

# Introduction to SPSS

---

*Dr Wan Nor Arifin*

*Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.*

*wnarifin@usm.my*



*Wan Nor Arifin, 2015. Introduction to SPSS by Wan Nor Arifin is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>.*

## Outlines

---

### Introduction

#### Data Editor

*Data View*

*Variable View*

*Menus*

*Shortcut Buttons*

#### Viewer

*Menus*

*Shortcut Buttons*

#### Practical Session

*Data Entry*

*Basic Data Analysis*

*Saving Dataset and Outputs*

## Introduction

---

- Stands for Statistical Package for the Social Sciences. Later, it is known as Statistical Product and Service Solutions.
- SPSS is a widely used statistical analysis application in the world.
- Commonly used statistical analysis software in Medical School (version 22).
- Easy to use with use of menu to perform statistical analysis vs command line (e.g. STATA, R).
- Shallow learning curve for simple statistical analysis.
- In general, the interface can be divided into two main parts: Data Editor and Viewer.

## Data Editor

---

- It consists of two tabs: Data View (Figure 1) and Variable View (Figure 2).

### *Data View*

- It is used for data entry.
- Columns consist of variable names.
- Rows consist of row number (not subject ID).

- One particular row represents one subject/patient/participant/case.
- One particular column represents characteristic/value of a variable for all subjects.
- Note in Figure 1, for column *age*, it represents age for subject number 1 to 7.

	age	marital	address	income	inccat	car	carcat	ed	employ
1	55	1	12	72.00	3.00	36.20	3.00	1	23
2	56	0	29	153.00	4.00	76.90	3.00	1	35
3	28	1	9	28.00	2.00	13.70	1.00	3	4
4	24	1	4	26.00	2.00	12.50	1.00	4	0
5	25	0	2	23.00	1.00	11.30	1.00	2	5
6	45	1	9	76.00	4.00	37.20	3.00	3	13
7	42	0	19	40.00	2.00	19.80	2.00	3	10

Figure 1: Data View

### Variable View

- It allows detailed specification of each variable.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	age	Numeric	4	0	Age in years	None	None	8	Right
2	marital	Numeric	4	0	Marital status	{0, Unmarrie...	None	8	Right
3	address	Numeric	4	0	Years at curren...	None	None	8	Right
4	income	Numeric	8	2	Household inco...	None	None	8	Right
5	inccat	Numeric	8	2	Income categor...	{1.00, Under...	None	8	Right
6	car	Numeric	8	2	Price of primary...	None	None	8	Right
7	carcat	Numeric	8	2	Primary vehicle...	{1.00, Econ...	None	8	Right
8	ed	Numeric	4	0	Level of education	{1, Did not c...	None	8	Right
9	emplnv	Numeric	4	0	Years with curr	None	None	8	Right

Figure 2: Variable View

- In this view, it has 11 columns for detailed specification of a variable (Figure 3).

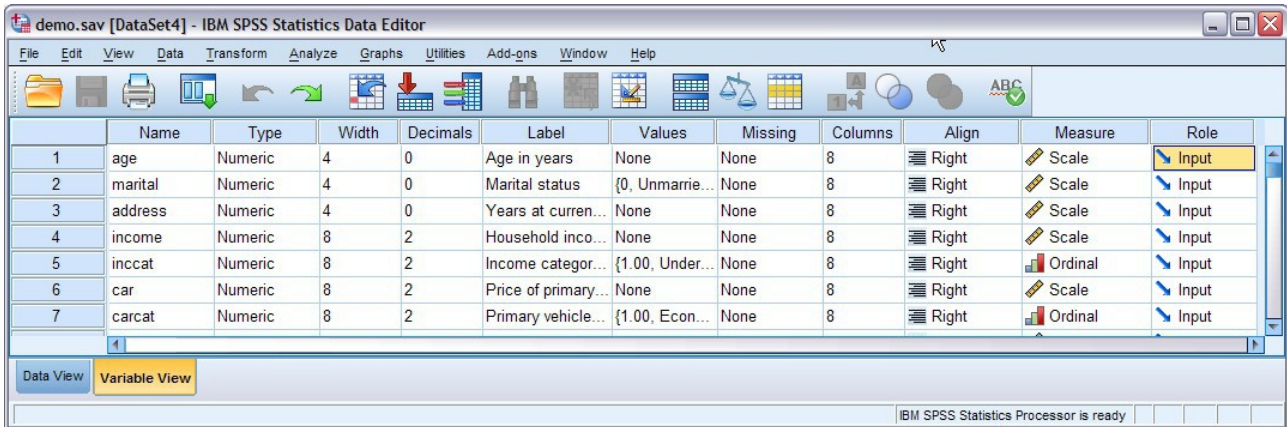


Figure 3: Variable View columns.

