
GIỚI THIỆU SƠ LƯỢC PHÂN TÍCH DỮ LIỆU VỚI SPSS

TS. LÊ VĂN HUY
Email: levanhuy@vnn.vn

NỘI DUNG CHÍNH

- Giới thiệu về SPSS
- Mã hóa dữ liệu
- Giao diện và nhập liệu
- Một số điểm thường sử dụng

SƠ LƯỢC VỀ SPSS

- SPSS là gì?

SPSS FOR WINDOWS (Statistical Package for Social Sciences)

- Tại sao là SPSS?

- ⊕ STATA

- ⊕ EIEWS

- ⊕ EXCEL

- ⊕ SAS

MÃ HÓA DỮ LIỆU (CODING)

- Các loại mã hóa

 - ⊕ Mã hóa trước

 - ⊕ Mã hóa sau

- Ví dụ

 - ⊕ 1: Nam

 - 0: Nữ

 - ⊕ 1: Đồng ý

 - 0: Không đồng ý

 - ⊕ 1: Rất không đồng ý

 - 7: Rất đồng ý

- Chú ý khi Coding

MÃ HÓA DỮ LIỆU

- Các loại mã hóa

 - ⊕ Mã hóa trước

 - ⊕ Mã hóa sau

- Ví dụ

 - ⊕ 1: Nam

 - 0: Nữ

 - ⊕ 1: Đồng ý

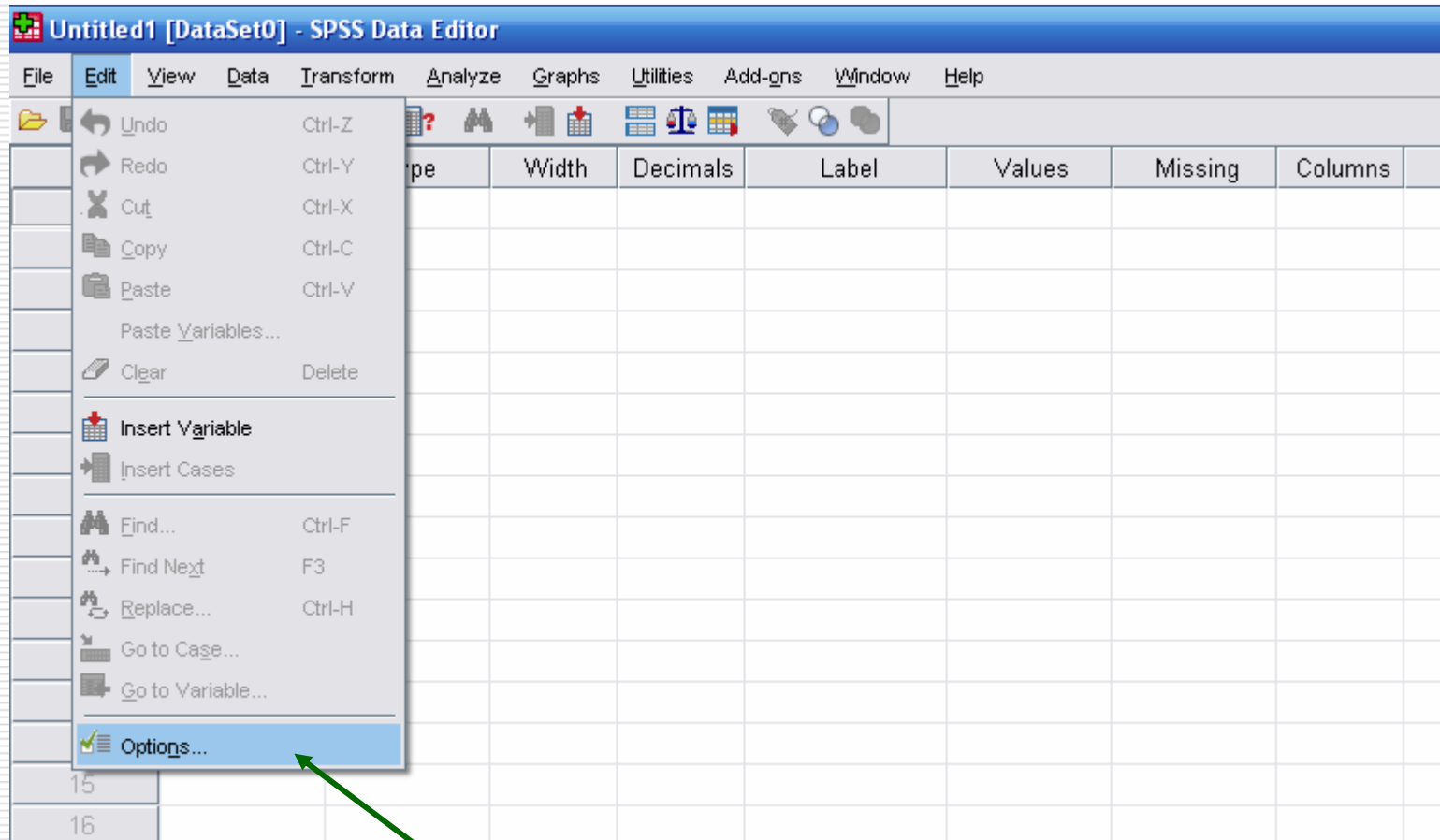
 - 0: Không đồng ý

 - ⊕ 1: Rất không đồng ý

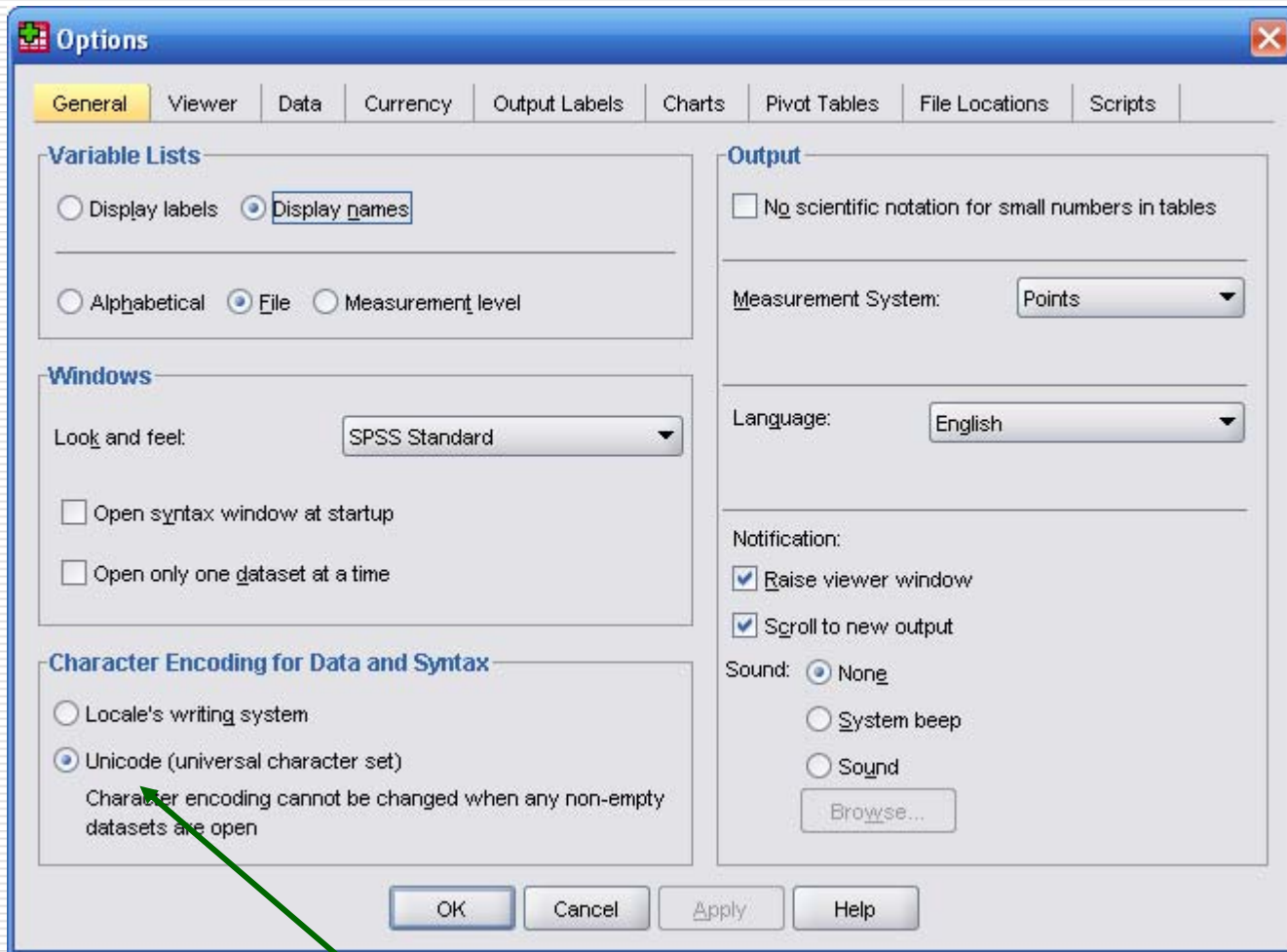
 - 7: Rất đồng ý

- Chú ý khi Coding

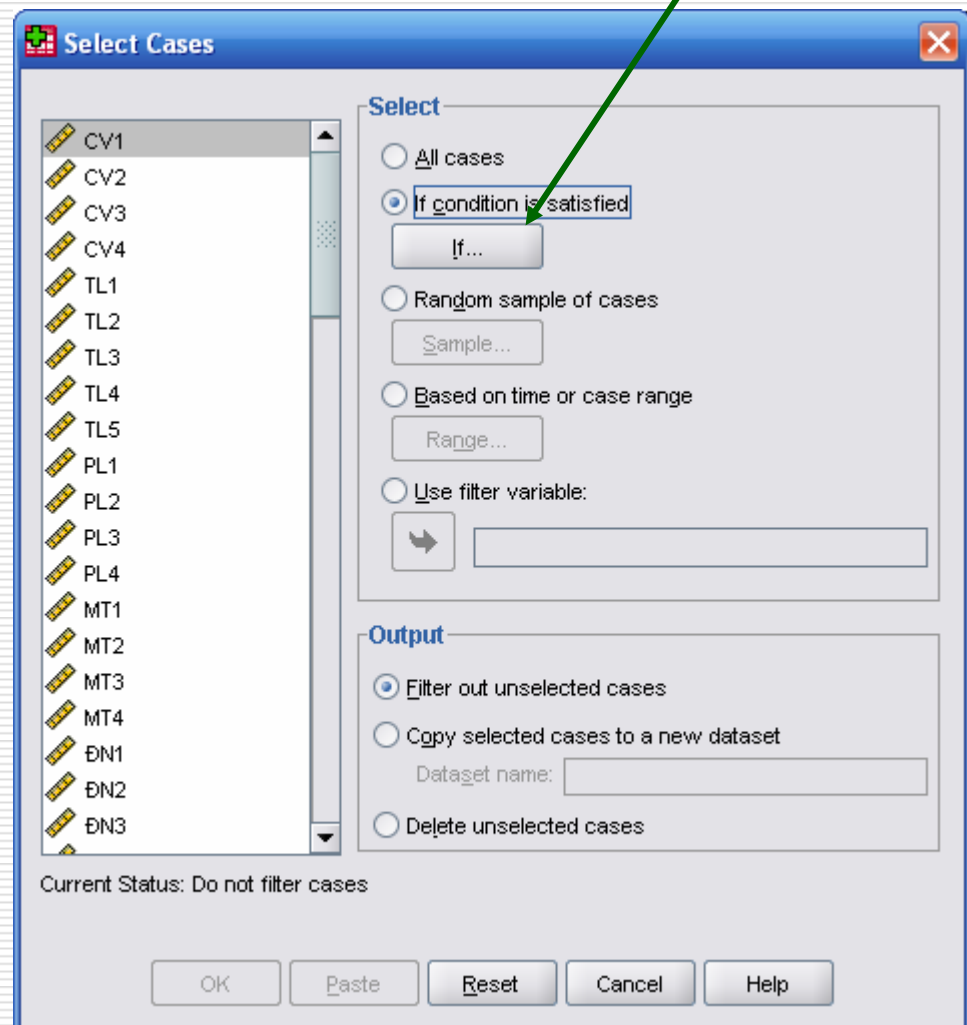
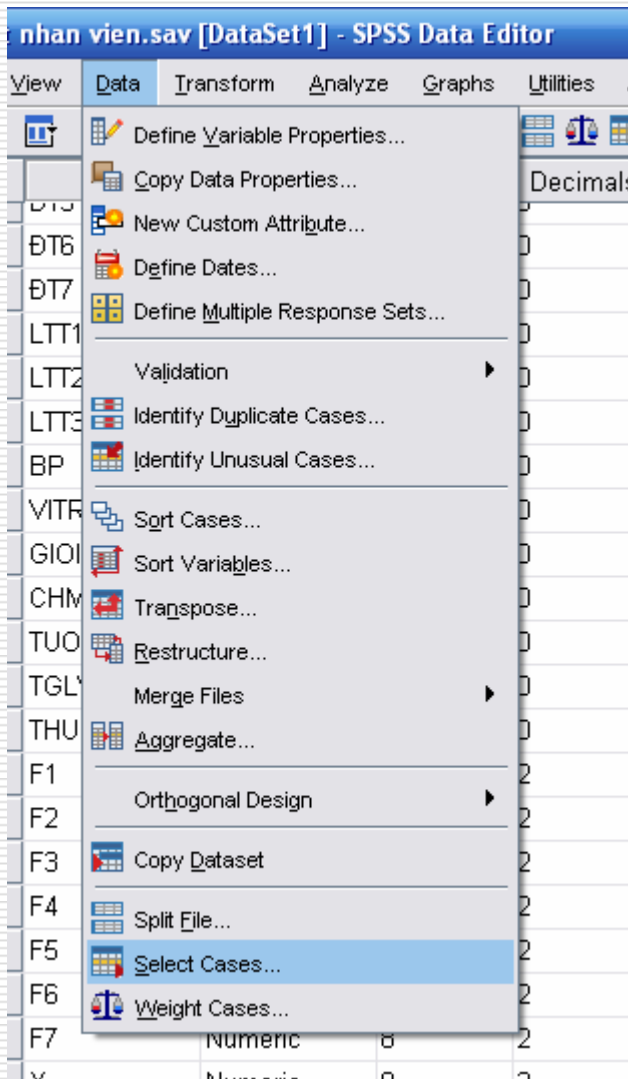
TIẾNG VIỆT (THEO UNICODE)



