

On Statistical Estimation: Some Views

Srilakshminarayana, G.

Assistant Professor - QM

SDMIMD, Mysore

lakshminarayana@sdmimd.ac.in

Abstract

Decision - making in a dynamic market environment emphasizes on studying the behaviour of key market parameters from time to time and keep a track on how the values of these parameters change with changing environment. In order to study these patterns, data has to be collected from the market at regular time intervals and processed using appropriate scientific tools. Theory of statistical estimation is a collection of such tools that is being used to draw reliable, accurate and precise conclusions regarding the market parameters. They help the researchers in identifying appropriate point estimates and then move on to constructing an interval estimate with an associated probability. These interval estimates form the set of values that the parameter under study is expected to assume and the associated probability gives the strength. Before estimating the parameters using the statistical procedures, the researchers have to consider certain aspects like definition of the population, type of survey, the associated sampling design, estimator, estimation space etc. and these form the center of our discussion in this paper.

Keywords: Population, Parameter, Estimator, Parametric estimation, Central limit theorem.

Introduction

Statistical Estimation is the process of studying and understanding the behaviour of the unknown characteristics of the population using statistical tools. These statistical tools not only help the researchers to provide a single value as an estimate from the data but also a range of values that form a set of possible estimates with an associated probability. It is a usual practice to start with a point estimate and end the estimation process with an interval estimate. These interval estimates are associated with the probabilities help and the researcher to determine the chance with which unknown parameter assumes values in the interval. The parameters are the unknown characteristics of the market under study. The markets under study are volatile and have to be monitored continuously. The data collected from the market helps the researchers to monitor the same and gives better estimates to the parameters under study. Not only the researchers, but also the organizations are involved in the process of estimation.

It is necessary for the organizations to have an understanding of the market they are into. The markets comprises of customers, competitors and other stake holders who are directly or indirectly associated with the performance of the organization. In order to understand how the customers behave it is necessary to continuously monitor the market research by adopting appropriate tools that give the accurate picture of how markets react to the policies of the organizations. It is well known that the

organization enter into the market via their products. Knowing customer talk about these products will help the organizations to improve the performance of the products in satisfying the needs of the customers. The behavior of the customers vary with the product. They look for a better product, each time they enter into the market that satisfies their needs. Understanding their behavior needs identifying important characteristics and estimating the same precisely using scientific tools so that appropriate decisions are taken on time. It is very important to select an estimator to study the behavior of these population characteristics. Also, care has to be taken while selecting the group of customers who respond to the queries posed by the researchers.

The selected estimator acts as a generator that generates the estimates and helps the researcher to study the patterns in the market. These patterns makes one to understand how the parameter reacts to the market and with the associated changes in the market. For example, if the parameter under study is the population average, the understanding is that the market is centered at a value and other observations that one makes are close to this value. Also, this center value acts as a benchmark when the organizations wishes to compare its performance with that of others. In other words the center value is the market standard that all the organizations accept to and acts close to it. This estimate changes as the sample changes and provides the researcher an opportunity to study how the

market center is changing as the market changes. This is called as location estimation. The located value decides how the market is moving and how it is centered at value, which decides the market standards. This centered value forms the base to measure how other observations are moving from the market standards. Usual practice that exists in the market is to start with assumption that the market is behaving as per normal standards. When the same is carried forward to the theoretical concepts it shapes to the assumption that the uncertainty that exists is being measured using normal probability law and the associated measures are used to draw inferences regarding the parameters of the market. This assumption of normality helps the researcher to understand the market behavior and any deviation from normality alerts him to take appropriate action in time to handle the situations that gives the results far from normal patterns.