- **Name:** Variable name.

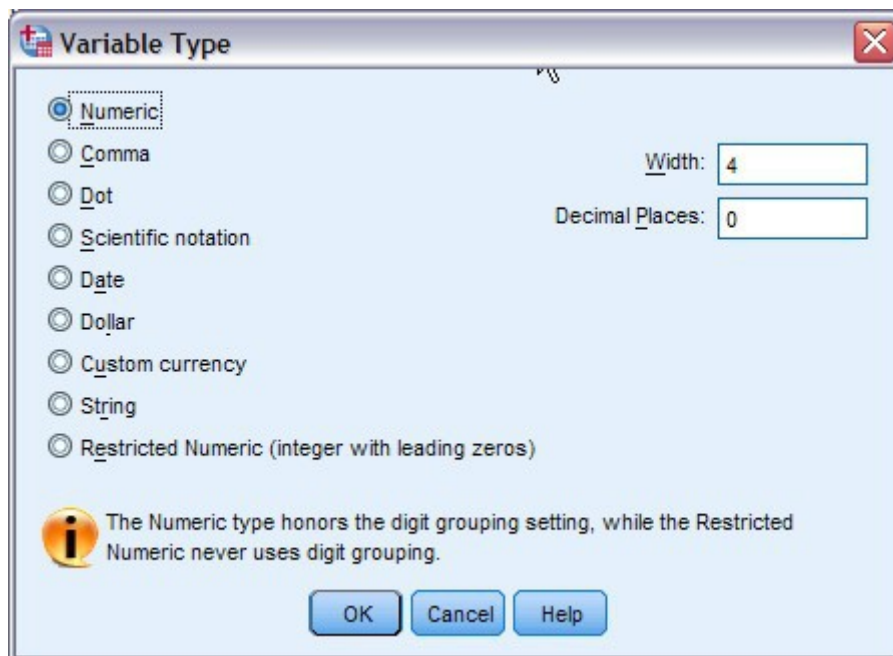


Figure 4: Variable type.

- **Type** (Figure 4): Type of variable. Of our concerns are **Numeric**, **String** and **Date**:
  - **Numeric** – Used for numerical variable and also categorical variable.
  - **String** – For variable that is meant for identification of subject only, not for analysis (e.g. name, ID).
  - **Date** – May use this variable type to enter date data, useful for conversion to other numerical data (e.g. age, time elapsed etc).
- **Width:** Number of characters visible, e.g. “abcde” string has a width of 5. “1000” numeric has a width of 4.
- **Decimals:** Number of decimal places visible.

- **Label:** Label for a variable. In this column you can write the description of each variable in length. It also serves as reminder for you and also when other people analyze your data.
- **Values:** Value labels/categories for categorical data. In statistics, categories are often given values/numbers, e.g. **Male = 1, Female = 2.**
- **Missing:** Special value for missing data/blank response/unavailable data. Assign a value that is out of range of data for a particular variable, e.g. age, set 999 as missing value as it is impossible for anyone to reach that age. Set as discrete value or as range.
- **Columns:** Width of respective column for a variable (unit is the character width).
- **Align:** Alignment for the data.
- **Measure:** Scale for numerical data, Nominal (name) for nominal categorical data, and Ordinal (ordered) for ordinal categorical data.
- **Role:** Default is **Input** for independent variable. You may specify a variable as **Target** for dependent variable. Usually left at default.

### Menus

- There are large number of options to choose from each of the main menu (Figure 5). The most useful/commonly used ones are highlighted here.

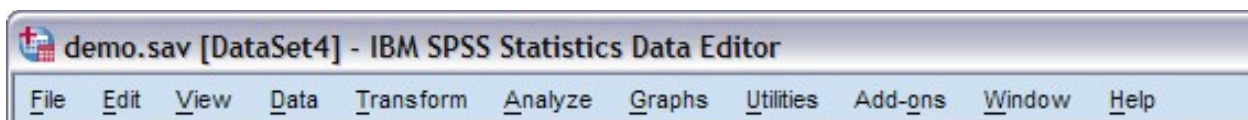


Figure 5: Menus.