TIẾNG VIỆT (THEO UNICODE)

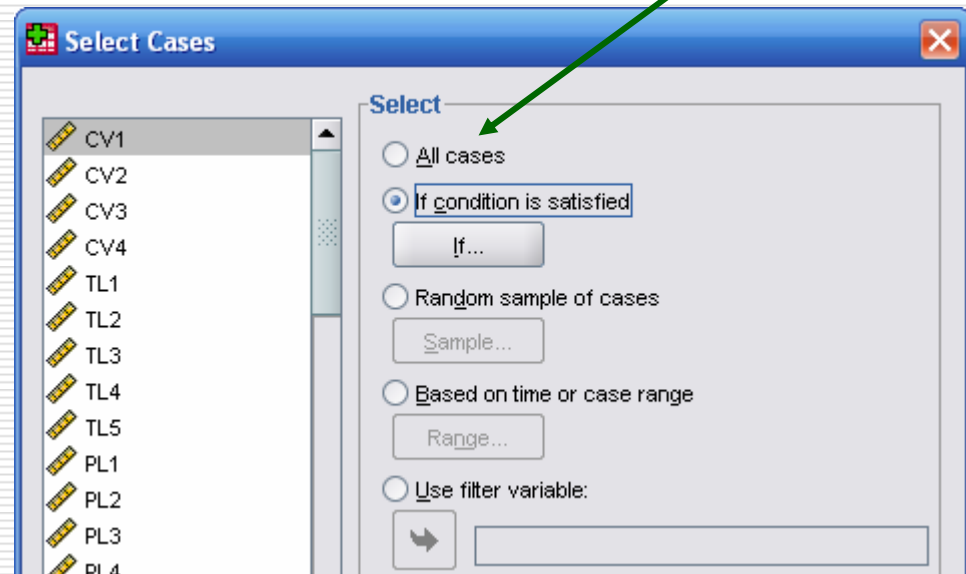


CHỌN DỮ LIỆU THEO ĐIỀU KIỆN

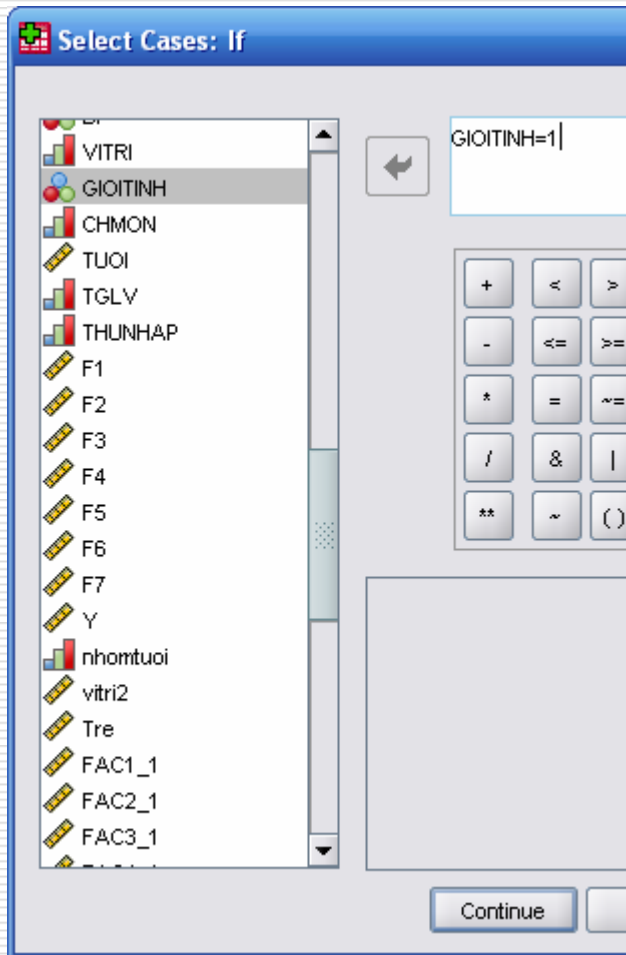


HÃY THỰC HÀNH

- Lựa chọn dữ liệu với điều kiện
 - ⊕ Giới tính = Nam
- Chú ý quan trọng: Sau khi chọn và phân tích sau, hãy trả về dữ liệu gốc khi chọn **All case**



CHỌN DỮ LIỆU THEO ĐIỀU KIỆN



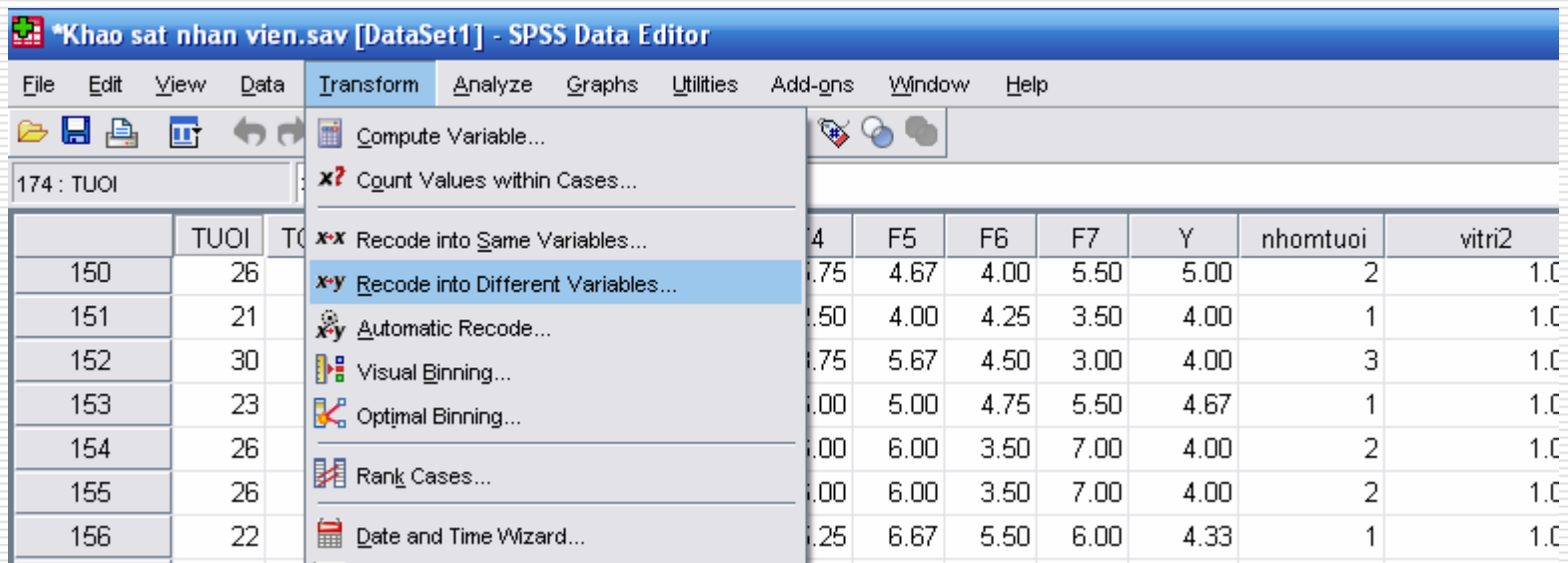
*Khao sat nhan vien.sav [DataSet1] - SPSS Data Editor

1 : CV1 4

	CV1	CV2	CV3	CV4	TL1	TL2
1	4	4	4	4	4	
2	4	4	4	3	3	
3	3	3	4	2	2	
4	5	6	5	5	4	
5	5	6	6	6	4	
6	7	7	4	7	4	
7	5	3	4	2	4	
8	5	4	5	6	4	
9	4	4	4	4	3	
10	6	6	7	4	3	
11	4	3	4	4	5	
12	3	3	5	3	3	
13	5	4	6	6	5	
14	6	5	4	7	5	
15	4	1	2	1	1	
16	3	3	2	5	5	
17	4	5	5	6	2	
18	5	5	6	3	4	

