus (CS 23) - Professions libérales (CS 31) - Clergé, religieux (CS 44) - Employés administratifs d'entreprise (CS 54) - Chauffeurs (CS 64) 	<ul style="list-style-type: none"> - Agriculteurs sur grande exploitation (CS 13) - Chefs d'entreprise de 10 salariés ou plus (CS 23) - Employés administratifs d'entreprise (CS 54) 	<ul style="list-style-type: none"> - Chefs d'entreprise de 10 salariés ou plus (CS 23) 	<ul style="list-style-type: none"> - Agriculteurs sur moyenne exploitation (CS 12) - Chefs d'entreprise de 10 salariés ou plus (CS 23) - Ouvriers agricoles (CS 69)

On

A noter : les postes identifiés ici comme fragiles correspondent aux postes de la nomenclature :

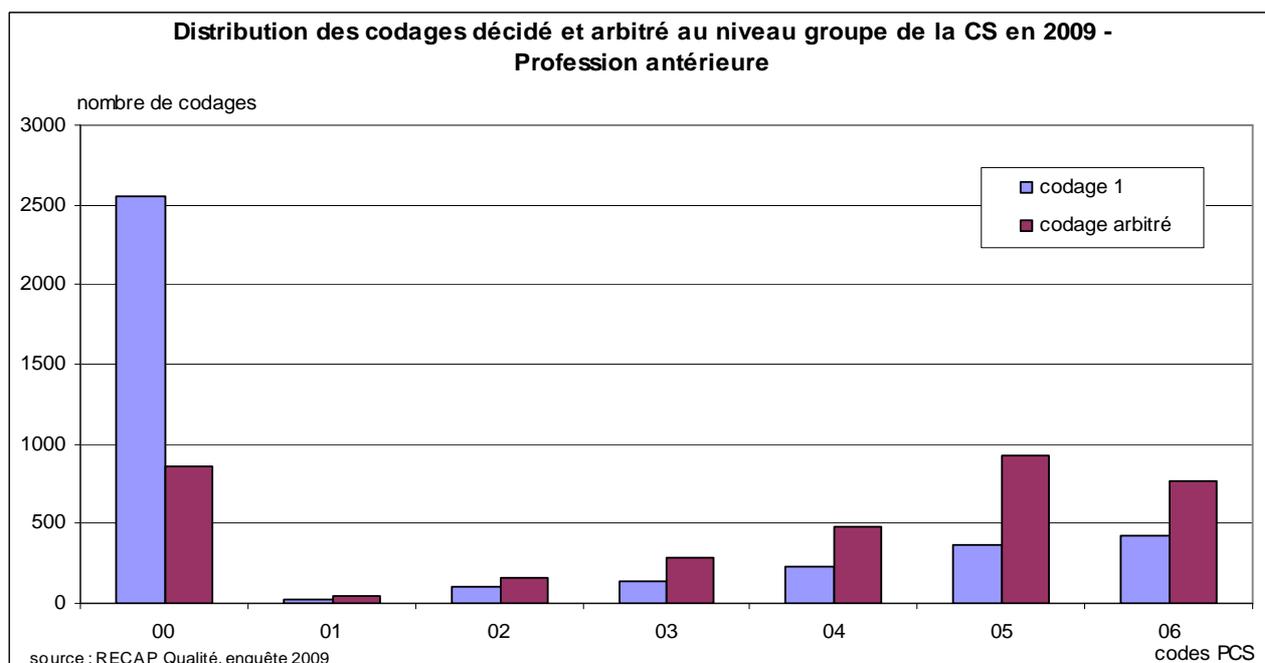
- dont la différence relative entre l'effectif initial et l'effectif arbitré est supérieure à **20%** (seuil retenu plus faible que pour l'activité, justifié par un moins grand nombre de poste - 31 postes d'actifs pour la CS versus 88 postes pour les divisions de l'activité), et donc un effectif moyen par poste plus élevé) ;
- dont l'effectif dans l'échantillon arbitré est au moins de 5 individus.

On ne retrouve pas exactement les mêmes catégories socioprofessionnelles d'une année sur l'autre. On observe néanmoins que la catégorie des Chefs d'entreprise de 10 salariés ou plus (CS 23) figure dans la liste pour les 4 campagnes.

Pour toutes les années (cf annexe 4), le nombre de codes à blanc (00) est toujours supérieur pour le codage arbitré que pour le codage initial¹⁵. Concrètement, cela signifie que des libellés profession sont codés dans le recensement alors qu'ils ne devraient pas l'être (par manque d'informations ou ambiguïté)¹⁶.

2.2.3. La profession antérieure

Au vu des échantillons de taille plus faible pour la profession antérieure, les distributions sont trop fragiles au niveau CS 42, elles sont donc comparées au niveau groupe. Par exemple en 2009 :



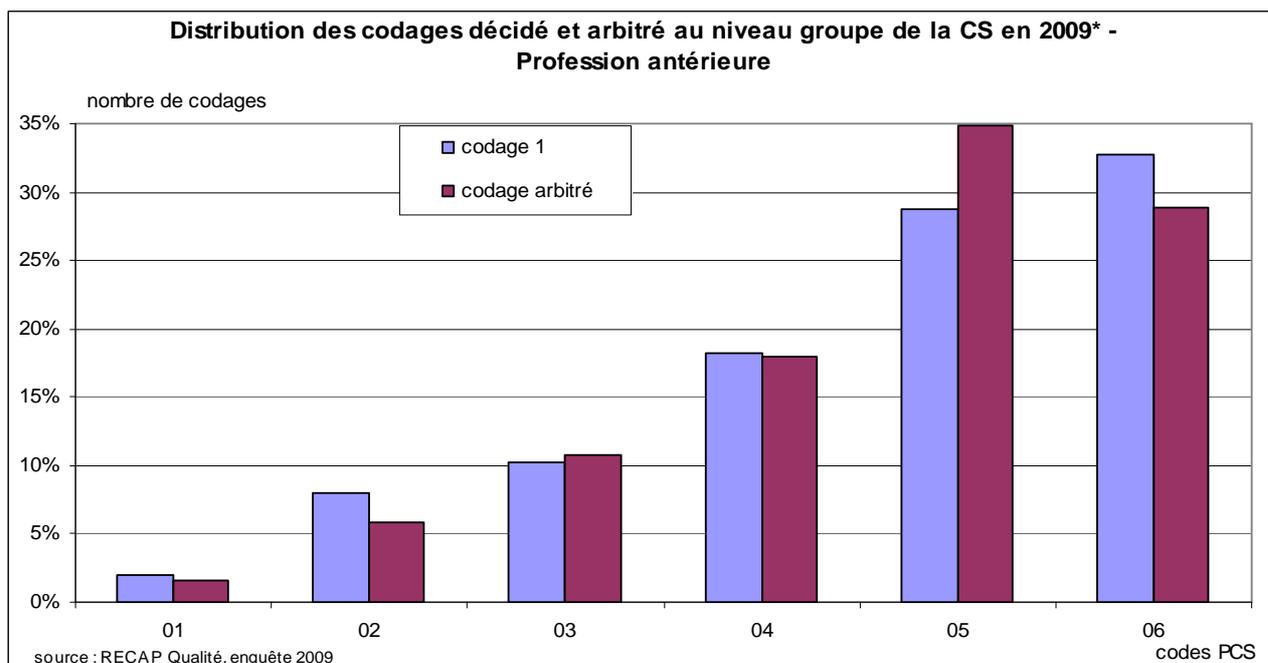
On note particulièrement le nombre important de codages à blanc lors du codage initial, qui diminue fortement lors de l'arbitrage. Le phénomène observé est donc inverse de celui qui prévaut pour les 2 autres variables : l'expert arbitre arrive à coder ce qui ne l'était pas en phase initiale. Il faut préciser que pour la profession antérieure, les codages sont possibles au niveau CS 42, contrairement à la profession actuelle.

En 2009, le taux de codes initiaux restant à blanc dans l'échantillon est de 62,2% contre 21,1% pour les codes arbitrés¹⁷. Pour cette raison, les mêmes distributions ont été réalisées en structure et en ne tenant pas compte des codages à blanc sur le graphique qui suit :

¹⁵ Dans l'échantillon 'profession actuelle' arbitré, les codages initiaux restant à blanc sont de 3% en 2006, 0,3% en 2007 et 2008 et 1,3% en 2009. Le taux de codages à blanc augmente lors de l'arbitrage : autour de 5,5% en 2006 et 2007 et de 2,5% en 2008 et 2009.

¹⁶ C'est aussi vrai pour l'activité, l'année 2006 mise à part. Par exemple en 2009, les codages initiaux restant à blanc représentent 5,1% contre 7,5% pour les codages arbitrés.

¹⁷ En 2007, on passe de 63,8% à 35,9%. En 2008, de 64,6% à 32%.



* en pourcentage et or codages à blanc

Un nombre important de codages à blanc sont finalement codés *in fine* parmi les employés (groupe 5 de la CS) : on passe de 369 cas en codage initial à 926 en codage arbitré. Parmi eux, ce sont les postes des « Personnels des services directs aux particuliers » (CS 56), des « Policiers et militaires » (CS 53), et des « Employés civils et agents de service de la fonction publique » (CS 52) qui comptabilisent le plus de différence entre les deux distributions.

