

Stat-3203: Sampling Technique-II

(Chapter-4: Sampling Errors and Non-sampling Errors)
(Section-B)

Md. Menhazul Abedin

Lecturer

Statistics Discipline

Khulna University, Khulna-9208

Email: menhaz70@gmail.com

Why have to know errors?

- Sample estimate will always be subject to deviation from parameter.

- \bar{y} is the unbiased estimate of \bar{Y} with **Standard deviation (Standard error)** $\sqrt{\frac{N-n}{Nn}} S$

(in case of Simple random sampling)

- This error mainly appeared from two sources
 - Due to sample
 - Other than sample
- It is important to understand common sources and types of errors so you can avoid them.

Sampling Errors and Non-sampling Errors

- The errors involved in collection, processing and analysis of the data in survey may be classified
 - Sampling error
 - Non-sampling error

Sampling Errors

- Sampling Errors(1): The error which arises due to only a sample being used to estimate the population parameter is termed sampling error or sampling fluctuation.
- Sampling Errors(2): Sampling error is the error that arises in a data collection process as a result of taking a sample from a population rather than using the whole population.

Sources of Sampling Errors

- Population specification error
- Sample frame error
- Selection error
- Non-response
- Sampling errors

Please visit

<https://www.qualtrics.com/blog/frequent-sampling-errors/>

Sampling Errors

- If the sample size n is equal to N we expect that sampling error will be zero.
- The decrease in sampling error is inversely proportional to the square root of the sample size.



Sampling Errors

- The amount of sampling error decreases with increase in the sample size but surprisingly it becomes otherwise in case of non-sampling error.
- The sampling errors are assigned to an estimate because it is based on a 'part' from the 'whole' while non-sampling errors are assigned because there is departure from the prescribed rules of the survey, such as survey design, field work, tabulation and analysis of data , etc.

How to remove?

- There is only one way to eliminate this error. This solution is to eliminate the concept of sample, and to test the entire population.
- In most cases this is not possible; consequently, what a researcher must do is to minimize sampling process error. This can be achieved by a proper and unbiased probability sampling and by using a large sample size.

Non-sampling error

- Non-sampling error can occur at any one or more stages of a survey i.e. planning, field work, and tabulation of survey data

Non-sampling error

- Besides sampling error, the sample estimate may be subject to other error which, grouped together, are termed non-sampling errors.
- Non-sampling error is the error that arises in a data collection process as a result of factors other than taking a sample.

Sources of Non-sampling error

- Non sampling errors broadly grouped are in three number
 - **Group A:** Errors resulting from inadequate preparation (**Non-response errors**)
 - **Group B:** Errors resulting in the stage of data collection or taking observation (**Response error**).
 - **Group C:** Errors resulting from data processing (**Tabulation errors**)

Sources of Non-sampling error

- Group A: Due to faulty sampling frame, biased method of selection units, inadequate schedule.
 - Ommision of duplication of units due to ambiguous definition of locale, units.
 - Inaccurate methods of interview and schedules
 - Difficulties arising due to unawarness on the part of respondents or faulty methods of enumeration/ data collection.

Sources of Non-sampling error

- Group B: These errors refer, the difference between the individual true value and the corresponding sample value irrespective of the reasons for discrepancy.
- Landholder says 10 factors
- Cadastral says 11 factors
- Response error occur...
- Main sources of these errors
 - Inadequate supervision and inspection of field staff
 - Inadequate trained and experienced field staff
 - Problems involved in data collection and other type of errors on the part of respondents.

Sources of Non-sampling error

- Group C: These errors can be assigned to a number of defective methods of editing, coding punching, tabulation, etc.
- Main sources
 - Inadequate scrutiny of basic data
 - Errors in data processing operations such as editing, coding, punching, listing, verification etc
 - Others errors committed or admitted during publication/ presentation of results.

Sources of Non-sampling error

- These are not exhaustive, appear may large number of errors in a survey
- It is difficult to ensure which of these errors are admitted, what is the frequency of their occurrence, and what are their effects on results?
- Need subtle investigation at every step of survey.

