Bayesian inference for sample surveys

Roderick Little

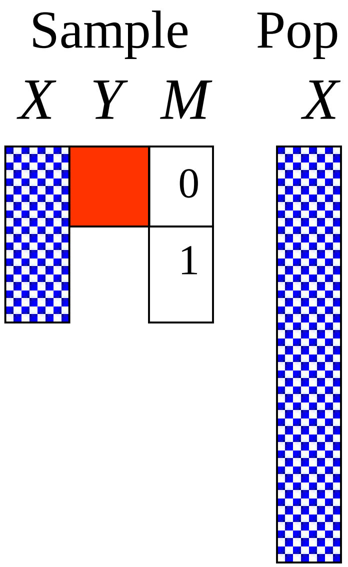
Trivellore Raghunathan

Module 15: missing data 2 -- unit nonresponse

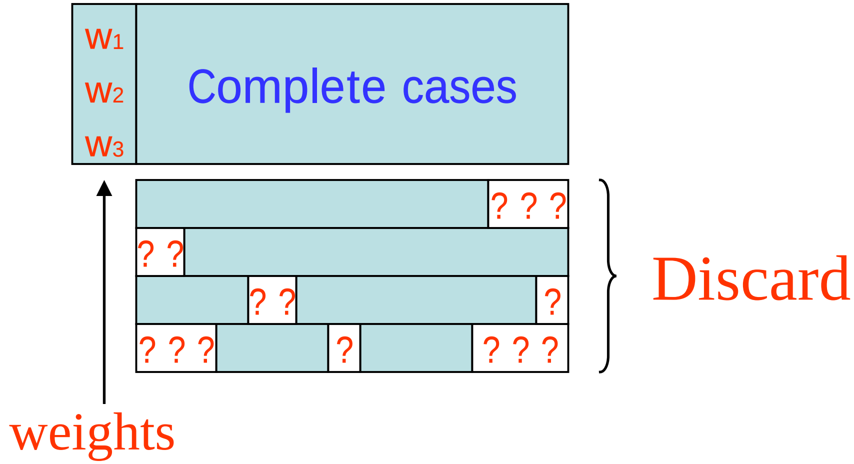
University of Michigan

Unit nonresponse

- Predict nonrespondents by regression on observed survey variables *X*
- For bias reduction, predictors *X* should be related to *M* <u>and</u> outcome *Y*
- In particular, consider categorical predictor *X* and flat priors, for inference about mean:
  - Bayes corresponds to weighting by inverse of estimated response rate in each category



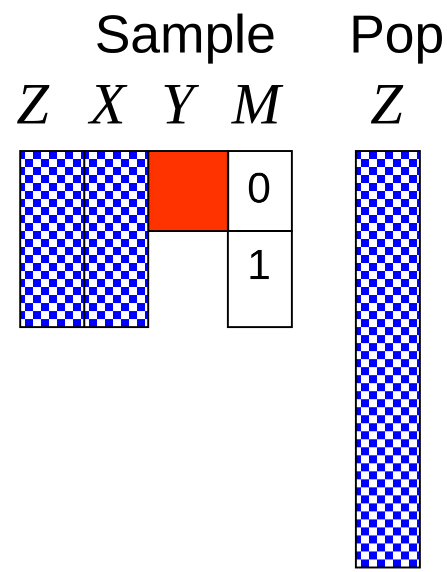
Design-based approach: weighting



- One way to reduce the potential bias of CC analysis is to weight respondents differentially e.g. a CC mean becomes a weighted mean
- Common for unit nonresponse in surveys
- "Quasi-randomization inference" : extends ideas of randomization inference in surveys

Unit nonresponse

- Suppose we have:
- Design variables *Z* observed for whole population
- Fully observed survey variables *X*, measured for respondents and nonrespondents
- Survey variables *Y* measured only for respondents (see diagram)
- Goal of weighting is to use information in *X* and *Z* to weight the respondents, improve estimates



Sampling weights

- In a probability survey, each sampled unit *i* "represents" w_i units of the population, where

$$w_i = \frac{1}{\text{Pr(unit i sampled)}}$$

w_i is determined by the sample design and hence *known*

- Extend this idea to unit nonresponse ...

