Food and Agriculture Organization
of the United Nations

GLOBAL STRATEGY TO IMPROVE AGRICULTURAL AND RURAL STATISTICS

TRAINING IN AGRICULTURAL STATISTICS

Module 3: Data processing, analysis and dissemination

2017

Plan of module 3

3.1 – General overview of current processing practices and limitations observed

3.2 – Areas and yields of pure and mixed crops

3.3 – Production

3.4 – Crop forecasting

3.5 – Analysis and dissemination

3.1

General overview of current processing practices and limitations observed

3.1 General overview of current processing practices and limitations observed

- Processing, analysis and dissemination of data from various sources (sample surveys and agricultural censuses):
  - This role is generally assigned to the National Statistical Office and the Ministry for Agriculture (in particular for agricultural data) (demand for skills)
- Training
  - Is provided at training centres or university statistics departments (supply of skills)
- Problems
  - Mismatch between the supply and demand for skills
    - shortage of qualified personnel
    - inadequacy of statistical methods
  - Methodological problems:
    - inconsistency in the production of core indicators to measure the development of agriculture
    - lack of modern equipment and poor data quality

3.1.1-Lack of qualified personnel

- Many developing countries suffer in particular from a lack of qualified staff in the processing, analysis and dissemination of agricultural data

- As confirmed by the following:
  - Vacant positions at various levels
  - High turnover of qualified staff
  - Lack of professional training

3.1.2 Inadequacy of statistical methods

- Current problems:
  - Survey design
  - Definition of methodologies
  - Dissemination of results

- For example,
  - there are relatively few data collection standards
  - the level of cover and comprehensiveness varies between surveys
  - fewer disaggregated data available on small administrative units and target groups
  - inability to provide reliable statistics on the increasingly large range of agricultural activities that generate income for rural households
  - the methods used for imputation of missing values and processing of outliers are not very transparent

3.1.3-Inconsistencies in the production of core indicators

- Difficulty in regularly producing core indicators for agriculture: the regular production of statistics is compromised by
  - The relatively small number of analysts capable of processing and analysing all the available data
  - Backlogs in planning collection operations
  - Data inconsistencies

- This remains a real constraint due to the cyclical nature of the indicators, owing to:
  - Climatic variations (weather, uncertainties, etc.)
  - Policy initiatives

- Example of cyclical indicators: area, quantities produced, yield, prices, etc.
- Accounting for uncertainties in the agricultural sector requires annual or seasonal data collection

3.1.4-Lack of modern equipment (1/2)

- Data collection: traditional approach (PAPI) vs modern approach (CAPI).
  - Traditional approach:
    - Census or sample survey data are captured manually in data processing software
    - The accuracy of the data is checked by double entry
  - CAPI approach:
    - New digitization technologies now help to save time and can greatly improve data accuracy.
    - These technologies also make it possible to access the results more quickly.

3.1.4-Lack of modern equipment (2/2)

- The Geographic information system (GIS), a computerized system that facilitates data processing, analysis and dissemination, has improved the scope and quality of georeferenced agricultural data
- Note that some countries still do not have modern equipment (e.g. GIS) for collecting data in agricultural censuses and surveys and for producing agriculture statistics

3.1.5 Poor data quality (1/2)

- Poor data quality hinders:
  - The implementation of agricultural and rural policies
  - The measurement of progress in terms of development
- Better quality statistics help to
  - assess the state of the nation
  - judge the government's performance
- ***Outlook***
  - Although improvements have been noted in the quality of agricultural data over the last three decades, efforts still need to be made in aspects relating to:
    - relevance
    - accuracy
    - timeliness
    - accessibility

3.1.5 Poor data quality (2/2)

- These gaps are the result of failures related to the data processing, analysis and dissemination methods
- Developing Strategic plans for agricultural and rural statistics (SPARS) and capacity building programmes help to solve these problems
- Particular importance should be paid in these SPARS to:
  - modernizing the production process
  - capacity building for statistics staff
- Priorities in capacity building are:
  - the collection design and the collection methodology
  - data processing and analysis

3.2 Areas and yields of pure and mixed crops

3.2.1-Data necessary for estimating yield or area:

- Subsections covered:

  a) Estimating areas from land register information

  b) Sample design with an area sampling frame for measuring areas

  c) Sample design to estimate areas with mixed crops for a household survey

  d) Useful information for measuring yield

  e) Calculating the level of parcel use

3.2.1-a-Estimating areas from land register information:

Availability of a parcel register with crops present by season (1/2)

- Can be done by means of stratified sampling taking the regions or districts as the stratum
- Sampling a fixed number of villages is done per stratum
- However, the type of villages sampled in terms of land development and type of parcel use (homogeneous, dispersed and by patch) is known only *a posteriori* (post-stratification)

3.2.1-a-Estimating areas from land register information:

Availability of a parcel register with crops present by season (2/2)

- To do this, areas are estimated in terms of parcel use using the domain estimation approach (Särndal et al., 1992)

