

## Questions réponses sur le panel européen (version française)

du 27/01/2000 au 13/09/2000

<b>SUIVI DES INDIVIDUS ET DES MENAGES</b>	<b>3</b>
Que signifient les variables TNOC, TPRESHA et TRINT ?	3
Quand un ménage A se sépare en deux ménages B et C; les deux nouveaux ménages gardent-ils l'identifiant IDENT de A? Et si oui comment les distingue t-on alors entre eux ? Sinon, comment sait-on qu'ils sont issus du ménage A ?	5
Comment est déterminé l'identifiant de deux ménages qui fusionnent ?	5
Est-il possible de savoir pour les personnes interrogées en 1994 la date d'arrivée dans le logement ? J'en ai besoin pour reconstituer la famille en 1993.	5
Comment reconstitue-t-on le ménage pour 1993, sachant qu'on a peu d'info sur les personnes ayant quitté le ménage avant l'enquête de 1994 (on connaît seulement la date de départ) ?	5
Pour reconstituer le ménage, quelle solution adopter pour les individus dont TNOC n'est pas connu ?	6
Comment sélectionner les individus répondants ?	6
Comment retrouver les enfants panel nés entre deux vagues (le test TNOC='22' semble peu satisfaisant) ?	7
Que penser des individus dont la date de changement de ménage est antérieure à la vague d'enquête précédente ?	7
Comment est définie la personne de référence, et est-ce elle qui répond au questionnaire ménage ? Et qui est présent ?	8
Les individus panels sont-ils représentatifs de la population, et a-t-on des variables de redressement spécifiques pour ces individus ?	10
Comment sont recueillies les dates d'entrée dans le nouveau ménage en cas d'éclatement de ménage (ainsi que les autres variables figurant sur le tableau permanent de composition du ménage) ?	10
<b>PRESENCE ET QUALITE DES VARIABLES - TABLE MENAGE</b>	<b>11</b>
Le principe de codification adopté en vague 1 pour le tableau F12 récapitulant les types de revenus perçus par les individus du ménage a changé pour les vagues suivantes.	11
Les variables de mois de perception d'aide sociale ne sont pas homogènes d'une vague à l'autre. Ainsi, la variable DE se nomme DEC en vague 3, et est soit à 2 caractères, soit à 3.	11
La variable VOIT (possession d'une voiture) est mauvaise en vague 2.	11
Le codage des variables QMA et PRESQM n'est pas homogène d'une vague à l'autre.	11
Comment définit-on la date de formation du ménage ?	11
<b>PRESENCE ET QUALITE DES VARIABLES - TABLE INDIVIDU</b>	<b>12</b>
Que signifie la variable QIA ?	12
Le NIDENT n'existe pas pour 38 individus.	12
Où est la variable TINPA (Individu panel) ?	13
Le code TNOC (mouvement de la personne) n'existe pas pour 676 individus en vague 2.	13
Est-ce normal d'avoir TPRES='2' (présence à l'enquête précédente='non') et TNOC='10' (personne présente à l'enquête et à l'enquête précédente) ?	13

La variable PVOIT (possession d'une voiture de société) a beaucoup de valeurs manquantes en vague 2.	13
La variable afinpa (année d'arrêt d'activité pour ceux qui ne travaillent plus) du fichier indiviv2 pose problème.	13
Les variables AMATRIE, ACOUPLE et ANACTU valent souvent '1900' en V1	13
Le codage de la variable PRESFR n'est pas homogène d'une vague à l'autre.	14
La variable LIEN est manquante en V2 pour 2 500 cas.	14
Quelle est la codification de la variable PCS (profession principale exercée) ? Et des variables PROPER et PROMER (dernières professions exercées par le père et par la mère) ?	14
Quel est le codage des variables OCHxD et PROFHxD concernant les enfants hors ménage ?	16
Les variables indiquant les changements d'entreprise (ADEBE et MDEBE) et d'établissement (TDETS) semblent parfois incohérentes.	16
<b>PRESENCE ET QUALITE DES VARIABLES - TABLE ACTIVITE</b>	<b>19</b>
Fichier activité : exiact est souvent à blanc et sur certaines lignes avec exiact='2', aucun mois d'activité n'est coché	19
Environ la moitié des variables du fichier activité de la vague 2 est en format numérique alors que toutes les variables des vagues 1 et 3 sont en format caractère.	19
NBACT : Nombre de lignes renseignées n'existe pas dans le fichier, est ce volontairement?	19
<b>PRESENCE ET QUALITE DES VARIABLES - TABLE REVENU</b>	<b>20</b>
En 1ère vague, on ne devrait pas connaître les mois de perception des revenus de 94. Dans le fichier les mois ont été remplis quand le nb total est 12.	20
Certains individus déclarant 12 mois de perception de revenu en 95 ou en 1996 ont des mois à vide.	20
Il semble que les montants de revenus déclarés inclus dans un autre revenu sont pris en compte dans le revenu individuel total calculé dans les tables de revenus simulés ?	20
Quel est le principe de l'imputation des revenus d'activité ?	21
Des revenus salariaux sont imputés à des non-salariés	21
Les revenus partiels sont-ils pris en compte pour les imputations de revenus ?	22
Certains revenus déclarés repris dans les tables de revenus simulés REVACVxS ne figurent pas dans la table REVENUV2	23

## **Suivi des individus et des ménages**

### **Que signifient les variables TNOC, TPRESHA et TRINT ?**

Ces variables, figurant dans le TABLEAU PERMANENT DE COMPOSITION DU MENAGE, permettent de suivre les mouvements des individus depuis la vague précédente (TNOC), de savoir si l'individu interrogé loge habituellement dans le ménage (TPRESHA), et quel a été le résultat de l'entretien (TRINT).

#### **Code A : Mouvement des personnes (TNOC)**

(colonne 11)

10. *Présent à l'enquête précédente, et toujours dans le ménage*

#### **Nouveaux membres (n appartenaient pas au ménage à la dernière enquête)**

21. *Personne arrivée dans le ménage depuis la dernière enquête (y.c. retour d'une personne ayant déjà appartenu au ménage)*

22. *Enfant (d'un des membres du ménage) né depuis la dernière enquête*

23. *Personne présente dans le ménage à la dernière enquête mais oubliée dans le Tableau de Composition du Ménage*

#### **Ne sont plus membres (appartenaient au ménage à la dernière enquête)**

31. *Parti dans un autre ménage ordinaire en France (y.c. D.O.M)*

32. *Parti dans un ménage collectif en France (y.c. D.O.M)*

33. *Parti dans une institution en France (y.c. D.O.M)*

34. *Parti à l'étranger dans un pays de l'Union Européenne*

35. *Parti à l'étranger dans un pays hors Union Européenne*

36. *Disparu, S.D.F., pas de nouvelle adresse connue...*

37. *Décédé*

40. *Ne fait pas partie du ménage, et n'en faisait pas partie à la dernière enquête*

**Attention**, la codification des codes TNOC tient compte des résultats des enquêtes téléphoniques qui ont eu lieu entre les vagues 1 et 2 (en février et en mai-juin), et entre les vagues 2 et 3 (mai-juin). Ainsi, il n'est pas anormal d'avoir un code TNOC égal à '40' (« ne fait pas partie du ménage et n'en faisait pas partie à la dernière enquête ») en vague 2, si l'individu était déjà parti avant l'enquête téléphonique précédant la deuxième enquête terrain. L'existence de ces vagues téléphoniques (dont le but était de ne pas perdre la trace des

individus panel) a en fait plutôt contribué à rendre plus complexe le suivi des individus d'une vague à l'autre, et créé des difficultés d'exploitation des fichiers relatifs aux enquêtes terrain.

