2 Data Entry and Import
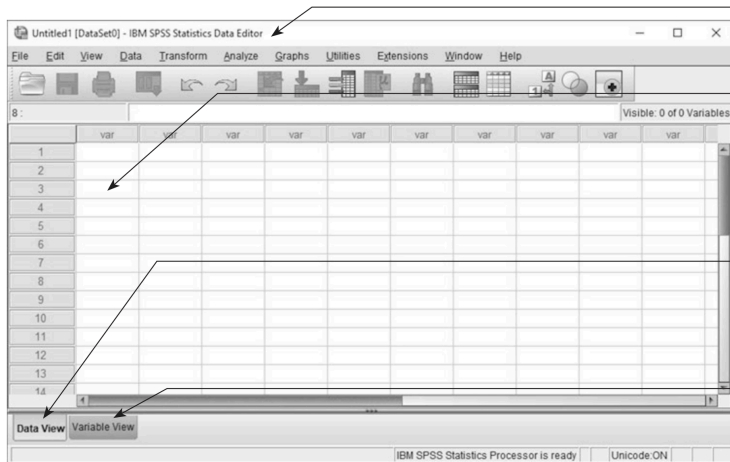## (PSY206) Data Management and Analysis

Md Rasel Biswas

Subsection 1

2.1 Data Entry in SPSS

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.
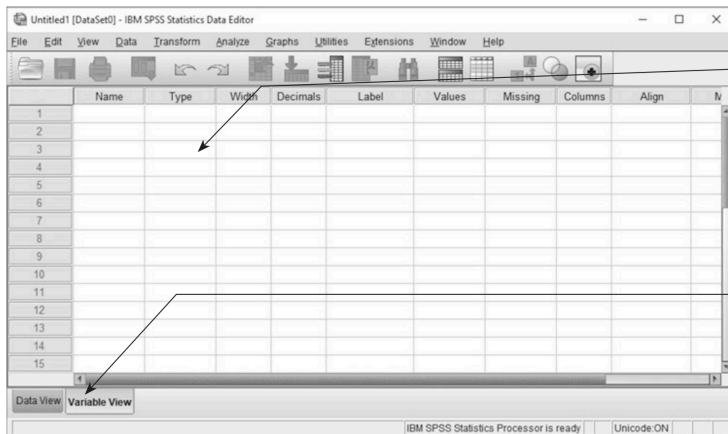
1. Data Entry Preparation

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.
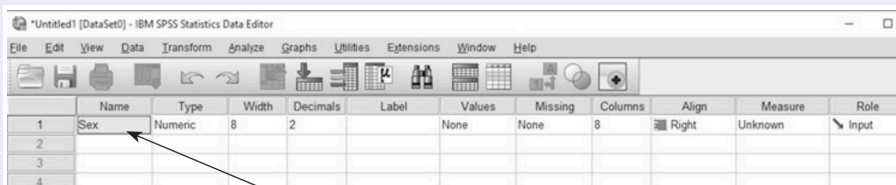
2. Defining Variables (in Variable View)

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.).

2. Defining Variables (in Variable View)

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.

2. Defining Variables (in Variable View)

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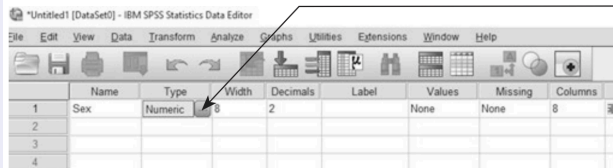
Click in the cell and then click on this button to call up the **Variable Type** dialogue box (see below).

