Techniques d'enquête, décembre 2006
Vol. 32, N° 2, pp. 135-136
Statistique Canada, N° 12-001-XPB au catalogue

Dans ce numéro

Ce numéro de *Techniques d'enquête* débute par le sixième article de la série d'articles annuels sollicités dédiée à Joseph Waksberg. C'est avec tristesse que nous soulignons sa disparition survenue en janvier 2005. Une brève biographie de Joseph Waksberg a été publiée dans notre numéro de juin 2001, qui présentait aussi le premier article de la série. Pour plus d'informations sur la vie et l'oeuvre de Joseph Waksberg, voir l'article de Statistical Science (Vol. 15, No 3) "A Conversation with Joseph Waksberg," par David Morganstein et David Marker disponible à http://projecteuclid.org/Dienst/UI/1.0/Home. J'aimerais remercier les membres du Comité de sélection – David Bellhouse, président, Gordon Brackstone, Sharon Lohr et Wayne Fuller – d'avoir choisi Alastair Scott comme auteur de l'article du prix Waksberg de cette année.

Dans son article intitulé « Études cas-témoins basée sur la population », Scott traite des études cas-témoins pour lesquelles les témoins sont obtenus à partir d'une enquête par sondage complexe. En utilisant l'exemple de la régression logistique, il démontre que les estimations pondérées de l'enquête peuvent être passablement inefficaces en raison du poids assez faible attribué aux cas. Établissant une analogie avec l'estimation du maximum de vraisemblance, il propose ensuite une solution de rechange simple et bien plus efficace, mais biaisée pour l'ordonnée à l'origine. À l'aide d'exemples, il démontre les propriétés d'efficacité et de robustesse. Finalement, il examine brièvement la question des études cas-témoins portant sur des familles.

Pour sa part, Kott examine la pondération par calage pour corriger les erreurs de non-réponse et de couverture. Il donne une description générale de l'estimation par calage et élargit l'approche de la forme fonctionnelle d'Estavo et de Särndal à la fonction générale de calage. Il analyse ensuite les propriétés de cette méthode de calage pour corriger les erreurs dues à la non-réponse des unités et les erreurs de couverture en fonction d'un modèle de quasi-randomisation. Il conclut avec un exemple empirique et une analyse de certaines questions importantes.

À l'aide d'une étude de simulation, Reiter, Raghunathan et Kinney étudient les conséquences de ne pas tenir compte des variables du plan de sondage lors de l'élaboration de modèles d'imputation dans un contexte d'imputation multiple. Ils montrent que les biais potentiels peuvent être réduits en incluant ces variables dans le modèle d'imputation au moyen d'un modèle à effets fixes ou à effets mixtes. Les auteurs concluent que les imputeurs auraient intérêt à inclure comme prédicteurs toutes les variables liées aux variables imputées, en particulier les variables du plan de sondage, de manière à satisfaire aux conditions de l'hypothèse habituelle selon laquelle la non-réponse est ignorable.

Dans leur article, Funaoka, Saigo, Sitter et Toida se penchent sur l'utilisation des estimateurs de variance bootstrap dans l'échantillonnage stratifié à plusieurs degrés lorsque les fractions de sondage sont grandes. Ils proposent une méthode bootstrap de type Bernoulli qui donne des estimations cohérentes de la variance bootstrap lorsque l'échantillonnage aléatoire simple sans remise est utilisé à chaque degré. La méthode proposée est simple à appliquer et peut être étendue à n'importe quel nombre de degrés d'échantillonnage sans trop de complications. La méthode est illustrée à l'aide d'une étude de simulation limitée qui utilise des données de l'Enquête nationale sur les prix menée en 1997 au Japon.

Dans leur article, Kozak et Verma comparent l'approche géométrique de stratification, proposée par Gunning et Horgan (2004), avec deux approches par optimisation : l'algorithme de Lavallé-Hidiroglou (Lavallée et Hidiroglou 1988) et l'algorithme d'optimisation proposé par Kozak (2004). Au moyen de cinq populations fictives de différentes tailles, les trois méthodes sont comparées selon deux scénarios : une comparaison du coefficient de variation (c.v.) résultant selon une taille d'échantillon fixe et une comparaison des tailles d'échantillon résultant selon un niveau de précision fixe.

Deville et Lavallée, pour leur part, présentent les fondements théoriques généraux de la méthode du partage des poids appliquée au sondage indirect. Ils définissent l'important concept d'une matrice de liens dans le sondage indirect, qui précise comment les éléments de la population échantillonnée sont liés à la population cible et qui pondère ces liens pour que l'estimation soit sans biais. Ils examinent d'importantes propriétés de la matrice de liens et définissent les conditions nécessaires et suffisantes à une matrice de liens optimale. Les auteurs démontrent leur théorie avec quelques exemples intéressants.

Deville et Maumy-Bertrand étudient la détermination d'un plan de sondage et d'une méthode d'estimation dans le cas d'une enquête sur la fréquentation touristique. Le problème principal posé par ce type d'enquêtes est l'absence d'une base de sondage permettant d'atteindre directement les touristes. Pour contourner ce problème, les auteurs suggèrent d'échantillonner des services destinés principalement aux touristes. On se trouve donc dans une situation de sondage indirect pour laquelle la méthode généralisée du partage des poids est utilisée pour obtenir les estimations des paramètres d'intérêt. Certaines extensions à la méthode sont nécessaires. Les auteurs portent leur attention plus particulièrement à l'une d'elles et la décrivent en détail.

Félix-Medina et Monjardin examinent une variante de l'échantillonnage par dépistage de liens. Pour ce faire, ils ont recours à une approche Bayesienne pour construire des estimateurs de taille de population. Cependant, pour que les inférences sur la taille de la population soient robustes aux spécifications erronées du modèle hypothétique, les auteurs font ces inférences selon une approche fréquentiste basée sur le plan. L'étude de simulation montre que les estimateurs proposés donnent de meilleurs résultats que les estimateurs du maximum de vraisemblance qui sont utilisés actuellement.

Dans leur article, Dorfman, Lent, Leaver et Wegman comparent la méthodologie de l'indice des prix à la consommation du Royaume-Uni et celle des États-Unis en s'appuyant sur les mêmes données scanographiques. Ils concluent que pour la population étudiée, l'approche du Royaume-Uni, qui suppose une stratification plus rigoureuse et, ce qui importe davantage, un échantillonnage intrastrate par choix raisonné plus restrictif que l'approche par échantillonnage probabiliste des États-Unis, donne un meilleur indice superlatif estimé cible. C'est le cas quel que soit l'estimateur (ratio des moyennes, moyenne géométrique ou moyenne des ratios) utilisé pour produire l'indice de prix de bas niveau.

Dans leur article, Thomas, Raghunathan, Schenker, Katzoff et Johnson ont recours à l'imputation multiple pour analyser des données qui, à cause d'un plan de sondage matriciel, présentent des valeurs manquantes. Dans le plan de sondage matriciel, seul un sous-ensemble de questions sont posées à chaque répondant afin de réduire le fardeau de réponse. Les auteurs ont élaboré une méthode permettant de créer des formulaires de sondage matriciel. Chaque formulaire contient un sous-ensemble de questions à poser à des répondants choisis au hasard. La méthode est conçue pour que chaque formulaire comprenne des questions qui permettent de prédire les réponses aux questions exclues, ce qui permet de déduire une partie de l'information exclue. La méthode proposée et l'imputation multiple sont évaluées à l'aide de données tirées de la National Health and Nutrition Examination Survey.

Harold Mantel, Rédacteur en chef délégué