Socio-Economic Status and Names

Relationships in 1880 Male Census Data
Rebecca Vick, University of Minnesota

Study Question
This research asks if there is a relationship between name commonness and socio-economic status. This question is important because it has implications for linked samples.

When linking historical records across two or more samples, the best way to avoid false links is to exclude records that match to more than one record. This usually occurs with common name like John Smith. Unique names, like Percy Pinkerton, typically have strong similarity to one and only one record, or no records at all (in the case of mortality, enumeration, or enumeration error).

Names are a personal identification method, something so basic that perhaps their relationship with other demographic characteristics are assumed to be random or benign. But is that assumption true? It behooves record linkers who use name data to know whether or not this is possible. I use 1880 U.S. Census data and the Duncan socio-economic index measure to examine this question.

The process of connecting records for the same individual across two or more data sources. Linked files are uniquely rich in information about individual life change such as migration, occupational mobility and household composition. Historical linked datasets contain new information that could solidify, enlighten or expand our knowledge of social science and demographic history.

Table 1. The Twenty Most-common First-Last Name Combinations in 1880 U.S. Census

Rufus Pinkerton
John Smith
Rufus Pinkerton
John Smith

Duncan Socio-Economic Index Score

The Duncan Socio-Economic Index score is the occupational standing measure used to measure socio-economic status for this study.

Name Processing
• Non-alphabetic characters, titles and non-pertinent characters removed
• Dictionary of standardized names applied to first name data to correct for abbreviations and nicknames.
• Although statistically significant, the SEI difference between common and uncommon names is very small (2.4 points). It is difficult to determine how this difference might affect a linked sample, but we can use this information to form an opinion. The proportion of working males with very common names is very small. The names deemed most common in this paper comprise a little over 1% of the overall study population. And the majority (69%) have uncommon names. The SEI difference between least common and most common names is 2.4 points—far less significant than typical differences. In my estimation, the threat to the quality of linked samples that stems from throwing out a disproportionate number of common names to avoid false links is likely negligible.

Socio-Economic Status and Names

Relationships in 1880 Male Census Data
Rebecca Vick, University of Minnesota

Analysis

Table 2. Name Commonness in Four Categories: 1880 Males Aged 20-50 with Occupational Responses

See Duncan's 1961 paper "A Socioeconomic Index for All Occupations"