**Code B : Présence habituelle (TPRESHA)**

(colonne 13)

1. *Vit ici régulièrement*

**Loge aussi ailleurs, habituellement un ou plusieurs jours par semaine :**

2. ... *dans une caserne*

3. °... *en internat, en cité universitaire*

4. °... *dans un foyer de travailleurs*

5. °... *dans un autre logement indépendant (dont il est locataire, propriétaire, ou occupant à titre gratuit)*

6. °... *chez quelqu'un (ami, famille,...) qui l'héberge*

7. °... *se déplace dans son travail, habituellement un ou plusieurs jours par semaine (loge à l'hôtel, personnel navigant, marin pêcheur,...)*

8. ° *Sans objet (personne absente pour plus de six mois pour des raisons de santé : hospitalisation,...)*

9. ° *Sans objet (personne absente pour plus de six mois pour d'autres raisons : prison,...)*

**Code C : Résultat de l'entretien (TRINT)**

(colonne 14)

**Questionnaire individuel renseigné**

11. *Entretien réalisé (même partiel)*

**Questionnaire individuel non renseigné**

20. *Personne de moins de 17 ans au cours de l'année*

21. *Personne incapable de répondre (maladie, incapacité,...)*

22. *Refus de coopérer*

23. *Personne de plus de 17 ans mais hors champ* Cette valeur ne doit logiquement pas exister pour les enquêtes « terrain » traitées dans ces fichiers (code créé exclusivement pour les enquêtes téléphoniques, car on n'interrogeait pas les individus de plus de 17 ans entrés dans le ménage depuis la vague terrain précédente)

31. *Personne temporairement absente (en congés,...)*

32. *Personne impossible à joindre malgré plusieurs essais*

33. *Autre raison*

**Quand un ménage A se sépare en deux ménages B et C; les deux nouveaux ménages gardent-ils l'identifiant IDENT de A? Et si oui comment les distingue t-on alors entre eux ? Sinon, comment sait-on qu'ils sont issus du ménage A ?**

=> Non, quand un ménage se sépare, on crée un "ménage éclaté", le code éclatement, qui est la dernière position de l'IDENT passe à 2 (puis 3...) pour le nouveau ménage.

Ex : en V1 existe le ménage 1100026001 composé de 4001484 et 4001485.

4001784 quitte le ménage en septembre 1995, soit entre les vagues 1 et 2.

On a donc dans la table ménageV2, deux ménages : l'ancien ménage 1100026001 (dans lequel ne reste que 4001485) et un nouveau ménage 1100026002 (avec 4001484).

Et dans la table indiviv2, on trouve les informations suivantes :

nident	ident	tnoc	tmoi	tan
4001485	1100026001	10 (présent à cette enquête et à la précédente)		
4001484	1100026001	31 (partant)	09	95
4001484	1100026002	21 (arrivant)	09	95

Ce qui indique que l'individu 4001484 a quitté le ménage en septembre 1995 pour aller créer un nouveau ménage. Bien sûr, si d'autres nouveaux ménages étaient créés à partir du même ménage de départ (ou du nouveau ménage créé), ils seraient indicés par 3 puis 4...

On peut donc suivre les individus soit à partir de leur nident, soit en regardant les premières positions de l'IDENT.

**Comment est déterminé l'identifiant de deux ménages qui fusionnent ?**

=> Le cas, peu probable, se produit effectivement dans le panel. En théorie, le nouvel ident est celui du chef du nouveau ménage (consigne d'Eurostat). En pratique, le nouvel ident est celui de l'individu que l'on est allé enquêter. Par exemple, il y a eu quelques fois le cas de jeunes (panel) qui retournaient vivre chez leurs parents (également panel), ce que l'on découvrait en allant enquêter les parents. Dans ce cas, on garde l'ident des parents, ce qui correspond également à la règle théorique. Mais dans d'autres cas (par exemple une mère et sa fille se mettent ensemble), il se pourrait que cela n'ait pas été le cas.

**Est-il possible de savoir pour les personnes interrogées en 1994 la date d'arrivée dans le logement ? J'en ai besoin pour reconstituer la famille en 1993.**

=> Oui, ces informations sont dans la table ménage, ce sont les variables ARR<sub>x</sub>M et ARR<sub>x</sub>A qui donnent la date d'arrivée dans le ménage de l'individu de noi NOI<sub>x</sub>A (si le ménage s'est créé en 1993 ou en 1994, tous les individus sont - en tout cas parfois - inscrits comme arrivants).

**Comment reconstitue-t-on le ménage pour 1993, sachant qu'on a peu d'info sur les personnes ayant quitté le ménage avant l'enquête de 1994 (on connaît seulement la date de départ) ?**

=> J'ai malheureusement bien l'impression que l'on ne sait effectivement rien sur ces individus (même pas s'il s'agit ou non d'enfants). Il faut donc sans doute faire une hypothèse pour le calcul d'unités de consommation.

### **Pour reconstituer le ménage, quelle solution adopter pour les individus dont TNOC n'est pas connu ?**

=> Ce problème ne se pose donc qu'à partir de la V2. En V2 seuls 13 individus ont un TNOC à blanc parmi ceux :

- qui ont un noi inférieur à '19' (c'est à dire en excluant les individus 21 et autres qui ont quitté le ménage et dont on ne collecte que les revenus),

- et qui ont également répondu (interro in ('R','P')).

Bien sûr, comme ces individus sont répondants, on peut considérer qu'ils sont présents dans le ménage ! Reste à savoir s'ils étaient déjà là avant. Je ne vois pas d'autre solution qu'aller récupérer l'info lors de la vague précédente (ou si l'on est en V2, regarder la tête du nident, s'il commence par un '4' -comme 1994- cela veut dire qu'ils étaient là depuis la V1.

### **Comment sélectionner les individus répondants ?**

=> Les adultes (plus de 17 ans) sont considérés comme répondants s'ils ont rempli un questionnaire individuel (même partiellement). Les enfants (moins de 17 ans), qui n'ont pas de questionnaire individuel à remplir, sont considérés comme répondants si leur ménage a été contacté (selon la recommandation d'Eurostat) : un ménage est contacté si une des variables TRESFA1 à TRESFA6 est comprise au sens large entre '11' et '17'. En pratique, on peut pour simplifier considérer qu'un enfant est répondant s'il appartient à un ménage répondant (PRESQM=1). De toute manière, déterminer si enfant est répondant ou non a peu d'intérêt car les enfants ne sont pas interrogés (on ne leur pose pas le questionnaire individuel, et ils n'ont donc pas de calendrier d'activité). Cependant, un enfant peut remplir une fiche revenus.

Pour déterminer les adultes répondants en V1 (i.e. les répondants avec questionnaire individuel), la variable RE peut être utilisée (RE='11'). Cette variable est collectée sur la fiche adresse. Elle fournit le renseignement demandé à 13 cas près, tandis que la variable INTERRO est moins bonne, et ne doit pas être utilisée en V1 (cette variable, créée après la collecte, sera recréée dans la prochaine version du CD-ROM).

En V2 et V3, INTERRO in ('R','P') permet de sélectionner exactement les adultes répondants (interrogés pour la première fois 'P' ou non 'R'), en laissant de côté les 4 jeunes interrogés (à tort) en V2 sur leurs études, et les deux personnes décédées qui ont un début de calendrier d'activité en V3.

Ce test sur INTERRO permet également de ne conserver que l'information relative au ménage d'arrivée lorsqu'un individu de plus de 17 ans apparaît deux fois (dans son ménage de départ et dans son ménage d'arrivée). Cela peut aussi se faire en ne conservant que les enregistrements pour lesquels le code TNOC (mouvement des personnes) est inférieur à '30'. Ainsi, pour l'individu 4001484 présent en V1 dans le ménage 1100026001 et en V2 dans le ménage 1100026002, deux enregistrements figurent dans la table individu V2 :

IDENT	NIDENT	TNOC	TMOI	TAN	INTERRO
1100026001	4001484	31	09	95	0
1100026002	4001484	21	09	95	R

Ne retenir que les enregistrements pour lesquels  $TNOC < '30'$  ou retenir ceux pour lesquels INTERRO in ('R', 'P') revient dans les deux cas à ne conserver que le second enregistrement pour cet individu (qui est l'enregistrement contenant les réponses au QI).

