
Sampling Methods

Objectives of the module

By the end of this module, you will:

- Understand the methodology of first and second stage sampling.
- Be able to decide when and how to apply each sampling method.
- Understand the purpose of applying PPS (Probability Proportional to Size).
- Be able to undertake segmentation when selecting households in the field.

3 Sampling Methods

There are 3 main probability sampling methods:

1. **Simple random sampling** → Households selected directly.
2. **Systematic random sampling** → Households selected directly.
3. **Cluster sampling** → (2 Steps)
 1. Select Clusters.
 2. Select Households from within those clusters.

1: Simple Random Sampling

Sampling units are selected randomly from a sampling frame.

- For small geographically concentrated populations.
- Requires a **complete and recent list** of basic sampling units. This is rare in emergency situations.
- Number each sampling unit then pick every HH or individual randomly using a random numbers table (or via ENA).

Random Selection of Sampling Units

- Pick a number from a bag or a hat.
- Random Numbers Table
- ENA software



Random Number Table

Range from to Numbers



Table de nombres aléatoire

Intervalle: 1 à 2000, Nombres: 200

639	1108	834	1570	191	660	1573	260	1169	1878	744	662
1609	697	265	1052	1836	963	1977	1243	1427	633	1553	481
577	1988	407	1396	1712	1060	1713	463	1686	1999	1126	
1955	1410	599	1281	849	264	1705	1688	657	1111	1707	100
645	296	1856	1420	1433	610	985	345	350	1733	581	746
1033	1363	17	1401	897	1772	117	1247	248	39	1463	810
1263	310	741	354	1005	1315	831	211	455	617	1809	450
1421	1462	1823	120	1082	869	1143	1975	1736	830	1956	
1191	1124	843	397	1574	1081	977	77	1203	1200	661	9
863	1589	1551	1676	1727	820	1519	1000	1738	1747	85	
1272	1348	106	763	1782	1035	1945	1390	1546	737	162	143
353	1523	526	281	1921	776	440	773	1469	503	774	288
388	769	693	1997	1323	451	1408	1399	1062	758	797	
1319	1414	751	881	826	1984	142	1312	295	1576	1716	
1193	1483	1020	1376	301	65	124	543	1096	1072	1911	
1721	647	232	1202	1099	1208	1615	1262	1631	921	980	73
172	1370	974	98	946	318	266	1182	724	1221	59	541
1531	1042										

Simple Random Sampling

(Draw sample of 3 households)

<u>Number</u>	<u>Household</u>
1	Aweno
2	Serrano
3	Coopsman
4	Franco
5	Oka
6	Parvanta
7	Roquefort
8	Moki
9	Stirling

Random number table

7648	2352	6959	1937
2554	6804	9098	4316
4318	2346	7276	1880
7136	9603	0163	3152
7000	2865	8357	4475
9804	0042	1106	7949
2932	9958	9582	2235
1140	1164	7841	1688

Simple Random Sampling

(Draw sample of 3 households)

<u>Number</u>	<u>Household</u>
1	Aweno
2	Serrano
3	Coopsman
4	Franco
5	Oka
6	Parvanta
7	Roquefort
8	Moki
9	Stirling

Random number table

7648	2352	6959	1937
2554	6804	9098	4316
4318	2346	7276	1880
7136	9603	0163	3152
7000	2865	8357	4475
9804	0042	1106	7949
2932	9958	9582	2235
1140	1164	7841	1688

Systematic Random Sampling

- Method can be used for:
 - ▣ Small scale surveys in geographically concentrated populations.
 - ▣ **With or without a list** of sampling units.
- If there is no list, dwellings must be organized in a **defined geometric order** (camps, streets). Total number must be known.

Systematic Random Sampling

1. Determine the sampling interval:

Sampling interval = $\frac{\text{total number of sampling units in population}}{\text{number of sampling units in sample}}$

2. Select the first sampling unit:

- ▣ Choose a number **between 1 and the sampling interval**.
- ▣ If the sampling interval contains a decimal, always **round down** for the selection of the **1st sampling unit**.

3. Select the following sampling units:

- ▣ Number of the previous sampling unit + sampling interval.

Systematic Random Sampling

(Draw sample of 3 households)

<u>Number</u>	<u>Household</u>
1	Aweno
2	Serrano
3	Coopsman
4	Franco
5	Oka
6	Parvanta
7	Roquefort
8	Moki
9	Stirling

$$SI = ?$$

$$SI = 9/3 = 3$$

Random number table

7648	2352	6959	1937
2554	6804	9098	4316
4318	2346	7276	1880
7136	9603	0163	3152
7000	2865	8357	4475
9804	0042	1106	7949
2932	9958	9582	2235
1140	1164	7841	1688

Systematic Random Sampling

(Draw sample of 3 households)

<u>Number</u>	<u>Household</u>
1	Aweno
2	Serrano
3	Coopsman
4	Franco
5	Oka
6	Parvanta
7	Roquefort
8	Moki
9	Stirling

(SI = 9/3 = 3)

1. Random Start Number: between 1-3 (**2**).
2. Add SI to Random Start Number.

Random number table

7648	2 352	6959	1937
2554	6804	9098	4316
4318	2346	7276	1880
7136	9603	0163	3152
7000	2865	8357	4475
9804	0042	1106	7949
2932	9958	9582	2235
1140	1164	7841	1688

Selection of the sampling units

- If you obtain a sampling interval with a decimal, use to the following rule of thumb:

0.0-0.2	Round down
0.3-0.7	Alternate
0.8-0.9	Round up

Rounding Example

Total HH : 340							
To choose: 25							
Sampling Interval: $340/25=13.6$							
Random start number: 11							
	Rounding Up	Rounding down	Alternating	Exact			
1st HH	11 14	11 13	11 13/14	11 13.6			
2nd HH	25	24	24	24.6			
3rd HH	39	37	38	38.2			
4th HH	53	50	51	51.8			
5th HH	67	63	65	65.4			
25th HH	347	323	335	337			

Systematic Random Sampling

Example:

The **village of Kawar** is composed of 3400 HH numbered sequentially from 1 to 3400. You need a sample of 250 HH.

The sampling interval (SI) = $\frac{3400}{250} = 13.6$

You ask someone to randomly choose a number between 1 and 13 and they choose **11**. This number is the equivalent to the 1st household to survey.

Subsequent HHs are chosen as follows:

Sampling Interval= 13.6

HH (Dumalag)	Calculations	HH Number
1st HH	11	11
2nd HH	11+13	24
3rd HH	24+14	38
4th HH	38+13	51
5th HH	51+14	65
6th HH	65+13	78
7th HH	78+14	92
8th HH	92+13	105
9th HH	105+14	119
etc.	etc.	etc.

Ex. Sampling Interval= 13.2

HH Rank	Calculations	HH Number
1st HH	11	11
2nd HH	11+13	24
3rd HH	24+13	37
4th HH	37+13	50
5th HH	50+13	63
6th HH	63+13	76
7th HH	76+13	89
8th HH	89+13	102
9th HH	102+13	115
etc.	etc.	etc.



5 min

Exercises

1. The city you are surveying contains 8400 HH. You need a sample of 556 HH. **What will be the sampling interval?** Choose the first and second HH to be surveyed.
2. A survey takes place in a city of 3800 HH. You need to select 680 HH for your sample. **What will be the sampling interval?** Choose the first and the following 3 HH to be surveyed.