2.3. L'influence régionale

Le taux d'erreur de codage manuel a été calculé pour chaque Direction Régionale afin de faire apparaître une éventuelle **influence régionale**.

Distribution des taux d'erreur de codage manuel de l'activité (niveau division) par DR				
	2006	2007	2008	2009
Maximum	41,9%	25,3%	18,4%	19,9%
3eme quartile	34,0%	23,6%	15,6%	18,1%
Mediane	32,3%	20,5%	14,4%	17,0%
1er quartile	30,6%	18,2%	13,1%	16,7%
Minimum	25,6%	15,5%	11,8%	13,1%

La distribution se « resserre » au fil des années : entre la DR qui a le plus faible taux d'erreur et celle qui a le plus fort, l'écart est de 16,3 points en 2006 et de 6,8 points en 2009.

Le taux d'erreur augmente légèrement en 2009 mais il ne faut pas oublier le changement de nomenclature.

En 2009, la moitié des DR ont un taux d'erreur compris entre 16,7% et 18,1%.

Distribution des taux d'erreur de codage de la profession actuelle (niveau groupe) par DR				
	2006	2007	2008	2009
Maximum	23,4%	18,3%	16,0%	14,5%
3eme quartile	17,1%	15,1%	12,2%	12,5%
Mediane	15,6%	13,2%	11,4%	11,6%
1er quartile	13,9%	12,2%	10,4%	10,4%
Minimum	12,2%	10,7%	7,8%	9,2%

Pour la profession actuelle, la même remarque peut être faite sur l'écart de la distribution : entre la DR qui a le plus faible taux d'erreur et celle qui a le plus fort, l'écart est de 11,2 points en 2006 et de 5,3 points en 2009.

En 2009, la moitié des DR ont un taux d'erreur compris entre 10,4% et 12,5%, soit entre deux points d'écart.

Distribution des taux d'erreur de codage de la profession antérieure (niveau groupe) par DR				
	2006	2007	2008	2009
Maximum		68,3%	70,0%	61,8%
3eme quartile		32,3%	41,7%	41,6%
Mediane		28,8%	36,8%	37,6%
1er quartile		24,2%	31,1%	32,9%
Minimum		15,9%	26,5%	26,1%

Les taux d'erreur de codage de la profession antérieure sont très importants pour certaines DR : 61,8% en 2009, 70% en 2008, ... Pour les trois années, les taux maximums sont à attribuer à trois DR différentes.

En 2007, l'écart est de 52,4 points entre la DR qui a le plus fort taux d'erreur et celle qui a le plus faible. En 2009, cet écart reste important - 35,7 points - et l'ajustement se fait plutôt vers le haut. Alors que le taux d'erreur le plus faible est de 15,9% en 2007, il passe à 26,1% en 2009.

A noter à nouveau toutes les précautions qu'il convient de prendre pour analyser les taux d'erreur de codage de la profession antérieure : moins de variables auxiliaires que pour la profession actuelle et un échantillon de taille 5 fois moins importante, donc des résultats plus fragiles.

Au vu de ces seuls résultats et la profession antérieure mise à part, on note la baisse des taux d'erreur de codage manuel entre 2006 et 2009, ce qui confirme l'amélioration du niveau d'expertise des gestionnaires de reprise en région.

CONCLUSION

Les quatre enquêtes annuelles RECAP-Qualité ont permis d'analyser les distributions macro des codages initial et arbitré. Au niveau de ces statistiques globales (cf partie II/B/), les distributions sont assez proches, la profession antérieure mise à part.

Toutefois, plus on descend à un niveau d'étude fin, plus les taux d'erreur de codage sont importants. Pour exemple, le taux d'erreur de codage de l'activité est supérieur de 3 points au niveau détaillé de la NAF par rapport au niveau division.

Les taux d'erreur diminuent régulièrement depuis 2006. En niveau NAF, le taux d'erreur de codage de l'activité diminue de 6,4 points entre 2006 et 2009. Pour la profession actuelle en niveau CS 42, on note une baisse de 8,5 points du taux d'erreur de codage.

Les taux d'erreur du codage de l'activité et de la profession actuelle sont, à tout niveau, inférieurs à 15% en 2009, soit 85% de codages initiaux corrects.

Quelques pistes de réflexion :

Afin de savoir si la tendance à la baisse se confirme, il faut attendre les résultats de l'enquête RECAP Qualité de 2011 qui seront disponibles courant 2012 (la campagne 2010 n'ayant pu être réalisée pour raisons budgétaires).

Le codage de la profession antérieure donnant des résultats bien à part (taux d'erreur de codage et codages à blanc plus importants que pour les autres variables), une attention particulière devra être portée sur le traitement de cette variable dans les campagnes futures de Recap.

En 2006, l'échantillon de la campagne RECAP-Qualité privilégiait les codages qui avaient été réalisés manuellement. Il était considéré que les taux d'erreur automatiques étaient forcément beaucoup moins importants qu'en reprise. Pourtant, au vu des résultats des campagnes suivantes, les taux d'erreur automatiques ne sont pas aussi faibles qu'attendu. Le contrôle des processus automatiques est donc important, et une amélioration de la qualité des processus automatiques est aussi un objectif à poursuivre .

Le codage automatique de l'activité est séparé en deux sous-processus : le codage automatique seul et le codage automatique complété par SICORE (Système Informatisé de COdage des Réponses aux Enquêtes). De 2006 à 2009, le taux d'erreur de codage est très sensiblement inférieur pour le deuxième sous-processus (cf. tableaux en annexe 3), ce qui laisse penser que le codage automatique devrait être complété plus souvent par SICORE. En effet, on a une différence de 1,7 points entre ces deux codages en 2006 et de 5,2 points en 2009 au niveau division.

Bibliographie

- [1] Bertrand Ph., Chauvet G., Christian B., Grosbras J.-M., « Les plans de sondage du nouveau recensement » et « Données produites par le recensement rénové de la population », JMS 2002
- [2] Durr J.-M., Dumais J., « La rénovation du recensement français », revue *Techniques d'enquête* de Statistique Canada, volume 28, n°1, juin 2002
- [3] Godinot A., « La rénovation du recensement de la population », INSEE, *Courrier des statistiques*, n°105-106, juin 2003
- [4] Cézard M., Lefebvre O., « La qualité du recensement », décembre 2008, disponible sur le site : <http://unstats.un.org/unsd/censuskb20/Attachment478.aspx>
- [5] CNIS - Commission nationale d'évaluation du recensement de la population, « Comprendre la population légale de votre commune, des enquêtes de recensement » : http://www.insee.fr/fr/publics/communication/recensement/presse/doc/Comprendre_poplegale_commu ne.pdf

ANNEXES

Annexe 1 : Bulletin individuel


Recensement de la population - 2011
Bulletin individuel


exemple : DUPUIS, épouse MAURIN

Nom : _____

Prénom : _____

Adresse : _____

Cadre à remplir par l'agent recenseur

commune

dépt. [] [] commune [] [] [] []

1 Sexe Masculin 1 Féminin 2 **SEXE**

2 Date et lieu de naissance

Né(e) le : JNAI MNAI ANAI
jour mois année

à : CNAI (cna_i_c_c)
commune (et arrondissement pour Paris, Lyon, Marseille)

DPNAI PNAI (pnai_c_c)
département code pays pour l'étranger, territoire pour les TOM

Si vous êtes né(e) à l'étranger, en quelle année êtes-vous arrivé(e) en France ? ARRIV
année

3 Quelle est votre nationalité ? INAT

- Française
 - Vous êtes né(e) français(e) 1
 - Vous êtes devenu(e) français(e) par exemple : par naturalisation, par déclaration, à votre mariage) 2
- ↳ Indiquez votre nationalité à la naissance : NATN (natn_c_c)
- Etrangère 3
- ↳ Indiquez votre nationalité : NATE (nate_c_c)

4 Êtes-vous inscrit(e) dans un établissement d'enseignement pour l'année scolaire en cours ? Y compris apprentissage ou études supérieures.

Oui 1 Non 2 **ETUD**

↳ Si oui, où est situé cet établissement d'enseignement ?

- Dans la commune où vous résidez (ou dans le même arrondissement pour Paris, Lyon, Marseille) 1
- Dans une autre commune (ou un autre arrondissement) 2
- ↳ Indiquez cette autre commune : CETU
commune (et arrondissement pour Paris, Lyon, Marseille) département code TOM

5 Où habitez-vous le 1^{er} janvier 2010 ? IRAN

Les enfants nés après ce date ne sont pas concernés.

