Journal of Studies in Social Sciences
ISSN 2201-4624
Volume 5, Number 2, 2013, 160-177

Snowball Sampling Completion

Irina-Maria Dragan, Alexandru Isaic-Maniu

Department of Statistics and Econometrics, University of Economics, Bucharest, Romania
University of Economic Studies, Faculty of Economic Cybernetics, Statistics and Informatics (ECSI), 15-17 Calea Dorobanți, District 1, room 2610, Bucharest Romania

Corresponding author: Alexandru Isaic-Maniu

Abstract. The snowball sampling, as a network research, presents numerous advantages in registering “hidden populations” such as: drug consumers, people with deviant sexual behavior, people with rare diseases which are unapproved by society, illegal immigrants, people working in the black market, but also presents a major disadvantage – the absence of objective, quantitative criteria in taking a decision to conclude the investigation. There is also a certain ambiguity on the number of interviewed persons, which cannot be established previously; thus, it depends on the researcher to stop the investigation when he considers the information gathered is satisfactory or sufficient relative to the objective of the investigation.
The purpose of this material is to adapt and develop the sequential method (Wald’s test) for the case of hidden populations, typical situations for sociologic researches where snowball sampling is adequate.
The results obtained in this material, innovative in their development, can contribute significantly to improving sociological research methods in a special area – field sociologic investigation developed in closed groups.
Keywords: Snowball sampling, network sampling, sequential survey, decision criteria, decisional risks, sample volume

1. Introduction

The “snowball” sampling is a research technique through survey and data registration which is usually used in sociology, psychology, or management studies, and are recommended when: the population cannot be strictly delimited or detailed (e.g. homeless people); the characteristics of the sample are rare (drug-addicts, people with rare diseases, unemployed youth, club members, elites, etc.); a good research method when the study is on behaviors, perceptions, customs, for the description of "typical" cases which cannot be generalized for an entire populations (a typical family in a rural area, school abandonment, the issue of malnourished children, health problems of mine workers, domestic violence, club members, political leaders in a certain area, the members of an organization, virtual communities, socially disgraced people, various elites, illegal immigrants, youths practicing exclusive sports (hockey, golf, ice skating, etc) or extreme sports, excessive alcoholics. The most frequent situations when this non-probabilistic method was used are: the research made on drug-addicts, people with rare diseases, children infected with AIDS, etc).

The characteristic of this type of sampling is that it is not used to estimate the characteristics of the general population but to estimate the characteristics of a network of "hidden" populations (rare, difficult to identify).

The term "hidden population", synonym to "*very seldom*" population or "*difficult to encounter*" population, is used to refer in general to populations on which there are not official information, or which represent less than 2% of the population. "Hidden" doesn't necessarily have the meaning of illegal, but unobserved, unregistered, or under the radar. In other words, due to their rarity, this type of population is difficult to identify, to study and to recruit for the imposed investigation, most of the times because of the attributed social stigmata, the legal status and the lack of visible consequences of the activity of their members. Starting from the presented definitions, it is clear that a "hidden" population is a population for which there isn't a predefined method of sampling and which cannot constitute the subject of probabilistic sampling methods as well as of most of the non-probabilistic ones. In time, there have been identified solutions to reduce the subjectivity of this method, such as: the ascendant and descendant methodology proposed by van Meter (1990), the inductive analysis – Adler (1990), quantitative methods – Strauss (1987), including the use of the computer – Fritz (1990).

2. Literature review

A few major themes have been the preferred subject for the authors preoccupied with the use of the "snowball sampling method" in survey researches.

