Sample size determination

$\mu : \bar{X} \qquad \sqrt{\dfrac{\sigma^2}{n}} \qquad \sqrt{\dfrac{S^2}{n}}$

$p : \hat{p} \qquad \sqrt{\dfrac{p(1-p)}{n}} \qquad \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

with replacement

$$\operatorname{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}, \quad i \neq j$$

$\mu : \bar{X} \qquad \sqrt{\dfrac{\sigma^2}{n}\left(1 - \dfrac{n-1}{N-1}\right)} \qquad \sqrt{\dfrac{S^2}{n}\left(1 - \dfrac{n}{N}\right)}$

$p : \hat{p} \qquad \sqrt{\dfrac{p(1-p)}{n}\left(1 - \dfrac{n-1}{N-1}\right)} \qquad \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}\left(\dfrac{n}{n-1}\right) \cdot \left(1 - \dfrac{n}{N}\right)}$

without replacement

n - sample size

N - population size

$n \geqslant 0{,}05\,N \;(0{,}1\,N) \Rightarrow$ correction

Sample Size Determination

$$\frac{\bar{X} - \mu}{\sigma^2 / \sqrt{n}} \sim N(0,1)$$

$$\bar{X} + \underbrace{z_{\alpha/2} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{s.e.}}}_{\text{margin of error}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$|\bar{X} - \mu| \leq E$$

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq E$$

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

$$\hat{p} + z_{\alpha/2} \cdot \sqrt{p(1-p)/n} \leq p \leq \hat{p} + z_{1-\alpha/2} \sqrt{p(1-p)/n}$$

$$z_{\alpha/2} \sqrt{p(1-p)/n} \leq E$$

$$n \geqslant \left(\frac{z_{\alpha/2}}{E}\right)^2 \cdot p(1-p)$$

★ If $n \geqslant 0{,}05\, N$ $(0{,}10)$

for μ and p $\quad n_c = \dfrac{nN}{N+n-1}$

n – min. sample size w.o. correction

n_c – –//– with correction

★ If error in percents:

$$E = R\mu$$

percentage

$$n = \frac{C^2 z^2_{\alpha/2}}{R^2}, \quad C = \frac{\sigma}{\mu}$$

Determining sample size for stratified sampling

- among heterogenous groups
- within homogenous

$$
\begin{array}{ccccc}
 & & \text{Population } N & & \\
 & \swarrow & & \downarrow & \\
N_1 & & \cdots & N_k & \quad \sum N_i = N \\
\downarrow & & & \downarrow & \\
n_1 & & \cdots & n_k & \quad \sum n_i = n
\end{array}
$$

$\bar{Y}_i = \frac{1}{N_i} \sum_j^{N_i} y_{ij}$ - pop. mean in ith strata

$\bar{y}_i = \frac{1}{n_i} \sum_j y_{ij}$ - sample mean in ith strata

$$\bar{Y} = \frac{1}{N} \sum N_i \bar{Y}_i = \sum w_i \bar{Y}_i \,, \quad w_i = \frac{N_i}{N}$$

$$\bar{y}_i = \frac{1}{n} \sum n_i \bar{y}_i$$

$$E(\bar{y}_i) = \frac{1}{n} \sum n_i E(\bar{y}_i) =$$

$$= \frac{1}{n} \sum n_i \bar{Y}_i \neq \bar{Y}$$

$$\bar{y}_{ST} = \frac{1}{N} \sum N_i \bar{y}_i$$

$$Var(\bar{y}_{ST}) = \sum W_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{S_i^2}{n_i}$$

$$\hat{Var}(\bar{y}_{ST}) = \sum \frac{W_i^2 s_i^2}{n_i} \qquad (SWR)$$

$$\hat{Var}(\bar{y}_{ST}) = \sum W_i^2 \left(\frac{N_i - n_i}{N_i n_i}\right) s_i^2$$

Allocation of sample sizes

$n_1, \ldots, n_k$

(i) minimise cost of survey

(ii) maximise precision

(1) Equal allocation

$n_i = n/k$

(2) Proportional allocation

$n_i \propto N_i$

$n_i = \delta N_i$

$\sum n_i = \delta$

$n = \delta N \quad \Rightarrow \quad \delta = \dfrac{n}{N}$

$n_i = \dfrac{n}{N} N_i$

(3) Neyman or optimal allocation

$n_i \propto N_i S_i \qquad E(s_i^2) = S_i^2$

$n_i = \delta N_i S_i$

$\sum n_i = \sum \delta N_i s_i$

$$n = \delta \sum N_i S_i$$

$$\delta = \frac{n}{\sum N_i S_i}$$

$$n_i = \frac{n \cdot N_i S_i}{\sum N_i S_i}$$

(4) Choice based on cost

$$C = C_0 + \sum C_i n_i$$

$$\mathcal{L} = Var(\bar{y}_{ST}) - \lambda^2 (C - C_0)$$

$$= \sum W_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) S_i^2 + \lambda^2 \sum C_i n_i - \dots$$

$$= \sum \left[\frac{W_i S_i}{\sqrt{n_i}} - \lambda \sqrt{C_i n_i} \right]^2 + \text{const}$$

$$\frac{W_i S_i}{\sqrt{n_i}} = \lambda \sqrt{C_i n_i}$$

$$n_i = \frac{1}{\lambda} \frac{W_i S_i}{\sqrt{C_i}}$$

λ ↗ min var. for fixed cost (a)

λ ↘ min costs for fixed var (b)

(a) $C = C_0^*$

$$\sum c_i n_i = C_0^*$$

$$\sum c_i \frac{1}{\lambda} \frac{w_i s_i}{\sqrt{c_i}} = C_0^*$$

$$\lambda = \frac{\sum \sqrt{c_i}\, w_i s_i}{C_0^*} \Rightarrow n_i$$

(b) $V = V_0$

$$\sum \left(\frac{1}{n_i} - \frac{1}{N_i} \right) w_i^2 s_i^2 = V_0$$

$$\sum \frac{\lambda \sqrt{c_i}}{w_i s_i} w_i^2 s_i^2 = V_0 + \sum \frac{w_i^2 s_i^2}{N_i}$$

$$\lambda = \frac{V_0 + \sum \frac{w_i^2 s_i^2}{N_i}}{\sum w_i s_i \sqrt{c_i}} \Rightarrow n_i$$

⑤ Proportional allocation for fixed cost or variance

a) $C = C_0 \qquad C_0 = \sum c_i n_i$

$$n_i = \frac{n}{N} \cdot N_i = n W_i$$

$$C_0 = n \sum W_i c_i$$

$$n = \frac{C_0}{\sum W_i c_i}$$

$$n_i = W_i \cdot n = W_i \cdot \frac{C_0}{\sum W_i c_i}$$

b) $V = V_0 \qquad$ (Bowley's allocation)

$$\sum \left(\frac{1}{n_i} - \frac{1}{N_i} \right) W_i^2 S_i^2 = V_0$$

$$\{ n_i = n \cdot W_i \}$$

$$\sum \frac{W_i^2 S_i^2}{n W_i} = V_0 + \sum \frac{W_i^2 S_i^2}{N_i}$$

$$n = \frac{\sum W_i S_i^2}{V_0 + \sum W_i^2 S_i^2 / N_i} \Rightarrow n_i = n \cdot W_i$$