Istat
Istituto Nazionale di Statistica

Microdata File for Research

Farm Structure Survey.
Year 2005

Survey methodology

July 2009

CONTENTS

1. Introduction 3
2. Target population 4
3. Sampling design 4
4. Data collection and data treatment 5
5. Weights calculation method 7
6. Dissemination of the survey results 8
7. Glossary 8
8. References 10
9. Contacts 10

1. Introduction

In 1967 Italian National Statistical Institute (Istat) carried out the first Italian sample survey on agricultural holdings aiming to provide a statistical tool able to draw a coherent and consistent picture on the primary sector's structure.

The following surveys were carried out in 1975, 1977, 1982, 1985, 1987, 1990, 1993, 1995, 1996, 1997, 1998, 1999, 2000, 2003.

Three of these surveys (1982, 1990, 2000) were carried out as full surveys (census) and they provided the frame for the following sample surveys.

Starting with 1993 edition the survey added to the structural scopes stated by Council Regulation (EEC) n.571/88 the short terms objectives stated by the following European normative:

837/90 (concerning statistical information on cereals production);

959/93, 2197/95, 296/2003 (concerning statistical information on crop products other than cereals);

93/16 (on statistical surveys of milk and milk products);

93/23 (on the statistical surveys to be carried out on pig production);

93/24 (on the statistical surveys to be carried out on bovine animal production).

93/25 (on the statistical surveys to be carried out on sheep and goats stocks);

milk.

Furthermore, some additional topics were surveyed by a specialized section of the questionnaire as shown in the following list:

- 1997 - fruit trees production;
- 1998 - environments;
- 1999 - rural development.

Since 2003 survey an approach closer to local administrations' purposes has been used. As it will be shown with more details in the next sections, sample design has been determined considering accuracy on variables of local interest and some items were introduced in the questionnaire in order to take into account some local needs.

In this paper we report the data as stemming from Regulation (CE) 2139/2004: in particular information about the sampling design, the data collection and dissemination.

2. Target population

Target population, or *reference population*, is defined as the set of farms with the following characteristics in 2005 agricultural year:

- the agricultural area utilized for farming is one hectare or more, or;
- the agricultural area utilized for farming is less than one hectare if they produce a certain proportion for sale (2.066 €) or if their production unit has exceeded certain physical threshold.

This definition of reference population is coherent with the definition used by 2000 census and by 2003 FSS.

Target population is a subset of the *Italian farms population* that contains farms whose size is below above-mentioned thresholds too.

Data have been collected with a random sample selected according to a stratified sample design with a *take all stratum* containing the biggest farms.

The theoretical sample size is 56,542 selected from the target population. Furthermore all farms resulting from a splitting or a merging of a sampling unit have been added to the sample by the interviewers. The number of respondents is 47780.

3. Sampling design

The sampling design is a stratified one with a *take all stratum* containing the biggest farms.

The theoretical sample size is 56,542 selected from the reference population. Furthermore all farms resulting from a splitting or a merging of a sampling unit have been added to the sample by the interviewers.

The stratification of units has been carried out as follow:

- First - the take all stratum has been defined using the *UAA*, *LSU* and *ESU* of each unit;
- Second - the reference population has been stratified according to location (region or province), dimension (*UAA*, *LSU* and *ESU*) and typology; furthermore a flag to identify public holdings have been used;
- Third - the remaining units of the population list have been stratified using the region code.

The sample size and its allocation have been defined aiming at achieving a given accuracy on the estimate totals of some variables of regional and national interest.

A coordinated selection (*positive coordination*) with previous FSS has been carried out looking for the maximum overlapping of the samples 2003-2005.

Definition of the take all stratum

Since the target population is much skewed (a lot of small farms and few big farms) it is advisable to put the most influential units in a stratum with a 100% sampling ratio.

