

New Data Editor in Version 10.0 of SPSS

Version 10 of SPSS introduces a very important difference from older versions — a dramatic retooling of the Data Editor.

The changes to the SPSS Data Editor in this new version are significant. The “spread sheet” — the entire view used when creating an SPSS data set — is different in version 10. However, all the concepts involved in creating a data set and defining variables are still valid. The mouse clicks and the new screens you see are, I feel, an improvement.

The new Data Editor provides two views of your data (see figures 1 and 2).

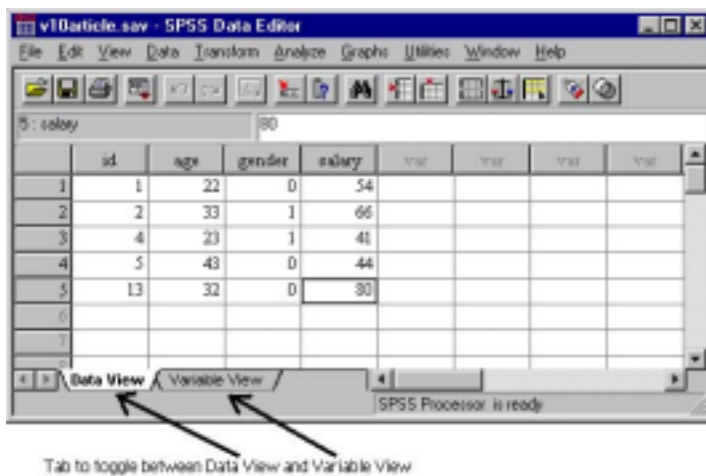
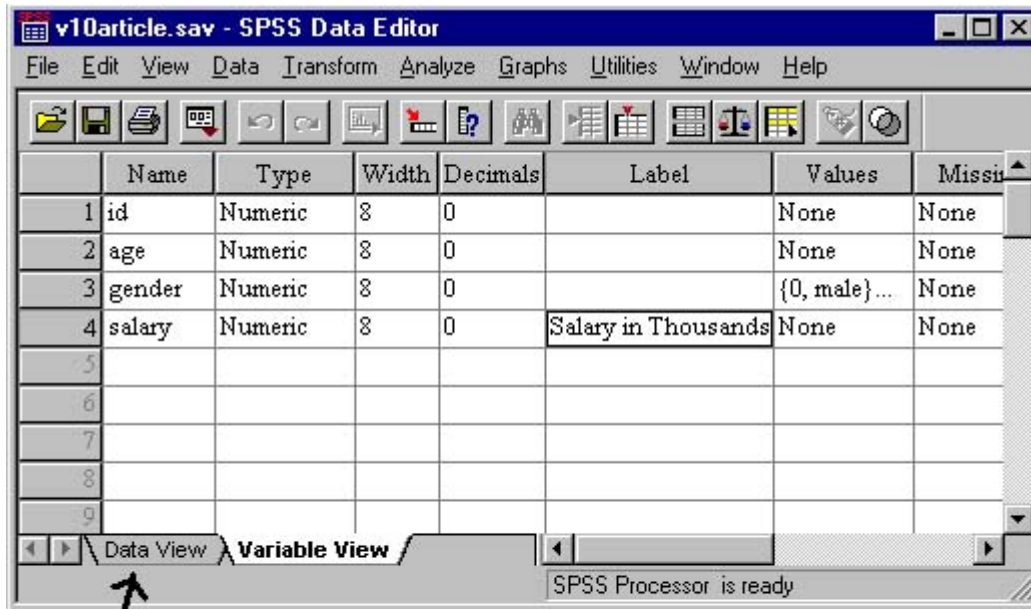


Figure 1. Data View of Data Editor

The Data View (figure 1, above) displays the data in a spreadsheet format. Each column is a different variable and each row across the spreadsheet is a case. A case is usually the unit of study in our research, such as a respondent’s answers to our questionnaire.



Click here to toggle back to Data View shown in figure 1.

