Tikrit Journal of Administrative and Economic Sciences, Vol. 20, No. 68, Part (2): 493-505
Doi: www.doi.org/10.25130/tjaes.20.68.2.26

Tikrit Journal of Administrative and Economics Sciences
مجلة تكريت للعلوم الإدارية والاقتصادية
EISSN: 3006-9149
PISSN: 1813-1719

The Sample Size Determination Strategy in the Simple Random Sampling Design

Hanaw Ahmed Amin*
College of Science/ University of Sulaimani

Abstract: This paper presents an overview for some strategies concerning the sample size determination including simple random sampling technique in the applied field of science. For this purpose, we introduce details of most commonly used formula for estimating the sample size in the case of dichotomous and continuous outcomes with presenting several effective criteria are affect the size determination. These factors are sampling error or (level of precision), confidence interval CI, and degree of variability (heterogeneity) between unites, proportion of event occurrence (p), population size, and study design. In general level of precision is pointed as ±10% for governmental politics voting survey, ±5% for marketing study, and ±1% for health and industrial purpose survey. This measure is affected by the sample size and can be denoted by SEM if descriptively computed and inferentially can be calculated by (t- test) statistics. In the case of dichotomous outcomes, the proportion of success (p) needs to be stated carefully, because the sample size directly affected by any increasing or decreasing changes volume. Beside these criteria presenting an outline of eight generally used techniques precisely to defined the sample size. We conclude that one technique for determining the sample size is not appropriate for solving all problems. A Specific research problem are needs some specific sample size techniques since initially they based on three different measurements of data (ordinal, nominal or ratio).

Keywords:
Simple Random Sampling, Sample Size Calculation, Sampling Error (e), Confidence Interval (CI), Degree of Variability, Sampling Proportion (P).

A R T I C L E I N F O

Article history:
Received 28 Jul. 2024
Accepted 18 Sep. 2024
Available online 31 Dec. 2024

*Corresponding author:
Hanaw Ahmed Amin
College of Science/ University of Sulaimani

Tikrit Journal of Administrative and Economic Sciences, Vol. 20, No. 68, Part (2): 493-505
Doi: www.doi.org/10.25130/tjaes.20.68.2.26

استراتيجية تحديد حجم العينة في تصميم المعاينة العشوائية البسيطة

هناو احمد امين
كلية العلوم/ جامعة السليمانية

المستخلص

في هذا البحث تم مراجعة واستعراض بعض المعايير والاستراتيجيات المتعلقة بطرق الاختيار العينة والهدف منها دراسة طرق مختلفة والمعايير المؤثرة في تحديد حجم العينة منها خطأ المعاينة، حدود الثقة، ونسبة الاختلاف او (عدم التجانس) للتوزيع المفردات المجتمع الدراسة. وبالتالي استعراض بعض للاستراتيجيات مختلفة في الاختيار حجم العينة عندما يكون اسلوب المعاينة بطريقة العشوائية البسيطة متعلقة بالبيانات من النوع الثنائي الحدين (Dichotomous) وفي حالة البيانات المستمرة (Continuous). تتأثر خطأ المعاينة (MSE) بنسبة حجم العينة ويمكن تقديرها من خلال قسمة انحراف المعياري عل جذر التربيعي للعينة (n)، اما في الاستدلال الاحصائي يمكن تقديرها باستخدام اختبار (t- test). وتختلف نسبتها حسب نوع الدراسة، في استبانة التصويت والمسائل السياسية يكون قيمهتا ±10%، في استبانة التسويق والاقتصاد يكون 5%±، اما في الدراسات الطبية والتصنيع يحدد قيمتها ب ±1%. بعد ذلك نوصي تقدير احتمالية نسبة النجاح الظاهرة (p) بدقة لأنها تتغير بشكل طردي مع تقدير الحجم العينة، اي تغير تصاعدي او تنازلي في نسبة النجاح تتغير معها الحجم بشكل تصاعدي او تنازلي. من هنا نرى اختيار حجم العينات تختلف مع نوع الدراسة. بجانب هذه المعايير، تم توضيح ثمانية اساليب مختلفة بدقة لتقدير حجم العينة. نستنتج انه لا يمكن استخدام تقنية واحدة لتحديد حجم العينة لان هناك انواع مختلفة من العوامل او المتغيرات داخلة في الدراسة، اما تكون الاسمية او الترتيبية او النسبية (المتغير من نوع المستمر) لذلك تتغير الطريقة تقدير حجم العينة مع تغير نوع المتغير او عدد العوامل في الدراسة حسب خطة البحث.

