

SURVEY ON FOREIGN TRAVELERS – METHODOLOGY AND IMPLEMENTATION

Rudi Seljak¹, Nataša Jokić²

¹Statistical Office of the Republic of Slovenia, e-mail: rudi.seljak@gov.si

²Statistical Office of the Republic of Slovenia, e-mail: natasa.jokic@gov.si

Abstract

In 2006 the Statistical Office of the Republic of Slovenia started to carry out the Survey on Foreign Travelers, which aimed at obtaining data on the structure of foreign travelers leaving Slovenia at road border crossings. The main emphasis was on foreign travelers who did not spend the night in Slovenia but came to Slovenia just for a same-day visit (same-day visitors) or they crossed Slovenia on their way to another country (transit travelers). The main target characteristics were the reasons for coming to Slovenia and data on their expenditure. Since the target population of the survey population was not known in advance and there was no fixed sampling frame available, the sample selection process and the estimation procedure were quite a challenge for our organization.

The aim of this paper is to describe the methodological solutions used as well as the concrete process of the execution of the survey. In the first part of the paper the sampling procedure is described, where the combination of stratified random sampling and quota sampling was used. The second part is devoted to the description of the estimation process with the emphasis on the usage of the external (auxiliary) data for the improvement of the final estimates. In the last part of the paper some concrete results and their coherence with the results of other surveys will be presented.

Keywords: flow sampling, estimation methodology, travelers statistics

1. Introduction

Data on tourism statistics are highly relevant not only for tourism policy making but also for other related fields of statistics such as the Balance of Payments. The importance of tourism in national economy is evaluated by Tourism Satellite Accounts (TSA).

Besides tourists¹, same-day visitors and transit travelers have an important role in tourism industry. In order to compile the data sources for Travel Balance of Payments on behalf of the Bank of Slovenia, the Statistical Office of the Republic of Slovenia (SORS) carried out the Survey on Foreign Travelers at Road Border Crossings. The main purpose of this survey was to obtain the data on the structure of foreign travelers leaving Slovenia at road border crossings in terms of their reasons for coming to Slovenia. The main emphasis was on foreign travelers who did not spend the night in Slovenia but came to Slovenia just for a same-day visit (same-day visitors) or they crossed Slovenia on their way to another country (transit travelers). Besides the structure of foreign travelers, the purpose of the survey was also to obtain the data on expenditure of foreign same-day visitors and transit travelers. The survey was conducted for the first time in summer 2006 only at the border with Croatia. The results of this survey were not published but were only used for calculating the Balance of Payments. In April

¹ Tourists are persons traveling to a place other than that of their usual environment, staying in a hotel or some other tourist accommodation in the place/country visited for at least one night (but not more than 365 nights consecutively) for leisure, business or other reasons, but not to make profit in the place visited (they receive no remuneration) (Applying Methodological Guidelines in Basic Tourism and Travel Statistics, 1996).

2007 and in the period between 15th July and 15th August 2007 the regular survey was conducted at all four borders (Italy, Austria, Croatia and Hungary).

The most challenging problem which had to be dealt with in the phase of survey planning was the question how to design an efficient and yet easily feasible sampling procedure. While in most of our sampling surveys the procedure follows the “classical” procedure of sampling frame construction and then the sample selection by one of the well known sampling designs, the situation here was quite different. Since the travelers crossing the border crossings are the classical example of flows, we had to deal with the problem of the “flow-sampling” and the direct consequence of that was the fact that there was no fixed sampling frame available. We could just use some auxiliary data to predict the size of the flow of the travelers in the different days at the different border crossings and in fact these auxiliary data played the crucial role in the sampling procedure. The whole sampling design was at the end the combination of multi-stage PPS sampling and quota sampling.

