

Italy 1995: Survey Information

Summary table

Generic information	
Name of survey	Survey on Household Income and Wealth – SHIW (<i>Indagine sui Bilanci delle Famiglie Italiane</i>)
Institution responsible	Bank of Italy
Frequency	Every two years
Survey year / Wave	1995
Collection period	May to September 1996
Survey structure	Cross-sectional and panel
Coverage	Private households in the whole territory
Geographic information	20 administrative regions (more detailed info on provinces is not available for external users)
Files delivered	Two sets of files, one for the historical database and one for the annual database, each of which comprising several files at different levels (household, individual, pensions, transfers, employment activities, real estates, etc.).
Sample size	
Households	8,135 households
Individuals	23,924 individuals (of which 14,699 income earners and 20,472 aged 15 and over)
Sampling	
Sampling design	<i>Initial sample (1987)</i> : two-stage stratified sampling, with the stratification of the PSUs (municipalities) by region and demographic size. <i>Subsequent samples</i> : households residing in panel municipalities that had participated in at least 2 surveys were all included in the sample; the remaining panel households were selected randomly from among those interviewed in the previous survey only; the non-panel households were selected randomly from municipal registers in both panel and non-panel municipalities.
Sampling frame	Municipal registry office records.
Questionnaires	Paper-based questionnaire with a modular structure: general part addressing aspects relevant to all households and a series of annexes with questions relevant to specific subsets of households.
Standard classifications	
Education	8 categories (only 6 available in historical database)
Occupation	6 categories for employees, and 6 for self-employed, used as labour force status and not occupation
Industry	10 sectors
Income	
Reference period	Income in the preceding calendar year (which coincides with the fiscal year)
Unit of collection	Mostly at the individual level, except for property income (household level)
Period of collection	Mostly monthly income with number of months, some annual.
Gross/net	All variables are recorded net of taxes and contributions.
Data editing / processing	
Consistency checks	Standard post-survey consistency checking procedure.
Weighting	Survey data can be grossed up to aggregate values thanks to appropriate weights assigned to each household according to its probability to be included in the survey.
Imputation	All the elementary variables that make up the aggregates are imputed; regression models are used to estimate the values to assign to the missing answers on the basis of other available information that is correlated with the missing data.

This document draws extensively upon the methodological Annex to the “I bilanci delle famiglie italiane nell’anno 1995”, *Supplementi al Bollettino Statistico – Note metodologiche e informazioni statistiche*, Bank of Italy, Year VII, No. 14, March 1997 (see

http://www.bancaditalia.it/statistiche/ibf/statistiche/ibf/pubblicazioni/boll_stat/supplemento_famiglie_1995_n.14_97.pdf).

Table of contents:

- A. General Characteristics
- B. Population, sample size and sampling methods
- C. Data collection and acquisition
- D. Definition of the survey units
- E. Contents
- F. Quality of data
- G. Uses of the survey

A. General characteristics

Official name of the survey/data source:

Survey on Household Income and Wealth – SHIW (*Indagine sui Bilanci delle Famiglie Italiane*)

Administrative Unit responsible for the survey:

Bank of Italy Research Department Divisione Rilevazioni e Metodi Statistici - R.M.S. Address: Via Nazionale 91, 00184 ROMA. WWW: http://www.bancaditalia.it e-mail: studi.indagini@insedia.interbusiness.it
--

The Survey on Household Income and Wealth (SHIW) began in the 1960s within the Research Department of the Bank of Italy with the aim of gathering data on the incomes and savings of Italian households. Over the years, the scope of the survey has grown and now includes wealth and other aspects of households' economic and financial behaviour such as, for example, which payment methods are used.

Until 1987 the Bank of Italy's survey of Italian household budgets was conducted with time-independent samples of households. In order to facilitate analysis of changes in the phenomena being investigated, since 1989 part of the sample has comprised households that were interviewed in previous surveys (panel households). The sample used in the most recent surveys (carried out generally every other year) comprises about 8,000 households (24,000 individuals), distributed over about 300 Italian municipalities and it is representative of the whole Italian population.

