Sampling Formulas

Parameters of Interest

Population average $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$

Population total $\tau = \sum_{i=1}^{N} y_i$

Population variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (y_i - \mu)^2$

Bias and MSE

$$Bias[\tilde{\mu}] = E[\tilde{\mu}] - \mu$$

$$\begin{aligned} MSE[\tilde{\mu}] &= E[(\tilde{\mu} - \mu)^2] \\ &= E[(\tilde{\mu} - E[\tilde{\mu}] + E[\tilde{\mu}] - \mu)^2] \quad \text{(add 0)} \\ &= E[(\tilde{\mu} - E[\tilde{\mu}])^2] + (E[\tilde{\mu}] - \mu)^2 + 2E[(\tilde{\mu} - E[\tilde{\mu}])(E[\tilde{\mu}] - \mu)] \\ &= V[\tilde{\mu}] + Bias[\tilde{\mu}]^2 \end{aligned}$$

Probability Sampling

Under SRSWOR

$$E(\tilde{\mu}) = \mu, \qquad Var\,(\tilde{\mu}) = \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n}$$

$$\widehat{Var}(\tilde{\mu}) = \left(1 - \frac{n}{N}\right)\frac{\hat{\sigma}^2}{n}$$

Under SRSWR

$$E(\tilde{\mu}) = \mu \quad \text{and} \quad Var(\tilde{\mu}) = \left(1 - \frac{1}{N}\right)\frac{\sigma^2}{n}$$

$$\widehat{Var}(\tilde{\mu}) = \left(1 - \frac{1}{N}\right)\frac{\hat{\sigma}^2}{n}$$

In general

$$\text{s.e.}(\hat{\mu}) = \text{SE}(\hat{\mu}) = \sqrt{\widehat{Var}(\tilde{\mu})}$$

Confidence Intervals

- 99%: 2.576 because $P(z \leq 2.576) = P(z \geq -2.576) = 0.995$
- 95%: 1.960 because $P(z \leq 1.960) = P(z \geq -1.960) = 0.975$
- 90%: 1.645 because $P(z \leq 1.645) = P(z \geq -1.645) = 0.950$

$$\hat{\mu} \pm c \times \text{s.e.}(\hat{\mu})$$

$$\hat{\tau} \pm c \times \text{s.e.}(\hat{\tau}) = N\hat{\mu} \pm c * N \times \text{s.e.}(\hat{\mu})$$

$$\hat{\pi} \pm c \times \text{s.e.}(\hat{\pi}) = \hat{\pi} \pm c \times \sqrt{\left(1 - \frac{n}{N}\right)\frac{\hat{\pi}(1-\hat{\pi})}{n-1}}$$

Sample Size Calculations

If the fpc is approximately 1, as is the case most of the time, we can rearrange the margin of error equation to solve for n_0:

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n_0}} \qquad \Longrightarrow \quad n_0 = z^2_{\alpha/2}\frac{\sigma^2}{e^2}$$

In cases where the fpc is small (i.e. when n is large compared with the population size) we would make the fpc adjustment:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{z^2_{\alpha/2}\sigma^2}{e^2 + \frac{z^2_{\alpha/2}\sigma^2}{N}}$$

In surveys on proportions, for large populations, $\sigma^2 \approx \pi(1-\pi)$, which is maximized when $\pi = 1/2$.

Strata Weights

We call $W_h = \frac{N_h}{N}$ the stratum weights

We call $w_{hj} = \frac{N_h}{n_h}$ the sampling weight

We call $w_h = n_h/n$ the sample strata weights

Stratum Mean and Variance

For each stratum $h = 1, 2, \ldots, H$ we have:

$$\hat{\mu}_h = \overline{y}_h = \frac{1}{n_h}\sum_{j\in s_h} y_{hj} \qquad \text{sample stratum mean}$$

$$\hat{\sigma}^2_h = \frac{1}{n_h - 1}\sum_{j\in s_h}(y_{hj} - \hat{\mu}_h)^2 \qquad \text{sample stratum variance}$$

$$\widehat{Var}(\tilde{\mu}_{\text{str}}) = \sum_{h=1}^{H} W_h^2\left(1 - \frac{n_h}{N_h}\right)\frac{\hat{\sigma}^2_h}{n_h}$$

$$\widehat{Var}(\tilde{\tau}_{\text{str}}) = \sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)N_h^2\frac{\hat{\sigma}^2_h}{n_h}$$

$$\hat{\mu}_{\text{str}} = \sum_{h=1}^{H} W_h\hat{\mu}_h$$

$$\hat{\mu}_{\text{str}} \pm z_{\alpha/2}\sqrt{\hat{V}(\tilde{\mu}_{\text{str}})}$$

$$\hat{\pi}_h = \hat{\mu}_h = \frac{1}{n_h}\sum_{j\in s_h} y_{hj} \qquad \hat{\sigma}^2_h = \frac{1}{n_h - 1}\sum_{j\in s_h}\hat{\pi}_h(1 - \hat{\pi}_h)$$

$$\hat{\pi}_{\text{str}} = \sum_{h=1}^{H}\frac{N_h}{N}\hat{\pi}_h$$

$$\hat{V}(\hat{\pi}_{\text{str}}) = \sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)\left(\frac{N_h}{N}\right)^2\frac{\hat{\pi}_h(1-\hat{\pi}_h)}{n_h - 1}$$

$$\hat{\tau}_{\text{str}} = \sum_{h=1}^{H} N_h\hat{\pi}_h$$

Emelie Gustafsson