THE USE OF RANKING SAMPLING METHOD WITHIN MARKETING RESEARCH

CODRUȚA DURA, IMOLA DRIGĂ *

ABSTRACT: *Marketing and statistical literature available to practitioners provides a wide range of sampling methods that can be implemented in the context of marketing research. Ranking sampling method is based on taking apart the general population into several strata, namely into several subdivisions which are relatively homogenous regarding a certain characteristic. In fact, the sample will be composed by selecting, from each stratum, a certain number of components (which can be proportional or non-proportional to the size of the stratum) until the pre-established volume of the sample is reached. Using ranking sampling within marketing research requires the determination of some relevant statistical indicators - average, dispersion, sampling error etc. To that end, the paper contains a case study which illustrates the actual approach used in order to apply the ranking sample method within a marketing research made by a company which provides Internet connection services, on a particular category of customers – small and medium enterprises.*

KEY WORDS: *market analysis; ranking sampling; proportional survey; non-proportional survey.*

JEL CLASIFICATION: *C83, M31.*

1. INTRODUCTION

Needs of rapidly growing information and the efficiency with which it must be obtained and analyzed have defined selective research as a quasi-general approach used in marketing studies. Selection offers the opportunity to obtain information relating to a general group by investigating only some of the components of the latter. Therefore, choosing and implementing an appropriate sampling technique, in relation to the objectives of each research project, becomes the key element that makes a selective research successful. The arguments which recommend selective research

* *Assoc. Prof., Ph.D., University of Petrosani, Romania, codrutadura@yahoo.com*
Lecturer, Ph.D., University of Petrosani, Romania, imola.driga@gmail.com

instead of the total research (like the population census) include substantial cost savings, time savings, human resource savings and other benefits for those who use the information obtained with the purpose to intervene in the economic life.

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 (Cătoiu, 2002).

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 (Dura, et al., 2010).

2. RANKING SAMPLING METHOD

Ranking sampling is based on grouping the investigated collectivity into subpopulations (stratum), i.e. into relatively homogeneous subgroups according to certain characteristics. The studied sample will be formed by random selection from each stratum, of a proportional number of elements, in order to achieve the predetermined volume of sample.

The use of ranking sampling method implies the following steps (Cătoiu, 2002):

- creating the sample frame and identifying the sampling criteria for the general collectivity;
- weighting and determining the size of each strata within the final sample in relation to the sampling scheme used (proportionate or disproportionate);
- random extraction, from each stratum, of a predetermined number of observation units according to the appropriately calculated volume of the sample.

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 (table 1); they are to be highlighted further on in the paper where there are made concrete references to the calculation of the indicators mentioned above.

Table 1. The calculation of the average value and of the dispersion for the analyzed characteristic in the general collectivity and in the sample

<table>
<tr><td colspan="2">NUMERICAL CHARACTERISTIC</td></tr>
<tr><td>The general collectivity (N)</td><td>The sample (n)</td></tr>
<tr><td>The average value: $M = \frac{\sum_{i=1}^{N} x_i}{N}$
Dispersion: $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - M)^2}{N}$
Mean square deviation: $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - M)^2}{N}}$</td><td>The average value: $m = \frac{\sum_{i=1}^{n} x_i}{n}$
Dispersion: $s^2 = \frac{\sum_{i=1}^{n}(x_i - m)^2}{n}$
Mean square deviation : $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m)^2}{n}}$</td></tr>
<tr><td colspan="2">ALTERNATIVE CHARACTERISTIC</td></tr>
<tr><td>The average value: π (the occurrence of state „yes” within the general collectivity)
Dispersion: $\sigma^2 = (1 - \pi)$
Mean square deviation : $\sigma = \sqrt{\pi(1-\pi)}$</td><td>Average: p (the occurrence of state „yes” within the sample)
Dispersion: $s^2 = p \cdot (1-p)$
Mean square deviation : $s = \sqrt{p(1-p)}$</td></tr>
</table>

The general population of size N (including consumers, users, distributors, voters etc.) must be analyzed according to characteristic *x* which can take individual values $\{x_1, x_2, ..., x_N\}$.

