10 Sampling design

After having defined research objectives, i.e. what information we need and how we obtain the information, the questions to be answered are:

- who provides the information, or in other words who is a respondent,
- where we find him/her,
- how we select the respondents
- how many of them we need to interview in order to get valid and reliable data.

The process associated with respondents' selection is called **sampling.** Sampling thus is an important tool of marketing research concerned with collecting data. Before describing a process of sampling, some of the terms used in sampling should be defined:

- **Population** is a group of people (e.g. customers, consumers, business partners) or objects (companies, shops, institutions) that have something in common and form the subject of the study in a particular research and can provide the information we want to obtain. Population has characteristics that can be estimated and classified according to the requirements of the research.
- **Census is** a procedure when all respondents (the whole population) related to the research topic are selected and interviewed. Census is expensive, rather slow and in case of marketing research rather rare. It can be used when population is small, e.g. in industrial research.
- **Sample** is a group of respondents drawn from a population and examined in some way. The information obtained from the sample is considered to represent the whole population. For this reason, the sample must be representative of the population so that valid conclusions about population can be inferred.
- **Elementary sampling unit** - an individual element of the population (a respondent) which is to be sampled.
- **Sampling frame** – a list or other records of a population from which a sample is to be selected, e.g. a list of all students registered at a school, a list of a company clients.

- **Statistic (estimator)** – refers to finding calculated from a sample to estimate a population parameter.
- **Parameter** - refers to the value of a variable calculated in the population, e.g. the average or a mean (see Figure 2).
- **Sampling error** – refers to the difference between a sample estimate and the value of the population parameter obtained by a complete count or census.

The theory of sampling is shown in Figure 1.

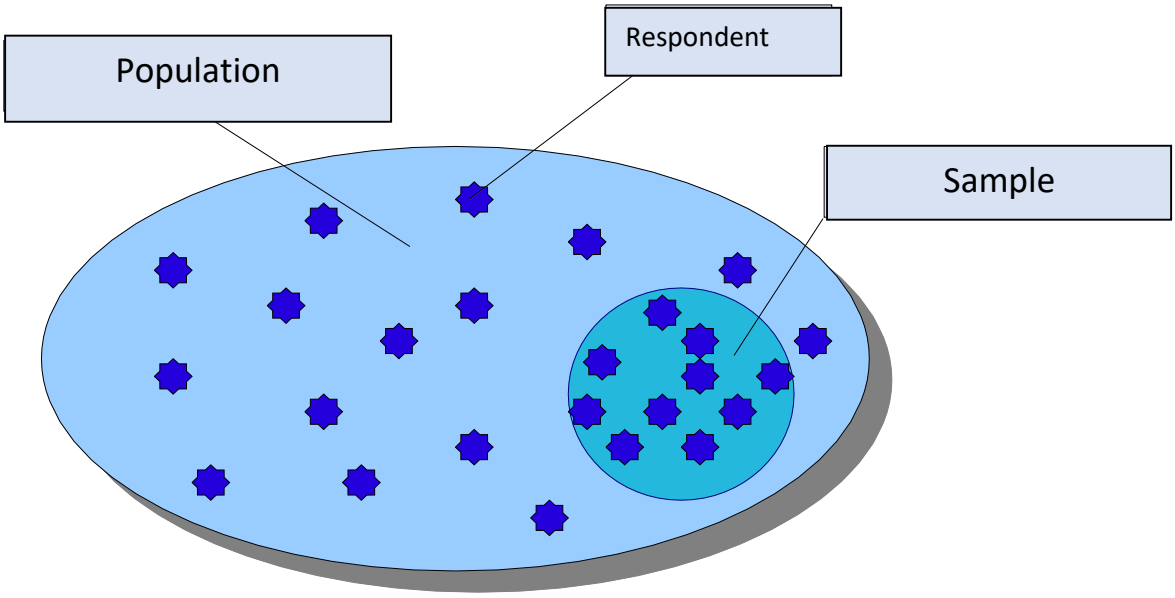


Figure 1 The theory of sampling

Source: Simova, 2010

Sampling theory is based on the study of relationship between a population and the sample selected from the population. Having collected data from the sample, certain conclusions can be drawn about the population from the sample. Findings (what we found from the sample) are estimations of population parameters and enables certain probability statements based on the sample. In other words, we try to generalize findings from the sample to the whole population being aware of a sampling error. Findings will always be estimations of the population parameters (Chisnall, p. 54-60).