- Dans le même logement que maintenant 1
- Dans un autre logement de la même commune (ou du même arrondissement pour Paris, Lyon, Marseille) 2
- Dans une autre commune (ou un autre arrondissement pour Paris, Lyon, Marseille) 3
- ↳ Indiquez cette autre commune : CRAN (cran_c_c)
commune (et arrondissement pour Paris, Lyon, Marseille)
- DRAN PRAN (pran_c_c)
département code pays pour l'étranger, territoire pour les TOM

6 La suite du questionnaire s'adresse aux personnes de 14 ans ou plus.

7 Vivez-vous en couple ? Oui 1 Non 2 **COUPLE**

8 Quel est votre état matrimonial légal ? MATR

- Célibataire (jamais légalement marié(e)) 1
- Marié(e) (ou séparé(e) mais non divorcé(e)) 2
- Veuf/veuve 3
- Divorcé(e) 4

9 Quel(s) diplôme(s) avez-vous ? DIPL

- Vous n'avez pas été scolarisé(e) 01
- Aucun diplôme mais scolarité jusqu'en école primaire ou au collège 02
- Aucun diplôme mais scolarité au-delà du collège 03
- CEP (certificat d'études primaires) 11
- BEPC, brevet élémentaire, brevet des collèges 12
- CAP, brevet de compagnon 13
- BEP 14
- Baccalauréat général, brevet supérieur 15
- Baccalauréat technologique ou professionnel, brevet professionnel ou de technicien, BEA, BEC, BEL, BEH, capacité en droit 16
- Diplôme de 1^{er} cycle universitaire, BTS, DUT, diplôme des professions sociales ou de la santé, d'infirmier(ère) 17
- Diplôme de 2^e ou 3^e cycle universitaire (y compris médecine, pharmacie, dentaire), diplôme d'ingénieur, d'une grande école, doctorat, etc. 18

10 Quelle est votre situation principale ? SITUAT

Ne cochez qu'une seule case.

- Emploi (salarial ou à votre compte, y compris aide d'une personne dans son travail) (à cocher puis passer en []) 1
- Apprentissage sous contrat ou stage rémunéré (à cocher puis passer en []) 2
- Études (libres, étudiant) ou stage non rémunéré 3
- Chômage (inscrit ou non au pôle emploi) 4
- Retraite ou préretraite (ancien salarié ou ancien indépendant) 5
- Femme ou homme au foyer 6
- Autre situation 7

11 Travaillez-vous actuellement ? SITRAV

Si vous avez un emploi occasionnel ou de très courte durée, ou si vous êtes en apprentissage ou en stage rémunéré, cochez « Oui ».

Si vous êtes en congé maladie ou de maternité, cochez « Oui ».

- Oui (à cocher puis passer en []) 1
- Non (à cocher puis passer en []) 2

continuez page suivante et n'oubliez pas de signer →

Imprimé n° 3

Annexe 2 : Les échantillons de la campagne RECAP Qualité

ECHANTILLONS DE LA CAMPAGNE RECAP QUALITE PAR VARIABLE, PAR ANNEE ET PAR PROCESSUS DE CODAGE					
ACTIVITE					
Processus de codage		2006	2007	2008	2009
Codage automatique	Automatique	2 000	6 000	6 000	6 000
	Automatique + SICORE	2 000	6 000	6 000	6 000
Codage manuel	Recap avec choix SIRET	2 000	6 000	6 000	6 000
	Recap avec choix d'un code NAF	46 181	33 000	33 069	33 079
TOTAL		52 181	51 000	51 069	51 079
PROFESSION ACTUELLE					
Processus de codage		2006	2007	2008	2009
Codage automatique		2 000	6 000	6 000	6 000
Codage manuel		48 520	33 376	33 587	33 593
TOTAL		50 520	39 376	39 587	39 593
PROFESSION ANTERIEURE					
Processus de codage		2006	2007	2008	2009
Codage automatique			1 000	1 000	1 000
Codage manuel			6 238	6 379	6 333
TOTAL			7 238	7 379	7 333

Annexe 3 : Tableaux synoptiques

TABLEAU SYNOPTIQUE 2006

ACTIVITE (NIVEAU DIVISION)

ECHANTILLON

Processus de codage		Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré ¹⁸	Nombre de mauvais codages	Taux d'erreur de codage de l'activité %	
Codage automatique	Automatique	2 000	131	Convergents	1 769	0	0	0,0	9,5
				Divergents	231	131	108	82,4	
				<i>dont NAF divergents</i>	159	131	108	82,4	
				<i>dont NAF convergents</i>	72	0	0	0,0	
	Automatique + SICORE	2 000	49	Convergents	1 617	0	0	0,0	7,8
				Divergents	383	49	20	40,8	
		<i>dont NAF divergents</i>	76	49	20	40,8			
		<i>dont NAF convergents</i>	307	0	0	0,0			
Codage manuel	Recap avec choix SIRET	2 000	178	Convergents	1 337	0	0	0,0	14,2
				Divergents	663	178	76	42,7	
				<i>dont NAF divergents</i>	312	178	76	42,7	
				<i>dont NAF convergents</i>	351	0	0	0,0	
	Recap avec choix d'un code NAF	46 181	10 738	Convergents	20 469	0	0	0,0	33,1
				Divergents	25 712	10 738	6 376	59,4	
				<i>dont NAF divergents</i>	5 028	2 836	2 019	71,2	
				<i>dont NAF convergents</i>	20 684	7 902	4 357	55,1	
Total	des convergents			25 192	0	0			
	des divergents			26 989	11 096	6 580			
	<i>des divergents dont NAF divergente</i>			5 575	3 194	2 223			
	<i>des divergnents dont NAF convergente</i>			21 414	7 902	4 357			
TOTAL :					52 181	11 096	6 580		

¹⁸ En 2006, tous les codages divergents ne sont pas arbitrés car la définition du niveau de codage pour les années suivantes n'est pas la même. En 2007, 2008 et 2009, on parle de convergence en niveau NAF et PCS alors qu'en 2006, la convergence était retenue à un niveau plus agé. Ainsi, les divergents qui apparaissent ci-dessus pouvaient être considérés comme convergents en 2006 et n'ont donc pas été arbitrés.

POPULATION TOTALE

Processus de codage		Population		Taux d'erreur de codage de l'activité par sous-processus %	Taux d'erreur de codage de l'activité par processus %	Taux d'erreur global %
Codage automatique	Automatique	594 086	37,1%	9,5	9,3	15,5
	Automatique + SICORE	94 733	5,9%	7,8		
Codage manuel	Recap avec choix SIRET	618 487	38,6%	14,2	20,2	
	Recap avec choix d'un code NAF	293 218	18,3%	33,1		

TOTAL 1 600 524

PROFESSION ACTUELLE (NIVEAU CS 42)

ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession %
Codage automatique	2 000	322	Convergenents	1 485	0	0	12,2
			Divergenents	515	322	152	
Codage manuel	48 520	14895	Convergenents	25 327	0	0	26,6
			Divergenents	23 193	14 895	8 285	
Total :	des convergenents			26 812	0	0	
	des divergenents			23 708	15 217	8 437	
TOTAL :				50 520	15 217	8 437	

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
Codage automatique	1 185 999	76,3%	12,2	15,6
Codage manuel	369 124	23,7%	26,6	

TOTAL 1 555 123

TABLEAU SYNOPTIQUE 2007

ACTIVITE (NIVEAU DIVISION)

ECHANTILLON

Processus de codage		Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de l'activité %	
Codage automatique	Automatique	6 000	819	Convergents	5 372	388	5	1,3	6,1
				Divergents	628	431	205	47,6	
				<i>dont NAF divergents</i>	449	314	199	63,4	
				<i>dont NAF convergents</i>	179	117	6	5,1	
	Automatique + SICORE	6 000	1 022	Convergents	5 049	350	15	4,3	6,0
				Divergents	951	672	101	15,0	
<i>dont NAF divergents</i>				241	177	75	42,4		
			<i>dont NAF convergents</i>	710	495	26	5,3		
Codage manuel	Recap avec choix SIRET	6 000	1 678	Convergents	4 108	289	6	2,1	7,8
				Divergents	1 892	1 389	279	20,1	
				<i>dont NAF divergents</i>	962	689	229	33,2	
				<i>dont NAF convergents</i>	932	700	50	7,1	
	Recap avec choix d'un code NAF	33 000	14 304	Convergents	14 562	1 001	85	8,5	23,1
				Divergents	18 438	13 303	4 612	34,7	
<i>dont NAF divergents</i>				3 729	2 654	1 151	43,4		
			<i>dont NAF convergents</i>	14 709	10 649	3 461	32,5		
Total	des convergents				29 091	2 028	111		
	des divergents				21 909	15 795	5 197		
	<i>des divergents dont NAF divergente</i>				5 381	3 834	1 654		
	<i>des divergents dont NAF convergente</i>				16 530	11 961	3 543		
TOTAL :					51 000	17 823	5 308		

POPULATION TOTALE

Processus de codage		Population		Taux d'erreur de codage de l'activité par sous-processus%	Taux d'erreur de codage de l'activité par processus%	Taux d'erreur global %
Codage automatique	Automatique	609 594	37,7%	6,1	6,1	9,8
	Automatique + SICORE	94 494	5,8%	6,0		
Codage manuel	Recap avec choix SIRET	621 801	38,4%	7,8	12,7	
	Recap avec choix d'un code NAF	291 668	18,0%	23,1		

TOTAL 1 617 557

TABLEAU SYNOPTIQUE 2007
PROFESSION ACTUELLE (NIVEAU CS 42)

ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré		Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession %	
Codage automatique	6 000	1 689	Convergents	4 790	479	19	4,0	10,2
			Divergents	1 210	1 210	421	34,8	
Codage manuel	33 376	17 193	Convergents	17 723	1 770	109	6,2	21,9
			Divergents	15 653	15 423	6 138	39,8	
Total :	des convergents			22 513	2 249	128		
	des divergents			16 863	16 633	6 559		
TOTAL :				39 376	18 882	6 687		

