Food and Agriculture Organization
of the United Nations

MASTER SAMPLING FRAME(MSF)

FOR

AGRICULTURAL STATISTICS

Module 3 – Session 3:

Sampling design and estimation when the MSF is an area frame

2017

Objectives of the presentation

- At the end of this session, the audience will:
  - Be able to know the kind of observation mode that should adopted for their survey
  - Be able to establish the formula of the the estimation of the variable of interest depending whether the sampling unit is a point or a segment
  - Understand how to avoid non-sampling errors when designing the survey

Outline

1) General concepts

2) Observation and reporting mode

3) Estimations when sampling unit is a segment

4) Estimations when sampling units is a point

5) Non-sampling errors and linking area frame with census or administrative data

6) Allocation methods

Introduction

- Very quick change of the population of production units in the agricultural sector
  - Difficult to keep list sampling frame up-to-date
- **Use of the territory to define an area sampling frame as an alternative**
  - more stable than the population of farms or of agricultural households.
  - boundaries usually constant
  - significant landscape changes generally easier to address than changes in the population of farms, as the stratification can simply be updated.

1 General concepts

Outline 1

1.1. General concepts

1.2. Main type of sampling units

1.3. Stratification in area frame

1.4. Multi-stage sampling

1.5. Multi-phase sampling

1.6. Systematic sampling

1.1. General concepts

- For an effective sampling, the sampling units should be associated to relevant characteristics (type of crops, type of land cover, etc.) related to the population of interest

- Typical definitions of strata:
  - « agricultural land > 60 percent », « agricultural land between 30 and 60 percent » etc.
  - using specific crops or crop groups: « irrigated crops > 50 percent », « permanent crops dominant » or « grassland mixed with cropland ».

1.2. Main types of sampling units

- Pieces of Land (segments);
- Points (in this case, the area frame may becalled point frame);
- Transects;

1.3. Stratification in area frame

- Stratification of an area frame generally based on the geo-referenced features observed on the land, possibly by means of image analysis or photo-interpretation
- Additional strata may be defined for specific crops or crop groups that are usually stable on the land
- The unit should be characterized by a set of variables when:
  - a division of the territory into geographic units is used as the basis for a Master Sampling Frame

1.3. Stratification in area frame

- Information on the variables likely to be far from perfect in practice:
  - Does not mean that the stratification will be inefficient

- The strata boundaries should follow the geometry of the segments, in particular for segments with a regular (square) shape

1.4. Multi stage sampling

- **First stage:**
  - A sample of area units (PSUs) is selected (example segments).

- **Second stage:**
  - A sample of points is drawn in each PSU selected in the first stage.

- In this case, the PSUs will be segments, but small **administrative units (EAs)** can be also used.

1.5. Multi-phase sampling

- Large sample is selected in the first phase and information is collected on the units. In a second phase, a subset of these units is selected based on the information collected.

- First-phase sample or **pre-sample:**
  - Generally stratified by means of a procedure that is more statistically efficient than a stratification system applied to the full area frame.
  - If the first-phase sample is sufficiently large, this large sample can be a good master sample, or a basis for one.

1.5. Multi-phase sampling

- Particularly useful for **non-clustered** point sampling
- Typical examples are the Italian AGRIT survey (since 2002) and the Eurostat LUCAS survey (since 2006)
- Subsampling of the pre-sample in the second phase with a rate that depends on the strata
  - Most of the final sample is concentrated on the strata with the highest priority

1.6. Systematic sampling

- Systematic sampling is often applied in list frames
- Elements of the frame sorted in accordance with a given criterion
- The sample contains the elements i, $i+k$, $i+2k$, etc., where i is random and the step k is adjusted to obtain the targeted sample size (i between 0 and k);
- The efficiency of systematic sampling depends on the degree of auto-correlation between neighboring elements

1.6. Systematic sampling

- Ordering PSUs in a serpentine arrangement in area frames of segments with physical boundaries

- **Main disadvantage**
  - No unbiased estimator of the variance;
  - However, the ensuing practical implications are limited.

