



SPSS introduction: Data input and import

ERM I – Topic of the Week 4

Claudio Alves da Silva
clal00001@stud.uni-saarland.de

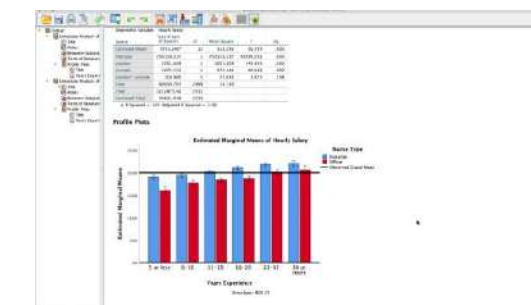
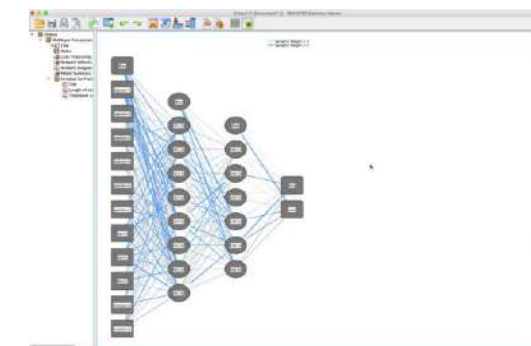
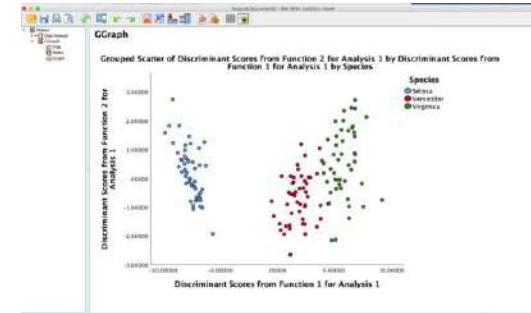



Topics covered:

- ◇ What is
- ◇ What is it used for
- ◇ How to use it
 - ◇ Importing data from Excel
 - ◇ Data view, variable view, syntax, output
 - ◇ Generating new variables
 - ◇ Defining missing values, labels, and missing values
 - ◇ Staying in control: variable names and labels
 - ◇ Non-parametric data in SPSS

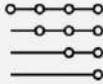
What is SPSS Statistics?

- ❖ **Statistical Package for the Social Sciences**
- ❖ Officially named "IBM SPSS Statistics"
- ❖ Current version 27 (Windows / macOS / Linux)
- ❖ [Video on how to install on Windows](#)





Easy to use


Perform powerful analysis and easily build visualizations and reports through a point-and-click interface, and without any coding experience.


Efficient data conditioning

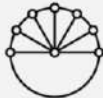
Reduce data preparation time by identifying invalid values, viewing patterns of missing data and summarizing variable distributions.


Quick and reliable


Analyze large data sets and prepare data in a single step with automated data preparation.


Comprehensive

Run advanced and descriptive statistics, regression and more with an integrated interface. Plus, you can automate common tasks through syntax.


Open source integration

Enhance SPSS syntax with R and Python using a library of extensions or by building your own.


Data security

Store files and data on your computer rather than in the cloud with SPSS that's installed locally.

Source of all images: <https://www.ibm.com/products/spss-statistics>

What is SPSS Statistics used for?

- ◇ Manipulation and statistical analysis of survey data
- ◇ In a quantitative research context:

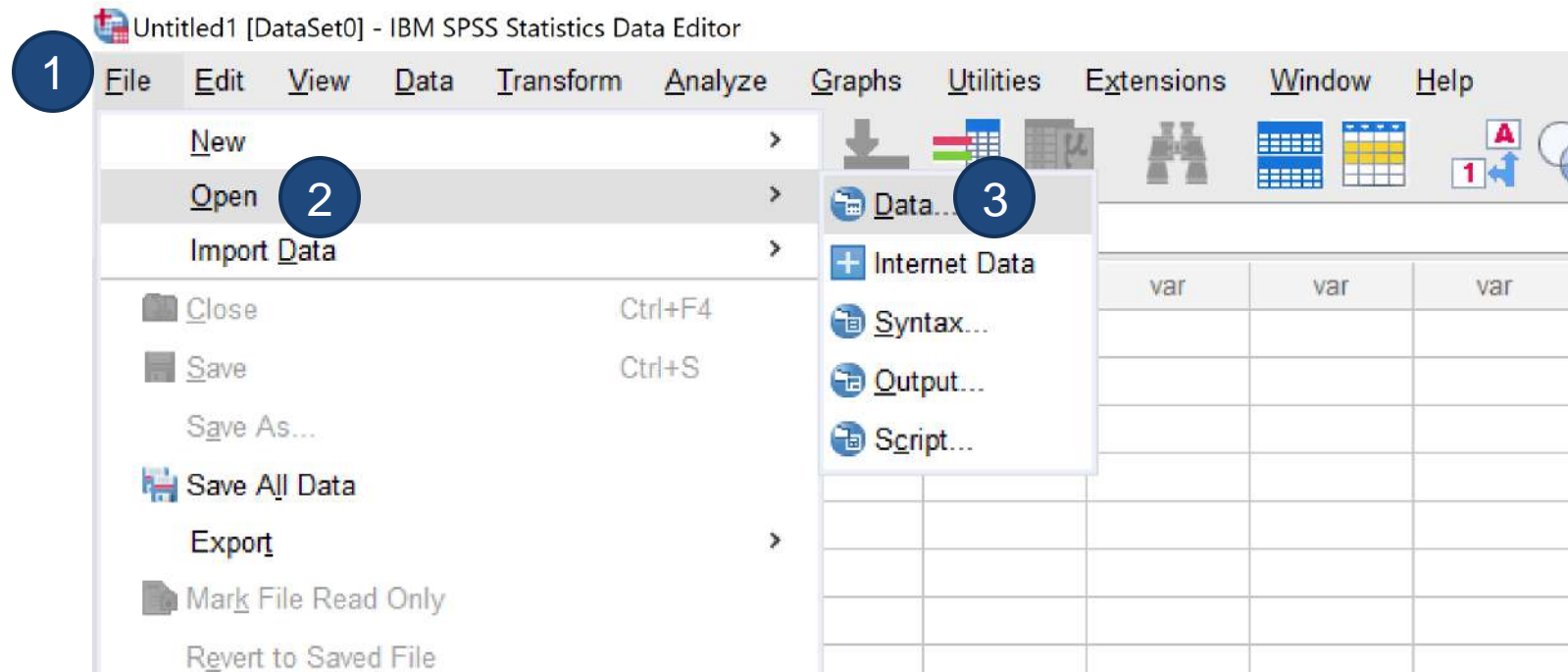


Image source: <https://www.youtube.com/watch?v=HL1H0d7IZWM>

Importing data from Excel into SPSS

◇ It is a piece of cake!