A sample research involves collecting necessary information from a number *n* of subjects which, most often is much smaller than the total population. The representativeness of sample *n* will depend on its size which in its turn is influenced by the dispersion of the characteristic studied.

Table 1 details the method of calculating the average and the dispersion of the characteristic studied, both in the case of the general collectivity and in the sample (Şerban, 2004).

The difference between the average of each sample and the real average (as determined for the entire population) is called estimation error limit (E) and it actually represents the maximum permissible error for a characteristic or an estimator, its size depending on both the size of the representativeness average error (σ_m) and the confidence of the forecasts. The average error of representativeness is nothing but an error committed when the researcher instead of considering all the N units of the general collectivity, investigates only a fraction of it - *n*.

In most cases, the parameters of the general collectivity (average dispersion etc.) are unknown to the researcher. Therefore, earlier judgments must be translated into probabilistic terms starting from an imaginary experience of a consecutive extraction of a series of samples of volume *n* from the total population N. In this case, we can determine *the selection dispersion* given by the average dispersion of each sample of volume *n* around the real average:

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

where:

σ_m^2 - the selection dispersion

σ^2 - the average dispersion of samples of volume *n*

σ - the mean square deviation of the general collectivity.

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

It is worth mentioning that σ_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 σ^2 the dispersion corresponding to the general collectivity, 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_{max}) and the minimum value (x_{min}) of the analyzed characteristic are known, then the relation $s \cong \frac{x_{\max} - x_{\min}}{6}$ leads to a rather good approximation of the square deviation.

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

$$\sigma_m^2 \cong \frac{\hat{s}^2}{n} \text{ or } \sigma_m = \frac{\hat{s}}{\sqrt{n}} \quad (3)$$

where:

$$\hat{s} = \sqrt{\hat{s}^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m)^2}{n}} \quad (4)$$

$\hat{s}$ - the constant square average deviation of the volume sample *n*.

For alternative characteristics, the average error of typicality noted σ_p^2 is calculated using the same formula:

$$\sigma_p^2 = \frac{\hat{s}^2}{\sqrt{n}} = \frac{p \cdot (1-p)}{\sqrt{n}} \quad (5)$$

$$\text{meaning } \sigma_p = \sqrt{\frac{p \cdot (1-p)}{n}} \quad (6)$$

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.

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, α is called *the significance level or threshold* and it corresponds to the probability that *m* is outside the confidence interval. Graphically, α results immediately from the sum of the areas of the two surfaces below the Gaussian curve, within its ends.

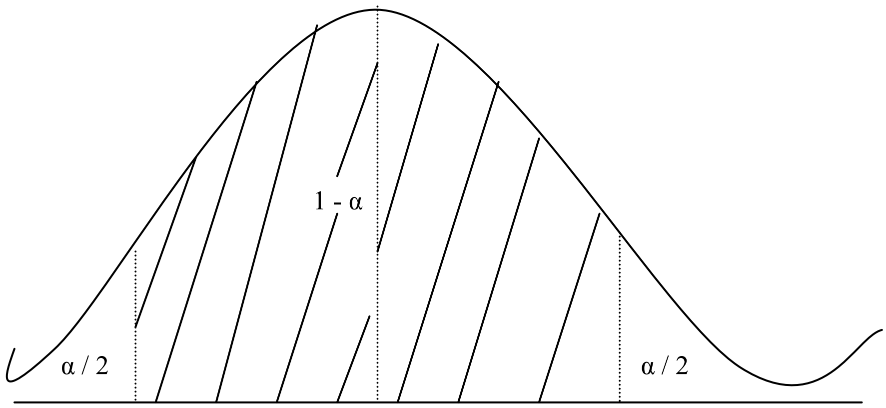
The confidence level 1 - α (and accordingly, the level of significance α) are chosen taking into account the specific problem to solve. The levels of confidence most frequently used in practice are the following values 90%, 95%, 98%, 99%, 99.9% which correspond to the significance levels 10%, 5% 2% 1% and 0, 1%.

