International Journal of Discrete Mathematics
2021; 6(2): 38-44
http://www.sciencepublishinggroup.com/j/dmath
doi: 10.11648/j.dmath.20210602.12
ISSN: 2578-9244 (Print); ISSN: 2578-9252 (Online)

Statistical Perspective Approach to Selection of Sample

Basavarajaiah Doddagangavadi Mariyappa[1, *], Bhamidipati Narasimha Murthy[2], Kyathanahalli Basavanthappa Vedamurthy[3]

[1]Department of Statistics, Dairy Science College, KVAFSU (B), Hebbal, Bengaluru

[2]Department of Biostatistics, National Institute of Epidemiology (NIE), Chennai, India

[3]Department of Economics, Dairy Science College, KVAFSU (B), Hebbal, Bengaluru

Email address:
sayadri@gmail.com (B. D. Mariyappa)

[*]Corresponding author

To cite this article:
Basavarajaiah Doddagangavadi Mariyappa, Bhamidipati Narasimha Murthy, Kyathanahalli Basavanthappa Vedamurthy. Statistical Perspective Approach to Selection of Sample. *International Journal of Discrete Mathematics*. Vol. 6, No. 2, 2021, pp. 38-44.
doi: 10.11648/j.dmath.20210602.12

Received: March 4, 2019; **Accepted**: April 13, 2019; **Published**: October 28, 2021

Abstract: Any research starts with the selection of a problem. Many characteristics or attributes may look for their problems of research *viz* novelty, interesting, importance, feasibility availability of data and hypothesis testing *etc*. In the essence of biological research we should make a formulated hypothesis at greater accuracy and precise level of significance 'α'. The research hypothesis is a presumptive statement of a proportion or a reasonable guess based upon the available evidences or attributes, which the researcher seeks to prove through his course of study. It is also driven from deductive reasoning from a scientific theory. The researcher may begin his study by selecting the sample size is very important dogma in his own area of interest. After selecting the particular theory, the researcher proceeds to derive the hypothesis from his theory. The required sample numerals (size of the sample) is very much concern for success of the research pedagogy and also how much data will require to make a correct decision about the population parameters. If we have accurate sampled data sets, then our decision will be more accurate and there will be less standard error of the parameters estimates of research concern. In this accord the present research paper address the basic principles adopted for sample size determination with respect to biological field.

Keywords: Sample Size, Hypothesis, Swat, Design of Experiments

1. Introduction-Statistical Dealing with Success of Good Research

This note monologue, at a prior level, extensive postulates that apply to many different disciplines of research (Medical, Agriculture, Biological and Veterinary Sciences). Anyone who has a research degree /empathy should be aware of them, whether or not they arise in their own research process. They give, also, pointers that may help in getting a clear view of where the researcher's project headed. Many researchers have sucuseesfully venture to the research cascade itself and in the salient examples. They are several reasons why researchers should take an interest in broad ranging multiple issues at the time of research planning and hypothesis concord. Many assumed hypothesis have been precluded that the following key points will narrate the story of research. The immediate research project may take twists and turns that are different from those for which earlier has been a preparation. This is especially likely for highly applied research projects, which typically demand a range of diverse skills and other variants affected during the study period. Those who acquire a wide range of research skills and knowledge are thereby better placed, after objective finalization to turn their hand to tasks different from those for which their immediate research training has equipped them. As broad based research skills will best equip to nurture for the researchers to respond to changing the subject area, as they move from one task to another. The designing and planning of research is very important tool *or* indicator the instrument panel of a research on large questioner domains may appear like an multifaceted problems arise. It can be really emphasize the critical and questioning role of scientific ways of thinking is the best solution for obtaining trust

worthy results. It does not much matter where you start practicing scientific thinking and project implementation *etc*. Because its explanatory power is so great deal to maintain open pedagogy, once you get the hang of scientific reasoning you are bound to start applying it everywhere during the course of action.

Finally, the scientific criticism and questioning are in tension with the openness to imaginative insight that is equally important to the research continuum process. Data may be in tension with the theoretical and practical insight that generates their collection. Robert Langkjaer B fellow of Royal statistical Society (2019) has quoted that "data are not just numbers without meaning or context. The issue of evidence is central there must be an assessment of the evidence in the literature that is the starting points for the research. There must be a research strategy that will bring together data sets that blissfully address the research questions and hypothesis. Statistical analysis will plausibly extract from the real data evidence must be integrated into the body of the earlier knowledge, creating a coherent account that will assimilation of a research project / articles / thesis and monographs *etc*.