3: Cluster Sampling

When:

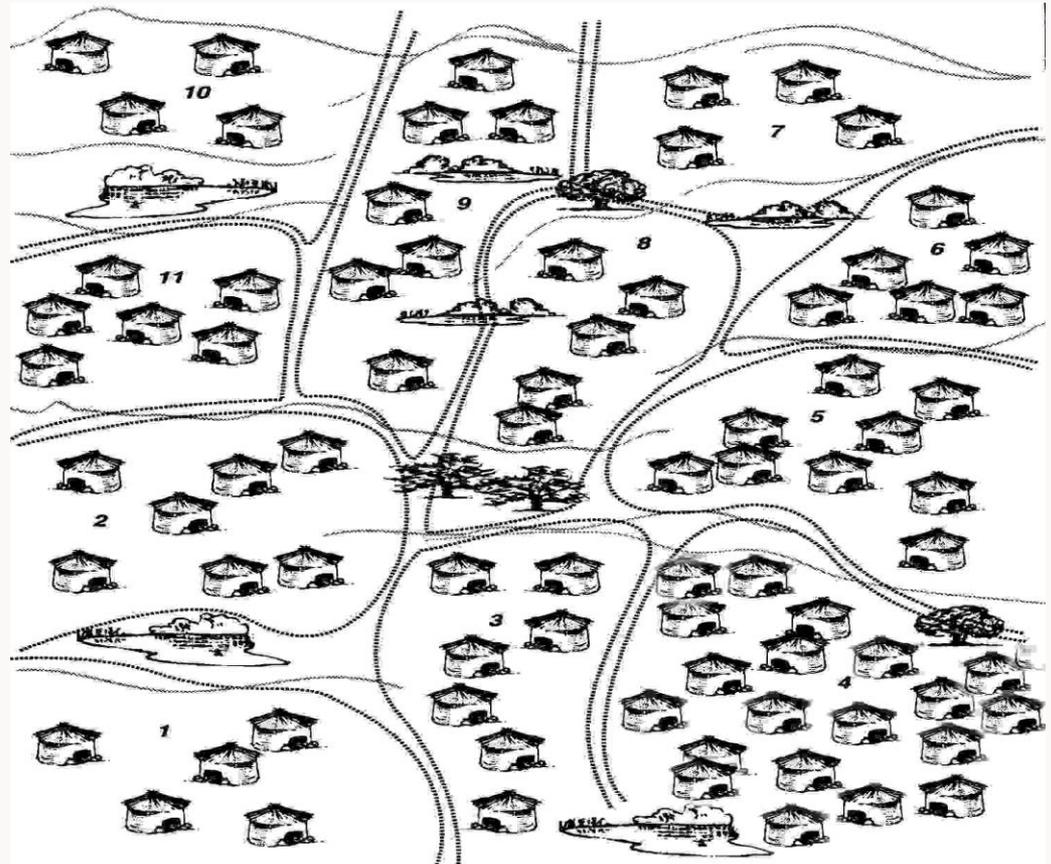
1. You **do not** have a **complete and updated list** of basic sampling units, or
2. Even if you have a list, but it is impractical to carry out systematic or simple random sampling because the survey population is **geographically dispersed**,

Use cluster sampling design

Cluster Sampling

Objective:

To choose smaller geographic areas (**clusters**) in which simple or systematic random sampling can be done.



3 Sampling Methods

Whole Sampling Population

1. Simple

2. Systematic

3. Cluster

1st: Clusters selected from whole population. (typhoon-affected municipalities of the Philippines).

3a. Simple

3b. Systematic

HH selected directly from whole population (population of Dumalag).

2nd: HH selected from within selected clusters (municipalities) (using simple or systematic).

Cluster Sampling Stages

Method applied in **two** stages:

Stage 1: Random selection of clusters from the total number of geographical units. Selection of clusters is done during the planning stage ***in the office***.

Stage 2: Random selection of households within the clusters. This stage is done ***in the field***.

Exercise- Definitions

Objective of your survey: To assess nutritional status of children 6-59 months and household access to safe water and latrines in typhoon-affected municipalities of the Maungdaw Tsp.

1. What is your sampling universe?
2. What is your sampling frame?
3. What is your sampling unit?
4. Who is your respondent?
5. Who is your survey subject?

Primary Sampling Unit

- **General**

- The sampling unit at the ***first*** stage sampling in cluster surveys.

- **Philippines**

- All typhoon-affected villages of the Maungdaw Tsp.

Basic (Elementary) Sampling Unit

□ General

- ▣ The sampling unit at the second stage sampling in cluster surveys.

➤ Maungdaw Tsp

- ▣ All households within the selected typhoon-affected villages of the Maungdaw Tsp.

Cluster sampling

Advantages

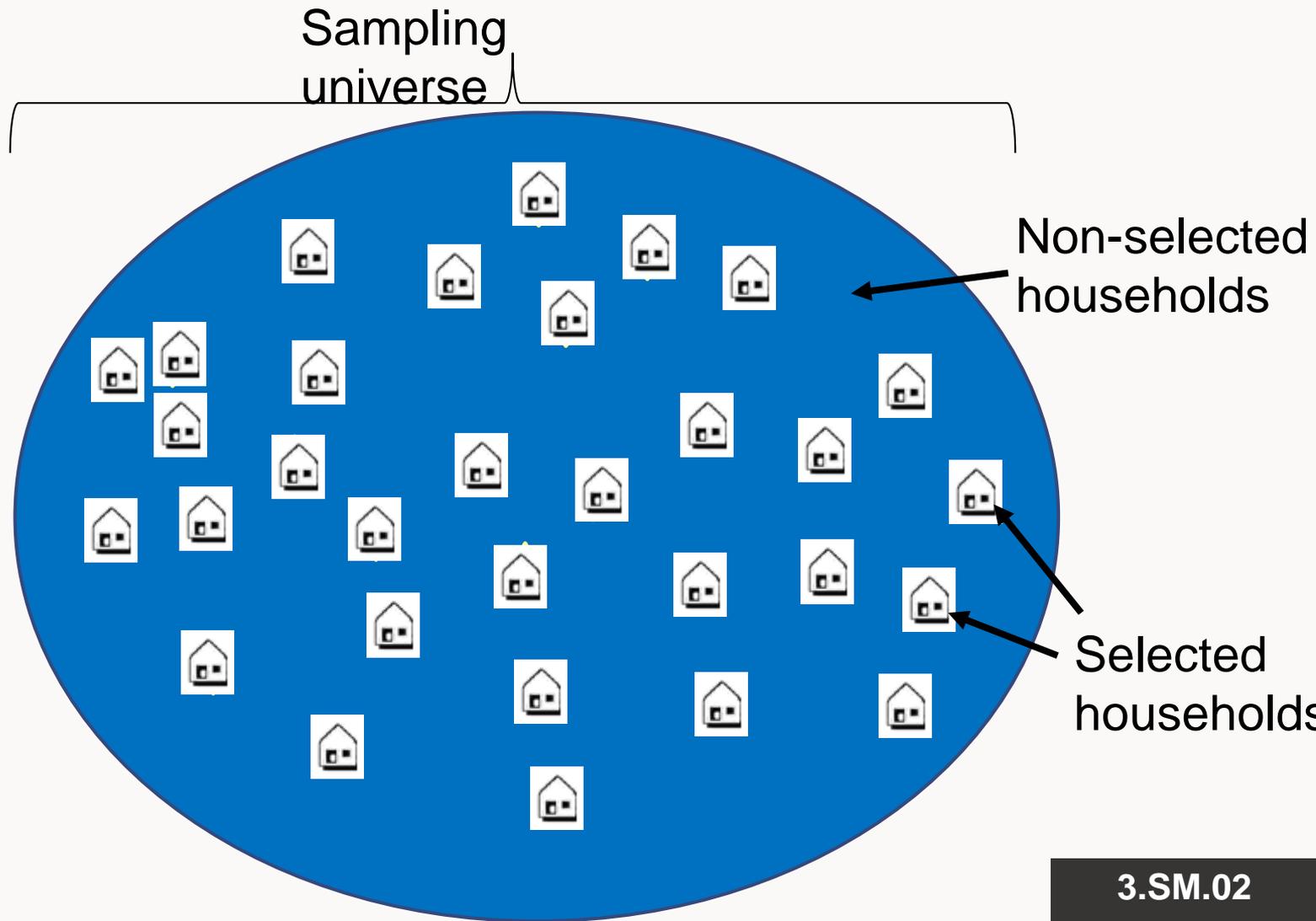
- Cheaper - basic sampling units closer together (practical).
- No need for a complete list of basic sampling units.
- Better data quality.

