

***PHÂN TÍCH SỐ LIỆU
NGHIÊN CỨU VỚI SPSS***



1. Mã hóa dữ liệu

1.1

Mã hóa dữ liệu

1.2

Làm sạch dữ liệu

1.1 MÃ HÓA DỮ LIỆU



MÃ HÓA DỮ LIỆU

* Nguyên tắc mã hóa dữ liệu:

- Đối với dữ liệu không phải ở dạng số (Ví dụ: Nữ, sinh viên, ...) cần phải tạo một con số để mã hóa.
- Đối với các thông tin/dữ liệu thu thập từ thang đo định lượng đã thể hiện dưới dạng số thì không cần mã hóa.

MÃ HÓA DỮ LIỆU

*Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Icons: Folder, Disk, Printer, Speech, Green Arrow, Undo, Redo, Paste, Download, Import, Export, Data, Missing, Split, Merge, Split, Merge, ABC, Checkmark

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											

1

Data View Variable View

IBM SPSS Statistics Processor is ready

MÃ HÓA DỮ LIỆU

Data View: kiểu nhìn dùng để nhập dữ liệu và thể hiện dữ liệu đã nhập.

Variable View: kiểu nhìn dùng để khai báo biến.

Name: Tên biến

Type: Kiểu biến

Label: Mô tả biến/Nhãn của biến

Values: Nhãn trị số của biến.

Measure: Thang đo của biến

MÃ HÓA DỮ LIỆU

Mã hóa dữ liệu với bảng hỏi:

- Mã hóa câu hỏi một câu trả lời đúng nhất.
- Mã hóa câu hỏi với thang đo Likert.
- Mã hóa câu hỏi với nhiều lựa chọn.

THỰC HÀNH MÃ HÓA VÀ NHẬP SỐ LIỆU
- PHIẾU THU THẬP THÔNG TIN DÀNH CHO KHÁCH DU LỊCH -

A. THÔNG TIN CHUYẾN ĐI

1. Mục đích chính của chuyến đi của ông (bà) là gì? (Hãy chọn 1 câu trả lời đúng nhất)

☐ Thăm bạn bè, người thân
☐ Công việc
☐ Nghỉ biển
☐ Y tế

☐ Văn hoá/Di sản
☐ Thiên nhiên
☐ Tour trọn gói
☐ Khác

2. Toàn bộ chuyến đi của ông (bà) kéo dài trong bao lâu?
_____ ngày

3. Ông (bà) vui lòng ước lượng chi phí cho một người/ngày trong chuyến đi của mình?
_____ đồng

4. Lý do ông (bà) đến Huế

1 - Không quan trọng/không phải là lý do → 5 - Rất quan trọng/là lý do chính

Biển
Văn hoá/ Di sản
Tự nhiên/Thiên nhiên
Bao gồm trong tour

1 2 3 4 5

☐
☐
☐
☐

5. Ông (bà) đã tìm hiểu về Huế qua các kênh thông tin nào?
(Có thể chọn nhiều câu trả lời)

