
**CORRECTION DES EFFETS DE MODE ET BIAIS DE SÉLECTION :
LES APPORTS D'UNE EXPÉRIMENTATION DE L'ENQUÊTE TIC (TECHNOLOGIES
DE L'INFORMATION ET DE LA COMMUNICATION) EN FACE-À-FACE**

Stéphane LEGLEYE^{1,2}, Louise VIARD-GUILLOT¹ & Amandine NOUGARET¹

1 Insee, Département des ressources et conditions de vie des ménages

2 : CESP, Inserm

stephane.legleye@insee.fr
louise.viard-guillot@insee.fr
amandine.nougaret@insee.fr

Mots-clés : Échantillonnage, collecte, multimode, effet de mode

Domaine concerné : Collecte, théorie des sondages aval

Résumé

L'enquête TIC auprès des ménages est la plus ancienne enquête ménages multimode mise en place à l'Insee. Elle interroge chaque année environ 30 000 individus sur leur équipement et leurs usages en matière d'Internet, d'informatique et de téléphonie fixe et mobile. Son protocole repose sur une collecte par téléphone et par Internet/papier.

Au quatrième trimestre 2021 aura lieu un **test de l'enquête TIC en face-à-face**, auprès d'un échantillon de 5 000 individus. L'objectif de cette expérimentation est de tester la présence de biais dans l'enquête TIC classique.

Les biais suspectés dans l'enquête TIC

Le questionnaire de l'enquête TIC ne comportant aucune question véritablement sensible (on ne trouve aucune interrogation sur des comportements déviants, des pratiques intimes, des croyances ou des actes illégaux), on suspecte *a priori* très peu de biais de mesure. En revanche, Internet étant à la fois un des thèmes et un des modes de collecte de l'enquête classique, on peut par conséquent suspecter un **biais de sélection**. Certains individus sélectionnés pour répondre par Internet pourraient avoir tendance à ignorer l'enquête lorsqu'ils sont peu à l'aise avec cet outil (bien qu'une solution de réponse par papier leur soit proposée) ou lorsque le sujet de l'enquête ne les mobilise pas. En effet, la précédente édition avait permis de montrer qu'environ 15 % de la population de 15 ans et plus n'utilisait pas ou très peu Internet et que 38 % des usagers dans l'année avait un manque de compétence dans le domaine, 17 % de la population pouvant ainsi être considérée en état d'illectronisme [1].

Or, si 45 % du sous-échantillon interrogé par Internet/papier dispose d'une adresse courriel, signe d'une potentielle utilisation d'Internet, ce n'est pas le cas des 55 % restants, soit environ 13 % de la population générale, *a priori* très peu utilisatrice d'Internet. Même si la possibilité de répondre par papier est mentionnée dès la première lettre-avis, remplir le questionnaire papier exige davantage d'efforts (les filtres doivent être appliqués par le répondant, les instructions doivent être lues avec attention, le questionnaire doit être posté...) : cela concourt à diminuer le taux de réponse. De plus, les ménages du sous-échantillon Internet/papier sont plus pauvres que les autres : 40 % des ménages de ce sous-échantillon gagnent moins de 20 000 euros par an, contre respectivement 25 % et 31 % pour les sous-échantillons téléphonique et méthodologique (individus disposant d'un numéro de téléphone mais interrogés par Internet/papier). Leur capacité à se mobiliser pour l'enquête pourrait être réduite. Et de fait, les taux de collecte sont très bas dans ce sous-échantillon : 39,3 % en 2021, dont 46 % pour les logements avec une adresse courriel et 34 % pour ceux qui en sont dépourvus.

Enfin, en plus du biais de non-réponse évoqué ci-dessus, la passation via le papier ou Internet ne permet pas de s'assurer de l'identité du répondant. Or l'enquête TIC repose sur le tirage d'un kish – individus de 15 ans ou plus du ménage dont la date de naissance est la plus proche du 1^{er} juin. On suspecte qu'au sein des ménages participant en auto-administré, la personne la plus intéressée par le sujet ou la plus à l'aise sur Internet, répondrait en lieu et place de la personne Kish, ce qui constitue un second biais de sélection. Ce dernier fait l'objet d'un traitement en standard à l'aval. Toutefois, il s'appuie sur une comparaison des sous-échantillons méthodologique et téléphonique, donc sur les sous-échantillons de ménages ayant un numéro de téléphone disponible et dont les caractéristiques sociodémographiques sont dissemblables de celles des individus de l'échantillon Internet sans numéro de téléphone.

Ce type de biais est proche d'un biais de non-réponse (ou de sélection) non-ignorable dans la mesure où ce sont bien les caractéristiques que l'on cherche à estimer qui causent une large partie de la non-réponse. Il est ignorable, donc corrigible par un redressement uniquement si l'on peut trouver des variables de la base de sondage (renseignées pour les répondants et non-répondants) suffisamment corrélées à la fois à la non-réponse totale et aux caractéristiques à estimer. La question est donc de savoir si la procédure de redressement actuelle de l'enquête assure bien ce travail. Au-delà des considérations statistiques propres aux techniques classiques de redressement [2,3], disposer d'un échantillon de contrôle permet d'apporter une réponse directe à la question de la fiabilité des estimations.

L'apport de la collecte en face-à-face

La collecte face-à-face permettrait d'obtenir un fort taux de collecte, en particulier pour les ménages sans téléphone, et ainsi de disposer d'un échantillon de contrôle monomode pour tous les sous-échantillons qui n'interfère pas avec la thématique de l'enquête et serait donc *a priori* moins biaisé que l'échantillon classique. Cela permettra d'évaluer la robustesse des niveaux des principaux indicateurs d'intérêt via un mode totalement indépendant de la thématique de l'enquête et de quantifier les biais de sélection éventuels dans l'enquête classique.

Les différents échantillons de répondants à l'enquête classique seront comparés à chaque étape du redressement avec l'échantillon de l'enquête face-à-face, afin de mettre en évidence le mécanisme global de réponse à l'enquête. Puis la comparaison des indicateurs d'intérêt obtenus dans l'enquête classique avec ceux obtenus en face-à-face (après les redressements des deux enquêtes) renseignera sur la qualité de l'enquête classique et de son redressement.

Bibliographie

- [1] Legleye S, Rolland A, Une personne sur six n'utilise pas Internet, plus d'un usager sur trois manque de compétences numériques de base, Insee Première, n°1780, 2019

- [2] Little R. J. A., Vartivarian S. Does weighting for nonresponse increase the variance of survey means?, Survey methodology 2005: 31: 161–168.

- [3] Haziza D., Lesage E. A discussion of weighting procedures for unit nonresponse, Journal of Official Statistics 2016: 32: 129-145.