The survey results are regularly published in the Bank's Supplements to the Statistical Bulletin. The data on the households is freely available, in an anonymous form, for further elaboration and research. Other than the annual datasets, a historical database has been put together comprising a slightly restricted selection of variables available in the normal annual files but consistent over all the years since 1977.

B. Population, sampling size and sampling methods

Coverage

The sample is representative of the whole Italian population.

Sample size

Table 1a shows the sample size used between 1987 and 2000, indicating the number of households interviewed in more than one survey. For example, of the 8,135 households that made up the sample in this survey, 126 had participated since 1987, 701 since 1989, 1,752 since 1991 and 1,066 since 1993. The remaining 4,490 were being interviewed for the first time.

Table 1a
Households interviewed in the 1987-2000 surveys

Year of first interview	Year of survey						
	1987	1989	1991	1993	1995	1998	2000
1987	8,027	1,206	350	173	126	85	61
1989		7,068	1,837	877	701	459	343
1991			6,001	2,420	1,752	1,169	832
1993				4,619	1,066	583	399
1995					4,490	373	245
1998						4,478	1,993
2000							4,128
Sample size	8,027	8,274	8,188	8,089	8,135	7,147	8,001

The overall size of the sample for the 1995 survey was 8,135 households. The proportion of panel households was 44.8 per cent.

Sampling design

The sample was drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and demographic size. Within each stratum, the municipalities in which interviews would be conducted were selected by including all municipalities with a population of more than 40,000 and randomly selecting smaller towns. The individual households to be interviewed were then selected randomly.

In order to form the panel, the municipalities were selected from among those already sampled in the 1993 survey (panel municipalities). Households residing in these municipalities that had participated in at least two surveys were all included in the sample (about 3,000 households); the remaining panel households to be interviewed were selected randomly from among those interviewed in the previous survey only.¹

In order to obtain information on intergenerational aspects, all households that had been established out of the original panel households were also contacted (these were normally new households formed by the children of the original household). There were a total of 102 such households.

The non-panel households were selected randomly from municipal registers in both panel and non-panel municipalities. Households were interviewed in 310 municipalities of which 281 were panel households and 29 non-panel households (Table 2a).²

Table 2a
Survey municipalities

Geographical area	Panel	Non-panel	Total
North	119	10	129
Centre	61	7	68
South and Islands	101	12	113
Total	281	29	310

C. Data collection and acquisition

Data collection

The interviews for the sample survey of Italian household budgets in 1995 were conducted between May and September 1996.

Interviews were conducted by a specialized company using professional interviewers. The interview stage was preceded by a series of meetings at which Bank of Italy officials and representatives of the company gave instructions directly to the interviewers. The households contacted for interviews, who are guaranteed complete anonymity, receive a booklet describing the purpose of the survey and giving a number of examples of the ways in which the data are used.³ The participating households may request a copy of the results of a previous survey.

¹ Differently from the previous surveys, panel households to interview were selected independently from the eventual availability to re-interview they had given in the preceding survey.

² Panel households that had changed their residence were, as far as possible, interviewed at their new address even if this was in a different municipality, as long as it was in Italy. This ultimately expanded the number of municipalities in which interviews were conducted to 326.

³ Households receive no compensation for interviews. When the results of the survey are published, the participants are sent a thank-you letter with copies of newspaper articles commenting on the survey.

Interviewers contacted 14,298 households, of which 57 per cent agreed to be interviewed (Table 3a).⁴ The sample was therefore composed of 8,135 households, of which 3,645 were panel households and 4,490 non-panel households. The participation rate was, as is normally the case, higher for panel households (77.8 per cent, compared with 46.7 per cent for non-panel households).

Table 3a
Households contacted and reason for non-participation (percentages)

Households:	Panel		Non-panel		Total	
	Number	Percentage	Number	Percentage	Number	Percentage
Interviewed	3,645	77.8	4,490	46.7	8,135	56.9
Refusals	779	16.6	2,874	29.9	3,653	25.5
Not contacted	259	5.5	2,251	23.4	2,510	17.6
Total	4,683	100.0	9,615	100.0	14,298	100.0
Ineligible (*)	150	3.1	1,158	10.7	1,308	8.4

(*) Households not at the address listed in the municipal register (wrong addresses, deceased, moved).

