Survey Data Analysis

Walter Leite, PhD
Assistant Professor
Research Evaluation & Methodology

What are sampling weights?

- Sampling weights are the number of individuals in the population each respondent in the sample is representing.
- A sample weight is the inverse of the probability of selection.
- For example, if my simple random sample is one tenth of the population size (i.e. my sampling fraction is 1/10), then each respondent in the sample is representing 10 people in the population.

Weights compensate for:

- Unequal probability of selection
- Unequal response rates
- Post-stratification (adjust the sample distribution for key variables of interest such as age, ethnicity, sex, to make it conform to a known population distribution)

How do weights work?

Score	Weight
4	1
2	2
1	4
5	1
2	2

Simple mean:

$$\frac{(4+2+1+5+2)}{5} = 2.8$$

Weighted mean:

$$\frac{(4x1)+(2x2)+(1x4)+(5x1)+(2x2)}{10} = 2.1$$

Weights are frequencies of each observation in the population.

Types of Weights

- Raw weights
- Relative or normalized weights
- Design effect adjusted weights

Raw weights or base weights

- Raw weights sum to the population size. They are the inverse of the probability of selection:

$$w_i = \frac{1}{p_i}$$

- For example, if the probability of selection of a unit is 1/50, its raw weight is 50.

Raw weights for multi-stage sampling

- The raw weight is the inverse of the product of the probabilities of selection at each stage.

$$w_i = \frac{1}{p_{1i} \times p_{2i}}$$

p_{1i} = probability of selection at stage 1

p_{2i} = probability of selection at stage 2

- For example, if the probability of selection at stage 1 (schools) is 1/5, and the probability of selection at stage 2 (teachers) is 1/20, the final probability of selection is 1/100 and the raw weight is 100.

Primary and secondary sampling units

- In multi-stage sampling, the units that are sampled at the first stage are the primary sampling units (PSU) (example: schools)

- The units sampled at the second stage are the secondary sampling units (SSU) (example: teachers)

Problems of using raw weights

- Estimates of means, proportions and standard errors obtained using raw weights will be based on the population size, not the sample size. The means and proportion estimates will be correct, but the test statistics will have too much power.

- Solution: Convert raw weights to normalized weights.

Normalized or relative weights

- Normalized weights sum to the sample size.
- With normalized weights in the analyses, the estimates of means, and proportions are correct. The estimates of standard errors are correct given a simple random sample or stratified sample.
- When a cluster or multi-stage sample is used, the estimation of standard errors will not be correct using only case weights. Special procedures such as Taylor-series approximation, bootstrapping or design effects need to be used to obtain correct standard errors.

Converting a raw weight to a normalized weight

- There are two ways of converting a raw weight to a normalized way:

1. Dividing the raw weights by the mean of the raw weights:

$$w_N = \frac{w_i}{\bar{w}}$$

2. Multiplying the raw weight by the overall sampling fraction:

$$w_N = w_i \frac{n}{N}$$

Adjusting for unit non-response

- Step 3: Calculate the weight for non-response, which is the inverse of the subgroup response rates.

$$w_{nr} = \frac{S_s}{n_s}$$

Number of sampled cases for the subgroup.

Number of responses obtained for the subgroup

- Calculate the non-response adjusted weight: It is the product of the original weight and the weight for non-response.

$$w_a = w_i \times w_{nr}$$