First steps in SPSS

Statistical Package for Social Science (SPSS) is a computer program, working with the Windows operating system, and is specialized in the classification, processing and analysis of statistic data. As its name suggests, it can be used mostly for analysis in the field of sociological and marketing research. But it is also applicable in engineering, mathematics and other natural sciences.

We start SPSS program under the Windows operational system as follows: Start → Programs → SPSS for Windows. After that we see the SPSS Data Edition window, which looks the following way:
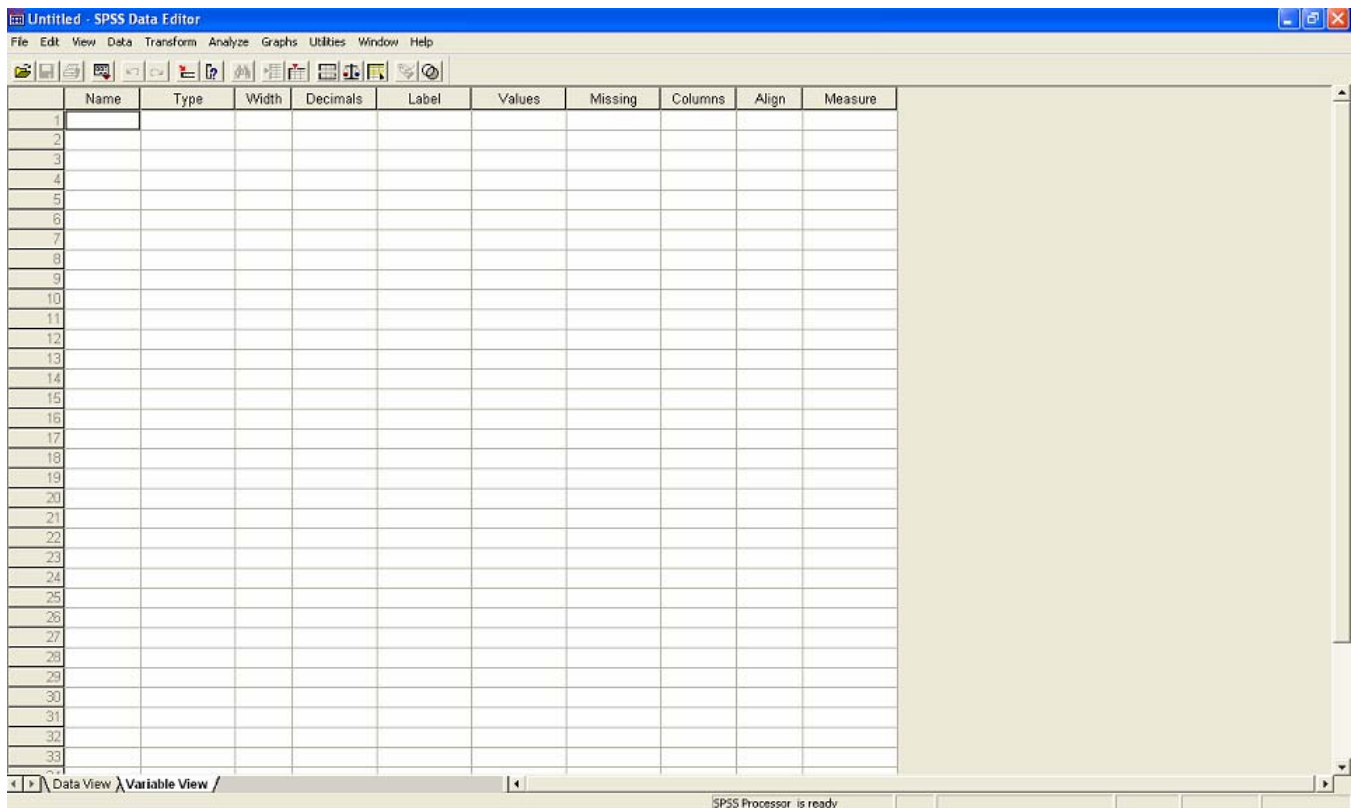

Figure 1

SPSS works with two windows: **Data Editor** – for input and editing of data and **Output Viewer** – for printing the results from the statistic processing – tables and graphs. We use the tabulators in the lower band on the screen to switch between the two windows. In the data window (SPSS Data Editor) the statistic data is defined, entered and edited. The viewer window is not displayed yet as it is automatically activated after the

fulfillment of the statistic processing. The calculated statistic indexes and graphic figures are displayed in it.

