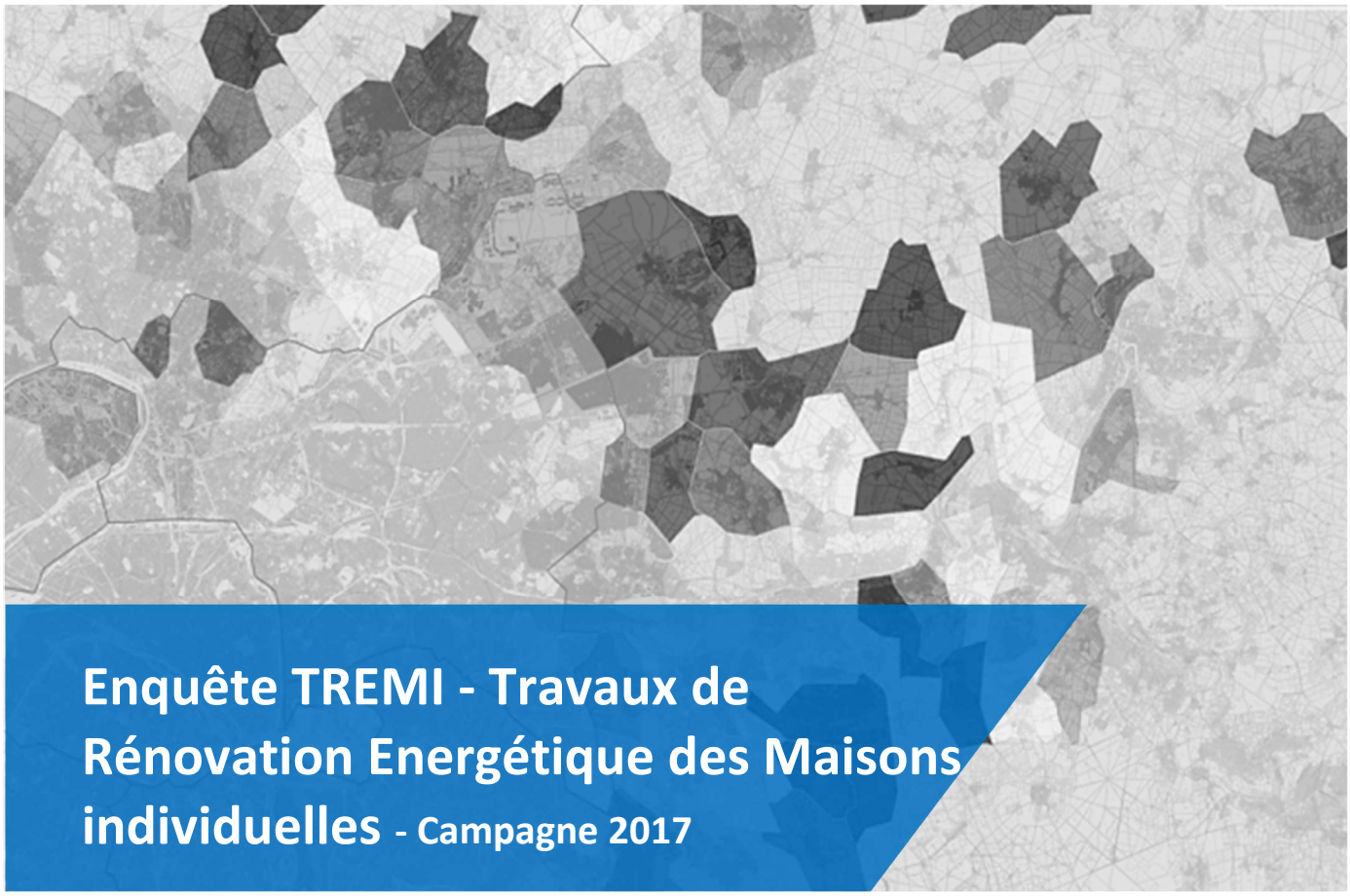

Enquête TREMI - Travaux de Rénovation Energétique des Maisons individuelles - Campagne 2017

Note méthodologique sur la Campagne 2017

Novembre 2018

énergies demain

KANTAR PUBLIC

POUGET Consultants

I. Echantillon et traitement statistique

I.1. Base de sondage et échantillonnage

Base(s) de sondage :

Population de référence = ménages vivant en France continentale.

