

2.1 Data Entry in SPSS

(PSY206) Data Management and Analysis

Md Rasel Biswas

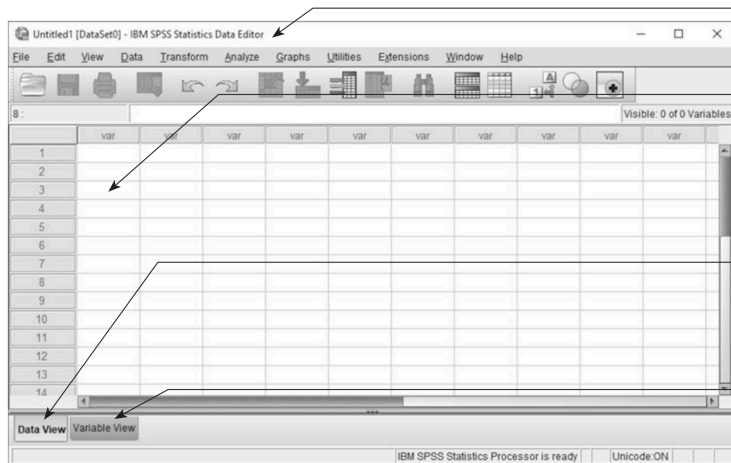
Subsection 1

1. Data Entry Preparation

1. Data Entry Preparation

- The **SPSS Data Editor** has two views:
 - ① **Data View** – displays the dataset in spreadsheet form.
 - ② **Variable View** – used to define variables, types, labels, and formats.

1. Data Entry Preparation



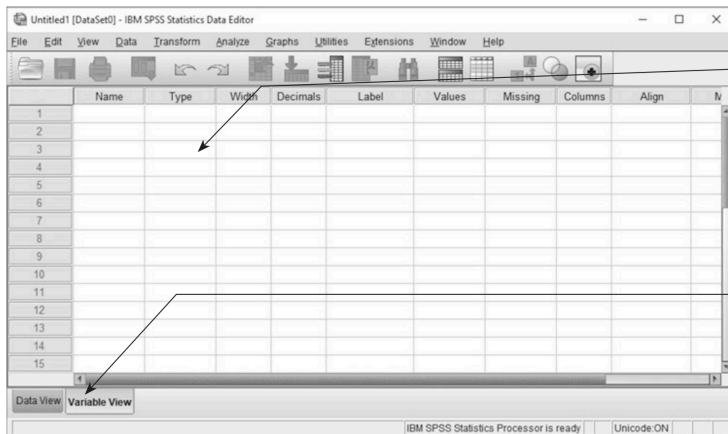
This is the Data View of the Data Editor window.

Typically, in psychology each row contains data from one participant and each column is a variable.

The **Data View** tab is highlighted. This tells you that you are looking at the Data View.

Click the **Variable View** tab to change to the Variable View (below).

1. Data Entry Preparation



This is the Variable View of the Data Editor window.

Note that the Variable View tab is highlighted to indicate you are in the Variable View.

Subsection 2

2. Defining Variables (in Variable View)

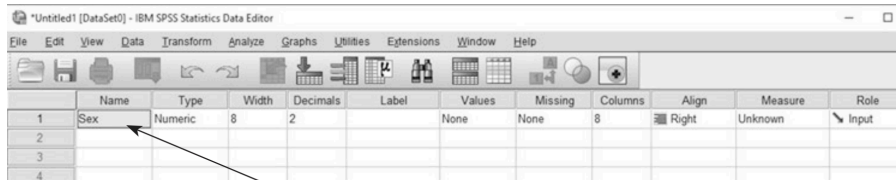
2. Defining Variables (in Variable View)

Before entering data, it is essential to define your variables properly:

- Choose clear variable names.
- Decide on coding schemes for categorical variables.
- Select appropriate formats (numeric, string, date, etc.).

Variable Name

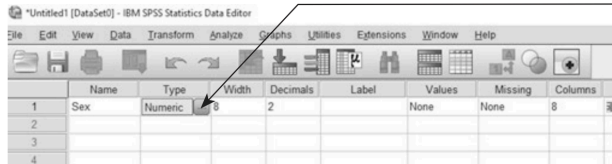
- Must start with a letter.
- No spaces allowed (use underscore _ instead).
- Example: age, gender, income_level.



