



## **CALCUL DES PONDERATIONS DE L'ENQUÊTE EPIC**

*Géraldine Charrance, Ined, 09/07/2015*



Le calcul des pondérations de l'enquête EPIC a été réalisé en deux étapes. Dans un premier temps, on procède à une correction de la non-réponse totale (à partir des informations disponibles dans la base de sondage, pour les répondants et les non-répondants), et, dans un second temps, on réalise un calage sur marges (marges obtenues à partir de sources externes considérées comme fiables).

### **PREMIERE ETAPE : MODELISATION DE LA NON-REPOSE TOTALE DANS L'ENQUETE EPIC**

#### **Phase 1 : Choix du champ retenu pour l'analyse**

---

Pour modéliser la participation individuelle à l'enquête, nous avons utilisé les fiches adresses (FA) dans le champ. Nous avons donc exclu de la base, pour ce travail, toutes les FA dont l'issue du terrain était codée comme suit :

- LOGEMENT VACANT
- LOGEMENT INCONNU
- REPERAGE NON FAIT
- ENQUÊTE NON ATTRIBUEE
- AUTRE HORS CHAMP
- INDIVIDU HORS CHAMP
- DEJA ENQUÊTE DANS UNE AUTRE ENQUÊTE

On compte donc dans la base utilisé 12 859 adresses (d'après le document, 12 861 FA dans le champ, il nous manque 1 FA « Refus » et une FA « Evitement »).

#### **Phase 2 : Prise en compte du mode de collecte**

---

La stratégie vise à produire un modèle unique de modélisation de la participation pour toutes les ZAE et échantillons utilisés durant l'enquête.

Toutefois, nous veillons à considérer la possibilité de tenir compte de spécificités liées au mode de contact utilisé dans chaque ZAE aux différentes étapes de l'enquête (sous-échantillons).

Parmi les FA utilisées pour modéliser la participation, voici la distribution selon le mode de collecte utilisée dans la zone :

- 5564 FA appartiennent à un sous-échantillon\*ZAE pour lequel tous les questionnaires ont été réalisés en face-à-face,
- 129 appartiennent à un sous-échantillon\*ZAE pour lequel tous les questionnaires ont été réalisés par téléphone
- 7166 où les deux modes de collecte ont été utilisés.

Des interactions seront introduites et testées afin de déterminer si les liens entre la participation et les variables explicatives retenues diffèrent selon le mode de collecte utilisé dans la zone (phase 4). Seules les interactions significatives seront retenues dans le modèle final.

### Phase 3 : Sélection des effets propres (ou simples)

La sélection des variables du modèle est ici faite sans la pondération initiale. En revanche, le modèle final (dont seront dérivés les coefficients de correction de la non-réponse) est réalisé en tenant compte des poids de sondage.

On modélise la participation à l'enquête selon les variables suivantes et selon quatre méthodes : sélection des variables dont la probabilité associée au test du  $\chi^2$  est inférieure à 0,05, sélection des variables pas à pas (stepwise), méthode descendante (backward) et stepwise sur des échantillons par bootstrap (50 réplifications, sélection des variables présentes dans au moins 60% des 50 modèles finaux).

Cette première phase cherche à être inclusive, quitte à éliminer des variables du modèle final.

Dans le tableau suivant, sont indiquées les variables retenues pour chaque méthode (les modalités de réponse de chaque variable sont disponibles en annexe).

	$P(\chi^2) < 0,05$	Stepwise	Backward	Bootstrap	Total
RR_Q1 (ZAE à faible taux de réponse (indicatrice))	X	X	X	X	4
RR_Q4 (ZAE à fort taux de réponse (indicatrice))	X	X	X	X	4
IDF (Résidence en Ile de France (indicatrice))		X	X		2
SEXE_CHEFMEN (Sexe du chef de ménage)					0
CLAGE_CHEFMEN (Classe d'âge du chef de ménage)			X	X	2
DIP_CHEFMEN (Diplôme du chef de ménage)	X	X	X	X	4
CS_CHEFMEN (Catégorie socioprofessionnelle du chef de ménage)					0
ACT_CHEFMEN (situation professionnelle du chef de ménage)		X			1
IMMIGRE_CHEFMEN (Statut d'immigration du chef de ménage)	X	X	X	X	4
MATRI_CHEFMEN (Statut marital du chef de	X		X	X	3

ménage)					
ZUS (Résidence en ZUS (indicatrice))	X	X	X		3
BATI_TYPE_HABITAT (Type d'habitation)					0
TAILLE_UU (Taille de l'unité urbaine de résidence)					0
NB_ACT_OCC_MEN (Nombre d'actifs occupés au sein du ménage)					0
NB_BAC_PLUS_3_MEN (Nombre de titulaires d'un diplôme de niveau BAC+3 ou plus au sein du ménage)					0
NB_HOMME_ELIG (Nombre d'hommes éligibles au sein du ménage)					0
NB_FEMME_ELIG (Nombre de femmes éligibles au sein du ménage)	X	X	X	X	4
TYPE_MDC (Mode de collecte utilisé dans la ZAE)					0