The general issue of social networks, developed by authors like: Martino, F., Spoto, A. (2006); Kuhar, R., Švab, A. (2008); Atkinson, R., Flint, J. (2001); Hanneman, R., A. (2001); Dodds, P., S., Muhomad, R., Watts, D., J. (2003); Adams, S., Carter, N., Hadlock (2008);

Migration of human trafficking, in the studies published by Salt, J. and Hogarth, J. (2000); Paspalanova, M: (2006); Salt, J., Clarke, J., Smidt, S. (2000);

The use of tobacco and other drugs is debated in depth in the research of Spreen, M., Coumans, M. (2003); Avico, U., Kaplan, C., Kockzak, D. ,Van Meter(1988); Biernacki (1990)Hendricks, V., M., Blanken, Adriaans, N. (1992); Griffiths, P., Gossop, M., Powis, B., Strang, J., Duncan, D., F., White, J., B., Nichalson, T. (2003); Etter, J., F., Perneger, T., V. (2000);

The issue and the forecast of homeless people: D`Onise, K., Wang, Y., McDermott, R. (2007); Skeldon, R. (2004);

Child abuse and the issue of teenagers: Barlow, J., Davis, H., McIntosh, E., Jarett, P., Mockford, C., Stewart-Brown, S. (2007); Day, C.,Davis, H. (2006); Kirkpatrick, S., Barlow, J., Stewart-Brown, S., Davis, H. (2007);

Research on sexual minorities using snowball sampling method is reflected in the studies of various authors such as: Švab, A., Kuhar, R. (2008); Diamond, M. (1993); Standfort, T. (1997); Aaron, D., J., Chang, Y-F., Markovic, N., LaPorte, R., E. (2003);

The statistic approach of the snowball sampling method and the methodology issue specific to this research can be encountered in Berg, S. (1988); Borgatti, S., P., Everett, M., G., Freeman, L., C. (2002); Johnson, T., P. (2005); Song Hoon Lee, Pan-Jum Kim, Hawoong Jeong (2006); Kolaczyk, E., D. (2009); Gabor, M., R. (2007); Goodman, L., A. (1961); Streeton, R., Cooke, M., Campbell, J. (2004), Kostinen, J., Snijders, T. (2007),Salganik ,M .,(2006) .

The principle of this sampling method includes the identification, done by the researcher, based on specific reasoning, of a number of respondents to be interviewed, and which in their turn, shall indicate (recommend) other respondents which will make the object of the research.

The methodology used and the initial interviewed subjects respectively, indicate additional subjects (random methods can be used to generate the first respondents), which is why the use of this method can decrease searching costs significantly, but can also become costly when introducing the systematic error, because the use itself reduces the probability that the survey has a good representativeness for the population.

The procedure would gain extra efficiency if the method proposed by the presented material had been implemented; the main objective of this paper is to adapt and develop the sequential analysis for the case of reduced populations, specific to social networks.

3. Brief history of sequential analysis

According to Irving I. Burr (1953) a so-called sequential test is a procedure through which after each phase (measurement, determining, testing etc) a certain hypothesis can be accepted, rejected, or additional information (proof, phase) can be requested. Precisely for these reasons, the size of the examined sample is random and not known in advance.

In some cases, this group is very small, or on the contrary, uneconomically large. The uncertainty status, referring to the situation of requiring additional evidences, can last more or less depending on the additional information brought by each additional individual.

The theoretical basis of the sequential analysis took place in the 1940s (20th century), while independent research had been taking place in Great Britain (see George A. Barnard, 1946, “Sequential tests in industrial statistics” published in the Journal of the Royal Statistical Society B Series, no 8, pg. 1-26) and in the U.S., the leader in this field being Abraham Wald, the one who managed to prove precisely the “critical points” of this methodology.

Abraham Wald (1902 – 1950) considered today the father of sequential analysis was born in Transilvania – Romania, in a Jewish family, in the period when this region was completely incorporated in the Austro-Hungarian Empire. He attended the University of Cluj, and in 1927 he went to Wien, starting a PhD program with the professor Karl Menger. He introduced him to a well-known Austrian economist and banker, Karl Schlesinger, who succeeds in taking Wald to the econometric research field – a developing field in that period. Until 1938 while he was in Wien, Wald's scientific interest and results are in this direction, as he succeeded to make a rigorous demonstration of a series of suppositions with which the economists of the time operated. In that period, Wald wrote in a series of papers in German, which were republished and translated in English after the World War II (e.g. On some systems of equations of mathematical economics", 1936, translated in Econometrica).

In the 1940s, at Columbia University in New York, a Statistical Research Group of Columbia University, also known under the name SRGCU, which, under the authority of the War Department attacked a series of issues of military interest and applicability, such as the optimal allocation of resources, the ammunition quality verification, the distribution of artillery hits around a fixed target, the method of comparing resistance at the penetration of various armors, etc (the U.S. as it is well-known, enters war with Japan in December 1941, after the Pearl Harbour episode).

