

**FARM STRUCTURE SURVEY 2007**

**NATIONAL METHODOLOGICAL REPORT**

**MEMBER STATE: SPAIN**

# Contents

Summary

## **1 Introduction**

- 1.1 History, scope
- 1.2 National legislation

## **2 Content**

- 2.1 Characteristics
- 2.2 Questionnaires

## **3 Survey methodology**

- 3.1 Survey organisation
- 3.2 Calendar (overview of work progress)
- 3.3 Preparing the survey operations
  - 3.3.1 Population and frame
  - 3.3.2 Survey design
  - 3.3.3 Pilot survey
  - 3.3.4 Informing and training the staff and respondents
- 3.4 Sampling, data collection and data entry
  - 3.4.1 Drawing the sample
  - 3.4.2 Data collection
  - 3.4.3 Utilisation of administrative data sources
  - 3.4.4 Control of the data
  - 3.4.5 Non-response
- 3.5 Data processing, estimation and analysis
  - 3.5.1 Methods for handling missing or incorrect data items
  - 3.5.2 Estimation and sampling errors
  - 3.5.3 Non sampling errors
  - 3.5.4 Evaluation of estimates

## **4 Publication and dissemination**

### **ANNEXES:**

Annexe I: Questionnaire

Annexe II: Formulas applied for estimation methods and calculating sampling errors

Annexe III: Scientific references

## Summary

The field work for the Farm Structure Survey 2007 (*Encuesta sobre la estructura de las explotaciones agrícolas*, EEA 2007) was conducted from 01.10.07 to 31.01.08.

The data refer to the 2007 agricultural campaign, i.e., the period 01.10.06 to 30.09.07, except for livestock, for which the reference date was the day of the interview.

We only investigated holdings with at least 1 ha of Utilised Agricultural Area (UAA) and holdings which, though below that threshold, met either of the following conditions:

- Farm holdings with at least 0.2 ha of UAA under cultivation of any of the following crop groups: vegetables, flowers and ornamental plants (grown outdoors or under low protective cover) or crops under glass or irrigated fruit trees (including citrus) or nurseries.
- Holdings which in the Agricultural Census 1999 had one or more Livestock Units (LSUs) and a Total Gross Margin (TGM) equal to or greater than 0.75 European Size Units (ESUs).

Data collection for EEA 2007 was carried out through INE's Provincial Offices, except in the Autonomous Communities of Catalonia and the Basque Country, where the respective regional statistical institutes, under cooperation agreements, took over the collation and data-entry of questionnaires in their territories.

A single hardcopy questionnaire model was used for data collection. In Catalonia and the Basque Country, which have their own co-official languages, questionnaires were bilingual.

Data were collected by personal interviews with farm holders or their appointees. Interviews were previously announced and arranged by a letter of appointment informing the respondent of the date, time and place of the interview. Generally, the appointment venue was the holder's (or his/her appointee's) address. In municipalities with a large number of sample holdings, however, interviews were held at offices made available for the purpose (premises temporarily loaned by local councils or other municipal bodies). Wherever possible, interviewers confirmed appointments with respondents by mobile telephone. Appointments were rescheduled if required.

### **The initial sample of 51,218 holdings for EEA 2007 was made up of:**

- Panel holdings and "daughter" holdings already interviewed in the 2005 survey, or were classed "discontinued" in the 2005 survey. They totalled 50,771 farms.
- Additional units totalling 447 holdings drawn from the livestock register and the central enterprise directory (Directorio Central de Empresas).

The reserve sample was not used for the EEA 2007 survey. As in the previous survey, however, additions to the sample were identified by the procedure of researching possible "daughter farms". A range of details on collection of the sample are provided below:

- Completed questionnaires were returned by 90.8% of holdings. 'Discontinued' holdings made for a further 4%. The remaining 7% was accounted for by incidents preventing obtainment of the questionnaire, among which refusal to respond represented less than 1%.
- In the course of field work, in order to **detect any new (daughter) holdings** not appearing in the 1999 Agricultural Census, we researched sample holdings where land

or livestock facilities had since been assigned or sold. The research was carried out under very specific processing criteria and enabled us to detect 1,443 daughter holdings, representing 3% of the initial sample.

For information editing, interviewers and Interviewer Inspector comprehensively reviewed questionnaires before recording them.

Questionnaire handling and recording was implemented using an interactive software application specifically developed by INE for this Survey. The application covered data validation by a range of logical controls to detect errors and inconsistencies, which were corrected by contacting the respondent over again where necessary. The application operates on a centralised database, so the information recorded at INE Provincial Offices was under ongoing control from Head Office. This enabled us to detect areas for correction or improvement and plan visits and inspections at specific Provincial Offices.

