# 1 Three Types of Nonresponse

# 2 Subsampling: a technique to reduce nonresponse (or produce unbiased estimates)

# 3 Randomized Response: a technique to reduce nonresponse

# 4 Nonresponse- & Poststratification-adjusted Sampling Weights

First, partition the population $\mathcal{U} = \{\, 1, 2, \ldots, N \,\}$ into $H$ disjoint *poststrata*, i.e. $\mathcal{U} = \bigsqcup_{\beta=1}^{H} \mathcal{P}_\beta$ . For each $i \in \mathcal{U}$, let $\mathcal{P}_{\beta(i)}$ be the unique poststratum that contains $i$. Let $N_\beta := \#(\mathcal{P}_\beta)$ be the number of units in the poststratum $\mathcal{P}_\beta$.

Next, let $\mathcal{S}$ be any realized sample.

**Response distribution conditional on the realized sample $\mathcal{S}$ :**

$$p_{\mathcal{S}} : \mathrm{PowerSet}(\mathcal{S}) \longrightarrow [0, \infty)$$

For each $i \in \mathcal{S}$, the *response probabilty* of $i$ is given by:

$$p_{\mathcal{S}}(i) \;:=\; \Pr(\, i \in \mathcal{R} \,|\, \mathcal{S}\,) \;=\; \sum_{\mathcal{R} \ni i} p_{\mathcal{S}}(\mathcal{R})$$

**We now make the assumption (i.e. model) on the response distribution $p_{\mathcal{S}}$ :**

Suppose $\mathcal{S} = \bigsqcup_{\alpha=1}^{A(\mathcal{S})} \mathcal{H}_\alpha$ can be partitioned such that the elements in each $\mathcal{H}_\alpha$ all have the same response probability.

The disjoint subsets $\mathcal{H}_\alpha, \alpha = 1, 2, \ldots, A(\mathcal{S})$, are called the *response homogeneity groups* of the realized sample $\mathcal{S}$. We emphasize that the number $A(\mathcal{S})$ of response homogeneity groups may vary from sample to sample.

Let $\mathcal{R} \subset \mathcal{S}$ be a response set, selected from $\mathrm{PowerSet}(\mathcal{S})$ according to the $\mathcal{S}$-conditional response distribution $p_{\mathcal{S}}$. Then, $\mathcal{R} = \bigsqcup_{\alpha=1}^{A(\mathcal{S})} \mathcal{R}_\alpha$, where $\mathcal{R}_\alpha := \mathcal{H}_\alpha \cap \mathcal{R}$.

For each $i \in \mathcal{R}$, let $\alpha(i) \in \{\, 1, 2, \ldots, A \,\}$ be the unique element such that $i \in \mathcal{R}_{\alpha(i)} \subset \mathcal{H}_{\alpha(i)}$. In other words, for each $i \in \mathcal{R}$, we let $\mathcal{H}_{\alpha(i)}$ denote the unique response homogeneity group that contains $i$, and $\mathcal{R}_{\alpha(i)} \subset \mathcal{H}_{\alpha(i)}$ is the respondent subset of $\mathcal{H}_{\alpha(i)}$.

$$
\begin{aligned}
W_{1,i} &:= \frac{1}{\pi_i}, \quad \text{for } \; i \in \mathcal{U} = \{\,1,2,\ldots,N\,\} \\
W_{2,i} &:= \begin{cases} W_{1,i} \cdot \dfrac{\displaystyle\sum_{j\in\mathcal{H}_{\alpha(i)}} W_{1,j}}{\displaystyle\sum_{j\in\mathcal{R}\cap\mathcal{H}_{\alpha(i)}} W_{1,j}}, & \text{if } \; i \in \mathcal{R} = \displaystyle\bigsqcup_{\alpha=1}^{A(\mathcal{S})} \mathcal{R}_\alpha \\ 0\,, & \text{if } \; i \notin \mathcal{R} \end{cases} \\
W_{3,i} &:= W_{2,i} \cdot \frac{N_{\beta(i)}}{\displaystyle\sum_{j\in\mathcal{S}\cap\mathcal{P}_{\beta(i)}} W_{2,j}}
\end{aligned}
$$

We let the final nonresponse- and poststratification-adjusted sampling weights be defined as:

$$W_i \;:=\; W_{3,i}$$

**IMPORTANT OBSERVATION:** The quantity $\displaystyle\sum_{j\in\mathcal{H}_{\alpha(i)}} W_{1,j}$ is simply the weight-derived "size" of the response homogeneity group $\mathcal{H}_{\alpha(i)}$. Similarly, $\displaystyle\sum_{j\in\mathcal{R}_{\alpha(i)}} W_{1,j}$ is the weight-derived "size" of the respondent subset $\mathcal{R}_{\alpha(i)} \subset \mathcal{H}_{\alpha(i)}$. Similarly, $\displaystyle\sum_{j\in\mathcal{S}\cap\mathcal{P}_{\beta(i)}} W_{2,j}$ is the weight-derived "size" of the set $\mathcal{S}\cap\mathcal{P}_{\beta(i)}$ of sampled units in the poststratum $\mathcal{P}_{\beta(i)}$. In other words, the "adjustments" described above are essentially based on "weight-derived sizes."

## 5 Imputation: techniques for substituting for missing data

1. Deductive Imputation.
2. Overall mean imputation.
3. Class mean imputation.
4. Hot-deck Imputation.
   - Sequential Hot-deck Imputation.
   - Distance Function Matching.
5. Cold-deck Imputation.
6. Regression Imputation.
7. Multiple Imputation.