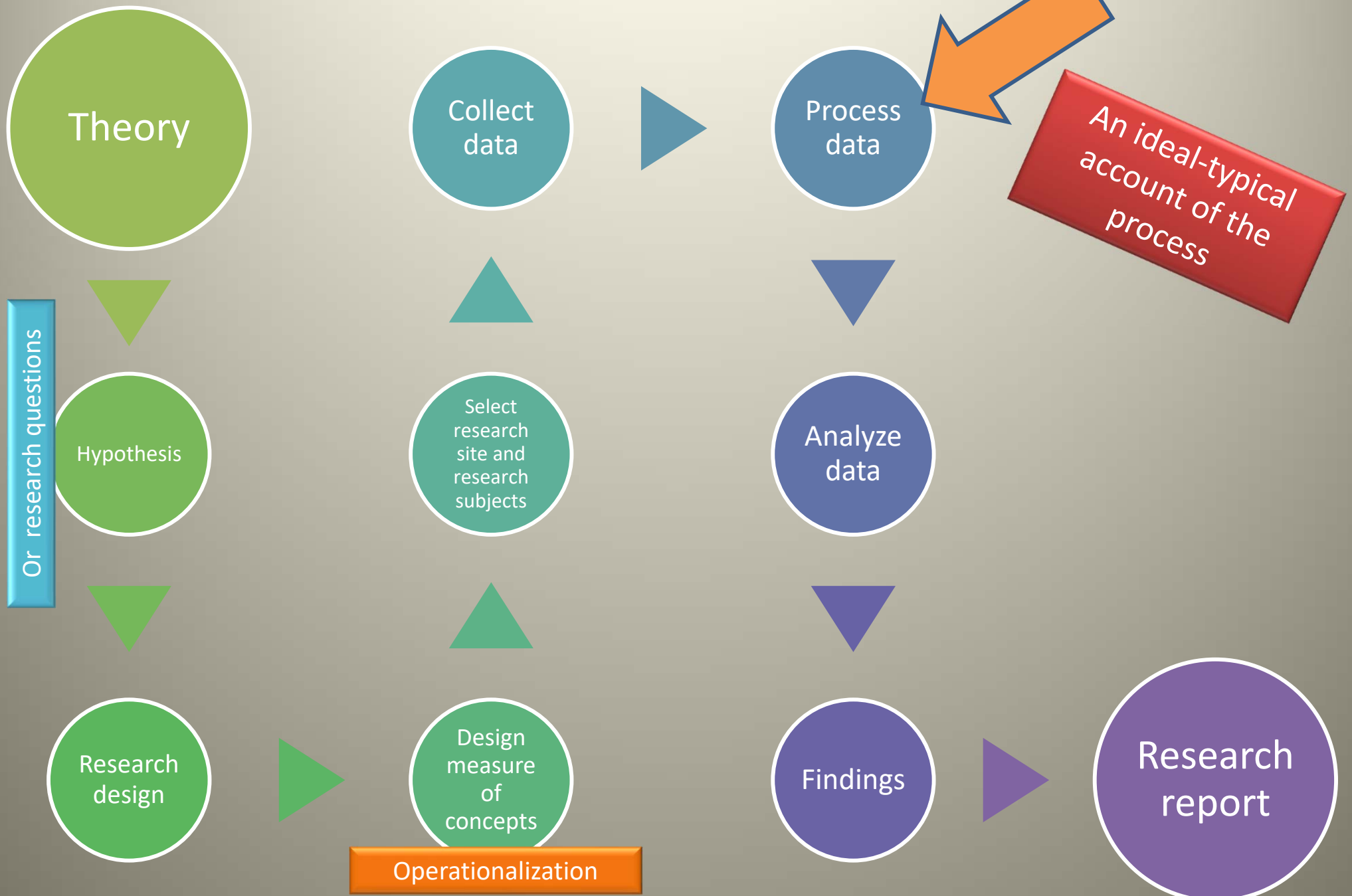# Introduction to SPSS

Processing data

Preparation for analysis

Quantitative research process

Getting started…

What is **SPSS**?

- A "statistical package for the social sciences";

- A computer programme written for personal computers;

- An intelligent database with built-in formulas for calculating correlations and degrees of significance;

- A report generator for graphs, charts and tables.

