

Enquête santé DOM 2019 : Calcul des pondérations

Cyril Favre-Martinoz (Insee)

Champ de diffusion géographique : il correspond à l'intégralité de la Guadeloupe, de la Martinique et de La Réunion. Dans le cas où certaines zones, regroupant une fraction de population faible, avaient été écartées au moment du tirage, c'est le calage sur marges qui permet d'assurer la diffusion sur ces zones. En Guyane, le champ de diffusion est limité aux 12 communes de la zone côtière élargie.

Codes résultat : Au total, 8 693 logements ont été échantillonnés mais du fait des éclatements de logements en cours de collecte, un total de 8 781 logements ont pris part à la collecte. Les codes résultats en fin de collecte sont les suivants :

Tableau 1 • Ventilation du nombre de logements par codes résultat agrégés

Hors-champ	1 180
Non-répondants	1 448
Répondants partiels	451
Répondants	5 702

Il y a donc un total de 6 153 ménages répondants (5 702+451). Ils contiennent 9 835 individus. Une étude de la non-réponse partielle dans l'enquête a été menée. Il a été défini une liste de 89 variables centrales dont 22 correspondent aux questions posées sous casque. On considère un individu comme non-répondant total s'il a plus de 30 non-réponses parmi ces 89 questions posées. Ainsi, 71 individus sont requalifiés en non-répondants. Finalement, on dispose de 6 139 ménages répondants avec 9 764 individus. Le traitement de la non-réponse par DROM va donc porter sur ces effectifs :

Tableau 2 • Ventilation du nombre ménages répondants/non-répondants par DROM

	Répondant	Non-répondant	Total
Guadeloupe	1 300	180	1 480
Martinique	1 195	311	1 506
Guyane	1 344	403	1 747
Réunion	1 078	406	1 484
Mayotte	1 222	162	1 384
Total tout DROM	6 139	1 462	7 601

Correction de la non-réponse totale au niveau du ménage :

Elle se fait par pondération en trois étapes :

- 1) Estimation DROM par DROM d'un modèle logit expliquant l'indicatrice de réponse des ménages par un ensemble de variables auxiliaires ;
- 2) Estimation de la probabilité de réponse pour l'ensemble des ménages grâce à ce modèle ;
- 3) Construction de groupes de réponse homogène (GRH) pour lisser les probabilités de réponse selon la méthode d'Haziza et Beaumont. Chaque ménage appartenant à un même groupe de réponse homogène aura la même probabilité de réponse à savoir le nombre de ménages répondants du groupe sur le nombre total de ménages du groupe.

Pour l'étape 1), la liste des variables figurant en annexe 1 a été testée puis un modèle a été sélectionné selon une méthode « backward » basée sur le critère d'information AIC.

Les modèles retenus par DROM sont les suivants, les notations renvoient à l'annexe 1 :

- Guadeloupe :

$rep \sim natloc + i_qpv + nb_pers_rg + i_sal_log + i_ben_ind_log + lgt_social + i_enf_6_10 + i_55_64 + i_75p$

- Martinique :

$rep \sim i_ASC + ZAE + natloc + i_pens_all_log + lgt_social + i_15_24 + i_25_34 + SSECH + i_pauvre60m + i_alfm + i_zfonm + i_zvamm + tx_men_pauv + tx_ind_65p$

- Guyane :

$rep \sim ZAE + ACHL_rg + i_chom_log + i_chom_succ + revdispm_decile_r + lgt_social + i_enf_0_14 + i_enf_11_14 + i_enf_6_10 + i_45_54 + i_55_64 + i_65_74 + i_75p + densite_tranche$

- La Réunion :

$rep \sim ZAE + natloc + nb_pers_rg + ACHL_rg + i_enf_11_14 + i_enf_3_5 + i_15_24 + i_65_74 + type_menf + i_aplm + i_minim + i_zvamm + tx_men_5ind + tx_ind_65p$

- Mayotte :

$rep \sim adresse + type + nb_niv + part_zeroact + part_deuxemp + part_logdur + part_internet + part_unepers + part_proprio$

À la suite des trois étapes, on obtient les poids corrigés de la non-réponse avec les GRH sur les répondants (tableaux 3 à 7).