### File

- **File** → **New** → **Data:** Create new blank data sheet.
- **File** → **Open** → **Data:** Open existing data. Note the formats compatible with SPSS (e.g. .xls, .xlsx).
- **File** → **Open** → **Output:** Open saved outputs. Useful to view back analyzed results.

### Edit

- **Edit** → **Options:** A number of tabs to change the way SPSS behave. The most useful options to change are under **General** tab and **Output Labels** tab. You may change the **Variable Lists** to **Display names** instead of **Display labels** depending on your preference. It is also useful when dealing with large number of variables with long labels. Similar change can be done on output under **Output Labels**.

### Data

- **Data** → **Identify Duplicate Cases:** Useful to find multiple entries of same subject from a large dataset.
- **Data** → **Split File:** Contrary to its name, it is not meant as a shortcut to split the dataset into

- pieces. Useful to analyze dataset in one go by groups (in a categorical variable).
- **Data → Select Cases:** Useful to do analysis only on selected subjects only. Usually the If condition is satisfied option is used. In the subsequent dialog box, you are required to enter logic operation, e.g. **gender=1** for if gender is male (if you code male as 1). Here you can see the usefulness of coding categorical variable as number. From this option also you can save dataset consisting of selected subjects only into new dataset.

## **Transform**

- **Transform → Compute Variable:** Useful to perform calculation based on many other variable and save the result into a new variable. For example, you can create **BMI** variable from **HEIGHT** and **WEIGHT** by specifying appropriate formula. There are also large number of built-in functions and special variables to be used. Requires some amount of programming instinct/MS Excel skills to master.
- **Transform → Recode into Different Variables:** Can be used to transform a numerical variable into categorical variable, e.g. to transform age to age categories.

## **Analyze**

- This is the bread and butter of SPSS. There are a number of sub-menus under this menu meant for data analysis. You will learn about this menu in lectures for statistical analysis.

## **Graphs**

- **Graphs → Legacy Dialogs:** Contains a number of menus for graphs and charts. The most useful ones are **Histogram**, **Scatter/Dot** and **Boxplot**.

## Shortcut Buttons

- Under the menus, there are a number of useful shortcut buttons (19 by default, Figure 6).



Figure 6: Shortcut buttons.

- Starting from left,

(1)	<b>Open data document</b>	(11)	<b>Insert cases</b>
(2)	<b>Save this document</b>	(12)	<b>Insert variable</b>
(3)	<b>Print</b>	(13)	<b>Split file</b>
(4)	<b>Recall recently used dialogs</b>	(14)	<b>Weight cases</b>
(5)	<b>Undo a user action</b>	(15)	<b>Select cases</b>
(6)	<b>Redo a user action</b>	(16)	<b>Value labels</b>
(7)	<b>Go to case</b>	(17)	<b>Use variable sets</b>
(8)	<b>Go to variable</b>	(18)	<b>Show all variables</b>
(9)	<b>Variables</b>	(19)	<b>Spell check</b>
(10)	<b>Find</b>		

- Of the shortcut buttons, the most frequently used are:

**(4) Recall recently used dialogs:** The most useful shortcut button is SPSS. Recalls recent used dialogs/menus. It saves time as you do not have to go through the menus to assess the same menu again and again.

**(16) Value labels:** If you already assign/code numbers to categorical data and specify labels for each, you can toggle between the numbers and the respective labels.

## Viewer

---

- It contains the outputs/results of analysis. It also logs every actions (e.g. clicking/selecting menus) in current session.
- **Viewer** is divided into two parts. The left panel are meant for lists of outputs in tree view. The right panel is meant for the outputs/results/logs.
- Most of the menus are similar to **Data Editor**. As such you can straight away continue your analysis without going back to **Data Editor** window and view the results.

