



Direction des Statistiques démographiques et sociales
Département des Prix à la consommation, des ressources et des conditions
de vie des ménages
Division "Conditions de vie des ménages"

Dossier suivi par :
Françoise DUMONTIER
Tél. : 01 41 17 55 04
Fax : 01 41 17 63 17
Messagerie : francoise.dumontier@insee.fr

PARIS, le 3 novembre 2004
N° 179/F340

Objet : Note au sujet du redressement des revenus dans l'enquête santé 2002-2003 et du calcul d'un revenu par unité de consommation

Sommaire

- A** Le relevé des revenus dans l'enquête santé
- B** Rappel sur les erreurs de collecte déjà corrigées et étude des revenus déclarés nuls
- C** Méthode choisie pour le redressement du revenu
- D** Correction des erreurs flagrantes sur REV2 (revenu en clair)
- E** Simulation d'un revenu pour REV3 (revenu en tranches) dans les tranches déclarées et pour les revenus non déclarés
- F** Bilan des corrections et calcul d'un revenu pour l'ensemble des ménages REVTOT
- G** Indicateurs de revenu ; revenu par unité de consommation
- H** Pondération ménage, décile de revenu par UC

ANNEXES

- 1 Résultats de la régression de log(Rev2)
- 2 Fréquences de la variable Testrev
- 3 Résultats de la régression qui linéarise Rev3-
- 4 Comparaison avec d'autres sources

A) LE RELEVÉ DES REVENUS DANS L'ENQUÊTE SANTÉ

On commence par recenser les sources éventuelles de revenus du ménage

Au cours des douze derniers mois, votre ménage a-t-il perçu des ressources provenant de :

REVS Salaires, traitements et primes y compris :

- 13^e mois, congés payés, heures supplémentaires, indemnités journalières
- rémunération des emplois temporaires, des activités secondaires, salaires des dirigeants salariés de leur entreprise
- intéressements et participations

1 oui

2 non

REVI Revenus d'une activité professionnelle indépendante

1 oui

2 non

RMI RMI

1 oui

2 non

ALLCHO Allocation de chômage

1 oui

2 non

RET Préretraites, retraites, pensions et rentes diverses

minimum vieillesse, aide aux personnes âgées, pensions d'invalidité, d'ancien combattant, pension alimentaire, bourse d'études

1 oui

2 non

DIVI Intérêts, revenus d'épargne, dividendes

1 oui

2 non

LOY Loyers et fermages

1 oui

2 non

ALLOC Prestations familiales, allocations de logement

(allocations familiales, complément familial, aide à la garde d'enfants, allocation de rentrée scolaire...)

1 oui

2 non

Après cette détection de sources de revenus éventuelles, on demande un montant global de revenu

REV2 En tenant compte de tous les différents types de revenu, indiquez quel est le **MONTANT TOTAL DES RESSOURCES perçues par tous les membres de votre ménage, sur les 12 derniers mois (EN EUROS)** (on tiendra compte uniquement du revenu NET (de cotisations sociales et de C.S.G.))

.....

Si Refus ou NSP à la question précédente (rev2=99999998 ou 9999999), on propose des tranches

REV3 Pouvez-vous nous indiquer ce montant à partir des tranches ci-dessous ?
(on tiendra compte uniquement du revenu net (de cotisations sociales et de C.S.G.)) (Tendre la carte-code n° 8)

01. pas de revenus (ou refus -consigne à l'enquêteur)
02. moins de 3 000 euros (19679F) par an
03. de 3 000 euros (19679F) à - de 4 600 (30174F) par an
04. de 4 600 euros (30174F) à - de 7 000 (45917F) par an
05. de 7 000 euros (45917F) à - de 9 000 (59036F) par an
06. de 9 000 euros (59036F) à - de 12 000 (78715F) par an
07. de 12 000 euros (78715F) à - de 15 000 (98394F) par an
08. de 15 000 euros (98394F) à - de 18 000 (118072F) par an
09. de 18 000 euros (118072F) à - de 23 000 (150870F) par an
10. de 23 000 euros (150870F) à - de 27 000 (177108F) par an
11. de 27 000 euros (177108F) à - de 37 000 (242704F) par an
12. de 37 000 euros (242704F) à - de 46 000 (301740F) par an
13. de 46 000 euros (301740F) à - de 68 000 (446051F) par an
14. 68 000 euros (446051F) et plus par an

CONSIGNE AUX ENQUETEURS

Si un ménage refuse ou ne sait pas donner une tranche, on affectera à REV3 la valeur 01
Ceci sera corrigé ultérieurement

B) RAPPEL DES ERREURS DE COLLECTE DEJA CORRIGÉES ET ETUDE DES REVENUS DECLARES NULS (43 ménages)

Rappel des corrections déjà effectuées

On a déjà corrigé, dans la table ménage, les 836 ménages qui avaient déclaré des tranches au lieu de revenus en clair(on leur a mis REV3=REV2 puis REV2=9999999) et 390 ménages qui avaient mis REV3='01', c'est à dire aucun revenu mais qui par ailleurs avaient déclaré des sources de revenus (dans les variables REVS à ALLOC). Pour ces ménages la valeur 01 de REV3 signifiait en fait un refus (cf. consigne aux enquêteurs du paragraphe précédent). On a donc mis REV3=99

NB 6 ménages ont abandonné l'enquête avant les questions sur les revenus. Aucun individu du ménage n'a répondu aux questions santé de visite 1 et ils auront de ce fait dans l'enquête des pondérations nulles. Le revenu de ces ménages est mis à valeur manquante (99)

Revenu en clair Nul :REV2=0 (36 cas)

Parmi ces ménages, on ne garde REV2=0 que pour les jeunes étudiants ou les jeunes chômeurs n'ayant jamais travaillé et ne déclarant aucune source de revenus (variables REVS à ALLOC toutes égales à 2). On suppose que leurs parents les aident.

Pour les autres on met Rev2 et Rev3 en valeur manquante (99) qu'il y ait ou non des sources de revenu déclarées.

Lorsqu'il n'y a pas de sources de revenu, on essaie de reconstruire ces sources avec les variables Occupa et Cs des individus du ménage

Pour deux ménages on n'a cependant pas su reconstruire de sources de revenu

idtmn= 11102117801 jeune femme de 30 ans mère d'un jeune enfant(0 ans) ayant été quelque temps domestique-(Cspr =56) chez des particuliers et qui ne dit avoir aucune source de revenus

idtmn=91100777202 couple d'étrangers inactifs, (occupa=8) de plus de 50 ans, diplômés, ayant travaillé, sans source de revenus déclarée

REV2 non déclaré et REV3=01 (7 ménages) et pas de sources de revenu déclarées

(le cas où Rev3 =01 avec des sources de revenu a déjà été traité en fin de collecte)

Même correction que précédemment:

S'il s'agit de jeunes étudiants ou chômeurs n'ayant jamais travaillé, on met rev2=0 et rev3=blanc (3 cas)

Sinon on met rev2=9999999 et rev3=99=NSP et on essaie de reconstruire des sources avec les variables Occupa et Cs des individus du ménage.(2 cas) quand on le peut