Unit Nonresponse Weights

- If probability of response was known, could obtain weight for units that are sampled and respond:

$$w_i = \frac{1}{\text{Pr(unit i is sampled and responds)}}$$

$$= \frac{1}{\text{Pr(i sampled)}} \times \frac{1}{\text{Pr(i responds|sampled)}}$$

$$= \text{(sampling weight)} \times \text{(response weight)}$$

Since prob of response is not known, we need to estimate it.

Adjustment Cell method

- Group respondents and nonrespondents into adjustment cells with similar values on variables recorded for both:
- e.g. white females aged 25-35 living in SW

100 in sample < 80 respondents / 20 nonrespondents

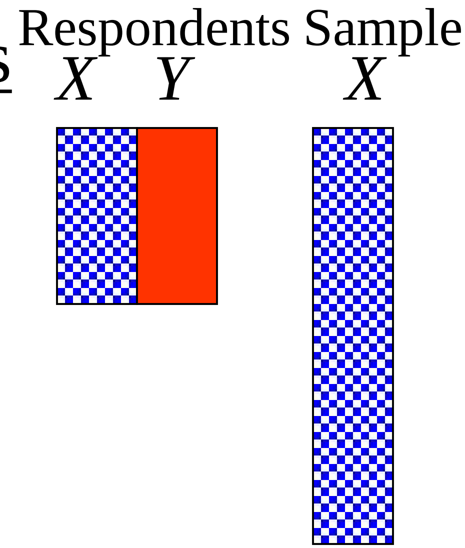
pr(response in cell) = 0.8

response weight = 1.25

Ex 1: nonresponse with categorical Z

- Population divided into J <u>adjustment cells</u>
- X is set of indicators, known for sample:

$$x_i = \begin{cases} 1, & \text{if unit } i \text{ is in cell } j; \\ 0, & \text{otherwise.} \end{cases}$$



- Simple random sampling

  of n_j units selected from cell j, r_j respond.

- Assume MAR: M and Y are independent given X.

Bayes inference for adjustment cell model

- Model for Y given X:

$$[y_i \mid x_i = j] \sim_{\text{ind}} N(\theta_j, \sigma_j^2)$$

- For simplicity assume σ_j^2 is known and the flat prior:

$$p(\theta_j \mid X) \propto const.$$

- Standard Bayesian calculations lead to

$$[\overline{Y} \mid Y_{\text{inc}}, X, \{\sigma_j^2\}] \sim N(\overline{y}_w, \sigma_w^2)$$

where:

$$\overline{y}_w = \sum_{j=1}^{J} p_j \overline{y}_j,\ p_j = n_j / n,\ \overline{y}_j = \text{ respondent mean in cell } j$$

Equivalent to weighting n_j / r_j, inverse of response rate in cell j

Bayes inference for adjustment cell model

- If X is a known for population (not just sample) posterior mean weights respondents by N_j/r_j (post)-stratification)
- Unlike stratified mean, the counts r_j are not under control of sampler, and weights can be very large if r_j is small
- Hence may want to put a proper prior on adjustment cell means, to pull in extreme weights
- Posterior variance accounts for uncertainty in cell proportions

Impact of weighting for nonresponse

$\mathrm{corr}^2(X,Y)$

$\mathrm{corr}^2(X,M)$

	Low	High
Low	---	var $\Downarrow$
High	var $\uparrow$	var $\Downarrow$ bias $\Downarrow$

Too often adjustments do this?

- Standard "rule of thumb" $\mathrm{Var}(\bar{y}_{\mathrm{w}}) = \mathrm{Var}(\bar{y}_{\mathrm{u}})(1+\mathrm{cv}(w))$ fails to reflect that nonresponse weighting can <u>reduce</u> variance
- Little & Vartivarian (2005) propose refinements

Weight the response rates?

