# 1. Introduction

Welcome to the series of E-learning modules on problems associated with sampling. In this module we are going to cover the Typical problems in sampling, sampling bias, selection and Non responsive error, errors due to sampling frame and the possible remedial actions.

By the end of this session, you will be able to:

- Explain sampling bias and Selection bias
- Explain the causes of problems
- Describe Measurement Errors and Response Bias
- Explain about typical problems of sampling
- Explain the types of sampling bias
- Describe the problems due to sampling frame
- Describe non response errors
- Explain possible remedial actions

Sampling is carried out in order to save time, energy and cost by studying only a part of a given population.

But sampling is not an easy task to perform. Even if utmost care has been taken in selecting sample the results derived from the sample study may not be exactly equal to the true value in the population.

The reason is that the estimate is based on a part and not on the whole and samples are seldom, and is a perfect miniature of the population.

Hence, sampling gives rise to certain fluctuations. These fluctuations would not be present in complete enumeration survey. The error in estimation arises solely from the random sampling variation that is present when units are measured from the sample of the population instead of complete enumeration.

We may come across with different types of problems such as: failure to measure some of the units in the chosen sample. This may occur with oversight or with human populations because of failure to locate some individuals or the refusal to answer the questions when located. The measuring device may be biased or imprecise. With human population the respondents may not possess accurate information or they may give biased answers.

## 2. Sampling Surveys and Problems

In sampling surveys there would naturally be certain amount of inaccuracy in the information collected; there is a tendency of a sample estimator to systematically over- or under-estimate a population parameter.

Causes of Problems: Problems may arise due to:

 Selection Bias: Errors often occur when the survey sample does not accurately represent the population. The bias that results from an unrepresentative sample is called selection bias. Some common examples of selection bias are as follows: Biased errors: These errors arise from any bias in selection, estimation etc. For example, if in place of Simple Random Sampling, Deliberate Sampling has been used in a particular case and some bias has been introduced in the result. Hence, such errors are called biased sampling errors.

Unbiased errors: These errors arise due to chance differences between the members of the population included in the sample and those not included.

Thus, the total sampling problems is made up of random sampling errors and errors due to bias if any. The essence of bias is that, it forms a constant component of error that does not decrease in a large population as the number in the sample increases. Such errors are also known as cumulative and non compensating error. The random sampling error on the other hand decreases on an average as the size of the sample increases. Such errors are also known as non cumulative or compensating error.

Under coverage: Under coverage occurs when some members of the population are inadequately represented in the sample.

A classic example of under coverage is the Literary Digest voter survey, which predicted that Alfred Landon would beat Franklin Roosevelt in the nineteen thirty six presidential election. The survey sample suffered from under coverage of low-income voters, who tended to be Democrats.

How did this happen? The survey relied on a convenience sample, drawn from telephone directories and car registration lists. In Nineteen thirty six, people who owned cars and telephones tended to be more affluent. Under coverage is often a problem with convenience samples.

Faulty selection of the sample may give rise to bias in number of ways, such as:

- a) Deliberate selection of representative sample.
- b) Conscious and Unconscious bias in the selection of a "random sample". The

randomness of selection may not really exist even though the investigator claims that he had a random sample if he allows his desire to obtain a certain result to influence his selection.

- c) Substitution: Substitution of an item in place of one chosen in random sample sometimes leads to bias. Thus, if it were decided to interview the fifty first or any other number in any place as the characteristics possessed by them, differ from those who were originally to be included in the sample.
- d) Non response: If all the items to be included in the sample are not covered there will be bias even though no substitution has been attempted.

Sometimes, individuals chosen for the sample are unwilling or unable to participate in the survey. The Literary Digest survey illustrates this problem. Respondents tended to be Landon supporters; and non respondents, Roosevelt supporters. Since only twenty five percent of the sampled voters actually completed the mail-in survey, survey results overestimated voter support for Alfred Landon.

The Literary Digest experience illustrates a common problem with mail surveys. Response rate is often low, making mail surveys vulnerable to non response bias. This fault particularly occurs in mailed questionnaires which are incompletely returned. Moreover the information supplied by the informants may also be biased.

