EduStat Software

Frequencies Distributions

The frequency procedure may compute, for one or more variables of the active EduStat database, the proportion represented by each of the observed values. Two frequency options are available: the calculation of standard frequencies and the compilation of distributed frequencies. In addition, the "Cross tabulation" procedure allows to distribute the frequency distribution for one or more variables according to the values of one or two other variables.

This document includes useful information for compiling frequencies distributions. The database "**SampleEN.edu**" will be used to perform the exercises. This database (it is also used in the presentation of the selection of a sample) includes 25 variables and 450 records. Note that it would also be possible to use the database "**StudentsEN.edu**" to produce frequency distributions. We will produce statistical compilations related to the following options:

- The compilation of standard frequencies
- The compilation of distributed frequencies
- The compilation of cross tabulations

Several documents are available when using the EduStat software. If you would like to have more information on how to proceed with frequencies distributions, 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 production of frequencies distributions by choosing the option of its choice: Slideshows, Videos, Procedural documents. It is also possible to consult the "User's Guide" (not available in English). We advise you to read some of these documents before carrying out the activities presented in this text. Here are some references.

User Guide – Not available in English

Viewing Slide Shows - "Standard frequencies"; "Cross tabulation"; "Distributed frequencies"

Viewing Videos – Not available in English

Viewing Procedural documents - "Standard frequencies"; "Cross tabulation"; "Distributed frequencies"
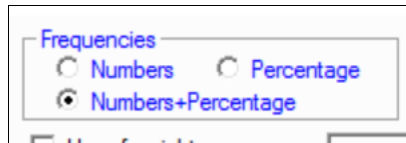
❖ *The compilation of standard frequencies*

By using this exercise, it will be possible to produce standard frequencies distributions using the following variables:

STATUS: Status of the school, that is to say: public school and private school

ZONE: Linking schools to urban and rural areas

Here is the approach to run.
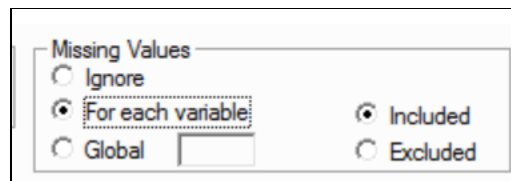
- Information provided

The following information on the frequencies produced can be obtained:
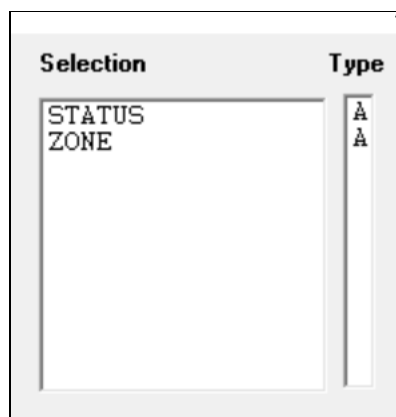
Frequencies in numbers only
Percentage frequencies only
Frequencies in numbers and percentage

- Processing missing values

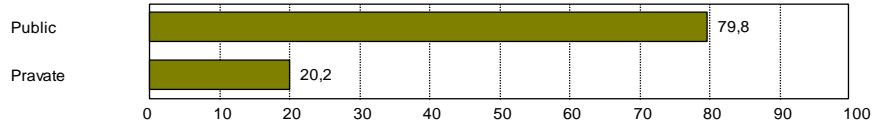- Choice of variables to be processed

Examples of reports:

School status

	N	Proportion
Public	359	79,8
Private	91	20,2
Total	450	100,0

School status

| Public | 79,8 |
| Pravate | 20,2 |

0 10 20 30 40 50 60 70 80 90 100

Area of the school

	N	Proportion
Urban	84	18,7
Rural	366	81,3
Total	450	100,0

Area of the school

| Urban | 18,7 |
| Rural | 81,3 |

0 10 20 30 40 50 60 70 80 90 100
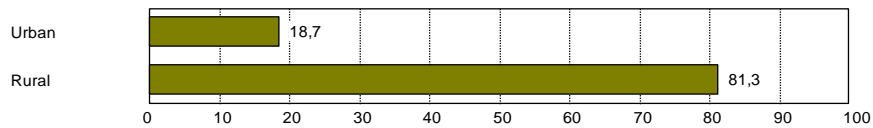
❖ *The compilation of distributed frequencies*
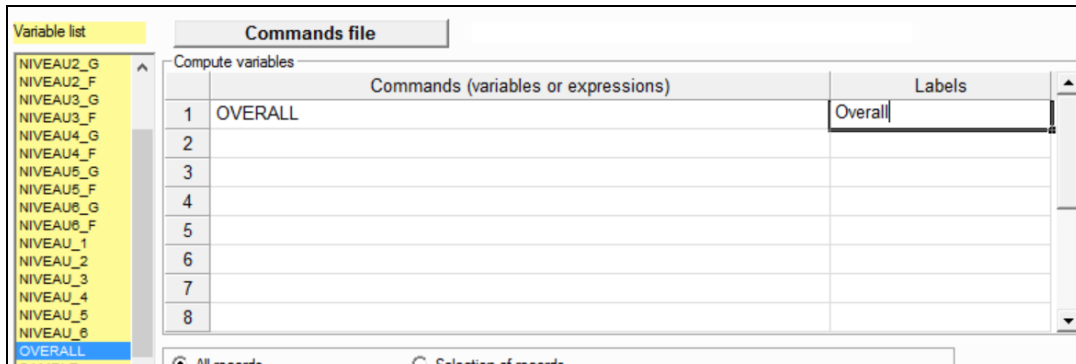
This statistical compilation option produces observed distributions according to a numeric format variable and two or three contextual variables.