Pour deux ménages cependant, on n'a pu corriger les sources de revenus

idtmn=11100148004 couple étranger lui inactif(occupa=8) ancien maçon et elle femme au foyer(occupa=7) et

idtmn=42100222703 ancien militaire irakien, inactif, venu en France récemment,

A l'issue de cette étape

En sortie, on n'a plus la modalité REV3=01 (pas de revenus). Il y a soit des revenus déclarés en clair (10719), soit en tranche (5710), soit des revenus non déclarés (419)

Répartition des 16 848 ménages

Nombre ménages	Rev2 déclaré en clair	Sinon, Rev3 déclaré en tranches
10 719	De 0 à 3 000 000	blanc
5 710	9999999 NSP ou refus	Tranches de 02 à 14
413	9999999 "	99
6 (abandon)	9999999 "	99

C) METHODE CHOISIE DE REDRESSEMENT DU REVENU

Pour plus de détails, voir l'article de Nicolas CHOPIN et Emmanuel MASSE intitulé " Imputation de l'enquête Budget de famille 2000", présenté aux JMS de 2002

On appelle méthode d'imputation économétrique une méthode reposant sur une modélisation paramétrique de type régression linéaire :

$$Y = X\beta + \sigma U \quad (1)$$

dans le cas continu, voire de type régression linéaire généralisée lorsque la variable s'y prête (régression logistique si Y est dichotomique). Dans l'équation précédente, $X\beta$ représente le produit scalaire entre les vecteurs X et β , U est une variable gaussienne centrée réduite, et σ est l'écart-type des résidus.

L'avantage d'une telle approche est que l'on modélise, de manière simple, les corrélations entre les variables X et la variable Y. La valeur imputée respectera de cette manière une certaine cohérence avec les valeurs prises par variables auxiliaires X, pour une unité donnée. Dans la pratique, l'imputation se fait en deux étapes :

1. Estimation du modèle : on estime les paramètres β et σ selon une méthode idoine (moindres carrés ordinaires par exemple). Soient $\hat{\beta}$ et $\hat{\sigma}$ les estimateurs obtenus.
2. Imputation : on remplace la valeur manquante Y_i pour le ménage i,
 - soit par la moyenne conditionnelle estimée $X_i\hat{\beta}$. (imputation par la moyenne conditionnelle) ;
 - soit par la valeur simulée $X_i\hat{\beta} + \hat{\sigma}U_i$, où U_i est un tirage aléatoire dans une loi normale centrée réduite (méthodes des résidus simulés).

Application dans l'enquête santé

La première méthode d'imputation servira, à détecter les erreurs les plus flagrantes de déclaration du revenu en clair des ménages, en comparant ce revenu REV2 à un revenu moyen estimé REVESTIM

La seconde méthode dite des "résidus simulés" qui est en général préférable à la première, car elle permet de mieux restituer la variabilité, servira à imputer des revenus aux ménages REVSIM

La variable effectivement modélisée dans les deux cas est le logarithme du revenu. On supposera donc une relation linéaire du type (1) entre Y, le logarithme du revenu considéré, et X les différentes variables disponibles dans l'enquête

Les ménages ayant refusé de donner le montant exact d'un revenu donné se sont vus proposés de préciser dans quelle tranche de revenus se situait ce montant (en égard à un système de tranches pré-établi). Ceci complique légèrement l'estimation des paramètres de la régression (1)

Le programme SAS utilisé, permet l'imputation par résidus simulés, en présence de données répondues en tranche. L'estimation des paramètres est obtenue par la procédure LIFEREG. Pour tous les individus ayant indiqué une tranche de revenu, la valeur imputée doit être simulée à l'intérieur de cette tranche. La solution retenue est de simuler en boucle des valeurs $X_i\hat{\beta} + \hat{\sigma}U_{i,j}$, où les $U_{i,j}$ sont des tirages gaussiens centrés réduits, jusqu'à obtenir une valeur comprise dans la tranche observée.

Si au bout de n simulations (n=1000) aucune valeur compatible avec la tranche répondue n'est obtenue, on impute en général le milieu de la tranche
Dans l'enquête santé, ce n'est pas le milieu de tranche qui a été retenu car il semblait évident que dans ces cas de non convergence, les tranches déclarées par les ménages étaient nettement trop

petites au vu des caractéristiques de ces ménages et correspondaient à des revenus mensuels et non annuels On a donc corrigé les tranches et repassé la procédure

C) CORRECTION DES ERREURS FLAGRANTES DE REV2, REVENU EN CLAIR

On fait une régression logarithmique sur les revenus en clair REV2 tels que $3\,000 < \text{REV2} < 300\,000$, les bornes retenues semblant être des revenus vraisemblables (cf. Annexe 1)

On exclut aussi de cette régression les jeunes PR(personnes de référence) étudiants âgés de moins de 26 ans et les autres jeunes de moins de 26 ans qui n'ont soit jamais travaillé, soit travaillé moins d'un an (397 cas), et qui sont caractérisés par la modalité 7 de la variable COREV2. (cf. plus loin)

Le revenu $R = \text{REV2}$ est analysé par un modèle linéaire en logarithme

$$\text{Log}R = \sum \alpha_i X_i + \varepsilon$$

avec ε suivant la loi normale $(0, \sigma^2)$

Donc l'espérance de revenu vaut :

$$E(R) = e^{\sum \alpha_i X_i} \cdot E e^\varepsilon$$

Calcul de l'espérance d'une loi lognormale

Le second terme est calculé comme l'espérance d'une loi lognormale soit de la façon suivante :

$$X \sim N(0, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

$$E(e^X) = \int_{-\infty}^{+\infty} e^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}(x^2 - 2\sigma^2 x)} dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}} \cdot e^{\frac{\sigma^2}{2}} dy$$

en posant $\boxed{y = x - \sigma^2}$ ($dy = dx$)

$$= e^{\frac{\sigma^2}{2}} \cdot \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} y^2} dy}_{=1}$$

$$\boxed{= e^{\sigma^2/2}}$$

D'où la formule

$$E(R) = e^{\sigma^2/2} \cdot e^{\sum \alpha_i X_i}$$

utilisé pour les imputations.