Once the field work was completed, the Provincial Offices transmitted the data to INE's Head Office, where the information was combined with the datasets from the Basque Country and Catalonia. At Head Office there ensued a new centralised review and control process of errors, which can be rectified using the same interactive software application as the Provincial Offices. After this second manual edit, an automatic data imputation phase was carried out to correct any remaining errors.

Afterwards, the phases of calculation of derived variables, data elevation and any operations required to obtain the final results are performed.

Once the final results are analysed and approved, the information will be tabulated and published in December 2008. Results will be stated at the national and Autonomous Community scale.

The tabulations obtained and our methodology (concepts, definitions, sampling, survey organisation, and so forth) are published on the INE website and made available to all users.

We shall also compile the EUROFARM dataset and cater for customised information requests.

# 1. Introduction

## 1.1 History, scope

The history of general surveys of farm structure in Spain began at INE with the first Agricultural Census of 1962. Further Agricultural Censuses were conducted in 1972 and 1982.

From Spain's accession as a full member to the European Community on 1 January 1986, INE joined the programme of 'Community surveys on the structure of agricultural holdings' created under Council Regulation No 70/66.

Spain first took part in the Community programme with the sampling survey of 1987. In accordance with Council Regulation (EC) No 571/88, on the organisation of Community surveys on the structure of agricultural holdings between 1988 and 1997, INE carried out an exhaustive survey, i.e., a census, in 1989, and sampling surveys in 1993, 1995 and 1997.

The Council Regulation (EC) No 2467/96 amended the earlier Regulation so as to cover the period 1998 to 2007. The Regulation requires all Member States to conduct a census in 1999 or 2000 and sampling surveys in 2003, 2005 and 2007. Therefore, INE conducted the Agricultural Census in 1999 and surveys in 2003, 2005 and 2007.

The Commission Regulation (EC) No 204/2006 adapts Council Regulation (EC) No 571/88 and amends Commission Decision 2000/115/EC with a view to the organisation of Community surveys on the structure of agricultural holdings in 2007.

The 2007 Survey pursues two key aims:

- a) To assess the state of Spanish agriculture, monitor structural change in agricultural holdings and produce results that may be compared across all European Union Member States.
- b) To comply with the Council Regulations, meet domestic statistical needs and cater for international requests for farm statistics.

To meet these goals and facilitate the comparability of series, the new 2007 Survey mostly keeps to the scheme of the latest censuses and surveys.

## 1.2 National legislation

Unlike censuses, surveys are not covered by specific legislation.

This is because under the *Ley 12/1989, de la Función Estadística Pública de 9 de mayo* ("the Public Statistics Act" or "LFEP"), which governs statistical activity in Spain, any statistic listed in the National Statistics Plan has the status of a statistic for Government use, and its compilation is thus mandatory.

Furthermore, additional provision 2 of the *Ley 13/1996, de 30 de diciembre, de Medidas Fiscales, Administrativas y del Orden Social* provides that a statistic is mandatory if European Community law requires Spain to compile it.

The Farm Structure Survey 2007 appears under programme number 4002 among the operations listed in the National Statistics Plan 2005-2008, enacted under Royal Decree 1911/2004 of 17 September 2004, and in Royal Decree 1575/2006 of 22 December 2006, which approves the 2007 Annual Programme of the National Plan, and Royal Decree 1756/2007 of 28 December 2007, which approves the 2008 Annual Programme of the Plan.

The 2007 Annual Programme sets out a brief description of the Survey effort:

- Bodies involved in the Survey: INE, Autonomous Communities
- Specific work scheduled for 2007: creation of questionnaires and survey design; sample selection; organisation of field work; Provincial Inspector training courses; data collection.
- Estimation of budget allocations required to finance the Survey, which must be allocated in the National Budget for the year.

The 2008 Annual Programme adds the work scheduled for 2008:

- Data entry (January-March 2008)
- Data editing, computer processing, tabulation and analysis (February-November 2008)
- Dissemination and publication of tables and micro-dataset (December 2008)
- Production of the EUROFARM micro-dataset for transmission to Eurostat (December 2008).

The Programme also specifies the budget credits needed for 2008.

Moreover, as stipulated by LFEP for all surveys and censuses, the Survey questionnaire must include the main legal points: nature, features and purpose of the Survey; statistical confidentiality; the duty to supply data; and penalties for violating the Act.

## **2. Content**

### **2.1 Characteristics**

The Survey investigates all the characteristics on the Community list (Regulation (EC) No 204/2006) except B.1.b, which, as the Regulation states, is non-significant (NS) in Spain.

