



Catalogue no. 11-522-XIE

**Statistics Canada International Symposium
Series - Proceedings**

**Symposium 2004: Innovative
Methods for Surveying
Difficult-to-reach Populations**

2004



Statistics
Canada

Statistique
Canada

Canada

INFLUENTIAL OBSERVATIONS FROM RARE SUBPOPULATIONS IN ESTABLISHMENT SURVEYS

John L. Eltinge¹

ABSTRACT

In the design-based analysis of establishment or enterprise survey data, one often encounters important issues arising from influential sample values. These influential values arise from several sources, including: (a) extreme true values; (b) gross errors in reporting; (c) severe inequalities in selection probabilities; (d) stratum jumping or other classification problems; and (e) unusual response pattern or other phenomena that lead to extreme weighting factors.

This paper reviews the general literature on robustness issues (a)-(e) and then considers two special problems that arise when one attempts to use this literature in the analysis of data from rare subpopulations. First, in the traditional robustness literature for (a)-(b), influence functions or other functions generally are based on the distance of an observation from a central or expected value. When one attempts to apply this idea to rare subpopulations in establishment surveys (e.g., rare groups based on industrial, occupational or geographical classifications), appropriate definitions of a central or expected value for calculation of an influence function or related diagnostic can be especially problematic.

Second, the original weighting scheme (based, e.g., on selection probabilities, nonresponse adjustments or poststratification) may not account explicitly for membership in the rare subpopulation(s) of interest. Consequently, for some of these rare subpopulations, weighting issues (c) - (e) become more problematic.

To address these issues, this paper develops an asymptotic framework intended to reflect trade-offs among the relative and absolute sizes of subpopulations of interest; and the effects of extreme observations and extreme weights on an idealized influence function. We use this asymptotic framework to compare alternative estimators of central or expected values to be used in the influence function used for units from a specific rare subpopulation. The resulting methods and diagnostics are motivated by, and illustrated with, data from the U.S. Current Employment Statistics program.

KEYWORDS: Current Employment Statistics Program; Downweighting; Main Effect Model; Outlier; Rare-Event Asymptotics; Stratum Jumpers.

¹ John L. Eltinge, Office of Survey Methods Research, U.S. Bureau of Labor Statistics; Eltinge_J@bls.gov