Go to **File (1) > Open (2) > Data (3)**

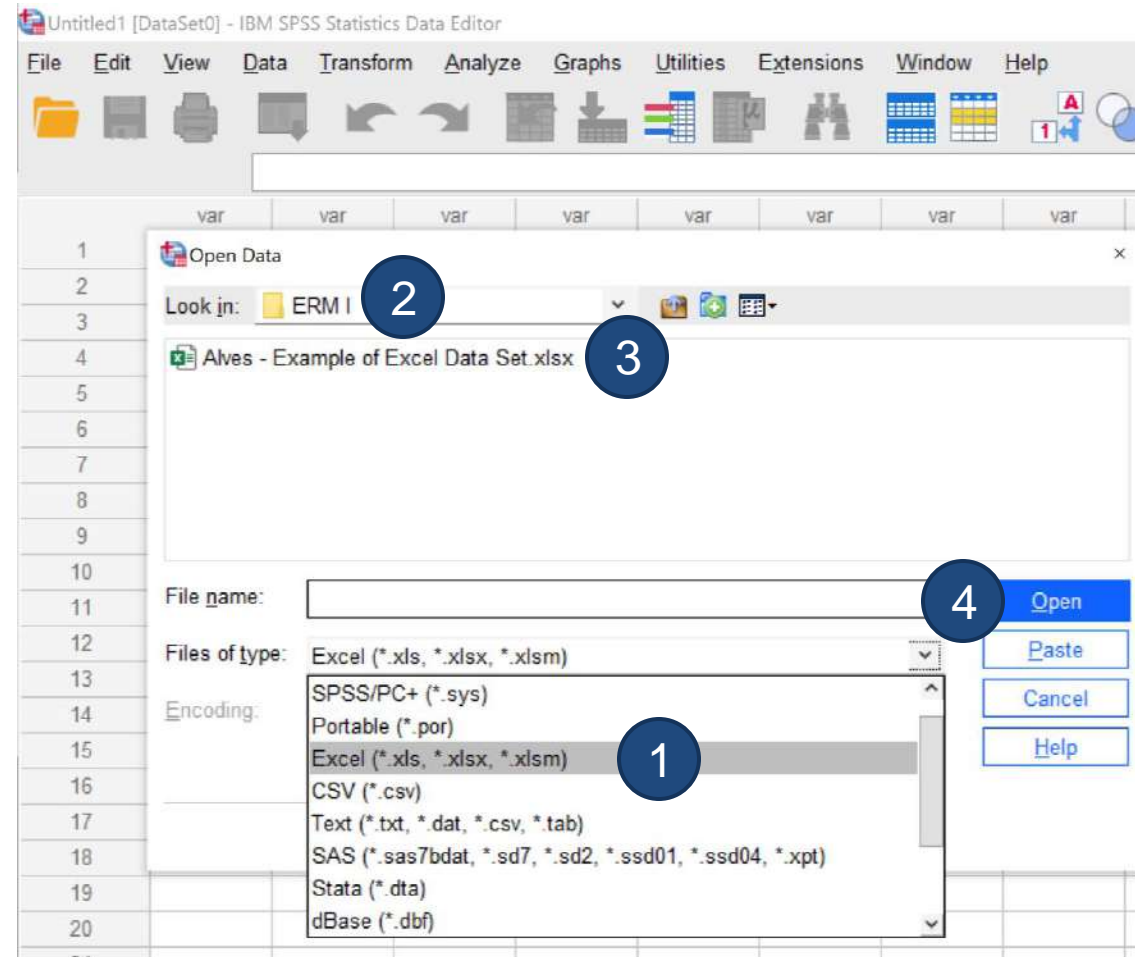


These instructions are validated for SPSS Version 25 to 27.

Importing data from Excel into SPSS

In the “Open Data” window, select the “Excel” option* (1) then navigate to the directory where the file is located (2), select the file (3) and click to open (4).

- If you do not specify the type of file that you wish to open, your file may not appear in the list of available files.



Importing data from Excel into SPSS

In the Read Excel File window choose the worksheet that contains your data* (1).

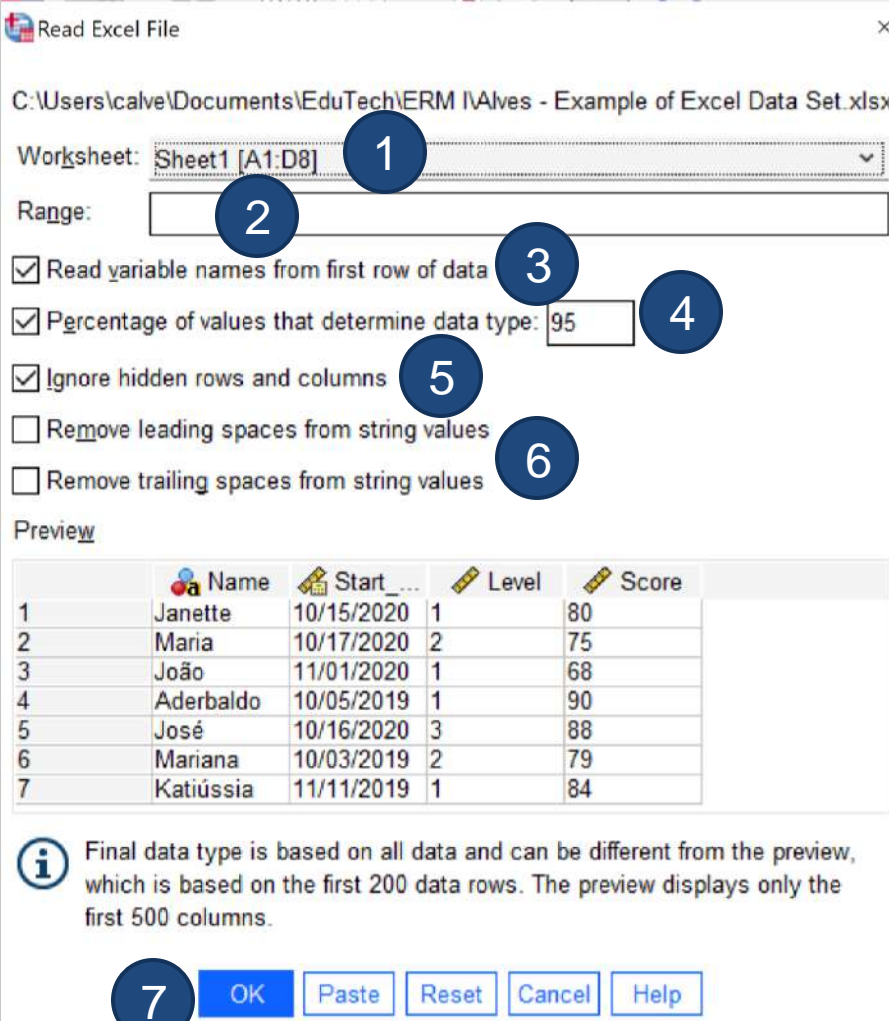
You may specify the range of rows / columns to import (2).