الكلمات المفتاحية: خطأ المعاينة (e)، حدود الثقة (CI)، مستوى التجانس, حجم العينة (n)، نسبة المعاينة (p).

1. Introduction

The collected data from the sample survey of observational or experimental study can be organized through a statistical method to help researcher making valid decision about inference of population. In applied research, there are important primitive concepts that must be understood before starting the analysis. First the descriptive statistical methods are used to measure the characteristics of unites. Including frequency measures such as rates, ratios and proportions. Furthermore, the sampling techniques are used for recovering evaluation at minimum cost and less time with more accuracy. Besides these points the accuracy selection of sample size is particularly important factor of any research, because in cases of too small or too large sample size the study would give more biased and complex estimation about the populations. (Dell, 2002). Too small size can't estimate

the main effects between variables or relations that leads to biased estimation, too large sample size gives complex estimation. Therefore, the selection sample size is a necessary case of any scientific research. Also, we not forget that the size is affected by the aspect and the correct plan of the study. (Jan and other 2020).

Observing the literatures, each researcher stated different criteria for the determination of sample size, principal guiding, formulas, and tables. There are different approaches for calculating sample size for any data research designs. We introduce and mentioning some correlated literature to our paper objective. (Singh, Masuku 2014).

Many different approaches are considered for estimating the sample size, relying on previous similar studies or experimental studies, calculation formula relating sample size estimation, use available tables, or basically predicted formula derived by (Cochran, 1963), (Israel, 1992) and (Nanjundes. & Shilpa 2021).

(Miaoulis & Michener, 1976: 17) argued that with the purpose of the study and population size effect there where three factors or criteria affect the sample size selection are degree of the precision, confidence interval CI, and the level of dispersion in the data.

(Lenth, 2001: 187) considered that due to theoretical, practical, economic, less time, accuracy results importance, scientific methods must be used to correctly determine the required sample size. In selecting the sample size before analysis, experts, specialists, and statistical analysts with scientific competence and skill can be consulted before conducting the study (Sathian et al.,2010: 4). An undersized size exposes objects to potentially unsafe treatments effect.

In medical study (Sathian et al. 2010: 4) discuss some factors that affect the sample size, level of significant, power of the test, event rate and study design. He exposed that level of significant and power of the study need to be stated previously before determining the sample size. (Singh and Masuku 2014: 124).

(Rao, 2012: 660) suggested in both studies descriptive and analytical statistics that the sample size is an important factor to the research plan or design. (Tejada & Punzalan, 2012: 129) they gives advice for researchers about Solvin's formula that used in sampling survey. This formula needs to be used carefully.

(Singh & Masuku, 2014: 124) they considered that in general the sample size is affected by five different factors: the sample size effect or consistency, predicted measure of variability, power of the test, level of significance, last one is the one- or two-tailed hypothesis statistical test analysis were desired".

In clinical studies (Chow, 2017) stated that the sample size calculation has a main effect on accurate analytical result, especially when patients respond to medications, follow-up must be done with caution.

(Jan, Paulo, Andhee and Bobby 2020), according to the simulation study analysis statistical modeling, they provide a new strategy for determining the sample size depending on an initial sample approach were applied to four different regression models specifically: linear, quadratic, cubic, and power regression model. From the coefficient of determination R^2 they decide that the regression model was the best for estimating the powerful size of sample.

Nanjundeswaraswamy & Divakar (2021) introduced that for the health science research the random probability sampling strategies are an appropriate and they considered population size, CI proportion for categorical sample data, standard deviation SD for ordinal sample data, and essential level of precision from the study as the best criteria for estimating the sample size.

This article, provides a summary of the most frequently used techniques calculation for estimating an adequate sample size processes. Including some special factors such as sampling error or (level of precision), confidence interval CI, and degree of variability (heterogeneity) between unites, proportion of occurrence (p), population size in the case of simple random sampling survey containing several specific methods commonly used in applied sciences.

2. **Sample Size Criteria:** Subsequent to the objective of the research with population proportions, three principles typically will have required for identifying the correct sample data: the amount of precision, the confidence level "CI", and the point of uncertainty in the characteristics determined (kish, 1965: 78). All of these is explained further down.