- **Advantage:**
  - More easily traceable

2

Observation and reporting mode

Outline 2

2.1. Direct observations

2.2. Sampling farms using area frame

2.1. Direct observations

2.1.1. Observation/reporting mode: Direct observations - Introduction

- **If you were instructed to provide the following information, which approaches would you use to collect data in the context of an area frame (5 min exercise)**
  - Distribution of the number of species (cattles, goats, sheeps) passing through a point (example well)... between 10 am and 11 am
  - Are farmers using Good Agricultural Environmental Practices (GAEP) in their cocoa farms?
  - Accurate land area under a specific crop

2.1.2. Observation/reporting mode: Direct observations

- Direct observations protect against subjectivity
  - **But limit the types of data that can be collected**

- Relatively easily performed for crop area estimation if:
  - Appropriate graphic material and GPS devices are provided to the enumerators

- In direct observation mode, the survey usually involves delineating fields using an **ortho-photo or satellite image** as a background

2.1.2. Observation/reporting mode: Direct observations

- In the case of point sampling:
  - Measurement of the field or plot area is not necessary for crop area estimation.
  - Only the area of the stratum and the proportion of points that fall in a certain crop are needed.

- Compare the area measured with **the holder-estimate** data if **the direct observations are combined with an interview with the holder**

2.1.2. Observation/reporting mode: Direct observations

- Direct observation is more problematic for yield estimation, especially when the yield is heterogeneous within a given field.

- **Authorization by the farmer** may be necessary to collect a crop sample, which requires an investment of time to locate the farmer.

- In some countries, direct observations **without previous contact with the farmer** may be perceived negatively by the population.

2.1.2. Observation/reporting mode: Direct observations

- **Potential biases:**
  - Wrong location of the enumerators on the ground
  - Wrong identification of crops
  - Unsuitable observation date:
  - A major bias can also be introduced if only one visit per year is made in areas where the proportion of double or multiple cropping is significant.

2.1.2. Observation/reporting mode: Direct observations

- **Potential biases:**
  - In some countries, difficulties in reaching the fields or even a point from which the field can be seen (e.g. in the case of large properties with restricted access)
  - Width of linear elements
  - Improper modifications of the sampling scheme

2.2. Sampling farms using an area frame

2.2.1. Sampling farms using an area frame

- Resources for estimations and design of an area frame:
  - Hendricks et al. (1965)
  - FAO, 1996 and 1998
  - Gallego et al. (1994 and 2013)

- **Difficulty:**
  - The identification of farmers linked with the points or segments sampled may be costly

2.2.2. Sampling farms using an area frame - segments

The open segment:

- A farm is selected if its headquarters is within a sampled segment
- The terms « headquarters » must be defined carefully. For agricultural households, the dwelling is a possible criterion
- The selection probability of the farm is the selection probability of the associated segment
- The number of farms per segment can be very hetegeneous

2.2.2. Sampling farms using an area frame - segments

Example of agricultural landscape with some farm headquaters (open segment)

2.2.2. Sampling farms using an area frame - segments

The open segment:

- When area frame are based on geometric grid, the open segment approach requires adoption of a particularly precise rule on **the reference point in locating the dwelling** (e.g. main entrance door)

- Otherwise, the headquarters may be split by the boundaries of two segments (example Farm 1 in the previous figure)

2.2.3. Sampling farms using an area frame - segments

The closed segment:

- The reporting unit is the tract, the part of the farm's fields inside the segment
- This approach is problematic because the farmer does not always have a **precise information on the target variables** referring to the field (s) inside the segment
- If the segment is defined by means of a geometric grid, the farmer's answers become more difficult
- Close segment is not recommended for surveys that involve an interview with the farmer

2.2.3. Sampling farms using an area frame - segments

Tracts inside a square segment (close segment)

2.2.4. Sampling farms using an area frame

The weighted segment

- When there is insufficient information on how an additive variable $\boldsymbol{Y}$ is distributed in a farm, the difficulty of reporting $\boldsymbol{Y}$ for a tract can be bypassed using the weighted segment approach.

- The area under a given crop, the production, the amount of fertilizer used, etc. are examples of additive variables. Yield is not an additive variable.

- The area of the tract corresponding to Farm k in the sampled segment is called T_{jk}. The total area of land operated by Farm k is A_k and y_k is the total for the farm of each variable being measured – area of maize, number of animals, income, etc.

2.2.4. Sampling farms using an area frame - segments

The weighted segment

- The weighted segment attributes to the tract a part of the additive variable that is proportional to the area

$$\boldsymbol{x_{ijk}} = \frac{\boldsymbol{T_{jk}}}{\boldsymbol{A_k}} \boldsymbol{y_k}$$

- This approach creates a fictitious variable $\boldsymbol{X}$ that is uniformly distributed in the farm area A_k and that has, by definition, the same total value as $\boldsymbol{Y}$ for each farm.

