

APPLICATION DE L'ESTIMATEUR DE LA VARIANCE JACKKNIFE AVEC SUPPRESSION D'UN GROUPE À L'ANALYSE DES DONNÉES D'UNE ENQUÊTE LONGITUDINALE COMPLEXE

Julia L. Bienias¹, Phillip S. Kott² et Denis A. Evans¹

RÉSUMÉ

L'estimateur de la variance jackknife avec suppression d'un groupe s'avère un outil fort utile de mesure de la variance en cas de plans d'échantillonnage complexes. Cette méthode consiste à diviser l'échantillon de premier degré en groupes mutuellement exclusifs et de variances presque égales, à supprimer un groupe à la fois pour créer un ensemble de répliques, et à procéder sur chaque réplique à des redressements par pondération analogues à ceux effectués sur l'échantillon dans son ensemble. L'estimation de la variance se fait selon la méthode usuelle (non stratifiée) du jackknife. Dans le présent document, nous illustrons l'application de cette méthode à une étude longitudinale complexe à plusieurs cycles, montrant son utilité pour la modélisation par régression logistique des facteurs de risque liés aux nouveaux cas de maladie.

MOTS-CLÉS : Enquêtes complexes; échantillonnage de Poisson; inférence en population finie; modèle.

1. INTRODUCTION

L'estimation de la variance dans le cas de plans d'échantillonnage complexes pose des défis particuliers. En recherche sociale, économique et sur la santé, on tire habituellement un échantillon d'individus pour un suivi plus approfondi lorsque le coût de la collecte de ce genre de données auprès de tous les sujets participant à l'étude est prohibitif. Lors de telles études, les paramètres étudiés sont habituellement les paramètres de régression ou d'autres estimations d'association, plutôt que des moyennes ou des totaux. Les plans d'échantillonnage complexes sont utilisés dans ce contexte pour assurer l'inclusion de certains sous-groupes de la population étudiée ou pour optimiser le plan par rapport à des variables explicatives connues ou soupçonnées d'intérêt.

L'inférence sous plan d'échantillonnage complexe peut être fondée sur un modèle (comme dans Valliant, Dorfman et Royall, 2000) ou sur le plan de sondage (comme dans Wolter, 1985). Lorsque les paramètres à estimer sont les paramètres du modèle, l'utilisation de techniques fondées sur le plan de sondage produisent souvent des inférences qui résistent à certaines formes d'échec du modèle. Voir, par exemple, Skinner (1989). Étant donné les observations sur un échantillon stratifié à plusieurs degrés avec probabilités de sélection au premier degré faibles, on peut utiliser la technique classique du jackknife stratifié pour déterminer simultanément la variance due au modèle et l'erreur quadratique moyenne de randomisation des paramètres estimés.

Dans le présent article, nous discutons d'une autre variante de la technique du jackknife, qui a des avantages pratiques par rapport au jackknife stratifié. Nous décrivons une application de l'estimateur de la variance *jackknife avec suppression d'un groupe (S-d-G)* à une étude longitudinale avec échantillonnage à plusieurs niveaux. Plus précisément, nous appliquons la méthode au Chicago Health and Aging Project (CHAP), qui est une étude longitudinale communautaire visant à examiner les facteurs de risque de problèmes de santé chroniques chez les personnes âgées. L'un des principaux objectifs de l'étude est d'étudier les facteurs de risque de la manifestation de la maladie

¹Rush Institute for Healthy Aging, Rush-Presbyterian-St. Luke's Medical Center, 1645 W. Jackson Blvd., Suite 675, Chicago, IL 60614, U.S.A. jbienias@rush.edu; denis_evans@rsh.net