- e) An appeal to the vanity of the person questioned may give rise to another kind of bias. For example: the question "are you a good student?" is such, that most of the students would succumb to pride and answer 'Yes'.
- ✓ Voluntary response bias: Voluntary response bias occurs when sample members are self-selected volunteers, as in voluntary samples. An example would be call-in radio shows that solicit audience participation in surveys on controversial topics (abortion, affirmative action, gun control, etc.). The resulting sample tends to over represent individuals who have strong opinions.
- 2) Faulty work during the collection: Any consistent error in measurement will give rise to bias and the danger of error is likely to be greater in sampling since the units measured are often smaller. Bias may arise due to improper formulation of the decision, wrongly defining the population, securing an inadequate frame and so on. Biased observations may result from a poorly designed questionnaire, an ill-trained interviewer, failure of a respondent's memory etc.
- 3) Faulty methods of analysis: Faulty methods of analysis may also introduce errors. Such bias may be avoided by adopting the proper methods of analysis.

## 3. Measurement Error

#### Measurement Error

A poor measurement process can also lead to trouble. In survey research, the measurement process includes the environment in which the survey is conducted, the way that questions are asked, and the state of the survey respondent.

Response bias refers to the bias that results from problems in the measurement process. Some examples of response bias are as follows:

Leading questions: The wording of the question may be loaded in some way to unduly favour one response over another. For example, a satisfaction survey may ask the respondent to indicate whether the respondent is satisfied, dissatisfied, or much dissatisfied. By giving the respondent one response option to express satisfaction and two response options to express dissatisfaction, this survey question is biased toward getting a dissatisfied response.

Social desirability: Most people like to present themselves in a favourable light, so they will be unwilling to admit unsavoury attitudes or illegal activities in a survey, particularly if survey results are not confidential. Instead, their responses may be biased toward what they believe is socially desirable.

Sampling Errors: A survey produces a sample statistic, which is used to estimate a population parameter. If we repeat a survey many times, using different samples each time, we might get a different sample statistic with each replication giving rise to varied estimates for the same population parameter.

If the statistic is unbiased, the average of all the statistics from all possible samples will equal to the true population parameter; even though any individual statistic may differ from the population parameter. But there will be variability among statistics from different samples which is called as sampling error.

For example: Population mean (mue) gross income is forty-two thousand and two hundred dollars. From the Sample 1 of size 400; a sample mean (x bar) is obtained as forty one thousand one hundred dollars. From the Sample 2 of size 400; a sample mean (x bar) is obtained as forty three thousand seven hundred dollars.

The meaning of sampling error can be easily understood from the following diagram:

Sampling error = Frame error + Chance error + Response error

Sampling errors can be controlled by (1) Careful sample designs, (2) Large samples, and (3) Multiple contacts to assure representative response.

Sampling Error: Error due to inappropriate selection of sample size. It can be minimized by choosing the appropriate sample size. As the sample keeps on increasing, the sampling error decreases.

Example: If a study is done amongst Maruti-car owners in a city to find the average monthly expenditure on the maintenance of car, it can be done by including all Maruti-car owners.

Increasing the sample size tends to reduce the sampling error; that is, it makes the sample statistic less variable. However, increasing sample size does not affect survey bias. A large sample size cannot correct for the methodological problems (under coverage, non response bias, etc.) that produce survey bias.

Sampling errors occurs randomly and are equally likely to be in either direction. The magnitude of the sampling error depends on the nature of the population; the more homogeneous universe the smaller the sampling error.

#### Figure 1



Typical Problems of sampling:

- Sample is of insufficient size
- The sample is biased
- The wrong variables were measured
- The sample is censored
- The data precision is low

Sampling Bias are the error resulting from taking a non-random sample of a population. <u>What is it?</u>

- Based on sampling method used, some members of a population are less likely to be included in the sample
- Reduces the ability for results to be generalized to a larger population
- More often, sampling bias occurs because of the inherent difficulty in obtaining a truly representative sample of a complex population

# 4. Types of Sampling Bias

Types of sampling Bias

1.Selection from only a specific area of the population (intentional "purposive", or accidental "convenience sample")

• Biased if certain members are under-represented or overrepresented relative to others in the population being generalized to

2.Self-selection bias

- Participants' decision to participate may be correlated with traits that affect the study, making the participants a non-representative sample
- 3. Pre-screening of or advertising for volunteers within particular groups
  - May also include selecting only certain kinds or groups of subjects in order to intentionally skew the sample toward a certain desired trait or characteristic
  - People who have strong opinions or substantial knowledge may be more willing to spend time answering a survey than those who do not
  - Online and phone-in polls are biased samples because the respondents are selfselected. This means that people with strong opinions are more likely to respond to the poll than people who have less strong opinions

#### Sampling Frame Error

Error in the specific list of population units, from which the sample for a study is being chosen, is known as sampling frame error.