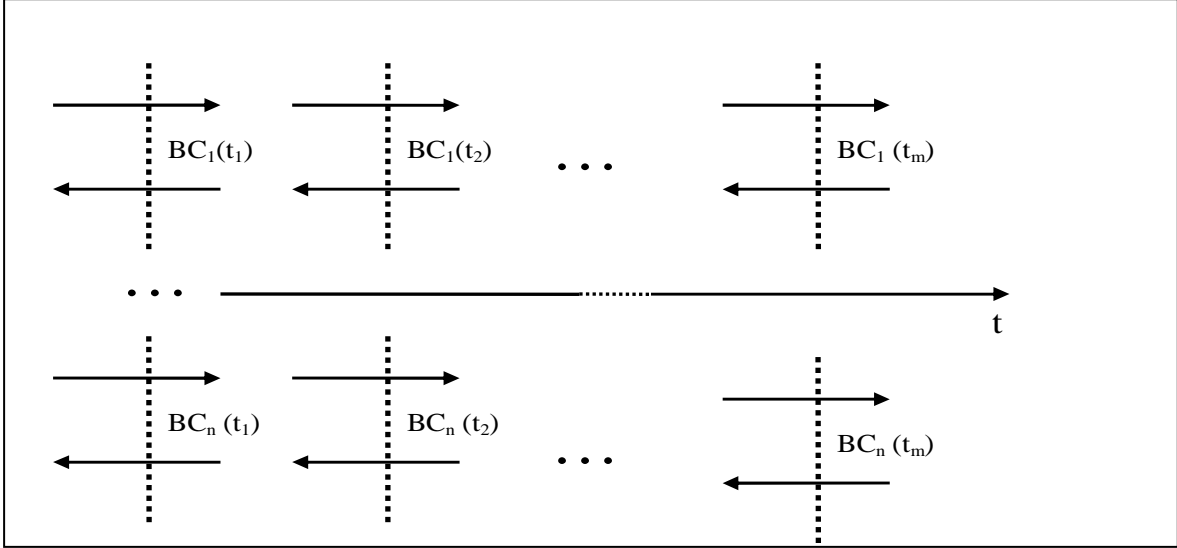
Since the target population was not represented through the fixed sampling frame, also the inferential procedure (mostly represented by the weights calculation) was quite challenging. Although the basis approach was still in the direction of the well-known Horvitz-Thompson estimator, some additional simplifications were needed.

In the first part of the paper the sampling design and the weighting procedure will be described. In the second part some main results will be presented and at the end some conclusions will be pointed out.

2. Sampling Design

The aim of the sampling procedure is to get the representative sample of the foreign travelers who have crossed one of the Slovenian border crossings within the reference month. Hence the sampling procedure has to deal with two dimensions: spatial (border crossings) and temporal (days, hours, etc.) dimension of the target population. We present the situation in the following chart where the abbreviation BC is used to designate the different border crossings and t represents the time dimension, assuming that we have n border crossings and m time-units in the target population.

Figure 1: Visual presentation of the target population



The arrows in both directions represent the incoming and outgoing flow of the travelers. In fact, only in the first execution of the survey both directions were sampled. Since the analyses of the results of

this first execution showed no significant difference according to the different direction, in all the subsequent executions only the outgoing flow was surveyed.

As already stated, our final target unit was the foreign traveler crossing one of our road border crossings. Since it is quite understandable that the list of vehicles or travelers was not available at the time of setting up the sampling procedure, the classical random sampling could not be used. Firstly, we had to define several stages of sampling, collect the auxiliary information for each of the stages and on the basis of the available information determine the final sampling design.

The most important auxiliary data used in the sampling procedure were the data on the number of vehicles crossing the border (in both directions) for every hour and every day of the previous year. These data were provided by the Slovenian Roads Agency (SRA) and obtained by the use of the automatic counters. Since it is quite obvious that the number of vehicles crossing the border is the subject of significant peaks through weekends and holidays, the counts from the previous years had to be correctly mapped to the current-year day. Hence, for each previous-year day d_i the count $(d_i + 366)$ was used as the estimated “flow size” for the current year. These data were in the sampling stage used to determine the size of the border crossing as well as for all the PPS procedures.

The sampling procedure could roughly be divided into two separate parts, where for the first part “classical” PPS random sampling and for the second part quota sampling was used. Since the list of the “real” observational units was not available at the time of sampling, we had to deal with the indirect units. The population of travelers could be conceptually considered as a two-dimensional flow $U(x, t)$, where x represents the spatial and t the temporal dimension of the flow. If we want to get at least approximately representative sample, we have to sample in both dimensions. The only possible sampling unit for the spatial dimension is clearly the border crossing; therefore m^2 border crossings for each of the directions were selected at the first stage. The population of all border crossings was firstly stratified according to the border state³ and according to the size. Since the border crossings are quite heterogeneous in the sense of the number of travelers, we took the “take all” approach, usually used in business surveys. The crossings that were put in the stratum of largest crossings were selected with certainty. From the “smaller crossings strata” the PPS random sample was selected.

