International Journal of Applied Mathematics &
Statistical Sciences (IJAMSS)
ISSN 2319-3972
Vol. 2, Issue 1, Feb 2013, 33-48

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

AN INVESTIGATION ON EFFECT OF BIAS ON DETERMINATION OF SAMPLE SIZE ON THE BASIS OF DATA RELATED TO THE STUDENTS OF SCHOOLS OF GUWAHATI

H. K. SARMAH[1], B. BORA HAZARIKA[2] & G. CHOUDHURY[3]

[1]Department of Mathematics, Gauhati University, Guwahati, Assam, India

[2]Department of Mathematics, Narangi Anchalik Mahavidyalaya, Guwahati, Assam, India

[3]Mathematical Science Division, Institute of Advanced Studies for Science and Technology, Guwahati, Assam, India

ABSTRACT

Survey studies are related to the inferences about a population characteristic under study. The sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inference about a population from a sample. If a sample is not true representative of the target population then it may lead to unreliable conclusions. So the determination of proper sample size using appropriate technique of sampling is vital in this type of studies. In this context, it is very much necessary to have an idea on the effect of bias on determination of sample size. Bias is the difference between population parameter and its estimated value. In the present paper a study on the appropriate method of sampling to gain maximum efficiency along with the effect of bias on the accuracy of an estimate is observed. The study is done on the basis of the students of different types of schools of Guwahati.

KEYWORDS: Methods of Sampling, Sample Size, Bias

INTRODUCTION

Research in every field and more so in the field of education is the demand of the day. Without systematic research, and its application, there would have been very little progress. Research is often defined as scientific thinking. In this regard the report of the University Education Commission of 1948 has the opinion that-"human civilization has derived great benefits from the efforts of specialists who have penetrated deeply into the secrets of nature and the motives and process of human behavior, individual and social. To a constantly increasing extent modern life is the outcome of research"[4]. In order to promote the scientific study of education, the activities like school surveys have great importance[3, 8, 17, 18]. In India the importance of educational research has been recognized rather late. Most of the research in the field of education was conducted after independence. Educational progress and national development go hand in hand. The Indian Education Commission (1964-66), rightly observed: "If the pace of national development is to be accelerated, there is need for a well-defined, bold and imaginative educational policy and for determined and vigorous action to utilize, improve and expand education" [6].

The present century has seen great advancement in scientific and technical knowledge as a result of exploration of knowledge. The rest of the world is marching ahead in every field of human activity. In order to keep pace with modernization, the commission (1964-66) suggested that Education should awaken curiosity, develop interests, attitudes, and build up essential skills as independent study and capacity to think and judge for one self.

In spite of the application of scientific method and refinement of research techniques, tools and designs, educational research has not attained the perfection and scientific status of physical sciences. Therefore, there is a great necessity to study properly about different tools and techniques of research methodology. While studying a particular phenomena, the researchers of this field face a problem at the beginning as what may be the representative sample. A good sample is one which is free from error due to bias and also from random sampling error. ‘Bias’ means flaws of data collection or analysis or of study design that undercut the basic assumptions of the study. Bias may creep into the process of selecting samples from populations and thereby leading to erroneous conclusions about a population which are drawn from biased samples. So, investigators should be on guard and verify if such biases are present in their study. In this context, various studies have been carried out in US education system to see the effect of bias in obtaining information about target population on the basis of responses of supplied questionnaires by the investigators to selected samples [15, 16]. However, very few investigations are carried out in this line in our part of the country.

Hence, in this article an attempt has been made to investigate the effect of bias on determination of a proper sample size through a proper method of sampling. Data related to the schools of Guwahati have been used in our investigation.

In survey studies, once data are collected, the most important objective of a statistical analysis is to draw inferences about the population using sample information. "How big a sample is required?" is one of the most frequently asked questions by the investigators. Sample size calculation for a study, estimating a population has been shown in many books e.g. Cochran (1977), Singh and Chaudhury (1985) and Mark (2005). The aim of the calculation is to determine an adequate sample size to estimate the population with a good precision. In other words one has to draw inference or to generalize about the population from sample data. The inference to be drawn is related to some parameters of the population such as the mean, standard deviation or some other feature like the proportion of an attribute occurring in the population. It is to be noted that a parameter is a descriptive measure of some characteristics of the population whereas if the descriptive measure is computed from the observations in the sample it is called a statistic. Parameter is constant for a population, but the corresponding statistic may vary from sample to sample. The statistical inference generally takes one of the two forms, namely, the estimation of population parameters or the testing of hypothesis.

The process of obtaining an estimate of the unknown value of a parameter by a statistic is known as estimation[7, 13, 14]. There are two types of estimations viz. point estimation and interval estimation.

If the inference about the population is to be done on the basis of the sample, the sample must conform to certain criteria: the sample must be representative [2, 12]. The question arises as to what is a representative sample and how such a sample can be selected from a population.

The computation of the appropriate sample size is generally considered the most important step in statistical study. But it is observed that in most of the studies this particular step has been overlooked. The sample size computation must be done appropriately because if the sample size is not appropriate for a particular study then the inference drawn from the sample will not be authentic and it might lead to some wrong conclusions [9].

The error which arises due to only a sample being used to estimate the population parameters is termed as sampling error or sampling fluctuations. Whatever may be the degree of cautiousness in selecting a sample, there will always be a difference between the parameter and its corresponding estimate. A sample with the smallest sampling error will always be considered a good representative of the population. Bigger samples have lesser sampling error. When the sample survey becomes the census survey, the sampling error becomes zero. On the other hand smaller samples may be easier to manage and have less non sampling error. Bigger samples are more expensive than smaller ones. The non sampling error increases with the increase in sample size [19]. Usually, the study on which researcher works is often based on a limited budget, so this also effects the sample size.