Les variables RR\_Q1 et RR\_Q4 sont créées à partir des données de collecte. On calcule pour chaque ZAE le taux de réponse réel (pondéré). Deux variables indicatrices sont ensuite créées indiquant si la FA se situe dans une ZAE à faible (RR\_Q1) ou fort (RR\_Q4) taux de réponse.

On conserve les variables qui ressortent sur au moins deux méthodes :

- RR\_Q1
- RR\_Q4
- IDF
- CLAGE\_CHEFMEN
- DIP\_CHEFMEN
- IMMIGRE\_CHEFMEN
- MATRI\_CHEFMEN
- ZUS
- NB\_FEMMES\_ELIG

Pour la phase suivante, on ajoutera à cette liste la variable TYPE\_MDC puisque nous allons introduire des interactions, il faut mettre dans le modèle l'effet propre.

#### **Phase 4 : Choix des interactions avec la variable TYPE\_MDC**

---

Pour vérifier que les variables introduites dans le modèle ne jouent pas différemment selon le mode de collecte utilisée dans la zone, nous introduisons toutes les interactions entre la variable TYPE\_MDC et les variables explicatives retenues.

D'après les résultats de la régression logistique, aucune interaction n'apparaît comme étant significative au seuil de 5%. Il n'est pas nécessaire de prendre en compte la variable TYPE\_MDC pour modéliser la participation à l'enquête.

## Phase 5 : Etablissement du modèle final

---

Pour modéliser la participation à l'enquête, on utilise les variables suivantes :

- RR\_Q1
- RR\_Q4
- IDF
- CLAGE\_CHEFMEN
- DIP\_CHEFMEN
- IMMIGRE\_CHEFMEN
- MATRI\_CHEFMEN
- ZUS
- NB\_FEMMES\_ELIG

<b>Modèle</b>	logit binaire
<b>Technique d'optimisation</b>	Score de Fisher

<b>Nombre d'observations lues</b>	12859
<b>Nombre d'observations utili</b>	12859

Response Profile		
Ordered Value	repondant	Total Frequency
1	1	7825
2	0	5034

La probabilité modélisée est repondant=1.

Class Level Information		
Class	Value	Design Variables
rr_q1	0	0
	1	1
rr_q4	0	0
	1	1
idf	0	0
	1	1
LZUS	0	0

Class Level Information						
Class	Value	Design Variables				
	1	1				
clage_chefmen	20-29ans	0	0	0	0	0
	30-39ans	1	0	0	0	0
	40-49ans	0	1	0	0	0
	50-59ans	0	0	1	0	0
	60-69ans	0	0	0	1	0
	70ans+	0	0	0	0	1
dip_chefmen	Aucun diplôme	0	0	0		
	BAC+3	1	0	0		
	BAC-BAC+2	0	1	0		
	BEPC,BEP,CAP	0	0	1		
immigre_chefmen_l	Immigré	1				
	Non-imm	0				
matri_chefmen_l	Divorcé	1	0	0		
	célibataire	0	0	0		
	marié ou sé	0	1	0		
	veuf	0	0	1		

#### Model Convergence Status

Critère de convergence (GCONV=1E-8) respecté.

#### Model Fit Statistics

Criterion	Intercept Only	Constante et Covariables
AIC	17217.734	16142.134
SC	17225.196	16276.446
-2 Log	17215.734	16106.134

#### Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DDL	Pr > Khi-2
Rapport de vrais	1109.6008	17	<.0001
Score	1077.4725	17	<.0001
Wald	998.9616	17	<.0001

Type 3 Analysis of Effects			
Effect	DDL	Wald Chi-Square	Pr > Khi-2
rr_q1	1	282.9312	<.0001
rr_q4	1	311.3055	<.0001
idf	1	4.7869	0.0287
LZUS	1	4.8574	0.0275
clage_chefmen	5	18.0616	0.0029
dip_chefmen	3	36.7227	<.0001
immigre_chefmen_l	1	14.7285	0.0001
matri_chefmen_l	3	8.5149	0.0365
nb_femme_elig	1	8.5487	0.0035