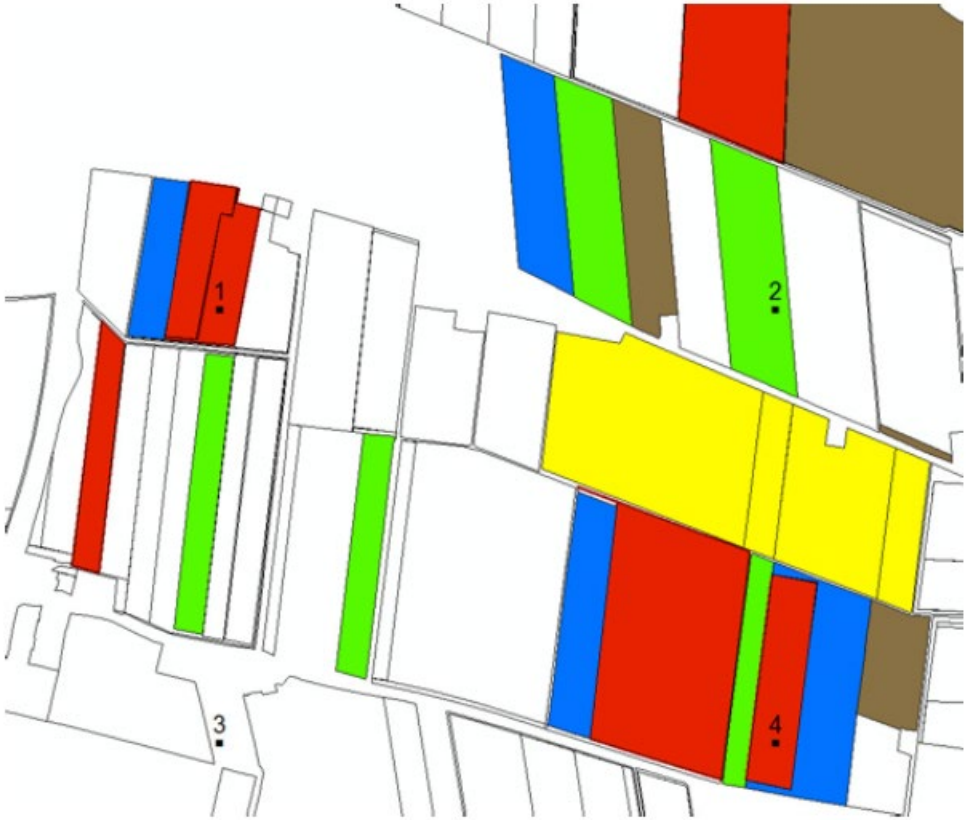
- The total values of $\boldsymbol{X}$ and $\boldsymbol{Y}$ also coincide for an administrative region if it is accepted that each farm has fields in only one administrative region. This assumption is usually not exactly true.

2.2.4. Sampling farms using an area frame - segments

Sampling farms (tracts) inside a square segment

2.2.4. Sampling farms using an area frame – point sampling

Sampling farms (tracts) by points



Plots highlighted with the same colour correspond to the same farm.

3 Estimation when the sampling unit is a segment

36

Outline 3

3.1. The open segment method

3.2. The weighted segment method

3.3. The closed segment method

3.4. Variance estimations

3. When the sampling unit is a segment: introduction

- For a given survey variable, such methods refer to the way the data of the reporting units (holdings or tracts) is aggregated to define the segment totals

- Three type of methods have been used: the closed segment, the open segment and the weighted segment methods

3. When the sampling unit is a segment: introduction

Sample unit	Reporting unit	Method of association
Segment	Holding	Weighted segment estimators
Segment	Holding	Open segment estimators
Segment	Tract	Close segment estimators

3.1. When the sampling unit is a segment: Close segment method

- The value of a variable in a segment is simply the **sum of its values in each of the tracts** of the segment
- Data on the totality of a holding is not needed except for holdings totally included in the segment
- It requires the collection of data (by direct observation or by personal interview with the holder) about agricultural activities that are found physically within the boundaries of the segment

3.1. When the sampling unit is a segment: Close segment method

- The close segment is the normally accepted method to estimate crop areas

- For livestock, the closed segment is only appropriate for those livestock controlled in pastures and/or corrals within the boundaries of the segment

- For estimating l**ivestock that can roam outside the segment**, the closed segment method should not be used

3.1. When the sampling unit is a segment: Close segment method

- The estimation formulas correspond to a single-stage design for simplification purposes.