²National Agricultural Statistics Service, USDA/NASS, Room 305, 3251 Old Lee Hwy., Fairfax, VA 22030-1504, U.S.A. phil_kott@nass.usda.gov

d'Alzheimer. Le plan de sondage courant du CHAP comprend deux composantes : 1) tous les trois ans, tous les membres survivants de la cohorte sont interrogés sur divers sujets liés à la santé; ces interviews incluent des mesures des fonctions cognitives et physiques et 2) lors de chacun de ces cycles de collecte des données, un échantillon de Poisson stratifié est sélectionné parmi les répondants à l'interview visant la population dans son ensemble afin de procéder à un examen clinique détaillé et à des tests neuropsychologiques. Pour étudier les facteurs de risque liés aux nouveaux cas de maladie, on définit une cohorte de personnes * exemptes de la maladie + au point précédent dans le temps et celle-ci forme une strate importante dans la base de sondage. Nous décrivons l'application de l'estimateur de la variance jackknife avec suppression d'un groupe à la modélisation des facteurs de risque de nouveaux cas de la maladie d'Alzheimer par régression logistique.

2. LE CHICAGO HEALTH AND AGING PROJECT

2.1 Objectifs du Chicago Health and Aging Project

Le Chicago Health and Aging Project (CHAP) est une étude longitudinale communautaire réalisée auprès des personnes âgées. Elle porte sur les facteurs de risque caractéristiques de maladies chroniques courantes de la vieillesse, en accordant une attention particulière à la maladie d'Alzheimer. Dépistée pour la première fois au début du vingtième siècle par le Dr Alois Alzheimer (Alzheimer, 1907), la démence de type Alzheimer est une maladie neurodégénérative évolutive particulièrement débilante, caractérisée par l'altération de la mémoire et d'autres processus cognitifs. On n'a pas encore réussi à en établir la cause ni à trouver une cure, mais il existe certains traitements symptomatiques partiels. En raison de sa longue durée et de l'évolution de la structure par âge de la population américaine et d'autres pays industrialisés, il est prévu que le coût du traitement et des soins prodigués aux personnes atteintes de la maladie d'Alzheimer fera peser un fardeau de plus en plus lourd sur le système de soins de santé (Hebert, Scherr, Bienias, Bennett et Evans, sous presse).

Jusqu'à présent, le CHAP a permis de fournir des éléments de réponse concernant l'effet de l'allèle E4 de l'apolipoprotéine E sur l'incidence de la maladie d'Alzheimer dans les populations biraciales (Evans et coll., 2003), la relation entre les activités cognitives et la prévalence ainsi que l'incidence de la maladie (Wilson et coll., 1999; Wilson et coll., 2002), l'évolution des fonctions cognitives au cours du temps (Wilson, Bennett, Bienias, Mendes de Leon, Morris et Evans, sous presse), l'association de la prise de vitamine E et d'autres nutriments antioxydants au risque de la manifestation de la maladie d'Alzheimer (Morris et coll., 2002a), l'altération du fonctionnement cognitif (Morris, Evans, Bienias, Tangney et Wilson, 2002b), ainsi que la relation entre l'apport alimentaire de matière grasse et l'incidence de la maladie (Morris et coll., 2003).

2.2 Aperçu du plan de sondage

La population cible du CHAP comprend toutes les personnes de 65 ans et plus résidant dans trois quartiers géographiquement définis du côté sud de Chicago, en Illinois (Morgan Park, Beverly et Washington Heights). Comme la ségrégation raciale est historique à Chicago, on a choisi ces quartiers parce qu'ils permettent d'inclure des nombres quasi-égaux de Noirs et de Blancs dans l'ensemble des trois quartiers (un faible pourcentage de personnes appartient à une autre race) et parce qu'ils diffèrent économiquement. Les employés responsables du projet ont d'abord procédé à un recensement porte-à-porte pour créer une liste de tous les résidents admissibles. Ensuite, ceux de 65 ans et plus ont été invités à participer à l'étude (7 826 personnes) et 78,7 % d'entre eux (N=6 158) ont participé à l'interview de référence.

Le plan de sondage de base est un plan de sondage par panel, ou cohorte, en vertu duquel tous les participants sont interviewés à leur domicile à peu près tous les trois ans. Cette *interview de la population* inclut des questions au sujet des antécédents de santé, des problèmes de santé courants, de la participation à des activités sociales et cognitives, de l'activité physique et du fonctionnement physique, ainsi que des mesures directes de la fonction cognitive et, lors de l'interview de référence, des questions au sujet des antécédents professionnels et des expériences vécues durant l'enfance. À l'heure actuelle, le troisième cycle complet de collecte des données du CHAP vient d'être terminé.

