

COMPARAISON DE TROIS MÉTHODES DE CORRECTION DE LA NON-RÉPONSE TOTALE UTILISANT LES PARADONNÉES DANS L'ENQUÊTE TÉLÉPHONIQUE FECOND

Géraldine Charrance¹, Stéphane Legleye² et Nicolas Razafindratsima³

¹ *Institut national d'études démographiques - geraldine.charrance@ined.fr*

² *Institut national d'études démographiques – stephane.legleye@ined.fr*

³ *Institut national d'études démographiques – razafind@ined.fr*

Résumé.

La baisse actuelle des taux de réponse, malgré les efforts mis en place lors de la collecte, peut conduire à biaiser et réduire la précision des estimateurs. Pour limiter le biais, il est courant de corriger l'échantillon par des méthodes de repondération (post-stratification ou calage).

Il est aussi possible d'utiliser les variables de la base de sondage, disponibles pour les répondants et les non-répondants, pour proposer un traitement préalable de la non-réponse totale. Dans les enquêtes téléphoniques par génération aléatoire de numéros, la base de sondage ne comporte pas d'autres informations que le numéro de téléphone. Cependant, les variables créées lors du processus de collecte, aussi appelées paradonnées, peuvent éventuellement être mobilisées pour corriger la non-réponse, puisque disponibles pour tous.

Dans cette communication, nous présenterons et comparerons trois méthodes de constitution de groupes homogènes de réponse à partir des paradonnées d'une enquête téléphonique, qui peuvent être utilisées pour corriger la non-réponse totale. La méthode classique est celle de la modélisation logistique de la participation. Les deux autres, plus originales, reposent sur l'élaboration de typologies de non-répondants : soit une analyse des composantes principales sur les paradonnées, puis une classification sur les premières composante, soit une analyse harmonique qualitative afin de tenir compte de la chronologie des événements, puis une classification ascendante hiérarchique sur les résultats.

Les performances de ces méthodes sont appréciées en termes d'hétérogénéité inter-classes du point de vue sociodémographique et des variables d'intérêt de l'enquête, ainsi que par la dispersion des poids finaux et la réduction du biais sociodémographique après correction de la non-réponse.

Mots-clés.

Traitement de la non-réponse ; enquête téléphonique ; paradonnées ; classification ; analyse harmonique qualitative.

L'objectif de cette communication est de comparer et d'évaluer différentes méthodes de constitution de groupes de réponse homogènes pour la correction de la non-réponse totale dans une enquête téléphonique en utilisant les données relatives aux issues d'appels (paradonnées) : par la modélisation de la probabilité de réponse, par la classification ascendante hiérarchique sur les composantes d'une ACP et par la classification sur les résultats d'une analyse harmonique qualitative des historiques d'appels. La première méthode est classique. Elle se concentre sur la probabilité de réponse. Elle peut comporter le désavantage de ne pas tenir compte de la structure des corrélations entre les variables et surtout d'agréger des individus disparates du point de vue des paradonnées, car leur regroupement n'est fait que du point de vue de la probabilité de réponse. La deuxième méthode tient compte des relations entre les paradonnées et mobilise donc plus d'informations que la première, mais ne tient pas compte du temps. La troisième tient compte explicitement du temps et donc de la relation entre toutes les issues d'appels successives dans le processus de collecte. Elle intègre donc a priori plus d'informations que les précédentes.

1. Eléments de contexte

La présence de non-réponse dans une enquête implique la production d'estimateurs potentiellement biaisés. Sans traitement et correction de la non-réponse, l'estimateur est non biaisé si l'on vérifie l'hypothèse que les non-répondants ont un comportement similaire aux répondants, que ces deux populations sont substituables l'une à l'autre. Or, dans la littérature scientifique, il a été montré que cette hypothèse est rarement vérifiée. Il est même très probable que la seule probabilité de participation cache des comportements d'intérêt très divers. En effet, un individu peut être difficile à enquêter parce qu'il est difficilement joignable en raison de son emploi du temps, parce qu'il filtre ses appels entrants ou les visites qu'il reçoit à son domicile, ou parce qu'il a d'abord refusé de participer avant d'accepter dans un second temps.