- The total number of segments is exactly known for each strata and PSU

- The probability of selection of each segment (constant within a stratum) is equal to its conditional probability in two stage sampling

3.1. When the sampling unit is a segment: Close segment method

- A direct expansion area frame estimate calculated by expanding tract level data for each holding with at least part of its land located in the sample area segment

$$Y_c = \sum_{h \in S} \sum_{j \in B_h} \sum_{k \in G_{hj}} e_{hjk} \sum_{m \in T_{hjk}} t_{hjkm}$$

Y_c	is the sample estimate of a total for the survey variable y
S	is the set of all land-use strata
B_h	is the set of all sub-strata in stratum h
G_{hj}	is the set of all segments in sub-stratum j of land-use stratum h
T_{hjk}	is the set of all tracts in segment k of substratum j or land-use stratum h
e_{hjk}	is the expansion factor for all tracts in segment k
t_{hjkm}	is the tract value for the variable y associated with tract m

3.1. When the sampling unit is a segment: Close segment method

Strengths

- Very precise estimates can be made, especially for the variables correlated with farmland and easily associated with the tract, such as crop areas
- Biases due to non-response can be decreased because tract level data can usually be observed fairly accurately
- Data collection efforts required are the simplest and easiest to follow
- Reported data for the tract are less susceptible to reporting errors than holding level reported data and are less sensitive for the respondent to report
- Duplication of data between segment is less likely to occur as when entire holding data is reported

3.1. When the sampling unit is a segment: Close segment method

Weaknesses

- Can not be used to estimate certain survey variables that are not associated easily with the tract, such as economic items
- Not efficient in estimating rare items or items concentrated in small geographic areas unless with an ad-hoc stratification
- Data collection costs will be higher than for the open segment estimator, resulting in two to three times as many interviews
- Response errors can occur through misidentification of segments and tracts due to obsolete maps and photos and response unfamiliarity with the tract reporting concept

3.1. When the sampling unit is a segment: Close segment method

Applications

- Areas of major crops that are commonly distributed
- Can be used for some common livestock item such as cattle

3.2. When the sampling unit is a segment: Weighted segment method

- The variable in each tract is defined as the value of the variable in the holding multiplied by a factor equal to the ratio between the area of the tract divided by the area of the holding
- Data is collected from each holding associated with any land area in the sample segment
- Data collected from each such holding is weighted by the proportion
  - Derived by dividing the area of the holding that is within the segment by the entire area of the holding (both inside and outside the segment)

3.2. When the sampling unit is a segment: Weighted segment method

- The weighted segment estimator can be used for all survey variables, since the holding is the reporting unit required
- It does not require that the holder, the location of his residence or the holding headquarters are precisely established
- However, it is obviously necessary to define the holding and find an informant that can provide data for the totality of the holding

3.2. When the sampling unit is a segment: Weighted segment method

- It simplifies the problem of urban segments if the cultural system in the country is such that each holder can be related to a define piece of agricultural land
- The urban segments can be observed and if there is no agricultural land or activity:
  - No enumeration is necessary since it is assumed that any holder that might live there will have the opportunity to be selected where his land is
- In countries where are many holders who **graze livestock on public or communal land** without any lease
  - Will be necessary to canvass city and urban segments if the weighted segment method is being used to record livestock belonging to those who profess to have no land

3.2. When the sampling unit is a segment: Weighted segment method

- A direct expansion area frame estimate calculated by expanding total holding data for each holding with land found in the sampled area segment after prorating the data by the proportion of the total holding land found in the segment

$$Y_w = \sum_{h \in S} \sum_{j \in B_h} \sum_{k \in G_{hj}} e_{hjk} \sum_{m \in T_{hjk}} w_{hjkm} t_{hjkm}$$

Y_w	is the sample estimate of a total for the survey variable y
S	is the set of all land-use strata
B_h	is the set of all zones in stratum h
G_{hj}	is the set of all segments in zone j of land-use stratum h
T_{hjk}	is the set of all tracts in segment k of zone j or land-use stratum h
e_{hjk}	is the expansion factor for all tracts in segment k
w_{hjkm}	is the weight used to prorate y_{hjkm}
t_{hjkm}	is the value for the variable y associated with tract m

3.2. When the sampling unit is a segment: Weighted segment method

Strengths

- Can be used to estimate all agricultural variables (characteristics) pertaining to the holding

- Usually the most precise of the three kinds of area estimates, especially for livestock and economic items