1.1. Research Perspectives in New Horizon

According to research prospective and ethical issues, there is an inherent frankness to new ideas and the ruthless criticism to which the scientific research process insists on extrapolating new idea. As well as research ethical issues, principles and methodological aspects to specify or describe the particular deciplines, there are general statistical principles and methodologies, though avoiding any attempt at rigid prescription of a acceptable scientific procedure. In order to criticize or address the research planning we would establish a frame work that is broad enough for the most of the research projects. The plan should include examination of existing knowledge, a decision on a research question or hypothesis a plan to follow in seeking answers, an analysis of the research data sets and an eventual report formulation.

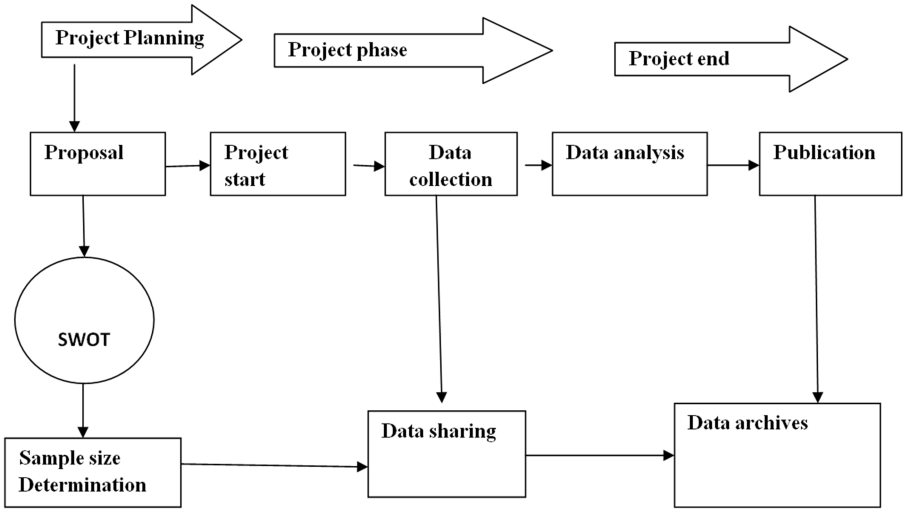

Figure 1. *Schematic Diagram of Research Project flow.*

Many researchers have cited in the research methodological intervention, certain elements have been elucidating the pure objective of research hypothesis. Firstly, the generation of new ideas that would be consider for the assessment or evaluation of data set, eventually the researcher openness to new ideas for solving the realistic problems and approaches to the new scientific interventions and methodological aspects. R. A Fisher cited new approach for conducting research noteworthy way back in 1926, as proved and disproved the hypothesis based on a existing data sets on hypothetical way. If the experimentation lead to have negativity (negative predictive value) we don't called as failure of experimentation. The researchers make quantify the reasons behind failure of the experimentation. Different types of study have been triggered the salient solution for manipulation of data sets through maximization of sample size and 'SWOT' analysis at inception stage. The creativity is very important us to make or draw right decision at right time based on the preliminary study and sample size. The Jextrapose of any project will begin with 'SWOT' analysis Figure 1.

Strength

The ability to widely distributed project information to vast audience.

Weakness

Lack of Knowledge technical information which can lead to misinformation being disseminated, the exposure of information before formal announcement are made.

Opportunities

Promote communication and engagement through various types of media source.

Threats

Disgrantled environmentalists contracting the media.

1.2. Statistical Thinking on Thematic Research Area

In the design of data collection, and in interpreting results, subject area insights should mesh with statistical and data analysis insights in ways that will vary from study to study. The researcher's challenge is to put together all the evidence – evidence from the literature, from the analysis of the researcher's own data, and less formal evidence that may not be amenable to statistical analysis, in a manner that presents a coherent story. This demand for coherence will appear repeatedly in these notes. This section is written from the point of view of a practising statistician who has often been involved in the research of others. A key emphasis is that there must be a correlation of statistical insights with application area insights. There must be shoe leather as well as statistical analysis. Careful planning will greatly increase the chances that, when your data analysis is complete, there will be a compelling story to tell. It is a fortunate researcher whose data tell a story that is as compelling as R. A. Fisher and Bayesian data, or as John Snow's data. Good planning of the project, and of the data collection, can greatly increase the chances of such good fortune.