The most common reason for non-participation was the unwillingness of the household (25.5 per cent; Table 3a). In 17.6 per cent of cases, the household could not be contacted by telephone or during the three visits paid by interviewers on different days and at different times.

The questionnaire

The questionnaire, which was based on that used in the previous survey, was subjected to the usual pretesting before the start of the general survey. In this survey, a new approach was experimented in the pretesting phase in order to reduce difficulties in comprehension and answering. The test survey, which involved about 100 households from around Italy, was conducted by pairs of interviewers. One conducted the interview proper, while the other noted difficulties on a special questionnaire.⁵ This exercise provided information that was helpful in reformulating certain questions more clearly.

The questionnaire used in the survey has a modular structure. It is composed of a general part addressing aspects relevant to all households and a series of annexes with questions relevant to specific subsets. In order to reduce the burden of answering, some sections were only administered to a random subset of the sample. In particular, households had to answer only one of the two sets of questions regarding working conditions and income expectations, depending on the year of birth (odd or even) of the head of household.⁶

⁴ The participation rate for the previous survey net of ineligible households (deaths, wrong addresses, change of residence) was identical (57 per cent).

⁵ The strategy adopted was broadly that described in L. Oksenberg, C. Cannel and G. Kalton, "New Strategies for Pretesting Survey Questions" in *Journal of Official Statistics*, vol. 7, no. 1, Statistics Sweden, Stockholm, 1991, pp. 349-365.

⁶ In addition to producing estimates based on a smaller sample, this approach makes it impossible to make joint use of the responses to the two sections. In this instance, the relationship between the two aspects involved was felt to be of little interest.

Interviews lasted an average of 56 minutes, compared with 64 minutes in the previous survey. However, there was considerable variability within the sample, which was positively correlated with income, wealth and number of household members.

The questionnaire for panel households reported also some information given by the same family in the previous interview, in order to help the interviewer to track and remedy, in presence of the household, inconsistencies in the responses as emerging from two different surveys. This approach, already used in the past, did not prove satisfying, because too complex. The only possibility to ensure a fully satisfying panel quality level for such a complex questionnaire, seems to be to include computer-assisted interviews (CAPI - *Computer-Assisted Personal Interviewing*). With this purpose, a test has been conducted in this survey with about 200 households (outside the sample), in order to verify the real possibilities offered by such an instrument, as well as to test its difficulties in the creation of the programs. The results obtained seem comforting in terms of data quality, albeit at the price of the more complex programming of the questionnaire.

D. Definition of the survey units

Household

The basic survey unit is the household, which is defined as “a group of individual linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes”. Are to be included in the household all persons that normally lived in the dwelling at 31 December 1995 who contributed at least part of their income to the household; this includes any members temporarily absent (e.g. on vacation, temporarily away for study, etc) and any non-relatives that lived stably in the household at 31 December 1995.

Head of household

The head of the household is defined at the survey stage as the person who says he/she has “the most responsibility for family finances”, which satisfies the need to determine the best informed person.

E. Contents

The survey contains information about:

- demographic characteristics of the household members (including education) and the parents of the head and spouse;
- employment status of the household members aged 15 or over (incl. employment opportunities and lifetime work experience);
- questions on working times (only for workers) and on future working perspectives for workers and job searchers (rotation section);

- incomes from payroll employment, self-employment, pensions, other transfers for all adult members of the household;
- payment instruments and forms of savings of the household;
- principal residence and other property of the household;
- consumption and other family expenditures;
- forms of insurance of the household (life insurances, private pensions and annuities, health insurances, accident insurances);
- information to be provided by the interviewer.