3.2. When the sampling unit is a segment: Weighted segment method

Weaknesses

- Data collection costs are the highest of the three kind of direct area sample estimators. Entire holding data must be collected for all tracts
- There is a tendency for holders to underreport total holding areas by omitting non-cropland areas and biases the weighted estimates upward
- Entire holding data for non-respondents can not be usually be observed within the segment.
- Coverage of holdings in densely populated areas can be poor due to the difficulty of screening for holders, multiple holders in a single household, and small holdings
- Entire holding data are susceptible to reporting errors if the respondent is not knowledgeable about the entire holding or misunderstands survey concepts relating to the entire holding

3.2. When the sampling unit is a segment: Weighted segment method

Applications

- All agricultural variables, including crop areas, stocks, and production, livestock inventories and economic items

3.3. When the sampling unit is a segment: Open segment method

- The open segment method also called **holding headquarters** method associates a segment to all holdings with headquarters included in the segment

- Clear rules have to be established to define a unique reference point for each holding, called **the headquarters**

- Several ways to do this, but the most common is to define the headquarters as the dwelling (residence) of the holder

3.3. When the sampling unit is a segment: Open segment method

- **The holder must be uniquely identified:**
  - Involves practical difficulties in cases where more than one person or household are involved in the operation of the holding
- **The urban areas have to be included in the area frame:**
  - An area sample can capture information from holders that cannot be related to a piece of rural land

3.3. When the sampling unit is a segment: Open segment method

- A direct expansion area frame estimate calculated by expanding total holding data for each holding in which the holder or Headquarters resides in the sampled area segment

$$Y_o = \sum_{h \in S} \sum_{j \in B_h} \sum_{k \in G_{hj}} e_{hjk} \sum_{m \in T_{hjk}} a_{hjkm} t_{hjkm}$$

Y_o	is the sample estimate of a total for the survey variable y
S	is the set of all land-use strata
B_h	is the set of all zones in stratum h
G_{hj}	is the set of all segments in zone j of land-use stratum h
T_{hjk}	is the set of all tracts in segment k of zone j or land-use stratum h
e_{hjk}	is the expansion factor for all tracts in segment k
a_{hjkm}	is 1 if the holding Headquarters for tract m in segment k is within the segment and zero otherwise
t_{hjkm}	is the value for the variable y associated with tract m

3.3. When the sampling unit is a segment: Open segment method

Strengths

- Can be used to estimate all agricultural variables (characteristics) pertaining to the holding

- Data collection costs can be lowered as only holders resident in their holding need to be contacted

3.3. When the sampling unit is a segment: Open segment method

Weaknesses

- Usually the least precise of the three types of area estimates
- Entire holding data for non-respondents can not usually be observed within the segment.
- Entire holding data are susceptible to reporting errors if the respondent is not knowledgeable about the entire holding or misunderstands survey concepts relating to the entire holding
- Usually results in an underestimation of the farm population due to the difficulty of screening for holders in densely populated areas
- The holder is more likely to refuse information about his entire holding than for a part of his holding

3.3. When the sampling unit is a segment: Open segment method

Applications

- All agricultural variables, particularly those easily associated with the entire holding such as livestock inventories, grain stocks, numbers of holdings, gross value of sales, and other economic characteristics.

3.4. When the sampling unit is a segment: Variance estimator

- The variance formula for the area sample estimators are nearly identical. Let the expanded tract level value for the variable Y be:
  - $Y_{hjkm}{}^{e} = e_{hjk} t_{hjkm}$ for the closed estimates
  - $Y_{hjkm}{}^{e} = e_{hjk} a_{hjkm} t_{hjkm}$ for the open estimates
  - $Y_{hjkm}{}^{e} = = e_{hjk} w t_{hjkm}$ for the weighted estimates
  - n_{hj}= number of segments in G_{hj}

Then the expanded value for the item Y for a segment is:

$$\boldsymbol{Y_{hjk}{}^{e} = \sum_{m \in T_{hjk}} Y_{hjkm}{}^{e}}$$

3.4. When the sampling unit is a segment: Variance estimator

- The variance is:

$$V=\sum_{h\in S}\sum_{j\in B_h}\left[\left(\frac{n_{hj}}{n_{hj}-1}\right)\left(\sum_{k\in G_{hj}}(Y_{hjk}{}^{e})^2-\frac{(\sum_{k\in G_{hj}}Y_{hjk}{}^{e})^2}{n_{hj}}\right)\right]$$