Select (3) if your variable names are in the first row of data, and (4) if you want SPSS to assign a data type to your variables**, you can also ignore hidden data (5).

The options to remove leading and trailing spaces from string values (6) removes any whitespace characters that appear at the beginning or the end of the string. Click OK (7) and the data will appear in SPSS.

* SPSS will import one sheet at a time.

** Any number above 50.



Read Excel File

C:\Users\calve\Documents\EduTech\ERM\Alves - Example of Excel Data Set.xlsx

Worksheet: Sheet1 [A1:D8] (1)

Range: (2)

Read variable names from first row of data (3)

Percentage of values that determine data type: 95 (4)

Ignore hidden rows and columns (5)

Remove leading spaces from string values

Remove trailing spaces from string values (6)

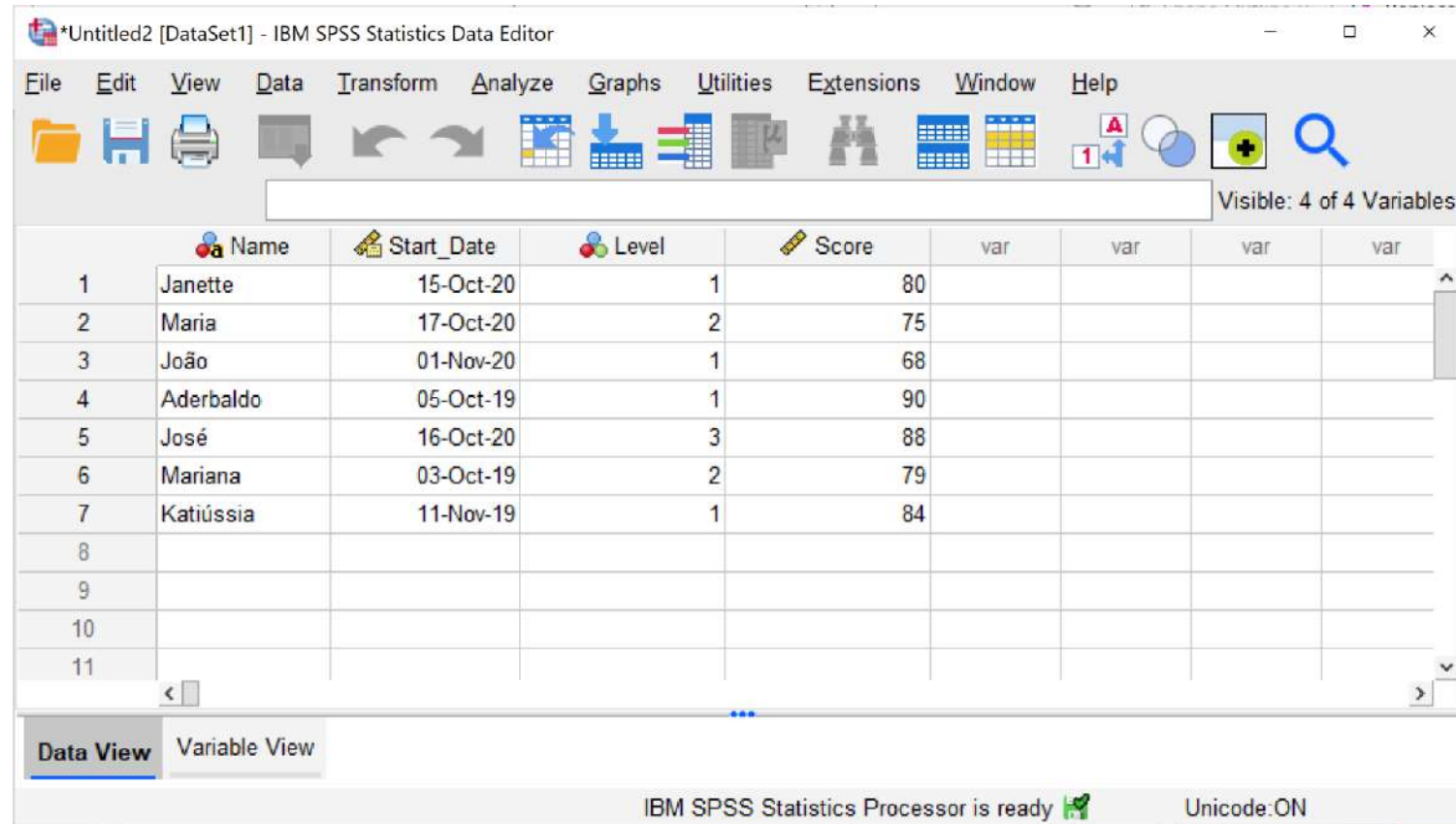
Preview

	Name	Start_...	Level	Score
1	Janette	10/15/2020	1	80
2	Maria	10/17/2020	2	75
3	João	11/01/2020	1	68
4	Aderbaldo	10/05/2019	1	90
5	José	10/16/2020	3	88
6	Mariana	10/03/2019	2	79
7	Katiússia	11/11/2019	1	84

i Final data type is based on all data and can be different from the preview, which is based on the first 200 data rows. The preview displays only the first 500 columns.