This group included distinguished scientists in the American scientific society such as Milton Friedman, Harold Freeman, W. Allen Wallis, and others. Wald was also included in this multidisciplinary group as a consultant in statistical issues, as he also had the advantage of speaking fluently other languages except English (German, Hungarian and French, but most probably Romanian too).

Friedman and Wallis asked for Wald's help, who dedicated his entire energy and talent to finding solutions to the problems raised by the "sequential experimentation" – as he correctly called it ever since 1943. Wald understood that two essential issues must be solved:

a) To prove the fact that the sequential process is completed after a finite number of steps – meaning that it cannot last indefinitely, and
b) To build a quantitative criterion in taking a decision.

Unlike the classical case, the sequential procedure implies three decisions:

i. Continuing the research
ii. Accepting the initial hypothesis
iii. Rejecting the initial hypothesis (and automatically accepting the alternative hypothesis)

In 1943, Wald writes a (secret) technical report entitled "Statistical Analysis of Statistical Data: Theory" which SRGCU transmits to the Defense Research Committee of War Department. In this document, Wald rigorously set the fundamentals of sequential analysis by creating the so-called SPRT (Sequential Probability Ratio Test). This SRPT becomes the practical work instrument in making sequential tests on statistical hypotheses.

After the war ended – as early as May 1945, SRGCU receives approval from the War department to eliminate the "secret" classification on his reports and papers, and thus, Wald and other members' results start to become known to the international scientific society, also being published in various profile journals, mainly American (The Annals of Statistics, Journal of the American Statistical Association, Industrial Quality Control etc).

Wald's most important book was published in 1947, at John Wiley Publishing House, entitled **Sequential Analysis**. The examples given by Wald in his book are from the area of quality, this proving the importance that the distinguished savant granted to this fully developing field. Also, Wald wrote his monograph and a brief historical research trying to discover similar preoccupations in the field in other sides of the world. Thus, he mentions the "sequential" experiments performed in Benegal (India) by Prasanta Chandra Mahalanobis (1893 - 1972) referring to the census of jute plantations, as well as the theoretical results obtained in certain particular problems in sequential analysis by the British statisticians C.M. Stokman and G.A. Barnard in 1944.

4. The sequential test – brief presentation

Let us consider a measurable variable X – or in statistical language, a random variable individualized by its density $f(x;\theta)$, where θ is an unknown parameter (or a vector of parameters) on which we focus our inference. Wald's construction is viable for discrete random variables as well, but to fix the principles, we operate with continuous random variables.

Thus, let us consider the hypothesis:

$$H_0 : \theta = \theta_0 \quad (1)$$

With the alternative hypothesis:

$$H_1 : \theta = \theta_1, (\theta_0 < \theta_1) \quad (2)$$

Here, θ_0 and θ_1 may have various interpretation – for instance, an average value.

Then, $f(x;\theta_0)$ represents the density of X when H_0 is true and $f(x;\theta_1)$ is the density of X when the alternative H_1 is just. Successive measurements are made on variable X, $x_1, x_2, x_3 \ldots$, and the likelihood functions associated to the two hypotheses:

$$P_{o,n} = f(x_1;\theta_0) \cdot f(x_2;\theta_0) \ldots f(x_n;\theta_0) \quad (3)$$

$$P_{1,n} = f(x_1;\theta_1) \cdot f(x_2;\theta_1) \ldots f(x_n;\theta_1) \quad (4)$$

Quantity

$$R_n = P_{1,n}/P_{o,n} \quad (5)$$

Is called likelihood ratio – term introduced by Sir Roland A. Fisher (1890 – 1962).

The sequential test is built as follows: for starters two constants are chosen A and B (depending on the associated risks α and β, corresponding to the two hypotheses), A, B > 0, and A>B.

At each experiment step (interviewing, measuring, examining, etc.), we compute R_n; if $B < R_n < A$, then the experiment continues, extracting another unit for examination. If $Rn \le A$ the process is stopped by accepting the alternative hypothesis (H1), automatically rejecting the null hypothesis (H_0).