Analysis of Maximum Likelihood Estimates						
Parameter		DDL	Estimate	Standard Error	Wald Chi-Square	Pr > Khi-2
Intercept		1	0.1113	0.1025	1.1804	0.2773
rr_q1	1	1	-0.8178	0.0486	282.9312	<.0001
rr_q4	1	1	0.8834	0.0501	311.3055	<.0001
idf	1	1	-0.1095	0.0501	4.7869	0.0287
LZUS	1	1	0.1689	0.0766	4.8574	0.0275
clage_chefmen	30-39ans	1	0.0544	0.0955	0.3247	0.5688
clage_chefmen	40-49ans	1	0.0517	0.0955	0.2935	0.5880
clage_chefmen	50-59ans	1	-0.00623	0.0977	0.0041	0.9492
clage_chefmen	60-69ans	1	-0.0635	0.1015	0.3912	0.5317
clage_chefmen	70ans+	1	-0.3949	0.1394	8.0322	0.0046
dip_chefmen	BAC+3	1	0.3744	0.0664	31.7978	<.0001
dip_chefmen	BAC-BAC+2	1	0.2604	0.0581	20.1144	<.0001
dip_chefmen	BEPC,BEP,CAP	1	0.1459	0.0561	6.7649	0.0093
immigre_chefmen_l	Immigré	1	-0.2376	0.0619	14.7285	0.0001
matri_chefmen_l	Divorcé	1	0.1349	0.0673	4.0149	0.0451
matri_chefmen_l	marié ou sé	1	0.1108	0.0480	5.3284	0.0210
matri_chefmen_l	veuf	1	0.2413	0.1098	4.8328	0.0279
nb_femme_elig		1	0.1260	0.0431	8.5487	0.0035

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
rr_q1 1 vs 0	0.441	0.401	0.486
rr_q4 1 vs 0	2.419	2.193	2.669
idf 1 vs 0	0.896	0.813	0.989
LZUS 1 vs 0	1.184	1.019	1.376
clage_chefmen 30-39ans vs 20-29ans	1.056	0.876	1.273
clage_chefmen 40-49ans vs 20-29ans	1.053	0.873	1.270
clage_chefmen 50-59ans vs 20-29ans	0.994	0.821	1.203
clage_chefmen 60-69ans vs 20-29ans	0.938	0.769	1.145
clage_chefmen 70ans+ vs 20-29ans	0.674	0.513	0.885
dip_chefmen BAC+3 vs Aucun diplôme	1.454	1.277	1.656
dip_chefmen BAC-BAC+2 vs Aucun diplôme	1.297	1.158	1.454
dip_chefmen BEPC,BEP,CAP vs Aucun diplôme	1.157	1.037	1.292
immigre_chefmen_I Immigré vs Non-imm	0.789	0.698	0.890
matri_chefmen_I Divorcé vs célibataire	1.144	1.003	1.306
matri_chefmen_I marié ou sé vs célibataire	1.117	1.017	1.227
matri_chefmen_I veuf vs célibataire	1.273	1.027	1.578
nb_femme_elig	1.134	1.042	1.234

Association of Predicted Probabilities and Observed Responses			
Pourcentage concordant	66.3	D de Somers	0.338
Pourcentage discordant	32.5	Gamma	0.342
Pourcentage lié	1.2	Tau-a	0.161
Paires	39391050	c	0.669

## Phase 6 : Calcul des coefficients de correction de la non-réponse totale

A partir des probabilités prédites par ce modèle (prenant en compte les poids de sondage), on constitue des groupes homogènes de réponse (déciles). Dans chaque groupe, on calcule le taux de réponse réel pondéré. On multiplie ensuite le poids de sondage de chaque individu par l'inverse du taux de réponse observé.

Avec cette étape de correction de la non-réponse totale, le rapport des poids max/min passe de 12,61 à 15,31 pour les ménages et de 40,97 à 42,85 pour les individus. La correction de la non-réponse n'affecte que faiblement la dispersion des poids.

## Phase 7 : Contrôle des résultats

Afin de contrôler que la correction de la non-réponse totale améliore la représentativité de l'échantillon des répondants, nous allons comparer la distribution de quelques variables de la base de sondage.

La distribution de référence est celle observée sur l'échantillon tiré pondéré par le poids de sondage (colonne « Echantillon PDS » dans le tableau ci-dessous). Nous allons la comparer à celle observée sur les répondants pondérés par le poids de sondage (colonne « Répondants PDS »), puis par le poids de sondage multiplié par le coefficient de correction de non-réponse (colonne « Répondants PDS x NR »).

Au vu du tableau ci-dessous, on constate rapidement que la structure de la population des répondants pondérés par le poids de sondage corrigé de la non-réponse semble plus proche de celle de l'échantillon que lorsque l'on utilise uniquement le poids de sondage sans aucune correction.