Au total, on obtient :

- 14333 individus adultes répondants en V1
- 13308 individus adultes répondants en V2
- 13052 individus adultes répondants en V3.

ATTENTION, les individus répondants n'ont pas forcément rempli un questionnaire revenu, et inversement, des individus n'ayant pas rempli de questionnaire individuel ont pu remplir une fiche revenus (cas des jeunes notamment).

### ***Comment retrouver les enfants panel nés entre deux vagues (le test $TNOC = '22'$ semble peu satisfaisant) ?***

=> Les enfants nés après la vague 1 ne sont panels que si leur mère l'est. Ils se voient alors affecter un poids égal à la moitié du poids longitudinal de leur mère. Ce qui nécessite bien sûr - pour que la mère ait un poids - que celle-ci soit répondante (et l'ait été lors des vagues précédentes). Mais il y a probablement quelques enfants nés entre deux vagues qui ne sont pas panel (mère non panel) ou qui n'ont pas de poids (mère interrogée en V1 et V3 et non interrogée en V2 par exemple).

La variable TNOC doit a priori valoir '22' pour tous les enfants nés entre deux vagues, mais il semble que les enquêteurs n'ont pas toujours respecté le codage.

Dans le suivi des individus que nous effectuons, nous trouvons :

- 158 naissances d'enfants panel "répondants" entre les vagues 1 et 2 (les enfants étant considérés comme répondants dès que le ménage auquel ils appartiennent a été contacté). Comme vous le signalez, seuls 71 ont un TNOC égal à 22, et 78 ont un TNOC égal à 10 !
- 188 naissances d'enfants panel "répondants" entre les vagues 2 et 3.

### ***Que penser des individus dont la date de changement de ménage est antérieure à la vague d'enquête précédente ?***

=> Une trentaine d'individus V3 ont changé de ménage avant l'enquête précédente. J'ai regardé les deux cas suivants pour avoir une idée :

- 4001200 : cet individu parti en 08/95 appartenait en V2 à un ménage sans individu répondant (code trint=22, refus de coopérer). Donc on ne pourra de toute façon pas reconstituer de l'information sur ce ménage en V2. (Sinon, on s'aperçoit en regardant la table individu qu'une autre personne a également quitté le ménage en 08/95 et que par ailleurs, les membres de ce nouveau ménage déclarent s'être mis en couple en 08/95 (AMATRI). Donc la date de 08/95 semble bonne).

- 4005002 : c'est un autre cas apparemment : l'individu, parti en 07/95 est répondant en V3, et ne l'était pas en V2 (absence temporaire). Le code 'absence temporaire' était vraisemblablement erroné en V2 (l'individu était déjà parti), car il est arrivé dans son nouveau ménage en même temps que Madame 6001574 (arrivée en 07/95) et il déclare aussi s'être mis en couple en 1995. Donc on peut considérer que le mouvement a bien eu lieu en 7/95.

**Comment est définie la personne de référence, et est-ce elle qui répond au questionnaire ménage ? Et qui est présent ?**

=> La définition de la personne de référence est commune à toutes les enquêtes Insee et est définie dans les Instructions aux enquêteurs sur le tronc commun des enquêtes, suivant la carte ci-dessous.

**CARTE 1**

**DÉTERMINATION de la PERSONNE de RÉFÉRENCE dans le LOGEMENT**

<p><b>A - Il existe un ou plusieurs couples ou de fait avec ou sans enfant</b></p>	→ 1 couple	→	Personne de référence : le conjoint masculin marié ou de fait
	→ 2 ou plusieurs couples	→	Personne de référence : parmi les conjoints masculins mariés ou de fait* le plus âgé ou, à défaut d'actif, le plus âgé
<p><b>B - A défaut de couple, il existe une famille monoparentale : une mère (ou père) sans conjoint avec des enfants célibataires, ces enfants étant eux-mêmes sans conjoint et sans enfant</b></p>	→ 1 famille monoparentale	→	Personne de référence : la mère (ou le père) des enfants
	→ 2 ou plusieurs familles monoparentales	→	Personne de référence : parmi les mères (ou les pères) des enfants, la personne active* la plus âgée, ou à défaut de personne active, la personne la plus âgée
<p><b>C - A défaut de couple et à défaut de famille monoparentale</b></p>	→ Toutes les personnes du logement sont considérées comme des personnes isolées	→	Personne de référence : parmi toutes les personnes « isolées », la personne active* la plus âgée ou à défaut de personne active, la



*personne la plus âgée*

\* y compris les chômeurs

**Cas particulier :**

- *Les salariés logés chez l'employeur, même s'ils constituent un couple ou une famille monoparentale, ne seront jamais retenus comme personne de référence. Choisir la personne de référence parmi le(s) autre(s) couple(s) ou autre(s) famille(s) monoparentale(s) ou autre(s) personne(s) isolée(s) du ménage.*
- *L'ascendant (grand-père) ou grand-mère) sans conjoint élevant seul des petits enfants ne constitue pas une famille monoparentale. Dans le cas, l'ascendant et ses petits-enfants seront assimilés à des « isolés ».*
- *En aucun cas, un enfant de moins de 15 ans ne peut être retenu comme personne de référence.*

**CODE - LIEN avec la PERSONNE de RÉFÉRENCE du MÉNAGE**

- 1- *Personne de référence*
- 2- *Conjoint de la personne de référence, marié ou de fait (la femme dans un couple)*
- 3- *Enfant de la personne de référence ou de son conjoint : fils, fille, gendre, bru, beau-fils, belle-fille, enfant adopté, enfant en tutelle*
- 4- *Petit-enfant de la personne de référence ou de son conjoint : petit-fils, petite-fille*
- 5- *Ascendant de la personne de référence ou de son conjoint : mère, père, beau-père, belle-mère, grand-père, grand-mère, etc.*
- 6- *Autre parent de la personne de référence ou de son conjoint : frère, soeur, neveu, nièce, cousin, cousine, beau-frère, belle-soeur, oncle, tante, etc.*
- 7- *Ami*
- 8- *Pensionnaire, sous-locataire, logeur, enfant en nourrice sans lien de parenté avec la personne de référence ou de son conjoint*
- 9- *Domestique ou salarié logé*

Les consignes aux enquêteurs pour l'enquête panel stipulent que le répondant doit être la personne de référence ou son conjoint. En vague 1, l'identification de la personne ayant répondu n'existe pas dans le fichier. En revanche, pour les vagues suivantes, on connaît le numéro d'individu dans le ménage du répondant (variable noirm en V2 et V3).

Pour trouver les personnes de référence du ménage, il faut regarder la variable lien (lien='1'), car ces personnes de référence n'ont aucune raison a priori d'avoir un noi égal à 01.

Par ailleurs, normalement, il pourrait y avoir des changements de personne de référence entre les vagues, mais on n'a bien sûr aucune garantie que ces changements aient bien été faits dans le panel. Les problèmes de non-permanence de la personne de référence ne sont pas tant liés

aux définitions retenues pour la personne de référence qu'à la mobilité des personnes et à l'évolutivité de la composition du ménage. Quelque soit la définition de la personne de référence retenue, on ne sera pas en mesure d'assurer sa permanence. Les discussions au niveau d'Eurostat ne visent d'ailleurs qu'à avoir une définition cohérente entre les différents pays de l'UE. Le changement de personne de référence n'est pas uniquement lié aux divorces comme pourrait le laisser entendre la question sur le taux de divorce. Les mariages, les pertes d'emploi des plus anciens lorsque plusieurs générations cohabitent ou encore les décès peuvent également être source de changements, etc.

