SAMPLE DESIGN AND SURVEY ERROR

Resource Paper for Sample Design Workshop
Seventh International Conference on Travel Survey Methods
Costa Rica, 1st – 6th August 2004

Mira Paskota
The Faculty of Transport and Traffic Engineering
University of Belgrade
Vojvode Stepe 305, Belgrade, Serbia
M.Paskota@sf.bg.ac.yu
tel +381 11 3091 330
fax +381 11 3096 704

Belgrade, July 2004
ABSTRACT
Sample design is one of the most critical phases in every survey. Various types of surveys influence the sample and errors arising from it in different ways. From all the errors that may take place as a consequence of decisions made in course of sample design process, only one of them - sample error - is quantifiable in a satisfactory manner, although all of them together influence the total survey error. Due to unknown population variance, determining of the sample size is also a problem. The situation is even more complicated in real, stratified and multistage samples that are used in travel surveys. In practice, these problems are usually solved subjectively, though the usage of multivariate statistics methods for those purposes is possible and often justified. The application of principal component analysis, cluster analysis and other multivariate statistics methods would provide solutions both for ‘typical’ variances and selection of significant variables. Precision of data is of utmost importance in travel surveys, since the data that have been collected that way are mostly used for modelling and planning in traffic. One should bear in mind that the error that has been propagated through the process of modelling in traffic is total survey error.
1. INTRODUCTION

All phases of survey research are closely related and in a way subordinated to the research goal. The same goes for the sample. We can easily say that survey results and the quality of collected data directly depend on the quality of sample; we can also say that sample design must be co-ordinated with the type of survey for a specific survey; that, again, depends largely on research goals.

Every survey is basically a very complex process. Travel surveys are even more complex than public opinion surveys, or surveys in marketing or some other areas. What makes travel surveys so complex is the larger number of dimensions than in other areas. The space, in other surveys used mainly to define geographic areas, where subpopulations do not have to be homogenous by indicator that are relevant for the survey, is given one new, essentially different meaning when it comes to travel surveys. Traffic is circulating in the area, hence, differently defined geographic areas or urban zones, as well as modes of their interconnections, become the subject of the survey.

However, let us go back to the beginning. Why the sample? What interests every researcher the most is to find out something new about the population, not about the sample. And yet, the sample is much more discussed than the population. There are very few cases where researcher has the opportunity to analyse the whole population, and such a survey is called a census, not the survey. In such cases, we are either talking about an extremely large, or extremely small population. National census of total population is conducted in most countries every ten years, and those are extremely extensive and expensive surveys, with years long preparations. Final results of a census are available to the public and are regularly used both when producing sample design, and when conducting weighting (poststratification) later on in the process. A census represents a basic survey, with a large number of indicators, usually financed by states for a large number of potential, beforehand unknown users, from various areas and with their diverse needs.

In surveys with much more narrowly defined goals, such as travel surveys, total population analysis is rarely required and truly justified. It could be justified only in another extreme case, an extremely small population, whereas with large populations census is not used for these purposes.

One should constantly keep in mind that every researcher, no matter what he does with the sample, actually wants to draw conclusions about the population. If the population were entirely homogeneous, it would be irrelevant as to which part of it we are examining, we would always get the same results. However, all real populations are far from being homogeneous, and it is of great importance for us that the part we are examining adequately describes the totality about which we want to talk, draw conclusions and perhaps make decisions. When we say adequate, we mean that the diversity of total population should show to the same degree in one selected part of it – the sample. And the sample that reaches this goal we call representative.

The sample is always a sort of compromise between the wish or need for the data as accurate as possible, and the costs of collecting them. When producing sample design, a statistician cannot be guided by purely theoretical reasons; he must take reality and the limitations it puts upon into account. The smallest possible sample error is of supreme importance for every survey; on the other hand, the budget is also limited, and all this must be taken into consideration.
2. **Basic Concepts Related to Sample Design**

2.1 **Influence of Survey Mode to the Sample**

Sample design is closely connected with the way research people intend to collect data, that is, with the chosen survey mode. Choice of survey mode is mainly influenced by the survey’s goal, target population characteristics, size and complexity of questionnaire, as well as limited resources, whether it's money, staff or equipment. Various forms of surveys are sensitive to different kinds of errors that can occur in the course of taking a survey, and they will be discussed in Chapter 2.5. All this has direct effects on the sample.

Traditional methods of interviewing are
- Mail survey
- Telephone survey
- Face-to-face survey

Due to the fast computer development, there have been significant changes in this field during the last 25 years. Some new methods of interviewing have been developed, and traditional methods have suffered significant changes. First to appear was CATI (Computer-Assisted Telephone Interviewing) as a modification of telephone survey and CAPI (Computer-Assisted Personal Interviewing) as a modification of face-to-face survey. Then came CASI (Computer-Assisted Self-Interviewing) and ACASI (Audio Computer-Assisted Self-Interviewing). Internet surveys have been very popular lately, and some experts claim them to be the survey method of the future. Also, a large number of various combinations of different interviewing types have appeared. All of them, along with other factors, have influenced sample design.

From the sampling point of view, what is crucial is the first contact and respondents' motivation to participate in the survey. Of course, the possibility of communication between the interviewer and the respondent depends directly on the type of survey. Mail as a type of interviewing is highly impersonal, so the communication possibility is minimal. As a consequence, with mail surveys the response rate is lowest. Face-to-face survey provides the interviewer with the best possibilities for communication, but only under the condition that the contact has been made. This problem cannot be discussed without mentioning huge differences that exist among particular countries.

2.1.1 **Advantages and Disadvantages of Some Survey Modes**

When listing advantages and disadvantages of the best-known survey modes, and taking the international character and this conference’s topic into consideration, we have paid special attention to the huge differences that exist among countries, and are in relationship with this.

**Mail survey**

This has been a dominant survey mode at the very beginning of survey researches, and has maintained the leading position for a long time. This was probably a consequence of the fact that people had more time, and their mailboxes were less occupied with spam than today.

The sole advantage of mail surveys is that they are relatively cheap and do not require a large number of trained staff. And a great disadvantage is a possibility of noncoverage error and low response rate. Address lists are difficult to obtain and often unreliable. These problems occur in many countries to a different extent and for various reasons; however, they are significant
everywhere. Lists of general population with addresses are in no country publicly available, so the researchers manage in various ways, by using phone books, membership lists from all sorts of associations, and similar. These sorts of lists miss a large part of population as a rule, and usually are out of date.

Some other flaws of mail surveys are: Respondent’s degree of interest for survey topic strongly influences response rate, and this introduces a great bias which is uncontrollable. There is no control about who fills in the questionnaire, if it is the person it was addressed to, or someone else. It is not possible to control whether the questionnaire is filled in correctly and completely. Also, one should not forget that in some countries there still exists the problem of illiteracy.

**Telephone survey**

In highly developed countries, where telephone penetration is extremely high, for years this has been, and still to a great extent is, a dominant survey mode. It is convenient for a number of reasons. These surveys are much faster both from face-to-face and mail surveys. Usage of RDD (Random Digit Dialing) system enables design of a high quality sample for the given frame, as well as anonymity. Interviewer’s participation brings significant control into the interviewing process. Usage of CATI system enables simultaneous control of the course of survey, data entry, data control and immediate accessibility in desired format.

