

SPSS TUTORIALS: DEFINING VARIABLES

Search this Guide

SEARCH

Defining Variables

Defining a variable includes giving it a name, specifying its [type](#), the values the variable can take (e.g., 1, 2, 3), etc. Without this information, your data will be much harder to understand and use. Whenever you are working with data, it is important to make sure the variables in the data are defined so that you (and anyone else who works with the data) can tell exactly what was measured, and how.

There are three ways of defining information about variables:

1. The Variable View column attributes.
2. Syntax.
3. The Define Variable Properties window.

We explain the different attributes that variables in SPSS have and how to define them in the sections below. We conclude with an example that demonstrates why it is important to define your variables—and especially why it will make working with your data and performing analyses much more straightforward.



Defining Variables in the *Variable View*

You can define information about your variables by accessing the **Variable View** tab (at the bottom of the Data Editor window). The **Variable View** tab displays information about the variables in your data. You can get to the Variable View window in two ways:

In the Data Editor window, click the Variable View tab at the bottom.

In the Data Editor window, in the Data View tab, double-click a variable name at the top of the column. This method has the advantage of taking you to the specific variable you clicked.

The **Variable View** tab displays the following information, in columns, about each variable in your data:

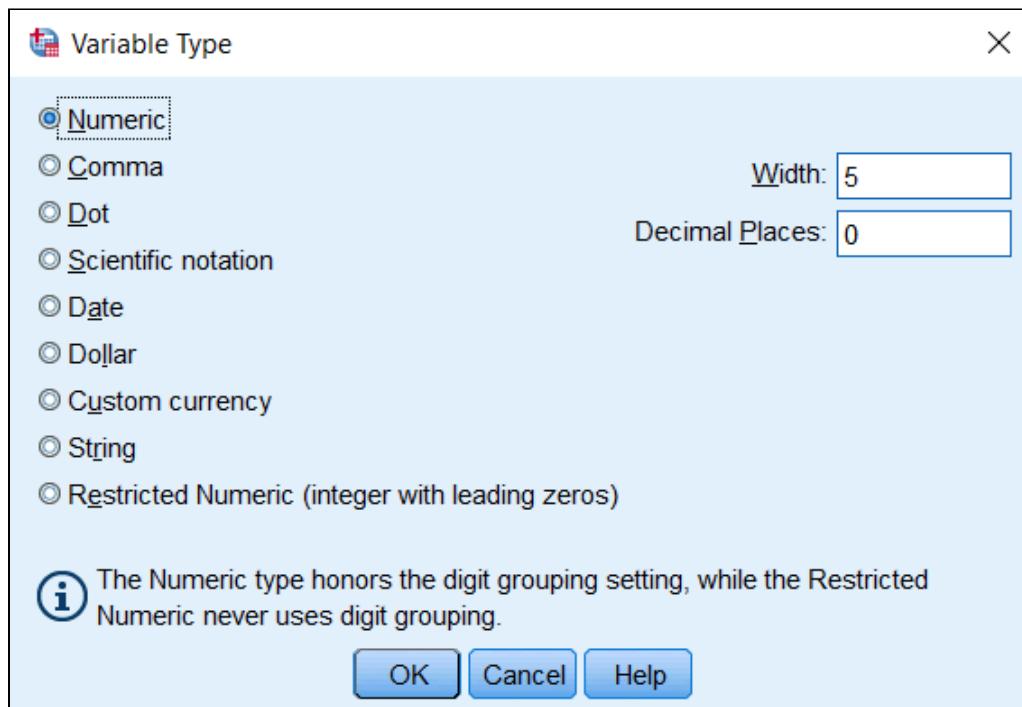
NAME

The name of the variable, which is used to refer to that variable in syntax. Variable names can not contain spaces. Note that when you change the name of a variable, it does not change the data; all values associated with the variable stay the same. Renaming a variable simply changes the name of that variable while leaving everything else the same. For example, we may want to rename a variable called *Sex* to *Gender*.

To change a variable's name, double-click on the name of the variable that you wish to re-name. Type your new variable name.

TYPE

The type of variable (e.g. numeric, string, etc.). (See the [Variable Types tutorial](#) for descriptions of the variable types in SPSS.)



To change a variable's type, click inside the cell corresponding to the "Type" column for that variable. A square "..." button will appear; click on it to open the Variable Type window. Click the option that best matches the type of variable. Click **OK**.

WIDTH

The number of digits displayed for numerical values or the length of a string variable.

To set a variable's width, click inside the cell corresponding to the "Width" column for that variable. Then click the "up" or "down" arrow icons to increase or decrease the number width.

DECIMALS

The number of digits to display after a decimal point for values of that variable. Does not apply to string variables. Note that this changes how the numbers are displayed, but does not change the values in the dataset.

To specify the number of decimal places for a numeric variable, click inside the cell corresponding to the "Decimals" column for that variable. Then click the "up" or "down" arrow icons to increase or decrease the number of decimal places.

Example: If you specify that values should have two decimal points, they will display as 1.00, 2.00, 3.00, and so on.



LABEL

A brief but descriptive definition or display name for the variable. When defined, a variable's label will appear in the output in place of its name.

Example: The variable *expgradate* might be described by the label "Expected date of college graduation".

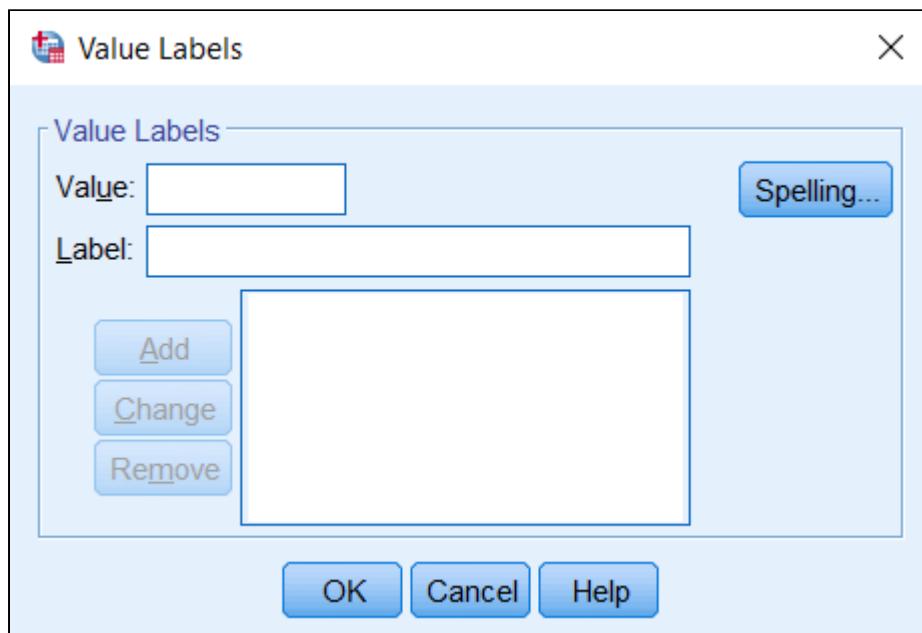
VALUES

For coded categorical variables, the value label(s) that should be associated with each category abbreviation. Value labels are useful primarily for categorical (i.e., nominal or ordinal) variables, especially if they have been recorded as codes (e.g., 1, 2, 3). It is strongly suggested that you give each value a label so that you (and anyone looking at your data or results) understands what each value represents.

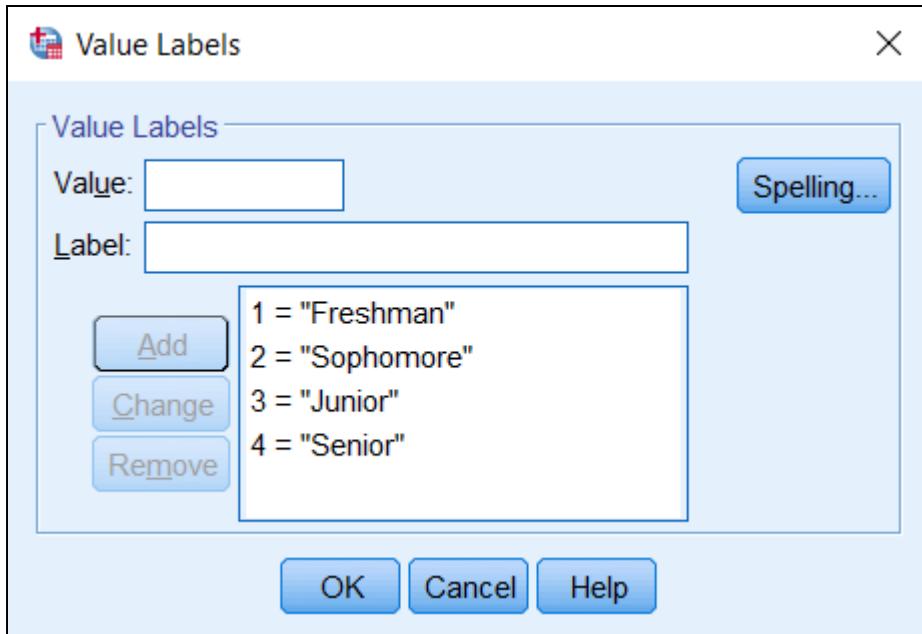
When value labels are defined, the labels will display in the output instead of the original codes. Note that defining value labels only affects the labels associated with each value, and does not change the recorded values themselves.