Pour confirmer ce constat, nous avons calculé la distance du  $\text{Khi}^2$  normalisée (calculée sans tenir compte de la taille des échantillons) pour l'échantillon de répondants pondéré par le poids de sondage et l'échantillon de répondants pondéré par le poids de sondage et corrigé de la non réponse, relativement à la population cible, celle de l'échantillon initial pondéré par le poids de sondage. Les distances sont nettement inférieures pour l'échantillon corrigé de la non réponse et les diminutions relatives de distances supérieures à 90% pour toutes les variables sauf le sexe (86.6%). Ceci prouve l'efficacité de la procédure de correction de la non-réponse totale par le modèle proposé.

**Tableau des distributions de quelques variables sociodémographiques pour l'échantillon initial, l'échantillon de répondants non corrigé et corrigé de la non-réponse totale par modélisation logistique.**

	Echantillon PDS	Répondants PDS	Khi <sup>2</sup> <sup>a</sup>	Répondants PDS × NR	Khi <sup>2</sup> <sup>b</sup>	Diff. Khi <sup>2</sup> <sup>c</sup> (a-b)/a
Sexe du chef de ménage						
Hommes	77.65	78.06	0.002	77.5	0.000	86.62 %
Femmes	22.35	21.94	0.008	22.5	0.001	
			0.010		0.001	
Classe d'âge du chef de ménage						
20-29ans	4.86	4.65	0.009	4.93	0.001	91.95%
30-39ans	19.93	20.08	0.001	19.77	0.001	
40-49ans	25.31	25.86	0.012	25.22	0.000	
50-59ans	25.72	25.92	0.002	26.08	0.005	
60-69ans	20.55	20.49	0.000	20.47	0.000	
70ans+	3.62	3.01	0.103	3.53	0.002	
			0.127		0.010	
Statut matrimonial du chef de ménage						
Célibataire	35.17	34.06	0.035	35.46	0.002	92.65%
Marié ou séparé	48.85	49.99	0.027	48.54	0.002	
Divorcé	12.12	12.05	0.000	12.12	0.000	
Veuf	3.86	3.9	0.000	3.89	0.000	
			0.062		0.005	
Diplôme du chef de ménage						
Aucun diplôme	19.07	17.41	0.144	19.18	0.001	98.72%
Bepc,bep,cap	33.91	34.5	0.010	34.04	0.000	
Bac-bac+2	30.51	31.51	0.033	30.4	0.000	
Bac+3	16.51	16.57	0.000	16.39	0.001	
			0.188		0.002	
Taille du ménage						
Une personne	24.47	22.99	0.090	24.31	0.001	



<b>Deux personnes</b>	30.39	30.45	0.000	30.29	0.000	
<b>Trois personnes</b>	19.41	19.46	0.000	19.36	0.000	
<b>Quatre personnes</b>	17.41	18.61	0.083	17.58	0.002	
<b>Cinq personnes ou plus</b>	8.32	8.49	0.003	8.46	0.002	
			0.176		0.006	96.86%
<b>Résidence en Ile-de-France</b>						
<b>Ile-de-France</b>	20.03	16.78	0.527	19.86	0.001	
<b>Province</b>	79.97	83.22	0.132	80.14	0.000	
			0.659		0.002	99.73%
<b>Taille de l'unité urbaine de résidence</b>						
<b>Rurales</b>	24.19	26.73	0.267	24.19	0.000	
<b>&lt; 20 000 ha</b>	14.46	14.52	0.000	14.35	0.001	
<b>20 - 100 000 hab</b>	11.51	12.1	0.030	11.78	0.006	
<b>&gt; 100 000 ha</b>	31.46	31.37	0.000	31.48	0.000	
<b>Uu paris</b>	18.38	15.27	0.526	18.2	0.002	
			0.824		0.009	98.91%
<b>Statut d'immigration</b>						
<b>Non immigré</b>	88.73	90.37	0.030	88.71	0.000	
<b>Immigré</b>	11.27	9.63	0.239	11.29	0.000	
			0.269		0.000	100.0%

<sup>a, b</sup> : Contribution à la distance du Khi<sup>2</sup> normalisée (indépendante de la taille des échantillons) et distance normalisée. Par exemple, la contribution pour le sexe masculin est  $(78.06 - 77.65)^2 / 77.65 = 0.002$  entre l'échantillon initial et l'échantillon de répondants pondérés tous deux par le poids de sondage.