2. Defining Variables (in Variable View)

Variable Width & Decimals

- **Width** defines how many characters are stored.
- **Decimals** sets the number of decimal places.
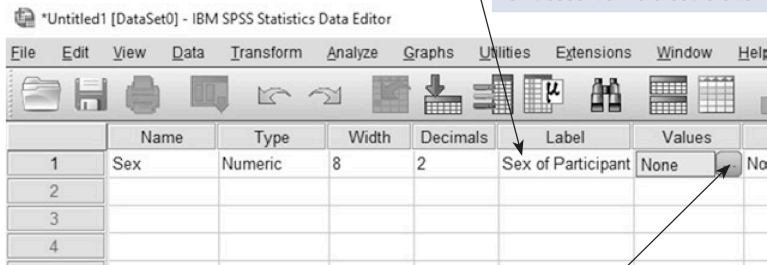- Defaults are usually sufficient unless specific formatting is required.
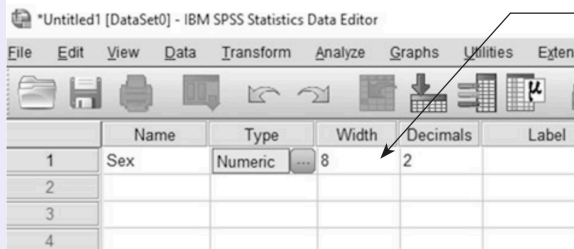
You can change the variable **Width** and number of **Decimals** places by changing these values. However, this only alters the look of the table, not the way the values are stored or used in calculations.

2. Defining Variables (in Variable View)
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.

*Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
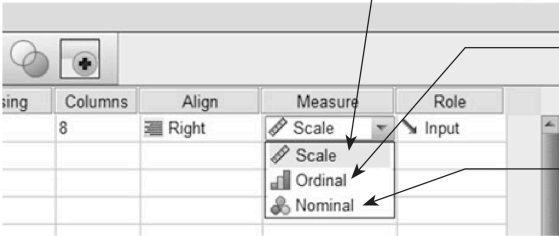
	Name	Type	Width	Decimals	Label	Values	
1	Sex	Numeric	8	2	Sex of Participant	None	No
2							
3							
4							

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).

2. Defining Variables (in Variable View)
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.

Value Labels ✕

Value Labels

Value: 2   Spelling...

Label: Female

1.00 = "Male"

Add
Change
Remove

OK Cancel Help

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.

2. Defining Variables (in Variable View)

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).

2. Defining Variables (in Variable View)

Measure

- This is used to specify the level of measurement for the variable.
- SPSS offers three options: Nominal, Ordinal and Scale
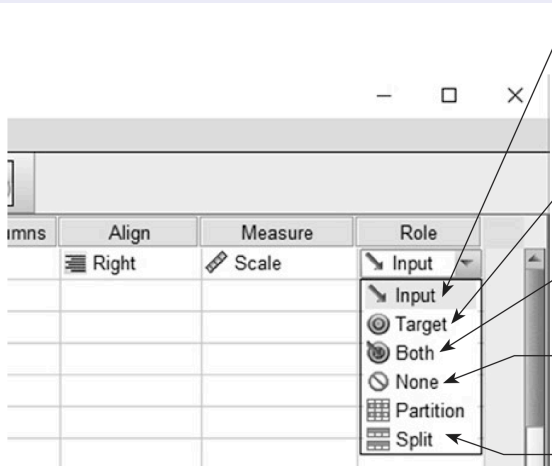
Select the **Scale** option for variables measured using either an interval or ratio scale.

Select the **Ordinal** option for variables measured using an ordinal scale.

Select the **Nominal** option for nominal variables (e.g. 'sex' or 'group').

2. Defining Variables (in Variable View)
Role

- The final column **Role** is recent addition to SPSS and is intended to help users select appropriate analyses.

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).

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.

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

File   Edit   View   Data   Transform   Analyze   Graphs   Utilities   Extensions   Window   Help

	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).

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.
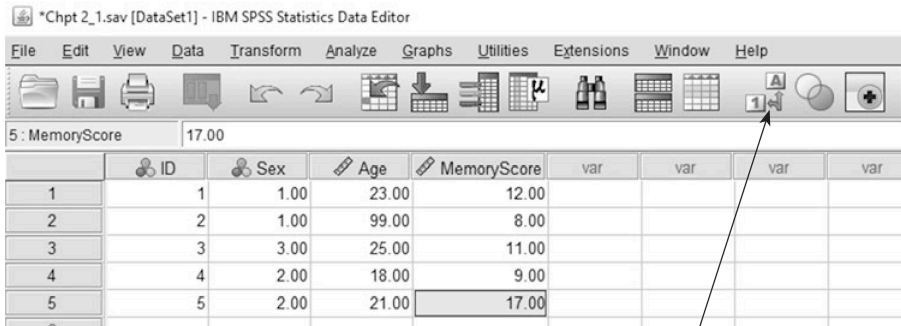
The Value Labels Button