The relationship between sample estimations and parameters of the population is shown in Figure 2.

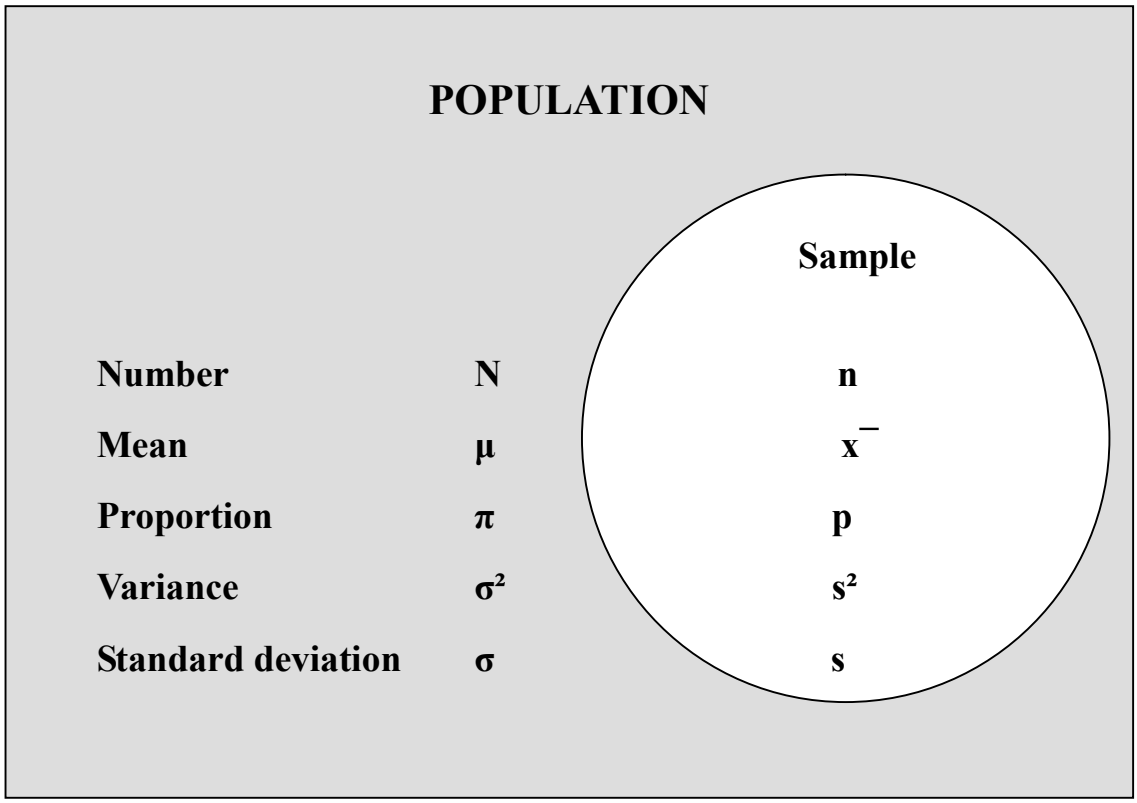


Figure 2 Relationship between sample estimators and parameters of the population

Source: Simova, 2010

10.1 Process of sampling

Sampling procedure is an important phase in marketing research as it considerably influences validity and reliability of data. The selection of sampling methods depends population that is a subject of the research.

The process of sampling refers to the following steps:

1. Definition of the population
2. Definition of the sampling frame
3. Selection of the sampling method
4. Defining a size of the sample
5. Sampling plan
6. Contacting and interviewing respondents

10.1.1 Definition of the population and a sampling frame

The objective of the definition of the population and a sampling frame is as follows:

- To state who is a respondent or an elementary sampling unit that posses the required information.
- To specify characteristics of the respondents in relation to the research objectives
- To define a place where we will contact the respondents
- To define the extent of the research geographically (Simova, 2010).

