USING STATISTICAL SURVEY IN ECONOMICS

Delia, Teselios[1]
Mihaela, Albici[2]

Abstract: *Statistical survey is an effective method of statistical investigation that involves gathering quantitative data, which is often preferred in statistical reports due to the information which can be obtained regarding the entire population studied by observing a part of it. Therefore, because of the information provided, polls are used in many research areas. In economics, statistics are used in the decision making process in choosing competitive strategies in the analysis of certain economic phenomena, the formulation of forecasts. The economic study presented in this paper has the role of illustrating how a simple random sampling is used to analyze the existing parking spaces situation in a given locality.*

Key words: *survey, standard deviation, standard score, mean.*

Classification JEL: *C10, C13.*

1. Introduction

The progress in the last 20 years, reflected on the concept of "information society" which was gradually accepted in academia both in political and economic areas, contributed to the development and deepening techniques used in economic statistics. The phenomena which produced radical changes (still in progress) in terms of increased demand for economic statistical information are [14]:

- changes occurred in the production processes thanks to the new information and communication technologies;
- globalization of social and economic systems;
- the concepts of "knowledge economy" and "digital economy".

The sources of obtaining statistical data, next to methods such as censuses, statistical reports, and statistical surveys represent a rapidly growing variant. Due to the efficiency and economy in obtaining data, the survey method is among the preferred procedures in gathering data. In a market economy based on private property, the survey is the predominant form of obtaining statistical data at the expense of statistical reports – a specific form of centrally planned economies [13].

Moreover, the need to obtain information with maximum efficiency, led to an expansion of using the statistical sampling method as a form of partial observation. In some practical situations the statistical survey is the only method of obtaining information. This is the case where the exhaustive research is being unreasonable, because of destructive nature of the research method – for example in controlling product quality, or uneconomic because of the high costs involved for the inclusion of all community units.

2. Research Sample

The selective research involves the collection and processing of the data related to a part of the general community and the results extend to the whole object of knowledge to achieve a satisfactory characterization. It is necessary for the analyzed part (sample, sample selection) to be representative, e.g. to reproduce at a reduced scale the essential features of the entire population [1].

The survey is a form of statistical survey conducted on a representative part of the population (general community) studied.

[1]PhD Lecturer, Constantin Brâncoveanu University, delia_teselios@yahoo.com.
[2] PhD Lecturer, Constantin Brâncoveanu University, mturmacu@yahoo.com.

All statements and conclusions based on the data from a survey are not determinist, having the nature of statistical statements and are therefore carried out under a certain probability of a given confidence level. The advantage of the survey method is that the level of this probability and of the errors is controllable and can be statistically estimated.

This type of method offers variants of survey and sampling techniques differentiated and adapted to different types of populations in order to ensure the randomness of the sampling units and, in the end, the sample is representative.

In a sample research, by two kinds of communities facing each other – the total community we want to know and the sample we record – meet a number of pair terms that have the same methodological content, but differ in terms of the information included.

The total community is also called female population or statistics reference population, native population, general population, and it represents all the component units of a community subject to statistical research. Defining the population (the total community) is done by listing all the component units. Their totality make up the volume of the community and it is denoted by *N*.

The partial community extracted from the total community in order to be observed is known under different names, depending on the scope of the survey research, namely: selection, sample, sampling, population observed. The sample represents a subset of units drawn from a community we want to know.

The definition of the sample is done by listing the component units of the extracted subassembly and observed from the population size *N*. The volume of the sample is denoted by *n*. The sample volume, mainly because of cost-related issues, is much smaller than the size of the population from which it has been extracted, for instance in the case of partial observation there is the following equation: $n < N$. If $n = N$, the case of complete observation.

The observed variable may be *quantitative* (individual income, number of employees of a company) or *qualitative*. The values of one *X* value are x_i.

Defining the statistical population requires, on the one hand, the statistical delimitation of the population, and on the other hand, checking the population's degree of homogeneity.

Checking the population's degree of homogeneity consists of the analysis of the variation of its characteristics definition, an operation which may be performed based on the data from a previous total observation or the data from surveys held successively. Checking homogeneity of the population is a prerequisite of the correct choice of the sampling procedures.

The results obtained based on the survey data are extrapolated to the whole population size. Extending the results from "part" to "total" does not have a deterministic nature, but a probabilistic one, so they are at risk of being wrong [3]. The advantage of the survey method is not so much eliminating the errors, but especially in their pre-sizing and the "control" of the statements' probability. The advantages of the survey research can be viewed in terms of costs, speed of obtaining the results, knowledge of difficult-to-access communities, etc.