Example: In the sample dataset, the variable Rank represents the student's class rank. The values 1, 2, 3, 4 represent the categories Freshman, Sophomore, Junior, and Senior, respectively. Let's define the category labels for the *Rank* variable in the sample data.

Under the column "Values," click the cell that corresponds to the variable whose values you wish to label. If the values are currently undefined, the cell will say "None." Click the square "... button. The Value Labels window appears.



Type the first possible value (1) for your variable in the **Value** field. In the **Label** field type the label exactly as you want it to display (e.g., "Freshman"). Click **Add** when you are finished defining the value and label. Your variable value and label will appear in the center box. Repeat these steps for each possible value for your variable. When all of the labels have been defined, the Value Labels window should look like this:



Click **OK** at the bottom of the window.

If you wish to change or remove a value and label that you have added to the center dialog box, do the following:

To change a specific value or label, highlight the value/label in the center text box in the Value Labels window. Now the selected value/label will be highlighted yellow. Make changes to the selected value or label as needed. Click **Change**. The changes will be applied to the value/label you highlighted.

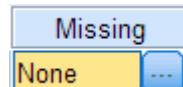
To remove a specific value/label, highlight the value/label in the center text box. Click **Remove**. The selected value/label will be removed from the center text box.



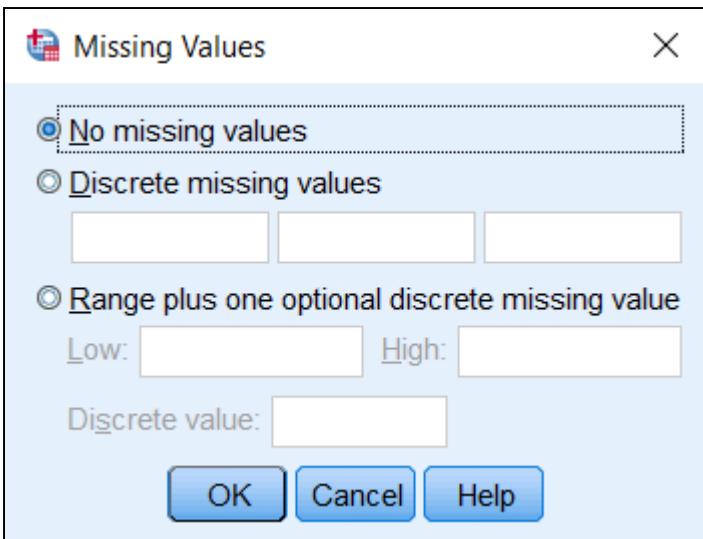
MISSING

User-defined data values (or ranges of values) should be treated as missing. Note that this property does not alter or eliminate SPSS's default missing value code for numeric variables ("."). This column merely allows the user to specify up to three unique missing value codes for the given variable; or, to specify a range of numbers to treat as missing, plus one additional unique missing value code.

To set user-defined missing value codes, click inside the cell corresponding to the "Missing" column for that variable. A square button will appear; click on it.



The Missing Values window appears.



Click the option that best matches how you wish to define missing data and enter any associated values, then click **OK** at the bottom of the window.

Note that you may enter numbers or letters as discrete missing value codes in the "discrete missing values" boxes.

Caution: If you have a dataset with string variables, blank cells are not automatically recognized as missing values. In order for blanks to be recognized as missing values, you can either:

⇑

- add a space character (`Spacebar` key) as a discrete missing value code (either in the Variable View or using syntax), or
- use the [Automatic Recode](#) procedure to recode the string variable into a labeled, numeric categorical variable with blanks recoded into a special missing value code.

The latter option works well if there are a limited number of unique string values, but is a poor option if there are many unique variations in the strings (e.g. capitalization, spelling, spacing).

COLUMNS

The width of each column in the Data View spreadsheet. Note that this is not the same as the number of digits displayed for each value. This simply refers to the width of the actual column in the spreadsheet.

To set a variable's column width, click inside the cell corresponding to the "Columns" column for that variable. Then click the "up" or "down" arrow icons to increase or decrease the column width.

ALIGN

The alignment of content in the cells of the SPSS Data View spreadsheet. Options include left-justified, right-justified, or center-justified.

To set the alignment for a variable, click inside the cell corresponding to the "Align" column for that variable. Then use the drop-down menu to select your preferred alignment: Left, Right, or Center.

MEASURE

The level of measurement for the variable (e.g., nominal, ordinal, or scale).

Some procedures in SPSS treat categorical and scale variables differently. **By default, variables with numeric responses are automatically detected as "Scale" variables.** If the numeric responses actually represent categories, you must change the specified measurement level to the appropriate setting.

To define a variable's measurement level, click inside the cell corresponding to the "Measure" column for that variable. Then click the drop-down arrow to select the level of measurement for that variable: Scale, Ordinal, or Nominal.



It is vital that you correctly define each variable's measurement level. This setting affects everything from graphs to internal algorithms for statistical analysis. **Incorrectly specifying measurement level can have unintended and potentially disastrous effects on your results.**

ROLE

The role that a variable will play in your analyses (i.e., independent variable, dependent variable, both independent and dependent).

Some options in SPSS allow you to pre-select variables for particular analyses based on their defined roles. Any variable that meets the role requirements will be available for use in such analyses. You can choose from the following roles for each variable:

Input: The variable will be used as a predictor (independent variable). This is the default assignment for variables.

Target: The variable will be used as an outcome (dependent variable).

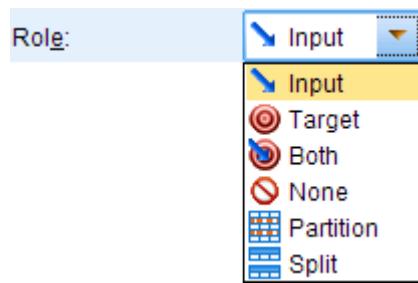
Both: The variable will be used as both a predictor and an outcome (independent and dependent variable).

None: The variable has no role assignment.

Partition: The variable will partition the data into separate samples.

Split: Used with the IBM® SPSS® Modeler (not IBM® SPSS® Statistics).

To define a variable's role in your analysis, click inside the cell corresponding to the "Role" column for that variable. Then use the drop-down menu to select the role that variable will take: Input, Target, Both, None, Partition, or Split.



Defining Variables using Syntax

VARIABLE NAMES

Change an existing variable's name

Rename one variable:

```
RENAME VARIABLES (oldname=newname) .
```

Rename more than one variable:

```
RENAME VARIABLES (oldname=newname) (oldname2=newname2)  
(oldname3=newname3) .
```

VARIABLE WIDTH

Set the width for one variable:

```
VARIABLE WIDTH var1 (10) .
```

Set the same width for multiple variables:

```
VARIABLE WIDTH var1 var2 var3 (10) .
```

Set different widths for multiple variables:

```
VARIABLE WIDTH var1 var2 var3 (10)  
/ var4 var5 (20)  
/ var6 (5) .
```



VARIABLE MEASUREMENT LEVEL

Set the measurement level (nominal, ordinal, or scale) for one or more variables at a time:

```
VARIABLE LEVEL var1 var2 var3 (SCALE) .
```

```
VARIABLE LEVEL var4 var5 (ORDINAL) .
```

```
VARIABLE LEVEL var6 (NOMINAL) .
```

Set more than one variable's measurement level at a time:

```
VARIABLE LEVEL var1 var2 var3 (SCALE)  
/ var4 var5 (ORDINAL) .  
/ var6 (NOMINAL) .
```