We have named the first variable 'Sex', because we are going to use this variable to code the sex of our participants. Note SPSS has added default values for the other columns. We will explain these next.

Variable Type

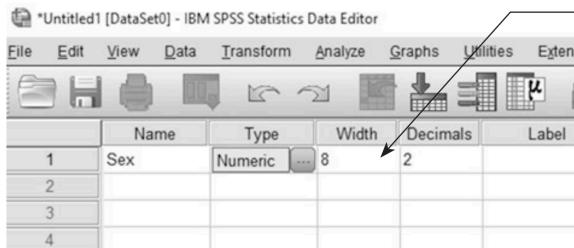
- Choose the correct type (numeric, string, date, etc.).
- Using **numeric codes** for categories is preferable (e.g., 1 = Male, 2 = Female) instead of using string variables like "M" and "F".
- **Avoid string variables** whenever possible, as they complicate analysis.



Click in the cell and then click on this button to call up the **Variable Type** dialogue box (see below).

Variable Width & Decimals

- **Width** defines how many characters are stored.
- **Decimals** sets the number of decimal places.
- Defaults are usually sufficient unless specific formatting is required.

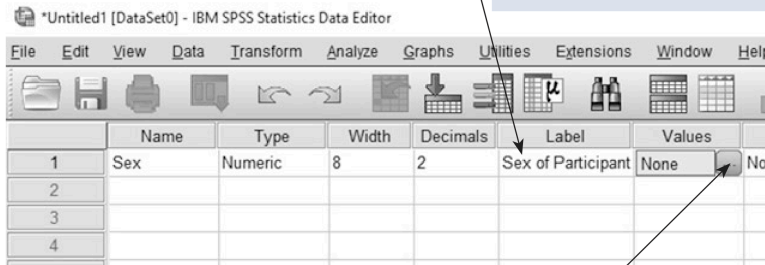


You can change the variable **Width** and number of **Decimals** places by changing these values. However, this only alters the look of the table, not the way the values are stored or used in calculations.

Variable Labels

- Add descriptive labels (e.g., *Respondent Age in Years*).
- These appear in SPSS output and help in interpretation.
- Not mandatory, but highly recommended for clarity.

Note that we have used a variable label. This isn't essential here but is often useful.



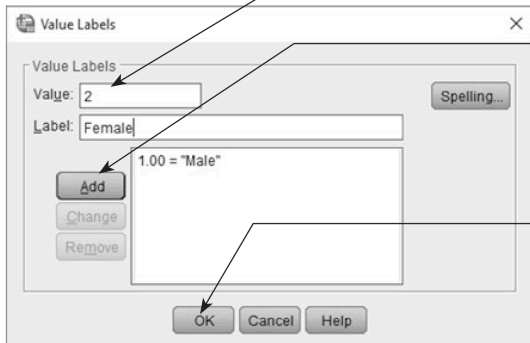
The screenshot shows the IBM SPSS Statistics Data Editor window. The variable list on the left shows four variables: 1. Sex, 2. (empty), 3. (empty), and 4. (empty). The main data grid has columns: Name, Type, Width, Decimals, Label, Values, and (empty). The first row (variable 1) has: Name: Sex, Type: Numeric, Width: 8, Decimals: 2, Label: Sex of Participant, Values: None. A callout box points to the 'Label' column for 'Sex' with the text: 'Note that we have used a variable label. This isn't essential here but is often useful.' Another callout box points to the 'Values' cell for 'Sex' with the text: 'To add value labels, click in the **Values** cell, then click on this button. This will call up the **Value Labels** dialogue box (see below).' The 'Values' cell for 'Sex' contains a button with a small icon.

	Name	Type	Width	Decimals	Label	Values
1	Sex	Numeric	8	2	Sex of Participant	None
2						
3						
4						

To add value labels, click in the **Values** cell, then click on this button. This will call up the **Value Labels** dialogue box (see below).

Value Labels

- Assign meaning to numeric codes.
- Example: 1 = Male, 2 = Female.