Selective research is cheaper, quicker and more accurate because it is made on a small number of items, and its results can be checked if necessary by another selective survey or by a complete research. The advantage in terms of costs arises from the fact that in a survey research the material, human resources and time costs are reduced compared to a total observation. Related to the human costs, by observing only a part of the community – a representative sample, the workload is reduced and may be covered by a small number of people, usually a specialized staff, thus enhancing the quality of the registration [1].

3. Economic Study

In order to make an analysis on the actual situation of the parking lots, a city hall performs a simple random sampling, choosing a representative sample of 30 parking lots of all 150 existing parking lots. The data resulted from the analysis are presented in Table 1.

Observation: the data from Table 1 have an illustrative role.

Table No. 1. Data from the survey

Parking	Parking capacity (spaces/parking)	Parking	Parking capacity (spaces/parking)
1	30	16	30
2	32	17	65
3	35	18	74
4	40	19	71
5	48	20	45
6	71	21	89
7	65	22	80
8	90	23	35
9	87	24	30
10	52	25	50
11	85	26	70
12	50	27	32
13	72	28	67
14	83	29	60
15	50	30	64

Source: Created by the author

In the first phase, the available data is to be collated.

In order to achieve this, we will note with x the grouping characteristic (parking capacity) and calculate the amplitude variation characteristic.

$$A_x = x_{max} - x_{min} = 90\text{-}30 = 60 \text{ spaces} \qquad (1)$$

With the help of Sturges's formula we first determine the group number ranges (variation) (noted with M) and then the group size range (variation) (denoted by d).

$$M = 1 + 3.22 * \log n, \; n = \text{number of measurements realized} \qquad (2)$$

$$\Rightarrow M = 1+3.22*\log 30 = 5.756 \approx 6$$

$$d = \frac{A_x}{M} = \frac{60}{6} = 10 \text{ spaces} \qquad (3)$$

Between x_{max} and x_{min} values, taking into consideration the values calculated M and d, the interval groups are set.

We centralized the data obtained in Table 2, where we consider the closed intervals to the left and the open ones to the right.

Table No.2. Systematic data

Variation intervals of the parking capacity (spaces/parking)	Number of parking spaces
30-40	7
40-50	3
50-60	4
60-70	5
70-80	5
80-90	6

Source: Created by the author

The systematic data are represented by the following histogram.

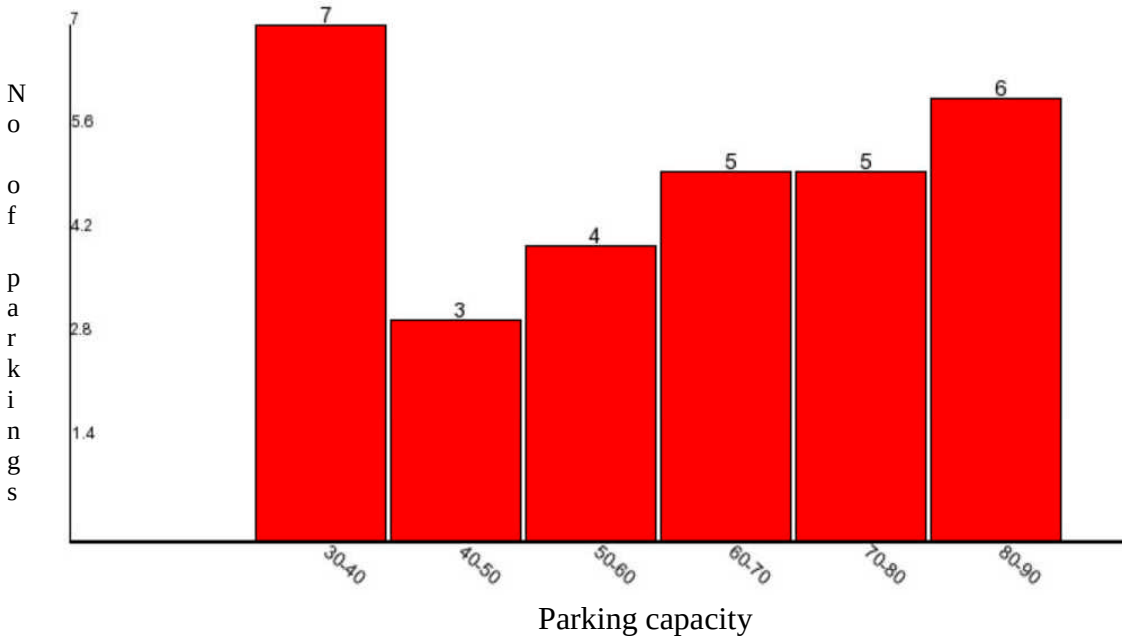

Figure 1. Graphical representation of data collated

Source: Created by the author

We calculate the average value in the sample, using the formula:

$$\bar{x}=\frac{\sum_{i=1}^{6} x_i n_i}{\sum_{i=1}^{6} n_i} \qquad (4)$$

and determine if it is representative for the selection community under review.

