

# ***Sources et méthodes***

## **Enquête sur la formation et la qualification professionnelle (FQP)**

Les enquêtes sur la formation et la qualification professionnelle constituent une des principales sources d'information sur la mobilité professionnelle, la mobilité sociale et sur les relations entre la formation initiale et professionnelle, l'emploi et les salaires.

Initiée en 1964, l'enquête a aussi eu lieu, sous des formes proches, en 1970, 1977, 1985, 1993 et 2003.

La méthodologie présentée ici concerne l'enquête de 2003. Les enquêtes précédentes, même si elles ont été conduites selon des protocoles un peu différents, en raison de l'évolution des méthodes d'enquêtes, restent comparables, dans leur contenu et en termes de comparabilité de la plupart des résultats, à l'enquête de 2003.

---

## **Les objectifs des enquêtes formation et qualification professionnelle**

---

Historiquement l'enquête sur la formation et la qualification professionnelle (FQP) est la seule à donner simultanément une information détaillée sur le salaire et la profession au moment de l'enquête, les positions professionnelles occupées à différents moments de la carrière, l'accès à la formation post-scolaire, la formation initiale et la position sociale des parents (*encadré*). Elle fournit en outre cette information selon un protocole très stable depuis plus de trente ans. Elle est ainsi une source d'informations clef sur les trajectoires sociales et le changement social en France depuis 1970.

Elle est l'enquête à partir de laquelle ont été réalisés la plupart des travaux de référence sur l'évolution du système éducatif, sur les transformations de la demande de travail par qualification, sur l'évolution de l'égalité des chances, sur les changements dans les profils des carrières ou encore sur l'évolution de la distribution de l'effort de formation continue et de son impact.

Chacun de ces travaux de référence a sollicité l'enquête dans sa spécificité, c'est-à-dire qu'ils n'ont été possibles que dans la mesure où l'enquête se répète dans le temps selon un protocole stable et aborde en même temps des sujets habituellement traités à part. Dans au moins cinq domaines, elle apporte des informations originales.

### **a) Le rendement des diplômés**

Les questions auxquelles l'enquête apporte des éléments de réponse sont nombreuses : est-ce bien le diplôme qui permet d'éviter le chômage et d'obtenir des salaires plus élevés, ou est-ce l'école qui ne décerne des diplômes qu'à des personnes qui de toute façon gagneraient plus sur le marché du travail ? Comment l'effet propre du diplôme évolue-t-il dans le temps ? Pour répondre à ces questions, il faut disposer de données détaillées sur le niveau scolaire, les revenus du travail, le chômage et l'emploi. L'enquête FQP apporte des informations sur tous ces aspects de la carrière.

### **b) Origine sociale et inégalité des chances**

Comment l'origine sociale des individus (qu'il s'agisse de ses composantes « culturelles » ou

bien « socio-économiques ») influe-t-elle sur la réussite scolaire et les carrières professionnelles ? Tenter de répondre à ce type de questions, de même qu'essayer de démêler les différentes dimensions de l'origine sociale, peut se faire à partir de quatre types d'information présents dans l'enquête : le diplôme et la profession des ascendants, le diplôme de l'individu, la situation professionnelle en début de carrière et la situation professionnelle à la date d'enquête.

### **c) La carrière et la mobilité professionnelle**

Les perspectives de progression en cours de carrière évoluent-elles avec le contexte macroéconomique ? Se modifient-elles de la même façon pour les diplômés et les non diplômés ? Quelle est la proportion d'actifs sans réelle perspective de mobilité ? Change-t-elle dans le temps et pourquoi ? Répondre à ces questions nécessite des informations sur la situation professionnelle à différents moments de la carrière : situation en début de carrière, cinq ans avant l'enquête, à la date d'enquête, calendrier professionnel au cours des cinq dernières années constituent un grand nombre d'informations sur la carrière professionnelle disponibles dans l'enquête.

### **d) Les transformations de la demande de travail par qualification**

Pourquoi le chômage des personnes sans qualification a-t-il autant augmenté au cours des vingt dernières années ? Il s'agit de comprendre l'impact du coût du travail et du progrès technique sur la demande de travail par qualification. A cet effet, il est nécessaire de connaître les structures des qualifications, des revenus salariaux et des diplômes, soit trois types de renseignements fournis par l'enquête : situation professionnelle, diplôme, revenus du travail.

