Published at http://www.ijate.net   http://dergipark.gov.tr   *Research Article*

Determination of Sample Size and Observation Units

Tülin Acar [ID][1,*]

[1] Parantez Education Research Consultancy Publishing, Ankara, Turkey

Abstract: The purpose of this study was to write programs to define sampling sizes and observation units by probability sampling methods and to provide an idea for software developers. The algorithms of the programs were written in Python 3. The programs may be run by double-clicking on the Windows operating system or by the command prompt of the DOS operating system. Each exe file has a memory space of 5 megabits on average. In this respect, the application files are very useful in terms of sharing by email and mobility by USB memory sticks.

1. INTRODUCTION

The purpose of science is to make explanations about "an object of research or its properties". The "method" is one of the principal issues in both creating scientific explanations that do not contract reason and logic and verifying or falsifying the existing explanations. Method is a kind of program for the things to do (Thomas, 2009). In this respect, acquisition of object(s), the number of objects observed, the representativeness of the objects selected, and whether objects are selected impartially are as critical as a scientific researcher's purpose, theory and hypotheses.

Since science has a language and a method, not all analyses, studies or observations are considered scientific, because the quality of being scientific implies a set of activities that do not rest on mere judgments but on proofs and methods. Both accuracy and ethics of the conclusions made in a study point to the soundness of the method of research. Therefore, a sound and strong method of scientific research rests on the knowledge of the researchers and their sensitivity of following the procedure and principles.

Testing/observing how many units/objects does it take to conduct a valid and sound research? Contrary to popular belief, the answer to this question is not related to a study being qualitative or quantitative. At this point, the object and subject matter of science or a scientific study takes priority. Science is what can be known by information. In other words, science is an effort to

---

explain/study a part of reality. Therefore, this effort to explain is possible by induction or deduction. Based on the principles of reasoning (deduction or induction) that must be employed in the effort to explain a part of reality, categorization of scientific studies as qualitative or quantitative research and association of certain sampling methods with the categorized types of research contradicts the nature of science. Just as the object of information in physics is not the same as that of anthropology, the nature of an object of information of a study cannot be associated with a researcher being qualitative or quantitative. The quality of being qualitative or quantitative is about the measurement and its result. Therefore, the answers to the question of "testing/observing how many units/objects does it take to conduct a valid and sound research?" are not about qualitative or quantitative research type.

One of the most essential elements to be distinguished in a scientific study that attempts to explain reality is to decide whether the sampling or a sample of the properties of the object of study should be analyzed. Sampling is a selection process. On the other hand, a sample is a specimen, a part of the whole. The question of "testing/observing how many units/objects does it take to conduct a valid and sound research?" will be inadequate when this difference is ignored. An initial principle will be necessary for a selection from defined units/objects/elements. The said initial principle may have several different sampling methods. In this case, sampling methods should inevitably have strengths and weaknesses compared to each other.

The method of sampling should not be considered merely a process of defining a number of units/objects/elements. The method of sampling is one of the premises of making a logical explanation about the subject matter of the object of research. For instance, it is not reasonable to take soil samples from every square meter of land using simple random sampling or purposeful sampling for the soil analysis of a field of 5 dunams since it is not compulsory to ensure that the sample is representative of the whole in such a study. On the other hand, for a research on reading comprehension of students in Turkey, it is essential for the students to be representative of the whole. Therefore, it is necessary to grasp the distinctive properties of all sampling methods. In the literature, sampling methods are principally classified in two categories, namely, probability sampling and non-probability sampling (Rubin & Babbie, 2010). Those sampling methods are summarized in Table 1.

In the probability sampling method, units/elements/items are selected from a defined universe based on a statistical probability. However, in the non-probability sampling method, the number of units/elements/items is not determined and units/elements/items are not selected based on a probability and sampling size. Defining a sampling size from the universe using non-probability sampling is accompanied by the selection of the observation units to be included in sampling. This selection process may be tedious for a researcher for studies with a broad universe in particular. Ways to save time and effort bring computer technologies to the mind. Computer technologies makes positive contributions to the research process particularly for the probability sampling methods. Software that is used to define the sample size and observation units is written by different programming languages including PHP, Java, C, C++, R, and Python.