☐ Internet
☐ Bạn bè, người thân
☐ Công ty du lịch/lữ hành

☐ Sách hướng dẫn du lịch
☐ Phim ảnh/ Chương trình tivi
☐ Khác

B. THÔNG TIN CÁ NHÂN

1. Tuổi

☐ ≤ 18
☐ 19 - 30
☐ 31 - 45
☐ 45 - 60
☐ > 60

2. Giới tính

☐ Nam
☐ Nữ

1.2 LÀM SẠCH DỮ LIỆU



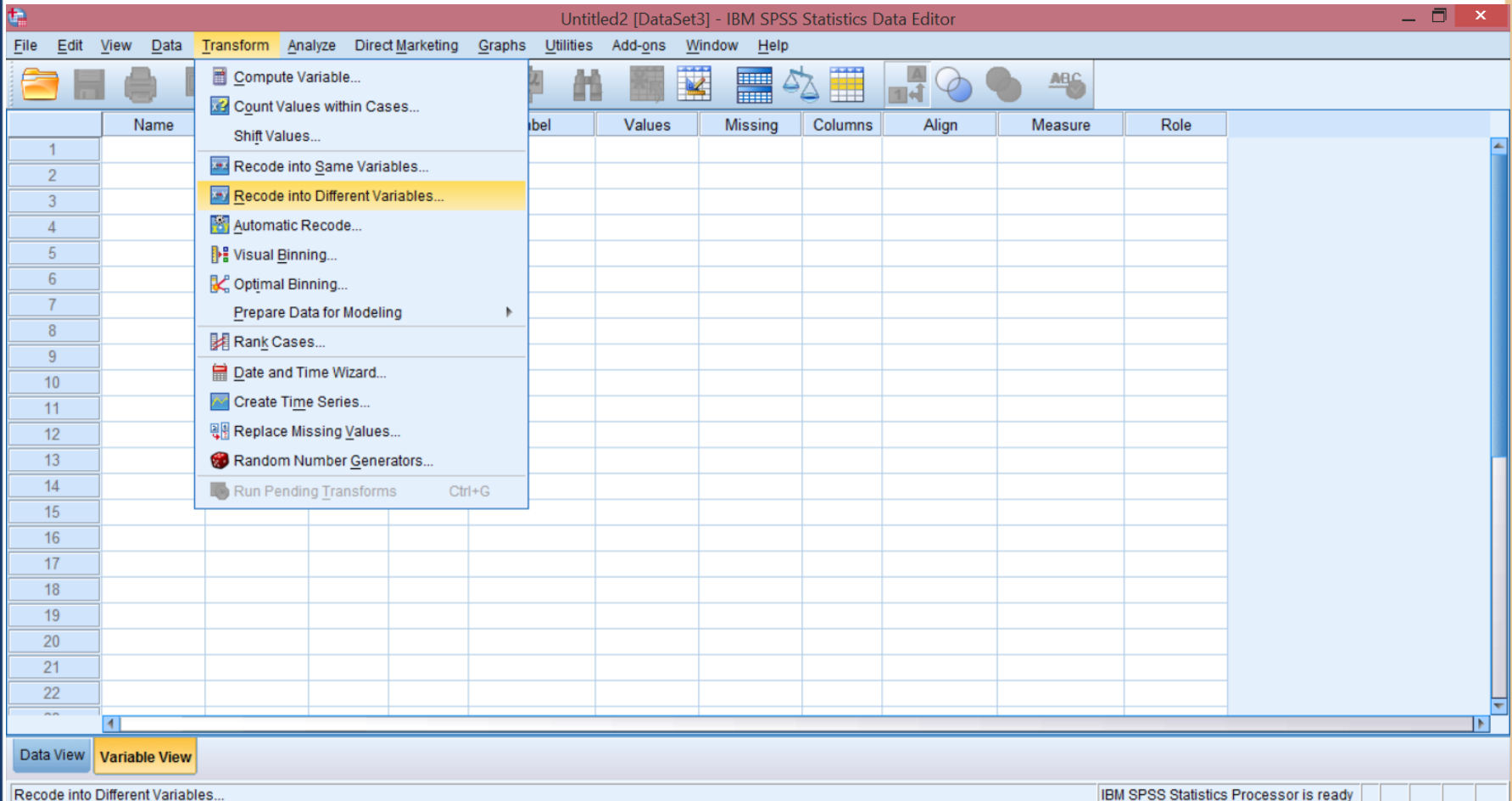
MÃ HÓA LẠI DỮ LIỆU (RECODE)

MÃ HÓA LẠI (RECODE)

- Mã hóa lại (Recode) dữ liệu nhằm chỉnh sửa giá trị của một biến.
- Ví dụ: Nhóm các dữ liệu của biến “Thời gian của chuyến đi” thành các nhóm dữ liệu.

MÃ HÓA LẠI (RECODE)

