Regional Training Course on Agricultural Cost of Production Statistics
21– 25 November 2016, Daejeon , Republic of Korea

Objectives and Use of Stratification in Sample Design

Contents

- Strata and Clusters
- Single-Stage Cluster sampling
  - *epsem* selection and estimation
  - Selection method (single stage) – *epsem* and PPSWR and PPS systematic
  - Estimation under PPS cluster sampling

Strata and Clusters

Clustering and Stratification in Sample Design

* Typically, sample surveys conducted by NSOs involve subdividing the population into strata and clusters.
* The sampler's objective is to get the right combination of stratification and clustering to get the required estimates at the desired level of accuracy with the given resources.

Strata and Clusters

* Both stratification and clustering involve subdividing the population into mutually exclusive groups.
* Sub-divisions of the population are called ‘clusters’ or ‘strata’ depending upon the sampling procedure adopted.
* The term ‘cluster’ is used in the context of cluster sampling and multi-stage (cluster) sampling.

Sample Design – Selection Plan

Sample Selection Plan
= Sampling System AND Sampling Scheme

Sampling system
- Element sampling
- Cluster sampling
- Stratified sampling
- Multi-stage sampling*

Sampling scheme
- Simple random sampling
- Systematic sampling
- PPS

*Select cluster first, then select elements within selected clusters

Choice of Strata

Stratification

* A powerful tool for improving efficiency.
* In complex surveys, the clusters (PSUs) are usually stratified.
* Often the ultimate-stage units (households / holdings) are also stratified.
* Permits independent selection and estimation for each stratum - at all stages of selection.
* Appropriate allocation of samples improve efficiency of the estimates.

Objectives of Stratification

* To obtain estimates of higher efficiency for given per unit of cost
* Providing separate estimates required for each sub-division of the population – "domain" estimates
* Using different sampling procedures for different sub-population, to

  (i) increase efficiency of the estimates

  (ii) organize the field work

Defining Strata

1. Choice of stratification variables (location, output etc.):
   * Homogeneous within strata; Heterogeneous across strata
   * Highly correlated with study variables (output with cropped area or irrigation status etc.)
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

Choice of Strata

To increase precision relative to SRS

- Form strata with stratum units homogeneous with respect to survey variable (homogeneous within stratum)
- Stratum means of characteristic of interest varies widely among strata (heterogeneous across strata)

To provide separate estimates for smaller domains

- Place each domain in a stratum or set of strata
- Apply appropriate sampling rate or sample design to achieve needed sample size and precision.

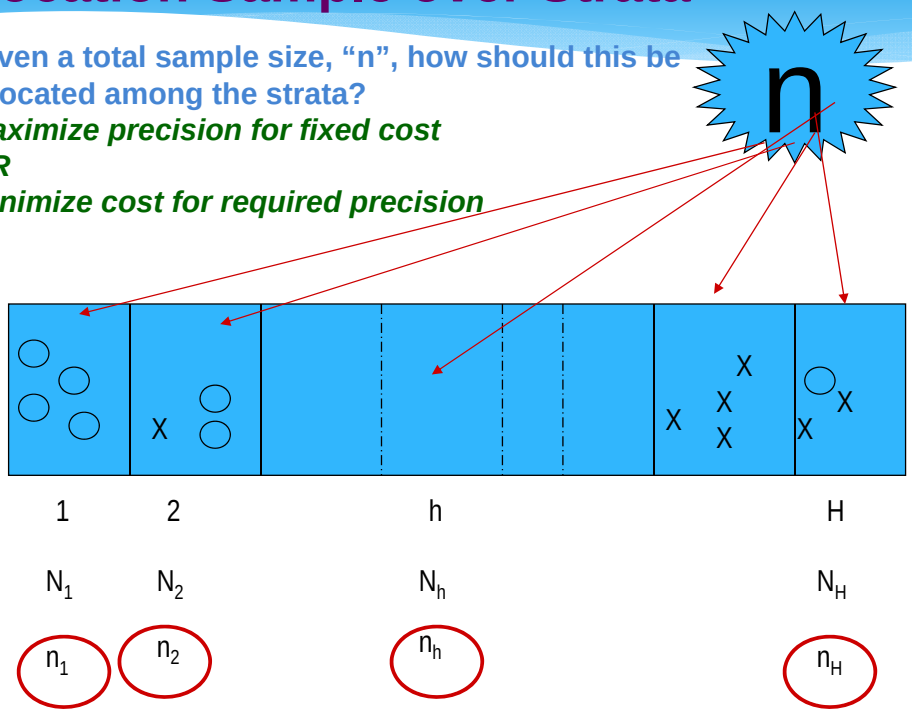
Stratification variables: Examples

* Household / Holdings surveys
  * At the First Stage
    * Urban/rural
    * Location: region; province
* At the Second Stage (for selection of households/ holdings)
    * Affluent / non-affluent
    * Households reporting cultivation, animal husbandry, poultry, horticulture etc.
    * Holdings of different size classes

Allocation of Sample Size to Strata

Allocation Sample over Strata

- Given a total sample size, “n”, how should this be allocated among the strata?
  Maximize precision for fixed cost
  OR
  Minimize cost for required precision



Sample Allocation to Strata

Alternatives Methods:

* Uniform or equal allocation
* Proportionate allocation
* Disproportionate allocation
  * Optimum allocation (minimum variance), fixed sample size
  * Cost optimum allocation (not discussed!)

Sample Allocation to Strata

* In *proportionate stratification*, an uniform sampling fraction is applied to each strata; that is, the sample size selected from each stratum is made proportionate to the population size of the stratum

* In *disproportionate stratification*, different sampling rates are used deliberately in different strata

Proportionate Allocation

In *proportionate stratification*,

$\frac{n_h}{N_h}$ is specified to be the same for each stratum.

This implies that the overall sampling fraction is $\frac{n}{N}$

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

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

$$n_h = (N_h)\frac{n}{N}$$

Proportionate Allocation

$$V_{SRS} \geq V_{prop}$$

Thus, for proportionate stratified *deff < 1*

For a given total variability in the population, the gain is greater if:

* the ***strata mean are more heterogeneous*** (more unequal strata mean)

OR

* ***the element values within the strata are more homogeneous***

Optimum Allocation

- Uses widely different sampling rates for the various strata.
- Objective: to achieve the *least variance for the overall mean* for the given sample size (Neyman's allocation); as well as given per unit of *cost in different strata.*
- Without cost consideration, the allocation is

$$n_h = n\frac{N_h\sigma_h}{\sum N_h\sigma_h}$$

- This gives better efficiency as compared to proportionate allocation:

$$V_{SRS} \geq V_{prop} \geq V_{opt}$$

Thanks