

AGRODEP
Household survey data course
Dakar, 8-10 October 2012

Stratification




HarvestChoice
BETTER CHOICES, BETTER LIVES




Stratified Sampling

- The population is divided up into subgroups or “**strata**”.
- A separate sample of households is then selected from each stratum.
- There are two primary reasons for using a stratified sampling design:
 - To potentially reduce sampling error by gaining greater control over the composition of the sample.
 - To ensure that particular groups within a population are adequately represented in the sample.
- The sampling fraction generally varies across strata.



These objectives are often contradictory in practice



Sampling weights need to be used to analyze the data

Estimation under stratified random sampling

- Each stratum is treated as an independent population
- Estimate of stratified total is sum of stratum totals
- Estimate of stratified mean is weighted combination of stratum means

Examples of Stratification

- **Establishment survey**
 - Stratification of establishments by economic activity and employment size
- **National household survey**
 - Geographic domains – regions, provinces
 - Urban/rural
 - Socio-economic groups
- **Agricultural survey**
 - Agro-ecological zones
 - Land use
 - Farm size

Design effects due to stratification

- Stratification generally increases the efficiency of sample design
 - Lowers the design effect
- No variance component between strata
- Example – stratified one-stage sample for establishment survey
 - Stratification by economic activity, number of employees
 - Design effects lower than 1

Estimation under stratified random sampling

Calculation of the Mean in a Stratified Sample

$$\bar{x} = \sum_{h=1}^L \frac{N_h}{N} \bar{x}_h$$

Calculation of the Variance in a Stratified Sample

$$Var(\bar{x}) = \sum_{h=1}^L (1 - n_h / N_h) \frac{N_h^2}{N^2} Var(\bar{x}_h) \approx \sum_{h=1}^L \frac{N_h^2}{N^2} Var(\bar{x}_h)$$

L = Number of strata
h = stratum number

N_h = Population size in stratum h
 n_h = sample size in stratum h

Sample allocation under stratified sampling

- Sample allocation procedure depends on objective of survey, and difference in variability and cost among the strata
- Three major types of sample allocation of sample units among the strata:
 - Proportional allocation
 - Equal allocation
 - “Optimum” allocation

Proportional allocation

- The sample allocated to each stratum is proportionally to the number of units in the frame for the stratum:

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

- Simplest form of sample allocation
- Provides self-weighting sample
- Efficient sample design for national-level results when variability is similar for the different strata

Equal allocation

- Each stratum is allocated an equal number of sample units:

$$n_h = \frac{n}{L}$$

- Used when same level of precision is required for each stratum
- Example: estimates of similar quality required for each region

Neyman allocation

- ❖ Provides minimum total error and minimum cost for a fixed sample size

$$n_h = n \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^L N_h S_h / \sqrt{c_h}}$$

- ❖ S_h = standard deviation in stratum h
- ❖ c_h = cost per unit in stratum h

Practical allocation criteria

- For national household surveys, sometimes allocation is a compromise between proportional, equal and Neyman allocation; e.g. we start with a proportional allocation and then we increase the sample size in the smaller regions
- In countries with high proportion of rural population, sometimes a higher sampling rate is used for the urban stratum, to increase the urban sample size and because of the lower cost of data collection in urban areas

Second Stage Stratification

- Sometimes it is desirable to stratify the sample in the last stage (household or individual level)
- Examples: male/female headed households, program beneficiaries, households with OVCs.
- Beware of the DANGERS! Second stage stratification increases the need for close supervision of field teams

Domains of analysis

- Subpopulation groups for which data will be tabulated and analyzed
- Sometimes domains will coincide with first levels of stratification
 - Example: regions, urban/rural
- Sample allocation by stratum should ensure sufficient sample size for each domain

Post-stratification domains

- Survey data tabulated by different subpopulation domains
- Example for households: type of housing unit, household expenditure level
- Example for persons: sex, age group, level of education
- For tabulations and analysis, necessary to ensure sufficient number of observations in each cross-tabulation cell

Weighting under stratified, multi-stage sample designs

- A proportionally allocated sample is self-weighted
- In non proportionally allocated samples, we must use weights to account for different sampling fractions by stratum