LBD Documentation: Auto Repair Shop Industry: Evaluation of Probabilistic Linking Algorithms

Javier Miranda
U.S. Census Bureau

CES Technical Notes
CES-TN-2002-02

Abstract

Probabilistic matching algorithms using name and address information are evaluated in a set of experiments that link establishments from the 1987 SSEL to the 1992 SSEL. To facilitate the computation of error match rates I focus on a narrow industry subset of all establishments in the SSEL; the Auto Repair Shop Industry (SIC 753). These algorithms will be used in the development of the Longitudinal Business Database, a database of longitudinally linked SSEL records. The algorithms tested generate a substantial number of links that would otherwise be missed using only numerical linking procedures. The error match rates achieved are significantly below those achieved in previous attempts using similar but less sophisticated procedures.

Relevant Datasets:
LBD

Required Dataset Permissions:
SSEL 1997 & 1992

CES Technical Notes may contain confidential data and, thereby, disclosure is prohibited. The opinions, conclusions and suggestions expressed in these notes are those of the authors and do not represent those of the U.S. Census Bureau. Researchers (CES staff, SSS and other Census employees) who wish to access the full text of a technical note or submit their own technical note should contact CES.Technical.Notes.List@census.gov.