OBJECTIVES OF THE STUDY

In the present study our aim is to analyze the method of determination of sample size along with the procedure of sampling in relation to our study entitled “factors effecting interest in mathematics among primary school students-a study on the basis of the students of Guwahati”.

There are various approaches for computing the sample size [1, 11, 20]. To determine the appropriate sample size the following basic factors are to be considered –

- The level of precision required by users
- The confidence level desired
- Degree of variability.

Level of Precision

The sample size is to be determined according to some pre assigned degree of precision. The degree of precision can be specified in terms of two criteria. The margin of permissible error between the estimated value and the population value. In other words it is the measure of how close an estimate is to the actual characteristic in the population. The level of precision, may be termed as sampling error, is the range in which the true value of the precision is estimated to be. According to W.G.Cochran (1977) precision desired may be made by giving the amount of error that are willing to tolerate in the sample estimates. The difference between the sample statistic and the related population parameter is called the sampling error. It depends on the amount of risk a researcher is willing to accept while using the data to make decisions. It is often expressed in percentage. If the sampling error or margin of error is ±5% ,and 70% unit in the sample attribute some criteria then it can be concluded that between 65% to 75% of units in the population have attributed that criteria.

High levelof precision require larger sample sizes and higher cost to achieve those samples.

Confidence Level Desired

The confidence or risk level is based on ideas encompassed under the Central Limit theorem. The main idea of the theorem is that when a population is repeatedly sampled, the average value of the attribute obtained by those samples is equal to the true population value. Furthermore, the values obtained by these samples are distributed normally about true value, with some samples having a higher value and some obtaining a lower score than the true population value.

Usually 95% and 99% of probability are taken as the two known degrees of confidence for specifying the interval within which one may ascertain the existence of population parameter (e.g. mean). 95% confidence level means if an investigator takes 100 independent samples from the same population, then 95 out of the 100 samples will provide an estimate within the precision set by him. Again, if the level of confidence is 99%, then it means out of 100 samples 99 cases will be within the error of tolerances specified by the precision.

Degree of Variability

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 required to be, to obtain a given level of precision. For less variable (more homogeneous) population, smaller sample sizes works nicely. Note that a proportion of 50% indicates a greater level of variability than either 20% or 80%. This is because 20% and 80% indicate that a large majority do not or do, respectively, have the attribute of interest. Because a proportion of 0.5 indicates the maximum variability in a population, it is often used in determining a more conservative sample size.

FORMULAE FOR DETERMINATION OF SAMPLE SIZE

There are different formulae for determination of appropriate sample size when different techniques of sampling are used. Here, we will discuss about the formulae for representative sample size when simple random sampling technique has been used. Simple random sampling is the most common and the simplest method of sampling. Each unit of the population has the equal chance of being drawn in the sample. Therefore it is a method of selecting n units out of a population of size N by giving equal probability to all units.

Cochran's Formula for Calculating Sample Size When the Population is Infinite

Cochran (1977) developed a formula to calculate a representative sample for proportions as

$$n_0 = \frac{z^2 pq}{e^2} \qquad (1)$$

where, n_0 is the sample size, z is the selected critical value of desired confidence level, p is the estimated proportion of an attribute that is present in the population, $q = 1 - p$ and e is the desired level of precision [5].

For example, suppose we want to calculate a sample size of a large population whose degree of variability is not known. Assuming the maximum variability which is equal to 50% (p =0.5) and taking 95% confidence level with ±5% precision, the calculation for required sample size will be as follows--

p = 0.5 and hence q =1-0.5 = 0.5; e = 0.05; z =1.96

So, $n_0 = \frac{(1.96)^2 (0.5)(0.5)}{(0.05)^2}$ =384.16=384

Again, taking 99% confidence level with ±5% precision, the calculation for required sample size will be as follows--

p = 0.5 and hence q =1-0.5 = 0.5; e = 0.05; z =2.58

So, $n_0 = \frac{(2.58)^2 (0.5)(0.5)}{(0.05)^2} = 665.64 = 666$

Cochran's Formula for Calculating Sample Size When Population Size is Finite

Cochran pointed out that if the population is finite then the sample size can be reduced slightly. This is due to the fact a very large population provides proportionally more information than that of a smaller population. He proposed a correction formula to calculate the final sample size in this case which is given below

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

where, n_0 is the sample size derived from equation (1) and N is the population size.

Now suppose, we want to calculate the sample size of a population where $N = 13191$, which is the population size of our study. According to the formula (1), the sample size will be 666 at 99% confidence level with margin of error equal to (0.05). If $\frac{n_0}{N}$ is negligible then n_0 is a satisfactory approximation to the sample size. But in this case, the sample size (666) exceeds 5% of the population size (13191). So, we need to use the correction formula to calculate the final sample size.

Here, N = 13191, n_0 = 666 (determined by using (1))

So, $$n = \frac{666}{1+\frac{(666-1)}{13191}} = 634.03 = 634$$

Yamane's Formula for Calculating Sample Size

Yamane (1967) suggested another simplified formula for calculation of sample size from a population which is an alternative to Cochran's formula. According to him, for a 95% confidence level and $p = 0.5$, size of the sample should be

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

Where N is the population size and e is the level of precision [20].

Let this formula be used for the population N =13191 with ±5% precision. Assuming 95% confidence level and p =0.5 we get the sample size as

$$n = \frac{13191}{1+13191(.05)^2} = 388$$

We want to mention here that though other formulae are also available in different literatures, the above two formulae are used extensively in comparison to the others.

