Nonresponse reduction and adjustment techniques

Raphael Nishimura
raphaeln@umich.edu

Outline

- Nonresponse & Nonresponse bias
- Weighting adjustments
  - Nonresponse weighting
  - Calibration
- Imputation
- Other adjustment methods

Nonresponse definition

- **Unit nonresponse**:
  - Failure to obtain survey measurements on sampled units
  - Example:

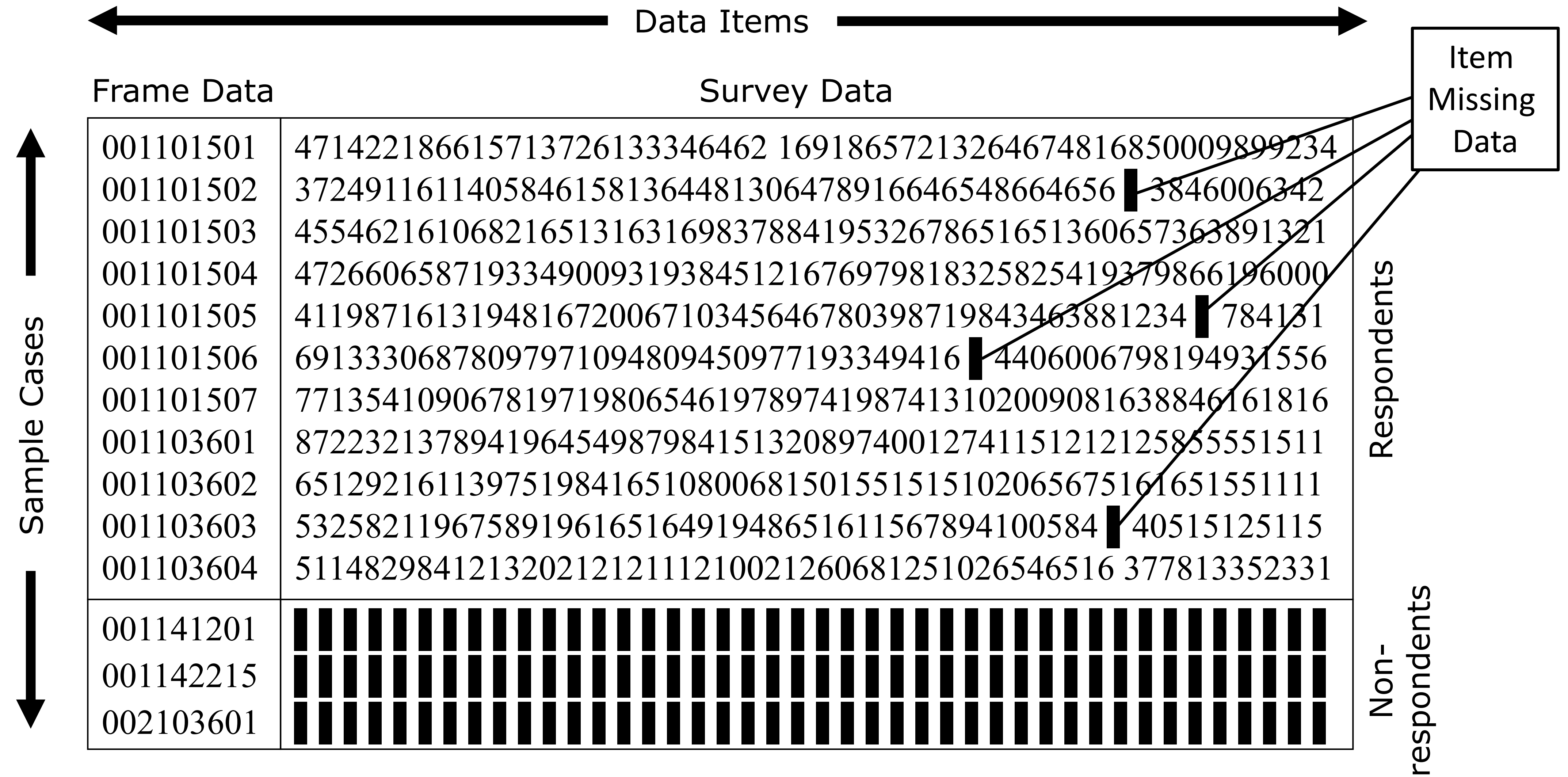
    Sample Person: "*I never participate in surveys. Please don't call me again.*"

- **Item nonresponse**:
  - Failure to obtain information for particular items in the survey
  - Example:

    Interviewer: "*What was your total income last year?*"

    Respondent: "*I don't know, my wife keeps those records*"

