The What, Where, Why, and How of Weighting Your Data

Do I Really Have to Do This?

Sylvia Dohrmann

Westat, Inc.

December 13, 2002

What are Weights?

- Weights are values which measure the number of people in the population a particular sampled case represents.
- Each sampled case will have a weight attached to it.
- The weights will be used in estimating survey items.

Why Weight?

- Allows us to produce estimates we would have obtained if the entire population were surveyed.
- Accounts for nonresponse to the survey.
- These features make weighted estimates more accurate and convincing when making statements about your reference population.

How are Weights Used?

- Example:

	Unweighted Value	Weight (People)	Weighted Value
	35	50	1,750
	65	20	1,300
	80	10	800
	45	40	1,800
Total	225	120	5,650
Mean	56.3		47.1
Standard Error	10.1		8.3

What do I Have to do Now?

- Not Much! Westat will perform the weighting.
- However, we need information from you to get the job done.

Weighting Steps

- Calculate base weights.
- Adjust for the people that did not respond to the survey (nonresponse adjustment).
- Adjust for under/over coverage of the survey as compared to auxiliary data (Poststratification or Raking).

Base Weights

- The base weight is the starting point (or "basis") for the weight calculations.
- It is calculated as the inverse of the selection probability (the probability of the case coming into the sample).
- In POMP, each sampled case will most likely have the same base weight.

Base Weights

- Example:
  - 100 people were selected out of 3,000, all with an equal chance of selection.
  - Selection probability for each case is $\frac{100}{3,000}$.
  - Weight for each case is $\frac{3,000}{100}=30$ people.

Base Weights

- In order to calculate the base weights, we need detailed information on how your sample was selected:
  - Total number of clients in the reference population.
  - Total number of clients selected.
  - How you selected the sample.

Nonresponse Adjustment

- If every person sampled agreed to do the survey, then weighted estimates using just the base weights would be sufficient to estimate the population values.
- However, every survey has some level of nonresponse.

Nonresponse Adjustment

- We can adjust for this nonresponse by spreading the weight of the nonrespondents to the respondents.
- The best defense against this is to make efforts to get people to respond.

Nonresponse Adjustment Steps

1. Determine what characteristics are related to response propensity.
   - These are commonly things like race, sex, age and possibly other things related to the subject matter of the survey. In the case of POMP, how long someone has been a client may be used.
   - Note that these must be items you have for responders and nonresponders, perhaps from the sample frame (i.e., client list).

Nonresponse Adjustment Steps

2. Separate the data into groups (or cells) defined by these characteristics.

3. Inflate the weights in each cell by the factor:

$$\frac{\text{Total weight of all cases in the cell}}{\text{Total weight of responders in the cell}}$$

Nonresponse Adjustment

- In order to perform this adjustment, we need the following:
  - The status of each interview:
    - Complete, partially complete
    - Refusal
    - Unable to contact
    - Ineligible
  - Demographic characteristics for *all* members of the sample including refusals, no contacts, and ineligibles.

Using Auxiliary Data

- A final adjustment can be made to bring the sample weights of who you did contact up to the level of the population.
- This is done when there are accurate population totals from a source other than the sample frame.

Using Auxiliary Data

- This is a way to account for under- or over-coverage of certain demographic groups.
- For example, we frequently under-cover black males 18-24. If we have good estimates of the population for this group, we can adjust the weighted total for the group up to the amount in the auxiliary data.

Summary

- For the weighting calculations, please provide Westat with the following:
  - Total number of clients in the reference population.
  - Total number of clients selected.
  - How you selected the sample.
  - The status of each interview.
  - Demographic characteristics for *all* members of the sample.