If you are using value labels, enter the value into the **Value** box, then enter the label for this value into the **Label** box.

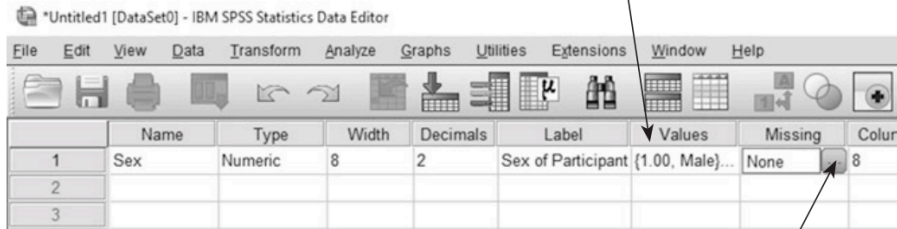
Then click on the **Add** button to add this value label to the list of labels for this variable. Repeat these steps to add additional values and labels.

When you have added all the values and labels for the variable, click on the **OK** button to close the dialogue box and return to the Variable View table.

Missing Values

- Define user-missing codes (e.g., 99 = missing).

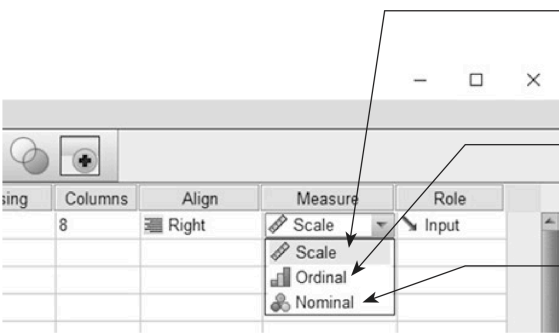
Note that we can see the value labels we added in the previous step.



Click on this button in the **Missing** cell to call up the **Missing Values** dialogue box (see below).

Measure

- This is used to specify the level of measurement for the variable.
- SPSS offers three options: Nominal, Ordinal and Scale



The screenshot shows the SPSS 'Measure' dropdown menu. The menu is open, displaying three options: 'Scale' (with a ruler icon), 'Ordinal' (with a bar chart icon), and 'Nominal' (with a pie chart icon). Three callout boxes with arrows point to these options:

- Scale**: Select the **Scale** option for variables measured using either an interval or ratio scale.
- Ordinal**: Select the **Ordinal** option for variables measured using an ordinal scale.
- Nominal**: Select the **Nominal** option for nominal variables (e.g. 'sex' or 'group').

The background shows a portion of the SPSS interface with columns labeled 'Columns', 'Align', 'Measure', and 'Role'. The 'Measure' column has a dropdown arrow, and the 'Role' column has the text 'Input'.

Role

- The final column of the Variable View table is called Role.
- This is a fairly recent addition to SPSS and is intended to help users select appropriate analyses.

The screenshot shows the SPSS Variable View table with the following columns: Name, Type, Width, Decimals, Align, Measure, and Role. The Role column dropdown menu is open, showing the following options: Input (selected), Target, Both, None, Partition, and Split. Arrows point from text boxes to each option:

- Input** is the default option and is used for independent variables and predictor variables.
- Target** is used for dependent variables.
- Both** is used for variables that can have either role.
- None** is used for variables that have no predefined role.
- For now don't use **Partition** or **Split** (see text).

Subsection 3

3. Entering Data (in Data View)

3. Entering Data (in Data View)

- We will enter the data from a simple study in which we have recorded each participant's
 - ▶ Sex (coded as 1 = Male, 2 = Female and 3 = Other),
 - ▶ Age (in years) and the
 - ▶ MemoryScore (number of words recalled from a list of 20) of each participant.
- In addition we have assigned each participant an ID number.
- Before we can enter these data, we need to define the four variables to be used (see the previous section for details of how to define a variable)

3. Entering Data (in Data View)

Defining variables in Variable View

We have set up four variables. All are numeric variables. We have left the width settings at their default values.