Unit and Item Nonresponse



Groves, et al. 2009, *Survey Methodology* Figure 2.3

Nonresponse bias

- Function of response rates and survey variables (Y)
- Specific to a statistic, not a survey
- Unadjusted estimate based only on the respondents ($\bar{y}_r$) is different from the estimate based on the entire sample (respondents + nonrespondents)
- What do you do about it?
  - Reduce it
    - Survey design and data collection protocols
    - "There is only one real cure for nonresponse and that is getting the response." (Benjamin King, 1996)
  - Adjust for it
    - Postsurvey adjustments (weighting and imputation)

Two theoretical perspectives to nonresponse bias

- **Deterministic perspective**
  - Assumes there are two types of people:
    - Those who will never respond to the survey
    - Those who will always respond to the survey

- **Stochastic perspective** :
  - Assumes that people have a response propensity, that is, a probability that they will respond to the survey (p_i)
  - As a result, even if same people in the sample, outcome might be different from across different survey realizations (under the same survey conditions)

Nonresponse bias: Deterministic perspective

$$Bias(\bar{y}_r) = \frac{M}{N}\left(\bar{Y}_R - \bar{Y}_{NR}\right)$$

Nonresponse rate

Difference between respondents and nonrespondents

$\bar{y}_r$: Unadjusted respondent mean
M: Number of nonrespondents in the population
N: Population size
Y_R: Respondent's population mean
Y_{NR}: Nonrespondent's population mean

Nonresponse bias: Stochastic perspective

$$Bias(\bar{y}_r) = \frac{\sigma_{py}}{\bar{p}}$$

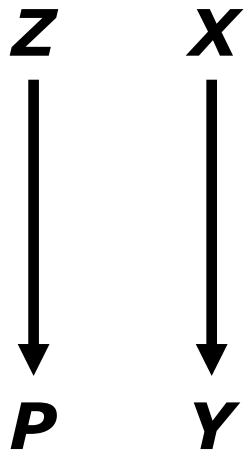
Covariance between response propensity and survey variable

Average response propensity

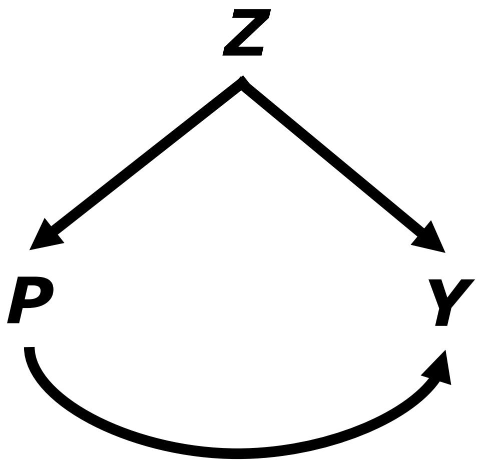
Response Rates vs. Nonresponse Bias

- Higher response rate is NO guarantee for lower nonresponse bias
- Example: Interested in measuring proportion of people who have anxiety symptoms
  - Survey protocol 1:
    - Response Rate: 75% (Nonresponse rate = $\frac{M}{N} = 0.25$)
    - Respondent proportion: $\bar{Y}_R = 0.10$
    - Nonrespondent proportion: $\bar{Y}_{NR} = 0.14$
    - Nonresponse bias: $Bias(\bar{y}_r) = 0.25 \times (0.10 - 0.14) = -0.01$
  - Survey protocol 2:
    - Response Rate: 90% (Nonresponse rate: $\frac{M}{N} = 0.10$)
    - Respondent proportion: $\bar{Y}_R = 0.10$
    - Nonrespondent proportion: $\bar{Y}_{NR} = 0.40$
    - Nonresponse bias: $Bias(\bar{y}_r) = 0.10 \times (0.10 - 0.40) = -0.03$

Causal models for nonresponse error (Groves, 2016)



Separate Cause Model



Common Cause Model

Survey Variable Cause Model

Missing Data Literature (Little and Rubin, 2019):

Missing Completely At Random (MCAR)