VARIABLE LABELS

Set label for one variable:

```
VARIABLE LABELS varname "Variable label".
```

Set labels for several variables:

```
VARIABLE LABELS var1 "Variable 1 label" var2 "Variable 2  
label" var3 "Variable 3 label".
```

VALUE LABELS

Define labels for one numeric variable's values:

```
VALUE LABELS var1 0 'No' 1 'Yes'.
```

Define labels for one string variable's values:

```
VALUE LABELS var2 'm' 'Male' 'f' 'Female'.
```



Define the same set of labels for more than one numeric variable (e.g. you have several 5-point Likert items that all use the same coding scheme):

```
VALUE LABELS service quality speed overall  
-2 'Very unsatisfied'  
-1 'Somewhat unsatisfied'  
1 'Somewhat satisfied'  
2 'Very satisfied'.
```

```
VALUE LABELS likert1 TO likert10  
1 'Strongly disagree'  
2 'Disagree'  
3 'Neither agree nor disagree'  
4 'Agree'  
5 'Strongly agree'.
```

Define more than one set of labels at a time:

```
VALUE LABELS married smoker 0 'No' 1 'Yes'  
/ sex 1 'Male' 2 'Female'.
```

MISSING VALUE CODES

Define one special missing value for a single numeric variable:

```
MISSING VALUES num1 (-999).
```

Define more than one special missing value for a single numeric variable:

```
MISSING VALUES num1 (-999, -888).
```

Define a set of missing value codes to be applied to several numeric variables:

```
MISSING VALUES num1 num2 num3 (-999, -888).
```



Define one special missing value for a single string variable:

```
MISSING VALUES string1 ('x').
```

Define a blank character as a special missing value (only applies to string variables):

```
MISSING VALUES string1 ("").
```

Define more than one special missing value for a single string variable:

```
MISSING VALUES string1 ('x', 'missing', '-999').
```

Define different sets of special missing values for different variables:

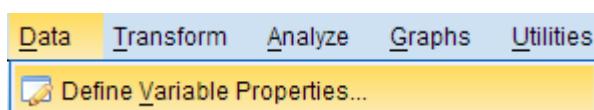
```
MISSING VALUES num1 num2 (-99, -88) string1 ('x',  
'missing') string2 ('-999').
```

Reset missing value codes for all variables:

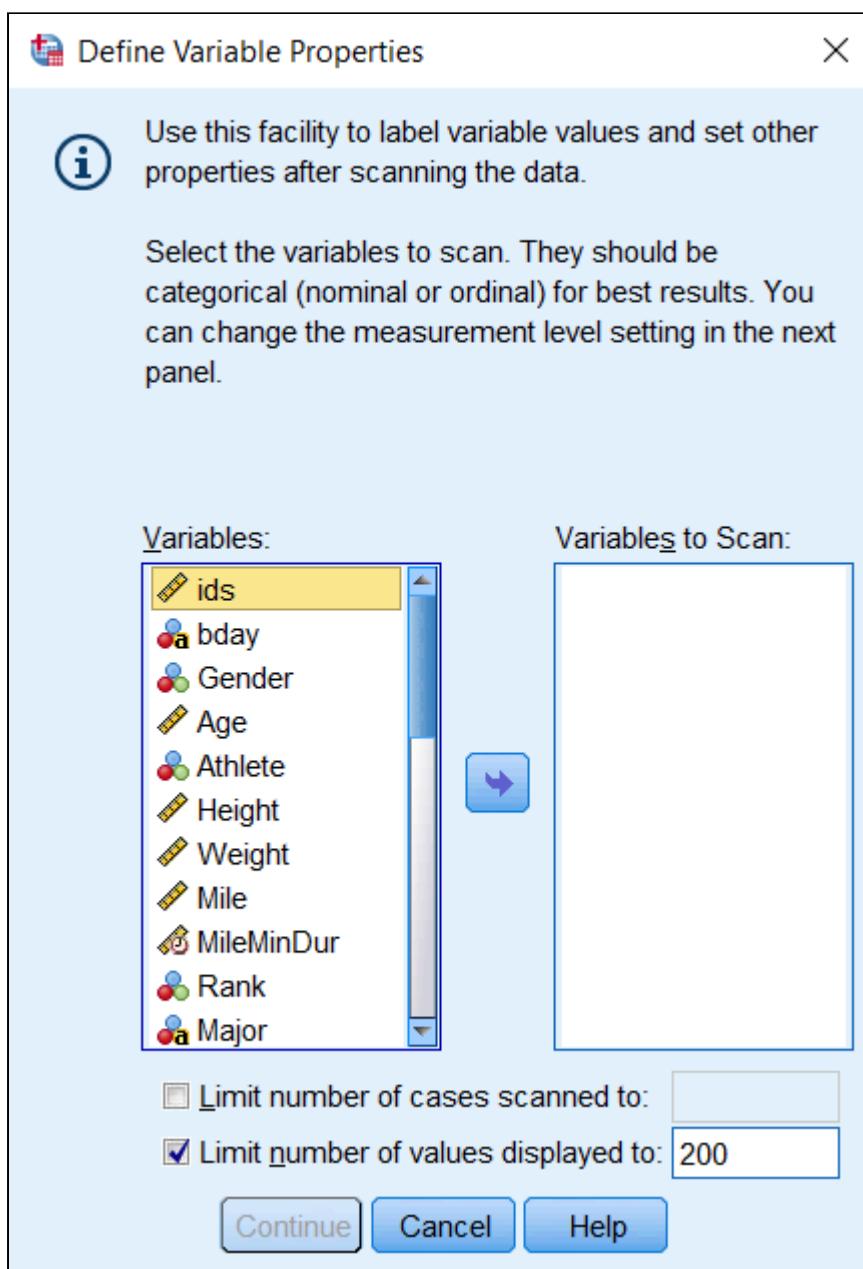
MISSING VALUES ALL () .

Defining Variables with Define Variable Properties

The Define Variable Properties window is an efficient way of defining many variables at once, or defining many variables that share the same formatting. Click **Data > Define Variable Properties**.

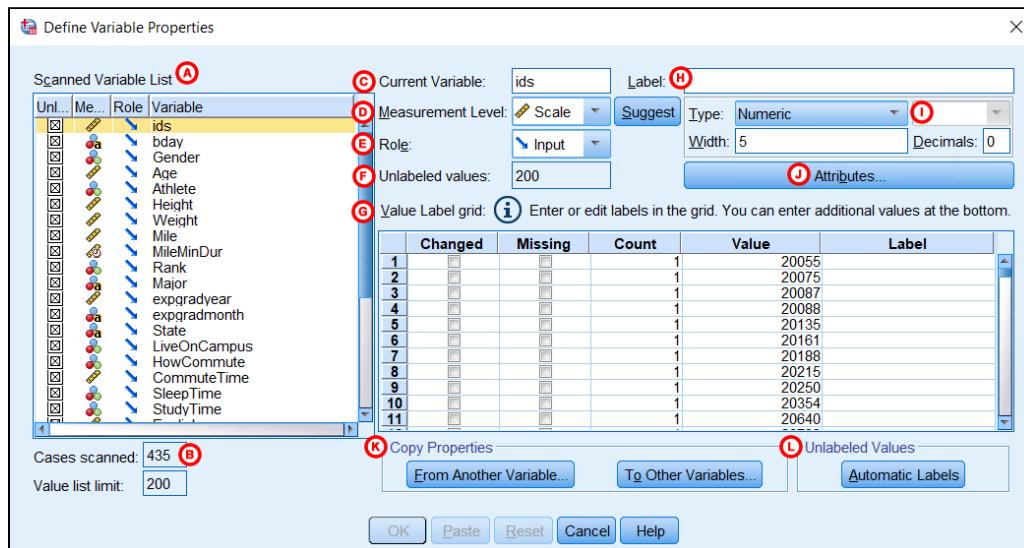


The Define Variable Properties window will open.



The left column displays all of the variables in your dataset. Select the variables you wish to define and move them to the right column using the arrow button. Note that you can specify the number of cases to scan, as well as the number of values that will display in the next step. Click **Continue** when you have finished selecting variables.

A second window will appear; this one allows you to define various properties for each variable you selected.



(A) Scanned Variable List: The “Scanned Variable List” column includes the variables selected in the previous step. Variables that do not have assigned value labels will have an X in the “Unlabeled” column. For example, if the variable Gender has potential values of “1” and “2” but these values are not labeled (e.g., “male” and “female”, respectively), the Unlabeled values check box will be selected for this variable. The current Measurement Level and Role for each variable is also displayed.

