

## Finland 2000: Survey Information

### 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. Quality of data

### A. General Characteristics

#### Name

Income Distribution Survey (IDS).

#### Administrative unit responsible for survey:

Statistics Finland / Social Statistics / Economic Welfare  
Phone Number: +358 9 1734 3680  
Fax Number: (09) +358 9 1734 3264  
E-MAIL Address: [toimeentulo.tilastokeskus@stat.fi](mailto:toimeentulo.tilastokeskus@stat.fi)  
Internet: <http://www.stat.fi/tk/el/tulo/tulonj.html>

#### Purpose and use of the survey

The main purpose of the survey is to measure household and individual income and its distribution.

#### History

The survey was conducted for the first time in 1977 and it is conducted every year. Micro data were available for the first time in 1986.

The Income Distribution Survey is funded from Statistics Finland's budget, i.e. government funding.

#### Main users of the survey

Ministry of Finance, Ministry of Social Affairs and Health, Government Institute for Economic Research (VATT), National Research and Development Center for Welfare and Health (STAKES), Ministry of Environment, other research institutes and universities (Åbo Akademi), labor market organizations (AKAVA, the Confederation of Unions for Academic Professionals in Finland), media. The Finnish IDS data is also in

the LIS (Luxembourg Income Study) data bank and it is used also in the EUROMOD, a European micro simulation model.

## **B. Population, sample size and sampling methods**

The sampling frame is the Central Population Register; therefore the sampling frame includes the total population of individuals. The Central Population Register includes population living in institutions. But people living in institutions were excluded before sampling (i.e. handled as over-coverage). The sampling frame includes members of the military living in military housing and military living with their families.

Once the sample was selected from the sampling frame, all units were eligible for interview. However, people living in institutions were excluded in the sample selection process.

### Sample design

Sample design is two-phase stratified PPS design. A rotating panel design is used so that each household is in the sample for two consecutive years.

First a large master sample of about 30,000 individuals is taken from Central Population Register. Dwelling units are then constructed by adding all other people living in the same address with the sampled person. At this stage the inclusion probability of a household depends on the number of persons over 15 years of age in the dwelling unit (PPS sampling).

This master sample is then merged with the most recent tax register and a preliminary socio-economic class (13 classes) is derived for each individual. Socio-economic class of a dwelling unit is that of the preliminary reference person (person with highest taxable income). The final sample is then drawn using stratification according to socio-economic class of the reference person so that each household is included in the sample via this reference person.

### Probability of selection

Sampling rates differ a lot from one stratum to another. Farmers, entrepreneurs and high-income wage-earner households have the largest sampling fractions.

### Sample size

Individuals: 5 105 187 (population excluding people in institutions)

Number of sampling units in the sampling frame, which was interviewed or for which sufficient data was obtained: N = 10 423 (unweighted) and 2 373 000 (weighted).

The sampling frame includes all geographic areas in the country surveyed.

### Sampling variability

Sampling variability is variation that occurs by chance because a sample was surveyed rather than the entire population. Standard errors are common measures of sampling variability although they also include some of the effect of non-sampling error.

Estimates of sampling variability, that is, standard errors (of mean) have been computed using the SUDAAN computer program and were published in the annual publication "Income Distribution Statistics 2000".

### Weighting procedure

Weights are just to adjust for non-response and to provide estimates of population totals.

Basic weights are the inverse of the inclusion probabilities. Preliminary non-response correction is then conducted to these weights using sample information (response rates in each of the 13 strata). These non-response weights are calibrated with SAS-macro CALMAR to improve estimates of the population structure and the income totals and to reduce non-response bias.

In 2000 weights were calibrated to match the following marginal distributions. Population structure: number of persons by sex and age in five year bands, regional distribution of dwelling units (13 classes), size distribution of dwelling units (8 classes); income totals: taxable income and wealth, entrepreneurial income, farm income, forestry income, property income. Range of weights was not constrained. Population marginal distributions for population structure were taken from the master sample, income totals were taken from tax totals.

The sum of the survey weights is equal to the total number of units in the sampling frame.

## **C. Data collection and acquisition**

### Method of data collection

The Income Distribution Survey (IDS) is a combination of interviews and administrative data. The dates below refer to the interviews (data collection for the survey 2000 started in January 2001 and was completed in May 2001); most of the register sources have been merged to the database in 2001.

For the 2000 survey the interviews were carried out as follows:

IDS 2000	New panel		Old panel	
	Households	%	Households	%
January 2001	7	0.1	2492	55.2
February 2001	1316	25.9	1836	40.7

March 2001	2253	44.4	185	4.1
April 2001	1382	27.2	-	-
May 2001	119	2.3	-	-
<b>Total</b>	<b>5077</b>	<b>100</b>	<b>4513</b>	<b>100</b>

First an advance letter (with some results from the previous year to the II panel) was sent to all respondents, followed by a call a few days later by the interviewer. A combination of CAPI (computer assisted personal interview) and CATI (telephone) interviews is used. In 2000, 43.7% of households in the new panel were CAPI-interviewed and 56.3% CATI-interviewed. In the old panel, corresponding figures were 3.3% and 96.7%.