Estimation of location is one such attempt to understand how the market is centered and ultimately set a benchmark that forms the normal standard. In general this center value is unknown and have to be estimated using the responses that one gathers from the market. Using these responses, an appropriate inference has to be drawn regarding how the individuals are moving, to what direction their thinking is moving and what is the center direction that all come to. In other words, it is to say that all merge into a common conclusion about the products that is being introduced in to the market. One has to understand that the group of customers who are in need and uses the product to satisfy the need decides what changes the organizations have to

make with respect to the quality standards of the products. Here, quality standards have to be understood that it is an aggregation of all those that makes the customer prefer the product that satisfies the needs. Note that the estimated center value helps what the group is expecting on average regarding the quality standards of the products. It is also important to understand that the center value changes as customers changes their perception about the product. Hence, it is necessary to have an understanding of how the center changes with the changes in the customer's perception. This can be done using interval estimation that gives the researcher a range of values that form the set of values that the parameter is expected to assume with certain confidence level. The intervals constructed are called as confidence intervals. They have two components. The first is the range of values and the other is the confidence level.

In this paper we present our views on some of the fundamental aspects that play an important role in the estimation process, that are being ignored when the researcher consider statistical estimation. An attempt is made to present a sequence of steps that a researcher can consider in his market research study. In section 2 we present some views on fundamental aspects that one need to consider in the process of estimation along with an empirical support.

The data considered in this paper is with respect to a famous hotel, whose name is not disclosed due to privacy constraints

Fundamental aspects in Statistical estimation

Population and a Sample

A population in estimation is seen as a source from where the responses are collected to estimate the unknown parameters. It can be defined as the group of individuals who possess the characteristics that meets the objectives of the study, well defined. In order to decide upon the population, one has to define clearly the characteristics of the individuals and accordingly define the population. For example, if one wishes to study the life style of working class women who are in the age group 20-35, average income level above 4.5 lakhs, then individuals with above mentioned requirements will constitute the population for the study.

Once a population is defined, the next step is to collect the responses and then estimate the parameters under study. If one questions whether the population can be varied based on the requirements, the answer is yes but it should be a rare case. Based on the objectives of the study population definitions can be changed but should not be done once the study has been designed. It is also possible that sometimes the objectives of the study may change and there by the characteristics of the individuals change. In such cases it is necessary to change the definition of the population. Non-availability of the respondents with the specified characteristics also makes the researcher to change the definition of the population. Though the definition of the population can be changed, it should not be changed unless there is a strong reason to change. To

avoid this it is always better to conduct a thorough pilot study of the population from where the sample is planned to collect. It helps to take decisions on necessary changes with respect to the target population. Clear definition of the population will help the researcher to focus more on the sampling.

Population can be finite, infinite, real, hypothetical. A finite population is the one which has finite number of units that can be enumerated and infinite population has units that are count ably infinitely many or un count ably infinitely many. For example, the number of students in a college considered for the study in finite. Whereas individuals who are interested to travel by city bus to their destination forms an infinite population. Note that before the actual survey starts, one has to clearly define whether the population is finite or infinite. A real population is the one which can be seen physically and measured properly for further mathematical treatment. A real population can be finite or infinite. A hypothetical population is the one which is assumed for the purpose of experimentation.

In a business context, population would be the market under study and the respondents are the customers that use the products, introduced by the organizations into the market. In order to find the satisfaction level, we approach the customers and obtain the feedback to estimate the average satisfaction level. In the process of estimation, defining a population clearly and completely will help the researcher to estimate the parameters with a

better confidence level. When a researcher clearly knows that group of respondents from whom a sample can be drawn, then he can choose a better confidence levels while estimating the parameters. Whereas, if the researcher is not clear on group of individuals to be considered for the study then there exists a level of ambiguity while choosing a sampling design and it effects the estimation process. This is very important if the researcher is conducting a market research. He should define the customer panels clearly before moving on to the next step of sampling design. He should get consent from the panel members before including them into the survey. This avoids non-response bias which will have an effect on the estimation process.

