



Thống kê mô tả bằng SPSS

1. Trình bày thông qua các tham số đặc trưng

1.1. Tổng quan

1.1. Các mức độ trung tâm

- Số bình quân (trung bình)
- Mốt (M_0)
- Số trung vị (M_e)

1.2. Phân vị

1.3. Độ phân tán

- Khoảng biến thiên: $R = X_{\max} - X_{\min}$
- Phương sai
- Độ lệch chuẩn
- Sai số chuẩn trung bình: $\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$
 - Gọi là: sai số bình quân chọn mẫu (LTTK 1)

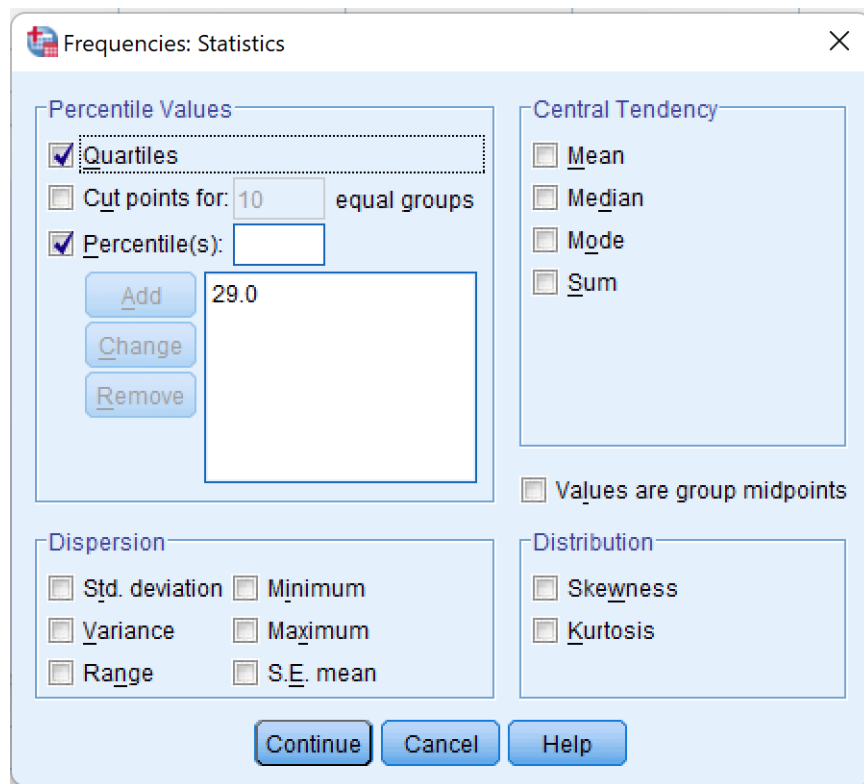
1.4. Phản ánh phân phối

- Hệ số bất đối xứng (Skewness): càng gần 0 thì càng chuẩn
- Hệ số độ nhọn: (Kurtosis): càng gần 3 (điều chỉnh thì càng gần 0) thì càng chuẩn

1.5. SPSS

▼ Cách 1: Frequencies

Analyze → Descriptive Statistics → Frequencies

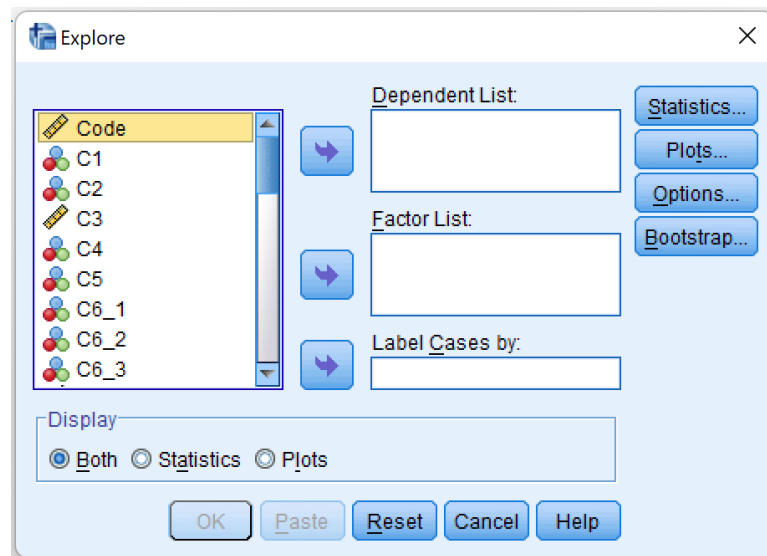


Nhớ tắt **display frequency tables**

▼ Cách 2: Explore

Analyze → Descriptive Statistics → Explore ...