If we didn't have SPSS...

$$t = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$F = \frac{MST}{MSE}$$

$$MST = \frac{\sum_{i=1}^{k}(T_i^2/n_i) - G^2/n}{k-1}$$

$$MSE = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}Y_{ij}^2 - \sum_{i=1}^{k}(T_i^2/n_i)}{n-k}$$

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

SPSS interface

SPSS Data Editor

This is the sphere of SPSS into which data are entered, edited and defined.

Output Viewer

Analysis and diagrams are displayed deposited here.

Data Viewer

This is the spreadsheet into which your data are inserted.

Variable Viewer

This spreadsheet displays information about each of the variables.

Variable Name

The name of a variable.

No spaces between the words

If you need a space between the words, use an underscore "_"

Type

- **Numeric: Data values that are recognized as numbers. We can sort them numerically or perform calculations.**
- Comma: Numeric variables that include commas (,) that delimit every three places. Decimals are delimited by a period (.).
- Dot: Numeric variables that include periods (.) that delimit every three places. Decimals are delimited by a comma (,).
- Scientific notation: Numeric variables whose variables are displayed with an E and power-of-ten exponent.
- **Date: Numeric variables that are displayed in any standard calendar date or clock time formats.**
- Dollar: Numeric variables that contain a dollar sign before numbers.
- Custom currency: Numeric variables that are displayed in a custom currency format. The currency must be defined in the variable type window.
- **String: Character variables that are treated as text. Even if they values contains numbers SPSS won't proceed in any computation.**
- Restricted numeric: Numeric variables whose values are restricted to non-negative integers. The values are displayed with leading zeroes and with reference to the width of the variable.

Open-ended questions

Variable Label

The label of a variable

Spaces can be used	It will appear in any output you generate

Width

- The width setting in the definition of the variable determines the number of character used to display the value.
- By default is set to 8.
- For numerical variables we don't usually need to bother with width.
- For string variables we need to set up the width to the corresponding number.
- Width range is between 1-40 characters for numerical variables.

Decimals

- The number of digits to display after the decimal point for values of that variable.
- By default is set to 2.
- Does not apply to string variables.
- Any changes made to the decimals display do not affect the values in the dataset.

Value Label

The label attached to a code that has been used when entering variable data

It is used for all types of variable except interval/ratio variables

It makes the interpretation and presentation of data easier

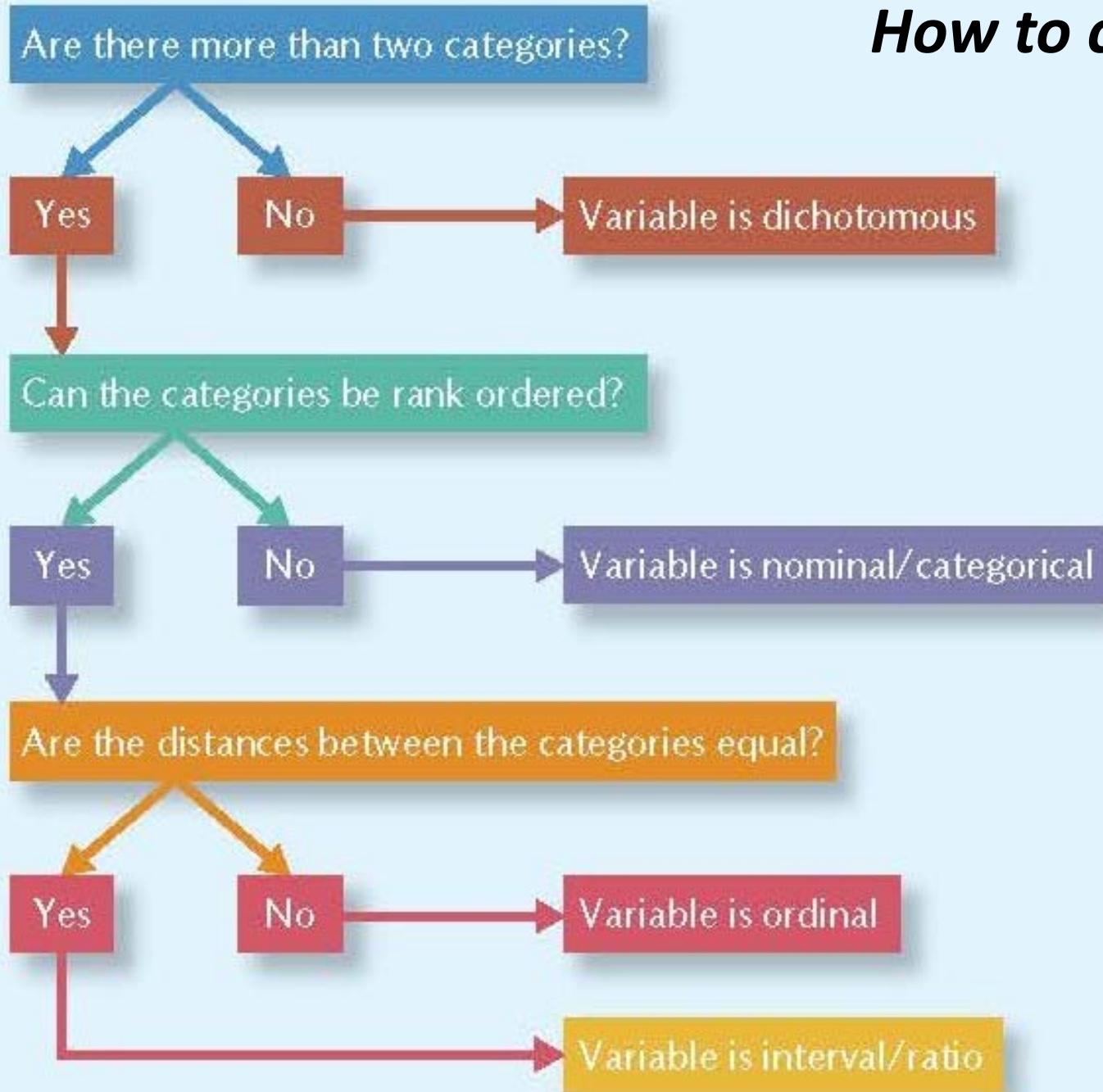Example:
Male (1)
Female (2)
Other (3)