Pour redresser un échantillon de façon efficace, sans augmenter la variance, il faut disposer d'une information auxiliaire possédant quatre propriétés (Little et Vartivarian, 2005 ; Sarndal et Lundstrom, 2005) :

1. Etre disponible sans valeur manquante sur les répondants et les non répondants
2. Etre mesurée sans erreur
3. Etre associée aux comportements d'intérêts mesurés dans l'enquête
4. Etre un prédicteur de la participation à l'enquête.

Dans une enquête téléphonique à génération aléatoire de numéros de téléphone, donc sans base de sondage, les historiques d'appels, comprenant pour chaque numéro la succession des issues des appels (décrites suivant une certaine typologie) vérifie le point 1, et assez largement le point 2, car une large partie de l'information est codée sans erreur par l'automate d'appel (ligne occupée ou non, rang de l'appel), et l'enquêteur (refus, sans réponse, demande de rendez-vous, début de passation). Il vérifie sans peine le point 4, presque par définition.

Ainsi, l'historique des appels reflète la disponibilité téléphonique et la bienveillance ou la perméabilité vis-à-vis du téléphone et des enquêtes. Il pourrait être utile au redressement si
Ainsi, une post-stratification ou calage sur les variables sociodémographiques omettent très vraisemblablement une information importante : la difficulté à joindre et à convaincre les individus
Pour limiter le biais des estimateurs, on peut procéder à la correction de la non-réponse totale avant calage.

Nous avons donc mobilisé ces variables pour corriger la non-réponse totale dans l'enquête FECOND. Cette enquête, conduite en 2010, porte sur les comportements sexuels et reproductifs des Français. Elle a été réalisée par téléphone et a permis de collecter 8 650 questionnaires. Lors de cette enquête, certaines mesures ont été prises pour améliorer le taux de participation : les numéros ont été appelés un grand nombre de fois et les refusants étaient rappelés une ou deux fois (Legleye et al., 2013). Malgré ces efforts, le taux de réponse s'élève à 44,8%, ce qui nous a amenés à considérer l'éventualité de corriger la non-réponse totale.

2. Méthodologie

Apurement et synthétisation des données

Lors de cette enquête téléphonique, un certain nombre de données ont été enregistrées à chaque appel, comme la date, l'heure, et le résultat de l'appel (codé automatiquement ou par l'enquêteur). Ces données servent, en premier lieu, à la gestion du terrain mais peuvent également être utilisées pour la correction de la non-réponse. Un travail d'apurement et de synthèse de l'information est néanmoins nécessaire. La base des historiques d'appels de l'enquête FECOND compte 1 100 000 appels, passés vers 92 516 numéros différents. Il faut synthétiser l'information, de sorte que la base soit au niveau « numéro » et non appel. Lors du travail exploratoire, nous avons vérifié la cohérence des différentes variables créées ainsi que la cohérence de la chronologie des appels. Les historiques d'appels comptant des issues « Hors cible » ont été exclus de l'analyse, cette issue n'étant présente (du moins en théorie) que chez les non-répondants. Puis, nous avons réduit le nombre de modalités d'issues d'appels. A l'issue de ce travail, chaque appel est codé selon une variable en 13 modalités (réductible à 4 modalités).

