

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

Session 2.3: Objectives and use of stratification

9 – 20 November 2015,
Jakarta, Indonesia



Why not SRS?

* Sampling units are often different in regards with:

- * *Unemployment* → *gender, education, age,...*
- * *Health* → *age, region, income,...*
- * *Crop* → *climate, soil, ...*

* Sample size in subpopulations is often matter

* Why not using our knowledge about the parameter?

- * *Divide sampling units into homogeneous groups*

How to make our sample more “representative”? and our estimates “more precise”?

Make sure each “group of similar units” is represented in the sample

3



4

Stratification

- * Divide the population into distinct groups (strata) based on **auxiliary information** (*stratification variables*)
- * The division of the population into strata is termed “**stratification**”
- * Each “**stratum**” is composed of **homogeneous units** in regards with stratifying variable.

5

Stratified sampling involves

- 1) Find correlates for the study variable
- 2) **Stratification of the population into homogeneous (similar) groups**
- 3) Selection of sampling units using a selection procedure
 - * *like SRS, SYS, etc within each stratum, and*
 - * *independent of the other strata.*
- 4) **Estimate parameter for each stratum**
- 5) Combine the estimates and inference about the population

6

In stratified sampling

- * Sampling fractions; and

- * Selection Procedure

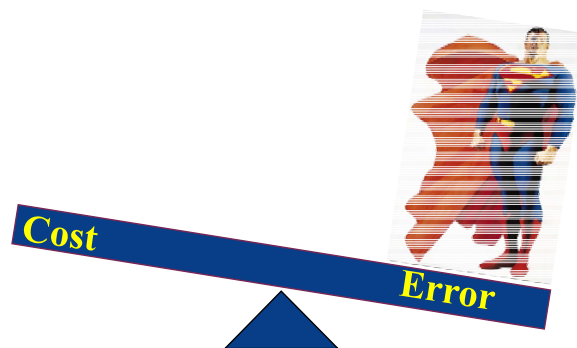
May be different across strata

- * The total sample size is distributed over all the strata (**allocation**)

- * The stratum results are **combined** to provide an estimate for entire population

7

Main goal for stratification



Reduce sampling error/ increase precision

8

Objectives of Stratification

- To obtain estimates of higher precision for given per unit of cost
- Providing separate estimates required for each stratum
- Using different sampling procedures for different strata, to (i) increase precision of the estimates (ii) organize the field work

9

Implicit stratification

- * Systematic sampling with the **units arranged** in a certain order
- * Prior to sample selection, all the units are sorted with respect to variable(s) that are **expected to be correlated** with the variable of interest. (*Normally geographical location of the units*)
- * It guarantees that units are **spread across the homogenous groups** (strata).
- * *Implicit stratification*: strata with size k (sampling interval)

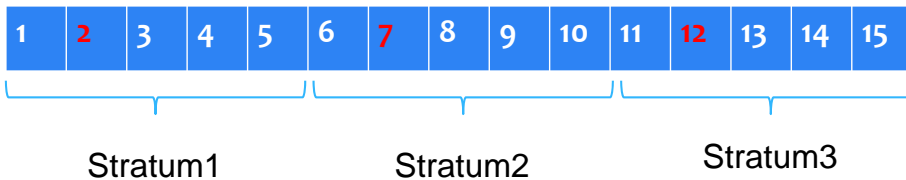
10

Implicit stratification

11

* $N=15$, $n=3$, $k=15/3=5$

$r=2$ (random number between 1 & 5)



Defining Strata

1. Choice of stratification variables :

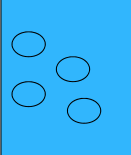
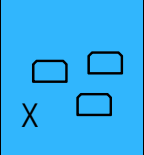
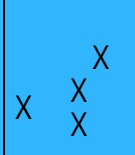
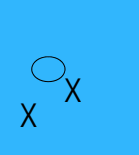
- * Homogeneous within strata; Heterogeneous across strata
- * Highly correlated with study variables (location with economic status, output with profit or number of employees etc)
- * More practical to choose naturally defined strata

2. Number of strata

- * Depends on availability of stratifying information in sampling frame: less information, fewer strata
- * At least two sampling units per stratum to be able to compute sampling error

12

Sample allocation to strata

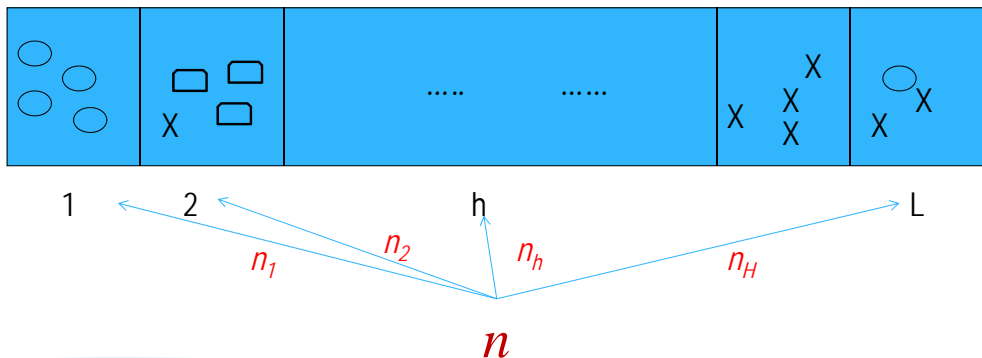
					
Stratum no.	1	2		h			L
Stratum size	N_1	N_2		N_h			N_H
Sample size	n_1	n_2		n_h			n_H

n

13

Sample Allocation to Strata

Maximize precision for fixed cost
OR
Minimize cost for required precision



14

Sample Allocation to Strata

Alternatives Methods:

- * Proportionate allocation
- * Disproportionate allocation
 - * Neyman allocation (minimum variance) (*discuss later*)
 - * Optimum allocation (cost and variance) (*not discussed here!*)

15

proportionate allocation

Share of each stratum from total sample is $\frac{N_h}{N}$, proportional to its size

The number of elements taken from the h^{th} stratum is

$$n_h = n \times \frac{N_h}{N}$$

Different view: sampling rate in each stratum is fixed and equal to $\frac{n}{N}$

$$n_h = N_h \times \frac{n}{N}$$

16

Probability of selection

Assuming SRSWOR in each stratum

Probability of selection may be different in each stratum

$$P_h = \frac{n_h}{N_h}$$