- The parcels (statistical units) in each sampled village are enumerated comprehensively and the areas collected according to the information recorded in the administrative register (as in India)

3.2.1-a-Estimating areas from land register information:

Register of available parcels, but absence of information relating to land use (1/2)

- In this configuration, the available land register does not provide information on parcel use

- A two-stage survey is therefore recommended:
  1) Villages or EAs are selected at the first stage
  2) Parcels are selected at the second stage

3.2.1-a-Estimating areas from land register information:

Register of available parcels, but absence of information relating to land use (2/2)

- Information collected at the parcel stage:
  - Areas
  - Additional data to confirm declarations:
    - quantity of inputs used per crop type
    - farm costs, etc.
- This declared information can be supplemented by real measurements made on a subsample of parcels
- Parcel use for each crop can be determined by a quick eye estimation

3.2.1-b-Sample design with an area sampling frame for measuring areas (1/2)

- The sampling frame is an area frame (of EAs, districts, villages, etc.)

- Use of a two-stage sample design
  1) <u>First stage:</u> elements in the sampling frame (EAs, districts, villages, etc.)
  2) <u>Second stage:</u> segments within the PSUs

3.2.1-b-Sample design with an area sampling frame for measuring areas (2/2)

- Information relating to the parcels identified in the sampled segments is then collected on a declarative basis from the holders concerned

- Area collection operations are then carried out on a subsample of parcels with the appropriate equipment (GPS, compass, tape measure, etc.).

3.2.1-c-Sample design to estimate areas with mixed crops for a household survey (1/3)

- Data collection through household surveys is necessary when parcels and land use type cannot be identified from any source of information
- In this case, the appropriate sample design is the two-stage design (with EAs/villages as primary sampling unit (PSU) and households as secondary sampling unit (SSU))
- Information on parcel use and to assess areas (quantity of seed used, etc.) per crop is collected by declaration

3.2.1-c-Sample design to estimate areas with mixed crops for a household survey (2/3)

- Area collection will be carried out on a subsample of households, taken from a subsample of PSUs
- The information will be collected for all eligible parcels
- Area collection is carried out by direct measurement using appropriate equipment (GPS, compass, tape measure, etc.)

3.2.1-c-Sample design to estimate areas with mixed crops for a household survey (3/3)

- The areas associated with each crop will be allocated in proportion to parcel use
- Parcel use is determined for each crop in the following cases:
  - Quick eye estimation
  - Estimation from additional information
  - Objective estimation by laying a density grid and counting the seedlings (mixed, with homogeneous density)
  - Estimation based on counting seedlings (dispersed or by patch)
  - The enumerators must collect all the information required for this evaluation (the data collection tools have been developed for this purpose)

3.2.1-d-Useful information for measuring yield

- The approach described below is valid regardless of the specific cases mentioned above
- The statistical unit is the parcel
- A list of parcels is selected from parcels identified from registers, or from the various sample segments or from households
- Two stages should be considered:
    1) Collecting information relating to production and area by declaration based on the holder's estimation/perception
    2) Laying yield grids on a sample of parcels initially used for area measurement

3.2.1-e-Calculating the level of parcel use

- Several different cases should be considered to determine the level of parcel use
  - Non-homogeneous arrangement of mixed crops
  - Homogeneous cover
  - Laying a density grid
  - Spacing between seedlings
  - Direct observation

3.2.1-e-Calculating the level of parcel use.
Non-homogeneous planting of crops:
i-Seeding rate (1/2)

- Use of seeding rate to assess the area of the various crops
  - Planted seeds might not all have germinated
  - However, this information is a proxy for having the information

3.2.1-e-Calculating the level of parcel use. Non-homogeneous planting of crops:

i-Seeding rate (2/2)

- Soient les cultures A et B
- Soient a et b la quantité de semence mise en terre respectivement pour les cultures A et B. Considérons, sans introduire de confusion, A et B les quantités recommandées de semences respectivement pour les cultures A et B
- Soit S la superficie de la parcelle sous association de culture et S_i la superficie associée à la culture i. Ainsi,

$$S_A = \frac{a/A}{(a/A + b/B)} * S$$

$$S_B = \frac{b/B}{(a/A + b/B)} * S$$

Pour illustration, considérons les paramètres ci-dessous :
S= 0.4 ha, a=50 kg, A = 120 kg/ha, b=1kg et B = 5 kg/ha.
Le taux de semis est : (50/120) pour A et (1/5) pour B.

$$S_A = (0.42/(0.42 + 0.2)) \text{ x } 0.4 \ = 0.27\text{ha}$$
$$S_B = (0.2/(0.42 + 0.2)) \text{ x } 0.4 \ = \ 0.13 \text{ ha}$$

3.2.1-e-Calculating the level of parcel use.
Non-homogeneous planting of crops:
ii-Crop density (1/2)

- Seedling density is used to assess areas
- Seedling density is determined objectively by laying a density grid at selected points at random
- The density is the number of seedlings per unit of area for each crop in the mixture
- The area of each crop in the mixture can be estimated by calculating the density ratio of the seedlings

