

Survey on Qualitative Characteristics of Residents Tourists is a sampling survey on households conducted yearly on a sample of approximately 9,500 households (consisting of approximately 20,000 household members). For each household a questionnaire is filled in pertaining to all the trips made by each household member during the reference year. For the reference year 2021, data from 6.984 households have been collected (size of the initial sample: 9,044 households).

The surveyed households are derived from the sample of the Labor Force Survey which is selected using a two-stage stratified sampling method. The primary unit of the survey is the surface area (one or more subsequent city blocks) and the secondary unit is the household.

### Stratification criteria

There are two levels of area stratification in the sampling design. The first level is the geographical stratification based on the partition of the total country area into thirteen (13) Regions corresponding to the European NUTS 2 level. The two former major city agglomerations of Greater Athens and Greater Thessaloniki constitute separate major geographical strata.

The second level of stratification entails grouping of municipal and local communes within each NUTS 2 Region by the degree of urbanization, i.e., according to their population size and to the following:

- Urban: Municipal communes with 10,000 inhabitants or more
- Semi-urban: Municipal and Local communes with 2,000 to 9,999 inhabitants
- Rural: Local communes up to 1,999 inhabitants

### Number of sampling stages

In the first sampling stage, from each created strata, according to the above criteria, the primary units (surfaces) are selected with probability of selection proportional to their size (number of households, according to the Census year 2011).

During the 2nd stage of sampling, in each selected primary sampling unit (area unit), the sample of households is selected with equal probability and by applying systematic sampling. In fact, in the second stage it is selected a sample of houses. However, in most cases, there is a match between household and residence. If the selected residence consists of more than one household, then all are surveyed.

### Estimation of the survey characteristics

A two-stage sampling scheme was applied, according to which in the final strata the areas (one or more unified blocks) were selected with probabilities proportional to their sizes and within the selected areas the households were selected with equal probabilities. Then the inclusion probabilities of households are defined, as follows:

$$p_{hij} = P_{hi} \cdot \frac{m_{hi}}{M_{hi}} \quad (1)$$

where:

$p_{hij}$ : Inclusion probability the sample ultimate unit (household) of order  $j$  ( $j = 1, \dots, m_{hi}$ ), that belongs to the selected area of order  $i$ , in stratum  $h$ .

$M_{hi}$ : Number of households in the  $hi$  area in the updated sampling frame (updated list of households)

$m_{hi}$ : Initial sample size of households in the  $hi$  area that were selected from the  $M_{hi}$  units (households)

Let  $dw_{hij}$  stand for the survey design weight attached to the sample ultimate unit (household) of order  $j$  ( $j = 1, \dots, m_{hi}$ ), that belongs to the selected area of order  $i$ , in stratum  $h$ . Then:

$$dw_{hij} = p_{hij}^{-1}$$

Non-response adjustment of the responding households was carried out within homogeneous classes that were created by unifying strata that belong the same Region (NUTS 2). In Region Attika, the classes were created by unifying strata that belong to the same Small Region (NUTS 3). More specifically, in each class the design weights of households were multiplied by the inverse of the response rate of the eligible households.

After the nonresponse adjustment, weight adjustment to external data sources is conducted. This involves the calibration of the household weights in conjunction with external sources. It enables the distribution of auxiliary variables at both household and individual level to coincide with the corresponding population distribution of the external data. The auxiliary variables used at household level are the household sizes of the total population (1, 2, 3, 4+ members) and at individual level the gender and age groups (0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+).

By applying calibration: a) the estimated population by gender and age groups conforms to the population projections for the reference period and b) the estimated households by size conform to the number of households of the reference period that results from projection that considers the population projection and the average number of households members from the Population Census 2011.