The portion of the population drawn using a sampling procedure for enumeration is called as a sample. It is assumed to possess all the characteristics of the population under study. It is said that the sample should be selected in such a way that it represents the entire population, a true representative of the population. Care has to be taken while including the respondents into the sample. Proper sampling design has to be adopted to draw the sample and each respondent has to be enumerated properly. In any study, the responses collected from the respondents will be analysed using appropriate statistical tools. These tools demand that the sample be a random sample. This is an important assumption as it will ensure that the researcher is not biased. Most of the studies depend on the results of the sample and try to conclude for the population. In such cases it is very important that the concept of probability has

to be used while drawing the sample. This gave rise to probability sampling. These techniques take the chance of selection that an individual has while selecting the sample and selects that individual. This process not only avoid bias but also assures that the error of choice is minimized. Some of the techniques assign equal probability whereas some other assigns unequal probabilities depending upon the importance the individual carries with him. It may also depend on the importance an individual carries in terms of the response. This can be commonly found in marketing research.

Apart from standard and advanced probability sampling techniques there exists non-probability sampling. Convenient, purposive, judgemental, snowball sampling techniques come under this category. When a sample is drawn using these techniques it is very important and necessary to check for randomness of the sample. In order to do this, the usual practice is to use a run test. This matters more when one moves onto hypothesis testing and with respect to estimation, it matters only when one moves onto interval estimation, where the sampling distribution comes into the picture. In cases where a parametric approach cannot be used a non-parametric approach or a robust approach can be used.

The population for the case considered is the groups of subsidiary hotels that are located in different parts of India. The size of the population is 2020, which is the case of a finite population. The division of the same is presented in the following table:

Table-1: Population details

| S.No. | Subsidiary | Places |
|--------------|-------------------|---------------|
| 1 | Subsidiary-1 | 133 |
| 2 | Subsidiary-2 | 169 |
| 3 | Subsidiary-3 | 1194 |
| 4 | Subsidiary-4 | 396 |
| 5 | Subsidiary-5 | 128 |

There are 5 subsidiary hotels and each is located in different places. For example subsidiary-1 is located at 133 different places and a similar explanation holds for other subsidiaries also. Note that each subsidiary is homogeneous within and heterogeneous between. Feedback forms were sent to the customers at each hotel during four different time intervals and similarly feedback forms received during the same time intervals were noted. The total feedback forms sent and total feedback forms received were considered as variables for the study.

The objective here is to estimate the average total feedback forms sent and the total average feedback forms received using point and interval estimation. The details of the forms sent and received has to be collected from managers of hotels.

Census Survey and a Sample Survey

A census survey is preferred when there is a demand for complete enumeration. Example would be the information

collected by government to decide upon providing basic requirements for the public. It is important to take responses from all the individuals if the study is related to providing facilities to the individuals in an educational institution, in a constituency etc.

Note that the size of the population should be finite or sufficient resources should be available to enumerate entire population. It is not advised to enumerate the entire population if the study can be concluded using a sample. Also by taking a sample one can enjoy the advantages of it like quick decision making, ease of drawing conclusions about the characteristics of the population and extracting more information from the respondents. If the population size is small, then it is not difficult to spend time in collecting the information from the respondents. But, if the population size is considerably large then it is better to take a sample as one can spend sufficient time with each respondent selected rather than spending little time with respondent and collecting incomplete data. This is because, a survey either census or sample should give the researcher sufficient, complete, reliable and accurate information.

In statistical estimation, a census survey gives good results provided only if the population is homogeneous. If a population is heterogeneous, it is not advised to consider a census survey and a stratified random sampling gives better estimates.

When the resources are limited or when one looks at optimizing the available resources, it is advised to study the population by drawing a sample from it. The sample has to be drawn in such a way that it gives optimum information about the population under study. A sample survey is a well-planned and designed way of obtaining information from the respondents. It includes only those who are selected and from whom the responses have to be collected. A focussed sample survey will result in responses free of bias and will make the survey end with minimum errors. This is because any survey will have errors and a sample survey should result in minimum errors.

