STATISTICAL LANDMARKS AND PRACTICAL ISSUES REGARDING THE USE OF SIMPLE RANDOM SAMPLING IN MARKET RESEARCHES

CODRUȚA DURA, IMOLA DRIGĂ, DORINA NIȚĂ *

ABSTRACT: The sample represents a particular segment of the statistical population chosen to represent it as a whole. The representativeness of the sample determines the accuracy for estimations made on the basis of calculating the research indicators and the inferential statistics. The method of random sampling is part of probabilistic methods which can be used within marketing research and it is characterized by the fact that it imposes the requirement that each unit belonging to the statistical population should have an equal chance of being selected for the sampling process. When the simple random sampling is meant to be rigorously put into practice, it is recommended to use the technique of random number tables in order to configure the sample which will provide information that the marketer needs. The paper also details the practical procedure implemented in order to create a sample for a marketing research by generating random numbers using the facilities offered by Microsoft Excel.

KEY WORDS: sampling; random number table technique; repeated survey; unrepeated survey; the average value; the dispersion; the square average deviation; the confidence interval; estimation limit error

JEL CLASSIFICATION: M 31

1. INTRODUCTION

Sampling is the process of extracting a number of subsets from a general frame, in order to find out its characteristics. Through logical inference, one can establish general rules for the whole frame, rules which have not been verified directly, but they derived from the information obtained from the sampling. There are two major types of sampling which can be applied in marketing researches: probability and nonprobability sampling. In the case of probability sampling, the elements are selected

* Assoc.Prof., Ph.D., University of Petrosani, Romania, codrutadura@yahoo.com
Lecturer, Ph.D., University of Petrosani, Romania, imola.driga@gmail.com
Lecturer, Ph.D., University of Petrosani, Romania, dorinamagda@yahoo.com
randomly and the probability of including them in the sample is well known (this probability does not have to be equal for all the elements of the frame). Because the selection of elements in the case of nonprobability sampling is nonrandom, the chance that an element should enter the sample can then be estimated.

**The sample size** is the number of items from which information will be collected. The statistical methods for sizing a sample establish a set of calculations for each type of sample (with replacement or without replacement, stratified etc.). Therefore, in order to determine the sample size one should take into account a number of quantitative restrictions, such as: the maximum allowable error, the confidence level of the results, the dispersion of the analyzed characteristic of the frame. The higher the degree of accuracy required is (and hence the maximum allowable error is lower), the greater the sample size is, as the targeted confidence level is better and the dispersion is greater. At the same time, the analysis of large samples falsely increases - most of the times - research costs, or it is well-known that in almost all cases the recipients of information have limited budgets. That is why proper sample sizing in marketing research is a critical operation, knowing that a significant increase of the sample is not reflected in a commensurate improvement in the accuracy of the information obtained (an increase, for example, by four times, of the sample size will result in only halving the sampling error).

Specialized literature also presents cases where sample sizing is based on using **non-statistical methods** (Gherasim and Gherasim, 2003; Cătoiu and Bâlan, 2002). They are based on empirical reasoning such as: sizing in accordance with past practices, sizing based on typical sizes* or magic numbers (such as, for example 20, 50 or 100), sizing based on a priori stratification of the frame (in which case the aim is to represent each segment of the total population with a minimum of observation units, thus the sample size results from their sum); sizing by meeting budgetary restrictions (the sample size is given by the ratio of the volume of available funds and the amount required to investigate a subject); sizing according to expert advice, etc.. Nonstatistical methods for sample sizing have a high degree of subjectivity, which is why the only area where their application can be useful is that of qualitative marketing research.

As a result, the quantitative restrictions explained above are not the only elements involved in the decision regarding the sample size. To these we must add a set of qualitative factors, which, together with the adopted sizing method, interfere more or less in the sampling process, namely: the nature and the objectives of the research, the number of variables investigated, the analytical methods used, the volume of samples used in similar marketing studies, the level of aggregation of the analysis, the rate of completing the interviews, available resources, etc..

2. **BRIEF THEORETICAL CONSIDERATIONS OF THE PROCESSES OF APPLYING SIMPLE RANDOM SAMPLING**

The main feature of **simple random sampling** is that it requires that each subunit belonging to the general frame have equal chances of being selected in the sample. For this purpose, one can use the "lottery" selection process and the random number table process as means of selection. The "lottery" selection process or the
"drawing of lots" method implies numbering the general frame elements from 1 to N, and then making a note, card, ball, etc. which gives the opportunity to register each number. These objects are then inserted in an urn from which, after careful mixing, one must pick as many subsets as necessary to constitute the sample size. The elements which correspond to the numbers indicated by the selected articles will enter the sample.