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
Codage automatique	1 190 978	75,8%	10,2	13,0
Codage manuel	380 795	24,2%	21,9	

TOTAL 1 571 773

PROFESSION ANTERIEURE (NIVEAU CS 42)

ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession %	
Codage automatique	1 000	164	Convergents	721	38	4	10,5	20,7
			Divergents	279	126	59	46,8	
Codage manuel	6 238	1 876	Convergents	2 561	26	3	11,5	35,1
			Divergents	3 677	1850	952	51,5	
Total :	des convergents			3 282	64	7		
	des divergents			3 956	1 976	1 011		
TOTAL :				7 238	2 040	1 018		

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
Codage automatique	937 410	82,2%	20,7	23,2
Codage manuel	203 483	17,8%	35,1	

TOTAL 1 140 893

TABLEAU SYNOPTIQUE 2008

ACTIVITE (NIVEAU DIVISION)

ECHANTILLON

Processus de codage		Echantillon	Echantillon arbitré		Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de l'activité %		
Codage automatique	Automatique	6 000	1 212	Convergents	5 319	531	2	0,4	5,0	
				Divergents	681	681	281	41,3		
				<i>dont NAF divergents</i>	495	495	277	56,0		
				<i>dont NAF convergents</i>	186	186	4	2,2		
	Automatique + SICORE	6 000	1 365	Convergents	5 150	515	3	0,6	2,2	
				Divergents	850	850	99	11,6		
				<i>dont NAF divergents</i>	214	214	73	34,1		
				<i>dont NAF convergents</i>	636	636	26	4,1		
Codage manuel	Recap avec choix SIRET	6 000	2 323	Convergents	4 085	408	1	0,2	5,0	
				Divergents	1 915	1 915	292	15,2		
				<i>dont NAF divergents</i>	1 013	1 013	272	26,9		
				<i>dont NAF convergents</i>	902	902	20	2,2		
	Recap avec choix d'un code NAF	33 069	19 798	Convergents	14 731	1 460	49	3,4	16,2	
				Divergents	18 338	18 338	4 870	26,6		
				<i>dont NAF divergents</i>	3 789	3 189	1 469	46,1		
				<i>dont NAF convergents</i>	14 549	14 549	3 401	23,4		
Total :	des convergents				29 285	2 914	55			
	des divergents				21 784	21 784	5 542			
	<i>des divergents dont NAF divergente</i>				5 511	4 911	2 091			
	<i>des divergents dont NAF convergente</i>				16 273	16 273	3 451			
TOTAL :					51 069	24 698	5 597			

POPULATION TOTALE

Processus de codage		Population		Taux d'erreur de codage de l'activité par sous-processus%	Taux d'erreur de codage de l'activité par processus%	Taux d'erreur global %
Codage automatique	Automatique	650 170	35,7%	5,0	4,6	7,0
	Automatique + SICORE	103 119	5,7%	2,2		
Codage manuel	Recap avec choix SIRET	726 576	39,9%	5,0	8,6	
	Recap avec choix d'un code	342 835	18,8%	16,2		

TOTAL 1 822 700

PROFESSION ACTUELLE (NIVEAU CS 42)

ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession %	
Codage automatique	6 000	1779	Convergenents	4 690	469	6	1,3	7,9
			Divergenents	1 310	1 310	414	31,6	
Codage manuel	33 376	17819	Convergenents	17 171	1 645	55	3,3	19,2
			Divergenents	16 416	16 174	5 791	35,8	
Total :	des convergenents			21 861	2 114	61		
	des divergenents			17 726	17 484	6 205		
TOTAL :				39 587	19 598	6 266		

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
Codage automatique	1 334 598	75,5%	7,9	
Codage manuel	433 331	24,5%	19,2	

TOTAL 1 767 929

PROFESSION ANTERIEURE (NIVEAU CS 42)

ECHANTILLON

Processus de codage	Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de la profession %	
Codage automatique	1 000	264	Convergenents	817	81	6	7,4	14,5
			Divergenents	183	183	84	45,9	
Codage manuel	6 379	4180	Convergenents	2 263	64	9	14,1	40,3
			Divergenents	4 116	4 116	2 250	54,7	
Total :	des convergenents			3 080	145	15		
	des divergenents			4 299	4 299	2 334		
TOTAL :				7 379	4 444	2 349		

POPULATION TOTALE

Processus de codage	Population		Taux d'erreur de codage de l'activité %	Taux d'erreur global %
Codage automatique	1 008 560	82,1%	14,5	
Codage manuel	219 802	17,9%	40,3	

TOTAL 1 228 362

TABLEAU SYNOPTIQUE 2009

ACTIVITE (NIVEAU DIVISION)

ECHANTILLON

Processus de codage		Echantillon	Echantillon arbitré	Qualité	Echantillon	Echantillon arbitré	Nombre de mauvais codages	Taux d'erreur de codage de l'activité %	
Codage automatique	Automatique	6 000	1 442	Convergents	5 064	506	1	0,2	8,1
				Divergents	936	936	476	50,9	
				<i>dont NAF divergents</i>	697	697	466	66,9	
	Automatique + SICORE	6 000	1 520	Convergents	4 977	497	3	0,6	2,9
				Divergents	1 023	1 023	144	14,1	
				<i>dont NAF divergents</i>	328	328	136	41,5	
Codage manuel	Recap avec choix SIRET	6 000	2 461	Convergents	3 932	393	1	0,3	5,5
				Divergents	2 068	2 068	320	15,5	
				<i>dont NAF divergents</i>	1 050	1 050	295	28,1	
	Recap avec choix d'un code NAF	33 079	19 481	Convergents	15 095	1 497	62	4,1	17,8
				Divergents	17 984	17 984	5 274	29,3	
				<i>dont NAF divergents</i>	3 777	3 777	1 666	44,1	
Total	des convergents				29 068	2 893	67		
	des divergents				22 011	22 011	6 214		
	<i>des divergents dont NAF divergente</i>				5 852	5 852	2 563		
	<i>des divergents dont NAF convergente</i>				16 159	16 159	3 651		
TOTAL :					51 079	24 904	6 281		

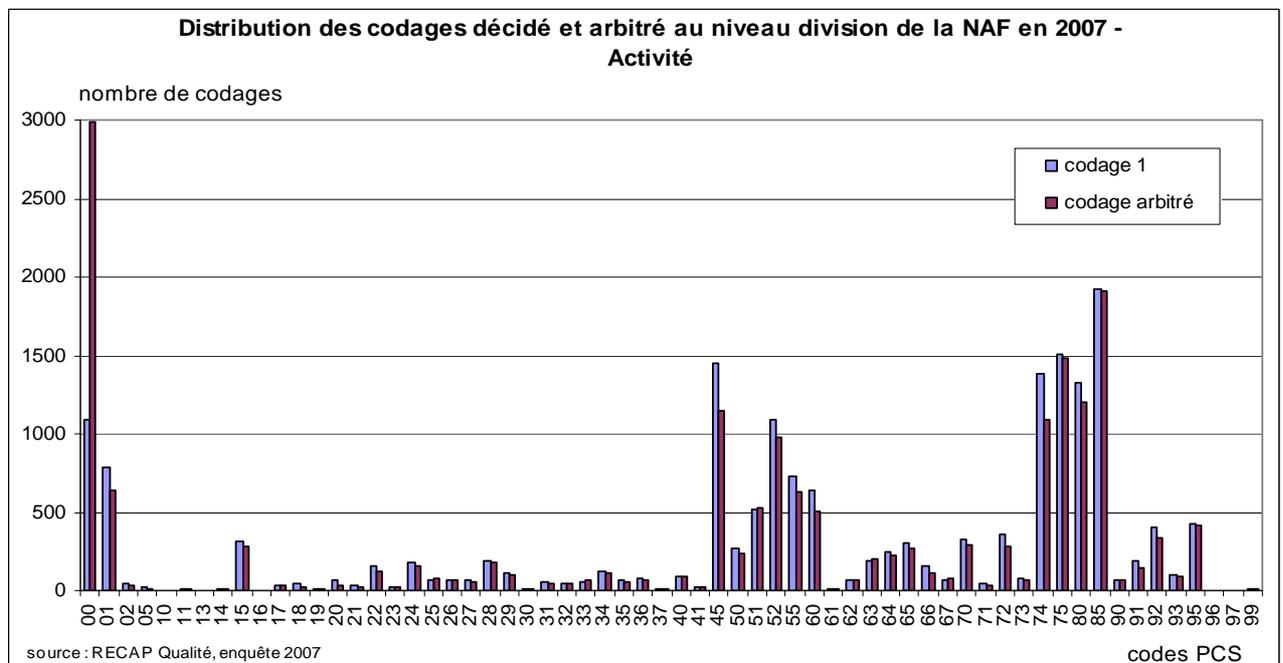
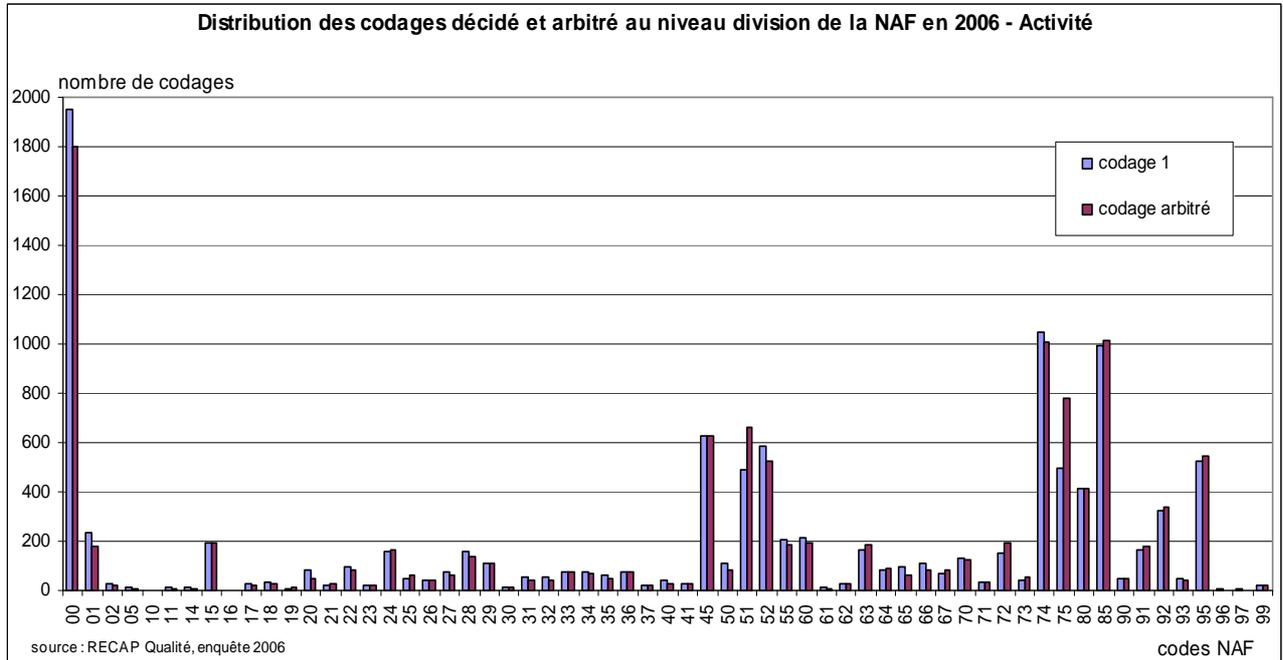
POPULATION TOTALE