2-1. Sampling Error or Level of Precision: Statistically, is the variety in which the exact value of the population is expected to be exist. This level is frequently stated in percentage scales, (± 5 percentage), for instance, if we

Doi: www.doi.org/10.25130/tjaes.20.68.2.26

stage on a ruler five times one after the other, a specific measure would provide us the similar value every time. Scientifically, level of precision of any values can be computed from the range of data, the mean deviation, or the standard deviation. (Yamane, 1967: 886).

2-2. The Confidence Interval: The confidence interval is constructed on concepts contained by the measurements of central tendency theorem. The main indication involved in the central tendency measurements is that when a population is sampled frequently, the mean score of the features achieved from samples is equivalent to the real population result. (Jan, Paulo and other 2020: 17) Also, the resells achieved from different sampling are dispersed normally around the real result estimated, when approximately selecting samples a larger or smaller size than the real population value. For normal distribution according to empirical rule nearly about 95% of the sample data are inside two SD of the arithmetic mean of the real population scale such as arithmetic mean. (Nanjun and Shilpa. 2021: 25)

This explain that, if a 95% confidence level is carefully chosen, 95 out of 100 samples will have the same population mean value that in section 2 identified before and (Figure 1). Sometime by chance have cases that a sample achieved does not denote the real population scale. In such case, data contains extreme outlier measures are represented in the shade regions in Figure 1. This riskiness is decreased when CI is increased to 99% and increased when CI decreased to 90%. (Cochran, 1963: 11).

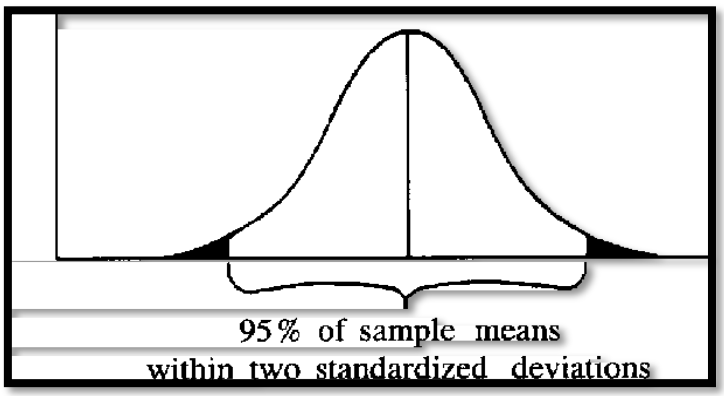

Figure (1): Repeated Sampling Distribution of Means

2-3 Degree of Variability: Describe the distribution of the individual in the population. This is relative to the degree of variability or sampling error, a population with a higher non- homogeneous, the large sample is necessary to achieve a certain degree of precision. A higher homogeneous population,

a small sample survey needs to a chive presided level of sampling error. (Nanjun, Shilpe 2021: 25).

3. **Sampling:** The history of sampling technique was stated in 1786, by Pier Simon Laplace predicting. The whole population of France depending on ratio predictors by sample size technique. And estimating probabilistic predictor of the error. At 1870 a sample survey was considered by Alexander Ivanovich Chuprov to Russian Imperial (Cochran 1963: 75) and (Robert et al 2004: 99). (Kish, 1965: 78) and (Kappor 1970: 78) defined sampling by different ways for selecting the observation from the population and own the characteristics and features of the population from which it was withdrawn. (Cochran, 1963: 75).

   Sampling is a method for selecting a sample from the population under the study, where sample is a small part and a best representative of the population. In order to obtain the best possible accuracy of the results, this requires choosing an appropriate sampling method to represents the population in the best way. In addition, the sample size must be within reasonable limits, so that it is not so large that it leads to spending a lot of time, effort and material resources, and it is not so small to achieve valid accuracy of the results. (Simth, 1983: 4).

4. **Types of Sampling Methods:** Sampling is generally divided into two main types: probability random sampling (random sampling) and non-probability random sampling (non- random sampling). (Sudman, 1976: 56)

   4-1. Probability Random Sampling Methods: The most important of random sampling methods are simple random sampling, stratified sampling, cluster sampling, systematic random sampling, multistage random sampling. In these methods the researcher has no any role in choosing this item over that one, i.e. there is a principle of equal chance of occurrence for any subject within this sample. (Singh and Masuku, 2014: 124).