3.2.1-e-Calculating the level of parcel use. Non-homogeneous planting of crops:

ii-Crop density

- Consider 3 mixed crops on an area estimated to be 0.8 ha
- Suppose that a density grid contains the following for each of the crops in the mixture:
  - 10 seedlings of A instead of 60 in a pure crop
  - 18 seedlings of B instead of 25 in a pure crop
  - 20 seedlings of C instead of 50 in a pure crop
- The density ratio is therefore

$$RS_A = \left(\frac{10}{60}\right) = 0{,}167$$

$$RS_B = \left(\frac{18}{25}\right) = 0{,}72$$

$$RS_C = \left(\frac{20}{50}\right) = 0{,}4$$

$$S_A = \frac{RS_A}{RS_A + RS_B + RS_C} * \mathrm{S} = \frac{0{,}167}{0{,}167 + 0{,}72 + 0{,}4} * 0{,}8\ \mathrm{ha} = 0{,}13 * 0{,}8\mathrm{ha} = 0{,}10\ \mathrm{ha}$$

$$S_B = \frac{RS_B}{RS_A + RS_B + RS_C} * \mathrm{S} = \frac{0{,}72}{0{,}167 + 0{,}72 + 0{,}4} * 0{,}8\ \mathrm{ha} = 0{,}56 * 0{,}8\mathrm{ha} = 0{,}45\ \mathrm{ha}$$

$$S_C = \frac{RS_C}{RS_A + RS_B + RS_C} * \mathrm{S} = \frac{0{,}4}{0{,}167 + 0{,}72 + 0{,}4} * 0{,}8\ \mathrm{ha} = 0{,}31 * 0{,}8\mathrm{ha} = 0{,}25\ \mathrm{ha}$$

3.2.1-e-Calculating the level of parcel use
iii-Interplanted crops: seedling line numbers

- Mixed crops, sown in lines (interplanted crop), but in separate rows:
  - The area under each constituent crop can be divided based on the number of lines of each constituent crop
  - The number of lines on a row of interplanted crops is counted at three places at random in the selected field to determine the average number of lines

3.2.1-e-Calculating the level of parcel use iii-Interplanted crops: seedling line numbers

- Consider two interplanted crops A and B, sown over 0.5 ha
- Suppose that the number of lines of crop A where the crops are mixed is 25, whereas the recommended density in a pure crop is 40
- For crop B, the number of mixed lines is 2 whereas the density is 5 in a pure crop
- The area ratio for A and B is therefore:

$$RS_A = \left(\frac{25}{40}\right) = 0{,}625$$

$$RS_B = \left(\frac{2}{5}\right) = 0{,}4$$

The areas thus allocated to crops A and B are:

$$S_A = \frac{RS_A}{RS_A + RS_B} * \mathrm{S} = \frac{0{,}625}{0{,}625 + 0{,}4} * 0{,}5\mathrm{ha} = 0{,}61 * 0{,}7\mathrm{ha} = 0{,}43\ \mathrm{ha}$$

$$S_B = \frac{RS_B}{RS_A + RS_B} * \mathrm{S} = \frac{0{,}4}{0{,}625 + 0{,}4} * 0{,}5\mathrm{ha} = 0{,}39 * 0{,}7\mathrm{ha} = 0{,}27\ \mathrm{ha}$$

3.2.1-e-Calculating the level of parcel use

iv-Interplanted crops: distance between seedlings

- Mixed crops, sown in lines (interplanted crop), but in separate rows
- The area under each constituent crop can be divided based on the distance between seedlings
- Consider two interplanted crops A and B, sown over 1 ha. Suppose that the average distances observed (information obtained at, at least, three points chosen at random on the parcel) are 2 m between seedlings of A and 3 m between seedlings of B
- The areas allocated to each crop are as follows:

$$S_A = \frac{E_A}{E_A + E_B} * \mathrm{S} = \frac{2}{2+3} * 1\ \mathrm{ha} = 0{,}4 * 1\mathrm{ha} = 0{,}4\ \mathrm{ha}$$

$$S_B = \frac{E_B}{E_A + E_B} * \mathrm{S} = \frac{3}{2+3} * 1\ \mathrm{ha} = 0{,}6 * 1\mathrm{ha} = 0{,}6\ \mathrm{ha}$$

3.2.1-e-Calculating the level of parcel use

iv-Interplanted crops: distance between seedlings (1/2)

- Note: Which approaches should be used for the cases described below?
  - In the case of a non-homogeneous, fairly complex mixture, the recommended approach is to determine the level of crop cover by subjective estimates (observation, declaration by the interviewee)
  - If crops cover a relatively marginal area (less than 10%), they should be ignored
  - A simple estimate consists in dividing the total parcel area by the number of mixed crops. This is a simple method, but it will give overestimates of plant production

3.2.1-e-Calculating the level of parcel use

iv-Interplanted crops: distance between seedlings (2/2)