* Cách thực hiện trên SPSS



Transform → Recode into Different Variables

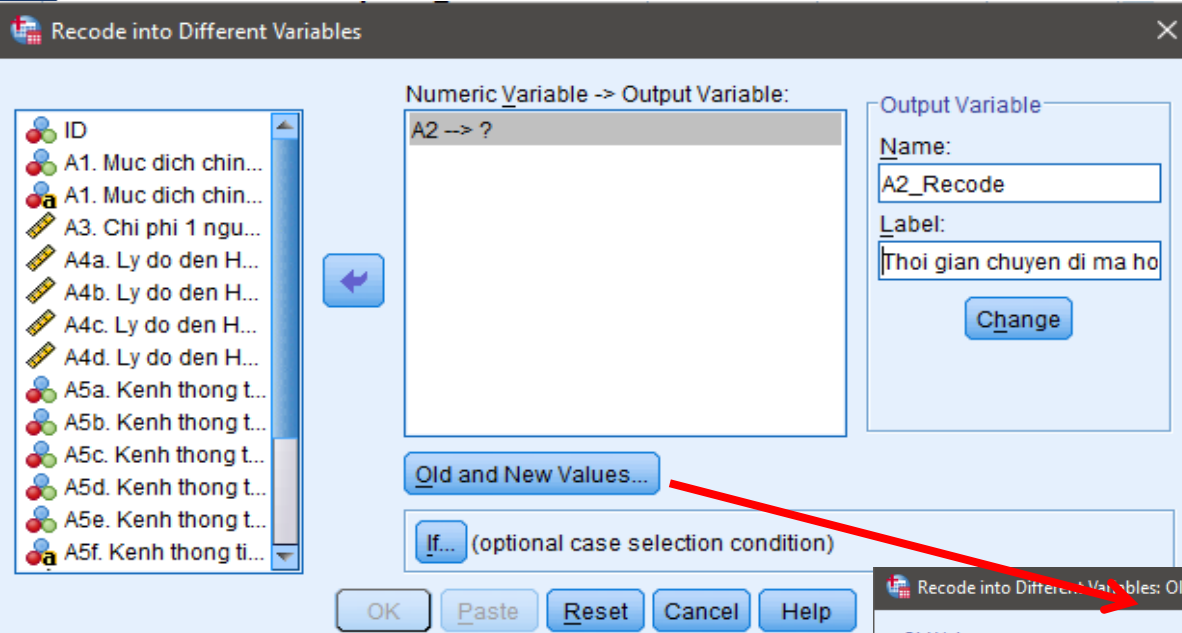
MÃ HÓA LẠI (RECODE)

* Cách thực hiện trên SPSS

Numeric Variable → Output Variable: Chọn biến cần mã hóa

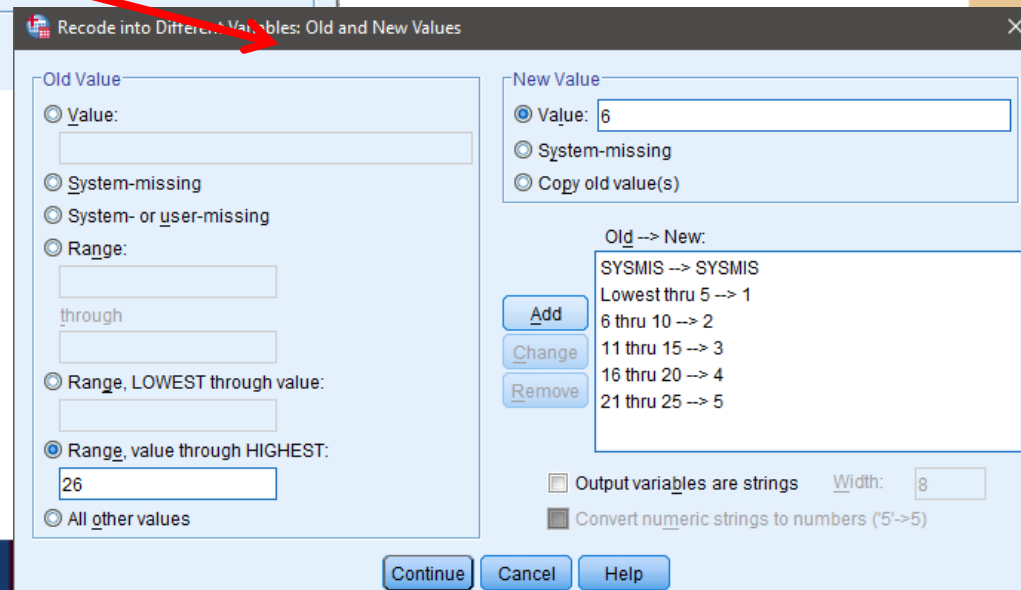
Name: Đặt tên mới cho biến mã hóa

Label: Đặt nhãn mới cho biến mã hóa



The main dialog box for recoding variables in SPSS. It features a list of variables on the left, including 'ID', 'A1. Mục đích chín...', 'A3. Chi phí 1 ngu...', 'A4a. Ly do den H...', 'A4b. Ly do den H...', 'A4c. Ly do den H...', 'A4d. Ly do den H...', 'A5a. Kênh thông t...', 'A5b. Kênh thông t...', 'A5c. Kênh thông t...', 'A5d. Kênh thông t...', 'A5e. Kênh thông t...', and 'A5f. Kênh thông t...'. A blue arrow points from the 'Old and New Values...' button to the 'Old and New Values' sub-dialog box. The 'Numeric Variable -> Output Variable:' section shows 'A2 --> ?'. The 'Output Variable' section has 'Name: A2_Recode' and 'Label: Thời gian chuyển đi ma họ' with a 'Change' button. At the bottom are 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' buttons.