Table 1. Sampling methods and descriptive properties

		Descriptive properties
A Probability Sampling	A.1. Simple Random Sampling	The probability of all units/elements/items being selected is equal and independent of one another.
	A.2. Systematic Sampling	The initial unit/element/item is selected randomly. Nevertheless, selection of other units/elements/items depends on the selection coefficient.
	A.3. Stratified Sampling	The universe is made up of sub-strata (i.e. universes) that are heterogeneous among each other but each homogeneous in itself. Units/elements/items are selected randomly, taking into consideration the strata's rate of representation of the universe.
B. Non-probability Sampling	B.1. Convenience Sampling	Units/elements/items that are easier to reach based with the prevailing conditions are studied.
	B.2.Purposive Sampling	Units/elements/items are studied based on defined and restricted properties.
	B.3. Quota Sampling	The universe is made up of sub-strata (i.e. universes) that are heterogeneous among each other but each homogeneous in itself. Units/elements/items are selected based on a quota.
	B.4. Panel Sampling	Units/elements/items that make the whole are homogeneous in their properties, regardless of the size of the universe. Therefore, the sample is taken from the whole.

1.1. Aim of the Study

The purpose of this study was to write programs to define sampling sizes and observation units by probability sampling methods. Also it was to provide an idea for software developers.

2. METHOD

The algorithms of the programs were written in Python 3. Python is an object-oriented, interpretive, modular, interactive, high-level programming language that is fast and easy to learn (Sahoo& Sahoo, 2016). Codes of the programs written were shared as open source, and user's guides were prepared for the application files. The py code and MIT licenses of the programs can be accessed via Github.

2.1 Determination of the Sampling Size and Observation Units by Simple Random Sampling

Sampling size for simple random sampling is calculated at two stages using the equation below (Royse, Thyer&Padgett, 2010).

$$Sample\ size\,(n) = \frac{n_0}{1+\dfrac{n_0}{N}} \qquad \text{and} \qquad n_0 = \frac{Z^2 * [P*(1-P)]}{c^2}$$

Where:

n_0= Initial size

N= Universe size

Z = Standard Z value for the level of reliability (e.g. ±1.96)

c = Acceptable amount of error (e.g. ±5)

P= Estimated rate of the sampling of interest in the universe (e.g. 50%)

The algorithm written in Python based on this calculation formula is available at https://github.com/totbicer/Simple-Random-Sampling. The application Random_sampling.exe is available at https://parantezanaliz.com/programs/Random_sampling.zip. Once the **Random_sampling.exe** is run, data is input in four steps as seen in Figure 1.

Figure 1. Screenshot of the simple random sampling program

Step 1: A numerical value is input with the defined and restricted size of the universe being a number. Assume that there are 500 units in the universe.

Step 2: A numerical value is input with the sampling error being an natural number between 1 and 100. Assume that the sampling error is 5.

Step 3: The confidence level is input. A numerical value of 95 or 99 is input. Assume that the confidence level is 95%.

Step 4: A numerical value is input with the probability or observation frequency of an event or fact of interest being a natural number. If no information is available as to the probability of happening or being observed, the value should be 50. Thus, the probability of happening or being for each of the observations is equal. For instance, 50 is input for a probability of 50%, 30 is input for a probability of 30%.

Figure 2. Result of the simple random sampling program

When the operation is continued, a simple random sampling is made to show the computed sampling size as shown in Figure 2. Based on the assumption that units in the universe are represented by ordinal numbers, ordinal numbers of the units randomly selected are listed on the screen.

2.2 Determination of Observation Units by Systematic Sampling

Based on the simple random sampling method or a sampling size determined hypothetically, the unit selection process is performed by the following equation (Babbie, 2008).

$$k = N / n \quad \text{and} \quad Ssn = \text{It is determined randomly between 1 and k value.}$$

Where:

N= Universe size

n= Sampling size

k= Selection coefficient

Ssn= Starting sequence number

The algorithm written in Python based on this calculation formula is available at https://github.com/totbicer/Systematic-Sampling. The application Systematic_sampling.exe is available at https://parantezanaliz.com/programs/Systematic_sampling.zip. Once the **Systematic_sampling.exe** is run, data is input in two steps as seen in Figure 3.

Step 1: A numerical value is input with the defined and restricted size of the universe being a number. Assume that there are 500 units in the universe.

