Scientific
Research
Publishing

An Original Solution for Completing Research through Snowball Sampling—Handicapping Method

Irina-Maria Dragan[1*], Alexandru Isaic-Maniu[2]

[1]Department of Statistics and Econometrics, The Bucharest University of Economic Studies, Bucharest, Romania
[2]Centre of Industry and Services Economy, National Institute of Economic Research, The Romanian Academy, Bucharest, Romania
Email: *irina.dragan@csie.ase.ro, alex.maniu@gmail.com

Abstract

Snowball sampling, as a network research, has a number of advantages in registering "hidden populations", such as: people with rare diseases and unrecognized by the community, illegal immigrants, drug users, people with deviant sexual behavior, undeclared workers, alcoholics, gambling or computer addicts, unemployed jobseekers, domestic and school violence, etc. The procedure is well known and widely applied, as are its drawbacks, the most important of which is the ambiguity regarding the size of the sample under investigation, as the number of interviewees cannot be predetermined, but it is up to the researcher to stop the survey when he/she considers the accumulation of information to be satisfactory or when the imposed limit of the research budget has been reached. The major disadvantage of this method of survey is the absence of objective, quantitative, scientifically based criteria, on the decision to stop the survey and this is precisely the aspect proposed to be resolved in the present paper. Our proposal, the handicapping method, adapts and develops the Wald test of the likelihood ratio for the case of investigating populations in terms of binary characteristics, a typical situation for sociological research. The procedure complements the snowball method, providing a completion under the conditions of assumed statistical risks. The results obtained in this original material can make a major contribution to the improvement of sociological research methods in a special field—sociological field survey/research in closed groups and non-visible populations.

Keywords

Snowball Sampling, Network Sampling, Sequential Survey, Handicapping Method, Decision Criteria, Statistical Risks and Errors

1. Introduction

Snowball sampling belongs to the group of non-random or non-probability sampling methods, sometimes also called directed, empirical, subjective, etc. surveys, but they all refer to the same principle, namely the rational and voluntary selection of the survey units, adapted to the type of sampling established. While for random sampling methods it was possible to calculate the probability that a unit in the population would be included in the sample, this is impossible in non-probability surveys. But, for the application of random surveys, there is a need for up-to-date data-bases and the costs of these sampling techniques are not to be neglected, as not infrequently, these sampling methods require a longer period for design- organization and are therefore less operational.

Non-random sampling, is less strict, easier to apply and does not involve consideration of representativeness as a desirable endpoint for describing the sample. These sampling methods usually leave it up to the researcher to decide which of the investigated community units will be selected and when to end the research. Therefore, schemes in this category are mainly used in exploratory surveys, carried out using data and qualitative research methods. The advantages of non-probability surveys include: they can be used successfully when there is no access to (or no list of) the population studied (e.g. there is no list of those who prefer a certain brand of cigarettes, beer, etc.); it is the only method that can be used when the target population is difficult to identify (e.g. alcoholics, drug users, etc.) or is very specific; a survey frame to design the research is not required; it is less costly than random surveys; it allows results to be obtained much more quickly than probability surveys.

Among the disadvantages of the method are: units are sampled in an arbitrary manner, so the probability of units entering the sample cannot be calculated, consequently the variance and the estimator shift cannot be calculated, so it is not possible to measure the precision of the indicators; there is no guarantee that all units of the population have an equal chance of entering the sample; the selection procedure depends a lot on the experience of the researcher, and quite often the resulting sample can be biased.

The major disadvantage of this method is the absence of objective, quantitative criteria for the termination of the survey. There is ambiguity about the size of the sample to be researched, and therefore the number of interviewees to ensure consistency and relevance of the information collected. This number cannot be calculated in advance, but depends on the experience of the researcher, and on the moment considered that the collection of information.

The present material attempts to counteract this shortcoming by providing a method that gives statistical consistency to the final decision on whether to terminate research under known statistical risks.

2. Literature Review

The literature review was conducted systematically on the main topics where the

snowball method is most commonly applied.

First, we review the treatment of some methodological problems of snowball sampling, found in Browne (2005) who emphasizes the interpersonal relationships between groups which help to build the sample, and Snijders (1992) who makes a critical review of the possibilities of sampling in a synthetically schematic network. A contribution to the development of the method is made by Berg (1988), and Naderifar et al. (2017) which establish chain research procedures. A mathematical formalization of the procedure is provided by Goodman (1961). Rather interestingly enough is the work developed by Lecy and Beatty (2012), who conducted an extensive review of specific literature in the world scientific literature. Dragan and Isaic-Maniu (2012, 2013) also contribute to the development of snowball sampling method. Estimating the number and identifying the health problems of homelessness is a topic developed by the Institute of Medicine, Committee on Health Care for Homeless People (1988) in Washington DC. Fisher (1994) estimates the number and mental health of groups of homeless people. The same important topic for urban agglomerations is addressed by D'Onise et al. (2007) to better calibrate the social services of city administrations. Child abuse is another topic well covered in the scientific literature by the snowball research method. The topic is extensively developed by Park-Higgerson et al. (2008), and ways to reduce this social scourge are re-searched by Ttofi and Farrington (2011). Menesini and Salmivalli (2017) as well as Walton (2005) discuss the prevalence, age and gender differences and different types of bullying. The consequences of violence on children are analysed by Moore et al. (2017), who insist on bullying actions and increased risk of suicidal behaviours in adolescents. A comparison between the level of bullying in state and private schools in Islamabad, Pakistan is conducted by Najam and Kashif (2018) through the medium o field research among 400 students, involving a series of interviews with children in grades 4 and 5, as well as with parents, teachers and school managers. Research with the same aim—identifying the causes and combating bullying, has also been conducted by Hymel and Swearer (2015), Patton et al. (2017), and Fullchange and Furlong (2016). Mishna (2004) organized a survey in order to find ways to reduce physical and verbal bullying. More recently, cyberbullying has become a serious public health problem faced especially by young people, a problem studied by Dennehy et al. (2020), who continue the research conducted by Vandebosch and Van Cleemput (2008) by organizing focus groups, as well as by Navarro and Serna (2016). Also, in the same sphere of concern, there is have the issue of the contradiction between the right to privacy and the expansion of the internet, namely the rise of Big Data, a topic developed by Hazarika et al. (2019).