4 Estimations when the sample unit is a point

62

Outline 4

4.1. Notations

4.2. Estimates based on Farm and Non-farm points

4.3. Estimates based only on Farm points

4.4. Farms with fields in different strata

4.1. When the sampling unit is a point: Notations

- We assume that the population Ω of segments is divided into strata $\Omega_h, \mathrm{h} = 1 \ldots . \mathrm{H}$

- The population size is N segments (N_h for stratum Ω_h)

- The sample size is n segments (n_h)

4.1. When the sampling unit is a point: Notations

- The sample size of points in each segment will be K_i (fixed); in general K_i=K
- F_i correspond to the farms on which these points fall
- Each segment «i» has a total of UAA (Utilized Agricultural Area) surface U_i

4.1. When the sampling unit is a point: Notations

- We have a two stage-sampling:
  - In the first stage, the segment i is selected with probability $\mathrm{p}_i = 1/\mathrm{N}_h$
  - In the second stage, the unit is not the farm but the tract (UAA in a segment, that belongs to the same farm)
  - The tract k of segment i has an area T_{ik} and the total UAA of the farm is A_{ik} over all segments. U_i is the sum of the tract T_{ik} in the segment i.

4.2. When the sampling unit is a point: Estimates based on farm and Non-Farm points

- There will be K-F_i observations with 0 corresponding to points outside the UAA

- Sampling through points makes that tracts are selected with replacement and with the probability p_{ik} proportional to the area T_{ik}/D_i (the knowledge of T_{ik} is not necessary) where D_i is the size of the segment determined by the frame design

- The joint selection probability that farms k and k' are in the sampling $p_{ikk'} \neq p_{ik} * p_{ik'}$ (sampling is made with replacement) and this case, it gives easier formulae for variance estimation

4.2. When the sampling unit is a point: Estimates based on farm and Non-Farm points

- W_{ik} will be an additive quantity for a farm, most often production or the area of a particular crop

- Since we have no information about how W_{ik} is distributed inside the farm, we create a fictitious variable X that is uniformly distributed, and that has, by definition, the same total as W for Each farm

$$X_{ik} = \frac{T_{ik}}{A_{ik}} * W_{ik}$$

4.2. When the sampling unit is a point: Estimates based on farm and Non-Farm points

- The two-stage version of Horvitz-Thompson estimation for the total of X in the stratum Q_h gives:

$$\hat{X}_h = \frac{1}{n_h}\sum_{i=1}^{n_h}\frac{\hat{X}_i}{p_i} = \frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{1}{K_i}\sum_{k=1}^{K_i}\frac{X_{ik}}{p_{ik}} = \frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{D_i}{K_i}\sum_{k=1}^{K_i}\frac{W_{ik}}{A_{ik}}$$

- This means that, even if the second stage sampling unit is a tract, we do not to know its area nor X_{ik}, but just the global information about the farm.

4.2. When the sampling unit is a point: Estimates based on farm and Non-Farm points

- The variance in stratum Ω_h can be estimated as (Cochran, 1977, section 11.6):

$$\hat{V}(\hat{X}_h) = \frac{\mathrm{N}^2{}_h}{\mathrm{n}_h}\left(1 - \frac{n_h}{\mathrm{N}_h}\right)\sum_{i=1}^{\mathrm{n}_h}\frac{(\hat{X}_h - \bar{X}_h)^2}{n_h - 1} + \frac{N_h}{\mathrm{n}_h}\sum_{i=1}^{\mathrm{n}_h}\frac{1}{K_i(K_i - 1)}\sum_{k=1}^{\mathrm{K}_i}\left(\frac{W_{ik}D_i}{A_{ik}} - \hat{X}_i\right)^2$$

- The estimates for the total are:

$$\hat{X} = \sum_{h=1}^{H}\hat{X}_h \qquad \hat{V}(\hat{X}) = \sum_{h=1}^{H}\hat{V}(\hat{X}_h)$$

4.2. When the sampling unit is a point: Estimates based on farm and Non-Farm points

- Crop areas are currently estimated from the segment survey with more objective ground data
- Farm surveys provide both area and production estimates but with more significant bias due to nonresponse and to sujective tendency of the farmer
- For cattle, the estimates are possible but the results will be bad if there are a substantial number of farms without any UAA