<sup>c</sup> : D Khi<sup>2</sup> : diminution relative de la distance du Khi<sup>2</sup> normalisée. Plus elle est élevée, plus importante est l'amélioration de la distribution de la variable relativement à l'échantillon initial pondéré par le poids de sondage.

PDS=poids de sondage.

## **DEUXIEME ETAPE : CALAGE SUR MARGES**

A l'issue de la correction de la non-réponse totale, on réalise un calage sur marges. Les marges utilisées sont les suivantes (les modalités de réponse des variables sont présentées en annexe):

Marges de niveau individu

- **Sexe\*âge**
- **Taille d'unité urbaine**
- **Diplôme le plus élevé obtenu**
- **ZEAT de résidence**
- **Catégorie socio-professionnelle**
- **Nationalité**
- **Résidence en ZUS**
- **Type de ménage**

Ces marges sont issues de l'EEC T1 à T4<sup>1</sup> 2013 et EEC T1 à T4 2012. En règle générale, les marges pour une année donnée sont calculées à partir des 4 enquêtes emploi trimestrielles. Toutefois, en 2013, suite à certains problèmes rencontrés dans l'enquête Emploi, les marges relatives à la catégorie socioprofessionnelle et à la taille de l'unité urbaine ont été obtenues en conservant les structures issues de l'EEC 2012 et en les appliquant aux niveaux issus de l'EEC 2013.

Avant de procéder au calage, nous avons comparé la distribution des variables de calage selon que l'on tient compte de la correction de la non-réponse totale ou non.

Pour trois des huit variables de calage, la distribution observée avec le poids de sondage corrigée de la non-réponse totale est moins proche de la cible que si l'on ne corrige pas la non-réponse totale (cf. tableau ci-dessous).

Cependant, ces distributions étant de toute façon contrôlées par le calage sur marges, il convient de conserver le poids corrigé de la non-réponse totale pour réaliser cette opération afin de corriger une part plus importante de la non-réponse et de prendre en compte plus de variables dans le redressement.

---

<sup>1</sup> Enquête Emploi en Continu Trimestre 1 à Trimestre 4.

Tableau des distributions des variables de calage pour l'échantillon des répondants non corrigé et corrigé de la non-réponse totale.

	Cible	Répondants	Khi <sup>2</sup> <sup>a</sup>	Répondants	Khi <sup>2</sup> <sup>b</sup>	Diff. Khi <sup>2</sup> <sup>c</sup> (a-b)/a
	EEC	PDS		PDS × NR		
<b>Sexe du chef de ménage</b>						
Hommes de 25 à 29 ans	5,68	3,84	0,60	3,98	0,51	
Hommes de 30 à 34 ans	5,92	4,57	0,31	4,47	0,35	
Hommes de 35 à 39 ans	6,13	4,87	0,26	4,73	0,32	
Hommes de 40 à 44 ans	6,59	5,77	0,10	5,65	0,13	
Hommes de 45 à 49 ans	6,56	5,95	0,06	6,02	0,04	
Hommes de 50 à 54 ans	6,28	6,03	0,01	6,12	0,00	
Hommes de 55 à 59 ans	5,94	6,26	0,02	6,17	0,01	
Hommes de 60 à 64 ans	5,81	6,09	0,01	6,04	0,01	
Femmes de 25 à 29 ans	5,87	5,33	0,05	5,31	0,05	
Femmes de 30 à 34 ans	6,20	6,49	0,01	6,41	0,01	
Femmes de 35 à 39 ans	6,20	6,64	0,03	6,43	0,01	
Femmes de 40 à 44 ans	6,80	7,57	0,09	7,66	0,11	
Femmes de 45 à 49 ans	6,74	7,65	0,12	7,54	0,10	
Femmes de 50 à 54 ans	6,60	7,14	0,04	7,16	0,05	
Femmes de 55 à 59 ans	6,36	7,91	0,38	8,11	0,48	
Femmes de 60 à 64 ans	6,33	7,89	0,38	8,21	0,56	
			2,47		2,74	-10,93%
<b>Taille de l'unité urbaine de résidence</b>						
Rural	23,96	28,24	0,76	25,69	0,12	
Agglomérations de moins de 20 000 habitants	16,77	14,73	0,25	14,61	0,28	
Agglomérations de 20 000 à 99 999 habitants	12,22	11,81	0,01	11,58	0,03	
Agglomérations de 100 000 à 1 999 999 habitants	29,78	30,45	0,02	30,59	0,02	
Agglomération parisienne	17,27	14,77	0,36	17,54	0,00	
			1,40		0,46	67,09%
<b>Diplôme le plus élevé obtenu</b>						
Diplôme supérieur	17,9	21,68	0,80	21,6	0,76	
BAC et BAC+2	31,2	29,45	0,10	28,96	0,16	
BEP, CAP et BEPC	31,7	32,78	0,04	32,5	0,02	
Sans diplôme et diplôme non déclaré	19,3	16,09	0,53	16,94	0,29	
			1,47		1,23	15,86%
<b>ZEAT de résidence</b>						
Ile-de-France et Bassin parisien	36,20	33,82	0,16	37,06	0,02	
Nord, Est et Ouest	28,50	31,36	0,29	28,81	0,00	
Sud-Ouest, Centre-Est et bassin méditerranéen	35,30	34,82	0,01	34,14	0,04	
			0,45		0,06	86,24%
<b>CS</b>						
Agriculteurs	1,75	2,11	0,07	1,95	0,02	
Indépendants et professions libérales	7,76	7,17	0,04	7,22	0,04	
Cadres et professions intermédiaires actifs	31,52	30,09	0,07	29,81	0,09	
Anciens cadres et professions intermédiaires	5,47	8,51	1,69	8,49	1,67	
Employés et ouvriers actifs	37,73	34,46	0,28	34,18	0,33	
Anciens employés et ouvriers	12,06	14,73	0,59	15,27	0,86	
Inactifs et chômeurs n'ayant jamais travaillé	3,70	2,93	0,16	3,08	0,10	
			2,91		3,11	-7,07%