We have specified a missing value for each variable. For example, we used 9 as the missing value for Sex – but we could have used any value that could not otherwise occur.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0		None	999	8	Right	Nominal	Input
2	Sex	Numeric	8	2		{1.00, Male}...	9.00	8	Right	Nominal	Input
3	Age	Numeric	8	2	Age in years	None	999.00	8	Right	Scale	Input
4	MemoryScore	Numeric	8	2	Mem /20	None	99.00	8	Right	Scale	Input

Although not necessary, we have set the number of Decimals to zero for the variable ID.

We have specified the measure for each variable. ID and Sex are Nominal as the values only represent a label. Age and MemoryScore are set to Scale. Note the different icons associated with these settings. These icons will appear next to the variable name in the Data View.

3. Entering Data (in Data View)

- Each row = one observation (e.g., one respondent).
- Each column = one variable (e.g., Sex, Age, Score).

*Chpt 2_1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities E

5: MemoryScore 17.00

	ID	Sex	Age	MemoryScore
1	1	1.00	23.00	12.00
2	2	1.00	999.00	8.00
3	3	3.00	25.00	11.00
4	4	2.00	18.00	9.00
5	5	2.00	21.00	17.00
6				

This shows that a value is currently being entered for the fifth participant in the column 'MemoryScore'.

This value represents the memory score for the fifth participant. As you type in the number, it appears both here and in the cell that was highlighted.

This participant declined to tell us their age so we have entered '999', which is the missing value we assigned for this variable.

This icon next to the variable name serves as a reminder of the Measure (the level of measurement) we specified. The three circles represent Nominal, and the ruler represents Scale.

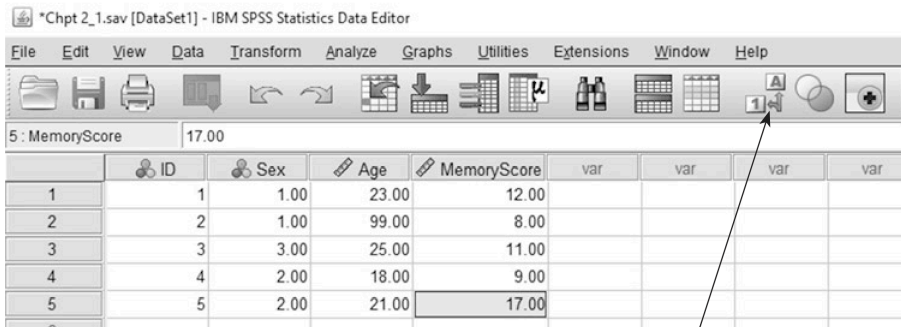
The Value Labels Button

*Chpt 2_1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

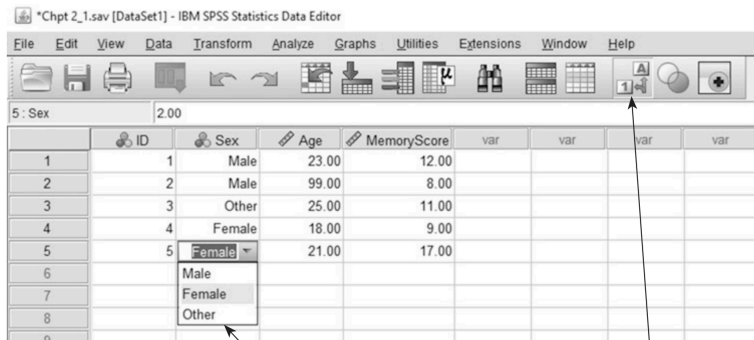
5 : MemoryScore 17.00

	ID	Sex	Age	MemoryScore	var	var	var	var
1	1	1.00	23.00	12.00				
2	2	1.00	99.00	8.00				
3	3	3.00	25.00	11.00				
4	4	2.00	18.00	9.00				
5	5	2.00	21.00	17.00				



Click on the **Value Labels** button to toggle between displaying the values entered (as shown here) and the associated labels (see below).