(B) Cases scanned: This section displays the number of cases that were scanned for each selected variable, as well as the number of values that are listed in the Value Label grid (G).

(C) Current Variable: Displays the variable that is currently selected from the Scanned Variable List (A).

(D) Measurement Level: Displays the level of measurement for the selected variable. You can change the level of measurement by clicking the menu arrow and choosing the desired measurement level from the listed options: Scale, Ordinal, Nominal. You can also

see the suggested level of measurement for your selected variable. To do this, click Suggest; this will open a new window that will display the currently selected variable, the current measurement level, and SPSS's suggested level of measurement. SPSS also provides an explanation for the suggestion, and a description of each possible type of measurement level (nominal, ordinal, scale) to help you make a decision.

E **Role:** Displays the role for the selected variable. Some options in SPSS allow you to pre-select variables for particular analyses based on their defined roles. Any variable that meets the role requirements will be available for use in such analyses. You can change the role by clicking the menu arrow and choosing the desired role from the listed options: Input, Target, Both, None, Partition, Split.

F **Unlabeled Values:** Specifies how many values do not have corresponding value labels.

G **Value Label grid:** Displays current information about the selected variable and updates the information based on any changes you make.

Label: Displays value labels that have already been specified for the variable. You can change value labels by clicking on cells beneath the "Label" column and typing labels for each value specified in the "Value" column. If there are values you wish to label that are not currently displayed, you can enter the values in the "Value" column below the last value listed.

Value: The values for the selected variable. Note: The values are based on the specified number of scanned cases (B).

Count: The number of times a value occurs. Note: The count is based on the specified number of scanned cases (B).

Missing: Defines values as missing data. To mark certain values as missing data, simply check the box under "Missing" for the associated value under the "Value" column. Note: If a variable already has defined missing values (e.g., -99), you

cannot change the missing values using the Define Variables Properties window. Instead, you will need to go to Variable View and specify any changes in the "Missing" column.

Changed: If you change the value label of a variable, the row associated with the changed value label will automatically be check-marked under the "Changed" column.

(H) Label: Allows you to add a label for the selected variable that describes more about what the variable is. This label is for the variable rather than for the values of the variable. For example, we might select the variable StudentID and give it the label "Student ID #".

(I) Type: Allows you to specify a particular kind of variable that helps SPSS know how to work with the variable during analyses. The types include numeric, comma, dot, scientific, date, dollar currency percent, string, and restricted numeric. Depending on the type you select for your variable, you may be asked to supply additional information. For example, if you select "Date" as the type, you will then be able to select the format of the date from a drop-down menu to the right. You can also set the width and may be asked to set the decimals for your variable. Notice that when you select a particular type for the variable, examples of how the variable would display in your data appear in the Value Label grid area under "Value."

(J) Attributes: Allows you to define custom attributes for variables. These attributes are supplementary information not otherwise specified by the variable's label, measurement labels, and missing values.

(K) Copy Properties: Allows you to copy properties from one variable to another variable. You can copy the properties from another variable to the currently selected variable, or copy the properties of the currently selected variable to one or more other variables. (For example, you may have several variables representing survey items, all of which use the value labels 0 = "No" and 1 = "Yes". After defining the value labels for the first item, you can use "Copy Properties" to quickly set the labels for the remaining survey item variables.)

L **Unlabeled Values:** Allows you to automatically label unlabeled values by clicking Automatic Labels.

When you are finished defining your variables, click OK at the bottom of the window to apply the changes to your data.

Example: Adding value labels

As we mentioned at the beginning of this tutorial, it is important to define the variables in your data so that you (and anyone else working with your data) can easily understand what was measured, and how. In this section, we provide an example of the confusion that can result when value labels are not defined, and how to correct it.

In the sample data, the variable *Gender* has two possible values: 0 and 1. The sample data file is not formatted with any value labels. Let's make a Frequency table of the *Gender* variable to see what the distribution of gender is in our sample. Click **Analyze > Descriptive Statistics > Frequencies**. Select the variable *Gender*, then click OK. (The Frequencies command will produce a frequency table.) The Output Viewer displays the following results:



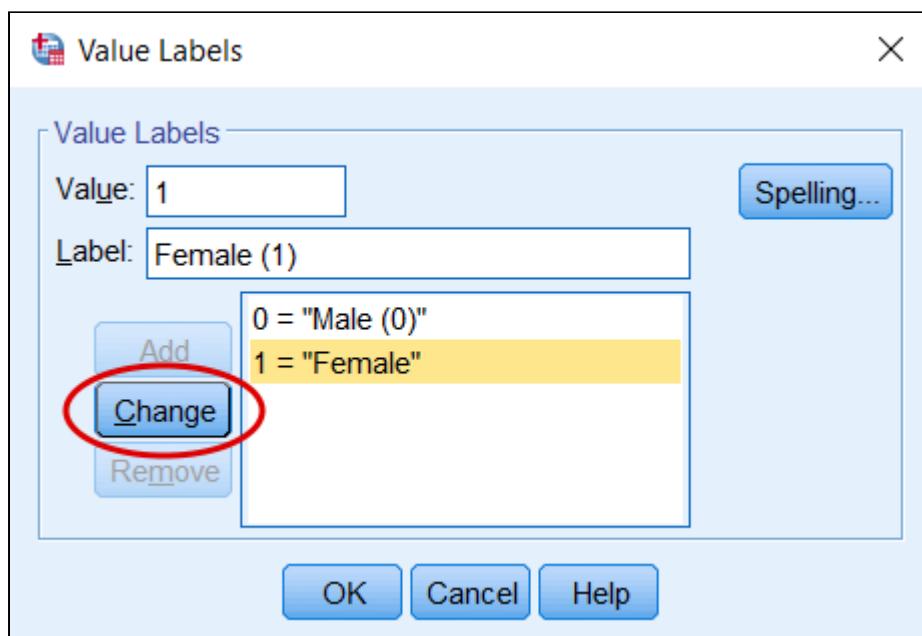
Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	204	46.9	47.9	47.9
	1	222	51.0	52.1	100.0
	Total	426	97.9	100.0	
Missing	System	9	2.1		
	Total	435	100.0		

This output shows frequencies for the variable *Gender*, which can take on values of "0" or "1." We see that value "0" has 204 cases and value "1" has 222 cases. But what do these values mean? Which values represent females, and which values represent males? There is no commonly accepted coding scheme for gender, so readers not familiar with the data can not be certain what is represented in this table.

In the sample data, 0 represents a Male, and 1 represents a Female. After defining the value labels (using the methods described above) and re-running the Frequencies command, the output is much easier for the reader to understand:

Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	204	46.9	47.9	47.9
	Female	222	51.0	52.1	100.0
	Total	426	97.9	100.0	
Missing	System	9	2.1		
	Total	435	100.0		

It may also be useful to rewrite the labels so that the numeric code is included with the label. In this situation, we could alter the label for "male" to "Male (0)", and alter the label for "female" to "Female (1)".



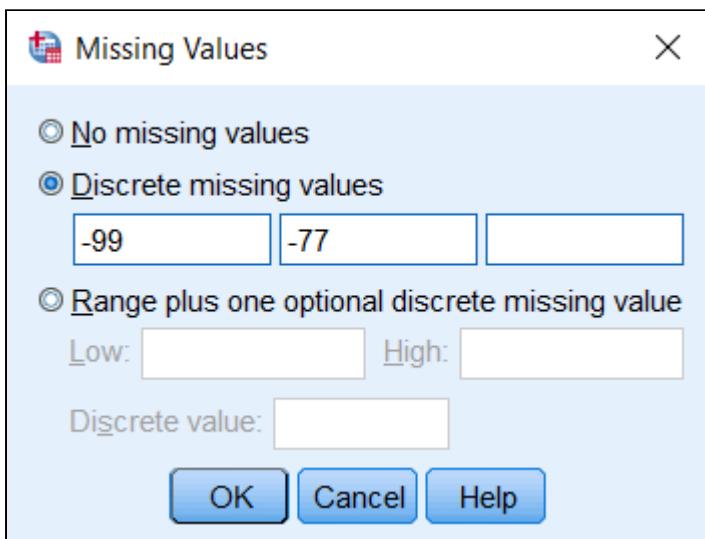
As you can see from this example, including value labels for each variable makes working with data and interpreting output much more straightforward. And remember: value labels are only one of many attributes that we can define for each variable. The more information you define about each variable, the easier it will be to navigate your data and interpret the output of analyses.