4.3. When the sampling unit is a point: Estimates based only on farm points

- In this case, we substitute K_i by F_i and D_i by U_i

$$\hat{X}_h = \frac{1}{\mathrm{n}_h}\sum_{i=1}^{\mathrm{n}_h}\frac{\hat{X}_i}{\mathrm{p}_i} = \frac{\mathrm{N}_h}{\mathrm{n}_h}\sum_{i=1}^{\mathrm{n}_h}\frac{1}{F_i}\sum_{k=1}^{\mathrm{K}_i} U_i\frac{W_{ik}}{\mathrm{A}_{ik}}$$

$$\hat{V}(\hat{X}_h) = \frac{\mathrm{N}^2{}_h}{\mathrm{n}_h}\left(1-\frac{n_h}{\mathrm{N}_h}\right)\sum_{i=1}^{\mathrm{n}_h}\frac{(\hat{X}_h-\bar{X}_h)^2}{n_h-1} + \frac{N_h}{\mathrm{n}_h}\sum_{i=1}^{\mathrm{n}_h}\frac{1}{F_i(F_i-1)}\sum_{k=1}^{\mathrm{K}_i}\left(\frac{W_{ik}U_i}{A_{ik}} - \hat{X}_i\right)^2$$

4.3. When the sampling unit is a point: Estimates based only on farm points

- Advantages and drawbacks of both approaches are not clear
- Using only farm points can increase the cost of the survey if the number of points per segment is to be keep constant
- But the non UAA points removed correspond to null values of W_{ik} and their removal can result in a reduction of the variance

4.4. When the sampling unit is a point: Farms with fields in different strata

- Perturbation due to farms with fields in different regions is expected small because of low proportion (generally under 1-2%)

- Compensation between the bias the bias due fields inside the region belonging to farms with the headquarters outside the region and vice versa

- W is calculated on the farms that have their headquarters inside the surveyed region.

NEXT……

How to link area frames with census or administrative information?

5

Non-sampling errors and linking area with census or administrative data

Outline 5

5.1. Non sampling errors in an area frame

5.2. Linking area frame with census or administratrive data

5.1. Non-sampling errors in an area sampling frame

5.1.1. Non-sampling errors in an area sampling frame

- As stated previously, an area sampling frame usually has a negligible undercoverage of farms that cultivate land or manage pastures (owned or rented).

- The main source of bias is linked to farms with **livestock that use common pastures – in particular nomadic livestock** – although this is not the only source of bias.

- However, an insufficiently careful management of the frame can introduce significant bias.

- The most important source of bias is probably the exclusion, from the sampling process, of strata labelled as purely non-agricultural.

5.1.2. Non-sampling errors in an area sampling frame

Closed segment

- Two types of non-sampling errors occurs for closed segment estimates:
  - The first type occurs because there are small areas of waste within the fields.
    - Because the scale of a photograph, accounting for areas of waste less than 1 acre is particularly difficult.
  - The second type of non-sampling error occurs in areas where cattle and sheep roam freely through open gates and cross tract or segment boundaries.

A major advantage of closed segment estimates is their ability to decrease any bias caused by non-response.

5.1.3. Non-sampling errors in an area sampling frame

Open segment

- Four types of non-sampling errors occurs for open segment estimates:
  - The first is caused by the incorrect application of the "headquarters" rule.
    - the identification of headquarters can become difficult for partnerships, corporations, and managed farms where enumerators need to ask a series of questions to eliminate potential duplicate reporting

5.1.3. Non-sampling errors in an area sampling frame

Open segment

- The second type is the underestimating of the farm population
  - because the headquarters of farm operators may be inadvertently missed in more densely population
- The third type is the underreporting of farm values
  - Farmers tend to forget about parcels of rented land not contiguous with the main part of their farm and;
  - about parcels of woodland or wasteland under their control, but considered "non-agricultural"
- The fourth type occurs when farmer report livestock data
  - A farmer tends not to report livestock on the farm that are owned by someone else.
  - A farmer also tends to report livestock owned by the farmer, but located on someone else's land

5.1.3. Non-sampling errors in an area sampling frame

- An example of risk management practice that can introduce significant bias is the « extended segment »
- A regular grid (in this case, square segment) is used to define sampling units
- But when a segment is sampled, the observation unit is defined considering the full parcels for all plots that intersect the square segment
- The overestimation is obvious when an expansion factor of the type (N/n) is used. The bias is less obvious but still significant if correction factors are not applied to compensate for the increase in segment size

5.1.3. Non-sampling errors in an area sampling frame

Example of an “extended segment”

5.2. Linking area frames with census or administrative data

5.2.1. Linking area frames with census or administrative information: the use of EAs

- Enumeration Areas (EAs) are generally associated with a delimited territory and can thus be considered to define an area sampling frame.
- However, if the specific characteristics of an area frame are not used, the set of EAs can be considered as a list frame.
- A sample of *n* EAs is selected in the first stage, usually with some type of PPS where size does not necessarily refer to the EA's geographical area.