### **e) La formation continue**

A qui et à quoi servent les efforts de formation post-scolaire ? Il s'agit en particulier d'étudier l'impact des formations proposées par les employeurs sur les carrières professionnelles et salariales. Evaluer l'impact des formations sollicite les parties formation continue, revenus du travail, situation professionnelle cinq ans avant l'enquête et calendrier professionnel.

## ***Le questionnaire de l'enquête formation et qualification professionnelle***

Le questionnaire actuel de l'enquête est très proche de celui de 1993 permettant ainsi une comparabilité dans le temps. Cependant des enrichissements ont été apportés notamment sur les aspects de mobilité professionnelle et de conditions de travail. L'enquête commence par une description générale du ménage, qui va fournir des informations biographiques sur la personne interrogée : sa date de naissance, son sexe, ses relations familiales avec d'autres membres du ménage. Le cœur de l'enquête est ensuite constitué des 11 parties suivantes :

### **Le module « complément biographique »**

Ce module fournit quelques informations supplémentaires : le pays ou le département de naissance, la nationalité, la date du premier mariage, le sexe et la date de naissance des enfants, en particulier ceux qui ne sont plus dans le ménage. Ce module se termine par une question sur la situation professionnelle à la date d'enquête qui conditionne l'entrée dans les modules suivants.

### **Le module « situation professionnelle à la date de l'enquête » : *questions supplémentaires sur les conditions de travail en 2003***

Pour les personnes en emploi au moment de l'enquête ou ayant travaillé moins de 5 ans auparavant, il s'agit de décrire précisément le dernier emploi occupé. Les questions sont reprises de l'enquête Emploi. Elles sont enrichies par une interrogation sur les conditions de travail : usage de l'informatique, contraintes pesant sur le rythme de travail, contact avec le public. Les indépendants précisent comment ils ont pris possession de leur entreprise. Enfin, quelques questions portent sur les changements de poste amorçant le questionnement du module « calendrier professionnel ».

### **Le module « situation professionnelle en 1998 »**

Ce module interroge la personne sur sa situation cinq ans auparavant (plus précisément, la date de référence en 1998 est fixée au trentième jour du mois où se déroule l'enquête). Dans le cas où la personne avait un emploi, on pose les mêmes questions que dans le module décrivant la situation professionnelle à la date d'enquête. Si la personne n'avait pas d'emploi cinq ans avant, seuls deux questions lui sont posées sur sa situation (chômage/inactivité) et sur son lieu de résidence à l'époque.

### **Le module « dernière situation professionnelle »**

Pour les personnes ayant occupé leur dernier emploi plus de cinq ans avant la date d'enquête, une version simplifiée du module « situation professionnelle à la date d'enquête » est proposée. L'objectif est de ne pas trop solliciter leur mémoire.

### **Le module « calendrier professionnel » : *module nouveau en 2003***

Pour les personnes n'occupant pas le même poste que cinq ans auparavant, le module « calendrier professionnel » retrace le parcours professionnel sur les cinq dernières années, repérant les phases d'emploi et de non-emploi et pendant les périodes d'emploi, les changements de poste. Pour chacune de ces périodes, des questions sont posées sur les circonstances et les conséquences des changements. Pour le premier changement d'emploi, on note la commune de l'établissement employeur.

### **Le module « formation initiale »**

Cette partie décrit les études initiales année scolaire après année scolaire (quelques questions concernent la scolarité en maternelle). Un calendrier indique pour chaque année scolaire le cours suivi avec sa spécialité et, s'il y a lieu, l'obtention d'un diplôme et la réalisation d'un stage. On repère les périodes de scolarité à l'étranger ou dans des sections spécialisées. Le stage le plus récent est décrit plus en détail et une dernière question porte sur le plus haut diplôme obtenu en dehors des études initiales.

### **Le module « premier emploi »**

Ce module permet de décrire le premier emploi. Il s'agit d'une version simplifiée du module « situation professionnelle ». On demande le lieu de résidence à cette époque.