MÃ HÓA LẠI BIẾN

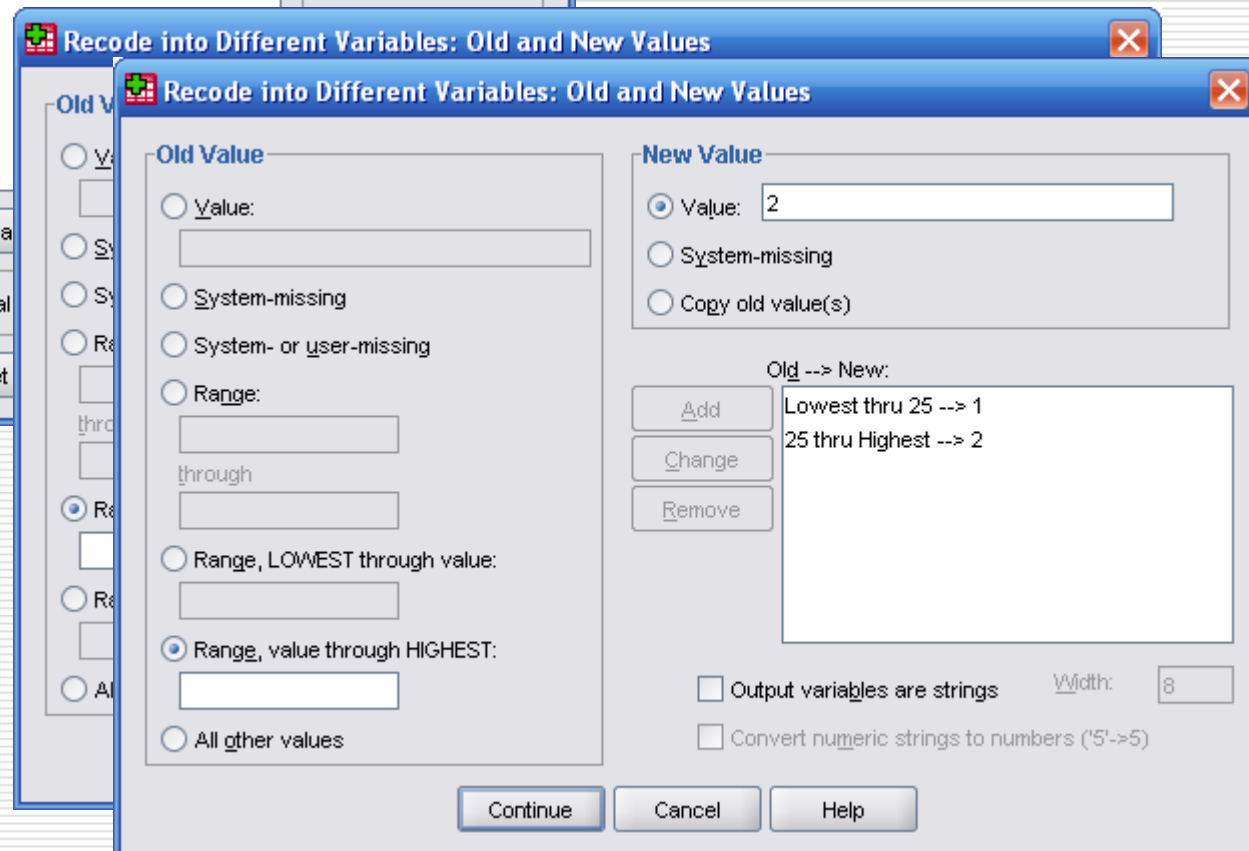
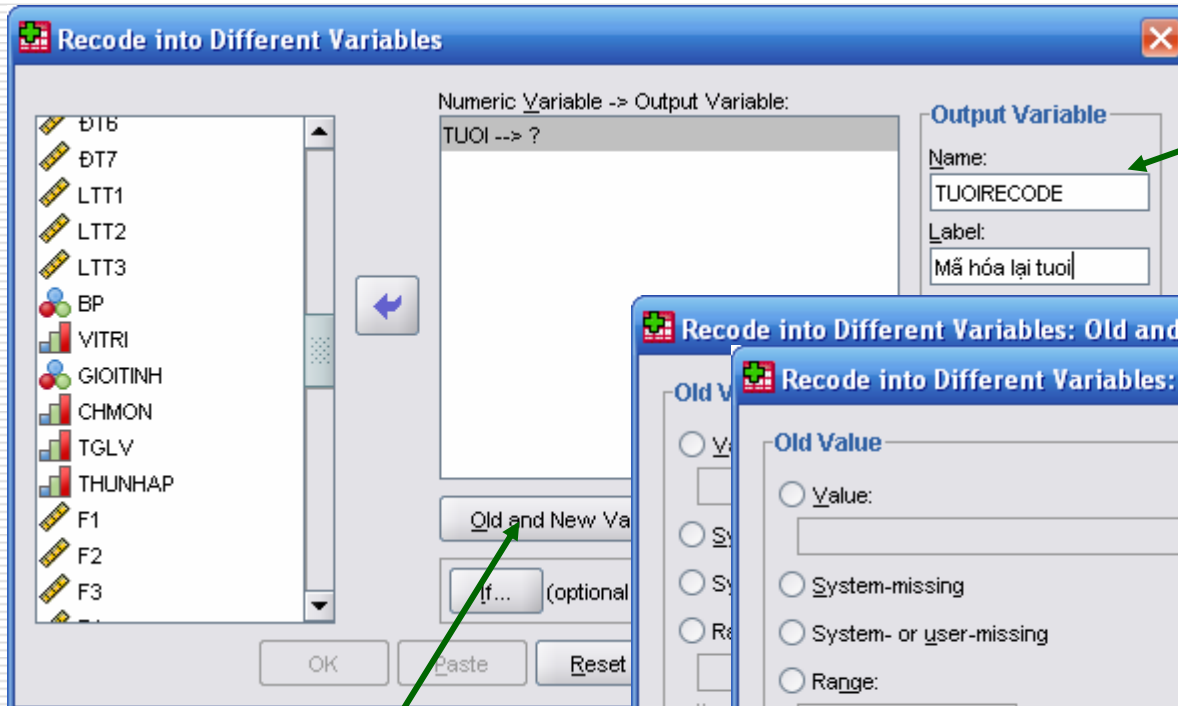
- Hãy mã hóa lại biến TUOI
 - ⊕ Tuổi < 25
 - ⊕ Tuổi > = 25
- Nên mã hóa và lưu lại trên **biến mới**



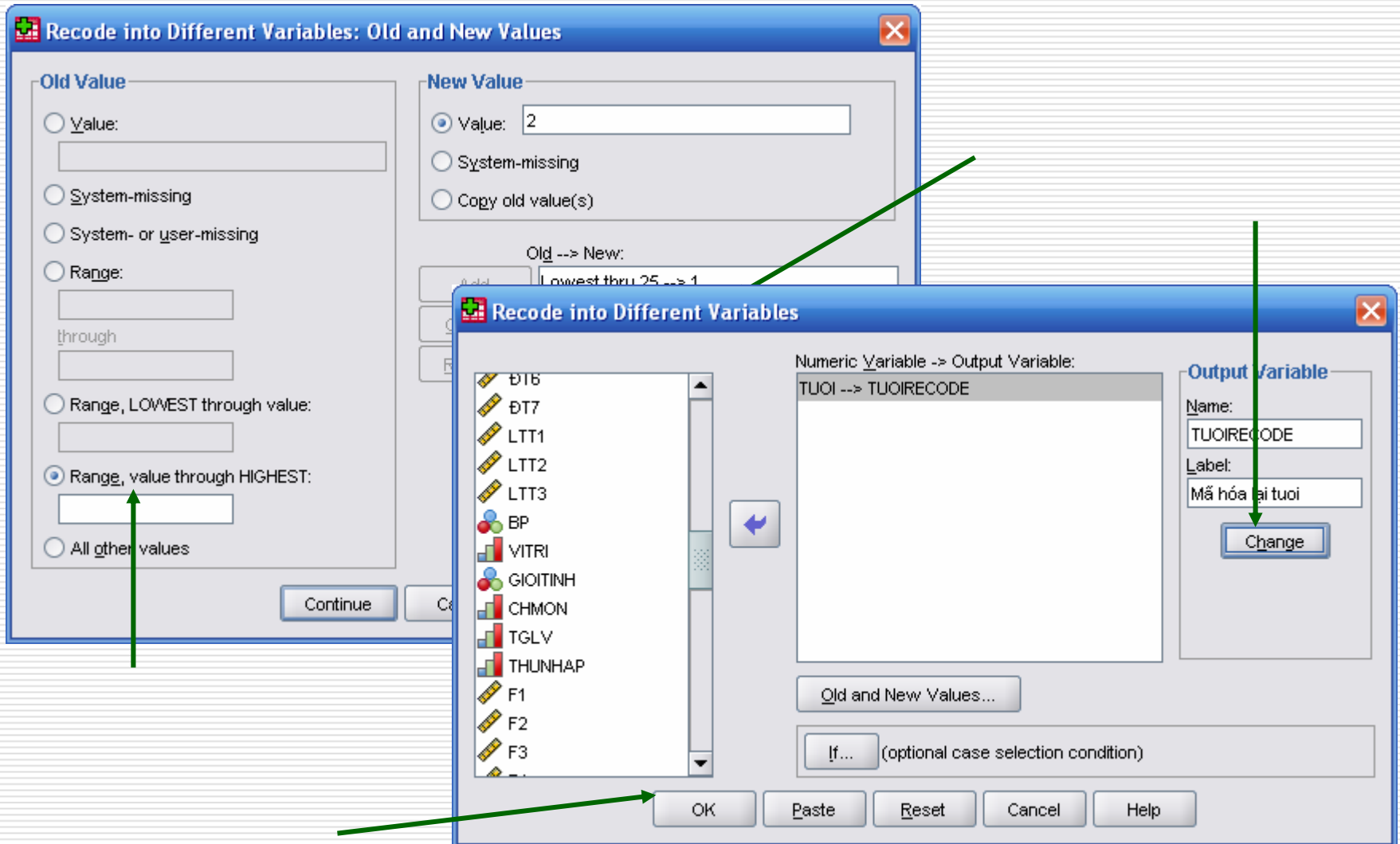
The screenshot shows the SPSS Data Editor interface for the file '*Khao sat nhan vien.sav [DataSet1]'. The 'Transform' menu is open, and the 'Recode into Different Variables...' option is selected. The dialog box shows the variable 'TUOI' being recoded into a new variable 'nhomtuoi'. The 'Old and New Values' table is partially visible, showing the mapping of original values to new values.

Old Value	New Value
4	F5
4.75	4.67
5.00	4.00
5.25	5.50
5.50	F7
5.75	Y
6.00	5.00
6.25	nhomtuoi
6.50	2
6.75	1
7.00	3
7.25	1
7.50	1
7.75	1
8.00	1
8.25	1
8.50	1
8.75	1
9.00	1
9.25	1
9.50	1
9.75	1
10.00	1
10.25	1
10.50	1
10.75	1
11.00	1
11.25	1
11.50	1
11.75	1
12.00	1
12.25	1
12.50	1
12.75	1
13.00	1
13.25	1
13.50	1
13.75	1
14.00	1
14.25	1
14.50	1
14.75	1
15.00	1
15.25	1
15.50	1
15.75	1
16.00	1
16.25	1
16.50	1
16.75	1
17.00	1
17.25	1
17.50	1
17.75	1
18.00	1
18.25	1
18.50	1
18.75	1
19.00	1
19.25	1
19.50	1
19.75	1
20.00	1
20.25	1
20.50	1
20.75	1
21.00	1
21.25	1
21.50	1
21.75	1
22.00	1
22.25	1
22.50	1
22.75	1
23.00	1
23.25	1
23.50	1
23.75	1
24.00	1
24.25	1
24.50	1
24.75	1
25.00	1
25.25	1
25.50	1
25.75	1
26.00	1
26.25	1
26.50	1
26.75	1
27.00	1
27.25	1
27.50	1
27.75	1
28.00	1
28.25	1
28.50	1
28.75	1
29.00	1
29.25	1
29.50	1
29.75	1
30.00	1