Making a parallel with gambling such as Lotto games, the rules of which are well known, we must draw attention to the conditions that ensure that the balls have equal chances of getting out: they must be perfectly spherical, similar in terms of mass, volume, density, etc. In the case of large collectivities which are involved, most often, in the marketing field investigations, to implement such a procedure while maintaining equal probabilities for all elements is practically impossible. However, the method is frequently mentioned in both marketing and statistics textbooks because they represent in fact the type of simple sampling, providing a theoretical selection scheme according to which the probability theory is issued; under the circumstances one should determine the relations for calculating sample size, selection error, confidence interval.

When we want to rigorously implement the simple random sampling, it is more appropriate to use a second method of reproducing the quasi-perfect conditions, the mathematical requirements namely, the use of random number tables. This technique is based on long series of random numbers which are usually found as appendices in the books of statistics and marketing research or they are printed separately as brochures or books for those directly interested. Fortuity refers to each number separately (because the component figures are selected in a random manner) as well as to the sequence of the numbers contained in such a table. Of course, lately, since the computer has become an indispensable tool for planning, organizing and conducting marketing research, the search for tables with random numbers is not a problem anymore as the researcher himself masters the process of generating random numbers using a computer program.

The use of random number tables is relatively simple and it involves going through the following steps: first establish the survey which should include all units of observation that form the general frame. Each unit will be assigned a code number from 1 to N (where N is the size of the survey base); next, use the computer to select, at random, a number of codes equal to the size of the default sample; then, configure the sample by identifying the unit of observation (individual, family, business) corresponding to each code selected in the previous step.

When the researcher uses the computer to generate random numbers, the implementation of this procedure is simple as it is able to establish since the beginning the limits of the random numbers, taking into consideration the objectives of the research. For example, if a study is undertaken to find out the consumers’ attitude toward a new product and the sample consists of 30 subjects from the first group of 200 buyers, the operator should simply program a random selection of 30 numbers between 1 and 200.

The classical method of random number tables is more laborious because it means adapting the strings of numbers from statistical tables to the requirements of the research. Thus, if the size of the sample base is expressed by a number composed of
"k" digits and the table columns contain less digits than "k", the number of digits necessary will be completed from the next column: similarly, if the numbers in the table contain too many digits, then the first or last numbers in each column can be eliminated. Moreover, in the case of large samples composed reading may be used (it is usually obtained from two readings of the numbers in the table); the final number is formed of combinations of individual readings - two digits from the first reading, three from the second reading or other combination methods that match that certain research.

3. POSSIBLE SAMPLING SCHEMES AND STATISTICAL DETERMINATIONS

Whatever the actual procedure used, simple random selection can be performed as a repeated survey or as an unrepeated survey. The two sampling methods find a theoretical correspondence in the scheme known as the Bernoulli’s ball. Thus, if the observation units selected during the sampling process are returned to the collectivity (the principle of the returned ball) the survey is a repeated one (returned). As opposed to this, in case the selected units are removed from the sample (the unreturned ball principle), we are dealing with an unrepeated random sampling scheme (without return). The actual method of determining the sample size, as well as the relations for calculating relevant statistical indicators - average, dispersion, sampling error, etc. - will be detailed by reference to earlier typology. We must also add that there are two categories of characteristics (variables) which may be the subject of investigations in marketing researches: quantitative characteristics - measurable or numeric (such as the average time between two consecutive purchases, the frequency of visiting an exhibition stand and others) and qualitative or alternative characteristics which evaluate the attributes of some elements of the frame by making grouping them into a relatively small number of classes (consumer / non-consumers, people who prefer / reject a product, etc.). In the case of alternative characteristics there are several features related to the calculation of the sample size, of the dispersion and of the selection error; they are to be highlighted further on in the paper where there are made concrete references to the calculation of the indicators mentioned above.

1. The repeated survey. In this case, the analyzed unit reenters the general collectivity, thus ensuring the stability of distribution of the analyzed characteristic. In addition, the calculated statistical indicators may vary from one sample to another and therefore they can be treated as random variables and they can be analyzed using the methods available, for this purpose, from mathematical statistics. The general frame of size N (made up of consumers, users, distributors, voters, etc.) must be analyzed in terms of characteristic X which can take individual values \( \{x_1, x_2, ..., x_N\} \). A sample research involves getting the information from a number n subjects, which is often much smaller than the total population. The typicality of the sample n will depend on its size, which in turn is influenced by the dispersion of the characteristic studied.

