SIAP

Statistical Institute for Asia and the Pacific

UNITED NATIONS
siap
Statistical Institute for
Asia and the Pacific

PROBABILITY PROPORTIONAL TO SIZE AND ONE STAGE CLUSTER SAMPLING

Regional Training Course on Sampling Methods for Producing Core Data Items for Agricultural and Rural Statistics

Jakarta, Indonesia ,29 Sep-10 October 2014.

UNITED NATIONS
siap
Statistical Institute for
Asia and the Pacific

LEARNING OBJECTIVES

At the end of this session participants are expected to:

1. Discuss the concept of probability proportional to size sampling;
2. Demonstrate knowledge of selection procedures for probability proportional to size sampling;
3. Explain the rationale of one stage clustering.

OVERVIEW OF THE PRESENTATION

- Probability selection
- Selection procedures
- Single stage clustering

Probability Proportional to Size Sampling

or PPS Sampling

SAMPLING WITH PROBABILITY PROPORTIONAL TO SIZE (PPS)

- Probability of selection is related to an auxiliary variable, Z, that is a measure of "size"

  Example

  Number of households

  Area of farms

- "Larger" units are given higher chance of selection than "smaller" units
- Selection probability of *i*th unit is

  i = 1,2, … , *N*

$$p_i = Z_i \Big/ \sum_{i=1}^{N} Z_i$$

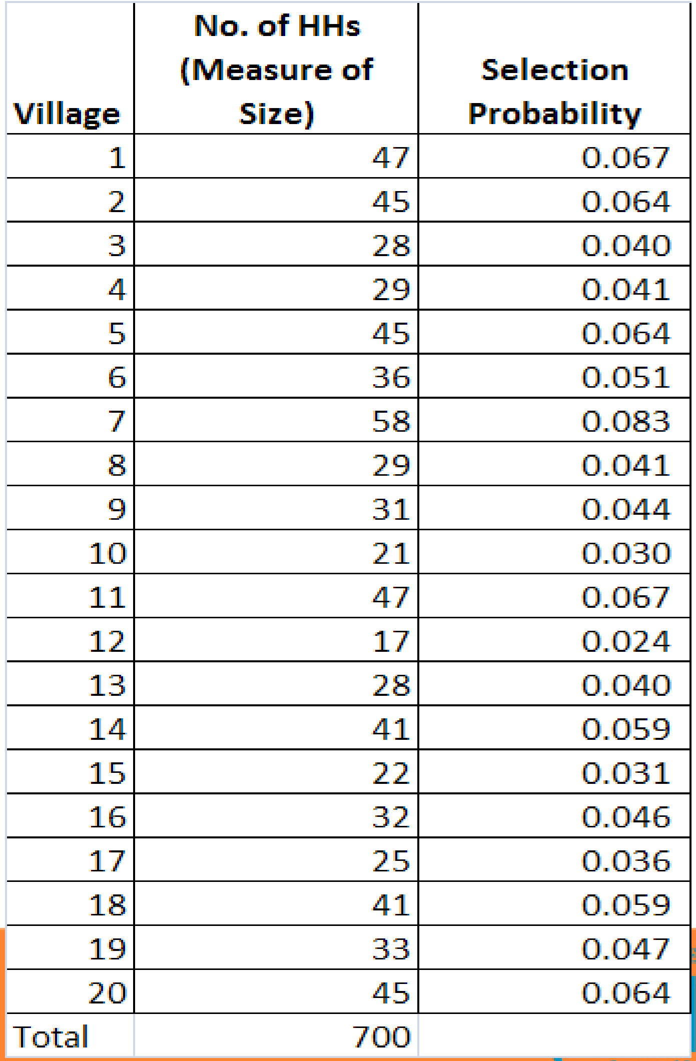
PPS SELECTION PROCEDURES

- Cumulative total method: with replacement
- Cumulative total method: without replacement
- PPS systematic sampling

PPS Selection

CUMULATIVE TOTAL METHOD

Select a sample of 5 villages using varying probability WR sampling, the size being the number of households