A partir des coefficients obtenus de la régression logarithmique sur les revenus en clair REV2 tels que $3\ 000 < REV2 < 300\ 000$ (dont la part de variance expliquée donnée par le R-Square est de 0.49), on peut donc estimer un revenu moyen REVESTIM pour tous les ménages (sauf pour les 397 jeunes cités précédemment), soit 16451 ménages.

$$REVESTIM = EXP(\hat{\alpha}X + \sigma^2 / 2)$$

Sachant que σ est donnée dans la régression par la variable `_RMSE_` les instructions SAS correspondantes sont

```
lrevenu =  $\sum \alpha_i X_i$  (calculé pour chaque ménage à partir des coefficients de la régression)
revestim=exp(lrevenu+(_RMSE_**2)/2)
```

Détection des erreurs

On va se servir de ce revenu moyen estimé uniquement pour détecter des erreurs flagrantes de déclaration en le comparant au revenu déclaré REV2

Pour ce faire, on calcule le ratio TESTREV = REV2/REVESTIM qui va de 0 à 88 pour tous les revenus déclarés non nuls, qu'on arrondit à un chiffre après la virgule (cf. Annexe 2)

```
if 0<rev2<9999990 then testrev=rev2/revestim; (instructions sas)
testrev=round(testrev,0.1)
```

Proposition de corrections

Pour détecter les invraisemblances, nous nous sommes aussi référés aux minima sociaux. En 2002, le RMI pour un ménage d'une personne était de 5000 euros, et le minimum vieillesse pour un vieux travailleur de 6000 euros environ Le SMIC était autour de 11 000 euros par an

Une fois détectées ces invraisemblances, soit on les corrige en émettant une hypothèse plausible d'erreur (francs au lieu d'euros, revenu mensuel au lieu d'annuel..), soit on leur affecte une valeur manquante lorsqu'on ne sait pas les corriger et ils seront simulés dans l'étape suivante par la méthode des résidus simulés.

Corrections sur REV2

```
corev2=0;***initialisation
rev2b=rev2;*** rev2b=revenu initial stocké;

*** revenus qui semblent trop grands, même si on les suppose donnés en
francs au lieu d'euros on les met à NSP (64 cas);
if testrev>8 then do;rev2=9999999;rev3=99;corev2=1;end;

*** revenus encore grands mais qu'on peut supposer avoir été donnés en
francs au lieu d'euros et qu'on corrige en conséquence.(91 cas)
if 5=< testrev=<8 then do;rev2=rev2b/6.56;corev2=2;end;

**** revenus trop petits par rapport à revestim et inférieur au RMI ;on
suppose qu'ils ont été déclarés mensuellement en francs au lieu d'annuels
en euros(122 cas);
if 0.1<testrev<0.6 and rev2b<5000 then do;
rev2=rev2*12/6.56;corev2=3;end;

**** revenus 10 fois trop petite et qui sont supposés avoir été déclarés
mensuels au lieu d'annuel (227 cas);
if testrev=0.1 then do; corev2=4;rev2=rev2b*12;end;
```

```

***** ménages avec testrev=0 , c'est à dire un revenu vraiment trop petit;
***la correction dépend d'occuppr( occupation de la PR)(53 ménages)
if testrev=0 then do; corev2=5;
    if occuppr in ('5','6') then rev2= max(rev2b*12,5000); (22 cas);
    else if occuppr='1' then rev2=99;(16 cas)
    else rev2=rev2b*12; (15 cas)
end;

```

E) SIMULATION D' UN REVENU : REVSIM

Une fois les REV2 en clair corrigés de ces invraisemblances criantes et en ajoutant les REV3, déclarés en tranche, on fait une régression (par la procédure SAS LIFEREG) sur l'ensemble des ménages (16 451), avec les mêmes variables explicatives que la régression précédente, pour linéariser la variable REV3 dans les tranches par la méthode des résidus simulés (Cf Annexe 3) La valeur de l'écart type de la régression appelé SCALE est de 0,43595)

Les tranches pour les REV2 se limitent à la valeur unique de REV2 (borne inférieur= borne supérieure =REV2)

On a limité les tranches entre 2000 et 300 000 euros puisqu'il n'y a plus de tranches 01 (pas de revenus). La tranche 02 va de 2 000 à 3 000 et la tranche 14 supérieure de 68 000 à 300 000

On a pensé que l'erreur "francs à la place d' euros" était moins vraisemblable que précédemment, car la carte présentée aux ménages mentionnait explicitement les deux échelles. Par contre l'erreur de "revenu mensuel au lieu d'annuel " devrait avoir la même probabilité, mais les tranches étant assez larges on préfère laisser le modèle détecter ces erreurs qu'on corrigera après

En effet, en sortie de la procédure lifereg, on trouve constate que le modèle, après 1000 itérations ne peut simuler un revenu dans la tranche donnée par les ménages pour 93 d'entre eux. On regarde ces ménages, et il semble qu'ils ont raisonné en tranches de revenu mensuel et non annuel. En effet, en divisant par 12 la variable REVESTIM (revenu moyen estimé par les coefficients de la régression sur REV2) on trouve un montant proche des tranches données par les ménages

On corrige les tranches de ces 93 ménages en conséquence et on refait tourner la régression par la procédure Lifereg (cf. Annexe 3 (2))

Cette nouvelle régression a un écart type(variable SCALE) plus petit que la première (0.41217 au lieu de 0.43595), ce qui fait que 24 nouveaux ménages ne peuvent être linéarisés dans leur tranche pour les mêmes raisons que précédemment

On corrige de nouveau les tranches pour ces 24 ménages et on refait tourner Lifereg ce qui donne encore un écart type plus petit (0,40789), et entraîne que 4 nouveaux ménages ne peuvent être linéarisés dans leurs tranches (cf. Annexe 3 (3)

On corrige les tranches de ces 4 ménages, on repasse la régression, et 2 ménages ne peuvent être par cette nouvelle régression, dont l'écart type est de 0,40709 (cf. Annexe 3 (4))

Ces deux ménages sont des ménages d'indépendants (cf. annexe 3 (5))

L'un (IDTMEN=82101020903, CS=23) a déclaré la tranche supérieure (rev3=14) on lui met pour revenu simulé la borne inférieure 68.000 euros (idtmen=82101020903, cs=23)

L'autre, (IDTMEN=31110474402, CS=22) ne correspond manifestement pas à sa tranche trop petite (REV3=4 de 4 600 à moins de 7000 euros), On lui affecte REVSIM=REVESTIM= 31 520 euros(montant moyen estimé par la régression sur REV2.)

In fine, on obtient pour les 16 451 ménages concernés une variable REVSIM (REVSIM n'est pas calculé pour les 397 jeunes de moins de 26 ans ayant peu ou pas travaillé)

The UNIVARIATE Procedure
Variable: revsim

Moments

N	16451	Sum Weights	16451
Mean	26412.3246	Sum Observations	434509152
Std Deviation	19276.8673	Variance	371597613
Skewness	3.16148057	Kurtosis	22.4633777
Uncorrected SS	1.75892E13	Corrected SS	6.11278E12
Coeff Variation	72.9843646	Std Error Mean	150.293537

Basic Statistical Measures

Location		Variability	
Mean	26412.32	Std Deviation	19277
Median	21952.00	Variance	371597613
Mode	18000.00	Range	333199
		Interquartile Range	19213

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 175.7383	Pr > t	<.0001
Sign	M 8225.5	Pr >= M	<.0001
Signed Rank	S 67662963	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	335000.00
99%	100000.00
95%	60000.00
90%	47200.00
75% Q3	33213.00
50% Median	21952.00
25% Q1	14000.00
10%	9384.00
5%	7432.67
1%	5300.00
0% Min	1801.10

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
1801.10	4887	260000	1568
2046.85	14037	260000	8843
2144.94	4551	300000	2997
2198.47	14270	300000	4988
2236.90	6248	335000	2407