### **Le module « formation continue »**

Trois types de formations suivies après la fin des études initiales sur les 5 dernières années font l'objet d'un questionnement approfondi : la dernière formation suivie en entreprise, toutes les formations de plus de 30 heures, sans autre contrainte (on pose des questions plus précises sur la plus ancienne), la formation en alternance la plus récente. Enfin, une dernière question recense le nombre total de formations suivies durant les 5 dernières années.

### **Le module « origine sociale et géographique » : questions supplémentaires sur les frères et sœurs en 2003**

Ce module, classique de l'enquête, recueille des informations sur l'origine sociale. Concernant le père et la mère, on recueille les éléments suivants : la date de naissance (et éventuellement de décès), le lieu de naissance (département ou pays), la nationalité à la naissance, la profession et les diplômes au moment où la personne interrogée a fini ses études.

Une originalité de l'enquête de 2003 consiste à fournir la date de naissance, le sexe, la profession et le diplôme le plus élevé obtenu d'un frère ou d'une sœur choisi au hasard. On demande également la profession des deux grands-pères paternels et maternels, comme cela avait été le cas en 1985. Le module interroge aussi sur la situation familiale pendant les études : divorce ou séparation de ses parents, remariage, personnes avec qui elle vivait, lieu de résidence des parents (à la fin des études de l'enquêté). Le module s'achève par une question subjective demandant à l'enquêté de comparer sa situation professionnelle actuelle à celle qu'avait son père, à l'époque où lui-même finissait ses études.

### **Le module « revenus du travail en 2002 »**

Le revenu professionnel fait l'objet du dernier module, dans lequel la personne indique le total de ses revenus professionnels (salaire ou bénéfices pour les indépendants) au cours de l'année précédant l'enquête, c'est-à-dire 2002.

---

## ***La méthodologie de l'enquête formation et qualification professionnelle 2003***

---

### **Le champ de l'enquête**

Le champ de l'enquête est constitué de l'ensemble des personnes âgées de 18 à 65 ans au moment de l'enquête (soit les personnes nées entre 1938 et 1985 inclus pour l'enquête FQP de 2003) appartenant à des ménages ordinaires et résidant en France métropolitaine. Un ménage ordinaire est constitué de l'ensemble des personnes qui vivent habituellement dans le même logement à caractère privé (on exclut les collectivités, comme les prisons, les maisons de retraite, les communautés religieuses ou les foyers de jeunes travailleurs). Il s'agit là du champ théorique de l'enquête.

En pratique, une autre condition a été ajoutée afin de faciliter la collecte : appartenir à un logement où il y avait une personne âgée de 14 à 61 ans au moment du recensement de 1999. Cette restriction dans l'échantillonnage a été introduite, après s'être assuré qu'elle n'entraînait pas de distorsion importante dans l'échantillon.

### **L'échantillon**

L'enquête se donne comme objectif d'avoir une vision fine à la fois du parcours scolaire des personnes interrogées et de leur profession, pour juger de l'adéquation entre

formation et emploi. Cette confrontation nécessite un échantillon de taille importante, pour que chaque groupe pertinent de professions et de diplômes soit suffisamment bien représenté. Un échantillon cible de 40 000 personnes répondantes a été retenu, permettant de s'assurer d'une bonne représentation de la plupart des groupes sociaux, sur lesquels porte à priori l'intérêt, à savoir les personnes sans diplôme, les indépendants, etc. Une telle taille d'échantillon permet de tirer l'échantillon sans procéder à des sur-représentations de certains groupes.

Cependant, pour parvenir à un échantillon d'une telle taille, en maintenant un budget raisonnable, il est nécessaire d'interroger deux personnes dans chaque logement échantillonné, dès que cela est possible. Ainsi, étant donné l'échantillon de départ de 39 877 logements (tirés dans l'échantillon-maître de l'Insee), et compte-tenu des refus, de la présence de logements occasionnels ou vacants et du fait que les trois quarts des ménages comportent au moins deux personnes de 18 à 65 ans, on obtient un échantillon exploitable d'environ 40 000 personnes répondants à l'enquête.