(7) OK Paste Reset Cancel Help

Importing data from Excel into SPSS



The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads '*Untitled2 [DataSet1] - 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, data manipulation, and analysis. A status bar at the top right indicates 'Visible: 4 of 4 Variables'. The main data grid is in 'Data View' and contains the following data:

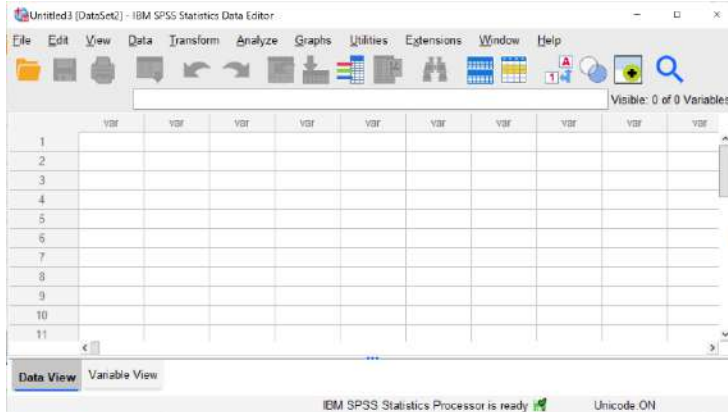
	Name	Start_Date	Level	Score	var	var	var	var
1	Janette	15-Oct-20	1	80				
2	Maria	17-Oct-20	2	75				
3	João	01-Nov-20	1	68				
4	Aderbaldo	05-Oct-19	1	90				
5	José	16-Oct-20	3	88				
6	Mariana	03-Oct-19	2	79				
7	Katiússia	11-Nov-19	1	84				
8								
9								
10								
11								

The bottom status bar indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode:ON'. The 'Data View' tab is selected at the bottom left.

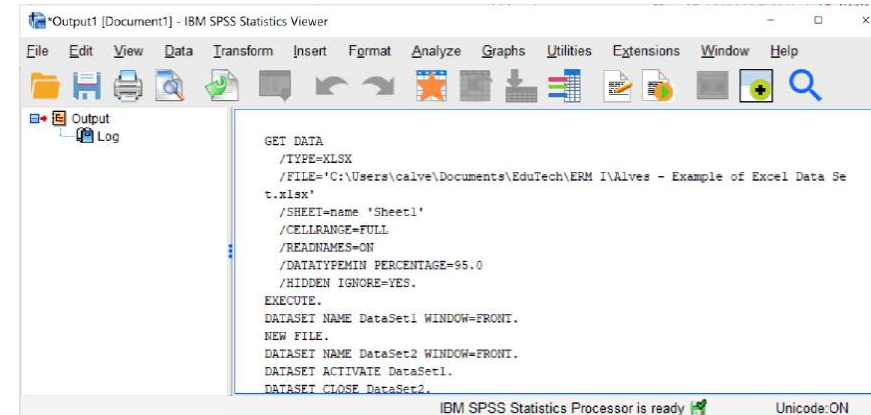
Example of how data appear in SPSS once the data have been imported.

SPSS Environment

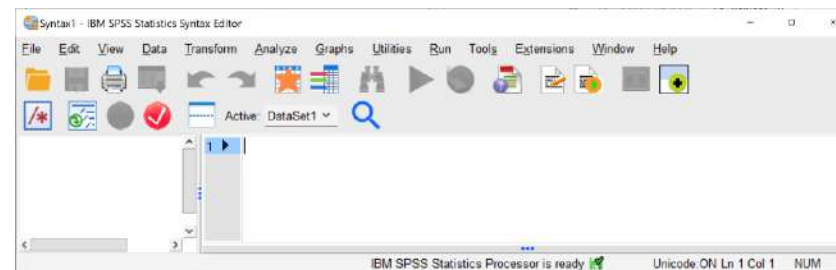
- ◇ Each SPSS “screen” is associated with specific tasks and types of SPSS files.



Data Editor



Output Editor



Syntax Editor

Data Editor

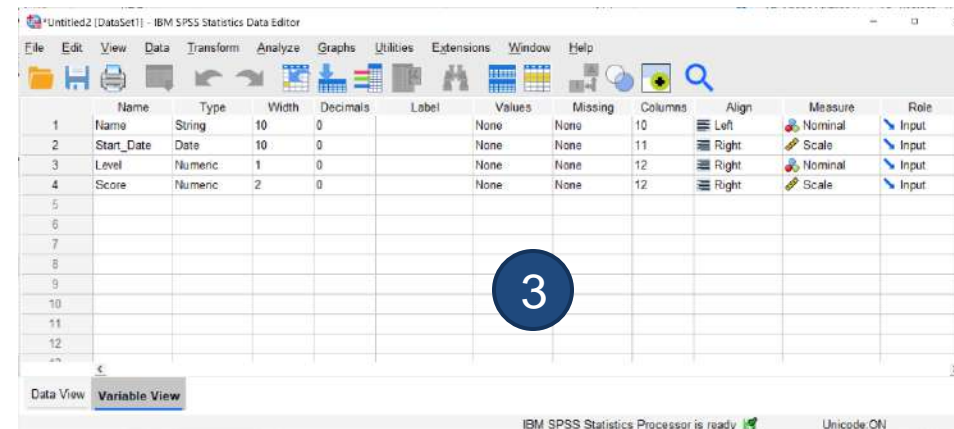
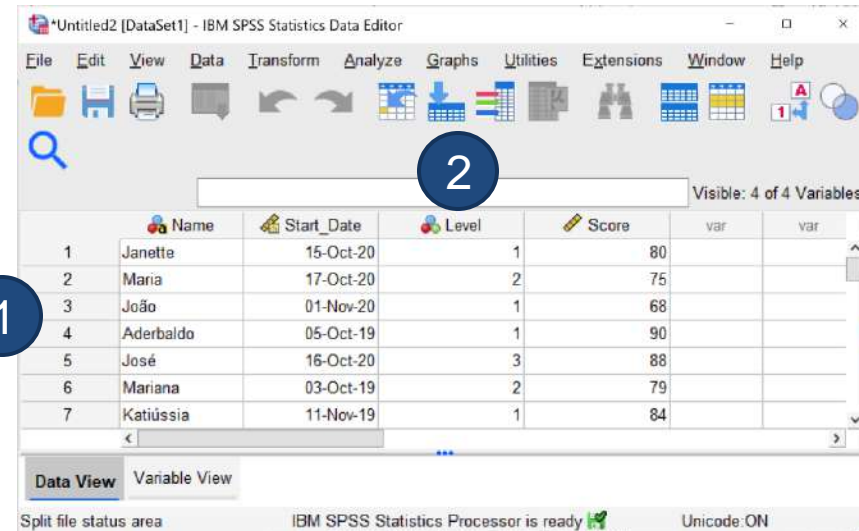
- ◇ Default window
- ◇ Displayed in spreadsheet format
- ◇ Contains 2 views: Data and Variable

Data View

- ◇ Columns represent variables (1)
- ◇ Rows represent cases (2)