Missing Values: When data are unavailable for a variable, you must specify how you are denoting missing values for that variable.

You will need to inform SPSS of the value that you have nominated for each variable to indicate a missing value.

In the Missing Values dialog box, enter the missing value below Discrete missing values and then click OK.

999	12	≡ Right	Nominal	↘ Input
999	12	≡ Right	Nominal	↘ Input
999	12			
None	12			
None	12			
None	12			
None	12			
None	12			
None	12			
None	12			
None	12			
None	12			
None	12			
None	12	≡ Right	Nominal	↘ Input

Missing Values

○ No missing values

◉ Discrete missing values

[999] [ ] [ ]

○ Range plus one optional discrete missing value

Low: [ ]   High: [ ]

Discrete value: [ ]

[OK] [Cancel] [Help]

Column	Align
• The width of each column in the Data View spreadsheet. It simply refers to the wideness of the actual column in the spreadsheet. • By default is set to 8.	• The alignment of content in the cells of the SPSS Data View spreadsheet. • By default is set to Right.

Measure in SPSS Variable View refers to the *different types of variable*. The four main types of variables are:

Interval/ratio variables: These are variables where the distances between the categories are identical across the range of categories. This is the highest level of measurement and a very wide range of techniques of analysis can be applied to interval/ratio variables. There is, in fact, a distinction between interval and ratio variables, in that the latter are interval variables with a fixed zero point.

Ordinal variables: These are variables whose categories can be rank ordered but the distances between categories are not equal across the range.

Nominal variables: These are variables comprise categories that cannot be rank. They are also called categorical or qualitative variables.

Dichotomous variables: These are variables contain data that have only two categories. They are treated are ordinary nominal variables.

Missing	Columns	Align	Measure	Role
None	8	Right	Scale	Input
None	8	Right	Scale	Input
None	8	Right	Ordinal	Input
None	8	Right	Nominal	Input
None	15	Right	Unknown	Input
None	8	Right	Unknown	Input
None	12	Left	Nominal	Input
None	17	Left	Nominal	Input
None	35	Left	Unknown	Input

How to categorize a variable

Are there more than two categories?

- Yes
- No → Variable is dichotomous

Can the categories be rank ordered?

- Yes
- No → Variable is nominal/categorical

Are the distances between the categories equal?

- Yes
- No → Variable is ordinal

Yes → Variable is interval/ratio

Bryman, A. (2012). *Social Research Methods, 4th edition.* Oxford University Press.