Disadvantages

- Decreased precision of estimate.
- Calculation of p-values and confidence limits more complicated.

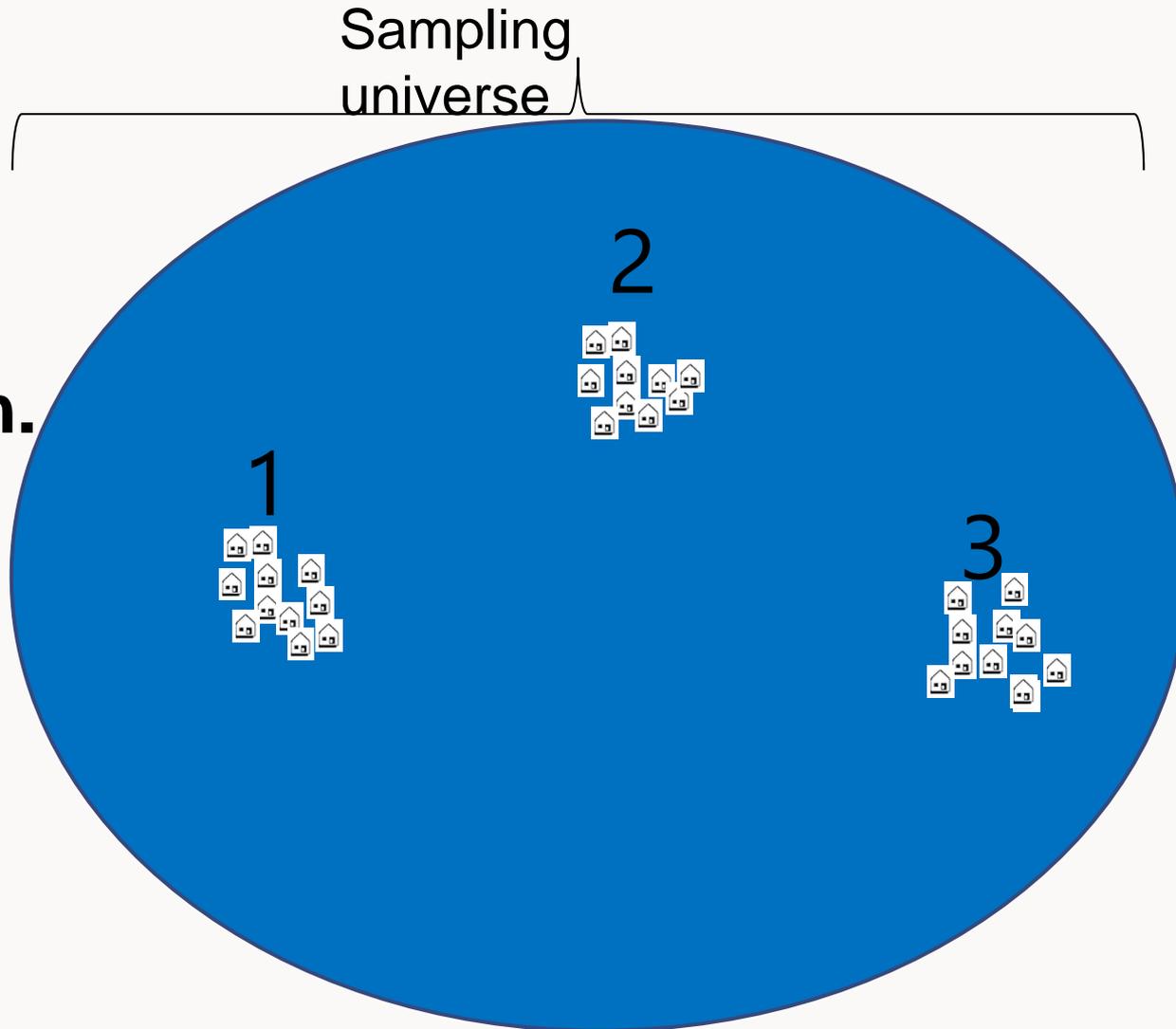
SRS vs. Cluster Sampling

**Simple
random
sampling:
30 HH.**



SRS vs. Cluster Sampling

Cluster sampling:
30 HHs in
3 clusters
of 10 each.



Exercise

There is a lottery to win 1 plane ticket to Canada.

Do you have a higher chance of winning if the lottery is played by the 30 people in this training, or if it played by the entire population of Yangon?

Cluster sampling: Stage 1

What do you do if the villages you are surveying are different sizes?

Use Probability Proportional to Size (PPS).

Cluster sampling

2 events need to happen in order for 1 child to be selected in Cluster sampling:

- 1. Their village (cluster) must be selected.**
- 2. Their household must be selected.**

$$P_{\text{village}} \times P_{\text{household}} = P_{\text{child being selected}}$$

Cluster Sampling with PPS

Probability of 2 events occurring together
= Probability of event 1 X probability of event 2

What is the probability of getting 2 heads when you flip 2 coins?



+



0.5

x

0.5

=

0.25

Probability Proportional to Size

- ▣ The probability of each **cluster (Stage 1)** being selected is **proportional to the size of its population (PPS)**.
- ▣ This assures that each **basic sampling unit (ex. HH) (Stage 2)** within the survey area has an **EQUAL (known and non-zero)** chance of being selected.

PPS

If we want all children to have the **SAME** probability of being selected, we must **adjust P_{village} using PPS.**

*Number of HH to visit per cluster does not change!

How to compensate for different P_{villages} ?

Using PPS in ENA, larger villages have a larger chance of being selected, and smaller villages have a smaller chance of being selected.