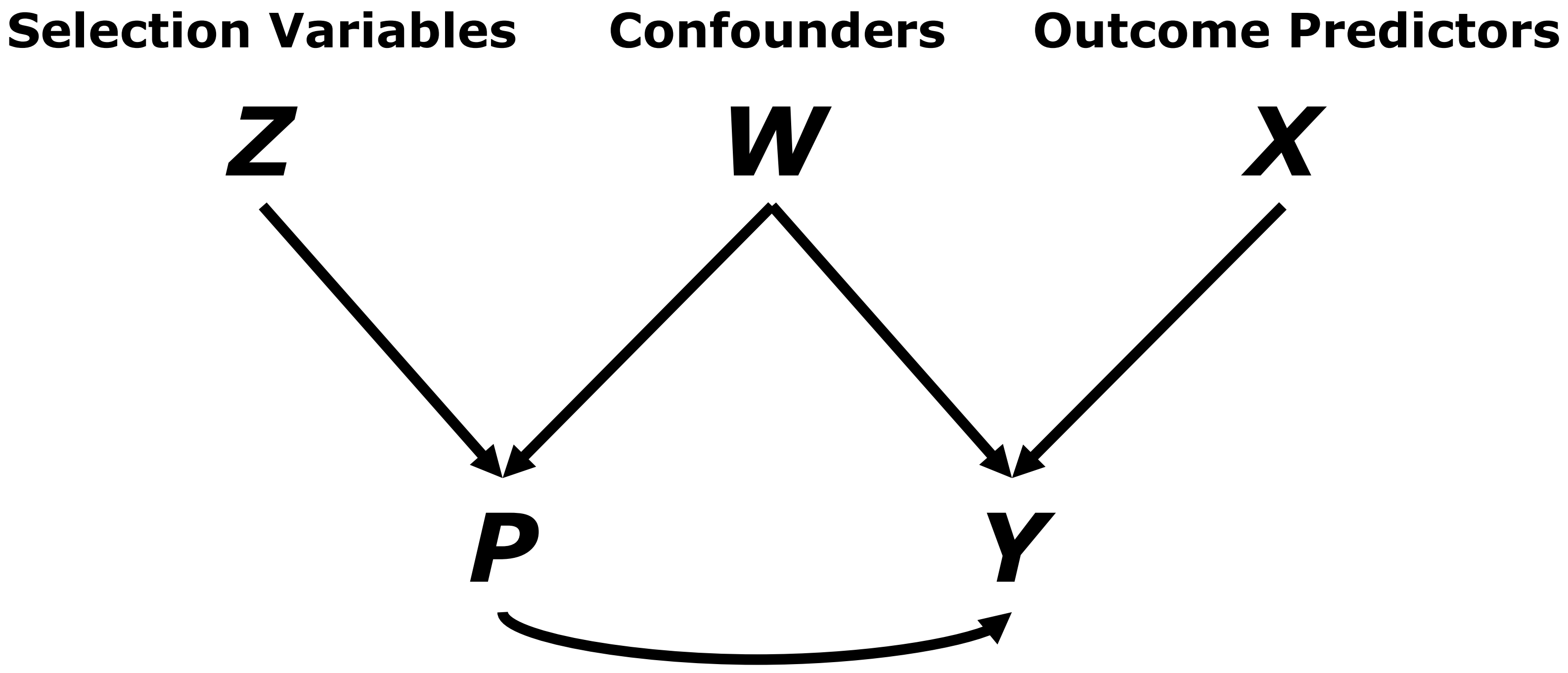
Missing At Random (MAR)

Missing Not At Random (MNAR)

Missing mechanisms/assumptions

- Separate Cause Model/Missing Completely at Random
  - Assumes that missingness is completely unrelated to survey outcome
  - Respondents are a Simple Random Sample of the sample
  - Very strong assumption, i.e., hardly realistic
- Common Cause Model/Missing at Random
  - Assumes that conditional to a set of auxiliary variables, missingness is unrelated to survey outcomes
  - Within sub-groups of the auxiliary variables, respondents are a random sample of the sample
- Survey Variable Cause Model/Missing Not at Random
  - Assumes that missingness is directly related to survey outcome

More general framework (Li et al, 2022)