Example: The objective of the research is to measure students satisfaction with living facilities (dormitory). The information we want to find will be provided by students, However, the research focuses on satisfaction with living in a dormitory. This is why only students living there will be contacted, The place to contact students is the dormitory.

A sampling frame represents a list of all respondents to select for an interview in the research, Sampling frame is necessary to have when using methods of a probability sampling. The condition that has to be fulfilled is that all respondents have the same probability (chance) of being selecting for questioning, There is no need for sampling frame when using non-probability sampling methods,

10.1.2 Sampling method selection

The choice of a sampling method (methods of selecting respondents and defining a size of the sample) depends on objectives of the research, research design, respondents as well as on methods used for data collection, costs in time and money.

Basically, sampling methods can be classified into two categories:

- **probability sampling using statistical methods**
- **non-probability sampling using non-statistical methods.**

10.1.3 Sampling plan

Sampling plan is a plan of selecting respondents for an interview. The plan includes place where respondents will be selected and contacted and a time schedule (days and daily time) of data collection.

10.2 Types of sampling

10.2.1. Probability sampling

Probability sampling known also as random sampling is based on an assumption that all respondents have the same chance (probability) of being selected in the sample, which means that probability of a respondent being selected can be calculated. The sample is representative, i.e. it represents the population. It is true that the sample is just a small part of the whole population and cannot perfectly represent the whole population. However, methods of probability sampling allow to calculate how precisely the findings of the research estimate a characteristic of the population. The results obtained from respondents selected by random sampling are statistically sounder and it is also possible to calculate the standard error of the mean. The bias caused by interviewing only the most easily available respondents is avoided. However, there is a problem that limits the use of methods based on probability sampling – very few lists of respondents exist in marketing practise (Simova, 2010).

There are several types of probability sampling:

- Simple random sampling
- Quasi-random sampling
- Stratified random sampling
- Cluster sampling
- Multistage sampling.

Simple random sampling

It is the simplest method where respondents are selected randomly. A good example of the method is a „lottery drawing“. Every unit (respondent) of the population is identified by a number. The numbers are mixed and withdrawn until the required sample size is achieved.

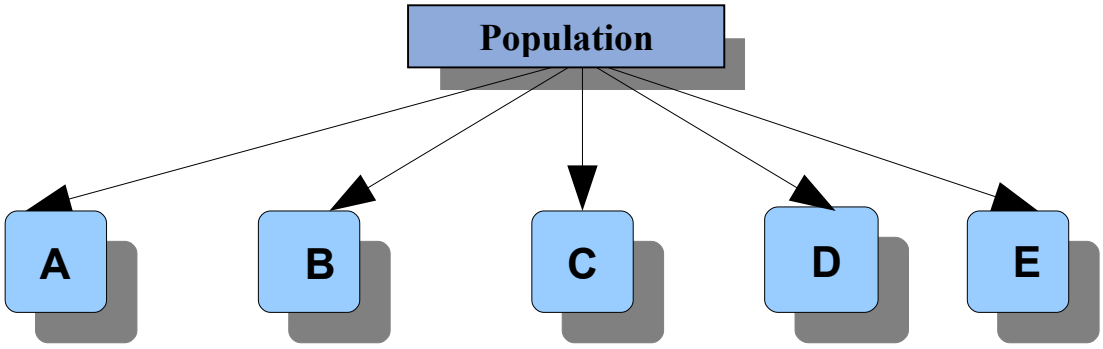
Quasi-random sampling

The method involves calculating the sample interval as the ration of the population to the sample (population/sample). A number from 1 to the sampling interval is drawn and used as a starting point for selection of respondents. Other respondents are chosen by adding the sampling interval to each succeeding number until the size of the sample is reached. The method is not truly random because respondents are not selected randomly as successive numbers are dependent on the pre-determined sampling interval.

Stratified random sampling

By this method, a population is divided into a groups (stratum) with similar characteristics which contributes to the accuracy of the sampling process. Stratification factors used to divide the population into groups should be related to the research. Quotas (number of selected respondents from each group) can be planned proportionally (representing the structure of the whole population) or disproportionally. Stratified random sampling tends to be more accurate than simple random sampling.

The principle of the stratified sampling is shown in Figure 3.



