
AGRÉGER LES ÉCHANTILLONS D'UNE ENQUÊTE MULTIMODE EN LIMITANT L'EFFET DE MESURE : UNE PROPOSITION D'IMPUTATION RAISONNABLE ET PRAGMATIQUE

Stéphane LEGLEYE (*), Tiaray RAZAFINDRANOVA (**), Gaël DE PERETTI (*)

(* Insee, Division Recueil et traitement de l'information

(**) Insee, Division Emploi

stephane.legleye@insee.fr

Mots-clés : enquête multimode, agrégation, effet de mesure, effet de sélection, imputation, score de propension

Résumé

Les enquêtes par Internet sont moins onéreuses qu'avec les modes administrés classiques et sont également censées offrir des réponses plus sincères aux questions sensibles [1], bien qu'elles puissent détériorer la qualité de recueil des questionnaires longs (biais de mesure) et qu'elles aboutissent généralement à des taux de participation inférieurs et à des profils de répondants différents (biais de composition).

Nous présentons un prolongement de travaux publiés sur l'enquête de victimation Cadre de vie et sécurité (CVS), enquête en face à face, et sur sa réplique sur Internet, Vols, violence et sécurité (VVS) [2]. Les enquêtes mobilisées avaient été réalisées indépendamment en 2013 (taux de participation de 63 % et 32 %).

Nous nous plaçons dans l'hypothèse fictive d'une agrégation des deux échantillons, la référence étant, historiquement, CVS. Nous proposons d'imputer les observations VVS porteuses de l'effet de mesure, laissant le biais de composition intact.

Les 6 variables cibles sont les déclarations d'un des événements suivants au cours des 24 derniers mois : vol avec violence ; vol sans violence ; cambriolage ; vol de véhicule ; violence physique ; menaces.

Les variables sociodémographiques (SD) utilisées dans le redressement de CVS sont au nombre de huit : sexe, âge, diplôme, PCS, type de logement, statut d'occupation du logement, taille d'unité urbaine, type de ménage.

Les variables auxiliaires (X) sont au nombre de quatorze (binaires ou ordonnées à quatre ou cinq modalités) et sont relatives aux opinions sur la police, la justice, et certains problèmes de société (chômage, sentiment d'insécurité, etc.).

Le biais de composition de VVS est important, notamment pour les variables auxiliaires X. Il pourrait être corrigé par repondération avant agrégation, mais cela aboutit à des poids calés trop dispersés (coefficient de variation CV=504) comparativement à ceux d'un calage direct (CV=93), que nous retiendrons.

Au sein de l'échantillon agrégé calé, nous modélisons par régression logistique l'appartenance à VVS. Sont intégrées les variables SD plus leurs interactions bivariées. Le support commun classique est défini par la plage de variation du score de propension commune aux deux sous-échantillons : seules 73 observations en sont exclues dans CVS et 126 dans VVS. À partir du score, nous apparions les observations de VVS à celles de CVS, 1-1 sans remise, avec choix aléatoire dans une plage d'ex-aequo de 0,2 écart-type du logit du score, en imposant l'égalité des variables SD les plus liées aux variables cibles : sexe, âge, PCS, taille d'unité urbaine, type de ménage, type de logement et statut

d'occupation du logement. Au total, 7957 paires d'observations CVS-VVS sont constituées (représentant 55 % de CVS, 62 % de VVS). En l'absence de biais de mesure, il ne devrait pas y avoir de différences de victimations Y en leur sein, tandis que les observations de VVS restantes représentent le biais de composition d'Internet du point de vue des variables SD.

Au sein des paires, 2512 (32%) présentent au moins une différence de Y, 1788 une seule et 169 au moins trois. Seule une fraction sera imputée : elle est tirée aléatoirement de façon à ce que la proportion imputée de chaque victimation égale celle dans CVS : entre n=149 –vols avec violence- et n=381 –vols sans violence-.

Après imputations stochastiques multiples sur l'ensemble des paires (n=5), l'effet de mesure (conditionnellement aux variables considérées) y est éliminé. Les victimations de CVS-VVS représentent entre 77 % et 92 % des victimations initiales, ou, par rapport à CVS seul, une augmentation moyenne de 11 % (entre -0,4 % pour les violences physiques et +35 % pour les vols avec violence), contre un écart initial de +30 %. La baisse est modeste, mais l'effet causal de l'interrogation sur Internet (méthode de régression/subclassification sur le score de propension, contrôlant les variables SD-X [3, 4]), est notablement réduit par rapport à la situation initiale : +0,48 % (p=0,008) vs +1,38 % (p=0,000) pour les vols avec violence, +0,68 % (p=0,016) vs +2,97 % (p=0,000) pour les vols sans violence, -0,29% (p=0,204) vs +0,13% (p=0,587) pour les violences physiques, -0,59 % (p=0,063) vs +1,28% (p=0,000) pour les cambriolages, -0,37% (p=0,264) vs 0,47% (p=0,167) pour les vols de véhicules et -0,52% (p=0,119) vs +1,14% (p=0,001) pour les menaces. La variance d'imputation est négligeable (<0,5%).

La méthode corrige l'effet de mesure et non l'effet de sélection dû au mode, qui peut lui aussi être entaché d'effet de mesure. Elle s'applique à toute enquête multimode, dès lors qu'il existe des sous-échantillons chevauchants du point de vue des variables SD-X. Elle est souple, les critères d'appariement et d'imputation pouvant être adaptés à l'objectif visé. Elle conserve les poids et est compatible avec une phase de redressement du biais de composition dont elle peut compléter l'effet, ou bien limiter la dispersion des poids, en permettant un choix plus parcimonieux des variables. Une des difficultés dans le cas présent est que les variables auxiliaires sont susceptibles d'être affectées par un effet de mesure important.

Bibliographie

- [1] De Leeuw, E.D., *Choosing the method of data collection*, in *International handbook of survey methodology*, E.D. De Leeuw, J.J. Hox, and D.A. Dillman, Editors. 2008, Lawrence Earlbaum Associates: New York. p. 117-135.
- [2] Razafindranovona, T., *Exploitation de l'enquête expérimentale Qualité de vie au travail*, in *Document de travail*. 2017: Paris. p. 58.
- [3] Razafindranovona, T., *Exploitation de l'enquête expérimentale Vols, violence et sécurité*, in *Document de travail*. 2016, Insee: Paris. p. 48.
- [4] Razafindranovona, T., *Exploitation de l'enquête expérimentale Logement internet/papier*, in *Document de travail*. 2014, Insee: Paris. p. 54.
- [5] Imbens, G.W. and D.B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. 2015, Cambridge, MA: Cambridge University Press
- [6] Rosenbaum, P.R. and D.B. Rubin, *Reducing Bias in Observational Studies Using Subclassification on the Propensity Score*. *Journal of the American Statistical Association*, 1984. **79**(387): p. 516-524.