After calculating the representative sample size the main aim of an investigator is to find the proper method of selecting sample. Sampling is simply the process of learning about the population on the basis of sample collected from the population. Sample is constituted by a part or fraction of the population. Thus, in the sampling technique instead of every unit of the population only a part of it is studied and the conclusions are drawn on that basis for the entire universe.

COMPARATIVE STUDY OF TWO DIFFERENT METHODS OF ALLOCATION

In our study, for selection of samples, stratified random sampling technique had been adopted. The three categories (i.e. strata) of schools such as Government and Govt. Provincialised schools under SEBA (Secondary Education Board of Assam), Permitted private schools under SEBA, Affiliated private schools under CBSE(Central Board of Secondary Education) of Guwahati were considered as the three strata. The samples from each stratum is taken through simple random sampling technique. The stratification is done to produce a gain in precision in the estimates of characteristics of the whole population.

The stratification was done following the principles that –

- The strata (i.e. categories of schools) are non-overlapping and together comprise the whole population.
- The strata (i.e. categories of schools) are homogeneous within themselves with respect to the characteristics under study

All the VIII standard students of govt., private including SEBA and CBSE schools of Guwahati formed the population of the study. Initially, we estimated the size of sample from a total of 13191 students of class VIII at 95% confidence level with $\pm$ 5% level of precision which was found to be 384. Thus, the sample size of 384 students of 13 selected schools to examine performance of students in mathematics is considered under the present study. This sample can be considered representative of the student population of Guwahati , with students coming from a wide range of socio-economic backgrounds and from each of the three categories of schools such as Co-Educational, only Boys and only Girls schools of Guwahati. The allocation of the sample to the different categories of schools was carried out through both the proportional allocation method and optimum allocation method of stratified random sampling procedure.

The Sample Size through Proportional Allocation Method is given by

The proportional allocation method was originally proposed by Bowley (1926). In this method , the sampling fraction , $\frac{n}{N}$ is same in all strata. This allocation was used to obtain a sample that can estimate size of the sample with greater speed and a higher degree of precision. The allocation of a given sample of size n to different stratum was done in proportion to their sizes. i.e. in the i^{th} stratum,

$$n_i = n\frac{N_i}{N} \qquad \text{where} \qquad i = 1, 2, 3.$$

n – total sample size, N_i – population size of the ith strata and N – total population size.

In our study, N = 13191; n = 384.

The Sample Size through Optimum Allocation Method is given by

The allocation of the sample to the different strata are determined with a view to minimize the variance for a specified cost of conducting the survey or to minimize the cost for a specified value of the variance. The cost function is given by

$$C = a + \sum_{i}^{k} n_i c_i$$

where, a is the observed cost which is constant, c_i is the average cost of surveying one unit in the i^{th} stratum.

Therefore, the required sample size in different stratum is given by

$$n_i = n\frac{\frac{N_i S_i}{\sqrt{c_i}}}{\sum_{i}^{k}\frac{N_i S_i}{\sqrt{c_i}}} \qquad (4)$$

where, n = sample size of the study, N_i = population size of the study, S_i =variance of the i^{th} stratum.

If the average cost of surveying per unit (i.e. c_i) are the same in all the strata, then the optimum allocation becomes the Neyman allocation. As cost of expenditure such as printing of questionnaires, sending and collecting of questionnaires etc. for different categories of schools during the survey by us are almost same, therefore, we can use Neyman allocation in order to determine size of sample for each categories of school. So, in our case, the sample size in different categories of schools is given by a simplified form of (4) which is given by

$$n_i = \frac{nN_iS_i}{\sum N_iS_i} ,$$

where $S_i^2 = \frac{N_i}{N_i - 1} P_iQ_i$ is the population variance of the i^{th} stratum.

N_i = population size of i^{th} stratum,

P_i = proportion of students who secured 50% or more mark in annual examination in i^{th}

stratum $= \frac{number\ of\ students\ in\ i^{th}\ category\ of\ school\ who\ secured\ 50\%\ or\ more\ marks\ in\ mathematics}{total\ number\ of\ students\ in\ i^{th}\ category\ of\ school}$

and $Q_i = 1 - P_i$.

Following table illustrates the distribution of the sizes of samples in different strata for proportional and optimum allocation methods which were calculated on the basis of above discussion.

Table 1: Distribution of Sample Students by Category of Schools

Categories of School		Total Students	
	N_I	N_I (Prop)	N_I (Opt)
Govt.(SEBA)	5609	163	181
Private(SEBA)	3498	102	106
Private(CBSE)	4084	119	97
TOTAL	13191	384	384

Calculation of Variances

The formula to calculate variances of mean for different sampling methods are given as:

For Simple Random Sampling

$$Var(\hat{\mu})_R = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

where, $s^2 = \frac{n}{n-1} pq$

p = proportion of Mark in annual examination who secured 50% and above in all the selected schools, $q = 1 - p$, N = population size, n = sample size.

For Stratified Random Sampling

$$Var(\hat{\mu})_{St} = \frac{1}{N^2}\sum N_i(N_i - n_i)\frac{S^2{}_i}{n_i}$$

where, $S_i^2 = \frac{N_i}{N_i - 1} P_iQ_i$

N =Total population size, N_i =population size of ith stratum, n_i =sample size of i^{th} stratum,

For Proportional Allocation

$$Var\left(\hat{\mu}\right)_{St(prop)} = \sum \frac{N_i}{N}\frac{S_i^2}{n}\left(\frac{N-n}{N}\right),$$

For Optimum Allocation

$$Var\left(\hat{\mu}\right)_{St(opt)} = \frac{\left(\sum w_i S_i\right)^2}{n} - \frac{\sum w_i S_i^{\,2}}{N}$$

where $w_i = \frac{N_i}{N}$

Following table shows the variances of all the schools together that is of the whole sample through different methods.