Biases and variables errors

- Relation between **biases** and **variables errors**

$$E(t - \theta)^2 = V(t) + (B(t))^2$$

- Total error can be written as

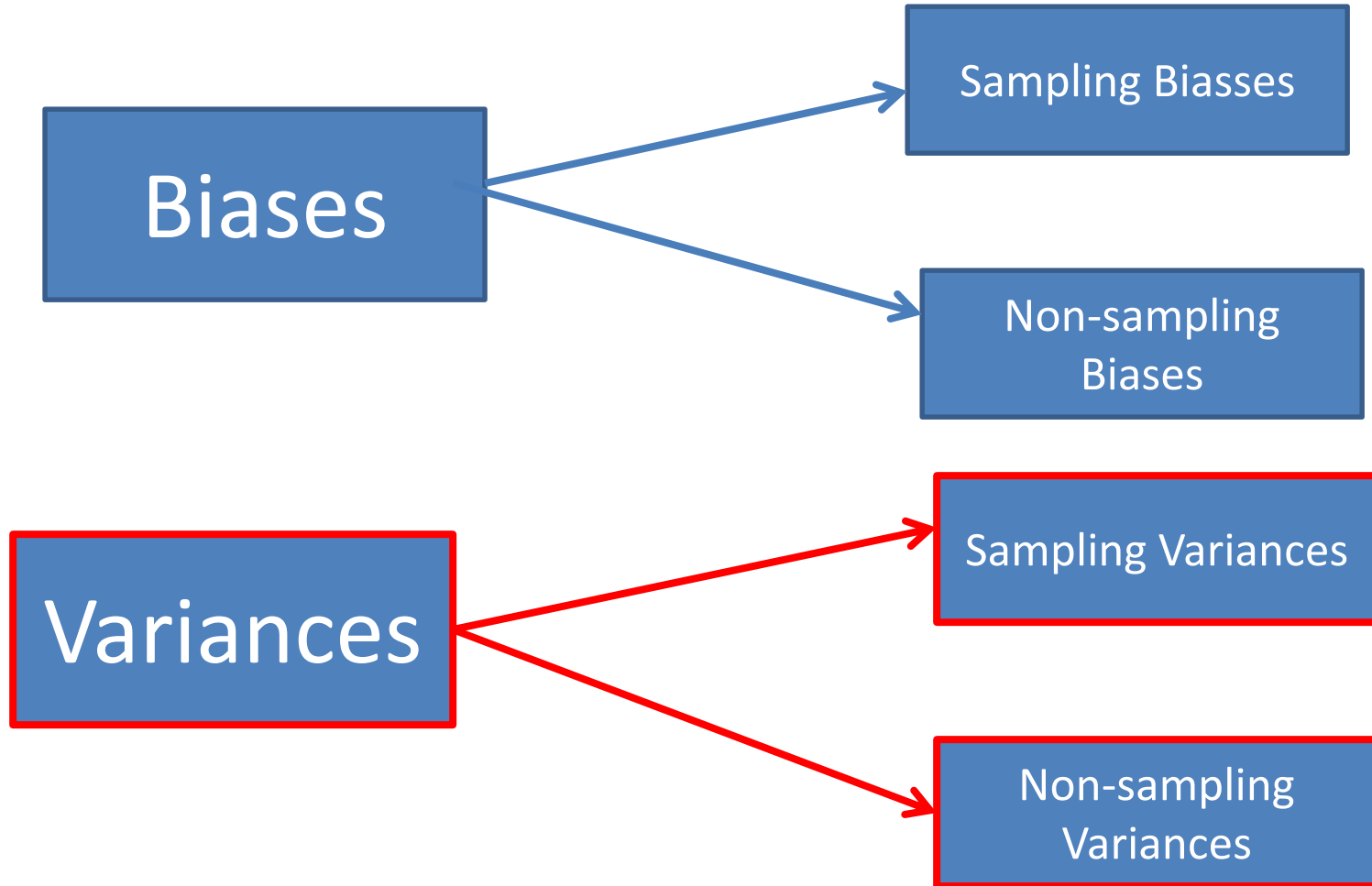
$$\textit{Total error} = [V(t) + (B(t))^2]^{\frac{1}{2}}$$

- It has been seen that sample values are subject to both sampling and non-sampling error.