Échantillonnage :

Tirage aléatoire des ménages sollicités au sein de chaque région.

44921 foyers ont été interrogés.

- Le nombre de foyers sollicités dans les bases de sondage utilisés n'est pas connu et aucun traitement de la non-réponse n'est fait à ce niveau.31% des foyers interrogés ont réalisé des travaux *(foyers ayant réalisé des travaux de rénovation dans leur logement entre 2014 et 2016 et dont au moins une partie des travaux s'est déroulée en 2016).*

	Ensemble		N'ont pas réalisé de travaux		Ont réalisé des travaux	
Nombre d'interviews réalisées	**44921**	**Dont maisons individuelles : 29253**	**30840**	**Dont maisons individuelles :19289**	**14081**	**Dont maisons individuelles : 9964**
	100%	**100%**	**69%**	**66%**	**31%**	**34%**
Grand Est (Alsace, Champagne-Ardenne, Lorraine)	4366	2728	3066	1835	1300	893
Nouvelle Aquitaine (Aquitaine, Limousin, Poitou-Charentes)	4348	3262	2991	2199	1357	1063
Auvergne, Rhône-Alpes	5150	2835	3610	1898	1540	937
Bourgogne, Franche-Comté	2354	1636	1616	1063	738	573
Bretagne	3089	2315	2191	1604	898	711
Centre Val de Loire	2343	1808	1627	1227	716	581
Ile-de-France	4680	1776	3131	997	1549	779
Occitanie (Languedoc-Roussillon, Midi-Pyrénées)	4602	3198	3142	2099	1460	1099
Hauts de France (Nord, Pas-de-Calais, Picardie)	4495	3509	2991	2255	1504	1254
Normandie (Basse-Normandie, Haute-Normandie)	2539	1823	1731	1194	808	629
Pays de la Loire	3113	2402	2151	1644	962	758
Provence-Alpes-Côte d'Azur	3842	1961	2593	1274	1249	687

I.2. Traitements statistiques

I.2.a. Cleaning des données

L'administration du questionnaire par internet, donc avec des guidages, limite fortement le risque d'erreur de déclaration. De plus, des fenêtres « pop up » ont été programmées pour s'ouvrir si les sondés entraient des valeurs aberrantes. Les valeurs ont été bornées pour certaines questions.

Toutefois, nous avons corrigé quelques déclarations sur les variables suivantes :

Demandes de l'ADEME	**Traduction**
Surface du logement	
Changer toutes les valeurs > 400m² par 400m² pour les appartements	Si (Q101 = 2 ou 3) et (Q41>400) alors mettre la valeur 400 en Q41
Changer toutes les valeurs < 40m² par 40m² pour les maisons	Si (Q101=1) et (Q41<40) alors mettre la valeur 40 en Q41
Nombre de fenêtres / Portes fenêtres / baies vitrées dans le logement avant travaux	
Pour les maisons individuelles : Si Q101=1 ou 3	Si (Q31_x > (Q41/10) * 3) alors (Q31_x = (Q41/10) * 3) Pour x=1,…,4 (cela signifie que le calcul est réalisé pour Q31-1 à Q31_4)
Pour les logements collectifs : Si Q101=2	Si (Q31_x > (Q41/10) * 2) alors (Q31_x = (Q41/10) * 2) Pour x=1,…,4
Nombre de fenêtres / Portes fenêtres / baies vitrées concernées par les travaux	
Pour q13A : plafonner avec la somme de Q31_1 et Q31_2	Si Q13a.1 (simple vitrage) > Q31.1+Q31_2 (tout type de vitrage) alors Q13a.1 = Q31.1+Q31_2
	Si Q13a.2 (double vitrage) > Q31.1+Q31_2 (tout type de vitrage) alors Q13a.2 = Q31.1+Q31_2
Pour q13B : plafonner avec la somme de Q31_3 et Q31_4	Si Q13b.1 (simple vitrage) > Q31_3+Q31_4 (tout type de vitrage) alors Q13b.1 = Q31_3+Q31_4
	Si Q13b.2 (double vitrage) > Q31_3+Q31_4 (tout type de vitrage) alors Q13b.2 = Q31_3+Q31_4
Montant des travaux	
Pour les montants des travaux par type de travaux : considérer comme valeurs manquantes les valeurs en dessous de 10 euros. Pour les montants des travaux par type de travaux : considérer comme valeurs manquantes les valeurs en dessous de 50 euros (et reconstruire le montant avec le modèle d'imputation).	Si Q20 < 10 mettre à blanc la réponse => les valeurs <10 ne sont donc pas imputées Si 10<Q20<50, mettre à blanc la réponse et imputer une valeur