For this survey the take all stratum has been defined using the method and the algorithm suggest by Hidiroglou (1996). This method helps in finding out an optimal threshold such that units exceeding the threshold should be included in the sample with certainty. The optimum threshold is defined aiming at a minimum sample size necessary to achieve a given accuracy in the estimate of the total of a variable of interest.

The algorithm has been implemented separately on three variables for this survey.

The first variable was *UAA* and the resulting threshold (273 hectares) has been obtained referring to the entire population list (2.6 million of units) and aiming at achieving an accuracy of 3%.

The second variable was *LSU* and the resulting threshold (446 *LSU*) has been computed considering the same set of units and the same accuracy used for *UAA*.

The third variable considered for the definition of the take all stratum was *ESU*. For this variable a threshold has been computed separately for each subpopulation listed in table 1.

The take all stratum of the 2003 FSS has been defined as the set of units exceeding at least one threshold.

Table 1: threshold (*ESU*) for take all stratum for principal types of farming

Principal types of farming	**Threshold**
Specialist dairying	404
Specialist cereals, oilseeds and protein crops	141
Various permanent crops combined	103
Specialist olives	70
General field cropping	144
Specialist fruit and citrus fruit	115
Mixed cropping	110
Specialist horticulture	267
Specialist vineyards	81
Other farms	400

The take all stratum has been defined as the set of agricultural holdings exceeding at least one of the three thresholds and including about 6.000 units.

4. Data collection and data treatment

Istat has sent to each regional office in charge of the survey the list of sampling units. Furthermore Istat has sent a personal mail to each sampled unit.

Each Region has assigned the sampling units to the recruited interviewers on provincial bases (around 30-40 units for each interviewer).

Each interviewer should have a first contact with the assigned units by phone in order to arrange a meeting with the farm holder for a "face-to-face" interview using the personalized paper questionnaire supplied by Istat.

If a sample unit splits in two or more farms the interviewer must fill in a questionnaire for each new unit other than one included in the original sample unit reporting the date of the split and the number of new farms.

Some efforts should be done contacting local farmer organizations or administrations by interviewer in order to up date wrong contacts references or to trace the new farms.

The filled questionnaires have been collected by a regional o provincial interviewer's coordinator in order to check the quality of the data and the work carried out by each interviewer. The coordinator could have corrected some data (using a red ink pen) and had to sign each recordable questionnaire.

Regional administrations are responsible to data entry using the software developed in BLAISE by Istat. The recorded data have been sent to Istat by web through a security procedure in line with the norms governing the treatment of personal information sending security

Two approaches have been used to handle missing or incorrect data items.

The first one is used to face the problem for influent farms (usually farms in the take all stratum or units with a high weight). In this case two actions are commonly carried out in order to obtain complete and correct questionnaires:

- a comparison between collected data and regional administrative data (if available);
- a telephone check performed by Istat's survey staff.

The second approach for data editing and imputation strategy is used for the other farms (non influent farms) and it is summarized as follows.

Data editing and imputation strategy

It has to be firstly underlined that errors regarding variables codes, unit codes, localization codes, sums and some incoherency of data are identified at the data entry stage (hard checks). This strategy guarantees the accuracy of entered data with respect to basic quality criteria.

However, some other errors or inconsistencies arising either at the data collection or at the data entry phase may affect the data. These errors may have a systematic or a random source and produce logically or mathematically not coherent and not acceptable information. These errors can be identified through the so called *query edits* (based on the analysis of observed distributions or data relations).

As a consequence of the complexity of the questionnaire and the possible errors, a mixed and hierarchical editing and imputation strategy has been designed, in which different approaches have been combined in order to deal with the different types of variables and potential errors (Di Zio, Luzi 2002).

As relating to the error localization phase, in addition to the traditional deterministic approach, adopted for either systematic errors or errors having a known source, a selective editing approach (Latouche et al., *1992*, Lawrence et al., *2000)* combined with graphical data representations has been adopted for the identification of influential outliers, while a probabilistic approach based on the Fellegi and Holt paradigm (Fellegi and Holt, 1976) has been used to identify errors that are not influent on target estimates and for which a random origin can be assumed.