General limitation for the telephone surveys lies in the fact that it is not applicable in case of very long questionnaires or diary surveys, which are relatively common in travel surveys. Other problems with telephone surveys are essentially different between developed and undeveloped countries. Whereas in undeveloped countries there is still a problem of noncoverage, i.e. low telephone penetration, especially in rural areas – a problem long time forgotten in developed countries, which is now coming back again, but in a quite a different form. In the last few years, an interesting phenomenon has been observed in highly developed countries – a growing number of households, mainly young couples and individuals, do not use stable phone at home, but only mobile phones. Since these individuals are much more mobile than average population, this kind of noncoverage could cause serious problems when travel surveys are considered. Experiences in some countries have shown that surveys on mobile phones are not practically applicable, making this issue of high current interest, and still not solved in a satisfactory manner.

**Face-to face survey**

Face-to-face survey was and has remained traditional and still unsurpassed mode of survey. This is the only way to take a survey among non-listed population. It enables completion of the longest and most complex questionnaires, possibly demanding some additional material to display to respondents (visual aids and similar). Interviewer’s presence ensures the correct filling in, and application of CAPI methodology and lap-top computers bring all the advantages of telephone CATI surveys. It is extremely convenient in cases of general population surveys and multistage samples.

One serious disadvantage of this mode of survey are big costs. The most common form of this mode of survey is FTF household survey. This mode of survey demands that the interviewer strictly observes the rules of household selection and an individual in it, as well as rigorous control of field work. Some of well-known problems related to FTF household surveys are: nonresponse differs drastically in urban and rural areas, smaller households are underestimated as a rule. There are huge differences among countries about the possibility of entering the household and establishing first
contact. In highly developed countries, especially in urban and exclusive residential parts, it is very difficult to enter the household, unlike less developed countries in which this problem does not exist.

**Internet survey**

Majority of so-called ‘surveys’ we can often see on the Internet have no sample control whatsoever, and cannot be regarded as serious researches. This, of course, doesn't mean that it would be impossible to organize a serious survey on the Internet, but such a project would take much more planning, rigorous control and much, much more work. Good results never appear by themselves, as some individuals appear to imagine when placing a question on a web page, and expect the results to “count themselves”!

Internet surveys call for extreme caution. At the moment, high quality Internet surveys are possible only in case of very narrow and specific populations, the ones we have appropriate frame for, and for which we are confident about their having regular access to Internet, such as university population or population of employees in a company. For general population surveys, and travel surveys belong to those, the Internet is sill not an appropriate survey method, even in most developed countries in the world, let alone the others.

2.1.2 **Convenient Combinations of Survey Modes**

In underdeveloped countries, due to telephone network’s low penetration, relatively low nonresponse with FTF surveys and easy access to households, face-to-face is the best survey mode for general population, provided that there is enough professional staff at disposal. In developed countries, the situation is quite different, and it is difficult to give general recommendation for an ideal mode of survey. Due to the disadvantages listed for the best-known modes of survey, the decision has to be made for each research individually. One of the options, which is most often resorted to, is combining various modes of data collecting in one survey.

Combinations of survey modes came as a result of efforts to use advantages and avoid disadvantages of all basic survey modes. For instance, with the telephone survey, the problem of noncoverage can be resolved by FTF interviewing of the respondents that have no telephone. Telephone is also often used to reduce nonresponse in mail survey. After a while, respondents that have not answered back are reminded by phone and asked to do so. Other combinations are also possible, such as mail invitation to participate in web survey, telephone announcement and scheduling FTF interviews etc.

Survey mode is directly related to the available sample frame. For telephone surveys, the most natural choice of frame is phone book, whereas in household FTF surveys it is usually not used. In multimode surveys, caution is advised in cases when the same questionnaire is used in one way on one part of population, and in another on the other part, because measurement error could differ significantly and bring in significant bias.

2.2 **Some Sample Classifications**

Various kinds of sample classifications are found in literature, and according to various criteria. One of the most common sample classifications is between probability samples and non-probability samples. Error is present in all types of samples, but only with probability samples it is possible to speak about estimation of error, or at least a part of it.
Probability samples are classified as follows:

- **Simple random sample.** In case of a simple random sample, an enumerated list of the total population is needed. Elements are chosen completely randomly, independent from one another and without replacement, that is, an element that has been drawn once cannot be chosen again. In the past, random number tables were used for simple random sample, and today computer random number generators are almost exclusively used.

- **Systematic sample.** Some authors consider systematic sample to be a variation of simple random sample, though it could be considered as a special case of cluster, even as a specified case of stratified sample. By dividing the population size by the sample size, we get the value called the step. It is sufficient to generate one random number smaller than the step; this value is called the start, and represents the ordinal number of the first element chosen for the sample. Then, the step value is added to the start value and the next sample element is determined. The procedure continues in this way, systematically determining the following elements all the way to the end - that is, to the last sample element.

- **Complex random samples** (Stratified sample, cluster sample, multistage sample)

Depending on whether an appropriate list of all sample elements exists, there is a sample classification to those whose frame is based on an exhaustive list of all elements, and to samples without the list. If there are clearly defined subpopulations (strata) in population, the sample is created to reflect each of those strata. This kind of sample is called stratified. Stratified sample can be with proportional probabilities, or with different probabilities of selection.

Depending on the number of stages the sample is projected to, there is a standard classification to one-stage and multistage samples. Multistage samples are the ones where in the first stage there is no sampling of the final sample elements, but only some of their groups, defined in various ways. In practise, these are usually two-stage and three-stage samples.

Samples that do not reflect properly of the population they have been selected from, and carry some degree of bias, are called biased and their error is large. Anyone who deals with surveys is aiming for a non-biased sample. A non-biased sample is also called a representative sample. Representativeness is a necessary condition for making conclusions about the population based on the sample. In order to have a representative nation-wide sample, it is necessary that it is large enough, and that its elements are scattered across the whole territory of the country as much as possible.

### 2.3 What is Necessary to Know Before Starting with Sample Design?

Each of the phases in the survey could be called critical for the process development, and that certainly goes for the sample. One of the reasons that has kept sample design taboo in survey methodology is most certainly statistics. On one hand, many researchers use survey on regular basis, without sufficient formal knowledge of statistics (“Statistics is bogeyman“ approach). On the other, again, statisticians dealing with samples have shown lasting tendency to stay within the limits of their own area, without explaining sample design to users (“Why bother explaining, they wouldn't understand anything anyway“ approach). Both approaches, extreme as they are, are bound to be wrong. Anyone dealing with survey design or with analysis of the collected data, has to be acquainted with all the phases in the survey, and degree of knowledge in statistics that is required for understanding of sample design basic principles is not too high. Also, with some good will, statisticians could adapt their terminology and explain in a simple manner what, how and for what purpose they are doing in course of sample design process.
As already said, survey’s aim is a starting point for all the phases of the survey, including sample design. Clearly defined target population and the instrument to be applied (questionnaire) also have to be known before sample designing. Every researcher's wish is to know sample size as soon as possible, and to try to fit it in the budget, or vice versa. However, as we shall see in Chapter 3, sample size is a very complex issue.

Generally, sample size is influenced by:

- Maximum error allowed.
- Population size (influences significantly when small populations are in question).
- Data variance.
- Identifying the smallest group (and its exact size) in the population for which we want an estimate based on the sample.
- What we intend to do with the collected data.

When asking data users about the permitted error level, one should check how clear they are about the concept of total survey error, if they can tell the difference between sampling and non-sampling errors and whether they are capable of determining the exact level of precision needed.

