18.1.2 Sampling design

The sampling design is a probability design (multi-stage stratified sampling).

Number of individuals interviewed in the selected household: One individual
Number of sampling stages: Two stages for households and three stages for individuals.

The primary sampling units are areas (one or more unified city blocks) participating in the EU-SILC of the years 2016 – 2022.

1) First stage – Stratification

The stratification variables of the first stage units (one or more unified city blocks) are:

1. Region (NUTS 2)
2. Degree of urbanization.

In each Region (NUTS 2), the stratification of primary units (first stage units) was conducted by allocating the Communes according to the degree of urbanization. Except for the former two Major City Agglomerations (Athens and Thessaloniki), the strata according to the degree of urbanization are:

1. Communes with 30.000 inhabitants or more
2. Communes with 5.000 to 29999 inhabitants
3. Communes with 1.000 to 4.999 inhabitants
4. Communes up to 999 inhabitants

The former Greater Athens Area was divided into 31 strata of about equal size (equal number of households), on the basis of lists of city blocks of the Communes that constitute it and taking into consideration socio-economic criteria. Similarly, the former Greater Thessaloniki Area was divided into 9 equally sized strata. Thus, the total number of strata of the survey was 90, that is the 'final strata'.

The sampling design is a probability design (multi-stage stratified sampling)

Number of sampling stages: Two stages for households and three stages for individuals.

The primary sampling units are areas (one or more unified city blocks) participating in the EU-SILC of the years 2016 – 2022.

1st stage of sampling

In this stage, for any ultimate stratum ('final strata'), say stratum h, a_h primary units were drawn with probabilities proportional to their sizes. The number a_h of draws is approximately proportional to the population stratum size N_h, as defined above.

The primary unit of order i in stratum h has probability of being drawn proportional to the population size as follows:

$$P_{hi}=\frac{a_h \cdot N_{hi}}{N_h} \quad (1)$$

where:

N_{hi} : Updated number of households from the EU-SILC in the hi primary unit.

N_h : Estimated number of households for the year 2021 in the stratum h .

The secondary sampling units are the households of the EU-SILC containing members belonging to the target population (individuals aged 16-74 years old). The third –and final- sampling unit is one person randomly selected among the household members of 16-74 years old.

In more detail:

2nd stage of sampling

In the hi primary unit, a sample of n_{hi} out of N_{hi} households was selected with equal probabilities. Each one of the n_{hi} households had the same chance to be selected, equal to:

$$\frac{n_{hi}}{N_{hi}} \quad (2)$$

The total number of households to be interviewed of the a_h sampling primary units is:

$$n_h = \sum_{i=1}^{a_h} n_{hi}$$

Within each primary sampling unit the calculation of the sampling interval $\delta_{hi} = \frac{N_{hi}}{n_{hi}}$ was carried out, so that the following two desired conditions to be satisfied.

a) The expectation of the fraction $\frac{n_h}{N_h}$ was constant in each stratum. That is:

$E\left(\frac{N_h}{n_h}\right) = \frac{1}{\lambda} = 2.96\ ‰$ (3) and

b) The estimator of the stratum total Y_h (for any characteristic) will be self-weighting. In other words, the estimate of the survey characteristics is derived as product of the sum of the values of the characteristics over the n_h sample households by the overall raising factor λ , which is equal in each stratum.

The conditions (a) and (b) are satisfied when:

$$\frac{1}{a_h} \cdot \frac{1}{P_{hi}} \cdot \frac{N_{hi}}{n_{hi}} = \lambda \quad (4) \Rightarrow \frac{1}{a_h} \cdot \frac{1}{P_{hi}} \cdot \delta_{hi} = \lambda \Rightarrow \delta_{hi} = \frac{N_{hi}}{n_{hi}} = \lambda \cdot a_h \cdot P_{hi} \quad (5)$$

From the relations (1) and (5)$\Rightarrow$

$$\frac{N_{hi}}{n_{hi}} = \lambda \cdot a_h \cdot \frac{N_{hi}}{N_h} \Rightarrow n_{hi} = \frac{N_{hi} \cdot N_h}{\lambda \cdot a_h \cdot N_{hi}} \Rightarrow n_{hi} = \frac{N_h}{\lambda \cdot a_h} \quad (6)$$

From the relation (3), it is deducted that: $\frac{1}{\lambda} = \frac{n_h}{N_h} \Rightarrow \lambda = \frac{N_h}{n_h}$ (7)

From the relations (6) and (7), we have: $n_{hi} = \frac{n_h}{a_h}$ (8)

3rd stage of sampling

In this stage from each household one individual (member of household belonging to the target population) was selected with equal probabilities.

Let p_{hij} is the selection probability of the hij individual, which belongs to the hi household.

As one individual was selected with equal probabilities out of m_{hi} members belonging to target population, the p_{hij} was defined as: $p_{hij} = \frac{1}{m_{hi}}$

The sampling fraction in each of the 90 strata (Stratum= Region x Degree of urbanization) is $f = \frac{1}{\lambda} = \frac{n}{N} \cong 0.0029$ where n =10,992 is the total sample size of households and N=3,711,296 is the estimated total number of households belonging to the target population.

The number of the sampling households in each of the 90 strata (let h) was defined by applying the proportional allocation as follows:

$$n_h = n \cdot \frac{N_h}{N}$$

where:

N_h : the population size of the stratum h .