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

$$E = z_\alpha \cdot \sigma_m = z_\alpha \cdot \frac{\hat{s}}{\sqrt{n}} \quad (7)$$

where:

z - is the coefficient corresponding to the confidence level predetermined by the researcher



Source: Prutianu, Ştefan, Anastasiei, Bogdan şi Jijie, Tudor (2005) Cercetarea de marketing. Studiul pieţei pur şi simplu, Iaşi, Polirom Publishing House

Figure 1. The distribution of the average values of *n* size samples by Gauss curve

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 characteristic, the formula for determining the estimation error is written, taking into account the actual way of expressing the mean square deviation - s, as follows:

$$E = \frac{z_\alpha^2 \cdot p \cdot (1-p)}{n^2} \quad (8)$$

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:

$$n = \left(\frac{z_\alpha \cdot s}{E}\right)^2 \text{ - for the numerical characteristic} \quad (9)$$

$$n = \frac{z_\alpha^2 \cdot p \cdot (1-p)}{E^2} \text{ - for the alternative characteristic} \quad (10)$$

3. APPLYING THE RANKING SAMPLING METHOD WITHIN MARKETING RESEARCH – A CASE STUDY

Managers of “PC Market”, a company which provides Internet connection services, demanded a marketing research on a particular category of customers - SMEs operating in the region. The objective of the research is to estimate the average monthly turnover of a SME, thus gathering the necessary information to substantiate the price policy within "PC Market". The error limit accepted by the beneficiaries of the research is 120 lei, and the confidence level desired is of 95.44%. The first stage of the research involves carrying out a survey on a control-sample consisting of 36 enterprises in order to estimate the average value and the dispersion of the turnover in the total collectivity (consisting of about 6000 active SMEs within the area). The information provided by the preliminary survey is summarized in Table 2.

Table 2. Information provided by the preliminary survey

Domain of activities	Number of SMEs	The Average Turnover (lei)	Mean Square Deviation (lei)
Production	1000	20000	1000
Comerce	3000	14000	1500
Services	2000	12000	1700
Total	6000	-	-

First, we assume that the survey is **proportional**. The proportional sampling is characterized by the fact that the number of observation units extracted from each stratum is appropriately determined so that the final sample structure should reproduce exactly the general collectivity structure in relation to the criteria used during the analysis.

In other words, the weight of each stratum *i* (k_i) is determined according to the report below:

$$k_i = \frac{N_i}{N} \tag{9}$$

where:
N_i – represents the number of elements from stratum “*i*” of the general collectivity;
N – the size of the general collectivity.

Starting from the general selection proportion (k_s):

$$k_s = \frac{n}{N}$$

(10)

the number of elements of the same type *i* (n_i), which will be taken in the sample, is determined based on the following relation:

$$n_i = k_i \cdot n \text{ or } n_i = \frac{n}{N} \cdot N_i \tag{11}$$

The corresponding weights of the three strata in each sample (enterprises within the productive sector, service firms and commercial companies) will meet the general collectivity structure:

$$k_1 = \frac{1000}{6000} \approx 0,17$$
$$k_2 = \frac{3000}{6000} = 0,50$$
$$k_2 = \frac{2000}{6000} = 0,33$$

The average value of the sample obtained through stratified sampling – *m* results from the following relation (Şerban, 2004):

$$m = \sum_{i=1}^{n} k_i \cdot m_i \qquad (12)$$

where:
m_i – the average value of the analyzed characteristic of the stratum *i*;
n – the number of strata corresponding to the general collectivity.

The average turnover for SMEs within the region can be determined using the relation:
m = 0,17 x 20.000 + 0,50 x 14.000 + 0,33 x 12.000 = 14.360 (lei)

The mean square deviation of the sample - *s* is determined as follows:

$$s = \sqrt{\sum_{i=1}^{n} k_i \cdot s_i^2} \qquad (13)$$

where:
s_i - the mean square deviation of the analyzed characteristic of stratum *i*.