En outre, lors de chaque cycle d'interviews de l'ensemble de la population étudiée, on sélectionne un échantillon aléatoire stratifié de personnes auxquelles on demande de se soumettre à une évaluation clinique détaillée, qui inclut des tests neuropsychologiques et un examen par un neurologue. Les examens en vue de dépister les cas de maladie d'Alzheimer sont structurés et uniformes, et les examinateurs ne connaissent pas les résultats des tests cognitifs auxquels les sujets ont été soumis durant l'interview visant l'ensemble de la population ni la strate d'échantillonnage. Des critères uniformes de détermination des cas de démence et de maladie d'Alzheimer sont appliqués. L'échantillon qui a été sélectionné à partir de l'échantillon de participants à l'interview de référence de la population représente un échantillon *de prévalence* pour l'estimation du taux de prévalence de la maladie. Lors des deuxième et troisième cycles, on a sélectionné un échantillon à partir d'un sous-ensemble de personnes chez lesquelles on avait établi l'absence de la maladie d'Alzheimer lors du premier cycle. Le présent article se concentre sur ces deux échantillons *d'incidence*.

2.3 Définition des cohortes * exemptes de maladie +

Pour l'échantillon stratifié sélectionné aux fins de l'évaluation clinique lors du deuxième cycle, la base de sondage était fondée sur une cohorte de personnes exemptes de la maladie d'Alzheimer dépistées lors du premier cycle. Ces personnes ont été dépistées de l'une des deux façons qui suivent : 1) elles ont obtenu une cote satisfaisante pour un ensemble de tests d'évaluation du fonctionnement cognitif que l'on a fait passer à toutes les personnes qui ont participé à l'interview visant la population, ou 2) elles ont été sélectionnées pour un examen clinique durant lequel on a constaté qu'elles ne souffraient pas de la maladie. Puis, à partir de cette cohorte synthétique, on a tiré un échantillon stratifié selon l'âge, le sexe, la race (Noir/non-Noir) et l'évolution du fonctionnement cognitif par rapport au cycle précédent (stable ou amélioré, légère altération, altération importante). À l'intérieur des strates, on a sélectionné les unités par échantillonnage de Poisson, parfois appelé aussi échantillonnage * indépendant de Bernoulli +.

L'avantage de cette méthode pour définir une cohorte exempte de la maladie est qu'elle donne à chaque sujet - y compris ceux dont la cote obtenue lors des tests de fonctionnement cognitif n'est pas bonne - la chance d'être sélectionné dans l'échantillon, donc d'éviter les biais dus aux tests de dépistage. Ce genre de biais se produit quand l'examen de dépistage est associé à des facteurs de risque étudiés ou à des covariables importantes. Dans le cas de l'étude de la maladie d'Alzheimer, les résultats des tests de la fonction cognitive sont, en fait, fréquemment associés à des variables importantes, comme le niveau de scolarité et la race.

Pour pallier l'érosion anticipée de l'échantillon due à la morbidité et à la mortalité au fil du temps, une caractéristique unique a été ajoutée au plan de sondage du CHAP lors du troisième cycle de collecte des données. Un deuxième échantillon de mesure de l'incidence a été sélectionné pour évaluer l'incidence de la maladie au moment du troisième cycle parmi les membres d'une nouvelle cohorte exempte de la maladie définie lors du deuxième cycle d'interviews de l'ensemble de la population. La définition d'* exemption de la maladie + était la même que celle appliquée lors de l'interview de référence. Ceci permettra d'utiliser plus efficacement la population étudiée et donnera plus de puissance aux analyses.

Le plan de sondage original du CHAP décrit plus haut a été étendu pour permettre l'entrée de cohortes d'âge successives de personnes vivant dans la collectivité à mesure qu'elles atteignent 65 ans, en répétant le plan de sondage général pour ces nouveaux participants. Le plan de sondage complet est décrit de façon plus détaillée dans Bienias, Beckett, Bennett, Wilson et Evans (sous presse), y compris la discussion de caractéristiques supplémentaires non décrites ici.