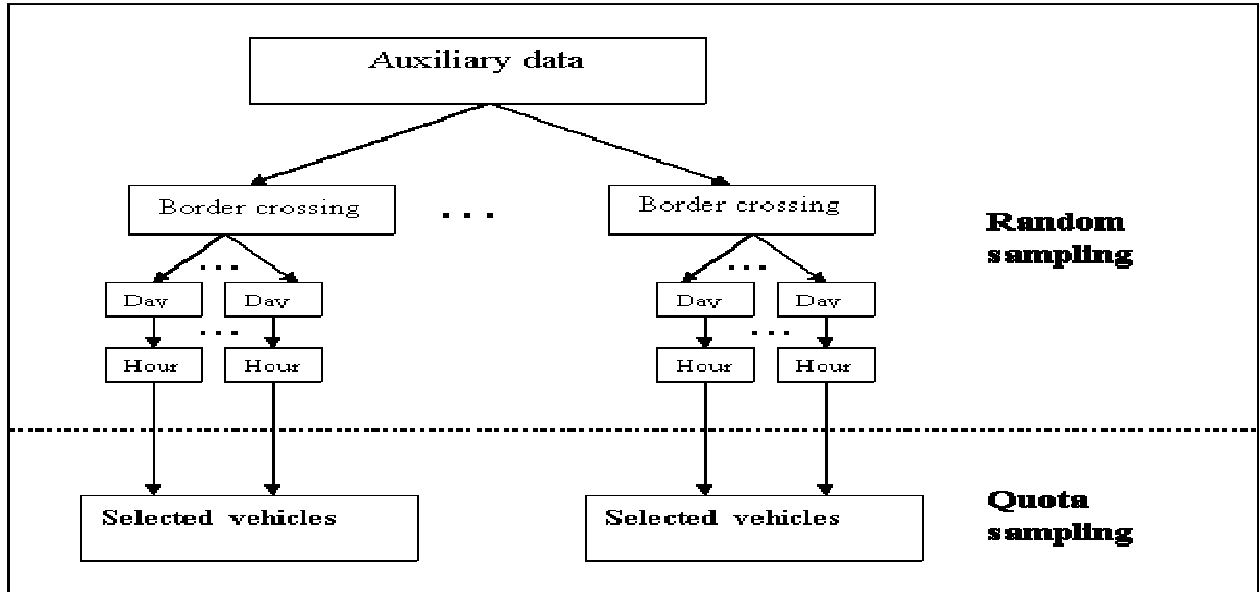
For the temporal dimension, the two stage sampling was used. At the first stage the sampling unit was the calendar day and we again used the PPS sampling to select the allocated number of days for each of the selected border crossings. As it was decided prior to the sampling selection, each selected day one interviewer should survey at the border crossing for eight hours. Therefore the second stage unit was the hour of the beginning of interviewers’ work and for each selected day the “starting hour” was randomly selected by using the PPS sampling again. So, the whole first part of the sampling procedure could be considered as a three stage sampling, where for each of the selected border crossings the survey days were selected and then as the third stage for each of the selected days the “starting hour” was selected.

In the second part of the sampling procedure the vehicles and then the travelers had to be selected. Since it was quite difficult to perform a correct random procedure in the situation like this, we decided to go for a quota sampling at this stage. Each interviewer was thus instructed to get at least 80 completed questionnaires for each of the 8-hours shifts. Later in the stage of weights construction, the quota sampling at the last stage was considered as the simple random sampling where all the vehicles had equal chance of being selected. We also present the system graphically.

² The number of selected border crossings slightly differs in different executions of the survey but it was generally between 20 and 25.

³ Slovenia has four border states: Austria, Croatia, Hungary and Italy.