Before starting the process of sample design, it is necessary to define sampling units. Sampling units are elementary units or groups of them, clearly defined, easily identifiable and suitable for sampling. Groups consisting of a larger number of elementary sampling units are usually called primary sampling units. In travel surveys, elementary sampling units are usually households or persons. With multistage samples, on higher sample levels, primary sampling units are always groups of elementary units. With households samples, we are usually talking about city blocks, local community centres, census areas or electoral districts.

Before starting sample designing, data availability is examined for the population structure, according to variables which influence the survey subject. At the very beginning, one should know if there is sufficient amount of data for stratification and, potentially, weighting.

Other information which is less often discussed, and which is directly implied by the survey's goal, is necessary before starting sample designing. What purpose will the gathered data serve, how will they be analysed, what statistical methods will be applied? It would be useful to, at least generally, know what you would like final tables in the report to look like. Will the multivariate statistics be applied, and what methods? If the data are to serve as an entry in the further process of planning in traffic and travel, this has to be known at the very beginning.

Sample design is always a compromise, aiming to achieve maximum final sample efficiency – achieving optimal precision by using all available means and taking the existing limitations into account.

Since travel surveys are mainly conducted among general population, or some large populations, differently defined, this paper’s starting point is that the population is big enough, and that the influence of finite population correction (fpc) could be disregarded. In practice, a population is considered to be large if its size is ten thousands or more. Other indicators that should be evaluated, such as error, variance and minimal population group for which an estimate is desired, are always important when creating sample design for travel survey.
2.4 Sample Frame
Sample frame is a group of elements (households or persons) that qualify for being chosen for the sample. The sample can be representative only if the frame equals population, that is, if there is no coverage error. Such ideal frames are seldom available. Probability sample, and that is the only kind we are dealing with in this paper, implies that each element (household or a person) has a known probability to be included in the sample. With simple random sample, those probabilities are equal for each element; however, with some other samples, as we are about to see, those probabilities are not necessarily equal. Therefore, the sampling frame consists only of the elements with known probabilities to be chosen for a sample.

Although all the surveys begin with an idea that the sample should represent the population, one should bear in mind that the sample can represent only sample frame, and that those elements that are outside the frame cannot be found in the sample either. When general populations are considered, and when household survey is done, the persons who are usually left outside the frame are: solders, individuals living in dormitories, prisoners, individuals living in nursing homes or some other social institution, and homeless people.

The frame usually represents the list of elements. In order to make a list good enough to be used as sample frame, several conditions have to be met:

- To cover population as well as possible.
- To exclude duplicates.
- To exclude redundant elements (meaning that all the elements of the list are at the same time population elements).

The frame could also be a geographic map or city map. If, along with geographically defined entities, there is also a list containing information about them, we are talking about standard frame. If these information about the entity are not given, only area probability sample is applicable.

In many cases, the lists are formed based on several sources, by combining a larger number of lists, containing similar information on different elements. This way of work is a result of efforts to reduce noncoverage, but could easily introduce another problem – duplicates and elements that do not exist in the list. Duplicates can be eliminated through a process of careful examination of the list, whereas with lists of elements that do not belong to the defined target population, the situation is somewhat different. Most commonly, the screening is used during the field work. After establishing first contact with respondents, the interviewer asks a group of selective questions, aiming to establish if the contacted person is a member of target population. If that is the case, the interviewer will continue further with questioning; otherwise, they will say thank you, and look for another respondent.

In some surveys, phone books of different areas, electoral lists, house owners lists, some club membership list and similar could be used as the frame; however, for general populations, such frames are basically wrong, since they leave a large part of the population uncovered and are often out of date. In some cases, with large surveys, it is possible that producing a list of households in the field would be only the first part of the survey, forming a frame that is used for sample afterwards; however, surveys with such level of organisation are very rare, due to high costs of the field work.
It is very important to list everything that has been used as the frame when producing sample design for each individual sample, thus focusing attention to a possible cause of coverage error. For the final sample estimate, details of the process are necessary, as well as the response rate.

As we’ve seen, obtaining an appropriate list for the frame could represent a problem; therefore, on some occasions it would make sense to ask if it would be better to choose a frame without the list. One of the alternatives to the frame based sample would be area probability sample. With this type of sample, geographically defined areas are used as basis, which are, due to the lack of other sorts of data, chosen with probabilities that are proportional to their area size.

### 2.5 Survey Proces and Possibility of Error

One of the first steps in making survey design, right after the survey goals are defined, is determining target population; that is, the population of interest for a specific survey. In this paper, we shall assume that target population has been previously clearly and decidedly defined.

Since the survey is, as already mentioned, a very complex process consisting of several phases, each of this phases gives possibility for an error to occur. Some of these errors are not related to the sample design, and such errors won’t be the subject of interest in this paper. However, as we are about to see, sample design is closely related to some of the phases of the survey; therefore, all the errors that can arise in the survey process will be of interest for us.

In the literature, four types of errors occurring in surveys are most frequently brought up: coverage error, sampling error, nonresponse error and measurement error (Fowler, 2002, Salant and Dillman, 1994). Classification into sampling and non-sampling errors is standard, and coverage error, nonresponse error and measurement error are usually listed under non-sampling errors. However, it is quite clear that only one of them, measurement error, is completely unrelated to design and sample realization process. From the sample statistician’s point of view, as well as for the purpose of this paper, it would be useful to make somewhat different and more complete classification of errors, which would include some errors that are rarely mentioned in survey literature. Survey errors classification is given in Table 1. Two basic groups of errors are: errors that are (directly or indirectly) sample dependent, and errors that are completely sample independent.

#### Table 1: Survey Errors Classification

<table>
<thead>
<tr>
<th>1. Sample dependent errors</th>
<th>2. Sample independent errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target population</td>
<td>1° Coverage error</td>
</tr>
<tr>
<td>Sampling frame</td>
<td>2° Sampling error</td>
</tr>
<tr>
<td>Projected sample</td>
<td>3° Nonresponse error</td>
</tr>
<tr>
<td>Realized sample</td>
<td>4° Adjustment error</td>
</tr>
<tr>
<td>Adjusted sample</td>
<td>5° Measurement error</td>
</tr>
<tr>
<td></td>
<td>6° Processing error</td>
</tr>
</tbody>
</table>
We shall discuss in detail only the errors that are in any way sample dependent.

2.5.1 Coverage Error
The first issue we have to face in sample projecting is the issue of frame. This is also the first opportunity for an error to occur. How well does the frame cover the population, and how big is the discrepancy between them? We have seen that even the best frame doesn't cover the population entirely, and the discrepancy between the target population and sampling frame is called coverage error. We can have an idea about the maximal coverage error we can allow in a specific survey, but it is highly questionable how reliable our estimation of the error can be. Estimation of coverage error represents a problem for numerous surveys, and one of the goals should be to reduce it as much as possible. If the discrepancy between target population and sample framework is too large, the sample statistician will frequently completely abandon the intended frame, and choose another type of frame, or possibly another solution.

2.5.2 Sampling Error
Sampling error occurs due to the fact that, instead of making conclusions about population based on all its elements, we take only one part of it. The reason for an error lies in the random nature of sampling, and probability theory is used to calculate it. Sampling error is the only error that can be estimated, and we shall mainly discuss it in this paper.