A well-designed sample survey gives better estimates for the parameters and also provides an opportunity to draw appropriate inferences on the parameters. This is because a well-designed sample survey selects respondents appropriately and also ensures that responses are obtained in a manner that will give better understanding of the market parameters. A sample survey that is free from bias error will give better estimates of the parameters.

A census survey here would be approaching all the hotels and obtain the required information. Note that doing the same needs sufficient time and other resources. Also it is not easy to collect the details from each and every manager. Hence, for the given objective it is planned to use a sample survey as it is easy to collect details from few homogeneous hotels under each subsidiary.

Pilot Survey

Before designing the final survey, it is very important to conduct a pilot survey in order to study and understand the dynamics of the population under study. As the name indicates this survey should be planned and executed immediately after the objectives have been identified. It help in testing the questionnaire designed for the study and also helps to know important aspects like cost for collecting responses, distance to be travelled, time to be spent etc. Apart from these, it also helps to understand the population characteristics which are quantitative in nature like mean, standard deviation etc. Most of the studies ignore the fact that a good pilot survey will give better estimates for the actual survey. For example, in order to estimate the sample size for the final study, one can get the estimates of the inputs in the formula like desired level of precision, estimate of the variance etc. Sometimes it is also used to identify missed or latent objectives, if any, and include them in the study. A good pilot survey should be planned in such a way that it provides all inputs necessary for the actual survey. In some instances, it also helps to design a sampling frame as well as a sampling design.

A pilot survey was conducted for the case to estimate necessary inputs for determining the sample size. The following table gives the details of the pilot survey:

Table-2: Pilot sample details

| S.No. | Subsidiary | Pilot Sample |
|--------------|-------------------|---------------------|
| 1 | Subsidiary-1 | 30 |
| 2 | Subsidiary-2 | 53 |
| 3 | Subsidiary-3 | 216 |
| 4 | Subsidiary-4 | 140 |
| 5 | Subsidiary-5 | 21 |

The variables under study are total feedback forms sent

Table-3: Descriptive statistics

| Subsidiary | Mean | | Standard deviation | |
|-------------------|-------------------|-----------------------|---------------------------|-----------------------|
| | Forms sent | Forms received | Forms sent | Forms received |
| 1 | 483 | 69 | 59 | 8 |
| 2 | 495 | 81 | 51 | 12 |
| 3 | 387 | 70 | 73 | 17 |
| 4 | 390 | 59 | 85 | 17 |
| 5 | 322 | 62 | 68 | 17 |

Sampling Frame

In order to draw a sample, the researcher has to design a sampling frame that has details of the individuals who form the population for the study. It should include at least the names of the individuals or any other details that are common between the units. Sampling frame is necessary to apply a sampling technique and ensures that only those individuals who will be part of the survey are included. In practice, it may be difficult to design a sample frame with accuracy. But one can try to design the sampling frame that contains at least few details of the population units. This is more mandatory if one decides to use a random sampling technique. Whereas if one decides to use a convenient sampling technique, judgement sampling etc., then sampling frame need not be mandatory.

The sampling frame for the case is the information on the place code the hotel belongs to and name of the subsidiary. Based on this the sample has been drawn using stratified random sampling.

Sample Size and Sampling Error

Determining the sample size in association with the objectives and the cost is an important step in the process of drawing the sample. It is also an estimation that helps the researcher know regarding the number of respondents required to estimate the average perception about the products that are introduced into the market. It gives the researcher an opportunity to plan his resources and check whether any adjustments are necessary in the budget

allocation. Also he can associate the desired precision in the estimation with the sample size. In other words he can decide with what precision the parameters should be estimated and to achieve the desired precision how many respondents have to be involved in the survey. Most of the times the researcher will be happy if he can meet reasonable number of respondents and interview them as per the objectives of the study. Through the sample size estimation methods, he gets an idea about the number of respondents required for the survey. It is to be noted here that as the sample size increases, the researcher will have the advantage of comprehending the market behavior with a better precision. This also decreases the sampling error which is inherent in the sampling surveys. Estimation of sample size depends on the sampling design chosen for the survey. Sometimes, the researcher can consider the power at which he wishes to estimate the parameter in estimating the sample size. Hence, estimation of the sample size has to be seen as an important step in the sample surveys. A good and exhaustive discussion on sample size determination can be found in Mace (1964).