   4-1-1. Simple Random Sampling: the simple random sampling method is a process of selecting a random sample from the population, such that any member of the population's individuals has the same chance of occurrence within the sample's individual. It is necessary to considering the condition of homogeneity among participants. (Singh and Masuku, 2012: 124). For example, if we want to study the causes of smoking among females, it would be noted here that the population under the study is homogeneous, since all members are female and smokers, this allows us to use a simple random

sampling. (Smith, 1983: 19). In order to review the below techniques for determining the sample size, we assume that the sampling method is a simple random method. This sampling provides the efficient unbiased predictor of parameters. (Cochran, 1963: 75).

4-2 Non- Probability Random Sampling: A non-random sampling methodology means that a sample of items selected from the study population in a technique that the researcher has a role in choosing this item rather than that one, and not on a random basis, due to considerations related to the nature of the study field. Among these quota sampling and purposive sampling. (Nanjun & Shilpa, 2021: 17)

5. **Determining Sample Size Strategies:** Her we present some strategies for determining the sample size. Consisting of using a small populations census, using similar studies sample size of similar, consuming published tables, and applying mathematical scientific formula for determining sample size. Details of each method are explained below.

5-1. Using a Small Population Census: A first strategy is a survey appropriate for small population taking all individuals in the population such as less than or equal to 200 and it is called "A Strategy for Small Population Census". These needs providing information for all populations' individual in the study to decrease sampling or chance error, degrees of precision and variability. But the disadvantage of this strategy is not efficient for large sample population, it requires a long time, more much cost for collecting information from all members of a population such as a questioner sampling, effort and precision. (Kish, 1965: 78).

5-2. Using a Similar Study Sample Size Strategy: Second strategy is using similar sample size strategies that studied before. To avoid repeating the same mistakes that occurred before in selecting the sample size, here needs to study all techniques and dimensions which are necessary according to the specializations. (Dell, 2002, 207).

5-3. Using a Published Tables Strategy: A third strategy is using a published table for determining the sample size according to a number of principles. (Israel, 1992: 95). As presented in Table 1 and Table 2 provides the sample sizes that would be appropriate for assumed criteria of CI, level of precision, and degree of variability. (Cochran, 1963: 75). These two tables are works due to two important points. First point, these sample size depends on the existing number of responses in the study, and not interested on

Doi: www.doi.org/10.25130/tjaes.20.68.2.26

sampling techniques were data collected via email or interview. This amount is usually increased to modify an incomplete response answers. Second points, the sample sizes in Table 2, suppose that the data distribution is normal or approximately normally distributed. In the case of violating this assumption, the whole survey may be used. (Yamane, 1967: 886).

Table (1): Sample Size for Sampling Error ± 3%, ± 5%, ± 7 % and ± 10%, CI = 95%, Proportion of Success (P = 0.5).

Size of Population	(n)Sample Size with Different Values of Sampling Error (e):			
	±3%	±5%	±7%	±10%
500	a[(1)]	222	145	83
600	A	240	152	86
700	A	255	158	88
800	A	267	163	89
900	A	277	166	90
1,000	A	286	169	91
2,000	714	333	185	95
3,000	811	353	191	97
4,000	870	364	194	98
5,000	909	370	196	98
6,000	938	375	197	98
7,000	959	378	198	99
8,000	976	381	199	99
9,000	989	383	200	99
10,000	1,000	385	200	99
15,000	1,034	390	201	99
20,000	1,053	392	204	100
25,000	1,064	394	204	100
50,000	1,087	397	204	100
100,000	1,099	398	204	100
>100000	1,111	400	204	100

(1) The main population should be sampled when of the condition of normality is poor or not met (Yamane, 1967: 886).

Doi: www.doi.org/10.25130/tjaes.20.68.2.26

Table (2): Sample size for Sampling Error ± 5%, ± 7%, and ± 10%, CI = 95% and Proportion of Success P = 0.50.

Population Size	(n)Sample Size with Different Values of Sampling Error (e):		
	±5%	±7%	±10%
100	81	67	51
125	96	78	56
150	110	86	61
175	122	94	64
200	134	101	67
225	144	107	70
250	154	112	72
275	163	117	74
300	172	121	76
325	180	125	77
350	187	129	78
375	194	132	80
400	201	135	81
425	207	138	82
450	212	140	82

5-4. Calculating Sample Size Strategies Formula: The above Tables 1 and 2 offers different sample size for a different combination of levels of sampling error or precision (e), CI, and (p) and degree of variability. In the same time there have applications of several calculating formulae for determining the sample size that will be the last approach presented in details below, which is very important for researchers needs to handle it carefully in the application. (Yamane, 1967: 886).

