



Community Innovation Survey (CIS2008)

2006-2008

Survey methodology

December 2010

CONTENTS

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

1. Introduction

The Italian Innovation Survey (Cis) covers innovation activities of the Italian enterprises with at least ten employees operative in industry and services. In particular, Cis collects information about new or significantly improved goods or services (product innovations) and new or significantly improved processes, logistics or distribution methods (process innovations), as well as about organizational and marketing innovation. Most questions refer to product and process innovations: to this regard, the survey provides a wide and articulated set of indicators on innovation activities, innovation expenditure, public funding, sources of information for innovation, innovation co-operation and innovation objectives. The last edition of the Italian Innovation Survey - the sixth Europe-wide Cis - has produced data with reference to the three-year period from 2006 to 2008 inclusive.

The survey is part of the Eu Community Innovation Survey (Cis), carried out on a two-year basis (from 2004 onwards) by all Eu Member States and candidate countries, plus Norway and Iceland. Cis provides information on a number of dimensions of innovation so as to monitor Europe's progress on innovation and, since 2000, Cis has also become one of the major sources of data for the *European Innovation Scoreboard*, and it has confirmed by the European Commission one of the flagship initiatives for measuring the performances of the Innovation Union within the Eu2020 strategy.

Since 2004, Cis has been conducted under the *Commission Regulation No 1450/2004*, which establishes the legal basis for innovation statistics and makes it compulsory to deliver data every two years on some key variables.

In order to ensure a sound comparability across countries, the Cis is carried out on the basis of a standard core questionnaire and a harmonised survey methodology developed by Eurostat, in close cooperation with the participating countries. The methodological basis of Cis is provided by the *Oecd Oslo Manual*.

In order to satisfy the users' needs, Istat disseminates micro-data files for scientific research. The dissemination of the statistical information is performed in full compliance with the regulations pertaining the privacy of respondents.

This document gives information on the survey methodology used for the Italian Cis 2008. In Section 2 the target population is described. In Section 3 the sampling design is illustrated. In Section 4 the data collection and treatment methods are presented, while in

Section 5 the methodology used for the weights computation is described. The last sections provide some information on the main channels for the data dissemination as well as the statistical definitions of innovation used in the survey and the key references for the survey development.

2. Target population

2.1 Economic activity - Nace Rev.2

In accordance with annex IV of the [Commission Regulation No. 973/2007](#) on innovation statistics, the following industries are included in the European core target population of the Cis 2008:

- mining and quarrying (NACE 05-09)
- manufacturing (NACE 10-33)
- electricity, gas steam and air conditioning supply (NACE 35)
- water supply; sewerage, waste management and remediation activities (NACE 36-39)
- wholesale trade, except of motor vehicles and motorcycles (NACE 46)
- transportation and storage (NACE 49-53)
- publishing activities (NACE 58)
- telecommunications (NACE 61)
- computer programming, consultancy and related activities (NACE 62)
- information services activities (NACE 63)
- financial and insurance activities (NACE 64-66)
- architectural and engineering activities; technical testing and analysis (NACE 71)

In addition, the Italian Innovation Survey has covered the following NACE 'non-core' industries: NACE Rev. 2 sections F and I, and NACE Rev. 2 divisions 45, 47, 59, 68, 72, 77.

2.2 Size-classes

According to the [Commission Regulation No. 1450/2004](#) on the production and development of Community statistics on innovation the statistics by size class are broken down into the following size classes:

- 10 - 49 employees
- 50 - 249 employees
- 250 + employees

The variable used for identifying size classes is the average number of persons employed in the year 2008.

2.3 Statistical units

The main statistical unit for CIS 2008 is the enterprise, as defined in the [Council Regulation 696/1993](#) on statistical units or as defined in the national statistical business register. [EU Regulation 2186/1993](#) requires that Member States set up and maintain a register of enterprises, as well as associated legal units and local units.

3. Sampling design

Data were collected through a combination of census and sample survey. The census concerned all the enterprises with more than 249 employees. For the rest of population, a stratified random sample was built. The sample survey was based on a stratified random sample with equal inclusion probabilities for all population units. At least 5 enterprises in each stratum were selected. In the case of less than 5 enterprises in a stratum, a full census was conducted. The target population was broken down into strata. The stratification is made taking mainly into account the study-domains for the output tabulation defined at European level.

The stratification variables used were:

1. the economic activities (in accordance with NACE Rev. 2). Stratification by NACE has been done at two-digit (division) level, except for sections F and I;
2. enterprise size according to the number of persons employed. The size-classes used were the following: between 10 and 49; between 50 and 249; 250 and more;
3. regional aspects. The breakdown of national territory into regions was performed on the basis of the NUTS level 2.