F) BILAN DES CORRECTIONS : CALCUL D'UN REVENU TOTAL REVTOT

REVSIM a été calculé pour l'ensemble des ménages sauf les 397 jeunes (corev2=7)

1) BILAN DES CORRECTIONS (variable COREV2)

corev2	Frequency	Percent	calcul de RECSIM
0	15772	93.61	REVSIM= REV2 ou si REV2=99 alors REVSIM simulé dans tranches
1	64	0.38	REV2 trop grand REVSIM simulé
2	91	0.54	REVSIM= REV2/6.56 (revenu déclaré annuel en francs)
3	122	0.72	REVSIM=REV2*12/6.56 (revenu déclaré mensuel en francs)
4	227	1.35	REVSIM= rEV2*12 (revenu déclaré mensuel en euros)
5	53	0.31	revenu trop petit si actif REVSIM simulé (16 cas) si retraité REVSIM=Max(REV2*12,5000) (22 cas) si autre REVSIM= REV2*12 (15 cas)
6	122	0.72	REVSIM simulé après avoir modifié les tranches de REV3 pour que REVSIM converge (sauf 2 ménages pour qui REVSIM est imputé)
7	397	2.36	pas de REVSIM (jeunes moins de 26 ans ayant peu ou pas travaillé)
total	16848	100.00	

On n'a ni estimé, ni simulé de revenu, pour les 397 ménages de jeunes ayant peu ou pas travaillé
En effet on n'a aucun renseignement dans l'enquête sur les probables transferts parents-enfants
On n'a pas non plus essayé de corriger les revenus déclarés ou les tranches extrêmes qui paraissent quelquefois douteuses

2) REVCOR (pour les 397 jeunes ménages)

On propose cependant de calculer un revenu REVCOR pour ces ménages de jeunes, égal au revenu qu'ils ont déclaré soit en clair, soit en tranches (en prenant le milieu de la tranche).
Pour les 7 d'entre eux qui n'ont rien déclaré (et pour qui REV3=99) on propose d'affecter un revenu qui sera la moyenne du revenu des 274 ménages qui ont déclaré leur revenu en clair non nul, soit 7876 euros

Rev2	Rev3	Nombre de ménages	REVCOR
0	blanc	18	=0
De 65 720000	blanc	274	=rev2
9999999	De 02 à 13	98	=milieu de tranches
9999999	99	7	=7 876
		397	

Moyenne des revcor sur les 397 jeunes ménages

N	Mean	Std Dev	Minimum	Maximum
397	8806.08	8713.45	0	72000.00

3) REVTOT REVENU POUR L'ENSEMBLE DES 16848 MENAGES

On peut calculer un revenu total REVTOT, sachant que pour les 397 jeunes il n'est pas toujours fiable
 = REVCOR si corev2=7
 = REVSIM si non

L'ajout des jeunes, à revenus plus faibles, fait que la moyenne et la médiane sont plus petites que celle de REVSIM

Variable: revtot

Moments

N	16848	Sum Weights	16848
Mean	25997.4577	Sum Observations	438005167
Std Deviation	19281.0298	Variance	371758112
Skewness	3.1244706	Kurtosis	22.1412822
Uncorrected SS	1.765E13	Corrected SS	6.26301E12
Coeff Variation	74.1650591	Std Error Mean	148.54432

Basic Statistical Measures

Location		Variability	
Mean	25997.46	Std Deviation	19281
Median	21500.00	Variance	371758112
Mode	18000.00	Range	335000
		Interquartile Range	19448

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 175.0148	Pr > t	<.0001
Sign	M 8415	Pr >= M	<.0001
Signed Rank	S 70816433	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	335000.0
99%	100000.0
95%	60000.0
90%	46904.1
75% Q3	32947.6
50% Median	21500.0
25% Q1	13500.0
10%	9000.0
5%	7000.0
1%	3800.0
0% Min	0.0

Extreme Observations

----Lowest----		-----Highest----	
Value	Obs	Value	Obs
0	14856	260000	1621
0	14832	260000	9034
0	14294	300000	3077
0	13577	300000	5111
0	13223	335000	2481

G) INDICATEUR DE REVENU : REVENU PAR UNITE DE CONSOMMATION

Pour les 16 848 ménages ; on peut calculer un revenu par unité de consommation tout en rappelant que pour les 397 jeunes il n'est pas très fiable (corev2=7)

L'unité de consommation UC est définie ainsi : $uc=1+(0.5*(adul-1))+(0.3*(nbper-adul))$;
Adul est le nombre de personnes de 18 ans et plus du ménage

revuc=revtot/uc sera le revenu par unité de consommation

Variable: revuc

Moments			
N	16848	Sum Weights	16848
Mean	16335.3149	Sum Observations	275217385
Std Deviation	11321.6444	Variance	128179632
Skewness	3.09736685	Kurtosis	19.4422831
Uncorrected SS	6.6552E12	Corrected SS	2.15944E12
Coeff Variation	69.3077819	Std Error Mean	87.2238663

Basic Statistical Measures			
Location		Variability	
Mean	16335.31	Std Deviation	11322
Median	13600.00	Variance	128179632
Mode	12000.00	Range	173333
		Interquartile Range	10657

Tests for Location: Mu0=0				
Test	-Statistic-	-----p Value-----		
Student's t	t 187.2803	Pr > t	<.0001	
Sign	M 8415	Pr >= M	<.0001	
Signed Rank	S 70816433	Pr >= S	<.0001	

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	173333.33
99%	60000.00
95%	36611.31
90%	28705.00
75% Q3	20000.00
50% Median	13600.00
25% Q1	9342.67
10%	6533.33
5%	5160.00
1%	3000.00
0% Min	0.00

Extreme Observations			
----Lowest----		-----Highest----	
Value	Obs	Value	Obs
0	14856	145652	2481
0	14832	150000	2784
0	14294	150000	5111
0	13577	173333	1621
0	13223	173333	9034

H) DECILE DE REVENU, PONDERATION DES MENAGES

Le revenu de l'enquête santé est un revenu global déclaré par les ménages et qui n'est pas, de ce fait aussi fiable qu'un revenu obtenu en sommant tous les revenus des individus du ménage comme le font certaines enquêtes spécifiques sur les revenus.