One of the most common misleading notions related to surveys is that it is possible to calculate error for every survey. Unfortunately, this is not the case. This is possible only with sampling error, and only with probability samples. It is a common practice for researchers to mention estimated error along with their survey results. But it is very seldom that they note that the mentioned error estimate is only a sampling error, and that is not the only source of errors in a survey. Even more rarely do they specify the type of the sample used. In surveys, quota sample is often used, which is non probability, therefore error cannot be estimated for it, but it doesn't bother some “researchers” to specify errors even for quota samples! And only they know how they'd got it. This line of thinking would lead us to problems of use and misuse of surveys, and that is not the subject of this paper.

Another misleading notion that has spread among non-experts is that “the bigger the sample, the smaller the error”, and that size of the sample is proportional to the size of population. This is partly true, since smaller samples always carry bigger error than big samples. However, very big samples are usually not needed, and are basically not rational. Enlargement of sample over a certain limit only increases costs, whereas the data quality increase is negligible. Sample size 500 will provide practically the same error for population size 10 000, and for population size 1 000 000. This “paradox” is caused, as we shall see in Chapter 3, by the fact that the relationship between error and sample size is not linear but quadratic. As a consequence, if we would want to reduce error k times, we would have to increase sample size k^2 times! One useful consequence of this insight is that, with nationwide surveys, the same sample size is sufficient both for, let's say, Austria and USA. Regardless of the fact that USA is several times bigger than Austria, the error would be almost the same. This opens the possibility of conducting the same type of surveys in various countries, on the sample of the same total size. However, everyday experience shows that in big countries bigger samples are still more common. Reason for this is not related to the size of total population, but to the number of entities for which the conclusions are drawn (number of strata).
Sampling error cannot be discussed in an abstract manner. Actually, sampling error is not completely precise expression, and is often misleading to a belief that it is possible to calculate sampling error per se, which is untrue. The error can be determined only for a specific measure or statistics calculated based on the sample, and is different for each of them. This is the reason why, among other things, it is necessary to know when determining sample size, what the collected data will be used for afterward. The type of survey (descriptive or analytical) will dictate the requested level of precision, and consequently the sample size.

Chapter 3 of this paper deals with sampling errors in detail.

2.5.3 Nonresponse Error
The realized sample is always different from the one that has initially been projected. This difference between them is called nonresponse error. Keeter et al. (2000) claim that nonresponse error is both the function of the nonresponse rate, and the differences between respondents and nonrespondents based on the variable of interest. Nonresponse can take place for a number of reasons, and it is common practise in surveys that diaries of nonresponse are kept in course of the field work, in order to obtain information on the nonresponse rate. Heterogeneity between respondents and nonrespondents remains mainly unknown.

It is common knowledge that some populations are harder to reach than the others. For instance, nonresponse is much higher in urban than in rural areas in all countries. In case of the household survey, it is much harder to find smaller, especially one person households. If the behaviour of these subpopulations in relationship to the survey subject were the same, the nonresponse problem would not exist; however, there are usually significant differences between easily and hardly reachable parts of the population, and significant attention should be focused on the nonresponse problem. With household travel surveys, the same problems arise as in other household surveys. However, the nonresponse error could have greater impact on data quality and reliability, than with some other types of surveys. Younger population and members of smaller households, regularly underestimated in household surveys, are most probably the extremely mobile part of population, therefore the ones that move around and travel the most, and are very relevant for any travel survey.

Experienced professionals can evaluate nonresponse rate subjectively, based on the characteristics of defined population and the projected questionnaire, and the sample is projected to be as large as possible, in order to make the size of the realised sample big enough. The only way to reduce nonresponse error is to obey strictly the procedure of respondents selection, and execute severe control of field work. If, after all, a significant error still takes place, the usual way to remove it is weighting.

2.5.4 Adjustment Error
After the data have been collected, and prior to their final analysis, there is always an effort to remove the errors that could have arisen in the process: coverage, sampling and nonresponse error. By performing analysis of the sample produced, and its control with the population structure, it would be possible to establish which parts of populations are represented more, and which less than they should be. Then the underrepresented elements are given more weight, and those that are overrepresented are given smaller weights. This procedure is called weighting, poststratification or sample adjustment. Basically, the idea for adjustment is to improve sample quality; however, the effect is not always the desired one. Choice of appropriate weights is a critical point, where a
mistake is easily made, and in such cases the consequences are significant. For the weights to be adequately determined, good knowledge of population structure is required, and if the secondary data are not sufficient or not precise enough, sometimes professional heuristic expertise is used.

3. SAMPLE SIZE AND ESTIMATION OF ERROR

We are often in the situation to have our client’s first sentence sound like this: "I want to do a travel survey. How big a sample do I need?" or "Is 400 elements a big enough sample?" The answer to both of them, though not very imaginative, always has to be – It depends!

Sample size is a complex problem that cannot be solved easily, and the answer to this question depends on many things. When you are dealing with a defined target population, in order to determine necessary size sample, the things you need to know are:

- What do you want to measure?
- How precise do you want it measured?
- Do you presume the population to be homogenous in relationship with that property or not?
- Do you intend to choose sample elements individually or in groups?

And, as we are about to see, even this is not enough.

Error and sample size are closely related to each other. Standard procedure is to have an idea about the error allowed for a specific survey. Then, taking this error into account, as well as numerous other parameters, the sample is projected. Only after the interviewing and data collecting is over, it is possible to give a final estimation of the sample error.

3.1 Simple Random Sample

Simple random sample is an ideal sample, we could even say - prototype of all other samples; however, we seldom have the opportunity to use it in practice. Regardless of this, thorough understanding of it is necessary for comprehension of more complicated samples.

Simple random sample gives each element of the sample frame equal chance to be selected. Theoretically speaking, when designing SRS, one should determine all possible samples without replacement of the size n, and than chose one of them randomly. This is never done in practice, because it is too time consuming for large populations.

Sample size also depends on the type of parameter you would like to calculate, or, to be precise, to estimate. Parameters that are most often estimated are the mean and proportion. Because of this, we shall show two examples for these two basic parameters, and explain the procedures for determining sample size.

3.1.1 SAMPLE SIZE REQUIRED FOR ESTIMATION OF MEAN

Mean carries a smaller error than proportion, and it takes smaller sample for its estimation than for proportion estimation. In travel surveys, we encounter continuous variables more often than in other survey types. An example of such a variable would be the following question:

*How long did your home - work trip last yesterday?*
Let us assume that the researcher intends to specify the mean of the trip duration in his report. How big a sample does he need? Standard deviation of the arithmetic mean is

$$\sigma_{\bar{x}} = \sqrt{\frac{1-f}{n}} \sigma,$$  \hspace{1cm} (1)

where \( n \) is the sample size, \( f = n/N \) fraction, and \( \sigma \) standard error of variable (in this case, the travel duration). The value \( 1-f = 1 - \frac{n}{N} = \frac{N-n}{N} \) is called finite population correction (fpc). In case of populations larger than several tens of thousands, the influence of small population is negligible, and \( 1-f \) can have value 1. In such cases, which is usual with travel and other household surveys, equation (1) gets its much better known form

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$  \hspace{1cm} (2)

Confidence interval for the mean is

$$P(\bar{X} - t_{\alpha} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha,$$  \hspace{1cm} (3)

where \( 1-\alpha \) denote the confidence level, and \( t_{\alpha} \) table value for risk \( \alpha \). Total confidence interval length is

$$l = 2d = 2t_{\alpha} \frac{\sigma}{\sqrt{n}},$$  \hspace{1cm} (4)

where \( d \) denote the maximum allowed positive or negative deviation from the mean. From the relation (4) the required sample size could be formulated as

$$n = \frac{t_{\alpha}^2 \sigma^2}{d^2}. \hspace{1cm} (5)$$

From the relation (5), we can see what we need to know in order to determine sample size: maximum allowed absolute error (d), variance (\( \sigma^2 \)) and confidence level (that implies \( t_{\alpha} \)). One of these values is usually unknown - the variance. To determine sample size for the mean of variable, such as time of travel, we should know the variance of that variable! We don't know the variance in advance and we can estimate it only after the data are collected. This is a serious problem. Do we have any other information that could be of help to us? Have similar surveys been done somewhere before, the results of which could provide guidelines on an approximate variance size?