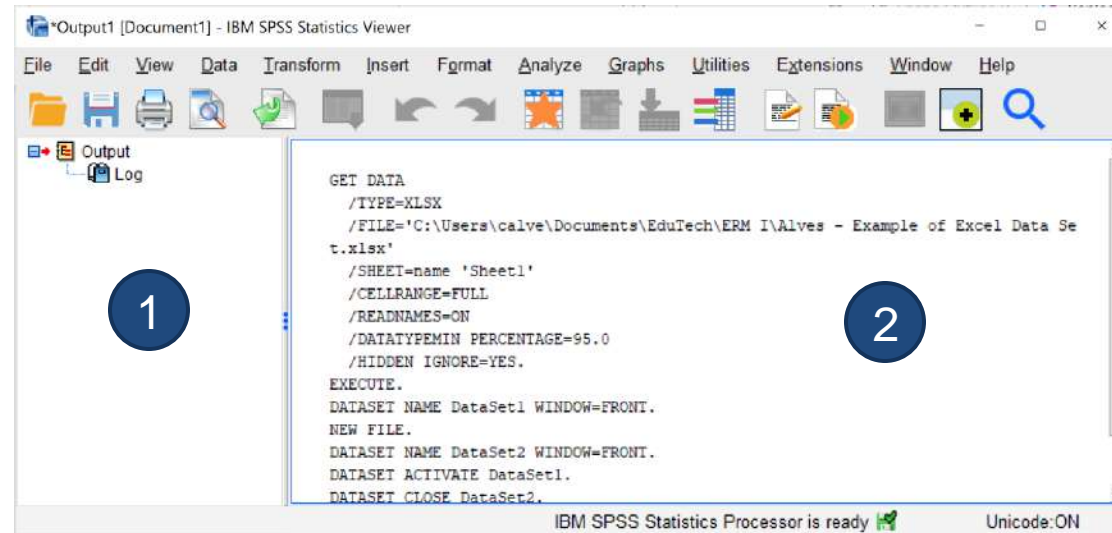
Variable View (3)

- ◇ Shows information about variables present in the open data (but not the data themselves).



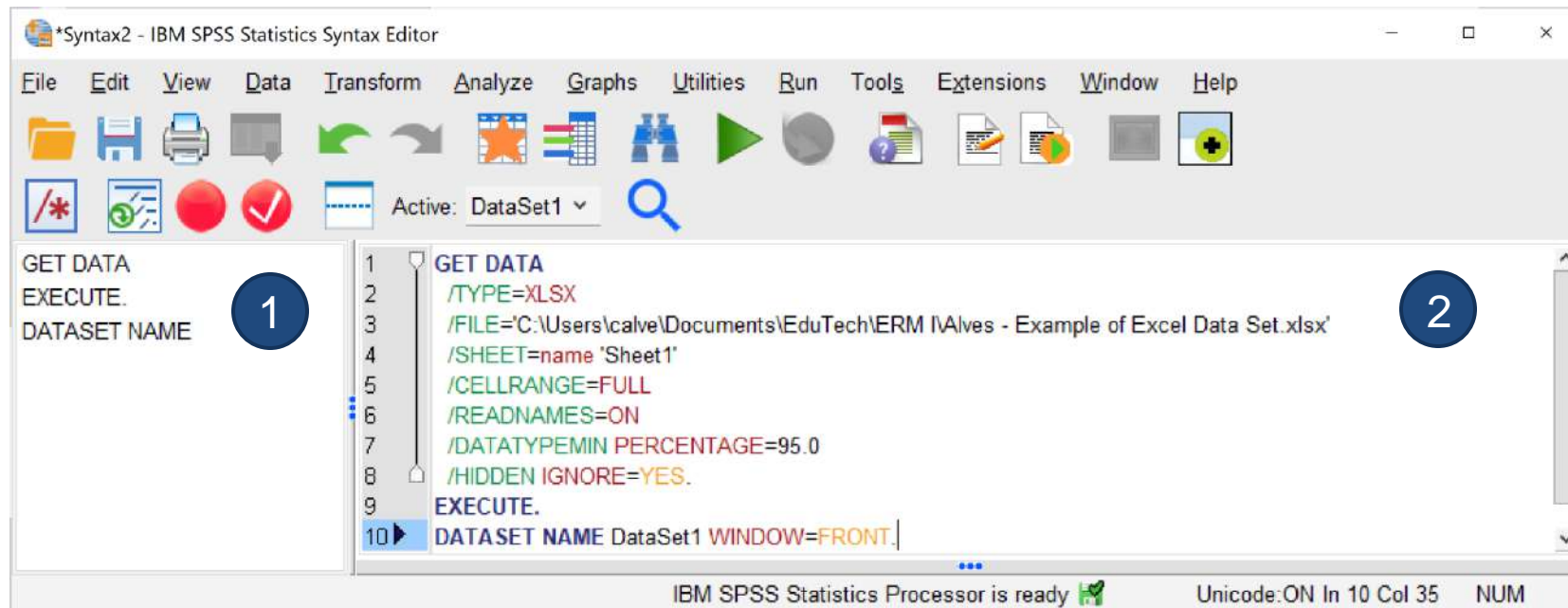
Output Viewer

- ◇ Opens automatically
- ◇ Display a log and output of the actions taken
- ◇ Shows the result of statistical analysis
- ◇ Contains the outline of the content in the viewer (1) and the actual output (2)
- ◇ The contents can be modified
- ◇ can be saved as a viewer file (*.spv) and other formats (e.g.: *.xlsx, *.doc, *.png)



