Intro to Nonresponse

Week 8 (8.1-8.4)

Stat 260, St. Clair

Quantifying nonresponse bias

Subpopulation	Size	Mean
Respondents	N_R	$\bar{y}_{RU}$
Nonrespondents	N_M	$\bar{y}_{MU}$
Everyone	$N = N_R + N_M$	$\bar{y}_U = \frac{N_R}{N}\bar{y}_{RU} + \frac{N_M}{N}\bar{y}_{MU}$

If estimate $\bar{y}_R$ is approx. unbiased est. of $\bar{y}_{RU}$, then nonresponse bias:

Bias: $E(\bar{y}_R) - \bar{y}_U \approx \bar{y}_{RU} - \left(\frac{N_R}{N}\bar{y}_{RU} + \frac{N_M}{N}\bar{y}_{MU}\right)$

$= \bar{y}_{RU} - \left(1 - \frac{N_M}{N}\right)\bar{y}_{RU} - \frac{N_M}{N}\bar{y}_{MU} = \frac{N_M}{N}\left(\bar{y}_{RU} - \bar{y}_{MU}\right)$

* No Nonresponse bias if $\bar{y}_{RU} = \bar{y}_{MU}$

↳ NR is not related to y

Quantifying nonresponse bias

Subpopulation	Size	Mean
Respondents	$N_R/N = 0.36$	$\bar{y}_{RU} = 36$
Nonrespondents	$N_M/N = 0.64$	$\bar{y}_{MU} = 50$

Bias $= (.64)(36 - 50) = -9$

If we use Respondents to est overall mean, the bias is -9 $\Rightarrow$ est. too low

$\Rightarrow \bar{y}_u = 36 + 9 = 45$

Types of nonresponse

Mechanism model:

$$R_i = \begin{cases} 1 & \text{if unit } i \text{ responds} \\ 0 & \text{if unit } i \text{ does not respond} \end{cases}$$

The probability of response is ("phi")

$$\phi_i = P(R_i = 1)$$

Variables:

- Response of interest y_i (only observed for respondents)
- Covariates x_i (known for both respondents and nonrespondents)
  - e.g. stratification variables

Types of nonresponse: MCAR

Missing Completely at Random (MCAR)

- ϕ_i does not depend on y_i or x_i (or design) $\Rightarrow$ $\phi = P(R_i = 1)$
- Nonresponse only lowers sample size
- No nonresponse bias since $\bar{y}_{R\mathcal{U}} = \bar{y}_{M\mathcal{U}}$

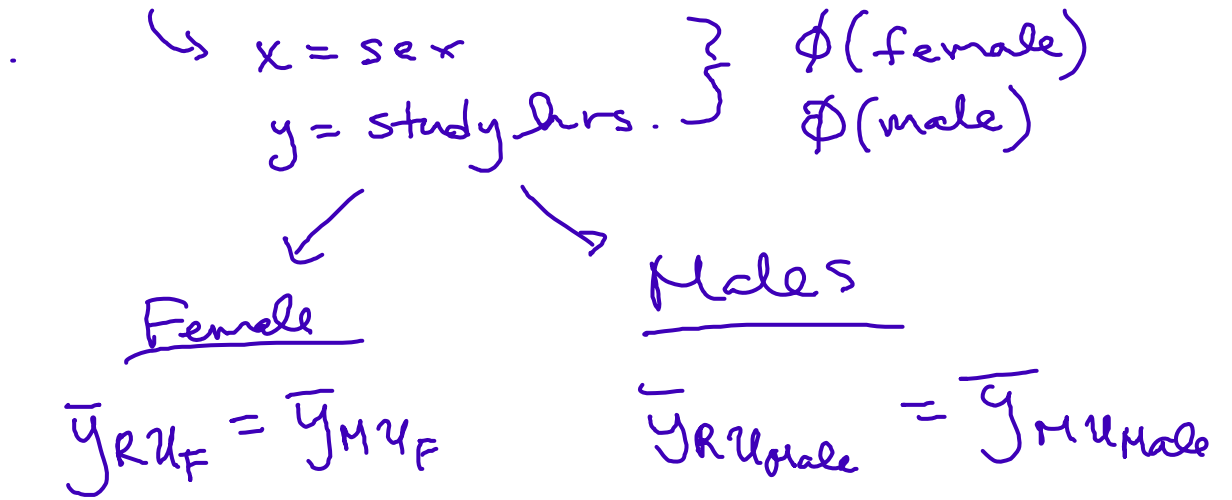
Types of nonresponse: MAR

Missing at Random given covariates (MAR)

- ϕ_i depends on x_i but not y_i (given x_i)

$$\phi_i(x_i) = P(R_i = 1 \mid x_i) = P(R_i = 1 \mid x_i, y_i)$$

- *ignorable* nonresponse: "Model" ϕ given x, then nonresponse can be ignored
  - e.g. poststratification



Types of nonresponse: NMAR

Not Missing at Random (MMAR)

- ϕ_i depends on x_i and y_i

$$\phi_i(x_i, y_i) = P(R_i = 1 \mid x_i, y_i)$$

- *nonignorable* nonresponse
- hard to distinguish MAR and NMAR
- most data is probably NMAR!

e.g. x = sex y = study hrs

Female

$\bar{y}_{R_F} \neq \bar{y}_{M_F}$