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

Module 2: Review of Basics of Sampling Methods: Probability Sampling, Sample Selection and Sample Design and Estimation

Session 2.5: Objectives of Multistage Sampling and Illustrative Designs

9 – 20 November 2015, Jakarta, Indonesia

UNITED NATIONS SIAP Statistical Institute for Asia and the Pacific

IMPROVING AGRICULTURAL STATISTICS Global Strategy

BPS, Statistics Indonesia

Sampling journey till now

Design(*representativeness*)

Element
- Costly
- No guarantee of representation

Strata
- Improve representativeness/precision
- But cost is high

Cluster
- Reduces the cost
- But also reduces precision

Selection method (*chance*)

SRS
- Good for theory development
- But Not practical

Sys
- Easy implementation
- But All units are not of same importance (size)

PPS
- Unequal probabilities (not epsem)

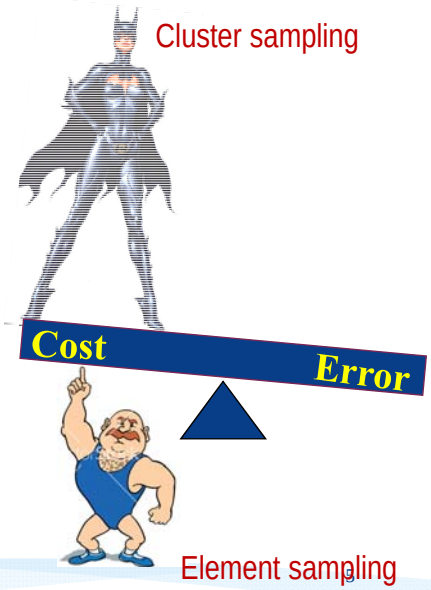
UNITED NATIONS SIAP

Stratification	Clustering
➢ Subdividing population into mutually exclusive groups	➢ Subdividing population into mutually exclusive groups
➢ Elements are homogenous in stratum	➢ Elements are heterogeneous in cluster
➢ We select an independent sample from each strum	➢ Clusters are sampling units
➢ Purpose is reducing sampling error	➢ Purpose is reducing cost

Main objective for stratification

Increase representativeness: *Reduce sampling error/ increase precision*

Cluster Sampling



Cost Solution

- Select a group of elements (Cluster)
- Then list elements only within the selected cluster

Problems in cluster sampling

In addition to cost:

- No control over sample size (clusters vary in size)
- Statistical inefficiency due to homogeneous clusters

Solution: sub-sampling

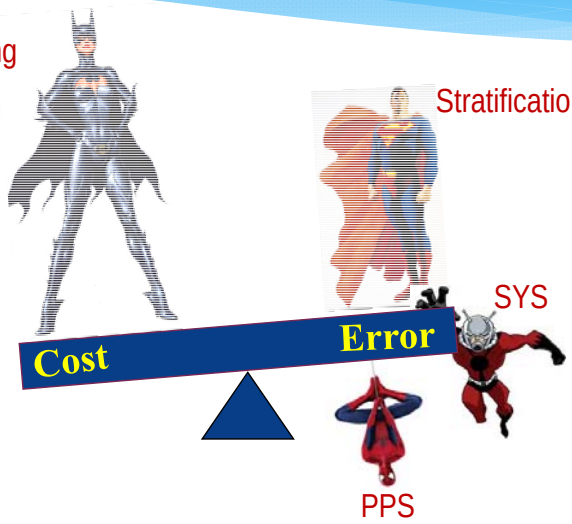
- Select more clusters and sub-sample of elements in each cluster
- Two possibilities:
  - Fixed sampling rate in all clusters (epsem)
  - Fixed sample size in all clusters (not epsem)

Two stage sampling

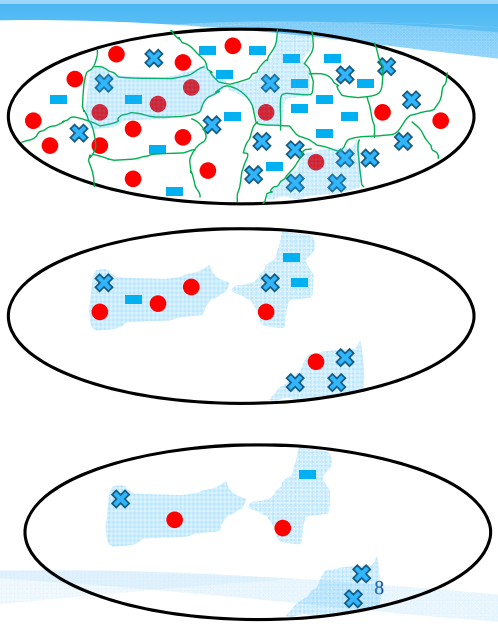
Clustering

Stratification

Complex designs:
Multi-stage stratified, with systematic PPS selection of PSUs



Two stage sampling (sub-sampling)



Population

clusters are primary sampling unites (PSUs)

1st stage sample

Elements are secondary sampling unites (SSUs)

2nd stage sample

In the context of agriculture surveys

Possible stratifications

- Space-time combinations *(province, urban/rural, zone, month, season)* in combination with a size variable;
  - Level of production *(normally a linear function of different products)*
  - Size of holdings *(normally sub-strata)*

Your country?

Possible clustering

- Space: *Farm/village/town/block/area…*
- Space-time: day-*site*
- Space-time-activity: *landing-time-site*

Use of systematic selection: *in PPS without replacement*

j	Size	Cumulative of size
1	Z1	Z1=T1
2	Z2	Z1+Z2=T2
3	Z3	Z1+Z2+Z3=T3
4	Z4	Z1+…+Z4=T4
5	Z5	Z1+…+Z5=T5
6	Z6	Z1+…+Z6=T6
7	Z7	Z1+…+Z7=T7
8	Z8	Z1+…+Z8=T8
9	Z9	Z1+…+Z9=T9
10	Z10	**Z1+…+Z10=T**

We need *n* sample

Assume *T= 100* and *n=4*

1. calculate sampling interval: *k=T/n=100/4=25*
2. Generate random number *r* between *1 and k (e.g 1< r =10<25)*
3. $X_i=r+(i-1)*I \quad (i=1…n)$
4. *j* is selected if $T_{j-1}<X_i<T_j$

Use of systematic selection: *in PPS without replacement*

Problem:

Any unit with size $Z>k$ may be selected more than once

Solution:

Put a restriction on measure of size: $Z_{max}=k=T/n$

In practice clusters are selected PPS, size being population of each cluster
We construct clusters to (1) meet above criterion; (2) manage enumerators' workload; (3) maintain heterogeneity in clusters