Example: Defining special codes for missing values

Suppose you have conducted a survey that has a time limit, and want to be able to distinguish respondents who refused to answer a question from respondents who ran out of time.

Respondents who refused to answer a survey item are coded as -99. Respondents who did not complete the survey item in the allotted time are coded as -77. All other missing responses were left blank.

To have SPSS recognize these special missing value codes, you'll need to enter these numbers as indicators of missing values under the **Variable View** tab. Click on the cell corresponding to the "Missing" column for the variable of interest to open the Missing Values window. Click **Discrete missing values**, then enter the two missing value codes.



TIPS

You can specify up to three different missing value codes.

You can apply value labels to missing value codes just like you would to valid categories. This is actually a good practice, because the names of missing value codes appear in the output.

Without Value Labels

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	32	20.5	23.7
	2	27	17.3	43.7
	3	14	9.0	54.1
	4	41	26.3	84.4
	5	21	13.5	100.0
	Total	135	86.5	100.0
Missing	-99	6	3.8	
	-77	11	7.1	
	System	4	2.6	
	Total	21	13.5	
Total		156	100.0	

With Value Labels

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	32	20.5	23.7
	Somewhat disagree	27	17.3	43.7
	Neither agree nor disagree	14	9.0	54.1
	Somewhat agree	41	26.3	84.4
	Strongly agree	21	13.5	100.0
	Total	135	86.5	100.0
Missing	Refused response	6	3.8	
	Out of time	11	7.1	
	System	4	2.6	
	Total	21	13.5	
Total		156	100.0	

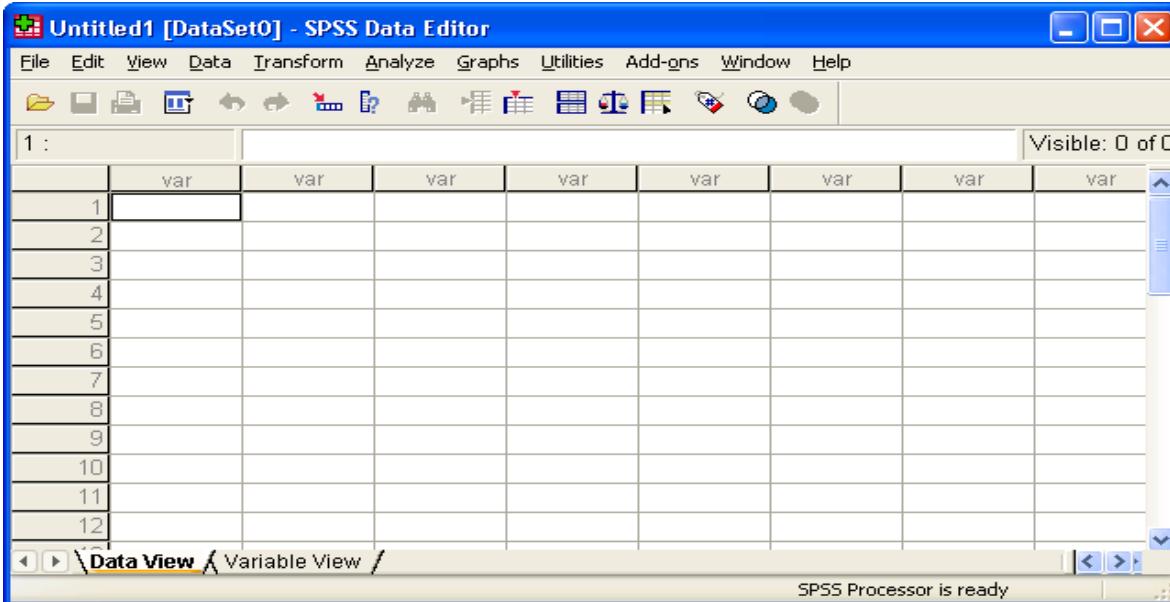


Variable Name	Description	Type (In Data File)	Measure (Actual)	Value labels
ids	ID number	Numeric	Nominal	
bday	Date of birth (mm/dd/yyyy)	Date, mm/dd/yyyy	Scale	
enrolldate	Date of college enrollment (mm/dd/yyyy)	String	Scale	
exgradate	Expected date of college graduation (mm/dd/yyyy)	String	Scale	
Rank	Class rank	Numeric	Ordinal	1=Freshman 2=Sophomore 3=Junior 4=Senior
Major	Major	String	Nominal	
Gender	Gender	Numeric	Nominal	0=Male 1=Female
Athlete	Is student an athlete?	Numeric	Nominal	0 = Non-athlete 1 = Athlete
Height	Height (inches)	Numeric	Scale	
Weight	Weight (pounds)	Numeric	Scale	
Smoking	Does student smoke?	Numeric	Nominal	0 = Nonsmoker 1 = Past smoker 2 = Current smoker
Sprint	35-meter sprint time (seconds)	Numeric	Scale	
MileMinDur	Mile run time (hh:mm:ss)	Datetime	Scale	
English	Score on English placement test (out of 100 points)	Numeric	Scale	
Reading	Score on Reading placement test (out of 100 points)	Numeric	Scale	
Math	Score on Math placement test (out of 100 points)	Numeric	Scale	
Writing	Score on Writing placement test (out of 100 points)	Numeric	Scale	
State	Is student in-state or out-of-state resident?	String	Nominal	
LiveOnCampus	Does student live on campus?	Numeric	Nominal	0 = Off-campus 1 = On-campus
HowCommute	How does student commute to campus?	Numeric	Nominal	1 = Walk 2 = Bike 3 = Car 4 = Public transit 5 = Other
CommuteTime	How long does it take you to commute to campus? (minutes)	Numeric	Scale	
SleepTime	About how many hours of sleep do you get per night?	Numeric	Scale	
StudyTime	About how many hours per week do you study?	Numeric	Scale	



Creating a New SPSS Dataset -- Variable Specification and Data Entry

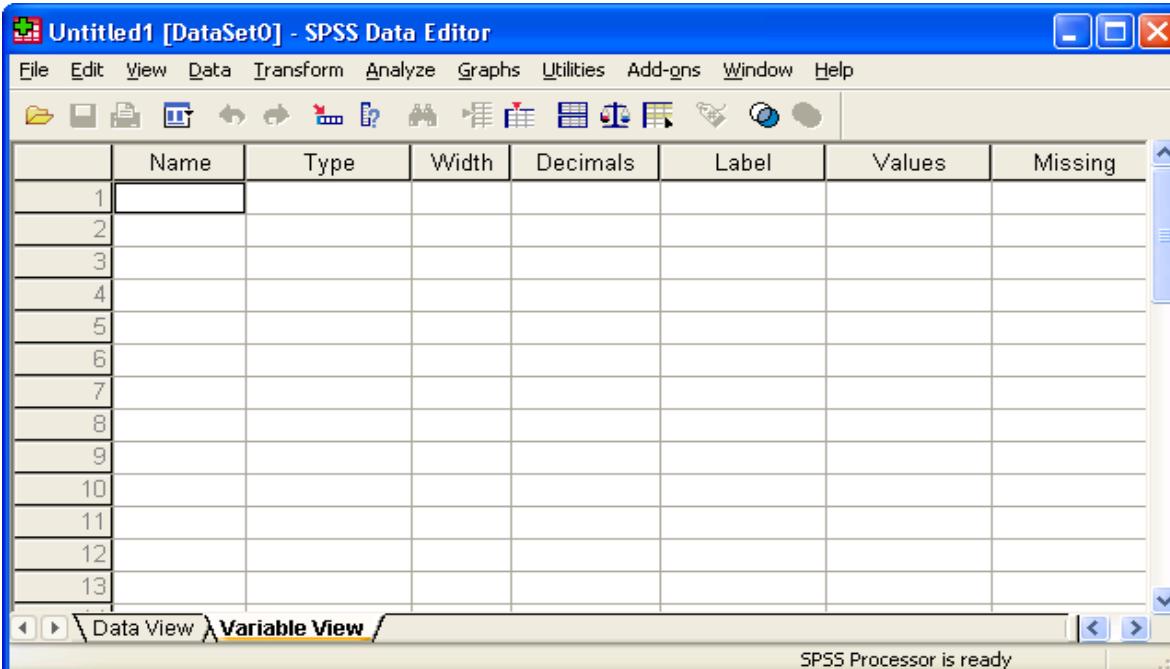
When you open SPSS you will see the Data Editor



This shows the **Data View** of the editor.

- Each row holds the data from a participant or “case”
- Each column holds the data for a variable

After specifying the information for each variable using the Variable View of the editor, we will return to this Data View to enter the data from each case.