Il y a pour les questionnaires individuels une indication sur les personnes présentes lors de l'interrogation sur la partie 'relations et responsabilités sociales'. Bien sûr, cela ne correspond pas forcément aux personnes présentes lors du questionnaire ménage. Mais peut-être est-il possible de faire des hypothèses (par exemple si tous les questionnaires individuels du même ménage ont lieu le même jour et en présence des mêmes personnes, on peut supposer que le QM est aussi passé avec ces mêmes personnes ?). Cette hypothèse reste osée car certaines expériences de collecte montrent que dans un ménage comportant un grand nombre d'individus, il peut y avoir beaucoup de va et vient pendant la collecte.

### ***Les individus panels sont-ils représentatifs de la population, et a-t-on des variables de redressement spécifiques pour ces individus ?***

=> Les individus panel sont les individus qui ont été interrogés lors de la première vague, et les nouveaux-nés de mère panel nés après. Ils doivent bien sûr être réinterrogés à chaque vague (sauf s'ils sont non-répondants deux vagues de suite). Sont également enquêtés à la vague courante tous les individus non panel qui partagent le logement d'un individu panel.

Il existe deux types de variables de redressement dans le panel :

- des variables de pondérations dites longitudinales (pondm en V1 dans la table ménage car en V1 pondération longitudinale et transversale sont identiques et identiques pour tous les membres du ménage, basev2 en V2 et basev3 en V3, dans les tables individu). Les pondérations longitudinales permettent de rendre les individus panel qui ont répondu à toutes les vagues jusqu'à la vague n représentatifs de la population dans le champ de l'enquête (i.e. hors ménages non ordinaires du type institution et hors nouveaux immigrants). Ainsi, si on veut étudier l'évolution d'une variable X entre la vague n et une des vagues précédentes, on utilise la pondération longitudinale de la vague n.

- des variables de pondération transversales, utilisées pour les analyses en coupe, qui permettent de rendre l'ensemble des individus interrogés (panel et non-panel) représentatifs de la population dans le champ de l'enquête cette année-là. Ces pondérations permettent donc d'utiliser l'information sur l'ensemble des individus enquêtés, et non uniquement les individus panels. Elles s'appellent PONDM, TRANSV2C et TRANSV3 et se trouvent dans les tables ménages.

### ***Comment sont recueillies les dates d'entrée dans le nouveau ménage en cas d'éclatement de ménage (ainsi que les autres variables figurant sur le tableau permanent de composition du ménage) ?***

=> Le nouveau Tableau Permanent de Composition du Ménage (correspondant au ménage qui se crée) est créé en Direction Régionale par des agents de l'Insee. La date d'entrée dans le nouveau ménage y est effectivement égale à la date de sortie du ménage précédent.

## **Présence et qualité des variables - Table ménage**

**Le principe de codification adopté en vague 1 pour le tableau F12 récapitulant les types de revenus perçus par les individus du ménage a changé pour les vagues suivantes.**

=> Effectivement, on a par exemple :

Vague 1 : RMI36fF et RMI46F représentent la perception du RMI en 1993 et 1994 respectivement (3 pour 93 et 4 pour 94) par l'individu NOI6F

Vague 2 et + : RMI36F et RMI46F sont remplacés par une seule variable RMI5F6. Le 6 n'est plus à sa place et le 5 est conservé à la vague 3.

Il faudra changer cela pour la prochaine version du CD-ROM

**Les variables de mois de perception d'aide sociale ne sont pas homogènes d'une vague à l'autre. Ainsi, la variable DE se nomme DEC en vague 3, et est soit à 2 caractères, soit à 3.**

=> Oui, il semblerait que l'on soit dans la situation suivante pour les variables de mois de perception d'aide des services sociaux :

vague	nom	valeurs possibles
V1	à 2 lettres (DE pour décembre)	numéro du mois (12 valeurs )
V2	à 2 lettres (DE pour décembre)	1 ou 2
V3	à >2 lettres (DEC pour décembre)	1 ou 2

Ce problème devrait être réglé dans le prochain CD-ROM !

**La variable VOIT (possession d'une voiture) est mauvaise en vague 2.**

=> En effet, deux régions de traitement sont concernées : le Nord Pas-de-Calais (Région 31) et les Pays de Loire (Région 52). Pour ces régions, la variable VOIT est renseignée au niveau ménage avec ce qui devrait être dans PVOIT au niveau individu. et l'information sur VOIT est perdue.

**Le codage des variables QMA et PRESQM n'est pas homogène d'une vague à l'autre.**

=> Effectivement, les années de passation du questionnaire ménage (QMA) sont codées sur des variables caractères à quatre positions en première vague (19xx), et sur des variables caractères à deux positions lors des deux vagues suivantes. La variable PRESQM (présence d'un questionnaire ménage) est codée en numérique sur trois positions en V1, puis en caractère sur une position lors des deux vagues suivantes (il en est de même de la variable PRESF dans la table individu). Ces points devraient être revus dans la prochaine version du CD-ROM.

**Comment définit-on la date de formation du ménage ?**

=> La date de formation du ménage est demandée dans le questionnaire ménage. Sa définition a fait l'objet de nombreux débats en comité de pilotage panel. Ainsi, doit-on considérer qu'on change de ménage quand quelqu'un arrive ? réponse oui. Même si c'est une naissance ?

réponse non. Et si c'est un enfant qui revient ? ça dépend notamment de l'âge de l'enfant. Doit-on considérer qu'on change de ménage quand quelqu'un part ? réponse oui. Même si c'est un décès ? réponse non. Et en cas de divorce ? oui. Et si c'est un enfant qui part de chez ses parents ? là c'est non. Change-t-on de ménage quand on change de logement ? oui et non.

Une partie du questionnement du questionnaire ménage fait explicitement référence à la date de création du ménage. Il était donc nécessaire d'avoir une définition de ce qu'est un nouveau ménage et cette définition devait être applicable sur le terrain (compte tenu des exemples ci-dessus ce n'était pas simple!). On souhaitait par ailleurs avoir une définition qui permette d'assurer une continuité dans le temps. On a donc considéré que le nouveau ménage avait pour date de création, la date de départ de l'individu panel de son ménage d'origine.

Encore un point sur ces notions de ménage et de date d'entrée dans le ménage. La notion de ménage Panel est étroitement liée aux individus panel composant ce ménage. Tant que les individus panel composant un même ménage restent ensemble, le ménage panel reste également le même. Ca semble simple comme définition mais les conséquences en sont les suivantes,

Prenons un exemple :

Un jeune homme qui est individu Panel vit avec une jeune femme qui n'est pas Panel.

Le jeune homme quitte la jeune femme. La jeune femme continue de résider à la même adresse et c'est le jeune homme qui part.

Le jeune homme en quittant la jeune femme retourne dans le ménage de ses parents (les parents n'étant pas panel).

Conséquences:

C'est la jeune femme qui part du ménage (la date de son départ correspond à la date à laquelle le jeune homme est effectivement parti du logement).

Les parents du jeune homme entrent ensuite dans le ménage du jeune homme.

---

## ***Présence et qualité des variables - Table individu***

### ***Que signifie la variable QIA ?***

=> Elle représente l'année de l'interview dans le questionnaire individuel. Un filtre dans le tableau de composition du ménage portait également le nom de QIA, mais a été changé en QIAC dans la base ménage.

### ***Le NIDENT n'existe pas pour 38 individus.***

=> Il s'agit de deux types d'individus :

- les individus qui ont un NOI (numéro d'individu) 21, c'est-à-dire ceux qui ne sont plus dans le ménage mais qui ont perçu des revenus. Pour ces individus, les seules informations disponibles sont les informations concernant la passation de la fiche revenu (date, durée...). En vague 1, ce type d'individu n'a pas été conservé dans la table individu (car en V1, il n'y avait de toute façon aucune information sur les conditions de passation de la fiche revenu !). On va réfléchir pour voir s'il faut les garder en V2.