In the national interest, as in earlier farm structure surveys, INE posits variables in addition to those required in EC law:

- For all crops making up the UAA, we determine whether the land is irrigated or not, including successive and associated crops.
- Greater disaggregation by crop regarding characteristics such as leguminous, industrial, foraging, citric, fruit, woody crop nurseries, crop association and successive crops.
- Livestock categories are broadened and a distinction drawn between animals owned by the holding and animals in "integration system" (animals belonging to an operator that supplies farms with feed and other inputs and remunerates farmers at a fixed price per animal).
- We research the main types of holder's legal personality.
- Regarding the introduction of the agricultural activity 'maintaining land in good agricultural and environmental conditions (GAEC), some areas which took part to the agricultural area (under F: F01+ F02) in 2005, have been now included under F03. Therefore, the total area under permanent grassland has not changed. Farms having only this kind of activity could not have appeared due to the design of the sample. The sample consists of the same panel of farms from 2003 to 2007. This sampling design could not very well measure the effect of new characteristics. In the next Census, this problem will disappear the real figure of this characteristic will be know.

The Survey data refer to the 2007 agricultural campaign, i.e., the period 01.10.06 to 30.09.07, except for livestock, for which the reference date was the day of the interview.

## 2.2 Questionnaires

Data collection was conducted by entry on a single hardcopy questionnaire model of data supplied by farm holders.

In the Basque Country and Catalonia questionnaires were bilingual: Basque/Spanish, Catalan/Spanish.

Interviewers collected data in the closing quarter of 2007 and January 2008, after the end of the 2007 agricultural reference period.

The Survey questionnaire is attached to this Report.

## 3. Survey methodology

### 3.1 Survey organisation

EEA 2007 was planned by INE's Head Office. The following units were involved:

Units involved	Functions
Agricultural Statistics Desk	Coordination of all survey performance operations. Editing of the farms directory of the 1999 Agricultural Census. Framing the method, setting rules of procedure and the schedule of work, editing, analysis and production of results.
Design and Sampling Desk	Sample design. Production of the sample, calculation of estimators and sampling errors.
IT Department	Development of the software application for handling, recording and validating survey data and coverage control. Programming for data processing. Tabulation of results. Creation of EUROFARM file
Coordination and Planning Desk	Coordination of cooperation agreements with the regional statistics offices of Catalonia and the Basque Country
Sub-Directorate General for Data Collection (Field Work Desk)	Field work planning. Coordination of collection units (Provincial Offices) and preparation of Survey materials (with the support of the Publications Unit so as to publish and print the various materials). Control and monitoring of collection work.
Secretariat General	Budgetary control and staff hirings.
Statistical Publications Department	Printing of questionnaires, collection manuals and other Survey materials. Publication and dissemination of results.

To do its work, the Agricultural Statistics Desk assigned the following staff from June 2006 onward:

- Three statisticians holding higher graduate qualifications
- Two statisticians holding graduate qualifications
- Two statistical assistants

Five temporary employees were engaged for work on editing and processing data from 26.03.08 through to completion of the project.

The Design and Sampling Desk hired a statistician holding higher graduate qualifications in June 2007.

From June 2007, the IT Department has employed a statistician holding graduate qualifications, a PL1 programmer and a TPL programmer, who have been working on all computer programming tasks.

The Field Work Desk, for the work described above, assigned the following staff from March 2007 to February 2008:

- Two statisticians holding higher graduate qualifications
- Two statisticians holding graduate qualifications
- One administration officer

The field work was done from Provincial Offices by temporary employees. Each Provincial Office assigned a statistician holding graduate qualifications to act as Provincial Inspector for the Survey.

In the Autonomous Communities of the Basque Country and Catalonia, INE signed a cooperation agreement with the respective regional statistics offices for the collation and recording of questionnaires in their territories.

The table below sets out the staff assigned to the Survey in all Provincial Offices, itemised by professional category and employment period:

Categories	Employment period	Staff numbers
<b>Provincial Inspectors (statisticians holding graduate qualifications)</b>	<b>Permanent civil servants</b>	<b>45</b>
<b>Interviewer Inspectors</b>	<b>17-sep 2007 – 31 Jan 2008</b>	<b>85</b>
<b>Interviewers</b>	<b>1 Oct 2007 – 31 Dec 2007</b>	<b>360</b>



