2018 Qualitative Characteristics of Resident Tourists Survey

Type of Survey: Sample survey

Sampling Frame

The **Qualitative Characteristics of Resident Tourists Survey** is an annual sample survey conducted on households. The sample of households for the 2018 survey derives from the sample of the Labour Force Survey which is a two-stage stratified sample survey with primary sampling unit the surface (one or more building blocks or a whole settlement) and final unit the household. The Labour Force Survey is a continuous survey producing quarterly estimations with a sample which follows a "rotational" pattern. Every quarter, the 5/6 of the sample remains the same, while the 1/6 is "rotated" and covers private households. The sample design of the Labour Force Survey was based on the data of the 2001 Population-Housing Census. The number of households and of persons per household is estimated by Region on the basis of the data from the 2011 Population-Housing Census.

The sample of the Labour Force Survey is the frame for the selection of the households of the sample of the 2018 Qualitative Characteristics of Resident Tourists Survey. The use of the sample of the LFS for the collection of the statistical data of the Qualitative Characteristics of Resident Tourists Survey facilitates the sample design and reduces the cost for the compilation of the lists.

Sampling Unit

The sampling unit is the private household and its members.

Sample Design

The sample of the 2018 Qualitative Characteristics of Resident Tourists Survey is a sub-sample of the LFS for the quarters Q4_2017-Q3_2018 and was selected among the survey households that were not surveyed anymore.

The Labour Force Survey is a sample survey covering the population living in private households. The selection of households to be surveyed follows a two-stage stratified sampling scheme with primary unit the surface (one or more building blocks) and secondary sampling unit the household.

This specific survey is a quarterly survey and the sample is renewed by its 1/6 every quarter. The reference population coincides with the reference population of the Qualitative Characteristics of Resident Tourists Survey and the survey design was based on the results of the 2001 Population-Housing Census and an adjustment was made on the basis of the data of the 2011 Population-Housing Census.

For the Qualitative Characteristics of Resident Tourists Survey, the primary units are the surface (one or more building blocks) and the secondary units are the households and all the persons living in them are surveyed.

Stratification criteria

The survey design follows a two-stage stratification scheme: a) geographical stratification which is based on the allocation of the Country into 13 Regions, while the agglomeration of the Capital and the agglomeration of Thessaloniki are two separate geographical strata, b) the second stage of stratification encompasses the grouping of the municipal and local community in each Regions and therefore of the surfaces belonging to them by degree of urbanization into 8 categories as follows:

Urban areas with 50,000 inhabitants and over
Urban areas with 30,000 – 49,999 inhabitants
Urban areas with 10,000 – 29,999 inhabitants
Urban areas with 5,000 – 9,999 inhabitants
Urban areas with 2,000 – 4,999 inhabitants
Rural areas with 1,000 – 1,999 inhabitants
Rural areas with 500 – 999 inhabitants
Rural areas with 1 – 499 inhabitants

The final number of strata in the 13 Regions amounts to 93, while the Regions of the Capital and of Thessaloniki were divided in 42 and 12 sub-strata, respectively, on the basis of the building blocks which consist the Municipalities. The final number of strata amounts to 147.

Size and criteria for the allocation of the sample

The size of the annual sample amounted to 8,705 households that belong to 1,118 primary sampling units.

1st stage of sampling

During the 1st stage of the sampling, for every (final) stratum h, a_h primary units were selected with a selection probability which was proportionate to its size. The number a_h of the selected primary units is approximated proportionally to the size of the stratum N_h in the population, as defined above.