- des individus arrivés dans le ménage lors des vagues téléphoniques entre la première et la seconde vague, et non-répondants lors de la seconde vague (INTERRO not in ('R','P')). Ces individus n'ont pas eu de nident. Il faut que l'on réfléchisse pour savoir si on les garde. De manière générale, il y a un gros travail à faire sur le panel pour savoir qui on garde dans les fichiers. Pour l'exploitation, ces problèmes ne sont pas trop gênants, car ils ne se posent que pour les non-répondants.

### ***Où est la variable TINPA (Individu panel) ?***

=> Elle a été enlevée car elle semblait mauvaise. Il vaut mieux la reconstituer. Les individus panel sont les individus répondants de la V1 (qui ont a priori un nident commençant par '4'), ainsi que les nouveau-nés des mères panels (ce qui est plus dur à caractériser, mais pas toujours très utile car les nouveaux nés ont rarement des revenus ou des calendriers d'activités très fournis).

### ***Le code TNOC (mouvement de la personne) n'existe pas pour 676 individus en vague 2.***

=> Cela concerne surtout des non-répondants (ou des individus partis du ménage). Il n'y a que 13 individus qui ont répondu pour lesquels on n'a pas de TNOC. Cela fait partie des erreurs 'normales' dans tout panel, que l'on ne corrigera sans doute pas.

### ***Est-ce normal d'avoir TPRES='2' (présence à l'enquête précédente='non') et TNOC='10' (personne présente à l'enquête et à l'enquête précédente) ?***

=> A priori non. Là encore, il y a toujours un taux d'erreur (une vingtaine de personnes seraient concernées). Ces erreurs pourraient provenir d'individus entrés en vague téléphonique.

N.B. La variable TPRES figure dans le tableau permanent de composition du ménage, et vaut '1' si l'individu était présent à l'enquête précédente, et '2' sinon.

### ***La variable PVOIT (possession d'une voiture de société) a beaucoup de valeurs manquantes en vague 2.***

=> En effet, tous les individus dont les dossiers ont été traités en Pays de Loire ou dans le Nord-Pas de Calais ont des valeurs manquantes. Pour ces deux régions, l'information de PVOIT a en fait été stockée dans la variable ménage VOIT (dont le contenu est perdu, Cf. plus haut question sur VOIT). Le contenu de PVOIT pourrait alors être reconstitué pour ces deux régions pour les individus seuls dans leur ménage (puisque VOIT est unique pour l'ensemble du ménage).

### ***La variable afinpa (année d'arrêt d'activité pour ceux qui ne travaillent plus) du fichier indiviv2 pose problème.***

=> En effet, comme vous le signalez, sur 719 observations renseignées, 221 ont afinpa=19, soit le tiers qui aurait arrêté de travailler en 1919 ! Et cette variable est codée avec 4 chiffres sur les 2 autres vagues (19XX). Le problème existe apparemment aussi sur les versions antérieures des fichiers (fichiers plats basev2 et totv2), je ne crois pas que l'on puisse faire quelque chose (sauf peut-être récupérer l'information dans une autre vague ?).

### ***Les variables AMATRIE, ACOUPLE et ANACTU valent souvent '1900' en V1***

=> Ces questions, posées dans la partie 'départ du foyer parental' du questionnaire individuel portent respectivement sur l'année du dernier changement d'état matrimonial légal, sur l'année de la première mise en couple, et sur l'année de début de la situation actuelle. AMATRIE a 24 valeurs à 1900, et ANACTU 43. On peut évidemment d'interroger sur la validité de ces informations. ACOUPLE a 2754 valeurs à '1900', correspondant (d'après le questionnaire) aux individus qui n'ont jamais vécu en couple. Le codage, malheureux, de cette dernière variable pourrait être revu.

***Le codage de la variable PRESFR n'est pas homogène d'une vague à l'autre.***

=> La variable PRESFR (présence d'une fiche revenu) est en effet codée en numérique sur trois positions en V1, puis en caractère sur une position lors des deux vagues suivantes (il en est de même de la variable PRESQM dans la table ménage). Ce point devra être revu.

***La variable LIEN est manquante en V2 pour 2 500 cas.***

=> Si l'on s'en tient aux adultes (plus de 17 ans), tous les individus qui ont rempli un questionnaire individuel (qui peuvent être repérés par la variable INTERRO in ('R','P') en V2 et V3) ont un lien rempli (à une exception près). Les valeurs manquantes pour des individus répondants ne peuvent donc concerner que des enfants. Après vérification, seule une trentaine d'enfants appartenant à un ménage répondant n'auraient pas de lien en V2, et une soixantaine en V3. Ces enfants sont en fait des enfants qui ont quitté le ménage (TNOC>'30'), qui n'ont donc pas de lien dans le ménage qu'ils quittent, mais en ont un dans le ménage dans lequel ils arrivent.

***Quelle est la codification de la variable PCS (profession principale exercée) ?  
Et des variables PROPER et PROMER (dernières professions exercées par le père et par la mère) ?***

(19/05/2000)

=> En ce qui concerne la variable 'PCS' :

La nomenclature utilisée est la PCS en 455 postes (la plus complète à 4 chiffres) à une exception près : les agriculteurs sont codifiés 1000 au lieu des codifications prévues 11.. à 13.. (pour pouvoir utiliser cette codification, il aurait fallu avoir du détail sur la surface de l'exploitation, le type de culture ou d'élevage). Le livre de référence est la 'Nomenclature des professions et catégories socioprofessionnelles 1994 PCS'. On y retrouve en page IX (Cf. tableau ci-dessous), la nomenclature en 24 postes, le plus gros du document étant consacré aux 455 postes. Les deux premiers chiffres de la variable PCS dans le panel correspondent aux deux chiffres du niveau détaillé en 42 postes (dernière colonne du tableau ci-dessous). Ils varient dans le panel entre 10 et 69.