The Value Labels Button



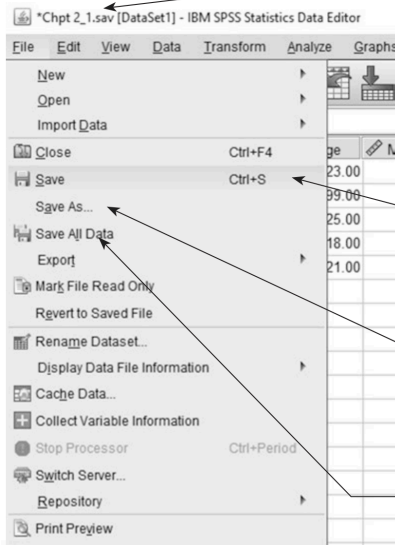
In this mode, if you double-click in a cell, SPSS will offer you a drop-down list of the value labels associated with this variable.

Once the **Value Labels** button is depressed, the values in the Sex variable are replaced by the associated labels.

Subsection 4

4. Saving a Data File

4. Saving a Data File



SPSS uses the file name 'Untitled1' for a file that has not been saved. If there are two unsaved data files open at the same time, the second will be called 'Untitled2'.

Once you save the file, your new name will appear here. We have called this file 'Chpt2_1'. SPSS adds the default suffix '.sav' to indicate a data file.

Click on **Save** to save the data file. If the file has been saved previously, it will be resaved using the same name. If not, you will be prompted for a name.

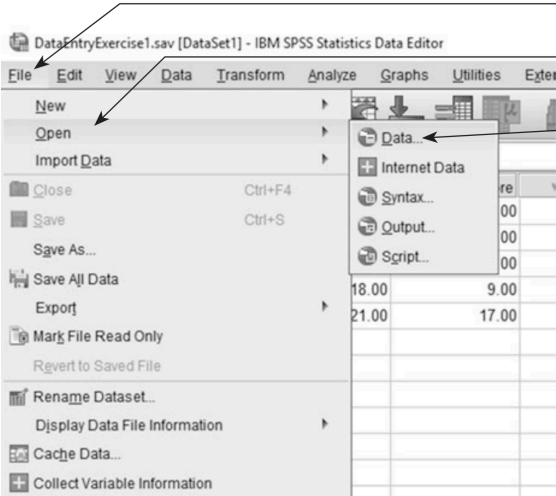
Click on **Save As** if the file has been saved before but you now want to save it under a different name.

Click on **Save All Data** if you have more than one file open, and you want to save them all.

Subsection 5

5. Opening a Data File

5. Opening a Data File



The screenshot shows the IBM SPSS Statistics Data Editor window with the title bar "DataEntryExercise1.sav [DataSet1] - IBM SPSS Statistics Data Editor". The "File" menu is open, displaying options such as "New", "Open", "Import Data", "Close", "Save", "Save As...", "Save All Data", "Export", "Mark File Read Only", "Revert to Saved File", "Rename Dataset...", "Display Data File Information", "Cache Data...", and "Collect Variable Information". The "Data..." option is highlighted in the "File" menu. A secondary menu is open for "Data...", showing options like "Internet Data", "Syntax...", "Output...", and "Script...". The background shows a data table with columns and rows of numerical data.

2. Click on the **File** menu.

3. Select **Open**.

4. Choose **Data**.

Subsection 6

6. Data Entry Exercise

6. Data Entry Exercise

RODENTS IN SPACE: A SIMPLE MEMORY EXPERIMENT

Twenty-one first-year undergraduates participated in a simple memory experiment designed to investigate the effect of a mnemonic strategy on memory for paired words. The participants were randomly divided into two groups. All participants were given two minutes to memorise a list of 20 words presented in pairs. All the participants were told to memorise the words, but those in the mnemonic instruction group were advised to try to form a mental image to link the two words in a pair (e.g. for the word pair Rocket - Hamster, a participant might imagine a small furry rodent being launched into space). The participants in the other group, the non-mnemonic group, were not given this instruction. After learning the words for two minutes, the participants were then required to complete some simple mental arithmetic problems for two minutes. Finally, all participants tried to recall the words in any order. The number of words correctly recalled was recorded. The data are summarised below.