<b>Nationalité</b>						
Nationalité française	92,90	94,33	0,02	93,50	0,00	
Autres	7,10	5,67	0,29	6,50	0,05	
			0,31		0,05	82,40%
<b>Résidence en ZUS</b>						
Non ZUS	94,20	94,34	0,00	93,80	0,00	
ZUS	5,80	5,66	0,00	6,20	0,03	
			0,00		0,03	-716,33%
<b>Nombre d'individus du champ appartenant à un ménage de type :</b>						
Ménages d'une personne	16,49	14,40	0,27	15,34	0,08	
Couples sans enfant	25,55	28,03	0,24	27,91	0,22	
Famille monoparentale	7,62	7,69	0,00	7,99	0,02	
Couples avec un enfant	17,18	17,29	0,00	17,12	0,00	
Couples avec deux enfants ou plus	29,31	29,54	0,00	28,47	0,02	
Autres	3,84	3,05	0,16	3,18	0,11	
			0,67		0,45	32,37%

A l'issue du calage sur marges, il faut contrôler les statistiques de poids, et notamment le rapport max/min (ci-dessous, les statistiques relatives aux 4 jeux de pondérations).

Le rapport max/min observé sur le poids de sondage est de 40,97. En corrigeant la non-réponse totale, il passe à 42,85. Après calage, il passe à 128,03 si on utilise le poids de sondage sans correction de la non réponse totale et à 98,01 en cas de correction de non-réponse totale.

Il convient donc d'introduire une phase de correction de la non-réponse totale avant de procéder au calage.

**Le poids retenu pour l'analyse est donc le poids résultant du calage après correction de la non-réponse totale.**

### POIDS DE SONDAGE

Moments			
N	7825	Somme des poids	7825
Moyenne	4163.20498	Somme des observations	32577079
Ecart-type	1725.83151	Variance	2978494.41
Skewness	1.30742983	Kurtosis	5.2678899
Somme des carrés non corrigée	1.58929E11	Somme des carrés corrigée	2.33037E10
Coeff Variation	41.4543968	Std Error Mean	19.5099529

Mesures statistiques de base			
Emplacement		Variabilité	
Moyenne	4163.205	Ecart-type	1726
Médiane	4286.705	Variance	2978494
Mode	5009.829	Intervalle	20596
		Ecart interquartile	2501

Tests de tendance centrale : Mu0=0				
Test	Statistic		p Value	
T de Student	t	213.3888	Pr >  t	<.0001
Signe	M	3912.5	Pr >=  M	<.0001
Rang signé	S	15309613	Pr >=  S	<.0001

Quantiles (Definition 5)	
Niveau	Quantile
100% Max	21111.40
99%	9940.11
95%	7080.13
90%	6038.29
75% Q3	5029.84
50% Median	4286.71
25% Q1	2528.71
10%	2234.56
5%	2062.81
1%	1683.64
0% Min	515.30

Observations extrêmes			
Le plus bas		Le plus haut	
Valeur	Obs	Valeur	Obs
515.300	168	15153.4	4258
564.270	76	17519.4	5606
564.270	71	19334.6	1285
588.959	47	19991.8	1109
596.394	198	21111.4	7013