**NOMENCLATURE DES CATEGORIES SOCIOPROFESSIONNELLES  
CORRESPONDANCE ENTRE LES NIVEAUX 8, 24 ET 42**

Niveau agrégé 8 postes dont 6 pour les actifs occupés	Niveau de publication courante (24 postes dont 19 pour les actifs)	Niveau détaillé (42 postes dont 32 pour les actifs)
1 Agriculteurs exploitants	10 Agriculteurs exploitants	11 Agriculteurs sur petite exploitation 12 Agriculteurs sur moyenne exploitation 13 Agriculteurs sur grande exploitation
2 Artisans, commerçants et chefs d'entreprise	21 Artisans	21 Artisans
	22 Commerçants et assimilés	22 Commerçants et assimilés
	23 Chefs d'entreprise de 10 salariés ou plus	23 Chefs d'entreprise de 10 salariés ou plus
3 Cadres et professions intellectuelles supérieures	31 Professions libérales	31 Professions libérales
	32 Cadres de la fonction publique, professions intellectuelles et artistiques	33 Cadres de la fonction publique 34 Professeurs, professions scientifiques 35 Professions de l'information, des arts et des spectacles
	36 Cadres d'entreprises	37 Cadres administratifs et commerciaux d'entreprises 38 Ingénieurs et cadres techniques d'entre- prises
4 Professions Intermédiaires	41 Professions intermédiaires de l'enseigne- ment, de la santé, de la fonction publique et assimilés	42 Instituteurs et assimilés 43 Professions intermédiaires de la santé et du travail social 44 Clergé, religieux 45 Professions intermédiaires administratives de la fonction publique
	46 Professions intermédiaires administratives et commerciales des entreprises	46 Professions intermédiaires administratives et commerciales des entreprises
	47 Techniciens	47 Techniciens
	48 Contremaîtres, agents de maîtrise	48 Contremaîtres, agents de maîtrise
5 Employés	51 Employés de la fonction publique	52 Employés civils et agents de service de la fonction publique 53 Policiers et militaires
	54 Employés administratifs d'entreprises	54 Employés administratifs d'entreprises
	55 Employés de commerce	55 Employés de commerce
	56 Personnels des services directs aux particuliers	56 Personnels des services directs aux particuliers
6 Ouvriers	61 Ouvriers qualifiés	62 Ouvriers qualifiés de type industriel 63 Ouvriers qualifiés de type artisanal 64 Chauffeurs 65 Ouvriers qualifiés de la manutention, du magasinage et du transport
	66 Ouvriers non qualifiés	67 Ouvriers non qualifiés de type industriel 68 Ouvriers non qualifiés de type artisanal
	69 Ouvriers agricoles	69 Ouvriers agricoles
7 Retraités	71 Anciens agriculteurs exploitants	71 Anciens agriculteurs exploitants
	72 Anciens artisans, commerçants, chefs d'entreprise	72 Anciens artisans, commerçants, chefs d'entreprise
	73 Anciens cadres et professions interméd- iaires	74 Anciens cadres 75 Anciennes professions intermédiaires
	76 Anciens employés et ouvriers	77 Anciens employés 78 Anciens ouvriers
8 Autres personnes sans activité professionnelle	81 Chômeurs n'ayant jamais travaillé	81 Chômeurs n'ayant jamais travaillé
	82 Inactifs divers (autres que retraités)	83 Militaires du contingent 84 Elèves, étudiants 85 Personnes diverses sans activité professionnelle de moins de 60 ans (sauf retraités) 86 Personnes diverses sans activité profes- sionnelle de 60 ans et plus (sauf retraités)

=> En ce qui concerne les variables 'Proper' et 'Promer' :

La nomenclature utilisée ne peut être trouvée intégralement parmi les codes habituels de l'INSEE.

Le point de départ est la PCS en 24 postes (à 2 chiffres, niveau de publication courante, deuxième colonne du tableau précédent) auquel des modifications ont été apportées :

\* rajout de codes pour les professions non clairement déterminées (intitulé pas assez précis pour permettre une codification correcte) :

20. Chefs d'entreprises sans indication sur le nombre de salariés

30. Cadres sans autre indication

40. Professions intermédiaires sans autre indication

50. Employés sans autre indication

60. Ouvriers sans autre indication.

\* regroupement des postes 81 et 82 en un seul poste : 80. Inactifs.

\* traitement différent des retraités : la nomenclature en 24 postes prévoit de codifier les retraités en tant que tel (postes 7 ...). Dans l'enquête panel, ce qui intéresse c'est de connaître, au moins la dernière profession exercée des parents : les retraités sont donc codifiés selon leur CS avant leur départ à la retraite.

### ***Quel est le codage des variables OCHxD et PROFHxD concernant les enfants hors ménage ?***

=> La variable OCHD décrit l'occupation de l'individu en 7 modalités. Le code se trouve dans le cahier des cartes, et vaut :

1 : exerce une activité professionnelle

2 : ne travaille pas actuellement mais a un emploi

3 : chomeur

4 : étudiant

5 : retraité, retiré des affaires

6 : femme au foyer

7 : autre inactif (sans emploi)

La variable PROFHD correspond à la nomenclature de la catégorie socioprofessionnelle à 8 postes donnée ci-dessus.

### ***Les variables indiquant les changements d'entreprise (ADEBE et MDEBE) et d'établissement (TDETS) semblent parfois incohérentes.***

En particulier, il semblerait qu'il puisse y avoir des changements d'entreprises sans qu'il y ait changement d'établissement (ou que la variable correspondant au changement d'établissement, TDETS, ne soit pas renseignée). En outre, est-il possible que pour des individus du panel ayant changé d'entreprise entre les vagues 1 et 2, ou 2 et 3, la partie correspondante du questionnaire ne soit pas renseignée ?

(18/07/00)



=> Normalement un changement d'entreprise entraîne le remplissage de la partie correspondante du questionnaire sur la description du nouvel employeur et de la nouvelle activité.

La variable TDETS est une variable de gestion posée par l'intermédiaire de l'enquêteur mais dont l'objectif était de faciliter le chiffrage par les agents des Directions Régionales des informations sur l'activité économique de l'établissement employeur.

Il n'y a pas eu de consignes particulières sur ce point au niveau national. Mais il est possible que localement (DR par DR) l'information soit plus ou moins bien remplie.

Toutefois les incohérences rencontrées sur ces variables peuvent également s'expliquer par des situations complexes.

Il est effectivement possible de changer d'employeur sans changer d'établissement (du moins au sens où le ressentent les enquêtés). C'est le cas lorsque l'établissement où travaille l'enquêté est repris par un nouveau "propriétaire", l'établissement reste le même alors que l'entreprise dans laquelle l'individu travaille change. J'ai sur ce point le souvenir d'un cas qui nous avait été soumis :

Un chef d'entreprise en difficulté a vendu son entreprise mais est resté en place en continuant (presque) le même travail qu'auparavant mais en changeant de statut (toute la partie sur la comparaison entre l'activité actuelle et l'ancienne devient caduque). Ce type de cas pose problème notamment pour les variables ADEBE et MDEBE. En effet pour l'enquêté son entrée dans l'entreprise est plus ancienne que le changement d'employeur.

Un autre cas que nous avons également rencontré est celui qui concerne les doubles activités. Ainsi un individu qui est salarié de deux entreprises en même temps, puis qui perd l'emploi qu'il considérait comme étant le principal parmi les deux mais conserve l'autre emploi, va avoir une situation très complexe dans le fichier Panel. En général dans ce cas son activité principale change (pas la même profession en activité principale), la partie concernant son employeur est décrite, cependant les dates ADEBE et MDEBE ne correspondent pas à la date de perte de son ancienne activité principale (elles correspondent au début de l'activité qui était auparavant considérée comme secondaire), et il ne change pas d'établissement (il est toujours employé par le même employeur; c'est simplement son activité secondaire qui devient principale).

Ces deux types de situations ont été rencontrées à plusieurs reprises. Elles peuvent au moins expliquer en partie les difficultés rencontrées sur les variables étudiées.

### ***Quelles variables permettent de décrire le secteur d'activité de l'individu ?***

=> Un enquêté *actif occupé* doit donner les coordonnées de son employeur (adresse ou SIRET), ce qui permet aux directions régionales de l'Insee de chiffrer le code activité APE (activité principale de l'établissement), qui figure dans la base du CD-ROM (variable APET). C'est un code à 3 chiffres et une lettre. La signification de ce code en 700 positions occupe une trentaine de pages, et est publiée dans un document officiel (Législation et Règlementation, « Nomenclatures d'activités et de produits françaises », les éditions des Journaux officiels, environ 700 pages). Cette nomenclature ainsi que les deux nomenclatures regroupées ci-dessous sont inchangées depuis le début du panel.