- Nonresponse weights are often computed using units weighted by their sampling weights $\{w_{1i}\}$

$$w_{2j}^{-1} = \left(\sum_{r_i=1, x_i=j} w_{1i}\right) / \left(\sum_{r_i=1, x_i=j} w_{1i} + \sum_{r_i=0, x_i=j} w_{1i}\right)$$

  – Gives unbiased estimate of response rate in each adjustment cell defined by X
  – Not correct from a prediction perspective

Arguments against weighting response weights

- Unnecessary if cells are created properly!
  - Adjustment cells should be homogeneous with respect to response propensity
  - In that case, weighting the nonresponse rates is unnecessary and adds variance to estimates
- Doesn’t work if cells are created improperly!
  - If adjustment cells are <u>not</u> homogeneous with respect to response propensity, then weighting RR’s does <u>not</u> yield unbiased estimates survey estimands.
- The right approach is to create adjustment cells based on classification of the observed variables <u>and</u> the survey design variables.
  - Then weighting is unnecessary

Simulation Study

- Simulations in Little & Vartivarian (2003 Statistics in Medicine) consider the variance and bias of estimators of weighted and unweighted rates and alternative estimators, under a variety of population structures and nonresponse mechanisms.
- Binary outcome Y, stratum Z, adjustment cell X to avoid distributional assumptions such as normality.
- 25 populations to cover the factor space

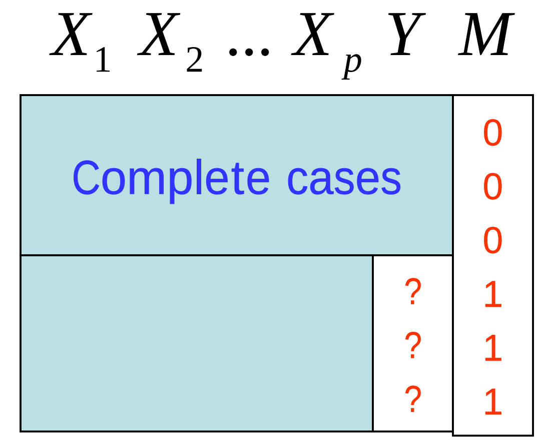
Simulation Results

- ML for the model used to generate the data is always best or close to best.
- Unweighted-RR(x, z) is best overall: form cells based on *X* and *Z*
- Additive model theoretically biased when the data-generating model includes *XZ* interaction, but in these simulations the bias is modest.
- Unweighted-RR (x) is biased when both *Y* and *R* depend on *Z*.
- Weighted RR (x) does not generally correct the bias in these situations: similar to Unweighted-RR(x) overall

Remarks

- Don't weight response rates! Rather
  - Condition on design variables when creating adjustment cells
- Too many strata? Response propensity stratification
- To improve efficiency: weight shrinkage by multilevel modeling

Response propensity stratification



- X = covariates observed for respondents and nonrespondents, Y missing
- M = missing-data indicator
  - nonrespondent = 1, respondent = 0
- (A) Regress M on X (probit or logistic), using respondent and nonrespondent data $\hat{p}(M=0 \mid X) = \text{propensity score}$
- (B1) Weight respondents by inverse of propensity score from (A), $1/\hat{p}(M=0 \mid X)$, or:
- (B2) form adjustments cells by categorizing $1/\hat{p}(M=0 \mid X)$
- Note that this method is only effective if propensity is also related to the outcome

Construction of Weights

- One should think about missing data at the design stage and collect information predictive of participation and key variables of interest
- Common sources for data
  - Administrative data from which the sample has been drawn
    - Medicare enrollment files
    - Driver license file
    - A large study may be used to sample subjects for a supplement or a smaller study.; Two-stage, nested case-control study etc.

