

Pondération de l'échantillon additionnel pour le canton Vaud Rapport technique

Erika Antal *

May 11, 2015

1 Introduction

Les pondérations transversales et longitudinales pour l'échantillon additionnel du canton Vaud ont été effectuées essentiellement de la même manière que ce que fait FORS pour le PSM. Les concepts relatifs à la pondération du PSM et les méthodes utilisées sont présentés dans le document y concernant. Les méthodes sont adaptées pour l'échantillon du canton de Vaud, seul le calcul des poids de sondage diffère légèrement, étant donné que l'échantillonnage, donc les plans de sondage des deux études sont différents. Le but de ce document est de présenter les détails de cette étape du calcul. Concernant les autres méthodes utilisées, voir la documentation du calcul des poids du PSM (<http://forscenter.ch/en/our-surveys/swiss-household-panel/documentationfaq-2/methods/weighting/>)

2 Poids de sondage et poids initiaux

Tout le système du calcul des poids se base sur les poids de sondage. Ce sont ces poids qui seront plus tard ajustés pour la non-réponse des différents niveaux - la grille, le questionnaire de ménages et le questionnaire individuel/ biographique - partagés aussi entre cohabitants du même ménage et finalement calés à des totaux connues de la population de référence. Donc, la première étape du processus est de déterminer les poids de sondage. Les poids de sondage sont les inverses des probabilités d'inclusion qui sont eux, entièrement déterminés par le plan de sondage. Le plan de sondage de l'échantillon vaudois est un plan simple des individus résidant dans le canton Vaud, stratifié par revenus et âges. Un échantillon de 800 personnes est ciblé, réparti entre les différents tranches/strates comme le tableau ci-dessous le montre.

*Swiss Centre of Expertise in the Social Sciences C/O Université de Lausanne - Bâtiment Géopolis - CH-1015 Lausanne - Switzerland emails: erika.antal@fors.unil.ch

Tranches		# de personnes
Tranche 1	11	30
	12	250
Tranche 2		280
Tranche 3		190
Tranche 4		50
Echantillon		800

La population de référence de l'échantillon vaudois est l'ensemble des personnes de plus que 14 ans, résidant dans le canton de Vaud. Afin de pouvoir estimer les tailles de ces différentes strates dans la population de référence, un premier échantillon de 15'000 adresses a été tiré. Pour différentes raisons, de ces 15'000 adresses, 507 étaient inutilisables. Le tableau 2 montre la répartition de cet échantillon.

Tranches		# d'adresse
Adresses inutilisable		507
Tranche 1	11	608
	12	1'139
Tranche 2		1'352
Tranche 3		9'622
Tranche 4		1'772
Echantillon		14'493

Comme le plan de sondage de ce premier échantillon était un plan simple, on peut supposer que la répartition de ces strates est la même dans la population que dans l'échantillon. Ainsi une première probabilité d'inclusion a été calculée. La taille de la population de référence est de 618'395 personnes, dont on estime 604'628 adresses utilisables. Avec le premier échantillon net de 14'493 la probabilité d'inclusion de la première phase est de 2.4%.

La deuxième phase de l'échantillonnage était de tirer un échantillon de 3'000 adresses à l'intérieur du premier. Le plan de sondage de cette deuxième phase est un plan simple stratifié. Afin d'assurer le nombre suffisant de personnes appartenant de strate cible de l'étude, différents taux de sondage par strate ont été appliqués. Les 3'000 adresses étaient réparties en deux groupes, 2'000 pour injecter au démarrage de la prise d'information et 1'000 en réserve. Finalement, 200 adresses ont été utilisées de la réserve, donc l'échantillon final est de 2'200 adresses. (π_{k2} représentent les probabilités d'inclusion de la deuxième phase, π_k celles de tout l'échantillonnage et w_k sont les poids de sondage.)

Tranches		# de personnes	π_{k2}	π_k	w_k
Tranche 1	11	92	15.132%	0.363%	275.706
	12	805	70.676%	1.694%	59.028
Tranche 2		717	53.033%	1.271%	78.666
Tranche 3		457	4.750%	0.114%	878.373
Tranche 4		129	7.280%	0.175%	573.065
Echantillon		2'200	15.180%	0.364%	274.831

Ce sont les poids de sondage de base, appliqués à toute les adresses tirées, indépendamment du fait si le ménage a accepté de participer à l'enquête ou non. Les poids des répondants (les participants de l'enquête et les répondants au questionnaire de non-réponse ou au questionnaire de refus) seront ensuite ajustés à la non-réponse totale.

A cette étape une probabilité d'inclusion générale était nécessaire à définir. La stratification de la population était inévitablement fait au niveau individu, donc une adresses/un ménage pouvait être sélectionné/e pour l'échantillon autant de fois que le nombre de personnes éligible du ménage. Cela peut mener d'un part à des duplications des adresses dans l'échantillon. Ce problème n'est pas très dérangent pour l'échantillon vaudois, vu que les taux de sondage sont relativement faibles dans chaque strate. D'autre part, le calcul de la probabilité d'inclusion d'un individu et également d'un ménage est affecté. Un membre éligible de ménage sera demandé à participer à l'enquête si au moins un des membres éligible du ménage est sélectionné pour l'échantillon. Mais à l'avance on ne connaît que l'appartenance de tranches concernant la personne sélectionnée (personne de référence). Celui des autres membres du ménage est inconnu, donc les probabilités d'inclusion de ces personnes ne sont pas calculables, en plus, celle de la personne de référence est également significativement sous-estimée.

Dans cette enquête la procédure appliquée pour surmonter ce problème est de recalculer les probabilités en utilisant une estimation d'une probabilité générale des personnes non-sélectionnées vivant dans un ménage d'un individu sélectionné pour l'échantillon. Comme l'appartenance des autres membres éligibles que la personne de référence du ménage n'est pas connu, il est estimé comme le nombre de ces personnes dans l'échantillon divisé par la taille de la population. Concernant l'enquête Vaud, il y avait 3'779 membres éligibles autres que la personne de référence d'un ménage participant à l'enquête, donc cette probabilité d'inclusion générale est de 0.625%. Les probabilités d'inclusion de base ont donc été corrigées (augmentées) par cette probabilité, en tenant également compte le nombre de personnes éligible du ménage, c'est à dire:

$$\pi_k^{corr} = \pi_k^{init} + (d - 1) \times \pi_k^{gen}$$

où π_k^{init} est la probabilité d'inclusion initiale de la personne de référence, d est le nombre de membres éligibles dans le ménage (y compris la personne de référence), π_k^{gen} est la probabilité générale et π_k^{corr} est la probabilité corrigée.

Cette probabilité d'inclusion était affectée à tous les membres éligibles du ménage. Les poids de sondage corrigés sont les inverses de ces probabilités

d'inclusion, puis les poids initiaux étaient calculés en ajustant les poids de sondage corrigés des participants (participants à l'enquête) pour la non-réponse, en utilisant le questionnaire de non-réponse ou le questionnaire de refus. Cet ajustement était également fait par la méthode "CHAID", comme tous les ajustements pour la non-réponse dans le calcul des poids du PSM.

3 Des étapes subséquentes

Une fois les poids initiaux calculés les différentes étapes du calcul, i.e. les ajustements pour la non-réponse au questionnaire individuel, niveau d'individu, ou au questionnaire de ménage au niveau ménage, le partage des poids (pas applicable la première année) et le calage sur marge était fait dans la même manière que pour le PSM, adapté évidemment à l'enquête Vaud. (<http://forscenter.ch/en/our-surveys/swiss-household-panel/documentationfaq-2/methods/weighting/>)