LECTURE 12: **9.1 Simple One-Stage Cluster Sampling**

DEFINITION Simple one-stage cluster sampling is a sampling in which clusters are chosen by drawing an SRS, and within each sampled cluster, all elements are selected.

How to draw a simple one-stage cluster sample?

Suppose we want to choose 11 clusters by drawing an SRS from among 97. We number clusters $01, 02, \ldots, 97$ and go through the table of random digits and choose 11 distinct two-digit numbers from this range.

38 40 38 93 10 79 45 02 19 32 40 40 24 67 91 10 96 67

Population Characteristics

N=population size,
M=number of clusters in the population,
X_{ij}= measurement for the jth element in the ith cluster, $j = 1, \ldots, N_i$, $i = 1, \ldots, M$,
$X_i = \sum_{j=1}^{N_i} X_{ij}$=total for the ith cluster,
$X = \sum_{i=1}^{M} \sum_{j=1}^{N_i} X_{ij}$=population total,
$\bar{X} = X/M$=mean per cluster,
$\bar{\bar{X}} = X/N$=mean per element,
$\sigma_{1X}^2 = \sum_{i=1}^{M} (X_i - \bar{X})^2/M$=variance over all clusters.

Sample Characteristics

m=number of clusters in the sample,
x_{ij}= measurement for the jth element in the ith sampled cluster, $j = 1, \ldots, N_i$, $i = 1, \ldots, m$,
$x_i = \sum_{j=1}^{N_i} x_{ij}$=total for the ith sampled cluster,
$x = \sum_{i=1}^{m} \sum_{j=1}^{N_i} x_{ij}$=sample total,
$\bar{x}_{clu} = x/m$=sample mean per cluster.

9.2, 9.3 Estimation of Population Characteristics

The **population total** X is estimated by $x'_{clu} = (M/m)x$. The standard error of the estimator is

$$SE(x'_{clu}) = \frac{M}{\sqrt{m}}\, \sigma_{1X} \sqrt{\frac{M-m}{M-1}}\,.$$

Thus, a $100(1-\alpha)\%$ CI for X is

$$x'_{clu} \pm z_{1-\alpha/2} \frac{M}{\sqrt{m}}\, \hat{\sigma}_{1X} \sqrt{\frac{M-m}{M-1}}$$

where

$$\hat{\sigma}_{1X} = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x}_{clu})^2}{m-1}} \sqrt{\frac{M-1}{M}}\,.$$

The **population mean per cluster** $\bar{X}$ is estimated by $\bar{x}_{clu} = x/m$. The standard error of the estimator is

$$SE(\bar{x}_{clu}) = \frac{1}{\sqrt{m}}\,\sigma_{1X}\,\sqrt{\frac{M-m}{M-1}}\,.$$

Thus, a $100(1-\alpha)\%$ CI for $\bar{X}$ is

$$\bar{x}_{clu} \,\pm\, z_{1-\alpha/2}\,\frac{1}{\sqrt{m}}\,\hat{\sigma}_{1X}\,\sqrt{\frac{M-m}{M-1}}\,.$$

The **population mean per element** $\bar{\bar{X}}$ is estimated by $\bar{\bar{x}}_{clu} = (M/mN)x$. The standard error of the estimator is

$$SE(\bar{\bar{x}}_{clu}) = \frac{M}{\sqrt{m}N}\,\sigma_{1X}\,\sqrt{\frac{M-m}{M-1}}\,.$$

Thus, a $100(1-\alpha)\%$ CI for $\bar{\bar{X}}$ is

$$\bar{\bar{x}}_{clu} \,\pm\, z_{1-\alpha/2}\,\frac{M}{\sqrt{m}N}\,\hat{\sigma}_{1X}\,\sqrt{\frac{M-m}{M-1}}\,.$$

9.4 How Large a Sample is Needed?

To estimate the **population total** X with $100(1-\alpha)\%$ confidence within $100\varepsilon\%$ of the true value, we need the number of clusters m that satisfies $z_{1-\alpha/2}\,SE(x'_{clu}) \le \varepsilon X$, or, equivalently,

$$z_{1-\alpha/2}\,\frac{M}{\sqrt{m}}\,\sigma_{1X}\,\sqrt{\frac{M-m}{M-1}} \le \varepsilon X.$$

Let $V_X = \sigma_{1X}/\bar{X} = \sigma_{1X}M/X$. We have

$$z_{1-\alpha/2}\,\frac{V_X}{\sqrt{m}}\,\sqrt{\frac{M-m}{M-1}} \le \varepsilon\,,$$

$$m \ge \frac{M z_{1-\alpha/2}^2 V_X^2}{(M-1)\varepsilon^2 + z_{1-\alpha/2}^2 V_X^2}\,.$$

It can be shown that this formula holds for the **population mean per cluster** $\bar{X}$, and the **population mean per element** $\bar{\bar{X}}$.