LECTURE 13: **10.1 Simple Two-Stage Cluster Sampling**

DEFINITION A simple two-stage cluster sampling is a sampling in which clusters are chosen at the first stage by drawing an SRS, and at the second stage, an SRS is drawn within each selected cluster.

We will consider first a special situation in which all population clusters have the same number $N_i,\ i = 1, \ldots, M$ of elements.

How to take a simple two-stage cluster sample?

Suppose a population has 20 clusters of size 25 each. We want to draw an SRS of 3 clusters, and within each sampled cluster take an SRS of 3 elements.

89 65 62 22 94 24 43 05 61 25 27 30 32 24

We enumerate the 20 clusters by $01-05$, $06-10$, $11-15$, $96-99, 00$ and use the table of random digits to chose 3 clusters.

We choose clusters 89 ($85/5 = 17, 89- > 18$), 65(65/5=13), 62(=13, skip), and 22 (=5).

Then number the 25 elements within each of the chosen clusters $01, \ldots, 04$, $05, \ldots, 08, \ldots, 97, 98, 99, 00$.

Within cluster 18 choose elements 94 (=24), 24(=6), 43(=11); within cluster 13, choose elements 05(=2), 61(=16), 25(=7); within cluster 5, choose elements 27(=7), 30(=8), 32(=8, skip), 24(=6)

Population Characteristics

N=population size,
M=number of clusters in the population,
$\bar{N} = N/M$ =(average) number of elements per population cluster,
X_{ij}= measurement for the jth element in the ith cluster, $j = 1, \ldots, \bar{N},\ i = 1, \ldots, M$,
$X_i = \sum_{j=1}^{\bar{N}} X_{ij}$=total for the ith cluster,
$X = \sum_{i=1}^{M} \sum_{j=1}^{\bar{N}} X_{ij}$=population total,
$\bar{X} = X/M$=mean per cluster,
$\bar{\bar{X}} = X/N$=mean per element.

Sample Characteristics

m=number of clusters in the sample,
n=total number of sampled elements,
$\bar{n} = n/m$=number of sampled elements in each cluster,
x_{ij}= measurement for the jth element in the ith sampled cluster, $j = 1, \ldots, \bar{n}$, $i = 1, \ldots, m$,
$x_i = \sum_{j=1}^{\bar{n}} x_{ij}$=total for the ith sampled cluster,
$x = \sum_{i=1}^{m} \sum_{j=1}^{\bar{n}} x_{ij}$=sample total,
$\bar{x} = x/m$=average per cluster in the sample,
$f_1 = m/M$=first-stage sampling fraction,
$f_2 = \bar{n}/\bar{N}$ =second-stage sampling fraction,
$f = f_1 f_2$=overall sampling fraction.

Estimation of Population Characteristics

The **population total** X is estimated by $x'_{clu} = x/f$. A $100(1-\alpha)\%$ CI for X is

$$x'_{clu} \pm z_{1-\alpha/2} \frac{M}{\sqrt{m}\, f_2} \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1}} \sqrt{\frac{N-n}{N}} .$$

The **population mean per cluster** $\bar{X}$ is estimated by $\bar{x}_{clu} = x'_{clu}/M$. A $100(1-\alpha)\%$ CI for $\bar{X}$ is

$$\bar{x}_{clu} \pm z_{1-\alpha/2} \frac{1}{\sqrt{m}\, f_2} \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1}} \sqrt{\frac{N-n}{N}} .$$

The **population mean per element** $\bar{\bar{X}}$ is estimated by $\bar{\bar{x}}_{clu} = x'_{clu}/N$. A $100(1-\alpha)\%$ CI for $\bar{\bar{X}}$ is

$$\bar{\bar{x}}_{clu} \pm z_{1-\alpha/2} \frac{1}{\bar{N}\sqrt{m}\, f_2} \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1}} \sqrt{\frac{N-n}{N}} .$$