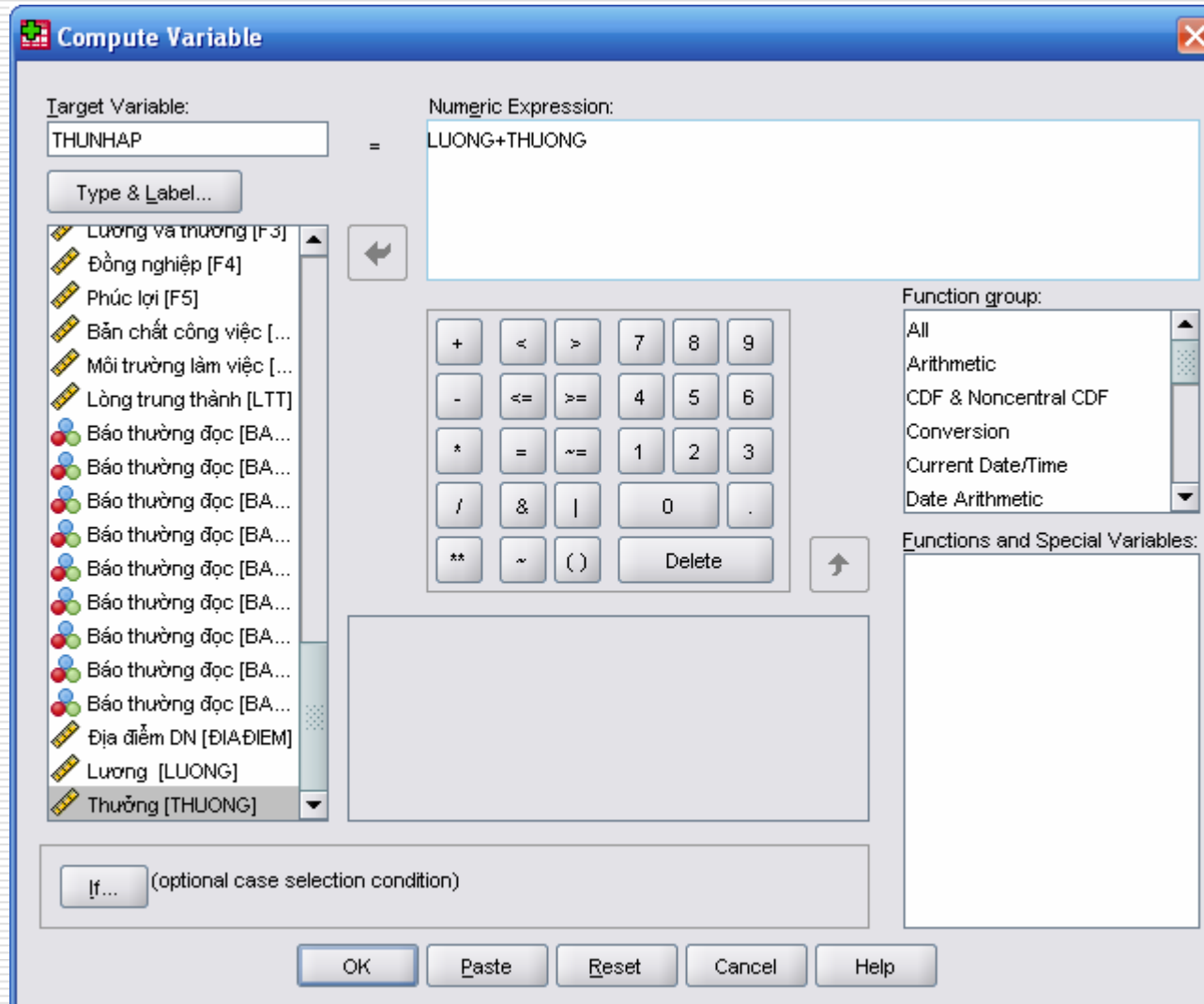
MÃ HÓA LẠI BIẾN



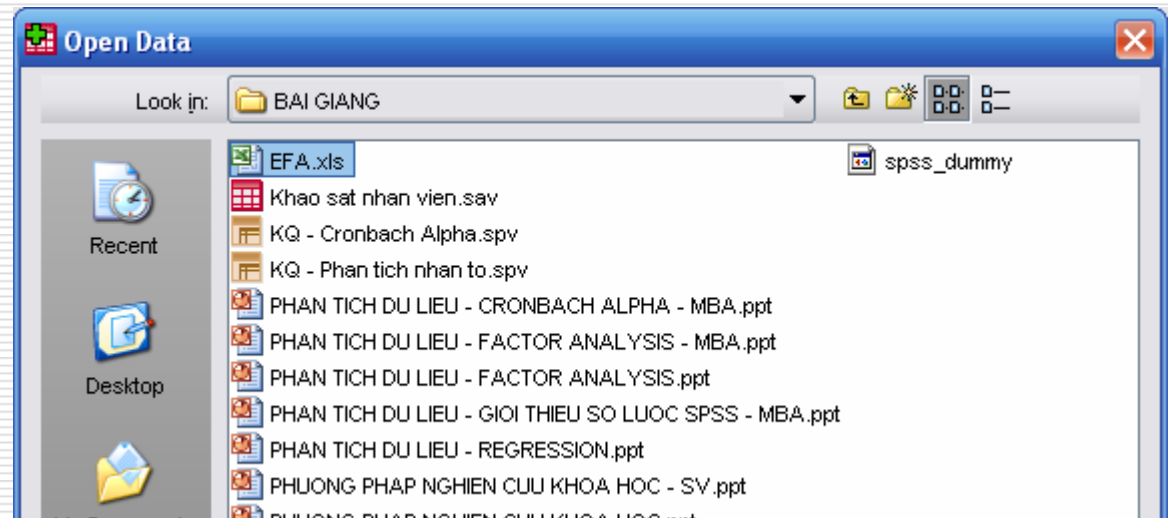
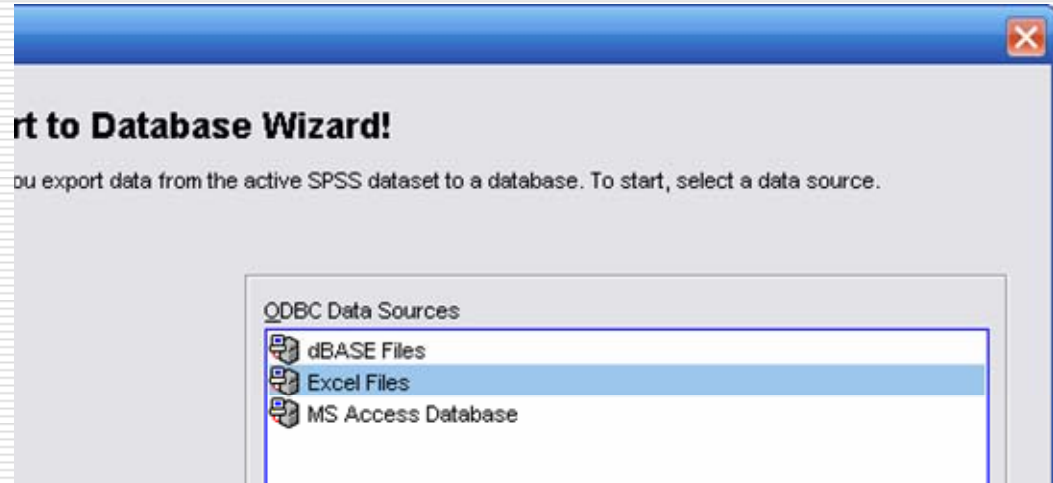
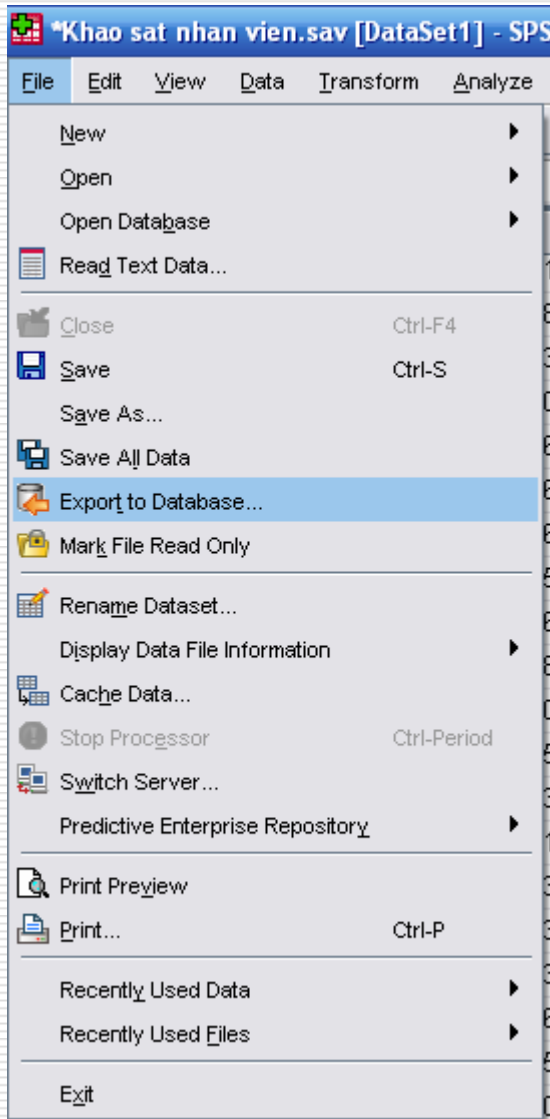
MÃ HÓA LẠI BIẾN



TÍNH TOÁN CÁC BIẾN



ĐỌC DỮ LIỆU TỪ EXCEL



PHÂN TÍCH MÔ TẢ

The screenshot shows the SPSS Data Editor interface. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The 'Analyze' menu is open, and 'Frequencies...' is selected under 'Descriptive Statistics'. The background shows a data table with columns TUOI, TGLV, and THUNHAP.

	TUOI	TGLV	THUNHAP
150	26	1	
151	21	1	
152	30	1	
153	23	1	
154	26	1	
155	26	1	
156	22	1	
157	27	1	
158	26	1	

The 'Frequencies' dialog box is shown. The 'Variable(s):' list contains GIOITINH, CHMON, and TUOI. The 'Display frequency tables' checkbox is checked. Buttons for 'Statistics...', 'Charts...', and 'Format...' are visible on the right. Green arrows point to the 'Statistics...' and 'Charts...' buttons.

Variable(s):
GIOITINH
CHMON
TUOI

PHÂN TÍCH MÔ TẢ

Frequencies: Statistics

Percentile Values

- Quartiles
- Cut points for: 10 equal groups
- Percentile(s):
- Buttons: Add, Change, Remove

Central Tendency

- Mean
- Median
- Mode
- Sum
- Values are group midpoints

Dispersion

- Std. deviation
- Variance
- Range
- Minimum
- Maximum
- S.E. mean

Distribution

- Skewness
- Kurtosis

Buttons: Continue, Cancel, Help

Thống kê xu hướng hội tụ

Đồ thị

Frequencies: Charts

Chart Type

- None
- Bar charts
- Pie charts
- Histograms:
 - With normal curve

Chart Values

- Frequencies
- Percentages

Buttons: Continue, Cancel, Help

Kiểm tra độ phân tán

Kiểm tra phân phối

PHÂN TÍCH MÔ TẢ

GIOITINH

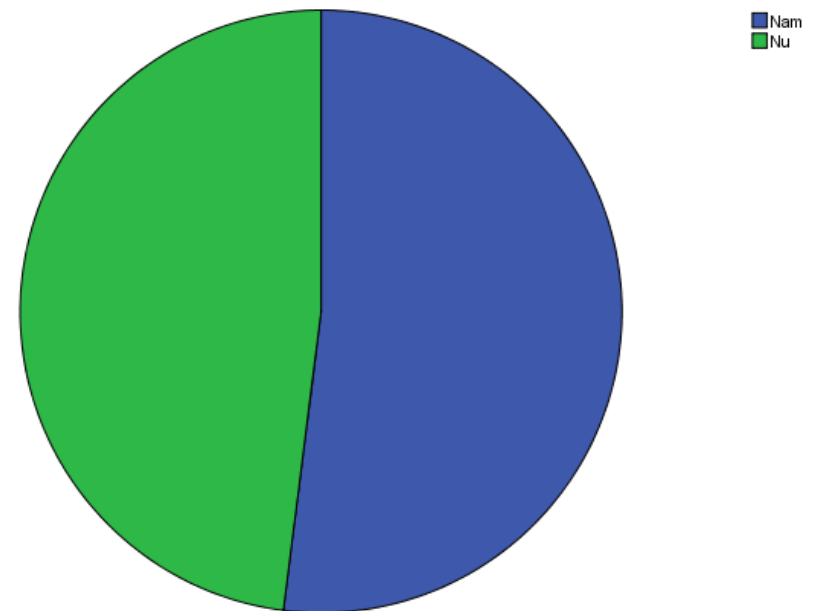
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	104	52.0	52.0	52.0
	2	96	48.0	48.0	100.0
	Total	200	100.0	100.0	

Statistics

TUOI

N	Valid	200
	Missing	0
Mean		26.47
Median		26.00
Minimum		21
Maximum		42
Sum		5294

Giới tính



NHẬP LIỆU CHO CÂU HỎI NHIỀU LỰA CHỌN

- Ví dụ:

- ⊕ Bạn hãy cho biết nhiều nhất 9 loại báo mà bạn thường đọc tại Cơ quan (nơi làm việc)

- ◆ Hà Nội mới
- ◆ SGGP
- ◆ Lao Động
- ◆ Tiền Phong
- ◆ Tuổi trẻ
- ◆ Phụ nữ Việt Nam
- ◆ Sài Gòn Tiếp thị
- ◆ Thế giới Phụ nữ
- ◆ An ninh thế giới

→ Có thể chọn nhiều hơn một phương án → làm thế nào để nhập liệu

NHẬP LIỆU CHO CÂU HỎI NHIỀU LỰA CHỌN

*THUC HANH SPSS - MBA.sav [DataSet1] - SPSS Data Editor

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

	Name	Type	Width	Decimals	Label
51	BAO1	Numeric	5	0	Báo thường đọc
52	BAO2	Numeric	6	0	Báo thường đọc
53	BAO3	Numeric	5	0	Báo thường đọc
54	BAO4	Numeric	5	0	Báo thường đọc
55	BAO5	Numeric	5	0	Báo thường đọc
56	BAO6	Numeric	5	0	Báo thường đọc
57	BAO7	Numeric	5	0	Báo thường đọc
58	BAO8	Numeric	5	0	Báo thường đọc
59	BAO9	Numeric	5	0	Báo thường đọc

Value Labels

Value:

Label:

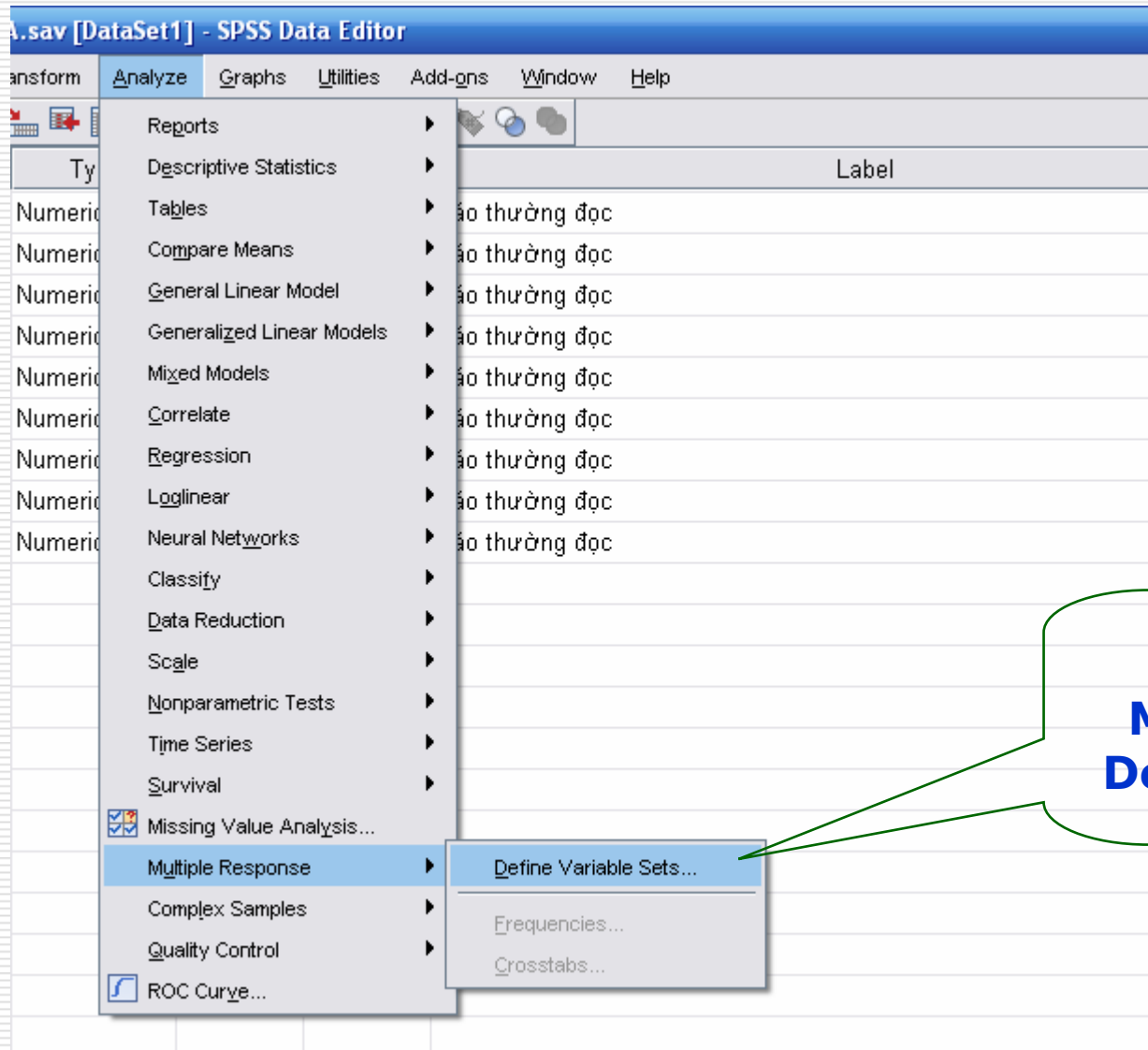
Spelling...

Add Change Remove

1 = "Hà Nội mới"
2 = "SGGP"
3 = "Lao Động"
4 = "Người Lao động"
5 = "Tiền Phong"
6 = "Thanh Niên"
7 = "Tuổi trẻ"

OK Cancel Help

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN



**Analyze
Multiple Response
Define Variable Sets**

NHẬP LIỆU CHO CÂU HỎI NHIỀU LỰA CHỌN

Define Multiple Response Sets

Set Definition

Thu nhập [THUNHAP]
Lãnh đạo [F1]
Cơ hội đào tạo và th...
Lương và thưởng [F3]
Đồng nghiệp [F4]
Phúc lợi [F5]
Bản chất công việc [...]
Môi trường làm việc [...]
Lòng trung thành [LTT]

Variables in Set:
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]

Multiple Response Sets:
\$BAO

Variables Are Coded As:
 Dichotomies Counted value:
 Categories Range: 1 through 18

Name: BAO
Label: Báo thường đọc

Note: Sets defined here are only available in the Multiple Response Frequencies and Crosstabs procedures. Use Define Multiple Response Sets on the Data menu for sets used elsewhere.

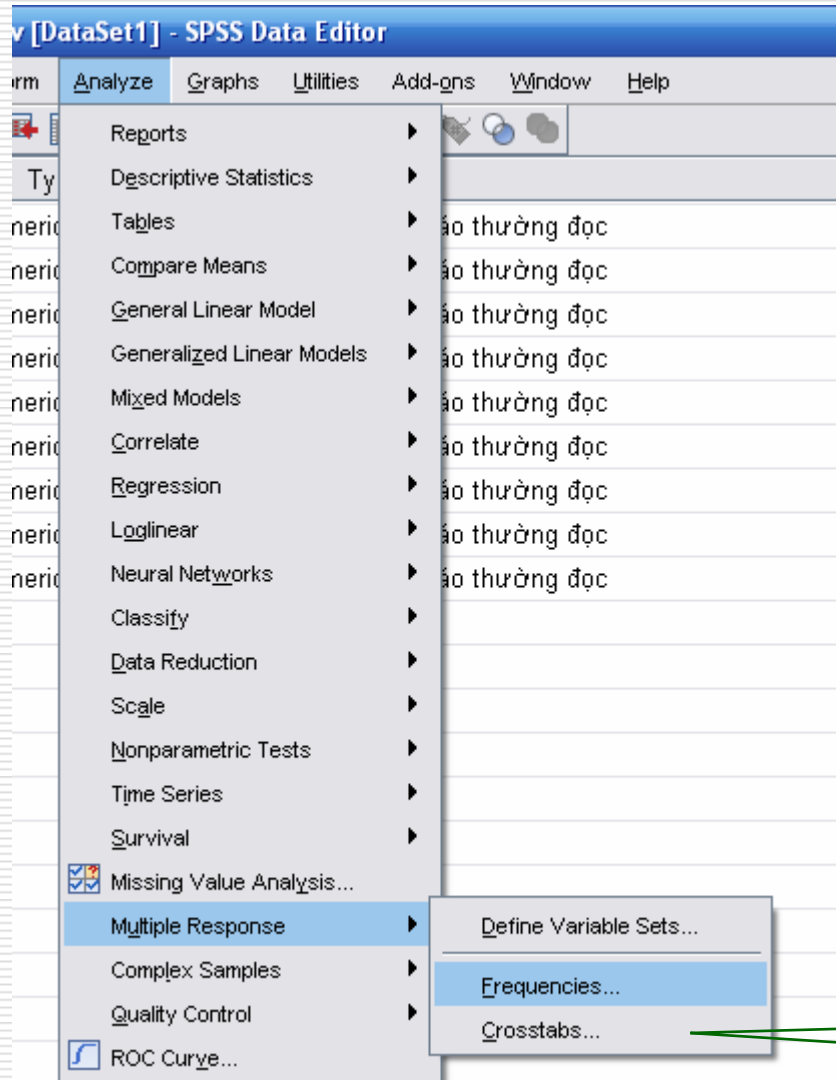
Close Help

Nhấn Add

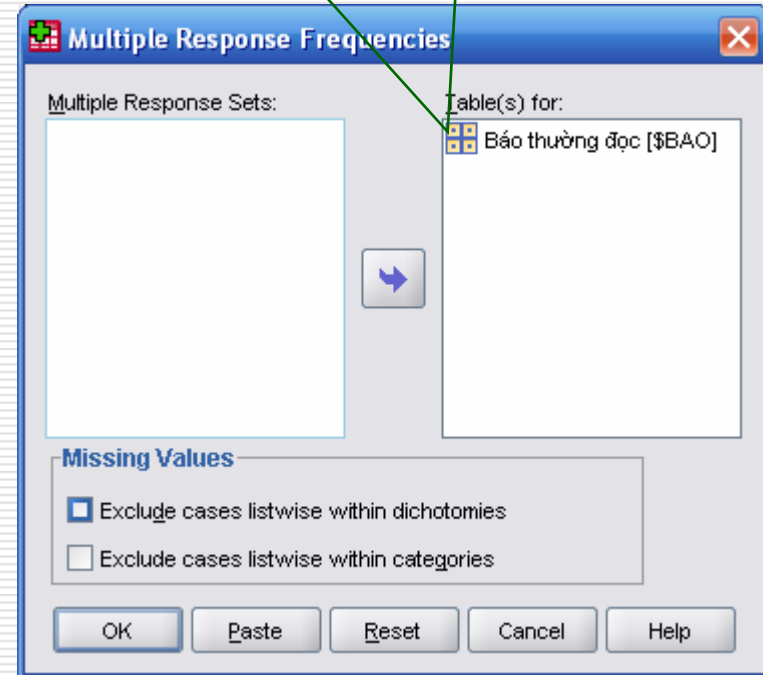
**Nhập kiểu định nghĩa
Dichotomies / Categories
Nhập số lượng lựa chọn**

**Nhập thông tin
Biến: BAO
Label: Báo thường
đọc**

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN



Chọn và chuyển tên biến cần tính



Mô tả tần suất

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN

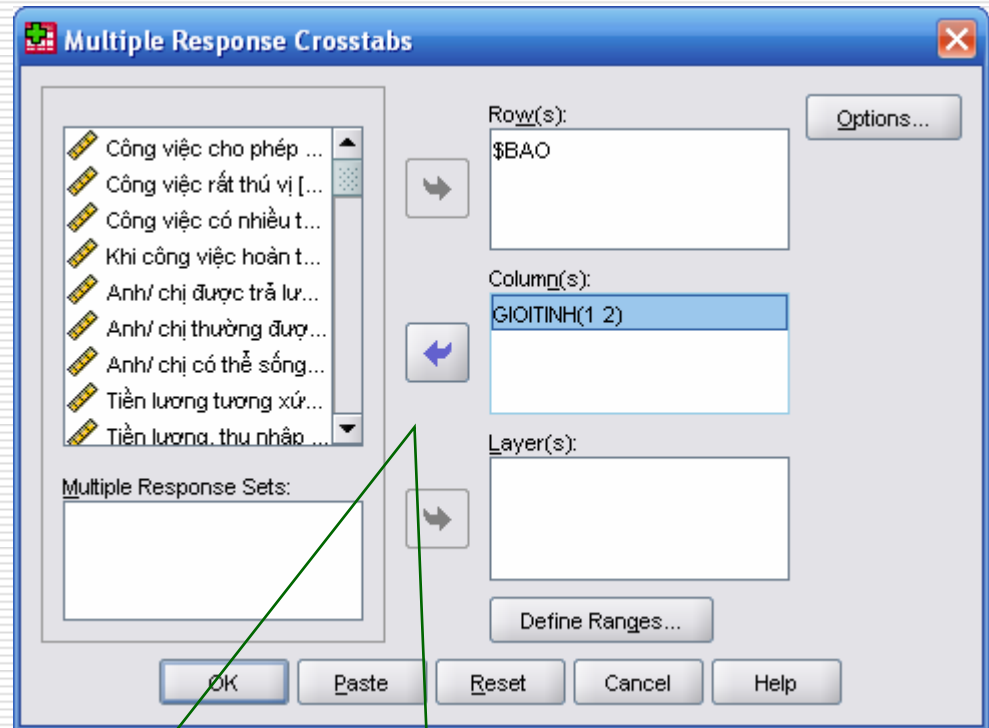
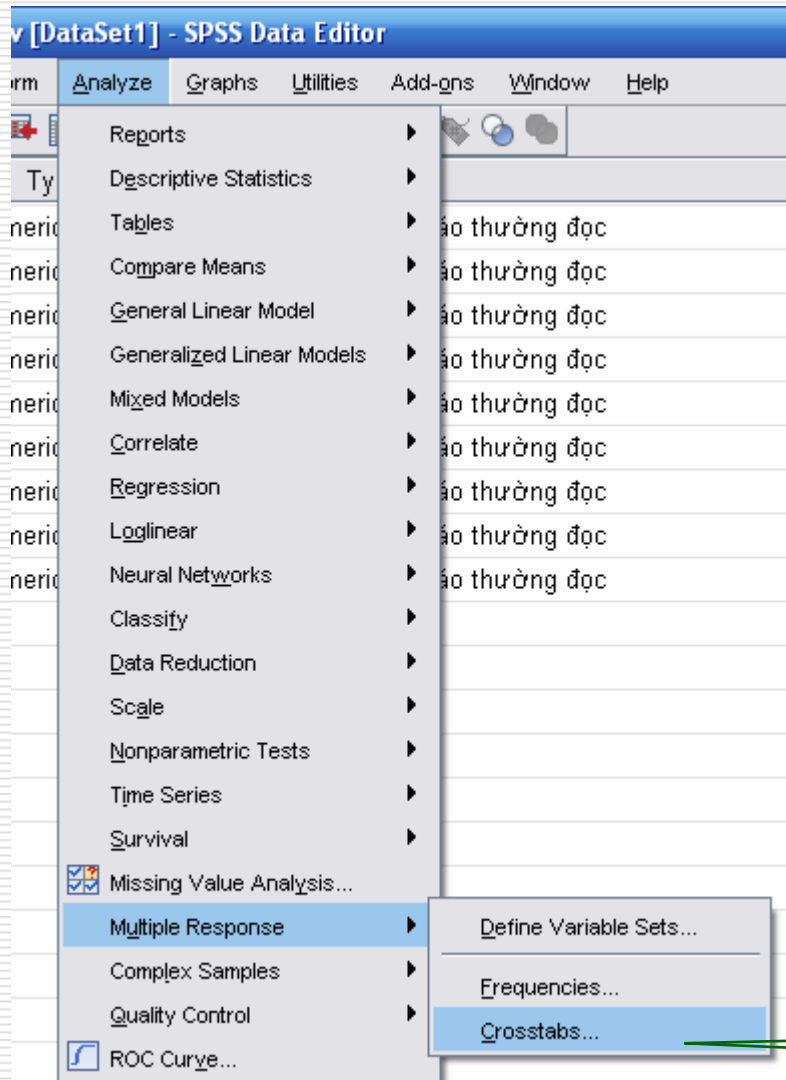
\$BAO Frequencies