**POIDS DE SONDAGE CORRIGE DE LA NON REPONSE TOTALE**

Moments			
N	7825	Somme des poids	7825
Moyenne	4163.20498	Somme des observations	32577079
Ecart-type	1989.38454	Variance	3957650.86
Skewness	1.79090413	Kurtosis	6.97848276
Somme des carrés non corrigée	1.6659E11	Somme des carrés corrigée	3.09647E10
Coeff Variation	47.7849289	Std Error Mean	22.4893325

Mesures statistiques de base			
Emplacement		Variabilité	
Moyenne	4163.205	Ecart-type	1989
Médiane	3863.524	Variance	3957651
Mode	3830.941	Intervalle	23246
		Ecart interquartile	2284

Tests for Location: Mu0=0				
Test	Statistic		p Value	
T de Student	t	185.1191	Pr >  t	<.0001
Signe	M	3912.5	Pr >=  M	<.0001
Rang signé	S	15309613	Pr >=  S	<.0001

Quantiles (Definition 5)	
Niveau	Quantile
100% Max	23801.975
99%	10652.286
95%	7845.471
90%	6617.511
75% Q3	4996.937
50% Median	3863.524
25% Q1	2713.329
10%	2038.379
5%	1779.980
1%	1464.685
0% Min	555.514

Observations extrêmes			
Le plus bas		Le plus haut	
Valeur	Obs	Valeur	Obs
555.514	76	17686.4	2860
694.646	71	17712.1	7013
734.863	168	21308.4	7322
896.820	284	21567.3	5606
908.065	88	23802.0	1285

**POIDS APRES CALAGE REALISE AVEC POIDS DE SONDAGE (SANS CORRECTION NON REPONSE)**

Moments			
N	7825	Somme des poids	7825
Moyenne	4163.20498	Somme des observations	32577079
Ecart-type	2188.72683	Variance	4790525.14
Skewness	2.11610176	Kurtosis	10.3637873
Somme des carrés non corrigée	1.73106 <sup>E11</sup>	Somme des carrés corrigée	3.74811 <sup>E10</sup>
Coeff Variation	52.5731219	Std Error Mean	24.742831

Mesures statistiques de base			
Emplacement		Variabilité	
Moyenne	4163.205	Ecart-type	2189
Médiane	3744.769	Variance	4790525
Mode	4229.258	Intervalle	28604
		Ecart interquartile	2528

Tests for Location: Mu0=0				
Test	Statistic		p Value	
T de Student	t	168.259	Pr >  t	<.0001
Signe	M	3912.5	Pr >=  M	<.0001
Rang signé	S	15309613	Pr >=  S	<.0001

Quantiles (Definition 5)	
Niveau	Quantile
100% Max	28828.821
99%	11707.957
95%	7882.650
90%	6698.157
75% Q3	5185.760
50% Median	3744.769
25% Q1	2657.840
10%	1997.107
5%	1635.510
1%	1022.281
0% Min	225.175

Observations extrêmes			
Le plus bas		Le plus haut	
Valeur	Obs	Valeur	Obs
225.175	3693	20187.7	1110
243.118	4552	23059.0	5606
298.939	4296	25098.6	1285
324.041	284	26429.2	7013
359.325	168	28828.8	1109

**POIDS APRES CALAGE REALISE AVEC POIDS DE SONDAGE CORRIGE DE LA NON REPONSE TOTALE)**

Moments			
N	7825	Somme des poids	7825
Moyenne	4163.20498	Somme des observations	32577079
Ecart-type	2416.84262	Variance	5841128.23
Skewness	2.30856153	Kurtosis	11.8243424
Somme des carrés non corrigée	1.81326E11	Somme des carrés corrigée	4.5701E10
Coeff Variation	58.052453	Std Error Mean	27.3216043

Mesures statistiques de base			
Emplacement		Variabilité	
Moyenne	4163.205	Ecart-type	2417
Médiane	3669.557	Variance	5841128
Mode	955.013	Intervalle	29746
		Ecart interquartile	2611

Tests for Location: Mu0=0				
Test	Statistic		p Value	
T de Student	t	152.3778	Pr >  t	<.0001
Signe	M	3912.5	Pr >=  M	<.0001
Rang signé	S	15309613	Pr >=  S	<.0001

Quantiles (Definition 5)	
Niveau	Quantile
100% Max	30052.266
99%	12325.969
95%	8584.181
90%	7060.856
75% Q3	5144.332
50% Median	3669.557
25% Q1	2533.584
10%	1807.630
5%	1487.173
1%	989.639
0% Min	306.625

Observations extrêmes			
Le plus bas		Le plus haut	
Valeur	Obs	Valeur	Obs
306.625	4296	24734.7	97
315.476	284	27622.3	1605
373.442	168	28984.6	1285
407.449	1760	29282.0	5606
413.466	3693	30052.3	7322



