Frequently Asked Questions

1. What do you mean by a sampling error and how it can be controlled?

Answer: The errors which are introduced due to the errors in the selection of the sample or the differences between population's parameters and estimates which are derived from a random sample. Sampling errors arise solely as a result of drawing a probability sample rather than conducting a complete enumeration. Sampling errors can be controlled by increasing the sample size.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.Sampling errors can be controlled by (1) Careful sample designs, (2) Large samples, and (3) Multiple contacts to assure representative response.

2. Explain Non-sampling errors.

Answer: An error in the sample estimates cannot be attributed to sampling fluctuations. The studies based on complete enumeration do not yield similar results in repealed enumerations. Such a difference occurs due to many errors which are termed as Non-sampling errors. Non-sampling errors are mainly associated with data collection and processing procedures.Non-sampling errors, therefore, arise mainly due to misleading definitions and concepts, inadequate frames, unsatisfactory questionnaires, defective methods of data collection, tabulation, coding, incomplete coverage of sample units etc. These errors are unpredictable and not easily controlled. Unlike in the control of sampling error this error may increase with increases in sample size. If not properly controlled nonsampling error can be more damaging than sampling error for large-scale household surveys.Non-sampling errors can be defined as errors arising during the course of survey activities rather than resulting from the sampling procedure due to mistakes and inaccuracies in the collection and processing of data. The non-sampling errors can never be eliminated.

3. What are the sources of non-sampling errors?

Answer:

1) Observational or Response errors

   If the observations are taken repeatedly on the same unit the observed values generally differ or otherwise even the same respondent is asked the same question repeatedly his response may differ.

2) Lack of preciseness in definition

This also adds to non-sampling errors. For eg: In judging the loss of crop due to disease like wilt or rust , will be subject to error due to definitions of what we call severely diseased, moderately diseased and low intensity of disease. More this measure o intensity will vary from person to person depending on the maturity, qualifications and training the person has.

3) Errors are also introduced in editing and tabulation of data

4. What are the factors which are responsible for non-sampling errors ?

Answer:

In general, non-sampling errors may arise from one or more of the following factors:

a. Data specification being inadequate and/or inconsistent with respect to objectives of the survey.

b. Duplication or omission of units due to imprecise definition of the boundaries of area units, incomplete or wrong identification particulars of units or faulty methods of enumeration.

c. Inappropriate methods of interview, observation or measurement using ambiguous questionnaires, definitions or instructions.

d. Lack of trained and experienced field enumerators including lack of good quality field supervision.

e. Inadequate scrutiny of the basic data.

f. Errors in data processing operations such as coding, keying, verification, tabulation etc.

g. Errors during presentation and publication of tabulated results.

5. State the components of non-sampling errors

Answer: Brieumer and Lyberg (2003) identify five components of non-sampling error, namely

- Specification error
- Coverage or frame error
- Non response
- Measurement and processing error
- Estimation error

6. Write a note on specification error.

Answer: The specification error occurs when the concept implied by the question is different from the underlying construct that should be measured. A simple question such as how many children do a person have can be subject to different interpretations in some cultures. In households with extended family member's biological children may not be distinguished from children of brothers or sisters living in the same household. In a disability survey, a general question asking people whether or not they have a disability can be subject to different interpretations depending on the severity of the impairment or the respondent's perception of disability. People with minor disabilities may perceive themselves to have no disability. Unless the right screening and filter questions are included in the questionnaire, the answers may not fully bring out the total number of people with disabilities.Specification Errors are the errors which arise due to faulty planning. They arise due to three causes

1) Wrong data specification which may be inconsistent with the objectives of survey or census

2) Sampling units may be inadequately defined or they may be wrongly identified. Sometimes units may be duplicated or may be overlapping. The methods of enumeration may be faulty.

3) The questionnaire/schedules may be faulty. The methods of interview or observation and measurement may be inappropriate.