If we assume that the home - work trip lasts about 40 minutes on the average, we would be 95% certain that the real value will be within the interval of \( \pm d \) minutes from the calculated mean. Calculated sample size for the example, depending on error and the variance (standard deviation), is given in Table 2.

Based on Table 2, one can get the impression about the influence of variance on sample size and about the importance of the correct estimation of its approximate size. If we underestimate the
variance and choose a smaller sample, we shall find ourselves in the position to be unable to estimate mean accurately enough. If, on the other hand, we overestimate the variance, sample size will be bigger than we really need, therefore putting us in the position to spend money needlessly.

Table 2: Sample size required for 95% confidence interval of the mean depending on error and standard deviation.

<table>
<thead>
<tr>
<th>Standard deviation σ</th>
<th>±4 minutes</th>
<th>±3 minutes</th>
<th>±2 minutes</th>
<th>±1 minute</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 minutes</td>
<td>25</td>
<td>44</td>
<td>98</td>
<td>392</td>
</tr>
<tr>
<td>15 minutes</td>
<td>56</td>
<td>98</td>
<td>221</td>
<td>882</td>
</tr>
<tr>
<td>20 minutes</td>
<td>98</td>
<td>175</td>
<td>392</td>
<td>1568</td>
</tr>
</tbody>
</table>

3.1.2 ERROR OF MEAN

The variance that is calculated based on value from the sample we denote with \( s^2 \). When the survey is over, this value can be used, based on the realised sample, to estimate error of the mean. By replacing \( \sigma^2 \) with \( s^2 \) in (5), we can transform it in the following way

\[
n = \frac{t_{\alpha}^2 s^2}{d^2}.
\]

We now know the size of the realised sample \( n \) and the sample variance \( s^2 \). From (6), we can determine absolute error \( d \) as

\[
d = t_{\alpha} \frac{s}{\sqrt{n}}.
\]

Standard way to present mean is \( \bar{x} \pm d \), though sometimes what is shown are confidence intervals \( (\bar{x} - d, \bar{x} + d) \), with the necessary mention of the confidence level \( 1 - \alpha \) or risk \( \alpha \).

3.1.3 SAMPLE SIZE REQUIRED FOR PROPORTION ESTIMATION

Although they contain more continuous variables than other surveys, travel surveys also contain a large number of extremely important variables with which the proportion is estimated. This means that for the final decision on sample size, proportions are crucial, not the mean.

Let us consider the following, relatively simple question:

\textit{Did you use the means of a public transport yesterday?}

We shall assume that the population is large enough, so we don’t have to use finite population correction (fpc). In that case, standard deviation of the proportion is

\[
\sigma_p = \frac{pq}{\sqrt{n}},
\]

with \( n \) denoting sample size, \( p \) expected proportion of respondents that use public transport during the day, and \( q=1-p \) the expected proportion of respondents that do not use public transport during the day. As we can see right at the beginning, standard deviation of proportion depends on the value
we are just trying to estimate. This problem is inevitable when we are talking about the sample, and is usually resolved by relying on data from some other source (from the past, or someone else’s data). The question the user is supposed to answer is whether there are some previous data, results of some previous surveys, secondary data and similar, that could provide us with information about what this proportion could be, at least roughly speaking. If such information is not available, the most conservative value \( p=q=0.5 \) must be accepted, and that is the value of proportion with highest variance.

The next thing one must know is the level of error that is acceptable for the user; let us denote this value with an \( e \). It is important to stress that no one can guarantee that the real value is within limits \( \pm e \) from the calculated proportion value, but it can be stated with high confidence. If we choose confidence of nearly 95\%, then the error we can tolerate is two standard deviations

\[
e = 2\sigma_p = 2\sqrt{\frac{pq}{n}}.
\]

Actually, with the error of two standard deviations, the exact confidence value is 95.46\%. If we take \( n \) from the relation (9), we shall get that the needed sample size is

\[
n = \frac{pq}{(e/2)^2} = \frac{4pq}{e^2}.
\]

In Table 3, the required sample sizes for various error values and proportions are listed.

**Table 3: The required sample size for 95.46\% confidence interval for proportion \( \pm 2\sigma_p \), depending on the error and the expected proportion**

<table>
<thead>
<tr>
<th>Proportion ( p )</th>
<th>±5%</th>
<th>±3%</th>
<th>±1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>144</td>
<td>400</td>
<td>3 600</td>
</tr>
<tr>
<td>0.2</td>
<td>256</td>
<td>712</td>
<td>6 400</td>
</tr>
<tr>
<td>0.3</td>
<td>336</td>
<td>934</td>
<td>8 400</td>
</tr>
<tr>
<td>0.4</td>
<td>384</td>
<td>1 067</td>
<td>9 600</td>
</tr>
<tr>
<td>0.5</td>
<td>400</td>
<td>1 112</td>
<td>10 000</td>
</tr>
</tbody>
</table>

Sometimes, in practice, confidence levels different from 95\% are used. Sample sizes required for some other confidence levels are not given in Table 3, but are easily calculated if in formula (9) 2 is replaced with an appropriate table value.

**3.1.4 Proportion Error**

Proportion error is expressed in (9). Proportion calculated based on sample we denote with \( \overline{p} \) and could use for final estimation of sample error

\[
e = 2\sigma_{\overline{p}} = 2\sqrt{\frac{\overline{p}q}{n}}.
\]

We have seen that, with proportion estimate we need a bigger sample than for the mean estimate. In all surveys, there is at least one variable of an extreme interest, that is presented in the form of
proportion; that is the reason why in some books on survey methodology, when discussing determining sample size, the mean is not even mentioned, and only the proportions are discussed. In textbooks on sample (Cochran, 1977, Murthy, 1967), the mean, its error and determining the size of the required sample always have their own chapter, but are, unfortunately, observed completely separately from the proportion.

If the results collected in the survey are analysed using one of advanced statistical methods, numerous other methods that have remained unmentioned above will appear as a result. For instance, the variance and covariance, the difference between the two values, correlation coefficient, regression coefficient, canonical correlation and similar. Modes of estimate of standard error of those parameters can always be found in the relevant statistics literature (Cochran, 1977, Murthy, 1967 etc.).

3.2 Stratified Sample
Equations listed in the previous chapter are related exclusively to simple random sample. As we know, when it comes to household surveys, this type of sample is used very rarely, almost never. Household surveys almost always require stratified sample.

Often we can divide population into easily distinctive groups we call subpopulations. General population of persons is naturally divided into male and female, older and younger subpopulation. If subpopulations do not overlap, that is, if they are mutually exclusive, then they are called strata. Variable according to which the population is divided into strata is called stratified variable. Examples of stratified variable, besides sex and age, could be race, religion, mother language, state or municipality of permanent stay, socio-economical status etc. Also, in cases when we are interested in travel phenomena, stratified variables can be defined in various ways. On households level, we separate those that own at least one motor vehicle from those who don’t, households living in their own apartment/house from those living in rented, urban and rural, etc. Individuals could be classified into those who have driving licence and those who don’t, whether they travel from home to work/school on daily basis, have they travelled abroad during the previous year or not, or in many different ways.

