Probability Proportional to Size Systematic Sampling

Presidency University, Kolkata

16th june,2022

Group Description

– Presentation By –

Kalpesh Chatterjee(Roll no.STAT02)

Ritwick Mondal (Roll no.STAT016)

Ankan Chakraborty(Roll no.STAT018)

Contents

Intruduction

Systematic sampling is a probability sampling method, where the sample units are chosen from a population by selecting a random starting point and then by selecting other members of the sample by fixing a sampling interval. Sampling interval is calculated by dividing the entire population size by the desired sample size.

Sampling scheme

Let we have total N unit of finite population denoted by $U_1, U_2, ..., U_N$ with respective size $X_1, X_2, ..., X_N$.
To select a sample of n unit we define sampling interval (k) which is
$= \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n}$

- First we calculate a table for cumulative total of X_i.

Sampling scheme

Let we have total N unit of finite population denoted by $U_1, U_2, \ldots.., U_N$ with respective size $X_1, X_2, \ldots..., X_N$.
To select a sample of n unit we define sampling interval (k) which is

$$= \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n}$$

- First we calculate a table for cumulative total of X_i.
- Next,We select a random number r from 1 to k

Sampling scheme

Let we have total N unit of finite population denoted by $U_1, U_2, \ldots.., U_N$ with respective size $X_1, X_2, \ldots..., X_N$.
To select a sample of n unit we define sampling interval (k) which is
$= \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n}$

- First we calculate a table for cumulative total of X_i.
- Next,We select a random number r from 1 to k
- In step 3 we define $T_i = \sum_{k=1}^{i} X_k$, $i = 1(1)N$. We select the units corresponding to the numbers $r + jk, j = 0(1)n - 1$

Sampling scheme

Let we have total N unit of finite population denoted by $U_1, U_2, ..., U_N$ with respective size $X_1, X_2, ..., X_N$.
To select a sample of n unit we define sampling interval (k) which is
$= \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n}$

- First we calculate a table for cumulative total of X_i.
- Next,We select a random number r from 1 to k
- In step 3 we define $T_i = \sum_{k=1}^{i} X_k,\ i = 1(1)N$. We select the units corresponding to the numbers $r + jk, j = 0(1)n - 1$
- In the last step *ith* population unit U_i is included in the sample if $T_{i-1} < r + jk \leq T_i, i = 1(1)N$ for some $j = 0(1)n - 1$

Circular Systematic Sampling

In most of the practical situation we'll see $\frac{X}{n}$ is not an integer.In that case k is taken as an integer nearest to X/n.In that case n varies from one sample to another .

This is known as **PPS Linear Systematic sampling**.

If we don't want n to vary from one sample to another we adopt another form of sampling scheme -**PPS Circular Systematic Sampling.**

- First,select a random number r from 1 to X.

Circular Systematic Sampling

In most of the practical situation we'll see $\frac{X}{n}$ is not an integer.In that case k is taken as an integer nearest to X/n.In that case n varies from one sample to another .
This is known as **PPS Linear Systematic sampling**.
If we don't want n to vary from one sample to another we adopt another form of sampling scheme -**PPS Circular Systematic Sampling.**

- First,select a random number r from 1 to X.
- We select the units corresponding to the numbers $r + jk, j = 0(1)n - 1$

Circular Systematic Sampling

In most of the practical situation we'll see $\frac{X}{n}$ is not an integer.In that case k is taken as an integer nearest to X/n.In that case n varies from one sample to another .
This is known as **PPS Linear Systematic sampling**.
If we don't want n to vary from one sample to another we adopt another form of sampling scheme -**PPS Circular Systematic Sampling.**

- First,select a random number r from 1 to X.
- We select the units corresponding to the numbers $r + jk, j = 0(1)n - 1$
- *ith* population unit U_i is included in the sample if $T_{i-1} < r + jk \leq T_i, i = 1(1)N$ provided $r + jk < X$.