Old and New Values: Đặt các giá trị cho từng nhóm trong biến mới



The 'Old and New Values' sub-dialog box. It has two main sections: 'Old Value' and 'New Value'. The 'Old Value' section has radio buttons for 'Value:', 'System-missing', 'System- or user-missing', 'Range:', 'Range, LOWEST through value:', 'Range, value through HIGHEST:', and 'All other values'. The 'New Value' section has radio buttons for 'Value:', 'System-missing', and 'Copy old value(s)'. Below these is a table for mapping old values to new values. The 'Add' button is highlighted. At the bottom are 'Continue', 'Cancel', and 'Help' buttons.

Old	New
SYSMIS	SYSMIS
Lowest thru 5	1
6 thru 10	2
11 thru 15	3
16 thru 20	4
21 thru 25	5

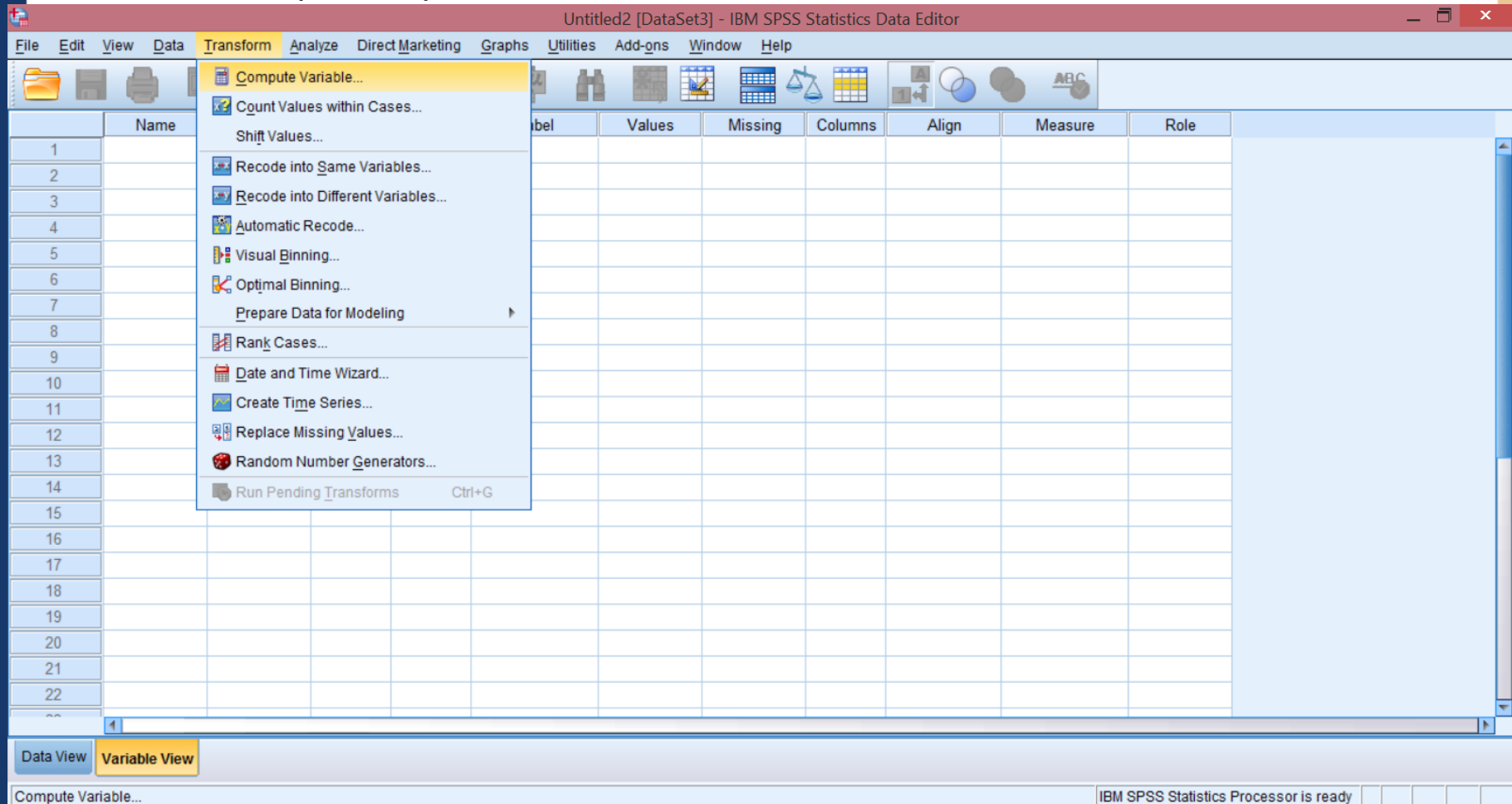
TÍNH TOÁN (COMPUTE)

