

## Article

# Dans ce numéro



Décembre 2009

## Dans ce numéro

Ce numéro de *Techniques d'enquête* commence par le neuvième article de la série annuelle d'articles sollicités du prix Waksberg, dédiée à Joseph Waksberg en reconnaissance à sa contribution à la théorie et à la pratique des techniques d'enquête. Le comité de rédaction remercie les membres du comité de sélection, composé de Bob Groves (président), Leyla Mohadjer, Daniel Kasprzyk et Wayne Fuller, d'avoir choisi Graham Kalton comme auteur de l'article du prix Waksberg de cette année.

Dans son article intitulé « Méthodes de suréchantillonnage des sous-populations rares dans les enquêtes sociales », Kalton donne un aperçu des méthodes d'échantillonnage de populations rares, que Kish avait appelées « domaines mineurs ». Après avoir abordé des questions générales, il décrit un certain nombre de méthodes, dont le dépistage, la stratification, l'échantillonnage à deux phases, les bases de sondage multiples, l'échantillonnage basé sur la multiplicité, l'échantillonnage d'emplacements ou des lieux de sociabilité et l'accumulation d'échantillons au fil du temps. Il discute des avantages et des inconvénients de chacune et donne de nombreux exemples de leur utilisation dans des enquêtes. En pratique, on a souvent recours à une combinaison de diverses approches.

Des stratégies de réponse aléatoire sont souvent utilisées afin de réduire les erreurs non dues à l'échantillonnage, comme la non-réponse et les erreurs de mesure. De telles stratégies peuvent aussi être utilisées dans le contexte du contrôle de la divulgation statistique dans les fichiers de microdonnées à grande diffusion. Dans son article, Quatember propose une standardisation des méthodes de réponse aléatoire. Il établit les propriétés statistiques de l'estimateur standardisé et applique la méthode proposée à une enquête sur la tricherie chez les étudiants.

Xu et Lavallée examinent le problème causé par la non-réponse de lien dans l'échantillonnage indirect lorsqu'on utilise la méthode généralisée du partage des poids. L'échantillonnage indirect consiste à sélectionner des échantillons dans une population qui n'est pas la population cible, mais qui y est reliée. Des estimations biaisées peuvent être obtenues quand on ne sait pas qu'une unité de la population échantillonnée est liée à une unité de la population cible. Les auteurs proposent plusieurs ajustements de la pondération pour surmonter le problème de la non-réponse de lien.

Dans le contexte des cas de non-réponse, les pondérations des répondants sont souvent ajustées par l'inverse de la probabilité de réponse estimée. Da Silva et Opsomer proposent de recourir à la régression par polynômes locaux pour estimer les probabilités de réponse. Ils présentent les résultats d'une étude par simulation qui confirment l'efficacité de la méthode.

L'article de Van den Brakel et Krieg porte sur un modèle de séries chronologiques structurel multivarié qui tient compte du plan de sondage de l'enquête sur la population active des Pays-Bas. Ils utilisent le modèle pour estimer les taux de chômage. Une étude empirique démontre que le modèle proposé accroît considérablement la précision des estimations.

Zhang étudie la production d'estimations croisées où l'une des marges du tableau croisé correspond à des petits domaines et où la non-réponse varie d'un domaine à l'autre. Il élabore une approche de modélisation mixte double qui combine les effets fixes et les effets aléatoires de domaine du modèle d'estimation sur petits domaines avec les effets aléatoires du mécanisme de création des données manquantes. Il calcule une approximation de l'erreur quadratique moyenne conditionnelle de prédiction sous la forme d'une décomposition en trois parties qui correspondent à une variance de prédiction naïve, une correction positive qui tient compte de l'incertitude hypothétique d'estimation des paramètres fondée sur les données complètes latentes et une autre correction positive pour la variance supplémentaire due aux données manquantes.

Souza, Moura et Migon proposent une application bayésienne d'estimation sur petits domaines fondée sur des modèles de croissance qui tiennent compte des relations hiérarchiques et spatiales. Il se sert de cette approche pour obtenir des prédictions de population pour les municipalités non échantillonnées dans l'enquête annuelle sur les ménages du Brésil et pour accroître la précision des estimations basées sur le plan de sondage obtenues pour les municipalités échantillonnées.

Shao et Thompson étudient le problème de l'estimation de la variance lorsque l'on procède à un ajustement de la pondération pour corriger la non-réponse dans les enquêtes-entreprises stratifiées. Ils dérivent deux estimateurs de variance par linéarisation convergents sous des hypothèses faibles. Les estimateurs de variance jackknife naïfs ne donnent pas de bons résultats à moins que la fraction d'échantillonnage soit négligeable, ce qui n'est pas le cas en présence de strates à tirage complet. Ils proposent un estimateur de variance jackknife modifié et convergent même en présence de strates à tirage complet, mais la fraction d'échantillonnage des strates à tirage partiel ne doit pas être élevée. Ils évaluent leur estimateur de variance empiriquement au moyen de données réelles et par une étude en simulation.

Dans son article, Preston étudie l'estimation de variance bootstrap pour les plans à plusieurs degrés quand les unités sont sélectionnées par échantillonnage aléatoire simple sans remise à chaque degré. Il propose une extension de l'estimateur bootstrap rééchantillonné souvent utilisé, qui repose sur l'hypothèse d'un échantillonnage avec remise ou de fractions d'échantillonnage négligeables au premier degré. Il compare l'estimateur proposé aux estimateurs bootstrap rééchantillonné et bernoullien.

Jang et Eltinge s'attaquent au problème de l'estimation du nombre de degrés de liberté sous des plans de sondage multivariés à plusieurs degrés quand un petit nombre d'unités primaires d'échantillonnage (UPE) sont tirées par strate. Étant donné le petit nombre d'UPE sélectionnées, l'estimation classique du nombre de degrés de liberté selon la méthode de Satterthwaite peut donner lieu à une sous-estimation grave. Les auteurs proposent un estimateur de recharge du nombre de degrés de liberté qui utilise les variances intra-UPE pour fournir de l'information auxiliaire sur la grandeur relative de la variance globale au niveau de la strate. Ils illustrent leur méthode au moyen de données provenant de la National Health and Nutrition Examination Survey (NHANES).

L'article de Wang et Bellhouse explore l'application de méthodes de régression non paramétriques pour étudier la relation entre la variable réponse et les covariables, ainsi que la prédiction en utilisant de l'information auxiliaire dans le contexte des enquêtes complexes. Leurs travaux prolongent ceux de Bellhouse et Stafford (2001), qui ont appliqué une fonction de régression non paramétrique simple au cas de plusieurs variables indépendantes, y compris les variables indicatrices qui figurent souvent dans l'analyse par régression portant sur des données d'enquête.

Et finalement, nous sommes heureux d'informer les lecteurs et auteurs que *Techniques d'enquête* sera bientôt citée par SCOPUS sur les bases de données Elsevier Bibliographic Databases, et ce à partir du numéro de juin 2008.

Harold Mantel, rédacteur en chef délégué