## **Le déroulement de l'enquête**

### **1. Le protocole de l'enquête**

Le protocole de l'enquête prévoit une ou deux interviews par logement : dès que trois personnes ou plus du ménage ont entre 18 et 65 ans, deux personnes sont tirées au sort pour répondre à l'enquête. Ce sont les deux premières dans la liste alphabétique des prénoms : des analyses préparatoires ont montré que cette procédure était extrêmement proche d'un choix purement aléatoire. Cette méthode facilite par ailleurs grandement le contact avec le ménage (ainsi, cette méthode est préférable à un tirage au sort par l'ordinateur car ce tirage nécessiterait l'allumage de l'appareil, ce qui n'est pas toujours souhaitable). Le recours à une tierce personne pour répondre à l'enquête est interdit, compte-tenu de la précision des questions posées et pour assurer une bonne qualité à l'enquête. Ces contraintes sont bien sûr sources de difficultés, mais la période de collecte assez longue qui a été prévue : 3 mois (d'avril à juin 2003) et la possibilité d'effectuer l'enquête en plusieurs visites permettent de faciliter le travail des enquêteurs.

La collecte se déroule en face-à-face, avec l'aide d'un ordinateur - méthode dite de Collecte Assistée Par Informatique ou encore CAPI). Des supports papiers sont prévus pour faciliter la saisie de certaines informations. C'est le cas, en particulier, pour la partie concernant le parcours professionnel récent : un calendrier professionnel sur papier permet de cerner rapidement la mobilité professionnelle de la personne interrogée dans les cinq années précédant l'enquête.

Le CNRS, financeur de l'enquête, a souhaité que des entretiens avec des sociologues soient possibles après l'enquête afin de valider de façon plus qualitative les informations standardisées du questionnaire. Il a donc été prévu de demander aux enquêtés s'ils acceptaient de recevoir un chercheur pour un entretien plus approfondi, six mois à un an après la réalisation de l'enquête elle-même.

### **2. Le déroulement sur le terrain**

La méthode d'enquête a été conçue de manière à éviter, dans la mesure du possible, les erreurs d'observation susceptibles d'être commises sur le terrain.

Les enquêteurs disposent d'une fiche-adresse où figure l'adresse du logement à enquêter. Avant le début de la collecte, les enquêteurs effectuent une tournée de repérage pour reconnaître les logements et les occupants concernés. Ensuite commencent les entretiens. Dans un premier temps, les enquêteurs renseignent le tableau de composition du ménage (TCM). L'enquêteur dresse la liste de tous les membres du ménage, en indiquant prénom, sexe, date de naissance et les relations conjugales et familiales entre les différentes personnes. Une ou deux personnes sont ensuite sélectionnées comme indiqué plus haut pour répondre au questionnaire proprement dit. Pour les autres, quelques questions sont posées sur le niveau d'études et la situation professionnelle.

## **La qualité de l'enquête**

La qualité de l'enquête dépend en premier lieu du bon déroulement de la collecte et en second lieu des traitements (redressements) effectués sur les données recueillies.

### **1. La qualité de la collecte**

Au stade de la collecte, le succès de l'enquête suppose le repérage correct des logements

échantillonnés, un questionnaire d'enquête adapté au sujet traité et au public interrogé, ainsi qu'un réseau d'enquêteurs qualifiés. Dans l'ensemble, ces conditions ont été remplies. La méthode d'échantillonnage, quant à elle, est éprouvée depuis longtemps. Le questionnaire a été mis à jour par rapport aux éditions précédentes et des efforts ont été faits pour améliorer la formulation des questions et l'adéquation du questionnaire aux nouveaux besoins d'information.

Malgré tout ces efforts, les résultats d'une enquête restent imparfaits : ils comportent une certaine part d'aléa, due au fait qu'ils sont obtenus à partir d'un échantillon. En outre, certains ménages refusent de répondre ou sont impossibles à joindre ; le dénombrement des logements ou de leurs occupants peut comporter des erreurs ou des omissions ; enfin, certaines questions (le salaire, la profession, l'adresse de l'employeur, le diplôme, etc.) peuvent être sujettes à des erreurs, des imprécisions ou même au refus de répondre de certaines personnes interrogées.

