

MODÉLISATION MULTI-NIVEAUX DE DONNÉES D'ENQUÊTE : IMPACT SUR LES ESTIMATIONS DÛ AUX POIDS DE SONDAGE DE NIVEAU 2 INTRODUITS DANS LA PSEUDO-VRAISEMBLANCE

Jean-Paul Lucas ¹

¹ UPE, CSTB, 84 avenue Jean Jaurès 77447 Marne-La-Vallée Cedex 2, France,

jean-paul.lucas@cstb.fr /

Université de Nantes - EA4275 Biostatistique, Recherche Clinique et Mesures Subjectives en
Santé, 1 rue Gaston Veil BP 53508, 44035 Nantes Cedex 1, France,

jean-paul.lucas@etu.univ-nantes.fr

La modélisation multi-niveaux sur données issues d'un plan de sondage à plusieurs degrés peut amener à des estimations biaisées lorsque les probabilités de sélection sont inégales (Pfeffermann *et al.*, 1998). Il est alors recommandé (Rabe-Hesketh et Skrondal, 2006) d'introduire des poids de niveau 1 (dernier degré) mis à une certaine échelle, dans la pseudo-vraisemblance. Nous proposons d'étudier l'impact que peuvent avoir, sur les estimations d'un modèle à 2 niveaux et à intercepte aléatoire, les poids de niveaux 2 introduits. Nos données étaient issues d'un plan de sondage à 4 degrés mais seules des informations des niveaux 1 et 2 entraient dans le modèle; les unités de niveau 1 avaient des probabilités de sélection conditionnelles toutes égales à 1. À partir de nos données, 5 types de poids de niveau 2 pouvaient être utilisés. Les résultats ont montré que certaines covariables avaient des coefficients estimés stables quelque soit le type de poids de niveau 2. A contrario d'autres covariables avaient des estimations substantiellement différentes voire même avec des signes opposés. De même la significativité des estimations pouvait ne pas être stable. Ces résultats montrent que les poids de niveau 2, souvent peu considérés, peuvent avoir un impact sur les estimations d'un modèle multi-niveaux. Ils permettent de mettre en garde les utilisateurs de données publiques où un seul type de poids de niveau 2 peut être fourni. La simulation d'une population permettra de déterminer quel type de poids de niveau 2 il alors est recommandé d'utiliser. Nos données étant issues d'un plan particulier et l'étude faite sur cas complets, il est alors nécessaire de poursuivre des travaux dans ce domaine afin d'étudier l'impact que pourraient avoir le nombre de degrés du plan, la stratification, et la présence de données manquantes.

Bibliographie

Pfeffermann, D., et Skinner, C.J., et Holmes, D.J., et Goldstein, H., et Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models, *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 60, 23-40.

Rabe-Hesketh, S., et Skrondal, A. (2006). Multilevel modelling of complex survey data, *Journal of the Royal Statistical Society : Series A (Statistics in Society)* 169, 805-827.