This shows the **Variable View** of the editor.

- Each row holds the specification for a variable
- Each column holds a different specification

It is important to specify key aspects of each variable. These specifications will make the statistical analyses and output clearer and easier to read as well as provide a “reminder” of the decisions you made when collecting and entering the data.

The initial form of most datasets is a listing of the variable values for each participant. Usually each row is a different participant with a column for each variable that was collected.

Below are the data for 12 participants. Each “participant” is a different pet store. The researchers visited each store and recorded information about eight variables. Those variables and details about each are shown on the right

1	1	2	12	1	6	32	3	34
2	2	8	14	2	5	41	3	38
3	2	9	15	2	3	31	3	45
4	2	7	12	2	3	38	3	32
5	1	4	7	1	7	21	1	12
6	1	7	4	2	9	13	1	11
7	2	4	10	2	9	17	2	22
8	1	4	4	1	8	22	1	9
9	1	5	14	1	6	24	2	20
10	2	9	4	1	9	11	1	8
11	2	7	5	2	7	17	2	19
12	1	2	10	1	8	20	1	8

Participant number -- usually consecutive numbers starting with 1, 001 or some similar numbering system

Type of reptile department -- “1” means the reptile displays are mixed in with other animal and product displays and “2” means that the store has a separate department

Rating of the quality of the reptiles -- using a 1-10 scale

Count of the number of reptiles in the store

Type if fish available -- “1” means only freshwater fish are available in the store and “2” means that both freshwater and saltwater fish are available

Rating of the quality of the fish -- using a 1-10 scale

Count of the number of fish in the store

The ownership of the store -- “1” means the store is a franchise of a petstore chain, ‘2’ means the store is privately owned, and “3” means the store is part of a petstore cooperative

Count of the number of mammals in the store

Naming Variables

- Click a cell under **Variable** and type in the variable name
- Variable names may have up to 8 characters -- letters, numbers, “-” and “_” work best -- the first character must be a letter
- When you enter a name for a variable a set of “default” values automatically appears in the other columns -- some of these we will change later
- The variable names for these data are shown to the right, along with the default values

Type, Width & Decimals

The default values for these specifications usually work very well.

- “Numerical” is the most common type for both qualitative & quantitative data
- “Width” tells the number of digits in the largest variable value -- must be at least one more than the number of decimal values
- “Decimals” tells how many digits of a variable value are decimal values

petstore.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Processor area SPSS Processor is ready

	Name	Type	Width	Decimals	Label	Values
1	strnum	Numeric	2	0	store number	None
2	reptdept	Numeric	1	0	type or reptile	{1, not sep ...}
3	reptgood	Numeric	1	0	rating of reptile	None
4	reptnum	Numeric	2	0	number of repti	None
5	fishdept	Numeric	1	0	type of fish ava	{1, freshwater}
6	fishgood	Numeric	1	0	rating of fish q	None
7	fishnum	Numeric	2	0	number of fish	None
8	chain	Numeric	2	0	type of store	{1, chain store}
9	mamlnum	Numeric	2	0	number of ma	None
10						

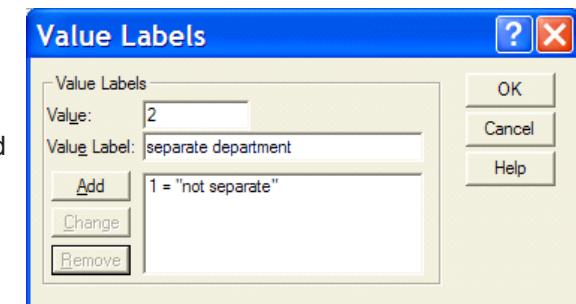
Data View Variable View

Variable Labels

- Click a cell under **Label** and type up to a 40-character description of the variable -- this label will appear in the output making it easier to understand
- giving a Label to every variable in the data set prevents forgetting what the 8-character variable names mean

Value Labels

- Specify Values for every qualitative variable -- these will appear in the output
- Click a cell under **Values**. Click the gray box that appears in the cell -- shown on the left for reptdept in row 2
- The Value Labels window will appear. In it type each "Value," its "Value Label", and click **Add** -- shown to the right for the value "2" of reptdept
- You can also highlight a specified value label and **Change** or **Remove** it later



petstore.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

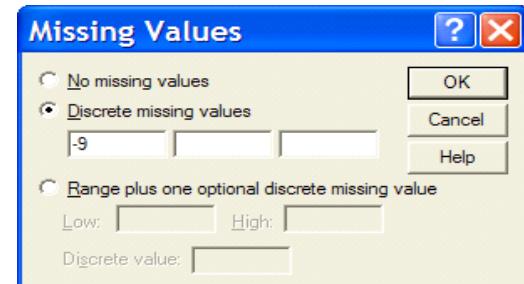
Processor area SPSS Processor is ready

	Name	Type	Width	Decimals	Label	Values	Missing
1	strnum	Numeric	2	0	store number	None	None
2	reptdept	Numeric	1	0	type or reptile	{1, not separat	None
3	reptgood	Numeric	1	0	rating of reptile	None	None
4	reptnum	Numeric	2	0	number of repti	None	None
5	fishdept	Numeric	1	0	type of fish ava	{1, freshwater}	None
6	fishgood	Numeric	1	0	rating of fish q	None	None
7	fishnum	Numeric	2	0	number of fish	None	None
8	chain	Numeric	2	0	type of store	{1, chain store}	None
9	mamlnum	Numeric	2	0	number of ma	None	None
10							

Data View Variable View

Missing Values

- leaving a cell in the dataset blank tells SPSS that value for that participant has a missing value -- a “.” will appear in that cell, signalling a “system missing value”
- you can also identify alternative “user-missing” values to indicate different types of missing data (e.g., -99 = did not respond, -98 = answer could not be scored -- up to 3 different missing values)
- Click the cell under **Missing**. Click the little gray box that appears in the cell -- shown on the left for reptdept in row 2
- Click the “Discrete missing values” button and enter the missing values you’ve chosen in the windows -- shown below for “-9” for reptdept
- Be sure to specify the meaning of each missing value in the “Values” column**



Entering Data

- Click the “**Data View**” tab in the bottom-left corner of the screen
- Enter each variable value -- in the correct column -- for each case or participant
- If you leave a space blank a “.” will appear, indicating that the value is missing -- called a “system missing” value

The completed data set for the petstores is shown below. These data will be used for many of the data analysis examples in this handbook

	strnum	reptdept	reptgood	reptnum	fishdept	fishgood	fishnum	chain	mamlnum
1	1	1	2	12	1	6	32	3	34
2	2	2	8	14	2	5	41	3	38
3	3	2	9	15	2	3	31	3	45
4	4	2	7	12	2	3	38	3	32
5	5	1	4	7	1	7	21	1	12
6	6	1	7	4	2	9	13	1	11
7	7	2	4	10	2	9	17	2	22
8	8	1	4	4	1	8	22	1	9
9	9	1	5	14	1	6	24	2	20
10	10	2	9	4	1	9	11	1	8
11	11	2	7	5	2	7	17	2	19
12	12	1	2	10	1	8	20	1	8
13									

Saving the Dataset

- File then **Save As** and proceed with the usual Windows “save” routine
- SPSS data sets have the extension “.sav”

