

LECTURE 11: 7.7 Determination of Sample Size for Ratio Estimation

Suppose we want to find a sample size n such that we can be $100(1 - \alpha)\%$ confident that the estimated ratio r is within $100\varepsilon\%$ of the true value of R . That is, we want to find n satisfying $z_{1-\alpha/2} SE(r) \leq \varepsilon R$. Equivalently,

$$z_{1-\alpha/2} (R/\sqrt{n}) (V_X^2 + V_Y^2 - 2\rho_{XY} V_X V_Y)^{1/2} \sqrt{\frac{N-n}{N-1}} \leq \varepsilon R$$

Suppose that from a pilot study we have the estimates \hat{V}_X^2 , \hat{V}_Y^2 , and $\hat{\rho}_{XY}$. Assume also that we have a good idea what N is. Then n is the smallest integer that solves

$$\begin{aligned} \frac{z_{1-\alpha/2}}{\sqrt{n}} (V_X^2 + V_Y^2 - 2\rho_{XY} V_X V_Y)^{1/2} \sqrt{\frac{N-n}{N-1}} &\leq \varepsilon, \\ \frac{N}{n} - 1 &\leq \frac{(N-1)\varepsilon^2}{z_{1-\alpha/2}^2 (V_X^2 + V_Y^2 - 2\rho_{XY} V_X V_Y)}, \\ n &\geq \frac{N z_{1-\alpha/2}^2 (V_X^2 + V_Y^2 - 2\rho_{XY} V_X V_Y)}{(N-1)\varepsilon^2 + z_{1-\alpha/2}^2 (V_X^2 + V_Y^2 - 2\rho_{XY} V_X V_Y)}. \end{aligned}$$

In practice, V_X , V_Y , ρ_{XY} are estimated from pilot or similar studies. Also a good guess of N should be available.

Chapter 8. Introduction to Cluster Sampling.

Suppose a population is divided into “natural” groups called clusters. In cluster sampling, a random sample of clusters is drawn first, and then an independent random sample of population elements is drawn within each selected cluster (or sometimes all elements within selected clusters are surveyed).

When cluster sampling is used?

1. When constructing a sampling frame of all population elements is difficult, expensive, or impossible.
2. When population is geographically dispersed.

Examples.

1. The target population is residents of nursing homes in U.S. It is cheaper to sample nursing homes (clusters) and then interview each resident in the selected homes than to interview an SRS of all nursing home residents in U.S. With an SRS of residents interviewers might have to travel to a nursing home just to interview one resident.
2. In an archaeological survey, researchers first choose regions at random and then examine all or an SRS of artifacts found in the selected regions.

They wouldnt examine just isolated chosen-at-random artifacts in different regions.

3. The sampling frame may consist of households in a city (clusters), and then every individual in the selected households is interviewed.

4. The sampling frame may be a list of farms in a region, and after an SRS of farms is drawn, each cow from selected farms is interviewed.

5. Long Beach Unified School District would like to take a random sample of sixgraders. It is more convenient to draw a sample of middle schools and then select a sample of individuals within each selected school than to obtain a list of all sixgraders within the school district and draw an SRS from this list.

6. Suppose it is known that people of certain nationality live on certain blocks in a city. It is impossible to procure the list of all residents who live on these blocks, but it is possible to draw a sample of several blocks, and then go door-to-door within the chosen city blocks and make a list of residents of this nationality who live there. And then draw an SRS from the created lists.