Postsurvey Adjustments

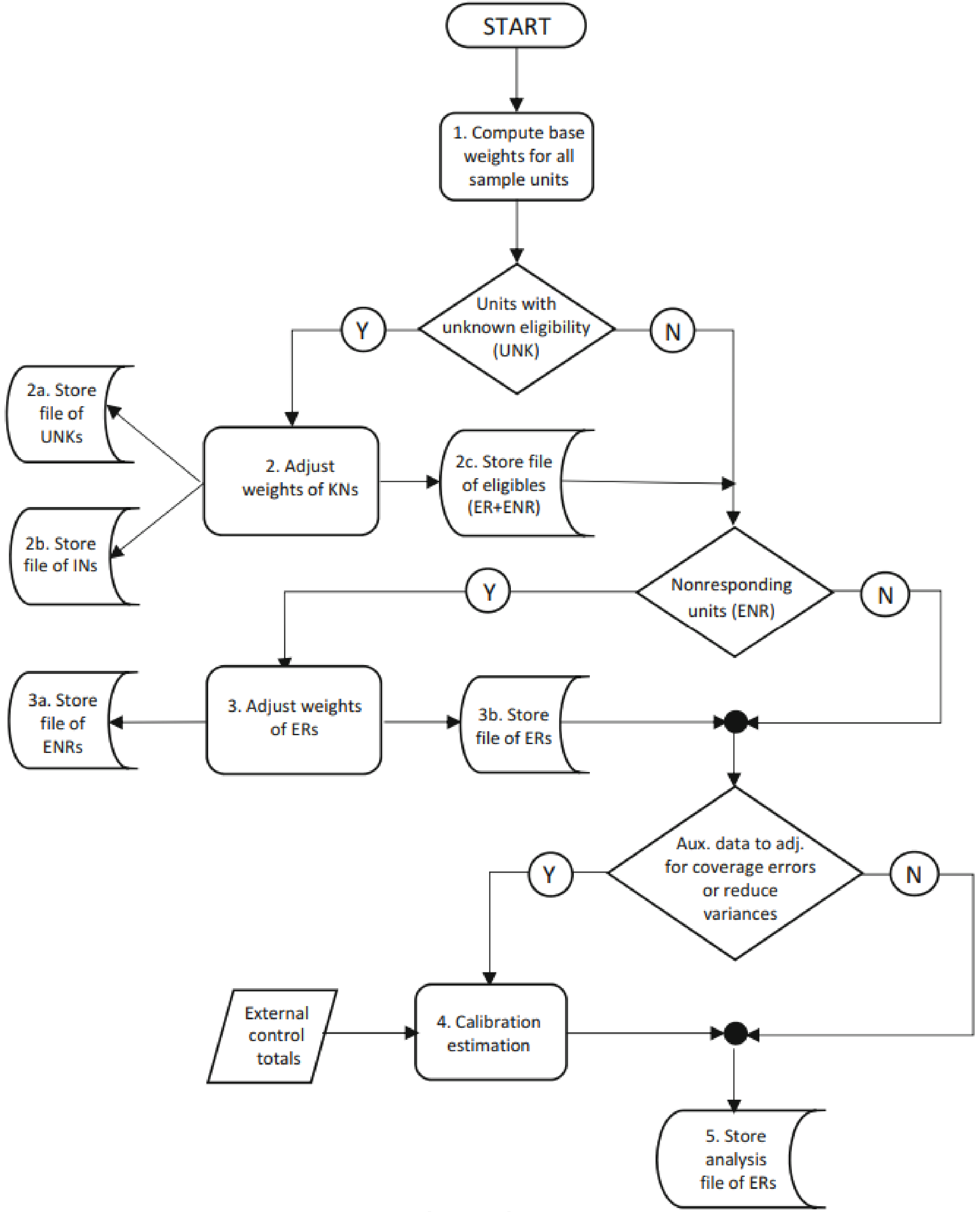
- Statistical adjustments done to address:
  - Unequal selection probabilities
  - Nonresponse
  - Issues with the coverage of the frame
  - Missing data

Survey weighting (unequal selection probabilities, nonresponse, issues with the coverage of the frame)

Data imputation (missing data)

Survey weighting

- In general, weights are designed to
  - Compensate for unequal selection probabilities
    - Base/design weights, adjustment for multiplicity and within household selection, frame integration
  - Adjust for unknown eligibility
  - Adjust for non-sampling errors
    - Nonresponse and coverage issues
  - Incorporate external data to improve the precision of the survey estimates and further mitigate for non-sampling errors
- Basic weighting steps are usually similar for different types of data collection modes



Source: Valliant et al. (2018)

Nonresponse weighting

- 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

- 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, we do not know what are 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 and the **missingness assumptions** (MCAR, MAR, MNAR)

Auxiliary data

- Auxiliary variables could be:
  - Demographic variables like gender, age, race/ethnicity, and education (in the frame or in administrative records)
  - Design or frame variables, like region
  - Variables observed during the data collection process (paradata)
- For nonresponse adjustment, we need such auxiliary variables observed for both *respondents* **AND** *non-respondents*

Nonresponse weighting: No auxiliary data

- 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: 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 $\boldsymbol{x}_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)$ available for each sample unit whether it responds or not
- Response propensities $\hat{\phi}_i$ estimated as inverse of response rate of each class

Nonresponse weighting: Class-based adjustment

R's	Male	Female	Total
0-19	10	15	25
20-65	24	36	60
66+	5	10	15
Total	39	61	100

Sample	Male	Female	Total
0-19	36	32	68
20-65	64	64	128
66+	13	17	30
Total	113	113	226

r_h n_h

Nonresponse weighting: Class-based adjustment

Response Rates	Male	Female
0-19	10/36=0.28	15/32=0.47
20-65	24/64=0.38	36/64=0.56
66+	5/13=0.38	10/17=0.59

Weights	Male	Female
0-19	36/10=3.6	32/15=2.1
20-65	64/24=2.7	64/36=1.8
66+	13/5=2.6	17/10=1.7

$$r_h / n_h$$

$$\hat{\phi}_h = n_h / r_h$$

- All respondents in the same cell receive the same nonresponse adjustment
- These weights are multiplied by the base/design weights

Nonresponse weighting: Propensity score adjustment

- What if we have both qualitative and quantitative variables, and maybe not interested in a fully saturated model?
- More general approach: fit model to predict response propensity
- 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
- Under a logistic regression, the estimated response propensity 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 coefficients

Nonresponse weighting: Propensity score adjustment

- After estimating the response probabilities, we can choose between:
  - Propensity weighting: Using the inverse directly as an adjustment for the weight
    - This adds more dependency on the response propensity model
  - Propensity stratification: Using to create adjustment classes (Little, 1986):
    - After estimating the response propensities $\hat{\phi}$, 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: Propensity score adjustment