Another important aspect that has to be taken in to consideration is sampling error. The sampling error is measured and based on the same appropriate actions are to be taken to minimize it. One way to minimize the error is to increase the sample size. Increasing the sample size implies that researcher is able to meet more number of respondents and gather appropriate information from them. This makes the researcher to have better comprehension about the market he is working with and

also have a better idea on the key parameters of interest. Spending more with the market will enable the researcher to decrease the error either at the mental level i.e., his perception or technical level, that considers the mathematical formulae for calculating the error. In either cases, the researcher ensures that the error level drops to an admissible level such that the precision of the survey increases. The admissible level has to be decided by the researcher before the survey begins. This can be measured using the level of precision decided by the researcher. The less the sampling error, the more precise the estimates obtained from the sample. The discussion of this end once the researcher decides on what has to be done to decrease the error that is inherent in the sample.

Sample size required for the sample survey was calculated using standard formulae and found to be 750 with a confidence level of 95%.

Sampling and Sampling design

To collect the group of individuals for the study, it is necessary to follow a scientific procedure that allows the researcher to remove the bias and ensure that the respondents are selected appropriately from whom the information has to be collected. Sampling is the way to collect the group for the study and base the inferences planned to draw on the characteristics of the population. It deals with steps involved in drawing a sample and helps the researcher to justify including the respondents for the study. It also takes into consideration auxiliary information available to the researcher.

An appropriate sampling design has to be chosen to draw the sample for the study. The design decides the way the respondents are to be approached to draw the responses. It decides where to start and where to end. Depending on the sample size estimated and the resources available, one has to decide upon the sampling design. If it is mandatory to draw a random sample, then it is necessary to adopt a sampling design that takes probability into consideration, while including the respondents for the survey. Even if it is not possible to adopt such a design, it is necessary to ensure that the sample is drawn using non-random sampling techniques and then tested for randomness. This is an important assumption necessary for most of the parametric methods of estimation. If the randomness assumption fails, then it is advised to use methods that give results with same precision as that of parametric methods. The behavior of the sample drawn decides the choice of the estimation method. It can be understood that the customers chosen for the study give the responses that are not random and deterministic. In such cases, the researcher has to be cautious as the deterministic responses cannot be used directly to infer about the population that has individuals with variability in terms of their tastes and choices. This happens when the sample chosen is not a random sample. Hence, it is necessary to choose an appropriate method that ensures that the estimated values can be used to draw inferences about the population. The alternative methods for parametric methods are non-parametric, semi-parametric, robust methods.

The most frequently used random sampling techniques include simple random sampling (SRS), Stratified random sampling (STRS), Systematic random sampling (SYRS), Cluster random sampling, Two stage sampling, Two phase sampling, Multi-stage sampling etc. The non-random sampling techniques include convenient sampling, judgemental sampling, purposive sampling, snow ball sampling etc. The discussion on these can be found in the book by Cochran (1977).

Stratified random sampling was used to collect the data and optimum allocation was used to allocate the sample sizes to different subsidiaries. Cost for collecting form each hotel was estimated and given in the following table

Table-4: Allocation of sample sizes to stratum

| S.No. | Stratum | Size | Standard deviation | Cost\$ | Optimal Allocation (Sample sizes) |
|-------|----------------------|------|--------------------|--------|-----------------------------------|
| 1 | Double Tree | 133 | 59 | 3 | 47 |
| 2 | Embassy Suites | 169 | 51 | 4 | 45 |
| 3 | Hampton Inn & Suites | 1194 | 73 | 6 | 368 |
| 4 | Hilton Garden Inn | 396 | 85 | 2 | 246 |
| 5 | Homewood Suites | 128 | 68 | 4 | 45 |