Processus de codage		Population		Taux d'erreur de codage de l'activité par sous-processus%	Taux d'erreur de codage de l'activité par processus%	Taux d'erreur global %
Codage automatique	Automatique	746 035	41,2%	8,1	7,8	8,7
	Automatique + SICORE	51 849	2,9%	2,9		
Codage manuel	Recap avec choix SIRET	687 395	37,9%	5,5	9,5	
	Recap avec choix d'un code NAF	326 088	18,0%	17,8		

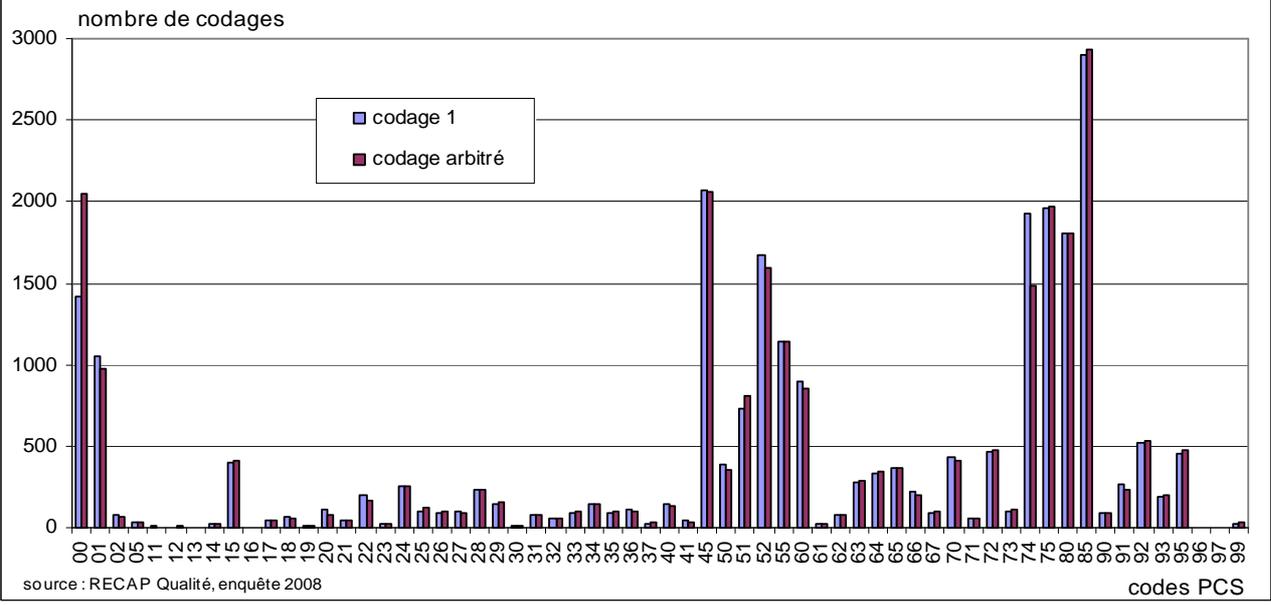
TOTAL 1 811 367

Annexe 4 : Distribution des codages initial (décidé) et arbitré

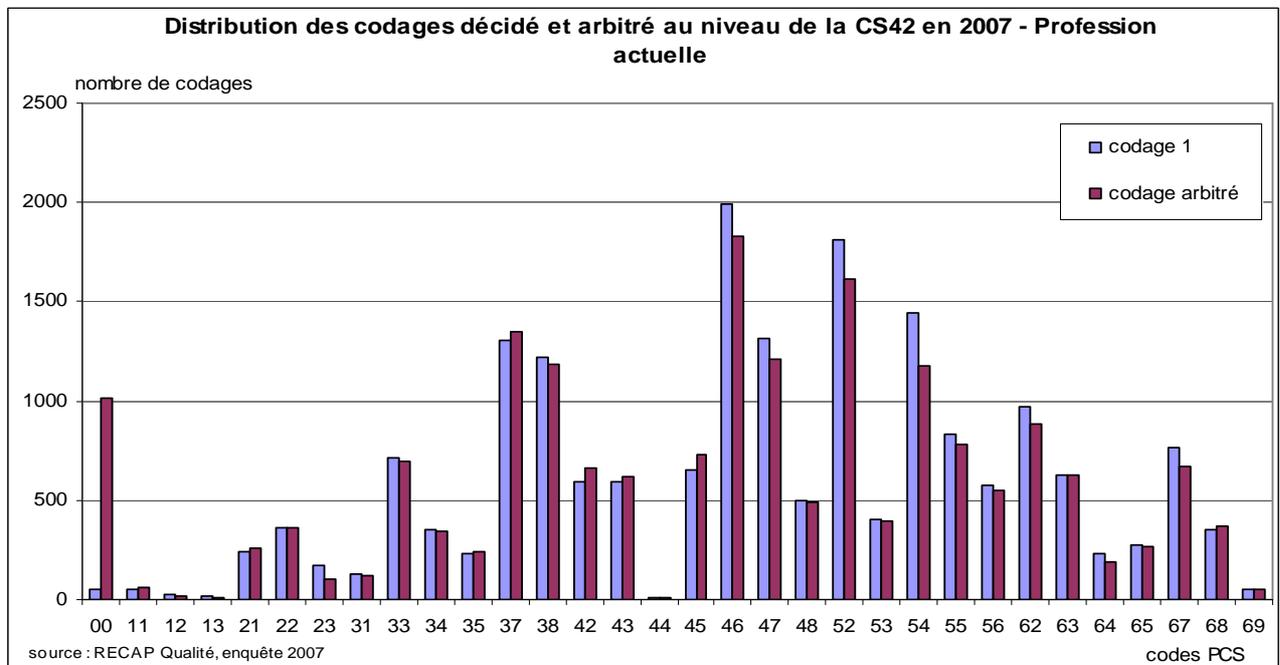
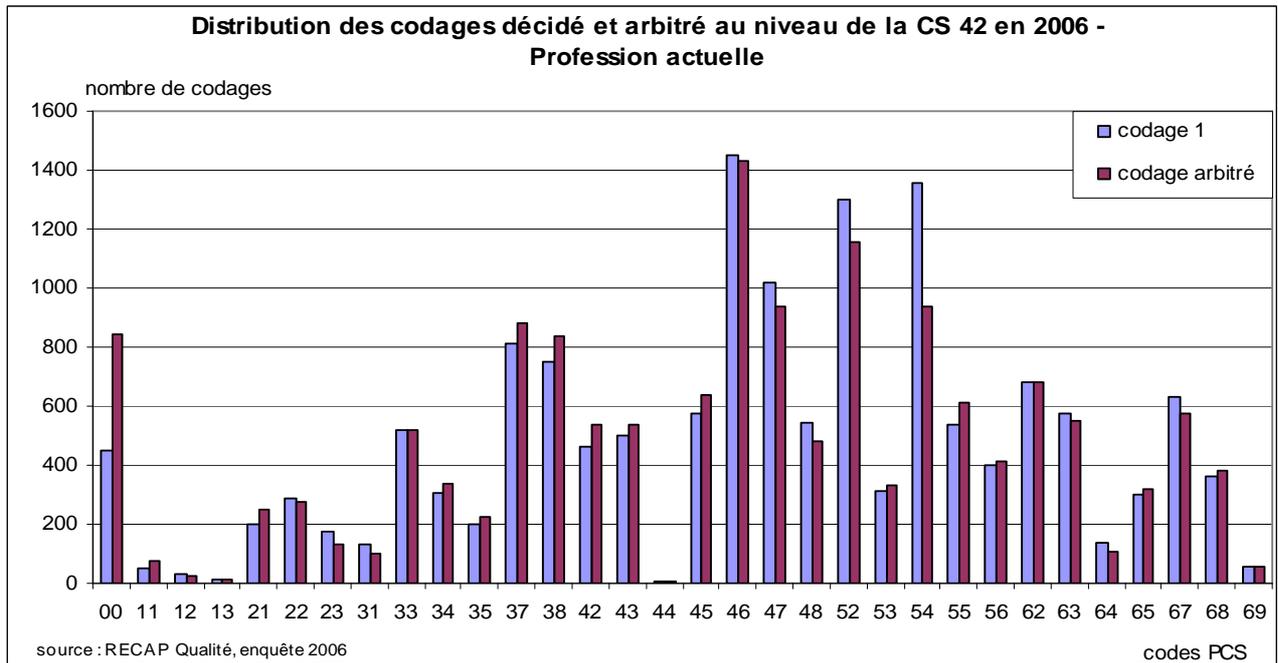
L'activité