Migration and trafficking in human beings is a continuation of the studies published by Salt and Almeida (2006), starting from the lack of data on migration flows, the authors initiate a field research, and Reichel and Morales (2017) study the issue of migration in European countries using both official data and field research, a topic also ad-dressed by Pastore and Roman (2020). The issue of

the health status of migrants arriving in waves is developed by Indatwa (2020). Dowle (2021), also using the snowball method, analyses the events of the 2015-2016 European migration "crisis" in each of the four Nordic countries: Denmark, Finland, Norway and Sweden as well as border management. The risks of different economic and social activities is the subject of concern of Mohammed Ameen and Mourshed (2017) who follow through a survey the consequences of accelerated urbanization development and the multiple effects of this process in terms of pollution, reduction of green spaces and carbon emissions. The process of hypertrophied growth of urban tourism and the social consequences of this process, rep-resent the subject of the concerns of Zmyślony and Kowalczyk-Anioł (2019) team, and Kraidi et al. (2020) aim to identify the risks in the management of pipeline systems for the transport of petroleum products, using a questionnaire distributed among different economic agents. McCarthy and Schurmann (2018) follow Australian farmers' perceptions of the risks associated with organic farming. Shanmugam et al. (2022) conduct a survey of financial risk awareness among the urban population in India. Stjepić et al. (2021) investigate the business success of SMEs in Business Intelligence Systems (BIS) and associated risks.

Various other economic issues. The environmental risk generated by excessive, uncontrolled urbanization with ignorance of environmental issues is analysed by Raed and Monjur (2017). Kaya et al. (2020) analyse the impact as well as the risks generated by the migration of IT infrastructure and applications to the cloud. Also, the snowball method is used by Rashid & Mohd Harif (2020) on major banks in Cairo-Egypt to identify risk factors in SME lending. Sequeira (2022) develops a topic of particular economic interest—poultry production and marketing in India and the spread of specific diseases. Topic also ad-dressed by Salman and Hassan (2020). The cross-border cooperation between Spain and Portugal is developed by González-Gómez and Estrella (2016) respectively González-Gómez and Estrella (2020), and Peck and Mulvey (2018) study cross-border work between England and Scotland.

Domestic Violence (DV) has become a major problem in contemporary society, and the snowball method lends itself to this area. Alhabib et al. (2009) conduct a research based on the investigation of databases, studies and field research on violence against women. Ruiz-Pérez et al. (2007) develop a problem related to domestic violence re-search by comparing different methods of conducting field research. The mental con-sequences of domestic violence are the subject of Hackett's (2011), Golding's (1999) and Trevillion's (2012) concern. The psychological causes of gender-based violence are analysed by Lucena et al. (2018). Shah et al. (2012) conduct a study on domestic violence by population segments, psychotic/non-psychotic and age categories of spouses. Among the effects generated by the COVID-19 pandemic there is a vigorous increase in Domestic Violence (DV), a situation analysed by Boserup et al. (2020). Also related to the pandemic, Debashish and Al-Khalifa (2022) develop the issue of volunteerism in Islamic society in Bahrain in population testing and vaccination

activities, and Wang et al. (2021) conduct a literature survey on the spread and control of the epidemic in different countries. Stevenson and Wakefield (2021) conduct an analysis of group behaviour under pandemic travel restrictions.

Other areas of application. Ruban (2017) investigates the limits of Cross-Border Co-operation (CBC) objective of the European Union, collecting interview data by using the snowball method. Jaisuekul and Teerasukittim (2017) identified the factors that can contribute to the promotion of medical and cosmetic tourism in Thailand by. Consumer behaviour is studied by Yoshida et al. (2013). A survey among gamblers to determine the impact of aggressive advertising on their behaviour is conducted by Killick and Griffiths (2022). The collection and entire process of data processing in marketing research is carried out by Gabor (2016). A study conducted by Khan and Bashir (2020) aimed at transferring practices from the commercial sector to non-profit activities through snowball sampling research.

3. Sequential Test—Short History

A so-called sequential test is a procedure whereby after each "test" (interview, measurement, determination, trial, etc.) a certain hypothesis can be accepted, the same hypothesis rejected or additional (additional) information (evidence) claimed (re-quested). It is for these reasons that the size of the sample examined is not known a priori, but is a random element. In some cases, this sample size is very small or, on the contrary, uneconomically large. The state of uncertainty, i.e. the situation when an additional sample is requested, may last longer or shorter depending on the additional information provided by each additional interviewee. The theoretical foundation of sequential analysis was laid in the 1940s (20th century), with independent research on the subject in the UK by Barnard (1946) and in the USA by Wald (1947), a leader in the field, who succeeded in demonstrating precisely the "critical points" of this methodology. The two, working independently, developed a procedure they called sequential analysis, in which the inference made on the population was carried out "step by step". Abraham WALD (1902-1950), was born in Cluj, Romania, studied in Cluj, then in 1927 in Vienna where he completed his doctorate with Karl Menger, and developed concerns related to econometric research. He emigrated to the USA after the Nazi expansion in Europe, where he headed the Statistical Research Group at Columbia University (SRGCU), which, under the aegis of the War Department, developed econometric problems of military interest and application, including the sequential test. In 1943, Wald wrote a technical report entitled "Statistical Analysis of Statistical Data: Theory", with results later published in the seminal work Sequential Analysis.

Synthesis of the Sequential Method

The essence of the method is given by the so-called Sequential Probability Ratio Test.