L'enquête PCV de l'INSEE "enquête sur les conditions de vie des ménages" a un questionnaire relativement semblable à celui de l'enquête santé. Elle pose les mêmes questions sur les sources de revenu mais ensuite demande des tranches de revenu sans passer par les revenus en clair

On trouve en Annexe 4 A les résultats non pondérés de cette enquête pour janvier 2003 qui sont assez voisins de ceux de l'enquête santé 2002-2003

Par contre l'enquête BDF 2000, dont le questionnaire sur les revenus est plus complet, donne des revenus nettement plus élevés, bien que cette enquête ait été réalisée deux ans plus tôt (Annexe 4 B)

Pour la comparer, avec santé, on a pondéré le REVTOT de santé par pondmen1 (Annexe 4 C)

On trouve un écart de 15 % environ entre les moyennes et les médianes des deux enquêtes, écart qui peut être majoré à 20%, pour tenir compte du fait que la collecte de BDF 2000 a eu lieu plus de 2 ans après celle de santé 2002-2003

Au lieu de prendre le RUC comme indicateur de niveau de vie, un indicateur plus robuste pourrait être le décile de RUC et dont le calcul est donné ci dessous (appelé ici DECRUC), qui a l'avantage de ne pas tenir compte du niveau du revenu mais de sa distribution

```
proc rank data=tot out=fin groups=10; var revucp;
ranks DECRUC;
data resul; set fin(keep=idtmn decruc);
```

Mais pour obtenir des déciles corrects, il faut d'abord pondérer le revenu par uc par une pondération ménage qui dépend de la pondération individu utilisée (revucp étant le revenu par uc pondéré)

Comme il y a dans l'enquête deux pondérations individuelles possibles, il y aura deux pondérations ménages possibles

Si on travaille au niveau de la première visite on utilisera la pondération PONDMEN1

PONDMEN1= (somme des pondérations PONDVIS1 des individus du ménage) / NBPERS

Et on calculera Revucp1=revuc*pondmen1)

NBPERS étant le nombre de personnes du ménage

Si on travaille au niveau de la troisième visite on utilisera la pondération PONDMEN3

PONDMEN3= (somme des pondérations PONDVIS3 des individus du ménage) / NBPERS

Et on calculera Revucp3=revuc*pondmen3

Sont données ci dessous les "proc univariate" des deux pondérations pondmen1 et pondmen3

REMARQUE

- On constate bien que la somme des observations pondérées sur ces deux proc univariate est bien le nombre de ménages de l'enquête emploi qui a servi à caler l'enquête. (24 737 732 ménages)

- Il y a 27 ménages pour qui pondvis1=0 car aucun de leurs individus n'a répondu aux questions santé de visite 1 et de même 2045 ménages pour qui pondvis3=0 car aucun de leurs individus n'a répondu aux questions santé de visite 3

Variable: POND MEN1

Moments

N	16848	Sum Weights	16848
Mean	1468.28895	Sum Observations	24737732.2
Std Deviation	659.026316	Variance	434315.685
Skewness	0.81356337	Kurtosis	1.86408416
Uncorrected SS	4.36391E10	Corrected SS	7316916340
Coeff Variation	44.8839663	Std Error Mean	5.07725037

Basic Statistical Measures

Location		Variability	
Mean	1468.289	Std Deviation	659.02632
Median	1369.479	Variance	434316
Mode	0.000	Range	5726
		Interquartile Range	1001

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 289.1898	Pr > t	<.0001
Sign	M 8410.5	Pr >= M	<.0001
Signed Rank	S 70740716	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	5725.651
99%	2933.981
95%	2521.718
90%	2335.536
75% Q3	1959.865
50% Median	1369.479
25% Q1	959.120
10%	698.087
5%	554.975
1%	400.621
0% Min	0.000

Extreme Observations

----Lowest----		-----Highest-----	
Value	Obs	Value	Obs
0	16825	5454.00	14726
0	14638	5470.81	15882
0	14572	5507.57	10773
0	13386	5635.53	14196
0	13368	5725.65	13442

Variable: POND MEN3

Moments

N	16848	Sum Weights	16848
Mean	1468.28895	Sum Observations	24737732.2
Std Deviation	899.069281	Variance	808325.571
Skewness	0.36042406	Kurtosis	0.64109331
Uncorrected SS	4.994E10	Corrected SS	1.36179E10
Coeff Variation	61.2324491	Std Error Mean	6.92658203

Basic Statistical Measures

Location		Variability	
Mean	1468.289	Std Deviation	899.06928
Median	1402.049	Variance	808326
Mode	0.000	Range	7017
		Interquartile Range	1199

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 211.9789	Pr > t	<.0001
Sign	M 7406.5	Pr >= M	<.0001
Signed Rank	S 54859946	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	7016.534
99%	3567.737
95%	2881.205
90%	2604.502
75% Q3	2092.645
50% Median	1402.049
25% Q1	893.548
10%	0.000
5%	0.000
1%	0.000
0% Min	0.000

Variable: POND MEN3

Extreme Observations

----Lowest----		-----Highest-----	
Value	Obs	Value	Obs
0	16842	6356.25	15894
0	16840	6378.20	15882
0	16835	6403.29	13442
0	16825	6429.95	8970
0	16820	7016.53	14415

ANNEXE 1 REGRESSION SUR Log(REV2)

The REG Procedure
 Model: MODEL1
 Dependent Variable: lrev2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	34	2233.69136	65.69680	280.14	<.0001
Error	10092	2366.74069	0.23452		
Corrected Total	10126	4600.43205			

Root MSE	0.48427	R-Square	0.4855
Dependent Mean	9.99593	Adj R-Sq	0.4838
Coeff Var	4.84466		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	9.16384	0.04031	227.34	<.0001
paris	1	0.13642	0.01271	10.73	<.0001
age2	1	0.15338	0.03366	4.56	<.0001
age3	1	0.26559	0.03394	7.83	<.0001
age4	1	0.22774	0.03969	5.74	<.0001
proploc1	1	0.11426	0.01142	10.01	<.0001
aidemen	1	0.17022	0.01649	10.32	<.0001
exrevs	1	0.22200	0.01794	12.37	<.0001
exrevi	1	0.09279	0.02166	4.28	<.0001
exrmi	1	-0.26680	0.03276	-8.14	<.0001
exallcho	1	-0.00754	0.02072	-0.36	0.7159
exdivi	1	0.09385	0.01204	7.80	<.0001
exloy	1	0.16230	0.01806	8.99	<.0001
exalloc	1	-0.15582	0.01476	-10.56	<.0001
typm2	1	0.51344	0.01600	32.09	<.0001
typm3	1	0.66089	0.01940	34.07	<.0001
typm4	1	0.75195	0.02270	33.13	<.0001
typm5	1	0.82525	0.02800	29.47	<.0001
typm6	1	0.27438	0.02249	12.20	<.0001
typm7	1	0.67468	0.03421	19.72	<.0001
diplo2	1	0.11653	0.01323	8.81	<.0001
diplo3	1	0.15376	0.01752	8.78	<.0001
diplo4	1	0.23511	0.01833	12.82	<.0001
csp1	1	-0.42608	0.02852	-14.94	<.0001
csp2	1	-0.09836	0.02249	-4.37	<.0001
csp3	1	0.25043	0.01692	14.80	<.0001
csp5	1	-0.12574	0.01608	-7.82	<.0001
csp6	1	-0.14542	0.01472	-9.88	<.0001
NBCHO	1	-0.17611	0.01819	-9.68	<.0001
NBETU	1	-0.03948	0.01280	-3.08	0.0021
NBRET	1	-0.02065	0.01590	-1.30	0.1943
NBINA	1	-0.19701	0.01361	-14.47	<.0001
NBCDD	1	-0.03823	0.01541	-2.48	0.0131
NBSTAGE	1	0.05539	0.04434	1.25	0.2116
NBTPAR	1	-0.13635	0.01469	-9.28	<.0001

