

No 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

**Symposium 2006 : Enjeux
méthodologiques reliés à la
mesure de la santé des
populations**



2006



Statistics
Canada

Statistique
Canada

Canada

Modèles de sélection des variables pour l'élagage des poids

Michael Elliott¹

Résumé

Dans le cas de l'échantillonnage avec probabilités de sélection inégales, les corrélations entre les probabilités de sélection et les données échantillonnées peuvent induire un biais. Pour le corriger, on fait souvent appel à des poids de sondage correspondant à l'inverse de la probabilité de sélection. Les plans d'échantillonnage fortement disproportionnels donnent lieu à des poids élevés, susceptibles d'introduire une variabilité superflue dans des statistiques telles que l'estimation de la moyenne de la population. L'élagage des poids consiste à réduire ceux dont la valeur est élevée à une valeur seuil fixée et à rajuster ceux dont la valeur est inférieure à ce seuil afin que leur somme demeure égale à celle des poids non élagués. Cet élagage réduit la variabilité, mais introduit un certain biais. Les méthodes habituelles ne sont pas « guidées par les données », c'est à dire qu'elles ne se fondent pas sur les données pour arriver au compromis biais variance approprié, ou bien elles le font de façon très inefficace. Cette communication décrit l'élaboration de méthodes bayésiennes de sélection des variables pour l'élagage des poids destinées à compléter les méthodes habituelles, ponctuelles, fondées sur le plan de sondage, lors de l'utilisation de plans à probabilités d'inclusion disproportionnelles où les variances dues aux poids de sondage excèdent la correction pour le biais. Ces méthodes sont utilisées pour estimer les paramètres de population provenant de modèles de régression linéaire et de régression linéaire généralisée dans le contexte de plans d'échantillonnage à probabilités de sélection connues stratifiés et stratifiés a posteriori. Des applications seront décrites dans le cas de données provenant d'enquêtes sur les blessures causées par des accidents de circulation, pour lesquelles des plans de sondage fortement disproportionnels sont souvent utilisés.

¹Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, États-Unis