2. Methods -Formulation and Frame Work of Research Project

The aim of the any new research project is to develop a framework that will be helpful in the later discussion of research projects. Since, it is fragile to get started at all unless there is a research question, or at least the beginnings of a research questions. It will be convenient to group the different components of a research project under the following headings:

1. Define the state variables -Assessment of the state of existing knowledge.
2. Implicit and explicit questions to presume your null hypothesis (Generation and honing of ideas).
3. Formulate proper design of experiment and execution of research that will explore or test specific ideas (Statistical methods for data analysis).
4. Analysis, interpretation and presentation of the resulting data sets.

While statistical ideas may not have much role in idea generation, they are certainly important *viz*., assessing existing knowledge, designing, executing research and data analysis and interpretation *etc*. In case of Veterinary and Medical Science Research emphasis on the review of existing knowledge, an area where the insights of experienced statisticians are sorely needed for guiding the research projects and Interim analysis work. An assessments of how effectively earlier workers have designed their study, and of how compelling their results are, may rely heavily on statistical insights or forethoughts. Even if the study design seems to stand up to critical scrutiny, the funding agency will ask whether the data interpretation is correct. Mistakes in the statistical analysis or in the interpretation of the analysis may lead to quite wrong conclusions, as in some of the examples that we give later. Sound statistical tool will curb the error of sample or population.

2.1. Sample Size Determination

Perhaps the most frequently asked question concerning sampling is, "What size sample do I need?" The answer to this question is influenced by a number of factors, including the purpose of the study, population size, the risk of selecting a "bad" sample, and the allowable sampling error. This section reviews criteria for specifying a sample size and presents several strategies for determining the sample size *etc*. In very important criteria for sample size determination is level of precision, level of confidence or risk and the degree of variability in the attributes being measured. Usually sample size determination has been followed five basic approaches.

Arbitrary approach –Rules of thumb (10% of the population).

Conventional approach-Average of similar studies (what others have done).

Cost basis approach –Availability of resources.

Statistical approach-Statistical consideration (adequate for sub group analysis).

Confidence Interval approach-Concept of variability, sampling decision and SE (x) (allows us to predetermine how precise our estimates are).

Different approaches of Sample size.

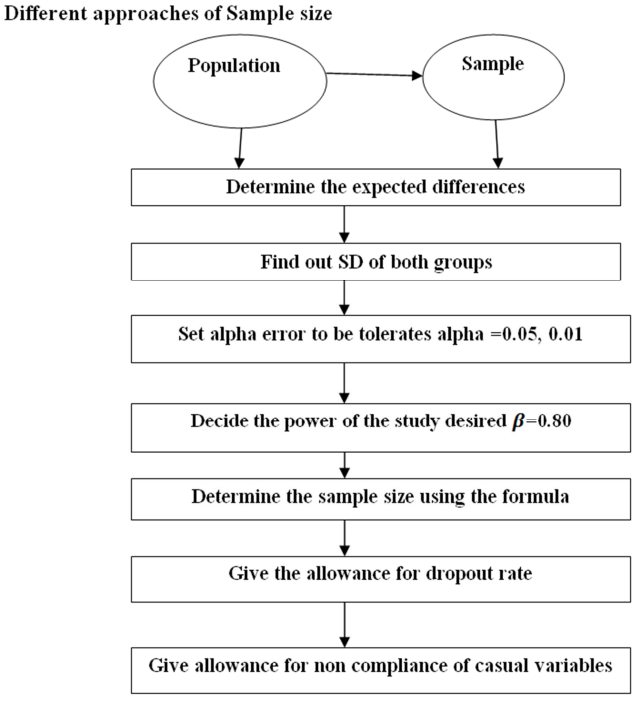

Figure 2. *Flow chart shows sample size determination wrt population and sample.*

