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 2.5: Objectives & Uses of Stratification

By
Chris Ryan
Statistician (ESCAP Pacific Office)

UNITED NATIONS SIAP Statistical Institute for Asia and the Pacific

FAO FIAT PANIS

UNITED NATIONS ESCAP Economic and Social Commission for Asia and the Pacific

Topics Covered

* What is stratification?
* How surveys can benefit from stratification
* Estimation formula for stratified samples

What is Stratification?

* Stratification is the process of taking a population and dividing it in to sub-populations. It is commonly used to take a heterogenous population and divided it in to M homogenous sub-populations
  * Heterogeneous: Units are all very different in nature
  * Homogenous: Units are similar to each other

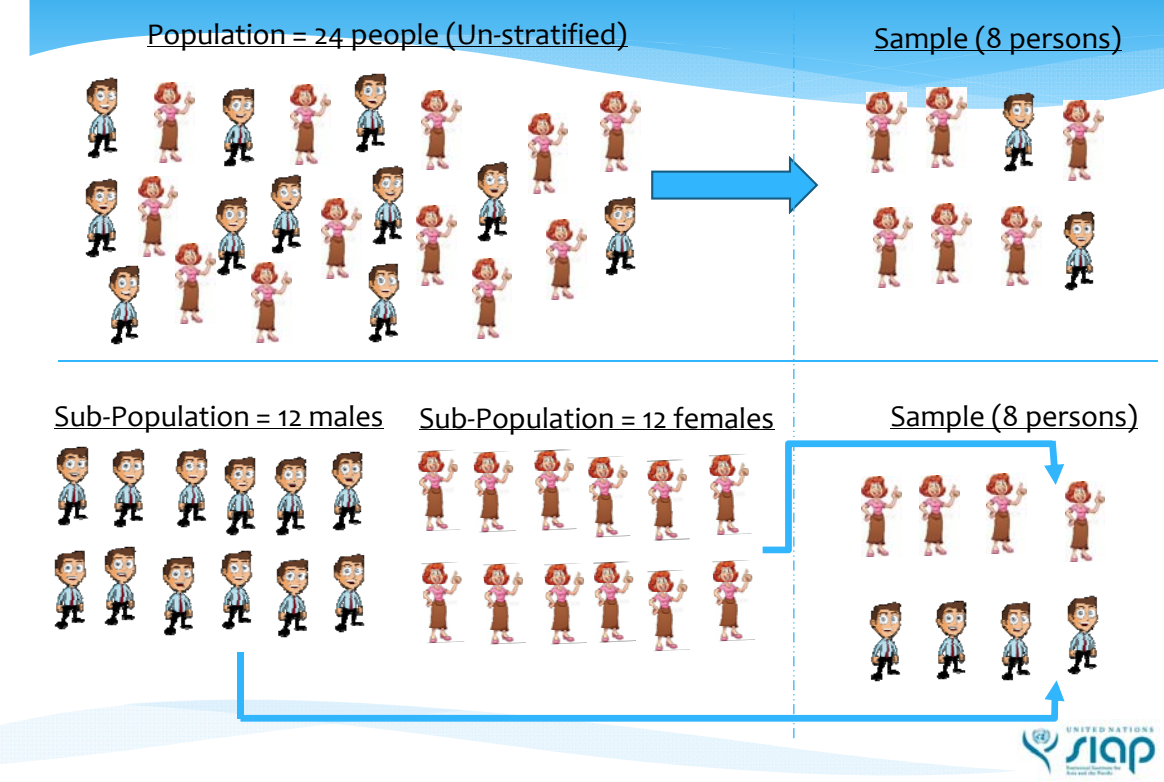
Why would we want to do this?

* Helps us control sample sizes for different sub-populations
* Helps us ensure a more representative sample is selected
* Allows us to use different sample strategies within each sub-population

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.

Example of stratifying a population



Handy things to know about stratification

- Strata can be created based on any auxiliary information available
  - Geography
  - Age or Sex
  - Farm size
- Sampling fractions can be different in different strata
- Sampling selection method can be different in different strata
- The total sample size is distributed over all strata
- Stratum results are combined to produce results for the entire population of interest

Advantages & Disadvantages – Stratified Sampling

* Advantages
  * Can control sample sizes for different sub-populations (eg, Regions)
  * Generally produces low SE's
  * Can adopt different sampling strategies for each stratum

* Disadvantages
  * Can be expensive in that you must sample from every stratum

Estimation in Stratified Sampling

* For stratified samples, the weights are determined separately within each stratum

Example (for 2 strata)

- Population (N) = 200, N_1 = 150, N_2 = 50
- Sample (n) = 15, n_1 = 10, n_2 = 5
  - (assume simple random sampling)
- Weights are derived as follows:

  $w_1 = N_1/n_1 = 150/10 = 15$

  $w_2 = N_2/n_2 = 50/5 = 10$

Estimation in Stratified Sampling (cont)

Estimate of a total

$$\hat{Y} = \sum_{h=1}^{H} \hat{Y}_h$$

where

$$\hat{Y}_h = N_h \times \sum_{i=1}^{n_h} \frac{y_{h,i}}{n_h} = \sum_{i=1}^{n_h} w_h y_{h,i} \quad \text{where} \quad w_h = \frac{N_h}{n_h}$$

Estimation in Stratified Sampling (cont)

Estimate of a mean

$$\hat{\bar{Y}} = \frac{\sum_{h=1}^{H} N_h \bar{y}_h}{N}$$