ANNEXE 2 TESTREV= REV2/REVESTIM

testrev	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	53	0.51	53	0.51
0.1	227	2.18	280	2.69
0.2	57	0.55	337	3.23
0.3	148	1.42	485	4.65
0.4	301	2.89	786	7.54
0.5	572	5.49	1358	13.03
0.6	915	8.78	2273	21.80
0.7	1180	11.32	3453	33.12
0.8	1399	13.42	4852	46.54
0.9	1323	12.69	6175	59.23
1	1133	10.87	7308	70.10
1.1	827	7.93	8135	78.03
1.2	666	6.39	8801	84.42
1.3	392	3.76	9193	88.18
1.4	276	2.65	9469	90.83
1.5	197	1.89	9666	92.72
1.6	132	1.27	9798	93.99
1.7	106	1.02	9904	95.00
1.8	62	0.59	9966	95.60
1.9	52	0.50	10018	96.10
2	27	0.26	10045	96.35
2.1	20	0.19	10065	96.55
2.2	23	0.22	10088	96.77
2.3	23	0.22	10111	96.99
2.4	11	0.11	10122	97.09
2.5	15	0.14	10137	97.24
2.6	8	0.08	10145	97.31
2.7	8	0.08	10153	97.39
2.8	7	0.07	10160	97.46
2.9	8	0.08	10168	97.53
3	6	0.06	10174	97.59
3.1	4	0.04	10178	97.63
3.2	3	0.03	10181	97.66
3.3	2	0.02	10183	97.68
3.4	3	0.03	10186	97.71
3.5	10	0.10	10196	97.80
3.6	6	0.06	10202	97.86
3.7	6	0.06	10208	97.92
3.8	5	0.05	10213	97.97
3.9	9	0.09	10222	98.05
4	1	0.01	10223	98.06
4.1	5	0.05	10228	98.11
4.2	5	0.05	10233	98.16
4.3	9	0.09	10242	98.24
4.4	8	0.08	10250	98.32
4.5	4	0.04	10254	98.36
4.6	3	0.03	10257	98.39
4.7	6	0.06	10263	98.45
4.8	7	0.07	10270	98.51
5	7	0.07	10277	98.58
5.1	3	0.03	10280	98.61
5.2	7	0.07	10287	98.68
5.3	5	0.05	10292	98.72
5.4	4	0.04	10296	98.76
5.5	1	0.01	10297	98.77
5.6	4	0.04	10301	98.81
5.8	4	0.04	10305	98.85
5.9	2	0.02	10307	98.87
6	2	0.02	10309	98.89
6.1	3	0.03	10312	98.92
6.2	3	0.03	10315	98.94
6.3	4	0.04	10319	98.98
6.4	2	0.02	10321	99.00
6.5	3	0.03	10324	99.03

6.6	1	0.01	10325	99.04
6.7	5	0.05	10330	99.09
6.9	2	0.02	10332	99.11
7	2	0.02	10334	99.13
7.1	2	0.02	10336	99.15
7.3	4	0.04	10340	99.18
7.4	1	0.01	10341	99.19
7.5	1	0.01	10342	99.20
7.6	6	0.06	10348	99.26
7.8	3	0.03	10351	99.29
7.9	5	0.05	10356	99.34
8	5	0.05	10361	99.39
8.1	1	0.01	10362	99.40
8.2	1	0.01	10363	99.41
8.3	3	0.03	10366	99.43
8.5	1	0.01	10367	99.44
8.6	5	0.05	10372	99.49
8.7	2	0.02	10374	99.51
8.8	4	0.04	10378	99.55
8.9	2	0.02	10380	99.57
9.1	1	0.01	10381	99.58
9.2	3	0.03	10384	99.61
9.3	1	0.01	10385	99.62
9.4	1	0.01	10386	99.63
9.5	1	0.01	10387	99.64
9.6	1	0.01	10388	99.65
9.7	1	0.01	10389	99.65
9.8	2	0.02	10391	99.67
9.9	2	0.02	10393	99.69
10	1	0.01	10394	99.70
10.1	1	0.01	10395	99.71
10.3	1	0.01	10396	99.72
10.6	1	0.01	10397	99.73
10.7	2	0.02	10399	99.75
11.1	1	0.01	10400	99.76
11.4	1	0.01	10401	99.77
11.5	1	0.01	10402	99.78
11.6	1	0.01	10403	99.79
12.1	1	0.01	10404	99.80
12.2	1	0.01	10405	99.81
12.3	1	0.01	10406	99.82
12.7	1	0.01	10407	99.83
13.1	1	0.01	10408	99.84
14	1	0.01	10409	99.85
14.3	1	0.01	10410	99.86
14.8	1	0.01	10411	99.87
14.9	1	0.01	10412	99.88
15.2	1	0.01	10413	99.88
15.8	2	0.02	10415	99.90
21.4	1	0.01	10416	99.91
23.5	1	0.01	10417	99.92
23.7	1	0.01	10418	99.93
26.2	1	0.01	10419	99.94
31.1	1	0.01	10420	99.95
38.2	1	0.01	10421	99.96
47.8	1	0.01	10422	99.97
59.8	1	0.01	10423	99.98
88.4	1	0.01	10424	99.99
88.7	1	0.01	10425	100.00

Frequency Missing = 6027

ANNEXE 3 LINEARISATION DE REV3 différents essais

ANNEXE 3 (1)

Premier tour de lifereg sur table Sas Brev3
(93 ménages n'entrent pas dans les tranches)

Model Information

```

Data Set                WORK.BREV3
Dependent Variable      Log(lower)
Dependent Variable      Log(upper)
Number of Observations  15957
Noncensored Values      10345
Right Censored Values   0
Left Censored Values    0
Interval Censored Values 5612
Missing Values          494
Name of Distribution     LNORMAL
Log Likelihood          -17106.11846
    
```

Algorithm converged.

Analysis of Parameter Estimates

Standard

Variable	DF	Estimate	Error	Chi-Square	Pr >	ChiSq	Label
Intercept	1	9.00226	0.02892	96921.5321	<.0001		Intercept
paris	1	0.15115	0.0090947	276.2243	<.0001		
age2	1	0.15042	0.02565	34.3995	<.0001		
age3	1	0.26909	0.02577	108.9913	<.0001		
age4	1	0.21170	0.02957	51.2588	<.0001		
proploc1	1	0.11392	0.0082717	189.6685	<.0001		
aidemen	1	0.16477	0.01163	200.8624	<.0001		
exrevs	1	0.21913	0.01274	295.6544	<.0001		
exrevi	1	0.09969	0.01451	47.1731	<.0001		
exrmi	1	-0.25471	0.02364	116.0556	<.0001		
exdivi	1	0.12398	0.0086282	206.4690	<.0001		
exloy	1	0.18199	0.01268	205.9791	<.0001		
exalloc	1	-0.13828	0.01054	172.0463	<.0001		
typm2	1	0.52148	0.01156	2035.1555	<.0001		
typm3	1	0.61825	0.01404	1939.8362	<.0001		
typm4	1	0.73190	0.01606	2077.1379	<.0001		
typm5	1	0.83217	0.01988	1752.4817	<.0001		
typm6	1	0.26779	0.01631	269.4935	<.0001		
typm7	1	0.62046	0.02371	684.8926	<.0001		
diplo2	1	0.12882	0.0094855	184.4294	<.0001		
diplo3	1	0.20005	0.01216	270.7349	<.0001		
diplo4	1	0.31924	0.01225	679.1433	<.0001		
csp1	1	-0.37879	0.01863	413.5541	<.0001		
csp2	1	-0.03539	0.01450	5.9536	0.0147		
csp3	1	0.31481	0.01161	735.0324	<.0001		
csp5	1	-0.03688	0.0099207	13.8191	0.0002		
NBCHO	1	-0.18244	0.01089	280.6765	<.0001		
NBETU	1	-0.04258	0.0089861	22.4546	<.0001		
NBRET	1	-0.03488	0.01102	10.0148	0.0016		
NBINA	1	-0.19690	0.0097250	409.9489	<.0001		
NBCDD	1	-0.05601	0.01084	26.6926	<.0001		
NBSTAGE	1	0.08228	0.03205	6.5902	0.0103		
NBTPAR	1	-0.12484	0.01037	144.9259	<.0001		
Scale	1	0.43595	0.0024987				Normal scale