If X is a continuous or discrete random variable individualized by the density $f(x;\theta)$ where θ is an unknown parameter, to check the statistical hypothesis

$$H_0 : \theta = \theta_0 , \tag{1}$$

with alternative

$$H_1 : \theta = \theta_1 \quad (\theta_0 < \theta_1) \tag{2}$$

The likelihood functions associated with the two hypotheses:

$$P_{on} = f(x_1;\theta_0) \cdot f(x_2;\theta_0) \cdots f(x_n;\theta_0) = \prod_{i=1}^{n} f_X(x_i;\theta_0) \tag{3}$$

$$P_{1n} = f(x_1;\theta_1) \cdot f(x_2;\theta_1) \cdots f(x_n;\theta_1) = \prod_{i=1}^{n} f_X(x_i;\theta_1) \tag{4}$$

$f(x;\theta_0)$ represents the density of X when H_0 is true, and $f(x;\theta_1)$ is the density of X when hypothesis H_1 is correct. Wald substantiated the sequential analysis by creating the Sequential Probability Ratio Test.

$$\gamma_n = \frac{P_{1,n}}{P_{0,n}} = \frac{\prod_{i=1}^{n} f_X(x_i;\theta_1)}{\prod_{i=1}^{n} f_X(x_i;\theta_0)} \tag{5}$$

where:

$x_1, x_2, \cdots, x_n$ are a succession of extractions from the population described by the distribution function;

$F_X(x,\theta)$ —density $f_X(x,\theta)$;

Ratio $\gamma_n = P_{1,n}/P_{0,n}$—likelihood ratio.

The sequential test is composed as follows: two constants A and B are chosen, both positive and $A > B$. At each extraction the ratio $\gamma_n = P_{1,n}/P_{0,n}$ is calculated, and if

$$B < \gamma_n < A \tag{6}$$

then the experiment continues, extracting a new unit. Whether

$$\gamma_n \geq A \tag{7}$$

the research process ends by accepting the alternative hypothesis (H_1) and obviously rejecting the null hypothesis (H_0). Whether

$$\gamma_n \leq B \tag{8}$$

the research ends by accepting the hypothesis H_0 and rejection H_1. The best approximations for limits A and B (Wald, 1947: pp. 44):

$$A \approx (1-\beta)/\alpha \tag{9}$$

respectively

$$B \approx \beta/(1-\alpha) \tag{10}$$

values that satisfy the theoretical requirements imposed by the precision of the method.

For pragmatic reasons it is much more convenient to work with the logarithm of the ratio $P_{1,n}/P_{0,n}$ than directly with the ratio itself (Girshick, 1946). The reason is that $\log(P_{1,n}/P_{0,n})$ can now be written as a sum of n terms.

Thus:

$$\sum \log \mathsf{Y}_n = \log \frac{f(x_1;\theta_1)}{f(x_1;\theta_0)} + \log \frac{f(x_2;\theta_1)}{f(x_2;\theta_0)} + \cdots + \log \frac{f(x_n;\theta_1)}{f(x_n;\theta_0)} \tag{11}$$

and noting

$$z_i = \log \frac{f(x_i;\theta_1)}{f(x_i;\theta_0)}, i = 1, 2, \cdots, n \tag{12}$$

we can write the decision rules as follows:

whether $\log B < \sum z_i < \log A$ the experiment continues; (13)

whether $\sum z_i \geq \log A$ hypothesis H_1 is accepted and H_0 is rejected; (14)

whether $\sum z_i \leq \log B$ hypothesis H_0 is accepted and H_1 is rejected. (15)

Because the sample size is random, it is necessary to evaluate its average volume in order to know how many extractions (in average value) are needed, in order that a final decision to accept or reject the null hypothesis could be made. Average Sample Number (ASN), noted $E_\theta(n)$, is given by (16).

$$\text{ASN} = E_\theta(n) = \left[L(\theta) \log B + (1 - L(\theta)) \log A \right] / E_\theta(z) \tag{16}$$

where:

θ is the true value of the considered parameter, and $E_\theta(z)$ is the average value of the variable $z = \log \left[f(x;\theta_0) / f(x;\theta_1) \right]$.

$L(\theta)$ it is an operational feature that gives the discrimination strength of a survey.

$$L(\theta) = \left[\left(\frac{1-\beta}{\alpha} \right)^h - 1 \right] \Big/ \left[\left(\frac{1-\beta}{\alpha} \right)^h - \left(\frac{\beta}{1-\alpha} \right)^h \right] \tag{17}$$

where h is given by

$$\int_D \left[f(x;\theta_1) / f(x;\theta_0) \right]^h \cdot f(x;\theta) \, dx = 1 \tag{18}$$

where D is the domain of definition of the variable X.

4. Sequential Validation of the Researched Fraction in the Total of a Population

The specific fraction or share of the population, which has the X characteristic followed in the research, in the sequential variant, is particularized, considering the hypothesis of a binomial distribution as follows: Sequential Probability Ratio Test, to check the statistical hypothesis

$$H_0 : p \leq p_0 \tag{19}$$

and

$$H_1 : p \geq p_1 \quad (\text{with} \quad p_0 < p_1) \tag{20}$$

also setting the quantities α, β—namely the statistical risks of type I and II that

accompany the test.

Let x_i be the result found in the questioned person of rank i, and the finding concerns a negative aspect, $x_i = 1$ and $x_i = 0$ are assigned, if the finding highlights a corresponding state, and of the n persons, $s(n)$ are those found in the situation $x = 1$. Then:

$$P_{1,n} = p_1^{s(n)} (1 - p_1)^{n-s(n)} \qquad (21)$$

respectively

$$P_{0,n} = p_0^{s(n)} (1 - p_0)^{n-s(n)} \qquad (22)$$

and the logarithm of the likelihood ratio becomes:

$$\log \frac{P_{1,n}}{P_{0,n}} = s(n) \log \frac{P_1}{P_0} (n - s(n)) \log \frac{1 - p_1}{1 - p_0} \qquad (23)$$

The investigation continues as long as

$$\log \frac{\beta}{1 - \alpha} < \log \frac{p_{1n}}{p_{0n}} < \log \frac{1 - \beta}{\alpha} \qquad (24)$$

If

$$\log \frac{P_{1,n}}{P_{0,n}} \geq \log \frac{1 - \beta}{\alpha} \qquad (25)$$

we reject H_0, accept H_1 and conclude the research, and if it is obtained

$$\log \frac{P_{1,n}}{P_{0,n}} \leq \log \frac{\beta}{1 - \alpha} \qquad (26)$$

the null hypothesis is accepted and the research is concluded.