PPS Method: $P_{\text{household}}$

$$P_{\text{village}} \times P_{\text{household}} = P_{\text{child being selected}}$$

ex: You need to select **20 HHs per cluster**:

Village A

200 HH

$$\begin{aligned} P_{\text{household A}} &= 20/200 \\ &= 1/10 \\ &= \mathbf{10\%} \end{aligned}$$

Village B

40 HH

$$\begin{aligned} P_{\text{household B}} &= 20/40 \\ &= 1/2 \\ &= \mathbf{50\%} \end{aligned}$$

PPS Method: $P_{\text{household}}$

$$P_{\text{village}} \times P_{\text{household}} = P_{\text{child being selected}}$$

Village A

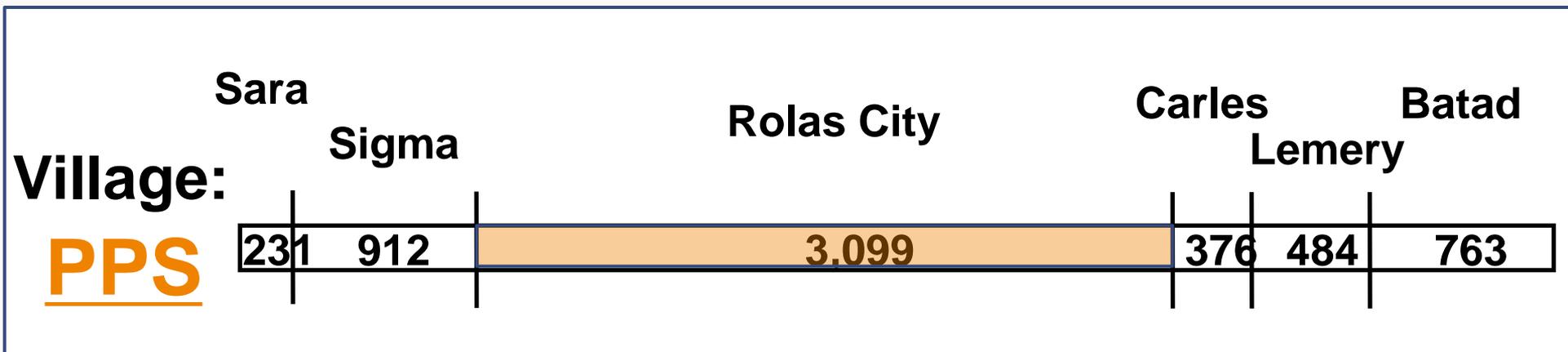
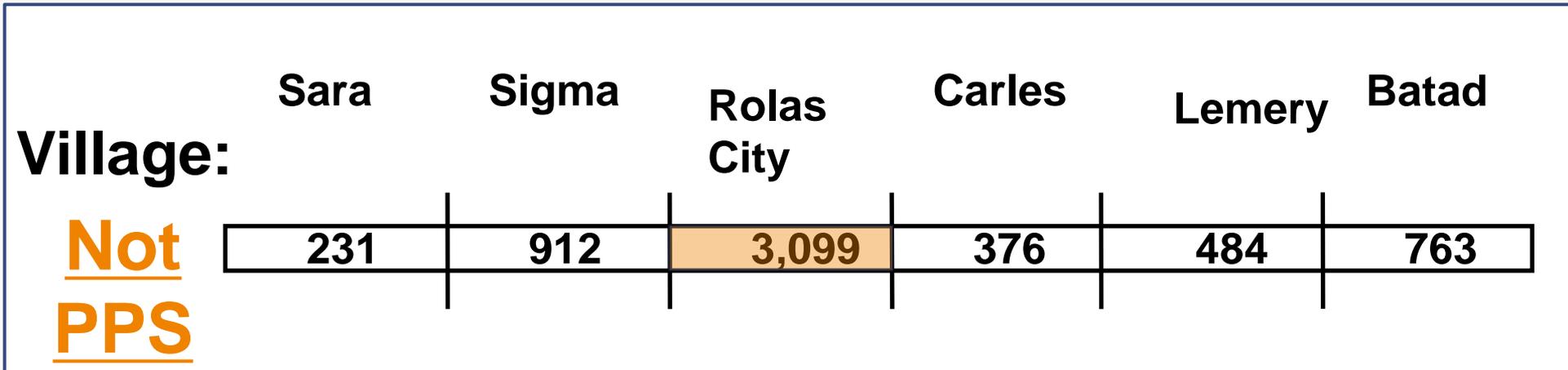
10%

Village B

50%

If the probability of selecting a village is equal, households in Village B are **5 times** more likely to be selected than households in Village A.

Probability Proportional to Size (PPS)



PPS Method: P_{village}

$$P_{\text{village}} \times P_{\text{household}} = P_{\text{child being selected}}$$

Village A (200HHs)

↑ chance of
being selected

$$5\% \times 10\% = \underline{0.5\%}$$

Village B
(40HHs)

↓ chance of
being selected

$$1\% \times 50\% = \underline{0.5\%}$$

Now $P_{\text{child being selected}}$ for both villages are EQUAL.

Stage 1: Applying PPS

1. **Create a list of villages** (or the smallest geographical area for which population data estimates are available) indicating their **population sizes**.
2. Calculate the **cumulative population**.
3. Apply the **systematic random sampling** method to determine the sampling interval & select a random start number.
4. **Select clusters** starting with the random start number and then by adding the sampling interval.

Stage 1- PPS : Creating a list of villages (municipalities)

VILLAGES	Estimated Total Population	Cumulative Population	Range Allocated
Altavas	500	500	1 - 500
Kalibo	400	900	501 - 900
Libacao	160	1060	901 - 1060
Dumalag	650	1710	1061 - 1710
Dumarao	520	2230	1711 - 2230
Pilar	640	2870	2231 - 2870
Dao	700	3570	2871 - 3570
Balasan	104	3674	3571 - 3674
Ormoc City	1480	5154	3675 - 5154
...
Banate	100	28,000	27,900- 28,000

With a total of 135 villages, the total population size is 28,000.

Stage 1- PPS: Assigning clusters

VILLAGES	Estimated Total Population	Cumulative Population	Range Allocated	CLUSTERS
Altavas	500	500	1 - 500	1
Kalibo	400	900	501 - 900	
Libacao	160	1060	901 - 1060	2
Dumalag	650	1710	1061 - 1710	
Dumarao	520	2230	1711 - 2230	3
Pilar	640	2870	2231 - 2870	4
Dao	700	3570	2871 - 3570	5
Balasan	104	3674	3571 - 3674	
Ormoc City	1480	5154	3675 - 5154	6-7

1. With 35 clusters to select, our sampling interval is $28,000/35= 800$.
2. Our random start number is 1-800 \rightarrow 242.

Second Stage Sampling

Cluster Sampling: Stage 2

- Several methods can be used to select households in the field:
 - ▣ Simple random sampling.
 - ▣ Systematic random sampling.
 - ▣ Segmentation followed by simple or systematic sampling.

Choice of method depends on what information you have in the field. Ideally, the same method should be used in each cluster. The method of selecting households is defined ***a priori***.