The questionnaire includes questions on both household and individual level and is divided in three main sections:

- first main section on main activities during a year and some employment related data for each person in the household;
- second main section (household level) concerns agriculture and forestry (for self-employed persons in agriculture);
- third main section (household level) concerns housing.

Participation was voluntary and privacy was assured.

## **D. Definition of survey units**

### Unit of observation

Survey unit is household defined as an economic unit.

### Household

A household consists of persons who live together and have their meals together or otherwise use their income together (income sharing) in private households. Persons living permanently in institutions (hospitals, prisons and so forth) are excluded.

People living in other location belong to the household if they participate in the acquisition of income. People conducting military service or equivalent belong to the household. Students living on their own or in other location do not belong to the household if they live mostly on their own income or on a student loan. Students (living together) in dormitories etc. each form their own household if they are not married or are not officially cohabiting.

Household as defined above is the basic unit of aggregation, but data on individuals is available as well.

### Household head

The head of the household was determined after the data collection according to several criteria such as age, activity and income.

Individual other than the household head can be distinguished. Their relationship to the sampled person (target person) is defined in the interview and to the head of the household after the data collection.

### Children

Children are all persons under the age of 18 and whose relationship to the sampled person is not spouse, mother or father, grandfather of grandmother, daughter-in-law or brother-in-law and he/she is not head of the household.

Since data includes information on individuals, it is possible to define children (e.g. age limits) differently.

It is possible to distinguish between children of the unit head and children of some other household member by using the data on relationship to the head and to the sampled person (see below).

### Spouses

Spouses are generally defined, as the person with the lowest income among cohabiting couples.

This is done using information on persons relationship to sampled person (asked in the interview) and to the head of the household.

### Person's relationship to the head of the household:

1) Head of household 2) Spouse 3) Child of head or spouse 4) parent of head or spouse 5) other relative (sister, brother, grandchild, grandparent) 6) other person.

### Person's relationship to the sampled person:

1) Sampled person 2) Spouse 3) Child 4) grandchild 5) father or mother 6) grandfather or -mother 7) sister or brother 8) brother-in-law or daughter-in-law 9) other

## **E. Quality of the data**

### Amount of data collected by proxy

Some other household member could provide answers, i.e. proxies were allowed. It was requested that respondent should be the person who knows best the economic situation of the household.

In about half of the sample the number of persons present in the interview does not equal the number of persons in the household, i.e. someone else has provided information on these members. However, practically all the income data are gathered from administrative files and registers and interview data are extensively checked and edited if necessary using register information. Therefore proxy answers relate principally to main activity during the year.

The minor income items, which are collected by interview may to some extent, be reported by proxy. Items that are asked at person level are: income from abroad, some daily allowances not subject to taxation, inter-household transfers, some grants and student allowances, and some insurance compensations not subject to taxation. The amount of income data collected by proxy is a very small proportion of the total amount.

Respondents were asked to consult or use pay records, tax returns, etc. to help provide the most accurate income information.

#### Non response

Non-response is higher than average for one-person households and in urban areas. Most importantly, non-response is known to be higher at the tails of the income distribution than in the middle.

#### Overall non response

The survey design is two-year rotating panel design. Therefore the response and non-response rates must be calculated separately. All figures are unweighted:

	Gross sample	Over-coverage	Net sample	Non-response	Final data	Response rate
New panel	6432	97	5077	1258	3819	75.2
Old panel	4777	48	4513	216	4297	95.2
<b>Total</b>	<b>11209</b>	<b>145</b>	<b>9590</b>	<b>1474</b>	<b>8116</b>	<b>84.6</b>

Response rate is calculated as follows:

1. Gross sample - Over coverage = Net sample
2. Net sample - Non response = Final data
3.  $100 \times \text{Final data} / \text{Net sample} = \text{Response rate}$

Weighted non-response rates for 2000 and their structure:

	Non-response rate	Structure of non-response		
		Refusals	Not reached	Other
New panel	19.9%	86.2%	13.2%	0.6%
Old panel	4.6%	64.8%	35.2%	0.0%

### Item non-response

There was no item non-response in income items derived from administrative records. Item non-response is a problem in some items that are still collected through interviews (especially interest income).

### Other quality measures

Internal checks for consistency errors or for data items have been made. In addition to that some adjustments were made as well. Logical checks and edits are done at data processing stage by comparing interview data with register data. If inconsistencies are found, interview data may be replaced or edited with register data.

### Imputations

Hot deck\*\* imputing for the interest income. Socio-economic status of the household and size of the household were used to determine the imputation classes.

\*\* (For example, missing values in one case may be replaced with the values of another case, which most closely corresponds for given indicators to the case which contains the missing values ("hot deck"))

Some values were replaced by consulting alternative data sources that could be matched on a case-by-case basis.

Some income items are estimated using regression models (e.g. home production) or other methods (net imputed rent using stratification method).

### Bottom coding

Bottom coding applies only to inter-household transfers (lower limit of 500 FIM). To reduce number of typing errors etc. top- and bottom codes are used in our Blaise-questionnaire, but interviewers can pass these.

## Other

Estimate of interest income was only about 25 % of known total amount (National Accounts). This is due to item non-response since interest income is one of the few items that still has to be collected through interviews. Other income items based on interviews (e.g. income from abroad, inter-household transfers) suffer from item non-response too, although their quality is much better than that of interest income.