Parameter, Statistic, Estimator, Estimate

The entire process of estimation deals with estimation of the parameters. A parameter is an unknown characteristic of the population or market under study. For example, it can be a population center, market variability etc. Sometimes it also can be the process that generates the responses. That is there exists a pattern in the market and that happens due to a process that prevails in the market. If the researcher is interested in finding the behaviour of the process, he can use estimation to understand the process better and take appropriate action against those events that are critical to his organization. Events can be estimated by conducting experiments by using standard designs. The definition of the parameter varies with the study and has to be defined before the study and the survey has to be carried such that precise, accurate and reliable estimates are obtained out of the sample drawn.

Statistic is a function of the sample collected and is a random quantity that is used to estimate the parameter identified for the study. It has to be noted that the statistic when used to estimate the parameter is called as an estimator. The value of the statistic is called as an estimate. The estimator has to be chosen such that it estimates the parameter accurately. This implies that the estimator possess all the properties of a good estimator. Also one has to note that the method used to derive the estimator should be selected properly. For example, maximum likelihood

method, method of percentiles etc. The estimator should be unbiased, efficient, consistent and sufficient. The sample mean is an unbiased, efficient, consistent, sufficient estimator whereas sample variance is not an unbiased estimator unless the denominator in the formula is $(n-1)$. The estimator is the mean square error. The details can be found in the paper of Deakin and Kildea (1999). Any systematic deviation from the actual value of the parameter on average is the bias that has to be minimized by selecting appropriate sample that forms a true representative of the population and of sufficient sample size. Here the sample size plays a role as it appears in the denominator of the formula used to calculate the sampling error. One can also note that non-sampling errors play a role in calculation of the estimates for the unknown parameters. Hence, the researcher has to be prepared to find means to handle the errors and set an admissible error rate before the estimation process. Usual practice is to decide upon these before the survey begins and this ensures that the sample drawn is also at the same confidence level. Another important aspect in estimation is the associated assumptions with the method used in the process. One has to ensure that these assumptions are satisfied by the data before they are used. For example, regression analysis is also used for estimating the

unknown values of the response at different levels of the repressors. The main assumption among others is that the sample is a random sample. Here, it is necessary to recall that the sample should possess the property of randomness. If the sampling design used is one of the random sampling techniques then the property of randomness is inherent. But, if the sampling technique used is non-random then, one can use run test to test for randomness. Run test will test whether the sample drawn has a pattern that is deterministic or random. The researcher has to take a decision on the method of estimation after the run test is used. The discussion ends once the method of estimation is chosen.

As per the objective, the parameter is the average total feedback forms sent and average total feedback forms received. Sample averages are used respectively to estimate these parameters.

Sampling distribution of the Statistic: Normal and Student-t

Once an estimator is chosen, one has to look upon the sampling distribution of the same in order adopt advanced analytical methods to extract more information from the sample drawn. For example, one may be interested to construct a confidence interval. To construct a confidence interval, it is very important to study the associated probability distribution. The estimator chosen will be used to obtain the point estimate and the same point estimate is

used to construct the interval estimate for the parameter under study. Note that any number of point estimates can be constructed with change in the samples. Hence, it is possible to study the patterns in the estimates using an appropriate probability law so that one can obtain the probability that the estimator generates the estimates between the limits. The advantage the researcher has by doing this exercise is that he can focus on those regions where the chances of concentration are high. If the parameter under estimation is population mean, then an interval will give the estimates for the center and the region around which it can fluctuate with a confidence level. It also helps the researcher to study the behavior of the estimator even before taking the sample with the information of few parameters associated with the estimator. The process of identifying the appropriate probability law depends on the type of estimator one chooses for estimation. Most of the times researchers try to show that the behaviour of the estimator after adjustment will be asymptotically normal. This is one of the reasons why normal distribution has been given prominent position in theory of estimation. Also the properties of normal distribution are vital in drawing conclusions related to the parameters of the population under study.

