I. Select a simple random sample of of clusters and then select all listing units within each cluster.

A. Example Table 9.1, page 233.
Data Collection, Page 232-234.

B. Notation Box 9.1, Box 9.2, pages 238- 239.

I. Estimation of Population Parameter

A Estimation of Population Mean $\overline{X}$ (Mean per cluster)

1. $\bar{x}_{clu} = \frac{x}{m}$, where x = sample total for characteristic X

m = # of clusters sampled.

Example X: # over 65 needing nurse; m = 2 (Developments 2 and 5)

$$\bar{x}_{clu} = \frac{11+12}{2} = 11.5$$

2. $\hat{SE}(\bar{x}_{clu}) = \frac{s_{1x}}{\sqrt{m}}\left(\frac{M-m}{M}\right)^{\frac{1}{2}}$ where M = # of clusters in the population,

and $s_{1x} = \left[\frac{\sum_{i=1}^{m}(x_i - \bar{x}_{clu})^2}{m-1}\right]^{\frac{1}{2}} = \hat{\sigma}_{1x}\left(\frac{M}{M-1}\right)^{\frac{1}{2}}$

is the sample standard deviation of cluster totals (x_i) about the mean per cluster ($\bar{x}_{clu}$).

Example

$$s_{1x} = \left[\frac{(11-11.5)^2 + (12-11.5)^2}{2-1}\right]^{\frac{1}{2}} = 0.7071$$

$$\hat{SE}(\bar{x}_{clu}) = \frac{0.7071}{\sqrt{2}}\left(\frac{5-2}{5}\right)^{\frac{1}{2}} = 0.3873$$

3. 100(1-α) Confidence Intervals $\overline{X}$

$$\bar{x}_{clu} \pm Z_{1-\frac{\alpha}{2}}\hat{SE}(\bar{x}_{clu})$$

Example 95% Confidence Intervals for $\overline{X}$

11.5 $\pm$ (1.96)(0.3873) = 11.5 $\pm$ 0.7591 = (10.74, 12.26)

B. Estimation of Population Total X

1. $\hat{X} = x'_{clu} = \left(\frac{M}{m}\right)x$

2. $\hat{SE}(x'_{clu}) = M\left(\frac{s_{1x}}{\sqrt{m}}\right)\left(\frac{M-m}{M}\right)^{\frac{1}{2}} = M\left[\hat{SE}(\bar{x}_{clu})\right]$

3. 100(1-α) Confidence Intervals X

$$x'_{clu} \pm Z_{1-\frac{\alpha}{2}}\hat{SE}(x'_{clu})$$

Example: Page 237.

C. Estimation of Population Ratio X/Y

1. $r_{clu} = \frac{x}{y}$, where x = sample total for X
   and y = sample total for Y.

2. $$S\hat{E}(r_{clu}) = (r_{clu}) \left\{ \frac{\left[S\hat{E}(\bar{x}_{clu})\right]^2}{\bar{x}_{clu}^2} + \frac{\left[S\hat{E}(\bar{y}_{clu})\right]^2}{\bar{y}_{clu}^2} - \left(\frac{2}{m}\right)\left(\frac{M-m}{M}\right)\frac{1}{\bar{x}_{clu}\bar{y}_{clu}}\left[\frac{\sum_{i=1}^{m}(x_i - \bar{x}_{clu})(y_i - \bar{y}_{clu})}{m-1}\right] \right\}^{\frac{1}{2}}$$

3. 100(1-α) Confidence Intervals $R = \frac{X}{Y}$

   $$r_{clu} \pm Z_{1-\frac{\alpha}{2}} S\hat{E}(r_{clu})$$

Example: x = 23 = total over 65 needing a nurse
y = 67 = total over 65.

$r_{clu} = \frac{23}{67} = 0.3433$ = proportion of persons over 65 needing a nurse.

Page 241 , $S\hat{E}(r_{clu}) = 0.0076$ and
95% Confidence Interval for R: (.3284, .3582)

III. Sample Size Determination

First define first stage variances σ_{1x}^2

$$\sigma_{1x}^2 = \frac{\sum_{i=1}^{M}(X_i - \overline{X})^2}{M} \qquad (9.1)$$

which measures variation among clusters with respect to the distribution of total levels of X. Recall, $\hat{\sigma}_{1x}^2 = s_{1x}^2\left(\frac{M-1}{M}\right)$.

Then, the required number of sample clusters, m, to estimate $\overline{X}$ (mean per cluster) within ε of the true value, with $(1-\alpha)100\%$ confidence, is given by

$$\boxed{m = \frac{Z_{1-\frac{\alpha}{2}}^2 M V_{1x}^2}{Z_{1-\frac{\alpha}{2}}^2 V_{1x}^2 + (M-1)\varepsilon^2}} \text{ where } V_{1x}^2 = \frac{\sigma_{1x}^2}{\overline{X}^2}$$

$$m \approx \frac{Z_{1-\frac{\alpha}{2}}^2 V_{1x}^2}{\varepsilon^2}, \text{ if } \frac{m}{M} \le .05,$$

Use a similar approach for totals and ratios, see Box 9.4, page 255.

IV. Strategies for Choosing a Sample design

A. Strategy I: Minimize field cost for fixed reliability of estimates

1. Determine ε (relative error for your estimate) for fixed α.
2. Determine the sample size required to meet the requirements in 1 for each considered sample design
3. Estimate field cost of obtaining the required sample size in 2 for each design

4. Choose the sample design with the lowest field cost.

B. Strategy II: Minimize the variance of the estimate at fixed cost.

1. Specify field cost according to your budget.
2. Calculate the sample size that can be taken with specified cost (1) using cost functions.
3. Calculate the variance of the estimate for each projected sample size.
4. Choose the sample design with the lowest variance in 3.