Role

Sets the intended role of a variable

- *Input:* The variable will be used as a predictor or an independent variable (Selected by default)

- *Target:* The variable will be used as a dependent variable

- *Both:* A variable used as an independent and dependent variable

- *None:* No role assigned

- *Partition:* A variable that breaks the data into groups (especially for experimental testing)

- *Split:* Offers compatibility with other applications of IBM SPSS Packages (such as IBM SPSS Modeler)

Defining variables

Variable Viewer

Provide a variable name: Click on the current variable name and type the name you want.

Provide a variable label: A more detailed description of the variable.

Value Labels: A small button with three dots will appear next to the variable of interest.

Entering data

Data Viewer

↓

The rows contain the answers

↓

Each row represents a case

↓

Each column represents a variable

Basic Operations

Recode

A procedure that allows codes or numbers to be changed

Example: Grouping people by age

Recode into same variable: The original variable will be replaced by new variable and values.
(Transform > Recode into Same Variables)

Recode into different variable: The original variable will be preserved.
(Transform > Recode into Different Variables)

Automatic recode: convert string and numeric values into consecutive integers *(Transform > Automatic Recode)*

Compute
A procedure that allows the combination of two or more variables to form a new variable
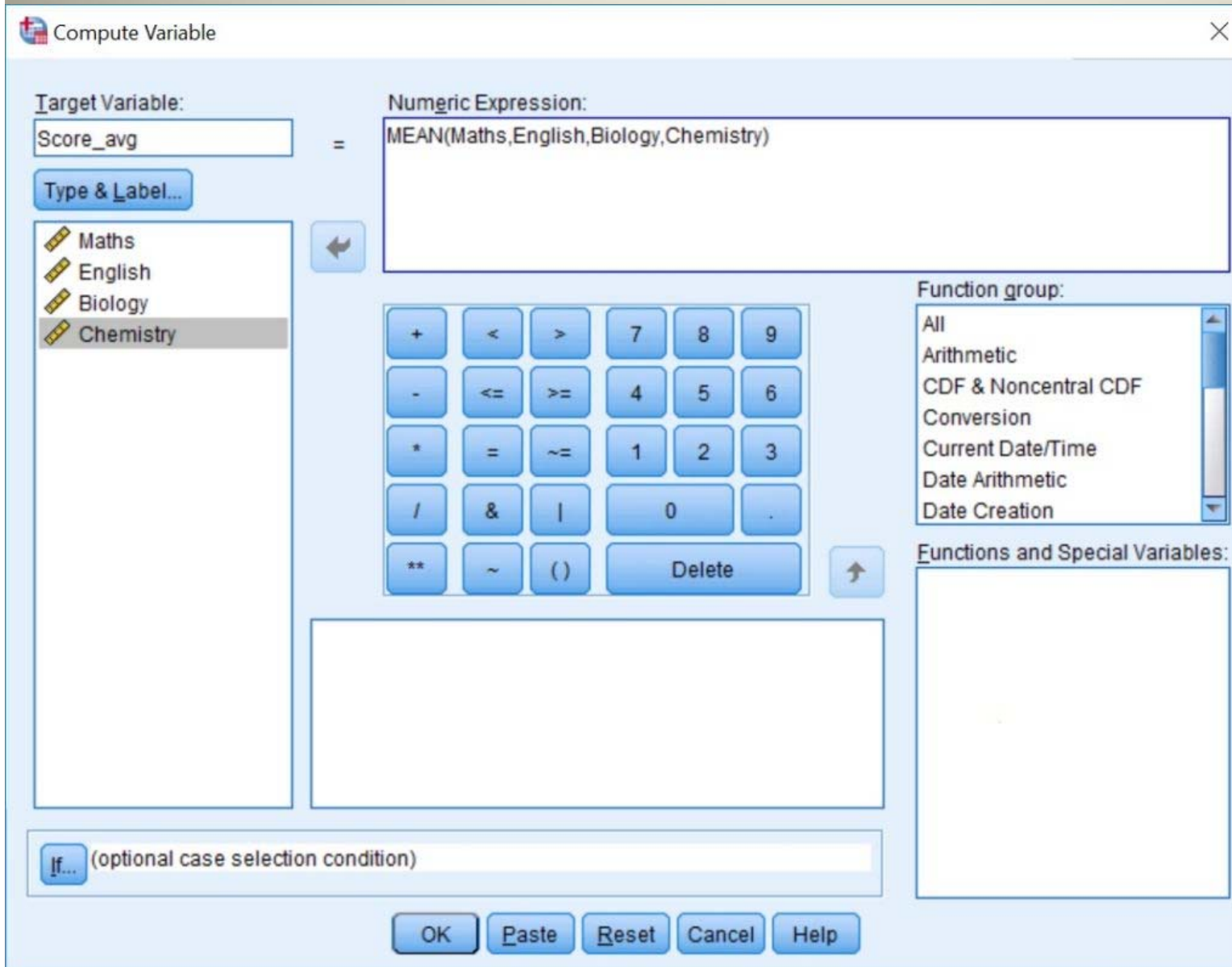
For example:
Compute a subscale score from items on a survey