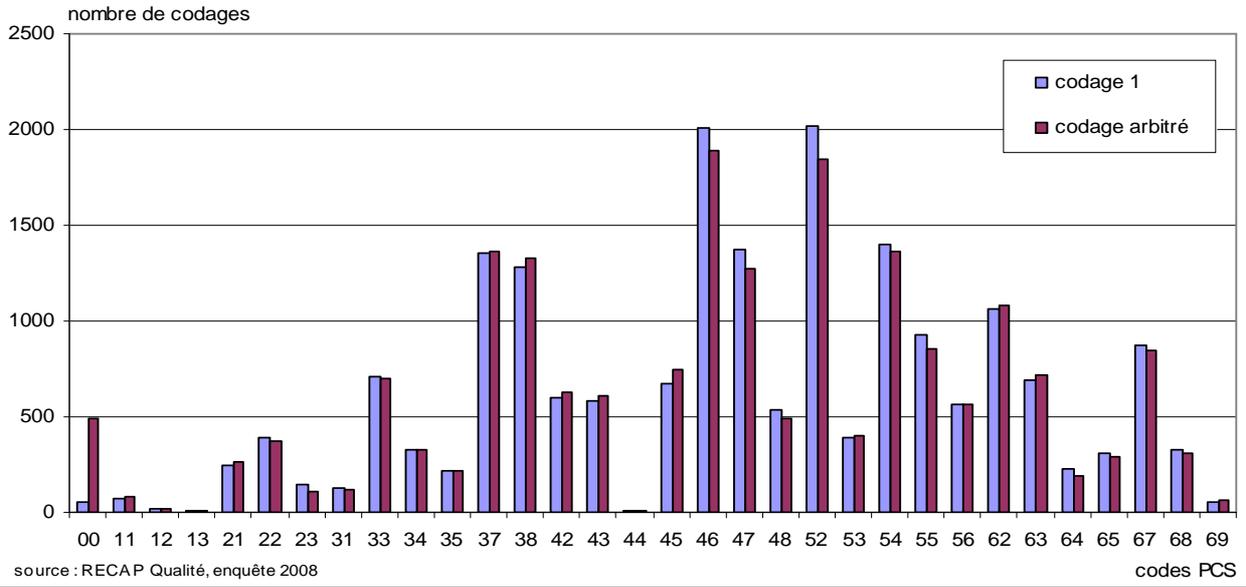
Distribution des codages décidé et arbitré au niveau division de la NAF en 2008 - Activité



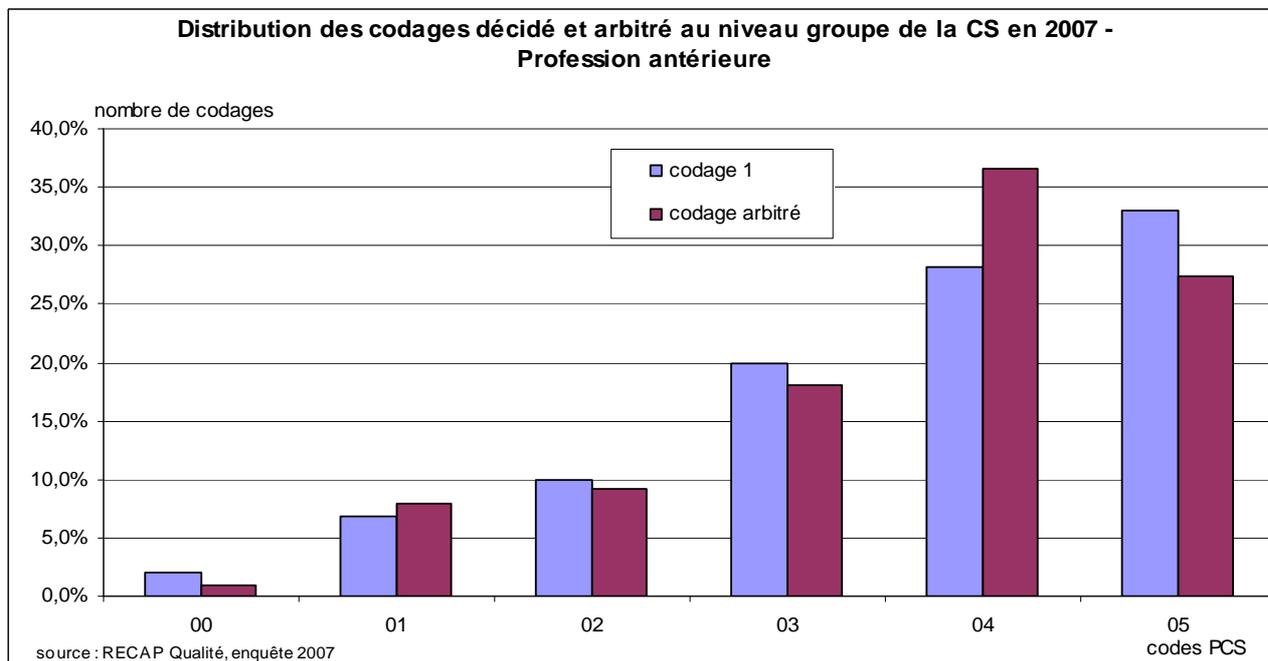
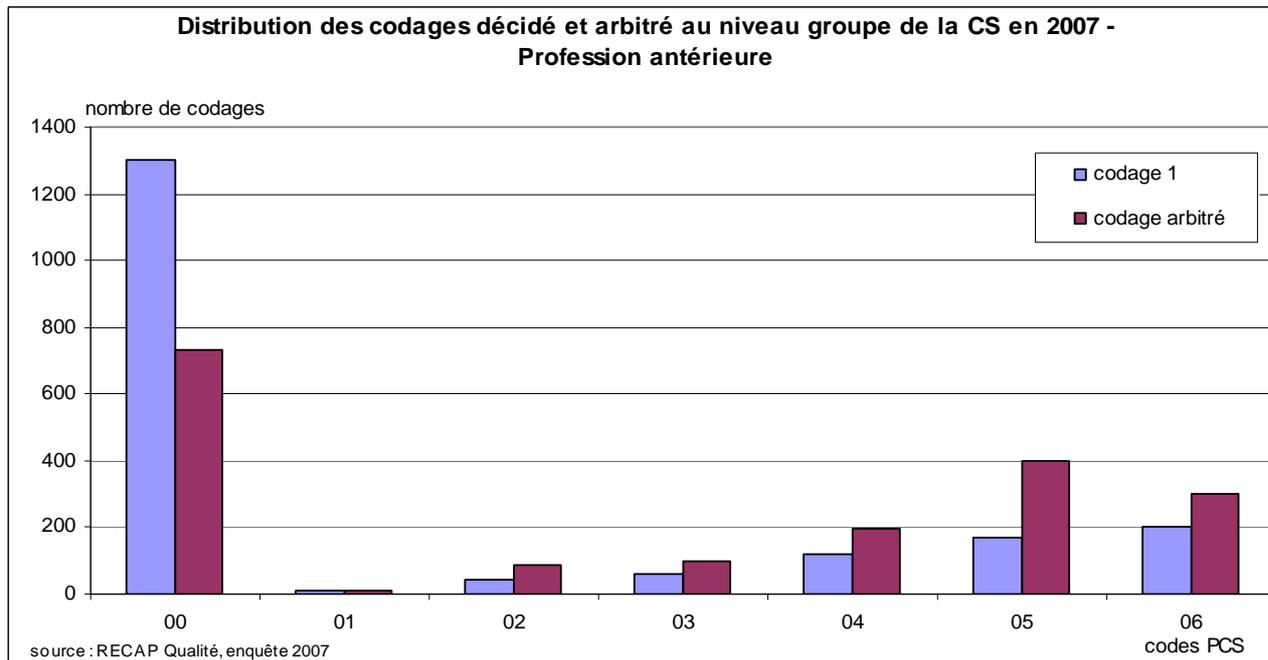
La profession actuelle (niveau CS42)



