



N° 11-522-XIF au catalogue

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

**Symposium 2004 : Méthodes
innovatrices pour enquêter
auprès des populations
difficiles à joindre**

2004



OBSERVATIONS INFLUENTES DANS LES SOUS-POPULATIONS RARES LORS D'ENQUÊTES AUPRÈS D'ÉTABLISSEMENTS

John L. Eltinge¹

RÉSUMÉ

Dans l'analyse basée sur le concept des données d'enquête auprès d'un établissement ou d'une entreprise, on peut souvent être confronté à des problèmes importants résultant de l'échantillon de valeurs atypiques. Ces valeurs atypiques proviennent de diverses sources, y compris : a) les valeurs réelles extrêmes; b) les erreurs grossières dans les rapports; c) les inégalités sévères dans les probabilités de sélection; d) le saut de strate ou autres problèmes de classification; et e) une structure de réponse peu commune ou autres phénomènes menant à des coefficients de pondération extrêmes.

Ce document examine la documentation générale sur les questions de robustesse a)-e) pour ensuite prendre en considération deux problèmes particuliers qui surviennent lorsque l'on tente d'utiliser cette documentation dans l'analyse de données provenant des sous-populations rares. D'abord, dans la documentation traditionnelle sur la robustesse pour a)-b), les fonctions influentes et autres fonctions sont généralement fondées sur la distance d'une observation à partir d'une valeur centrale ou attendue. Lorsqu'on essaie de mettre en pratique cette idée auprès des sous-populations rares dans les enquêtes auprès des établissements (p. ex., les groupes rares fondés sur les classifications industrielles, par type de profession ou géographiques), les définitions appropriées de la valeur centrale ou attendue pour le calcul d'une fonction d'influence ou d'un diagnostic connexe peuvent s'avérer problématiques.

Deuxièmement, la structure de pondération originale (fondée par exemple sur les probabilités de sélection, les ajustements aux non-réponses ou la poststratification) ne peut justifier de manière explicite l'appartenance aux sous-populations d'intérêt rares. Par conséquent, pour certaines de ces sous-populations rares, les questions de pondération c)-e) deviennent plus problématiques.

Pour traiter ces problèmes, ce document élabore un cadre de travail asymptotique voulant refléter les échanges parmi les dimensions relatives et absolues des sous-populations d'intérêt; et les effets des observations et des pondérations extrêmes sur une fonction d'influence idéalisée. Nous utilisons ce cadre de travail asymptotique pour comparer les estimateurs alternatifs des valeurs centrales ou attendues qui seront utilisés dans la fonction d'influence employée pour les unités provenant d'une sous-population particulière rare. Les méthodes et diagnostics qui en résultent sont justifiés par les données, et sont illustrés à l'aide de ces mêmes données, provenant du U.S. Current Employment Statistics program.

MOTS CLÉS : Current Employment Statistics Program, erreurs aberrantes, événements rares asymptotiques, modèle de l'effet principal, « sauteuse » de strate, sous-pondération.

¹ John L. Eltinge, Office of Survey Methods Research, Bureau of Labor Statistics des États-Unis; Eltinge_J@bls.gov