Circular Systematic Sampling

In most of the practical situation we'll see $\frac{X}{n}$ is not an integer.In that case k is taken as an integer nearest to X/n.In that case n varies from one sample to another .
This is known as **PPS Linear Systematic sampling**.
If we don't want n to vary from one sample to another we adopt another form of sampling scheme -**PPS Circular Systematic Sampling.**

- First,select a random number r from 1 to X.
- We select the units corresponding to the numbers $r + jk, j = 0(1)n - 1$
- *ith* population unit U_i is included in the sample if $T_{i-1} < r + jk \leq T_i, i = 1(1)N$ provided $r + jk < X$.
- If $r + jk > X$ then U_i is included in the sample if $T_{i-1} < r + jk - X \leq T_i, i = 1(1)N$

Estimation Procedure

An ***Unbiased Estimator of Population Total*** $Y = \sum_{i=1}^{N} Y_i$ *has same form as PPSWR scheme method.*

$$\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} \; (\pi_i = np_i)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{X_i/X} \; (p_i = X_i/X)$$
$$= k \sum_{i=1}^{n} \frac{y_i}{X_i}$$

An Example

We have a dataset regarding different crops harvested at different plots of West Bengal in the time period of 2017 - 2018. And a small part of that data is extracted here for demonstration purpose.

Data Source

The real life agricultural data is taken from : http://desagri.gov.in/wp-content/uploads/2021/04/P2017-18.xlsx

Example

Table - 1: Showing population data[1] of yield area (in Ha.)

No.	Crop	No. of Plots	Cropped Area (Ha.)
1	Wheat	5	0.95
2	Gram	8	1.76
3	Masur	35	8.4
4	Mustard	253	70.84
5	Jute	164	60.68
6	Paddy	1954	683.9
7	Moong	10	1.2
8	Sesamum	63	17.64
9	Potato	220	68.2
	Total	2712	913.57

[1] Source: http://desagri.gov.in/wp-content/uploads/2021/04/P2017-18.xlsx

Example

Problem Statement

Here we want to estimate population cropped area (in Ha.) Y, say.

Way of approach

We would like to estimate population cropped area (in Ha.) $\hat{Y}_{PS}$ by its unbiased sample estimate in two different ways:

1. Linear Systematic Sampling and
2. Circular Systematic Sampling.

Example

Problem Statement

Here we want to estimate population cropped area (in Ha.) Y, say.

Way of approach

We would like to estimate population cropped area (in Ha.) $\hat{Y}_{PS}$ by its unbiased sample estimate in two different ways:

1. Linear Systematic Sampling and
2. Circular Systematic Sampling.

Example

Problem Statement

Here we want to estimate population cropped area (in Ha.) Y, say.

Way of approach

We would like to estimate population cropped area (in Ha.) $\hat{Y}_{PS}$ by its unbiased sample estimate in two different ways:

1. **Linear Systematic Sampling** and
2. Circular Systematic Sampling.

Example

Problem Statement

Here we want to estimate population cropped area (in Ha.) Y, say.

Way of approach

We would like to estimate population cropped area (in Ha.) $\hat{Y}_{PS}$ by its unbiased sample estimate in two different ways:

1. Linear Systematic Sampling and
2. Circular Systematic Sampling.

Linear Systematic Sampling

- Let us firstly conduct **linear systematic sampling** draw a sample of size $n = 5$.

Linear Systematic Sampling

- Let us firstly conduct **linear systematic sampling** draw a sample of size $n = 5$.
- Here, X_i = number of plots whre i^{th} crop is cultivated; $i = 1(1)9$. And $X = \sum_{i=1}^{9} X_i = 2712$.

Linear Systematic Sampling

- Let us firstly conduct **linear systematic sampling** draw a sample of size $n = 5$.
- Here, X_i = number of plots whre i^{th} crop is cultivated; $i = 1(1)9$. And $X = \sum_{i=1}^{9} X_i = 2712$.
- Y_i = cultivated or cropped area (in Ha) of i^{th} crop; $i = 1(1)9$.

Linear Systematic Sampling

- Let us firstly conduct **linear systematic sampling** draw a sample of size $n = 5$.
- Here, X_i = number of plots whre i^{th} crop is cultivated; $i = 1(1)9$. And $X = \sum_{i=1}^{9} X_i = 2712$.
- Y_i = cultivated or cropped area (in Ha) of i^{th} crop; $i = 1(1)9$.
- So our sampling interval $k = {}^{X}/_{n} = {}^{2712}/_{5} = 542.4 \approx 542$. Hence we flatten the Y_i values and organize in a matrix form where number of columns is **sampling interval**, $k = 542$.

Linear Systematic Sampling

- Let us firstly conduct **linear systematic sampling** draw a sample of size $n = 5$.
- Here, X_i = number of plots whre i^{th} crop is cultivated; $i = 1(1)9$. And $X = \sum_{i=1}^{9} X_i = 2712$.
- Y_i = cultivated or cropped area (in Ha) of i^{th} crop; $i = 1(1)9$.
- So our sampling interval $k = X/n = 2712/5 = 542.4 \approx 542$. Hence we flatten the Y_i values and organize in a matrix form where number of columns is **sampling interval**, $k = 542$.
- We draw a random number between 1 and $k = 542$; and let that be $r = 2$.