The primary unit i in the stratum h has a probability to be selected which is proportionate to the population size as follows:

$$P_{hi} = \frac{N_{hi}}{N_h} \tag{1}$$

where:

N_{hi}: is the number of households in the primary unit i of stratum h

2nd stage of sampling

During the 2nd stage of sampling, in the primary unit hi a sample n_{hi} was selected from N_{hi} households with equal probabilities of selection. Each one of the n_{hi} households had an equal probability to be selected which was:

$$\frac{n_{hi}}{N_{hi}} \tag{2}$$

The total number of households for the a_h primary sampling units is: $n_h = \sum_{i=1}^{a_h} n_{hi}$

In each primary sampling unit the sampling fraction $\delta_{hi} = \frac{N_{hi}}{n_{hi}}$ was estimated in such a way so that the following two conditions are met:

(a) The expected value of the fraction $\frac{n_h}{N_h}$ remains constant in each stratum. That is:

$$E\left(\frac{n_h}{N_h}\right) = \frac{1}{\lambda} = 2.11‰ \quad (3)$$

and

(b) The estimator of the total at the level of the stratum Y_h (for each variable) is auto-weighted, i.e., the estimation of the survey variables can derive as the sum of the product of the variables' value in the selected households n_h multiplied by the coefficient λ, which is the same in each stratum.

The conditions (a) and (b) are met if:

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

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

On the basis of the formula (1) and (5) $\Rightarrow$

$$\frac{N_{hi}}{n_{hi}} = \lambda \cdot a_h \cdot \frac{N_{hi}}{N_h} \Rightarrow n_{hi} = \frac{N_{hi} \cdot N_h}{\lambda \cdot a_h \cdot N_{hi}}$$

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

On the basis of the formula (3), it is concluded that:

$$\frac{1}{\lambda} = \frac{n_h}{N_h} \Rightarrow \lambda = \frac{N_h}{n_h} \quad (7)$$

On the basis of the formula (6) and (7), it is concluded that:

$$n_{hi} = \frac{n_h}{a_h} \quad (8)$$

Estimation of the survey characteristics

Making the assumption that h is the final stratum of households (Final stratum = Region x Urbanity degree), then it will have the following values: $h = 1, 2, ..., H$ (where $H = 147$). In each of the final strata (let us say h), if the statistical information was selected from a sample of n'_h households, the weighting coefficient of the household j that belongs to the primary unit i is defined as follows:

$$w_{hij} = \frac{N_h}{a_h \cdot N_{hi}} \cdot \frac{N_{hi}}{n_{hi}} \cdot \frac{1}{r_h} \cdot t_{hij} = \frac{N_h}{a_h \cdot n_{hi}} \cdot \frac{1}{r_h} \cdot t_{hij} \quad (9)$$

On the basis of the formula (8) and (9), the following applies:

$$w_{hij} = \frac{N_h}{a_h \cdot \frac{n_h}{a_h}} \cdot \frac{1}{r_h} \cdot t_{hij} \Rightarrow w_{hij} = \frac{N_h}{n_h} \cdot \frac{1}{r_h} \cdot t_{hij} \quad (10)$$

where:

N_h: is the size of the population in the stratum h in accordance with the 2011 Population and Housing Census and the population projections

n_h: is the initial sample in the stratum h

$\frac{N_h}{n_h}$: is the reverse probability of selection of the selected sample of households in the stratum h, so as the estimator of the total Y_h, (for each variable) in the stratum is auto-weighted,

$r_h = \frac{n'_h}{n_h}$: is the reverse response rate in the stratum h

t_{hij}: is the adjustment factor of the weighing coefficients of households, so as their sample allocation is in line with the allocation of the population at the level of the Large Geographical Area (NUTS 1). The classes that have been used were the total of households by household size class (1, 2, 3, 4+ members) and the total of persons by gender and age group (0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+), on the

basis of the data from population projections and the data of the 2011 Population and Housing Census.

w_{hijk} (>0) is the weighting coefficient for the person k ($k = 1,...,m_{hij}$) that belongs to the surveyed household j ($j = 1,...,n_{hi}$) and which belongs to the primary sampling unit i of the stratum h.

y_{hijk} is the variable y of the persons k of the household j, which belongs to the primary sampling unit hi (one or more building blocks). Moreover, Y is the variable for the total population, which is the sum of the variables y for all the persons belonging to the strata h.

On the basis of the above, the formula of the estimator of the two-stage survey is as follows:

$$Y = \sum_{h=1}^{H} \sum_{i=1}^{a_h} \sum_{j=1}^{n_{hi}} \sum_{k=1}^{m_{hij}} w_{hijk} \cdot y_{hijk} \qquad (11)$$