However if $R_n \geq B$ the experiment stops by accepting H_0 and rejecting H_1.

In his book (1947 edition, pg. 44 and Bârsan-Pipu - 2002), Wald shows that it is difficult to obtain exact values for A and B, but that the approximations $A \approx (1-\beta)/\alpha$ and $B \approx \beta(1-\alpha)$ satisfy the theoretical requests imposed by the sequential method.

Wald also makes the following remark (idem op. cit. pg. 38):

"Also for reasons related to the easiness in computations, it is much more convenient to work with the logarithm of $P_{1,n}/P_{0,n}$ than with the fraction itself. The reason is that $\log(P_{1,n}/P_{0,n})$ can also be written as a sum of terms".

Indeed, we have:

$$\log R_n = \log \frac{f(x_1;\theta_1)}{f(x_1;\theta_0)} + \log \frac{f(x_2;\theta_1)}{f(x_2;\theta_0)} + \ldots + \log \frac{f(x_n;\theta_1)}{f(x_n;\theta_0)} \qquad (6)$$

and noting:

$$z_i = \log \frac{f(x_1;\theta_1)}{f(x_1;\theta_0)}, \quad i = 1, 2, \ldots, n \qquad (7)$$

We can write the decision rules thus:

(1) if $\log B < \sum z_i < \log A$ (8)

The experiment continues;

(2) if $\sum z_i \geq \log A$ (9)

Hypothesis H_1 is accepted and H_0 is rejected;

(3) if $\sum z_i \leq \log B$ (10)

Hypothesis H_0 is accepted and H_1 is rejected.

5. Sequential research for measurable characteristics

At the basis of building the control plan, are the principles of verifying the statistical hypotheses. The sequential sampling must validate one of the hypotheses: $H_1: P \leq P_1$, the effective weight of cases is under an admissible threshold, versus

$H_2: P \geq P_2$, the weight of cases is over a threshold defined as tolerable.

As the research is made using surveys, the decisions will be affected by errors such as: type I error (α), which measures the probability of rejecting hypothesis H1, and type II error (β), which measures the probability of accepting the false hypothesis. The case proposed in research assumes a normal distribution of the variable under study, this having a superior limit.

We say that a continuous random variable X follows a normal law if the probability density is:

$$f(x;m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \quad (11)$$

Where *x* and *m* are real numbers, and 6 is a positive real number.

Parameters *m* and σ^2 completely determine the normal law and have the significance of the average and the theoretical variance of the random variable X respectively.

The normal distribution with parameters $m = 0$ and $\sigma = 1$ is called **standardized normal distribution**, and the distribution function for the **standardized normal variable** $U = \frac{x-\mu}{\sigma}$ is given by:

$$f_0(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \text{ and } F_0(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt \text{ respectively} \quad (12)$$

The quality characteristic has a superior limit. From the known expression we obtain:

$$P_1 = \frac{1}{2} - \Phi\left(\frac{LS - m_1}{\sigma}\right) \quad (13)$$

In the case that hypothesis $H_1 : P \leq P_1$ is just and respectively:

$$P_2 = \frac{1}{2} - \Phi\left(\frac{LS - m_2}{\sigma}\right) \quad (14)$$

In the case that $H_2 : P \geq P_2$ is just.

We shall note: m_1 – the characteristic average when $P \leq P_1$ and m_2 – the characteristic average when $P \geq P_2$.

It results that, in hypothesis H_1 the standard deviation is $z_1 = \frac{LS - m_1}{\sigma}$, and in hypothesis $H_2 : z_2 = \frac{LS - m_2}{\sigma}$.

Thus, in hypothesis H_1 the characteristics average m_1 can be written as $m_1 = LS - z_1 \cdot \sigma$, while in hypothesis $H_2 : m_2 = LS - z_2 \cdot \sigma$.

Obviously, when the characteristic has a superior limit, $m_1 \leq m_2$, the decision of accepting or rejecting hypothesis H_1 is made based on the probability report test $\mathbf{P}_{2,n}/\mathbf{P}_{1,n}$.

