Proceedings of Statistics Canada Symposium 2002
Modelling Survey Data for Social and Economic Research

BRIDGING MULTIPLE-RACE RESPONSES IN THE U.S. CENSUS TO SINGLE-RACE CATEGORIES FOR THE CALCULATION OF VITAL RATES

Nathaniel Schenker, Jennifer D. Parker, and Deborah D. Ingram[1]

ABSTRACT

In 1997, the U.S. Office of Management and Budget issued revised standards for the collection of race information within the Federal statistical system. One revision allows individuals to choose more than one race group when responding to Federal surveys and other Federal data collections. This change presents challenges for analyses that involve data collected under both the old and new race reporting systems, since the data on race are not comparable.

Since most people under both systems report only a single race, a common proposed solution is to try to bridge the transition by assigning a single-race category to each multiple-race reporter under the new system, and to conduct analyses using just the observed and assigned single-race categories. Thus, the problem can be viewed as a missing-data problem, in which single-race responses are missing for multiple-race reporters and need to be imputed.

The Office of Management and Budget suggested several simple bridging methods to handle this missing-data problem. Schenker and Parker (Statistics in Medicine, forthcoming) analyzed data from the National Health Interview Survey of the U.S. National Center for Health Statistics, which allows multiple-race reporting but also asks multiple-race reporters to specify a primary race, and found that improved bridging methods could result from incorporating individual-level and contextual covariates into the bridging models.

The current paper considers a specific bridging problem that arises when vital rates are calculated by race. Such rates are frequently used in epidemiologic and other studies. Beginning in 2000, data from the U.S. decennial census, which are used to calculate the denominators for rates, were collected under the new race reporting system. In contrast, birth and death records, which provide the data for the numerators, are implementing the change to the new system over the next several years. Thus, numerators will often be available under the old single-race categories, whereas denominators will be available under the categories of the new system.

To bridge the census data with the vital event data, we build on the approach of Schenker and Parker (Statistics in Medicine, forthcoming) and fit bridging models to data from the National Health Interview Survey. We then apply the resulting bridging methods to the 2000 census to create denominators under the old race reporting system. Estimates of vital rates will be used to demonstrate the results of the methodology and to compare these results with those from simpler approaches.

While Schenker and Parker (Statistics in Medicine, forthcoming) discussed only three large multiple-race groups, the current application requires predicting single-race categories for several small multiple-race groups as well. Thus, problems of sparse data arise in fitting the bridging models. We address these problems by building combined models for several multiple-race groups, thus borrowing strength across them. These and other methodological issues will be discussed.

KEY WORDS: Logistic regression; Missing data; Imputation; Survey; Vital records.

[1] U.S. National Center for Health Statistics, U.S.A., nschenker@cdc.gov