- Propensity stratification: Using $\hat{\phi}_i$ to create adjustment classes:
  - After creating the classes, there are several options for computing a single adjustment in each class c:
    - unweighted average estimated propensity
    - weighted average estimated propensity
    - unweighted response rate
    - weighted estimate of response rate
    - unweighted median estimated propensity

Nonresponse weighting: classification algorithms

- Objective is to classify elements as respondents or nonrespondents based on covariates available for all sample cases
- Input data is the same as for propensity score adjustment
- Classification algorithms:
  - Classification and Regression Tree (CART)
  - Chi-squared Automatic Interaction Detection (CHAID)
  - Random Forests
- Advantages over propensity score adjustment:
  - Variable selection (main and interaction effects) done automatically
  - Continuous variables are categorized automatically

What variables to use for nonresponse adjustment?

Table 1

Effect of Weighting Adjustments on Bias and Variance of a Mean, by Strength of Association of the Adjustment Cell Variables with Nonresponse and Outcome

	Association with outcome	
Association with nonresponse	Low	High
Low	Cell 1 Bias: --- Var: ---	Cell 3 Bias: --- Var: ↓
High	Cell 2 Bias: --- Var: ↑	Cell 4 Bias: ↓ Var: ↓

Source: Little & Varitivarian (2005)

Calibration

- The general idea is to use auxiliary variables to improve the efficiency of estimators (Deville & Särndal, 1992)
  - Create weights such that weighted estimates matches the population with respect to auxiliary variables
- **Auxiliary data**: information available for the entire frame or target population, either for each individual population unit or in aggregate form
  - May come from the frame, administrative records, published statistics or other sources
  - Large, high-quality surveys: counts of persons in groups defined by age, race/ethnicity, and gender may be published from a census or from population projections that are treated as highly accurate
  - Unlike in nonresponse adjustments, for calibration we do not need for the nonrespondents

Calibration

- Auxiliary variables are typically included in survey questionnaire:
  - Make sure to include variables that are predictive of the survey outcomes
- Among the potential benefits of calibration are:
  - Decrease in sampling variances
  - Bias correction for frame coverage and other frame errors
  - Adjustment for nonresponse
- Most common calibration methods:
  - Post-stratification
  - Raking
  - Generalized Regression (GREG)

Post-stratification

- Implemented within weighting classes formed by crossing all categories of the qualitative variables
- For example, post-stratified by gender and age:
  - Gender: Male, Female
  - Age: 0-19, 20-65, 66+
- Cross-classification leads to $G = 6$ classes that could be used as post-strata
- Using many important auxiliary variables for post-stratification can reduce bias and improve precision
- But may result in empty weighting classes or ones with a small number of respondent cases
  - Results in unstable estimates of the population controls and adds unnecessarily variability to the final weights

Post-stratification

Respondents	Male	Female	Total
0-19	10	15	25
20-65	24	36	60
66+	5	10	15
Total	39	61	100

Population	Male	Female	Total
0-19	15	14	29
20-65	29	30	59
66+	5	7	12
Total	49	51	100

$$r_h / r \qquad N_h / N$$

Post-stratification

Respondents	Male	Female	Total
0-19	10	15	25
20-65	24	36	60
66+	5	10	15
Total	39	61	100

Population	Male	Female	Total
0-19	15	14	29
20-65	29	30	59
66+	5	7	12
Total	49	51	100

Create weights to make the distribution of respondents match the population distribution

- To make the distribution of males aged 20-65 match the population, create a weight that is the ratio of the population percentage to the respondent population: 29/24
- For Females aged 20-65, use: 30/36

Post-stratification

Weights	Male	Female
0-19	15/10 = 1.5	14/15 = 0.933
20-65	29/24 = 1.208	30/36 = 0.833
66+	5/5 = 1.0	7/10 = 0.70

- All respondents in the same cell receive the same poststratification adjustment
- These weights are multiplied by the base/design weights or nonresponse adjusted weights

Raking

- Also known as Iterative Proportional Fitting
- Raking is an adjustment procedure in which estimates are controlled by marginal population totals
  - Implicitly assumes that the interactions between the calibration variables are not important to explain the survey variable(s)
- The main advantage of raking over post-stratification is that raking potentially allows the use of more auxiliary information (only needs the marginal totals not the cross-classified categories totals)
- Also decreases issues of sparse or empty weighting cells
- In the next example, with raking only the marginal sex and age control counts are needed

Raking

Respondents	Male	Female	Total
0-19	10	15	25
20-65	24	36	60
66+	5	10	15
Total	39	61	100