Le code APE peut être regroupé en une nomenclature économique de synthèse en 16 postes (NES16), utilisée par exemple pour la publication des résultats de l'enquête Emploi. Voici le programme de passage de la variable APET à la NES 16 :

```

if "011A"<=apet<="050C" then nes="EA";
else if "151A"<=apet<="160Z" then nes="EB";
  else if ("181Z"<=apet<="193Z" ! "221A"<=apet<="223E" ! "244A"<=apet<="245C"
! "297A"<=apet<="297C" ! apet="323Z" ! "334A"<=apet<="335Z" ! "361A"<=apet<="366E")
then nes="EC";
else if "341Z"<=apet<="343Z" then nes="ED";
else if ("281A"<=apet<="283C" ! "291A"<=apet<="296B" ! "300A"<=apet<="311C"
! "322A"<=apet<="322B" ! "331A"<=apet<="333Z" ! "351A"<=apet<="355Z")
  then nes="EE";
else if ("131Z"<=apet<="132Z" ! "141A"<=apet<="145Z" ! "171A"<=apet<="177C"
! "201A"<=apet<="212L" ! "241A"<=apet<="243Z" ! "246A"<=apet<="275G" !
"284A"<=apet<="287P" ! "312A"<=apet<="321B" ! apet="371Z" ! apet="372Z")
then nes="EF";
else if ("101Z"<=apet<="120Z" ! "231Z"<=apet<="233Z" ! "401Z"<=apet<="410Z")
  then nes="EG";
else if ("451A"<=apet<="455Z" ) then nes="EH";
else if ("501Z"<=apet<="527H") then nes="EJ";
  else if ("601Z"<=apet<="634C" ) then nes="EK";
  else if ("651A"<=apet<="672Z") then nes="EL";
  else if ("701A"<=apet<="703E") then nes="EM";
  else if ("641A"<=apet<="642B" ! "711Z"<=apet<="748K" ! "900A"<=apet<="900C")
then nes="EN";
  else if ("551A"<=apet<="555D" ! "921A"<=apet<="950Z" )
then nes="EP";
    else if ("801Z"<=apet<="853K" ) then nes="EQ";
    else if ("751A"<=apet<="753C" ! "911A"<=apet<="990Z" ! apet="990Z") then nes="ER";
run;

```

et le format correspondant (p. 716 du document cité ci-dessus) :

```

proc format;
value $fnes
'EA'='EA. agriculture, pêche'
'EB'='EB. industries agricoles et alimentaires'
'EC'='EC. industrie des biens de consommation'
'ED'='ED. industrie automobile'
'EE'='EE. industrie des biens équipement'
'EF'='EF. industrie des biens intermédiaires'
'EG'='EG. énergie'
'EH'='EH. construction'
'EJ'='EJ. commerce'
'EK'='EK. transports'
'EL'='EL. activités financières'
'EM'='EM. activités immobilières'
'EN'='EN. services aux entreprises'
'EP'='EP. services aux particuliers'
'EQ'='EQ. éducation, santé, action sociale'
'ER'='ER. administration';
run;

```

Le code APE peut également être regroupé en une nomenclature de 17 sections utilisée au niveau européen, mais non au niveau français (en effet, cette nomenclature ne distingue notamment pas les entreprises qui produisent des biens intermédiaires des autres, alors que cette distinction est importante pour les comptes nationaux par exemple).

Le programme permettant de passer au code APE à la nomenclature en 17 sections est le suivant :

```

ape94=substr(apet,1,2);

if "01"<=ape94<="05" then ape=1 /*agriculture- peche*/;
else if "10"<=ape94<="41" then ape=2 /*ind extratives et manufac*/;
/*else if "15"<=ape94<="37" then ape=3 ind manufacturières;
else if "40"<=ape94<="41" then ape=4 prod et distri eau'*/;
else if ape94="45" then ape=5 /*construction*/;
else if "50"<=ape94<="52" then ape=6 /*commerce*/;
else if ape94="55" then ape=7 /*hotels et restaurants*/;
else if "60"<=ape94<="64" then ape=8 /*transports et communications*/;
else if "65"<=ape94<="67" then ape=9 /*activites financières*/;
else if "70"<=ape94<="74" then ape=10 /*immobilier services entrepri*/;
else if ape94="75" then ape=11 /*administration publique*/;
else if ape94="80" then ape=12 /*education*/;
else if ape94="85" then ape=13 /*sante action sociale*/;
else if "90"<=ape94<="93" then ape=14/* services collectifs*/;
else if ape94="95" then ape=15 /* services domestiques*/;

```

```
else if ape94="98" then ape=16 /*activites extra-territoriales*/;
```

et voici le format correspondant (cf. p30. de l'ouvrage cité ci-dessus) :

```
proc format;
value $fnace
'A'='A. agriculture'
'B'='B. peche'
'C'='C. industrie extractive'
'D'='D. industrie manufacturière'
'E'='E. production et distribution eau'
'F'='F. construction'
'G'='G. commerce'
'H'='H. hôtel restaurant'
'I'='I. transport et communication'
'J'='J. activites financières'
'K'='K. immobilier, service aux entreprises'
'L'='L. administration publique'
'M'='M. education'
'N'='N. santé et action sociale'
'O'='O. services collectifs sociaux et personnels'
'P'='P. services domestiques'
'Q'='Q. activites extra territoriales';
run;
```

Le code SIRET (qui ne figure pas dans les fichiers du CD-ROM de diffusion pour des questions d'anonymat) et qui permettrait d'avoir accès à des informations supplémentaires est relativement peu renseigné (une fois sur deux ?), et semble donc peu exploitable.

---

## **Présence et qualité des variables - Table activité**

**Fichier activité : exiact est souvent à blanc et sur certaines lignes avec exiact='2', aucun mois d'activité n'est coché**

=> J'ai regardé un peu les lignes avec exiact='2' et aucun mois derrière. Il n'y en a pas beaucoup (de l'ordre d'une centaine). Pour celles-là, on peut essayer de récupérer le calendrier de revenu (suivant une suggestion d'un des membres du groupe d'exploitation).

De même, il y a très peu de variables exiact à blanc, sauf pour les CODACT 11, 13 et 17 pour lesquels EXIACT n'a pas de sens (et pour ces CODACT là, seules les variables de nombre de jours sont remplies, ce qui peut donner l'impression que les calendriers d'activité sont vides). Pour mémoire, il y a environ une vingtaine d'exiact à blanc en v1 et une vingtaine en v3, et peut-être 80 en V2 (si on ne compte pas les codact 11,13 et 17).

**Environ la moitié des variables du fichier activité de la vague 2 est en format numérique alors que toutes les variables des vagues 1 et 3 sont en format caractère.**

=> Cela devra être modifié dans la prochaine version du CD-ROM.

**NBACT : Nombre de lignes renseignées n'existe pas dans le fichier, est ce volontairement?**

=> Oui, en fait cette variable devait permettre aux DR de vérifier la qualité de la saisie (comparaison entre NBACT et le nombre de lignes saisies). Mais le contenu de la variable n'avait pas l'air très bon, et de plus, on a eu des pb avec l'équipe informatique (qui mettait la

variable alternativement en table individu, activité...). On a donc décidé de supprimer la variable.

## **Présence et qualité des variables - Table revenu**

**En 1ère vague, on ne devrait pas connaître les mois de perception des revenus de 94. Dans le fichier les mois ont été remplis quand le nb total est 12.**

=> Oui, il faudra que l'on corrige ce point.

**Certains individus déclarant 12 mois de perception de revenu en 95 ou en 1996 ont des mois à vide.**

=> Oui, seulement pour des types de revenus pour lesquels on ne demande pas de calendrier (Cf les types de revenus concernés pour la V2 ci-dessous).

CODFR			Cumulative	Cumulative
	Frequency	Percent	Frequency	Percent
12	1	0.4	1	0.4
13	7	3.0	8	3.4
14	6	2.6	14	6.0
15	40	17.1	54	23.1
16	78	33.3	132	56.4
17	5	2.1	137	58.5
18	19	8.1	156	66.7
19	1	0.4	157	67.1
42	56	23.9	213	91.0
43	13	5.6	226	96.6
49	5	2.1	231	98.7
50	3	1.3	234	100.0

**Il semble que les montants de revenus déclarés inclus dans un autre revenu sont pris en compte dans le revenu individuel total calculé dans les tables de revenus simulés ?**

=> A priori, l'existence pour une même ligne de revenu d'un code inclusion et d'un montant est anormale (la documentation aux enquêteurs de la première vague indique que les codes inclusions doivent être utilisés lorsqu'il n'est pas possible de séparer des montants de revenus différents). On peut donc a priori considérer soit que le code inclusion est mauvais (et donc qu'il faut tenir compte des montants déclarés pour obtenir le revenu total), soit que le montant indiqué n'aurait pas dû l'être. Pour les CD-ROM diffusés, nous avons fait le premier choix, mais bien sûr, cela reste discutable. Suite à votre question, nous avons toutefois vérifié que les implications de ce choix restent limitées. En effet, en vague 2, seuls 93 individus ont au moins une catégorie de revenu salarial ayant à la fois un montant et un code inclusion (sur environ 6000 à 7000 individus ayant des revenus salariaux). Pour ces individus, tenir compte ou non de ces montants entraîne une différence en général inférieure à 10000 F, qui représente une part modérée du revenu total (moins de 10 % dans une très grosse majorité des cas). En vague 3, il n'y a que 31 individus pour lesquels le problème se pose.