Legend:

A, B, C, D, E……. groups (strata) defined by stratified factors of the research. Respondents are selected from the groups randomly either proportionally or disproportionally.

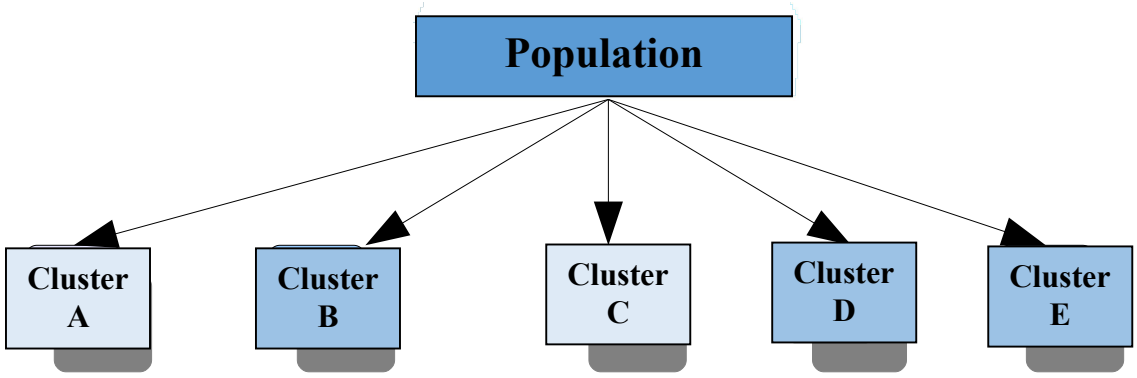
Figure 3 Stratified random sampling

Source: Simova, 2010

Cluster sampling

Respondents from the population are concentrated into a relatively small number of groups or clusters, which are selected at random. Within the selected clusters, every respondent is sampled and interviewed. This way of sampling is useful where the populations under the research are widely dispersed. The disadvantage is that it can miss complete section of the population or respondents within the cluster may not be homogenous.

Principles of cluster sampling are shown in Figure 4.



Legend:

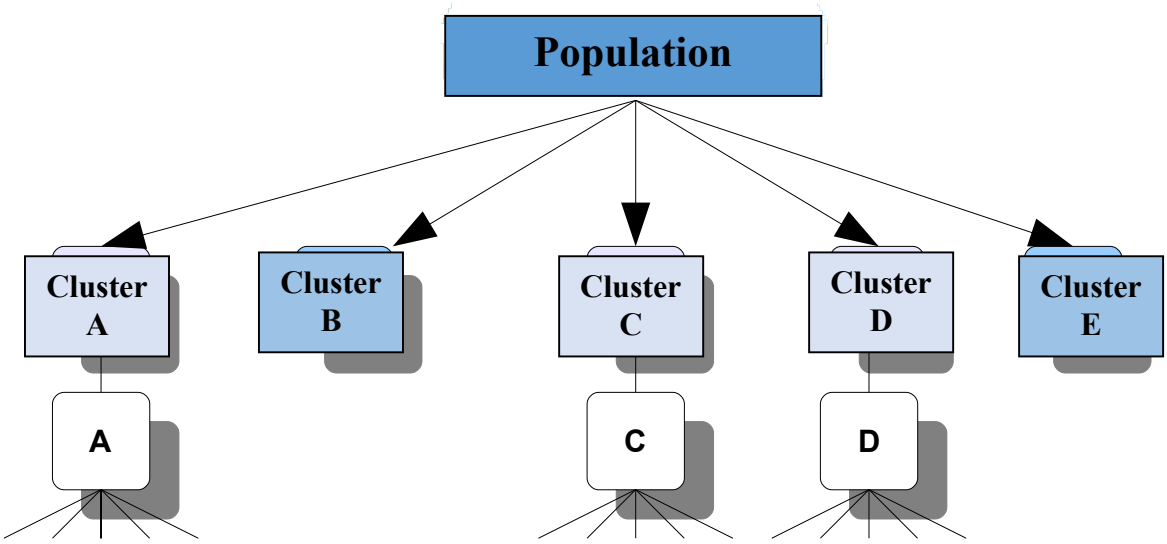
Clusters A, C …… randomly selected clusters. Respondents are selected randomly from the selected clusters or every respondent from the selected clusters is interviewed

Figure 4 Cluster sampling

Source: Simova, 2010

Multi-stage sampling

The process of selection of respondents is done at two or more successive stages. It is similar to cluster sampling. Multi-stage sampling occurs where there is subsampling within the clusters selected at the first stage. If clusters are individually sampled, cluster sampling take place. If a random sample within the clusters is used, this is a simple multi-stage sampling. The process of multi-stage sampling is shown in Figure 5.