F. Quality of data

Quality of estimates

1. Non-participation

Households that could not be interviewed were replaced by others selected randomly in the same municipality. Obviously, this technique does not eliminate the risk of obtaining samples in which the less-cooperative segments of the population are underrepresented, thus generating biased estimates (*selection bias*). One indication of the extent of the phenomena is provided by the number of contacts needed to obtain an interview. In order to conduct the 8,135 interviews, interviewers made a total of 14,019 contact attempts, including 11,923 personal visits and 2,096 telephone calls (the latter were made solely to fix an appointment).⁷ The difficulty of obtaining an interview increased with income, wealth and the educational qualification of the head of household. It was less difficult to obtain interviews in smaller municipalities, with households of small size and where the head of household was retired or female. In spite of the fact that the post stratification on the basis of certain individual characteristics of the interviewees makes it possible to take into account some of these factors, it cannot be excluded that the various segments of the population are not mis-represented within the sample; studies of the data from the 1989 survey have nevertheless suggested that the bias of the estimates due to non-participation is small, thanks in part to the measures taken.⁸

2. Response reliability

An additional aspect that can influence the quality of estimates is the reluctance of households to report their sources of income or the real or financial assets they hold.⁹ Although participation in the survey is voluntary and the content of the survey is known to the interviewee before the start, it is possible that respondents are not entirely truthful in their responses to the more “sensitive” questions, such as those regarding income or wealth. In order to assess the extent of such phenomena, which by their very nature are difficult to investigate, interviewers were asked to express a summary evaluation of the

⁷ A total of 13,127 contact attempts were made for the 7,471 households that were not interviewed.

⁸ See L. Cannari and G. D’Alessio, *Mancate interviste e distorsione degli stimatori*, Temi di Discussione del Servizio Studi, no. 172, Banca d’Italia, Rome, June 1992. With reference to the 1989 survey, the authors estimate that household income was understated by 5 per cent owing to non-participation.

⁹ Moreover, it is not unreasonable to believe that certain sorts of liability might be deliberately understated by interviewees.

presumed reliability of the responses immediately following the interview, basing their judgement on the correspondence between the information provided and objective evidence available to them (zone and type of dwelling occupied by the household, standard of living implied by quality of furnishings, etc.).¹⁰ Although the reliability level was satisfactory on average, it was not homogeneous across the sample. The highest ratings were given to households with heads who were young, had a high educational qualification, were employees and resided in the North. Slightly lower ratings were given to households with heads who were elderly, had a low educational qualification, were self-employed or retired and resided in the South or the Islands. Reliability increased as the income and wealth reported in the survey increased.

Checking data and imputing missing data

The questionnaires are checked first of all to verify that no annexes are missing, no questions have been skipped and that there are no editing errors. In this phase the codification of open-answer fields is carried out (i.e. the answer “other – please specify”). Data are subsequently entered into magnetic support and automatic checks are carried out to verify the consistency of single fields or correlated fields.¹¹

Once the checks were completed, work began on imputing missing answers, which could have been due to reticence on the part of the interviewee or difficulties that respondents had in replying to the question.¹² It is necessary to impute answers for all the elementary variables that make up the aggregate, since the absence of even one component would prevent calculation of the aggregate (for example, it is necessary to impute fringe benefits such as lunch coupons in order to calculate income from salaried employment).

The amount of imputed data is generally small, on the order of a few dozen cases for most variables. For more complex questions that require the respondent to estimate amounts, such as fringe benefits for salaried workers, depreciation for the self-employed, the value of dwellings or business equity, imputed rents, other property and furnishings, between 200 and 300 cases must be imputed.

Regression models are used to estimate the values to assign to the missing answers on the basis of other available information that is correlated with the missing data. In order to avoid an excessive concentration around average values, a random component is added, extracted from a normal variable, with a mean of zero and a variance equal to that of the residuals in the regression model. This preserves the mean and variance of the data actually measured.

Weighting: the sample estimates

¹⁰ The interviewers' evaluations were expressed on a scale from 1 (completely unreliable) to 10 (completely reliable).

¹¹ In presence of inconsistencies or anomalies, households were contacted telephonically to verify those information and eventually correct them.

¹² Nevertheless, while not answering was possible for some questions, the failure to indicate sources of income or the most significant components of wealth resulted in the invalidation of the interview.