Step 2: A numerical value is input with the sampling size being a natural number. Assume that sampling size is 50.


```
S Y S T E M A T I C    S A M P L I N G    P R O G R A M
                      Vesion: (P).2.1
by Tülin (Otbiçer) ACAR © 2019 (P)arantez Education Research Consultancy and Publishing
https://parantezanaliz.com & totbicer@gmail.com

1# The size of the universe defined and limited (N): 500
2# Sample size (Ss) : 50

 >>> C A L C U L A T E D    R E S U L T <<<
 The number of universes are   500  and The sample size is   50
Systematic sample selection coefficient (k) = 10 and Start sequence number (Ssn) = 10

The sequence numbers of the units to be sampled according to the systematic sampling method are listed below.
10      20      30      40      50      60      70      80      90      100     110     120     130 140 150
160     170     180     190     200     210     220     230     240     250     260     270     280 290 300
310     320     330     340     350     360     370     380     390     400     410     420     430 440 450
460     470     480     490
```

Figure 3. Screenshot of the systematic sampling program

As can be seen in Figure 3, the selection coefficient and the randomly selected initial number are shown on the screen. Based on the assumption that units in the universe are represented by ordinal numbers, ordinal numbers of the units selected by the systematic sampling method are listed on the same screen.

2.3 Determination of the Sizes of the Strata with Stratified Sampling

In the stratified sampling method, if a sampling size is not available, the sampling size is determined firstly by the simple random sampling method. Based on a known or determined sampling size and the number of strata, observation units are selected by the equations below (Kalton, 1983).

$$W_h = N_h / N \qquad \text{and} \qquad F_h = W_h * Ss$$

Where:

h= Number of stratum

N_h= h. sampling size of stratum

N= Universe size, $\sum N_h$

W_h = Representation rate of the stratum of the universe, $\sum W_h = 1$

Ss= Sampling size

F_h= h. sampling size for the stratum

The algorithm written in Python based on this calculation formula is available at https://github.com/totbicer/Stratified-Sampling. The application Stratified_sampling.exe is available at https://parantezanaliz.com/programs/Stratified_sampling.zip. Once the **Stratified_sampling.exe** is run, data is input in four steps as seen in Figure 4.

Step 1: A numerical value is input with the defined and restricted size of the universe being a number. Assume that there are 500 units in the universe.

Step 2: A numerical value is input with the sampling size being a natural number. Assume that sampling size is 217.

Step 3: A numerical value is input with the number of strata being a natural number. For instance, assume that there are 3 strata in the universe.

Step 4: A numerical value is input with the number of units/observations for each stratum being a natural number. For instance, assume that there are 150, 50 and 300 observations for each stratum.

```
S T R A T I F I E D   S A M P L I N G   P R O G R A M
                    Version: (P).3.1
by Tülin (Otbiçer) ACAR © 2019 (P)arantez Education Research Consultancy and Publishing
https://parantezanaliz.com & totbicer@gmail.com

1# The size of the universe defined and limited (N) : 500
2# The sample size (Ss) : 217
3# Number of stratum (h) : 3
4# Enter the unit/observation number on the 1. stratum (Nh): 150
4# Enter the unit/observation number on the 2. stratum (Nh): 50
4# Enter the unit/observation number on the 3. stratum (Nh): 300

 Universe representation rate= 0.3, number of sample for stratum (Fh)= 65

 Universe representation rate= 0.1, number of sample for stratum (Fh)= 21

 Universe representation rate= 0.6, number of sample for stratum (Fh)= 130
```

Figure 4. Screenshot of the stratified sampling program.

The rate of representation of each stratum of the universe and the number of units to be sampled for the strata are output on the screen. The process of selecting observation units of a specified number for the strata may also be performed randomly or systematically.

3. CONCLUSION

Execution files in exe format written in Python 3 were prepared to determine the sampling size and observation units for the probability sampling methods. Each program was shared as as open source in trial version for a year. Updates were made during the trial period to ensure that the language and the expressions were clear and comprehensible. No software error was reported during the trial period.

The programs may be run by double-clicking on the Windows operating system or by the command prompt of the DOS operating system. Each exe file has a memory space of 5 megabits on average. In this respect, the application files are very useful in terms of sharing by email and mobility by USB memory sticks.

ORCID

Tülin ACAR 🆔 https://orcid.org/0000-0001-7976-5521

4. REFERENCES

Babbie, E. (2008). *The basics of social research.* USA: Thomson Coperaion

Kalton, G. (1983). *Introduction to survey sampling.* , USA: Sage Publications Ltd

Royse, D., Thyer, B. A. & Padgett, D. (2010). *Program evaluation: An ıntroduction.* USA: Cengage Learning

Rubin, A. & Babbie, E.R. (2010). *Essential research methods for social work.* USA: Cengage Learning

Sahoo, R. & Sahoo, G. (2016). *Computer science with python.* India: Newsaraswati House Pvt. Ltd.

Thomas, G. (2009). *How to do your research project*: A *guide for students in education and applied social sciences*. Londan: Sage Publications Ltd