- Data collected during the study conduct phase ("Paradata")
  - Statements made by sampled subjects to the interviewer
  - Interviewer observations
- Data at Neighborhood level
  - Block or Block-group level characteristics

- Example
  - Assets and Health Dynamics (AHEAD) Study
  - Primary outcome variable: 5 point Self-reported health status
  - Interviewer noted the following four variables
    - Time delay statements
    - Negative statements about participation in study
    - Positive statements
    - Statements about being Old to participate etc

- Subject level variable
  - Age and Sex
- Neighborhood level variables
  - Large urban area (yes/no)
  - Barriers to contact (yes/no)
  - Block: Persons per sq-mile
  - Block : % persons 70 or older
  - Block: % Minority populations
  - Block: % Multi-unit structures (10+)
  - Block: % Occupied Housing Units
  - Block: % Single person Hus
  - Block: % vacant Hus
  - Block: Persons per occupied Hu
- Sample size: n=10,173, respondents=8,212 (80.7%)

Comparison of “Paradata” between respondents and nonrespondents

Variable	Respondent	Nonrespondent	Effect Size
Age	77.41 (6.81)	77.83 (6.12)	0.065
Sex (% Female)	63.3%	63.6%	0.006
Mention age or illness	10.2%	16.1%	0.175
Negative Statement	9.5%	37.7%	0.703
Positive Statement	11.6%	2%	0.388
Time delay statement	8.1%	13.8%	0.183

$Effect\ size$

$$D = \frac{|\bar{x}_R - \bar{x}_{NR}|}{\sqrt{(s_R^2 + s_{NR}^2)/2}}$$

For proportions, $s^2 = p(1-p)$

$Small : D \leq 0.25$

$Medium : 0.25 < D \leq 0.5$

$Large : 0.5 < D \leq 0.75$

$Very\ Large : D > 0.75$

Comparison of Neighborhood level data between respondents and nonrespondents

Variable	Respondents	Nonrespondents	Effect size
% from Large urban areas	15.4%	22.7%	0.19
% with Barriers to contact	12.7%	20.0%	0.20
Block: Persons per sq-mile	6162.20 (14318.64	7750.92 (15086.18)	0.11
Block : % persons 70 or older	10.98 (8.18)	11.16 (7.96)	0.02
Block: % Minority populations	22.76 (30.16)	24.60 (31.94)	0.06
Block: % Multi-unit structures (10+)	10.96 (20.24)	12.65 (21.74)	0.08
Block: % Occupied Housing Units	33.50 (23.76)	35.43 (24.68)	0.08
Block: % Single person HUs	24.48 (11.54)	25.17 (11.75)	0.06
Block: % vacant HUs	9.05 (9.18)	8.82 (8.90)	0.03
Block: Persons per occupied Hu	2.64 (0.47)	2.62 (0.48)	0.03

Building a Response Propensity Model

- Goal is to obtain a well fitting (logistic or probit) model that balances the covariates between respondents and nonrespondents

$$e(x) = \text{Propensity score}$$

$$X \perp R \,|\, e(x)$$

- One useful measure of goodness of fit: Hosmer-Lemeshow test
  - Create deciles based on the estimated propensity score , compare the observed and expected frequencies /counts in these 10 classes
  - Chisquare statistics with 9 degrees of freedom as a measure of lack of it

- Checking the balancing property
  - Once satisfied with the logistic or probit model create classes (or strata or groups) based on quartiles or quintiles or deciles (depending upon your sample size) of the estimated propensity score.
  - In each class compute the effect size as we did for the overall sample.
  - Good balancing is indicated by the effect size being very small in each class
- Developing the propensity score model and checking its balancing property is an iterative process

- One can start with the main effect model and check the Hosmer-Lemeshow chisquare statistic
- Add two factor interactions to reduce the value of chisquare statistic
- Use stepwise or other selection procedure to reduce the model complexity
- Add higher order interactions if necessary
- Transform the covariates or standardized the continuous covariates (i.e. subtract the overall mean and divide by the overall standard deviation)