The estimation procedure, which is similar to that used in the last survey, consists of three stages:

a) Calculation of the sampling weights for households

Each member of the household is assigned an initial weight defined as the inverse of his/her probability of inclusion in the sample. Given the sample design, the coefficient is constant at the municipality level and is equal to:

$$(1) w_{hi} = \begin{cases} \frac{P_h P_{hi}}{\tilde{P}_h n_{hi}} \\ \frac{1}{m_h} \frac{P_h}{n_{hi}} \end{cases}$$

respectively for municipalities with more than 40,000 inhabitants and for municipalities with up to 40,000 inhabitants, where P_h , \tilde{P}_h and m_h are respectively the resident population, that of the municipalities in the survey and the number of sample municipalities in the h^{th} stratum, and P_{hi} and n_{hi} are respectively the population and the number of respondents in the i^{th} municipality of the h^{th} stratum.¹³

b) Post-stratification of the panel households

The socio-demographic characteristics of the panel households may differ somewhat from those of the entire sample in 1993, mainly owing to missing interviews. In order to correct for this possible source of distortion in the estimates, the panel section of the sample is post-stratified on the basis of a number of characteristics of the previous survey (geographical area, income classes, professional status of head of household) so as to modify the initial weight of this subset of households.

c) Estimation of aggregates

An unbiased estimator of the mean of variable x is given by the Horwitz-Thompson estimator:

$$(2) \quad \bar{x} = \frac{\sum x_j w_j}{\sum w_j} \quad j = 1, \dots, N$$

However, if the values of variable x measured on two successive waves are correlated, an optimal estimator of the mean is given by:¹⁴

¹³ The probability of a household being extracted in a selected municipality is approximately equal to n_{hi} / P_{hi} . For municipalities with more than 40,000 inhabitants, which are all included in the theoretical sample, we need to bear in mind that for organizational purposes it is not always possible to conduct interviews in all the municipalities in the stratum. The first term of equation (1) therefore allows us to take account of this circumstance. Municipalities with fewer than 40,000 inhabitants are selected with a probability proportional to their size (PPS). The selection probability of the i^{th} municipality in the h^{th} stratum is therefore equal to $m_h * P_{hi} / P_h$. The probability of a household being included in the sample can therefore be written as $m_h * n_{hi} / P_h$.

¹⁴ See L. Fabbri, "L'indagine campionaria", La Nuova Italia Scientifica, Rome, 1989.

$$(3) \quad \bar{x}_t^* = \mathbf{a}\bar{x}_t^q + (1-\mathbf{a})\bar{x}_t^p + (1-\mathbf{a})\mathbf{r}(\bar{x}_{t-1} - \bar{x}_{t-1}^p)$$

$$(4) \quad \text{with } \mathbf{a} = \frac{Q(1-r^2Q)}{1-r^2Q^2}$$

where \bar{x}_t and \bar{x}_{t-1} are respectively the means of variable x at time t and time $t-1$, \bar{x}_t^p and \bar{x}_t^q are the means of variable x at time t for the panel and non-panel parts of the sample respectively, \mathbf{r} is the correlation coefficient between \bar{x}_t and \bar{x}_{t-1} and Q is the share of non-panel households.

The estimator (3) is not a simple weighted average of the values measured at time t , since, in addition to the correlation coefficient, it refers to the values of x from the previous survey for the panel and the total sample. However, following the post-stratification described above, the main variables approximately satisfy:

$$(5) \quad \bar{x}_{t-1} = \bar{x}_{t-1}^p$$

and the last term of (3) disappears. In addition, given that the correlation coefficients for the main variables examined are between 0.6 and 0.7, giving \mathbf{r} the intermediate value $\tilde{\mathbf{r}} = 0.65$, it is possible to approximate the estimator (3) by way of:

$$(6) \quad \bar{x}_t^+ = \mathbf{a}\bar{x}_t^q + (1-\mathbf{a})\bar{x}_t^p$$

which is obtained as the mean of the data measured at time t , weighted with coefficients equal to:

$$(7) \quad w_{hij}^* = \begin{cases} w_{hij} \frac{1-\mathbf{a}}{1-Q} \\ w_{hij} \frac{\mathbf{a}}{Q} \end{cases} \quad \text{respectively for panel households and for non-panel}$$

households. This estimator differs from (2) since, being based on the positive correlation between the data gathered from the same households in successive surveys, it gives a higher relative weight to the panel segment of the sample than the share of panel interviews actually conducted (48.1 per cent compared with 44.8 per cent), with a corresponding reduction in the weight assigned to the non-panel households. Since this reweighing could change the structure of the sample, the final sample is modified to assume the same characteristics as the population with regard to sex, age group, geographic area and size of municipality of residence.