- Note: Which approaches should be used for the cases described below?
  - A rough estimate involves allocating the total seeded area to each crop in the mixture: the method is very rough and can therefore lead to overestimated areas
  - If temporary crops (seasonal and annual crops) are sown at the same time in a crop mixture and harvested at different times, the parcel is treated as double harvested. The total area is recorded under each crop in the mixture for seasons when they are harvested

3.2.1-e-Calculating the level of parcel use

iv-Allocation of area when temporary crops are sown with permanent crops

- Two-stage allocation:
  - Estimate the area occupied by permanent crop seedlings
  - This area is then inferred from the total area of the parcel. The resulting area is allocated to temporary crops (generally sown as interplanted crops)
- The area under permanent crops is the area under the canopy. ”The canopy is the upper layer of the forest, directly influenced by solar radiation”**
- The area can be estimated by simply multiplying the average area of the canopy (per tree) by the number of permanent crop seedlings
- The average area of the canopy is calculated based on the average area of cover (πr^2), assessed under three or five trees selected at random, where r is the radius of the canopy

https://fr.wikipedia.org/wiki/Canopée

3.2.2-Calculating areas from field data (extrapolation)

- The total area of a holding is the sum of the areas of the parcels on the holding
- The area is estimated at stratum level from the total where the variable *y* represents the area (see next slide)

3.2.2-Calculating areas from field data (extrapolation): reminder of annotations

- S: total area of a given crop (maize, rice, etc.) in a given stratum
- i: PSU index
- j: SSU index
- M: number of PSU in the stratum considered
- m: number of PSU sampled in the stratum considered
- N_i: Total number of SSU in PSU i
- n_i: Number of SSU sampled in PSU i
- N: total number of SSU sampled in the stratum considered
- y_{ij}: area of the crop considered in SSU j of PSU i of the stratum considered
- π_i: probability of inclusion of PSU i
- π_{ij}: probability of inclusion of SSU j of PSU i

3.2.2-Calculating areas from field data (extrapolation): reminder of annotations

- The total area S is given by:

$$S = \frac{N}{m} \sum_{i=1}^{m} \frac{1}{n_j} \sum_{j=1}^{n_j} y_{ij}$$

With probabilities in the form

$$\pi_i = \frac{m\ N_i}{N} \qquad \pi_{ij} = \frac{n_j}{N_i}$$

The difficulty in applying this formula lies in determining y_{ij} (the areas per agricultural holding for a given crop) from the collected data

For a given holding, if the crop concerned is mixed on some plots with other crops, the area allocated to the crop cannot in fact be clearly determined. There are several methods of doing this and the free application of any of these methods makes any area comparison inappropriate

Calculating area: specific case

Box 18: Current methods of dividing the area of a parcel containing a main crop and a secondary crop (1/2)

- The area of the parcel is counted twice: it is allocated at the same time to the main crop and to the secondary crop, and these are then called developed areas: the sum of cropped areas per crop is greater than the total physical area, but the sum of areas per crop is equal to the total cropped area
- The area of the parcel is counted only for the main crop, and is then referred to as *physical areas:* the areas per crop tend to be underestimated
- The area of the parcel is divided between the two crops in a proportion routinely fixed at *1 for main crops and 0.5 for secondary crops:* the total cropped area is overestimated

Calculating area: specific case

Box 18: Current methods of dividing the area of a parcel containing a main crop and a secondary crop (2/2)

- The area of the parcel is *evenly divided between the two crops*, the sum of the area under crops is equal to the total cropped area

- Densities method: this method aims to allocate the area of a parcel between its various crops using the ratio between its pure crop density and its mixed crop density for each crop. Unfortunately, calculated ratios do not lend themselves well to this method in practice

3.2.3-Calculating yields from field data

- **Yield in primary sampling units**
  - Calculating average yield is related to the selection method used to obtain yield plots
  - There are two methods of determining yield in primary sampling units (villages, enumeration areas or enumeration sections):
    1) ***Yield plots are selected with equal probability***
    2) ***Yield plots are selected with probability proportional to size***

3.3 Production

3.3- Production (1/3)

- **Production as a product of area and yield*:***
  - Knowing the area and yield of a crop, the production of the stratum is obtained:
    - by multiplying the two values expressed in compatible units
    - If the **yield calculation** has taken into account the **type of mixture** (whether the crop is pure, main or secondary or according to the number of mixed crops),
      - ✓ the yield of each type of mixture will be applied to the corresponding area
    - If, **however,** the types of mixture have not been taken into account
      - ✓ the average yield obtained will be applied to the total area

3.3- Production (2/3)

- **Production as extrapolated data**
  - If production is known for each holding, determining the production of the stratum
    - involves estimating a total where y represents production
      - ✓ estimation method used only in developing countries as it is impossible to obtain the production of each holding

3.3- Production (3/3)

- These methods of estimating agricultural production can result in 2 types of crop forecasting techniques:
  1) **forecasting by density grids**: this uses cropping areas and yield forecasting
  2) **forecasting by interview**: this uses declaration by the farmer

3.4 Crop forecasting

3.4.1- Forecasting from cropping areas
Scope and principle (1/2)

