

# CALAGE SUR COMPOSANTES PRINCIPALES APPLIQUÉ À LA MESURE D'AUDIENCE INTERNET

*Camelia GOGA*<sup>1</sup>, *Aurélie VANHEUVERZWYN*<sup>2</sup>,  
*Eléna VOUGE*<sup>2</sup>

Lancée en 2010 par Médiamétrie, la mesure de l'audience de l'Internet Mobile vise à quantifier et qualifier l'audience des sites et applications disponibles à partir des réseaux de télécommunications français.

La dimension quantitative est assurée par une mesure automatique et exhaustive à la manière d'un recensement. Les trois principaux opérateurs télécom français envoient chaque jour l'ensemble des logs de connexion de leurs abonnés. Chaque ligne de log représente l'affichage d'une page mobile ou la connexion à une application mobile. Après traitement et cumul sur un trimestre, ces logs fournissent le nombre de visiteurs, de pages vues et le temps passé sur les sites et applications mobiles.

Cette mesure exhaustive est complétée par une mesure par panel dans le but de déterminer le profil socio-démographique et d'équipement des visiteurs des sites mobiles. Ce panel est composé de 10 000 individus de 11 ans et plus recrutés par téléphone et via Internet. Les connexions des panélistes sont directement extraites des données exhaustives via un système d'anonymisation des logs. Les mêmes indicateurs peuvent donc être calculés dans les deux sources.

Comme tout échantillon, le panel présente des déformations par rapport à la structure théorique de la population des mobinautes qui est apportée par une enquête de cadrage réalisée par téléphone. Un premier redressement est effectué sur le panel pour assurer a posteriori sa représentativité sur des critères sociodémographiques (sexe, âge, catégorie socioprofessionnelle, région d'habitation,...) ainsi que des critères d'équipement en téléphonie mobile (opérateur téléphonique, type d'abonnement mobile, type de téléphone,...). Par ailleurs, le recrutement étant établi au travers d'une population déclarant se connecter à l'Internet Mobile, il existe des différences comportementales assez importantes entre les usages relevés dans la mesure exhaustive et ceux observés sur le panel. Il convient donc de caler le panel sur des critères d'usage.

D'un point de vue théorique, disposer d'information auxiliaire est a priori un atout, cette information auxiliaire pouvant être utilisée au niveau de l'échantillonnage ou au niveau du calcul des estimateurs comme démontré par Ardilly (1994). Dans le cas présent, les données exhaustives nous fournissent une telle quantité d'information que la méthode d'estimation par calage classique peut s'avérer inefficace.

L'objectif des travaux présentés dans ce papier est d'étudier une extension de la méthode d'estimation par calage pour prendre en compte une information auxiliaire plus riche et ainsi améliorer la précision des estimations. Il s'agit d'une nouvelle classe d'estimateurs basés sur la régression en composantes principales qui permettra de réduire la dimension de l'information auxiliaire tout en gardant le maximum d'information.

<sup>1</sup> *IMB, Université de Bourgogne, 9 avenue Alain Savary, Dijon*

<sup>2</sup> *Médiamétrie, 70 rue Rivay, 92532 Levallois-Perret Cedex*

*email : camelia.goga@u-bourgogne.fr, avanheuverzwyn@mediametrie.fr, evouge@mediametrie.fr*

## Bibliographie

Ardilly P. (1994), *Les techniques de sondage*, Editions Technip, Paris.

Deville J.-C. et Särndal C.-E. (1992), Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, Vol. 87, n° 418, pp. 376-382.

Deville J.-C. et Särndal C.-E. et Sautory O. (1993), Generalized Raking Procedures in Survey Sampling, *Journal of the American Statistical Association*, Vol. 88, n° 423, pp. 1013-1020.

Sautory O. (1993), La macro CALMAR Redressement d'un échantillon par calage sur marges, *Série des documents de travail de la Direction des Statistiques Démographiques et Sociales*, F 9310.

Silva P.L.N. et Skinner C. (1997), Variable selection for regression estimation in finite population, *Survey Methodology*, 23, 23-32.