Standard errors

The standard errors cannot easily be determined with the usual analytical methods. The presence of stages b) and c) as described in the previous section render useless - except

with a large margin of inaccuracy - the equations for calculating standard errors of the means in a two-stage sampling with stratification of the first-stage units. For this reason, the standard errors were calculated using simulation methods that take account of the original design of the sample and subsequent adjustments. In particular, 100 bootstrap samples of equivalent size to the actual sample were replicated (drawing the units with replacement in both stages). The mean values of the main variables were obtained by performing the full estimation process.¹⁵ The variability of the estimators was approximated analysing the distribution of simulated mean values.

The standard errors of the means of the main variables are shown in Table 8a. The table reveals the limited variability of the means for the demographic variables, which is mainly attributable to the post-stratification carried out in stage c. As regards the main economic variables, it can be noted that the standard errors in the means for consumption and income are significantly smaller than the standard error for net wealth. The standard errors in estimates at the level of geographical area are naturally larger than those for the sample as a whole.

Table 8a
Standard errors in the estimation of the means for the main variables
(units, euros, percentages)

Variable	North		Centre		South and Islands		Total sample	
	Value	% of estimate	Value	% of estimate	Value	% of estimate	Value	% of estimate
Mean number of members	0.032	1.2	0.044	1.5	0.046	1.5	0.017	0.6
Mean age	0.40	0.7	0.59	1.1	0.45	0.8	0.19	0.4
Household income	817	1.7	1,202	2.5	797	2.4	505	1.2
Household consumption	543	1.5	800	2.2	526	2.0	328	1.0
Net wealth	11,343	4.4	13,866	4.9	7,020	4.3	6,199	2.7

G. Uses of the survey

Publications

The results are regularly published in the Supplements to the Statistical Bulletin of the Bank of Italy. A whole bibliography of the research carried out using data from the SHIW is available (in Italian) from the Bank of Italy web-site (<http://www.bancaditalia.it/statistiche/ibf/statistiche/ibf/pubblicazioni/altre/biblio.pdf>). A list of the Economic Research Papers of the Bank of Italy concerning the SHIW is reported here:

¹⁵ This has also allowed to evaluate the effect on variability of the post-stratification estimated adopted; it was observed that standard errors of incomes and consumption are about 20 per cent inferior than those relative to the estimators that take only the sampling plan (phase a) into account.

E. Battistin, R. Miniaci and G. Weber (2003), What do we learn from recall consumption data?, Bank of Italy, *Temi di Discussione*, N. 466.

Giovanni D'Alessio and Ivan Faiella (2002), Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth, Bank of Italy, *Temi di Discussione*, N. 462.

Silvia Magri (2002), Italian households' debt: determinants of demand and supply, Bank of Italy, *Temi di Discussione*, N. 454.

Guido de Blasio and Sabrina Di Addario (2002), Labor market pooling: evidence from Italian industrial districts, Bank of Italy, *Temi di Discussione*, N. 453.

A. Brandolini, P. Cipollone and P. Sestito (2001), Earnings dispersion, low pay and household poverty in Italy, 1977-1998, Bank of Italy, *Temi di Discussione*, N. 427.

Andrea Brandolini and Piero Cipollone (2001), Multifactor Productivity and Labour Quality in Italy, 1981-2000, Bank of Italy, *Temi di Discussione*, N. 422.

Piero Cipollone (2001), Is the Italian Labour Market Segmented?, Bank of Italy, *Temi di Discussione*, N. 400.

G. D'Alessio e L. F. Signorini (2000), Disuguaglianza dei redditi individuali e ruolo della famiglia in Italia, Bank of Italy, *Temi di Discussione*, N. 390.

Andrea Brandolini (1999), The Distribution of Personal Income in Post-War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality, Bank of Italy, *Temi di Discussione*, N. 350.

Poverty and income distribution

According to the publication "I bilanci delle famiglie italiane nell'anno 1995", *Supplementi al Bollettino Statistico – Note metodologiche e informazioni statistiche*, Bank of Italy, Year VII, No. 14, March 1997, the Gini coefficient of concentration is 0.362 for the distribution of household income.