Denoting by $A_{(n)}$ the quantity we will call the line of acceptance

$$A_{(n)} = \frac{\log \frac{\beta}{1 - \alpha}}{\log \frac{p_1}{p_0} - \ln \frac{1 - p_1}{1 - p_0}} + n \cdot \frac{\log \frac{1 - p_0}{1 - p_1}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} \qquad (27)$$

and with $R_{(n)}$ the quantity we will call the line of rejection

$$R_{(n)} = \frac{\log \frac{1 - \beta}{\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} + n \cdot \frac{\log \frac{1 - p_0}{1 - p_1}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} \qquad (28)$$

then the decision can be expressed as follows:

If $A_{(n)} < s_{(n)} < R_{(n)}$, then the research must be continued by questioning the a $(n + 1)$ person;

If $s_{(n)} \geq R_{(n)}$, then the research is concluded by rejecting the H_0 hypothesis;

If $s_{(n)} \leq A_{(n)}$, then the research is concluded by accepting the H_0 hypothesis.

It can be seen that the quantities $A_{(n)}$ and $R_{(n)}$ depend only on p_0, p_1 and on the α and β, risks and, therefore, they can be calculated before the investigation process of the targeted population, and the field survey can be organized much more efficient. If, from calculations, $A_{(n)}$ is not a whole number, then we round to the largest integer less than $A_{(n)}$ and if $R_{(n)}$ is not a whole number, then we

round to the smallest integer greater than $R_{(n)}$, obviously as we refer to a number of people.

Relationships that give $A_{(n)}$ and $R_{(n)}$ appear geometrically as two lines (the independent variable being n), as they have the same slope ρ and the ordinates χ_0 and χ_1

$$\rho = \frac{\log \dfrac{1-p_0}{1-p_1}}{\log \dfrac{p_1}{p_0} - \log \dfrac{1-p_1}{1-p_0}} \qquad (29)$$

Therefore, decision lines have the next equations

$$A_{(n)} = \chi_0 + n \cdot \rho \qquad (30)$$

and respectively

$$R_{(n)} = \chi_1 + n \cdot \rho \qquad (31)$$

The symbol "log" usually means the decimal logarithm but, because in the continuous case, many densities are exponential, then the application of the natural logarithm leads to convenient forms from a practical point of view.

5. Handicapping Method and Case Study

The proposed procedure is practical and easy to implement, and the presentation is made directly in an applicable manner, on the concrete case of homeless people in Bucharest, Romania. The appearance of homeless people and street children occurred in Romania after 1990, and the authorities' concern for these people was especially visible after 2007, with the country's accession to the European Union and the emergence of new rules on supporting these people. At European level there is a European Federation of National Organizations Working with the Homeless—FEANTSA (FEANTSA, 2021, 2022) which focuses on the urban dimension of EU policy, through cooperation with local authorities in the European Forum to Combat the Phenomenon and for the collection of data on homelessness. FEANTSA has developed a European Typology of Homelessness and housing exclusion—ETHOS. The number of homeless people in Romania is very difficult to establish, and the situation is very volatile, the available data being unreliable. According to European Social Policy Network (ESPN, 2019), the estimated number would be 15,000 people living rough, refers to the year 2008, this figure being estimated by the Samusocial Foundation and Médecins Sans Frontières survey (KE-02-19-507-EN-N-pdf, p. 32). According to the 2011 Census, the number of homeless people in the country was of 162,375, and the 2021 census has been postponed to 2022, when newer figures are expected. The report of the *Samusocial Romania* association (Samusocial Romania, 2013) for Bucharest indicates the number of 5000 homeless people in 2010. The Quality of Life journal (Dan & Dan, 2005: pp. 101-122) shows the same number of 5000 people, and the local authorities in Bucharest (DGASMB, 2021) communicate that they take care of about 900 people. This homeless population is

generated by the depopulation of villages, the loss of homes, but also by the fact that about 3000 young people, who turn 18, leave childcare institutions every year, start life on their own, and as some of them, fail to integrate into the society (White Paper on Homeless Youth in Romania, 2017), increase the ranks of homeless people.

As part of an action to assess the health of homeless people present in the North Station area, Bucharest, and of the subscribers of the "Sfantul Andrei" Care Centre, the abusive consumption of alcohol was monitored, a factor likely to aggravate or trigger various diseases. The World Health Organization (WHO) has defined harmful consumption as a regular average consumption of more than 40 g of alcohol/day in women and 60 g of alcohol/day in men (WHO, 2018). It was considered that a proportion of up to 1% of alcohol consumers above the limit recommended by the WHO is acceptable at the group level ($p_0 \leq 0.01$), and that a weight of 10% or more leads to an appreciation of alcohol abuse at community level ($p_1 \geq 0.10$). The interviews took place individually, successively. The number of people consuming alcohol above the limit recommended by the WHO denoted as $s_{(n)}$ is compared with the decision limits $A_{(n)}$, respectively $R_{(n)}$, given by Relations (30) and (31). The statistical risks of type I and II, associated with the procedure, were established at the usual level of sample surveys ($\alpha = 5\%$ and $\beta = 10\%$). If the number of people identified as consuming alcohol is above the recommended limit ($n) \leq A_n$, the research ends with the acceptance of H_0 hypothesis, and if $s_{(n)} \geq R_n$ the research ends by rejecting the H_0 hypothesis and accepting the H_1 hypothesis, so there is alcohol abuse at the community level. We will not use the classic sequential method, but the handicapping method, starting from Relations (30) and (31), which can also be written in the form

$$n - \frac{s_{(n)}}{\rho} > \frac{-\chi_0}{\rho} \tag{32}$$

respectively

$$n - \frac{s_{(n)}}{\rho} + \frac{\chi_1}{\rho} < 0 \tag{33}$$

The proposed procedure, called by us Handicapping method, a name inspired by the sports terminology, was initially applied in the races of sailing ships, in order to allow the participation in the same race of boats of different classes. Handicap *noun* (competition) is a disadvantage given to a person taking part in a game or competition in order to reduce their chances of winning, or a sports event in which such disadvantages are given (Cambridge Dictionary, 2022). The process has been standardized and extended to other sports (Jensen, 2017). The participants in the competition, different in terms of the previous performances or size classes, are assigned a *score* or a *time*, so the competitors have a different *Go* at the start. This *handicap* aims to equalize the chances in competition for participants different in terms of force. Over time, the process has expanded and

applied in Horse racing, in Chess, GO, Shogi, Polo, Gliding, Sailing, Tennis, Motorcycle speedway, Golf, etc.