6. Data Entry Exercise

Memory scores (out of 20) for participants in the mnemonic instruction group:

20, 18, 14, 18, 17, 11, 20, 18, 20, 19, 20

Memory scores (out of 20) for participants in the non-mnemonic group:

10, 20, 12, 9, 14, 15, 16, 14, 19, 12

6. Data Entry Exercise

Using these data, attempt to do the following:

- ➊ Set up an SPSS data file to record these data. Give appropriate names to the variables you are using.
- ➋ Apply value and variable labels where appropriate and specify the level of measurement and the missing values for each variable.
- ➌ Enter and check the data, then save the file using an appropriate file name.
- ➍ Ensure that you can reopen the file.
- ➎ Compare the data file you have constructed to the one illustrated in Section 7.

6. Data Entry Exercise

*DataEntryExIndGrps.sav [DataSet2] - IBM SPSS Statistics Data Editor

	ParticipantID	Condition	MemoryScore
1	1	1	20.00
2	2	1	18.00
3	3	1	14.00
4	4	1	18.00
5	5	1	17.00
6	6	1	11.00
7	7	1	20.00
8	8	1	18.00
9	9	1	20.00
10	10	1	19.00
11	11	1	20.00
12	12	2	10.00
13	13	2	20.00
14	14	2	12.00
15	15	2	9.00
16	16	2	14.00
17	17	2	15.00
18	18	2	16.00
19	19	2	14.00
20	20	2	19.00
21	21	2	12.00
22			

We have called this variable 'MemoryScore'. Being lazy, we have left the variable width as 8 with 2 decimal places, but we have increased the Columns setting to 10 so that the variable name fits. This isn't essential.

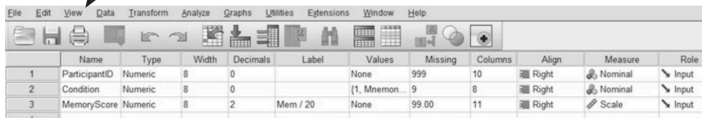
We have used the name 'Condition' for this variable. Because it is a nominal variable, we have used value labels to show that the value 1 indicates the participant was in the mnemonic condition, and the value 2 that they were in the non-mnemonic condition (see screenshot of the Value Labels dialogue box below). We have set Decimals to 0 (to display no decimal places), but this isn't essential.

For this screenshot, we did not have the **Value Labels** button depressed; if it is depressed on your system, then your screen will show the value labels rather than the values.

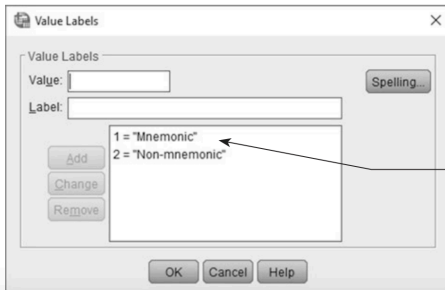
We have included a variable called 'ParticipantID', which gives each participant a number. This is not essential but is good practice; we will explain why later. We have also increased the column width so the variable name fits in.

6. Data Entry Exercise

This is the Variable View of our data file. Note that we have set Missing values for each variable and have set the Measure for each. Although not critical, we have also set some Decimals to zero and have changed the Columns for some of these variables.

A screenshot of the SPSS Variable View window. It shows a table with columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. There are four rows of variables: ParticipantID, Condition, MemoryScore, and an unlabeled row. Arrows point from the explanatory text above to the 'Missing' and 'Measure' columns.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ParticipantID	Numeric	8	0		None	999	10	Right	Nominal	Input
2	Condition	Numeric	8	0		{1, Mnemon...	9	8	Right	Nominal	Input
3	MemoryScore	Numeric	8	2	Mem / 20	None	99.00	11	Right	Scale	Input
4											

A screenshot of the 'Value Labels' dialog box in SPSS. It has fields for 'Value' and 'Label'. Below them is a list box containing '1 = "Mnemonic"' and '2 = "Non-mnemonic"'. There are buttons for 'Add', 'Change', 'Remove', 'Spelling...', 'OK', 'Cancel', and 'Help'. An arrow points from the explanatory text on the right to the list box.

Value Labels

Value:

Label:

1 = "Mnemonic"

2 = "Non-mnemonic"

Add Change Remove Spelling...

OK Cancel Help

These are the value labels we used for the variable 'Condition'.