Biases and variables errors

- θ : true value
- θ_p : Expected survey value
- θ' : estimated parametric value
- Thus the total error can be written as
$$(t - \theta) = [t - E(t)] + [E(t) - \theta_p] + [\theta_p - \theta'] + [\theta' - \theta]$$
 - Sampling error
 - Non-sampling error
- Both variable errors and biases can arise either from sampling or non sampling operations

Biases and variables errors



Biases and variables errors

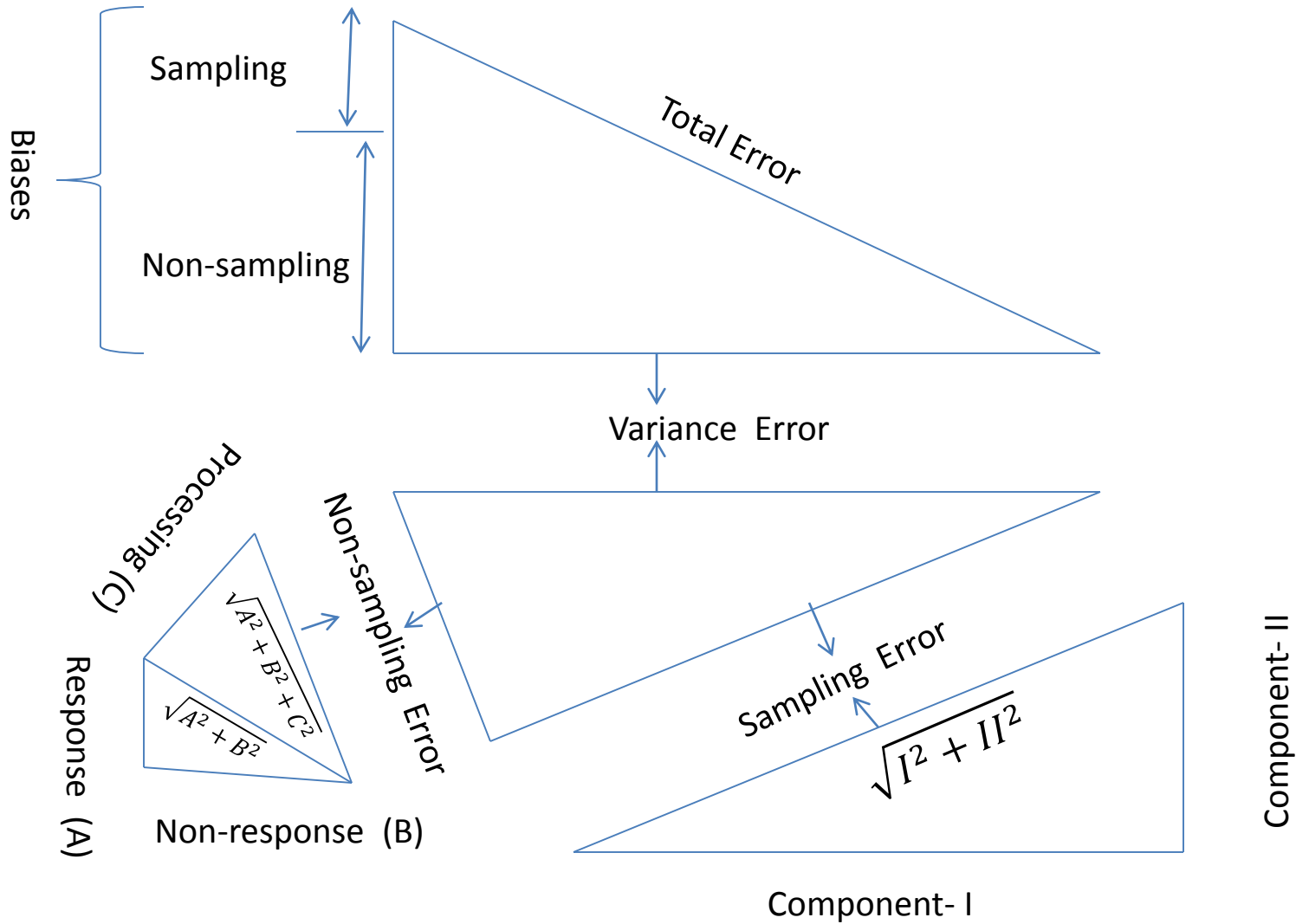
- Generalization form of biases and variables errors