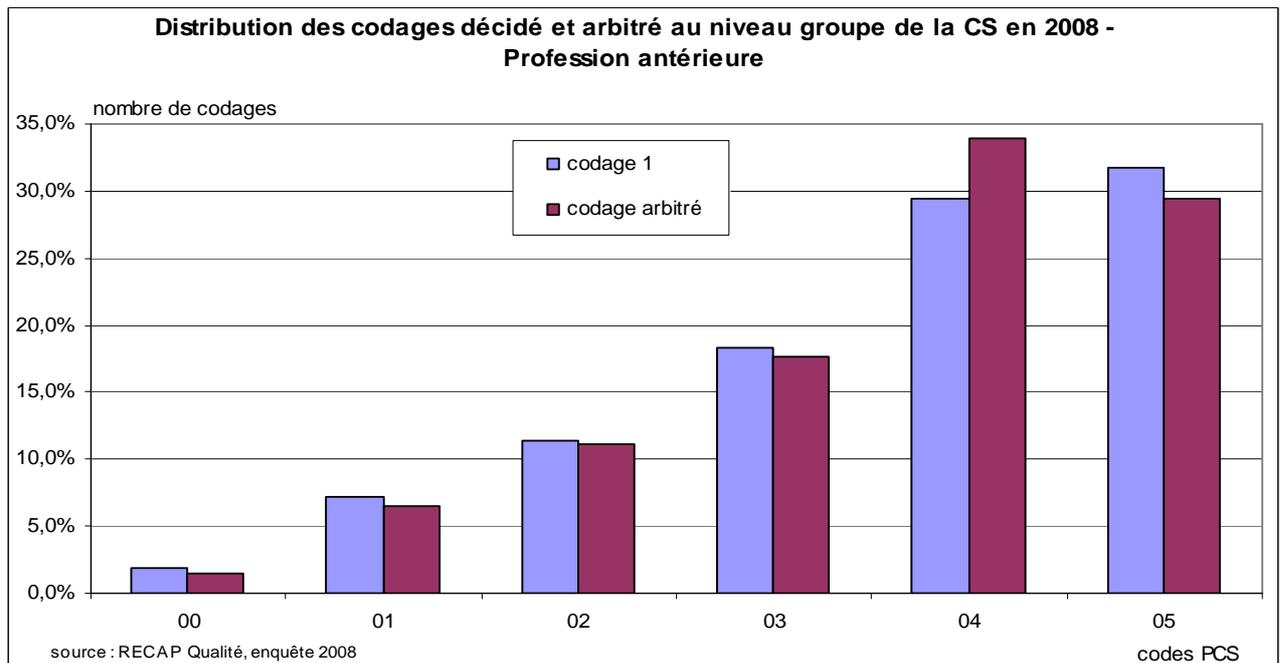
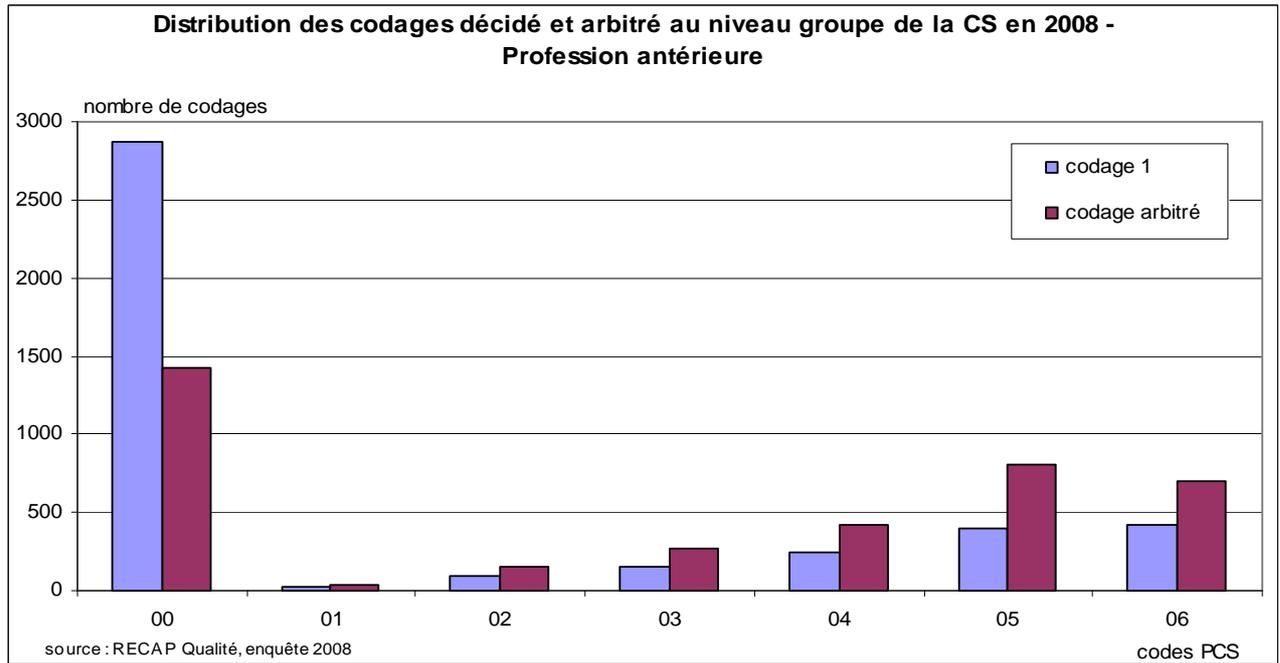
**Distribution des codages décidé et arbitré au niveau de la CS42 en 2008 -
Profession actuelle**



La profession antérieure (niveau groupe)



* en pourcentage et or codages à blanc



* en pourcentage et or codages à blanc

ANNEXE 5 : Nomenclatures activité et profession

Activité : NAF révision 1 - Niveau division (62 postes)

NAF rév. 1		Niveau 60 - Liste des divisions	
Section	Libellé des sections	Code division	Intitulé
A	AGRICULTURE, CHASSE, SYLVICULTURE	01	Agriculture, chasse, services annexes
		02	Sylviculture, exploitation forestière, services annexes
B	PÊCHE, AQUACULTURE, SERVICES ANNEXES	05	Pêche, aquaculture, services annexes
C	INDUSTRIES EXTRACTIVES	10	Extraction de houille, de lignite et de tourbe
		11	Extraction d'hydrocarbures ; services annexes
		12	Extraction de minerais d'uranium
		13	Extraction de minerais métalliques
		14	Autres industries extractives
D	INDUSTRIE MANUFACTURIÈRE	15	Industries alimentaires
		16	Industrie du tabac
		17	Industrie textile
		18	Industrie de l'habillement et des fourrures
		19	Industrie du cuir et de la chaussure
		20	Travail du bois et fabrication d'articles en bois
		21	Industrie du papier et du carton
		22	Edition, imprimerie, reproduction
		23	Cokéfaction, raffinage, industries nucléaires
		24	Industrie chimique
		25	Industrie du caoutchouc et des plastiques
		26	Fabrication d'autres produits minéraux non métalliques
		27	Métallurgie
		28	Travail des métaux
		29	Fabrication de machines et d'équipements
		30	Fabrication de machines de bureau et de matériel informatique
		31	Fabrication de machines et appareils électriques
32	Fabrication d'équipements de radio, télévision et communication		
33	Fabrication d'instruments médicaux, de précision, d'optique et d'horlogerie		
34	Industrie automobile		
35	Fabrication d'autres matériels de transport		
36	Fabrication de meubles ; industries diverses		
37	Récupération		
E	PRODUCTION ET DISTRIBUTION D'ÉLECTRICITÉ, DE GAZ, ET D'EAU	40	Production et distribution d'électricité, de gaz et de chaleur
		41	Captage, traitement et distribution d'eau
F	CONSTRUCTION	45	Construction
G	COMMERCE ; RÉPARATIONS AUTOMOBILE ET D'ARTICLES DOMESTIQUES	50	Commerce et réparation automobile
		51	Commerce de gros et intermédiaires du commerce
		52	Commerce de détail et réparation d'articles domestiques
H	HÔTELS ET RESTAURANTS	55	Hôtels et restaurants
I	TRANSPORTS ET COMMUNICATIONS	60	Transports terrestres
		61	Transports par eau
		62	Transports aériens
		63	Services auxiliaires des transports
		64	Postes et télécommunications
J	ACTIVITÉS FINANCIÈRES	65	Intermédiation financière
		66	Assurance
		67	Auxiliaires financiers et d'assurance
K	IMMOBILIER, LOCATION ET SERVICES AUX ENTREPRISES	70	Activités immobilières
		71	Location sans opérateur
		72	Activités informatiques
		73	Recherche et développement
		74	Services fournis principalement aux entreprises
L	ADMINISTRATION PUBLIQUE	75	Administration publique
M	EDUCATION	80	Education
N	SANTÉ ET ACTION SOCIALE	85	Santé et action sociale
O	SERVICES COLLECTIFS, SOCIAUX ET PERSONNELS	90	Assainissement, voirie et gestion des déchets
		91	Activités associatives
		92	Activités récréatives, culturelles et sportives
		93	Services personnels
P	ACTIVITÉS DES MÉNAGES	95	Activités des ménages en tant qu'employeur de personnel domestique
		96	Activités indifférenciées des ménages en tant que producteurs de biens pour usage propre
		97	Activités indifférenciées des ménages en tant que producteurs de services pour usage propre
Q	ACTIVITÉS EXTRA-TERRITORIALES	99	Activités extra-territoriales

