

CALAGE SUR COMPOSANTES PRINCIPALES *versus* CALAGE PÉNALISÉ APPLIQUÉ À LA MESURE D'AUDIENCE HYBRIDE INTERNET

Camelia Goga ¹ & Aurélie Vanheuverzwyn ² & Lila Zydorczak ²

¹ *IMB, Université de Bourgogne, 9 avenue Alain Savary, Dijon*

camelia.goga@u-bourgogne.fr

² *Médiametrie, 70 rue Rivay, 92532 Levallois-Perret Cedex*

avanheuverzwyn@mediametrie.fr, lzydorczak@mediametrie.fr

L'internet en tant que média numérique offre deux principales possibilités de mesure. La première dite mesure site centric est fondée sur l'insertion de marqueurs ou tags et permet une analyse exhaustive du trafic et de la fréquentation d'un site. La seconde mesure est une mesure d'audience qui repose sur un panel et permet quant à elle de quantifier et de qualifier les visiteurs d'un site sur des critères socio-démographiques. Ces deux outils apportent ainsi une perspective différente et complémentaire sur l'analyse des usages internet que la mesure hybride souhaite réconcilier au sein d'une mesure tirant parti du meilleur de chacune. La méthodologie envisagée se traduit en pratique par l'injection de résultats issus de la mesure site centric au sein de la mesure d'audience issue du panel. Ainsi, le redressement du panel ne se restreint plus aux seuls critères socio-démographiques classiques (sexe, âge, CSP, etc.) et une dimension quantitative est ajoutée par la mesure site centric qui délivre, à la manière d'un recensement, le nombre de visites pour un certain nombre de sites. Cependant, la mesure site centric nous fournit une telle quantité d'information que la méthode d'estimation par calage classique peut s'avérer inefficace. Le calage pénalisé (Rao and Singh, 2009) et le calage sur composantes principales (Goga *et al.*, 2011) proposent tous deux de relâcher les contraintes de calage mais de façon différente. Dans le premier cas, on utilise une contrainte quadratique sur les poids de calage. Pour la distance de chi-deux, cela revient à faire une régression de type ridge. La deuxième méthode consiste à réaliser un calage classique sur un nombre réduit de nouvelles variables, les composantes principales de la matrice de variables de calage. De cette façon, on réduit la dimension tout en gardant le maximum d'information. L'objectif des travaux de cette présentation est d'appliquer et de comparer ces deux méthodes dans le cadre de la mesure hybride Internet afin de prendre en compte l'information auxiliaire plus riche et ainsi d'améliorer la précision des estimations.

Bibliographie

Goga, C., Shehzad, M.-A. and Vanheuverzwyn, A. (2011), Principal Component Regression with Survey Data. Application on the French Media Audience, Proceedings of the 58th ISI World Statistics Congress - Dublin 2011.

Rao, J.N.K. and Singh, A. C. (2009). Range Restricted Weight Calibration for Survey Data Using Ridge Regression. *Pakistan Journal of Statistics*, 25(4), 371-384.