AHEAD example

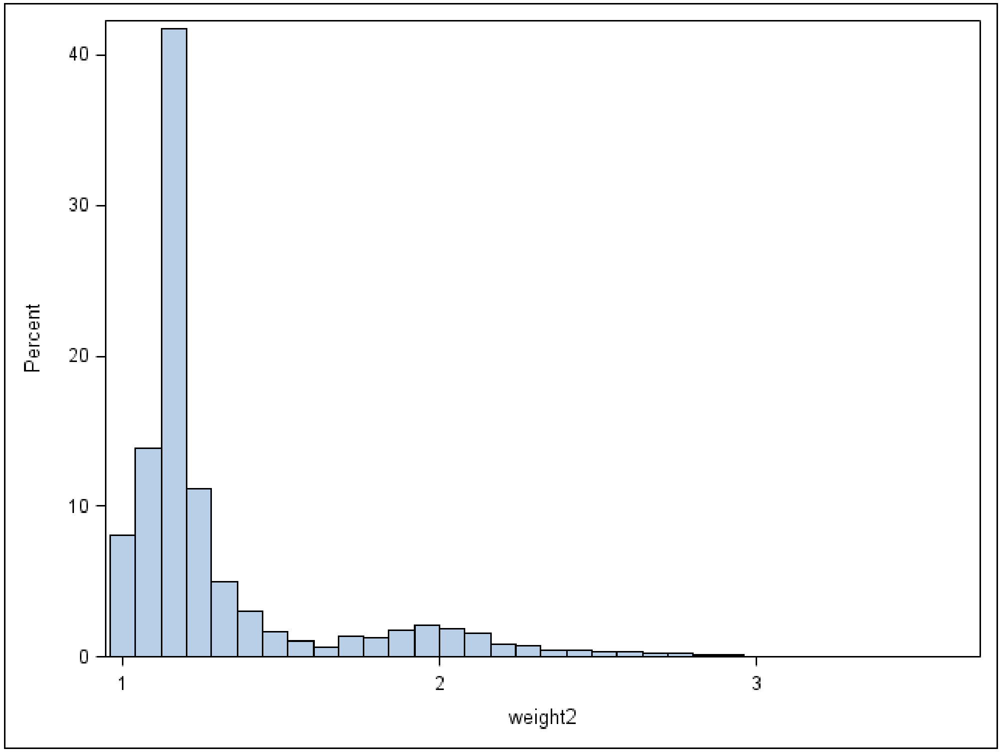
- First model with 16 main effects results in Hosmer-Lemeshow chisquare statistic of 28.68 with the p-value of 0.0004
- Next tried the model with all two factor interactions which reduced the chisquare statistic as 17.33 with the p-value of 0.03.
- As is typical, the full interaction model may be overmatching and result in poor fit.
- Next tried stepwise selection with different entry and exit probabilities. Finally with 0.5 as the probability for both entry and exit, obtained the model with Chisquare statistic of 4.75 with the p-value 0.784

Checking the Balance

- Created 4 groups (strata or classes) based on the quartiles of the estimated propensity squares.
- Computed the effect sizes for the 16 variables in each class. All were smaller than 0.05 indicating good balance.
- Propensity score stratification weights

Group	Respondents	Nonrespondents	Weight
1	1505	1037	1.6890
2	2100	445	1.2119
3	2237	306	1.1368
4	2370	173	1.0730

- Inverse probability weighting : weight=1/estimated propensity score
- Histogram of the nonresponse adjustment weight



Weighted analysis

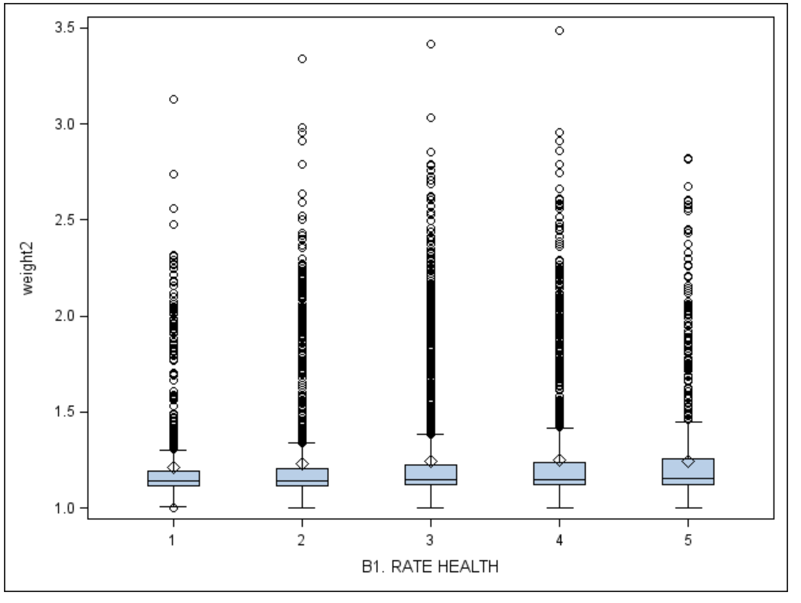
- Survey analysis software need to be used to correctly estimate the standard error. SAS procedures surveymeans, surveyfeq, surveyreg and surveylogistic are a few examples in SAS.
- In STATA, use svy commands
- In R, Thomas Lumley has developed packages and can be downloaded from the r library.