Sur les 39 979 logements de l'échantillon effectivement identifiés sur le terrain (certains des logements de la base de sondage ont pu être éclatés en plusieurs logements, aussi le nombre de logements identifiés sur le terrain peut-être légèrement supérieur au nombre de logements tirés dans l'échantillon), 16% n'ont pas pu donner lieu à une enquête car ils ont disparu ou ne sont pas habités de façon

régulière durant la période de collecte (cf. tableau). Pour 1% des autres logements, l'enquête n'a pu être réalisée car la personne occupant le logement était inapte à répondre (problèmes mentaux graves notamment). Par ailleurs, des logements se retrouvent hors-champ par rapport à la tranche d'âge concernée par l'enquête : seules des personnes âgées de moins de 18 ans ou de plus de 65 ans sont présentes.

En pratique, l'enquête n'était possible que dans 31 577 logements, soit 79 % de l'échantillon total. Parmi eux, environ un sur cinq n'a pas donné lieu à une enquête du fait d'un refus ou d'une impossibilité de joindre le ménage. Des refus ont également été constatés après le passage du « tronc commun à tous membres du ménage », en particulier quand une seule des deux personnes possibles avait répondu au questionnaire. Quand il n'y avait qu'une personne dans le champ, le taux d'acceptation est de 99 % (7 809/7 896). Quand il y avait deux personnes dans le champ, on observe deux refus dans moins de 1 % des cas (148/16 981), un refus et une acceptation dans 13 % des cas (2 163/16 981) et deux acceptations, sinon.

Au total, l'enquête a pu être réalisée dans 24 642 logements (plus de 90 % des logements identifiés sur le terrain), dont plus de la moitié ont donné lieu à deux enquêtes, soit un total de 39 312 personnes interrogées.

### Résultat de la collecte

	Effectif	%
<b>(1) : Fiches Adresses initiales</b>	<b>39 979</b>	
(2) : Logements détruits, absorbés, etc.	714	2
(3) : Résidences secondaires ou occasionnelles, logements vacants,	5 518	14
(4) : Inaptes	497	1
(5) : Logements ne comportant aucune personnes de 18-65 ans	1 673	4
<b>(6) : N'ont pas pu donner lieu à une enquête = (2+3+4+5)/(1)</b>	<b>8 402</b>	<b>21</b>
(7) : Impossible à joindre et absents de longue durée	4 109	13
(8) : Refus au niveau du ménage	2 591	8
<b>(9) : Les refus de ménages et les personnes injoignables = (7+8)/(1-2-3-4-5)</b>	<b>6 700</b>	<b>21</b>
(10) : deux refus individuels (sur 2)	148	
(11) : un refus individuel (sur 1)	87	
(12) : un refus et une acceptation (sur 2)	2 163	
(13) : une acceptation (sur 1)	7 809	
(14) : deux acceptations (sur 2)	14 670	
(15) : Acceptations et refus individuels de l'enquête et taux d'acceptation = (12+13+14x2)/(10x2+11+12x2+13+14x2) *	24 877	94
<b>(16) : Echantillon exploitable (nombre de personnes répondantes) = 12+13+14x2</b>	<b>39 312</b>	

\* contrairement aux autres taux, celui-ci est calculé au niveau individuel et non des ménages. Il permet d'apprécier le taux d'acceptation individuel.



## 2. Les redressements

Comme pour toutes les enquêtes réalisées auprès des ménages, le redressement des résultats de l'enquête a deux objectifs principaux : compenser le biais introduit par les non-réponses et réduire autant que possible les fluctuations d'échantillonnage.

### a. Le redressement pour corriger la non-réponse

Certains ménages concernés par l'enquête n'ont pu être interrogés par suite d'absence de longue durée ou de refus de répondre. Dans la pratique, les non-répondants ne se répartissent pas aléatoirement dans l'échantillon : le taux de non-réponse pour les ménages d'une seule personne est plus élevé que celui pour les ménages comportant plusieurs personnes, il est également plus élevé dans les grandes villes que dans les campagnes, dans les immeubles que dans les maisons individuelles. L'existence d'un digicode est un facteur entravant l'enquête dans les immeubles collectifs, tandis que les maisons individuelles apparaissent plus accessibles. Les refus sont aussi plus fréquents dans les logements les plus anciens. Les personnes habitant un logement, classé résidence occasionnelle ou secondaire ou vacante au recensement, refusent l'enquête un peu plus souvent que la moyenne. Parmi les résidences déjà occupées au recensement, le taux de réponse croît avec le nombre de pièces.

