STA 304H1F-1003HF Fall 2019
Surveys, Sampling and Observational Data
Section: L0101

Professor Tounkara

Week 9 - Systematic Sampling, Chapter 7

Introduction

1-in-k systematic sampling

Definition (1-in-k systematic sampling):

1-in-k systematic sampling is a sampling method in which one of the first k members of the population is selected at random. Then beginning with that member, every kth member is selected.

Example:

- For example, if k = 10, then one of the first 10 members is selected at random (using **sample(10,size=1)**).
- Suppose member #6 is selected.
- Then beginning with member #6, every 10th member is selected. That is, members 6, 16, 26, 36, and so on.

Why Systematic sampling?

The principal reasons for using Systematic sampling rather than SRS are:

1. Simple to implement (easier to perform)

2. May be started without a complete listing frame

(say, interview of every 9th patient coming to a clinic).

3. Can provide greater information per unit cost than SRS can for populations with certains patterns in the arrangment

How to draw 1-in-k systematic sample ?

1 Number the members of the population 1 through N.

2. Using **sample(k,size=1)**, choose a random starting point in the first block of size k. That represents the first member of the sample.

3. From that starting point, put every kth member in the sample.

How to draw 1-in-k systematic sample in R

A systematic sample can be selected in R as follows.

Suppose we want to sample rainfall at a site every 50th day over a year. Thus, M = 365 and we wish to select a 1 in k systematic sample, with k = 50.

```
M <- 365
k <- 50
start <- sample(1:k, 1)
s <- seq(start, M, k)
s
```

```
[1]  46  96 146 196 246 296 346
```

An example: 1-in-10 systematic sample

Let N=100, n=10, then k=100/10.

Then the random start r is selected between 1 and 10 (say, r=7).

- So, the sample will be selected from the population with serial indexes of:

$$7, 17, 27, ..., 97$$

$$\text{i.e.}, r, r+k, r+2k, ..., r+(n-1)k$$

What could be done if k=N/n is not an integer?

Selection of systematic sampling when sampling interval (k) is not an integer

Consider, n=175 and N=1000. So, k=1000/175 = 5.71

- One of the solution is to make k rounded to an integer, i.e., k=5 or k=6.
- Now, if k=5, then n=1000/5=200; or,
- If k=6, then n=1000/6 = 166.67 $\sim$ 167.

Which n should be chosen?

Solution

1. if k=5 is considered, stop the selection of samples when n=175 achieved.

2. if k=6 is considered, treat the sampling frame as a circular list and continue the selection of samples from the beginning of the list after exhausting the list during the first cycle.

3. An alternative procedure is to keep k non-integer and continue the sample selection as follows:

- Let us consider, k=5.71, and r=4.
- So, the first sample is 4th in the list.
- The second = $(4+5.71) = 9.71 \sim$ 9th in the list, the third $= (4+2*5.71) = 15.42 \sim$ 15th in the list, and so on.
- (The last sample is: $4+5.71*(175-1) = 997.54 \sim$ 997th in the list).
- Note that, k is switching between 5 and 6.

Types of population

A population is random if the elements of the population are in random order.

- **Example:** A rondom population may occur in an alphabetical listing of student grades on exam. There is no reason why students at the beginnig should have lower or higher grades than those at the end.

A population is ordered if the elements of the population have values that trend upward or downward when they are listed.

A population is periodic if the elements of the population have a values that tend to cycle upward and downward in a regular pattern when listed.

More details: See FIGURE 7.1, 7.2, and 7.3 on pages 222, and 223

Estmation Based on Systematic Sample

Estimation of the population Mean, I

Suppose we have a systematic sample of size n from a ramdom population of size N.

Estimator of μ:

$$\widehat{\mu}_{sy} = \bar{y}_{sy} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Estimated variance of $\hat{\mu}_{sy}$:

$$\boxed{\widehat{V}(\widehat{\mu}_{sy}) = \left(1 - \frac{n}{N}\right)\left(\frac{s^2}{n}\right)}$$

Estimation of the population Mean, I

Note: Assuming ramdomly ordered population, the estimated variance is identical to the estimated variance of $\overline{y}$ obtained by SRS.

This does not imply that the variance are the same:

$$\mathbf{V}(\overline{\mathbf{y}}_{\mathbf{sy}}) = \frac{\sigma^2}{\mathbf{n}}[\mathbf{1} + (\mathbf{n}-\mathbf{1})\rho] \neq \frac{\sigma^2}{\mathbf{n}}\frac{\mathbf{N}-\mathbf{n}}{\mathbf{N}-\mathbf{1}} = \mathbf{V}(\overline{\mathbf{y}}_{\mathbf{sy}})$$

Where:

ρ : is the a measure of the correlation between pairs of elements within the same systematic sample.