Table 2: Table Showing Variances

Method	$Var\left(\hat{\mu}\right)_R$	$Var\left(\hat{\mu}\right)_{St(prop)}$	$Var\left(\hat{\mu}\right)_{St(opt)}$
Variances	0.00060839	0.0004673	0.00046

Gain in Efficiency (GE) in Stratified Random Sampling over Simple Random Sampling without Replacement

In order to observe how the sample size gets affected due to different types of allocation, an analysis on gain in efficiency (GE) due to different types of allocations is utmost required.

Gain in Efficiency (GE) Due to Proportional Allocation

$$GE_{prop} = \frac{Var(\hat{\mu})_R - Var(\hat{\mu})_{(St)\,prop}}{Var(\hat{\mu})_{(St)\,prop}} = \frac{0.00060839 - 0.0004673}{0.0004673} = 0.3017333 = 0.30$$

Gain in Efficiency (GE) Due to Optimum Allocation

$$GE_{opt} = \frac{Var(\hat{\mu})_R - Var(\hat{\mu})_{(St)opt}}{Var(\hat{\mu})_{(St)opt}} = \frac{0.00060839 - 0.00046}{0.00046} = 0.3223913 = 0.32$$

From the above results it can be said that optimum allocation provides a little better estimates as compared to proportional allocation. But, the most serious drawback of optimum allocation is the absence of the knowledge of the population variances i.e. S_is of different strata in advance. In that case, the calculations are carried out by performing a pilot survey and by drawing simple random samples without replacement from each stratum as suggested by P.V. Sukhatme (1935). Due to the above mentioned drawback, the allocation of sample size to different strata for our study has been calculated by proportional allocation method. As shown above, by using this method we have gained an efficiency of 0.30 over the simple random sampling.

COMPARATIVE STUDY OF EFFECT OF BIAS IN THE CONTEXT OF DATA OF OUR STUDY

After examining the gain in efficiency (GE) for allocation of sample size to each category of school, students were selected randomly from different schools within that category. In the present study , students were selected from each schools by using Cochran formula at 95%confidence level with ±15% margin of error. Out of these 13 schools 6 are from Govt. SEBA; 3 are from Pvt. SEBA and 4 are from Pvt. CBSE schools. In the case of Pvt. CBSE schools total sample size is 119. But when students of 4 schools are taken into consideration it becomes 131. Hence, to make it 119 from each of the 4 schools three students were not taken into account. Following table illustrates the distribution of the sample by gender and category of schools.

Table 3: The Distribution of Sample Size for Class VIII Students of Different Schools of Guwahati

Category of Schools	Sl. No.	Name of School	Population Size	Sample Size (max)	Allotted Sample Size		
					Boys	Girls	Total
SEBA (Govt.)	1	Ulubari H.S.	95	30	16	14	30
	2	Dispur Vidyalaya	88	29	16	13	29
	3	Ganesh Mandir Vidyalaya	112	31	17	14	31
	4	Noonmati M.E. School	79	28	12	16	28
	5	Uzan Bazaar Girls' School	43	22	_	22	22
	6	Arya Vidyapeeth High School	46	23	23	_	23
SEBA (Pvt.)	7	Nichol's School	125	32	22	10	32
	8	Asom Jatiya Vidyalaya	200	36	26	10	36
	9	Holy Child School	170	34	_	34	34
CBSE(Pvt.)	10	Gurukul Grammar School	154	34	14	17	31
	11	Maharishi Vidya Mandir School	160	34	17	14	31
	12	Sarala Birla Gyan Jyoti	115	31	13	15	28
	13	Shankar Academy	118	32	17	12	29
Total					**193**	**191**	**384**

It is well known that during the collection of sample units, both sampling and non-sampling errors creep into the process. The non sampling errors occur because the procedures of observation (data collection) are not perfect and their contribution to the total error of survey may be substantially large, which may affect survey results adversely. On the other hand, the sampling errors arise because a part (sample) from the whole(population) is taken for observation in the survey. Since in our study sample size is 384, which is quite large, hence, by virtue of the Central Limit Theorem (CLT) we can use normal probability table to calculate the effect of bias for the questionnaires used in order to collect the data of marks in mathematics.

The total error is expressed as:

$$Total\ Error\ (TE) = \sqrt{Mean Square Error\,(MSE)} = \sqrt{Variance of\ mean + Square of\ Bias}$$

Again, Bias is the difference between the estimated value of population mean and sample mean.

Even with estimators that are un-biased in probability sampling, errors of measurement and non response may produce biases in the numbers that we are able to compute from the data.

To examine the effect of Bias, let us suppose that the estimate $\hat{\mu}$ is normally distributed about a mean m that is at distance B from the true population value μ. Therefore, the amount of bias is $B = m - \mu$. As a statement about the accuracy of the estimate, we declare that the probability is 0.05 that the estimate $\hat{\mu}$ is in error by more than 1.96σ.

This can be calculated with the help of following transformation

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu+1.96\sigma}^{\infty} e^{\frac{-(\mu-m)^2}{2\sigma^2}}\, d\mu = \phi(\mu+1.96\sigma)$$

Now putting $\mu - m = \sigma t$ in above integral, we get lower limit of the range of integration for 't', will be

$$\frac{\mu-m}{\sigma}+1.96 = 1.96-\frac{B}{\sigma},$$

where $B = m - \mu$ is the amount of bias occur for adjusting the sample size for each strata.