5.2.1. Linking area frames with census or administrative information: the use of EAs

- In this context, size may refer to the number of farms or to the agricultural area. If it is desired to sample EAs with a probability proportional to the geographic area or to the agricultural land, an area frame approach can be used.
- For example, a systematic sample of points (a square grid with a 10 km or a 20 km step may be an option) can be chosen; this will ensure a relatively homogeneous geographic distribution.

5.2.2. Linking area frames with census or administrative information: the use of EAs

Several sampling rules are available:

- All the EAs on which a grid point falls are selected.
- A more reasonable criterion may be sampling with a probability proportional to the agricultural land.
- It may be preferred to assign a higher probability to EAs with a greater share of high-value agricultural products, e.g. irrigated land or permanent crops.
- An EA can be considered as a large segment of an area frame. However, due to the EA's large size, the use of the traditional open, closed, and weighted segment methods without subsampling of farms is likely to be very inefficient. Subsampling farms thus becomes necessary.

5.2.2. Linking area frames with census or administrative information: the use of EAs

- One of the advantages of using EAs as PSUs both as list frames and as area frames is the frequent existence of relevant data for all EAs in the country, such as expert reports on crop area.

- These figures may be subjective, to a certain extent; however, they constitute a valuable source of covariates that can be used with more objective data estimated for a sample of EAs.

- The US was one of the first countries to successfully use an area frame in agricultural surveys. Its experience is briefly described in Annex C (in the handbook), along with those of Brazil, China, Ethiopia, Guatemala, and the EU's Joint Research Centre MARS and LUCAS projects.

6 Allocation methods

90

6. Allocation method

- In the case that prior information are available before the sampling (former surveys, variance of the variable of interest etc.), Neyman allocation could be adopted

- Otherwise, the allocation method could be discussed be the experts of the domain

- Another approach is to do the allocation proportionnaly to the size of the strata (the example of brazil)

Conclusion

- Estimations methods should be applied related to the sampling design (segments or points)

Sample unit	Reporting unit	Method of association
Segment	Holding	Weighted segment estimators
Segment	Holding	Open segment estimators
Segment	Tract	Close segment estimators

- Allocations methods could be discussed with the experts if prior information is not available

- The observation mode should be related to the indicators that need to be computed

References

- Carrie, D. 2009. *Area frame design for agricultural survey*. United States Department of Agriculture. Washington, USA.

- Cristiano, F. 2014. *Improving methods for linking area frames with list frames in agricultural surveys*. Global Strategy to improve agricultural and rural statistics. FAO, Rome, Italy.

- Delince, J. 2016. *Preliminary results on the Master Sampling Frame Field-testing*. Outreach Workshop on Agricultural Surveys. Rome, Italy

- FAO. 1996. *Multiple frame agricultural surveys: current surveys based on area and list sampling method*s. Rome, Italy.

References

- Gallego, F.J. 1995. *Sampling Frames of Square Segments*. Report EUR 16317. European Commission Joint Research Centre Publication, Ispra, Italy.

- Gallego F.J., Delincé J., & Carfagna E. 1994. *Two-Stage Area Frame Sampling on Square Segments for Farm Surveys*. Survey Methodology, 20(2): 107-115.

- Gallego, J.F., Herve, K., & Olivier, L. 2015. *Using GIS and Remote Sensing to built Master Sampling Frames for Agricultural Statistics*. JRC Technical Reports. European Commission.

- Global Strategy to improve agricultural and rural statistics., 2017. *Master sampling frames for agriculture: supplement on selected country experiences.* Rome, Italy

References

- Global Strategy to improve agricultural and rural statistics., 2015. *Hanbook of Master Sampling Frame for Agricultural Statistics: Frame development, Sample design and Estimation*., Rome, Italy

- Global Strategy to improve Agricultural and Rural Statistics. 2016. *Improving the Methodology for collecting and using administrative data in an agricultural statistics system*. Rome, Italy.

- Global Strategy to improve Agricultural and Rural Statistics. 2014. *Technical report on Identifying the most appropriate sampling frame for specific landscape types*. Rome, Italy.

- National Agricultural Statistics Service. 1987. *Area frame design for agricultural surveys*. U.S. Departement of Agriculture

Thank You