ANNEXE 3 (2)

Deuxième tour sur table Brev4=table Brev3 corrigée (24 nouveaux ménages n'entrent plus dans les tranches)

The LIFEREG Procedure

Model Information

Data Set	WORK.BREV4
Dependent Variable	Log(lower)
Dependent Variable	Log(upper)
Number of Observations	15957
Noncensored Values	10345
Right Censored Values	0
Left Censored Values	0
Interval Censored Values	5612
Missing Values	494
Name of Distribution	LNORMAL
Log Likelihood	-16251.07667

Algorithm converged.

Analysis of Parameter Estimates

Variable	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Label
Intercept	1	9.00431	0.02735	108384.232	<.0001	Intercept
paris	1	0.15486	0.0086039	323.9447	<.0001	
age2	1	0.15807	0.02426	42.4614	<.0001	
age3	1	0.28016	0.02438	132.0657	<.0001	
age4	1	0.22396	0.02797	64.1223	<.0001	
proploc1	1	0.11094	0.0078246	201.0434	<.0001	
aidemen	1	0.16394	0.01100	222.1385	<.0001	
exrevs	1	0.21708	0.01206	324.2589	<.0001	
exrevi	1	0.09581	0.01373	48.6799	<.0001	
exrmi	1	-0.25128	0.02237	126.2263	<.0001	
exdivi	1	0.11784	0.0081623	208.4255	<.0001	
exloy	1	0.17564	0.01200	214.3509	<.0001	
exalloc	1	-0.14897	0.0099732	223.1226	<.0001	
typm2	1	0.52411	0.01093	2297.4803	<.0001	
typm3	1	0.63047	0.01328	2254.1493	<.0001	
typm4	1	0.74660	0.01519	2414.9588	<.0001	
typm5	1	0.83493	0.01881	1971.1914	<.0001	
typm6	1	0.26901	0.01543	303.9342	<.0001	
typm7	1	0.63306	0.02243	796.7433	<.0001	
diplo2	1	0.13407	0.0089722	223.2874	<.0001	
diplo3	1	0.20189	0.01150	308.1623	<.0001	
diplo4	1	0.32083	0.01159	766.5525	<.0001	
csp1	1	-0.37759	0.01762	459.1182	<.0001	
csp2	1	-0.02248	0.01372	2.6854	0.1013	
csp3	1	0.31584	0.01099	826.5613	<.0001	
csp5	1	-0.03652	0.0093842	15.1425	<.0001	
NBCHO	1	-0.18329	0.01030	316.6051	<.0001	
NBETU	1	-0.03976	0.0085021	21.8751	<.0001	
NBRET	1	-0.03668	0.01043	12.3777	0.0004	
NBINA	1	-0.20011	0.0091993	473.1875	<.0001	
NBCDD	1	-0.05657	0.01026	30.4276	<.0001	
NBSTAGE	1	0.07186	0.03032	5.6177	0.0178	
NBTPAR	1	-0.12608	0.0098106	165.1512	<.0001	
Scale	1	0.41217	0.0023593			Normal scale

ANNEXE 3 (3)

Troisième tour sur table Brev5= table Brev4 corrigée (4 nouveaux ménages n'entrent plus dans les tranches)

Model Information

Data Set	WORK.BREV5
Dependent Variable	Log(lower)
Dependent Variable	Log(upper)
Number of Observations	15957
Noncensored Values	10345
Right Censored Values	0
Left Censored Values	0
Interval Censored Values	5612
Missing Values	494
Name of Distribution	LNORMAL
Log Likelihood	-16094.63386

Algorithm converged.

Analysis of Parameter Estimates

Variable	DF	Estimate	Standard		Chi-Square	Pr >	ChiSq	Label
			Error	Chi-Square				
Intercept	1	9.00813	0.02707	110755.161	<.0001			Intercept
paris	1	0.15532	0.0085153	332.7090	<.0001			
age2	1	0.15560	0.02401	42.0060	<.0001			
age3	1	0.27583	0.02413	130.6978	<.0001			
age4	1	0.22055	0.02768	63.4905	<.0001			
proploc1	1	0.11051	0.0077439	203.6621	<.0001			
aidemen	1	0.16287	0.01089	223.8145	<.0001			
exrevs	1	0.21701	0.01193	330.8463	<.0001			
exrevi	1	0.09984	0.01359	53.9550	<.0001			
exrmi	1	-0.24347	0.02213	120.9868	<.0001			
exdivi	1	0.11611	0.0080783	206.6006	<.0001			
exloy	1	0.17369	0.01187	214.0042	<.0001			
exalloc	1	-0.14916	0.0098705	228.3774	<.0001			
typm2	1	0.52572	0.01082	2360.0772	<.0001			
typm3	1	0.63282	0.01314	2318.5117	<.0001			
typm4	1	0.74660	0.01504	2465.4398	<.0001			
typm5	1	0.83598	0.01861	2017.5053	<.0001			
typm6	1	0.26768	0.01527	307.2400	<.0001			
typm7	1	0.63841	0.02220	827.2674	<.0001			
diplo2	1	0.13622	0.0088796	235.3408	<.0001			
diplo3	1	0.20291	0.01138	317.8046	<.0001			
diplo4	1	0.31955	0.01147	776.3841	<.0001			
csp1	1	-0.37384	0.01744	459.4563	<.0001			
csp2	1	-0.02453	0.01358	3.2627	0.0709			
csp3	1	0.31725	0.01087	851.3468	<.0001			
csp5	1	-0.03646	0.0092874	15.4129	<.0001			
NBETU	1	-0.04017	0.0084145	22.7898	<.0001			
NBRET	1	-0.03469	0.01032	11.3024	0.0008			
NBINA	1	-0.19950	0.0091045	480.1687	<.0001			
NBCDD	1	-0.05813	0.01015	32.7945	<.0001			
NBSTAGE	1	0.07002	0.03001	5.4441	0.0196			
NBTPAR	1	-0.12543	0.0097096	166.8731	<.0001			
Scale	1	0.40789	0.0023338					Normal scale

ANNEXE 3 (4)

Quatrième tour sur table Brev6 =tableBrev5 corrigée (2 nouveaux ménages n'entrent plus dans les tranches)

On corrigera ces ménages sans refaire tourner lifereg

The LIFEREG Procedure

```

Model Information
Data Set          WORK.BREV6
Dependent Variable Log(lower)
Dependent Variable Log(upper)
Number of Observations    15957
Noncensored Values      10345
Right Censored Values      0
Left Censored Values      0
Interval Censored Values   5612
Missing Values           494
Name of Distribution      LNORMAL
Log Likelihood           -16065.73609
    
```

Algorithm converged.