Thus we require to calculate bias by consulting the normal probability table with the help of following :

$$\frac{1}{\sqrt{2\pi}}\int_{1.96-\frac{B}{\sigma}}^{\infty} e^{\frac{-t^2}{2}}\, dt = \phi\left(1.96-\frac{B}{\sigma}\right)$$

In the table the effect of a bias B on the probability of an error greater than 1.96σ has been shown in tabular form. The calculations were carried out using the normal probability table.

Table 4: Effect of a Bias B on the Probability of an Error Greater than 1.96σ

B/σ	Probability of Error		Total
	<-1.96σ	>1.96σ	
0.01	0.0244	0.0256	0.0500
0.03	0.0233	0.0268	0.0501
0.05	0.0222	0.0281	0.0503
0.07	0.0212	0.0294	0.0506
0.09	0.0202	0.0307	0.0509
0.10	0.0197	0.0314	0.0511
0.25	0.0136	0.0436	0.0572
0.40	0.0091	0.0594	0.0685
0.55	0.0060	0.0793	0.0853
0.70	0.0039	0.1038	0.1077
0.85	0.0025	0.1335	0.1360
1.00	0.0015	0.1685	0.1700
1.50	0.0003	0.3228	0.3231

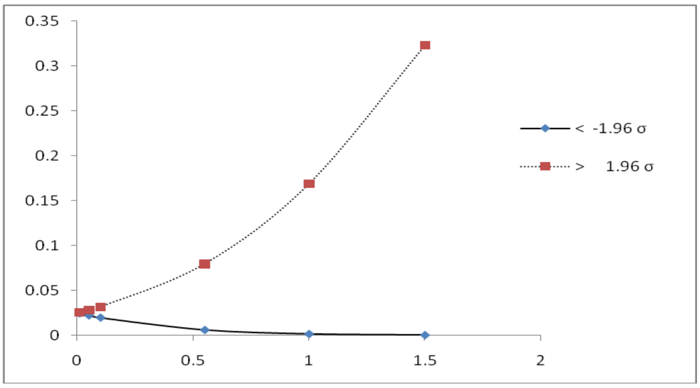

Figure 1: B/σ Values (x-axis) vs Praobability of Error (Less than -1.96 σ) and (Greater than 1.96 σ) (Generated from the Above Table)

It is known that, in order to compare a biased estimator with an unbiased estimator, or two estimators with different amounts of bias, a useful criterion is the mean square error(MSE)of the estimate, measured from the population value that is being estimated.

The relationship between MSE and Bias is given by

$$MSE(\hat{\mu}) = Variance\ of\ (\hat{\mu}) + (Bias)^2$$

In the following tables variances for different schools of Guwahati included in the sample and different categories of schools are given. In our total sample size and sample size in different strata has been calculated with margin of error ±0.05.

But, while calculating the sample sizes in the 13 selected schools the margin of error was taken to be ±0.15; because greater precision requires large sample size which is not practicable in case of selection of sample from different schools. For this difference in precision some bias may occur in the process and hence it becomes very important to calculate the bias and its effect.

Table 5: For the Schools

Sl. No. of Schools	Population Size N	Sample Size n	No. of Students Securing 50 or More Marks in Mathematics	P	Q	Variances
1	95	30	2	.07	.93	.00153593
2	88	29	11	.38	.62	.0056414
3	112	31	9	.29	.71	.00496366
4	79	28	17	.61	.39	.00568819
5	43	22	8	.36	.64	.00535814
6	46	23	8	.35	.65	.00517045
7	125	32	25	.78	.22	.0041184
8	200	36	31	.86	.14	.0028208
9	170	34	18	.53	.47	.00603879
10	154	31	28	.90	.10	.0023961
11	160	31	30	.97	.03	.000782063
12	115	28	21	.75	.25	.00525362
13	118	29	23	.79	.21	.00446886

Table 6: For Different Categories of Schools

Strata	Sample Size n	No. Students Securing 50 or More	P	Q	Variances
SEBA Govt.	163	55	.34	.66	.00134493
SEBA Pvt.	102	74	.73	.27	.00189458
CBSE Pvt.	119	102	.86	.14	.000990608

In the following tables probability of an absolute error ≥1√MSE and 1.96√MSE for different categories of schools are given. Below each table, graphs of MSE, 1√MSE and 1.96√MSE versus B/σ values (in x axis) are shown

Probability of an absolute error ≥ 1√MSE and 1.96√MSE

Table 7: For SEBA Govt.: V=0.00134493, p=0.34, q=0.66

$\frac{B}{\sigma}$	MSE	$1\sqrt{MSE}$	$1.96\sqrt{MSE}$
0.01	0.00384493	0.0620075	0.121535
0.03	0.00384493	0.0620075	0.121535
0.05	0.00384493	0.0620075	0.121535
0.07	0.00394493	0.0628087	0.123105
0.09	0.00394493	0.0628087	0.123105
0.10	0.00394493	0.0628087	0.123105
0.25	0.00444493	0.0666703	0.130674
0.40	0.00604493	0.0777492	0.152388
0.55	0.00864493	0.0929781	0.182237
0.70	0.0129449	0.113776	0.223001
0.85	0.0198449	0.140872	0.276109
1.00	0.0302449	0.173911	0.340865
1.50	0.105745	0.325184	0.637362