In the case of "PC Market", the mean square deviation of the turnover results from the relation:

$$s = \sqrt{\sum_{i=1}^{3} k_i \cdot s_i^2} =$$
$$= \sqrt{0,17 \cdot 1000^2 + 0,50 \cdot 1500^2 + 0,33 \cdot 1700^2} \approx 1500(lei)$$

We now have the necessary information to calculate *the confidence interval* of the estimation:

$$I = (m \pm E) = (m \pm \frac{z_\alpha \cdot s}{\sqrt{n}})$$

$$I = (14360 \pm 2 \cdot \frac{1500}{\sqrt{36}}) = (14360 \pm 500) = (13860;\ 14860)\ (\text{lei})$$

As a consequence, the turnover falls between 13860 and 14860 lei per month, with a probability of 95.44%. This precision provided by the pilot-sample is not satisfactory for the managers of the company, because the error of the turnover estimation ($E = \frac{z_{\alpha} \cdot s}{\sqrt{n}} = 500$ lei) is four times greater than desired.

The sample size, which might ensure an estimation limit error of 120 lei results from:

$$n = \left(\frac{z \cdot s}{E}\right)^2 = \left(\frac{2 \cdot 1500}{120}\right)^2 = 625\ \ (\text{enterprises})$$

If the permissible limit error was of 60 de lei, the sample would be of size:

$$n = \left(\frac{2 \cdot 1500}{60}\right)^2 = 2500\ (\text{enterprises})$$

Therefore, improving estimation accuracy by reducing by half the admissible limit error requires increasing by four times the number of companies taken in the sample.

The volume of the proportional sample is calculated using the same formula:

$$n = \left(\frac{z \cdot s}{E}\right)^2 = \frac{z^2}{E}\left(\sum_{k=1}^{n} k_i \cdot s_i^2\right) \tag{14}$$

$$n = \frac{2^2}{120^2}\left(0,17 \cdot 1000^2 + 0,50 \cdot 1500^2 + 0,33 \cdot 1700^2\right) =$$
$$= 624,63 \approx 625\ (\text{enterprises})$$

The sample size that provides an estimation of the monthly turnover average with a maximum permissible error of 120 lei and a confidence level of 95.44% is of 625 enterprises, and it is equal to the unstratified sample. However, the elements of the sample will be selected in a way that the structure of the general collectivity should be reproduced exactly. The size of each stratum results from the relationship:

$$n_i = k_i \cdot n$$

$n_1 = 0,17 \times 625 \approx 106$ (enterprises)
$n_2 = 0,50 \times 625 \approx 313$ (enterprises)
$n_3 = 0,33 \times 625 \approx 206$ (enterprises)

The research sample includes 106 companies from the productive sector, 313 commercial enterprises and 206 enterprises specialized in services.

In the case of **non-proportional survey**, the total population structure cannot be kept in the sample composition, either because the dispersion of the analyzed characteristic is different from one stratum to another, or because some strata are more important for the decision makers than others. Regardless of the situation, the non-proportional ranking procedure leads to estimates very close to reality, while research costs are generally low.

The weight of a stratum in the sample (k_i) is determined by the dispersion of the characteristic of that stratum:

$$k_i^{'} = \frac{N_i \cdot \sigma_i}{\sum_{i=1}^{n} N_i \cdot \sigma_i} \approx \frac{N_i \cdot \hat{s}_i}{\sum_{i=1}^{n} N_i \cdot \hat{s}_i} \quad (15)$$

where:
σ_i – the mean square deviation from stratum i;
$\hat{s}_i$ – the mean square deviation estimator corresponding to stratum *i*.

The number of elements forming stratum *i* ($n_i^{'}$) from the final sample can be calculated as the product of:

$$n_i^{'} = k_i^{'} \cdot n = \frac{N_i \cdot s_i}{\sum_{i=1}^{n} N_i \cdot s_i} \cdot n \quad (16)$$

In the example above, the weights of the strata within the sample can be determined with the relation:

$$k_i^{'} = \frac{k_i \cdot s_i}{\sum_{i=1}^{3} k_i \cdot s_i}$$

(17)

$$k_1^{'} = \frac{0,17 \cdot 1000}{0,17 \cdot 1000 + 0,50 \cdot 1500 + 0,33 \cdot 1700} = 0,11$$

$$k_2^{'} = \frac{0,50 \cdot 1500}{0,17 \cdot 1000 + 0,50 \cdot 1500 + 0,33 \cdot 1700} = 0,51$$

$$k_3^{'} = \frac{0,33 \cdot 1700}{0,17 \cdot 1000 + 0,50 \cdot 1500 + 0,33 \cdot 1700} = 0,38$$

