Controlling Sampling and Non-sampling errors and precision of estimates

(Indian Experience)

1. Sampling error

1.1 The error arising due to drawing inferences about the population on the basis of observations on a part or sample of it is termed **sampling error** .Clearly, the sampling error in this sense is non-existent in case of a complete enumeration or census.

2. Non-sampling error

The errors mainly arising at the stages of data collection and processing of data are termed **non-sampling errors** .Thus non-sampling errors are common to both censuses and sample surveys.

3. Behaviour of sampling and non-sampling errors with sample size

3.1 The magnitude of sampling and non-sampling errors has opposite relationship with the sample size. While the sampling error generally decreases, the non-sampling errors do increase with the increase in sample size. Under many situations, the sampling error is found to be inversely proportional to the square root of the sample size.

3.2 There is a drastic reduction in the sampling error with increase in sample size at the initial stage. After a certain stage, the gain in reduction of sampling error may not be commensurate with the increase in cost due to increase of sample size. Thus, there is a strong case for resorting to sampling rather than census or complete enumeration for providing estimates with permissible margin of error.

4. Measures of error

4.1 Since a probability scheme gives rise to different samples, the estimates based on the sample observations will, in general, differ from sample to sample and also from the value of parameter under consideration. The difference between the estimate based on a particular sample and the parameter may be called the error of the estimate. Thus, sampling error varies from sample to sample.

4.2 An average measure of the divergence is given by the expected value of the squared error, which is: $M(t) = E\,(t-\Theta)^2$

The above measure is known as **mean square error** (mse) of the estimator t of Θ, the population parameter under study .

5. MSE

5.1 The mean square error of the estimator 't' can be written as:

$$M(t) = E\,(t-\Theta)^2 = \sum (t-\Theta)^2\, p_i = V(t) + [B(t)]^2$$

This Courseware has been received with thanks from the National Academy of Statistical Administration (NASA), CSO, India for use in the SAARCSTAT Web portal. Necessary editing has been done, wherever needed, without distortion of facts and figures.

which shows that the mse is the sum of the sampling variance V(t) and the square of the bias B(t).

6. RSE (Relative standard error)

6.1 If 't' is **an unbiased estimator** of the parameter Θ , the mse and the sampling variance become the same. The square root of the sampling variance is termed the **standard error** of the estimator 't'.

6.2 The ratio of the standard error of the estimator to the expected value of the parameter is known as **relative standard error** (rse) or **coefficient of variation** (CV) of the estimator. Conventionally, CV is expressed as a percentage.

7. Sources of non-sampling errors

7.1 Non-sampling errors can occur at every stage of planning and execution of the census or survey. Some of the major factors contributing to non-sampling errors are the following:

- data specification being inaccurate and inconsistent;

- omission or duplication of units

- inaccurate/inappropriate methods of interview, observation or measurement;

- lack of trained and experienced investigators

- recall and other types of errors;

- lack of adequate inspection and supervision of work;

- inadequate scrutiny of the basic data;

- errors in data processing operations such as coding, verification, tabulation;

- errors in printing and presenting of tabulated results;

- use of defective frames and faulty selection of sampling units

8. Measurement and control of non-sampling errors

8.1 The adoption of suitable methods for assessing and controlling non-sampling errors require careful consideration even before the initiation of the main census or survey operation . Just as a substantial increase is required to effect even marginal reductions in the sampling error after a certain stage. Enormous cost and effort would be needed if the last few traces of non-sampling errors were to be removed .

8.2 A rational approach to the problem of controlling non-sampling errors will be to try to reduce them as much as possible to levels at which the results will be usable but not to such an extent as will render the efforts and costs to become incommensurate with the improvements achieved .

Some of the common procedures of controlling non-sampling errors are as under:

- Providing detailed guidelines for data collection and data processing

- Imparting proper training to the field workers and data processing personnel;

- Introducing consistency checks

- Performing sample check

- Carrying out post-census and post-survey checks

- Performing external record check

- Introducing the scheme of interpenetrating sub-samples