Normal probability law can be derived as a limiting case of many discrete or continuous probability laws. For example, limiting case of Binomial and Poisson laws can be shown as normal. This happens as the sample size increases and approaches to a limit that makes the behaviour of these

laws asymptotically a closer to the behavior of a normal law. This can be experienced when the researcher is working with the market center. One can use central limit theorem to study the behavior of the sample average. When the sample size increase, the behaviour of sample mean will be normal. But, the condition for this to happen is that the variance of the population is known. The subtle understanding of the assumption is that the researcher has an understanding of the market and how the market fluctuates with respect to the fluctuations in factors affecting the market. Note that the asymptotic distribution is standard normal law that is centered at 0 and symmetric. About other properties relating to normal law can be found in Patel and Read (1982).

When the variance is unknown the alternative law to normal law is student-t. William Gosset in his paper "probable error of mean" (1908) proposed the "t" distribution that has similar behaviour as that of normal law. It is centered at 0, symmetric and has attains the behavior of normal law as the sample size increases.

Both normal law and Student-t law are used to study the behaviour of the sample average. The discussion on this ends after the researcher identifies appropriate probability law as the sampling distribution of the estimator. The same can be used to construct the confidence interval and further hypothesis testing on the parameters. The important aspect of either of the cases is that the assumptions associated with the student-t should

be satisfied by the sample drawn. One such assumption is that the population follows normal law. Even though the population do not follow normal, we assume that it follows normal and use student-t law. Bradley (1968) discusses the use of student-t under non-normal conditions. Evarist, Friedrich and Mason (1997) give conditions under which student-t statistic will be asymptotically normal. The literature available on the usage of student-t under non-normal conditions make student-t robust.

The sampling distribution for the present case is the student-t distribution as the population variance is unknown. Here we assume that the population follows normal distribution.

Estimation

The researcher knows well that purpose of the sampling is to study the key parameters associated with the market by adopting the estimation methods that are available in the literature. Estimation itself a special branch and a research area is a collection of various methods that can be chosen based on the objectives of the study. An important step in the process is to identify the key parameters associated with the study. This is because most of the times the researcher miss key parameters and proceed in the survey and after a point of time realize that parameters was missed out from the estimation process. To overcome this, the researcher has to conduct a well-designed and implemented pilot study that captures all the parameters whose values affect the market and the decisions of the researcher. The process

includes both point and interval estimation. Parametric, semi-parametric, non-parametric, Bayesian, and robust statistical methods form the class of all methods that are being used to estimate the parameters. The process also takes the quality of the sample drawn from the population into consideration, as the results obtained from the methods depends on precision of the sample. Recall that sample size estimation takes into consideration the precision with which the parameter is estimated and associated cost.

From the above details, data has been collected and the point estimates for the required parameters is presented in the following table:

Table 5: Point Estimates

| Subsidiary | Mean | | Standard deviation | |
|------------|------------|----------------|--------------------|----------------|
| | Forms sent | Forms received | Forms sent | Forms received |
| 1 | 482 | 68 | 62 | 10 |
| 2 | 486 | 80 | 94 | 17 |
| 3 | 382 | 68 | 95 | 22 |
| 4 | 343 | 52 | 80 | 18 |
| 5 | 321 | 58 | 65 | 14 |

Now using these point estimates, interval estimates for mean were constructed and presented in the following table:

Table 6 : 95% Confidence intervals

| Subsidiary | 95% Confidence Interval Forms sent | | 95% Confidence Interval Forms received | |
|------------|---------------------------------------|-------------|---|-------------|
| | Lower Limit | Upper Limit | Lower Limit | Upper Limit |
| 1 | 448 | 515 | 62 | 73 |
| 2 | 460 | 512 | 75 | 85 |
| 3 | 370 | 394 | 66 | 70 |
| 4 | 322 | 363 | 55 | 49 |
| 5 | 301 | 341 | 53 | 63 |