Determining the probabilities $\mathbf{P}_{1,n}$ and $\mathbf{P}_{2,n}$ is made based on the normal law model, as follows:

In the hypothesis $H_1 : P \leq P_1$, i.e. the hypothesis that the characteristic X average is m_1, the probability that after n successive samplings values $x_1, x_2, \ldots, x_n$ are obtained is:

$$P_{1,n} = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left(- \frac{\sum_{i=1}^{n} (x_i - m_1)^2}{2\sigma^2} \right) \tag{15}$$

Similarly, in case of hypothesis H_2, we have

$$P_{2,n} = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left(- \frac{\sum_{i=1}^{n} (x_i - m_2)^2}{2\sigma^2} \right) \tag{16}$$

As the risks α and β are fixed through the report

$$\frac{P_{2,n}}{P_{1,n}} = \frac{\frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left(- \frac{\sum_{i=1}^{n} (x_i - m_2)^2}{2\sigma^2} \right)}{\frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left(- \frac{\sum_{i=1}^{n} (x_i - m_1)^2}{2\sigma^2} \right)} \tag{17}$$

The three conditions in taking decisions are obtained from the relations:

- Acceptance condition

$$\frac{\exp\left(-\frac{\sum_{i=1}^{n}(x_i - m_2)^2}{2\sigma^2}\right)}{\exp\left(-\frac{\sum_{i=1}^{n}(x_i - m_1)^2}{2\sigma^2}\right)} \le \frac{\beta}{1-\alpha} \tag{18}$$

- Rejection condition

$$\frac{\exp\left(-\frac{\sum_{i=1}^{n}(x_i - m_2)^2}{2\sigma^2}\right)}{\exp\left(-\frac{\sum_{i=1}^{n}(x_i - m_1)^2}{2\sigma^2}\right)} \ge \frac{1-\beta}{\alpha} \tag{19}$$

- Control continuance condition

$$\frac{\beta}{1-\alpha} \angle \frac{\exp\left(-\frac{\sum_{i=1}^{n}(x_i - m_2)^2}{2\sigma^2}\right)}{\exp\left(-\frac{\sum_{i=1}^{n}(x_i - m_1)^2}{2\sigma^2}\right)} \angle \frac{1-\beta}{\alpha} \tag{20}$$

Applying a logarithm, we obtain the acceptance and rejection numbers respectively, by the linearization of the expressions

$$\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - m_1)^2 - \sum_{i=1}^{n}(x_i - m_2)^2\right] \le \ln\frac{\beta}{1-\alpha}, \tag{21}$$

$$\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - m_1)^2 - \sum_{i=1}^{n}(x_i - m_2)^2\right] \ge \ln\frac{1-\beta}{\alpha} \tag{22}$$

These expressions can also be written as

$$-\frac{m_1-m_2}{\sigma^2}\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i+\frac{n}{2\sigma^2}\left(m_1^2-m_2^2\right)\le \ln\frac{\beta}{1-\alpha} \qquad (23)$$

$$-\frac{m_1-m_2}{\sigma^2}\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i+\frac{n}{2\sigma^2}\left(m_1^2-m_2^2\right)\le \ln\frac{1-\beta}{\alpha} \qquad (24)$$

The average is estimated based on the data, $\bar{x}=\sum_{i=1}^{n}x_i/n$, which is why the values for which hypothesis H1 can be accepted or rejected must be established. Because in these relations we have $\sum_{i=1}^{n}x_i$, it is more convenient to find acceptance and rejection conditions for this sum and not for the average, thus obtaining

$$\sum_{i=1}^{n}x_i\le\underbrace{\frac{\sigma^2}{m_1-m_2}\ln\frac{1-\alpha}{\beta}}_{h_1}+\underbrace{\frac{m_1-m_2}{2}}_{k} \qquad (25)$$

or:

$$\sum_{i=1}^{n}x_i\le h_1+kn \qquad (25')$$

Respectively

$$\sum_{i=1}^{n}x_i\ge\underbrace{\frac{\sigma^2}{m_1-m_2}\ln\frac{\alpha}{\beta}}_{h_1}+n\underbrace{\frac{m_1-m_2}{2}}_{k} \qquad (26)$$

This leads to:

$$\sum_{i=1}^{n}x_i\ge h_2+kn \qquad (26')$$

So the *acceptance* and *rejection numbers* are given by the expressions

$$A_n=h_1+kn$$

$$R_n=h_2+kn$$

Where n = 1, 2, 3 … represent the order numbers for the products extracted from the lot and under control.