Weighted and Unweighted frequencies

Self-rated health	Unweighted	Propensity score Class weighted	inverse probability weighted
1	10.75	10.51 (0.34)	10.52 (0.34)
2	22.80	22.53 (0.47)	22.61 (0.47)
3	30.35	30.41 (0.52)	30.50 (0.52)
4	23.08	23.29 (0.48)	23.31 (0.48)
5	13.03	13.26 (0.38)	13.06 (0.38)

Weighted and Unweighted analysis

- The effect of weighting will depend upon the correlation between weight and the survey variable of interest



- Weighting is not that effective

Inference from Weighted Data

- Role of weights in analytical inference (regression, factor analysis, …) is controversial
- Can use packages for computing standard errors for complex sample designs -- but often these do not take into account sampling uncertainty in weights
- Bootstrap/Jackknife of weighting procedure propagates uncertainty in weights – but weights need to be recalculated on each BS/JK sample

Penalized Spline of Propensity Prediction (PSPP)

- PSPP (Little & An 2004, Zhang & Little 2009, 2011).
- Regression imputation that is
  - Non-parametric (spline) on the propensity to respond
  - Parametric on other covariates
- Exploits the key property of the propensity score that conditional on the propensity score and assuming missing at random, missingness of Y does not depend on other covariates
- This property leads to a form of double robustness.
- Similar to previous PSPP for continuous stratifier, with propensity for selection replaced by estimated propensity to respond

PSPP method

Estimate: Y^{*}=logit (Pr(M=0|$X_1,\ldots,X_p$))

Impute using the regression model:

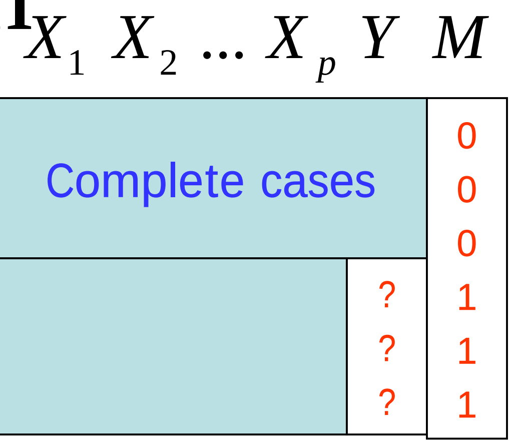
$$(Y \mid Y^{*}, X_1, \ldots, X_p; \beta) \sim$$

$$N(s(Y^{*}) + g(Y^{*}, X_2, \ldots, X_p; \beta), \sigma^2)$$

- Nonparametric part
- Need to be correctly specified
- We choose penalized spline

- Parametric part
- Misspecification does not lead to bias
- Increases precision
- X_1 excluded to prevent multi-collinearity

Alternative method: Predictive mean stratification



- X = covariates observed for respondents and nonrespondents
- Y = outcome with missing data
- (A) Regress Y on X (linear, other as appropriate) using respondent data only
- (B) Form adjustments cells with similar values of predictions from the regression $\hat{Y}(X)$

- This method has potential to reduce both bias and variance
- Note that the adjustment cells depend on outcome, so this method yields different cells for each outcome, hence is more complex with many outcomes with missing values

Conclusion

- Unit nonresponse: standard design-based approach is to weight by inverse of estimated response propensity
- Bayes for adjustment cells yields a posterior mean that is equivalent to weighting
- More generally, Bayes can use estimated propensity to respond as a predictor, as in PSPP