Table 1 details the calculation of average and of the dispersion of the analyzed characteristic, both in the general frame and in the sample. The difference between the average of each sample and the real average (determined for the whole population) is called estimation limit error (E) and it actually represents the maximum allowable
error for a feature or an estimator, its size depending on both the size of the average
error of typicality and on the confidence level of the estimation. The average error of
typicality is no more than an error committed by a researcher when he chooses to
examine only a fraction of the general frame - n instead of considering all the N units
of the general collectivity. Most of the times, the general collectivity parameters
(average, dispersion, etc.) are unknown to the researcher. Therefore previous reasoning
should be transposed into probabilistic terms starting from the imaginary experience of
a subsequent extraction of a series of samples of volume n out of the total population
N. In this case, we can determine a selection scattering given by the dispersion of the
average of all volume samples n around the real average:

$$\sigma_{m^2} = \frac{\sigma^2}{n} \quad \text{or} \quad \sigma_m = \frac{\sigma}{\sqrt{n}} \quad (1)$$

where:

- $\sigma_{m^2}$ - the selection dispersion
- $\sigma^2$ - the dispersion of the average volume samples n
- $\sigma$ - the square average deviation recorded in the general collectivity

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{M})^2}{N}}$$

It is worth mentioning that $\sigma_m$ - the square average deviation of the selection is
frequently used as a measure unit of the average error of typicality.

In order to approximate $\sigma^2$ the dispersion corresponding to the general frame,
the researcher has several options (Prutianu et al., 2002):
- to use the results of a similar study conducted in a prior period of time (if
  available),
- if there are no such recent studies, a preliminary investigation will be conducted on
  a pilot sample established by a random method,
- if the maximum ($x_{\text{max}}$) and the minimum value ($x_{\text{min}}$) of the analyzed characteristic
  are known, then the relation $s \geq \frac{x_{\text{max}} - x_{\text{min}}}{6}$ leads to a rather good approximation
  of the square deviation.

Using one or another of the three processes, we obtain an estimator $s$ of the
dispersion of the characteristic which enables the approximation of the selection
dispersion $\sigma_{m^2}$ with the help of the following relations:

$$\sigma_{m^2} \approx \frac{s^2}{n} \quad \text{or} \quad \sigma_m = \frac{s}{\sqrt{n}} \quad (2)$$
where:
\[
\hat{s} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - m)^2}{n - 1}}
\]
\[
\hat{s} - \text{the constant square average deviation of the volume sample } x.
\]

Table 1. The calculation of the average and of the dispersion of the analyzed characteristic in the general frame and in the sample

<table>
<thead>
<tr>
<th>NUMERICAL CHARACTERISTIC</th>
<th>The general collectivity (N)</th>
<th>The sample (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average: ( M = \frac{\sum_{i=1}^{N} x_i}{N} )</td>
<td>Average: ( m = \frac{\sum_{i=1}^{n} x_i}{n} )</td>
<td></td>
</tr>
<tr>
<td>Dispersion: ( \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - M)^2}{N} )</td>
<td>Dispersion: ( s^2 = \frac{\sum_{i=1}^{n} (x_i - m)^2}{n} )</td>
<td></td>
</tr>
<tr>
<td>Square average deviation: ( \sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - M)^2}{N}} )</td>
<td>Square average deviation: ( s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - m)^2}{n}} )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALTERNATIVE CHARACTERISTIC</th>
<th>Average: ( \pi ) (the occurrence of state „i” within the total frame)</th>
<th>Average: ( p ) (the occurrence of state „i” within the sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion: ( \sigma^2 = (1 - \pi) )</td>
<td>Dispersion: ( s^2 = p \cdot (1 - p) )</td>
<td>Square average deviation: ( s = \sqrt{p(1 - p)} )</td>
</tr>
<tr>
<td>Square average deviation: ( \sigma = \sqrt{\pi(1 - \pi)} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the dispersion in the general collectivity is known (a situation rarely encountered in practice), the calculation of the average error of typicality will be based on this.