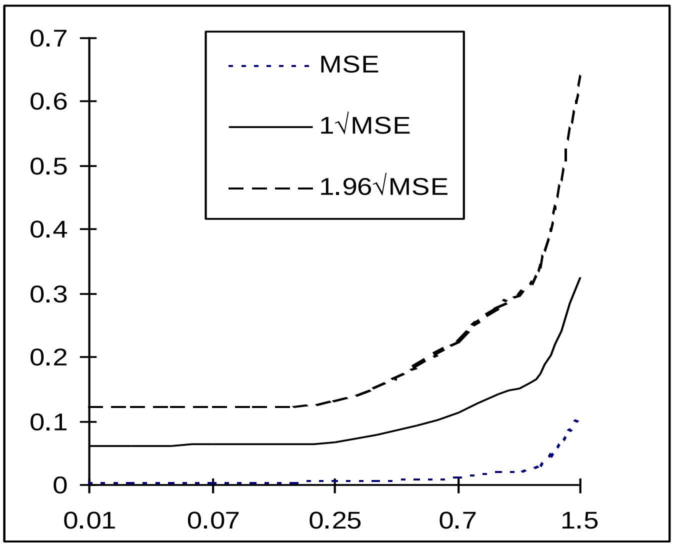

Figure 2: For SEBA Govt.: $\frac{B}{\sigma}$ **Values (x-axis) vs** MSE **,** $1\sqrt{MSE}$ **and** $1.96\sqrt{MSE}$

Table 8: For SEBA Pvt. : V=0.00189458, p=0.73, q=0.27

$\frac{B}{\sigma}$	MSE	$1\sqrt{MSE}$	$1.96\sqrt{MSE}$
0.01	0.00439458	0.0662916	0.129932
0.03	0.00439458	0.0662916	0.129932
0.05	0.00439458	0.0662916	0.129932
0.07	0.00449458	0.0670416	0.131402
0.09	0.00449458	0.0670416	0.131402
0.10	0.00449458	0.0670416	0.131402
0.25	0.00499458	0.0706723	0.138518
0.40	0.00659458	0.081207	0.159166
0.55	0.00919458	0.0958884	0.187941
0.70	0.0134946	0.116166	0.227686
0.85	0.0203946	0.14281	0.279907
1.00	0.0307946	0.175484	0.343948
1.50	0.106295	0.326028	0.639016

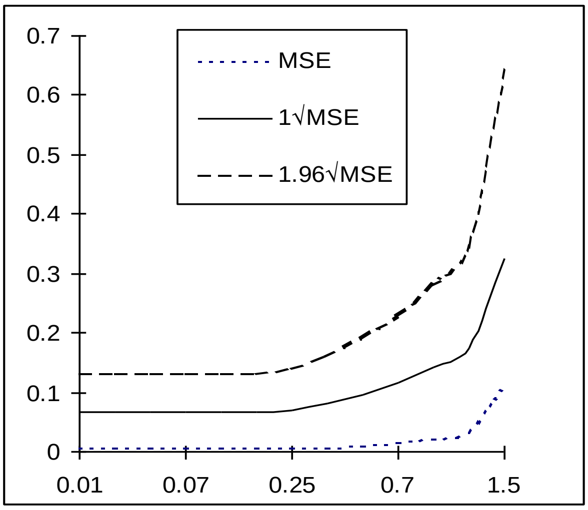

Figure 3: For SEBA Pvt.: $\frac{B}{\sigma}$ Values (x-axis) vs MSE, $1\sqrt{MSE}$ and $1.96\sqrt{MSE}$

Table 9: For CBSE Pvt.: V=0.000990608, p=0.86, q=0.14

$\frac{B}{\sigma}$	MSE	$1\sqrt{MSE}$	$1.96\sqrt{MSE}$
0.01	0.00349061	0.0590814	0.115799
0.03	0.00349061	0.0590814	0.115799
0.05	0.00349061	0.0590814	0.115799
0.07	0.00359061	0.0599217	0.117447
0.09	0.00359061	0.0599217	0.117447
0.10	0.00359061	0.0599217	0.117447
0.25	0.00409061	0.0639579	0.125357
0.40	0.00569061	0.0754361	0.147855
0.55	0.00829061	0.0910528	0.178463
0.70	0.0125906	0.112208	0.219927
0.85	0.0194906	0.139609	0.273633
1.00	0.0298906	0.172889	0.338862
1.50	0.105391	0.324639	0.636293

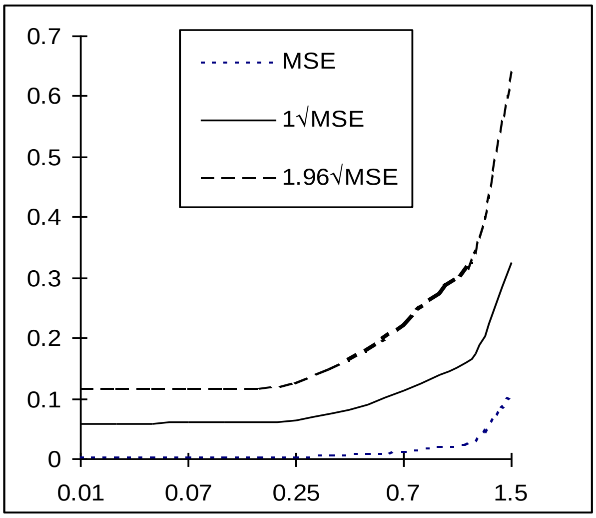

Figure 4: For CBSE Pvt.: $\frac{B}{\sigma}$ Values (x-axis) vs MSE, $1\sqrt{MSE}$ and $1.96\sqrt{MSE}$