Correction de la non-réponse totale au niveau individuel :

Cette étape permet de prendre en compte la sélection des individus Kish et leur non-réponse éventuelle. La non-réponse des individus Kish a été très faible, mais il y a eu quelques cas de ménages avec un unique Kish répondant alors que deux au sein du ménage avaient été tirés.

Pour prendre en compte la sélection des individus interrogés dans chaque logement, il suffit de déterminer à partir du tableau des habitants du logement, le nombre de personnes dans le ménage relevant du champ de l'enquête (15 ans ou plus) :

- si le nombre de membres du ménage appartenant au champ de l'enquête est inférieur ou égal à 2, ils sont tous interrogés ;
- sinon, deux membres du ménage sont sélectionnés suivant la méthode de la date d'anniversaire¹ qui est assimilable à un sondage aléatoire simple sans remise de deux personnes parmi l'ensemble des membres du ménage dans le champ de l'enquête.

Finalement, les 6 139 ménages répondants à l'enquête contiennent 16 969 occupants, dont 12 262 relevant du champ de l'enquête et 10 099 sont échantillonnés pour répondre à l'enquête. Parmi ces 10 099 personnes échantillonnées, on dénombre 9 764 répondants, soit un taux de réponse individuel de l'ordre de 96 %.

La correction de la non-réponse (CNR) au niveau individuel a suivi la même méthode que pour la correction de la non-réponse au niveau du ménage. Compte-tenu du très faible taux de non-réponse, le nombre de variables auxiliaires introduites dans les modèles logit est limité : le sexe de l'individu, son statut d'activité et son âge quinquennal sont les seules variables retenues. Elles sont issues de la collecte de la partie ménage du questionnaire (le tronc commun des ménages) avant tirage des Kish. Les poids obtenus sont les poids individuels après CNR ménages et individus (tableaux 3 à 7).

Calage sur marges :

Le calage consiste alors à modifier les poids de l'échantillon pour que les estimateurs des totaux des variables auxiliaires concernées appelées variables de calage, soient égaux aux valeurs connues par ailleurs dans la population. La population cible de l'enquête, comme précisé plus haut, est l'ensemble des individus de 15 ans et plus habitant un logement ordinaire à titre de résidence principale dans le champ géographique de diffusion. Le calage permet de réduire la variance pour les variables mobilisées dans le calage mais aussi pour celles qui leur sont corrélées. Il permet aussi de réduire le biais de

¹ Dans les quatre DROM historique, les Kish sélectionnés étaient ceux dont la date de naissance était la plus proche du 1^{er} juillet après le 1^{er} juillet. Le 1^{er} juillet était proche de la date de lancement de l'enquête. À Mayotte, il a été préféré faire un Kish purement aléatoire car les personnes qui ne connaissaient pas leur date de naissance s'attribuent le plus souvent une date de naissance au 1^{er} janvier.

couverture de l'enquête. Ce biais porte sur le champ géographique puisque certaines zones n'ont pas été échantillonnées alors qu'elles figurent dans l'objectif de diffusion. Il porte aussi sur les résidences qui se sont construites ou sont devenues des résidences principales entre le millésime de la base de sondage (1^{er} janvier 2018) et la collecte (2019).

Les marges retenues pour les quatre DROM historiques sont précisées ci-dessous. Elles ont été calculées au 1^{er} juillet 2019, proche de la date de lancement de l'enquête et sont calculées sur la totalité des départements sauf en Guyane où seule la zone côtière élargie est visée en termes de diffusion.

- La répartition de la population par sexe et âge ;
- Le lieu de naissance (né dans un DROM/en France métropolitaine/à l'étranger) ;
- Le niveau de diplôme le plus élevé (sans diplôme/CAP-BEP/Bac/diplôme du supérieur) ;
- La catégorie socio-professionnelle (sur une position) ;
- Le type de logement (individuel/collectif) ;
- La répartition estimée de la population de 15 ans ou plus par micro-région ;
- La répartition de la population de 15 ans ou plus par type de ménage (personne seule/famille monoparentale/couple sans enfant/couple avec un enfant/couple avec plusieurs enfants/Autre) ;
- La répartition de la population de 15 ans ou plus par statut d'occupation du ménage (locataire ou logé gratuitement/propriétaire).

Pour Mayotte, les marges sont moins nombreuses : répartition de la population par sexe et âge et répartition de la population par lieu de naissance (France hors Mayotte/Mayotte/étranger).