- Ưu điểm: có thể phân loại theo factor
- Nhược điểm: ít lựa chọn thống kê hơn

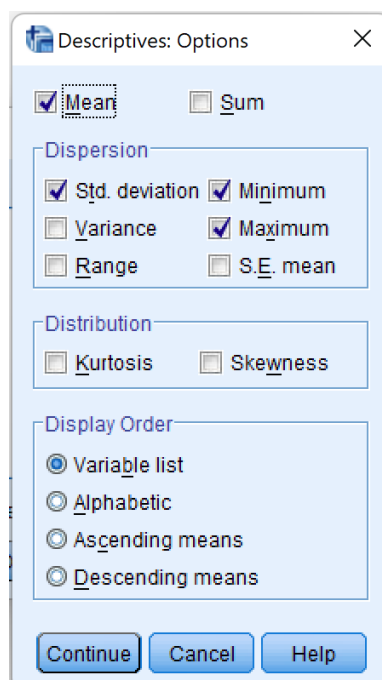


95% Confidence Interval for Mean	Lower Bound	13231.88
	Upper Bound	14487.36

Giá trị này chênh lệch càng lớn thì càng có lượng biến động xuất

▼ Cách 3: Descriptives

Analyze → Descriptive Statistics → Descriptives ...

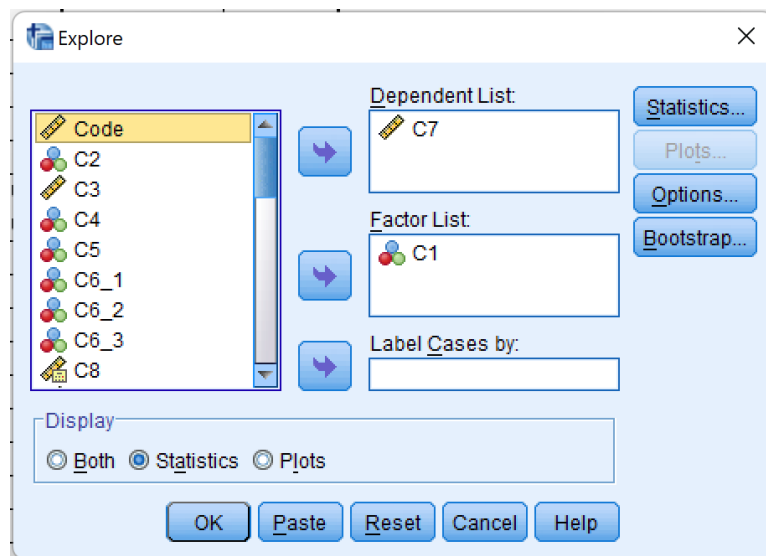


▼ Bài tập thực hành

1. Tính tiền lương trung bình? Mốt và Trung vị về tiền lương

2. Tính tiền lương trung bình cho từng khu vực

▼ Giải

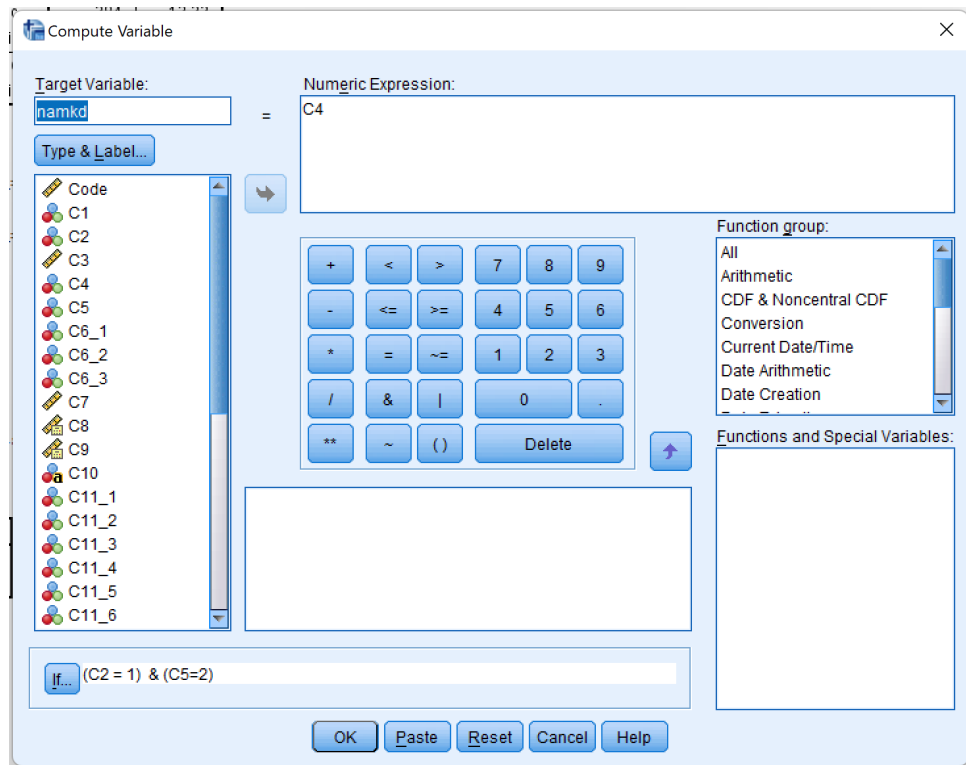


3. Tính số năm đi học trung bình với các khách hàng nam là nhân viên kinh doanh?

▼ Giải

Cách 1: dùng select case

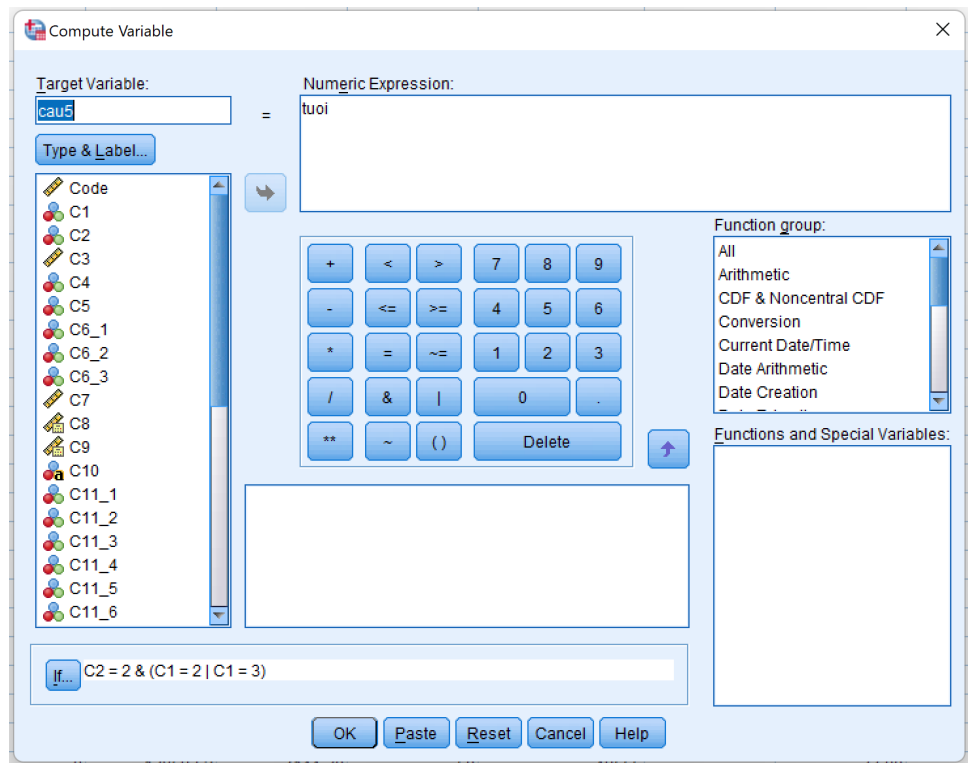
Cách 2: dùng compute



4. Tính tuổi trung bình của khách hàng nữ sống ở miền Nam và miền Trung

▼ Giải

- Cách 1: Dùng select case
- Cách 2 (tốt hơn): Dùng compute



5. Tính phân vị mức 30 của tuổi

▼ Giải

Statistics

Tuoi cua khach hang