Graphs

A drop-down menu illustrating all different type of charts we can use

Chart Editor: When a graph is produced, it can be edited with the Chart Editor. To activate the Editor, double-click anywhere in the graph.
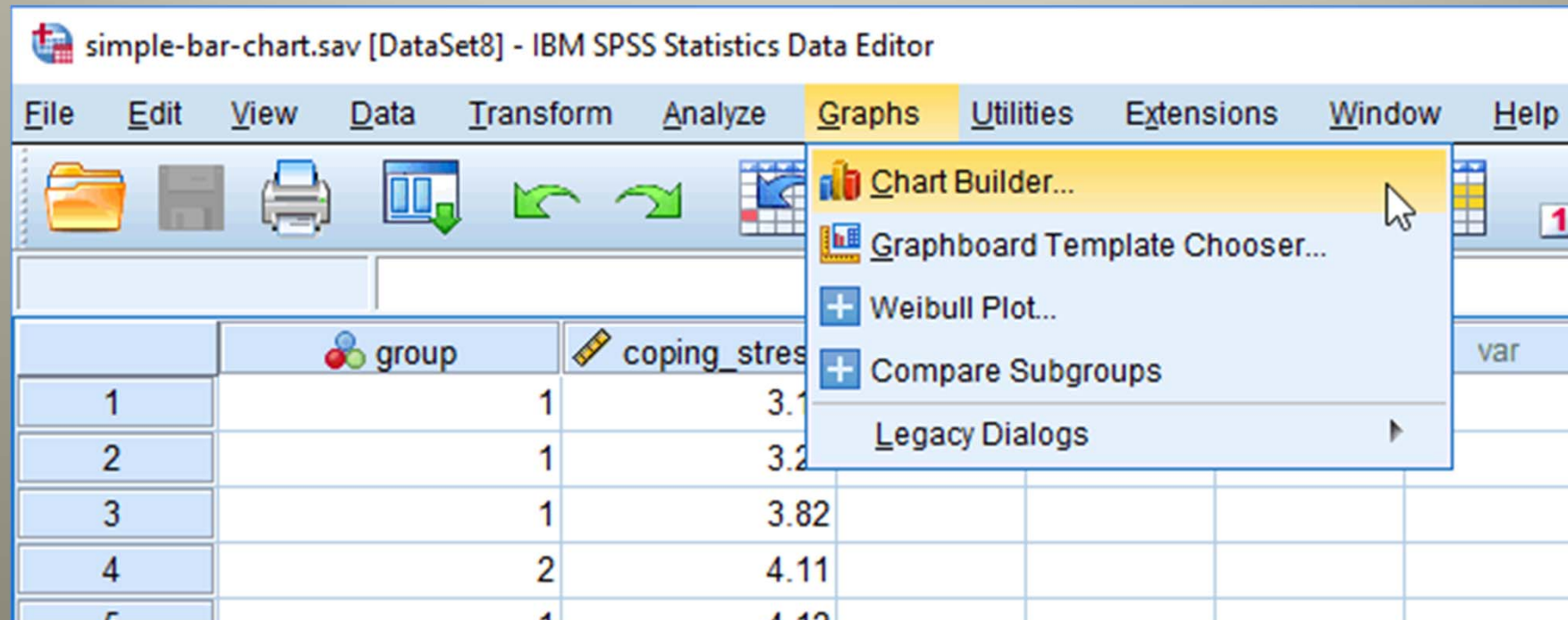
Analyze
A dropdown menu where all techniques of analysis are presented.