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

*Chpt 2_1.sav [DataSet1] - IBM SPSS Statistics Data Editor
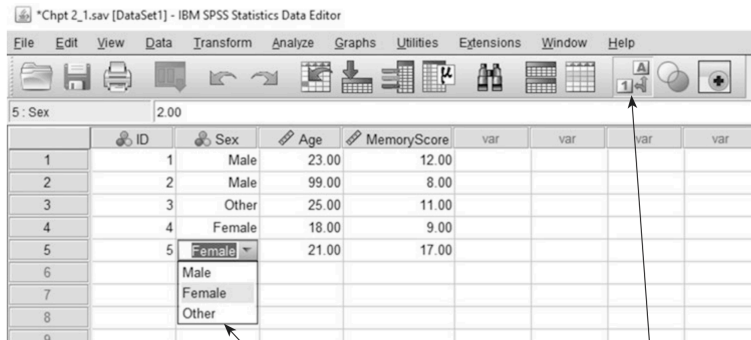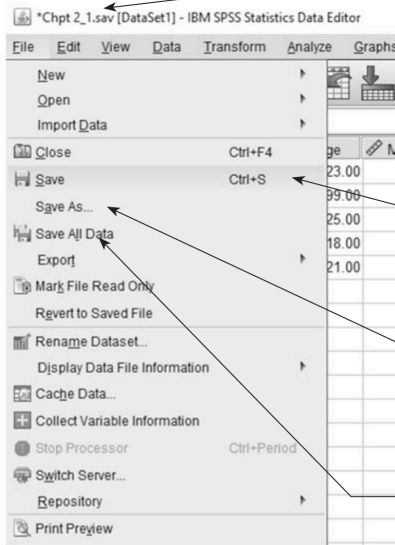
In this mode, if you double-click in a cell, SPSS will offer you a drop-down list of the value labels associated with this variable.

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

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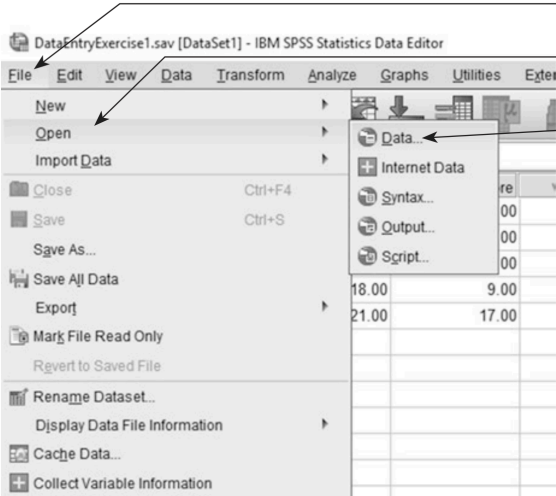
Once you save the file, your new name will appear here. We have called this file 'Chpt2_1'. SPSS adds the default suffix '.sav' to indicate a data file.

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.

5. Opening a Data File

DataEntryExercise1.sav [DataSet1] - IBM SPSS Statistics Data Editor

| File | Edit | View | Data | Transform | Analyze | Graphs | Utilities | Exte |

New

Open

Import Data

Close   Ctrl+F4

Save   Ctrl+S

Save As...

Save All Data

Export

Mark File Read Only

Revert to Saved File

Rename Dataset...

Display Data File Information

Cache Data...

Collect Variable Information

Data...

Internet Data

Syntax...

Output...

Script...

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

3. Select **Open**.

4. Choose **Data**.

re

00

00

00

18.00   9.00

21.00   17.00

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:

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

6. Data Entry Exercise

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.
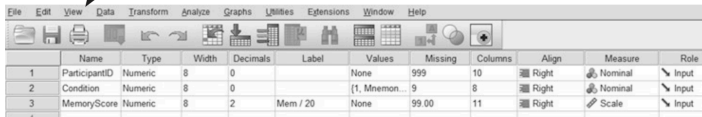
We have used the name 'Condition' for this variable. Because it is a nominal variable, we have used value labels to show that the value 1 indicates the participant was in the mnemonic condition, and the value 2 that they were in the non-mnemonic condition (see screenshot of the Value Labels dialogue 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.
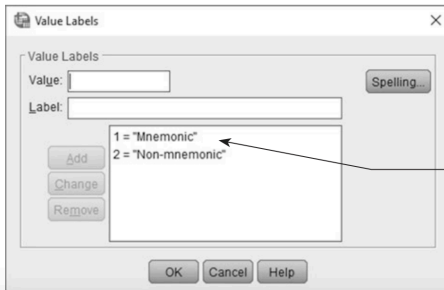
We have included a variable called 'ParticipantID', which gives each participant a number. This is not essential but is good practice; we will explain why later. We have also increased the column width so the variable name fits in.