Figure 2: Visual presentation of the sampling procedure



3. Weighting

Since all the data were collected on the basis of (mostly) randomly selected sample, all the data should be properly weighted in the estimation procedure. The weights calculation should as strictly as possible follow the selection procedure in order to assure that the weights and consequently the final results should properly reflect the effective sample as well as some exogenous information on the target population. We describe the concrete procedure through the following steps:

- Weight for the selected day.** In the first step of the procedure the sampling weight for each of the selected days was calculated. Following the selection procedure this weight should be calculated as the product of the two “Horvitz-Thompson” weights; the weight reflecting the selection probability of the border crossing and the weight reflecting the selection probability of the selected day for the particular border crossing. Hence, for each day d_j for the selected border crossing BC_i , the weight is calculated as: $w_{ij}^d = (1/\pi(BC_i)) \cdot (1/\pi(d_j))$, where $\pi(\cdot)$ denotes the selection probability.
- Weight for the selected traveler.** While in the first stage of the weights calculation only the information derived directly from the sampling procedure was used, for the second stage, which should reflect the quota sampling, some additional assumptions were needed. For the sake of simplicity, we assumed that the quota sample was in fact a simple random sample of all foreign travelers crossing a particular border on a particular day. So, the weight at this stage should be calculated as the ratio between the number of all the foreign travelers crossing the border and the selected foreign travelers. Since the number of selected travelers was of course known from the survey, the population count should somehow be estimated. Again here the “automatic counters data” were used but in this case already the data from the current year were available, hence the counts for the exact date of survey execution could be used. We then combined these data together with the estimations of the share of the foreign vehicles for each surveyed border crossing which we obtained from the Cross Border Survey. Since these estimations gave us only the number of foreign vehicles, we still needed to “expand” these counts to the number of foreign travelers. For this purpose we estimated the average number of travelers in different types of vehicles (cars, buses, vans, etc.) directly from the survey data and then used these averages to get the wanted estimates. The weight for the particular traveler

on the particular day j at the particular border crossing i was hence calculated by the formula:

$$w_{ij}^T = \frac{\sum_{k=1}^V N_{kij} \cdot T_{ki} \cdot f_{ki}}{n_{ij}}$$

where V is the number of different types of vehicles, N_{kij} number of counted vehicles of particular type, T_{ki} average number of travelers per vehicle, f_{ki} estimated share of foreign vehicles and n_{ij} number foreign travelers interviewed. The final sampling weight for the selected traveler interviewed on the day j at the border crossing i was then calculated as

$$w_{ij}^s = w_{ij}^d \cdot w_{ij}^T.$$

- **Calibration.** Since the first step of the sampling weights calculation was entirely based on the population counts from the previous year, also the final sampling weight partly reflects the population from the previous year. Therefore in the last step the weights were calibrated to the current year population counts. The calibration was performed on the basis of the estimated population counts of the foreign travelers for each week of the "surveyed month". Using the denotations from above, for each week w these estimates were obtained by using the formula

$$N_w = \sum_{j=1}^7 \sum_{i=1}^{BC_w} \sum_{k=1}^V N_{kij} \cdot T_{ki} \cdot f_{ki}.$$

If n_w is the number of the foreign travelers interviewed during

the particular week, the final weight is then calculated as:

$$w = \frac{N_w}{\sum_{i=1}^{n_w} w_i^s} \cdot w_i^s.$$

The above given description described the procedure for the calculation of the weights referring to the selected traveler. In fact, in the concrete procedure besides the weights for the travelers also the weights for the selected vehicles were calculated. The calculation procedure for these weights was mostly in accordance with the procedure for the travelers weights, the only difference was that here the usage of the average number of travelers per vehicle was omitted.

The weights, calculated as described, were used as the basic tool, which enabled the correct inferential procedure for the calculation of the statistical estimates. Although the specific situation of the "flow-sampling" demanded some simplifications which caused a slight departure from the strict theoretical concepts, all the analyses proved that for the most of the estimated figures a sufficient degree of reliability has been assured. In order to separate the estimates with lower degrees of reliability, we also estimated standard errors for each estimated statistics and then in accordance with the general policy of SORS (see Seljak(2008)), some estimates with the largest standard errors were not disseminated and some estimates were marked as the estimates of lower precision.