A multi-variable and multi-domain sample allocation was used. The adopted procedure was an application of the Bethel algorithm (Bethel, 1989). It was an optimum allocation since it aimed at minimizing survey costs under the constraint that sampling errors of estimates of each variable of interest didn't exceed the given upper bounds assigned to each of them. For the pursuit of the best allocation, three auxiliary variables were used: number of persons employed, turnover and total innovation expenditure. In particular, previous Cis 2006 results were used, all referring to the year 2006.

The type and the number of domains of interest are shown in the following table:

Types and number of domains of interest of the Italian CIS 2008

Code	Type of domain (partition of population of interest)	Number of domains (in the partition)
DOM1	<i>NACE Rev.2 Division (2-digit), except for sections F and I</i>	55
DOM2	<i>NACE Rev.2 Section (distinguishing in Eu core services from non Eu core services) by size-class</i>	48
DOM3	<i>NUTS 2 region by size-class (2: SMEs/large enterprises) and macro-sector (3: Industry/Eu core services/non Eu core services)</i>	126

To keep the response burden down, a coordinated selection technique (Jales sampling) was adopted in order to avoid the inclusion of the same enterprises in the sample over time (Ohlsson et al., 1995).

The official, up-to-date, statistical business register, called Asia (Archivio Statistico delle Imprese Attive - statistical business register of active enterprises) was used as sampling frame. It provides both the key variables for the stratification (number of employees, Nace economic activity, Nuts geographical information) and the identification characters (enterprise name, address, etc.). With reference to the Cis 2008, the survey universe consisted of all the profit enterprises and independent professional units active at least for six months in the reference year.

The net initial sample (just the eligible units) included 38,200 enterprises, with an overall sample rate of 18.3 per cent of the target population.

4. Data collection and data treatment

The Italian CIS 2008 was mainly a web survey. Data capturing makes use of an Istat Web site (INDATA), which hosts the majority of Istat surveys: <https://indata.istat.it>. Respondents - through their browsers - can access an electronic questionnaire in PDF format, put on the Web site, and fill in it online.

Since it was the first time that the CIS used the electronic questionnaire, a self-compiled paper form sent via postal mail has been associated to the Web mode.

After the launch of the survey in February 2009, in order to minimise unit non-response, two reminder letters were sent out by mail to non respondents in May and September 2009, plus a special recall to the largest enterprises (500 + employees) in November 2009. Data collection ended in December 2009.

Enterprises included in the final sample are 19,904, with an overall response rate¹ of 52.1%. The e-response rate (defined as 'electronic valid responses/total valid responses') is 69.8% .

During the processing cycle, there has been a sustained follow up with the responding enterprises in order to obtain the missing information and to keep at a minimum items non-response. Several data quality checks both at micro-level and macro-level were done before the results are finally processed. An updating of a SAS software appositely designed by Eurostat for Cis has been used for data processing.

The types of checks can be grouped into five categories:

- Completeness checks. These refer to the questionnaires that are not fully completed. Contact were made with the surveyed units to get the missing information.
- Out of scope checks concerning the units that do not belong to the target population (i.e. wrong NACE, wrong size etc). The out of scope units were dropped from data processing.
- Domain or coding checks, aimed at verifying if the given answers were permissible (i.e. the answer is within the range of answers allowed). Such errors were corrected to bring it into line with the range allowed.
- Consistency checks to verify if there are some situations in which a combination to several questions is possible or if relationships between two variables are within specific bounds.
- Routing errors' checks concerning the right/wrong comprehension of the sequencing of questions.

Errors are corrected, firstly by contacting respondents or using administrative, historical (e.g. the CIS 2006 survey) and other available data sources (such as R&D survey), and then by using different statistical methods of imputation.

A first type of imputation was the logical (or deductive) imputation, that was applied to correct inconsistencies or for imputing some item non response, based on the answers provided in related questions. After having cleaned all the data, the SAS imputation programmes went on correcting metric variables separately from ordinal variables.

¹ The overall (unit) response rate is defined as (the number of units with a response) / (the total number of eligible and unknown eligibility units in the sample). Eligible are the sample units which indeed belong to the target population. Frame imperfections always leave the possibility that some sampled units may not belong to the target population. Moreover, when there is no contact with sample units and no other way to establish their eligibility they are characterised as 'unknown eligibility units'.

As far as the metric variables are concerned, a weighted mean of each metric variable, by NACE and size class, was calculated (after having removed outliers) and applied as a ratio to the enterprises with the missing values within each stratum.