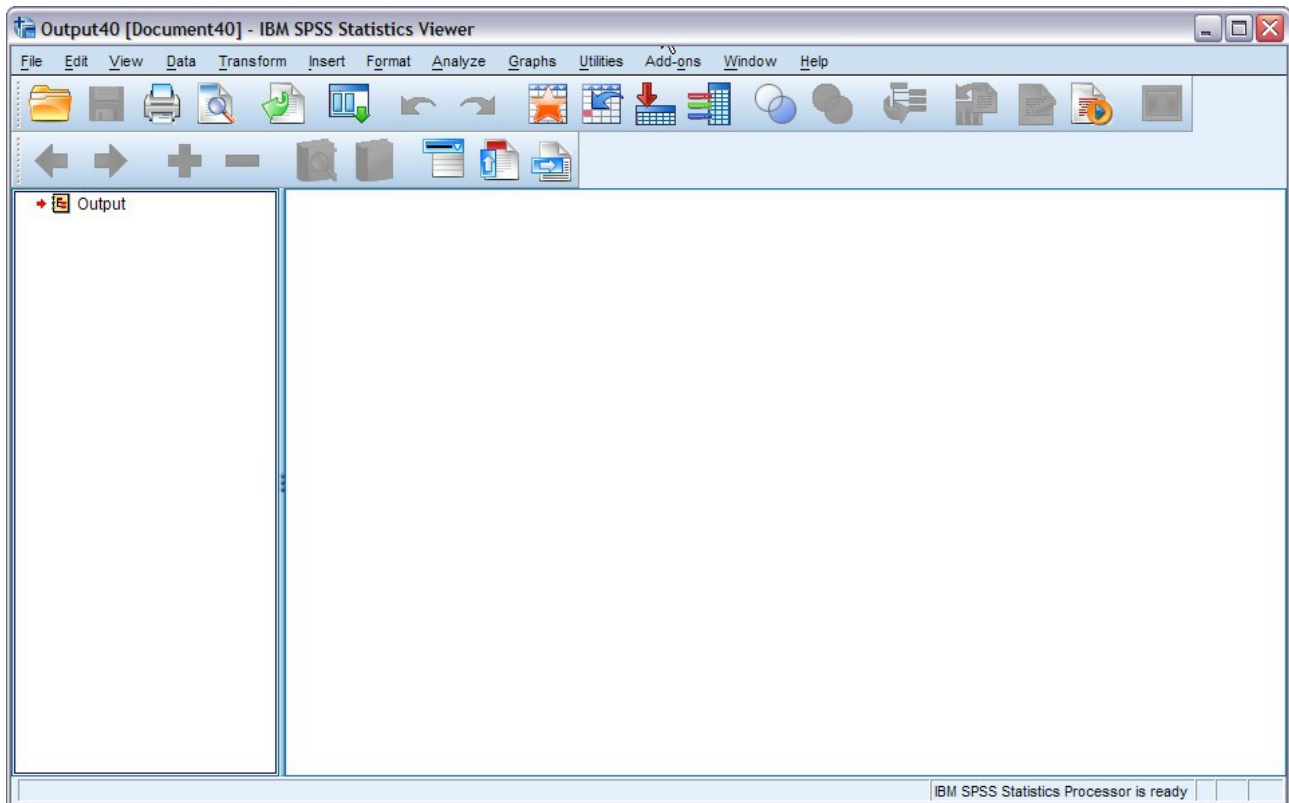


Figure 7: Viewer.

- To see some results on **Viewer**, please open demo.sav:

**File** → **Open** → **Data** → C:\Program Files\IBM\SPSS\Statistics\20\Samples\English\demo.sav

- Once you open **data.sav**, from menus:

**Graphs** → **Legacy Dialogs** → **Histogram...**

- In **Histogram** dialog box, select *Age in years (age)* on left panel and click on the arrow button next to **Variable** text box. Click on **Display normal curve**, then **OK**.