Simple Random Sampling

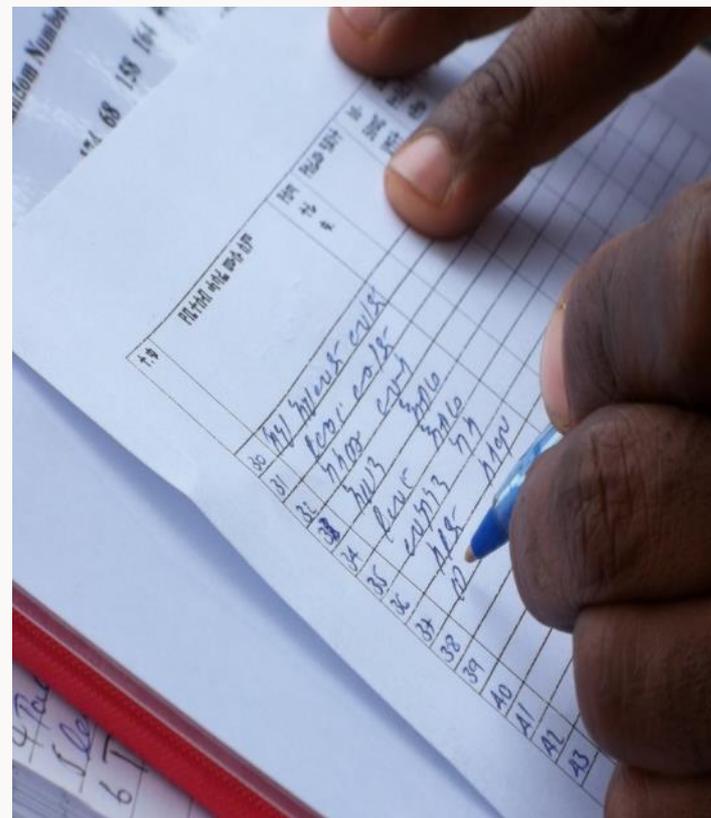
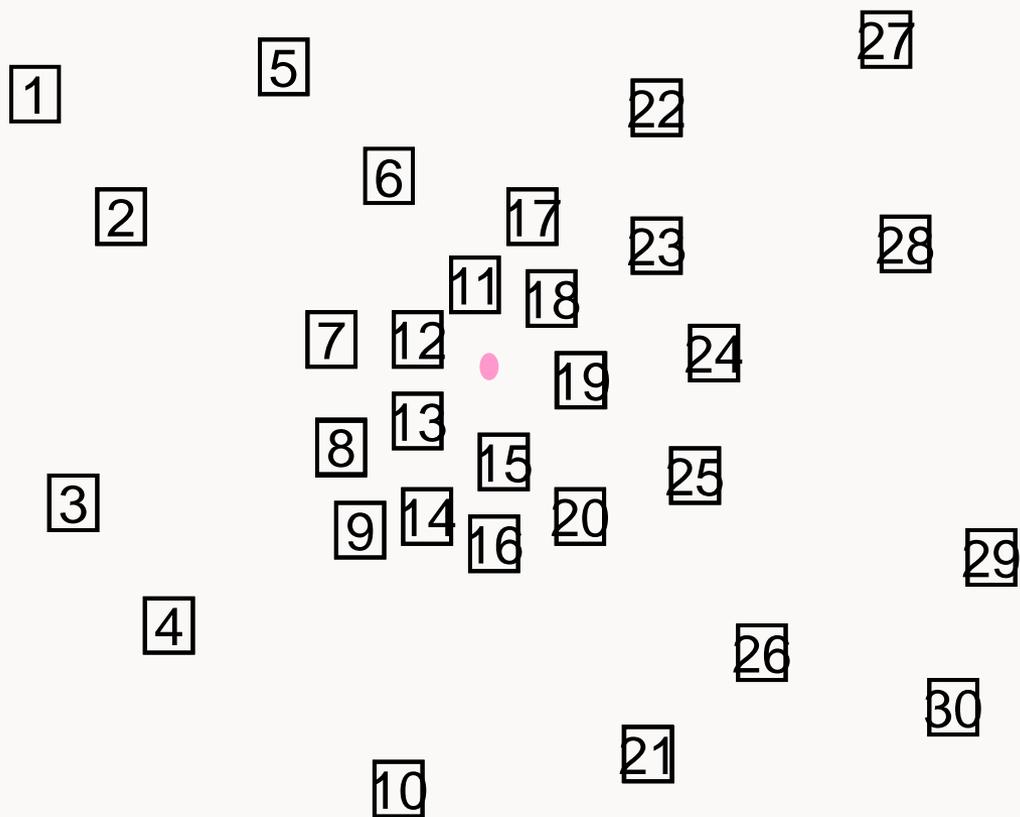
When:

- ▣ A **complete and updated list** of all the households in the cluster is **available**.

OR/

- ▣ A **list can be created** using key informants and walk-around techniques in the cluster (usually less than 200 households).

Random Selection of Households



Systematic Random Sampling

When:

- ▣ **Complete list** of basic sampling units (HH) (as in Simple Random).

or/

- ▣ When households are arranged in a **defined geometric order & total number HH** in the village is known/can be estimated.

→ This is an **advantage** over simple random sampling.

Systematic Random Sampling- Selection of HHs

1. Determine how many HHs in your village/cluster.

2. Determine the sampling interval:

$$\square \text{ Sampling interval} = \frac{\text{total number of HHs in cluster}}{\text{number of HHs to be sampled}}$$

3. Select first household randomly between 1 and the sampling interval.

4. Select subsequent households:

$$\square (\text{Previous Household Number}) + (\text{Sampling Interval}).$$

Segmentation

When? If the population is **dispersed** or villages are **large** (> approximately 250 HHs).



If you have a **complete list of > 250 households** and they are **very close together**, segmentation does **NOT** need to be applied.

Segmentation

→ **Makes survey area more manageable.**

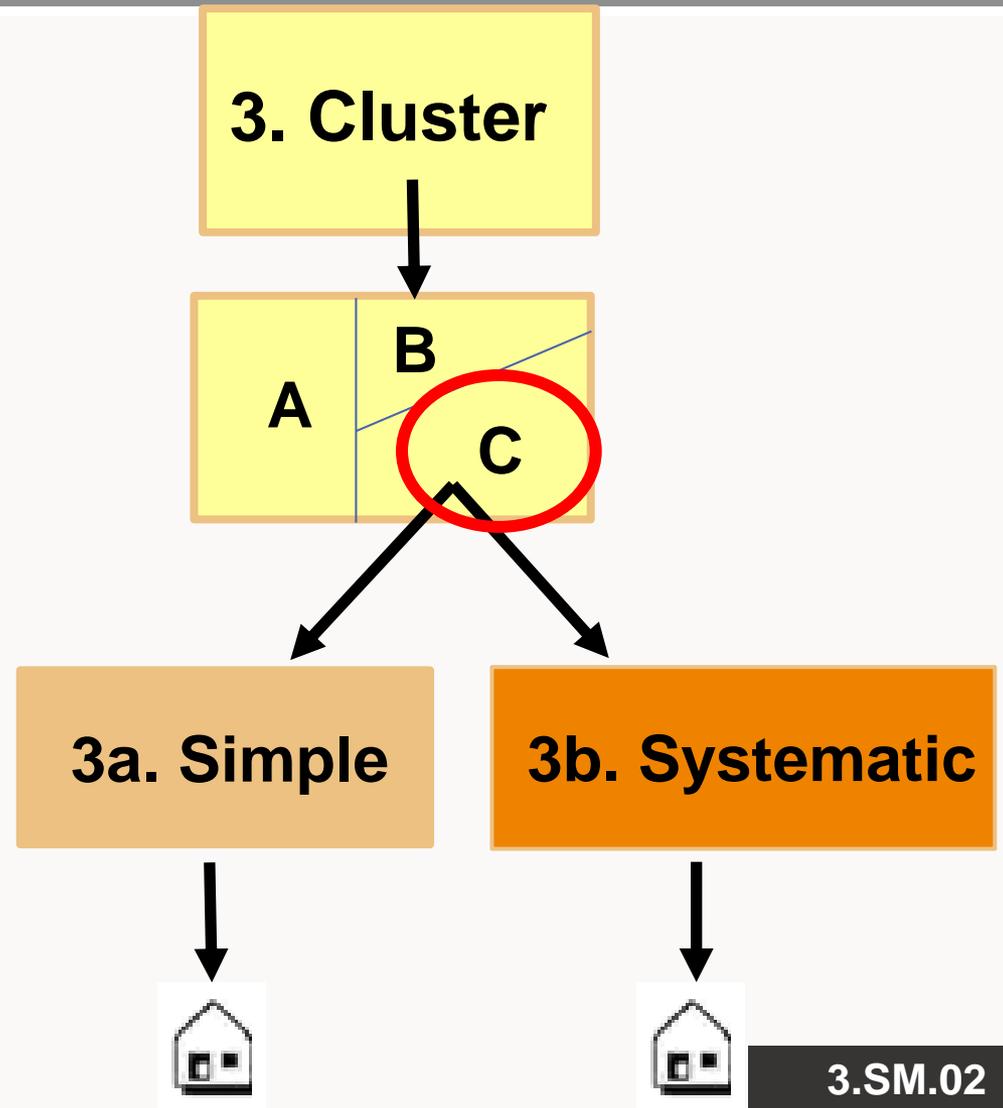
- **Divide the village according to:**
 - Existing administrative sub-divisions.
 - Natural barriers: river, road, mountain, etc.
 - Public places: market, schools, churches, mosques, etc.
- **Selection depends on the relative size of the segments.**