2.2. Results

Table 1. *Matrix shows various formulas to determine sample size.*

Types	Sample	Population
Proportion	$n = \frac{Z_\alpha{}^2 pq}{e^2}$ Z_α = Table value for ND p = Estimated proportion of sample $q = (1-p)$ e = precision of the experiment (0.05)	$n = \frac{n_0}{1+\frac{n_0-1}{N}}$ n = sample size n_0 = slected population small in size N = population
Mean	$n = \frac{Z_\alpha{}^2 \sigma^2}{e^2}$ Z_α = Table value for ND σ = SD of the sample e = precision of the experiment (0.05)	$n = \frac{Z_\alpha{}^2 \mu_\sigma{}^2}{Me^2}$ Z_α = Table value for ND μ_σ= SD of the population Me^2 = Marginal error
	$ME = Z_\alpha\sqrt{\frac{P(1-P)}{n}}$ $ME = 1.96\sqrt{\frac{P(1-P)}{n}}$ Formula based on t score $ME = t\left(\frac{\sigma}{\sqrt{n}}\right)$ $n = \left(\frac{\sigma t}{ME}\right)^2$ $N = \frac{N}{1+Ne^2}$ $n = the\ sample\ size, N = the\ Population\ size, e = the\ acceptable\ SE$ $n = \frac{\left(\frac{Range}{2}\right)^2}{\left(\frac{Accuracylevel}{Confidencelevel}\right)^2}$ $size = \frac{\chi^2 np(1-p)}{d^2(n-1)+\chi^2 p(1-p)}$ Where $\chi^2 = table\ value\ of\ chisquare$ N = population size P = population proportion d = degree of accuracy (expression as proportion) Pocock sample size formula $n = \frac{[P_1(1-P_1)+P_2(1-P_2)](Z_\alpha+Z_\beta)^2}{(P_1-P_2)/2}$ n = required sample P_1 = estimated proportion of study outcome in the exposed group P_2 = estimated proportion of study outcome in the un exposed group Z_α = represents the desired level of significance 0.05 Z_β = represents the desired power 0.84 Calculating sample size for independent t test $n = \frac{(M_1-M_2)-(\mu_1-\mu_2)}{(\mu_1-\mu_2)^2}$ $n = \frac{N\sum N_h {S_h}^2}{N^2D^2+\sum N_h {S_h}^2}\ D = \frac{e^2}{t^2}$ $n = sample\ size$ $N_h = Number\ of\ respondents\ in\ acceible\ population$ $S_h = SD\ of\ each\ stratum$ $D^2 = Desired\ variance$ $e = Permitted\ error$ $t = table\ t\ value$	

Source: Kish, Leslie. 1965. *Survey Sampling.*

Table 2. *Sample size for ±5%, ±7% and ±10% Precision Levels Where Confidence Level is 95% and P=0.5.*

Size of the population	±5%	±7%	±10%
100	81	67	51
125	66	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

Size of the population	±5%	±7%	±10%
425	207	138	82
450	212	140	82

Table 3. *Sample size for ±3%, ±5%, ±7% and ±10% Precision Levels Where Confidence Level is 95% and P=0.5.*

Size of the population	Sample size for precision 'e'			
	±3%	±5%	±7%	±10%
500	A	222	145	83
600	A	240	152	86
700	A	255	158	88
800	A	267	163	89
900	A	277	166	90
1000	A	286	169	91
2000	714	333	185	95
3000	811	353	191	97
4000	870	364	194	98
5000	909	370	196	98
6000	938	375	197	98
7000	959	378	198	99
8000	976	381	199	99
9000	989	383	200	99
10000	1000	385	200	99
15000	1034	390	201	99
20000	1053	392	204	100
25000	1064	394	204	100
50000	1087	397	204	100
1000000	1099	398	204	100
>1000000	1111	400	204	100

a=assumption of ND is poor (Yamane, 1967). The entire population should be sampled.
Source data: University of Florida press 1992 fact sheet report, Medical Science -single arm and dual arm study, the table is available in cited ref (7).

Table 4. *Comparison of Independent sample groups based power (Group I&II).*

Percentage of group II	Percent of group I									
	0	10	20	30	40	50	60	70	80	90
10	74									
20	34	199								
30	21	62	293							
40	15	32	81	356						
50	11	20	39	93	387					
60	8	13	23	42	97	387				
70	6	10	14	23	42	93	356			
80	5	7	10	15	23	39	81	293		
90	4	5	7	10	14	20	32	62	19	
100	2	4	5	6	8	11	15	21	34	74

Note: $\alpha = 0.05, \beta = 0.80$.