Using decimal logarithms, we have

$$h_1 = 2{,}3026 \frac{\sigma^2}{m_1 - m_2} \lg \frac{1-\alpha}{\beta} \quad (27)$$

$$h_2 = 2{,}3026 \frac{\sigma^2}{m_1 - m_2} \lg \frac{\alpha}{1-\beta} \quad (28)$$

and $K = \frac{m_1 + m_2}{2}$.

We mention that in this case, if for the controlled characteristic only a superior limit is fixed (LS), the size of h_1 is negative.

6. Numerical example and conclusion

In a college, the management is preoccupied with reducing alcohol consumption, starting from the connection between excessive use and delinquency. In this respect, a research was initiated among the pupils in the last to grades. The control characteristic is "weekly dosage of refined alcohol", which, according to nutritional standard mustn't exceed the value of 60 + 0.10 grams.

The investigation method used is the sequential alternative developed for the case of "snowball". The formulated hypotheses consider the proportions if alcohol consumers: P_1=0.001; P_2=0.009, and the risks of erroneous appreciations 0.03 and 0.06.

We compute the theoretical averages for the case of the two hypotheses H_1 and H_2 :

$$k = \frac{m_1 + m_2}{2}$$

$$m_1 = LS - z_1\sigma = 60.1 - 3.09 \cdot 0.025 = 60.02275$$

$$m_2 = LS - z_2\sigma = 60.1 - 2.37 \cdot 0.025 = 60.04075$$

The values 3.09 and 2.37 represent the Laplace function values for variables P_1 and P_2 for the alcohol consumers' weights.

The average of averages m_1 and m_2 is: $k = 60.03175$

We establish variables h_1 and h_2:

$$h_1 = 2.3026 \frac{(0.025)^2}{-0.018} \lg \frac{0.97}{0.06} = -0.09657$$

$$h_2 = 2.3026\frac{(0.025)^2}{-0.018}\lg\frac{0.03}{0.94} = 0.11953$$

The decision equations are:

A_n = -0.09657 + 60.03175n

R_n = 0.11953 + 60.03175n

The investigation is performed as follows: from the population of pupils, we interview a pupil at a time, who declares the level of consumption. The cumulated registered values $\left(\sum_i x_i\right)$ are compared to the acceptance and rejection numbers corresponding to the order number of the interviewed pupils.

Thus, after subject with number 133, there is the acceptance decision, because $\sum_{i=1}^{33} x_i < A_{33}$, as can be seen in the table below, which presents the computations results.

The values of acceptance and rejection limits

n	1	2	3	...	33
A_n	59.935	119.970	179.999	...	1981.038
R_n	60.151	120.183	180.215	...	1981.167

Thus, $\sum_{i=1}^{33} x_i = 1966.2 < 1981.038$, and in the end we accept the first hypothesis: the weight of alcohol consumers within the maximum dosage limit is under 1%.

We can appreciate that the proposed statistical method can complete the sociologic research procedure of the “snowball” through extra precision which accompanies the field researcher’s intuition in continuing or stopping an investigation, introducing a investigation procedure with probabilistic decision (risk affected), but objective relative to the initial opinion. The management of such investigations will generate additional precision as well as an optimal investigation expenses level.

References

[1] Aaron, D., J., Chang, Y., F., Markovic, N., La Porte, R., E. (2003): Estimating the lesbian population: a capture-recapture approach, JECH online, BMJ Publishing, 57, page 207 - 209.

[2] Adams, S., Carter, N., Hadlock, Ch., (2008): Change in Connectivity in a Social Network over Time: A. Bayesian Perspective, CONNECTIONS, Vol. 28, Issue 2, page 17 - 27.