The menu of both windows is the same, but the icons are different. The menu of SPSS contains some of the basic submenus for Windows programs, as well as some special functions. In the Data Editor window this looks as follows:
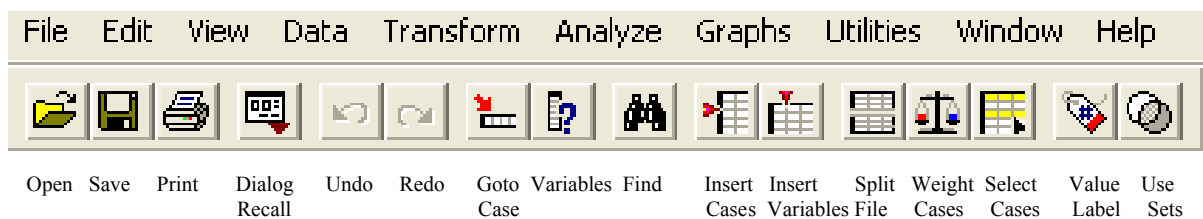

File Edit View Data Transform Analyze Graphs Utilities Window Help

Open Save Print Dialog Undo Redo Goto Variables Find Insert Insert Split Weight Select Value Use
Recall Case Cases Variables File Cases Cases Label Sets

Figure 2

Example: In Table 1 are displayed the basic indexes of UniCredit Bulbank. Each year is considered as a separate unit, and the indexes as separate quantities. The matrix of the data, which is presented in the table, has dimension 7×5.

Table 1

Basic indexes (in millions of levs) of UniCredit Bulbank for the period 2000-2006

№	Year	Net profit	Own funds	Assets	Client deposits	Credits
		X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}
1	2000	160,065	602,776	2 559,476	1 692,270	316,380
2	2001	68,912	490,479	2 731,686	2 021,634	362,353
3	2002	79,130	513,249	2 721,980	2 049,957	563,935
4	2003	89,752	550,026	2 825,439	2 177,781	916,634
5	2004	86,279	574,112	3 614,696	2 917,169	1 393,968
6	2005	96,116	609,609	3 474,829	2 618,771	1 706,858
7	2006	120,654	630,781	4 346,594	3 336,875	2 131,577

If the different variables are presented in different measures and scales, they have to be standardized. For example, when variables of different type – interval and nominal, take part in the cluster analysis, in order to use them we have to standardize them into one type – interval. Standardization in SPSS can be done automatically, by using different methods. The most common method is Z scores, in which all the variables are transferred to one range from -3 to +3 and then are changed to interval. With the Z-

method the standardized values have zero intermediate and standard deviation 1. Other methods range in the interval from -1 to 1 or from 0 to 1, as well as in standardization with regard to maximum value, intermediate quantity or standard deviation. If the variables are presented in the same measures and scales their standardization is not recommended. In the theory only the symbol X_{ij}, was used for the separate observations (the elements of matrix X), having in mind that they can refer to non- standardized or standardized variables in accordance with the reasons mentioned above.

It is accepted in statistics that variables (indications), the values of which are given in words, are called **character** (non-numeric). Most variables in inquiries are of this type, because they are questions, to which verbal answers are given. The other type of variables is numeric. Their values are given in numbers.

It is preferable to use numeric variables. It makes statistic processing and analysis easier. That is why the tendency is to artificially convert character indications into numeric. For example, the variable „gender", which is a character, is artificially converted into numeric by coding men with 1 and women with 2. In this way inquiries are prepared for statistic processing and analysis.