La base de données livrée à l'Ademe comporte pour les variables corrigées la version d'origine et la version finale.

I.2.b. Redressement et extrapolation

Redressement

Le redressement vise à corriger l'échantillon enquêté de ses éventuelles déformations par rapport à la population cible de l'enquête (les ménages vivant en France continentale).

Nous avons appliqué la méthode du **calage sur marge** et essayé successivement plusieurs redressements.

La macro SAS CALMAR (CALage sur MARges) permet de redresser un échantillon provenant d'une enquête par sondage, par repondération des individus, en utilisant une information auxiliaire disponible sur un certain nombre de variables, appelées variables de calage. Le redressement consiste à remplacer les pondérations initiales (ou "poids de sondage") par de nouvelles pondérations telles que :

- pour une variable de calage catégorielle (ou "qualitative"), les effectifs des modalités de la variable estimés dans l'échantillon, après redressement, seront égaux aux effectifs connus sur la population ;
- pour une variable numérique (ou "quantitative"), le total de la variable estimé dans l'échantillon, après redressement, sera égal au total connu sur la population.

Nous avons utilisé deux sources pour les données de redressement :

Source : Recensement de la population 2014 (RP 2014)

- Age personne de référence du ménage (en 5 tranches)
- PCS personne de référence du ménage (en 3 modalités)
- Nombre de personnes au foyer (4 tranches)
- Taille agglo (5 tranches)
- Statut logement (propriétaire, locataire HLM, locataire hors HLM)
- Type de logement (maison, appartement)

Source : Filocom 2015

- Année de construction du logement (7 tranches)

Nous avons appliqué un redressement par région sur les 7 variables et avons pondéré les sous-échantillons de chacune des régions afin de leurs redonner leur poids réel au niveau national.

Il a été décidé, après avoir procédé au redressement des données sur les 44921, ménages de ne garder dans l'univers OPEN que les maisons individuelles.

Le redressement appliqué en amont n'avait pas pour but de redresser spécifiquement cet univers plus restreint de maisons individuelles. Rappelons que celui-ci a été appliqué aux 44921 interviews, donc sur l'échantillon total, et non sur le seul périmètre finalement retenu (MI).

Toutefois, en comparant la structure des maisons individuelles ainsi obtenue aux statistiques du recensement de la population (sur les critères statut logement, nombre de personnes au foyers, catégorie socio-professionnelle et tranche d'âge du chef de ménage, agglomération), les écarts observés étaient très faibles et le redressement a donc été validé par le SDES, présent lors du COPIL.

Résultats du redressement : pour rappel, un redressement a des conséquences sur la précision des résultats à cause de la disparité des poids. Il faut s'assurer que le redressement n'aboutit pas à une trop grande dispersion des poids. Le tableau suivant donne des informations sur cette dispersion.

base	44921	**Exemples de lecture :**
moyenne	1	
somme	44921	

poids_min	0,25	
poids_max	4	
percentile_5	0,31	
percentile_10	0,34	
quartile_1	0,42	25% des poids sont <0,42
médiane	0,67	50% des poids sont supérieurs à 0,67 et 50% sont inférieurs
quartile_3	1,24	25% des poids sont supérieurs à 1,24 et 75% sont inférieurs
percentile_90	2,31	
percentile_95	2,92	5% des poids sont supérieurs à 2,92
efficacité	59%	