Population	Male	Female	Total
0-19	?	?	29
20-65	?	?	59
66+	?	?	12
Total	49	51	100

Raking

- Iterative process:
  - Create weights to match the distribution of one variable
  - Recalculate the cell counts and create new totals
  - Then, create weights to match the distribution of the other variable, using the new cell counts
  - Iterate through this procedure, alternating between the two variables until weights and marginal counts stabilize
- Also known as iterative proportional fitting or multiplicative weighting
  - A log-linear model in which you want the complete cross-classification and are given marginal distributions

Raking: First iteration

Respondents	Male	Female	Total
0-19	10	15	25
20-65	24	36	60
66+	5	10	15
Total	39	61	100

Raking 1	Male	Female	Total
0-19	12.564	12.541	25.105
20-65	30.154	30.098	60.252
66+	6.282	8.361	14.643
Total	49	51	100

- Create first set of weights, matching on gender distribution:
  - Male: 49/39 = 1.25641
  - Female: 51/61 = 0.83607
- Multiply internal cells by these weights and calculate new totals
  - E.g., Male 0-19: 10*1.25641=12.564
  - Female 0-19: 15*0.83607=12.541
  - Etc.

Raking: Second iteration

Respondents	Male	Female	Total
0-19	12.564	12.541	25.105
20-65	30.154	30.098	60.252
66+	6.282	8.361	14.643
Total	49	51	100

Raking 2	Male	Female	Total
0-19	12.512	12.488	25
20-65	30.028	29.972	60
66+	6.435	8.565	15
Total	48.974	51.026	100

- Create second set of weights:
  - 0-19: 25/25.105 = 0.99581
  - 20-65: 60/60.252 = 0.99581
  - 66+: 15/14.643 = 1.02440
- Multiply internal cells from first iteration by these weights and calculate new totals:
  - E.g., Male 0-19: 12.564*0.99581=12.512
  - Female 0-19: 12.541*0.99581=12.488
  - Etc.

Raking: Keep going

- Continue the iterations, alternating between gender and age distributions until the weights and cell counts stop changing or until you hit an acceptable amount of change at some predetermined decimals place (e.g., change only at the 5th decimal)
- Our example stabilized at the 6th iteration

Raking: Final distribution and final weights

Final counts	Male	Female	Total
0-19	14.459	14.541	29
20-65	29.416	29.584	59
66+	5.126	6.874	12
Total	49	51	100

Final raking weights	Male	Female	Total
0-19	1.446	0.969	--
20-65	1.226	0.822	--
66+	1.025	0.687	--
Total	--	--	--

- Final distribution is the product of each of the iterations
- Final weight is the product of each of the weights from each iteration

Post-stratification vs. Raking

Final PS weights	Male	Female
0-19	1.500	0.933
20-65	1.208	0.833
66+	1.000	0.700

Final raking weights	Male	Female
0-19	1.446	0.969
20-65	1.226	0.822
66+	1.025	0.687

- Poststratification assumes an interaction effect
- Raking assumes two marginal effects, but no interaction effect
- The stronger the interaction effect, the more the weights created by raking and by poststratification will differ

When does calibration work?

- Similar to take-away message from Little & Varitivarian (2005) for nonresponse weighting adjustment:
  - If auxiliary variables are correlated with survey outcomes:
    → improve estimates' precision (decrease standard errors)
  - If auxiliary variables are correlated with both the survey outcomes and nonresponse:
    → improve estimates' precision (decrease standard errors)
    → reduce non-sampling bias

Software

- Different calibration methods are available in the main statistical software:
  - R
    - survey package (Lumley, 2020): calibrate, postStratify, rake
    - sampling package (Tillé & Matei, 2021): calib, poststrata, ratioest, regest
  - Stata
    - svycal command (Valliant & Dever, 2018)
  - SAS
    - rake_and_trim macro (Izrael et al, 2017)
- Weights should be accounted in both point and sampling variance estimation (Taylor series or replication methods)

Weighting non-probability samples

- Three approaches, all involve modelling (Elliott & Valliant, 2017):
  - Quasi-randomization weighting
    - Combine non-probability sample with a reference probability sample that covers the entire target population
    - Estimate pseudo-inclusion probabilities based on auxiliary variables, similar to propensity score adjustment
    - Use inverse of pseudo-selection probabilities as weights
  - Superpopulation modelling
    - Fit model for survey outcome *y* based on auxiliary variables
    - Weights are based on model
    - Use model weights for estimation
    - Ultimately, equivalent to calibration
  - Doubly robust

Imputation

- "Fix" item missing data
  - Don't knows
  - Refusals
  - Interviewer errors
- As with unit nonresponse, there are deterministic approaches and stochastic approaches
  - However, these correspond to the actual act of filling in the missing data, rather than differences in theoretical conceptualization of the mechanism

Types of imputations

- Complete Case Analysis
- Mean value imputation
  - Mean imputation with random residual
- Hot deck imputation (Andridge & Little, 2010)
  - Sequential
  - Hierarchical
- Regression imputation
  - Deterministic and stochastic
- Sequential regression imputation (Raghunathan et al., 2011; Van Buuren et al., 2011)