Segmentation

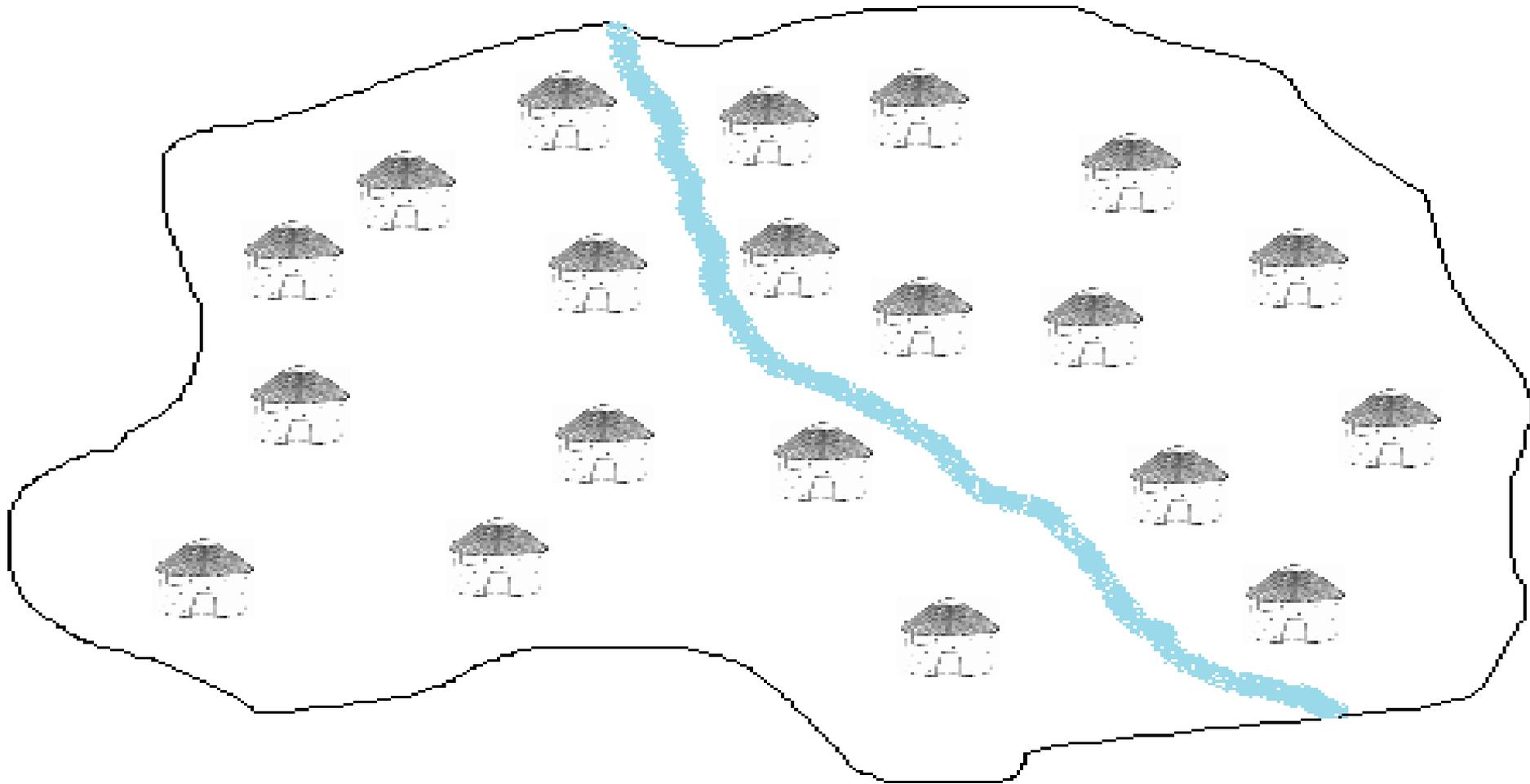
A cluster is selected from whole population.

That cluster is segmented.