3. L'ESTIMATEUR DE LA VARIANCE JACKKNIFE AVEC SUPPRESSION D'UN GROUPE

3.1 Formule générale d'estimation de la variance fondée sur la répétition

Le but des méthodes par répétition est de reproduire le plan de sondage dans une série de sous-échantillons qui sont ensuite utilisés pour refléter l'échantillon global. Chaque sous-échantillon, ou réplique, retient toutes les caractéristiques de sondage du plan original. La variance des estimations calculées à partir de cette série de sous-échantillons est utilisée pour estimer la variance de l'estimateur d'intérêt. La forme générale de l'estimateur par répétition est :

$$\hat{v}(\hat{\theta}) = c \sum_{k=1}^G (\hat{\theta}_k - \hat{\theta})^2, \text{ où} \quad (1)$$

- G nombre de répliques
- c constante qui dépend de la méthode de répétition
- θ estimation à partir d'une réplique donnée ou à partir de l'échantillon complet.

Dans le cas de répliques répétées par la technique du jackknife, l'idée générale est * supprimer * une unité +, repondérer le reste, estimer, répéter pour l'unité suivante + (rappelons les * résidus jackknife + en régression), où * une unité + dépend des détails du plan de sondage (Wolter, 1985; Lee, Forthofer et Lorimor, 1989).

3.2 Échantillonnage de Poisson c. échantillonnage aléatoire stratifié type

Dans le cas d'un échantillon aléatoire simple stratifié, la probabilité de choisir une unité dans une strate de l'échantillon influe sur la probabilité de choisir une deuxième unité d'échantillonnage à partir de la strate. Par exemple, supposons qu'une strate contient 100 unités, et que nous souhaitons sélectionner 10 unités à partir de l'échantillon. La probabilité de choisir le membre *A* de la strate à partir de l'échantillon, $P(A)$ est égale à 1/10. La probabilité de choisir *A*, étant donné que le membre *B* de la strate a déjà été sélectionné, $P(A|B)$, est $9/99 = 1/11$, qui est légèrement inférieure à 1/10. Donc, les mécanismes de sélection des deux unités ne sont pas indépendants. Dans le cas de l'échantillonnage de Poisson, on applique une probabilité de sélection à une unité de population indépendamment de ce qui s'est passé auparavant dans le processus d'échantillonnage. Mathématiquement, $P(A|B) = P(A)$. Le compromis est que, si une strate contient 100 unités associées chacune à une probabilité de sélection de 1/10, rien ne garantit que l'on sélectionnera exactement 10 unités dans l'échantillon, bien que ce résultat soit celui auquel on s'attend * en moyenne +.

Le point essentiel en ce qui concerne l'estimation de la variance est que, à cause de cette caractéristique, les * strates + n'ont, en réalité, pas d'importance. Elles offrent simplement un moyen de se souvenir de la façon dont les probabilités de sélection ont été attribuées aux unités individuelles. En réalité, la sélection de *chaque* unité est indépendante de celle de chaque autre unité, entre les strates ou à l'intérieur de celles-ci.

3.3 Estimateur de la variance jackknife avec suppression d'un groupe

Dans les conditions de l'échantillonnage de Poisson, nous pourrions laisser tomber une seule unité pour définir chaque réplique, mais en pratique, cela serait vraisemblablement irréalisable. Nous souhaitons limiter le plan de sondage en définissant chaque réplique et les probabilités de sélection constituent la caractéristique unique de ce plan. Donc, nous devons créer une série de répliques qui incluent chacune certaines personnes donnant (approximativement) tous les niveaux de probabilités de sélection. Pour cela, nous pouvons ordonner les unités d'après la probabilité de sélection qui leur est associée (ou, de façon équivalente, d'après les poids d'échantillonnage), puis passer systématiquement à travers la liste pour définir un ensemble de groupes. Une réplique donnée serait alors définie en supprimant un groupe de l'échantillon.