Sélection des variables auxiliaires

Quatre jeux de données ont ainsi été constitués, du plus simple au plus complet :

- Jeu n°1 (2 variables) : nombre d'appels et nombre de refus
- Jeu n°2 (5 variables) : nombre d'appels, nombre de refus, de non contacts, de rendez-vous et nombre d'interviews (partielles ou complètes)
- Jeu n°3 (13 variables) : nombre d'appels, nombre d'appels indiquant que la ligne est en dérangement, nombre d'appels codés « occupé », nombre d'appels sans réponse, nombre de répondants, nombre de refus au contact, nombre de refus de niveau ménage, nombre de refus de niveau individu, nombre de rendez-vous avant sélection, nombre de rendez-vous après sélection par la personne contact, nombre de rendez-vous après sélection par la personne sélectionnée, nombre d'abandons en cours de passation de questionnaire, nombre de suspension de questionnaire, nombre d'interviews complétées
- Jeu n°4 (20 variables) : par bloc d'appels (1-3, 4-6, 7-9, 10-19, 20-50), pourcentages de refus, de non contacts, de rendez-vous et nombre d'interviews (partielles ou complètes).

Le jeu n°1 comprend les variables les plus simples et les moins sujettes aux erreurs de mesures : il est donc le plus robuste. Les travaux déjà publiés sur le sujet montrent son utilité. Le jeu n°2 est une agrégation simple des issues d'appels détaillées possibles. Le jeu n°3 est la version détaillée des issues des appels. Le jeu n°4 intègre spécifiquement la chronologie des événements mesurés dans le jeu n°2 : c'est donc le plus complet.

Afin de choisir l'information la plus judicieuse, nous avons réalisé une analyse factorielle multiple

sur les quatre nomenclatures possibles pour choisir celle à retenir pour la correction de la non-réponse totale. L'objectif était de voir quelles dimensions étaient en commun dans les différents jeux de nomenclatures testés.

Les résultats montrent que les jeux sont très corrélés entre eux (au sens des coefficients de liaison R_v et L_g), et que le jeu n°2 est le plus corrélé à l'analyse d'ensemble ($R_v=0.94$) ainsi qu'à chacun des autres groupes et en particulier le jeu n°4 ($R_v=0.87$).

Méthodes de formation des groupes homogènes de réponse

A partir d'un même jeu de variables auxiliaires, différentes méthodes peuvent être utilisées pour constituer des groupes homogènes de réponse : la modélisation de la participation ou la classification de la population des numéros.

Première méthode : la modélisation logistique

La première méthode est également appelée la méthode des scores. Elle repose sur la modélisation de la participation à l'enquête à partir de variables auxiliaires (ici, c'est le jeu de parodonnées n°2 qui a été utilisé). On estime, pour chaque individu présent dans l'échantillon, sa probabilité de réponse par régression logistique où les variables explicatives sont celles du jeu 2 (aucune interaction entre les variables du jeu 2 n'a été considérée). On découpe ensuite l'échantillon (répondants et non-répondants) en quantiles et on calcule, dans chaque groupe, le taux de réponse réel. Les répondants verront leur poids multiplié par l'inverse du taux de réponse de leur groupe. Quatre groupes homogènes ont été constitués, sur la base des quartiles de probabilité prédite de participer à l'enquête.

Deuxième méthode : classification par ACP

La méthode consiste à réaliser une analyse des composantes principales sur les parodonnées (ACP), puis une classification sur les premières composantes. C'est encore le jeu de parodonnées n°2 qui a été utilisé. Cette méthode est mise en œuvre sur les non-répondants : les répondants sont alors affectés aux groupes de non-répondants les plus proches. L'intérêt de cette méthode relativement à la modélisation est de tenir compte des corrélations entre les variables, ce qui enrichit potentiellement l'analyse et le pouvoir prédictif. Nous avons toutefois montré qu'il était possible d'opérer directement sur la base des répondants et non-répondants : le coefficient Kappa mesurant l'accord inter-classifications était de 0.96 pour classer répondants et non répondants, de 0.79 pour classer les répondants et de 0.97 pour classer les non répondants.

Troisième méthode : classification par analyse harmonique qualitative (AHQ)

Pour le jeu de parodonnées n°4, nous avons réalisé une analyse harmonique qualitative afin de tenir compte de la chronologie des événements, puis une classification ascendante hiérarchique sur les résultats. Là encore, cette méthode permet de tenir compte des corrélations temporelles des variables. Pour ces deux méthodes de classification, quatre classes ont été produites.