Mean value imputation

- Replace missing values with mean for the variable
  - Equivalent to complete case analysis
  - Distorts distribution, with “spike” at one value
  - Consider also mode or median values
- Class mean replaces values with mean for classes
  - Reduces distortion to distribution
- Prefer stochastic element: add random residual to class mean value
  - Select residual from empirical distribution, or theoretical distribution

Sequential hot deck imputation

- Widely used method
  - Sort data by predictors of value to be imputed
  - Retain last non-missing value, and impute to next missing value
  - Draw non-missing value randomly within imputation class
- Problems
  - Multiple donation
  - Weak within group correlations
- Sort boundaries (see illustration)

i	Gender	Educ	Reported family income	Hot Value	Imputation flag	Final value
1	M	9	23	*51*	0	23
4	M	11		*23*	1	23
2	M	12		*23*	1	23
3	M	12	43	*23*	0	43
7	M	12	35	*43*	0	35
8	M	12	42	*35*	0	42
5	M	16	75	*42*	0	75
6	M	16	88	*75*	0	88
16	F	10		*88*	1	88
15	F	12	28	*88*	0	28
17	F	12	31	*28*	0	31
18	F	12	35	*31*	0	35
19	F	12	30	*35*	0	30
22	F	12		*30*	1	30
13	F	14	67	*30*	0	67
14	F	15	56	*67*	0	56
21	F	15	72	*56*	0	72
20	F	18	66	*72*	0	66

Hierarchical hot deck

- Flexible match
  - Group respondents and non-respondents into classes
  - Select donor at random within class
  - Collapse classes to achieve match
- Improved donor and recipient match
- Reduced multiple donations

Hierarchical hot deck

Gender	Education					
	<12		12		>12	
	R	M	R	M	R	M
Male	23	[…]	43 35 42	[…]	75 88	
Female		[…]	28 31 35	[…]	67 56 66	

Regression imputation

- Generalization of hot deck methods
- Deterministic and stochastic versions
- Estimate model among respondents:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi} + \varepsilon_i$$

- Predict values for missing using estimated coefficients
- Add a random residual…
  - Select value from
  - Select respondent and compute deviation between observed and predicted as residual $N(0, \sigma^2)$

Sequential Regression Imputation

- Extension of regression imputation
  - Most survey data sets don't have only normally distributed variables
    - Continuous, categorical, count, etc.
  - Most survey data sets have multiple variables that need to be imputed
- Sequential regression
  - Do a number of regressions to impute missing data, adding stochastic residuals where appropriate
  - Start with the variable with the least amount of missing data
    - Use an appropriate regression form (e.g., logistic regression for binary data)
  - Use imputed values from that variable to impute the next variable
  - Keep going until you have gone through the entire data set
  - When all of the variables have been imputed, start over, retaining the previously imputed values
  - Cycle through the data set at least 5 times

Multiple imputation

- Imputation adds variance to estimates
- May be viewed as two-phase sampling
- Under-estimate variances using imputed data sets, because of larger sample sizes
- Multiply impute (Rubin, 1987)
  - Create multiple data sets (at least 5), each imputed separately
  - This is a Bayesian procedure, so you select values from the posterior distribution
  - Imputing multiple times allows you to adequately account for the variance from imputation

Variance estimation

- Impute m times
- Compute estimate $\bar{y}_\gamma$ for each imputed data set and compute

$$\bar{y} = \frac{1}{m}\sum_{\gamma=1}^{m} \bar{y}_\gamma$$

- The variance is a sum of within and between replicate variance:

$$\text{var}(\bar{y}) = \frac{1}{m}\sum_{\gamma=1}^{m} \text{var}(\bar{y}_\gamma) + \left(\frac{m+1}{m}\right)\left(\frac{1}{m-1}\right)\sum_{\gamma=1}^{m}(\bar{y}_\gamma - \bar{y})^2$$

Illustration

Estimate	Imputation			
	1	2	3	Total
$\bar{y}_\gamma$	47.5	45.7	46.7	46.6
$\text{var}(\bar{y}_\gamma)$	21.5	23.5	24.1	23.0
$\left(\frac{m+1}{m}\right)\left(\frac{1}{m-1}\right)(\bar{y}_\gamma - \bar{y})^2$	0.751	0.803	0.001	1.037
Overall	--	--	--	24.04

Software

- Various routines for single and multiple imputation, and analyzing multiple imputed datasets:
  - R
    - Packages: mi, mice, smcfcs, hot.deck
  - Stata
    - mi command
  - SAS
    - PROC MI and PROC MIANALYZE

What if nonresponse is nonignorable (NMAR)?