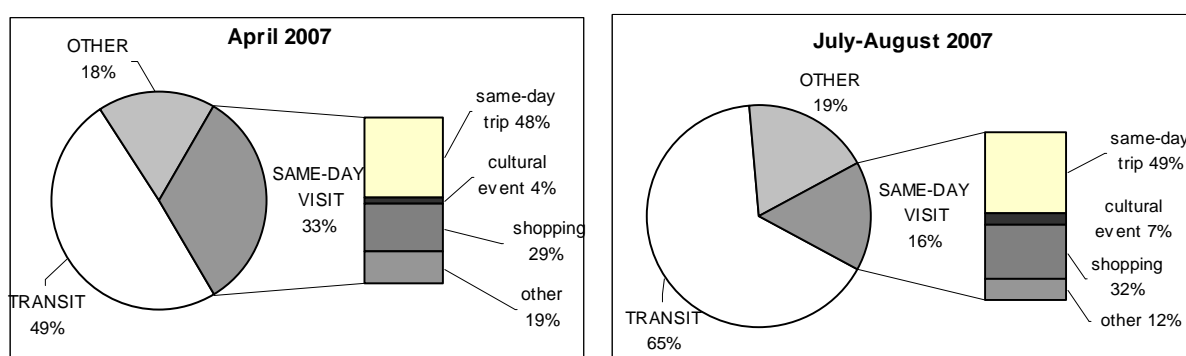
4. Results

The questionnaire was divided into two parts: the first part related to some main characteristics (country of residence, type of vehicle, entering Slovenia at which border, main reason for coming to Slovenia). If the main reason for coming to Slovenia was a same-day visit or transit, the foreign travelers were asked about their expenditures during the visit in Slovenia. The data on expenditure referred to the whole family or traveling party with whom the respondent traveled and shared costs.

As regards country of residence, most of the foreign travelers came from countries that border with Slovenia (Italy, Austria, Hungary and Croatia). In April 2007 three quarters of foreign travelers were from those countries. In the summer months the national structure of foreign travelers was more diverse (55% of travelers were from neighboring countries, 17% from Germany and 28% from other countries).

The share of foreign travelers according to the main reason for coming to Slovenia differed in the spring months from those in the summer months. While in April less than half of foreign travelers were in transit, in summer months the share of transit travelers was about two thirds. In April one third (33%) of foreign travelers came to Slovenia for a same-day visit, while in July and August 16% of foreign travelers came for that reason. The share of foreign travelers who stated other reasons for coming to Slovenia (stay of several days, work, school, daily or ordinary business) was in April and in summer months approximately the same (about 18%).

Figure 3: Foreign travelers by main reason for coming to Slovenia, April, July-August 2007



Source: SORS, Survey on Foreign Travelers at Road Border Crossings, Slovenia, April, July-August 2007

We asked same-day visitors about the main reason for their same-day visit to Slovenia. Less than half of them came on a same-day trip (e.g. for entertainment, sports, sight-seeing, walking, lunch, etc.), less than one third (29% in April and 32% in July and August) came for shopping, while less than one tenth in July and August stated other reasons (above all visits to relatives and friends, business meetings). Most same-day visitors were from Italy, followed by same-day visitors from Croatia and Austria.

Same-day visitors (spending about 35 euros per person in April and about 43 euros per person in July and August) spent on average about four times more than transit travelers (about 9 euros per person per day in April and about 11 euros per person per day in July and August). As regards the main reason for same-day visit, most money was spent by same-day visitors that came to Slovenia for shopping.

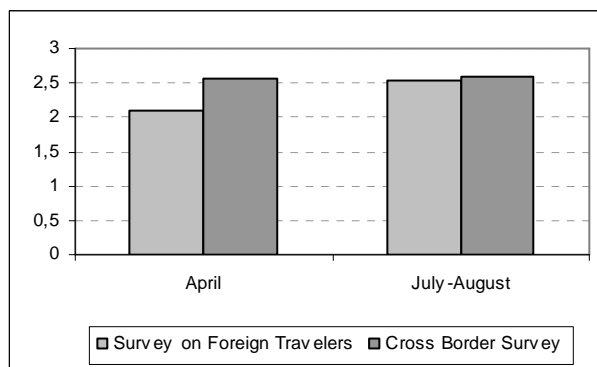
4.1 Coherence of the results with the other survey

Many European countries are aware of the importance of same-day visitors and they carry out surveys to collect information on them (some countries for inbound, some for outbound same-day visits and some for both). Even though there are some basic guidelines about the definition of same-day visits, countries are usually using slightly different definitions of same-day visitors considering their specific situation. Countries are collecting data on same-day visitors for different reference periods. Considering this and the fact that country breakdown in some surveys is not sufficient, mirror statistics is hard to use. Countries usually publish same-day visits of domestic population divided into domestic same-day visits and same-day visits abroad and do not break down same-day visits abroad according to country of destination. In many cases breaking down data on same-day visits abroad according to country visited makes data for certain countries less reliable.

Due to specific characteristics of transit traffic, a comparison with other countries is not possible.

With the intention of finding out the coherence of the data the results of the Survey on Foreign Travelers at Road Border Crossings were compared with the results of the Cross Border Survey which monthly monitors cross-border traffic of vehicles, passengers and goods by road. Entering Slovenia in the Schengen area made it possible to carry out the survey only at the border with Croatia. That is why we compared data between these two surveys only for the border with Croatia. This comparison showed certain distinction considering the number of foreign travelers crossing the border in April. According to the Survey on Foreign Travelers, the number of foreign travelers leaving Slovenia at road border crossings was in April 2007 13% lower than the number of foreign travelers according to data of the Cross Border Survey. In July-August 2007 the distinction was negligible (less than 1%). We assume that this difference is on one hand the result of the use of a slightly different methodology for collecting the data. The data on road cross-border traffic of vehicles and passengers are provided by the border police who at some border crossings estimate the number of passengers using information on the number of vehicles crossing the border and certain coefficient about average occupancy of the vehicle. In the Survey on Foreign Travelers at Road Border Crossings interviewers had to count the number of travelers in the vehicles that were included in the survey. Comparing average occupancy rate of the vehicles shows that the Survey on Foreign Travelers ascertains certain distinction between April and summer months, while according to the Cross Border Survey the average occupancy rate was almost the same in both observed periods of the year. We assume that data on average occupancy rate of the Cross Border Survey were a bit overestimated in the month outside the high tourist season (Figure 4).

Figure 4: Average occupancy rate of private vehicles⁴ at the border with Croatia, April, July-August 2007 (data for the Cross Border Survey are unofficial and are only for the purpose of comparison in this paper).



Source: SORS, Survey on Foreign Travelers at Road Border Crossings, Slovenia, April, July-August 2007
SORS, Cross Border Survey, April, July, August 2007

On the other hand, data from the Survey on Foreign Travelers at Road Border Crossings are based on a sample, which means that certain sample errors always occur.

5. Conclusions

The Survey on Foreign Travelers, which has been carried out by SORS since 2006, meant a challenging task at the time of planning and designing as well as later in the execution phase. Since

⁴ For this comparison as private vehicles were considered private cars, vans, campervans.

several characteristics of the survey (especially the target population of the flow of travelers) were quite different from other regular surveys conducted at SORS, some new approaches were needed. In the first part of the paper we described the sampling procedure, which could roughly be divided into two parts. In the first part we used the previous year counts to predict the present year flow size and then on the basis of this prediction used PPS sampling design to select border crossings, days and starting hours for the field interviewers. In the second part of the procedure the interviewers were instructed to obtain a certain quota of completed questionnaires by interviewing the foreign travelers in different types of vehicles. The collected data were then statistically processed; the grossing-up weights were calculated and with the use of these weights the population estimates were produced and finally disseminated.

Two basic preconditions were needed to assure estimates with a sufficient degree of accuracy. The first is the assumption that the flow-sizes do not differ much according to different years. The second one is the assumption that the interviewers strictly follow the instructions, especially the instructions on the allocated survey date and the starting hour on the selected day. Any departure from the planned sampling scheme requires use of the adjusting procedures in the weighting procedure and can cause a lower degree of accuracy.

One problem for conducting the border survey in the future is the fact that Slovenia entered the Schengen area with the other new Member States of the European Union in the second half of December 2007. In our opinion the most important data as the output from conducting the survey at all national borders in 2007 are the estimates of the transit traffic flow and data on the number of transit travelers (we estimate that most of the transit traffic is oriented to the border with Croatia). Even though conducting the survey only at the border with Croatia will be sufficient for certain time, we are planning to carry out surveys near other national borders – with the difference that interviewers will move from border control points to the nearest gas stations, parking areas or supermarkets. Data on same-day visitors will also have to be completed with a survey at important tourist attractions around the country.

References

- [1] Kish Lesley (1965) Survey sampling: John Wiley & sons
- [2] Sarndal Carl-Erik (1992) Model Assisted Survey Sampling: Springer-Verlag
- [3] Seljak Rudi (2008) Standard Errors Presentation and Dissemination at the Statistical Office of the Republic of Slovenia, European Conference on Quality in Official Statistics, Rome
- [4] Applying the Eurostat Methodological Guidelines in Basic Tourism and Travel Statistics: A Practical Manual (revised version): March 1996