Voici l'idée générale de la méthode du jackknife avec suppression d'un groupe que l'on peut appliquer à l'échantillonnage de Poisson : grouper les données de façon raisonnable, puis supprimer un groupe à la fois; le complément du groupe supprimé forme une réplique. Puis, procéder à l'estimation de la variance de la façon habituelle : estimer les quantités d'intérêt d'après chaque réplique et calculer la variance sur l'ensemble de ces estimations pour obtenir une estimation de la variance en échantillon complet. Kott (1998a; 1998b) décrit la méthode; l'annexe D de Kott (1998a) donne une preuve pour une application de ce genre. Dans l'équation (1), *c* est $(G-1)/G$, où *G* est le nombre de groupes définis. Pour un test statistique *univarié*, le nombre de degrés de liberté est *G*-1.

La méthode s'est avérée utile dans le cas d'enquêtes fondées sur l'échantillonnage de Poisson (Kott, 1998a; Kott, 1998b; Kott et Bailey, 2000), ainsi que lors de l'étude de la non-réponse à une enquête (Kott et Bellow, 1999).

La méthode du jackknife avec suppression d'un groupe décrite plus haut peut poser un problème lors de l'estimation des paramètres en population finie. Elle ne tient pas compte de l'effet sur la variance des grandes fractions

d'échantillonnage appliquées à la première phase. Toutefois, lors de l'estimation des paramètres d'un modèle, la question ne se pose pas, car l'objectif réel n'est pas de faire des inférences au sujet de la population cible, à proprement parler, mais au sujet d'une population conceptuellement infiniment grande à partir de laquelle la population cible est elle-même sélectionnée aléatoirement (Fuller, 1975).

4. APPLICATION AU CHAP

4.1 Comparaison avec le jackknife stratifié

Dans les analyses réalisées sur les données provenant du premier échantillon de mesure de l'incidence, les estimations de la variance ont été calculées par la technique conventionnelle du jackknife stratifié sous un plan de sondage à deux UPE par strate (p. ex., Wolter, 1985). Comme le premier échantillon de mesure de l'incidence du CHAP était effectivement un échantillon aléatoire stratifié, nous avons commencé par créer des pseudo-UPE, ou * unités de calcul de l'erreur-type +(UCET) en divisant aléatoirement chaque échantillon de strate en deux. Puis, nous avons défini chaque réplique en passant à travers les strates, en supprimant dans chacune l'une des deux UCET et en repondérant l'autre. Ensuite, nous avons ajusté des modèles de régression linéaire et logistique et des modèles mixtes à effets aléatoires dans SAS® (SAS, 2000) et nous avons estimé les variances dans SAS en utilisant un logiciel personnalisé que nous avons écrit (Bienias, 2001). Sous des hypothèses faibles, cette approche produit des estimations asymptotiquement sans biais de la variance.

Dans une première étape en vue de déterminer l'utilité de l'estimation de la variance jackknife avec suppression d'un groupe (S-d-G) pour le CHAP, nous avons réestimé certains modèles de régression logistique clés provenant des analyses du premier échantillon de mesure de l'incidence en utilisant 100 groupes S-d-G, puisque l'échantillon contenait 91 strates originales. Nos estimations de la variance étaient presque identiques.

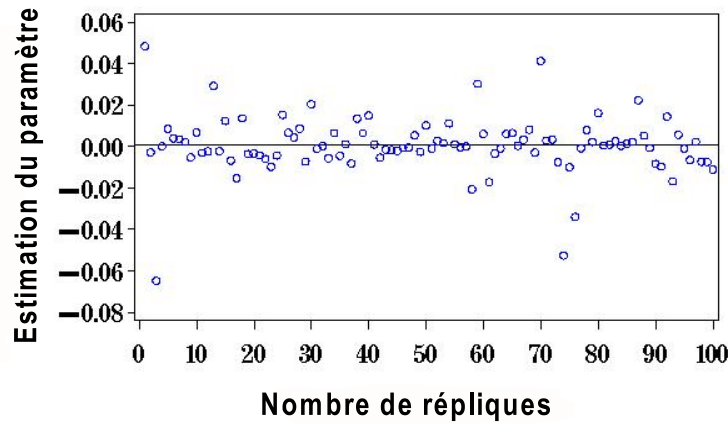
4.2 Choix du nombre optimal de groupes de variances

