Sampling Procedures

Survey Panel Data

The YouGov panel, a proprietary opt-in survey panel, comprises 1.2 million U.S. residents who have agreed to participate in YouGov Web surveys. At any given time, YouGov maintains a minimum of five recruitment campaigns based on salient current events.

Panel members are recruited by several methods and on various topics to help ensure diversity in the panel population. Recruiting methods include Web advertising campaigns (public surveys), permission-based email campaigns, partner sponsored solicitations, telephone-to-Web recruitment (RDD based sampling), and mail-to-Web recruitment (Voter Registration Based Sampling).

The primary method of recruitment for the YouGov Panel is Web advertising campaigns that appear based on keyword searches. In practice, a search in Google may prompt an active YouGov advertisement soliciting opinion on the search topic. After the short survey concludes, respondents are invited to join the YouGov panel to receive and participate in additional surveys. After a double opt-in procedure, where respondents must confirm their consent by responding to an email, the database checks to ensure the newly recruited panelist is new and that the address information provided is valid.

Sampling and Sample Matching

Sample matching is a methodology for selecting “representative” samples from non-randomly selected pools of respondents. It is ideally suited for Web access panels but could also be used for other types of surveys, such as phone surveys. Sample matching starts with an enumeration of the target population. For general population studies, the target population is all adults and can be enumerated through the decennial Census or a high-quality survey, such as the American Community Survey. In other contexts, this is known as the sampling frame, though, unlike conventional sampling, the sample is not drawn from the frame. Traditional sampling, then, selects individuals from the sampling frame at random for participation in the study. This may not be feasible or economical as the contact information, especially email addresses, is not available for all individuals in the frame. Refusals to participate increase sampling costs in this way.

Second, we select one or more matching members from our pool of opt-in respondents for each member of the target sample. This is called the matched sample. Matching is accomplished using a large set of available variables in consumer and voter databases for both the target population and the opt-in panel.

Third, for each member of the target sample, we select one or more matching members from our pool of opt-in respondents. This is called the matched sample. Matching is accomplished using a large set of variables that are available in consumer and voter databases for both the target population and the opt-in panel.

The purpose of matching is to find an available respondent similar to the selected member of the target sample. The result is a sample of respondents who have the same measured characteristics as the target sample. Under certain conditions described below, the matched sample will have similar properties to a true random sample. That is, the matched sample mimics the characteristics of the target sample.

When choosing the matched sample, it is necessary to find the closest matching respondent in the panel of opt-ins to each member of the target sample. YouGov employs the proximity matching method to find the closest matching respondent. For each variable used for matching, we define a distance function, d(x,y), which describes how "close" the values x and y are on a particular attribute. The overall distance between a member of the target sample and a member of the panel is a weighted sum of each attribute's individual distance functions. The weights can be adjusted for each study based upon which variables are thought to be important for that study, though, for the most part, we have not found the matching procedure to be sensitive to small adjustments of the weights. On the other hand, a large weight forces the algorithm toward an exact match on that dimension.