3. Comparaison des méthodes

La comparaison des méthodes s'est faite suivant trois critères mesurés chez les répondants : l'hétérogénéité inter-classes de la structure sociodémographique (mesurée par la distance du K_{hi}^2), des prévalences des comportements d'intérêt de l'enquête (idem). Par ailleurs, nous avons également testé l'existence d'écarts de prévalence résiduels après contrôle du profil sociodémographique (mesurés dans des régressions logistiques ajustées sur les variables

sociodémographiques). Enfin, nous avons mesuré la dispersion des poids finaux et la réduction du biais sociodémographique après correction de la non-réponse (par la distance du Khi^2 à la structure de la population cible).

Les classes constituées à partir de l'analyse harmonique qualitative offrent les résultats les plus satisfaisants en termes d'hétérogénéité interclasses sur les variables sociodémographiques comme sur les variables d'intérêt. Pour la structure sociodémographique (décrite à travers 8 variables), l'AHQ obtient les Khi^2 les plus élevés dans 5 cas sur 8, l'ACP dans 3 cas sur 8. Pour les 9 variables d'intérêt, l'AHQ obtient les Khi^2 les plus élevés dans 5 cas sur 9, l'ACP dans 4 cas.

Dans les modélisations logistiques des 9 variables d'intérêt, ajustées sur les variables sociodémographiques et les classes (modélisation logistique, ACP et AHQ), on observe des écarts de prévalence significatifs (au seuil de 5%) entre classes pour 2 variables sur 9 pour la modélisation logistique, 4 pour l'ACP et 5 pour l'AHQ.

Pour étudier la qualité de la correction de la non-réponse, on peut comparer les biais sociodémographiques résiduels. En corrigeant la non-réponse totale, on espère améliorer la structure sociodémographique de l'échantillon de répondants. On cherche donc un jeu de poids qui réduise la distance du Khi^2 entre l'échantillon et la population cible. Parmi les 8 variables sociodémographiques étudiées, l'ACP et l'AHQ permettent d'améliorer la distribution de 5 d'entre elles (âge en 7 tranches, situation professionnelle –actif occupé ou non-, mode de vie –en couple, chez les parents, autre-, taille du ménage –seul, deux, trois-quatre, cinq et plus-, zone de résidence - bassin parisien, Nord, Sud- contre seulement une pour la modélisation logistique (la situation professionnelle, mais très marginalement).

Une comparaison plus fine entre les méthodes ACP et AHQ montre que l'ACP est meilleure pour le sexe, l'âge, le mode de vie, la zone de résidence et le lieu de naissance -France, étranger-.

Pour les poids, on observe que le rapport poids max/poids minimum est plus faible pour l'AHQ (55.37 contre 8.59 pour la pondération initiale avec le poids de sondage), devant l'ACP (78.81) et la modélisation logistique (109.84). Les coefficients de variations sont ordonnés de façon différente : la modélisation arrive en tête (61.27 vs 49.08 pour la pondération initiale) devant l'AHQ (93.63) et l'ACP (137.72).

4. Conclusion et limites

A travers cette étude, nous avons montré :

- qu'il était possible de retenir une typologie simple des issues d'appels téléphonique pour travailler sur les parodonnées ;
- que la prise en compte de la chronologie des événements dans les issues d'appels téléphonique était aussi possible sur la base de cette typologie simple ;
- que la correction de la non-réponse par classification (tenant compte de la chronologie ou non) semble apporter de meilleurs résultats que la modélisation logistique, très souvent utilisée malgré tout.

Cependant, ces résultats peuvent être nuancés. Pour chaque méthode, nous avons découpés la population des numéros en seulement quatre groupes homogènes de réponse. Pour l'ACP et l'AHQ, l'analyse étant conduite sur les non-répondants uniquement, augmenter le nombre de classes aurait conduit à des profils de non-répondants très spécifiques, ne pouvant trouver leur «réplique» parmi les répondants. Une alternative à la classification ascendante hiérarchique pourrait être les K-means, dans lesquels il est possible de spécifier un nombre de classes avec un effectif minimal, ce qui pourrait permettre de limiter les risques de se confronter à ce problème.