In *snowball sampling* as well, the proposed procedure requires that the value of the handicap $((1 − ρ)/ρ)$ penalize, during the series of extractions, each element/person classified in the negative category, so it delays the decision with the size of this handicap.

Next, the decisional relations are rewritten, in order to highlight the number of people outside the negative criterion ($n − s_{(n)}$)

$$\frac{\chi_1}{\rho} + \left(n - s_{(n)}\right) - s_{(n)} \frac{1-\rho}{\rho} > \frac{\chi_1 - \chi_0}{\rho} \tag{34}$$

$$\frac{\chi_1}{\rho} + \left(n - s_{(n)}\right) - s_{(n)} \frac{1-\rho}{\rho} < 0 \tag{35}$$

Relation (34) gives the condition of acceptance of the H_0 hypothesis, so the completion of the survey with the conclusion that the proportion of non-alcoholic persons does not exceed the threshold established as admissible, and Relation (35) represents the condition of acceptance of the H_1 alternative. The handicapping method parameters are:

$$\frac{\chi_1}{\rho} \tag{36.1}$$

$$\frac{\chi_1 - \chi_0}{\rho} \tag{36.2}$$

$$\frac{1-\rho}{\rho} \tag{36.3}$$

The value $\frac{\chi_1}{\rho}$ represents *the value of the handicap*, and the development of the research require a certain algorithm. First, the interview is conducted with the first person identified from the target group. If he/she is non-alcoholic, number 1 is added to the initial value $\frac{\chi_1}{\rho}$, and if he/she is in the abusive consumer category, the value $\frac{1-\rho}{\rho}$ is subtracted from $\frac{\chi_1}{\rho}$. The interview with the second person identified in the group is carried out—if he/she is non-alcoholic, the number 1 is added to the new value of the handicap ($\frac{\chi_1}{\rho}+1$), and if he/she is an abusive consumer, $\frac{1-\rho}{\rho}$ is subtracted from the increased value of the handicap ($\frac{\chi_1}{\rho}+1$). The survey thus continues, interview after interview, and if a number higher by 1 than the limit $\frac{\chi_1 - \chi_0}{\rho}$ is obtained, the research ends with the acceptance of the H_0: $p \leq p_0$ hypothesis, so in the case of the study, the conclusion that the group does not contain alcohol consumers in a proportion

greater than the established threshold ($p_0 = 1\%$), obviously a statement affected by the two decisional risks α, respectively β. If a negative number is obtained in the sequence of interviews, the survey is stopped with the acceptance of the H_1: $p \geq p_1$ hypothesis, and the final conclusion is that in the researched group alcohol abuse is above the limit established as the research hypothesis (10% in this case).

Based on the Relations (27), (28) and (29) we obtain: ρ = 0. 04318, $\chi_0 = -0.93886$, respective $\chi_1 = 1.205379$, consequently the decisional Lines (30) and (31) are

$$A_{(n)} = \chi_0 + n \cdot \rho = -0.93886 + n \cdot 0.04318$$

$$R_{(n)} = \chi_1 + n \cdot \rho = 1.205379 + n \cdot 0.04318$$

Carrying out the field investigation, according to this procedure, involves successive interviews, and after each result is compared the cumulated number of registered values with $A_{(n)}$, respectively $R_{(n)}$, and then the decision is made to accept or reject the advanced hypotheses. Thus, the difficulty and slowness of the process highlights the advantages of handicapping method, which involves the calculation of Indicators (36.1), (36.2), (36.3). In this study, the values (rounded, being people) are obtained: $\frac{\chi_1}{\rho} = 27.9152 \cong 28$, $\frac{\chi_1 - \chi_0}{\rho} = 49.658 \cong 50$,

$\frac{1-\rho}{\rho} = 22.1588 \cong 22$.

Following the conduct of the survey, the first interviewed subject was not an abusive consumer, so $\frac{\chi_1}{\rho} + 1 = 29$. Then the next 18 subjects were non-abusive consumers, and thus the value reaches 47. Obviously, after each recalculation, the result is compared with $\frac{\chi_1 - \chi_0}{\rho} + 1$, this being the limit of acceptance of the null hypothesis. Subject 18 fell into the type of abusive consumer, so from 47 we substract ($\frac{1-\rho}{\rho}$) and the cumulative value reaches 23. The next subject is non-alcoholic, therefore 24, as well as the following 28. Thus, the cumulative value 51 is greater than $\frac{\chi_1 - \chi_0}{\rho} \cong 50$, so with the assumed risks (α = 0.05, β = 0.10) the survey can be concluded with the acceptance of H_0 hypothesis, so the proportion of alcohol abusers in the investigated community does not exceed the threshold of 1%, and contradicts the general opinion on this issue.

6. Discussions and Conclusion

Snowball sampling, as non-random sampling method, is less strict, easier, compared to probabilistic sampling methods, easier to apply, as it does not involve considering representativeness, for the description of the sample, and leaves the researcher to decide which components of the investigated communities will be selected when the research is completed. The method can be used successfully when there is no database, nor a list of the studied population, and the results can

be obtained faster than in the case of probabilistic surveys. It should also be noted that the units are included in the sample in an arbitrary manner, so the probabilities of being included in the sample cannot be calculated, thus the variance and displacement of the estimator cannot be calculated, and consequently, the accuracy of the indicators cannot be measured. The major disadvantage is the absence of objective, quantitative criteria regarding the decision to terminate the survey. The number of people who will be part of the survey cannot be previously calculated, but depends on the researcher's decision.

We also consider it necessary to specify that, in the article, the main goal was not to analyze the situation of homeless people in Bucharest, but to present the handicapping method, as a simple procedure to complete the snowball sampling usable in a wide range of practical cases and reflected in the literature research. Thus, the case of homeless people in Bucharest had only the role of a practical illustration of the handicapping method.

