

UTILISATION DE MODÈLES À FONCTION GÉNÉRALISÉE DE LA VARIANCE POUR L'INFÉRENCE À PARTIR DE DONNÉES PROVENANT D'ENQUÊTES SOCIALES ET ÉCONOMIQUES

J. L. Eltinge¹, D. S. Jang², M.J. Chol

RÉSUMÉ

Les analyses de données provenant d'enquêtes sociales et économiques s'appuient parfois sur des modèles à fonction généralisée de la variance pour approximer la variance due au plan de sondage des estimateurs ponctuels des moyennes et proportions de population. Les analystes peuvent utiliser les estimations résultantes de l'erreur-type pour calculer les intervalles de confiance ou les variables à tester pour les moyennes et les proportions étudiées. Comparativement aux estimateurs de la variance basés sur le plan de sondage calculés directement à partir des microdonnées d'enquête, les modèles à fonction généralisée de la variance peuvent offrir plusieurs avantages, dont la simplicité des opérations, une plus grande stabilité des erreurs-types et, dans le cas où l'on utilise des ensembles de données à grande diffusion, la réduction des problèmes de limitation de la divulgation des renseignements personnels que pose la grande diffusion d'indicateurs de strates et de grappes.

Cependant, plusieurs problèmes d'inférence peuvent annuler en partie ces avantages éventuels. Premièrement, les propriétés des statistiques inférentielles fondées sur des fonctions généralisées de la variance (par exemple, taux de couverture et largeur des intervalles de confiance) dépendent fortement de l'importance empirique relative des composantes de la variabilité associée, respectivement, à :

- a) la sélection aléatoire d'un sous-ensemble d'items utilisés pour estimer le modèle à fonction généralisée de la variance;
- b) la sélection d'unités d'échantillonnage conformément à un plan d'échantillonnage complexe;
- c) le mauvais ajustement du modèle à fonction généralisée de la variance;
- d) la génération d'une population finie sous les conditions d'un modèle de superpopulation.

Deuxièmement, sous certaines conditions, on peut lier chacune des composantes (a) à (d) à diverses mesures empiriques de l'adéquation prédictive d'un modèle à fonction généralisée de la variance. Par conséquent, ces mesures d'adéquation prédictive peuvent fournir certains éclaircissements sur la mesure dans laquelle un modèle à fonction généralisée de la variance donné convient à l'inférence dans des applications particulières.

Certains tests et diagnostics proposés sont appliqués aux données de la U.S. Survey of Doctoral Recipients et de la U.S. Current Employment Survey. Pour la Survey of Doctoral Recipients, les composantes (a), (c) et (d) sont celles dont il faut se préoccuper principalement. Pour la Current Employment Survey, on accorde surtout de l'attention aux composantes (b), (c) et (d), et la disponibilité de microdonnées de population permet le développement de modèles particulièrement détaillés pour les composantes (b) et (c).

MOTS CLÉS : Current Employment Survey; inférence basée sur le plan de sondage; effet de spécification incorrecte; ensemble de données à grande diffusion; superpopulation; Survey of Doctoral Recipients.

¹ U.S. Bureau of Labor Statistics; Eltinge_J@bls.gov

² Mathematica Policy Research, É.-U.