TÍNH TOÁN (COMPUTE)

- Tính toán (Compute) là tính toán giá trị mới từ biến có sẵn.
- Ví dụ: Tạo biến mới có giá trị chính là giá trị trung bình của các lý do mà du khách đến Huế

TÍNH TOÁN (COMPUTE)

* Cách thực hiện trên SPSS



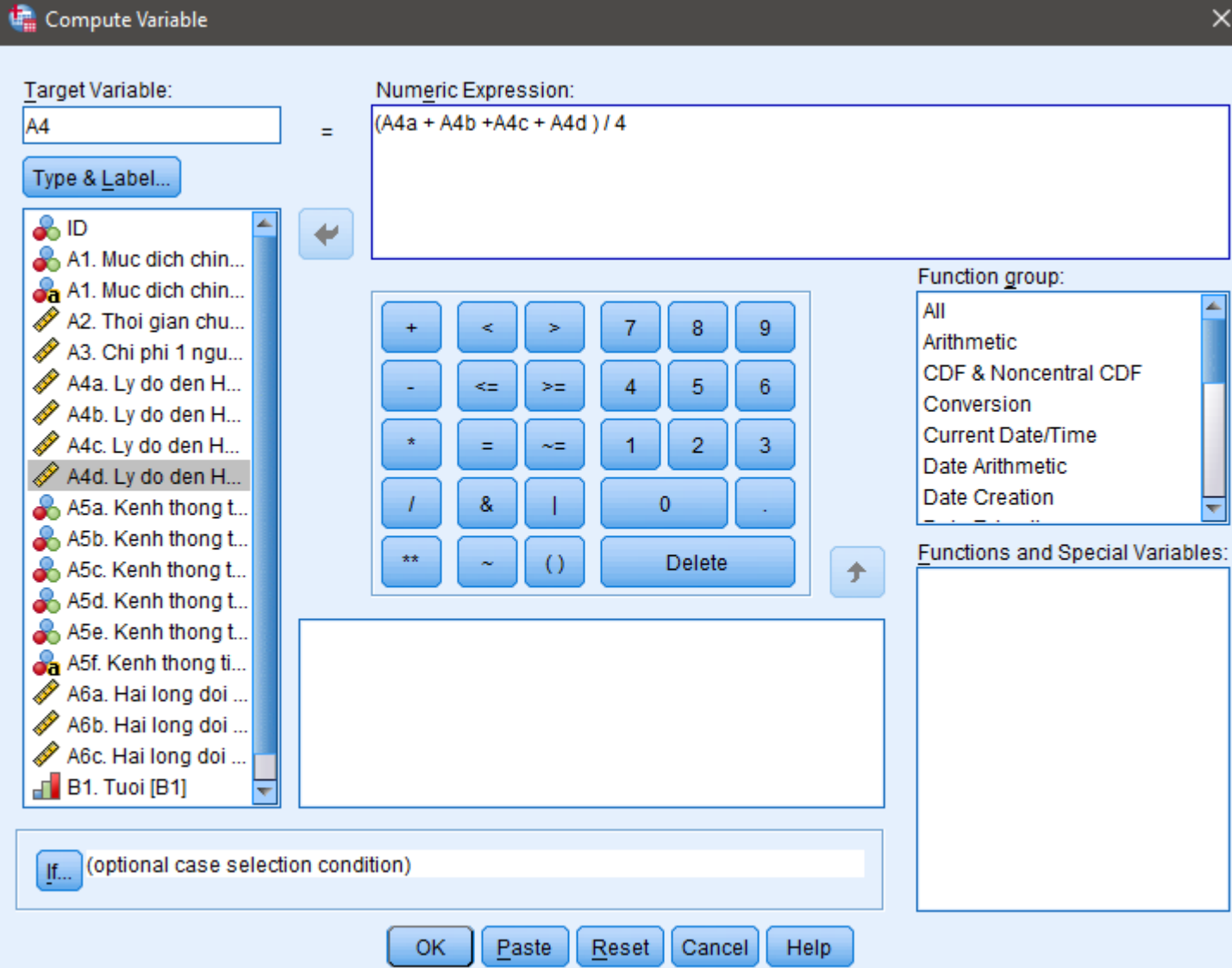
Tranform → Compute Variables

TÍNH TOÁN (COMPUTE)

* Cách thực hiện trên SPSS

Target Variable:
Đặt tên biến mới

Numeric Expression: Công thức tính



The image shows the 'Compute Variable' dialog box in SPSS. The 'Target Variable' field contains 'A4'. The 'Numeric Expression' field contains the formula $(A4a + A4b + A4c + A4d) / 4$. A list of variables is on the left, with 'A4d. Ly do den H...' selected. A calculator keypad is in the center. On the right, there are lists for 'Function group' and 'Functions and Special Variables'. At the bottom, there is an 'If...' field for optional case selection conditions and buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

Compute Variable

Target Variable: A4

Numeric Expression: $(A4a + A4b + A4c + A4d) / 4$

Type & Label...

Variables:

- ID
- A1. Muc dich chin...
- A1. Muc dich chin...
- A2. Thoi gian chu...
- A3. Chi phi 1 ngu...
- A4a. Ly do den H...
- A4b. Ly do den H...
- A4c. Ly do den H...
- A4d. Ly do den H...
- A5a. Kenh thông t...
- A5b. Kenh thông t...
- A5c. Kenh thông t...
- A5d. Kenh thông t...
- A5e. Kenh thông t...
- A5f. Kenh thông ti...
- A6a. Hai long doi ...
- A6b. Hai long doi ...