- This method
  - applies to cereals
  - is not suited to other crops
  - is based on the same principle as estimating final production
    - Determining an **average yield** to be applied to the various cropping areas

3.4.1- Forecasting from cropping areas
Scope and principle (2/2)

- **Yield forecasting** obtained **with the aid of density grids** and not yield grids as for final production
  - ✓ Yield grids used to estimate final yields
    - are also used in yield forecast estimates
    - are then known as density grids
      - ➢ A density grid is in some cases **part** of the yield grid with the aim of facilitating the field work

3.4.1- Forecasting from cropping areas - Reminders (1/3)

- **Reminders**
  - **The method of estimating yield from yield grids** involves:
    - placing a grid **at random** in a subsample or in all of the parcels of sample holdings
    - in order to determine an average yield after harvest

3.4.1- Forecasting from cropping areas - Reminders (2/3)

- **Reminders**
  - **The method of estimating yield forecasts** consists in:
    - counting the number of potential cobs in the yield grids
      - ✓ before harvest
      - ✓ when crop forecasting takes place

3.4.1- Forecasting from cropping areas - Reminders (3/3)

- Forecast production is calculated in the same way as final production
  - yield of grids = yield forecasts
- The yield forecast of the grid is obtained by multiplying
  - the number of potential cobs **by** mean cob weight

3.4.1- Forecasting from cropping areas: mean cob weight (1/5)

- **Determination of mean cob weight,** mean cob weight:
  - is a parameter recorded in the field and entered in survey forms
  - is determined for a given crop from
    - the **random collection** of a sample of cobs, where possible from the **previous year**
    - from each sample holding

3.4.1- Forecasting from cropping areas: mean cob weight (2/5)

- **Purpose of determining mean cob weight**
  - to apply it to the number of potential cobs present in the density grids
  - to determine the yield forecast

- As some factors influence cob weight, the ideal approach would be to apply to the number of cobs:
  - the mean weights corresponding to cobs produced under the same conditions as the environment of the grid

3.4.1- Forecasting from cropping areas: mean cob weight (3/5)

- Three (3) cases can be envisaged and the application of these
  - depends on how the variation in mean cob weight is assessed:
    - <u>Assessment 1</u>
      - ✓ Using the mean cob weight of the holding to which the parcel where the grid is placed belongs
      - ✓ Collecting a sample
        - of the previous year's harvest
        - from all the holdings where a density grid is placed
      - ✓ Finally, calculating the mean weight of the cobs collected by crop variety

3.4.1- Forecasting from cropping areas: mean cob weight (4/5)

- <u>Assessment 2</u>
  - ✓Applying a single mean weight, that of the village, to all the density grids in the same village
  - ✓Collecting a sample of cobs by variety
    - from the previous year's harvest
    - from sampled holdings where a density grid is placed
  - ✓Finally, calculating mean cob weight in the PSU in question

3.4.1- Forecasting from cropping areas: mean cob weight (5/5)

- <u>Assessment 3</u>
  - ✓Applying the same mean weight, that of the stratum, to potential cobs in the same stratum
  - ✓Collecting a sample of cobs
    - from the previous year's harvest
    - from the whole of the stratum in question
  - ✓Finally, calculating mean cob weight in the PSU in question

3.4.1- Forecasting from cropping areas
Yield forecast (1/3)

- **Yield forecast estimate**
  - Purpose of the method
    - To forecast production for each yield grid
      - To do this, the number of potential cobs is multiplied by the mean cob weight
      - The difference between the forecast and the real production of the grid thus depends on two elements:
        1) the difference between the number of potential cobs and the number of cobs at harvest
        2) the difference between the mean weight applied and the real weights of the cobs harvested

3.4.1- Forecasting from cropping areas
Yield forecast (2/3)

- Number of potential cobs
  - This number is at least equal to the number of cobs actually collected
  - Not all the cobs in the yield grid will necessarily reach maturity for various reasons
  - **Because of the collection period** of data for forecasts (e.g. September for Sahelian countries), **the majority of potential cobs** found **must** be able **to reach maturity**
  - The number of potential cobs is closer to the number of cobs harvested, the nearer the forecasting period is to the harvest
  - Keeping to the scheduled periods for counting potential cobs is therefore vital in this forecasting method

3.4.1- Forecasting from cropping areas
Yield forecast (3/3)

- Mean cob weight to be applied to cobs in the density grid:
  - can create great **differences** between the forecast production of the grid and its final production
    - ✓ The difference between the forecast production and the final production of the grid depends on the differences between
      - the mean cob weight (to be applied)
      - and the real weights of the various cobs

3.4.2- Forecasting by interview
Definition and process

- Forecasting by interview is based on the **assessment** by **producers** of the forecast production of their cereal parcels
  - The sampled producer declares for each cereal crop:
    - the previous year's production
    - the forecast production
  - The forecast growth rate of production is determined from these
    - the rate applied to the final production of the previous year obtained by the survey for forecast production
  - The results of the previous year are provided by an administrative or other division
    - the rates are calculated according to the same bodies