Le poids obtenu est le poids calé (tableaux 3 à 7). Il s'agit du poids définitif utilisé pour la publication des premiers résultats et mis à disposition dans les données. Ce sont des poids individuels.

Finalement, les distributions de poids calés sont très correctes. Le facteur multiplicatif entre le 1^{er} et le 9^e décile n'excède pas 5 pour la Guadeloupe, Martinique, la Réunion et Mayotte et 8 pour la Guyane. Pour la Guyane, la dispersion des poids est plus élevée notamment car la taille des ménages est en moyenne plus élevée. Par conséquent, les poids des individus Kish, qui sont calculés comme le nombre de personnes éligibles divisés par 1 ou par 2 en fonction du nombre de personnes éligibles du ménage, sont plus dispersés. En effet, si on se réfère au tableau 5, une grande partie de la dispersion est déjà observable sur les poids corrigés de la non-réponse au niveau ménage et individu.

L'étude des valeurs influentes de poids met en avant que dans tous les DROM les 25 % des individus qui ont les plus gros poids représentent de l'ordre de 45 % des estimations, ce qui semble tout à fait raisonnable sauf en Guyane où le taux s'élève à 57 %. On constate par ailleurs, que 5 % des répondants avec les poids les plus élevés comptent pour 13 % à 21 % des estimations selon les départements.

Tableau 3 • Distribution des poids aux différentes étapes de la repondération (Guadeloupe)

	Poids initiaux répondants-non répondants	Poids initiaux répondants	Poids corrigés de la non-réponse sur les répondants	Poids corrigés de la non-réponse avec GRH sur les répondants	Poids corrigés de la non-réponse ménages et individus sur les répondants	Poids calés
Somme des observations	141 110	123 823	141 114	141 107	242 072	307 282
Moyenne	95	95	109	109	122	155
Écart-type	24	24	28	29	39	85
Coefficient de variation	0,2	0,2	0,3	0,3	0,3	0,5
Min	79	79	82	81	82	42
10 %	79	79	85	83	87	77

25 % Q1	79	79	88	88	92	99
50 % Médiane	80	80	94	94	102	132
75 % Q3	130	129	141	140	148	190
90 %	132	132	152	151	168	259
Max	134	134	217	177	455	1014

Tableau 4 • Distribution des poids aux différentes étapes de la repondération (Martinique)

	Poids initiaux répondants-non répondants	Poids initiaux répondants	Poids corrigés de la non-réponse sur les répondants	Poids corrigés de la non-réponse avec GRH sur les répondants	Poids corrigés de la non-réponse ménages et individus sur les répondants	Poids calés
Somme des observations	147 407	116 473	147 330	147 337	251 440	298 190
Moyenne	98	97	123	123	138	164
Écart-type	20	19	36	36	49	92
Coefficient de variation	0,2	0,2	0,3	0,3	0,4	0,6
Min	84	84	85	88	88	49
10 %	84	84	91	91	93	81
25 % Q1	84	84	97	97	101	98
50 % Médiane	85	85	110	111	125	139
75 % Q3	125	125	143	140	162	199
90 %	127	128	169	169	203	279
Max	129	129	381	258	445	1035

Tableau 5 • Distribution des poids aux différentes étapes de la repondération (Guyane)

	Poids initiaux répondants-non répondants	Poids initiaux répondants	Poids corrigés de la non-réponse sur les répondants	Poids corrigés de la non-réponse avec GRH sur les répondants	Poids corrigés de la non-réponse ménages et individus sur les répondants	Poids calés
Somme des observations	51 655	39 591	51 628	51 684	99 287	167 808
Moyenne	30	29	38	38	46	77

Écart-type	24	24	34	34	39	78
Coefficient de variation	0,8	0,8	0,9	0,9	0,8	1,0
Min	20	20	20	20	20	8
10 %	20	20	21	20	22	20
25 % Q1	20	20	22	23	25	31
50 % Médiane	20	20	25	25	31	53
75 % Q3	20	20	31	31	49	92
90 %	88	88	101	103	103	163
Max	94	94	286	229	354	848

Tableau 6 • Distribution des poids aux différentes étapes de la repondération (La Réunion)