- A6c. Hai long doi ...
- B1. Tuổi [B1]

Function group:

- All
- Arithmetic
- CDF & Noncentral CDF
- Conversion
- Current Date/Time
- Date Arithmetic
- Date Creation

Functions and Special Variables:

If... (optional case selection condition)

OK Paste Reset Cancel Help

TÍNH TOÁN (COMPUTE)

* Cách thực hiện trên SPSS

*Thuc hanh 1.sav [DataSet1] - IBM SPSS Statistics Data Editor

	A4a	A4b	A4c	A4d	A4	var
1	3,00	5,00	4,00	1,00	3,25	
2	5,00	4,00	3,00	1,00	3,25	
3	3,00	4,00	4,00	2,00	3,25	
4	2,00	5,00	4,00	1,00	3,00	
5	3,00	4,00	4,00	2,00	3,25	
6	5,00	2,00	3,00	2,00	3,00	
7	2,00	3,00	3,00	1,00	2,25	
8	3,00	4,00	2,00	.	.	
9	4,00	4,00	4,00	3,00	3,75	
10	3,00	.	3,00	1,00	.	
11	4,00	4,00	4,00	5,00	4,25	
12	4,00	5,00	4,00	5,00	4,50	
13	5,00	3,00	.	3,00	.	
14	4,00	4,00	4,00	3,00	3,75	
15	3,00	5,00	5,00	.	.	
16	3,00	3,00	4,00	.	.	
17	1,00	5,00	1,00	1,00	2,00	
18		1 00				

Biến A4 mới được tạo ra. Các phần tử của biến A4 chính là giá trị trung bình của các biến tạo nên biến A4.

KIỂM TRA LẠI DỮ LIỆU

KIỂM TRA LẠI DỮ LIỆU

Sự cần thiết của kiểm tra lại dữ liệu do dữ liệu sau khi nhập xong thường chưa đưa vào xử lý và phân tích vì còn nhiều lỗi do:

- Chất lượng phỏng vấn: phỏng vấn viên hiểu sai câu hỏi, thu thập số liệu sai, ghi chép nhầm, người được phỏng vấn hiểu sai ý....
- Nhập dữ liệu: sai, sót, thừa.

KIỂM TRA LẠI DỮ LIỆU

- * Các cách kiểm tra lại dữ liệu:
 - Sử dụng bảng tần số.
 - Sử dụng lệnh Sort Cases.