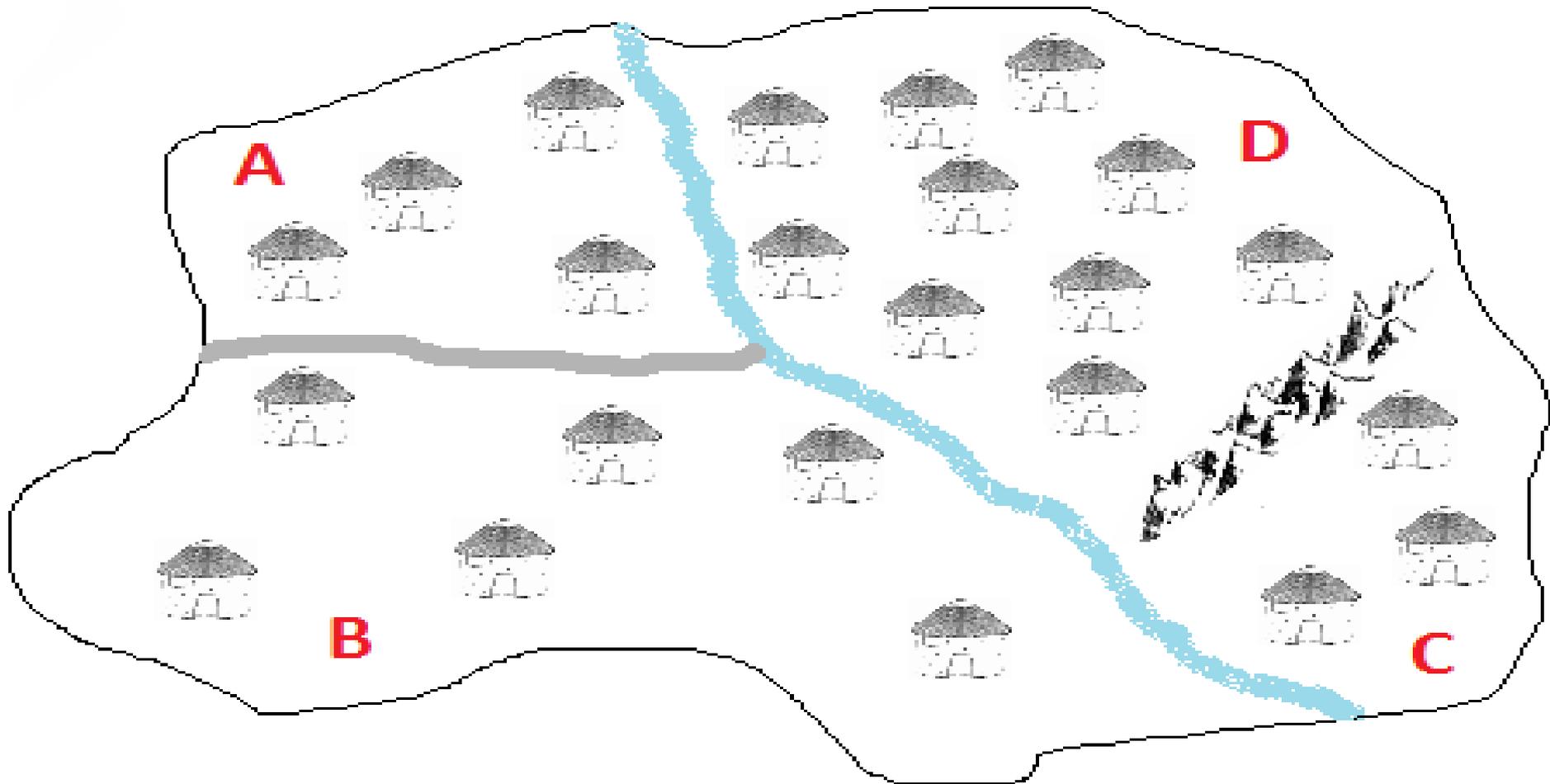
ONE segment is chosen, and HH selected from within that segment.



Segmentation: Equal Parts



Segmentation: Unequal Parts



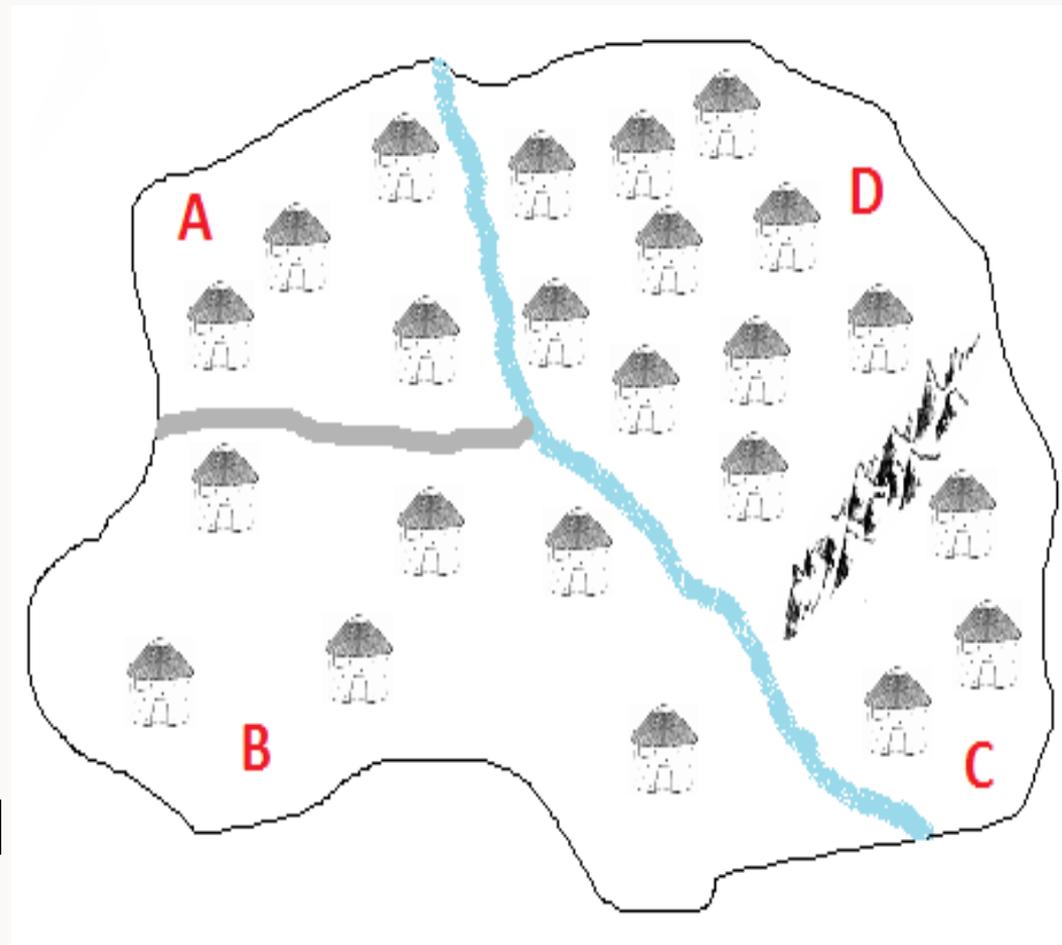
Segmentation: Unequal Parts

Example:

- A = 70 HH
- B = 100 HH
- C = 30 HH
- D = 190 HH

TOTAL = 390 HH

- Cluster size = 15 HH



Unequal parts: apply PPS

Segment	Population (HH)	Cumulative population	Intervals
A	70	70	1 – 70
B	100	170	71 - 170
C	30	200	171 - 200
D	190	390	201 - 390

Pick a 3-digit number from: 001- 390.

E.g. We picked: 167.

This number is within which segment?