Dans les enquêtes réalisées par le U. S. National Agricultural Statistics Service (NASS) (voir Kott, 1998a), on a utilisé 15 groupes S-d-G qui ont donné 15 répliques et 14 degrés de liberté pour les tests statistiques univariés. Ce nombre est raisonnable, étant donné que les enquêtes du NASS visent principalement à estimer des moyennes et des totaux. Dans le cas du CHAP, nous voulons estimer les coefficients de modèles complexes, parfois individuellement, mais parfois en groupe. Nous pourrions vouloir réaliser un test F pour un modèle de régression logistique complet comptant 15 variables explicatives. Pour cela, 15 groupes est un nombre manifestement inadéquat (voir Korn et Graubard, 1990). Au lieu de cela, nous avons commencé par envisager 100 groupes, parce que ce nombre correspondait au nombre de strates dans le premier échantillon de mesures de l'incidence. Notre taille d'échantillon pour l'analyse combinée était de 1 134 et les probabilités finales de sélection des unités (tenant compte à la fois des premières et des deuxièmes opérations d'échantillonnage) variaient d'environ 0,05 à 1,0, avec une valeur moyenne de 0,65. Nous avons commencé par examiner un modèle simple d'intérêt établissant la relation entre l'incidence de la maladie et l'âge de la personne au début de la période d'observation, puisque l'âge est le prédicteur le plus important de l'incidence de la maladie d'Alzheimer au stade actuel de nos connaissances. En outre, nous voulons tenir compte du temps écoulé entre le début de la période d'observation d'une personne donnée et le moment où le résultat, c'est-à-dire l'incidence de la maladie, est déterminé (la * durée de l'observation +), parce que plus la période d'observation est longue, plus la probabilité d'observer le résultat est forte, toutes les choses étant égales par ailleurs. Donc, nous avons modélisé :

$$E[\text{logit}(\text{Pr}(\text{maladie d'Alzheimer}))] = \beta_0 + \beta_1 \text{durée de l'observation} + \beta_2 \text{âge.} \quad (2)$$

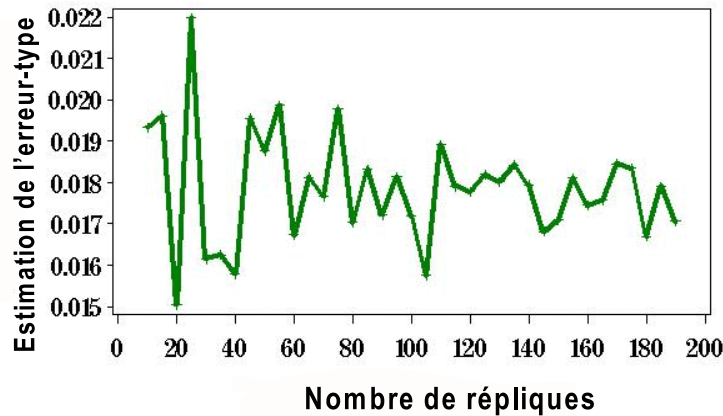
La figure 1 montre les estimations répétées pour la variable * durée de l'observation + pour ce modèle simple (2) établissant la relation entre la cote exprimant le risque d'observer un cas de maladie d'Alzheimer et la durée de l'observation ainsi que l'âge. La droite de référence horizontale passe par le point de donnée estimé pour l'échantillon complet, qui était 0,00070918. En pratique, l'analyse de nos données n'a pas indiqué que la durée de l'observation était un prédicteur significatif, et il en est de même ici. Comme le montre la figure, les estimations sont assez stables avec 100 groupes. L'erreur-type de l'estimation du paramètre de durée de l'observation est de 0,1401.

Figure 1. Estimations des paramètres pour chaque réplique pour le paramètre * durée de l'observation +. La droite de référence correspond à l'estimation pour l'échantillon complet.