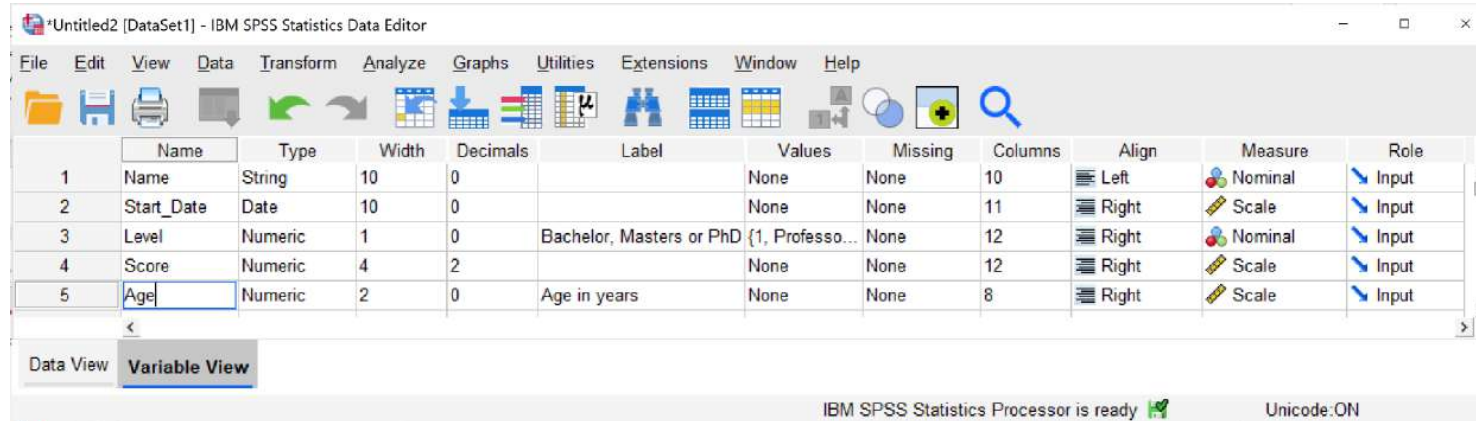
Syntax Editor

- ◇ SPSS syntax is a programming language (alternative to menus)
- ◇ Users can write, debug and run
- ◇ Shows an outline of commands (1) and the editor (2)
- ◇ Can be saved as an *.sps file



Variables

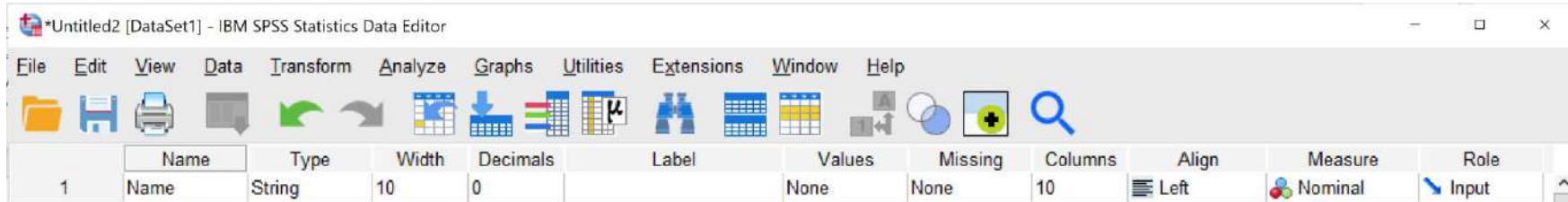
- ◇ You can manage variables using the Variable View, once you enter the variable name, all the fields will be filled with default values that you can adjust if needed.



- ◇ Best practices for naming variables in SPSS:

- ◇ Each variable name must be unique.
- ◇ Keep variable names up to 64 characters long.
- ◇ No spaces, you may include alphanumeric characters, non-punctuation characters. Also periods and underscores within (not at the end).
- ◇ You can use upper and lower cases (Camel case - e.g. typeOfcar).
- ◇ SPSS has reserved keywords, you can't use them as a variable name (i.e.: ALL, AND, BY, NOT, OR, TO or WITH).
- ◇ You can add any type of character and quantity at the "Label", but also keep it short to keep the output more "readable".

Variables



The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Extensions, Window, and Help. The toolbar contains various icons for file operations, data manipulation, and analysis. Below the toolbar is a table with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The table contains one row for a variable named 'Name'.

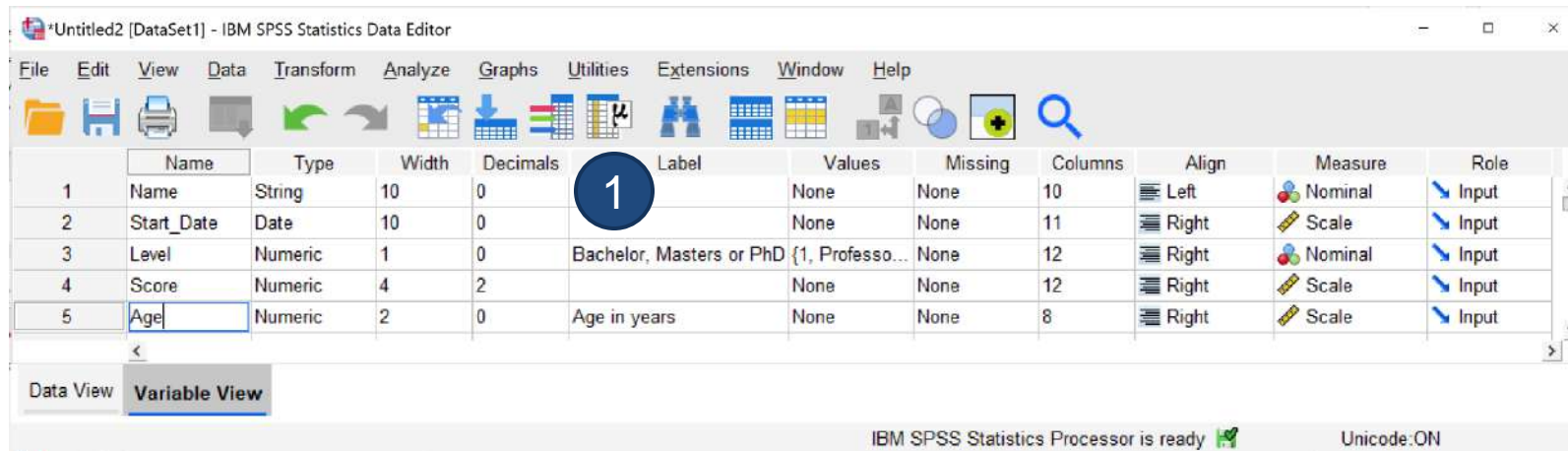
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Name	String	10	0		None	None	10	Left	Nominal	Input

- ◇ **Type:** type of variable, e.g.: string, numeric, comma, Date)
- ◇ **Width:** number of digits displayed (for numeric values) or the number of characters (for a string)
- ◇ **Decimals:** number of digits (just apply to decimals)
- ◇ **Columns:** the width of the actual column in the Data View (not the number of digits).
- ◇ **Align:** the alignment of content of the cell.
- ◇ **Measure:** level of measurement for the variable nominal, ordinal, or scale (interval or ratio). This setting affects everything from graphs to internal algorithms for statistical analysis.
- ◇ **Role:** how the variable will be used in your analysis (e.g.: input for IV or target for DV)

Variables

◇ Labels (1)

- ◇ Brief and descriptive display name for the variable.
- ◇ Labels are going to be printed in the output.



