Privacy and Security Experiences of Low-Socioeconomic Status Populations

Survey Methodology Report

Princeton Survey Research Associates International for
Data & Society Research Institute

January 2016

WEIGHTING AND ANALYSIS

Weighting is generally used in survey analysis to adjust for effects of the sample design and to compensate for patterns of nonresponse that might bias results. The weighting was accomplished in multiple stages to account for the disproportionately-stratified samples, the overlapping landline and cell sample frames, household composition, and differential non-response associated with sample demographics.

The first stage of weighting corrected for different probabilities of selection associated with the number of adults in each household and each respondent's telephone usage patterns.[1] This weighting also adjusts for the overlapping landline and cell sample frames and the relative sizes of each frame and each sample. Since we employed a disproportionately-stratified sample design, the first-stage weight was computed separately for each stratum in each sample frame.

The first-stage weight for the ith case from stratum h can be expressed as:

$$WT_{hi} = \left[\left(\frac{S_{LLh}}{F_{LLh}} \times \frac{1}{AD_{hi}} \times LL_{hi}\right) + \left(\frac{S_{CPh}}{F_{CPh}} \times CP_{hi}\right) - \left(\frac{S_{LLh}}{F_{LLh}} \times \frac{1}{AD_{hi}} \times LL_{hi} \times \frac{S_{CPh}}{F_{CPh}} \times CP_{hi}\right)\right]^{-1}$$

Where S_{LLh} = the size of the landline sample in stratum h

F_{LLh} = the size of the landline sample frame in stratum h

S_{CPh} = the size of the cell sample in stratum h

F_{CPh} = the size of the cell sample frame in stratum h

AD_{hi} = Number of adults in household i of stratum h

LL_{hi}=1 if respondent i of stratum h has a landline phone, otherwise LL_{hi}=0.

CP_{hi}=1 if respondent i of stratum h has a cell phone, otherwise CP_{hi}=0.

[1] i.e., whether respondents have only a landline telephone, only a cell phone, or both kinds of telephone.

This first-stage weight was used as an input weight for the demographic raking. The data was first divided into three groups – African-Americans, Hispanics and others. Each group was raked separately to population parameters for sex, age, education, region and number of adults in the household.

After the raking by race/ethnicity, the combined dataset was raked to total adult population parameters for sex, age, education, region, number of adults in the household, household telephone usage, population density and race/ethnicity.

The telephone usage parameter was derived from an analysis of recently available National Health Interview Survey data[2]. The population density parameter is county-based and was derived from Census 2010 data. All other weighting parameters were derived from an analysis of the 2013 American Community Survey 1-year PUMS file.

Each stage of weighting incorporated previous weighting adjustments. Raking was accomplished using SPSSINC RAKE, an SPSS extension module that simultaneously balances the distributions of all variables using the GENLOG procedure. The rakings correct for differential non-response that is related to particular demographic characteristics of the sample. The weight ensures that the demographic characteristics of the sample closely approximate the demographic characteristics of the target population. Table 1 compares weighted and unweighted total sample demographics to population parameters.

[2] Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December, 2014. National Center for Health Statistics. June 2015.

Table 1. Sample Demographics

	Parameters	Unweighted	Weighted
Sex			
Male	48.2%	52.4%	48.7%
Female	51.8%	47.6%	51.3%
Age			
18-29	20.9%	16.3%	20.1%
30-49	34.7%	24.6%	32.6%
50-64	26.0%	28.8%	25.4%
65+	18.4%	27.0%	18.6%
Education			
LT HS	13.3%	12.8%	12.6%
HS graduate	28.0%	27.4%	27.8%
Some college	31.0%	24.0%	30.0%
College graduate	27.7%	34.6%	28.7%
Region			
Northeast	18.0%	13.7%	17.3%
Midwest	21.3%	13.6%	20.3%
South	37.3%	46.2%	38.5%
West	23.4%	26.6%	23.8%
# of adults in HH			
1	16.5%	27.4%	17.5%
2	51.9%	48.9%	51.3%
3+	31.6%	23.8%	31.2%
HH phone use			
LLO	7.4%	4.7%	6.4%
Dual	44.8%	55.5%	45.2%
CPO	47.8%	39.8%	48.4%
Population Density			
1-Lowest	19.9%	30.3%	20.7%
2	20.0%	18.6%	20.0%
3	20.1%	14.6%	19.9%
4	20.0%	13.3%	19.4%
5-Highest	20.0%	23.3%	20.0%

Continued…

Table 1. Sample Demographics (continued)

	Parameters	Unweighted	Weighted
Race/ethnicity			
White, not Hispanic	65.8%	58.1%	62.8%
Black, not Hispanic	11.5%	14.0%	11.8%
Hispanic, native born	7.5%	8.9%	7.8%
Hispanic, foreign born	7.5%	9.7%	7.8%
Other, not Hispanic	7.6%	6.7%	7.4%

EFFECTS OF SAMPLE DESIGN ON STATISTICAL INFERENCE

Specialized sampling designs and post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. PSRAI calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called "design effect" or *deff* represents the loss in statistical efficiency that results from a disproportionate sample design and systematic non-response. PSRAI calculates the composite design effect for a sample of size *n*, with each case having a weight, w_i as:

$$deff = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}$$

In a wide range of situations, the adjusted standard error of a statistic should be calculated by multiplying the usual formula by the square root of the design effect ($\sqrt{deff}$). Thus, the formula for computing the 95% confidence interval around a percentage is:

$$\hat{p} \pm \left(\sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

where $\hat{p}$ is the sample estimate and *n* is the unweighted number of sample cases in the group being considered.

The survey's margin of error is the largest 95% confidence interval for any estimated proportion based on the total sample — one around 50%. For example, the margin of error for the total sample is ±2.7 percentage points. This means that in 95 out every 100 samples using the same methodology, estimated proportions based on the entire sample will be no more than 2.7 percentage points away from their true values in the population. It is important to remember that sampling fluctuations are only one possible source of error in a survey estimate. Other sources, such as measurement error, may contribute additional error of greater or lesser magnitude.