- Specific statistical models with very strong (and in most cases untestable) assumption about the distribution of the nonrespondents
  - Heckman selection model (Heckman, 1976)
  - Pattern-mixture models (Little, 1993; Andridge & Little, 2011)
- In most cases, not possible to adjust the estimates
  - Sensitivity analysis to assess potential nonresponse bias according to various scenarios
  - Measures of bias (Andridge et al., 2019; Little et al., 2020)

Summary

- Nonresponse adjustment methods:
  - Unit nonresponse → Survey weighting
  - Item nonresponse → Data imputation
- Nonresponse adjustments rely on assumptions
  - Typically, it is assumed a Missing at Random (MAR) mechanism
  - Hard to impossible to assess validity of these assumptions in practice

Summary

- A variety of weighting and imputation methods
  - More important than the which method is what auxiliary variables to use in these adjustments
  - Preferably use variables correlated with survey outcomes
    - Reduce both bias and variance
  - Make sure to plan in advance and collect such types of variables in the survey
- Both weighting and imputation can impact the variance of the estimates
  - Ideally use methods that accounts for that:
    - Weighting → Replicated methods (Bootstrap, Jacknife, BRR, etc)
    - Imputation → Multiple imputation

Recommended textbooks

- **Missing data:**
  - Little, R. J., & Rubin, D. B. (2019). Statistical Analysis with Missing Data. John Wiley & Sons.
- **Weighting:**
  - Valliant, R., Dever, J. A., & Kreuter, F. (2018). Practical tools for designing and weighting survey samples. New York: Springer.
  - Valliant, R., & Dever, J. A. (2018). Survey weights: a step-by-step guide to calculation. College Station, TX: Stata Press.
- **Imputation:**
  - Carpenter, J. R., Bartlett, J., Morris, T., Wood, A., Quartagno, M., & Kenward, M. G. (2023). Multiple imputation and its application. John Wiley & Sons
  - Van Buuren, S. (2018). Flexible imputation of missing data. CRC press. https://stefvanbuuren.name/fimd/
- **Survey Data Analysis:**
  - Heeringa, S., West, B., Berglund, P. (2017). Applied Survey Data Analysis, 2nd edition. Boca Raton, FL: CRC Press.

References

- Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. International statistical review, 78(1), 40-64.
- Andridge, R. R., & Little, R. J. (2011). Proxy pattern-mixture analysis for survey nonresponse. Journal of Official Statistics, 27(2), 153.
- Andridge, R. R., West, B. T., Little, R. J., Boonstra, P. S., & Alvarado-Leiton, F. (2019). Indices of non-ignorable selection bias for proportions estimated from non-probability samples. Journal of the Royal Statistical Society. Series C, Applied statistics, 68(5), 1465.
- Battaglia, M., D., Izrael, and S. Ball. (2017). Tips and tricks for raking survey data with advanced weight trimming. SESUG Paper SD-62-2017.
- Boonstra, P. S., Little, R. J., West, B. T., Andridge, R. R., & Alvarado-Leiton, F. (2021). A simulation study of diagnostics for selection bias. Journal of official statistics, 37(3), 751-769.
- Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. Journal of the American statistical Association, 87(418), 376-382.
- Elliott, M. R. & Valliant, R. (2017). Inference for Nonprobability Samples. Statistical Science 32 (2) 249 – 264.

References

- Frankel, M., & King, B. (1996). A conversation with Leslie Kish. Statistical Science, 11(1), 65-87.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey Methodology. John Wiley & Sons.
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In Annals of economic and social measurement, volume 5, number 4 (pp. 475-492). NBER.
- Li, Y., Irimata, K. E., He, Y., & Parker, J. (2022). Variable inclusion strategies through directed acyclic graphs to adjust health surveys subject to selection bias for producing national estimates. Journal of official statistics, 38(3), 875-900.
- Little, R. J. (1986). Survey nonresponse adjustments for estimates of means. International Statistical Review/Revue Internationale de Statistique, 139-157.
- Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association, 88(421), 125-134.
- Little, R. J., & Rubin, D. B. (2019). Statistical Analysis with Missing Data. John Wiley & Sons.

References

- Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means?. Survey Methodology, 31(2), 161.
- Little, R. J., West, B. T., Boonstra, P. S., & Hu, J. (2020). Measures of the degree of departure from ignorable sample selection. Journal of survey statistics and methodology, 8(5), 932-964.
- Lumley, T. (2020) survey: analysis of complex survey samples. R package version 4.0.
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey methodology, 27(1), 85-96.
- Rubin, D.B. (1987), Multiple Imputation for Nonresponse in Surveys, New York: John Wiley & Sons, Inc.
- Tillé, Y. and Matei, A. (2021). sampling: Survey Sampling. R package version 2.9. https://CRAN.R-project.org/package=sampling
- Valliant, R., Dever, J. A., & Kreuter, F. (2018). Practical tools for designing and weighting survey samples. New York: Springer.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of statistical software, 45, 1-67.

Thank you!

raphaeln@umich.edu