	Poids initiaux répondants-non répondants	Poids initiaux répondants	Poids corrigés de la non-réponse sur les répondants	Poids corrigés de la non-réponse avec GRH sur les répondants	Poids corrigés de la non-réponse ménages et individus sur les répondants	Poids calés
Somme des observations	312 854	224 419	311 910	312 321	612 302	659 085
Moyenne	211	208	289	290	347	374
Écart-type	68	66	141	139	180	222
Coefficient de variation	0,3	0,3	0,5	0,5	0,5	0,6
Min	174	174	176	181	185	102
10 %	175	175	185	183	198	178
25 % Q1	176	176	200	204	216	229
50 % Médiane	178	177	231	230	295	315
75 % Q3	179	179	332	339	420	461
90 %	343	342	462	441	542	670
Max	397	397	1039	859	2804	3523

Tableau 7 • Distribution des poids aux différentes étapes de la repondération (Mayotte)

	Poids initiaux répondants-non répondants	Poids initiaux répondants	Poids corrigés de la non-réponse sur les répondants	Poids corrigés de la non-réponse avec GRH sur les répondants	Poids corrigés de la non-réponse ménages et individus sur les répondants	Poids calés
Somme des observations	53 861	47 516	53 757	53 869	112 956	154 343
Moyenne	39	39	44	44	56	76
Écart-type	0	0	10	11	24	39
Coefficient de variation	0,0	0,0	0,2	0,2	0,4	0,5
Min	39	39	39	39	39	34
10 %	39	39	40	40	40	40
25 % Q1	39	39	40	41	41	54
50 % Médiane	39	39	41	41	44	66
75 % Q3	39	39	44	44	63	85
90 %	39	39	49	46	88	123
Max	41	41	138	106	250	429

Annexe 1. Variables testées dans les modèles logit d'estimation de la probabilité de réponse en vue de la repondération de l'enquête Santé DOM

Variables testées pour les quatre DROM historiques :

NOM VARIABLE	LIBELLE	Commentaire
achl	année d'achèvement du logement regroupé selon les tranches du RP	
i_asc	Indicatrice de présence d'un ascenseur	
code_base	logement connu à la TH, requalifié en res princ (2) or crée (3)	
i_alfm	indicatrice de perception d'allocations logement à caractère familial en n-1	
i_alm_log	perception d'une pension alimentaire au niveau logement	
i_alsm	Indicatrice de perception de l'allocation logement à caractère social en n-1	Uniquement 972,974
i_aplm	indicatrice de perception d'APL en n-1	Uniquement 972,974
i_alfm	indicatrice de perception d'allocations familiales en n-1	Uniquement 972,974
i_ben_agri_log	Perception de bénéfiques agricoles au niveau logement	
i_ben_ind_log	perception de bénéfiques individuels	
i_ben_non_com_log	Perception de bénéfiques non commerciaux	
i_chom_log	indicatrice de perception du chômage en l'année n au niveau logement	
i_chom_succ	indicatrice de perception du chômage en l'année n et n-1 au niveau logement	
i_enf_0_14	indicatrice de présence d'enfants de moins de 15 ans au 1 ^{er} janvier 2018	
i_enf_0_2	indicatrice de présence d'enfants de moins de 3 ans au 1 ^{er} janvier 2018	
i_enf_11_14	indicatrice de présence d'enfants de 11 à 14 ans au 1 ^{er} janvier 2018	
i_enf_3_5	indicatrice de présence d'enfants de 3 à 5 ans au 1 ^{er} janvier 2018	
i_enf_6_10	indicatrice de présence d'enfants de 6 à 10 ans au 1 ^{er} janvier 2018	
i_15_24	indicatrice de présence de personnes âgées entre 15 et 24 ans au 1 ^{er} janvier 2018	
i_25_34	indicatrice de présence de personnes âgées entre 25 et 34 ans au 1 ^{er} janvier 2018	
i_35_44	indicatrice de présence de personnes âgées entre 35 et 44 ans au 1 ^{er} janvier 2018	
i_45_54	indicatrice de présence de personnes âgées entre 45 et 54 ans au 1 ^{er} janvier 2018	
i_55_64	indicatrice de présence de personnes âgées entre 55 et 64 ans au 1 ^{er} janvier 2018	
i_65_74	indicatrice de présence de personnes âgées entre 65 et 74 ans au 1 ^{er} janvier 2018	
i_75p	indicatrice de présence de personnes âgées de plus de 75 ans au 1 ^{er} janvier 2018	
i_minim	indicatrice de perception de minima sociaux en n-1	Uniquement 972,974
i_pauvre60m	indicatrice de ménage pauvre à 60 % (non disponible pour 971 et 973)	Uniquement 972,974
i_pens_all_log	indicatrice informant si tous les individus du logements perçoivent une pension	
i_pens_log	perception d'une pension l'année n dans le logement	
i_pfamm	perception d'allocation familiale dans le logement	Uniquement 972,974
i_qpv	indicatrice de quartier prioritaire de la ville	
i_res_secondaire	Un des cohabitants de l'individu cible possède d'une résidence secondaire	
i_sal_log	perception d'un salaire dans le logement l'année n	
i_sal_log_succ	perception d'un salaire dans le logement l'année n-1 et l'année n	
i_zavlm	perception de pensions alimentaires	Uniquement 972,974
i_zfonm	perception de revenus fonciers	Uniquement 972,974
i_zvamm	perception de produits de placement revenus fixes ou de revenus non soumis au prélèvement libératoire	Uniquement 972,974
lgt_social	indicatrice de logement social	
micro_region	micro_région des DOM	
natloc	type de local	
nb_pers	Nombre de personnes dans le logement	
nbpi	nombre total de pièces	
nivviem_decile	niveau de vie en décile	
retrouve_filosofi	indicatrice ménage retrouvé dans Filosofi	Uniquement 972,974
revdispm_decile	revenu disponible l'année n découpé en déciles	
ssech	numéro du sous-échantillon	
stocd	statut d'occupation type recensement	
surf	surftot regroupé selon les tranches du recensement	
type_menf	type de ménage fiscal	
type_res1	indicatrice résidence principal au sens RP	
zae	sae dans les DOM : secteur action enquêteur	
Variables externes envisagées		
densite_tranche	Densité de population au carreau ou à l'iris pour la Guadeloupe	
Variables Filosofi		
tx_men_pauv	Taux de ménages pauvres	Uniquement 972,974
tx_men_5ind	Taux de ménages de 5 individus ou plus	
tx_men_prop	Taux de ménage propriétaires	
tx_men_fmp	Taux de ménages familles monoparentales	
tx_men_coll	Taux de ménages en logements collectifs	
tx_log_soc	Taux de logements sociaux	