SPSS TUTORIALS: DATA CREATION IN SPSS

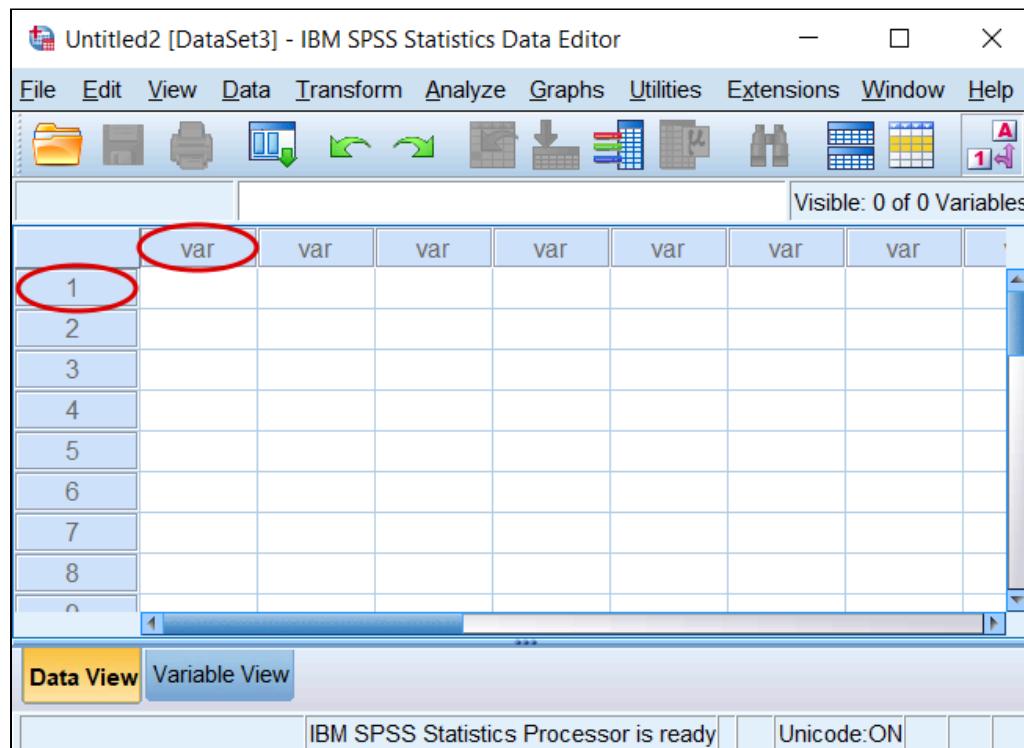
Search this Guide

SEARCH

Data Creation in SPSS

When you open the SPSS program, you will see a blank spreadsheet in Data View. If you already have another dataset open but want to create a new one, click **File > New > Data** to open a blank spreadsheet.

You will notice that each of the columns is labeled "var." The column names will represent the variables that you enter in your dataset. You will also notice that each row is labeled with a number ("1," "2," and so on). The rows will represent cases that will be a part of your dataset. When you enter values for your data in the spreadsheet cells, each value will correspond to a specific variable (column) and a specific case (row).



Follow these steps to enter data:

1. Click the **Variable View** tab. Type the name for your first variable under the **Name** column. You can also enter other information about the variable, such as the type (the default is “numeric”), width, decimals, label, etc. Type the name for each variable that you plan to include in your dataset. In this example, I will type “School_Class” since I plan to include a variable for the class level of each student (i.e., 1 = first year, 2 = second year, 3 = third year, and 4 = fourth year). I will also specify 0 decimals since my variable values will only include whole numbers. (The default is two decimals.)

The screenshot shows the SPSS Variable View interface. A single row is selected in the table, corresponding to the entry '1' in the 'Row' column. The columns are labeled 'Name', 'Type', 'Width', 'Decimals', 'Label', and 'Values'. The 'Name' column contains 'School_Class', 'Type' is 'Numeric', 'Width' is 8, 'Decimals' is 0 (highlighted in yellow), 'Label' is 'None', and 'Values' is empty. Below the table, the tabs 'Data View' and 'Variable View' are shown, with 'Variable View' being the active tab.

Row	Name	Type	Width	Decimals	Label	Values
1	School_Class	Numeric	8	0	None	

2. Click the **Data View** tab. Any variable names that you entered in *Variable View* will now be included in the columns (one variable name per column). You can see that *School_Class* appears in the first column in this example.

The screenshot shows the SPSS Data View interface. There are 10 rows of data, indexed from 1 to 10 in the first column. The columns are labeled 'School_Class', 'var', and 'var'. The 'School_Class' column contains the values 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 respectively. Below the table, the tabs 'Data View' and 'Variable View' are shown, with 'Data View' being the active tab.

	School_Class	var	var
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

3. Now you can enter values for each case. In this example, cases represent students. For each student, enter a value for their class level in the cell that corresponds to the appropriate row and column. For example, the first person's information should appear in the first row, under the variable column *School_Class*. In this example, the first person's class level is “2,” the second person's is “1,” the third person's is “1,” the fourth person's is “3,” and so on.

	School_Class
1	2
2	1
3	1
4	3
5	1
6	1
7	4
8	4
9	1
10	1
11	4
12	1
13	3

Data View Variable View

4. Repeat these steps for each variable that you will include in your dataset. Don't forget to periodically save your progress as you enter data.

Inserting or Deleting Single Cases

Sometimes you may need to add new cases or delete existing cases from your dataset. For example, perhaps you notice that one observation in your data was accidentally left out of the dataset. In that situation, you would refer to the original data collection materials and enter the missing case into the dataset (as well as the associated values for each variable in the dataset). Alternatively, you may realize that you have accidentally entered the same case in your dataset more than once and need to remove the extra case.

INSERTING A CASE

To insert a new case into a dataset:

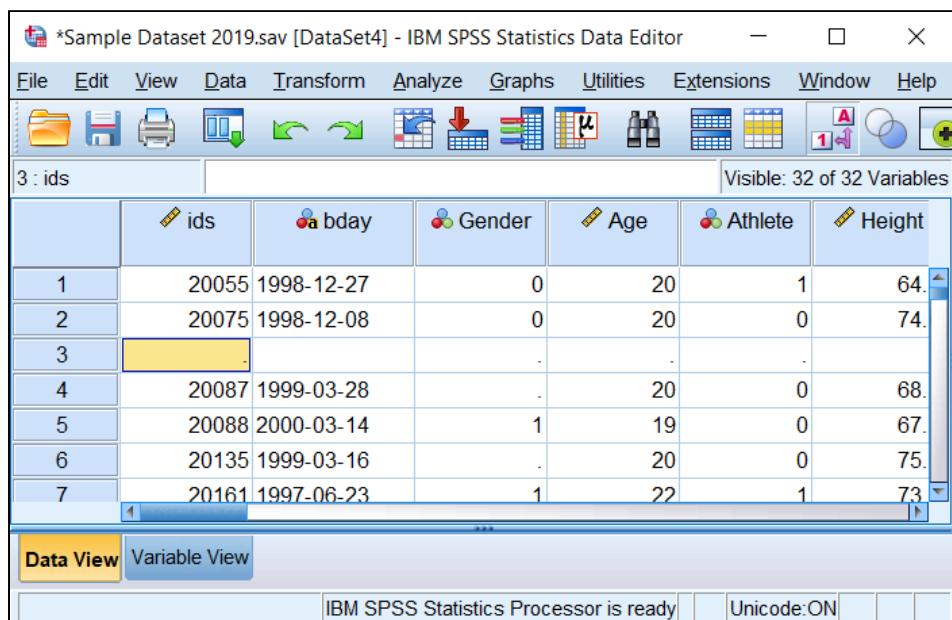
1. In Data View, click a row number or individual cell below where you want your new row to be inserted.
2. You can insert a case in several ways:

Click **Edit > Insert Cases**;

Right-click on a row and select **Insert Cases** from the menu; or

Click the **Insert Cases** icon ().

3. A new, blank row will appear above the row or cell you selected. Values for each existing variable in your dataset will be missing (indicated by either a “.” or a blank cell) for your newly created case since you have not yet entered this information.



	ids	bday	Gender	Age	Athlete	Height
1	20055	1998-12-27	0	20	1	64.
2	20075	1998-12-08	0	20	0	74.
3
4	20087	1999-03-28	.	20	0	68.
5	20088	2000-03-14	1	19	0	67.
6	20135	1999-03-16	.	20	0	75.
7	20161	1997-06-23	1	22	1	73.

4. Type in the values for each variable in the new case.

DELETING A CASE

To delete an existing case from a dataset:

1. In the Data View tab, click the case number (row) that you wish to delete. This will highlight the row for the case you selected.
2. Press **Delete** on your keyboard, or right-click on the case number and select “Clear”. This will remove the entire row from the dataset.

Inserting or Deleting Single Variables

Sometimes you may need to add new variables or delete existing variables from your dataset. For example, perhaps you are in the process of creating a new dataset and you must add many new variables to your growing dataset. Alternatively, perhaps you decide that some variables are not very useful to your study and you decide

to delete them from the dataset. Or, similarly, perhaps you are creating a smaller dataset from a very large dataset in order to make the dataset more manageable for a research project that will only use a subset of the existing variables in the larger dataset.