Legend:

Clusters A, C, D …….. selected by random process. From each cluster a certain number of respondents is selected by random

Figure 5 Multi-stage sampling

Source: Simova, 2010

10.2.2. Non-probability sampling

Non-probability sampling (also referred to as judgement sampling) is a sampling where selection of the sample is dependent on human judgement and not on the rigorous application of probability theory. It is used more often since it does not require the sampling frame (a list of respondents). A sample is selected by needs of the research or by judgement, experience or opinion of experts (researchers). The weakness of this method is that representativeness of the sample cannot be guaranteed.

Methods based on non-probability sampling are as follows:

- quota sampling
- snowball sampling
- sampling by accessibility (availability) of respondents
- sampling by purpose.

Quota sampling

It is a sampling in which biases caused by non-probability method of respondent selection are eliminated to some extent by stratification and the setting quotas for each stratum.

The population is divided into groups or strata according to the requirements of the research, usually by demographics (age, gender and social class) or other classifications. Quotas are stated by researcher in order to match the structure of the population to get as representative sample as possible. Researchers must tend to achieve distribution over various subgroups in proportion to their importance in the total population. Actual respondents are selected by interviewers instead of using a probability process, which leads to bias. Despite its weaknesses, the quota sampling is widely and frequently used, it is cheap, fast, efficient, simple and convenient. However, it is not possible to estimate the sampling error and evaluate the reliability of estimates. There are also some limitations to statistical analysis. Some researchers argue that if the method is done properly and with control, it can bring fairly accurate results.

Snowball sampling

By this way of sampling, the first respondent is selected by researcher´s judgement, the other respondents are selected by the reference of the previous respondents. This method is suitable for a small population with specific characteristics (Simova, 2010).

10.3 Sample size

There are two ways of stating the size of the sample. The sample size can be computed or estimated. If the respondents are selected by using **statistical methods** based on random technique, the size of the sample can also be stated by applying statistical methods. Otherwise **non-statistical methods** based on judgement can be used.

10.3.1. Statistical methods used to state a sample size

The size of a sample depends on the characteristics of population, research objectives, type of collected information and the costs. The sample should be designed in a way to obtain the right quality and quantity of information. There is an opinion that a sample size should be a

certain percentage of the population (e.g. 10 percent). It is accepted that the larger the size of the sample, the greater its precision or reliability, but also a higher demand on time, staff and cost. The size of the sample is possible to mathematically calculate if random sampling is used in the research. Computing the size of the sample is rather a complex process. It can be done by using a standard error and required accuracy of the results. Standard error of the mean is actually the standard deviation of a sampling distribution. It is a difference between an estimation of a sample statistics $\overline{x}$ and a real value of the population parameter μ. The larger the sample, the smaller the standard error. If the standard error is too large, the sample estimate of the population mean will vary greatly.

The sample size depends on:

- variability of the explored characteristic, expressed usually by standard deviation
- size of the standard error (required accuracy of the results)
- reliability of the sample estimates.

Standard error of the mean in population can by calculated as:

$$SE_{(x)} = \frac{\sigma}{\sqrt{n}} \qquad (1)$$

where

$SE_{(x)}$ …… standard error of the mean
σ …… standard deviation
n ……. sample size (number of respondents)

The bigger the sample, the smaller standard error.