		Responses		Percent of Cases
		N	Percent	
Báo thường đọc ^a	Hà Nội mới	134	15.3%	67.0%
	SGGP	6	.7%	3.0%
	Lao Động	54	6.2%	27.0%
	Người Lao động	18	2.1%	9.0%
	Tiền Phong	36	4.1%	18.0%
	Thanh Niên	24	2.7%	12.0%
	Tuổi trẻ	36	4.1%	18.0%
	Phụ nữ Việt Nam	80	9.1%	40.0%
	Thời báo KTSG	30	3.4%	15.0%
	Thời báo KT Việt Nam	6	.7%	3.0%
	Sài Gòn Tiếp thị	42	4.8%	21.0%
	Thế giới Phụ nữ	48	5.5%	24.0%
	Tiếp thị và cuộc sống	18	2.1%	9.0%
	Mua và bán	12	1.4%	6.0%
	An ninh thế giới	156	17.8%	78.0%
	An ninh thủ đô	170	19.4%	85.0%
	Công An TP HCM	6	.7%	3.0%
Total		876	100.0%	438.0%

Kết quả thống kê

a. Group

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN



**Chọn các biến vào phân tích
Nhân Define Ranges để định
nghĩa biến nominal ở cột**

Kết quả thống kê

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN

\$BAO*GIOITINH Crosstabulation

			Giới tính		Total
			Nam	Nu	
Báo thường đọc ^a	Hà Nội mới	Count	71	63	134
	SGGP	Count	1	5	6
	Lao Động	Count	33	21	54
	Người Lao động	Count	3	15	18
	Tiền Phong	Count	20	16	36
	Thanh Niên	Count	12	12	24
	Tuổi trẻ	Count	21	15	36
	Phụ nữ Việt Nam	Count	37	43	80
	Thời báo KTSG	Count	15	15	30
	Thời báo KT Việt Nam	Count	1	5	6
	Sài Gòn Tiếp thị	Count	18	24	42
	Thế giới Phụ nữ	Count	26	22	48
	Tiếp thị và cuộc sống	Count	12	6	18
	Mua và bán	Count	6	6	12
	An ninh thế giới	Count	79	77	156
	An ninh thủ đô	Count	88	82	170
	Công An TP HCM	Count	3	3	6
Total	Count	104	96	200	

Percentages and totals are based on respondents.

a. Group

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)

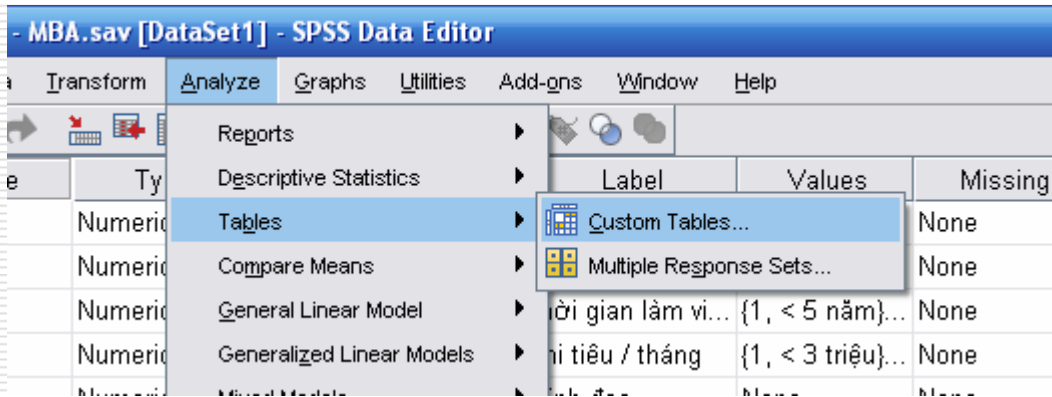
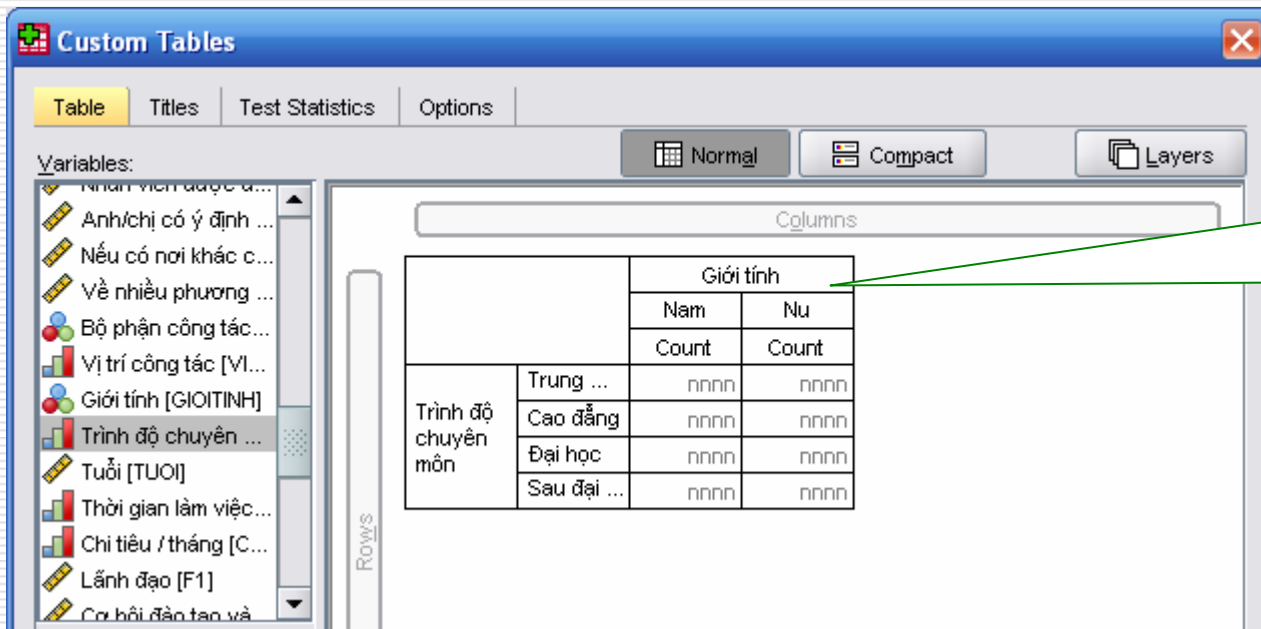


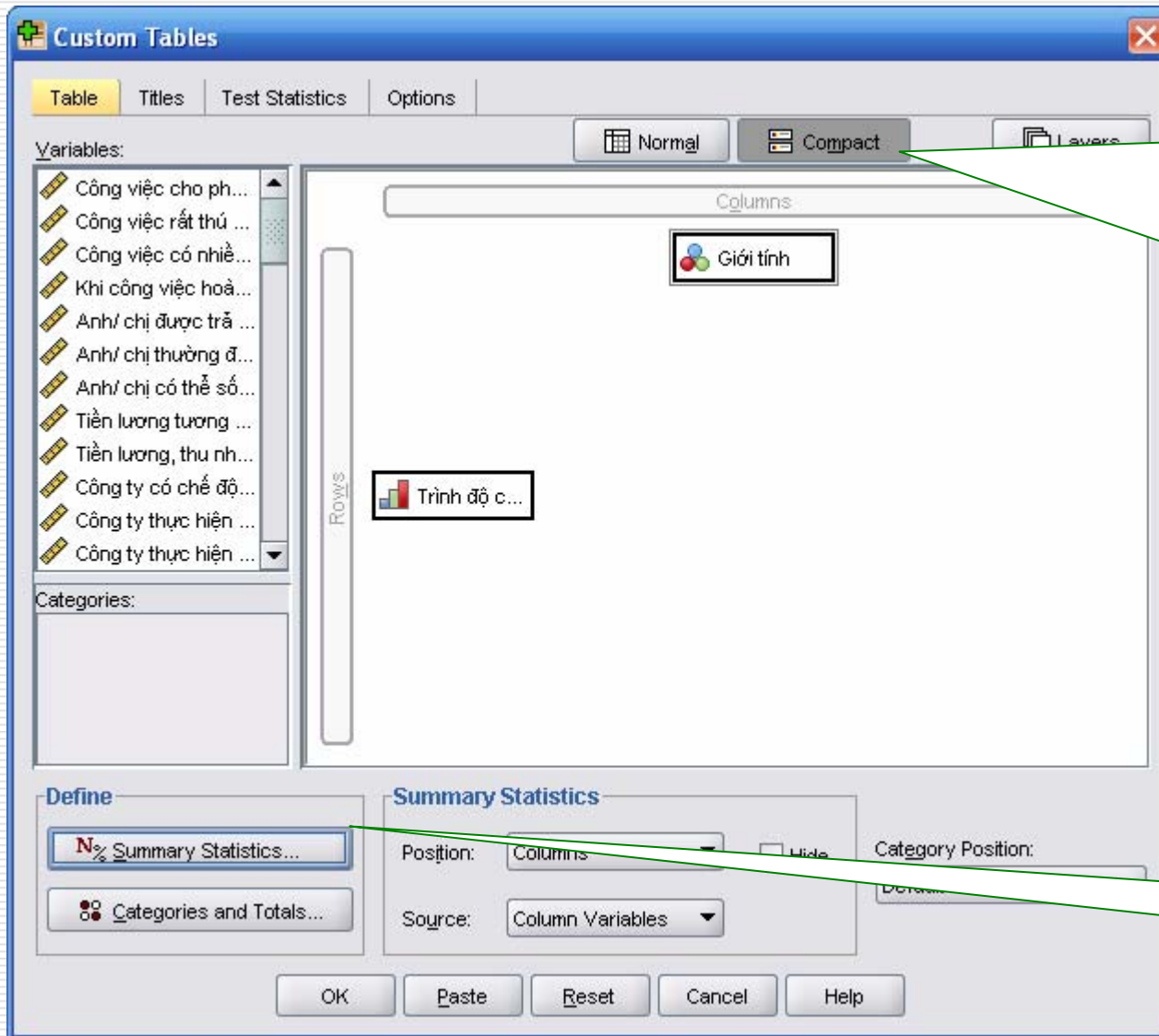
Table 1

		Giới tính	
		Nam	Nu
		Count	Count
Trình độ chuyên môn	Trung cấp	11	3
	Cao đẳng	3	8
	Đại học	87	82
	Sau đại học	3	3



**Kéo biến GIOITINH
sang ô cột, biến
TRINHDO sang ô
dòng**

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)

