# *Data Processing*

Before we can capture the information collected, it is necessary to have prepared the description of the database that will contain the gathered information. So there can be an "empty" database prior to data entry. The database is used to define the input window or the relationship between the input mechanisms of data and files to be created. Following the data entry operations, several activities usually take place in order to prepare the databases that will be used to perform the required compilations. EduStat software can be used to perform data processing in relation to steps usually retained; this is the following:

- the data entry by batch;
- the import of batch files in the database;
- the quality control;
- the first phase of compilation;
- the merging of information from different sources;
- the creation of databases according to different formats.

## A.  Data entry by batch

Data is usually entered on computer 'by batch', that is to say a set of documents divided into several groupings of modest size to each of which corresponds an input file. It will be easier to carry out monitoring of each of these sets of forms throughout the operations that lead to the formation of the necessary databases to perform the required compilations. Once a batch has been codified and the rating was performed, it is possible to entrust the entry to the operators concerned.

## B.  Importation of batches in the database

As soon as batch files are available, it is possible and desirable to incorporate the data in the database. The gradual establishment of the database will perform quality control on a daily basis. If problems are detected (e.g., presence of non-predicted values, format does not expect value), it will be easier to resolve the problem immediately. First compilations (e.g., frequency distributions) may be performed; these transactions may provide preliminary information on the collected data and adjust the most appropriate value labels to reality.

## C.  Quality control

Some activities may be undertaken to monitor the quality of entered data and included in the database. In addition to periodically produce frequency distributions to detect possible abnormalities, it may be useful to perform compilations that reflect the rigorousness of coding and marking operations. We can mention here the mechanism for comparing the values of the database in preparation with a base formed of a sample of records inputted by another data entry team.

## D.  First phase of compilation

After incorporating all the scheduled batch files, it is possible to make some preliminary compilations affecting variables that will be used to group the records for analysis. For example, background information of students, teachers and other respondents should be available to perform certain treatments. If there is no information or information not intended for the variables "sensitive", it will be appropriate to consult the collected tests or questionnaires and make corrections if this proves to be possible. It is sometimes possible to make connections between certain information to check the consistency of the provided information.

The implementation of various information processing procedures will finalize the database. It may be necessary to code certain variables to make information more consistent and easier to deal with when we draw comparisons with other data sets.

## E.  Merging information from different sources

Information gathered as part of an evaluation activity usually come from several sources. One can, for example, administer tests to students, asking them to answer a complementary questionnaire probing their teachers or other staff. We may also collect contextual information from different sources. At the time of data analysis, it will be desirable to merge information from these different sources. It will therefore be "twinning" of information. For this, it will be provided with identification numbers to establish the necessary links.

## F.  Creating databases according to different formats

Finally, the collected data is usually treated using various software. It is rare that one software is able to perform all the desired compilations. So we must be able to produce compatible file formats for software used to process all scheduled compilations.