Linear Systematic Sampling

- Let us firstly conduct **linear systematic sampling** draw a sample of size $n = 5$.
- Here, X_i = number of plots whre i^{th} crop is cultivated; $i = 1(1)9$. And $X = \sum_{i=1}^{9} X_i = 2712$.
- Y_i = cultivated or cropped area (in Ha) of i^{th} crop; $i = 1(1)9$.
- So our sampling interval $k = {}^{X}/_{n} = {}^{2712}/_{5} = 542.4 \approx 542$. Hence we flatten the Y_i values and organize in a matrix form where number of columns is **sampling interval**, $k = 542$.
- We draw a random number between 1 and $k = 542$; and let that be $r = 2$.
- So we choose 2^{nd} column of that designed matrix form and those will be of $(r + jk)^{th}$ positions; $j = 0(1)n - 1$.

Linear Systematic Sampling (Contd.)

Table - 2: Showing sample of size $n = 5$ is selcted by linear systematic sampling :

Sample Sl. No.	1	2	3	4	5
Values of j	0	1	2	3	4
Values of $(2 + j \times 542)$	2	1086	1628	2170	2712
Units selected	Wheat	Paddy	Paddy	Paddy	Potato
$p_j = x_j/x$	0.0018	0.7205	0.7205	0.7205	0.0811
Cropped area (Ha.) (y_j)	0.95	683.9	683.9	683.9	68.2

So, an **unbiased estimator** of **total cropped area, Y** is :

$$\hat{Y}_{PS} = \frac{1}{5}\sum_{j=1}^{5}\frac{y_j}{p_j} = 843.2641 \text{ (in Ha.)}$$

Circular Systematic Sampling

- Let us now conduct **circular systematic sampling** draw a sample of size $n = 5$.

Circular Systematic Sampling

- Let us now conduct **circular systematic sampling** draw a sample of size $n = 5$.
- All the notations are as before. But here we flatten the Y_i values and organize in the perimeter of an imaginary circle with number of points in perimeter is nothing but the **sampling interval**, $k = 542$.

Circular Systematic Sampling

- Let us now conduct **circular systematic sampling** draw a sample of size $n = 5$.
- All the notations are as before. But here we flatten the Y_i values and organize in the perimeter of an imaginary circle with number of points in perimeter is nothing but the **sampling interval**, $k = 542$.
- We draw a random number between 1 and $k = 542$; and let that be $r = 5$.

Circular Systematic Sampling

- Let us now conduct **circular systematic sampling** draw a sample of size $n = 5$.
- All the notations are as before. But here we flatten the Y_i values and organize in the perimeter of an imaginary circle with number of points in perimeter is nothing but the **sampling interval**, $k = 542$.
- We draw a random number between 1 and $k = 542$; and let that be $r = 5$.
- So we start from 5^{th} point of perimeter of that circle and select those units having position :

$$\begin{cases} (r + jk) & \text{if } (r + jk) \leq X \\ (r + jk - X) & \text{if } (r + jk) > X \end{cases} \quad \text{where, } j = 0(1)n - 1$$

Circular Systematic Sampling (Contd.)

Table - 3: Showing sample of size $n = 5$ is selcted by circular systematic sampling :

Sample Sl. No.	1	2	3	4	5
Values of j	0	1	2	3	4
Values of $(5 + j \times 542)$	5	1089	1631	2173	2715
Values of $(5 + j \times 542 - X)$	-	-	-	-	3
Units selected	Wheat	Paddy	Paddy	Paddy	Wheat
$p_j = x_j/x$	0.0018	0.7205	0.7205	0.7205	0.0018
Cropped area (Ha.) (y_j)	0.95	683.9	683.9	683.9	0.95

So, an **unbiased estimator** of **total cropped area, Y** is :

$$\hat{Y}_{PS} = \frac{1}{5}\sum_{j=1}^{5}\frac{y_j}{p_j} = 780.6323 \text{ (in Ha.)}$$

Example (Contd.)