KIỂM TRA LẠI DỮ LIỆU

* Sử dụng lệnh Sort Cases

*Thuc hanh 1.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View **Data** Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

1 : A4a

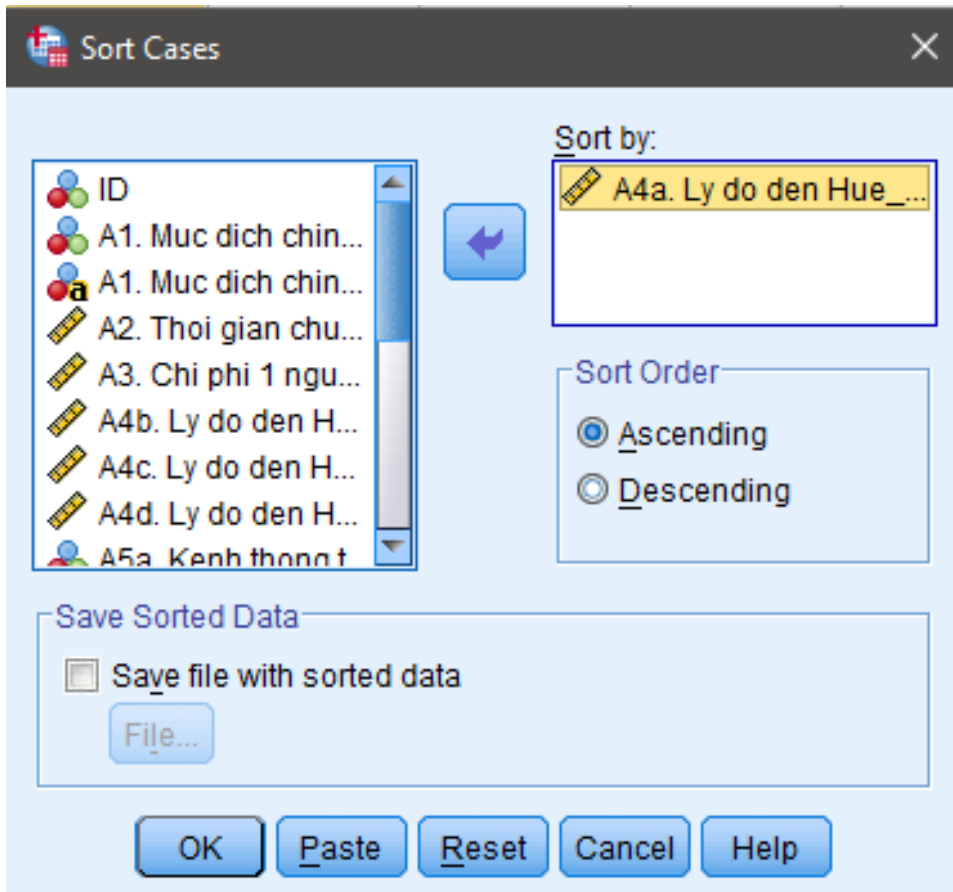
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Sort Cases...
Sort Variables...
Transpose...
Merge Files
Restructure...

	A3	A4a	A4b
7	300	3,00	5,00
3	1000	5,00	4,00
7	500	3,00	4,00
7	600	2,00	5,00
5	700	3,00	4,00
4	2000	5,00	2,00
0	.	2,00	3,00
0	300	3,00	4,00
3	500	4,00	4,00
2	800	3,00	.
2	700	4,00	4,00
3	900	4,00	5,00

Data → Sort Cases

KIỂM TRA LẠI DỮ LIỆU

* Sử dụng lệnh Sort Case



Sort by: Chọn biến cần sắp xếp

Sort Order: Chọn kiểu sắp xếp

Ascending: Sắp xếp tăng dần

Decending: Sắp xếp giảm dần

- Thông thường, do điều kiện hạn hẹp về chi phí và thời gian, chúng ta đều dựa trên tiêu chuẩn 5:1 của Bollen (1998) và Hair & ctg (1998).

Để thỏa mãn yêu cầu về dữ liệu của phân tích định lượng, một biến cần có 5 quan sát tương ứng với 5 đáp viên. Bảng câu hỏi đưa ra có 30 biến, nên mẫu tối thiểu là 150 ($30 \times 5 = 150$) người

THANK YOU!