Analysis of Parameter Estimates

Variable	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Label
Intercept	1	9.00789	0.02702	111180.251	<.0001	Intercept
paris	1	0.15488	0.0084988	332.1269	<.0001	
age2	1	0.15591	0.02396	42.3378	<.0001	
age3	1	0.27536	0.02408	130.7619	<.0001	
age4	1	0.22091	0.02763	63.9485	<.0001	
proploc1	1	0.11132	0.0077289	207.4467	<.0001	
aidemen	1	0.16308	0.01087	225.2738	<.0001	
exrevs	1	0.21693	0.01191	331.8641	<.0001	
exrevi	1	0.09927	0.01357	53.5501	<.0001	
exrmi	1	-0.24449	0.02209	122.4817	<.0001	
exdivi	1	0.11560	0.0080627	205.5530	<.0001	
exloy	1	0.17565	0.01185	219.7259	<.0001	
exalloc	1	-0.14875	0.0098514	228.0050	<.0001	
typm2	1	0.52554	0.01080	2367.5666	<.0001	
typm3	1	0.63221	0.01312	2322.9917	<.0001	
typm4	1	0.74578	0.01501	2469.5431	<.0001	
typm5	1	0.83514	0.01858	2021.2577	<.0001	
typm6	1	0.26908	0.01524	311.6823	<.0001	
typm7	1	0.63773	0.02215	828.6967	<.0001	
diplo2	1	0.13695	0.0088624	238.7880	<.0001	
diplo3	1	0.20280	0.01136	318.6826	<.0001	
diplo4	1	0.31988	0.01145	781.0306	<.0001	
csp1	1	-0.37487	0.01741	463.7815	<.0001	
csp2	1	-0.02392	0.01355	3.1147	0.0776	
csp3	1	0.31747	0.01085	855.8358	<.0001	
csp5	1	-0.03705	0.0092694	15.9724	<.0001	
NBCHO	1	-0.18341	0.01017	324.9561	<.0001	
NBETU	1	-0.04024	0.0083982	22.9545	<.0001	
NBRET	1	-0.03453	0.01030	11.2416	0.0008	
NBINA	1	-0.19927	0.0090869	480.8884	<.0001	
NBCDD	1	-0.05815	0.01013	32.9525	<.0001	
NBSTAGE	1	0.07006	0.02995	5.4714	0.0193	
NBTPAR	1	-0.12552	0.0096909	167.7767	<.0001	
Scale	1	0.40709	0.0023290			Normal scale

ANNEXE 4 AUTRES SOURCES

A) PCV janvier 2003 (le plus voisin en mode de collecte), mais on ne demande que des tranches et pas de revenus en, clair

Non pondérée

The UNIVARIATE Procedure
Variable: revsimeuro

Moments

N	6007	Sum Weights	6007
Mean	25048.989	Sum Observations	150469277
Std Deviation	17372.6989	Variance	301810666
Skewness	2.42094217	Kurtosis	11.6274871
Uncorrected SS	5.58178E12	Corrected SS	1.81267E12
Coeff Variation	69.3548905	Std Error Mean	224.149862

Basic Statistical Measures

Location		Variability	
Mean	25048.99	Std Deviation	17373
Median	21371.49	Variance	301810666
Mode	2332.47	Range	205576
		Interquartile Range	19064

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 111.7511	Pr > t	<.0001
Sign	M 2998	Pr >= M	<.0001
Signed Rank	S 8989503	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	205576.22
99%	88827.15
95%	55681.33
90%	44325.13
75% Q3	32239.86
50% Median	21371.49
25% Q1	13175.82
10%	8610.76
5%	6530.11
1%	3117.17
0% Min	0.00

Variable: revsimeuro

Extreme Observations

----Lowest----		-----Highest----	
Value	Obs	Value	Obs
0	11	153140	6003
0	10	156149	6004
0	9	177469	6005
0	8	205454	6006
0	7	205576	6007

B) SOURCE ENQUÊTE BUDGET DE FAMILLE 2000 PONDEREE

Variable: REVTOT (TOTAL REVENUS HORS RESSOURCES EXCEPT. EN EUROS)

Freq: POND MEN (PONDERATION MENAGE)

Moments

N	24522571	Sum Weights	24522571
Mean	27534.4404	Sum Observations	6.75215E11
Std Deviation	20391.8857	Variance	415829003
Skewness	4.02527406	Kurtosis	41.3259438
Uncorrected SS	2.87889E16	Corrected SS	1.01972E16
Coeff Variation	74.059561	Std Error Mean	4.11788665

Basic Statistical Measures

Location		Variability	
Mean	27534.44	Std Deviation	20392
Median	23135.00	Variance	415829003
Mode	14635.00	Range	463764
		Interquartile Range	20237

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 6686.546	Pr > t	<.0001
Sign	M 12261286	Pr >= M	<.0001
Signed Rank	S 1.503E14	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	464450
99%	102101
95%	61376
90%	49322
75% Q3	34758
50% Median	23135
25% Q1	14521
10%	9798
5%	7652
1%	4340
0% Min	686

C) ENQUÊTE SANTE PONDEREE (par pondération menage visite 1)

Variable: revtot

Weight: pondmen1

Weighted Moments

N	16848	Sum Weights	24737732.2
Mean	24713.5869	Sum Observations	6.11358E11
Std Deviation	689535.763	Variance	4.7546E11
Skewness	2.67678055	Kurtosis	15.7855947
Uncorrected SS	2.31189E16	Corrected SS	8.01007E15
Coeff Variation	2790.10799	Std Error Mean	138.636266

Weighted Basic Statistical Measures

Location		Variability	
Mean	24713.59	Std Deviation	689536
Median	20130.00	Variance	4.7546E11
Mode	18000.00	Range	335000
		Interquartile Range	18104

Weighted Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 178.2621	Pr > t <.0001

Weighted Quantiles

Quantile	Estimate
100% Max	335000.00
99%	91000.00
95%	57000.00
90%	45000.00
75% Q3	31000.00
50% Median	20130.00
25% Q1	12895.84
10%	8624.25
5%	6770.60
1%	3800.00
0% Min	0.00

Extreme Observations

----Lowest----		-----Highest----	
Value	Obs	Value	Obs
0	14856	260000	1621
0	14832	260000	9034
0	14294	300000	3077
0	13577	300000	5111
0	13223	335000	2481