The input of data in SPSS is done by entering the names of the variables first. They are to be found in the first row of the table, using **Data View**, where in each cell is written **var**. After that you switch over to **Variable View**, where basically you have to write the name, the type and the label (explanation) of the variable.

UniCredit Bulbank - EN.sav - SPSS Data Editor

File   Edit   View   Data   Transform   Analyze   Graphs   Utilities   Window   Help

1 : net_profit   160,065

	net_profit	own_funds	assets	client_deposits	credits	var
1	160,065	602,776	2559,476	1692,270	316,380	
2	68,912	490,479	2731,686	2021,634	362,353	
3	79,130	513,249	2721,980	2049,957	563,935	
4	89,752	550,026	2825,439	2177,781	916,634	
5	86,279	574,112	3614,696	2917,169	1393,968	
6	96,116	609,609	3474,829	2618,771	1706,858	
7	120,654	630,781	4346,594	3336,875	2131,577	
8						
9						
10						

Figure 3

Figure 4

In **Name** you write the name of the variable in short, and it must be unique, that is to say that two variables can't have the same name. After that you select using the mouse cursor **Type** and you select the type of the variable. In our case it is **Numeric**. You can change the width of the column through **Width** and the decimal signs after the comma through **Decimal Places**. The field **Label** is used for entering the label of the variable, its extended name, and **Values** - for a label of the variable's values.

When a numeric variable is converted to character, the codes and the labels of the categorical variables are entered in the column **Values**. After you click on the grey button, in the right part of the column the dialogue box **Value Labels** appears. In the field Value you enter the code, with which the category of the indication is marked, for example 1, and in Value Label you write its corresponding label in letters - „man". Clicking on Add attributes 1="man" and then it is entered into the field. After that the same is repeated for women – in Value enter 2, and in Value Label „woman" and again Add. The procedure is repeated for all the categories that the indicator may have, after which it is confirmed with OK.
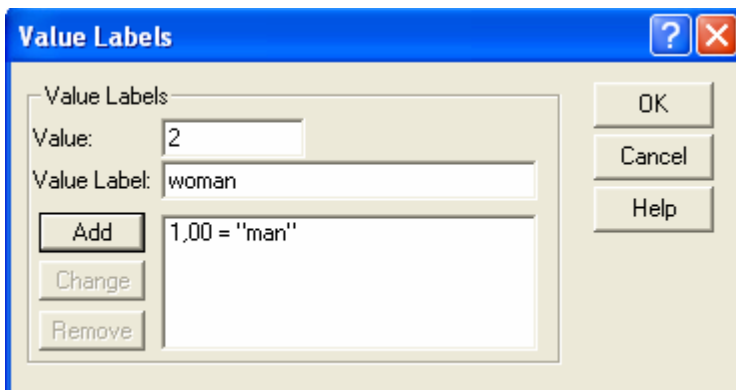

Figure 5

When you enter the values for numeric variables, in each cell write the corresponding numerical value, and when you enter values for a character variable – the corresponding code. For example, for the character indicator gender, if the first inquired person is a woman, write code 2. This can be done more easily – by activating the window labeled with from the menu. Go to the corresponding cell for the character indicator and choose an option from the falling list with possible definitions of the indicator – in this case „man'' and „woman''. The same is done for all character variables. In particular, labels are not defined for numeric variables because there are no codes.

Through **Missing Values** you can define some values as missing, include them in others and analyze them separately. You can use this for coding questions with answers such as "I don't know", which shouldn't be lost, and are not necessary for the analysis. The width of the columns can be defined from **Columns**, the alignment from **Align**, and from **Measure** you can define the scale of the variables. The numerical values are interval/scale, and those with symbols - nominal. It is possible to define some numeric variables, which are codes as nominal, and others, for which the order of the values is of importance, as ordinal. After you have defined the variable, it shows up in the first column of the table. The rest of the variables can be defined in the same way.

Author: Dessislava Vojnikova,
             Plovdiv University,
             4th year Bachelor program in Applied Mathematics
Supervisor: Snezhana Gocheva-Ilieva