INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN: 2455-0620 Volume - 6, Issue - 7, July – 2020
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87 Impact Factor: 6.719
Received Date: 07/07/2020 Acceptance Date: 25/07/2020 Publication Date: 31/07/2020

Sample size calculation in Bio statistics with special reference to unknown population

Dr. Ashwin Shete*, Ph D (Scholar)
Asst. Prof ,Dept of Rasashastra Evam Bhaishajya Kalpana, Dr D Y Patil College of Ayurved & R C , Pimpri, Pune-18., Email - ashwin3485@gmail.com

Dr. Avinash Ashok Shete, Ph D (Scholar),
MS (Streeroga evam Prasutitantra), Dept of Pratutitantra evam Stree roga, Dr D Y Patil College of Ayurved & R C , Pimpri, Pune-18. Email - dr.avinashshete@gmail.com

Mr. Som Prakash Dube,
Chief statastician, , Dr D Y Patil College of Ayurved& R C , Pimpri, Pune-18.
Email: dubeysom88@gmail.com

Dr. Arati P Dubewar,
Ph. D Guide,Prof & HOD , Dept Of Rasashastra Evam B K, Dr D Y Patil College of Ayurved& R C , Pimpri,Pune-18. MobNo.9850349125,aratidubewar@gmail.com.

Corresponding author : **Dr. Ashwin Shete***, Email - ashwin3485@gmail.com

Abstract: *Research related with medical field requires biostatistics to find appropriate results. Moreover for survey and clinical study, sampling is being used. To determine Sample size is major issue in research. Most of the time population is unknown as well as data is not available for specific territory for specific time. Population is variable at any point of time, as migration, deaths and birth are altering it. Whereas authentic statistical agencies publish data at specific interval.*

To overcome this, biostatistics has formula for sample size. Medico or bio-researchers are facing problem for sample size. Solution to this problem as statistical methods being used is discussed and demonstrated in this article. So that all researchers can be able to find out sample size by their own. The inferences drawn from the sample are applied to whole sample.

Keywords: *Population, Sample size, unknown population, sampling*

1. INTRODUCTION:

In research related with population or human being, sampling is done. As one is unable to study all individual in population, so some individuals are being studies as samples. For example, while cooking rice in home, the cook checks one rice grain by smashing it on finger. If it is smashed it is assumed that rice is well cooked. In this example, one rice smashed by fingle is sample whereas whole rice being cooked is population.

Sampling is the process of inferring something about a large group of elements by studying only a part of it is called as Sampling.[1] Sampling refers to the part of the population so that some inference about the population can be made by stydying the sample. Sampling is used in survey as well as clinical trials in human beings.

Sampling involves selecting a number of units. Sample represents the population. Various sampling methods are used according to research study design, but sample size is the basic requirement. Moreover population or incidence or prevalence is unknown. To overcome this problem , statistical methods being used is discussed and demonstrated here.

2. NEED OF THE STUDY:

Researchers are facing more problems to find sample size, when population is unknown, or SD or μ is not defined. So that this statistical mehod is demonstrated with example.

3. LITERARY REVIEW:

To calculate sample size some terminologies must be known

- Population: Population is a group of units defined accordingly to the aims and objects of the survey or stydy.[2]
- Population Size: Total population being studied is called as population size. To find out, who does and doesn't fit into your group. For example: To find out onset of menarche in school going girls at specific time interval in specific area, one should exclude total female, females above 16years etc. Due to specific inclusion and exclusion criteria, population size become specific and peculiar. It's common to have an unknown number or an estimated range.
- Sample: A finite measurable subset of Statistical individuals in a population is called a sample.[3]
- Margin of error (confidence interval): Errors are inevitable – the question is how much error you'll allow. The margin of error, AKA confidence interval, is expressed in terms of mean numbers. You can set how much difference you'll allow between the mean number of your sample and the mean number of your population. If you've ever seen a political poll on the news, you've seen a confidence interval and how it's expressed. It will look something like this: "68% of voters said yes to Proposition Z, with a margin of error of +/- 5%."[3] This is also known as allowable error. Normally denoted by m and mostly assumed as 5%=0.05.

- Confidence level: It deals with how confident you want to be that the actual mean falls within your margin of error. The most common confidence intervals are 90% confident, 95% confident, and 99% confident[4]. In bio statistics , 95% CI is used. Z value is determined by CI. Its given in below table.

Table 1: Confidence interval with Z Values.

Sr No	Confidence interval	Z Values
1.	90%	1.645
2.	95%	1.96
3.	99%	2.576

- Standard deviation
  This step asks you to estimate how much the responses you receive will vary from each other and from the mean number. A low standard deviation means that all the values will be clustered around the mean number, whereas a high standard deviation means they are spread out across a much wider range with very small and very large outlying figures. Since you haven't yet run your survey, a safe choice is a standard deviation of 0.5 which will help make sure your sample size is large enough.[5]
  Population proportion is assumed to be 50%=0.5 which is denoted by p in formula. Sometimes population proportion is given in formula , as 0.5.

4. MATERIALS AND METHODS :

To find out sample size for unknown population

Formula as

Necessary Sample Size $S = (Z\text{-score})^2 * StdDev*(1-StdDev) / (margin\ of\ error)^2$

$$S = (Z\text{-score})^2 * p*(1-p) / (m)^2$$

S= Sample size for infinite population

Z= Z score

Std deviation=also presented as p in some formula denotes population proportion, assumed to be 50%=0.5

m= Margin of error=0.05

Now in medical statistics Confidence Interval is 95%, so Z value is 1.96.

Putting values in formula,

$$S = (Z\text{-score})2 * p*(1-p) / (m)^2$$

$$S = (1.96)2 * 0.5*(1-0.5) / (0.05)^2$$

$$S = 3.8416 * 0.25 / 0.0025$$

$$S = 3.8416 * 100$$

$$S = 384.16$$

Sample size for unknown population is 384.16. So that approximately 385 sample size can be utilized for unknown population.

5. DISCUSSION:

Sample size for unknown population is calculated with above formula. With this formula, sample size is used for any population for any study. As many new pandemic diseases or life style disorders are found in this era. No definite population prevalence, standard deviation or population mean is not seen. So that this formula gives ray of light for medical researchers.

6. CONCLUSION :

With this formula sample size for unknown population is 384.16

REFERENCES:

1. Dr Yogesh B. Shinde, Chimne, Pawar,Simplified Research methodology and biostatistics, Arihant Excel Book House, Aurangabad,First edition,2019, pg 169
2 Dr Sachin Sarpotdar, Dr Santosh Bhor, Dr Ila Bhor,Edited by Dr Ranade,Research Methodology and Medical Statistics, Manakarnika Publication, Pune, Fourth Edition Feb 2017, / Ranade 269
3, 4, 5 https://www.qualtrics.com/au/experience-management/research/determine-sample-size/

Previous work done

Ralph B Dell, Steve Holleran,Rajshekhar, Ramkrishnana,Sample Size determination, ILAR J. 2002; 43(4): 207–213
https://www.ncbi.nlm.nih.gov/pubmed/?term=Holleran%20S%5BAuthor%5D&cauthor=true&cauthor_uid=12391396