[3] Adler P.,(1990): Requirements for Inductive Analysis , NIDA Research Monograph 98, Rockville, MD 20857

[4] Atkinson, R., Flint, J., (2001): Accessing Hidden and Hard-to-Reach Populations: Snowball Reseaech Strategies.: Social Research Update, No. 33, http://www.soc.surrey.ac.uk/sru/SRU33.html (19. 3. 2007).

[5] Avico, U., Kaplan, C., Korczak, D. and Van Meter, K., (1988): Cocaine epidemiology in three European Community cities: a pilot study using a snowball sampling methodology. Brussels, European Communities Health Directorate.

[6] Barlow, J., Davis, H., McIntosh, E., Jarrett, P., Mockford, C., Stewart-Brown, S., (2007): Role of home visiting in improving parenting and health in families at risk of abuse and neglect: results of a multicentre randomized controlled trial and economic evaluation. Archives of Diseases in Childhood, 92, page 229 - 233.

[7] Bârsan-Pipu, N., Isaic-Maniu, Al. Vodă, V., (2002): Decision procedures in multivariate statistical process control. Economic Computation and Economic Cybernetics Studies and Research, vol. 36, no. 1 - 4, page 77 - 85.

[8] Berg, S., (1988): Snowball sampling, (in Kotz, S. and Johnson, N., L., Encyclopedia of Statistical Sciences, Vol. 8) John. Wiley and Sons, NY.

[9] Borgatti, S., P., Everett, M., G. and Freeman, L., C., (2002): Ucinet - Software of Social Network Analysis User`s Guide. Harvard: Analytic Technologies, Inc.

[10] Bur, Iv., (1953): Fundamental Principles of Sequential Analysis. Industrial Quality Control, vol. 9

[11] D`Onise, K., Wang, Y., McDermott, R., (2007): The Importance of Numbers: Using Capture-recapture to make the Homeless Count in Adelaide: „Australian Journal of Primary Health, 13(1), page 89 - 96.

[12] Day, C. & Davis, H., (2006): The effectiveness and quality of routine child and adolescent mental health care outreach clinics. British Journal of Clinical Psychology, 45, page 439 - 452.

[13] Diamond, M., (1993): Homosexsuality and Bisexuality in Different Populations. Archives of Sexual Behaviour, 22 (4), page 291 - 310.

[14] Dodds, P., S., Muhamad, R., Watts, D., J., (2003): An experimental study of search in global social networks. Science, 301, page 827 - 829.

[15] Duncan, DF, JB White & Nicholson, T., (2003): Using internet-based surveys to reach hidden populations: Case of nonabusive illicit drug users`, American Journal of Health Behavior, 27(3), page 208-218. http://www.duncan-associates.com/hiddenpop

[16] Etter, J., F. , Perneger, T., V., (2000): Snowball sampling by mail: application to a survey of smokers in the general population. International Journal of Epidemiology, 29 (1), page 43 - 45.

[17] Fritz, R., B,(1990).: Computer Analysis of Qualitative Data, NIDA Research Monograph 98, Rockville, MD

[18] Goodman, L., A., Snowball sampling. Annals of Mathematical Statistics, 32, page 148 - 170.

[19] Griffiths, P., Gossop, M., Powis, B., Strang, J., (1993): Reaching hidden populations of drug user by privileged access interviewers: methodological and practical issues, Addiction, vol. 88, page 1617 - 1626.

[20] Hanneman, R., A., (2001): Introduction to Social Network Methods, http://faculty.ucr.edu/hanneman/SOC157/

[21] Hendricks, V., M., Blanken, P. and Adriaans, N., (1992): Snowball Sampling: A Pilot Study on Cocaine Use. Rotterdam, IVO.

[22] Isaic-Maniu, Al. Vodă, V., (2009): The Poisson property in the case of some continuous distributions. Romanian Statistical Review, No. 2, page 56 - 64.

[23] Isaic-Maniu , Al.; Voda V.Gh. (2008) : Some Comments on an Entropy – Like Transformation of Soleha and Sewilam : „Economic Computation and Economic Cybernetics Studies and Research, nr. 1-2 , ,pag.5-11

[24] Isaic-Maniu , Al.; Voda V.Gh (2009): Some Inferences On The Ratio Average Lifetime/Testing Time In Acceptance Sampling Plans For Reliability Inspection; "Electronic Journal Of International Group On Reliability" ; vol.2 No.3 (14) ,september; pp.51-58

[25] Johnson, T., P., (2005): Snowball Sampling. Encyclopedia of Biostatistics, John Wiley et Sons.