$$E(t - \theta)^2 = \left(\sum_i B_i \right)^2 + \sum_i S_i^2 / n_i a_j$$

Where

- B_i stands for the bias
- S_i stands for varince
- n_i stands for sample size
- a_j Stands for the term for the sampling design used in the survey

Error decomposition



Non-response errors

- Non response errors arises due to various causes
 - Not-at-home
 - Refusal
 - Lost schedule

Adjustment for Non-response

- Population N size divided into two parts
 - N_1 (Response class) and
 - N_2 (Non-response class)
- Response class mean = \bar{Y}_1
- Non-response class mean = \bar{Y}_2
- Population mean $\bar{Y} = \frac{N_1\bar{Y}_1 + N_2\bar{Y}_2}{N} = W_1\bar{Y}_1 + W_2\bar{Y}_2$
- $W_1 + W_2 = 1$

Adjustment for Non-response

- Response class sample size = n_1
- Non-response class sample size = n_2
- Response class sample mean = \bar{y}_1
- Non-response class sample mean = \bar{y}_2
- \bar{y}_1 is a **biased** estimator of population mean = \bar{Y}
- Amount of bias

$$B(\bar{y}_1) = E(\bar{y}_1) - \bar{Y} = \bar{Y}_1 - \bar{Y} = W_2(\bar{Y}_1 - \bar{Y}_2)$$

Adjustment for Non-response

- \bar{y}_2' mean of subsample of size n_2' ($n_2 > n_2'$)
- Since the population mean \bar{Y} is expressed in terms of unknown parameters N_1, N_2, \bar{Y}_1 and \bar{Y}_2 .
- Find unbiased estimator of N_1 and N_2
- $\widehat{N}_1 = \frac{n_1}{n} N$ and $\widehat{N}_2 = \frac{n_2}{n} N$

Adjustment for Non-response

- Hansen and Hurwitz technique
 - Take a random sample, wor , of n respondent and mail a schedule to all of them.
 - When the dead line of reply calculate non response
 - Select a subsample n_2' from n_2 and collect information from personal interview.
 - Pool the results from both the classes to estimate the population values.

Adjustment for Non-response

- Pooled estimator of the population mean \bar{Y}_2 is

$$\bar{y}_w = \frac{1}{n} (n_1 \bar{y}_1 + n_2 \bar{y}_2')$$

and variance is

$$V(\bar{y}_w) = (1 - f) \frac{S^2}{n} + \frac{1 - k}{n} W_2 S^2_2$$

Where $k = \frac{n_2}{n_2'}$, S^2 is as usual S^2_2 is the mean square in the non-response class.