Village	No. of HHs (Measure of Size)	Selection Probability
1	47	0.067
2	45	0.064
3	28	0.040
4	29	0.041
5	45	0.064
6	36	0.051
7	58	0.083
8	29	0.041
9	31	0.044
10	21	0.030
11	47	0.067
12	17	0.024
13	28	0.040
14	41	0.059
15	22	0.031
16	32	0.046
17	25	0.036
18	41	0.059
19	33	0.047
20	45	0.064
Total	700	

Solution

- Sampling unit: **village**
- Measure of size: **number of households in village**

- Selection probability:

$$p_i = \frac{\text{number of HHs in village i}}{\text{total number of HHs}}$$

PPS Selection

Cumulative Total Method (Contd.)

- Write down cumulative total for the sizes Z_i, i=1,2..N
- Choose a random number r such that 1 $\leq r \leq Z$
- Select ith population unit if
- $T_{i-1} \leq r \leq T_i$ where

$$T_{i-1} = Z_1 + Z_2 + .. + Z_{i-1}$$

and

$$T_i = Z_1 + Z_2 + .. + Z_i$$

Village	No. of HHs (Measure of Size) (Z_i)	Cumulative Size (T_i)	Assigned Random Numbers
1	47	47	1 - 47
2	45	92	48 - 92
3	28	120	93 -120
4	29	149	121 - 149
5	45	194	150 - 194
6	36	230	195- 230
7	58	288	231 - 288
8	29	317	289 - 317
9	31	348	318 - 348
10	21	369	349 - 369
11	47	416	370 - 416
12	17	433	417 - 433
13	28	461	434 - 461
14	41	502	462 - 502
15	22	524	503 - 524
16	32	556	525 - 556
17	25	581	557 - 581
18	41	622	582 - 622
19	33	655	623 - 655
20	45	700	656 - 700
Total	700		