Equation for the size of the sample:

$$n = \frac{\sigma^2}{SE^2} \qquad (2)$$

Note: *If we do not know the variance σ^2 of the population, we have to estimate it from the variance s^2 of the sample. For this reason, we conduct a survey with a few respondents selected by random and compute the variance.*

Standard error of the proportion in population

The standard error of the proportion in a population can be obtained by formula:

$$SE_{(p)} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}} \qquad (3)$$

where

$SE_{(p)}$ …… standard error for estimation of sample proportion

p ... percentage (proportion) of a sample possessing attribute defined by research objectives

q ... percentage (proportion) of a sample who do not have the attribute under study

The relationship between p and q is as follows:

$$p + q = 1 \qquad \text{respectively} \quad p + q = 100\ \% \qquad (4)$$

The sample size can be computed by formula:

$$n = \frac{pq}{SE^2} = \frac{p(1-p)}{SE^2} \qquad (5)$$

Reliability and standard error we are willing to accept are important for stating the size of the population. Computation of the sample size is demonstrated by following examples (Simova, 2010).

In case we need to calculate a sample size for an estimation of the proportion (percentage) in population, the following formula can be used.

Example 1:

A travel agency would like to conduct a research trying to find how much money an average family spends annually for a holiday. From the research conducted a few years ago, it is known that on average a family paid CZK 3200 ± 750 a year (standard deviation was CZK 750.

We would like to make an estimation with 95 % confidence interval and a maximum standard error of CZK 50.

How big sample do we need to interview to reach the required accuracy and reliability?

It is known that:

$\bar{x} = 3200$ Kč, *s = 750 CZK, standard error = 50 CZK.*

$$n = \frac{\sigma^2}{SE^2}$$

$$n = \frac{750^2}{25^2} = \frac{562500}{625} = 900$$

To get results of the required accuracy and reliability, the sample should consist at least 900 respondents.

Note: *It should be pointed that in terms of the sampling theory and standard error concept, the normal curve or distribution is widely used in studying statistical data. If standard deviations (σ) are drawn at interval of one standard deviation from the mean, the proportion of the area under the curve will be as follows:*

- *Approximately 68 % of the units will be within ± 1* σ *of mean* μ
- *Approximately 95 % of the units will be within ± 2* σ *of mean*
- *Approximately 99 % of the units will be within* $\langle \mu - 3\sigma ; \mu + 3\sigma \rangle$. *(Simova, 2010)*

Having 95 % confidence interval (± 2 σ): 2*SE* = 50 CZK of which *SE* = 25 CZK.

For 99 % confidence interval (3*SE* = 50 CZK), the size of the sample would be:

$$n = \frac{750^2}{16{,}66^2} = 2025{,}13$$ i.e. approximately 2050 respondents.

Example 2

Foreign company entering the Czech market would like to make sure that having an advertising campaign would increase the share of people knowing its products. For this reason, it would like to conduct the research. Company knows that at this moment about 20 % of people knows its products. How many respondents should a company interview if the results should be at 95 % confidence interval and SE =± 4 %?

For 95 % confidence interval $2SE = 4$ %

$$n = \frac{20.80}{2^2} = 400$$

The sample should consist of 400 respondents.

For SE= ± 2 % ($2SE = 2$ %), the size of the sample would count:

$$n = \frac{20.80}{1^2} = 1600 \text{ respondent ů}.$$

Having 99 % confidence interval and SE= ± 4 % ($3SE = 4$ %) the size of the sample should be:

$$n = \frac{20.80}{1{,}3^2} = 947 \text{ respondent ů}.$$

10.3.2. Non-statistical methods used to state a sample size

In practise, non-statistical methods are very often used to state the size of the sample. The size of the sample can be stated by:

- judgement or experience of the researcher
- cost or budget we can afford or want to spend on the research
- the size of the sample of the similar research conducted in the past
- access to respondents (Simova, 2010).

Summary

The process associated with selection of respondents is called **sampling. Sampling procedure** is an important phase in marketing research as it considerably influences validity and reliability of data. The selection of sampling methods depends population that is a subject of the research. Two categories of sampling methods - probability sampling using statistical methods and non-probability sampling using non-statistical methods can be used in the research.

References:

1. CHISNALL, P. M.: Marketing Research. McGraw-Hill Book Company (UK) Ltd. 4th ed., 1992. ISBN 07-084010-5.
2. SIMOVÁ, J.: *Marketingový výzkum*, 2nd ed. Technická univerzita v Liberci, 2010. ISBN 978-80-7372-662-1.