(Theorem 13.7.1)

Adjustment for Non-response

- Proof:

$$\begin{aligned} E(\bar{y}_w) &= E_1 E_2 (\bar{y}_w | n_1, n_2) \\ &= E_1 E_2 \left(\frac{n_1 \bar{y}_1}{n} \mid n_1 \right) + E_1 E_2 \left(\frac{n_2 \bar{y}_2'}{n} \mid n_2 \right) \end{aligned}$$

Again

$$E_1 E_2 \left(\frac{n_1 \bar{y}_1}{n} \mid n_1 \right) = \frac{N_1}{N} \bar{Y}_1$$

And

$$E_1 E_2 \left(\frac{n_2 \bar{y}_2'}{n} \mid n_2 \right) = \frac{N_2}{N} \bar{Y}_2$$

Thus

$$E(\bar{y}_w) = \bar{Y} \text{ (unbiased)}$$

Adjustment for Non-response

- Variance:

$$\begin{aligned}V(\bar{y}_w) &= V_1 E_2(\bar{y}_w) + E_1 V_2(\bar{y}_w) \\ &= V_1(\bar{y}) + E_1[V_2(\bar{y}_w | n_1, n_2)]\end{aligned}$$

- Here $V_1(\bar{y}) = (1 - f) \frac{S^2}{n}$
- And $E_1[V_2(\bar{y}_w | n_1, n_2)] = \frac{1-k}{n} W_2 S^2_2$
- Combine them

(Proved)

Adjustment for Non-response

- Politz- Simmons' technique.
 - Study yourself

Important surveys in Bangladesh

- First census 1872
- 2nd census 1881
- Since then regular interval of 10 years
- Occasional sample surveys are also being conducted alongside to supplement the short-time needs between censuses.

Imporant Surveys in Bangladesh

- Surveys are divided in four broad categories
 - Contraceptive prevalence survey (CPS)
 - Demographic survey
 - Demographic and health survey (DHS)
 - Nutrition and health survey including goiter and (iodine deficiency disorder) IDD prevalence survey

Contraceptive prevalence survey (CPS)

- Beginning 1979 funded USAID as apart of USAID's global CPS project
- Agreement BD gov and World Health System
- Designed to provide rapid feedback to improve family planning programe performance by collecting information on contraceptive use.
- CPSs became an important management tool for monitoring levels and trends of family planning program performance for quite a long time in Bangladesh.
- 3rd Stage Cluster sampling and PPS

Demographic survey

- Demographic data at the national level were lacking both in quality and coverage in Bangladesh.
- Absence of demographer
- A number of demographic survey were conducted

Demographic survey

- A number of demographic survey were conducted
 - Demographic Survey in East pakistan:1961-1962
 - Population Growth Estimation(PGE):1962-65
 - National Impact Survey (NIS): 1968-69
 - Population Growth Survey(PGS):1968-70
 - Bangladesh Retrospective Survey of Fertility and Mortality (BRSFM):1974
 - Bangladesh Fertility Survey (BFS):1975 and 1989

Demographic and health survey (DHS)

- Demographic and health survey (DHS) is the first of this kind in Bangladesh.
- It is part of the worldwide Demographic and Health Survey(DHS).
- Designed to collect data on fertility, mortality, family planning and maternal and child health.
- Conducted by Mitra and Associates under the authority of the National Institute Research and Training, Under the Ministry of Health and Family Welfare, Government of Bangladesh.

Demographic and health survey (DHS)

- The BDHS: 1993-94
- The BDHS: 1996-97
- The BDHS: 1999-2000
- The BDHS: 2004
- The BDHS: 2007
- The BDHS: 2011
- The BDHS: 2014

Demographic and health survey (DHS)

Nutrition and health survey

- East Pakistan Nutrition Survey:1962-64
- Nutrition Survey of Rural Bangladesh:1975-1976
- National Goiter Prevalence Survey:1981-82
- Bangladesh National Nutrition Survey: 1995-96
- National Iodine Deficiency Disorder Survey
 - 1993
 - 1999

Bangladesh Child Nutrition Survey

- The BBS (Bangladesh Bureau of Statistics) has been periodically conducting Child Nutrition Survey (CNS) as part of its regular activities since 1985
- Funding UNICEF (United Nations International Children's Emergency Fund)
- First conducted 1985-86
- 1989-90, 1995-96, 2000
- Objective: Nutritional status of pre school children aged 6-71 month