# I. <u>One in k systematic sampling</u>

Randomly selecting one element from the first k elements in the frame and every $k^{th}$ element thereafter, where $k = \frac{N}{n}$ for a population size N and a sample size n.

## A. Point estimates for population parameters are calculated using the same formulas for simple random sampling.

1. Sample Total: $x = \sum_{i=1}^{n} x_i$
2. Sample Mean: $\bar{x} = \frac{x}{n}$
3. Sample Proportion: $p_y = \frac{y}{n}$, for $y_i = \begin{cases} 1 & if \quad yes \\ 0 & if \quad no \end{cases}$
4. Estimated Population Total: $x' = \left( \frac{N}{n} \right) x$

## B. If $n = \frac{N}{k}$ is an integer, the estimated means, totals, and proportions are unbiased for systematic sampling.

## C. Variances of the estimates depend on the type of populations sampled from. In many situations VAR($\bar{x}$) cannot be determined using only one 1 in k systematic sample of size n. Thus an approximate 95% bound cannot be determined unless we use <u>Repeated Systematic Sampling</u>.

1. <u>Example</u>: Tables 4.2, 4.3 - pages 88, 89.
   a. Unbiased estimates – each sample has equal n.
   b. SE's differ under SRS and SYS.

$$SE(\bar{x})_{SYS} = \sqrt{\frac{\sum_{i=1}^{5} \left[ \bar{x} - E(\bar{x}) \right]^2}{5}} = 2.90 \qquad \text{(page 89).}$$

$$SE(\bar{x})_{SRS} = \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{67.91}{5} \left( \frac{25-5}{24} \right)} = 3.36$$

(from table 2.1, page 17, N=25, $\sigma_X^2 = 67.91$).
Thus, SRS methods will overestimate VAR($\bar{x}$) in this case.

2. <u>Example</u>: Table 4.4, page 90.

   1-in-6 systematic sample $\Rightarrow n = \frac{25}{6} = 4.167$.

   n non integer $\Rightarrow$ biased estimates, since physicians do not have equal impact on each sample. $E(\bar{x}) = 4.79, \bar{X} = 5.08$

   The variances of the estimates depend on the type of population sample from.

# II. Types of Populations

A. **<u>Random Population</u>**: one in which the elements are listed in random order with regard to the size of the variate.
If sampling from a random population, a systematic sample can be treated as a simple random sample,

i.e. VAR( $\bar{x}$ ) $_{SYS}$ = VAR( $\bar{x}$ ) $_{SRS}$ .

B. **<u>Ordered Population</u>**: one in which the elements are listed in the frame according to the size of the variate of interest.
Systematic sampling tends to yield more precise estimates than simple random for an ordered population,
i.e. VAR( $\bar{x}$ ) $_{SYS}$ < VAR( $\bar{x}$ ) $_{SRS}$ , since
1. <u>homogeneous</u> elements will be listed together;
2. a systematic sample will be <u>heterogeneous</u>.

C. **<u>Periodic Population</u>**: one in which the elements are listed in the frame in such a way that a cyclic pattern is established with regard to the variate of interest.
In general, systematic sampling is less efficient than simple random sampling when sampling from a periodic population,
i.e. VAR( $\bar{x}$ ) $_{SYS}$ > VAR( $\bar{x}$ ) $_{SRS}$ .

**III.** Variances of Population Estimates Under Systematic Sampling

A. $VAR(\hat{d})_{SYS} = VAR(\hat{d})_{SRS}\left[1+(n-1)\delta_x\right]$,
where $\delta_x$ is **the <u>intraclass correlation coefficient</u>**. $\delta_x$ is a measure of correlation between pairs of elements within the same systematic sample.

<u>Total</u> $x'$ :
$$VAR(x') = \left(\frac{N^2\sigma_x^2}{n}\right)\left[1+(n-1)\delta_x\right] \qquad (4.2)$$

<u>Mean</u> $\bar{x}$ :
$$VAR(\bar{x}) = \left(\frac{\sigma_x^2}{n}\right)\left[1+(n-1)\delta_x\right] \qquad (4.3)$$

<u>Proportion</u> $p_y$ :
$$VAR(p_y) = \left(\frac{P_Y(1-P_Y)}{n}\right)\left[1+(n-1)\delta_x\right] \qquad (4.4)$$

where $n = \frac{N}{k}$,
$$\delta_x = \frac{2\sum_{i=1}^{k}\sum_{j=1}^{N}\sum_{l<j}\left(X_{ij}-\bar{X}\right)\left(X_{il}-\bar{X}\right)}{nk(n-1)\sigma_X^2} \qquad (4.5)$$

and $X_{ij}$ represents the $j^{th}$ element in the $i^{th}$ systematic sample.

B. Variance relationships and intraclass correlation

1. $\delta_x = 0 \Rightarrow VAR(\bar{x})_{SYS} = VAR(\bar{x})_{SRS}$ (random population)
2. $\delta_x < 0 \Rightarrow VAR(\bar{x})_{SYS} < VAR(\bar{x})_{SRS}$ (ordered population) (SYS more efficient)
3. $\delta_x > 0 \Rightarrow VAR(\bar{x})_{SYS} > VAR(\bar{x})_{SRS}$ (periodic population) (SRS more efficient)

**IV. Repeated Systematic Sampling Notation:**

m: the number of repeated 1 in $k'$ systematic samples of size $n'$

$n'$: sample size of each sample

$k'$: sampling interval for each repeated sample, $k' = \frac{N}{n'}$.

N: population size

n: total sample size, n = $n'm \Rightarrow n' = \frac{n}{m}$

k: sampling interval for one systematic sample, $k = \frac{N}{n} = \frac{M}{m}$

M: the number of possible 1 in $k'$ systematic samples that can be drawn from a population of size N. M = $k'$.

$\bar{x}_i$ : sample mean for the $i^{th}$ repeated sample, i = 1,2,3,…,m.

$\bar{\bar{x}}$ : sample mean for all n elements. $\bar{\bar{x}} = \frac{\sum_{i=1}^{m} \bar{x}_i}{m}$

**V. Repeated Systematic Sampling Procedure**

A. Choose m random numbers between 1 and $k'$.

B. Choose m 1 in $k'$ systematic samples beginning with each random number.

**VI. Repeated Systematic Sampling Estimation of $\mu_X = \bar{X}$:**

$$\hat{\mu}_X = \frac{\sum_{i=1}^{m} \bar{x}_i}{m} = \bar{\bar{x}} \qquad (4.8)$$

$$V\hat{A}R\left(\bar{\bar{x}}\right) = \frac{s_{\bar{x}}^2}{m}\left(\frac{M-m}{M}\right) = \left(\frac{1}{m}\right)\frac{\sum_{i=1}^{m}\left(\bar{x}_i - \bar{\bar{x}}\right)^2}{m-1}\left(\frac{M-m}{M}\right) \qquad (4.9)$$

<u>(1-$\alpha$) 100% Confidence Interval for</u> $\mu_X$

$$\bar{\bar{x}} \pm Z_{1-\frac{\alpha}{2}}\sqrt{V\hat{A}R\left(\bar{\bar{x}}\right)} \qquad (4.10)$$

**VII. Example – Table 4.17, page 104.**

N = 162, n = 18, m = 6 samples,