

# **BA (H) GEOGRAPHY**

## **SEMESTER IV**

### **SEC: RESEARCH METHODS**

**Unit 2:** – Data Collection: Type and Sources of Data; Methods of Collection; **Input and Editing**

#### **Introduction**

Data editing is defined as the process involving the review and adjustment of collected survey data. The purpose is to control the quality of the collected data. This process is divided into four (4) major sub-process areas. These areas are: -

- 1) Survey Management – The survey management functions are: a) completeness checking, and b) quality control including audit trails and the gathering of cost data.
- 2) Data Capture – Data capture is the conversion of data to electronic media. The data may be key entered in either a heads down or heads up mode.
  - i) Heads down data entry refers to data entry with no error detection occurring at the time of entry.
  - ii) Heads up data entry refers to data entry with a review at time of entry.
- 3) Data Review – Data review consists of both error detection and data analysis.
  - i) Manual data review may occur prior to data entry.
  - ii) Interactive data review involves immediate review of the questionnaire after adjustments are made.
  - iii) Data review (error detection) may occur at many levels.
    - a) Item level - Validations at this level are generally named "range checking". Since items are validated based on a range.
    - b) Questionnaire level - This level involves across item checking within a questionnaire

- c) Hierarchical - This level involves checking items in related sub-questionnaires. Data relationships of this type are known as "hierarchical data" and include situations such as questions about an individual within a household.
- iv) Across Questionnaire level edits involve calculating valid ranges for each item from the survey data distributions or from historic data for use in outlier detection.  
questionnaire level edits are generally grouped into two types: statistical edits and macro edits.

Statistical Edits use the distributions of the data to detect possible errors. These procedures use current data from many or all questionnaires or historic data of the statistical unit to generate feasible limits for the current survey data.

Macro Edits are a review of the data at an aggregate level. Inconsistencies are traced to the individual records involved

#### 4) Data Adjustment (Data Editing and Imputation)

- a) Manual data adjustment is when the selection of a more reasonable value is done by a person. It may involve writing down, for key entry, the adjustments to be posted to the survey data file using a batch procedure.
- b) Automated data adjustments occur as a result of computer actions. A desirable option in any system allowing computer actions is to allow for the overriding of those actions at some level. Automated imputations generally fall into one of five categories. a. Deterministic - where only one correct value exists, as in the missing sum at the bottom of a column of numbers. A value is thus determined from other values on the same questionnaire. b. Model based - use of averages, medians, regression equations,

etc. to impute a value. c. Deck - A donor questionnaire is used to supply the missing value.

- c) Mixed - In most systems there is usually a mixture of categories used in some fixed ranked fashion for all items
- d) Expert Systems - Expert systems are only recently being applied to data editing and much research is beginning in this area

## Editing methods

---

### Selective editing

*Macro editing: involves two methods: Aggregation method and Distribution method*

### Automatic editing