Table 5. *Matrix for Medical /Veterinary study for sample size determination.*

Population	CI-95%									
	10	20	30	40	50	60	70	80	90	95
60	37									
70	41	18								
80	44	19	19							
90	47	19	18	16						
100	49	20	20	17	16					
110	52	20	20	17	16	15				
120	54	20	20	17	16	14	14			
130	55	20	20	17	16	14	14	13		
140	57	21	20	17	16	14	14	12	10	
150	59	20	20	17	18	14	14	12	10	9
160	60	21	20	17	18	14	14	12	10	9
170	62	21	20	18	18	14	14	12	10	9
180	63	21	20	18	19	14	13	12	10	9
190	64	21	20	18	19	13	12	12	10	9
200	65	22	20	18	19	14	12	12	10	9

Population	CI-95%									
	10	20	30	40	50	60	70	80	90	95
210	66	22	20	18	19	14	12	12	10	9
220	67	22	21	19	19	12	12	12	10	9
234	68	22	21	19	20	13	12	12	10	9
240	69	22	21	19	20	12	12	12	10	10
250	70	22	21	19	20	12	12	12	10	10
>250	71	24	22	19	20	12	12	12	10	11

2.3. Merits of Sample Size Determination

Low cost of sampling –Cost minimization (Reduced cost).

Less consuming in Sampling –Consistency result.

Scope of sampling is high (Error of sample is very less).

Maintain good accuracy and precision of the experiment.

Suitable in very limited resources *wrt* geographical location.

Better rapport of research insight.

2.4. Demerits of Sample Size Determination

Chance of bias at induction or end of the experiment.

Difficulties in selecting random true representative samples during the study period.

Need for subject expertisation.

Changeability of sampling unit due to intrinsic and extrinsic factors.

Plausible changes of sampling unit due to accidental causes (bottle neck effect).

2.5. Level of Precision

The level of precision, sometimes called sampling error (se), is the range in which the true likelihood values of the population is estimated to be. This range is often expressed in percentage numerals (eg. ±5, ±10 percent), in the same way that results for survey on QOL are reported by the Basavarajaiah et al. Opined that 60% of the farmers who are rearing animal husbandry have adopted a recommended scientific practices with precision rate of ±5%, then he can concludes that between 55% and 65% of farmers in the population have adopted the practice.

2.6. The Confidence Level

Central Limit Theorem clearly states that, any state variables (quantitative/qualitative) can be normally distributed with parameters μ and common variance σ^2. If the variable not normally distributed the confidence boundary as lead to be skewness (positive/negative). However, the researcher addressed the issues of variable of interest. As per the theorem the key idea encompassed in the Central Limit Theorem is that, when a large population is repeatedly sampled with large entries, the average value of the attributed data obtained by those samples is equal to the true population mean value μ. Furthermore, μ values were obtained by these samples are distributed normally $N(\mu, \sigma^2)$ about the true value, with some samples having a higher value and some obtaining a lower score than the true population value is accord. In a normal distribution approximation, approximately 95% of the sample values are within two standard deviations of the true population value (e.g. mean). In other words, this means that, if a 95% confidence level is selected, 95 out of 100 samples will have the true population value within the range of precision specified earlier (repetitions of the experimental value). There is always a chance that the sample you obtain does not represent the true population value. This risk is reduced for 99% confidence levels and increased for 90% (or lower) confidence levels.

$$CI\ 95\% = \bar{\mu} + Z_{\alpha\ 1/2} < ND < \bar{\mu} - Z_{\alpha\ 1/2}$$

2.7. Using a sample Size of a Similar Study

Another salient approach is to use the sample size as those of studies similar to the one you plan. Without reviewing the procedures employed in these studies you may run the risk of repeating errors that were made in determine the sample size for another study. However, a review of the literature in your discipline can provide guidance about typical sample sizes which are used in the experimentation.

2.8. Degree of Variability

The important third criteria, the degree of variability in the attributes being measured refers to the distribution of attributes in the population, The more heterogeneous a population, the larger the sample size is required to obtain a given level of precison. The less variable (more homogeneous) a population, the smaller the sample size. The point to be noted that, the proportion of 50% indicates a greater level of variability than either 20% or 80%. This is because 20% and 80% indicate that a larger majority do not or do, respectively, have the attribute of interest. Because a proportion of population is necessary. The research study indicates the maximum variability in a population; it is often used in determining a more conservative sample size *ie* The sample size may be larger than if the true variability of the population attributes were used.

A third way to determine sample size is to rely on published tables which provide the sample size for a given set of criteria. The sample size would be necessary for given combinations of precision, the confidence level and variability have been considered for determination of sample size it was presented in tables 1 & 2 The following note has been inclusion for sample size reflect *viz*., the number of obtained responses and not necessarily the number of surveys mailed or interview planned (this number is often increased

to compensate for non responsive). Second, the sample sizes in table 2 presume that the attributes being measured in various characteristics and also measurements are distributed normally or nearly so. If this assumptions can't be met, then the entire population may need to be surveyed with cost minimization.