For alternative characteristics, the average error of typicality noted \( \sigma_p^2 \) is calculated using the same formula: \( \sigma_p^2 = \frac{\hat{s}^2}{\sqrt{n}} = \frac{p \cdot (1 - p)}{\sqrt{n}} \), that is to say
\[
\sigma_p = \sqrt{\frac{p \cdot (1 - p)}{n}}.
\]

For a given estimation error \( E \) we can determine a range of selection averages from the overall average with the help of which we can measure accuracy of the estimation: \( I = (M - E, M + E) \), where \( I \) represents the confidence interval. If we
represent graphically the distribution of the average volume samples \( n \) as a normal curve (Gauss-Laplace) - Figure 1, the confidence interval will be highlighted by the shaded surface area.

\[
\alpha / 2 \quad 1 - \alpha \quad \alpha / 2
\]

**Source:** Prutianu, Ștefan, Anastasiei, Bogdan și Jiție, Tudor (2005) Cercetarea de marketing. Studiul pieței pur și simplu. Iași: Editura Polirom

**Figure 1. Distribution of the average volume samples \( n \) as a normal curve**

The probability \( P(M - E < m < M + E) = 1 - \alpha \) is called confidence level and it reflects the safety with which it can be said that the average is inside the confidence interval. Its complement, \( \alpha \) is called the significance level or threshold and it corresponds to the probability that \( m \) is outside the confidence interval. Graphically, \( \alpha \) results immediately from the sum of the areas of the two surfaces below the Gaussian curve, within its ends.

The confidence level \( 1 - \alpha \) (and accordingly, the level of significance \( \alpha \)) are chosen taking into account the specific problem to solve. The levels of confidence most frequently used in practice are the following values 90%, 95%, 98%, 99%, 99.9% which correspond to the significance levels 10%, 5% 2% 1% and 0, 1%. For example, ensuring the results with a probability of 95% (-5% significance level) means that in only 5 cases out of 100 surveys, the sample average will be placed outside the confidence interval.

The estimate error \( (M - m) \) can be assessed using standardized normal variable values - \( z \) corresponding to the level of significance \( \alpha \). For this purpose the following relationship is used:

\[
E = z_\alpha \cdot \sigma_m = z_\alpha \cdot \frac{\hat{s}}{\sqrt{n}}
\]

where:

\( z \) - is the coefficient corresponding to the confidence level predetermined by the researcher.
The value of $z$ is taken from the relevant statistical tables. In some cases (the case of samples of small volume) $z$ – the argument of Gauss-Laplace function is replaced by $t$ – the argument of the Student function. The value of $t$ corresponding to the desired probability of guaranteeing the results of the research will be sought in this case, in the statistical tables of the Student distribution.

For an alternative feature, the formula for determining the estimation error is written, taking into account the actual way of expressing the square average deviation - $s$, as follows:

$$I = (\pi \pm E) = (0,4 \pm 0,05) = (0,35;0,45)$$ (4)

If it is considered a level of the limit error (E) set at the beginning of the research, one can obtain the required sample size by using the relationship:

- $n = \left(\frac{z_0 s}{E}\right)^2$ - for the normal characteristic;
- $n = \frac{z^2 \cdot p \cdot (1 - p)}{E^2}$ - for the alternative characteristic.

2). Unrepeated survey. This sampling scheme is designed to eliminate the distortions induced by the inclusion, for several times, of the same unit of observation in the selection frame. Therefore, the elements of the general frame are not re-included in the sample and the main consequence is that the selections of observation units within the sample are no longer equi-probable events. Indeed, the probability that the selection variable $x_1$ should take the actual value $x_1$ is $P(x_1 = x_1) = \frac{1}{N}$, but the probability of the next event ($x_2 = x_2$) will be conditioned by the fact that after the event $x_1 = x_1$, the observation unit $x_1$ was excluded from the frame. Thus, $P(x_2 = x_2 / x_1 = x_1) = \frac{1}{N - 1}$, that is to say that the probability of the second event depends on the previous one. In this case, it is shown that the dispersion of the selection average/criterion is given by (Baron et al., 1996):

$$\sigma^2_m = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1}$$ (5)

and the square average deviation as a measurement for the average selection error is:

$$\sigma_m = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N - n}{N - 1}} \approx \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$$ (6)

When the volume of the frame is large as compared to the sample size, $(N-1)$ is approximated with $N$ and the previous relationship becomes:
\[
\sigma = \frac{s}{\sqrt{n}} \left( \sqrt{1 - \frac{N}{n}} \right)
\]

(7)

where: \( k = \left( \sqrt{1 - \frac{N}{n}} \right) \) - is called the correction factor.