Si le refus de répondre est corrélé avec certaines des variables que l'on cherche à mesurer, on introduit une distorsion dans l'échantillon en faisant abstraction des non-réponses. Le principe du redressement consiste à remplacer les ménages défectueux par des ménages présentant autant que possible les mêmes caractéristiques. Cela revient à faire l'hypothèse que pour les caractéristiques choisies, les ménages ayant accepté l'interview ne diffèrent pas de ceux qui l'ont refusée ou n'ont pas pu être joints par l'enquêteur. Ces caractéristiques doivent donc être connues à la fois pour les répondants et pour les non-répondants.

Pour les logements neufs (1 644 logements), le modèle de correction de la non-réponse repose sur une seule caractéristique, la tranche d'unité urbaine, faute d'autre information pertinente.

Pour les logements recensés en 1999, on prend en compte :

- la taille de l'agglomération,
- la date de construction du logement,
- les facilités d'accès à l'immeuble (présence d'un digicode),
- la distinction entre « logements occasionnels » et « résidences secondaires »,
- le nombre de pièces du logement,
- le nombre de personnes vivant dans le logement,
- l'existence ou non du logement au moment du recensement de 1999 (repérage des logements neufs).

Le redressement de la non-réponse est réalisé par la méthode « de calage sur marges » (ou calage sur des marges de référence). Cette méthode consiste à modifier les poids des personnes répondant de telle sorte que le total de certaines variables (en l'occurrence celles citées ci-dessus) estimé à partir des répondants de l'échantillon soit égal au vrai total connu par ailleurs.

### b. Le redressement pour fluctuations d'échantillonnage

Lorsque l'on tire au sort un échantillon pour une enquête, il arrive fréquemment que l'échantillon obtenu ne soit pas exactement représentatif de la population de référence, au regard de certains critères importants pour le sujet de l'enquête. Par exemple, on sait qu'en 2003, il y avait 50,6 % de femmes dans la population métropolitaine française. La proportion de femmes ayant répondu à l'enquête, après correction de la non-réponse, est de 51,4 %. De légères distorsions existent aussi selon l'âge : la proportion de personnes de 45 ans ou plus est plus élevée dans l'enquête que la proportion réelle. Cette différence entre la vraie valeur et la valeur fournie par l'enquête s'appelle une fluctuation d'échantillonnage.

Ces fluctuations d'échantillonnage peuvent aussi être redressées par une modification des poids des individus de l'échantillon (repondération). Corriger les fluctuations d'échantillonnage suppose d'avoir des informations issues d'une autre source que l'enquête et de meilleure qualité, ou mieux encore de la « vraie » valeur. Les possibilités de correction se trouvent de ce fait limitées.

Les critères sur lesquels s'effectuent la repondération sont l'âge au 31 décembre de

l'année (en tranches d'âge quinquennal) et le sexe. Le redressement modifie donc le poids de chaque répondant en sorte que le nombre de personnes de même sexe et de même groupe d'âge quinquennal estimé par l'enquête soit égal au nombre de personnes de même sexe et de même âge quinquennal connu par ailleurs (informations démographiques issues de l'actualisation par les données de l'état civil des données du recensement de la population).

Un calage sur les marges de référence a été effectué (à l'aide de la macro CALMAR - CALage sur les MARGes - développée par l'Insee). Le rapport entre le poids final et le poids avant calage est en moyenne de 1,11. Le minimum est de 0,91 et le maximum de 1,23.

### **c. La combinaison des deux types de redressement**

Les deux types de pondération, pour correction de la non-réponse et pour fluctuations d'échantillonnage, peuvent se faire soit successivement (transformation des poids pour correction de la non-réponse, puis modification des poids obtenus pour correction des erreurs aléatoires), soit simultanément. Dans l'enquête, on a procédé en deux étapes, d'abord une correction de la non-réponse, ensuite une correction des fluctuations d'échantillonnage.