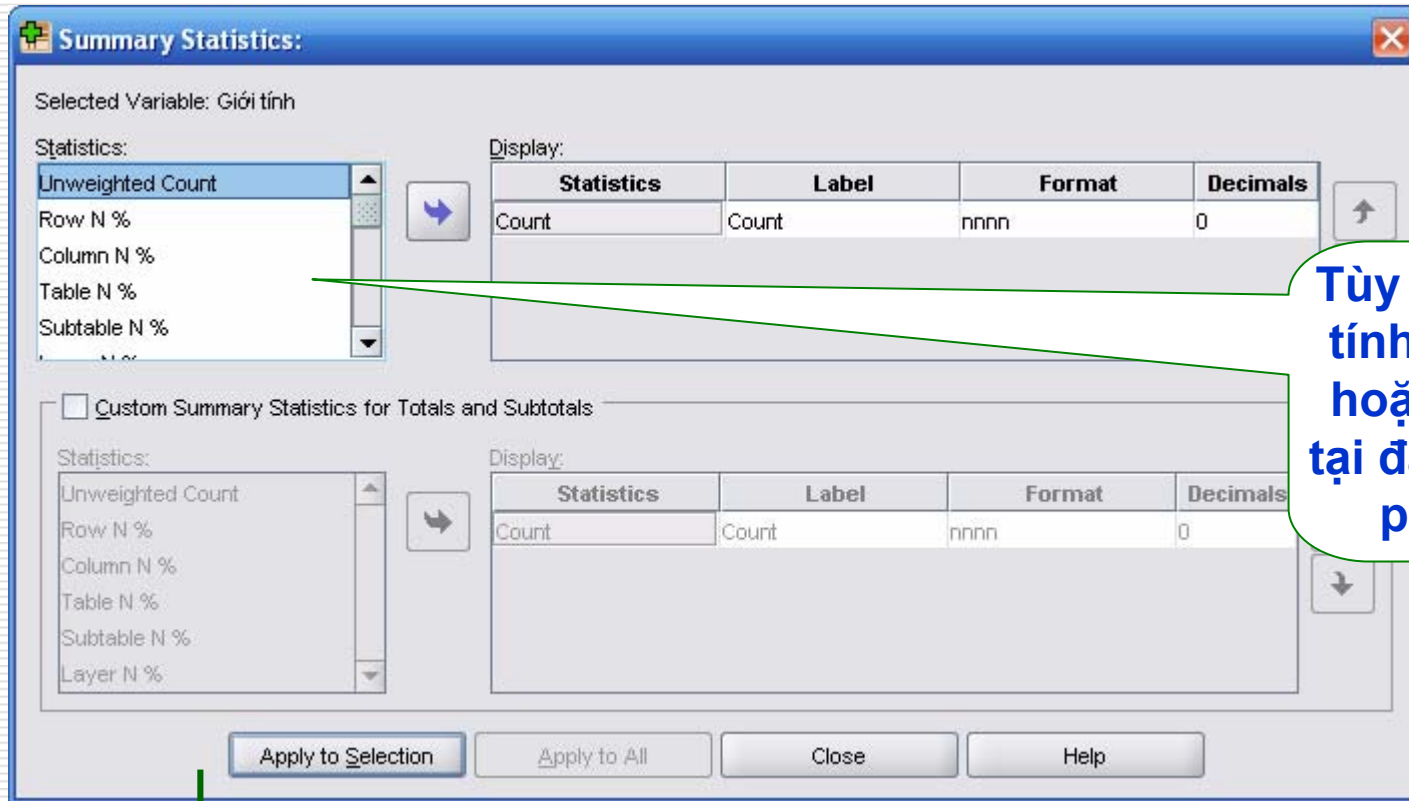


**Kéo biến GIOITINH
sang ô cột, biến
TRINHDO sang ô
dòng**

**Nếu tính thêm %
thì chọn Compact**

**Sau đó chọn
Summary Statistics**

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)



Tùy theo mục đích tính % theo hàng, hoặc cột thì chọn tại đây và kéo sang phần Display

Table 1

		Giới tính					
		Nam			Nu		
		Count	Row N %	Column N %	Count	Row N %	Column N %
Trình độ chuyên môn	Trung cấp	11	78.6%	10.6%	3	21.4%	3.1%
	Cao đẳng	3	27.3%	2.9%	8	72.7%	8.3%
	Đại học	87	51.5%	83.7%	82	48.5%	85.4%
	Sau đại học	3	50.0%	2.9%	3	50.0%	3.1%

Sau đó chọn Apply to Selection

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)

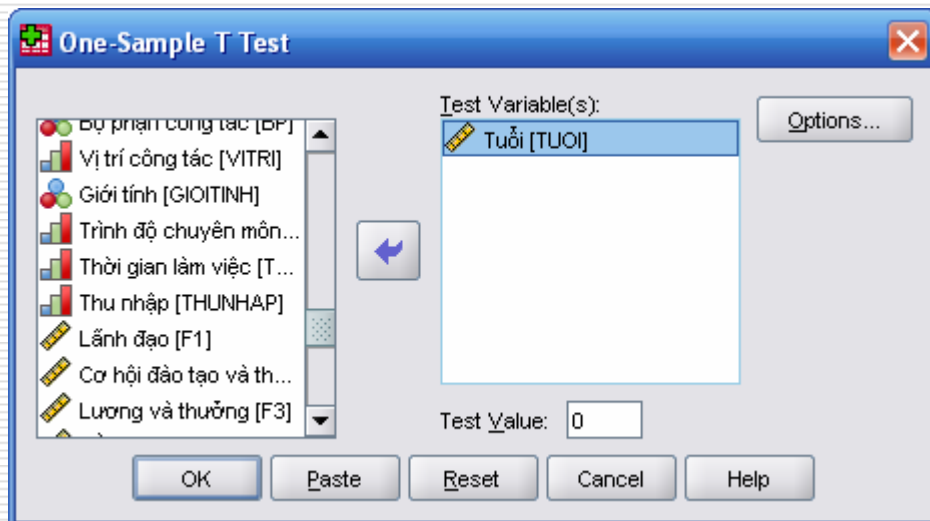
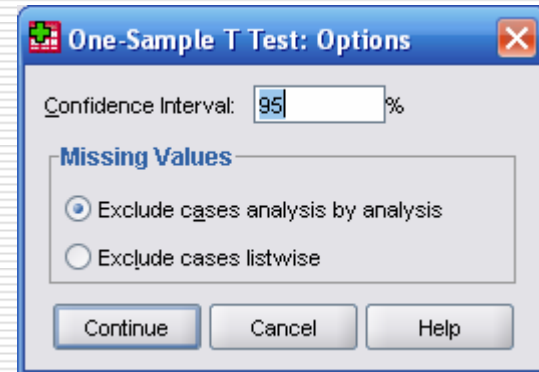
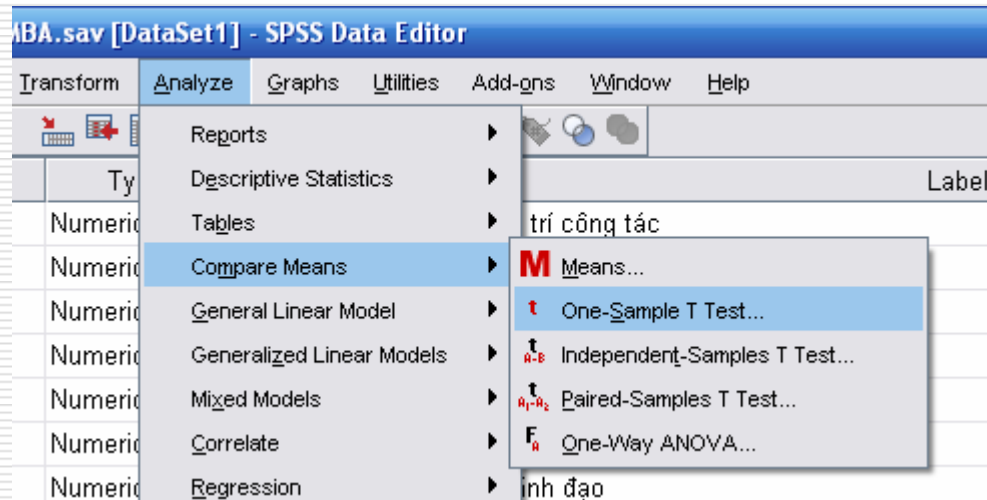
Table 1

		Giới tính					
		Nam			Nu		
		Count	Row N %	Column N %	Count	Row N %	Column N %
Trình độ chuyên môn	Trung cấp	11	78.6%	10.6%	3	21.4%	3.1%
	Cao đẳng	3	27.3%	2.9%	8	72.7%	8.3%
	Đại học	87	51.5%	83.7%	82	48.5%	85.4%
	Sau đại học	3	50.0%	2.9%	3	50.0%	3.1%

ƯỚC LƯỢNG THAM SỐ TRUNG BÌNH

ƯỚC LƯỢNG KHOẢNG TIN CẬY CỦA GIÁ TRỊ TRUNG BÌNH	
Đã biết phương sai tổng thể (σ^2)	Chưa biết phương sai tổng thể (σ^2)
<p><i>Phân phối chuẩn</i> σ: Độ lệch chuẩn của tổng thể μ: Trung bình tổng thể $\alpha = 1 - \gamma$: Mức ý nghĩa</p> $\bar{x} - \frac{\sigma}{\sqrt{n}} U_{1-\frac{\alpha}{2}} < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} U_{1-\frac{\alpha}{2}}$	<p><i>Phân phối Student với (n-1) là bậc tự do</i> s': Độ lệch chuẩn hiệu chỉnh mẫu</p> $\bar{x} - \frac{s'}{\sqrt{n}} t_{\frac{\alpha}{2}}^{n-1} < \mu < \bar{x} + \frac{s'}{\sqrt{n}} t_{\frac{\alpha}{2}}^{n-1}$
	<p><i>Phân phối chuẩn</i> s': Độ lệch chuẩn hiệu chỉnh mẫu</p> $\bar{x} - \frac{s'}{\sqrt{n}} U_{1-\frac{\alpha}{2}} < \mu < \bar{x} + \frac{s'}{\sqrt{n}} U_{1-\frac{\alpha}{2}}$

ƯỚC LƯỢNG THAM SỐ TRUNG BÌNH



ƯỚC LƯỢNG THAM SỐ TRUNG BÌNH

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Tuổi	200	26.47	3.161	.223

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Tuổi	118.442	199	.000	26.470	26.03	26.91

SO SÁNH 2 THAM SỐ TRUNG BÌNH

- Ước lượng tham số trung bình:
 - ⊕ Dữ liệu: định lượng (liên tục)
 - ⊕ Độ tin cậy (1-mức ý nghĩa)
- So sánh trung bình hai mẫu độc lập:
 - ⊕ H_0 : Không khác nhau giữa độ tuổi trung bình giữa người nữ và người nam trong tổng thể
 - H_1 : Có sự khác nhau giữa độ tuổi trung bình giữa người nữ và người nam trong tổng thể
 - ⊕ Điều kiện ứng dụng:
 - ◆ Dữ liệu định lượng (liên tục)
 - ◆ Dữ liệu của hai nhóm phải tuân theo quy luật phân phối chuẩn

SO SÁNH 2 THAM SỐ TRUNG BÌNH

KIỂM ĐỊNH THAM SỰ KHÁC NHAU HAI TRUNG BÌNH TỔNG THỂ

(dựa trên sự phân phối từng cặp)

1. Giả thiết và đối thiết:

	Đôi xứng	Phải	Trái
Giả thiết	$H_0: \mu_x - \mu_y = D_0$	$H_0: \mu_x - \mu_y \leq D_0$	$H_0: \mu_x - \mu_y \geq D_0$
Đối thiết	$H_1: \mu_x - \mu_y \neq D_0$	$H_1: \mu_x - \mu_y > D_0$	$H_1: \mu_x - \mu_y < D_0$

