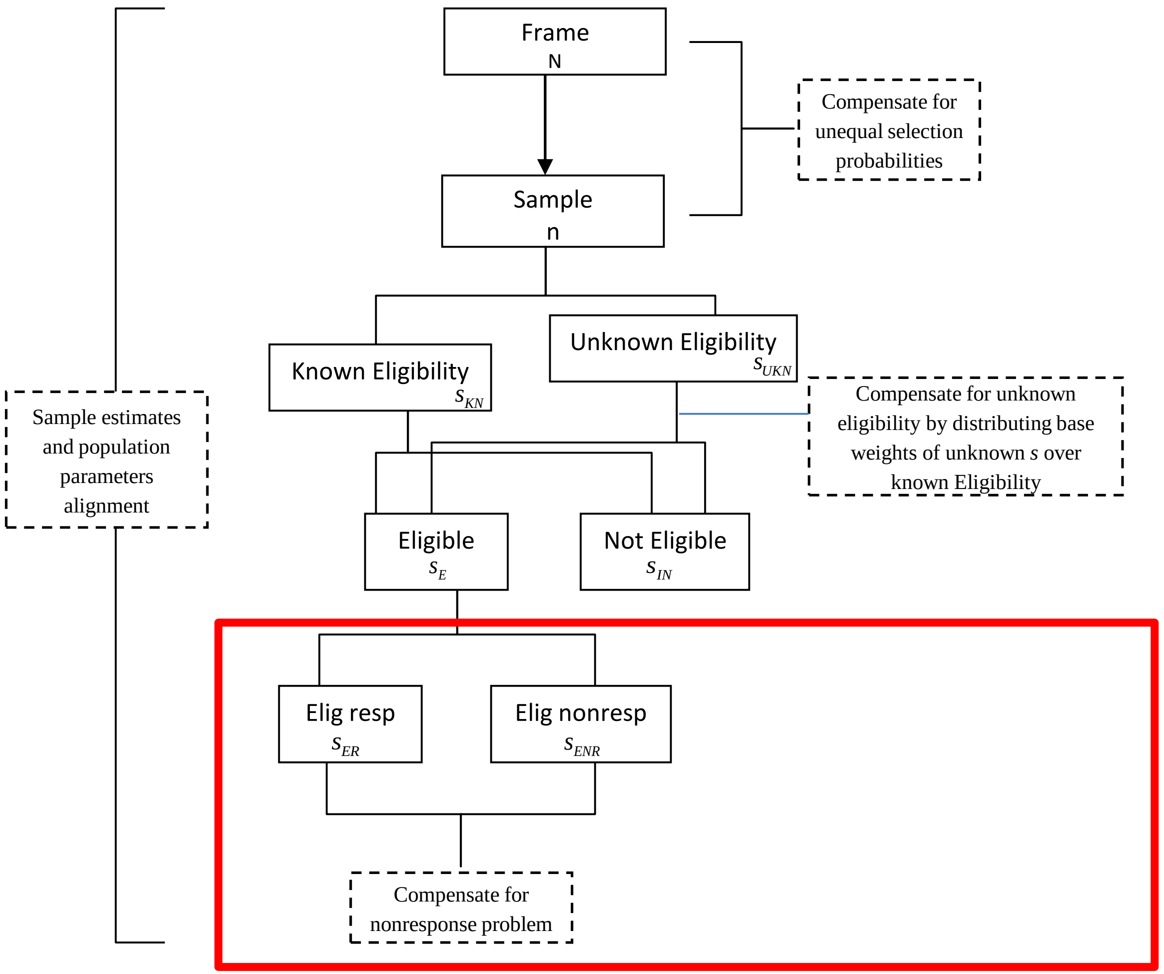
Video 18 of 21: Weighting Nonresponse adjustment

Sampling

Nonresponse weighting (I)

- In an ideal survey, all the units in the population are in the sample frame and all those in the sample participate in the survey
  - In practice, however, neither of these conditions occurs
- Some units are not included in the frame (undercoverage) and some of the sampled units do not respond (nonresponse)
- Nonresponse weighting is generally applied to reduce the bias cause when some sampled units do not response (unit nonresponse)
- The respondents set *r* can be seen as a subset of the sample *s*

General steps used in weighting: Nonresponse weighting



Nonresponse weighting (II)

- Define the following sampling and response indicators, respectively:

$$I_i = \begin{cases} 1, & \text{if unit } i \text{ selected in the sample} \\ 0, & \text{otherwise} \end{cases}$$

$$R_i = \begin{cases} 1, & \text{if unit } i \text{ responds} \\ 0, & \text{otherwise} \end{cases}$$

- The probability of being in the sample *s* is $P(I_i = 1) = \pi_i$
- The probability of responding (being in respondents set *r*) given that unit *i* is in the sample *s* is $P(R_i = 1 | I_i = 1) = \phi_i$

Nonresponse weighting (III)