## **3. Recodage et traitement des données**

### **a. Codage de la profession et de l'activité de l'entreprise**

L'enquête recueille un grand nombre d'informations sur les professions exercées à divers moments de la carrière ou par d'autres membres de la famille (parents, frère ou sœur, grands-parents). Ces informations sont en partie recueillies en clair, c'est-à-dire que l'intitulé de la profession exercée est enregistré tel que le donne l'enquêté, sans chercher à se placer dans une quelconque nomenclature. L'ensemble de ces informations permet de classer chaque profession décrite dans la nomenclature française des professions et catégories professionnelles (dite PCS). Cette nomenclature comporte 486 postes. La majeure partie de la codification des réponses dans la nomenclature PCS ou dans le niveau agrégé de cette nomenclature (comportant 42 postes) s'est faite automatiquement. Pour cela, le logiciel SICORE (Système Informatique de

CODification des Réponses aux Enquêtes), l'outil de codification automatique développé par l'Insee, a été utilisé.

Les libellés non codés automatiquement l'ont été manuellement. Cette étape de codification manuelle a été réalisée par un groupe de 8 personnes spécialisées dans ce type de travail, localisé dans la direction régionale d'Aquitaine ; une autre équipe localisée dans la direction du Nord - Pas-de-Calais a aussi participé à cette codification manuelle. De même, le secteur d'activité de l'établissement employeur a été codé par appariement avec les fichiers SIRENE, à partir de la raison sociale de l'employeur et l'adresse déclarées par l'enquêté. En cas d'échec, le codage était manuel, à partir du secteur d'activité déclaré par l'enquêté.

### **b. Les revenus**

La dernière partie du questionnaire interroge sur les revenus professionnels perçus au cours de l'année précédent l'enquête. Le concept retenu est celui du revenu déclaré pour le calcul de l'impôt sur le revenu : ainsi, l'enquêté doit normalement inclure la CSG imposable à son revenu professionnel. Cependant, dans certains cas, l'enquêté n'a pas pu ou voulu indiquer le montant exact de son salaire, mais il a accepté d'indiquer l'intervalle dans lequel son revenu du travail ou son salaire se situait. Dans certains cas, il a refusé de répondre à ces questions. On a imputé un revenu du travail ou un salaire aux personnes ayant refusé de répondre à cette question, de même qu'à celles ayant indiqué seulement un intervalle.

## **4. L'interprétation des résultats**

Les données de l'enquête sont, comme toutes les enquêtes par sondage, affectées d'erreurs qui peuvent être mesurées par l'écart-type de chacune des estimations. L'estimation de ces erreurs aléatoires repose sur deux éléments :

- d'une part, le mode d'élaboration des données, c'est-à-dire le type d'échantillonnage décrit précédemment ;
- d'autre part, le fait que l'enquête est recalée sur une information qui lui est exogène et qui est considérée comme certaine (redressements).

L'échantillon utilisé pour faire cette enquête est l'un des multiples échantillons qu'il aurait été possible de constituer avec le plan de sondage utilisé. Chacun de ces échantillons aurait



donné des chiffres plus ou moins différents. La technique d'échantillonnage utilisée est sans biais, c'est-à-dire que pour un chiffre donné, la moyenne des estimations de chacun des échantillons que l'on aurait pu constituer par cette technique est égale à ce qu'on obtiendrait en menant l'enquête auprès de l'ensemble de la population. L'estimation de ce chiffre par notre échantillon particulier risque par contre de s'écarter de la moyenne. En particulier, si on étudie une population très peu nombreuse, cette population peut être quasi-absente de l'échantillon alors qu'elle existe dans la population totale. C'est pourquoi, il faut interpréter avec prudence les faibles effectifs

des tableaux. Ainsi, il faut garder en mémoire qu'un effectif de 2 000 dans un tableau ne correspond qu'à environ 2 personnes dans l'échantillon. Une case vide ne signifie pas qu'aucun individu en France ne répond aux critères étudiés, mais plutôt qu'aucun ne se trouve dans l'échantillon (ce qui indique malgré tout que les personnes de ce type sont probablement très peu nombreuses en France métropolitaine).