2. Xác định mức ý nghĩa




3. Phương pháp kiểm nghiệm sự khác nhau của hai trung bình tổng thể - Bảng phân phối chuẩn hoặc T-student (nếu $n < 30$)

4. Tiêu chuẩn kiểm định T hoặc U:

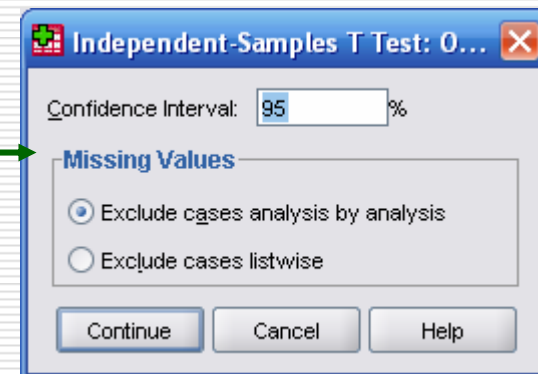
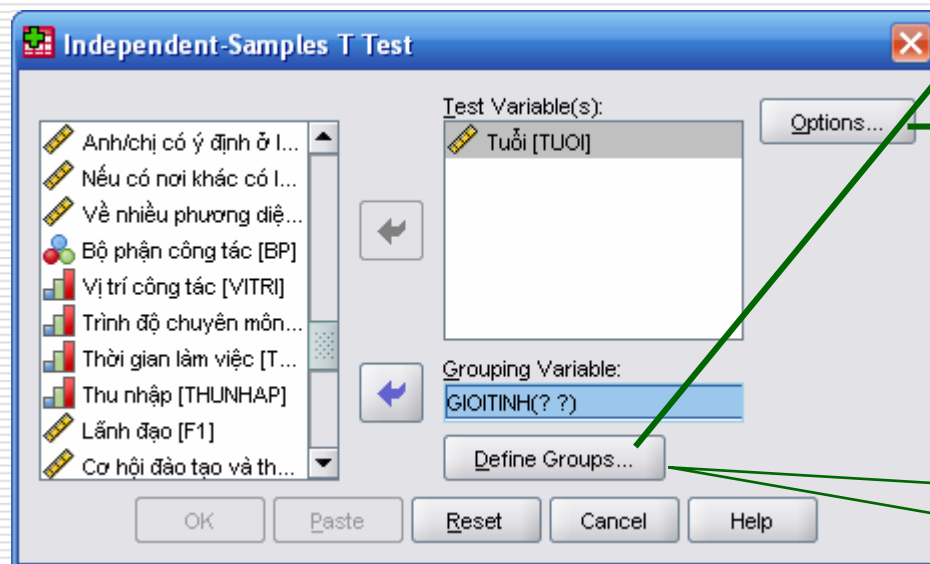
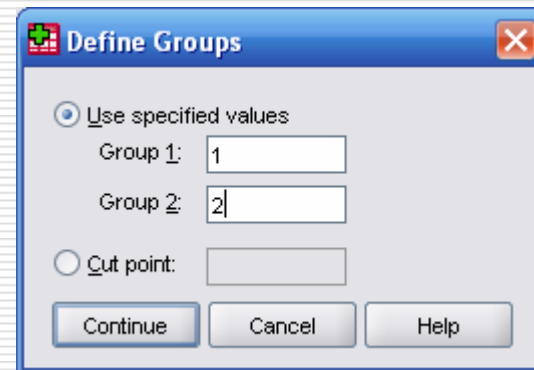
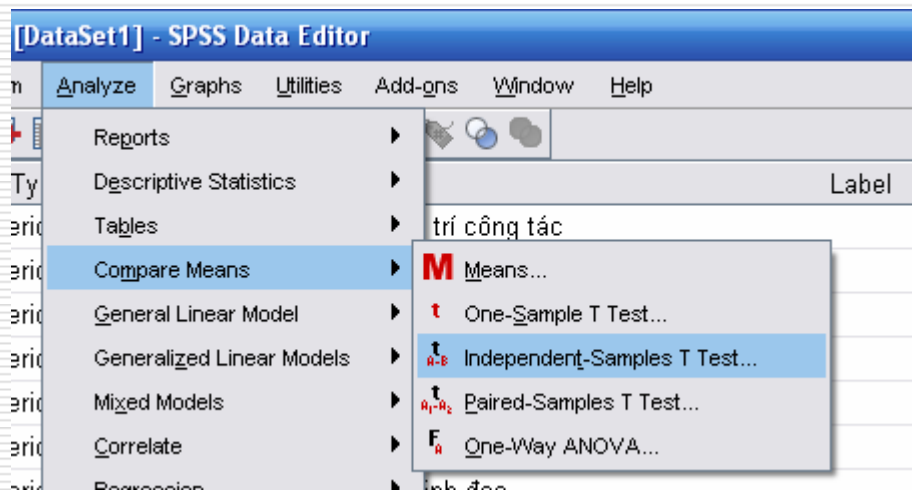
$$k_{qr} \equiv T = \frac{(\bar{x} - D_0)\sqrt{n}}{s'_d}$$

\bar{x} và s'_d là trung bình và độ lệch chuẩn điều chỉnh của n khác biệt

5. Điểm tới hạn và miền bác bỏ:

	Đôi xứng	Phải	Trái
Điểm tới hạn	$-T_{(n-1);1-\alpha/2}$ và $T_{(n-1);1-\alpha/2}$	$T_{(n-1);1-\alpha}$	$-T_{(n-1);1-\alpha}$
Miền bác bỏ	$D < -T_{(n-1);1-\alpha/2}$ và $D > T_{(n-1);1-\alpha/2}$	$D > T_{(n-1);1-\alpha}$	$D < -T_{(n-1);1-\alpha}$
Mô hình			

TRÌNH TỰ THỰC HIỆN



**Chọn các biến phân tích
Nhấn Define Group để Đ.ng**

KẾT QUẢ PHÂN TÍCH

Group Statistics

	Giới tính	N	Mean	Std. Deviation	Std. Error Mean
Tuổi	Nam	104	27.06	3.353	.329
	Nu	96	25.83	2.820	.288

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Tuổi	Equal variances assumed	.345	.558	2.783	198	.006	1.224	.440	.357	2.092
	Equal variances not assumed			2.802	196.340	.006	1.224	.437	.363	2.086

Nếu sig. trong kiểm định phương sai > 0,05 thì phương sai giữa hai mẫu bằng nhau, ta sẽ dùng kết quả kiểm định t ở dòng thứ 1

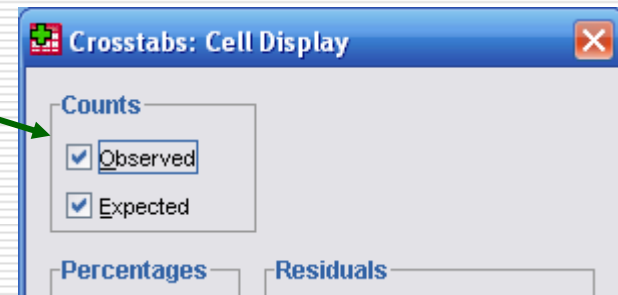
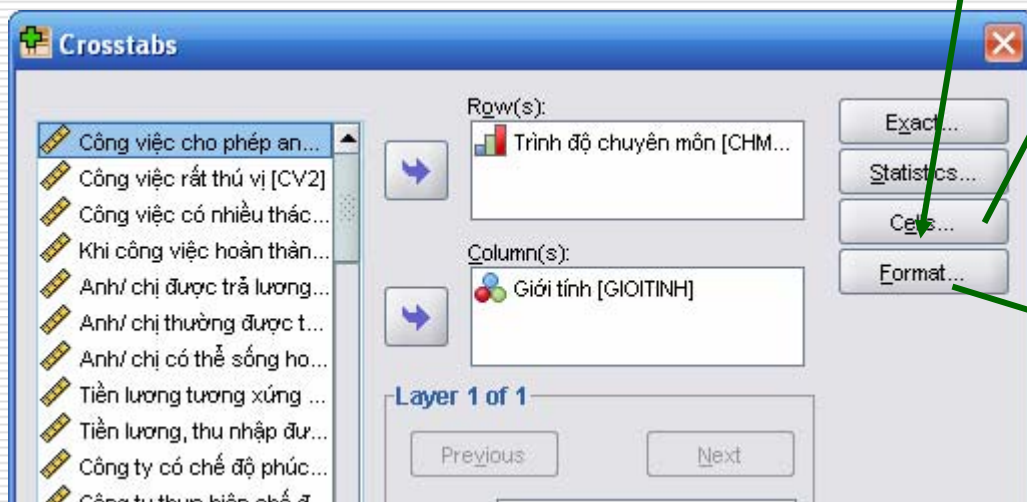
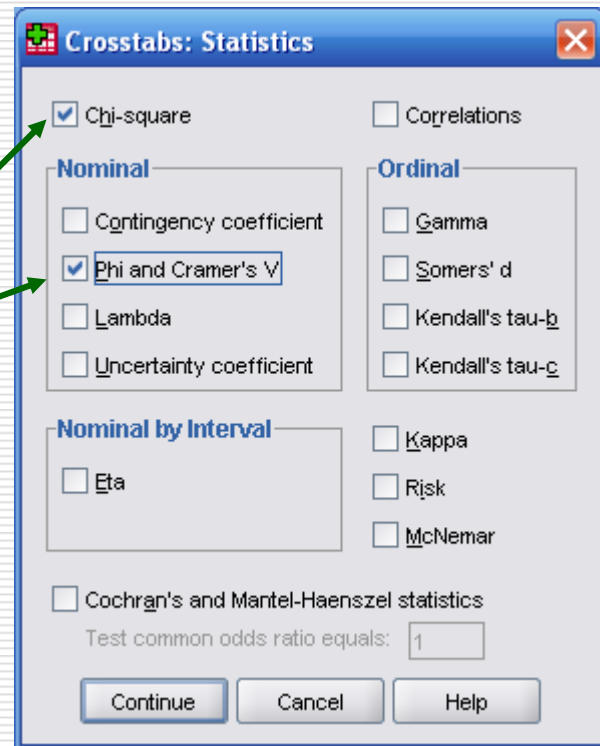
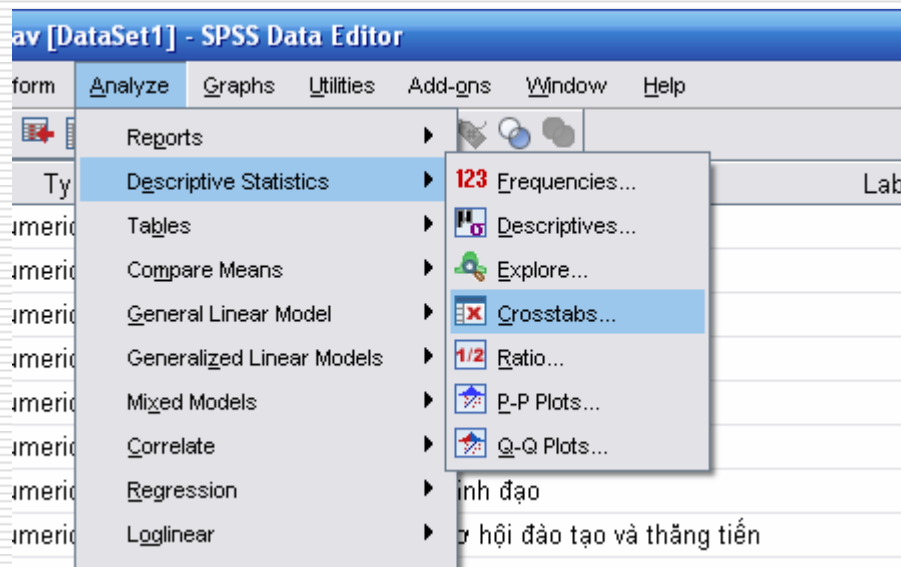
Giá trị t của kiểm định

p-value của giá trị t

PHÂN TÍCH MQH GIỮA 2 BIẾN ĐỊNH TÍNH

- Điều kiện (hai biến định tính)
- Các giả thuyết:
 - H_0 : Hai biến độc lập với nhau (hai biến không có MQH)*
 - H_1 : Hai biến có liên quan với nhau*
- χ^2 được thiết lập để xác định có hay không một mối liên hệ giữa hai biến, nhưng nó không chỉ ra được cường độ của mối liên hệ đó.
- χ^2 cho phép tìm ra những mối liên hệ phi tuyến tính
- Cramer-V: Cường độ của nó biến động từ 0 đến 1.

PHÂN TÍCH MQH GIỮA 2 BIẾN ĐỊNH TÍNH



PHÂN TÍCH MQH GIỮA 2 BIẾN ĐỊNH TÍNH

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.183	.083
	Cramer's V	.183	.083
N of Valid Cases		200	

Trình độ chuyên môn * Giới tính Crosstabulation

			Giới tính		Total
			Nam	Nu	
Trình độ chuyên môn	Trung cấp	Count	11	3	14
		Expected Count	7.3	6.7	14.0
	Cao đẳng	Count	3	8	11
		Expected Count	5.7	5.3	11.0
	Đại học	Count	87	82	169
		Expected Count	87.9	81.1	169.0
	Sau đại học	Count	3	3	6
		Expected Count	3.1	2.9	6.0
Total		Count	104	96	200
		Expected Count	104.0	96.0	200.0