

YouGovPolimetrix

Voter Validation Variables

matched - The record matched the voter file (1 if yes, 0 if no)

matchState - The voter file state matched to the record. In most cases, matchState is the same state that the respondent recorded in the survey. However, the Polimetrix matching algorithm looks for matches by (1) residential address, (2) mailing address, and (3) NCOA address. For example, the algorithm will detect if a panelist has moved and submitted NCOA information. If the NCOA address matches voter registration from another state, the respondent will be matched to the address in the new state (note: NCOA is only used on the respondent's address, not the voter file addresses). Similarly, if the respondent's address in the survey is a mailing address rather than the registration address, the match will also determine the actual registration address based on the information given. Wisconsin and North Dakota do not have voter files; these are the only two states that were not matched.

G2006 - Y/NULL for whether the respondent voted in the 2006 general election. Note that a null may mean that the respondent did not vote, the respondent wasn't registered yet, or the state, county or precinct is missing that information.

G2006status describes the completeness of vote history for a state. The values are:

1: There is no vote history for 2006 at all

2: There is some vote history for 2006, but it is incomplete. Often this means that many counties have complete information, but some are missing turnout entirely. For example, in New Jersey the state level Voter File Turnout over Actual Turnout is 75.8% (status 2), but on a county-by-county basis there are 4 counties (out of 21 total) with exactly 0 votes in 2006. Presumably if these 4 counties were disregarded, the numbers would be much closer to 100%

3: There is complete vote history for 2006. In this case the Voter File Turnout / Actual Turnout is between 95% and 105%. Most of the states in this category are between 98-99%.

YouGovPolimetrix

Voter Registration Data Matching with PoliMatch

To do these kinds of matches, YouGov Polimetrix has developed PoliMatch, software for name and address parsing and fuzzy matching that is tailored to one specific application: matching to voter registration lists. Voter lists have some unusual characteristics that can make matching somewhat difficult, but, if exploited properly, can increase the accuracy of the process. For example, many households have husbands and wives on the voter list with identical last names, similar birth years, but different genders. In this case, PoliMatch's algorithms must avoid mismatching two registrants with the same last name and address and possibly identical (or close) birth years. To achieve this, first name and gender are given additional weight in determining the best match. Similarly, many households have fathers and sons (or mothers and daughters) with identical last names, genders, and (sometimes) first names, but substantially different birth years and (sometimes) name suffixes. In this case, different criteria are used to avoid mismatches.

PoliMatch uses a series of steps to find the best matching record. In the first stage, records are parsed and standardized into a common set of fields. Then records are indexed into "blocks" that will be searched for matching records. It is beyond current computational capabilities to do pair-wise comparisons of all records with a database this large. Instead, the software defines blocks that are almost certain to contain a matching record if it exists in the other file. Next, PoliMatch conducts pair-wise comparisons of all records within corresponding blocks of each file. The comparisons use a set of customized fuzzy string matching algorithms to compute a probability-like score that each pair of records match.

Based upon YouGov Polimetrix's internal testing and validation studies, we have found that PoliMatch delivers match rates superior to database joins (and similar to manual matching) for large lists at substantially lower cost and with a relatively high degree of accuracy. For example, in a case where the first name of the panelist is listed as "Jeff" and the first name on the voter registration record is "Jeffrey," but all other name and address fields match, a simple database join would fail to recognize the match, whereas a manual match and a fuzzy match done by PoliMatch would accept that both of these records were for the same person.