Figure 2. Variable View associated with Data View from Figure 1

The Variable View (figure 2) is also a spreadsheet, but it lists our variables rather than data. Each row across names a variable, such as ID, AGE or SALARY. It then includes other information about the variable, such as each variable's attributes. In older versions, each variable's information was presented in a separate "Define Variables" window. Now, in Version 10, you see the information about all the variables at the same time.

So the old Define Variables window — which was used to allow us to define and change the variables' names and to define the four attributes used when creating an SPSS data set — is gone! Now we look at a second spreadsheet, the Variable View, to view all the information we put into defining the variables.

In the Data Editor window, you move between the Data View and the Variable View by clicking on their tabs on the lower left in either view; they are pointed out in figures 1 and 2.

Defining Variables in a New Data Set.

Each row in the Variable View spreadsheet names and defines the attributes for each variable in your data set. The columns in the Variable View spreadsheet are assigned values that determine the attributes for each variable. The first column

is used to name the variable being defined. The Name field is limited to eight characters without spaces.

After the Name column, the remaining nine columns of the Variable View are used to manage the attributes for that variable:

- Type – click within the cell and an icon appears (see figure 3).

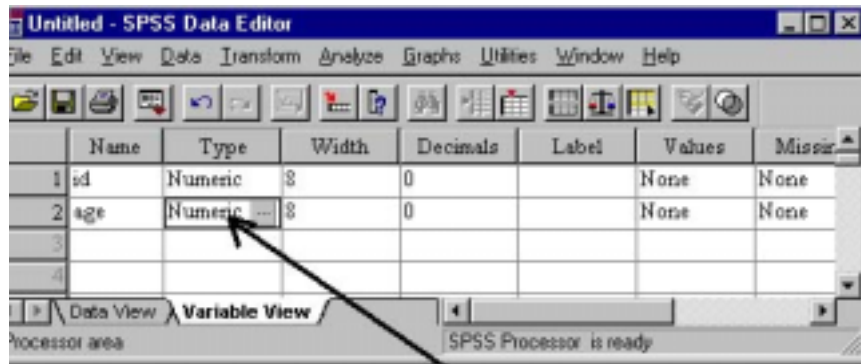


Figure 3. Managing the "Type" attribute in the Variable View

Click in the "Type" cell for age to bring up the icon shown

Click on that icon and a window opens where you select the type of variable (see figure 4) such as "numeric," "date" or "string."

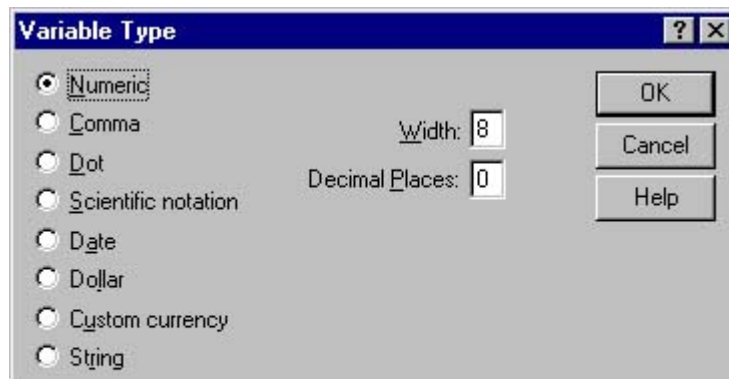


Figure 4. Action window opens up to allow "Type" to be modified. The default is Numeric.

Choose the length for the variable with the Width column. Use Decimals to change the number of significant integers for numeric variables.

- Label. SPSS only lets you use eight characters (without spaces) to name a variable. MOTHERS_INCOME is a great name, but it is too long. So you can name the variable M_INCOME, and then label it "Mother's Income" in the Label field. This label is used in statistical procedure output. For example, if you make a table involving M_INCOME, the label "Mother's

Income” is used. That variable label becomes part of the data set. When you give the data set to someone else, these labels are included. The data set becomes self-documenting.