Le coefficient d'efficacité représente l'accroissement de variance dû à la déformation de l'échantillon durant le redressement. Il mesure l'effet négatif entrainé par la distorsion des poids, par le fait d'avoir des poids très différents. Pour le calculer on utilise la somme des poids au carré. Ce taux est d'autant plus faible que les écarts entre les poids sont importants.
Exemple 1
On a 500 individus avec poids=0.8 et 500 individus avec poids=1.2
Base brute =1000 / base redressée = 500x0.8 + 500 x 1.2= 400+600=1000
Efficacité redressement = 1000 / (500x0.8x0.8 + 500x1.2x1.2)=1000/ (320+ 720)=1000/1040=0.96 efficacité=96 %

Exemple 2
On a 500 individus avec poids=0.25 et 500 individus avec poids=1.75
Base brute =1000 / base redressée = 500x0.25 + 500x1.75= 125+875=1000
Efficacité redressement = 1000 / (500x0.25x0.25 + 500x1.75x1.75)=1000/ (31.25+ 1531.25)=1000/1562.5=0.64 efficacité=64 %

Dans notre cas, l'efficacité sur le redressement de 59% sur l'échantillon de 44921 ménages. Cela signifie que la base ne « vaut », statistiquement parlant, que pour 26503 ménages (base effective). (= 44921 x 0.59). Cela se traduit par une diminution de la précision car c'est cette base qui est utilisée pour mesurer l'intervalle de confiance des résultats obtenus.

Cas particuliers de la question Q70 posée à un échantillon plus restreint:

Pour la question Q70 (« *Parmi ces aides à la réalisation de travaux de rénovation énergétique, lesquelles connaissez-vous, ne serait-ce que de nom ?* ») une personne sur cinq a été interrogée afin de diminuer la longueur du questionnaire, soit 6168 individus. Afin de se recaler sur une base de 44921, chacun des 6168 individus a un poids de 5 (14081 +(6168x5) =44921 correspondant à l'échantillon global). En redressé, comme nous avons des différences de poids, nous obtenons les chiffres suivants : 13038 + (6316,6 x 5) =44621. Il faut utiliser la variable poidsq70 au lieu de la variable poids pour avoir des résultats redressés et poidsq70_extra pour avoir des résultats extrapolés.

Ces remarques sont également valables pour la question Q80, avec laquelle il faut utiliser les mêmes poids « poidsq70 » et « poidsq70_extra ».

Extrapolation

Une extrapolation est le principe par lequel on estime que les résultats d'une enquête effectuée sur un échantillon peuvent être généralisés à la population étudiée dans son ensemble (en l'occurrence dans cette enquête il s'agit des ménages ayant réalisé des travaux dans leur logement principal sur la période définie).

Nous extrapolons les chiffres de chaque région de façon à pouvoir fournir des effectifs « réels » sur les critères qui nous intéressent.

Extrapolation à 27 903 024 ménages (équivalent à « résidence principale » - RP 2014).

Le coefficient d'extrapolation est donc de 621,16. Calcul : 27 903 024 /44 921.

I.2.c. Imputations statistiques

Une imputation consiste à imputer une valeur appropriée en cas de valeur manquante.

Il s'agit ici de combler les valeurs manquantes (5%) pour le montant des travaux (question Q20 du questionnaire). Il s'agit de 5% des gestes sachant que la base des gestes est d'environ 34 000 gestes.

Ce sont uniquement les montants des travaux qui ont fait l'objet d'une imputation (cf bas du tableau paragraphe 1.2. a qui parle des imputations réalisées), les autres variables descriptives du ménage et des travaux réalisés devant être précisées par le répondant et étant utilisées pour imputer les coûts manquants.

La méthode d'imputation utilisée est « l'estimateur du plus proche voisin ».