7. What do you mean by Coverage or frame error?

Answer: In most area surveys primary sampling units comprise clusters of geographic units generally called enumeration areas (EAs). It is not uncommon that the demarcation of EAs is not properly carried out during census mapping. Thus households may be omitted or duplicated in the second stage frame. Frame imperfections can bias the estimates in the following ways: If units are not represented in the frame but should have been part of the frame, this result in zero probability of selection for those units omitted from the frame. On the other hand if some units are duplicated, this results in over coverage with such units having larger probabilities of selection. Errors associated with the frame may, therefore, result in both over coverage and under coverage. The latter is the most common in large-scale surveys in most African countries.In multi-stage household surveys, which are common in the Southern African Development Community region, sampling involves a number of stages, such as selection of area units in one or more stages; listing and selection of households; and listing and selection of persons within selected households. Coverage error can arise in any of these stages.Non coverage denotes failure to include some sample units of a defined survey population in the sampling frame. Because such units have zero probability of selection, they are effectively excluded from the survey results. It is important to note that we are not referring here to deliberate and explicit exclusion of sections of a larger population from survey population. Survey objectives and practical difficulties determine such deliberate exclusions. For example attitudinal surveys on marriage may exclude persons under the minimum legal age for marriage. Residents of institutions are often excluded because of practical survey difficulties. Areas in a country infested with landmines may be excluded from a household survey to safeguard the safety of field workers. When computing non coverage rates, members of the group deliberately and explicitly excluded should not be counted either in the survey population or under

noncoverage. In this regard defining the survey population should be part of the clearly stated essential survey conditions.

8. What do you mean by non coverage?

Answer: Non coverage is often associated with problems of incomplete frames. Examples are to omissions in preparing the frame but also missed units, implying omissions due to faulty execution of survey procedures. Thus non coverage refers to the negative errors resulting from failure to include elements that would, under normal circumstances, belong in the sample. Positive errors of over coverage also occur due to inclusion in the sample of elements that do not belong there. The term gross coverage error refers to the sum of the absolute values of non coverage and over coverage error rates. Most household surveys in developing countries suffer mainly from under coverage errors. Most survey research practitioners agree that in most social surveys under coverage is a much more common problem than over coverage. Corrections and weighting for non-coverage are much more difficult than for non responses, because coverage rates cannot be obtained from the sample itself, but only from outside sources.The non coverage errors may be caused by the use of faulty frames of sampling units. If the frames are not updated or old frames are used as a device to save time or money, it may lead to serious bias. For example, in a household survey if an old list of housing units is not updated from the time of its original preparation say 10 years prior the current survey, newly added housing units in the selected enumeration area will not be part of the second stage frame of housing units. Similarly, some disbanded housing units will remain in the frame as blanks.

9. Briefly explain Non response.

Answer: Non response refers to the failure to measure some of the sample units. Thus failure to obtain observations on some units selected for the sample. It is instructive to think of the sample population as split into two strata, one consisting of all sample units for which measurements can be obtained and the second for which no measurements could be obtained.In most cases non response is not evenly spread across the sample units but is heavily concentrated among subgroups. As a result of differential non response, the distribution of the achieved sample across the subgroups will deviate from that of the selected sample. This deviation is likely to give rise to non response bias if the survey variables are also related to the sub-groups. The non response rate can be accurately measured if accounts are kept of all eligible elements that fall into the sample. Response rate for a survey is defined as the ratio of the number of questionnaires completed for sample units to the total number of sample units. Reporting of non response is good practice in surveys. Non response can be due to respondents not being -at-home, refusing to participate in the survey, being incapacitated to answer questions and to lost schedules/ questionnaires. All categories of non-response refer to eligible respondents and should exclude ineligibles. For example, if a survey is on fertility, the ultimate frame in the selected EAs will comprise only women in the reproductive age groups, thus excluding, for instance, young females who are not in this group.