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.

File   Edit   View   Data   Transform   Analyze   Graphs   Utilities   Extensions   Window   Help

	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											

Value Labels ✕

Value Labels

Value: [   ]   [Spelling...]

Label: [   ]

[Add]   1 = "Mnemonic"
[Change]   2 = "Non-mnemonic"
[Remove]
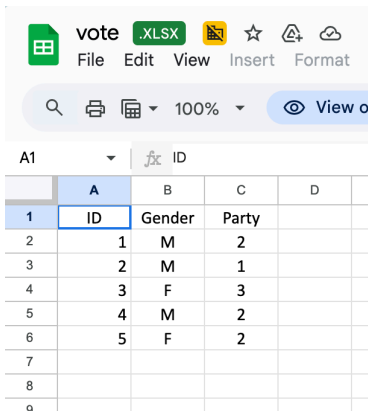
These are the value labels we used for the variable 'Condition'.

[OK] [Cancel] [Help]

Subsection 2

2.2 Data Import

Importing Excel Data

- Suppose we have an Excel file `vote.xlsx` (download) containing data like the following:

Importing Excel Data

Steps

1. Open SPSS and go to `File > Open > Data`.
2. A new window will appear asking you to select the **File Type**.
3. Choose **Excel (*.xls, .xlsx)** from the drop-down menu.
4. Browse to the folder where the file is saved and select your file.
5. Click **Open**.
6. A new window will appear as shown below:
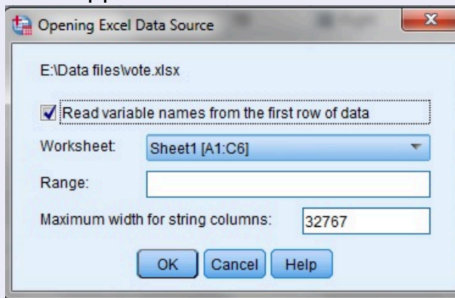
Importing Excel Data

Selecting Data from the File

- In the import dialog box:
  - ▸ Select the appropriate **worksheet** (if the Excel file has multiple sheets).
  - ▸ By default, SPSS imports the **entire sheet**.
  - ▸ To import only a portion of the data, specify a **Range** (e.g., `A1:B6`).

Example: To import only the variables *ID* and *Gender* from the first 6 rows, enter `A1:B6` in the Range field.

Importing Excel Data

Practice

- Try importing the following files:
    - `hs0.sav` (download)
    - `hs0.xls` (download)

Reflection Question

- What are the **advantages of importing a `.sav` file** instead of an Excel file?
  (Hint: Consider variable names, labels, formats, and missing value handling.)

Importing Other Data (Stata, SAS, etc.)

- Besides Excel, SPSS can directly open datasets created in **other statistical software** such as:
  - **Stata** (.dta)
  - **SAS** (.sas7bdat)
  - **SPSS portable files** (.por)

This feature is useful when collaborating with researchers who use different software.

Importing Other Data (Stata, SAS, etc.)

General Steps

1. Go to `File > Open > Data`.
2. From the **File Type** drop-down list, choose the relevant format (e.g., Stata, SAS).
3. Browse to the folder where the dataset is saved.
4. Select the file and click **Open**.
5. The dataset will appear in SPSS with variable names, labels, and some metadata preserved.

Importing Other Data (Stata, SAS, etc.)

Practice

- Try importing a dataset from another software, e.g.,
    - `bdhs2014.dta` (download) from Stata or
    - `airline.sas7bdat` (download) from SAS.

Summary

- **Excel files**: Good for raw data entry and sharing, but limited in metadata.
- **Stata / SAS files**: Retain more metadata and are useful for collaboration.
- **SPSS (.sav)**: Native format with full compatibility (best for long-term work).