WEIGHTING SUMMARY

FOR ISPU'S 2022 ILLINOIS MUSLIM SURVEY

Prepared for the
Institute for Social Policy and Understanding

SSRS Contact:
Jordon Peugh, EVP
jpeugh@ssrs.com

ssrs
research. refined.

FEBRUARY 24, 2022

SSRS Job #V1021

Introduction

On behalf of the Institute of Social Policy and Understanding (ISPU), SSRS created benchmarks and a weighting variable for survey data collected by ISPU. ISPU provided SSRS with an Excel file containing responses from 1,620 Muslim respondents living in Illinois. These data were collected by ISPU and their partner through convenience sampling approaches.

The following document describes the procedures that SSRS undertook to identify benchmarks for adult Muslim residents of Illinois and weight the data provided by ISPU.

Weighting Procedures

The data consists of Illinois Muslim adults of age 18+ years old. The data were weighted to correct for systematic nonresponse along population parameters. See Table 1 for population parameters and source.

Table 1: Calibration Variable Sources

Dimension	Source
Gender	Multilevel Regression and Poststratification (MRP) Estimates
Age	
Education	
Voter registration	SSRS Omnibus[1] and Opinion Panel[2] Estimates

Due to the small population, compared to the national population, parameters were estimated using an MRP model. MRP is a small area estimation method that can improve the precision of estimates within small domains by leveraging correlations that are observed in a larger population. First, a multilevel logistic regression model was estimated using SSRS Omnibus and Opinion Panel data. The model's dependent variable was an indicator of whether the respondent identified as Muslim. Predictors included gender, age, education, race, and marital status; an indicator for whether the respondent lived in Illinois; and interactions between the Illinois indicator and all demographics. The model was then applied to a poststratification table from the 2019 American Community Survey (ACS) for the national adult (18+) population. Predicted probabilities obtained from the model were multiplied by the ACS population estimates to estimate the number of Muslims within cells defined by all possible combinations of the model predictors. These estimates were then used to estimate the percentage distribution of the above demographic dimensions among Muslims in Illinois. MRP was not used to estimate voter registration as it was not available in ACS 2019; thus, this dimension was directly estimated from SSRS proprietary data.

Although race and marital status were included as predictors in the MRP model to help obtain accurate benchmarks for the weighting variables, these two dimensions were not themselves used in this weighting. Race was excluded from the weighting because the racial categories available in the survey data differed

[1] https://ssrs.com/ssrs-omnibus-surveys/
[2] https://ssrs.com/opinion-panel/

substantially from the SSRS proprietary and ACS data used in the modeling. Upon reviewing the benchmarks, marital status was dropped from the weighting because the resulting benchmark differed substantially from other sources, and it could not be determined which source was the most accurate.

After the parameters for gender, age, and education were estimated using MRP, and the parameter for voter registration using SSRS proprietary data, the survey data were weighted to those parameters. Weighting was accomplished using SPSSINC RAKE, an SPSS extension module that simultaneously balances the distributions of all variables using the GENLOG procedure.[3] Weights were trimmed at the 2nd and 98th percentile to prevent individual interviews from having too much influence on the final results. After trimming, there continue to be some individual respondents who have a fairly high weight. There are 50 individuals who have a weight factor that is over 5. These individuals could, therefore, exert particular influence on results. We chose not to trim further because doing so would pull data further off of the benchmarks and final data would be less reflective of the expected population of Muslims in Illinois.

Tables 2 compares unweighted and weighted sample distributions to target population benchmark distributions.

Table 2: Weighting Dimensions, Benchmarks, Unweighted and Weighted Distribution

Dimension	Value Label	Benchmark	Unweighted	Weighted
Sex	Male	60.6%	45.6%	58.1%
	Female	39.4%	54.4%	41.9%
Age	18-35	53.7%	27.7%	50.0%
	36-55	32.1%	51.1%	34.5%
	56+	14.2%	21.2%	15.5%
Education	High school grad or less	34.3%	8.4%	30.9%
	Some college	29.0%	11.6%	29.1%
	College grad	36.7%	80.0%	39.9%
Voter Registration	Registered	71.2%	94.1%	77.5%
	Not Registered	28.8%	5.9%	22.5%

Margin of Sampling Error

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. SSRS 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. The total sample design effect for this survey is 2.89.

SSRS calculates the composite design effect for a sample of size n, with each case having a weight, w, as:[4]

$$deff = \frac{n\sum w^2}{\left(\sum w\right)^2}$$

[3] To handle missing data among some of the demographic variables we employ a technique called hot decking. Hot deck imputation replaces the missing values of a respondent randomly with another similar respondent without missing data. These are further determined by variables predictive of non-response that are present in the entire file. We use an SPSS macro detailed in 'Goodbye, Listwise Deletion: Presenting Hot Deck Imputation as an Easy and Effective Tool for Handing Missing Data' (Myers, 2011).

[4] Kish, L. (1992). Weighting for Unequal Pi. Journal of Official Statistics, Vol. 8, No.2, 1992, pp. 183-200.

Margins of sampling error are calculated to provide a reasonable range for the error that may exist in an estimate due to random sampling fluctuations. Margins of sampling error are meaningful only if it can be assumed that selection into the sample is random and that each unit's probability of being sampled would remain the same if the sample were repeated many times. These assumptions are less realistic for non-probability online samples than for probability-based samples, because we cannot observe or control the factors that determine whether a given unit is included in a non-probability online sample. We provide estimated margins of error here to provide a general assessment of error ranges that may be associated with the data, given the sample size. However, margins of error for non-probability online samples should always be interpreted with caution, as the underlying assumptions cannot be verified.

The survey's margin of error is the largest 95% confidence interval for any estimated proportion based on the total sample — the one around 50%. For example, the margin of error for the entire sample is ± 4.0 percentage points. This means that in 95 out of every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 4.0 percentage points away from the currently reported estimate. Margins of error for subgroups will be larger.

It is important to remember that the sampling fluctuations captured in the margin of error are only one possible source of error in a survey estimate. Other sources, such as respondent selection bias, questionnaire wording, and reporting inaccuracy, may contribute additional error of greater or lesser magnitude.