### 3.2 Calendar (overview of works progress)

Survey phases and tasks	Unit responsible	Dates
<b>Draft project and other initial tasks</b>	<b>Agricultural Statistics Desk</b>	
Methodology		Sept 06- March 07
Questionnaire		Sept 06- March 07
Validation and programming rules		March-Oct 2007
<b>Sample design, sample selection and calculation of sampling errors and estimates</b>	<b>Sampling Unit</b>	<b>June 07- Dec 2008</b>
Sample design		
Sampling selection		
Resolution of incidents in collection phase		
Calculation of sampling errors and estimates		
<b>Preparation of survey material</b>	<b>Field Work Unit</b>	<b>April - Sept 2007</b>
Edition of questionnaires		
Route sheets and other field work material		
Advertising posters		
Letters of introduction and appointment		
Preparation of training material (manuals, handbooks, etc.)		
<b>Centralised IT processes</b>	<b>IT Unit</b>	<b>from June 2007</b>
Pre-programming for design		June-Sept 2007
Detection of errors		Sept-Dec 2007
Automatic imputation		Jan-April 2008
Calculation of aggregate variables		Feb-June 2008
Tabulation of results		April - June 2008
Programming of EUROFARM dataset		junio-2008
<b>Collection and recording software</b>	<b>IT Unit</b>	<b>from May 2007</b>
Development of first phase (up to directory editing)		May-July 2007
Development of second phase (up to dispatch of letters of appointment)		agosto-2007
Development of third phase (recording, validation, coverage control)		Sept-Oct 2007
Maintenance, helpdesk, etc.		July 2007- Feb 2008
Creation of end of survey national dataset		Feb-March 2008
<b>Coordination, assistance, monitoring and follow-up tasks of the central field work unit</b>		
Training course for 45 Provincial Inspectors for the Survey	In cooperation with the Agricultural Statistics Desk and IT Unit	junio-2007
Application for temporary staff to be recruited at provincial offices		junio-2007
Application for mobile phones for use by Interviewers		junio-2007
Distribution of material to provincial offices		septiembre-2007
Guidance during collection		Sept 2007 - Jan 2008
Monitoring and follow-up of collection		bi-weekly from Oct 2007 to Feb 2008
<b>Collection of survey</b>	<b>Provincial Offices</b>	<b>from Oct to Jan 2008</b>
Prior editing of identification and localisation data		July - Sept 2007
Recruitment of Interviewer Inspectors		17 Sept 2007- 31 Jan 2008
Training course for Interviewer Inspectors		17-21 Sept 2007
Recruitment of Interviewers		1 Oct -31 Dec 2007
Training course for Interviewers		1-3 Oct 2007
Dispatch of letter of introduction to farm holders		17-21 Sept 2007
Dispatch of letters of first appointment and, where applicable, second		from 1 October 2007
Conduct of interviews		from 8 Oct to 31 Dec 2007
Questionnaire editing, recording and validation		from 8 Oct 2007
Dispatch of thank you letters		from 15 Oct 2007
Creation of provincial end of survey datasets		Dec 2007- Jan 2008
<b>Information processing</b>	<b>Agricultural Statistics Desk</b>	<b>Feb to Dec 2008</b>
Review of provincial end of survey datasets		
Editing of information		
Automatic imputation programme		
Aggregate variable calculation		
Analysis of results		
Dispatch of EUROFARM dataset		

### 3.3 Preparing the survey operations

#### 3.3.1 Population and frame

##### Population

The population surveyed conforms to the Community definition, in line with the following criteria:

- All farms with at least 1 ha of Utilised Agricultural Area (UAA).
- All farm holdings with at least 0.2 ha of UAA under cultivation of any of the following crop groups: vegetables, flowers and ornamental plants (grown outdoors or under low protective cover) or crops under glass or irrigated fruit trees (including citrus) or nurseries.
- Holdings which in the 1999 Agricultural Census had one or more Livestock Units (LSUs) and a Total Gross Margin (TGM) equal to or greater than 0.75 European Size Units (ESUs).

The three criteria are independent: at least one must be met for a holding to be eligible for the Survey.

This definition of agricultural holding is used for all surveys since 1987. The number of holdings in the Agricultural Census of 1999 meeting these criteria was 1,287,418.

To meet national statistical needs, the 1999 Census – like earlier censuses – increased the scope of holdings under study to a total of 1,790,162 holdings.

In censuses, the study population is defined by one of the following two conditions:

- Farms the total area of which is equal to or greater than 0.1 ha, in one or more plots, which need not be contiguous.
- Farms with less than 0.1 ha but possessing one or more head of cattle; two or more head of horse, mule or ass; six or more head of sheep or goat; two or more head of hogs; fifty or more poultry, including chicken, turkey, duck, goose, guinea-fowl, pigeon, quail, pheasant and partridge bred in captivity; thirty or more breeding doe rabbits; ten or more beehives. This livestock may be reared in rural or urban areas.

##### Frame

Frame is a list of holdings. It is obtained as a result of the Agricultural Census 1999 and the threshold indicated above. Administrative sources had been also used.

**Time frame of reference and updating process of the frame:** year 1999. We built a farm panel (retaining the same sample throughout all occasions). This panel is investigated and updated every two years (at the time of the survey).

#### 3.3.2. Survey design

**Sampling design** is a probabilistic sampling: stratified simple random sample.

The sample consists of a panel of farms or holdings. Only on the first time we carry out the survey (the year 2003) we admit extra farms when the sample unit is non-response. The target is to maintain the initial sample size over time. Initial sample size is obtained applying optimum allocation. Having a limited maximum sample size of about 55,000 units, we vary the coefficient of variation of some key variables that are under control to compute

minimum sample size into strata. The initial total sample size is 53,859. The disappearance rate between 2005 to 2007 has been about 3% .

The sample is retained across time (until a new census). We update it using the method of the daughter farms. When a farm from our panel sample is death, we investigate if this farm (it is called mother farm) has generated a birth (daughter farm). In this case, we include the new farm into the sample.

**Stratification:** In each geographical area (NUTS 2), strata are built by community typology (two digit categories) and five size classes. To define the five size classes we consider five key variables: economic size, annual working unit, livestock unit, utilised agricultural area and total area taken by arable land, permanent crops and kitchen gardens. The cum  $\sqrt{f}$  rule is applied separately to each key variable. Lastly, we define size classes of a unit as the maximum of the size classes obtained for these five key variables in this unit.

**Allocation method:** This is the optimum allocation using the Bethel algorithm.

**Strata where the sampling ratio is 100%:** First of all, before stratification, we decide what farms will be in an exhaustive stratum. The sigma-gap rule is applied separately to each key variable using all farms having a non-zero value for the variable in question.

### 3.3.3 Pilot survey

No pilot survey was conducted.

### 3.3.4 Informing and training the staff and respondents

Survey staff was trained in accordance with the following plan:

- June 2007: training course for Provincial Inspectors imparted by the Agricultural Statistics, the Field Work and the IT Desks (all are Head Office Units).
- September 2007: training course for Interviewer Inspectors, run at each Provincial Office by the respective Provincial Inspector.
- October 2007: training course for Interviewers run at each Provincial Office.

Methodological points and field work procedures were documented in a Collection Manual.

The following measures were taken to promote the Survey among respondents:

- Brochures providing the key figures of the 2005 Farm Structure Survey were issued to all respondents.
- A 2008 calendar was created using a design relating to agricultural products and seasons for provision as a gift to cooperative respondents.
- Advertising posters were printed in Spanish, Valencian, Galician and Majorcan for display in public places. Catalonia and the Basque Country published their own posters.
- Information was sent to the main farmers' organisations to seek their help and support in promulgating the Survey and encouraging farm holders to cooperate.
- All respondents received a letter of introduction to the Survey stating Survey goals and the variables to be investigated.
- Provincial Offices asked Municipalities – particularly those with a large number of sample holdings – to make available premises or offices for conducting interviews.
- Some Provincial Offices announced the Survey on local radio stations.

### **3.4 Sampling, data collection and data entry**

#### **3.4.1 Drawing the sample**

The sample is obtained systematically by ordering the units of the frame prior to obtaining the sample by variable TL (Arable land plus Kitchen gardens and permanent crops) for predominantly cropping TF2 units, by variable LSU (livestock units) for predominantly livestock TF2 units, and by variable AA (agricultural area) for predominantly grazing TF2 units.

To select the sample, we used customised programs developed with the SAS statistics package.

#### **3.4.2 Data collection and entry**

##### **Data collection organisation**

Each INE Provincial Office (except in Catalonia and the Basque Country, owing to the cooperation agreements with the respective regional statistics offices, mentioned above) collected data for the holdings in its province.

At each Office, a Government Statistics graduate (Diploma-holder) was appointed Survey Provincial Inspector, being the technical officer in charge of the Survey in that province. The Provincial Inspector headed a team of temporary staff comprising Interviewers, in charge of conducting interviews and completing questionnaires, and one or more Interviewer Inspectors, whose role was to oversee Interviewers, review questionnaires for errors and support the Inspector.

360 Interviewers and 85 Interviewer Inspectors were engaged for all INE Provincial Offices.

Field work was chiefly conducted from October to December 2007. In January 2008 questionnaire validation work was completed.

Data were collected by personal interviews with farm holders or their appointees.

The field work procedure was as follows:

- 1) Before the start of the Survey, towards the close of September 2007, all farm holders selected in the original sample were sent a letter of introduction, announcing the Survey, stating its goals, and stressing the importance of respondents' cooperation. With the letter we enclosed a brochure of the 2005 Farm Structure Survey data.
- 2) The Provincial Inspector scheduled the field work, assigned time frames to the sample and allocated work to Interviewers.
- 3) Fortnightly or weekly, farm holders to be interviewed in the following week(s) were sent a letter. The letter stated the date, time and venue of the interview (generally the holder's or his/her appointee's address, although some interviews were conducted at municipal offices) and listed the key facts that respondents would be asked about. Details were also given of a contact person whom the farm holder could approach for any queries.
- 4) Prior to the interview date, wherever possible, the Interviewer would telephone the holder to confirm that he/she had received the letter and the interview could proceed as planned. All Interviewers were provided with mobile telephones to confirm appointments.

5) The Interviewer conducted his/her allocated interviews and completed questionnaires with data supplied by respondents. Before the end of each interview, he/she would carry out a first editing of errors by checking the consistency of data.

6) If due to the respondent's absence an interview could not be held on the arranged date, the Interviewer would leave a second letter announcing a second visit if the interview had been scheduled at the holder's address, or the Provincial Office would send the holder a second letter of appointment by post.

7) If difficulties emerged in the course of field work, they were noted in the relevant route sheet (sample replacements were not effected). Where land or livestock facilities had been wholly or partly transferred, potential daughter farms were investigated and documented in the relevant route sheet.

8) On a weekly basis the Interviewer would turn in to the Interviewer Inspector a report on the previous week's work, completed questionnaires and the respective route sheets. The Interviewer Inspector would then review all this material and check for errors exhaustively before handing it over to the Provincial Inspector. At least fortnightly, each Interviewer's weekly work reports were recorded using the Survey software application.

9) The data entry and validation process was performed according to the Survey calendar at all times.

### **Data collection methods**

Questionnaires were completed by personal interview with farm holders (or their appointed respondents).

Questionnaire data were recorded using a software application designed and developed by INE specifically for the Survey. A range of controls and checks were conducted during data entry itself. Later on, the application subjects entered data to a set of rules for internal consistency of questionnaires (partial absence of data, inconsistent data and range control). A total 103 logical and preventive controls were in place.

### **Completion time per questionnaire**

Completion time was dependent on the size of the agricultural holding, the diversity of crops and livestock, the extent to which the respondent was cooperative and organisational considerations, among other factors. Therefore, completion time could vary substantially from one respondent to another.

In the route sheet of EEA 2007, questionnaire length was specified. This included questionnaire completion and investigation into potential daughter holdings. The questionnaire lasted an average of 20 to 24 minutes.

### **Data entry modes**

A software application was designed using the ORACLE language to manage field work. The program's functions covered the following six main areas:

- 1) Prior editing of the directory;
- 2) Survey work allocation;
- 3) Generation of letters and route sheets;
- 4) Recording of coverage, route sheets and daughter questionnaires – reference to census;
- 5) Work report to Head Office;

## 6) Production of Survey completion datasets.

The data-entry program allows the user to enter codes and values as he/she enters data, or choose not to use codes. Questionnaires are recorded by keying in each code to be entered (with the respective datum), and the application lets the user record or edit the code value in the appropriate format.

### **Processes and techniques of the data transmission to the national statistical office/final centre.**

The Survey completion dataset is created directly at INE Head Office, though a copy can be produced at the Provincial Office, plus a dataset of recorded remarks.

### **3.4.3 Utilisation of administrative data sources**

No administrative data sources were used.

### **3.4.4 Control of the data**

#### **Data control at Provincial Offices**

- 1) Prior to recording data, Interviewers would check that each questionnaire was complete and consistent.
- 2) Prior to recording data, Interviewer Inspectors review 100% of questionnaires.
- 3) After data entry, data were validated automatically by the software application. Error lists were produced, divided into 'type I errors' (clearly incorrect, and therefore must be amended) and 'type II errors' (reflect an anomalous or unusual situation, but are not necessarily incorrect; the anomaly must be duly accounted for).

Provincial Offices corrected all type I errors and verified type II errors.

#### **Control at INE Head Office**

Once the Survey completion dataset has been produced, data control is centralised at INE Head Office, where questionnaires are sorted and filed and staff is deployed for data control work.

The computer-driven process was divided into several stages for each province or dataset:

Stage 1: Detection and correction of coverage errors and problems (duplicates, mistaken identification, etc.).

Stage 2: Detection and correction of errors in the internal consistency of questionnaires. New controls relating to organic farming, irrigation and labour were added to the logical and preventive controls already applied at Provincial Offices.

Stage 3: Automatic Residual-Error Imputation Program.

Stage 4: Provisional calculation of derived variables (without imputation of non-responding holdings)

Stage 5: Imputation of holdings and final calculation of derived variables

Stage 6: Incorporation of elevation factors Production of final datasets

Stage 7: Conversion of the format of datasets produced at the previous stage for use in the following stage

Stage 8: Tabulation of results

Stage 9: Generation of dataset for Eurostat

Manual correction of stage 1 and 2 errors was conducted on screen using the field-work software application.

To correct these errors, questionnaire data were checked against data from other sources. Where necessary, respondents were contacted by telephone to correct or confirm data, as applicable.

After manual correction, all questionnaires underwent an Automatic Imputation Process (AIP) at stage 3.

### 3.4.5 Non-response

The reasons preventing questionnaire completion (final defaults, including non-response) were as follows:

- **Discontinued ('AB')**: a holding at which none of the land or livestock facilities were dedicated to any agricultural or livestock activity, although the agricultural orientation of the farm remains in place; no European Community subsidy is received for such relinquishment.

- **Other purposes ('OF')**: a holding at which all the land and livestock facilities held by the holder were used for non-agricultural purposes.

- **Erroneous inclusion in sample ('EI')**: a holding which, though agricultural, does not meet the minimum thresholds to enter the Survey sample.

- **Holding transferred ('CE')**: the farm holder states that he/she no longer owns any part of the holding (land or fixed livestock facilities) due to having sold or assigned the entire holding.

- **Duplicate ('DU')**: a holding included in the sample more than once.

- **Uncontactable ('IL')**: the holder or an appropriate respondent for the holding could not be contacted.

- **Refusal ('NE')**: a holding for which the holder or his/her appointed respondent refused to provide the required data, despite attempts by Survey officers of various grades. This incident represented less than 0.4% of the initial sample.

In Farm Structure Surveys 2005 and 2007, unlike 2003, non-responses in collection were not dealt with by using replacements from the reserve sample. For IL and NE non-responses, questionnaires were imputed with data from the previous survey.

Potential daughter farms were investigated for all sample holdings (except for units coded IL, DU and NE).

The table below summarises the final data collection result of the EEA 2007 Survey.

EEA 2007	Total	% compared to initial sample
<b>Initial sample</b>	<b>51,218</b>	<b>100.0%</b>
Questionnaires collected (RE <sub>t</sub> )	46,490	90.8%
Incidents in initial sample	4,728	9.2%
AB	1,999	3.9%
OF	164	0.3%
EI	592	1.2%
CE	1,305	2.5%
DU	105	0.2%
IL	366	0.7%
NE	197	0.4%

We took the following measures to minimise non-responses:

- All selected sample units were sent a letter of introduction briefly setting out the aims of the Survey. Respondents were told that a later letter would indicate the date and venue of interview. A summary of the data that would be asked about was also included. We reminded respondents that their data would be protected by Statistical Confidentiality, and that response to the Survey was mandatory.
- Farm holders (or their appointed respondents) were sent a letter of appointment stating the date, time and place of the interview. In each province interview rounds were generally organised by municipality. Interviews were usually conducted at the respondent's address, though specific interview offices were set up in some municipalities. As far as possible, holders living in a province other than where their holding was located were offered alternatives to interview (for example, self-completion of the questionnaire with telephone guidance, or appointment of a proxy respondent).
- If the holder or appointee failed to attend the interview, he/she was sent a second letter establishing a further appointment.
- Before each interview, interviewers confirmed the appointment wherever possible.
- Respondents were appropriately informed that response to the Survey was mandatory. Failure to respond may involve incurring a fine. However, given the low proportion of refusals to respond, the penalties procedure was only initiated in a minimal number of cases.

### 3.5. Data processing, analysis and estimation

#### 3.5.1. Methods for handling missing or incorrect data items

Besides manual correction as mentioned above, we used a custom-designed Automatic Imputation Program (AIP) to handle missing data in questionnaires.

The AIP comprises fourteen blocks, each of which performs a specific function. Blocks are applied in the sequence 1 to 14 to each holding; each block basically conducts three types of operation:

- Queries to detect inconsistencies.
- Queries to acquire information from the questionnaire itself where inconsistencies have been detected.
- Imputations as necessary.



If there are no inconsistencies, the block makes no imputations to the holding in question and moves on to the next holding. If a block applies imputations to a holding, the amended data are final; therefore, queries made by subsequent blocks refer to the updated rather than the original data. This also applies within each block. Queries made after application of one or more imputations always refer to the updated state of data, even where modified within the process of that same block.

Imputations are of two general types. One type is deduced from the information in the questionnaire itself by applying given criteria, while the second type requires recourse to external information to make up for missing data in the questionnaire. Imputations of the first type, where they relate to arithmetic inconsistencies, squaring up sums for instance, generally operate by imputing new data in proportion to those appearing in the questionnaire the sum of which verifies the desired consistency condition. Imputations of the second kind refer to information drawn from a set of Hot Deck (HD) matrices designed for the purpose, which record data for holdings processed previously. The information is classified by size (total surface area) of holdings; when used to remedy arithmetic inconsistencies, new data are imputed following the same criterion of proportional allocation. HD matrices have to be set up prior to the process in respect of each province, based on the highest probability as given by earlier surveys and the agricultural statistics annual.

Owing to the battery of prior controls, the AIP made corrections to only 1.65 % of the total data.

In the cases of non-response total, we impute questionnaires until the theoretical sample size using data from the last census is reached.

### 3.5.2. Estimation and sampling errors

**Estimation method:** We use simple expansion estimators.

**Composition of extrapolation factor:** In general, this consists of basic weight, inverse probability selection. It is seldom adjusted due to incorrect classification.

**Variance estimation:** We use the standard formula.

**Sampling error:** We compute and disclose relative standard sampling error of the key variables by geographical area.

To calculate these errors we use an SAS program. The SAS programs are tailor-made. Formula are provided in annexe II.

### 3.5.3 Non sampling errors

As a frame error, we incurred an excess coverage rate of almost 8%. As the sample is panel-based, to preserve the initial selection probabilities we only made stratification changes in influential units. We made 388 size stratum changes in the 2007 Farm Structure Survey. The response and non-response rates are detailed at section 3.4.5.

### 3.5.4 Evaluation of results

At the error-editing stage, comparisons are made at the micro-data level with the 2005 Survey and other administrative sources, such as the Organic Farms Registry.

At the macro level, results are compared to censuses and farm structure surveys and various sources of the Ministry of the Environment and Rural and Marine and other bodies, such as data on arable land, livestock surveys, the Organic Farms Registry, rural property registries, etc.

## 4.– Publication and dissemination

In 2007, before field work began, we published the 2007 Farm Structure Survey Design. The Design set out all the Survey definitions and concepts, the sample design, how field work was to be organised, the questionnaire and the publications schedule. Hence from the start users were provided with all the information needed to interpret the Survey properly. The Design is publicly available on INE's website, along with other statistical projects.

The first results are published on the Internet in late 2008. As for earlier surveys, general data will be released on land use, form of ownership or tenancy, legal personality, labour force, livestock, risk and rural development, at both Autonomous Community and the national levels. The results are classified by utilised agricultural area and farm type.

We also created a register of anonymised individual data available to users requesting it.

We cater for customised information requests.

# Annexes

## Annexe II: Formulas applied for estimation methods and calculating sampling errors

### Estimators

Let:

$N_h$ =Frame population in stratum h.

$d_h$ =Original-sample units of stratum h defaulting under any of the following codes:

–'EI': erroneously included

–'CE': holding assigned without creation of daughters

–'AB': holding discontinued without creation of daughters

–'OF': holding used for other purposes, without creation of daughters

$e^2_h$ =Original-sample or reserve-sample units of stratum h defaulting under any of the following codes:

–'CE': holding assigned, with creation of daughters

–'AB': holding discontinued, with creation of daughters

–'OF': holding used for other purposes, with creation of daughters

$n_h$ =Units of the theoretical initial sample in stratum h.

$n^e_h$ =Units of the main or reserve sample in stratum h which are not daughter holdings

$n^v_h$ =Daughter holdings of holdings in stratum h.

Definitions:

$$\hat{N}_h = N_h \left(1 - \frac{d_h}{n_h}\right)$$

$$F_h = \hat{N}_h / (n^e_h + e^2_h)$$

The estimator of the total of a variable Y in stratum h is:

$$\hat{Y}_h = \sum_{i=1}^{n^e_h + n^v_h} F_h * Y_{hi}$$

The total of Y for a given Autonomous Community is estimated as the sum of all strata of that Autonomous Community. The national total is estimated as the sum of all estimated strata.

### Estimation of sampling errors

Let Y be the variable for which we wish to estimate the variance of the estimator of its total in stratum h.

The variable is defined as:

$$Y'_{hi} = \sum_{m=0}^{K_{hi}} Y_{hi_m}$$

where:

– $i$  varies from 1 to  $n^e_h + d_h + e^2_h$  (let this value be  $n^t_h$ )

– $k_{hi}$  is the number of daughter holdings of holding  $i$ .

– $Y_{him}$  is the value of  $Y$  for the  $m$ -th daughter holding of holding  $hi$ . In particular,  $Y_{hi0}$  is the value of  $Y$  for the mother holding.

Then:

$$\hat{V}(\hat{Y}_h) = \hat{V}(\hat{Y}'_h) =$$

$$= \frac{N_h^2 * (1 - f'_h) \sum_{i=1}^{n^t_h} (y'_{hi} - \bar{y}_h)^2}{n_h^t (n_h^t - 1)}$$

$$f'_h = \frac{n^t_h}{N_h}$$

The variance of the total of a variable for a given sum of strata is given by the sum of variances of the strata.

### **Annexe III: Scientific references**

- 1.- Bethel, J. (1989), Sample Allocation in Multivariate Surveys. Survey Methodology, June 1989. Vol 15, N°1, pp 47-57**
- 2.- Brion, P. et Hoffmann, A, (1985), La base de sondage Permanente. Service Central des enquêtes et études statistiques, N° 11, Ministère de l'Agriculture (France)**
- 3.- Cochran, W.G. (1977), Sampling Techniques, New York: Wiley**
- 4.- Cox, B.G. et al. (1995), Business Survey Methods, Wiley**
- 5.- Julien, C. And Maranda, F. (1990), Sample Design of the 1988 National Farm Survey. Survey Methodology, June 1990. Vol 16, N° 1, pp 117-129**
- 6.- Valliant, R and Dorfman, A.H and Royall, R.M, (2000), Finite Population Sampling and Inference, Wiley**