INSERTING A VARIABLE

To insert a new variable into a dataset:

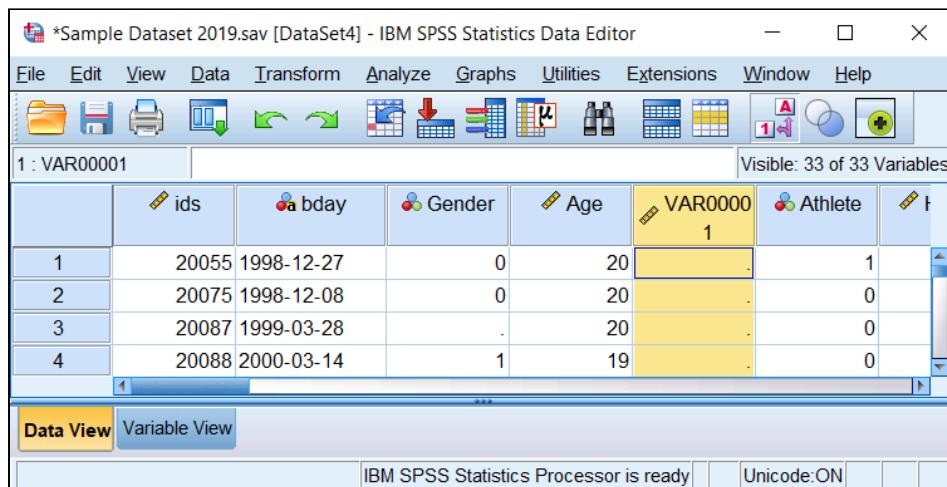
1. In the Data View window, click the name of the column to the right of where you want your new variable to be inserted.
2. You can now insert a variable in several ways:

Click **Edit > Insert Variable**;

Right-click an existing variable name and click **Insert Variable**; or

Click the **Insert Variable** icon ().

A new, blank column will appear to the left of the column or cell you selected.



The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads "*Sample Dataset 2019.sav [DataSet4] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Extensions, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main Data View window displays a table with 4 rows and 8 columns. The columns are labeled: 1, ids, bday, Gender, Age, VAR00001, Athlete, and t. The first row is a header. The 'VAR00001' column is highlighted in yellow. The 'ids' column contains dates like 20055, 1998-12-27, etc. The 'Age' column contains integers like 0, 20, etc. The 'Athlete' column contains binary values like 1, 0, etc. The bottom navigation bar shows 'Data View' is selected, along with 'Variable View' and other status indicators.

1	ids	bday	Gender	Age	VAR00001	Athlete	t
1	20055	1998-12-27		0	20	.	1
2	20075	1998-12-08		0	20	.	0
3	20087	1999-03-28		.	20	.	0
4	20088	2000-03-14		1	19	.	0

New variables will be given a generic name (e.g. VAR00001). You can enter a new name for the variable on the Variable View tab. You can quick-jump to the Variable View screen by double-clicking on the generic variable name at the top of the column. Once in the Variable View, under the column "Name," type a new name for the variable name you wish to change. You should also define the variable's other properties (type, label, values, etc.) at this time.

All values for the newly created variable will be missing (indicated by a “.” in each cell in Data View, by default) since you have not yet entered any values. You can enter values for the new variable by

clicking the cells in the column and typing the values associated with each case (row).

Tip

Is it possible to insert a variable using syntax?

Technically, there's no direct syntax command to do so. Instead, you'll need to use two syntax commands. You'll first use the **COMPUTE** command to initialize the new variable. You'll then use the **MATCH FILES** command to actually re-order the variables. Suppose we want to insert a new column of blank values into the sample dataset after the first variable, *ids*. We can use this syntax to perform these tasks:

```
/*Compute new variable containing blanks  
(system-missing values).*/  
COMPUTE newvar=$SYSMIS.  
EXECUTE.  
  
/*Reorder the variables to place the new  
variable in the desired position.*/  
MATCH FILES  
FILE = *  
/KEEP = ids newvar ALL.
```

In the **MATCH FILES** command, **FILE=*** says to act on the current active dataset. The **/KEEP** statement tells SPSS the specific order of the variables you want: we list the variables by name, in the order we want, separated by spaces, on the right side of the equals sign. The **ALL** option at the end of the line says to retain all remaining variables in their current order. The **ALL** option can only be used at the end of the line; the code will fail if you try to put it before other variable names. If we do not include **ALL**, SPSS will throw out any variables not named in the **/KEEP** statement.

DELETING A VARIABLE

To delete an existing variable from a dataset:

1. In the **Data View** tab, click the column name (variable) that you wish to delete. This will highlight the variable column.

2. Press **Delete** on your keyboard, or right-click on the selected variable and click "Clear." The variable and associated values will be removed.

Alternatively, you can delete a variable through the *Variable View* window:

1. Click on the row number corresponding to the variable you wish to delete. This will highlight the row.
2. Press **Delete** on your keyboard, or right-click on the row number corresponding to the variable you wish to delete and click "Clear".

You can also delete variables using [command syntax](#).

```
/*Delete one variable*/
DELETE VARIABLES var1.

/*Delete several variables*/
DELETE VARIABLES var1 var2 var3.
```

ID Variables versus Row Numbers

Now that you know how to enter data, it is important to discuss a special type of variable called an *ID variable*. When data are collected, each piece of information is tied to a particular case. For example, perhaps you distributed a survey as part of your data collection, and each survey was labeled with a number ("1," "2," etc.). In this example, the survey numbers essentially represent ID numbers: numbers that help you identify which pieces of information go with which respondents in your sample. Without these ID numbers, you would have no way of tracking which information goes with which respondent, and it would be impossible to enter the data accurately into SPSS.

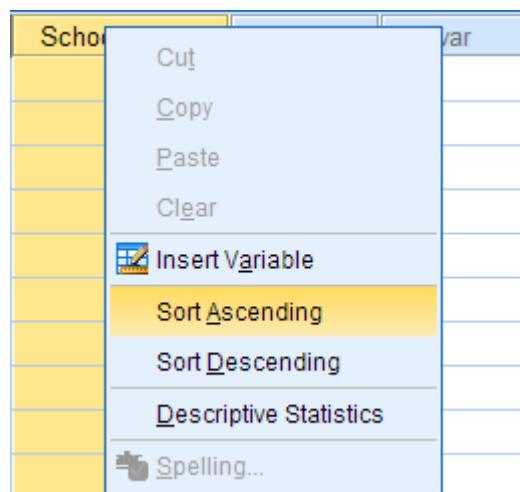
When you enter data into SPSS, you will need to make sure that you are entering values for each variable that correspond to the correct person or object in your sample. It might seem like a simple solution to use the conveniently labeled rows in SPSS as ID numbers; you could enter your first respondent's information in the row that is already labeled "1," the second respondent's information in the row

labeled “2,” etc. However, you should *never* rely on these pre-numbered rows for keeping track of the specific respondents in your sample. This is because the numbers for each row are visual guides only—they are not attached to specific lines of data, and thus cannot be used to identify specific cases in your data. If your data become rearranged (e.g., after sorting data), the row numbers will no longer be associated with the same case as when you first entered the data. Again, **the row numbers in SPSS are not attached to specific lines of data and should not be used to identify certain cases.** Instead, you should create a variable in your dataset that is used to identify each case—for example, a variable called *StudentID*.

Here is an example that illustrates why using the row numbers in SPSS as case identifiers is flawed:

Let's say that you have entered values for each person for the *School_Class* variable. You relied on the row numbers in SPSS to correspond to your survey ID numbers. Thus, for survey #1, you entered the first respondent's information in row 1, for survey #2 you entered the second person's information in row 2, and so on. Now you have entered all of your data.

But suppose the data get rearranged in the spreadsheet view. A common way of rearranging data is by sorting—and you may very well need to do this as you explore and analyze your data. Sorting will rearrange the rows of data so that the values appear in ascending or descending order. If you right-click on any variable name, you can select “Sort Ascending” or “Sort Descending.” In the example below, the data are sorted in ascending order on the values for the variable *School_Class*.



But what happens if you need to view a specific respondent's information? Or perhaps you need to double-check your entry of the data by comparing the original survey to the values you entered in SPSS. Now that the data have been rearranged, there is no way to identify which row corresponds to which participant/survey number.

The main point is that you should not rely on the row numbers in SPSS since they are merely visual guides and not part of your data. Instead, you should create a specific variable that will serve as an ID for each case so that you can always identify certain cases in your data, no matter how much you rearrange the data. In the sample data file, the variable *ids* acts as the ID variable.

Tip

If you do not have an ID variable in your dataset, a convenient way to generate one is to use the system variable \$CASENUM. You can use the [Compute Variables](#) procedure (simply enter \$CASENUM in the Numeric Expression box), or by running the following syntax after all of your data has been entered:

```
COMPUTE id=$CASENUM.  
EXECUTE.
```