Survey will be conducted in segment B

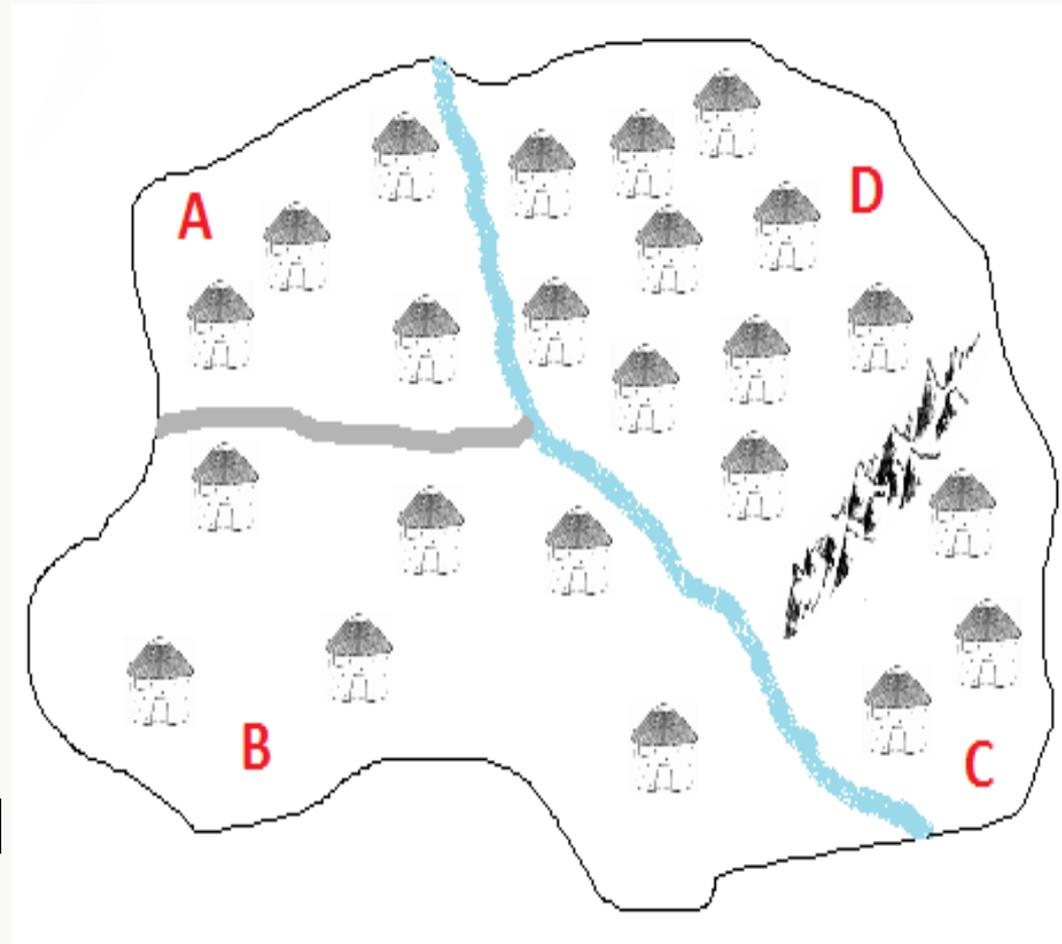
2 Clusters assigned to same village

Example:

- A = 220 HH
- B = 200 HH
- C = 80 HH
- D = 190 HH

TOTAL = 690 HH

- Cluster size = 15 HH



Recap on Sampling Methods

Type of Sampling	Requirements
Simple Random	<ul style="list-style-type: none">• <u>Complete and recent list</u> of basic sampling units.
Systematic Random	<ul style="list-style-type: none">• <u>With or without a list</u> of basic sampling units.• If <u>no list</u>, dwellings organized in a <u>defined geometric order</u>.
Cluster	<ul style="list-style-type: none">• <u>No list</u> of basic sampling units.• <u>If there is a list</u>, can still use cluster if population is geographically dispersed.

Recap on Sampling Methods

PPS:

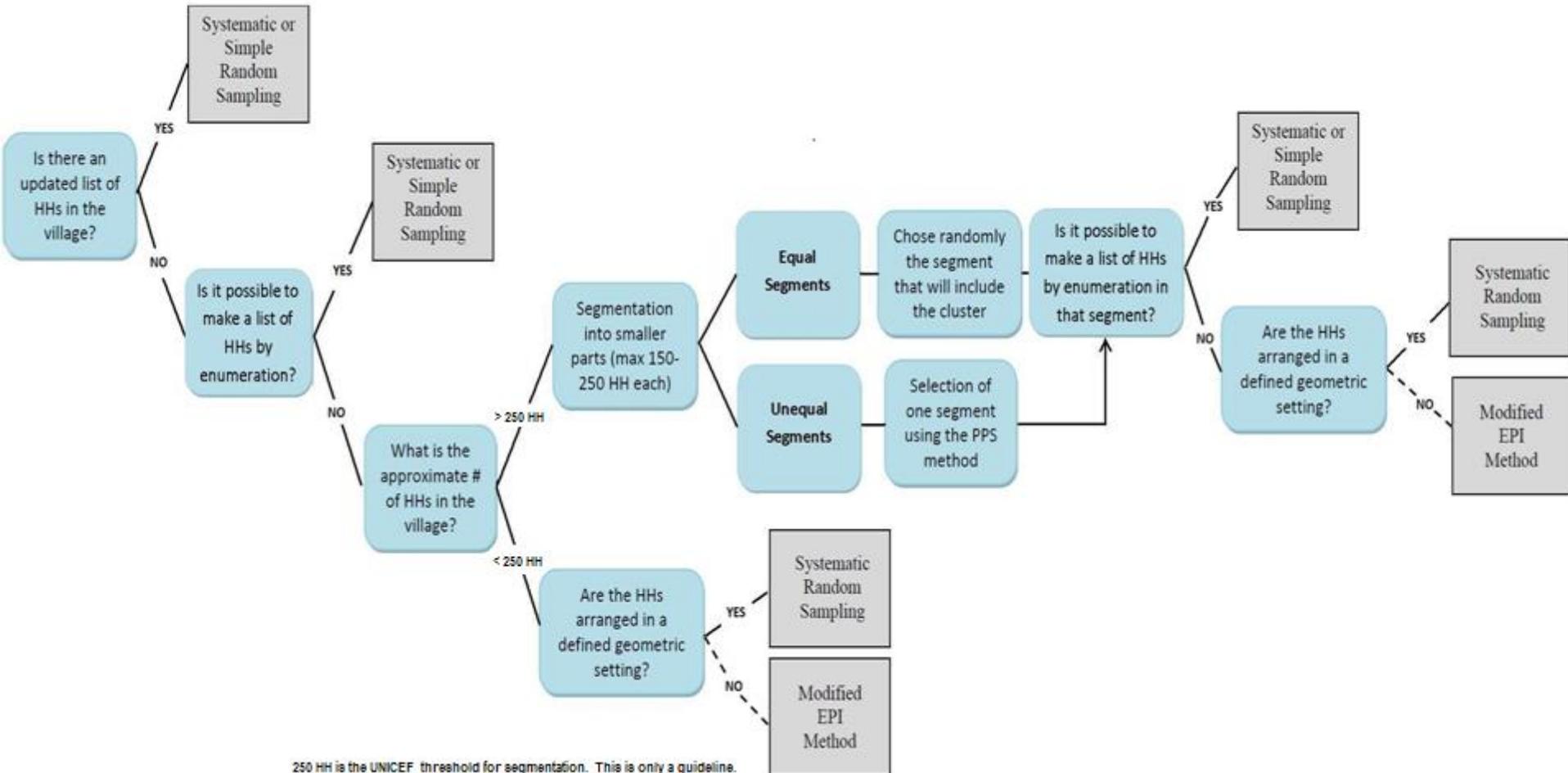
- Cluster sampling (ENA).
- Segmentation (if unequal segments).

Segments (# HH)	How to choose Segment for survey
Equal	Random
Unequal	PPS

Sampling – Decision Tree

- To use at the last stage of sampling.
- Questions to consider:
 - ▣ Do we have a list of the households in the village?
 - ▣ Is it possible to build a list?
 - ▣ What is the approximate number of households in the village?
 - ▣ What is the geometric setting of households in the village?

Sampling – Decision Tree



250 HH is the UNICEF threshold for segmentation. This is only a guideline.