5-5. A Strategy for Computing a Sample Proportions: Cochran's formula explained in Equation 1 for selecting a proportion sample from a large proportion, as stated below: (Cochran 1963: 75)

$$n_0 = \frac{Z^2 pq}{e^2} \dots\dots\dots\dots\dots\dots (1)$$

Which is useable where n_0 is the sample size, Z^2 is the X- axis of the normal distribution curve which incisions of interval (α) at the tailpiece (1 – α equals the selected CI, given by 95%), e is the sampling error of precision, (p) is the expected proportion of a point that is existed in the population, and q is (1-p) proportion of failure or non- existence. The value for Z is estimated

Doi: www.doi.org/10.25130/tjaes.20.68.2.26

using statistical standard normal tables which contain the area under the normal curve.

For example, in the study of systolic blood pressure in the case of unknown variability in the proportion between patient's gender (women and men) who were screened for high blood pressure; so, assume (p = 0.50) (higher variability). Also, assume we want a 95% confidence interval and ±5% precision. The illustration result is:

$$n_0 = \frac{z^2 pq}{e^2} = \frac{(1.96)^2(0.5)(0.5)}{(0.05)^2} = 385 \text{ patient}$$

5-6. Finite Population Correction for Proportions: In the case of small population considerably the sample size be decreased. Usually, all samples selected from small populations contains accuracy details more than obtained from a large population. The sample size n_0 is modified in Equation 2. (Singh and Masuku, 2012: 124).

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} \dots\dots\dots\dots\dots (2)$$

Where n is the sample size and N is the population size. In a clinical survey study of systolic blood pressure for 2000 patients who were screened for high blood pressure. The required sample size is explained below. (Simth, 1983: 11)

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} = \frac{385}{1 + \frac{(385 - 1)}{2000}} = 323 \text{ patients}$$

This correction, "named the finite population adjustment" significantly minimize the required sample size for small populations.

5-7. A Simplified Strategy for Proportions: The calculation formula for the sample size in Tables 2 and 3 were the simplified strategy for (Yamane, 1967:886) used with a 95% of confidence intervals CI and (P= 0.5) as stated in Equation 3.

$$n = \frac{N}{1 + N(e)^2} \dots\dots\dots\dots\dots (3)$$

Knowing that n and N are sample and population size, and e is the sampling error or precision. For calculation:

$$n = \frac{N}{1 + N(e)^2} = \frac{2000}{1 + 2000(0.05)^2} = 333 \text{ patientes}$$

Doi: www.doi.org/10.25130/tjaes.20.68.2.26

5-8. The Sample Size Mean Strategy: The above discussion of tables and strategies measured the proportion for determining the sample size. There are two formulas to calculate sample size for discrete or continuous variables. The first formula is to integrate members into two groups then calculate the sample size depending on proportion (Smith, 1983: 19). The second formula is to use the strategy arithmetic mean for the sample size. This strategy is analogous to the first formula of the proportion, including the measure of variability σ^2 instead of (pxq) and it is formulated by Equation 4.

$$n_0 = \frac{Z^2\sigma^2}{e^2} \dots\dots\dots\dots\dots\dots (4)$$

Where n_0 is the sample size, z is the X- axis of the normal distribution curve that cuts off an region $\propto$ at the tails, (e) is the preferred level of precision (in the equal part of measure as the variance), and σ^2 is the variability of the characteristic in the population. This strategy needs a precis calculation of population variance. Frequently, an accurate evaluation of the variance is not estimated, this is being a disadvantage point of this strategy. Also, we know that almost the sample size can be changed according to the characteristic of the response with unequal and different variance. This is another disadvantage of this strategy, so frequently the formula of proportion is preferred.

6. **Conclusion:** Under the assumption of the simple random sampling design methodology, this article discusses the basis of most useful criteria and strategies for determining the sample size for real- life data application with regards to medical, health, social, educations, agriculture and economic. There are further topics needs to be taken in to concentrations.

A. an alternative way to these strategies and techniques were mentioned above, a simulation study analysis can be used to determine the sample size such as using a different orders of regression analysis model based on the coefficient of determination R2.