In case of stratified sample design, it is necessary to know the size of the stratum in population. In order to obtain such information, external and secondary sources are used. For demographic variables, these information can be obtained from census, for the number of registered motor vehicles or drivers, appropriate official data basis can be used, etc.

There are many reasons for stratification. Stratification is needed if we want to make conclusions about each subpopulation independently; however, one of the most common reasons for stratification is reduction of the total variance, which also represents an attempt to reduce the sample size. Sometimes, reasons for stratification are solely administrative.

If simple random sample is taken from each stratum, such sample is called stratified random sample. Appropriate stratification can reduce the variance in relationship with random sample, thus increasing precision of the estimate of parameters for total population. Of course, a good stratification that accomplishes this effect is only the one that manages to separate strata that are essentially different from one another, and rather homogenous within themselves.
After the strata in a population have been defined, and information about their size obtained, the next step in designing stratified sample is allocation. Allocation represents distribution of total sample size into strata. The most common and the simplest allocation is the one in which every strata is given, or allocated sample that is proportional to its size. Such allocation is called allocation with probabilities proportional to size (PPS).

In sample design and statistics in general, it is usual to denote the population size with $N$, and sample size with $n$. Ratio between those two values $n/N$ is called fraction and is denoted with $f$. If we assume that there are a total of $k$ strata, we shall denote the size of each of them with $N_i$, and the size of an appropriate sample with $n_i$. With proportional (PPS) allocation, we can see that

$$\frac{N_i}{N} = \frac{n_i}{n} = w_i, \ i=1, ..., k. \tag{12}$$

Quotients $w_i$ are called stratum weights. Equation (12) could also be expressed as

$$\frac{n_i}{N} = \frac{n}{N} = f, \ i=1, ..., k. \tag{13}$$

Samples with the constant $n_i/N_i$ ratio equal to the fraction are called selfweighting samples. With samples where this is not the case, if the allocation is not proportional, posterior data weighting is needed, to bring back each strata the probability which is proportional to its size. This kind of weighting procedure is called poststratification.

### 3.2.1 Variance with Non-Proportional Allocations

Many authors have dealt with the problem of variance estimate with stratified samples in details (Golder and Yeomans 1973, Cochran, 1977, Kish and Anderson, 1978, Jolliffe, 2003). For the purpose of better understanding of the variance estimate issue and determining size in case of complex samples, in this paper only basic concepts are given.

In case of stratified sample, mean of the whole set could be calculated based on strata means as weighted mean. The variance of the mean with the stratified sample, in cases of very large populations ($n << N$), when finite population correction doesn’t have to be used, equals

$$\sigma^2 = \sum_{i=1}^{k} \frac{w_i^2 \sigma_i^2}{n_i} = \frac{1}{n} \sum_{i=1}^{k} w_i \sigma_i^2, \tag{14}$$

where $k$ stands for the number of strata, $n_i$ for sample size in each strata, $w_i$ for weights of each stratum and $\sigma_i^2$ for the variance in strata.

Based on (14), determining the size of stratified sample is simple, $n = \frac{\sum_{i=1}^{k} w_i \sigma_i^2}{\sigma^2}$, yet implies that variances of all subsamples are known, which is a theoretical assumption in most cases. With proportion the situation is similar, the variance of which is mainly a theoretical assumption. The situation is similar with proportion, the variance of it is
\[ \sigma_p^2 = \sum_{i=1}^{k} \frac{w_i^2}{n_i} p_i q_i = \frac{1}{n} \sum_{i=1}^{k} w_i p_i q_i. \]  

(15)

In cases of nonproportional allocation, the formula to calculate mean and the variance is even more complicated, but since all these formulae can be found in relevant literature (Cochran, 1977, Kish and Anderson, 1978), this paper won’t be needlessly burdened with addition of more and more new formulae.

Optimal allocation is a procedure for determining subsample sizes, which minimize total variance of mean when total sample size is given. For each stratum, subsample size is determined as

\[ n_i = n \frac{w_i \sigma_i}{\sum w_i \sigma_i} = n \frac{N_i \sigma_i}{\sum N_i \sigma_i}. \]  

(16)

This type of allocation is also called Neyman allocation, and it provides a stratified sample with the lowest variance. If we denote the variance of mean with \( \sigma_{sr}^2 \) in case of simple random sample, with \( \sigma_{pps}^2 \) the variance of mean in case of stratified sample with allocation with probabilities proportional to size (pps) and with \( \sigma_{opt}^2 \) the variance of mean in case of stratified sample with optimal (Neyman) allocation, then the following holds

\[ \sigma_{opt}^2 \leq \sigma_{pps}^2 \leq \sigma_{sr}^2. \]  

(17)

As Golder and Yeomans (1973) explain, contribution of proportional stratification in relationship with simple random sample can be measured accurately in the following manner

\[ \sigma_{sr}^2 = \sigma_{pps}^2 + \frac{\sigma_{bs}^2}{n}, \]  

(18)

where \( \sigma_{bs}^2 \) denotes the between-strata variance. Thus, the variance with proportional stratification is always \( \sigma_{bs}^2/n \) smaller than simple random sample variance.

Optimal allocation is optimal in the matter of the variance; however, in the matter of allocation, it contains a serious problem. In some cases (small stratum and large total sample size), optimal allocation could produce sample size for a stratum bigger than the population size in that stratum! Of course, such allocation is not possible, and in such cases 100% allocation is used in stratum.

Theoretically ideal variable for stratification is precisely the variable we are measuring, that is, the one of special interest for the survey, since stratification according to it would produce a very small variance. Of course, stratification according to that variable is not possible, since we have no previously knowledge about it. Because of this, it is common practice, when choosing among variables for which we have data on population, to pick for stratification those we assume to be in most ways related to the measuring subject. This introduces an amount of subjectivity into stratification process, but anyway, it would be impossible to avoid with sample design. Since a well chosen stratified variable significantly improves the quality of estimate, the choice of it made by an experienced researcher is completely justified.
3.2.2 Determining the Sample Size Based on Multiple Variables

As Žarkovic (1961) claims, the reality is even more complicated than the examples given in textbooks. In most surveys, there is more than one variable of interest for the study; also, another problem arises, the problem of determining optimal sample size that would take all of them into account. If all the variables had the same variance, and if that variance were known to us, there would be no problem and sample size could be determined in a very simple formula. However, variables that are of interest to us have various distributions and various variances. Applying the procedure as described above would be hard work and mostly a waste of time, because we would get a large number of different values for sample size, from extremely small to extremely large, and again we wouldn’t know which one to choose.

Common practice has found a compromise solution. In most cases, the first step is to narrow down the number of variables, based on which sample size is determined. It is necessary to choose several, not too many variables, which are of an utmost importance for the study. If their variances are similar, sample sizes attained for each of them will also be similar. It would suffice to choose one convenient number of a suitable representative size. This procedure provides that a small number of variables, but the ones of highest interest for the survey, get to be measured with the desired precision. The precision of other variables will be lower or higher than this one, depending on their variance.