Handicapping method, proposed as a customized version of the sequential method, aims to eliminate the uncertainty of completing the research, introducing a quantitative, objective element, in conditions of probability and decisional statistical risk. The results obtained are unique and can make a major contribution to improving the methods of sociological research in a special field—the sociological survey field, conducted in closed groups and non-visible populations. The major advantage of the handicapping method is given by the introduction of a quantitative, objective criterion for completing the survey. The fact that it does not require special statistical knowledge is not insignificant. The specific parameters can be established prior to the field start of the survey and there is a wide range in which the options for designing the handicapping method are placed, from the values of the proportions considered acceptable or unacceptable (p0 and p1) to specific errors such as I and II (α and β), depending on the desired accuracy and the aspect of the investigation, but also on the nature of the pursued indicators.

Staff conducting the field survey, in addition to conducting the interview, have to perform simple arithmetic addition or subtraction operations and comparisons with the previously calculated limits. The results obtained by the survey indicated, contrary to the general perception, that *most homeless people* were not heavy drinkers, so the general opinion spread was not confirmed, at least in the community of residents in the area of Bucharest North Station.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Alhabib, S. et al. (2009). Domestic Violence against Women: Systematic Review of Prevalence Studies. *Journal of Family Violence, 25,* 369-382.

https://doi.org/10.1007/s10896-009-9298-4

Barnard, G. (1946). Sequential Tests in Industrial Statistics. *The Journal of the Royal Statistical Society, 8,* 1-26. https://www.jstor.org/stable/2983610

Berg, S. (1988). Snowball Sampling. In Kotz, *et al.* (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 8, pp. 528-532). Wiley.

Boserup, B. et al. (2020). Alarming Trends in US Domestic Violence during the COVID-19 Pandemic. *The American Journal of Emergency Medicine, 38,* 2753-2755. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7195322

Browne, K. (2005). Snowball Sampling: Using Social Networks to Research Non-Heterosexual Women. *International Journal of Social Research Methodology, 8,* 47-60. https://doi.org/10.1080/1364557032000081663

Cambridge Dictionary (2022). https://dictionary.cambridge.org/dictionary/english/handicap

D'Onise, K. et al. (2007). The Importance of Numbers: Using Capture-Recapture to Make the Homeless Count in Adelaide. *Australian Journal of Primary Health, 13,* 89-96. https://www.publish.csiro.au/py/PY07012 https://doi.org/10.1071/PY07012

Dan, A. N., & Dan, M. (2005). Homeless People in Romania—An Estimation of Their Number. *Quality of Life/Calitatea Vieții*, **1-2**, 101-122. https://www.revistacalitateavietii.ro/journal

Debashish, S., & Al-Khalifa, D. (2022). Motivations of Young Women Volunteers during COVID-19: A Qualitative Inquiry in Bahrain. *Administrative Sciences, 12,* Article 65. https://www.mdpi.com/2076-3387/12/2/65

Dennehy, R. et al. (2020). Young People's Conceptualizations of the Nature of Cyberbullying: A Systematic Review and Synthesis of Qualitative Research. *Aggression and Violent Behavior, 51,* Article ID: 101379. https://doi.org/10.1016/j.avb.2020.101379

DGASMB (2021). *Raport-anual-de-activitate-2020-DGASMB*. https://www.dgas.ro/wp-content/uploads/2017/07/Raport-anual-de-activitate-2020-DGASMB-22.02.2021-final1.pdf

Dowle, L. J. (2021). Toward a (Co)relational Border? Order, Care, and Chaos at Nordic Borders. *Political Geography, 90,* Article ID: 102455. https://doi.org/10.1016/j.polgeo.2021.102455

Dragan, I. M., & Isaic-Maniu, A. (2012). Snowball Sampling Developments Used in Marketing Research. *International Journal of Arts and Commerce, 1,* 214-223. https://ijac.org.uk/vol-1-no-6-november-2012

Dragan, I. M., & Isaic-Maniu, A. (2013). Snowball Sampling Completion. *Journal of Studies in Social Sciences, 5,* 160-177. https://www.infinitypress.info/index.php/jsss/article/view/355

ESPN (European Social Policy Network) (2019). *Fighting Homelessness and Housing Exclusion in Europe: A Study of National Policies*. https://op.europa.eu/en/publication-detail/-/publication/2dd1bd61-d834-11e9-9c4e-01aa75ed71a1/language-en

FEANTSA (2021). *European Federation of National Organisations Working with the Homeless.* https://www.feantsa.org/en/about-us/what-is-feantsa

FEANTSA (2022). *Working Together to End Homelessness in Europe.* https://www.feantsa.org/en/report/2022/05/18/gendered-approach-to-youth-homelessness-study-session-report

Fisher, N. (1994). Estimating Numbers of Homeless and Homeless Mentally Ill People in

North East Westminster by Using Capture-Recapture Analysis. *BMJ Clinical Research, 308,* 27-30. https://www.researchgate.net/signup.SignUp.html
https://doi.org/10.1136/bmj.308.6920.27

Fullchange, A., & Furlong, M. (2016). An Exploration of Effects of Bullying Victimization from a Complete Mental Health Perspective. *SAGE Open, 6,* 1-12.
https://doi.org/10.1177/2158244015623593

Gabor, M. R. (2016). *Analiza si inferenta statistica a datelor de marketing.* C.H. Beck.
https://www.amazon.com/Analiza-interferenta-statistica-marketing-Romanian/dp/6061806329

Girshick, A. M. (1946). Contributions to the Theory of Sequential Analysis-I. *Annals of Mathematical Statistics, 17,* 123-143.
https://projecteuclid.org/download/pdf_1/euclid.aoms/1177730976

Golding, J. M. (1999). Intimate Partner Violence as a Risk Factor for Mental Disorders: A Meta-Analysis. *Journal of Family Violence, 14,* 99-132.
https://link.springer.com/article/10.1023/A:1022079418229

González-Gómez, T., & Estrella, G. (2016). Reporting a Bottom-Up Political Process: Local Perceptions of Cross-Border Cooperation in the Southern Portugal-Spain Region. *European Urban and Regional Studies, 23,* 468-480.
https://doi.org/10.1177/0969776413518781

González-Gómez, T., & Estrella, G. (2020). Disclosing the Relational Structure of Institutional Cross-Border Cooperation in Two Cross-Border Regions in Europe. *Journal of Borderlands Studies, 35,* 113-129. https://doi.org/10.1080/08865655.2017.1399810

