LECTURE 4- PROBLEMS/LIMITATIONS/DISADVANTAGES OF SURVEYS (ERRORS IN SURVEYS)

- The **public may not cooperate (legality)** as well in a sample survey as in a national census, which receives a great deal of publicity with attendant patriotic appeal

- **Among disadvantages** of surveys are **sampling** and **nonsampling errors**

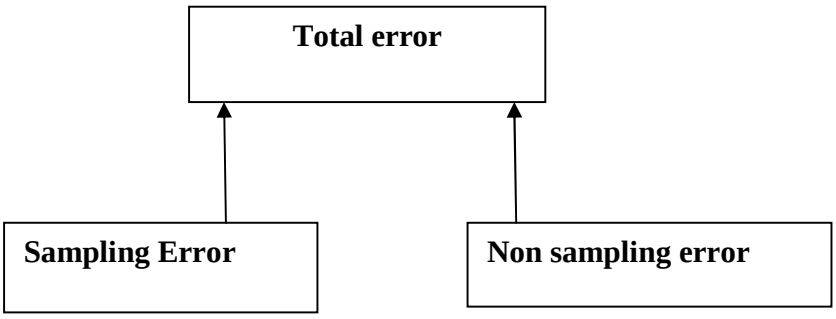
- All **survey data** are **subject to error** from various sources

- The **quality of a sample estimator** of a **population parameter** is a **function of total survey error,** comprising both **sampling** and **non-sampling errors**

- **Total survey error** is equal to **sampling error** plus **non sampling error**

- **Sampling error**- is as **a result of selecting a sample instead of canvassing the whole population**

- **Non-sampling errors-** is mainly due to **adopting wrong procedures** in the system of data collection and/or processing



SAMPLING ERRORS

- **Defects (biases)** in the **sample design**

- **How will the sample be determined?**

- Decide on whether **the results will be generalized** to the population

- Make a **choice between** the use of **probability** and **non – probability** sampling

- This is **possible if probability methods are used** in selecting the sample

- **Good survey practice includes calculation and/or estimation of errors**

- That is; **margin of error** and/or **amount of variability**

THE MARGIN OF ERROR

- This is a **measure of the difference** between the **sample estimate** and the **actual population value**

- The **better the design**, the **less is the margin of error** and the **more precise the estimates.**
- **Design=(** stratification, clustering, and proportionate sampling e.t.c)

- **E.g. (**OPINION POLL) **VOTES THAT THE CANDIDATES FOR A PARTICULAR OFFICE ARE EXPECTED TO GET**, THE RESULTS COULD BE STATED THAT CANDIDATE A'S **VOTES ARE ESTIMATED AT 57 PERCENT WITH THE ERROR UNLIKELY TO BE MORE THAN 3 PERCENT**, SO THAT THIS CANDIDATE'S VOTES ARE EXPECTED TO FALL IN THE RANGE OF 54-60 PERCENT.

NON SAMPLING ERRORS

1. **SOURCES OF NON SAMPLING ERRORS**

In general, ***nonsampling* 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 inaccurate 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**.

2. COMPONENTS/TYPES OF NON SAMPLING ERRORS

- Brieumer and Lyberg (2003) identify five components of *nonsampling error*, namely **specification**, **frame**, **nonresponse**, **measurement** and **processing error**. We may add that **estimation error** is another error, which should be considered

SPECIFICATION ERROR

- This 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 does 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.

COVERAGE OR FRAME ERROR

- **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 **results in zero probability** of selection for those units omitted from the frame.

- On the other hand if some **units are duplicated**, this results in **overcoverage** with such units having larger probabilities of selection.

- Errors associated with the frame may, therefore, result in both ***overcoverage*** and ***undercoverage*****.**

- The **latter is the most common in large-scale surveys in most African countries**.

- ***Noncoverage*** 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

- **Noncoverage is often associated with problems of incomplete frames**.

- Thus **noncoverage** refers to the **negative errors** resulting from failure to include elements that would, under normal circumstances, belong in the sample.

- **Positive errors** of **overcoverage** also occur due to inclusion in the sample of elements that do not belong there.

- **Noncoverage** errors **differ** from **nonresponse**.

NONRESPONSE

- **Nonresponse** refers to the **failure to measure some of the sample units**. **Thus failure** to **obtain observations** on some units selected for the sample.

- 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.**

- In summary **the types of nonrespondents** include:

a. **Not-at-homes**: prospective respondents who may not be at home when enumerators visit their households.
b. **Refusals**: respondents who refuse to give information for whatever reasons.
c. Not identifiable respondents

MEASUREMENT ERROR

- 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 **content** of the survey such as **definition of survey objectives**, their **transformation into usable questions,** and the **obtaining** of responses.

- These errors concern the accuracy of measurement at the level of individual units.

1. 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 interviewers **leading** to **inaccuracies** in the collected data

2. **Inadequate instructions to field staff** are another source of error. For some surveys instructions are **vague** and **unclear** leaving interviewers to use their **own judgment** in carrying out fieldwork.
3. The interviewers themselves can be a source of error. At times the **information collected** on a given item for all units **may be wrong**; this is mainly due to **inadequate training of field workers.**

4. **Age reporting** in **Africa** is a **common measurement problem** through **age heaping** and **digital preference**. These and other examples of measurement error may be **attributable to respondents** or **interviewers** or both. At times there may be **interaction between the two**, which may contribute to **inflating such errors**.

Respondents can introduce errors because of the following reasons:

− Failure to understand the survey question(s)

− **Careless** and **incorrect answers** due to, for example, **lack of adequate understanding** of the objective(s) of the survey; in particular, the respondent may not give sufficient time to think over the questions.

− **Desire to "cooperate"** by 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 stigmatized diseases.

− **Memory lapses** if there is a long reference period, a case in point being the collection of information of non-durable commodities in expenditure surveys.

PROCESSING ERRORS

Processing errors comprise:

- **Editing** errors.
- **Coding** errors.
- **Data entry** errors.
- **Programming** errors etc

Both sampling and *nonsampling errors* need to be controlled and reduced to a level at which their presence does not defeat or eliminate the usefulness of the final sample results

END