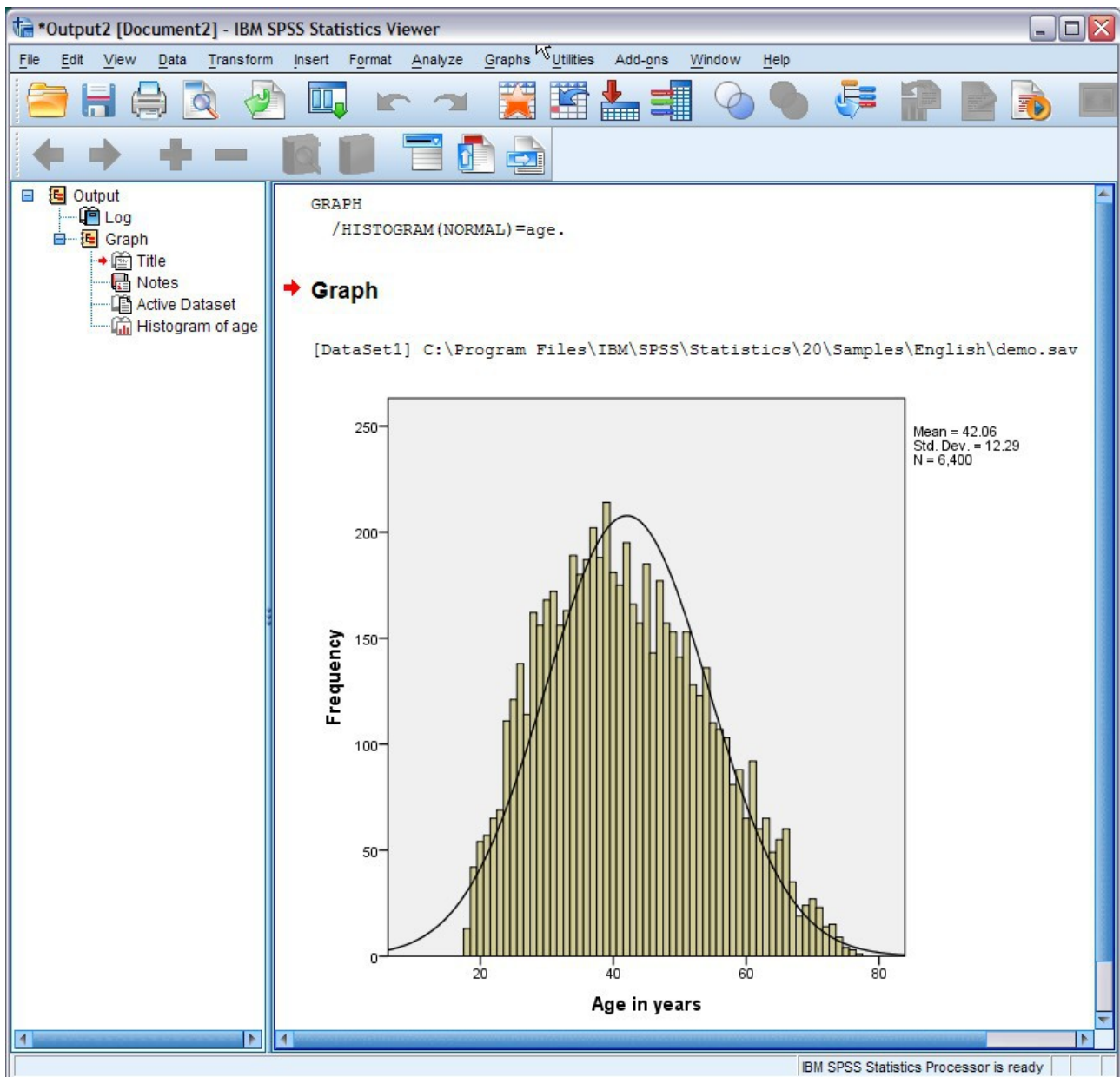


Figure 8: Example of results.

- In Figure 8, notice the tree view of the output, consisting of **Log**, **Graph**, **Title**, **Notes**, **Active Dataset** and **Histogram of age**. If you click on any of the heading on the left panel, it would highlight the respective result on the right panel.
- Notice that **Log** actually contains the SPSS syntax for the analysis we did just now (i.e. **Histogram**). **Active Dataset** part displays the active dataset we are currently using. It serves as reminder when you have multiple data opened and it is worth checking that you are doing analysis on the right dataset.



## Practical Session

---

### Data entry

- Please enter the following data into SPSS [15 minutes]:

Patient ID	Name	Gender	Age	Income (RM)	Education Level
ABC001	Abu	Male	43	9400	PhD
ABC002	Lee	Male	32	4300	Degree
ABC003	Ravichandran	Male	38	7300	Master
ABC004	Mei Lee	Female	33	4800	Degree
ABC005	Siti	Female	26	2300	Secondary School
ABC006	Letchumi	Female	37	6600	Master
ABC007	Leong	Male	31	4000	Degree
ABC008	Minah	Female	28	2600	Secondary School
ABC009	Ali	Male	20	1000	Secondary School
ABC010	Sarah	Female	34	5300	Degree

- What were the problems that you encountered while entering the data?

### Basic Data Analysis

- Frequency by gender and education levels
- Age distribution: Histogram, box-and-whisker plot
- Age: Mean, standard deviation, median, inter-quartile range
- Income distribution: Histogram, box-and-whisker plot
- Income: Mean, standard deviation, median, inter-quartile range
- Education level by gender

### Saving Dataset and Output

- Dataset: In form of .sav
- Output: In form of .spv

### References

---

IBM Corp. (2011). *IBM SPSS 20 built-in help*. Chicago IL: IBM Corp.

SPSS (2013). In *Wikipedia*. Retrieved February 18, 2013, from <http://en.wikipedia.org/wiki/SPSS>