Regional Training Course on Agricultural Cost of Production Statistics
21– 25 November 2016, Daejeon , Republic of Korea

Systematic Sampling

UNITED NATIONS
SIAP
Statistical Institute for
Asia and the Pacific

IMPROVING
AGRICULTURAL
STATISTICS
Global Strategy

Contents

- **Linear systematic sampling**

- **Circular systematic sampling**

- **Implicit Stratification**

UNITED NATIONS
SIAP

Systematic Sampling Schemes

Systematic sampling: A method in which the sample is obtained by selecting every k^{th} element of the population, where **k** is an integer > 1.

Often the units are ordered with respect to that auxiliary data.

3

Systematic Sampling

* This is also an equal probability selection method (*epsem*)

* Two kinds:
  - Linear systematic
  - Circular systematic

Systematic Sampling

* Systematic Sampling (SYS), like SRS, involves selecting **n** sample units from a population of **N** units

* Instead of randomly choosing the **n** units in the sample, a skip pattern is run through a list (frame) of the **N** units to select the sample

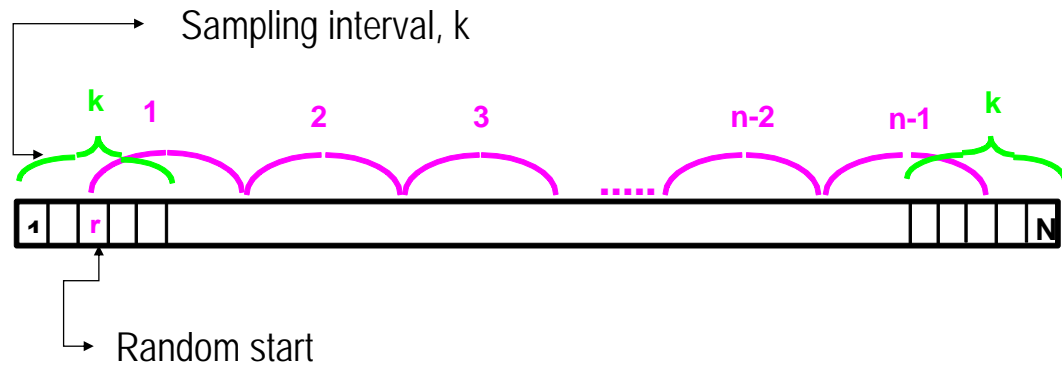* The *skip* or *sampling interval*, **k** = N/n

5

Linear Systematic Sampling

6

Linear Systematic Sampling

Sampling interval, k

Random start

7

Selection Procedure - *Linear Systematic Sampling*

Steps involved:

1. Form a **sequential list** of population units

2. Assign sampling serial number to each unit.

3. Decide on a sample size *n* and compute the skip (***sampling interval***), ***k*** = ***N/n***

4. Choose a random number, ***r*** (***random start***) between 1 and ***k*** (inclusive)

8

Selection Procedure - *Linear Systematic Sampling*

Steps involved:

5. Select the unit with serial number **r** as the first unit in the sample.

6. Add "**k**" to selected random number to select the second unit and

7. continue to add "**k**" repeatedly to previously selected unit number to select the remainder of the sample, till **n** units are selected in the sample.

9

Problem - *Linear Systematic Sampling*

If **N** is a multiple of **n**, then the number of units in each of the **k** possible systematic samples is **n**.

In this case systematic sampling amounts to grouping the **N** units into **k** samples of exactly **n** units each in a systematic manner and selecting one of them with probability 1/**k**.

In this case, the sampling scheme is ***epsem***.

10

Problem - *Linear Systematic Sampling* (Contd.)

But, if **N**/**n** is not an integer, then the number of units selected systematically with the sampling interval **k** [= nearest integer to **N**/**n**] – no longer **epsem**.

Question: Is this a probability sampling scheme?

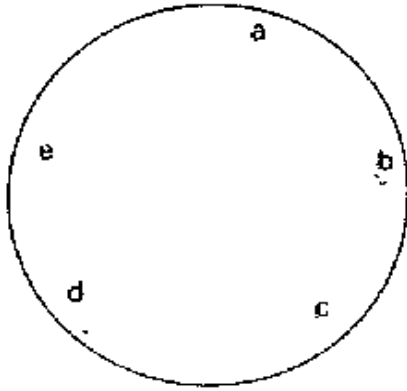This problem may be overcome by adopting a device, known as *circular systematic sampling*.

Circular Systematic Sampling

Circular Systematic Sampling

Systematic Sampling

K=5/2=2.5

a) If k=2 possible samples are:

  ac; bd; ce; da and eb

b) If k=3 possible samples are:

  ad; be; ca; db and ec.

Circular Systematic selection

* Useful when **N**/**n** is not integer
* Determine the interval **k** – rounding <u>down</u> to the integer nearest to **N**/**n**

  [If **N** = 15 and **n** = 4, then **k** is taken as 3 and not 4]
* Take a random start between 1 and **N**
* Skip through the circle by **k** units each time to select the next unit until **n** units are selected
* Thus there could be **N** possible distinct samples instead of **k**
* This method is termed *Circular Systematic Sampling* (CSS)

Systematic Sampling – Important Features

* Often used as an alternative to SRS.
* Requires ordering of the population units
  * Ordering enables SYS sample to be more representative
  * Ordering done by geographical location (say of dwellings) ensures fair spread of sample
  * Ordering done by industry type ensures fair representation of industries
* Ensures each population unit equal chance of being selected into sample

15

Advantages

* Operationally convenient - easier to draw a sample.
* SYS distributes the sample more evenly over the population – thus likely to be more efficient than SRSWOR,

  particularly when the ordering of the units in the list is related to characteristics of the variable of interest.

16

Advantages and Disadvantages

Advantages:

∗ Operationally convenient - easier to draw a sample.

∗ SYS distributes the sample more evenly over the population – thus likely to be more efficient than SRSWOR, particularly when the ordering of the units in the list is related to characteristics of the variable of interest.

Disadvantages :

∗ Requires complete list of the population.

∗ A bad arrangement of the units may produce a very inefficient sample

∗ **Variance estimates cannot be obtained from a single systematic sample.**

17

Systematic Sampling – Implicit Stratification

18

Implicit Stratification

* This refers to a systematic sampling with the units arranged in a certain order.

* Prior to sample selection, all the units are sorted with respect to one or more variables that are deemed to have a high correlation with the variable of interest.

* Implicit stratification guarantees that the sample of units will be spread across the categories of the stratification variables.

19

Thanks