B. All of the strategies previously explained depend on the random sampling method collection. However, there are more complex methods such as, stratified random samples, strata, or clusters, it is necessary to calculate the variance of the subpopulations.

C. A sufficient sample size that necessary for multiple regression, analysis of variance (ANOA), or log- linear analysis is 200-500. In the case of descriptive statistics any number of data can be sufficient.

D. In the case of main and sub- groups comparison study the minimum suggested size is 100 observations for the main group or subgroup and for minor group 20 – 50 is required.

E. Under the assumption of normality, the size of (30 – 200) is required if the elements are 20% to 80% proportion of the time, when the curve of data distribution is skewed from the normal distribution a large sample size (N > 200) is required.

F. The most common criteria for determining the sample size in statistical analysis studies is based on level of precision or chance error in the statistical testing hypothesis, confidence interval CI, less variability or more homogeneity between observation, level of significance 5% study design, and existence of normality assumption.

G. The use of the proportion sampling variability determined by (P=0.5) in the computation of the sample size for the ratio usually will yield additional protective large sample size than will be computed by the mean sample size strategy.

H. Usually, we can add a 30% percent of observations to the data in the case of non-responses members in the survey. Also improved by 10% percent in the case inaccessibility connecting to participant. Therefore, increasing 10% and 30% percent of responses to the survey from mailed or interview sampling that makes the expected values of confidence interval and sampling error smallest.

References:

1. Cochran, W. G. 1963. Sampling Techniques, 2nd Ed., New York: John Wiley and Sons, Copyright © 1977, Inc, pp. 1 - 75. ISBN 0-471-16240-X.
2. Dell R, S. Holleran, R. Ramakrishnan (2002). Sample Size Determination. ILAR J 43:207-213.
3. Israel, Glenn D. 1992. “Determining Sample Size”, Fact Sheet PEOD-6, November 1992, University of Florida. National Science Foundation, Research and Development in Industry: 1992, NSF 95-324. Arlington, VA.
4. Jan Homar T A., Paulo V. Cenas, Andhee M. Jacobe, Bobby F. Roaring. 2020. Sample Size Estimation Using Preliminary Sample as Basis: A Simulation Analysis. PSU-Journal of Engineering Technology and Computing Sciences (JETCS). Vol. No. 2, Issue 1 pp. 17-25, June 2020. ISSN 2599-4638.
5. Kish, Leslie. 1965. Survey Sampling. New York: John Wiley and Sons, Inc, pp (78 - 94).

6. LENTH, Russell V. 2001. Some Practical Guidelines for Effective Sample Size Determination. The American Statistician, August 2001, Vol. 55, No. 3. Pp (187 – 193).
7. Miaoulis, Gregore & Michener R. D. 1976. An Introduction to Sampling Dubuq, lowa:endall/Hunt Publishing Company. pp 17. SBN: 9780840316035, 0840316038.
8. Singh AS, MB Masuku. 2012. An Insight Statistical Techniques and Design in Agricultural and Applied Research, World Jr. of Agricultural Sciences, 8(6), 568-584.
9. Singh AS, MB Masuku (2012): Fundamentals of Applied Research and Sampling Techniques, International Jr. of Medical and Applied Sciences, 2(4), 124-132.
10. Singh Ajay S. and Masuku Micah B. 2014 Sampling Techniques & Determination of Sample Size in Applied Statistics Research: An overview. International Journal of Economic, Commerce and Management. United Kingdom, Vol. II, Issu 2(11), pp. (1-22): November 2014. ISSN 23480386.
11. Smith, M. F. 1983. Sampling Considerations in Evaluating Cooperative Extension Programs. Florida Cooperative Extension Service Bulletin PE-1. Institute of Food and Agricultural Sciences. University of Florida. Viii 64 pages; 23 cm.
12. Sudman, Seymour. 1976. Applied Sampling. New York: Academic Press. pp. 56. ISBN-13:978-0126757507.
13. T S Nanjundeswaraswamy, Shilpa divakar. 2021. Determination of Sample Size and Samling Methods in Applied research. Proceeding on Engineering Science. Vpl. 03, No. 1. (2021). 25-32, doi:10.24874/PES03/01.003.
14. Yamane, Taro. 1967. Statistics: An Introductory Analysis, 2nd Ed., New York: Harper and Row, pp 886.