Table 10: For All the Schools: V=0.0004673, p=0.60, q=0.40

$\frac{B}{\sigma}$	MSE	$1\sqrt{MSE}$	$1.96\sqrt{MSE}$
0.01	0.0029673	0.0544729	0.1067668
0.03	0.0029673	0.0544729	0.1067668
0.05	0.0029673	0.0544729	0.1067668
0.07	0.0030673	0.0553832	0.108551
0.09	0.0030673	0.0553832	0.108551
0.10	0.0030673	0.0553832	0.108551
0.25	0.0035673	0.0597268	0.1170645
0.40	0.0051673	0.0718839	0.1408924
0.55	0.0077673	0.0881322	0.1727391
0.70	0.0120673	0.1098512	0.2153083
0.85	0.0189673	0.1377218	0.2699347
1.00	0.0293673	0.1713689	0.335883
1.50	0.1048673	0.3238322	0.6347111

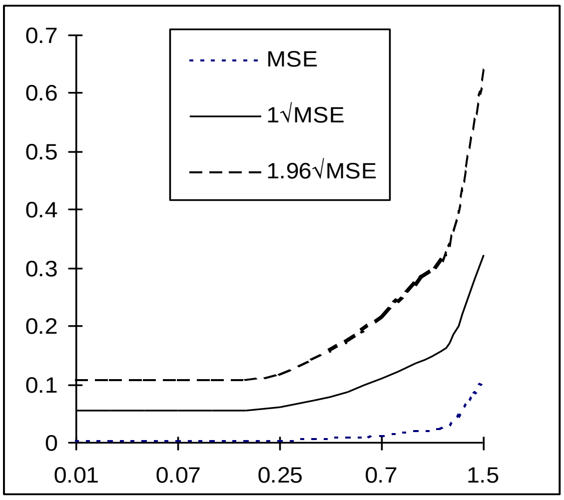

Figure 5: For All the Schools: $\frac{B}{\sigma}$ Values (x-axis) vs MSE, $1\sqrt{MSE}$ and $1.96\sqrt{MSE}$

The following figure shows the comparison between the MSE of different categories mentioned above.

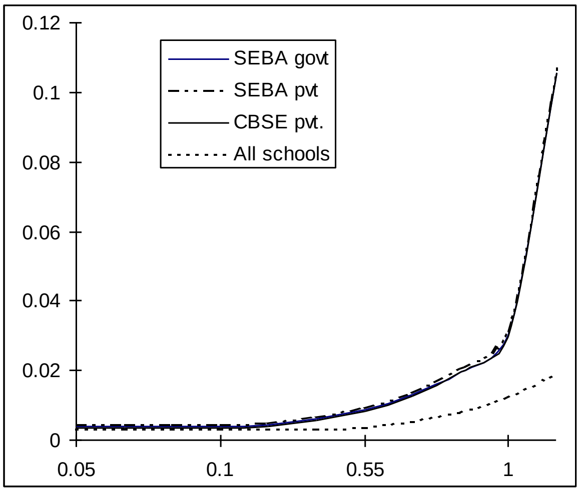

Figure 6: For All the Schools: $\frac{B}{\sigma}$ Values (x-axis) vs MSE of All Categories of Schools

The following table shows the Probability of an absolute error ≥ 1√MSE and 1.96√MSE for different schools:

Table 11: Probability of an Absolute Error ≥ 1√MSE and 1.96√MSE for Different Schools

Sch	p,q,V		$\frac{B}{\sigma}$					
			0.01	0.10	0.25	0.55	1.00	1.50
1	p =0.07	MSE	0.00403593	0.00413593	0.00463593	0.00883593	0.0304359	0.105936
	q=0.93	1√MSE	0.063529	0.0643112	0.0680877	0.0939997	0.174459	0.325478
	V=0.00153593	1.96√MSE	0.124517	0.12605	0.133452	0.184239	0.34194	0.637937
2	p =0.38	MSE	0.0081414	0.0082414	0.0087414	0.0129414	0.0345414	0.110041
	q=0.62	1√MSE	0.0902297	0.0907821	0.0934954	0.11376	0.185853	0.331725
	V=0.0056414	1.96√MSE	0.17685	0.177933	0.183251	0.22297	0.364272	0.650181
3	p =0.29	MSE	0.00746366	0.00756366	0.00806366	0.0122637	0.338637	0.109364
	q=0.71	1√MSE	0.0863925	0.0869693	0.0897979	0.110741	0.184021	0.330702
	V=0.00496366	1.96√MSE	0.169329	0.17046	0.176004	0.217053	0.360681	0.648175
4	p =0.61	MSE	0.00818819	0.00828819	0.00878819	0.0129882	0.0345882	0.110088
	q=0.39	1√MSE	0.0904886	0.0910395	0.0937453	0.113966	0.185979	0.331795
	V=0.00568819	1.96√MSE	0.177358	0.178437	0.183741	0.223373	0.364519	0.650319
5	p =0.36	MSE	0.00785814	0.00795814	0.00845814	0.0126581	0.0342581	0.109758
	q=0.64	1√MSE	0.0886461	0.0892084	0.919681	0.112508	0.18509	0.331298
	V=0.00535814	1.96√MSE	0.173746	0.174848	0.180258	0.220516	0.362776	0.649343
6	p =0.35	MSE	0.00767045	0.00777045	0.00827045	0.0124705	0.0340705	0.10957
	q=0.65	1√MSE	0.0875811	0.0881502	0.090942	0.111671	0.184582	0.331014
	V=0.00517045	1.96√MSE	0.171659	0.172774	0.178246	0.218876	0.36178	0.648788
7	p =0.78	MSE	0.0066184	0.0067184	0.0072184	0.0114184	0.330184	0.108518
	q=0.22	1√MSE	0.0813535	0.0819658	0.0849612	0.106857	0.18171	0.329421
	V=0.0041184	1.96√MSE	0.159435	0.160653	0.166524	0.20944	0.356151	0.645666
8	p =0.86	MSE	0.0053208	0.0054208	0.0059208	0.0101208	0.0317208	0.107221
	q=0.14	1√MSE	0.0729438	0.0736261	0.0769467	0.100602	0.178103	0.327446
	V=0.0028208	1.96√MSE	0.14297	0.144307	0.150816	0.19718	0.349083	0.641794
9	p =0.53	MSE	0.00853879	0.000863879	0.00913879	0.0133388	0.0349388	0.110439
	q=0.47	1√MSE	0.0924056	0.0929451	0.095597	0.115494	0.186919	0.332323
	V=0.00603879	1.96√MSE	0.181115	0.182172	0.18737	0.226368	0.366362	0.651354
10	p =0.90	MSE	0.0048961	0.0049961	0.0054961	0.0096961	0.0312961	0.106796
	q=0.10	1√MSE	0.0699722	0.0706831	0.0741357	0.0984688	0.176907	0.326797
	V=0.0023961	1.96√MSE	0.137145	0.138539	0.145306	0.192999	0.346738	0.640522
11	p =0.97	MSE	0.00328206	0.00338206	0.00388206	0.00808206	0.0296821	0.105182
	q=0.03	1√MSE	0.0572893	0.0581555	0.0623062	0.0899003	0.172285	0.324318
	V=0.000782063	1.96√MSE	0.112287	0.113985	0.12212	0.176205	0.337678	0.635663
12	p =0.75	MSE	0.00775362	0.00785362	0.00835362	0.0125536	0.0341536	0.109654
	q=0.25	1√MSE	0.0880547	0.0886207	0.0913982	0.112043	0.184807	0.33114
	V=0.00525362	1.96√MSE	0.172587	0.173697	0.17914	0.219604	0.362222	0.649034
13	p =0.79	MSE	0.00696886	0.00706886	0.00756886	0.0117689	0.0333689	0.108869
	q=0.21	1√MSE	0.0834797	0.0840765	0.0869992	0.108484	0.182671	0.329953
	V=0.00446886	1.96√MSE	0.16362	0.16479	0.170518	0.212629	0.358036	0.646708

