**EduStat Software**

# Sample selection

This document deals with sampling. The database "**SampleEN.edu**" will be used to carry out the proposed exercises. It focuses on sampling (Note that another training module provides for the presentation of sampling more fully). This database contains 25 variables and 450 records. Note that the records included in this database represent primary schools. In addition to dealing with sample selection, it is also possible to produce frequency distributions as is the situation for most databases.

Several documents are available when using EduStat. If you would like more information on how to select samples, the following documents can be found in the documentation. To access it, choose the "Documentation" option from the "Help" menu. A window with five options appears. It is possible to open one of the three options allowing to have information on the selection of samples by choosing the option of its choice: Slideshows, Videos (note: not available in English), Procedural documents. It is also possible to consult the "User's Guide" (note: not available in English). We advise you to read some of these before carrying out the activities presented in this document. Here are some references.

*User Guide* – Not available in English


*Viewing Slide Shows*

      Sampling: "Sampling Design and Selection"


*Viewing Slide Shows* – Not available in English
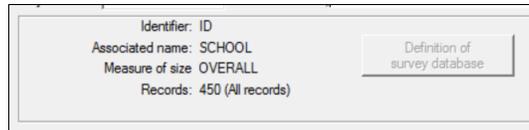

*Viewing Procedural Documents*

      Sampling: "Design and Selection"

The following four sections present information on sampling

- *Definition of the sampling design*

Three variables must be identified in order to be able to define the sampling design. It is also necessary to indicate if we wish to keep all records belonging to the active database (here 450 schools) or a selection of records.

Here is the window showing the sampling plan information. We find that the identification variable is "ID" which is associated with the variable "SCHOOL". The size measure (here the number of students per school) is "OVERALL". Finally, it is recalled that all the schools on the base were retained.



- *Determining the target population*

It was decided that schools with fewer than 50 students will be excluded. The following table indicates that 38 schools will be a priori excluded involving the non-participation of 1418 pupils if the pupils will be chosen later. This is the target population.



- *Drawing*

  Size: 50 schools

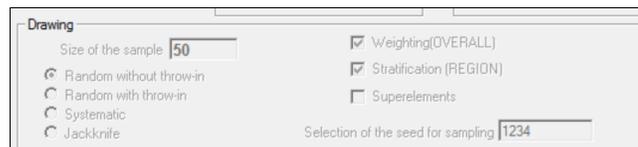  Type: random without throw-in

  Weigh: in relation to the number of schools

  Stratification: in relation to the regions

  Drawing root: 1234

  Number of groups: 1

The following table provides the parameters selected for the desired drawing.

- *Viewing operations*

After selecting the "Draw" button, it is possible to produce different types of report. At any time, a report indicating the choices made is accessible ("See" button).

```
                      File : C:\DemoComplet_Fra\BasesDonnees\SampleEN.edu  (450 rec.)
                 Identifier: ID
            Measure of size: OVERALL
                       Name: SCHOOL
                  Selection: All records
                       Draw: Random without throw-in
                             Weighted  (OVERALL)
                             Stratified (REGION)
                             Seed: 1234

wo-Step Sampling Design with 1 Group at the Second Step

                         Elements          Measure of size
      Sampling base:    450( 100%)         88019( 100%)
          Excluded:      38( 8,4%)          1418( 1,6%)
Target population:      412(91,6%)         86601(98,4%)
            Sample:      50(11,1%)Total    14169(16,1%)Total
                         50(11,1%)Target   14169(16,1%)Target
```

It will be possible to manage the sampling design then available. Functions will allow you to perform several tasks, including:

- Indicate the withdrawals and their replacements
- Enter the participation
- Produce reports