Few Additional Aspects

Central Limit Theorem

The most important development in statistics that attracted many corporates and academicians is central limit theorem (CLT). The main application of this theorem is in studying the behaviour of the sample mean which is used more frequently in estimating the behavior of market average. It also helps in studying the same even if the

market behavior is far from normal behavior. Hans Fischer (2011) discusses the history starting from classical to modern theory. This book gives the reader a good idea on the research on central limit theorem and its usage.

The role of CLT in estimation is important as it helps the market researcher to construct the interval estimates for the market average. The sample average which is a random variable has to be modelled using appropriate probability law in order to obtain the areas of concentration with accurate probability. When the population is normal, the sampling distribution of the sample mean is also normal. But, when the population is non-normal, one can study the asymptotic behaviour of sample mean using CLT, under the condition that the variance of the market is known. That is if the researcher has an idea about the market fluctuation then he can study the behaviour of the sample average using CLT. Understanding the market fluctuation is possible if the researcher has prior information regarding the market he is working or records on similar studies in the market are available. This implies that information on market fluctuation should be available to the researcher if he wishes to use CLT to study the behaviour of the sample average. If the market fluctuation is available from the secondary records then before using the same, he has to test whether the sample drawn from the current market support the same. Chi-square test for significance of population variance can be used for this purpose. If the sample do not support the same, one has to use student-t law.

Degrees of Freedom

When the population is non-normal, it is very important to ensure that sample average will be modelled using student-t law. This law is characterized by the quantity called as degrees of freedom. Degrees of freedom is calculated based on the number of restrictions on the measurements, number of parameters to be estimated and number of missing observations. Degrees of freedom also gives the strength of the sample in estimating the parameters and also estimating given situations or processes. That is, if there are more missing observations in the sample, then it is necessary to estimate them using the remaining observations. In such cases, we lose the degrees of freedom and this weakens the power of the sample in drawing precise conclusions on the parameters. Suppose that the study is aimed at understanding the process that is generating bubbles about the product in the market. If, more parameters need to be estimated in order to understand the behavior of the process then, the degrees of freedom are reduced based on the parameters estimated. This is not a desired situation because as the number of parameters increase, dependency on the sample to estimate these parameters increases and estimating each parameter with desired accuracy is not an easy task. Ultimately one has to understand that it is better to have more degrees of freedom while estimating the parameters as it decreases the sampling error.

Conclusion

We end the discussion with a note that care has to be taken while using statistical tools in studying the market behaviour. Also, proper estimating tools with good properties and algorithms to compute the same has to be selected appropriately.

The points mentioned and the subsequent data analysis is only an attempt to highlight the subtle aspects in estimation. Either the points stated or the analysis done need not be exhaustive and can vary with problem under study and data.

Acknowledgement

We thank the referee for his valuable comments that helped us to present the discussion in a better way.

References

- Cochran, W.G. (1977): Sampling techniques. John Wiley & Sons.
- Patel, J.K. and Read, C.B. (1982): Normal Distribution. Marcel Dekker, Inc.
- Evarist, G., Friedrich and Mason, D.M. (1997): When is the Student t-Statistic asymptotically standard normal. The Annals of Probability, Vol. 25, No. 3, pp-1514-1531.
- Fischer, H. (2011): A History of the Central limit theorem- From Classical to Modern Probability Theory. Springer.

Deakin, R.E. and Kildea, D.G. (1999): A note on Standard deviation and RMS. The Australian Surveyor, Vol.44, No.1, pp-74-79.

Bradley, E. (1969): Student's t-test under non-normal conditions. Journal of American Statistical Association. Vol. 64, No. 328, pp-1278-1302.

Mace, E.A. (1964): Sample Size Determination. Reinhold Publishing Corporation.