Ensuite, nous avons ajusté le modèle à l'équation (2) plusieurs fois, en estimant chaque fois la variance jackknife avec suppression d'un groupe en utilisant un nombre différent de groupes. Rappelons qu'il existe une association biunivoque entre le nombre de groupes de variances et le nombre de répliques formées pour l'estimation de la variance, puisque, pour chaque réplique, nous supprimons un groupe de l'échantillon. La figure 2 montre les estimations de l'erreur-type pour l'estimation du coefficient mesurant l'effet de l'âge au début de l'étude. Comme il faut s'y attendre, les estimations de l'erreur-type sont plus stables à mesure que le nombre de groupes augmente. Notons, toutefois, que ces résultats ne sont produits que pour un seul échantillon.

Figure 2. Estimations de l'erreur-type fondées sur des nombres différents de groupes de variances. Les segments de droite sont tracés pour illustrer plus clairement la tendance sur les répliques.



5. CONCLUSIONS

L'estimateur de la variance jackknife avec suppression d'un groupe est très utile dans des conditions où les techniques ordinaires ne sont pas applicables ou difficiles à mettre en œuvre. Dans nos conditions, la difficulté était due au plan d'échantillonnage double (voir, par exemple, Hidiroglou, 2001). Les probabilités individuelles de sélection sont devenues trop complexes pour qu'on puisse les calculer et les décrire, parce que les individus pouvaient être associés à de nombreux cheminements différents à travers les données du premier cycle au troisième. Par conséquent, nous ne disposions plus de * strates + bien définies comme fondement de l'application des méthodes plus courantes d'estimation

de la variance. Comme nous avons utilisé l'échantillonnage de Poisson, nous avons pu passer directement à l'application de la formule du jackknife avec suppression d'un groupe une fois que le plan de sondage est devenu plus complexe.

Nous continuerons d'étudier l'effet de la variation du nombre de répliques, et nous examinerons également l'utilité de l'estimateur de la variance jackknife avec suppression d'un groupe dans des conditions d'échantillonnage aléatoire stratifié avec ou sans remise.

REMERCIEMENTS

Ces travaux ont été financés par une subvention du United States National Institute on Aging, National Institutes of Health (Grant RO1 AG11101). Les opinions exprimées dans le présent article sont celles des auteurs et ne reflètent pas forcément celles du U. S. National Agricultural Statistics Service. Les auteurs remercient Todd Beck et Woojeong Bang de leur aide pour la programmation.

RÉFÉRENCES

- Alzheimer, A. (1907), "Über eine eigenartige Erkrankung der Hirnrinde," *Allg. Z. Psychiatr. Psych. Gerichtl. Med.*, 64, pp. 146-148; translated by Jarvik, L., and Greenson, H., 1987, *Alzheimer Disease and Associated Disorders*, 1, pp. 7-8.
- Bienias, J. L. (2001), "Replicate-Based Variance Estimation in a SAS® Macro," *Proceedings of the Fourteenth Annual Meeting of the NorthEast SAS Users Group, Statistics, Data Analysis, and Econometrics Section*, pp. 727-735. [Available from <http://www.nesug.org>.]
- Bienias, J. L., Beckett, L. A., Bennett, D. A., Wilson, R. S., and Evans, D. A. (in press), "Design of the Chicago Health and Aging Project (CHAP)," *Journal of Alzheimer's Disease*.
- Evans, D. A., Bennett, D. A., Wilson, R. S., Bienias, J. L., Morris, M. C., Scherr, P. A., Hebert, L. E., Aggarwal, N., Beckett, L. A., Joglekar, R., Berry-Kravis, E., et Schneider, J. (2003), "Incidence of Alzheimer's Disease Incidence in a Biracial Urban Community: Relation to Apolipoprotein E Allele Status," *Archives of Neurology*, 60, pp. 185-189.
- Fuller, W. A. (1975), "Regression Analysis for Sample Survey," *Sankhya*, Series C, 37, pp. 117-132.
- Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A., et Evans, D. A. (in press), "Alzheimer's Disease in the US Population: Prevalence Estimates Using the 2000 Census," *Archives of Neurology*.
- Hidiriglou, M. A. (2001), "Double Sampling," *Survey Methodology*, 27, pp. 143-154.
- Korn, E. L. et Graubard, B. (1990), "Simultaneous Testing of Regression Coefficients With Complex Survey Data," *American Statistician*, 44, pp. 270-276.
- Kott, P. S. (1998a), *Using the Delete-A-Group Jackknife Variance Estimator in NASS Surveys*, RD Research Report No. RD-98-01, U. S. Department of Agriculture, National Agricultural Statistics Service.
- Kott, P. S. (1998b), "Using the Delete-a-Group Jackknife Variance Estimator in Practice," *Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods*, pp. 763-768. Alexandria, VA: American Statistical Association.
- Kott, P.S. et Bailey, J. T. (2000), "The Theory and Practice of Maximal Brewer Selection," *Proceedings of the Second International Conference on Establishment Surveys, Invited papers*, pp. 269-278.