As relating to the correction and imputation phase, both the interactive treatment (for outliers and influential errors) and the automatic approach are used. In the latter the non parametric *hot-deck nearest neighbor donor* imputation technique is adopted to predict not acceptable or missing values together with other methods based on the use of known statistical relations between variables.

5. Weights calculation method

The survey estimates of totals for national and regional domains will be produced using the following estimator

$$\tilde{\tilde{Y}} = \sum_{k \in s_{r,d}} y_k w_k$$

where $s_{r,d}$ is the set of respondent and new farms belonging to domain *d*, w_k is the final weight of unit *k* and y_k. is the variable of interest. The final weights will be obtained as a product of three factors

$$w_k = d_k \, \gamma_{1k} \, \gamma_{2k}$$

where $d_k = \frac{1}{\pi_k}$ is the sampling weight and it is the result of sample design (for sake of simplicity the stratum index is skipped), γ_{1k} is the factor used to avoid problems arising with total non response and γ_{2k} is the factor used to include some auxiliary information at the estimation stage.

The second factor, γ_{1k}, is computed as the inverse of the response rate on each stratum. In few cases the factor has been computed collapsing two similar strata.

The third factor, γ_{2k}, is used to achieve the consistency of sample estimates with respect to some known totals of the population. It is computed using the calibration theory explained in Estevao, Hidiroglou and Särndal (1995).

The variance estimator for regional and national estimates is defined by formula of Estevao, Hidiroglou and Särndal (1995) that is implemented in the software *GENESEES* (available on Istat's web site: http://intranet.istat.it/doutdes/SW_statistico_1/repositary/index.html).

) for the sampling error estimation step.

6. Dissemination of the survey results

Plan of dissemination of the 2005 FSS results forecasts:

- Transmissions of validated individual data and quality report to Eurostat before 31st October 2006 in line with the time limit fixed by Commission Regulation (EC) n.2139/2004.
- Transmissions of validated individual data to the requesting Regions before 30 November 2006.
- Dissemination of data tables and meta-date on the ISTAT website available for all users: http://www.istat.it/dati/dataset/20070613_00/ and press release available at: http://www.istat.it/salastampa/comunicati/non_calendario/20061227_00/.
- Publication of the data tables in the Yearly Agricultural Statistics book during 2007. http://epp.eurostat.ec.europa.eu/portal/page/portal/agriculture/data/database.

7. Glossary

Agricultural holding: A single unit both technically and economically, which has single management and which produces agricultural products. The holding may also provide other supplementary (non-agricultural) products and services.

The holder: The holder of the holding is that natural person, group of natural persons or the legal person on whose account and in whose name the holding is operated and who is legally and economically responsible for the holding, i.e. who takes the economic risks of the holding. The holder can own the holding outright or rent it or be a hereditary long-term leaseholder or a usufructuary or a trustee.

Manager of the holding: The natural person responsible for the normal daily financial and production routines of running the holding concerned.

Utilised agricultural area: The total area taken up by arable land, permanent grassland, permanent crops and kitchen gardens.

Owner-farmed agricultural area: Agricultural land being farmed by the holding being surveyed and which is the property of the holder or farmed by him as usufructuary or heritable long-term leaseholder or under some other equivalent type of tenure.

Tenant-farmed agricultural area: Land rented by the holding in return for a fixed rent agreed in advance (in cash, kind or otherwise), and for which there is a (written or oral) tenancy agreement. An area is allocated to only one holding. If an area is rented out to several holdings during the

reference year, it is normally allocated to the holding with which it is associated for the longest time period during the reference year.

Share-farmed agricultural area: Land (which may constitute a complete holding) farmed in partnership by the landlord and the sharecropper under a written or oral share-farming contract. The output (either economic or physical) of the share cropped area is shared between two parties on an agreed basis.

Colonia parziaria of complete holdings. In colonia parziaria of complete holdings the grantor entrusts a farm to the head of a family, who undertakes to carry out with the aid of the members of his family (the farm family) all the work required on the holding and himself to bear part of the outgoings and to divide the farm's production with the grantor in fixed proportions.

Organic farming: Information is to be collected on whether or not the holding practises agriculture according to certain set standards and rules specified in Council Regulation (EEC) No 2092/91 (1), as last amended by Commission Regulation (EC) No 473/2002 (2) or, where applicable, the most recent legislation, on .organic production of agricultural products and indications referring thereto on agricultural products and foodstuffs., and/or the equivalent national rules for organic production.

Total area: The total area of the holding (D to H) consists of the agricultural area utilised by the holding and other land**.**

Arable land: Land worked (ploughed or tilled) regularly, generally under a system of crop rotation

Kitchen gardens: areas devoted to the cultivation of agricultural products intended for consumption by the holder and his household.

Permanent grassland and meadow: Land used permanently (for five years or more) to grow herbaceous forage crops, through cultivation (sown) or naturally (self-seeded) and that is not included in the crop rotation on the holding.

Permanent crops: Crops not grown in rotation, other than permanent grassland, which occupy the soil for a long period and yield crops over several years.

Other land: Other land. includes unutilised agricultural land (agricultural land which is no longer farmed, for economic, social or other reasons, and which is not used in the crop rotation system) and land occupied by buildings, farmyards, tracks, ponds, quarries, infertile land, rock, etc.

Livestock: Number of production animals that on the day of the survey are in the direct possession or management of the holding. The animals are not necessarily the property of the holder. These animals may be on the holding (on utilised areas or in housing used by the holding) or off the holding (on communal grazing or in the course of migration, etc.).

Farm labour force: The statistical information on farm labour force is collected in a way that makes it possible to compose tables where the data (e.g. on age and working time) from the different farm labour force categories and classes are crossed with each other and/or with any other

survey characteristics. This means that each person doing farm work on the holdings is to be classified according to all classes required for the category.

Rural development: Information is collected on whether the holder and/or the spouse or other family members or one or more partners on a group holding carry out any gainful activities that do not comprise any farm work as defined under L/1 to L/6 but which are directly related to the holding and have an economic impact on the holding.

8. References

Ballin M., Falorsi P.D., Scepi G. e Vicard P. (1998). "New tools for planning business sampling surveys". Proceeding of the international seminar on New Techniques and Technologies for Statistics, Sorrento (Italy) 4/6 November 1998, pp. 49-53

Bethel, J. (1989), "Sample allocation in multivariate surveys", *Survey Methodology*, 15, 47-57

Di Zio, M., Luzi, O. (2002), *Combining Methodologies in a Data Editing Procedure: an Experiment on the Survey of Balance Sheets of Agricultural Firms*, Statistica Applicata, 14, 1, pp. 59-80.

Estevao, Hidiroglou and Särndal (1995) "Methodological Principles for a Generalized Estimation System at Statistics Canada", *Journal of Official Statistics*, vol. 11, n.2, pp181-204

Fellegi, I.P., Holt, T.D. (1976), *A Systematic Approach to Edit and Imputation*, Journal of the American Statistical Association,71, 17-35.

Hidiroglou M.A. (1996) "The Costruction of Self-Representing Stratum of Large Units in Survey Design". *The American Statistician*, 40, 27-31.

Latouche, M., Berthelot, J.M. (1992), *Use of Score Functions to Prioritise and Limit Recontacts in Editing Business Surveys*, Journal of Official Statistics, 8, 3, Part II.

Lawrence, D. and R. McKenzie (2000), *The General Application of Significance Editing*, Journal of Official Statistics.

9. Contacts

Massimo Greco
msmagrec@istat.it
Tel +390659524570

Curatori dei capitoli

Massimo Greco

Daniela Fusco