Here is the approach to use.

- Selection of variables (maximum: 3)

Distribution Variables					
STATUS		Selection	Labels		
	1	REGION	Region		Cancel
	2	ZONE	Zone		
	3				
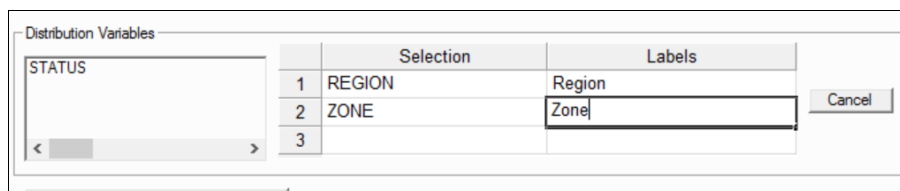
- Identification of the calculation variable

	Commands file		
	Compute variables		
	Commands (variables or expressions)		Labels
1	OVERALL		Overall
2			
3			
4			
5			
6			
7			
8			

Variable list:
NIVEAU2_G
NIVEAU2_F
NIVEAU3_G
NIVEAU3_F
NIVEAU4_G
NIVEAU4_F
NIVEAU5_G
NIVEAU5_F
NIVEAU6_G
NIVEAU6_F
NIVEAU_1
NIVEAU_2
NIVEAU_3
NIVEAU_4
NIVEAU_5
NIVEAU_6
OVERALL

Report

Region	Status	Enrollment
Region 10	Public	5435
	Private	1016
	Total	6451
Region 11	Public	4403
	Private	199
	Total	4602
Region 20	Public	3982
	Private	169
	Total	4151
Region 16	Public	5267
	Private	883
	Total	6150
Region 17	Public	10518
	Private	273
	Total	10791
Region 14	Public	3354
	Private	110
	Total	3464
Region 21	Public	3367
	Private	759
	Total	4126
Region 18	Public	6245
	Total	6245
Region 12	Public	8070
	Private	3467
	Total	11537
Region 22	Public	5008
	Private	4068
	Total	9076
Region 15	Public	6765
	Private	969
	Total	7734
Region 19	Public	8917
	Private	1139
	Total	10056
Region 13	Public	3296
	Private	340
	Total	3636
Overall		88019

This report distributes the 88,019 pupils attended by the 450 schools in the database in relation to each region and according to the status of the schools. For example, there are 3636 students in "Region 13"; there are 3296 pupils in public schools and 340 in private schools.

❖ *The compilation of cross tabulations*

This compilation option allows to establish links between two or three variables. A statistical index - chi-square - shows the statistical importance of these links.

We will compile a table using the following two variables: administrative regions and school status. Here's what to do:

Identification of the first variable - the variable "REGION"
Identification of the second variable - the variable "STATUS"
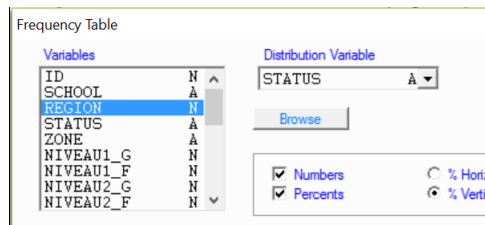
We must also indicate that we want the calculation of chi-square. Here is the report.

Report

Status of the schools

Administrative regions	Public		Private		Total	
	Number	%	Number	%	Number	%
Region 10	26	7,2	7	7,7	33	7,3
Region 11	37	10,3	4	4,4	41	9,1
Region 12	18	5,0	3	3,3	21	4,7
Region 13	43	12,0	0	0,0	43	9,6
Region 14	32	8,9	18	19,8	50	11,1
Region 15	41	11,4	12	13,2	53	11,8
Region 16	27	7,5	3	3,3	30	6,7
Region 17	24	6,7	1	1,1	25	5,6
Region 18	18	5,0	1	1,1	19	4,2
Region 19	18	5,0	7	7,7	25	5,6
Region 20	33	9,2	9	9,9	42	9,3
Region 21	16	4,5	21	23,1	37	8,2
Region 22	26	7,2	5	5,5	31	6,9
Overall	359		91		450	

Chi square: 62.636; Degrees of freedom:12; Probabilite 0.000

The chi-square value (62.636) indicates that there is a statistically significant relationship between the two variables (probability index less than 0.05). The proportion of public and private schools varies according to administrative regions.