Goodman, L. A. (1961). Snowball Sampling. *Annals of Mathematical Statistics, 32,* 148-170.
https://www.jstor.org/stable/2237615?seq=1
https://doi.org/10.1214/aoms/1177705148

Hackett, M. (2011). Domestic Violence against Women: Statistical Analysis of Crimes across India. *Journal of Comparative Family Studies, 42,* 267-288.
https://doi.org/10.3138/jcfs.42.2.267

Hazarika, B. B. et al. (2019). The Role of Service Recovery in Online Privacy Violation. In Information Reso Management Association (Ed.), *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 1498-1527). IGI-Global.
https://doi.org/10.4018/978-1-5225-8176-5.ch076
https://www.igi-global.com/chapter/the-role-of-service-recovery-in-online-privacy-violation/224644

Hymel, S., & Swearer, S. (2015). Four Decades of Research on School Bullying. *American Psychologist, 70,* 293-299. https://doi.org/10.1037/a0038928
https://www.apa.org/pubs/journals/releases/amp-a0038928.pdf

Indatwa, A. N. (2020). *Health Literacy and Diabetes among Refugee Women Residing in Arizona* (p. 8393). Walden Dissertations and Doctoral Studies, Walden University.
https://scholarworks.waldenu.edu/dissertations/8393

Institute of Medicine, Committee on Health Care for Homeless People (1988). *Homelessness, Health, and Human Needs.* The National Academies Press.
https://www.ncbi.nlm.nih.gov/books/NBK218232/
https://doi.org/10.17226/1092

Jaisuekul, M., & Teerasukittim, C. (2017). The Study of Thailand's Cosmetic Surgery Market and Attitudes of Surgeons and Foreign Patients towards Cosmetic Surgery in Thailand. *Journal of Community Development Research (Humanities and Social Sciences), No. 10,* 16-26. http://www.journal.nu.ac.th/JCDR/article/view/1717/1050

Jensen, C., & Petersen. M. B. (2017). The Deservingness Heuristic and the Politics of Health

Care. *American Journal of Political Science, 61,* 68–83.
https://doi.org/10.1111/ajps.12251

Kaya, F. et al. (2020). The Banking Industry Underestimates Costs of Cloud Migrations. In *22nd IEEE International Conference on Business Informatics* (pp. 300-309). IEEE.
https://ieeexplore.ieee.org/document/9140248
https://doi.org/10.1109/CBI49978.2020.00039

Khan, I., & Bashir, T. (2020). Market Orientation, Social Entrepreneurial Orientation, and Organizational Performance: The Mediating Role of Learning Orientation. *Iranian Journal of Management Studies (IJMS), 13,* 673-703. http://ijms.ut.ac.ir

Killick, E., & Griffiths, M. A. (2022). A Thematic Analysis of Sports Bettors' Perceptions of Sports Betting Marketing Strategies in the UK. *International Journal of Mental Health and Addiction, 20,* 800-818. https://doi.org/10.1007/s11469-020-00405-x

Kraidi, L. et al. (2020). Using Stakeholders Judgment and Fuzzy Logic Theory to Analyze the Risk Influencing Factors in Oil and Gas Pipeline Projects: Case Study in Iraq, Stage II. *International Journal of Critical Infrastructure Protection, 28,* 14-22.
https://doi.org/10.1016/j.ijcip.2020.100337

Lecy, J., & Beatty, K. (2012). Structured Literature Reviews Using Constrained Snowball Sampling and Citation Network Analysis. *SSRN Electronic Journal,* 1-15.
https://doi.org/10.2139/ssrn.1992601

Lucena, K. et al. (2018). Gender and Violence: Contributions to the Debate. *Journal of Human Growth and Development, 28,* 109-112.

McCarthy, B., & Schurmann, A. (2018). Risky Business: Growers' Perceptions of Organic and Biodynamic Farming in the Tropics. *Rural Society, 27,* 177-191.
https://doi.org/10.1080/10371656.2018.1504734

Menesini, E., & Salmivalli, C. (2017). Bullying in Schools: The State of Knowledge and Effective Interventions. *Psychology, Health & Medicine, 22,* 240-253.
https://doi.org/10.1080/13548506.2017.1279740

Mishna, F. (2004). A Qualitative Study of Bullying from Multiple Perspectives. *Children & Schools, 26,* 234-247. https://doi.org/10.1093/cs/26.4.234

Mohammed Ameen, R. F., & Mourshed, M. (2017). Urban Environmental Challenges in Developing Countries. A Stakeholder Perspective. *Habitat International, 64,* 1-10.
https://www.sciencedirect.com/science/article/abs/pii/S0197397516307093#preview-section-cited-by
https://doi.org/10.1016/j.habitatint.2017.04.002

Moore, E. S. et al. (2017). Consequences of Bullying Victimization in Childhood and Adolescence: A Systematic Review and Meta-Analysis. *World Journal Psychiatry, 7,* 60-76.
https://doi.org/10.5498/wjp.v7.i1.60

Naderifar, M. et al. (2017). Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research. *Strides in Development of Medical Education Journal (SDME), 14,* e67670. https://doi.org/10.5812/sdme.67670

Najam, H., & Kashif, I. (2018). Impact of Bullying on Students' Behavioral Engagement. *Journal of Health Education Research & Development, 6,* Article ID: 1000261.
https://doi.org/10.4172/2380-5439.1000261

Navarro, R., & Serna, C. (2016). Spanish Youth Perceptions about Cyberbullying: Qualitative Research into Understanding Cyberbullying and the Role That Parents Play in Its Solution. In R. Navarro, S. Yubero, & E. Larrañaga (Eds.), *Cyberbullying across the Globe* (pp. 193-218). Springer. https://doi.org/10.1007/978-3-319-25552-1_10

Park-Higgerson, H.-K. et al. (2008). The Evaluation of School-Based Violence Prevention

Programs: A Meta-Analysis. *Journal of School Health, 78,* 465-479.