3.4.2- Forecasting by interview Methods (1/2)

- Two collection methods are distinguished in the interview with the farmer for a given crop:
  - <u>1st case</u>
    - The farmer is asked for **the production of** each of **the farm's parcels** for the previous year **and their forecast production**
    - <u>Drawback</u>: here the farmer is far more likely to give:
      - ✓ the total production for the previous year
      - ✓ than parcel by parcel production

3.4.2- Forecasting by interview Methods (2/2)

- <u>2nd case</u>
  - The farmer is asked for **the farm's total production** (all parcels together) for the previous year and **the forecast production of each parcel**
  - <u>Drawback:</u> asking farmers for the total production of the holding involves a risk
    - ✓ that they will only limit themselves to the production of common parcels
    - ✓ ignoring that of individual parcels
  - <u>Solution:</u>
    - ✓ Using only the common parcels on the holding to carry out the exercise

3.4.2- Forecasting by interview illustrations: annotations (1/2)

- Pour illustrer les deux cas de figure, pour une région (province, département) comportant m strates, on note:
- y la production prévisionnelle d'une culture donnée selon la déclaration paysanne
- y' la production de la dernière campagne d'une culture donnée selon la déclaration du producteur
- i numéro d'identification des strates $i = 1,2,3 \dots, m$
- n_i nombre d'unités primaires de la strate i
- y'_i la production définitive de la strate i selon l'enquête agricole
- p_i la taille de la strate i (habitants, ménages, concessions ou exploitations
- j numéro d'identification des unités primaires

3.4.2- Forecasting by interview
illustrations: annotations (2/2)

- P_{ij} la taille de l'unité primaire j de la strate i
- m_{ij} nombre d'unités secondaires de l'unité primaire j (village, SE ou ZD) de la strate i
- M_{ij} le nombre d'exploitations de l'unité primaire j de la strate i
- k numéro d'identification des unités secondaires (exploitation)
- p_k nombre de parcelles de l'exploitation k portant la culture étudiée au cours de la présente campagne
- p'_k numéro d'identification des unités primaires k ayant porté la culture étudiée au cours de la campagne précédente

3.4.2- Forecasting by interview illustrations: case 1 (1/2)

- ***1er cas :*** L'exploitant déclare pour chaque parcelle (y compris les parcelles abandonnées), la production de la campagne précédente et la production prévisionnelle
- La production prévisionnelle de l'exploitation k de l'unité primaire (village, SE ou ZD) j de la strate i est:
  - $y_{ijk} = \sum_{l-1}^{l-p_k} y_l$
- sa production de la campagne précédente est :
  - $y'_{ijk} = \sum_{l-1}^{l-p'_k} y'_l$
- La production prévisionnelle de strate i est:
  - $y_i = \frac{p_i}{n_i} \sum_{j-1}^{j-n_j} \frac{M_{ij}}{P_{ij}} \frac{\sum_{k-1}^{k-m_j} y'_{ijk}}{m_{ij}} = \sum_{i-1}^{i-n} \sum_{j-1}^{j-m_i} \frac{p_i}{n_i} \frac{M_{ij}}{P_{ij}} \frac{y'_{ijk}}{m_{ijk}} = \sum_{i-1}^{i-n_i} \sum_{j-1}^{j-m_j} c_{ij} y'_{ijk}$
- Où c_{ijk} est le coefficient d'extrapolation de la strate j

3.4.2- Forecasting by interview illustrations: case 1 (2/2)

- Le calcul du taux d'accroissement prévisionnel $T_i = \frac{y_i}{y_i'}$ de la production de la strate i entre la campagne précédente et la campagne s'en suit et, la production prévisionnelle de la strate sera
- $\hat{Y}_i = T_I * Y_i'$

3.4.2- Forecasting by interview illustrations: case 2 (1/2)

- ***2e cas :*** L'exploitant donne le total de sa production lors de la dernière campagne puis pour chaque parcelle exploitée, il donne une production prévisionnelle.
  - Les calculs se ramènent à ceux du 1er cas, la seule différence est
    - qu'ici les y_{ijk} ne sont pas calculés comme ci-dessus,
    - mais donnés directement par l'exploitation.
  - Cette méthode
    - a l'avantage d'être **simple au niveau de la collecte des données** (interview)
    - est beaucoup **plus apte à donner une tendance** (plutôt grossière) qu'un chiffre de prévision très proche de la réalisation,
      - surtout quand la campagne n'est pas uniforme dans son évolution par rapport à l'année de référence

3.4.2- Forecasting by interview illustrations: case 2 (2/2)

- the growth rate to be applied to the previous year's production is often just a guide because:
  - shortcomings of local units of measurement as units of measuring weight
  - inaccuracies in forecast declarations by the holder

3.5 Analysis and dissemination

3.5.1- Analysis techniques: Census and sample survey integration (small domain estimation)

- Two (2) main sources of data on quality of life:

    1) **Census**

    2) **Surveys of households**

3.5.1- Analysis techniques: Census

- **The census**
  - concerns all households
  - can be used to calculate **reliable estimates** according to very detailed disaggregation, at the level of towns and small villages
  - does not collect detailed information relating to
    - income or consumption
    - quality of life level and its variation,
      - ✓ e.g. poverty levels or indicators of inequality

3.5.1- Analysis techniques: Household surveys

- **Household surveys**
  - often include a detailed section on income or consumption expenditure
  - Only give generally representative information
    - in geographic sectors
    - in larger regions of the country

3.5.1- Analysis techniques: small area techniques (1/2)

- Small area techniques
  - combine sample survey and census data
  - need data from at least two sources:
    - a detailed household survey
      - ✓ which includes a quality of life indicator, generally consumption per inhabitant
    - a national census or, if unavailable, a large-scale national survey or administrative data targeting a significant part of the country's population

3.5.1- Analysis techniques: small area techniques (2/2)

….Small area techniques:

- make use of detailed data derived from
  - sample surveys of households
  - the comprehensive cover of census

- Basic principle:
  - ”to borrow strength” from related domains
  - by means of indirect estimators based
    - ✓on implicit or explicit matching models
    - ✓with the aid of additional information

3.5.1- Analysis techniques: Spatial analysis, definition (1/3)

- In statistics, **spatial analysis** or **spatial statistics**
  - covers all formal techniques that allow the study of entities based on their **topological and geometric properties**
  - involves:
    - examining places and attributes
    - studying the relations of characteristics in spatial data
    - using other analytical techniques to answer a question or produce useful knowledge

3.5.1- Analysis techniques:
Spatial analysis, definition (2/3)

- creates knowledge from spatial data
  - which broadly speaking amounts to answering the question: “what is happening where?”
- covers numerous disciplines
- uses methods which vary according to the specific questions studied, whether it involves:
  - making forecasts
  - processing large volumes of data generated by GPS and satellite remote sensing

3.5.1- Analysis techniques: Spatial analysis, definition (3/3)

- Like evaluation of the development of agriculture, spatial analysis
  - is related to numerous issues of spatial distribution
  - has spatial data analysis techniques that are important for
    - processing
    - analysing agricultural data
- Advantage:
  - summarizing complex spatial trends

3.5.1- Analysis techniques: Spatial analysis, domain

- Domain
  - agriculture
  - geology
  - soil
  - hydrology
  - the environment
  - ecology
  - mineral extraction
  - oceanography
  - forestry
  - air quality
  - remote sensing
  - epidemiology
  - disease mapping

3.5.1- Analysis techniques: Geographic information system (1/3)

- Definition: The GIS can be defined as a system designed to:
  - capture
  - record
  - handle
  - analyse
  - manage
  - present all types of geographic data

3.5.1- Analysis techniques:
Geographic information system (2/3)

- It can be considered a system which combines:
  - mapping
  - statistical analysis
  - data processing
- It can process large quantities of data obtained by remote sensing
- Geographic information systems and the related science have a significant effect on spatial analysis

3.5.1- Analysis techniques: Geographic information system (3/3)

- Remote sensing, GPS and GIS technologies are useful for:
  - mapping
    - agricultural yields
    - crops
    - harmful organisms, etc.
  - carrying out relevant analyses
- **N.B.** Professionals responsible for processing agricultural data should be capable of using GIS technologies

3.5.2- Metadata: definition

- Metadata
  - is defined as any information which helps users to:
    - find
    - understand
    - and use data and information
  - helps to
    - judge the quality of a survey
    - determine whether it meets expected needs
- The partial or total lack of metadata causes discrepancies in the analysis of results

3.5.2- Metadata: scope (1/2)

- Metadata covers the following items:
  - Concepts and definitions
  - Objective measured
  - Reference period of the data
  - Target population
  - Calculation methods
  - Sources of raw data
  - Relevant level of disaggregation

3.5.2- Metadata: scope (2/2)

- Any adjustments made to statistical sources and estimation procedures
- Shortcomings and proposed improvements
- Contact person(s) for updating the indicator
- All useful documents concerning the preparation and conduct of the collection operation
  - enumerator's manual,
  - study protocol, etc.

3.5.2- Metadata: standard

- There is no single better standard for all types of metadata, the two main ones are:
  - **Data Documentation Initiative (DDI)**: an international initiative
    - ideal for survey data and describing social science data
  - **Statistical Data and Metadata eXchange (SDMX)**: an international initiative for
    - establishing and
    - improving statistical information exchange standards
    - providing standard formats
      - ✓for data and metadata
      - ✓content directives
      - ✓IT architecture for data and metadata exchange

3.5.3- Archiving: definition and significance (1/2)

- Data archiving involves storing data in the long term
- The organization generating the data must have
  - a repository of official copies of all information made public, such as archives of records from agricultural surveys and census
- Archives
  - are:
    - an institutional memory
    - a **systematic, reliable history** of the experience of the organization which can be **consulted for planning and evaluation purposes**
  - serve to store data for future use
  - are essential for official reference purposes