However, if sample sizes that are determined by this narrower set of variables differ significantly from one another, we are forced to accept a conservative approach and take the highest value for the sample size, thus providing sufficient precision even for the variable with an extreme variance. Of course, in that case other variables will be measured with a higher precision than actually required. If this is the case, than we are in the position to drastically increase sample size because of one variable, which automatically increases the costs. In such cases perhaps it would be wise to reconsider the importance of that variable, and its role in the survey. Final decision cannot be made without consulting the ultimate user of the data.

One kind of sample that is commonly used in travel surveys is stratified cluster sample. Cluster sample basically increases intergroup correlation. Also, correlation between stratifying variable and the variable we are examining is larger in cluster sample than in simple random sample. Because of this, when using cluster sample, in order to avoid increase of the variance due to interclass correlation and design effect, the number of elements in one cluster is usually not too large.

With samples of households and persons, much fewer stratifying variables are available, than with the sample of administrative units, such as states, cities, municipalities, companies, and similar.

With most stratifications, though not for all of them, their primary goal is reduction of the total sample variance. A gain from stratifying variables depends on their correlation with the variable we are studying, as well as their interclass correlation. Perhaps it would be better for the discussion, and in relationship with introducing potential default variances and choice of typical variables to determine sample size, to be directed to the problem of improving stratification. Stratification deals with both issues and successfully unites them. Thus, it would be wise to see what the others have done in that area so far.

In solving problem of stratification, multivariate statistics has been used by a number of authors in different ways. The simplest, and probably because of that the first that has been used for these
purposes, is principal components analysis (PCA). Principal components analysis is applied on a set of stratifying variables, and then stratification is done according to the first, or according to the first and the second principal components. The procedure is simple and quite attractive, but has some shortcomings. Kish and Anderson (1978) name the following:

1. Principal components are not suitable for categorical (especially nominal) stratifying variables.
2. Quantitative stratifying variables that are not linear won’t contribute to variance reduction.
3. Strata obtained in this manner are difficult to interpret.
4. Method refers more to the internal structure of variables, and not so much on their effect on survey variables.

The same authors also state that bivariate stratification is always better than the stratification across only the first principal component.

Another interesting idea that has caused a lot of comments and discussions is stratification by cluster analysis (this shouldn’t be mistaken for cluster sample, these are completely different ideas, the only thing they have in common is the word cluster in their names). Golder and Yeomans (1973) have used cluster analysis for sample stratification for a travel survey on use of seat belts in cars. Taking a larger number of stratifying variables as a starting point, cluster analysis is used to reveal latent, otherwise undetectable clusters in population. Then the stratification is done across the clusters thus obtained. Unlike principal components, in this type of statistical analysis, the variables are not of primary importance, but the elements or objects. In some cases of cluster analysis use, it is possible to set in advance the number of clusters/strata, whereas in other cases this number doesn’t have to be fixed.

Heeler and Day (1975) do not criticise the idea of applying cluster analysis in stratification, but they criticise the choice of measure Golder and Yeomans used. Since it is generally known that cluster analysis results depend on choice of metrics, as well as on the choice of grouping algorithm, and both of them are usually determined subjectively, that is, based on researcher’s previous experience, clearly this way of stratification is extremely flexible, and hardly subjective to some serious standardisation. For each specific case of travel survey it is possible to find at least one suitable out of various possible metrics, which, if combined with appropriate grouping algorithm, could provide us with good results.

Interesting point of view about multivariate and multipurpose stratification was presented by Kish and Anderson (1978). Their work’s accent is on multipurpose stratification. Although they criticise both principal components analysis and cluster analysis, these two, and even some other multivariate statistics methods certainly deserve attention, when it comes to multivariate stratification. Probably, most stratification problems related to travel surveys are to some degree possible to standardize. Certainly, results of only one survey would not suffice for something like that. However, it would make sense to consider an idea of making such an attempt, based on same survey conducted in several countries. The same methods of multivariate statistics should be applied to the data collected in such a manner. Comparative analysis of the results of such a complex survey could provide us with an idea about a suitable stratification mode.

Another idea worth considering is applying other multivariate statistics methods in stratification. The usual situation in travel surveys is to have secondary data only on demographic characteristics of population, whereas data on travel variables stay unknown as a rule. Use of canonical correlation or discriminate analysis to the results of previous travel survey could solve the problem of
determining the set of demographic variables that are most closely related to travel variables, and could be used for stratification in some future travel survey.

In some cases of stratification, the primary goal is not the variance reduction, but to create domain for analysis, that is, appropriately defined subpopulations from which we can draw conclusions about the phenomena we are dealing with. This is especially the case with multivariate stratification. A larger number of stratifying variables are especially convenient with multipurpose surveys. Each of variables we are analysing has a different best stratifying variable, and stratification across several of them could be a good idea. Of course, with this multivariate stratification, there is always a limitation in the sense of the total numbers of strata allowed for the sample. Depending on the survey’s goal, type of population and the budget, total sample size could vary, but the number of strata even for the largest sample is seldom over ten, and often is much smaller.

As already mentioned, in sample design all the elements need not necessarily have the same probability of selection. Whether these probabilities are going to be the same or not often depends on the survey goal - whether we are interested in general population, or we would want to find out more about its mobile members, that is, about the ones that travel frequently and more often than the others. If, for example, the passengers’ sample we are making is based on a list of tickets some airplane company has sold during the previous year, probability is higher for passengers that have travelled more often and bought more tickets – more mobile passengers. If we do not want different choice probabilities, then we shall allocate equal probabilities to all elements in a posterior weighting process.

3.3 Choosing Respondents in Households

In some travel surveys, households are final sampling units. In such cases, information on the whole household and all household members are gathered (a sort of census on household level). However, in most cases, final sample element in travel surveys is an individual, and in such cases only one person should be chosen from a household. In travel surveys, as well as in other household based surveys of persons, two methods for respondents selection in households are most common: Kish tables and first birthday. Both methods, provided they are consistently conducted, generate uniform distribution, thus preserving population structure by sex and by age. Some researchers more prefer one of those methods, some the other, but the crucial factor of selection is not differences among results they produce; it’s the possibility to ensure their accurate realisation, that is, an estimation about what the interviewers will conduct more successfully on field, and where they will find weaker resistance by respondents. Years of experience of this author indicate that both methods give very similar results, and that, if consistently conducted, in neither case posterior weighting by sex or age is needed.

Hoffmeyer-Zlotnik (2003) describes and criticises the practise that has recently came out in Germany. A change is introduced In the last stage of probability sample – selection of the household member. Instead of traditional Kish tables or first birthday methods, interviewers get the list of households and quotas with detailed instructions how to folow them. When creating this new hybrid, the idea was to reduce interviewers’ field work, thus reducing total costs of survey realisation, and to achieve better results than with standard quota samples. Unfortunately, this last stage makes otherwise appropriate sample into a non-probability sample, for which error estimate is not possible. According to Hoffmeyer-Zlotnik (2003), this attempt didn’t succeed in removing flaws.
from standard quota samples, and samples realised in this way differ significantly in structure from population.

4. **SAMPLE DESIGN FOR BELGRADE 2002 SURVEY**

In autumn 2002, a very complex travel survey was conducted (Paskota, 2002). The survey goal was to collect sufficient amount of data that would help solving problems of public transport and city traffic in general. The whole survey was multipurpose and took place on multiple levels, and, because of that, a specially created composite measure instrument was used – a questionnaire with questions addressing different target groups. Apart from the demographic data, information was collected on habitation, motor and other vehicles ownership, parking problems, public transportation use, all daily trips in the city, attitudes on traffic and public transport etc.