The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Variable View' tab is active, displaying a table of variables. A blue circle with the number '1' highlights the 'Label' column header.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Name	String	10	0		None	None	10	Left	Nominal	Input
2	Start_Date	Date	10	0		None	None	11	Right	Scale	Input
3	Level	Numeric	1	0	Bachelor, Masters or PhD	{1, Professo...	None	12	Right	Nominal	Input
4	Score	Numeric	4	2		None	None	12	Right	Scale	Input
5	Age	Numeric	2	0	Age in years	None	None	8	Right	Scale	Input

Variables

◇ Defining values

- ◇ For coded categorical variables E.g.: (2="Master", 3="PhD")
- ◇ Helps to understand what each value represents
- ◇ The value labels will display in the output instead of the original codes

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays the Variable View for a dataset named 'Untitled2 [DataSet1]'. The variable 'Level' is selected, and its 'Values' column is highlighted. A pink arrow points from the 'None' value in the 'Values' column to the 'Value Labels' dialog box. The dialog box is open, showing the 'Value' field set to '3' and the 'Label' field set to 'PhD'. Below these fields, there is a list of existing value labels: '1 = "Bachelor"' and '2 = "Masters"'. The 'Add' button is highlighted, and the 'Spelling...' button is visible in the top right corner of the dialog box. The 'OK', 'Cancel', and 'Help' buttons are at the bottom of the dialog box.

	Name	Type	Width	Decimals	Label	Values
1	Name	String	10	0		None
2	Start_Date	Date	10	0		None
3	Level	Numeric	1	0	Bachelor, Masters or PhD	None
4	Score	Numeric	4	2		None
5	Age	Numeric	2	0	Age in years	None

Variables

◇ Defining values

- ◇ After defining values, in the Data View (1), you can use the option Value Labels (2) to visualize the corresponding level.

IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window

Visible: 5 of 5 Variables

	Name	Start_Date	Level	Score	Age	var	var
1	Janette	15-Oct-20	Bachelor	80.00	.		
2	Maria	17-Oct-20	Masters	75.00	.		
3	João	01-Nov-20	Bachelor	68.00	.		
4	Aderbaldo	05-Oct-19	Bachelor	90.00	.		
5	José	16-Oct-20	PhD	88.00	.		
6	Mariana	03-Oct-19	Masters	79.00	.		
7	Katiússia	11-Nov-19	Bachelor	84.00	.		

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

Variables

◇ Defining missing values

- ◇ User added value to treat as missing data
- ◇ Value that do not occur on the real data
- ◇ Numeric variables when missing are attributed a default value "."
- ◇ String variables when blank are not automatically recognized as missing values

Variables

◊ Recoding missing values in Data View

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Transform' menu is open, and 'Recode into Same Variables...' is highlighted. The data view shows a table with columns for Name, Start_Date, Level, Score, and Age. The 'Score' column has values 80.00, 75.00, and 90.00 for the first three rows.

Name	Start_Date	Level	Score	Age
Janette	15-Oct-20	1	80.00	
Maria	17-Oct-20	2	75.00	
João	01-Nov-20	1		
Aderbaldo				
José				
Mariana				
Katiússia				

The screenshot shows the 'Recode into Same Variables: Old and New Values' dialog box. The 'Old Value' section has 'System missing' selected. The 'New Value' section has 'Value: -1' selected. The 'Old -> New:' list shows 'SYSMIS -> -1'. The 'Continue' button is highlighted.

Name	Start_Date	Level	Score	Age
Janette	15-Oct-20	1	80.00	
Maria	17-Oct-20	2	75.00	
João	01-Nov-20	1		
Aderbaldo				
José				
Mariana				
Katiússia				

Variables

◇ Defining missing values in the Variable view

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays the Variable View for a dataset named 'Untitled2 [DataSet1]'. The variable 'Score' is selected, and the 'Missing' column is highlighted. The 'Missing Values' dialog box is open, showing the 'Discrete missing values' option selected. The first discrete missing value is set to '-99'. The 'OK' button is highlighted.

	Name	Type	Width	Decimals	Label	Values	Missing
1	Name	String	10	0		None	None
2	Start_Date	Date	10	0		None	None
3	Level	Numeric	1	0	Bachelor, Masters or PhD	{1, Bachelor...	None
4	Score	Numeric	4	2		None	None
5	Age	Numeric	2	0	Age in years	None	None

IBM SPSS Statistics Processor is ready Unicode:ON

Level of measurement and roles in SPSS

- ◇ Level of measurement (1) for the variable: nominal, ordinal, or scale (interval or ratio). Important! as SPSS may treat numeric as "Scale" by default and this info is vital to analyze your data.
- ◇ Set the role (2) the variable plays in you analyzes: Input (IV), Target (DV), Both, None, etc.

The screenshot displays the IBM SPSS Statistics Data Editor interface. The main window shows a list of variables with their types and widths. A 'Variables' dialog box is open, showing the 'Variable Information' for the selected variable 'Name'. The 'Measure' is set to 'Nominal' and the 'Role' is set to 'Input'. A table on the right side of the dialog shows the measurement levels and roles for several variables, with '1' and '2' circled in blue to indicate the focus of the slide.

Measure	Role
Nominal	Input
Scale	Input
Nominal	Input
Scale	Input
Scale	Input

Non-parametric data in SPSS

- Work with your data when it is in form of ranks, small sample sizes, etc.