## ANNEXES

### Noms, libellés en clair et modalités de réponse des variables utilisées pour la modélisation de la non-réponse totale

Nom de la variable	Libellé en clair	Modalités
RR_Q1	Appartenance à une ZAE à faible taux de réponse	Oui Non
RR_Q4	Appartenance à une ZAE à fort taux de réponse	Oui Non
IDF	Résidence en Ile-de-France	Oui Non
SEXE_CHEFMEN	Sexe du chef de ménage	Homme Femme
CLAGE_CHEFMEN	Classe d'âge du chef de ménage	20-29 ans 30-39 ans 40-49 ans 50-59 ans 60-69 ans 70 ans et plus
DIP_CHEFMEN	Diplôme le plus élevé obtenu par le chef de ménage	Aucun diplôme BEP,CAP et BEPC BAC-BAC+2 BAC+3 et plus
CS_CHEFMEN	Catégorie socioprofessionnelle du chef de ménage	Indépendants Cadres et professions intellectuelles sup. Professions intermédiaires Employés Ouvriers Chômeurs ou inactifs
ACT_CHEFMEN	Situation professionnelle du chef de ménage	Actifs occupés Chômeurs Inactifs
IMMIGRE_CHEFMEN	Statut d'immigration du chef de ménage	Immigré Non immigré
MATRI_CHEFMEN	Statut marital du chef de ménage	Célibataires Mariés ou séparés Divorcés veufs
ZUS	Résidence en ZUS	Oui Non
BATI_TYPE_HABITAT	Type de logement	Logement individuel Logement collectif
TAILLE_UU	Taille de l'unité urbaine	Communes rurales Communes des unités urbaines de moins de 20000 hab. Communes des unités urbaines entre 20 000 et 99 999 hab. Communes des unités urbaines de 100 000 hab ou plus Communes des unités urbaines de Paris
NB_ACT_OCC_MEN	Nombre d'actifs occ. dans le ménage	Variable continue (val. Min : 0 / val max : 5)
NB_BAC_PLUS_3_MEN	Nombre de titulaires d'un diplôme de niveau bac+3 ou plus au sein du ménage	Variable continue (val. Min : 0 / val max : 5)
NB_HOMMES_ELIG	Nombre d'hommes éligibles au sein du ménage	Variable continue (val. Min : 0 / val max : 5)
NB_FEMMES_ELIG	Nombre de femmes éligibles au sein du ménage	Variable continue (val. Min : 0 / val max : 4)
TYPE_MDC	Mode de collecte de la ZAE	Exclusivement en face-à-face Exclusivement par téléphone Deux modes : face-à-face et par téléphone

**Noms, libellés en clair et modalités de réponse des variables utilisées pour le calage sur marges**

Nom de la variable	Libellé en clair	Modalités
SEXE_AGE	Sexe et âge du répondant	Hommes de 25 à 29 ans Hommes de 30 à 34 ans Hommes de 35 à 39 ans Hommes de 40 à 44 ans Hommes de 45 à 49 ans Hommes de 50 à 54 ans Hommes de 55 à 59 ans Hommes de 60 à 64 ans Femmes de 25 à 29 ans Femmes de 30 à 34 ans Femmes de 35 à 39 ans Femmes de 40 à 44 ans Femmes de 45 à 49 ans Femmes de 50 à 54 ans Femmes de 55 à 59 ans Femmes de 60 à 64 ans
TR_UU	Taille de l'unité urbaine de résidence	Rural Agglomérations de moins de 20 000 hab. Agglomérations de 20 000 à 99 999 hab. Agglomérations de 100 000 à 1 999 999 hab. Agglomération parisienne
DIPLOME	Diplôme du répondant	Diplôme supérieur Bac à Bac+2 BEP, CAP et BEPC Sans diplôme et diplôme non déclaré
ZEAT	Région de résidence	Ile-de-France et bassin Parisien Nord, Est et Ouest Sud-Ouest, centre-Est et bassin méditerranéen
CS	Catégorie socioprofessionnelle du répondant	Agriculteurs Indépendants et professions libérales Cadres et professions intermédiaires actifs Anciens cadres et professions intermédiaires Employés et ouvriers actifs Anciens employés et ouvriers Inactifs et chômeurs n'ayant jamais travaillé
NATIO	Nationalité du répondant	Nationalité française Autre nationalité
ZUS	Résidence en ZUS	Oui Non
TYP_MEN	Type et taille du ménage	Ménage d'une personne Couple sans enfant Famille monoparentale Couple avec un enfant Couple avec deux enfants ou plus Autres