[26] Johnson, J. M.,(1975):. Doing Field Research., New York: Free Press, 1975.

[27] Johnson, B.D.; Goldstein, P.J.; Preble, E.; Schmeidler, J.; Lipton, D.S.; Spunt, B.; and Miller, T. (1985):Taking Care of Business: The Economics of Crime by Heroin Abuser., Lexington: Lexington Books

[28] Kirkpatrik, S.,Barlow,J.,Stewart-Brown, S.,Davis, H., (2007): Working in partnership: user perceptions of intensive home visiting. Child Abuse Review, 16, page 32 - 46.

[29] Kolaczyk, E., D., (2009): Statistical Analysis of Network Data. Hardcover, ISBN: 978.0-387-88145-4.

[30] Koskinen, J., Snijders, T., (2007): Bayesian inference for dynamic social network data. "Journal of Statistical Planning and Inference", 137 (2007), page 3930 - 3938.

[31] Kuhar, R., Švab, A., (2008): Researching hard-to-reach social groups: the example of gay and lesbian population in Slovenia. Social Research - Journal for General Social Issues, 3, page 529 - 552.

[32] Martino, F., Spoto, A., (2006): Social network analysis: a brief theoretical review and further perspectives in the study of information technology. University of Padova, Psychology Journal, 4 (1), page 53 - 86.

[33] Lambert,Y.,Elz.(1990): The Collections and Interpretation of Data from Hidden Populations;NIDA Research Monograph 98,Rockville MD.

[34] Pârțachi, I., Isaic-Maniu, Al., Vodă, V., (2008): Controlul statistic al calității. ASE-Bucharest, ASEM, Chişinău.,Moldova Republic.

[35] Paspalanova, M., (2006): Identifying and assessing the development of populations of undocumented migrants: the case of undocumented Poles and Bulgarian in Brussels, Catholic University of Leuven, Belgia.

[36] Peterson, E., (2005): Evaluation of digital libraries using snowball sampling, First Mondey, volume 10, number 5.

[37] Salganik, M., (2006): Variance Estimation, Design and Sample Size Calculations for Respondent - Driven Sampling. "Journal of Urban Health", , Nov. 83 (Suppl. 1), page 98 - 112.

[38] Salt, J.; Hogarth, J., (2000): Migrant trafficking and human smuggling in Europe - A review of the evidence in IOM Migrant Trafficking and Human Smuggling in Europe, IOM, Geneva.

[39] Skeldon, R., (2004): Migration, the Asian financial crisis and its aftermath, "International Migration and the New Millennium". Aldershat, Ashgate, page 57 - 74 (http://books.google co.uk/books).

[40] Song Hoon Lee, Pan-Jum Kim, Hawoong Jeong, (2006): Statistical properties of sampled networks. Physical Review.

[41] Spreen, M., Coumans, M., (2003): A note on network sampling in drug abuse research. Maastricht University & Addiction Research Institute Rotterdam, INSNA.

[42] Standfort, T., (1997): Sampling male homosexuality. "Researching sexual behavior: Methodological issuer", page 261 - 275, Bloomington: Indiana University Press.

[43] Streeton, R., Cooke, M. & Campbell, J., (2004): Researching the researchers: using a snowballing technique, Nurse Researcher, 12 (1), page 35-47.

[44] Van Meter ,K.M. (1990) : Methodological and Design Issues Techniques for Assessing the Representatives of Snowball Samples; IDA research Monograph 98; www.informaworld.com/index/713660263.pdf

[45] Van Meter, K.M. Block-(1985) :Modeling and cross-classification techniques for estimating population parameters from network data., Workshop on the Methodology of Applied Drug Research, European Community Health Directorate, Luxembourg

[46] Wald, A., (1936): On some systems of equations of mathematical economics. Economica, vol. 19, October, page 368 – 463(translation by Otto Eckstein).

[47] Wald, A., (1947): Sequential Analysis, John Wiley, New York.