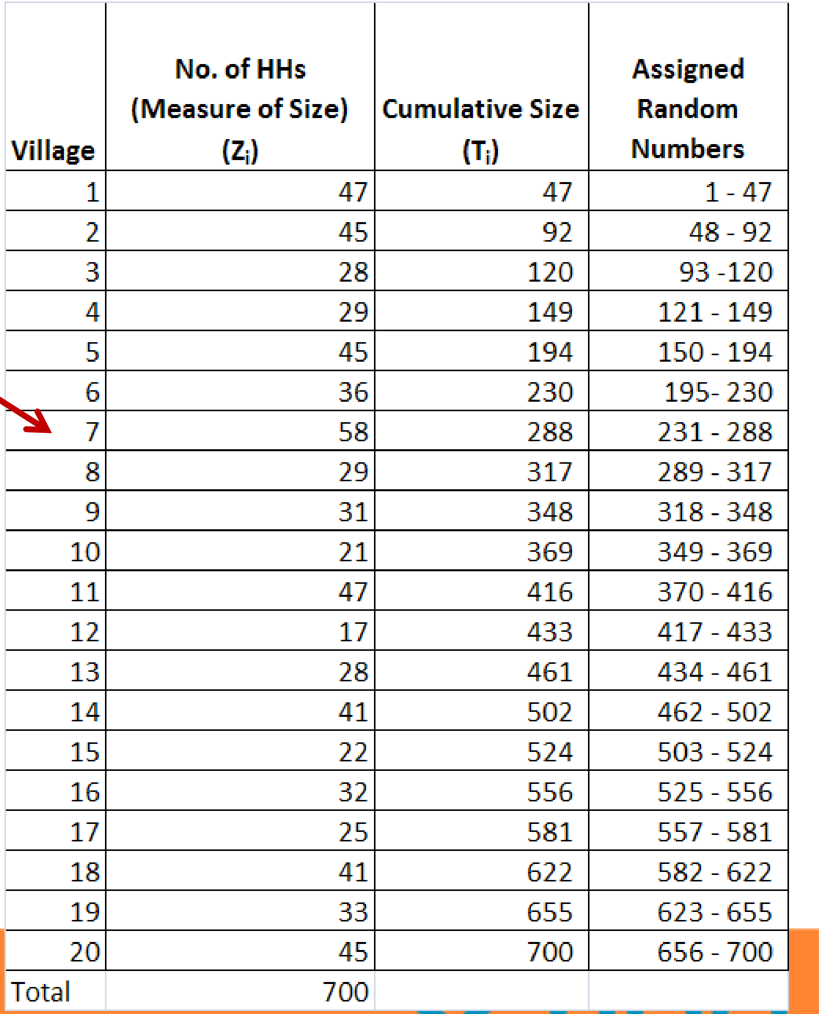
PPS Selection

Cumulative Total Method (Contd.)

- To select a village, a random number *r*, 1 ≤ *r* ≤ 700, is selected.
- Suppose *r* = 259,

  Since 231 ≤ 259 ≤ 288, the 7th village is therefore selected. The next 4 random numbers to be considered are 548, 170, 231, 505. Hence the required sample selected using PPS with replacement are 16th, 5th, 7th, 15th .

Note: The 7th village is selected twice.



Village	No. of HHs (Measure of Size) (Z_i)	Cumulative Size (T_i)	Assigned Random Numbers
1	47	47	1 - 47
2	45	92	48 - 92
3	28	120	93 -120
4	29	149	121 - 149
5	45	194	150 - 194
6	36	230	195- 230
7	58	288	231 - 288
8	29	317	289 - 317
9	31	348	318 - 348
10	21	369	349 - 369
11	47	416	370 - 416
12	17	433	417 - 433
13	28	461	434 - 461
14	41	502	462 - 502
15	22	524	503 - 524
16	32	556	525 - 556
17	25	581	557 - 581
18	41	622	582 - 622
19	33	655	623 - 655
20	45	700	656 - 700
Total	700		

PPS Selection

Cumulative Total Method (Contd.)

- For a PPSWR selection therefore the sample would be: 16th, 5th, 7th, 15th , with 7th village repeated.
- For a PPSWOR selection, we have to continue further to get 5 distinct units in the sample.
- Suppose the next random selected is *r* = 375,

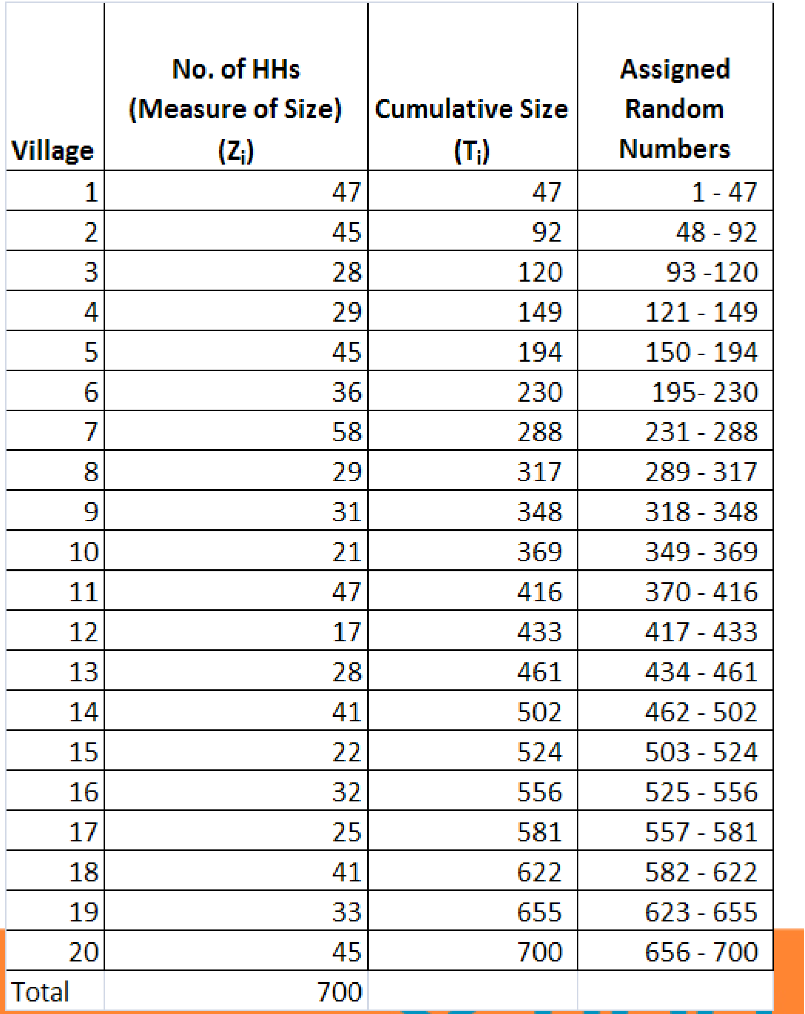
The required PPSWOR sample would be 16th, 5th, 7th, 15th & 11th

Village	No. of HHs (Measure of Size) (Z_i)	Cumulative Size (T_i)	Assigned Random Numbers
1	47	47	1 - 47
2	45	92	48 - 92
3	28	120	93 -120
4	29	149	121 - 149
5	45	194	150 - 194
6	36	230	195- 230
7	58	288	231 - 288
8	29	317	289 - 317
9	31	348	318 - 348
10	21	369	349 - 369
11	47	416	370 - 416
12	17	433	417 - 433
13	28	461	434 - 461
14	41	502	462 - 502
15	22	524	503 - 524
16	32	556	525 - 556
17	25	581	557 - 581
18	41	622	582 - 622
19	33	655	623 - 655
20	45	700	656 - 700
Total	700		

PPS Selection

PPS Systematic

- Derive cumulative totals for the sizes Z_i, i=1,2..N, and allot random numbers to different units.
- Calculate interval $k = Z_N / n$ (in this case 700/5 = 140)
- Select a random number r (say 101) from 1 to k; and obtain $r+k$, $r+2k$, $r+3k$, …, $r+(n-1)k$
- In this case, the selected cumulative sizes are 101, 241, 382, 523 & 664.



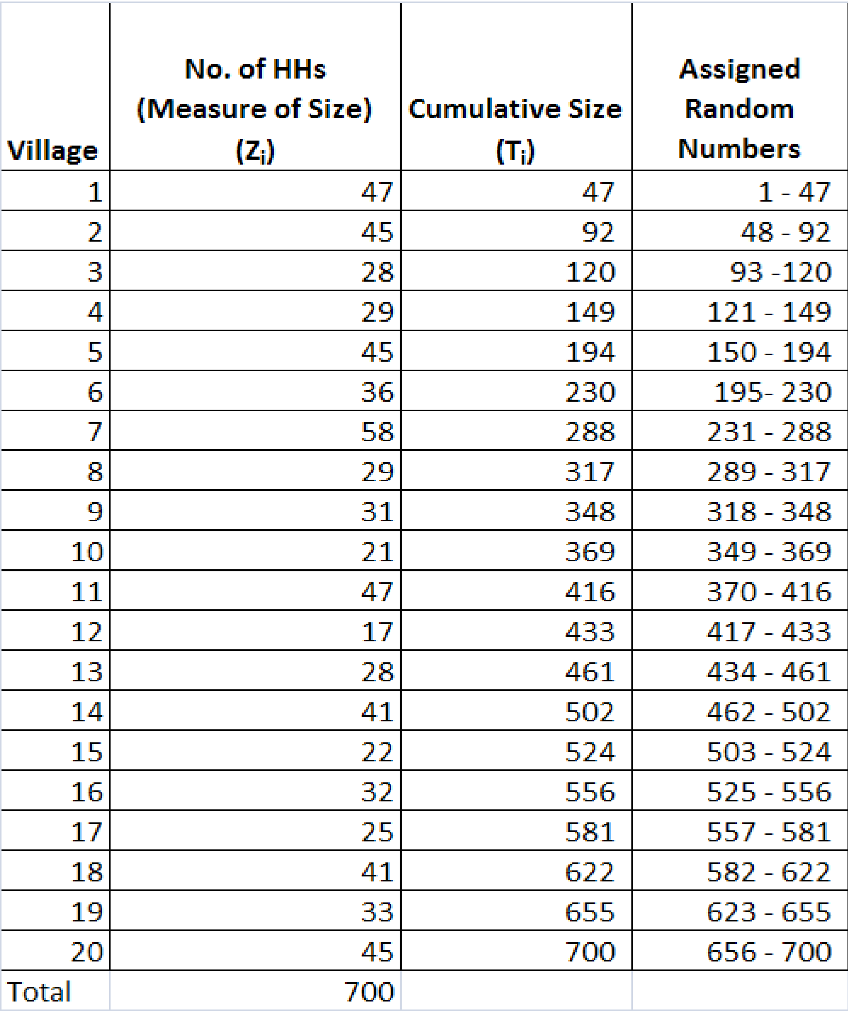
Village	No. of HHs (Measure of Size) (Z_i)	Cumulative Size (T_i)	Assigned Random Numbers
1	47	47	1 - 47
2	45	92	48 - 92
3	28	120	93 -120
4	29	149	121 - 149
5	45	194	150 - 194
6	36	230	195- 230
7	58	288	231 - 288
8	29	317	289 - 317
9	31	348	318 - 348
10	21	369	349 - 369
11	47	416	370 - 416
12	17	433	417 - 433
13	28	461	434 - 461
14	41	502	462 - 502
15	22	524	503 - 524
16	32	556	525 - 556
17	25	581	557 - 581
18	41	622	582 - 622
19	33	655	623 - 655
20	45	700	656 - 700
Total	700		

PPS Systematic (Contd.)

- Thus the selected units are:

  3rd (for 101),
  7th (for 241),
  11th (for 382),
  15th (for 523) &
  20th (for 664)

- ***Note:*** If any unit has size greater than *k*, it may be selected more than once.



Village	No. of HHs (Measure of Size) (Z_i)	Cumulative Size (T_i)	Assigned Random Numbers
1	47	47	1 - 47
2	45	92	48 - 92
3	28	120	93 -120
4	29	149	121 - 149
5	45	194	150 - 194
6	36	230	195- 230
7	58	288	231 - 288
8	29	317	289 - 317
9	31	348	318 - 348
10	21	369	349 - 369
11	47	416	370 - 416
12	17	433	417 - 433
13	28	461	434 - 461
14	41	502	462 - 502
15	22	524	503 - 524
16	32	556	525 - 556
17	25	581	557 - 581
18	41	622	582 - 622
19	33	655	623 - 655
20	45	700	656 - 700
Total	700		

Cluster Sampling

CLUSTER SAMPLING

Cluster sampling - selection of a sample of clusters and survey all the units of each selected clusters.

This is also called ‘Single-stage cluster sampling’.

‘Multi-stage cluster sampling’ or simply ‘multi-stage sampling’: Instead of doing survey of all the units of selected clusters, only a sample of units are taken from each selected clusters.

Cluster Sampling

SELECTING A (SINGLE-STAGE) CLUSTER SAMPLE

- Required sampling frame: list of all the clusters.
- From the list, a sample of clusters is selected - this using a selection scheme (e.g., SRS, Systematic)
- All population units within the selected clusters are listed
- The information is then collected from all the units of the selected clusters

CLUSTER SAMPLING - ADVANTAGES

Main advantage

- Exact knowledge of the size of the sub-divisions (clusters) not required, unlike that for stratified sampling.
- Often a complete list of clusters - defined by location or as social entities or by institutions – is available, but frame of population units is not available or is costly to obtain.

  In such cases, cluster sampling can be adopted.
- Reduced cost if personal interviews, particularly when the survey cost increases with the distance separating the sampled units.

CLUSTER SAMPLING - DISADVANTAGES

Main disadvantage

Increased sampling error due to a less representative sample, since:

in practice, units are typically homogeneous within normally defined clusters

and the composition of clusters can not be altered, as they are pre-defined.

THANK YOU