In order to determine *the sample volume*, we can use the mean square deviation estimated with the help of the pilot survey of 36 companies. The formula used in the case of the non-proportional survey is (Şerban, 2004):

$$n = \frac{z^2}{E^2} \cdot \left(\sum_{i=1}^{3} k_i \cdot s_i \right)^2$$

$$n = \frac{4}{14.400} \left(0,17 \cdot 1000 + 0.50 \cdot 1500 + 0,33 \cdot 1700 \right)^2 \approx 609 \text{ (enterprises)}$$

Each stratum of the sample including 609 enterprises shall be in direct proportion to the dispersion of the monthly turnover of the companies within. Thus, 67 productive enterprises (11% of the sample), 311 commercial enterprises (51% of the sample) and 231 service providers (38% of the sample) shall be analyzed.

As we can see, when the number of criteria used for stratification is limited, the ranking sampling method is highly efficient, ensuring an improved representativeness of the samples and high accuracy of estimates made in terms of reasonable research costs.

4. CONCLUSION

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. Sample research had a steady increase in scope, precisely because it offers clear advantages for analysts and decision makers.

Ranking sampling is particularly effective in situations where the general collectivity is focusing on a set of general criteria which allow a detailed characterization - geographic, demographic, economic, behavioural, etc. From this perspective, ranking sampling is similar to the procedures of market segmentation; the differences between the two are especially noticeable in the objectives pursued: collecting information in the first case and marketing strategy formulation in the second case.

The main advantage of this method is the improved representativeness obtained in comparison with using simple random sampling. In addition, stratified sampling is suitable for comparison between the characteristics of different subpopulations (from this perspective we can analyze, for example, the votes of young people as compared with those of the elderly, the electoral preferences of those living in rural areas as compared with those of the townspeople etc.).

REFERENCES:

[1]. **Balaure, V.** (2002) *Marketing*, Editura Uranus, Bucureşti

[2]. **Baron, T.; Biji, E.** (1996) *Statistică teoretică şi economică*, Editura Didactică şi Pedagogică R.A., Bucureşti

[3]. **Cătoiu, I.** (2002) *Cercetări de marketing*, Editura Uranus, Bucureşti

[4]. **Dura, C.; Drigă, I.; Niță, D.** (2010) *Statistical Landmarks and Practical Issues Regarding the Use of Simple Random Sampling in Market Researches*, Annals of the University of Petroşani - Economics, Vol. X

[5]. **Foltean, F.** (2000) *Cercetări de marketing*, Editura Mirton, Timişioara

[6]. **Gherasim, A.** (2005) *Studiul pieței*, Editura Economică, Bucureşti

[7]. **Gherasim, T.; Gherasim, A.** (2003) *Cercetări de marketing*, Editura Economică, Bucureşti

[8]. **Isaic–Maniu, A.; Mitruț, C.; Voineagu, V.** (1999) *Statistică pentru managementul afacerilor*, Editura Economică, Bucureşti

[9]. **Jaba, E.** (2002) *Statistică*, Editura Economică, Bucureşti

[10]. **Munteanu, V.** (2006) *Marketing. Ştiința şi arta afacerilor*, Editura Tipomoldova, Iaşi

[11]. **Niculescu, E.** (2000) *Marketing modern*, Editura Polirom, Iaşi

[12]. **Porojan, D.; Ciocănel, B.** (2007) *Bazele sondajului*, Institutul Irecson, Centrul de Informare Tehnologică, Bucureşti

[13]. **Prutianu, Ş.; Anastasiei, B.; Jijie, T.** (2002) *Cercetarea de marketing. Studiul pieței pur şi simplu*, Editura Polirom, Iaşi

[14]. **Şerban, D.** (2004) *Statistică pentru studii de marketing şi administrarea afacerilor*, Editura ASE, Bucureşti