

Méthodologie

Odilon Dakpakete

2025-02-12

Les variables sur lesquelles nous travaillons sont décrites comme suit :

- X0 une matrice des variables auxiliaire connue pour les répondants et non-répondants ;
- X1 une matrice des variables sociodémographique connue uniquement pour les répondants ;
- Y la variable d'intérêt, connue uniquement pour les répondants : Y est expliquée par X0 et X1, * mais aussi U
- U une variable non mesurée affectant la probabilité de réponse à Internet, au téléphone et Y, calculée pour les répondants et les non-répondants
- Pi la probabilité de répondre sur Internet
- Pt la probabilité de répondre au téléphone

Soit Y une variable numérique On s'intéresse à la moyenne de Y définit par la formule suivante :

$$\mu = \frac{1}{N} \sum_{i=1}^N Y_i$$

Cette moyenne n'étant pas calculable, il nous faut donc l'estimer par le biais d'un sondage Posons s_i la variable aléatoire prenons la valeur 1 si l'individu i est sélectionné dans le plan de sondage et 0 sinon. Le plan de sondage est défini par le vecteur $S = (s_1, s_2, \dots, s_N)$.

L'estimateur d'Horvitz-Thompson de la moyenne de Y est donné par la formule suivante :

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\pi_i} s_i$$

Ici, nous faisons l'hypothèse que la probabilité d'inclusion $\pi_i = \pi$ est la même pour tous les individus i. C'est à dire que $\pi = \frac{n}{N}$.

Dans notre, il s'agit d'une enquête multimode séquentiel (internet, puis téléphone). Il nous faut donc définir deux variables instrumentale qui identifient les réponses par internet et les réponses par téléphone.

Soit z_i la variable aléatoire prenant la valeur 1 si l'individu i répond par internet et 0 sinon.

Soit w_i la variable aléatoire prenant la valeur 1 si l'individu i répond par téléphone et 0 sinon. Sur les n réponses, on compte t réponses par internet et n-t réponses par téléphone.

Ainsi l'estimateur de la moyenne de Y par internet est donné par la formule suivante :

$$\hat{\mu}_1 = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\pi_i} s_i (z_i + w_i)$$

Cette d'estimateur est biaisé, car d'après le document de travail (CASTELL, SILLARD, 2021), $\mathbb{E}[s_i(z_i + w_i)] \neq \pi$ On pose $r_i = (z_i + w_i)$

Ainsi, on cherche à estimer :

$$\hat{\mu}_2 = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\pi_i \hat{\rho}} s_i r_i$$

Où $\hat{\rho}$ sera à estimer pour que $\hat{\mu}_2$ soit sans biais.