Comme son nom l'indique, avec cette méthode, un individu n'ayant pas répondu prend la valeur de l'individu le plus proche ayant répondu. Pour déterminer quel est l'individu le plus proche, un score est calculé pour chacun via une régression logistique. Plus les scores sont proches, plus les individus le sont également.

La régression logistique est faite en prenant comme variable à modéliser le fait d'avoir répondu ou non à la Q20 (variable à imputer) et comme variables explicatives la liste ci-dessous. Le score issu de cette régression permet de trier les individus par proximité de réponses sur l'ensemble des variables explicatives. Ainsi, pour chaque non-répondant, on peut trouver ainsi un individu ayant le maximum de caractéristiques communes. Le montant donné par cette personne est alors reporté sur l'individu n'ayant pas répondu.

Les variables explicatives suivantes ont été utilisées :

- Q1 type de travaux
- Q41 surface habitable
- Q100 Statut du logement (locataire propriétaire)
- Q101 type du logement (appartement / maison)
- Q102 année de construction
- RS6 profession
- RS182 revenus du ménage
- Q18 qui a réalisé les travaux

Et pour certains travaux, ces variables ont également été pris en compte :

- Q10 type d'isolant
- Q11 part de la surface isolation
- Q12 épaisseur isolant
- Q13 nombre de fenêtres / porte-fenêtres
- Q14 système de chauffage
- Q15 système eau chaude
- Q16 système ventilation
- Q17 système de climatisation

Pour valider la méthode avant de l'appliquer, nous avons calculé un ratio entre la moyenne avant et la moyenne après l'imputation et vérifier que celui-ci n'était pas trop éloigné de 1. Nous avons réalisé ce test sur 3 échantillons aléatoires différents, les 3 échantillons comprenant l'ensemble des interviewés et l'ensemble des gestes réalisés, mais en mettant à blanc 5% des valeurs[1] à chaque fois et en effectuant l'imputation.

Pour déterminer si les résultats sont satisfaisants, on utilise la moyenne, min et max des ratios.

[1] Ce sous-échantillon de 5% est différent sur les 3 échantillons testés.

Les ratios servent à valider l'imputation. Sur chacun des 3 échantillons testés, le ratio est réalisé entre la valeur déclarée et la valeur imputée pour chacun des 20 gestes. On obtient donc, pour chacun des 3 échantillons testés, 20 ratios et on fait la moyenne des 20 ratios.

Rappelons que l'imputation a été faite sur l'ensemble des interviewés et l'ensemble des gestes réalisés.

On obtient :

- Version utilisée : moy=1.0075 min=0.9829 max=1.0744
- Test1 moy=1.0029 min=0.9391 max=1.0428
- Test2 moy=0.9882 min=0.8939 max=1.0308
- Test3 moy=0.9988 min=0.9241 max=1.0742

Ces tests ont donné des résultats satisfaisants car proches de 1. Etant donné la bonne qualité des résultats obtenus lors des 3 tests nous nous sommes arrêtés à 3 tests.

L'imputation est intégrée dans la table de données brutes transmises à l'Ademe.

Le fichier de données comporte les variables finales qui tiennent compte des imputations et des variables d'origine.

Recommandation pour une prochaine vague d'enquête :

- Bien définir en amont le type de logements que l'on souhaite étudier
- Bien définir en amont le type de travaux rentrant dans le champ de l'étude
- Si l'on souhaite privilégier une seule source de données pour le redressement, poser les questions avec des modalités (tranches) compatibles avec celles de la source statistique choisie
- Pour la question de la notoriété des aides : si l'on souhaite la comparer à la notoriété obtenue dans d'autres études en ligne, garder la même formulation pour les questions
- Si l'on veut un seul poids pour le traitement de l'étude, poser à tous les questions sur la notoriété des aides (dans la mesure où le temps de passation prévu du questionnaire le permet) afin de ne pas créer un poids spécifique pour cette question

ANNEXE : Document Excel nommé « TREMI 2017-Note_methodo-stat_annexe.xls »

Onglet 1 : Questions utilisées pour les variables de redressement

Onglet 2 : Statistiques pour le redressement

Onglet 3 : Données redressées