Pastore, F., & Roman, E. (2020). Migration Policies and Threat-Based Extraversion. Analysing the Impact of European Externalisation Policies on African Polities. *Revue Européenne des Migrations Internationales, 36,* 133-152. https://doi.org/10.4000/remi.14591

Patton, U. D. et al. (2017). A Systematic Review of Research Strategies Used in Qualitative Studies on School Bullying and Victimization. *Trauma, Violence, & Abuse (TVA), 18,* 3-16. https://doi.org/10.1177/1524838015588502

Peck, F., & Mulvey, G. (2018). Cross-Border Collaboration in Economic Development: Institutional Change on the Anglo-Scottish Border. *Journal of Borderlands Studies, 33,* 69-84. https://doi.org/10.1080/08865655.2016.1257365

Raed, A., & Monjur, M. (2017). Urban Environmental Challenges in Developing Countries—A Stakeholder Perspective. *Habitat International, 64,* 1-10. https://doi.org/10.1016/j.habitatint.2017.04.002

Rashid, S., & Mohd Harif, M. A. A. (2020). Influential Financial Factors of Employee Engagement Outcomes within SME Business in Malaysia: A Qualitative Study. *Journal of Business Management and Accounting, 6,* 111-125. https://e-journal.uum.edu.my/index.php/jbma/article/view/8853

Reichel, D., & Morales, L. (2017). Surveying Immigrants without Sampling Frames—Evaluating the Success of Alternative Field Methods. *Comparative Migration Studies, 5,* Article No. 1. https://doi.org/10.1186/s40878-016-0044-9

Ruban, T. (2017). Using Traditional Faiths and Knowledge Are Secret of Science and Technology by the Tamil Communities in Batticaloa, Sri Lanka. *Advances in Crop Science and Technology, 5,* 321-329.

Ruiz-Pérez, I. et al. (2007). Methodological Issues in the Study of Violence against Women. *The Journal of Epidemiology and Community Health, 61,* 26-31. https://jech.bmj.com/content/61/Suppl_2/ii26 https://doi.org/10.1136/jech.2007.059907

Salman, S., & Hassan, R. (2020). Motivating Factors for Consumers to Participate in Takāful: A Survey by Indian Insurance Policyholders. *Journal of Islamic Accounting and Business Research, 11,* 1871-1888. https://doi.org/10.1108/JIABR-12-2018-0217

Salt, J., & Almeida, J. C. (2006). International Migration in Europe. *Revue Européenne des Migrations Internationales, 36,* 155-175. https://journals.openedition.org/remi https://doi.org/10.4000/remi.2828

Samusocial Romania (2013). *Real Life Estate.* https://www.samusocial.ro/proiecte/proiecte-in-derulare/atelierul-ocupational-proiect

Sequeira, S. A. (2022). *Analysis of Poultry Trade Networks to Improve Risk-Based Surveillance: A Survey Study in Gujarat, India.* Universidade de Lisboa. http://hdl.handle.net/10400.5/24370

Shah, H. S. et al. (2012). Perception and Prevalence of Domestic Violence in the Study Population. *Industrial Psychiatry Journal, 21,* 137-143. https://pubmed.ncbi.nlm.nih.gov/24250047

Shanmugam, K. et al. (2022). Relationship between Big-Five Personality Traits, Financial Literacy and Risk Propensity: Evidence from India. *IM Kozhikode Society & Management Review.* https://doi.org/10.1177/22779752221095282

Snijders, A. T. (1992). Estimation on the Basis of Snowball Samples: How to Weight? *Bulletin of Sociological Methodology, 36,* 59-70. https://doi.org/10.1177/075910639203600104

Stevenson, C., & Wakefield, J. (2021). Collectively Coping with Coronavirus: Local Community Identification Predicts Giving Support and Lockdown Adherence during the

COVID-19 Pandemic. *British Journal of Social Psychology, 60,* 1403-1418.

Stjepić, A. et al. (2021). Exploring Risks in the Adoption of Business Intelligence in SMEs Using the TOE Framework. *Journal of Risk and Financial Management, 14,* Article No. 58. https://www.mdpi.com/1911-8074/14/2/58
https://doi.org/10.3390/jrfm14020058

Trevillion, K. et al. (2012). Experiences of Domestic Violence and Mental Disorders: A Systematic Review and Meta-Analysis. *PLOS ONE, 7,* e51740.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530507

Ttofi, M., & Farrington, D. (2011). Effectiveness of School-Based Programs to Reduce Bullying: A Systematic and Meta-Analytic Review. *Journal of Experimental Criminology, 7,* 27-56. https://doi.org/10.1007/s11292-010-9109-1

Vandebosch, H., & Van Cleemput, K. (2008). Defining Cyberbullying: A Qualitative Research into the Perceptions of Youngsters. *CyberPsychology & Behavior, 11,* 499-503.
https://doi.org/10.1089/cpb.2007.0042

Wald, A. (1947). *Sequential Analysis.* Chapman & Hall.
https://www.amazon.com/Sequential-Analysis-Abraham-Wald-dp-0486615790/dp/0486615790/ref=mt_other?_encoding=UTF8&me=&qid

Walton, G. (2005). Bullying Widespread: A Critical Analysis of Research and Public Discourse on Bullying. *Journal of School Violence, 4,* 91-118.
https://doi.org/10.1300/J202v04n01_06

Wang, X. et al. (2021). Community Identity Profiles and COVID-19-Related Community Participation. *Journal of Community & Applied Social Psychology, 32,* 398-410.

White Paper on Homeless Youth in Romania (2017). *Carta Albă a Tinerilor fără Adăpost din România 2017.*
https://www.younginitiative.org/wp-content/uploads/2015/09/WhitePaper_RO_TineriFaraAdapost.pdf

WHO (World Health Organization) (2018). *Global Status Report on Alcohol and Health 2018.* https://apps.who.int/iris/handle/10665/274603

Yoshida, T. et al. (2013). Snowball Sampling Consumer Behaviour Research to Characterise the Influence of Market Mavens on Social Networks. *International Journal of Intelligent Systems Technologies and Applications, 12,* 268-282.

Zmyślony, P., & Kowalczyk-Anioł, J. (2019). Urban Tourism Hypertrophy: Who Should Deal with It? The Case of Krakow (Poland). *International Journal of Tourism Cities, 5,* 247-269. https://doi.org/10.1108/IJTC-07-2018-0051