10. Distinguish between non response and non coverage?

Answer: Non coverage errors differ from non response. The latter, results from failure to obtain observations on some sample units, due to refusals, failure to locate addresses or find respondents at home and losses of questionnaires. The extent of non response can be measured from the sample results by comparing the selected sample with that achieved. By contrast the extent of non coverage can only be estimated by some kind of check external to the survey operation, Sample selection and implementation errors

11. Explain the two types of non response.

Answer: There are two types of non-responses:

- ✓ Unit non response and
- ✓ Item non response.

Unit non response implies that no information is obtained from certain sample units. This may be because respondents refuse to participate in the survey when contacted or they cannot be contacted.

Item non response refers to a situation where for some units the information collected is incomplete. Item non response is therefore, evidenced by gaps in the data records for responding sample units. Reasons may be due to refusals, omissions by enumerators and incapacity.

The magnitude of unit (total) non response, among other reasons, is indicative of the general receptivity, complexity, organization and management of the survey. The extent of item non response is indicative of the complexity, clarity and acceptability of particular items sought in a questionnaire and the quality of the interviewer work in handling those items.

12. Discuss the measurement error or observed error.

Answer: These errors arise from the fact that what is observed or measured departs from the actual values of sample units. These errors centre on the sustentative content of the survey such as definition of survey objectives, their transformation into usable questions, and the obtaining, recording, coding and processing of responses. These errors concern the accuracy of measurement at the level of individual units.For example at the initial stage wrong or misleading definitions and concepts on frame construction and questionnaire design lead to incomplete coverage and varied interpretations by different enumerators leading to inaccuracies in the collected data.Inadequate instructions to field staff are another source of error. For some surveys instructions are vague and unclear leaving enumerators to use their own judgement in carrying out fieldwork. At times sample units in the population lack precise definition, thereby resulting in defective and unsatisfactory frames. The enumerators themselves can be a source of error. At times the information on items for all units may be wrong, this is mainly due to inadequate training of field workers.

For example, age reporting in Africa is another common measurement problem through age heaping and digital preference. Depending on the type and nature of enquiry or information collected, these errors may be assigned to respondents or

enumerators or both. At times there may be interaction between the two, which may contribute to inflating such errors. Likewise, the measurement device or technique may be defective and may cause observational errors. Reasons for such errors are:

- Inadequate supervision of enumerators.

- Inadequately trained and experienced field staff.

- Problems involved in data collection and other type of errors on the part of respondents.

- Failure to understand the question.

13. What are the errors that creep in during data collection? Explain

Answer:

These are the errors that creep in during field work. These can be classified into the following:

1) The investigators may lack training and experience. They may also lack the will and enthusiasm to collect the data properly. Sometimes they are likely to replace one sampling unit by another, easily approachable. Even if they are honest there are likely to be errors of observations and measurement
2) There may be practical difficulties in the collection of the correct data. For example in the collection of data on farm yields there are likely to be genuine errors of measurement. In the case of mailed questionnaire the respondents since they are left to themselves in supplying data, may introduce errors due to wrong understanding of the questions. The errors due to non response may also be included here.
3) There may be accidental loss of information or recording of incorrect data
4) Field inspection and supervision may be inadequate

   At times respondents may introduce errors because of the following reasons:

   - Careless and incorrect answers from respondent due to, for example, lack of adequate understanding of the objective(s) of the survey

   - Respondents answering questions even when they do not know the correct answer.

   - Deliberate inclination to give wrong answers, for example, in surveys dealing with sensitive issues, such as income and stigmatised diseases.

   - Memory lapses if there is along reference period

   14. Discuss the concept of classification and tabulation errors.

   Answer: These are the errors in processing the data before they are subjected to mathematical analysis. We may classify them as follows

1) Errors due to insufficient scrutiny of the collected data

2) Errors in coding , punching, verification and tabulation
3) Errors may also be introduced while programming, presenting the data, in printing and other operation

15. Explain the concept of errors in estimation.

Answer:

These arise in the process of extrapolation of results from the observed sample units to the entire target population. This group of errors centres on the process of sample design, implementation and estimation.

Biases of the estimating procedure may either be deliberate, due to the uses of a biased estimation procedure or it may be due to inadvertent use of wrong formula.