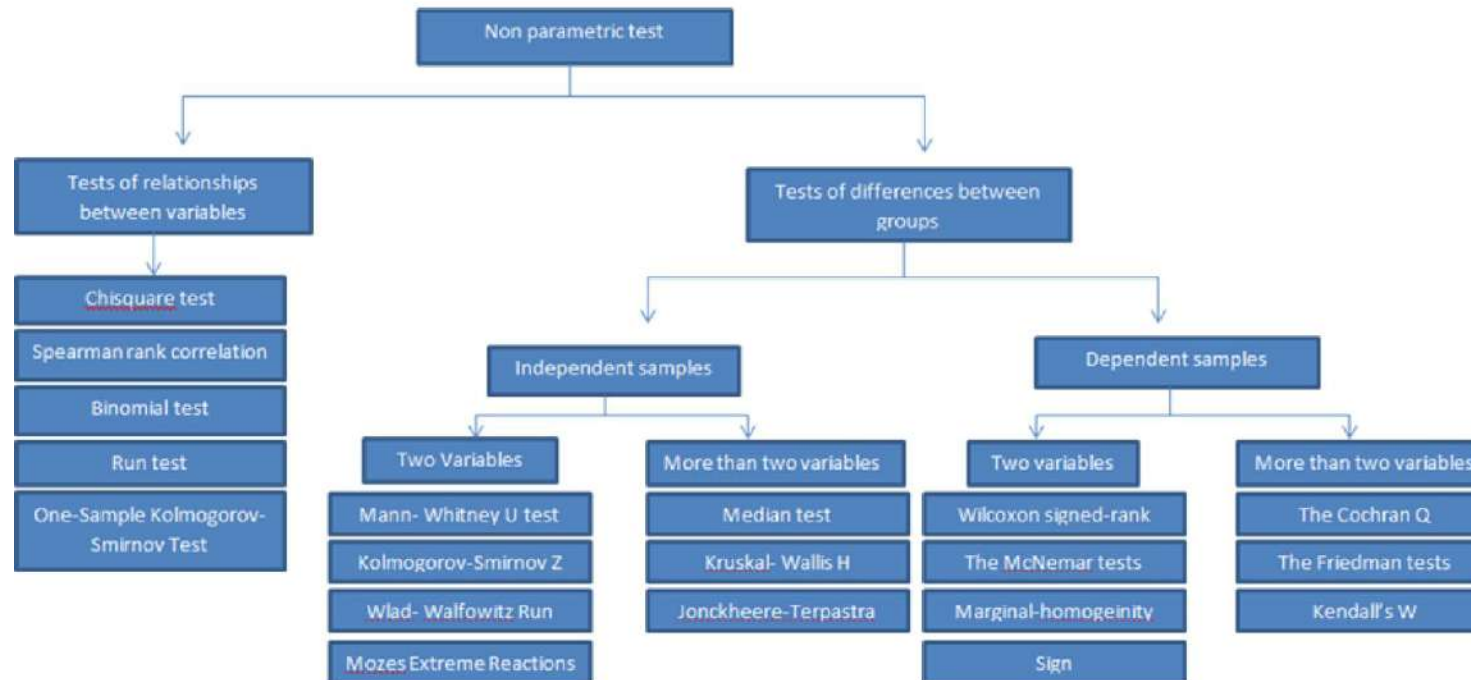
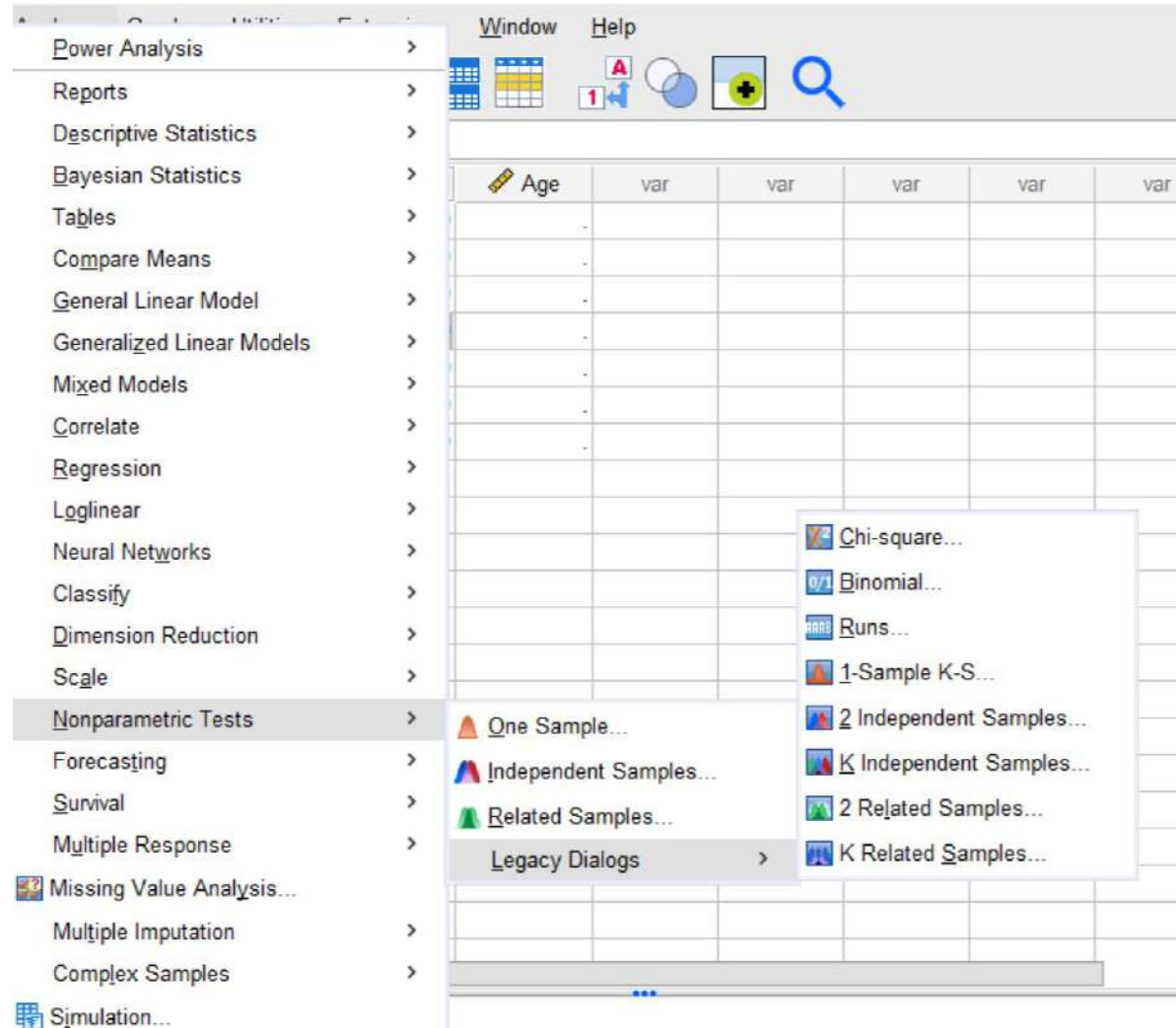


Image source: http://www.kiran.nic.in/pdf/Social_Science/e-learning/Non_Parametric_Test.pdf

Non-parametric data in SPSS



References

EZ SPSS Tutorials (n.d.). *Importing Data into SPSS from Excel*. Retrieved from <https://ezspss.com/importing-data-into-spss-from-excel/>

Field, A. (n.d.). *Importing Data into SPSS from Excel*. YouTube. Retrieved from <https://www.youtube.com/watch?v=b163iBByycw>

IBM. (n.d). IBM SPSS Statistics 27 documentation. Retrieved from https://www.ibm.com/support/knowledgecenter/SSLVMB_27.0.0/

Kent State University Libraries. (2017). SPSS tutorials. Retrieved from <https://libguides.library.kent.edu/SPSS/home>

TheRMUoHP Biostatistics Resource Channel. (n.d.). *Home [YouTube channel]*. YouTube. Retrieved from <https://www.youtube.com/watch?v=NoRXJI5WgdI>