Table - 4: Showing summary above findings in systematic sampling :

Population cropped area (Ha.)	Estimated cropped area (Ha.)	
913.57	By Linear Sys.	By Circular Sys.
	843.2641	780.6323
Bias in estimation	70.3059	132.9377

Interpenetrating subsampling

- Technique of interpenetrating sub-sampling was originally developed by **P.C. Mahalanobis** in 1936 as **Inter-Penetrating Network of Sub-samples (IPNS)**.

Interpenetrating subsampling

- Technique of interpenetrating sub-sampling was originally developed by **P.C. Mahalanobis** in 1936 as **Inter-Penetrating Network of Sub-samples (IPNS)**.
- It was in the context of studying correlated errors within same samples arising due to enumerators' effect in large scale surveys conducted in India.

What is Interpenetrating Sampling?

- Interpenetrating sampling technique consists of drawing the samples from the same universe

What is Interpenetrating Sampling?

- Interpenetrating sampling technique consists of drawing the samples from the same universe
- In the form of two or more samples, selected in an identical manner and

What is Interpenetrating Sampling?

- Interpenetrating sampling technique consists of drawing the samples from the same universe
- In the form of two or more samples, selected in an identical manner and
- Each capable of providing a valid estimate of the population parameter

What is Interpenetrating Sampling?

- Interpenetrating sampling technique consists of drawing the samples from the same universe
- In the form of two or more samples, selected in an identical manner and
- Each capable of providing a valid estimate of the population parameter
- In some studies Mahalanobis applied different treatments to different sets of sub-samples to compare effects of treatments

What is Interpenetrating Sampling? (Contd.)

- Suppose a sample is selected in the form of k subsamples where each subsample being selected by the same sampling procedure.If each subsample gives a valid estimate of population parameter θ Then the network of subsamples forms **interpenetrating subsampling.**

What is Interpenetrating Sampling? (Contd.)

- Suppose a sample is selected in the form of k subsamples where each subsample being selected by the same sampling procedure.If each subsample gives a valid estimate of population parameter θ Then the network of subsamples forms **interpenetrating subsampling.**
- If the subsamples are drawn independently then an unbiased estimator of the sampling variance is obtained.

Estimator of the sampling variance

Let $t_1,\ t_2, \ldots\ldots\ldots\ldots.,\ t_k$ be unbiased estimator of θ such that $E(t_i) = \theta\ .\ \forall\theta, i = 1(1)k$.
Then the overall estimator of θ is given by $\bar{t} = \frac{1}{k}\sum_{i=1}^{k} t_i$
$E(\bar{t}) = E(\frac{1}{k}\sum_{i=1}^{k} t_i) = \theta\ \forall\theta$
An U.E of $Var(\bar{t})$ is given by $\widehat{Var(\bar{t})} = \frac{1}{k(k-1)}\sum_{i=1}^{k}(t_i - \bar{t})^2$
In replicated sampling $t_i's$ are **uncorrelated** due to independent selection .However, estimation of this variance becomes extremely simple in replicated sampling .

Interpenetrating Sampling - Uses

- Provide control in data collection and processing stage.

Interpenetrating Sampling - Uses

- Provide control in data collection and processing stage.
- Provide a non-parametric method of computing.

Interpenetrating Sampling - Uses

- Provide control in data collection and processing stage.
- Provide a non-parametric method of computing.
- Provides basis of analytical studies.

Interpenetrating Sampling - Uses (Contd.)

- Provide correction for bias in ratio type estimator

Interpenetrating Sampling - Uses (Contd.)

- Provide correction for bias in ratio type estimator
- Examine factors for different sources of variation, e.g, enumerators, filed schedules, different methods of data collection and processing .

Interpenetrating Sampling - Uses (Contd.)

- Provide correction for bias in ratio type estimator
- Examine factors for different sources of variation, e.g, enumerators, filed schedules, different methods of data collection and processing .
- Probability of sub-sample range covering median of estimator

Main Purpose of Interpenetrating Sampling

- Interpenetrating sub-samples can be used to secure information on non-sampling errors such as difference arising from differential interviewer bias, different method to elicit information, etc.

Acknowledgement

We would like to thank and express out heartiest gratitude to our respected **Prof. Biswajit Roy** sir for providing us with this presentation and help usto learn some interesting result and facts regarding the topic. We shall remain ever grateful to him for pushing our limit into real life working field.

References

M_N_Murthy_Sampling_Theory_and_Methods_Statistical_

References

- M_N_Murthy_Sampling_Theory_and_Methods_Statistical_
- https://www.unsiap.or.jp/e-learning/el_material/5_Agri/rap_Sampling_Indonesia/12_M