Value labels, the other kind of label, allow you to record the meaning of categorical values of a variable. For example, you might have a variable COB (Country of Birth) where 1 means USA, 2 is Argentina, 3 is Brazil, etcetera. These values can be recorded at the time you create the data set. These labels are also used in output presentations.

- Missing Values. We hardly ever get a complete data set. Rarely is every question answered on a questionnaire. But analysts can't throw away an entire questionnaire if the person won't answer one question out of a hundred. Using codes to designate missing values allows us to deal with missing data. This is a major topic in statistical analysis and is outside the scope of this short article.
- Columns and Align control the display properties of a variable, such as the width of the column in the spreadsheet display.

Example: Creating a Sample Data Set

To introduce you to variable definition, we'll walk through creating a simple data set using information on age, gender and salary.

Define a variable named ID, which will serve as a unique identifier or key for our cases.

Be sure you are on the Variable View within the Data Editor window. Enter **id** in the first Name field. The default description usually describes a numeric variable and so is probably suitable for the ID variable. Be sure there are no decimals in the definition. We are finished defining the ID variable.

Define a variable named Age.

Enter **age** in the second row for the variable Name field. Again, the default description is probably okay.

Define a variable named Gender as a numeric variable with value labels.

Enter **gender** in the third row Name field for the third variable.

The variable name Gender is sufficiently descriptive, so we will not add a variable Label. But we will add value labels.

Single click on the cell for the Values column. An icon appears in the cell (see figure 5). Single click on that icon and a Values Label window appears (see figure 6). Enter **0** in the Value field and **Male** in the Value Label field. Click the Add button. The phrase 0="Male" appears in the lower section of the dialog box. Now, enter **1** in the Value field and **Female** in the Value Label field. Click on the Add button again.

We are using the values 0 and 1 to represent the two genders, but the choice of numbers is arbitrary. Numeric values are often assigned to a categorical variable when a questionnaire is designed. When that is the case, use the reassigned numbers to define value labels.

Click on the OK button. The Define Labels dialog box closes.

NOTE: Do not to click the OK button until all of your labels appear in the lower section of the dialog box. If you do not click on the Add button after entering a Value and a Label, the information will be lost.

Define a variable named Salary with a variable label "Salary in Thousands."

Enter **salary** in the fourth variable Name field.

In the Labels cell, type in **Salary in Thousands**.

The fourth column now contains the labeled salary variable.

All of the variables in our sample data set are now defined. Save the data set and you are ready to enter data. Enter the data into the Data View spreadsheet one cell at a time. Again, toggle between the Data View and the Variable View using the tabs at the lower left of the SPSS Data Editor window.

Other functionality, outside the scope of this article, will quickly unfold to the experienced Windows user. For example, as you experiment with cutting and pasting these definitions within the Variable View, you will see how to create other variables with the same data definitions from one variable definition. We could always cut and paste within the data spreadsheet; now it is possible in the Variable View spreadsheet as well.

To conclude, the new Data Editor for Version 10 offers much to the SPSS programmer. The Variable View allows us to see and manipulate data definitions for the entire data set in one window. Older versions only allowed us to inspect

one variable definition at a time. For more complete information on the new Data Editor, see the *SPSS Base 10.0 User's Guide*, published by SPSS, Inc.

This article is a partial update to the NYU/ITS *SPSS for Microsoft Windows*, V.9.0 document (12/99: <http://www.nyu.edu/its/socsci/Docs/SPSSwin9.pdf>). To this end, users of V.10 should refer to these pages for information about the Data Editor. This article replaces pages 3 through 7 of the aforementioned document beginning with the section titled *Part 2. SPSS Basics*, section *Creating a New Data Set*, subsection *Defining Variables*.

Otherwise, the SPSS for Windows Version 9 document, <http://www.nyu.edu/its/socsci/Docs/SPSSwin9.pdf> (also available at the Tisch Hall, LC8 Academic Computing PC site) is still useful to the new SPSS user. The skills needed to perform data analysis in SPSS are accurately detailed in the Version 9 document.