

Data Processing

After collection, the data has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis.

Data processing is the conversion of data into usable and desired form. This conversion or “processing” is carried out using a predefined sequence of operations either manually or automatically. Most of the processing is done by using computers and thus done automatically. The output or “processed” data can be obtained in various forms. Example of these forms include image, graph, table, vector file, audio, charts or any other desired format. The form obtained depends on the software or method of data processing used. When done itself it is referred to as automatic data processing.

Data processing is done in three distinct stages:

1. Data editing
2. Coding of data
3. Tabulation of data

Data Editing

The data which has been collected from various primary/secondary sources is RAW in nature; this means that there are likely chances of errors and inconsistencies in it.

Since data collected is of pivotal importance to policy and decision makers everywhere be it governmental departments, business organizations, health or educational institutions etc it would be better to have a team of experts at hand who know how to scrutinize, review and **edit** this data before it is finally fed into the data bases and the required statistics are generated.

The process through which the data is reviewed to check for consistency, adequacy, detect errors and outliers (values that are either too big or too small from the rest of the data) and the correction of errors within the data in order to maximize its usefulness for the purpose for which it was collected is called **data editing**.

Purposes and objectives of Data Editing

The basic purpose served by data editing is that it improves the quality, accuracy and adequacy of the collected data thereby making it more suitable for the purpose for which the data was

collected. The following can therefore be identified as the main objectives of the data editing process:

- Detection of errors in the data that otherwise affect the validity of outputs.
- Validation of data for the purposes it was collected.
- Provision of information that that would help assess the overall level of accuracy of the data.
- Detection and identification of any inconsistencies in the data and outliers and to make adjustments for them.

Electronic Editing

In recent years, most of the researchers prefer to submit electronic questionnaires wherever it is possible. Electronically sent questionnaires are easy to edit, because in the electronic questionnaire you can set some parameters. The computer can edit the questionnaire itself and the job of the editor becomes easy. You can avoid inconsistencies in the electronic questionnaire. The logical errors can be completely avoided. No response answers are few in electronic questionnaires.

Types of Data Editing

There are different types of data editing:

- **Validity and completeness of data:** refers to correctness and completeness of obtained responses. This helps ensure that there are no missing values or empty fields in the data bases.
- **Range:** verifies that data within a field fall between the boundaries specified for the particular field.
- **Duplicate data entry:** this helps ensure that there is no repetition or duplication of data and each unit on the data base or register was filled only once.
- **Logical consistency:** through this type of editing connections between data fields or variables are taken into account.
- **Outliers:** this type of editing helps detect values that are too extreme or unusual so that they can be verified and checked.

The following practical **guidelines** may be handy while editing the data:

- The editor should have a copy of the instructions given to the interviewers.
- The editor should not destroy or erase the original entry. Original entry should be crossed out in such a manner that they are still legible.
- All answers, which are modified or filled in afresh by the editor, have to be indicated.
- All completed schedules should have the signature of the editor and the date.

Editing is done in two stages:

1. **Field Editing:** It is done immediately after the instrument has been administered. It removes any glaring errors and omissions in the questionnaire form.
2. **Central Editing:** It should be carried out when all the forms of schedules have been completed and returned to the headquarters.

Coding

Coding is the process by which items or groups of data are assigned codes. These may be used to simplify and standardize the data for analytical purpose, as when characteristics like sex, marital status or occupation are replaced by numbers (e.g. replacing 'male' by '1' and 'female' by '2'). It may involve some reduction in the quantity of the data, as when ages, locations or attitudes are categorized into a limited number of groups, with each group then assigned its own numerical identity (e.g. categorizing ages as "under 21", '21-64' and '65 and over', and then replacing these by '1', '2' and '3' respectively).

Data coding is the process of deriving codes from the observed data. In qualitative research the data is either obtained from observations, interviews or from questionnaires. The purpose of data coding is to bring out the essence and meaning of the data that respondents have provided. The data coder extract preliminary codes from the observed data, the preliminary codes are further filtered and refined to obtain more accurate precise and concise codes. Later, in the evaluation of data the researcher assigns values, percentages or other numerical quantities to these codes to draw inferences. It should be kept in mind that the purpose of data coding is not to just to eliminate excessive data but to summarize it meaningfully. The data coder should ascertain that none of the important points of the data have been lost in data coding.

Data coding in research methodology is a preliminary step to analyzing data. The data that is obtained from surveys, experiments or secondary sources are in raw form. This data needs to be refined and organized to evaluate and draw conclusions. Data coding is not an easy job and the person or persons involved in data coding must have knowledge and experience of it.

The following are the broad **general rules for coding**:

- Each respondent should be given a code number (an identification number).
- Each qualitative question should have codes. Quantitative variables may or may not be coded depending on the purpose. Monthly income should not be coded if one of the objectives is to compute average monthly income. But if it is used as a classificatory variable it may be coded to indicate poor, middle or upper income group.
- All responses including “don’t know”, “no opinion”, “no response”, etc. are to be coded.

Code Book

A code book or coding scheme contains each variable in the study and specifies the application of coding rules to the variable. The code book guides the researcher in assigning numerical codes to different response categories. The code book serves the following purpose:

- It helps in identifying and locating different variables.
- Data entry is done more effectively and is less prone to errors.
- A code book generally contains information on (a) Question number (b) Variable number (c) Column number (d) Code description and (e) Variable name.