Activité : NAF révision 2 - Niveau division (88 postes)

NAF rév. 2

Liste des divisions

Section	Libellé des sections	Code Division	Intitulé
A	AGRICULTURE, SYLVICULTURE ET PÊCHE	01	Culture et production animale, chasse et services annexes
		02	Sylviculture et exploitation forestière
		03	Pêche et aquaculture
B	INDUSTRIES EXTRACTIVES	05	Extraction de houille et de lignite
		06	Extraction d'hydrocarbures
		07	Extraction de minerais métalliques
		08	Autres industries extractives
		09	Services de soutien aux industries extractives
C	INDUSTRIE MANUFACTURIÈRE	10	Industries alimentaires
		11	Fabrication de boissons
		12	Fabrication de produits à base de tabac
		13	Fabrication de textiles
		14	Industrie de l'habillement
		15	Industrie du cuir et de la chaussure
		16	Travail du bois et fabrication d'articles en bois et en liège, à l'exception des meubles ; fabrication d'articles en vannerie et sparterie
		17	Industrie du papier et du carton
		18	Imprimerie et reproduction d'enregistrements
		19	Cokéfaction et raffinage
		20	Industrie chimique
		21	Industrie pharmaceutique
		22	Fabrication de produits en caoutchouc et en plastique
		23	Fabrication d'autres produits minéraux non métalliques
		24	Métallurgie
		25	Fabrication de produits métalliques, à l'exception des machines et des équipements
		26	Fabrication de produits informatiques, électroniques et optiques
		27	Fabrication d'équipements électriques
		28	Fabrication de machines et équipements n.c.a.
29	Industrie automobile		
30	Fabrication d'autres matériels de transport		
31	Fabrication de meubles		
32	Autres industries manufacturières		
33	Réparation et installation de machines et d'équipements		
D	PRODUCTION ET DISTRIBUTION D'ÉLECTRICITÉ, DE GAZ, DE VAPEUR ET D'AIR CONDITIONNÉ	35	Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
E	PRODUCTION ET DISTRIBUTION D'EAU ; ASSAINISSEMENT, GESTION DES DÉCHETS ET DÉPOLLUTION	36	Captage, traitement et distribution d'eau
		37	Collecte et traitement des eaux usées
		38	Collecte, traitement et élimination des déchets ; récupération
		39	Dépollution et autres services de gestion des déchets
F	CONSTRUCTION	41	Construction de bâtiments
		42	Génie civil
		43	Travaux de construction spécialisés
G	COMMERCE ; RÉPARATION D'AUTOMOBILES ET DE MOTOCYCLES	45	Commerce et réparation d'automobiles et de motocycles
		46	Commerce de gros, à l'exception des automobiles et des motocycles
		47	Commerce de détail, à l'exception des automobiles et des motocycles
H	TRANSPORTS ET ENTREPOSAGE	49	Transports terrestres et transport par conduites
		50	Transports par eau
		51	Transports aériens
		52	Entreposage et services auxiliaires des transports
		53	Activités de poste et de courrier
I	HÉBERGEMENT ET RESTAURATION	55	Hébergement
		56	Restauration
J	INFORMATION ET COMMUNICATION	58	Édition
		59	Production de films cinématographiques, de vidéo et de programmes de télévision ; enregistrement sonore et édition musicale
		60	Programmation et diffusion
		61	Télécommunications
		62	Programmation, conseil et autres activités informatiques
		63	Services d'information
K	ACTIVITÉS FINANCIÈRES ET D'ASSURANCE	64	Activités des services financiers, hors assurance et caisses de retraite
		65	Assurance
		66	Activités auxiliaires de services financiers et d'assurance
L	ACTIVITÉS IMMOBILIÈRES	68	Activités immobilières
M	ACTIVITÉS SPÉCIALISÉES, SCIENTIFIQUES ET TECHNIQUES	69	Activités juridiques et comptables
		70	Activités des sièges sociaux ; conseil de gestion
		71	Activités d'architecture et d'ingénierie ; activités de contrôle et analyses techniques
		72	Recherche-développement scientifique
		73	Publicité et études de marché
		74	Autres activités spécialisées, scientifiques et techniques
		75	Activités vétérinaires

N	ACTIVITÉS DE SERVICES ADMINISTRATIFS ET DE SOUTIEN	77	Activités de location et location-bail
		78	Activités liées à l'emploi
		79	Activités des agences de voyage, voyagistes, services de réservation et activités connexes
		80	Enquêtes et sécurité
		81	Services relatifs aux bâtiments et aménagement paysager
O	ADMINISTRATION PUBLIQUE	82	Activités administratives et autres activités de soutien aux entreprises
		84	Administration publique et défense ; sécurité sociale obligatoire
P	ENSEIGNEMENT	85	Enseignement
Q	SANTÉ HUMAINE ET ACTION SOCIALE	86	Activités pour la santé humaine
		87	Hébergement médico-social et social
		88	Action sociale sans hébergement
R	ARTS, SPECTACLES ET ACTIVITÉS RÉCRÉATIVES	90	Activités créatives, artistiques et de spectacle
		91	Bibliothèques, archives, musées et autres activités culturelles
		92	Organisation de jeux de hasard et d'argent
		93	Activités sportives, récréatives et de loisirs
S	AUTRES ACTIVITÉS DE SERVICES	94	Activités des organisations associatives
		95	Réparation d'ordinateurs et de biens personnels et domestiques
		96	Autres services personnels
T	ACTIVITÉS DES MÉNAGES EN TANT QU'EMPLOYEURS ; ACTIVITÉS INDIFFÉRENCIÉES DES MÉNAGES EN TANT QUE PRODUCTEURS DE BIENS ET SERVICES POUR USAGE PROPRE	97	Activités des ménages en tant qu'employeurs de personnel domestique
		98	Activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre
U	ACTIVITÉS EXTRA-TERRITORIALES	99	Activités des organisations et organismes extraterritoriaux

Profession : Nomenclature PCS 2003 - Niveau des catégories socioprofessionnelles en 42 postes, avec le regroupement en 8 groupes sociaux (dont 6 groupes d'actifs)

Niveau agrégé (groupe) 8 postes dont 6 postes pour les actifs	PCS 2003 Code	Niveau 3 - Liste des catégories socioprofessionnelles détaillées en 42 postes Libellé
1 Agriculteurs exploitants	11 12 13	Agriculteurs sur petite exploitation Agriculteurs sur moyenne exploitation Agriculteurs sur grande exploitation
2 Artisans, commerçants et chefs d'entreprise	21 22 23	Artisans Commerçants et assimilés Chefs d'entreprise de 10 salariés ou plus
3 Cadres et professions intellectuelles supérieures	31 33 34 35 37 38	Professions libérales Cadres de la fonction publique Professeurs, professions scientifiques Professions de l'information, des arts et des spectacles Cadres administratifs et commerciaux d'entreprise Ingénieurs et cadres techniques d'entreprise
4 Professions intermédiaires	42 43 44 45 46 47 48	Professeurs des écoles, instituteurs et assimilés Professions intermédiaires de la santé et du travail social Clergé, religieux Professions intermédiaires administratives de la fonction publique Professions intermédiaires administratives et commerciales des entreprises Techniciens Contremaîtres, agents de maîtrise
5 Employés	52 53 54 55 56	Employés civils et agents de service de la fonction publique Policiers et militaires Employés administratifs d'entreprise Employés de commerce Personnels des services directs aux particuliers
6 Ouvriers	62 63 64 65 67 68 69	Ouvriers qualifiés de type industriel Ouvriers qualifiés de type artisanal Chauffeurs Ouvriers qualifiés de la manutention, du magasinage et du transport Ouvriers non qualifiés de type industriel Ouvriers non qualifiés de type artisanal Ouvriers agricoles
7 Retraités	71 72 74 75 77 78	Anciens agriculteurs exploitants Anciens artisans, commerçants, chefs d'entreprise Anciens cadres Anciennes professions intermédiaires Anciens employés Anciens ouvriers
8 Autres personnes sans activité professionnelle	81 83 84 85 86	Chômeurs n'ayant jamais travaillé Militaires du contingent Elèves, étudiants Personnes diverses sans activité professionnelle de moins de 60 ans (sauf retraités) Personnes diverses sans activité professionnelle de 60 ans et plus (sauf retraités)