- Kott, P. S., et Bellow, M. (1999), "Capturing and Evaluating the Effect of Reweighting for Unit Nonresponse," paper presented at the International Conference on Survey Nonresponse, Portland, Oregon.
- Lee, E. S., Forthofer, R. N., et Lorimor, R. J. (1989), *Analyzing Complex Survey Data* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-071), Newbury Park, CA: Sage Publications.
- Morris, M. C., Evans, D. A., Bienias, J. L., Tangney, C. C., Bennett, D. A., Aggarwal, N., Schneider, J., et Wilson, R. S. (2003), "Dietary Fats and the Risk of Incident Alzheimer's Disease," *Archives of Neurology*, 60, pp. 194-200.
- Morris, M. C., Evans, D. A., Bienias, J. L., Tangney, C. C., Bennett, D. A., Aggarwal, N., Wilson, R. S., et Scherr, P. A. (2002a), "Dietary Intake of Antioxidant Nutrients and the Risk of Incident Alzheimer's Disease in a Biracial Community Study," *Journal of the American Medical Association*, 287, pp. 3230-3237.
- Morris, M. C., Evans, D. A., Bienias, J. L., Tangney, C. C., et Wilson, R. S. (2002b), "Vitamin E and Cognitive Decline in Older Persons," *Archives of Neurology*, 59, pp. 1125-1132.
- Skinner, C. J. (1989), "Domain Means, Regression, and Multivariate Analysis" in C. J. Skinner et al. (eds.), *Analysis of Complex Surveys*, New York: Wiley, pp. 59-87.
- SAS Institute Inc. (2000), "SAS/STAT® User's Guide, Version 8," Cary, NC: SAS Institute Inc.
- Valliant, R., Dorfman, A. H., et Royall, R. M (2000), *Finite Population Sampling and Inference: A Prediction Approach*, New York: Wiley.
- Wilson, R. S., Bennett, D. A., Beckett, L. A., Morris, M. C., Gilley, D. W., Bienias, J. L., Scherr, P. A., et Evans, D. A. (1999), "Cognitive Activity in Older Persons from a Geographically Defined Population," *Journal of Gerontology: Psychological Sciences*, 54B, pp. P155-P160.
- Wilson, R. S., Bennett, D. A., Bienias, J. L., Aggarwal, N. T., Mendes de Leon, C. F., Morris, M. C., Schneider, J. A., et Evans, D. A. (2002), "Cognitive Activity and Incident Alzheimer's Disease in a Population-Based Sample of Older Persons," *Neurology*, 59, pp. 1910-1914.
- Wilson, R. S., Bennett, D. A., Bienias, J. L., Mendes de Leon, C. F., Morris, M. C., et Evans, D. A. (in press), "Cognitive Activity and Cognitive Decline in a Biracial Community Population," *Neurology*.
- Wolter, K. M. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.

SAS et SAS/STAT sont des marques de commerce déposées du SAS Institute Inc. aux États-Unis et dans d'autres pays. ® indique l'enregistrement aux États-Unis.