Therefore there is a relation between the appropriate selection dispersions of the repeated survey and that of the unrepeated survey:

\[
\frac{\sigma^2_{n, rep}}{\sigma^2_{n, nerep}} = \frac{N - 1}{N - n} \geq 1 - \frac{N}{n}
\]

(8)

where:

\( \sigma^2_{n, rep} \) - the dispersion of the average volume sample \( n \) around the real average, in the case of the repeated survey

\( \sigma^2_{n, nerep} \) - the dispersion of the average volume sample \( n \) around the real average, in the case of the unrepeated survey.

Knowing that the ratio \( \sqrt{\frac{N - 1}{N - n}} \) is always a proper fraction it is clear that the estimation error of the unrepeated survey will always be smaller than the error associated with repeated survey.

Since \( E = z_\alpha \cdot \sigma_m \), in the case of the unrepeated sampling one can write:

\[
E = z_\alpha \cdot \frac{\sigma_m}{\sqrt{n}} \cdot \sqrt{\frac{N - n}{N - 1}}
\]

(9)

If we consider that \( s^2 \) is a good estimator of \( \sigma^2_m \), the previous relationship becomes:

\[
E = z_\alpha \cdot \sqrt{\frac{s^2}{n}} \cdot \frac{N - n}{N - 1}
\]

(10)

In order to determine the volume of the sample \( n \), one uses consecutive deductions:

\[
E^2 \cdot n \cdot (N - 1) = z^2_\alpha \cdot s^2 \cdot N - z^2_\alpha \cdot s^2 \cdot n
\]

(11)

where:

\[
n \left[ (N - 1) \cdot E^2 + z^2_\alpha \cdot s^2 \right] = z^2_\alpha \cdot s^2 \cdot N\]
\[ n = \frac{Z_\alpha^2 \cdot s^2}{(N - 1) \cdot E^2 + Z_\alpha^2 \cdot s^2} \]

Generally, the \( \frac{N}{N - 1} \) ratio is used when operating with large collectivities and the relationship to calculate the sample size becomes:

\[ n = \frac{Z_\alpha^2 \cdot s^2}{E^2 + \frac{Z_\alpha^2 \cdot s^2}{N}} \tag{12} \]

Of course, the previous formula is valid only for numeric characteristics. For alternative characteristics, we must take into account \( s^2 = p \cdot (1 - p) \) and the sample size will be given by:

\[ n = \frac{Z_\alpha^2 \cdot p \cdot (1 - p)}{E^2 + \frac{Z_\alpha^2 \cdot p \cdot (1 - p)}{N}} \tag{13} \]

In the case of small samples (with less than 30 units of observation), in all previous calculations the value \( Z_\alpha \) is replaced with \( t \) corresponding to the Student distribution.

A further observation is also required in this case: if the volume of the general collectivity \( N \) is large and the size of the sample \( n \) is comparatively small, then the ratio \( \frac{N - n}{N - 1} \rightarrow 1 \) and the results estimating the average dispersion of the selection for the unrepeated survey are practically identical with those of the repeated survey. Similarly, when calculating the sample size for the unrepeated survey

\[ n = \frac{Z_\alpha^2 \cdot s^2}{E^2 + \frac{Z_\alpha^2 \cdot s^2}{N}} \]

the ratio \( \frac{Z_\alpha^2 \cdot s^2}{N} \rightarrow 0 \) for higher values of \( N \) and again, get to

\[ n = \frac{Z_\alpha^2 \cdot s^2}{E^2} \]

Therefore, the differences between the results obtained in situations when there was applied a repeated sampling procedure and an unrepeated sampling procedure, become noticeable in case the ratio of the sample volume and of the general frame
volume is high enough (generally, it is considered that a value of the ratio \( \frac{n}{N} \) greater than 0.2 enables to highlight that certain distinction).

For example, we consider that a business manager, with 150 directly productive workers involved in the production of electric motors, wishes to estimate the unused productive time by its employees in order to adopt efficiency measures. A study conducted earlier showed an average daily unused productive time of 50 minutes per employee, with a mean square deviation of 10 minutes. The maximum accepted error is 3.5 minutes and the desired confidence level is 95% (\( z = 1.96 \)).