The first group of questions was related to the household as a whole, the second was about all the household members age 6+, whereas the third group of questions were answered by only one chosen member of each household, age 15 or more. Collected data were analysed on 5 levels:

1. Households
2. All household members 6+
3. Daily trips
4. Daily trips’ segments (trip legs)
5. Chosen individuals 15+

For each level, a different sample was realized.

From common basis, different samples were designed for each of the three levels:

- On the household level, two-stage stratified household sample.
- On the individual level, for questions on daily trips, three-stage stratified cluster sample of persons.
- On the individual level, for questions on attitudes in traffic and public transport, three-stage stratified random sample of persons.

**Target population**

For the part of the survey about daily trips, the target population was defined as urban population of Belgrade, age 6 or more. For the part about attitude on traffic and public transport, target population was defined as urban population of Belgrade, age 15 or more.

**Population structure**

Population structure regarding age and sex was taken from the last population census, dated 31.03.2002.

**Frame**

Since a reliable list of all households was not available, the list of electoral units from the urban city territory was taken. Each electoral unit in the city consists of around 200 households or around 600 persons.

**Stratification**

Administrative municipal areas were used for stratification. Out of 10 municipalities belonging to city centre, 8 of them are inhabited with urban population exclusively, whereas in 2 municipalities a smaller part of population is rural. Purely urban population was observed, and for these two...
municipalities, strata were defined as their urban part. Allocation of the number of PSU across strata was proportional to their urban parts sizes. 

In order to reduce interclass correlation as much as possible, number of SU (households) in one PSU (electoral unit) was 10. The projected size was 2600 households or 260 PSU, but, based on previous experiences with difficulties in the field, the sample was projected with the size of 2750 (275 PSU).

Two-stage stratified household sample. 
Units of the first stage were the electoral units, defined by streets and house numbers they consist of. PSUs were chosen with probabilities proportional to the number of voters. Units of the second stage were households. In the second stage, systematic sample with given start, step and direction for each PSU was used.

Three-stage stratified cluster sample of persons (questions on daily trips). 
The first two stages are the same as with household sample. In the third phase, all household members satisfying the criteria of being over 6 years old are listed.

Three-stage stratified random sample of persons (questions on attitudes on city traffic and public transport). 
The first two stages were the same as with household sample. In the third stage, only one person age 15+ was selected from the household. For selection of respondents in households, Kish tables method was used.

Possible sources of error
Bias caused by non coverage could appear for two reasons: out of date electoral lists and the difference between voters and 6+ population. Lists of voters in Belgrade, in autumn of 2002, were significantly updated from that of two years earlier, many “dead souls” were excluded and new listed voters added. The biggest problem with this list wasn’t the lack of persons, it was the surplus. Certain number of people are on it because they are registered on their old addresses, and have left the country years ago (economic emigration from Serbia has been significant since the beginning of 1990’s). Since the emigrants were young people, mostly single or couples that didn’t own apartments of their own and lived with their parents, the assumption was that this hasn’t significantly affected the total number of households; but affected age structure of the central city parts. New age structure, changed in comparison with the 1991 census, was known from the results of the 2002 census. Another problem could have been significant only if there was a big difference in representation of the 6-18 years population in strata. Through control of age structure and household sizes in strata, it has been established that such differences really exist, but not to the extent of affecting the desicion about the choice of lists of electoral units for sample frame. Household size in municipalities varied form 2.24 to 2.71, not too big a deviation from the city mean, which is 2.62.

Field work
Field work was performed during a two-week period, Tuesday – Friday afternoons and evenings + the whole day Saturday. Questions about daily trips were always referring to the previous, necessarily working day. 105 interviewers took part, 10% of the sample was subject to random control, households with problematic or incomplete questionnaires were additionally controlled. Data entry lasted for 10 days, and 44 persons took part in doing it.
Realised sample size
2 650 households, 7 852 household members 6+, 16 181 daily trips, 28 272 trip legs, 2650 persons 15+

Poststratification
Realized sample slightly changed the sample size in strata; in poststratification process were used weights with value from 0.8 to 1.11.

Sample error estimate
Simple random sample error for proportion doesn’t go over 2% for households and 1.2% for all household members, and in most cases is even much lower. Values for the simple random sample estimated errors for different proportions are shown in Table 4.

<table>
<thead>
<tr>
<th>Proportion p</th>
<th>Sample size 2650</th>
<th>Sample size 7852</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1.943%</td>
<td>1.128%</td>
</tr>
<tr>
<td>0.2</td>
<td>1.554%</td>
<td>0.903%</td>
</tr>
<tr>
<td>0.05</td>
<td>0.847%</td>
<td>0.492%</td>
</tr>
</tbody>
</table>

Since we had administrative stratification, not according to some variable that is significant for traffic, that is, the sample error for those variables. For example, the number of households with at least one motor vehicle (whose proportion evaluated from the sample is 0.541 or 54.1%), in case of SRS sample size 2650 would have an error 1.94%, and the estimated error in the realized stratified sample for the same variable would be 1.92%. Therefore, with confidence higher than 95%, we could state that in urban parts of Belgrade, the number of households that own at least one motor vehicle equals 54±2%, that is, somewhere between 52% and 56%. From the example shown, we can see that in case of this survey, as upper error limit we can freely take values for SRS.

5. CONCLUSIONS
Sample error is historically the first error from various survey errors to be identified. Many authors that have dealt with this issue, and a solid statistical apparatus, have contributed that this sort of error is best explained and precisely quantified. For a number of years, the researchers have known that the main cause of error in surveys is not sample error at all. Problem of sample design is mainly solved in a satisfying, though not always standard manner. If in some survey sample error is extremely high, the reason for this shouldn’t be looked for in a lack of methodology, but rather in a lack of will to pursue it consistently in practice. Travel survey practice in most countries shows, that nonsample errors, primarily nonresponse error among them, are much more common source of errors. Although from year to year growing attention has been paid to them, it seems that, for a number of reasons, these errors have constant tendency to grow.

As for the travel survey sample design, standards are possible only for standardised survey. If we wish to conduct the same survey in a number of countries, gather same kind of data, use the same or similar questionnaire, and define population in the same manner, than it would be necessary to
consider standardisation of the sample design. Before that we have to be sure that we are measuring same things on same objects in the same way. Even if we use different types to collect same data (face to face, telephone or mail), this alone could cause problems.

Information that is collected in the course of most travel surveys is most commonly used for modelling and planning in traffic, and that is the reason why the precision of these data is of much greater importance than in case of other surveys. One should bear in mind that the error propagated through the process of modelling in traffic is the total survey error. A well projected sample with a low sampling error is necessary, but it still does not guarantee high data precision. For data precision, good quality sample realization is just as important. We shouldn’t forget that nonsampling error behave conversely from sampling error, it grows with the increase of sample. Because of this, an unnecessary sample increase, apart from the costs increase, could also lead to increase of total survey error.

In weighing all these reasons, a measure of subjectivity must be used. Many things are done subjectively in survey researches, and even in statistics. The very selection of the subject of measurement is subjective. The same is the case with maximum allowed error – some will think that 5% is quite all right, and the other will demand 3% or even lower error. Why is this? Did he read somewhere that this would be the desired error, did he see other people also use 5%? Does the one strictly demanding 1.5% sample error know how big are the other types of error he cannot measure, but are almost certainly higher than this percentage?

The usual phrase that is met in literature is that in numerous situations like this we rely on experts’ knowledge. Is that so bad? Why fear it?
REFERENCES