- Inverse of the selection probabilities, π_i^{-1}, to adjust for the different selection probabilities
- Use the inverse of the responding probabilities, ϕ_i^{-1}, to adjust for the non-response
  - The final weight would then be $(\pi_i \times \phi_i)^{-1}$
- However, it is not possible to use the actual responding probabilities ϕ_i.
  - Estimates for the responding probabilities, $\hat{\phi}_i$, are used instead
- The type of method use to estimate these response propensities will depend on the amount and type of auxiliary variables available for both *respondents* **AND** *non-respondents*, and the missingness assumptions
- For missing data terminology, see Little & Rubin (2019)

Nonresponse weighting (IV)

- In the absence of any auxiliary variable, the simplest approach is to assume a Missing Completely at Random (MCAR) mechanism, in which the response propensities can be estimated by the overall response rate, $\hat{\phi}_i = RR$

Nonresponse weighting (V)
Class-based adjustment

- Class-based adjustment assumes that we can create classes where either all units have about the same probability of response or about the same *y* values -- *Missing at Random* (MAR) mechanism
  - The estimated responding probabilities $\hat{\phi}_i$ can be modeled if we have a set of auxiliary variables $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ available for each sample unit whether it responds or not
- Auxiliary variables could be:
  - Demographic variables like gender, age, race/ethnicity, and education,
  - Design or frame variables, like region
  - Variables observed during data collection (Paradata)
- Ideally, these covariates should be related to both the *response propensity* **AND** the *y*'s being measured

Nonresponse weighting (VI)
Class-based adjustment

- The nonresponse adjustment factor for units in class *c* can be computed using the inverse of the unweighted response rate:

$$a_{4,c} = \frac{n_{c,E}}{n_{c,ER}}$$

- Or the inverse of the weighted response rate:

$$a_{4,c} = \frac{\sum_{i \epsilon s_{c,E}} d_{3i}}{\sum_{i \epsilon s_{c,ER}} d_{3i}}$$

- The non-response weight is then the product of the design weight and the non-response adjustment factor:

$$d_{4i} = d_{3i} \times a_{4i}$$

Nonresponse weighting (VII)
Class-based adjustment example

- Consider regions as classes for the non-response class-based adjustment and the following response distribution:

Region	$n_{c,ER}$	$n_{c,ENR}$	$n_{c,E} = n_{c,ER} + n_{c,ENR}$	$a_{4,c} = n_{c,E}/n_{c,ER}$
West	60	38	98	98/60 = 1.6333
Midwest	73	34	107	107/73 = 1.4658
Northeast	98	63	161	161/98 = 1.6429
South	57	42	99	99/57 = 1.7368

Nonresponse weighting (VIII)
Propensity score adjustment

- Model nonresponse using logistic regression, probit or complementary log-log models
- The response indicator R_i works as a dependent variable, and the available auxiliary variables (for respondents and non-respondents) work as the independent variables
- The estimated response probability can be written as

$$\hat{\phi}(\boldsymbol{x}_i) = \frac{exp\left(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}\right)}{1 + exp\left(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}\right)}$$

- where $\boldsymbol{x}_i = \left(x_{i1}, \dots, x_{ip}\right)$ is the vector of independent variables and $\widehat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p\right)$ are the estimated logistic regression coefficients

Nonresponse weighting (IX)
Propensity score adjustment

- After estimating the response probabilities, we can choose between:
  - Propensity weighting: Using the inverse $\hat{\phi}_i$ directly as an adjustment for the weight
    - This adds more dependency on the response propensity model
  - Propensity stratification: Using $\hat{\phi}_i$ to create adjustment classes (Little, 1986):
    - After estimating the response propensities $\hat{\phi}(\boldsymbol{x}_i)$, sort the file in ascending order by $\hat{\phi}(\boldsymbol{x}_i)$
    - Then, form classes with about same number of initial (respondents and non-respondents) sample units in each
    - Breaking the response propensities $\hat{\phi}(\boldsymbol{x}_i)$ variable by the quintiles or deciles could be a good option as a grouping technique

Nonresponse weighting (X)
Propensity score adjustment

- Propensity stratification: Using $\hat{\phi}_i$ to create adjustment classes (cont.):
  - After creating the classes, there are several options for computing a single adjustment in each class *c*:
    - unweighted average estimated propensity: $\hat{\phi}_i = \sum_{i \epsilon s_c} \hat{\phi}(\boldsymbol{x}_i)/n_c$; where n_c is the unweighted number of cases in class *c*
    - weighted average estimated propensity: $\hat{\phi}_i = \sum_{i \epsilon s_c} d_i \hat{\phi}(\boldsymbol{x}_i) / \sum_{i \epsilon s_c} d_i$; where d_i is the input weight to the NR step and $\sum_{i \epsilon s_c} d_i = \hat{N}_c$, the estimated number of population units in class *c*
    - unweighted response rate: $\hat{\phi}_i = n_{cR}/n_c$; where n_{cR} is the unweighted number of respondents in class *c*
    - weighted estimate of response rate: $\hat{\phi}_i = \sum_{i \epsilon s_{cR}} d_i / \sum_{i \epsilon s_c} d_i$
    - unweighted median estimated propensity: $\hat{\phi}_i = median\left[\hat{\phi}(\boldsymbol{x}_i)\right]_{i \in s_c}$

END OF VIDEO 18