

Overview of KnowledgePanel[®] Statistical Weighting Protocol

November 11, 2009

Prepared by Charles DiSogra
cdisogra@knowledgenetworks.com
Chief Statistician
Knowledge Networks

In a fashion similar to other probability-based samples, interview study samples drawn from KnowledgePanel[®] require a series of statistical adjustments to account for the probabilities of selection and also to render the sample more representative of the population of interest. These weights support analysis of the survey data for the purpose of estimating accurate and generalizable measures. These adjustments are accomplished through the application of statistical weights that function as a unique multiplier for each case in the study's interview data file. The computation of these weights is a complex task and addresses several features that are the necessary result of construction of the panel from recruitment through the assigned study-specific sample and finally the response to the study survey itself (see Callegaro and DiSogra, 2008). KnowledgePanel weights thus take into account what are in effect two distinct sampling segments: 1) panel recruitment and construction which are handled in a panel base weight, and 2) elements related to the study sample incorporated in study-specific weights.

Panel Base Weight

The recruitment of all panel members is based on a probability sample that had been traditionally sourced through a random-digit dial (RDD) sample frame but is now more recently sourced through an address-based sample frame (ABS) (see DiSogra, Callegaro, Hendarwan, 2009; Dennis, 2009). The national ABS recruitment sample provides superior coverage of the U.S. population including the rapidly growing proportion of households that do not have landline telephone service (almost entirely replaced by cell phones). Current KnowledgePanel members available for study samples had either been recruited from RDD methods or more recently from ABS mail methods. This makes the existing panel a dual-frame composition.

The weighting computation takes these differing frames and their related design features into account. This includes the different selection probabilities associated with the RDD- and ABS-sourced samples since a number of efficiency strategies are used in the respective recruitment samples, such as the oversampling of minority communities and of Spanish-language dominant Hispanic areas.

Additionally, there are several sources of error that are an inherent part of any complex survey process, for example, non-response to panel recruitment and panel attrition among recruited panel members. These and other sources of sampling and non-sampling error are corrected with a post-stratification adjustment using demographic distributions from the most recent data from

the Current Population Survey (CPS), the monthly population survey conducted by U.S. Bureau of the Census that is the gold standard for measuring demographic and other trends in the U.S.. Other reliable benchmarks (not available in CPS) are employed for Spanish language usage and for household Internet access adjustments. The demographic variables used in this post-stratification task are gender, age, race/ethnicity, education, U.S. Census region, metropolitan area, Internet access, and language spoken at home (English/Spanish).

The results of all these adjustments go into the panel base weight that is used to make KnowledgePanel representative of the U.S. population of adults. This panel base weight is then employed in a probability proportional to size (PPS) selection method for drawing specific study samples from KnowledgePanel.

Study-Specific Weights

Once a study sample has been drawn, fielded and the data compiled from all KnowledgePanel respondents, a sample-specific post-stratification process is carried out to adjust for survey non-response and for any elements related to the study-specific sample design (such as subgroup oversamples). An iterative raking procedure starting with the panel base weight is used for this task. Demographic and geographic distributions for the population ages 18+ from the most recent CPS provide the majority of the benchmarks for this adjustment. The demographic variables used are again gender, age, race/ethnicity, education, U.S. Census region, metropolitan area, Internet access, and, if appropriate for the study sample, language spoken at home (English/Spanish).

The distribution of these calculated weights are generally examined to identify and trim outliers at the extreme upper and lower tails of the weight distribution. These trimmed weights can be scaled to the nominal sample sizes to optimize statistical analyses as well as scaled to the generalizable population size for population count estimates.

References

Callegaro, Mario & Charles DiSogra (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*. 72(5) pp. 1008-1031.

Dennis, J. Michael (2009). Summary of KnowledgePanel[®] Design. Available at Other methodological papers related to KnowledgePanel are available at <http://www.knowledgenetworks.com/ganp/reviewer-info.html>.

DiSogra, Charles, Mario Callegaro, & Erlina Hendarwan (2009). Recruiting Probability-Based Web Panel Members Using an Address-based Sample Frame: Results from a pilot study conducted by Knowledge Networks. Presented at the 2009 Joint Statistical Meetings. Full paper available at <http://www.knowledgenetworks.com/ganp/reviewer-info.html>.