CONCLUSIONS

Use of the MSE as a criterion of the accuracy of an estimator amounts to regarding two estimates that have the same MSE are equivalent. It has been shown by Hansen, Hurwitz and Madow [10] that if for $\frac{B}{\sigma}$, MSE is less than one half, then the estimator can be considered almost identical with its true value. The tables 7, 8, 9, 10 and their corresponding graphs in figure 2, 3, 4 and 5 highlights this criterion in case of our study.

So, we can conclude that the effect of bias in our study is negligible and the estimations derived from the selected samples will be in good agreement with their corresponding values for the whole population.

REFERENCES

1. Bantelett, J.E.II; Kotrlik, J. W; Higgins, C. C. (2001). Organizational Research: Determining Appropriate Sample Size In Survey Research, Information Technology, Learning and Performance Journal, Vol.19, No.1, pp (43-50).
2. Best, John W. and Kahn, James V.(1996), *Research in Education*,7th edition. New Delhi: Prentice Hall of India Pvt.
3. Brush, L.R.(1985). Mathematics Anxiety, Mathematics Achievement, Gender and Socio-economic Status among Arab Secondary students in Israel. 3 *Math.Edu.SCI.Technol*.21(2),319-327.
4. Buch,M.B.(1983-88): 4th *Survey of Research in Education*, vol 1, NCERT.
5. Cochran, W.G.(1977). Sampling Techniques (3rd edition). John Wiley &Sons, New York.
6. Pandey,.R.S.(2002-03).Development of Indian System of Education. Vinod Pustak Mandir, Agra.
7. Garrett, H. E. (1981). Statistics in Psychology and Education.Vakils, Feffer and Simons LTD.
8. Government of India (1986). National Policy on Education(1986). Programme of action, New Delhi: Department of Education,MHRD.
9. Guadagnoli, E., & Velicer, W. F. (1988). Relation of Sample-Size to the Stability of Component
10. Hansen, M. H; Hurwitz, W.N and Madow, W. G : Sample Survey Methods and Theory. John Wiley and Sons, New York, Vol. I and II.
11. Israel, Glenn, D. (1992). Determining Sample Size. Program Evaluation and Organizational Development, Florida Co-operative Extension Service. IFAS, University of Florida. PEOD-6. November.
12. Kasunic, Mark (2005). Small sample techniques- Designing an Effective Survey. Hand book. CMU/SEI-2005-HB-004. September.
13. Krejcie, V.R; Morgan, D. W. (1970). Educational and Psychological Measurement, 30. Pp (607-610).
14. Medhi, J (2006). Statistical Methods, An introductory text. New Age International Publishers.
15. Meknigh, C; Magid, A ; Murphy, T. J and Meknigh, M : Mathematics Education Research : A Guide for Research Mathematician, American Mathematical Society, Providence, Rhode Island.
16. National Center For Education Statistics, U.S Department of Education, 'Using TIMSS to analyse Correlates of Performance Variations in Mathematics', Working Paper Series, Working Paper No. 2001-2005, June 2001, (Edited by M. M. Millen and R.Lee) 1990 K. St. NW. Washington DC 2006
17. NCERT(2005). National curriculum Framework, New Delhi.
18. NCERT. Position Paper 1.2, National Focus Group on Teaching of Mathematics.
19. Singh, D.; Choudhary, F. S. (1986). Theory and Analysis of Sample Survey Designs. Wiley Eastern Limited.
20. 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 .
21. Yamane, Taro, (1967). Statistics, An Introductory Analysis, 2nd ed., New York: Harper and Row.