Variables testées pour Mayotte :

NOM VARIABLE	LIBELLE
com	commune
geog	zone géographique
adresse	type d'adresse (monologement, adresses moyennes, grandes adresses)
type	type de bâti (dur, tôle)
i_qpv	Indicatrice d'appartenance à un QPV
nb_niv	Nombre d'étages
groupe_vil	Typologie des villages
groupe	Typologie des communes
Variables externes construites à partir des données du RP 2017 agrégées à l'îlot	
part_zeroact	part des ménages sans individu actif dans l'îlot
part_logdur	part des logements en dur dans l'îlot
part_pas65ans	part des ménages sans individu de 65 ans ou plus dans l'îlot
part_nonsco	part des ménages sans personne scolarisée dans l'îlot
part_65ans	part des ménages avec au moins une personne âgée de 65 ans ou plus dans l'îlot
part_paselec	part des logements sans électricité dans l'îlot
part_unepers	part des ménages d'une personne dans l'îlot
part_grandmen	part des ménages de grande taille dans l'îlot
part_zeroemp	part des ménages sans individu en emploi dans l'îlot
part_LOGSEUL	part des logements individuels dans l'îlot
part_pas_eau	part des logements sans eau courante dans l'îlot
part_pas16ans	part des ménages sans personne de moins de 16 ans dans l'îlot
part_surfagri	part des logements avec une parcelle agricole dans l'îlot
part_etudiant	part des ménages sans jeune de 14 ans ou plus en étude dans l'îlot
part_proprio	part des ménages propriétaires dans l'îlot
part_deuxemp	part des ménages avec deux personnes en emploi dans l'îlot
part_telfixe	part des ménages ayant un téléphone fixe dans l'îlot
part_nonacces	part des ménages sans accès routier à moins de 50 mètres dans l'îlot
part_pas24ans	part des ménages sans individu de 24 ans ou moins dans l'îlot
part_internet	part des ménages ayant accès à internet dans l'îlot
part_petitmen	part des ménages de petite taille dans l'îlot