Table No. 3. Values used in the sample average formula

	x_i	n_i	$x_i * n_i$
	35	7	245
	45	3	135
	55	4	220
	65	5	325
	75	5	375
	85	6	510
Total	360	30	1810

Source: Created by the author

According to the calculations from Table 3,

$$\bar{x} = \frac{1810}{30} = 60,(3) \text{ spaces/parking} \qquad (5)$$

In order to determine the scatter degree of the data around the central value, we compute the standard deviation, measured here in spaces/parking.

$$\sigma=\sqrt{\sigma^2}=\sqrt{\frac{\sum_{i=1}^{6}(x_i-\bar{x})^2\cdot n_i}{\sum_{i=1}^{6} n_i}}$$ =18.39 spaces/parking, where σ^2 = dispersion. (6)

In other words, the lower limit of the normal variation of the data is 41.94 and the maximum limit is 78.72. By approximation, we can say that there are between 42 and 79 spaces/parking.

The coefficient of variation is

$$v=\frac{\sigma}{\bar{x}}\cdot 100\approx 30\% \qquad (7)$$

Therefore, the analyzed selection community is homogeneous and the average $\bar{x}\approx 60$ spaces/parking is representative for the collectivity.

To the extent to which only a sample of the population was analyzed, the confidence interval must be determined, using the properties of the normal law (Gauss's law).

To calculate the confidence interval of the mean, we apply the following formula:

$$\bar{x}-z_\alpha\cdot\frac{\sigma}{\sqrt{n}}\le\ \mu\ \le\bar{x}+z_\alpha\cdot\frac{\sigma}{\sqrt{n}}, \quad (8)$$

where μ = real average and z_α = materiality

For a symmetrical test, generally used for computing the error, the normal law values are different for different significance thresholds.

Table no. 4 Standard Scores

	1%	**5%**	**10%**	**20%**
z_α	2.57	1.96	1.64	1.28

Source: Made by the author using z-score counter available at http://www.measuringusability.com/pcalcz.php

$$60.3-1.96\cdot\frac{18.39}{\sqrt{30}}\le\mu\le 60.3+1.96\cdot\frac{18.39}{\sqrt{30}}$$

$$\Leftrightarrow\ 53.71\le\mu\le 66.88 \text{ spaces/parking.} \qquad (9)$$

4. Conclusions

The result obtained signifies that the real average of the parking spaces has 95% chances to be in the interval [54,67] spaces.

If we want the maximum permissible error to be reduced to three-quarters (from 6.58 to 4.93) and the results obtained to be guaranteed with a 99% probability (z_α =2.57), the sample selected must have a larger volume.

We calculate this new volume using the formula:

$$n^*=\frac{(2.57)^2\cdot\sigma^2}{(4.93)^2}=\frac{6.60\cdot 338.22}{24.35}=91.74\approx 92 \text{ parkings} \quad (10)$$

Bibliography

1. Albici M. (2010), *Statistică economică. Matematici aplicate în economie*, Editura Universitaria, Craiova

2. Biji E., Baron T. (1991), ş.a, *Statistică teoretică şi economică*, Editura Didactică şi Pedagogică, Bucureşti

3. Isaic-Maniu A. (2001), *Tehnica sondajelor şi anchetelor*, Editura Independența Economică, Piteşti

4. Isaic-Maniu A., Korka M., Mitruț C., Voineagu V. (1998), *Statistică,* Editura Independența Economică, Brăila

5. Isaic-Maniu A., Mitruț C., Voineagu V., *Statistică generală*, disponibilă la http://www.biblioteca-digitala.ase.ro/

6. Popa, M., *Scoruri standard. Curba normala (Gauss)*, disponibilă la http://www.mpopa.ro/statistica_licenta/St1_05_z_gauss.pdf

7. Porojan D. (1993), *Statistică şi teoria sondajului*, Editura Șansa

8. Rotariu T., Iluț P. (1997), *Ancheta sociologică şi sondajul de opinie*, Editura Polirom, Iaşi

9. Voineagu, V., Mitrut, C., ș.a, *Statistică*, disponibilă la http://www.biblioteca-digitala.ase.ro/

10. http://math.tutorvista.com/statistics/z-score-table.html#

11. https://statistics.laerd.com/statistical-guides/standard-score.php

12. http://www.measuringusability.com/pcalcz.php

13. http://simionmija.wordpress.com/2009/10/02/sondajul-statistic/

14. http://www.oecd.org/std/41746710.pdf