The imputation of the ordinal and nominal was done after the metric estimation. The technique used is nearest-neighbour hot decking. Nearest-neighbour imputation uses data from clean records (free of logical inconsistency) to impute missing values of recipient records. The donors were chosen in such a way that the distance between the donor and the recipient is minimised. The distance measure is a multivariate measure based on the reported data.

After running the estimation procedure for nominal and ordinal variables, the next step was to run macro-level checks in order to check if the CIS 2008 results are in line with other published results. In particular, these checks were carried out by comparing CIS 2008 and SBS data, on one side, and CIS 2008 and CIS 2006 data, on the other.

5. Weights calculation method

Calibration estimators methodology, currently applied at Istat, was used for the estimation process (Deville and Särndal, 1992). It can be applied to the extent that the known totals of some auxiliary variables, strictly correlated to the variables of interest, are available. These calibration estimators have the following properties: they are more efficient than the direct estimators because of the auxiliary information used; they reduce the bias effect due to the non response and under coverage; they produce estimates of auxiliary variables that equal the known totals of such variables.

The final weights are obtained by adopting the following procedure: an initial weight is assigned to each sampled unit with reference to the sampling plan as the reciprocal of the inclusion probability. Two correction factors for initial weights are then calculated: a first one is the unit non response factor; a second one is to satisfy equality between estimation of auxiliary variables and known totals from the Register. The final weights are thus obtained as the result of the product between initial weights and correction factors. For CIS, as well as for most of the business surveys, number of enterprises and number of employees were used as auxiliary variables, according to the information provided by the Italian Official Business Register ASIA.

The software used was GENESEES, a generalised software implemented in SAS language by ISTAT researchers, available for all users from ISTAT website:

http://www.istat.it/strumenti/metodi/software/produzione_stime/genesees/index.html

6. Dissemination of the survey results

Besides the microdata file for scientific research, several statistical information products are available.

The Italian CIS 2008 aggregated indicators are disseminated mainly through the ISTAT web-site: see http://www.istat.it/salastampa/comunicati/non_calendario/20101209_00/

The aggregated indicators of the European survey may be found at Eurostat web-site: <http://epp.eurostat.ec.europa.eu>.

The researchers might analyse the non-perturbed microdata file at the Italian or European Research Data Centers. More information is available at the ISTAT and EUROSTAT web-sites: respectively, <http://www.istat.it/dati/microdati/adele.html> and <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>.

7. Glossary

The questionnaire, which is reproduced in file ISTAT_MFR_CIS_Questionnaire_2008.pdf, provides the definitions of many of the concepts specific to innovation. The main ones are repeated here, in so far as they are used to categorise enterprises as having innovation activity or not, and to determine different kinds of innovators.

Innovation: a new or significantly improved product (good or service) introduced to the market or the introduction within an enterprise of a new or significantly improved process. Innovations should be new to the enterprise concerned. Innovation must be new to the enterprise, but it does not need to be new to its sector or market. It does not matter if the innovation was originally developed by your enterprise or by other enterprises.

Product innovation: a new good or service or a significantly improved good or service with respect to its capabilities, such as improved software, user friendliness, components or sub-systems. The simple resale of new goods purchased from other enterprises and changes of a solely aesthetic nature are not included.

Process innovation: a new or significantly improved: 1) methods of manufacturing or producing goods or services; 2) logistics, delivery or distribution methods; 3) supporting activities for own processes, such as maintenance systems or operations for purchasing, accounting, or computing. Purely organisational or managerial changes are not included.

Innovation activities: activities aimed at developing or introducing new or significantly improved products (goods or services) or implement new processes, including R&D activity. They also include still on-going innovation activities by the end of 2008 or abandoned innovation activities during 2006 to 2008.

Innovation expenditure: expenditure sustained by the enterprise for activities aimed at developing or introducing new or significantly improved products (goods or services) or implement new processes. It includes expenditure on: Research and development (both in-house and performed outside); Acquisition of machinery, equipment and software; Acquisition of other external knowledge, such as patents and non patented inventions, know-how, etc.; Training; Market introduction of innovations; Design and Procedures and technical preparations to implement innovations.

8. References

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

Deville, J.C. and Särndal, C.E. (1992) "Calibration estimators in survey sampling", *Journal of the American Statistical Association* 87, 367.382

OECD (2005) "Oslo-Manual. Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition", Organisation for Economic Co-Operation and Development, Paris.

Ohlsson, E., B. G. Cox, D. A. Binder, B. N. Chinnappa, A. Christianson, M. J. Kott e P. S. Colledge (eds.). "Coordination of samples using permanent random numbers". In *Business Survey Methods*. Wiley, New York, 1995.

9. Contacts

For questions on the survey, please contact Valeria Mastrostefano, mastrost@istat.it .

10. Authors

The document was prepared by Valeria Mastrostefano.