3.5.3- Archiving: definition and significance (2/2)

- **Efficient** archiving **corresponds to** the organization of data production work; **files** must be:
  - named
  - indexed
  - protected
  - backed up appropriately
  - stored in appropriate digital archives at the end of the whole process

3.5.3- Archiving: cost

- All or almost all the **information relating to any data collection operation** can be **archived cheaply**, including:
  - planning files
  - operational documents
  - questionnaires
  - monitoring forms
  - data sets
  - final results
  - evaluation files

3.5.3- Archiving: contents (1/2)

- Archives should contain:
  - All information made public
  - Individual records containing all the data
    - from census or sample surveys
    - with their metadata to process them.
- Archives of sketch maps:
  - Possible when generated by geographic information software

3.5.3- Archiving: contents (2/2)

- The main constraints on the volume to be archived:
  - lie in the ability of the organization in question to:
    - store documents as they are produced
    - index them so that the information can be readily retrieved in the future

3.5.3- **Archiving:** Protection and maintenance (1/3)

- Archived data must be protected
  - Against loss, corruption and unauthorized access
  - and by setting up a **regular** program to copy them to a different medium

- When choosing the type of data dissemination, consider
  - security
  - and privacy

3.5.3- Archiving: Protection and maintenance (2/3)

- Security problems
  - digital media **more exposed** to risks than the traditional paper media
    - Measures are available
      - ✓to reduce these risks and render them negligible
      - ✓but it is essential to understand the risks and to implement these measures

3.5.3- Archiving: Protection and maintenance (3/3)

- For **Personal information** about individuals, storage in a safe place must be provided:
  - **For data lists on paper**
    - lock away when not in use
    - limit access to a small number of people
  - **For data stored on computer**
    - create individual password-protected user accounts
    - set automatic exit from the screen or the connection (ONU-CEA-CAS)

3.5.4- Database and CountrySTAT: database (1/3)

- Database
  - Definition
    - It is a structured dataset organized for easy, quick storage and recovery of data
  - Generally **created when** agricultural census and surveys generate **a large number of data**
- Numerous tools have been developed by international institutions

3.5.4- Database and CountrySTAT: database (2/3)

- Security principles when creating a database
  - provide reliable, accurate data
  - while preventing the disclosure of personal information on the individuals interviewed

- The data processed and converted into useful information is presented
  - in the form of tables
  - in the required tabulation or report formats

3.5.4- Database and CountrySTAT: database (3/3)

- Must also be included in the database:
  - The metadata that
    - describes the data
    - provides essential information about the data for interpreting it and using it wisely

- International institutions have developed numerous tools to:
  - encourage countries to make data available to users

3.5.4- Database and CountrySTAT: tools developed

- The main tools in current use over the previous 2 decades are:
  - 2gLDB (World Bank)
  - DevInfo (UNICEF jointly with the United Nations system)
  - **CountrySTAT** (FAO)
  - StatBase (ECA)
  - CensusInfo (United Nations Statistics Division, in partnership with UNICEF and UNFPA)
  - IMIS-Redatam (ECLAC)
  - NADA (National Data Archive), released by IHSN, the international household survey network, with the operational support of the World Bank, PARIS21 and the OECD
  - Data Portal (ADB)

3.5.4- Database and CountrySTAT: CountrySTAT (1/2)

- FAO's intended aim:
  - to harmonize data from different sources
    - according to international standards
    - and at the same time
  - to guarantee the quality, reliability and comparability of the data

3.5.4- Database and CountrySTAT: CountrySTAT (2/2)

- Definition: CountrySTAT is a web-based information system:
  - for food and agriculture statistics at a national and subregional level
  - which allows decision-makers to access statistics across thematic areas such as:
    - ✓ production
    - ✓ prices
    - ✓ trade
    - ✓ consumption

3.5.5- Safeguarding data-dissemination systems (1/3)

- The data management system performs three functions:
  1) access to official statistics for dissemination purposes
  2) storage and extraction of survey results
  3) access to
     - holding data
     - household data
     - geo-referenced data for research work

3.5.5- Safeguarding data-dissemination systems (2/3)

- The data management system
  - covers the various data sources necessary for preparing
    - supply/utilization accounts
    - food balance sheets
    - various economic and environmental accounts and all derived statistics
  - also enables agriculture statistics to be incorporated into the national statistical system

3.5.5- Safeguarding data-dissemination systems (3/3)

- Data dissemination mechanisms:
  - publications (reports, summaries, brochures, etc.) in printed or electronic versions
  - web sites
  - workshops, etc.
- Persons responsible for analysing agricultural data must be familiar with:
  - database management methods
  - data security issues
  - existing mechanisms for disseminating agricultural information

Exercises

- **Exercise 19: laying a yield grid**
- **Exercise 20: yield**
- **Exercise 21: area/yield/density**
- **Exercise 22: density & area**
- **Exercise 23: area**
- **Exercise 24: yield and production: specific case of plantain**

THANK YOU