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 2.2: Sampling & Estimation Basics

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

- An introduction to sampling
  - Sample v Census
  - Different types of sample approaches
- An introduction to estimation
  - What is an estimate
  - The Concept of a weight
  - What an estimation formula looks like

Census v Sample

* There are two types of survey available to the survey designer
  * Census: Select all units the population of interest
  * Sample Survey: Select a sample of units in our population of interest

Advantages of a Sample Survey

* Reduces cost
* Reduces time
* Data can be collected more accurately
  * less non-response
  * less collection/processing error
* More detailed questions can be asked

Disadvantages of a Sample Survey

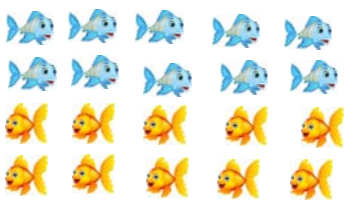
* Results are subject to sampling error
* Detailed cross-tabulations and results for small geographical areas and sub-populations may be too inaccurate
* Difficulty communicating results to users

The key to a good sample

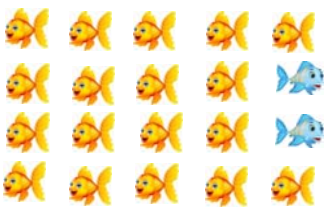
Samples are considered good if the sample represents the population.

Population = 40 Gold Fish & 40 Blue Fish

Good Sample: 10 Gold Fish & 10 Blue Fish



Bad Sample: 18 Gold Fish & 2 Blue Fish



Sampling methods

* Probability Sampling – each unit in the population has a known, non-zero probability of selection
* Quota Sampling – Establish a quota of units from different sub-populations that will be in the sample
  * Eg, select 50 males 15-24
    select 50 males 25+
    select 50 females 15-24
    select 50 females 25+
  * People are selected from a crowd, often on the street until the quota is met

Sampling methods (cont)

* Haphazard Sampling – use whatever sample is available
  * Eg, a sample of meteors would be restricted to meteors that have fallen to earth
* Purposive Sampling – Select units that are believed to be representative
  * Eg, outer islands D and F are considered the best representatives of all outer islands
* Volunteers
  * Eg: TV or radio phone-in polls

Sampling methods (cont)

* For nearly all credible surveys, Probability Sampling is used
  * This is for the desirable feature of knowing what the probability of selection was for each unit which can be used in applying accurate estimation techniques
  * It also gives all units in our population of interest a chance of being selected in our survey
* Each of the other methods discussed can often be heavily biased to such an extent that the results are meaningless, and are often used just to create publicity

What is estimation ?

* When we select a sample of units from our population, we don't know the true value for our variables of interest
  * We therefore have to estimate what these true values might be from the sample

  Estimation is the process of compiling an approximate measure for a characteristic of a population, based on a sample, with a known level of accuracy

Concept of a weight

* If only a sample of units are being used to estimate the characteristics for a population, then those units must represent more than only themselves
* The factor that determines how many units in the population a sampled unit represents, is known as the weight
* This weight can differ for different units in the sample

"Probability of selection" v "Weight"

* The weight is equivalent to the inverse of the probability of selection
* For example
  * You have a population with 10 persons (N)
  * You select a sample of 5 persons (n)
  * Pr (selection) is 5/10 = 0.5
  * The weight is 1/Pr(selection) = 1/0.5 = 2
* We are therefore saying each person in the sample represents 2 people

Dealing with formulas - an example

$$\hat{\bar{Y}} = \sum_{i=1}^{n} \frac{y_i}{n}$$

$\hat{\bar{Y}}$

- Indicates it's an estimate
- Indicates it's a mean
- Indicates the variable

y_i

- Indicates the variable
- Indicates the observation

$\sum_{i=1}^{n}$ → Indicates you are summing from the 1st observation until the nth observation (n = sample size)

Different types of estimates

Sample surveys can produce a range of different types of estimates with the most common being:

An estimate of mean (or average)

EG: Average quantity of fish caught in a week per household

An estimate of total

EG: Total quantity of Taro produced in a year

An estimate of proportion

EG: Proportion of household engaged in raising cattle