One approach is to use the entire population as the sample. Although cost considerations make this impossible for large populations a census is attractive for small populations (eg. 200 or less respondents) Table 4. A census eliminates sampling error and provides data on all the individuals in the population. In addition some costs such as questionnaires design and developing the sampling frame are fixed. They will be the same for samples of 50 or 200 Table 3. Finally, virtually the entire population would have to be sampled in small populations to achieve a desirable level of precision of the experiments or research *etc*.

3. Discussion

As per the previous studies, sample size is very important domain for drawing correct decision about the population and solves the real world problems at larger extent, the insight of study will describe the characteristics of parameters tested. Too small a sample size is more likely to generate inconclusive, incorrect or spurious results. This is because a smaller sample size will generate estimates which have higher variation. These estimates will then be less useful in modelling and understanding the real underlying hypothesis of interest. Secondly, studies which more likely to fail due to inadequate sample size are considered unethical [1-6]. This is because exposing human subjects or lab animal/patients recruitment for the study to the possible risks associated with research is only justifiable if there is a realistic chance that the study will yield useful information [6]. Additionally, a study which is too large faces the same ethical problem and will also waste scarce resources such as money, subjects and time [9]. When conducting research, quality sampling may be characterized by the number and selection of subjects or observations. Obtaining sample size that is appropriate in both regards is critical for many reasons. Most importantly, a large sample size is more representative of the population, limiting the influence of outliers or extreme observations. Sufficiently large sample size is also necessary to introduce results among variables that are significantly different [9]. For qualitative studies, where the goal is to reduce the chances of discovery failure. A large sample size broadens the range of possible data and forms a better picture for analysis. Sample size is also very important for economic, physical, psychological and ethical reasons. As Russell Lenth from the University of lowa explains "An under sized study can be a waste of resources for not having the capability to produce useful results, while an oversized one uses more resources than necessary. In an experiment involving human or animal subjects, sample size is a pivotal issue for ethical reasons [8]. An undersized experiment exposes the subject to potentially harmful treatments without advancing knowledge [7]. In an oversized experiment, an unnecessary number of subjects are exposed to a potentially harmful the experiments or research of interest.

4. Conclusion

In designing experiments or study, sample size calculation is very important for methodological, ethical reasons, as well as for reasons for human and animal intervention and also financial resources. When conducting a research, the researcher should be alert to ascertain that, the study is subjected to sample size determination. In the absence of the sample size determination, the findings of the study should be interpreted with serious caution and elucidate with probable realistic threat. An appropriate sample size renders the research more efficient, data generated are reliable, resource investment is as limited as possible, while conforming to ethical and research perspectives. The use of sample size determination directly influences the research intuition and findings. Researcher would have confine to validate the samples at inception of research study, as a result, eliminate the biasness and hypothetical errors.

References

[1] Fisher RA. 1935. *The Design of Experiments*. Oliver and Boyd.

[2] Cohen I B. 1984. Florence Nightingale. *Scientific American* 250: 98-107.

[3] Nelder J A. 1999. From statistics to statistical science. *Journal of the Royal Statistical Society*, Series D; 48: 257-267.

[4] Sudman, Seymour. 1976. Applied Sampling. New York: *Academic Press*.

[5] Yamane T. 1967. *Statistics*, An Introductory Analysis, 2nd Ed., New York: Harper and Row.

[6] Kish L. 1965. *Survey Sampling*. New York: John Wiley and Sons, Inc.

[7] Basavarajaiah D M. 2017. *Recent statistical techniques in clinical research*. India: Educreation, *Inc.*

[8] Kirby A, Gebski V, Keech AC. Determining the sample size in a clinical trial. *Med J Aust*. 2002; 177: 256–7.

[9] Larsen S, Osnes M, Eidsaunet W, Sandvik L. Factors influencing the sample size, exemplified by studies on gastroduodenal tolerability of drugs. *Scand J Gastroenterol*. 1985; 20: 395–400.

[10] Prakesh B, Babu SR, Sureshkumar K. Response of Ayurvedic therapy in the treatment of migraine without aura. *Int J Ayurveda Research*. 2010; 1: 29–35.

[11] Cady RK, Sheftell F, Lipton RB, O'Quinn S, Jones M, Putnam G, et al. Effect of early intervention with sumatriptan on migraine pain: Retrospective analyses of data from three clinical trials. C*lin Ther*. 2000; 22: 1035–48.