In case of a repeated survey, the sample is determined using the following statistical formula:

\[
 n = \left( \frac{z \cdot s}{E} \right)^2 = \left( \frac{1.96 \cdot 10}{3.5} \right)^2 \approx 31 \text{ (individuals)}. 
\]

In case of an unrepeated survey, the sample is determined based on the following formula:

\[
 n = \frac{Z^2 \cdot s^2}{E^2 + \frac{Z^2 \cdot s^2}{N}} = \frac{196^2 \cdot 10^2}{3.5^2 + \frac{196^2 \cdot 10^2}{150}} = 25.93 \approx 26 \text{ (individuals)}. 
\]

Since it has been shown that a survey without re-sampling leads to forecasts with higher accuracy, it is sufficient to make observations on a smaller sample, consisting of 26 people.

The confidence interval of the estimation is:

\[
 I = (50 \pm E) = (50 \pm 3.5) = (46.5; 53.5) 
\]

Therefore, we can state with a probability of 95% that the average daily unused productive time per worker is between 46.5 and 53.5 minutes.

The following stages must be covered in order to create the sampling for marketing research:

1. drawing up a list of directly productive 150 workers (their names are arranged in alphabetical order), assigning a code between 1 and 150 for each name on the list – the first worker will receive the code 001, the second will get 002 and so forth;

2. drawing up a random number table using the RAND function in Microsoft Excel program. The function has the following syntax = RAND () and it generates a real random number between 0 and 1. To build a random integer number table between 1 and 150 (suitable for marketing research undertaken for the problem to be solved), a combination of functions can be used: = Roundup (150 * RAND (), 0). The Roundup Function is used to round a number through addition and it has two arguments: the real number to be rounded (in our case 150 * RAND ()), and the number of decimal places up to which the random number is to be rounded. Since we determined that the second argument of the Roundup function equals 0, the result of each cell of the random number table created in Excel will be an integer number;

3. further on, we shall use the INDEX function to perform a random selection of 26 codes in the table with random numbers, which was drawn up in the previous step. The INDEX function returns the value of an item from a table or from a matrix selected by a number indicating the row and column. More precisely, the syntax is:
= INDEX (array, row number, column number), where the matrix argument represents a cell zone in Excel (in our case the area is bounded by the table of random numbers from figure 2). We used the compound function: INDEX ($A$1:$T$30;ROUNDUP(30*RAND();0);ROUNDUP(20*RAND();0)); the Excel program was ordered to choose at random a line number and a column number out of the table (matrix) with 30 rows and 30 columns in order to indicate the position in the table of the number that is to be selected in the sample. If by any chance, the computer would draw a number that would repeat itself among the 26 that are to form the sample of employees, that certain number would have been skipped, thus, continuing the selection procedure of random numbers until the desired sample would be obtained. By implementing the compound function mentioned above, we can obtain the following code numbers associated with persons who will form the sample under research: 146, 82, 127, 90, 33, 21, 98, 142, 41, 109, 55, 128, 10, 49, 50, 28, 96, 4, 75, 134, 64, 68, 88, 125, 2 and 80;
4. the sample is completed after each code is assigned the name and surname of the employee who fills out the marketing research enquiry.

Figure 2. Using Microsoft Excel functions to create samples for research
4. CONCLUSIONS

The need for information and the efficiency with which it has to be obtained and analyzed have established that the selective approach is a quasi-general approach used in marketing researches. Selection gives the opportunity to obtain information on a general community by investigating only some of the components within. Among the arguments which recommend the selective research instead of the total research (such as the population census) we should include substantial cost savings, time, human resources, as well as the advantages of knowing and applying the information obtained in order to intervene in the economic life. Therefore, the choice and the application of an appropriate sampling method for the objectives of each research project are the key elements that make a success out of a selective survey.

Using statistical survey for market analysis, as well as for other research domains, is due to the fact that sampling theory is based on the law of large numbers. This statistical rule asserts, with a sizeable probability (which is closer to one), that statistical indicators which characterize the sample are very few different from statistical indicators belonging to the statistical population, provided that the sample is large enough. Marketing and statistical literature available to practitioners provides a wide range of sampling methods that can be implemented in the context of marketing research. If in case of probabilistic methods the calculation of sampling error is possible, in case of non-probabilistic methods they remain unknown. In order to choose between a probabilistic or non-probabilistic sampling technique it should be taken into consideration if a random procedure provides higher value information than a non-probabilistic one, at a certain level of cost. This decision is taken according to: costs, nature of information to be obtained (in case of generalizing the results to the entire population), desired accuracy of estimation, estimated effect of sampling error on results, homogeneity of population. Despite of relatively high costs involved, the probabilistic model remains one of the most rigorous designed research models for both macroeconomic phenomena and for microeconomic level: attitudes, opinions and behaviours of consumers, operators or managers.

REFERENCES:


