

# Pacific Training on Sampling Methods for Producing Core Data Items for Agricultural and Rural Statistics

13-17 August, Suva, Fiji

## Module 2: Review of Basics of Sampling Methods Session 2.1: Terminology, Concepts and Definitions

By  
Chris Ryan  
Statistician (ESCAP Pacific Office)



## Topics Covered

- \* What is PPS Sampling?
- \* What is 2-stage Sampling
- \* Incorporating PPS sampling in to 2-stage sampling
- \* Estimation formula for 2-stage sampling with PPS

# Probability Proportional to Size Sampling (PPS)

- \* This sampling technique is a little more complex than others used in the Pacific
- \* The size of the unit is used to determine its chance of selection
- \* Examples of the size of a unit include:
  - \* The number of households in a village
  - \* The number of students in a school



## Example of PPS sampling

- \* Step 1: Determine number of islands to select (below = 3)
- \* Step 2: Compute the skip =  $2,100 / 3 = 700$
- \* Step 3: Select a random start between 0 and 700 =  $\text{rand}() * 700$  (below = 467)
- \* Step 4: Select the 3 islands by adding the skip to the random start

	# dwellings	Cum # dwell	Sel'n number
Island 1	458	458	
Island 2	347	805	467
Island 3	294	1099	
Island 4	255	1354	1167
Island 5	232	1586	
Island 6	167	1753	
Island 7	144	1897	1867
Island 8	102	1999	
Island 9	59	2058	
Island 10	42	2100	



## Example of PPS sampling 2

EA	# HHs	Cum # HHs	Selection	EA	# HHs	Cum # HHs	Selection
<b>100101</b>	<b>43</b>	<b>43</b>	<b>15</b>	<b>100301</b>	<b>48</b>	<b>743</b>	<b>699</b>
100102	81	124		100302	38	781	
100103	52	176		100303	71	852	
100104	61	237		100304	55	907	
<b>100105</b>	<b>44</b>	<b>281</b>	<b>243</b>	<b>100305</b>	<b>51</b>	<b>958</b>	<b>927</b>
100106	38	319		100306	41	999	
100201	72	391		100307	49	1048	
100202	49	440		100308	73	1121	
<b>100203</b>	<b>47</b>	<b>487</b>	<b>471</b>	<b>100309</b>	<b>48</b>	<b>1169</b>	<b>1155</b>
100204	33	520		100310	39	1208	
100205	61	581		100311	32	1240	
100206	63	644		100312	67	1307	
100207	51	695		100313	58	1365	

Number of HHs	1365
Number of EAs	26
Number of EAs to select	6
Skip	227.5
Random Start	14.04546
Seln 1	15
Seln 2	243
Seln 3	471
Seln 4	699
Seln 5	927
Seln 6	1155



## Probability Proportional to Size Sampling (PPS)

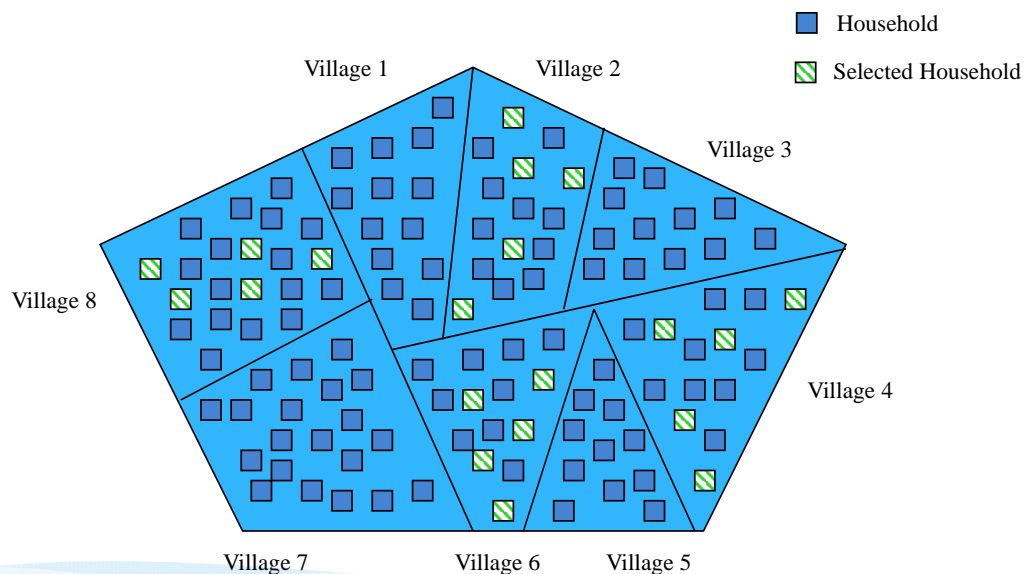
- \* The ordering of the units in the population before selection can be important
- \* If the ordering is done by geographical position then better geographical representativeness can be achieved
- \* If the ordering is done by the size of the unit, then better representativeness can be achieved from a size perspective



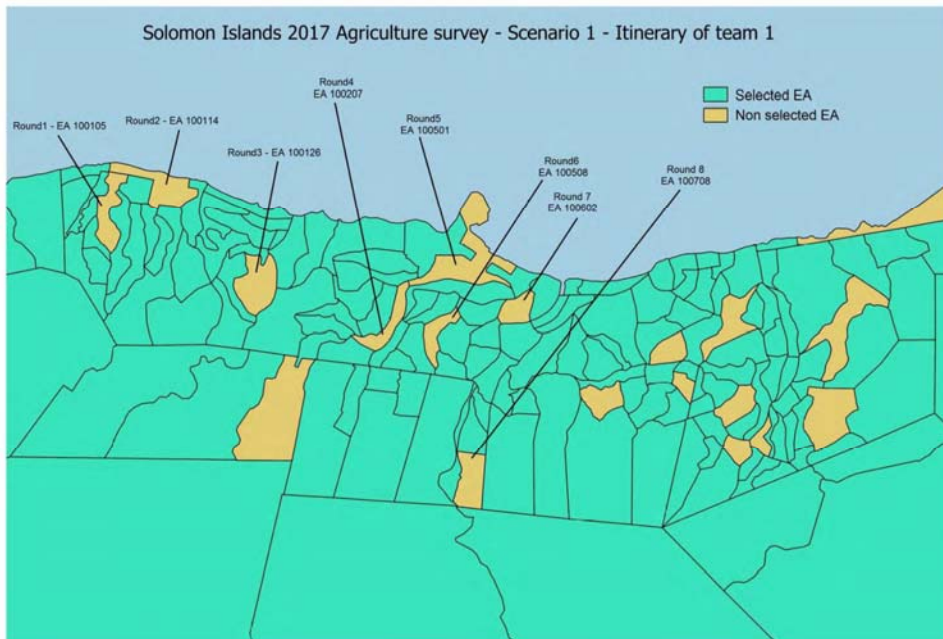
# Multi-stage sampling

- \* As the name suggests, this sampling approach involves more than one stage (most commonly two)
- \* At each stage, a different level of unit is selected, egs:
  - a) 1st stage : Villages                      2nd stage : Households
  - b) 1st stage : Schools                      2nd stage : Students
- \* At each stage, different sampling techniques may be adopted (eg, SRS, Systematic, PPS)

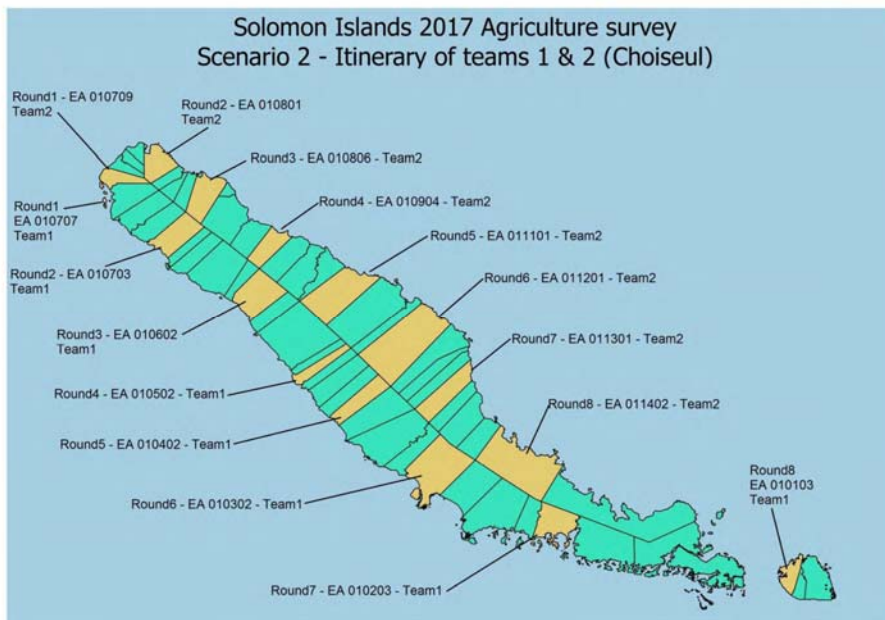
## Example of 2-stage sampling



# SI Agriculture Survey – EG of stage 1 selection



# SI Agriculture Survey – EG of stage 1 selection (cont)



# PPS sampling in 2-stage sampling

The most common form of household sampling adopted in the Pacific involves:

- \* Stratifying the population in to core reporting groups
  - \* Provinces, Divisions, etc
  
- \* Two stages of sampling
  - \* Stage 1: Selecting small geographical areas using PPS
  - \* Stage 2: Selecting a fixed number of households using systematic sampling



# PPS sampling in 2-stage sampling

Why is this approach so commonly adopted?

- a) It is cost effective, and avoids having to travel to households which could be located anywhere in the country
  - \* That is, straight Simple Random Sample is not being applied
  
- b) It still ensures coverage of all key domains of interest
  
- c) Within each stratum, each household has a known and similar chance of being selected in the survey
  - \* Important quality for deriving weights, and ensuring a representative sample



# PPS sampling in 2-stage sampling

Probability of selection for a dwelling - example

	# dwellings	Cum # dwell	Sel'n number
Island 1	458	458	
Island 2	347	805	467
Island 3	294	1099	
Island 4	255	1354	1167
Island 5	232	1586	
Island 6	167	1753	
Island 7	144	1897	1867
Island 8	102	1999	
Island 9	59	2058	
Island 10	42	2100	

## Stage 1 - Probability of Selection of Island

$\text{Pr}(\text{sel}) = \# \text{ dwell on island} / \text{skip}$

$\text{Pr}(\text{Island 2 selected}) = 347 / 700$

$\text{Pr}(\text{Island 4 selected}) = 255 / 700$

$\text{Pr}(\text{Island 7 selected}) = 144 / 700$

## Stage 2 - Probability of Selection of Dwell

$\text{Pr}(\text{sel}) = \# \text{ dwell sel} / \# \text{ dwell on island}$

$\text{Pr}(\text{dwell on Is 2 sel}) = 20 / 347$

$\text{Pr}(\text{dwell on Is 4 sel}) = 20 / 255$

$\text{Pr}(\text{dwell on Is 7 sel}) = 20 / 144$



# PPS sampling in 2-stage sampling

Example continued

- \* Prob of a dwelling being selected in the survey from the start  
= Prob (island selected) x Prob (dwelling selected)

$\text{Pr}(\text{sel}) \text{ for Island 2} = (347 / 700) \times (20 / 347) = (20 / 700) = 0.02857$

$\text{Pr}(\text{sel}) \text{ for Island 4} = (255 / 700) \times (20 / 255) = (20 / 700) = 0.02857$

$\text{Pr}(\text{sel}) \text{ for Island 7} = (144 / 700) \times (20 / 144) = (20 / 700) = 0.02857$

$\text{Weight} = 1 / \text{Pr}(\text{sel}) = 1 / 0.02857 = 30$





## Estimation for a two-stage process (without stratification)

*Estimate of a total*

$$\hat{Y} = \sum_{i=1}^n wy_i$$

*Estimate of a mean*

$$\hat{\bar{Y}} = \frac{\sum_{i=1}^n wy_i}{N}$$

## Estimation for a two-stage process (with stratification)

*Estimate of a total*

$$\hat{Y} = \sum_{i=1}^n wy_i$$

*Estimate of a mean*

$$\hat{\bar{Y}} = \frac{\sum_{i=1}^n wy_i}{N}$$