Example: Assume that a bank wants to contact the people belonging to a particular profession over phone to market a home loan product.

The sampling frame in this case is the telephone directory. This sampling frame may pose several problems: 1) People might have migrated. 2) Numbers have changed. 3) Many numbers were not yet listed. Residents who are included in the directory are likely to differ from those who are not included.

The record of nineteen ninety eight cannot be used for two thousand eight or nine as it becomes outdated. This is one problem

Secondly the name and addresses mentioned are often inappropriate and the duplication of errors can also occurs.

Thirdly there is no record of those who has left an area and those who has come to live here. These are the some problems relating to the sampling frame.

## 5. Non Response Error

#### Non response error

The two major non response issues in sampling are improving response rates and adjusting for non response. Non response error arises when some of the potential respondents included in the sample do not respond.

The primary causes of low response rates are refusals and not-at-homes. Refusals, which result from the unwillingness or inability of people included in the sample to participate, result in lower response rates.

Attempts to lower refusal rates:

- 1. Prior notification Potential respondents are sent a letter notifying them of the coming up mail, telephone, personal or internet survey.
- 2. Motivating the respondents The interviewer starts with a small request such as 'Will you please take five minutes to answer five questions?' which is followed by a larger request (foot-in-the-door strategies). In the reverse strategy, the initial request is relatively large, followed by a smaller request (door-in- the-face strategy).
- 3. Incentives Offering monetary as well as nonmonetary incentives to potential respondents.
- 4. Questionnaire design and administration A well designed questionnaire can decrease the overall refusal rate as well as refusals to specific questions.
- 5. Follow-up Contacting the non respondents periodically after the initial contact or by sending a letter to remind non respondents to complete and return the questionnaire.
- With non coverage and non response the most important consequences is that estimates may become biased because the part of the population that is not reached may differ from the part that is sampled. There is now ample evidence that these biases vary considerably from item to item and from survey to survey being sometimes negligible and sometimes large.
- A second consequence is of course that the variances of estimates are increased because the sample actually obtained is smaller than the target sample. This factor can be allowed or at least approximately in selecting the size of the sample.
- Errors of measurement that are independent from unit to unit within the sample and average to zero over the whole population are properly taken into account. These are taken into account in the usual formulae for computing the standard errors of the estimates provided that finite population correction terms are negligible. Such errors decrease the precision of the estimates and it is worthwhile to find out whether this decrease is serious.
- If errors of measurement on different units in the sample are correlated the usual formulae for the standard errors are biased. The standard errors are likely to be too

small since the correlations are mostly positive in practice. This type of disturbance is easily overlooked and may often have passed unnoticed.

• A constant bias that affects all units alike is hardest of all to detect. No manipulations of the same data will reveal this bias.

Hence sampling theory has got its own problems which may be briefly outlined as follows:

- Proper care should be taken in the planning and execution of the sample survey otherwise the results obtained may be inaccurate and misleading
- Sampling theory requires the services of experienced, trained and qualified personnel and sophisticated equipment for its planning, execution and analysis, in the absence of these the results of the sample survey are not trust worthy. In India shortage of experts in the sampling field is a serious hurdle in the way of reliable statistics
- At times the sampling plan may be so complicated that it requires more time, labour and money than the complete count. This is so, if the size of the sample is a large proportion of the total population and if the complicated weighted procedures are used. With each additional complication in the survey, the chances of error multiply and a greater care must be taken which in turn means more time and labour
- However if time and money are not important factors or if the population is not too large a complete census may be better than any sampling method

Kaplan's "paradox of sampling"

On the one hand, the sample is of no use if it is not truly representative of its population, if it is not a "fair" sample. To know that it is representative, we must know what the characteristics of the population are, so that we can judge whether the sample reflects them properly; but in that case, we have no need of the sample at all.

Selecting a representative sample is very necessary for the validity of our data. If our sample do not represents the opinion of its population then it is biased and not reliable. A study of the problems sampling is slow and difficult.

One has to work hard in order to overcome the problems associated with sampling. Lesser the sampling problems the more representative our sample is, the more representative our sample is the more valid and reliable our estimates will be.

Here's a summary of our learning in this session:

- Sampling bias and Selection bias
- Causes of problems
- Measurement Errors and Response Bias
- Typical problems of sampling
- Types of sampling bias
- Problems due to sampling frame
- Non response errors
- Possible remedial actions