Nous aurions pu tester les trois méthodes sur le jeu de données n°3. Toutefois, celle-ci est plus complexe à mettre en œuvre et nous avons préféré nous limiter à une typologie plus simple et plus facilement applicable dans d'autres enquêtes.

Il semble par ailleurs inévitable de procéder à un calage sur marges de l'échantillon redressé pour la correction de la non réponse totale : cela n'a pas été présenté ici. De même, aucune analyse de variance ni de biais des estimateurs produits par les différentes méthodes de redressement n'est proposée ici. Cela fera l'objet d'une analyse ultérieure par simulation, qui permettra de comparer l'efficacité des méthodes entre elles (dont le calage direct).

Enfin, la qualité des analyses dépend de la qualité des données d'appels. S'il existe relativement peu de doutes sur les typologies retenues pour les issues d'appels, l'éligibilité de tous les numéros non joints ou des refusants n'est pas toujours connue. Il est très difficile d'obtenir de l'information à ce sujet, mais il est à noter que 28% des numéros n'ont pu être contactés à l'issue des 20 tentatives d'appels. Nous n'avons pas ôté ces numéros de l'analyse, bien qu'il semble qu'une partie non négligeable pourrait l'être, car les numéros correspondent à des dispositifs domotiques ou industriels, ou ne sont pas aptes à recevoir ou passer des appels (numéros de box Internet, par exemple).

Bibliographie

[1] Legleye, Stéphane, Géraldine Charrance, Nicolas Razafindratsima, Aline Bohet, Nathalie Bajos and Caroline Moreau. 2013. "Improving Survey Participation: Cost Effectiveness of Call-Backs to Refusals and Increased Call Attempts in a National Telephone Survey in France." *Public Opinion Quarterly*, vol. 77, Issue 3, pp. 666-695

[2] Legleye, Stéphane, Géraldine Charrance, Nicolas Razafindratsima et l'équipe INSERM FECOND, 2012. L'effet du rappel des refusants et du rang d'appels sur la représentativité d'une enquête téléphonique. L'expérience de l'enquête FECOND. Acte de colloque, Colloque francophone sur les sondages, novembre 2012

[3] Maitland A., Casas Cordero C. et Kreuter F.. 2008. An exploration into the use of paradata for nonresponse adjustment in a health survey, *JSM 2008, Section on survey Research Methods*

[4] Olson Kristen, 2013. Paradata for nonresponse adjustment, Sociology department, Faculty publications. Paper 203.

[5] Kreuter F. et Olson K., 2011. Multiple auxiliary variables in nonresponse adjustment. *Sociological methods & Research*, 40 (2), pp. 311-332

[6] Beaumont J.-F., 2005. L'utilisation de renseignements sur le processus de collecte des données pour traiter la non-réponse total au moyen de l'ajustement de poids, *Statistique Canada*, volume 31, n°2, pp. 249-254

[7] Little, Roderick J. A., and Sonya Vartivarian. 2005. Does weighting for nonresponse increase the variance of survey means? *Survey Methodology* 31 (2): 161–68

[8] Sarndal, Carl-Erik, and Sixten Lundstrom. 2005. *Estimation in surveys with nonresponse*. Chichester, UK: John Wiley

[9] Deville J.C., 1974, « Méthodes statistiques et numériques de l'analyse harmonique », Annales de l'INSEE 15, 3-101

[10] Deville J.C., Saporta G., 1983, « Correspondence analysis, with an extension towards nominal time-series », Journal of Econometrics 22, 169-189

[11] Haziza, D. (2006). Traitement de la non-réponse dans les enquêtes, Ensai, support de cours FCDA.

[12] Caron, N. (2005). La correction de la non-réponse par repondération et par imputation, Insee, Document de travail, n° M0502.