

## European Union - Statistics on Income and Living Conditions (EU-SILC)

### TYPE

Sampling survey

### Sample design

#### Type of sample design

The two-stage area sampling was applied for the EU-SILC survey.

#### Sampling units

The sample of private households was selected in two stages. The primary units are the areas (one or more unified building blocks) and the ultimate sampling units selected in each sampling area are the households.

#### Stratification and sub-stratification criteria

There are two levels of area stratification in the sampling design. The first level is geographical stratification based on the partition of the total country area into thirteen standard regions corresponding to the European NUTS II level. The two major city agglomerations of Greater Athens and Greater Thessalonica constitute separate major geographical strata.

The second level of stratification entails grouping municipalities and communes within each NUTS II region by degree of urbanization, i.e., according to their population size. The scaling of urbanization was finally designed in four groups:

- $\geq 30.000$  inhabitants
- 5.000 - 29.999 inhabitants
- 1.000 - 4.999 inhabitants
- 0 - 999 inhabitants

The number of final strata in thirteen (13) administrative regions was 50. The Greater Athens Area was divided into 31 strata of about equal size (equal number of households) on the basis of the lists of city blocks of the Municipalities that constitute it and taking into consideration socio-economic criteria. Similarly, the Greater Thessaloniki Area was divided into 9 equally sized strata. The two Major City Agglomerations account for about 40% of total population and for even larger percentages in certain socio-economic variables. Thus, the total number of strata of the survey was 90.

#### Sample size and allocation criteria

The initial sample size is 8.000 households (the sampling fraction is about 2‰). This fraction was the same in each administrative region.

The regions (NUTS II) in Greece are 13 in number. However, throughout this study the 2<sup>nd</sup> administrative region (Central Macedonia) was considered without Greater Thessaloniki and the 9<sup>th</sup> administrative region (Attica) without the Greater Athens area, while either of these two major agglomerations was treated as a geographical region.

#### Sample selection schemes

##### 1<sup>st</sup> stage of sampling

In this stage, from any ultimate stratum (crossing of Region with the degree of urbanization), say stratum  $h$ ,  $n_h$  primary units were drawn, where the number  $n_h$  of draws was approximately proportional to the population size  $X_h$  of the stratum (number of households according to the last population census of the year 2001). Each area unit (primary unit) of the stratum had a selection probability proportional to its size. So, if  $X_{hi}$  was the number of households according to the 2001 population census of the unit in the sample of order  $i$ , then the probability of being drawn was:

$$P_{hi} = \frac{X_{hi}}{X_h}$$

The total number of the primary sampling units is 1.056 areas.

As in each year the 25% of the sample households is replaced, the new households belong to different primary sampling units.

## 2<sup>nd</sup> stage of sampling

In this stage from each primary sampling unit (selected area) the sample of ultimate units (households) is selected. Actually, in the second stage we drew a sample of dwellings. However, in most cases, there is one to one relation between household and dwelling. If the selected dwelling constitutes of one or more households then all of them are interviewed.

Let  $M_{hi}$  be the number of households during the survey period in the  $i$  selected area of the stratum  $h$ . Out of them a systematic sample of  $m_{hi}$  households is selected with equal probabilities. Each of  $m_{hi}$  households has the same chance to be included

in the survey, equal to:  $\frac{m_{hi}}{M_{hi}}$

In any selected primary unit, the determination of the sample size  $m_{hi}$  remains. The total number of households to be interviewed of the  $n_h$  selected primary sampling units

will be  $m_h = \sum_{i=1}^{n_h} m_{hi}$

i.e. finally by applying the two stage sampling procedure, from the stratum is drawn the percentage of households  $\frac{m_h}{M_h}$ .

In repeated sampling, the numerator of this fraction will vary from sample to sample, to be more specific the fraction  $\frac{m_h}{M_h}$  will be a random variable. Within each primary

sampling unit the calculation of the sampling interval  $\delta_{hi} = \frac{M_{hi}}{m_{hi}}$  will be carried out, so that the following two desired conditions to be satisfied.

The expected result  $\frac{m_h}{M_h}$  should be the predetermined over sampling fraction  $\frac{1}{\lambda}$  in

each region (NUTS II):  $E\left(\frac{m_h}{M_h}\right) = \frac{1}{\lambda} = 2\%$

The estimator of the stratum total  $Y_h$  (for any characteristic) should be self-weighting. In other words, the estimation calculated is the result derived from the sum of the values of the characteristic over the  $m_h$  sample households by the overall raising factor  $\lambda$ , which is the same in each administrative region. The conditions above conditions are satisfied when:

$$\frac{1}{n_h} \cdot \frac{1}{P_{hi}} \cdot \frac{M_{hi}}{m_{hi}} = \lambda \Rightarrow$$

$$\frac{1}{n_h} \cdot \frac{1}{P_{hi}} \cdot \delta_{hi} = \lambda \Rightarrow$$

$$\delta_{hi} = \frac{M_{hi}}{m_{hi}} = \lambda \cdot n_h \cdot P_{hi}$$

#### Renewal of the sample: rotational groups

The survey is a *simple rotational design* survey (which means once the system is fully established). The sample for any year consists of 4 replications, which have been in the survey for 1-4 years. With the exception of the first three years of survey, any particular replication remains in the survey for 4 years, each year one of the 4 replications from the previous year is dropped and a new one is added. Between year T and T+1 the sample overlap is 75%; the overlap between year T and year T+2 is 50%; and it is reduced to 25% from year T to year T+3, and to zero for longer intervals.

#### Estimation of the survey characteristics

##### Weightings - Design factor

In order to estimate the survey characteristics, the statistical information concerning each person was multiplied with an extrapolation factor. The extrapolation factor is the product of the following three elements:

(a) The household design weight is defined as the inverse of its probability of selection.

$$\frac{1}{n_h} \cdot \frac{1}{P_{hi}} \cdot \frac{M_{hi}}{m_{hi}} = DW_{hi}$$

$M_{hi}$  = the number of households in the updated sampling frame in **hi** area (primary unit).

$m_{hi}$  = the number of selected households in **hi** area (primary unit).

$n_h$  = the sample size of primary units in **h** stratum.

$P_{hi}$  = the selection probability of **hi** primary unit.

(b) The reverse of the response rate of households in the primary unit.

(c) A factor defined in a way that the following statements stand:

- i. In each geographical area, the estimated population by sex and age conforms to the population calculated by projecting data of the reference period coming from vital statistics (population census, births, deaths, migration).
- ii. The estimated households by size and state of ownership conform to the number of households resulting from projection of the trend observed between the population censuses of 1991 and 2001.