Par ailleurs, sur quelques individus choisis au hasard, il semble délicat de choisir entre inclure ou non les montants déclarés à l'aide d'un code inclusion à la seule vue des tables de revenu des années précédentes ou suivantes. Suite à votre question, nous avons calculé pour ceux des 93 individus de la V2 pour qui se posaient le problème et qui étaient de plus présents les trois vagues (soit 78 individus) un indicateur permettant de savoir si l'évolution des revenus étaient plus régulière si l'on incluait ou si l'on n'incluait pas les revenus avec codes inclusion dans le revenu total (pour cela nous avons simplement calculé pour chaque individu  $(\text{revenusalarialV1}-\text{revenusalarialV2})^2 + (\text{revenusalarialV2}-\text{revenusalarialV3})^2$ ). En moyenne sur l'ensemble des individus, on arrive à des écarts légèrement plus élevés en calculant les revenus salariaux totaux comme nous l'avons fait (i.e. en tenant compte des montants précédés d'un code inclusion) (l'écart quadratique moyen entre deux années est de 26900 F dans le premier cas, et de 24000 F dans l'autre). Ce test reste donc peu conclusif (et de plus, repose implicitement sur l'hypothèse que les revenus salariaux des individus concernés sont constants dans le temps, et il serait trop coûteux d'aller vérifier si tel est bien le cas en fonction des calendriers d'activité...). In fine, aucune des deux solutions n'apparaît indiscutablement supérieure à l'autre. Le choix entre retenir ou non les montants avec un code inclusion pour le calcul des revenus totaux reste donc largement arbitraire. La question méritait en tout cas d'être posée (voire rediscutée : dans le prochain CD-ROM, il est vraisemblable que l'on revienne sur le premier choix). De manière générale, compte tenu des autres anomalies relevées dans la table revenus, il pourrait être préférable de privilégier un traitement plus global de la validité des revenus (examens des évolutions aberrantes, utilisations de modèles appropriés...) que de passer trop de temps sur des individus minoritaires.

### ***Quel est le principe de l'imputation des revenus d'activité ?***

=> Les procédures d'imputation sont décrites dans les documents de travail diffusés sur le CD-ROM. En voici le principe :

L'imputation des revenus d'activité se fait en deux étapes :

- il y a repérage de la "probabilité" de perception en utilisant certaines variables de la vague : les revenus et les calendriers d'activité déclarés, des variables indicatrices d'existence de revenus, en utilisant aussi des variables de la vague précédente : l'existence du type de revenu concerné (montant et indicatrice), calendrier d'activité (cf. algorithme de repérage de la probabilité de perception à la question ci-dessous pour les non-salariés).

A partir de ces variables, il est établi une typologie de la probabilité d'existence du type de revenu dont les différentes modalités peuvent être regroupés en 3 catégories :

revenus déclarés,

revenus non déclarés mais dont l'existence est probable, ils sont donc à imputer,

revenus non déclarés et considérés comme ne pouvant pas exister.

- les montants sont ensuite imputés par régression stochastique pour la deuxième catégorie. Cette régression inclut parmi les variables explicatives le montant annuel des revenus de l'année précédente.

### ***Des revenus salariaux sont imputés à des non-salariés***

=> Voici le programme qui nous a servi à déterminer quels étaient les individus à qui l'on devait imputer un revenu (individus avec  $\text{VMSAL} > 0$ ). Si vous trouvez des individus non

salariés avec un salaire imputé, c'est sans doute que vous avez une règle différente de la nôtre pour déterminer qui a pu percevoir un salaire.

\*\*\* type de valeurs manquantes pour les revenus d activité salariée : variable VMSAL

0 : valeur déclarée

pas de valeur déclarée alors qu activité salariée en 95 et pas de revenus  
d activité secondaire ni de revenus d indépendants, et ...

1 : revenu salarié déclaré en Vague 2

2 : existence de revenu déclarée en F12(V2) et F12(V3)

3 : existence de revenu déclarée en F12(V2) seul ou F12(V3) seul

4 : aucun signe de revenu salarié ni en V2 ni en V3, à part l activité

5: pas de valeur déclarée, pas d activité salariée, pas de revenus d activité  
secondaire ni de revenus d indépendants et salaires déclarés en V2

\*\*\*\*\*;

```
if tot1>1 then vmsal=0; * tot1 est le total des revenus salariaux de l'individu ;  
else if tot1<=1 & dcdi95+dcdd95>0 & r20<=1 & r21<=0.5 then do; /* dcdi95 et dcdd95  
testent l'existence dde CDI ou de CDD dans le calendrier d'activité en 1995 ; r20 et r21 sont  
respectivement les revenus d'activité indépendante et secondaire */
```

```
if sum(of m11-m19)>0 then vmsal=1;
```

```
else if f121=1 & fi1=1 then vmsal=2;
```

```
else if f121=1 ! fi1=1 then vmsal=3;
```

```
else vmsal=4;
```

```
end;
```

```
else if tot1<=1 & dcdi95+dcdd95<=0 & r20<=1 & r21<=1 & sum(of m11-m19)>0  
then vmsal=5;
```

### ***Les revenus partiels sont-ils pris en compte pour les imputations de revenus ?***

=> Pour l'imputation des revenus concernant l'année complète dans une vague donnée, il est tenu compte dans l'équation économétrique de montant des revenus annuels de l'année précédente, mais pas des revenus partiels pour l'année en cours (aussi déclarés à la vague précédente). Mais effectivement, pour les individus qui déclarent des revenus partiels, on aurait pu préférer calculer les revenus annuels à partir de ces revenus partiels. Cela aurait sans doute nécessité de faire des hypothèses sur le nombre de mois travaillés, le revenu mensuel moyen, voire sur le temps de travail hebdomadaire. Toutefois, une stratégie mixte aurait aussi pu être retenue, comme par exemple conserver le maximum du revenu imputé selon la méthode actuelle et du revenu partiel (ou du revenu partiel équivalent annuel). Ce point peut être rediscuté.

***Certains revenus déclarés repris dans les tables de revenus simulés REVACVxS ne figurent pas dans la table REVENUV2***

=> Vous notez avec raison que le fichier revacv3s indique pour le nident='4005342' un revenu salarial partiel déclaré V2 de 70533 en 95, et que ce montant n'apparaît pas dans le fichier revenuv2.

Effectivement, il y a eu un problème dans les traitements informatiques effectués sur la deuxième vague. Les revenus (partiels ou pour l'année complète) pour lesquels le code d'existence de revenu était manquant dans les premières versions de fichiers plats ont visiblement été supprimés dans des versions ultérieures. Cela concerne 34 revenus de la ligne 11, 23 revenus de la ligne 20 et 61 de la ligne 21 pour ne retenir que les lignes les plus importantes, soit environ 200 observations au total. Nous essaierons de réparer cette erreur dans la prochaine version du CD-ROM (ce qui ne va pas complètement de soi car certains montants qui ont disparu ont été supprimés avec raison, par exemple si l'individu avait déclaré un salaire à la place d'un revenu d'indépendant).