N	Valid	500
	Missing	0
Percentiles	30	34.0000

6. Tính tuổi trung bình cho 30% KH có tuổi thấp nhất

▼ Giải

Statistics

cau6

N	Valid	161
	Missing	339
Mean		31.5342

7. Tính tiền lương trung bình cho 20% KH có tuổi thấp nhất

▼ Giải

Statistics

cau7

N	Valid	127
	Missing	373
Mean		11313.1024

8. Tính số năm đi học trung bình cho 10% khách hàng có tiền lương cao nhất

▼ Giải

- Phân vị mức 90: 24240

Statistics

cau8

N	Valid	50
	Missing	450
Mean		17.9400

2. Sắp xếp và phân tổ thống kê

- Sắp xếp dữ liệu

- Data → Sort Cases

- Đưa biến nào trước thì nó sẽ ưu tiên sắp xếp theo biến đó trước
 - Mặc định theo chiều tăng dần

- Sơ đồ thân lá

- Analyze → Descriptive Statistics → Explore

- Box plot

- Analyze → Descriptive Statistics → Explore

- Phân tổ thống kê

- ▼ Đối với biến thuộc tính hoặc biến định lượng có ít biểu hiện

- **Analyze** → **Descriptive Statistics** → **Frequencies** ...
- Chọn display frequency table (chọn biến thuộc tính)

Khu vực sinh sống

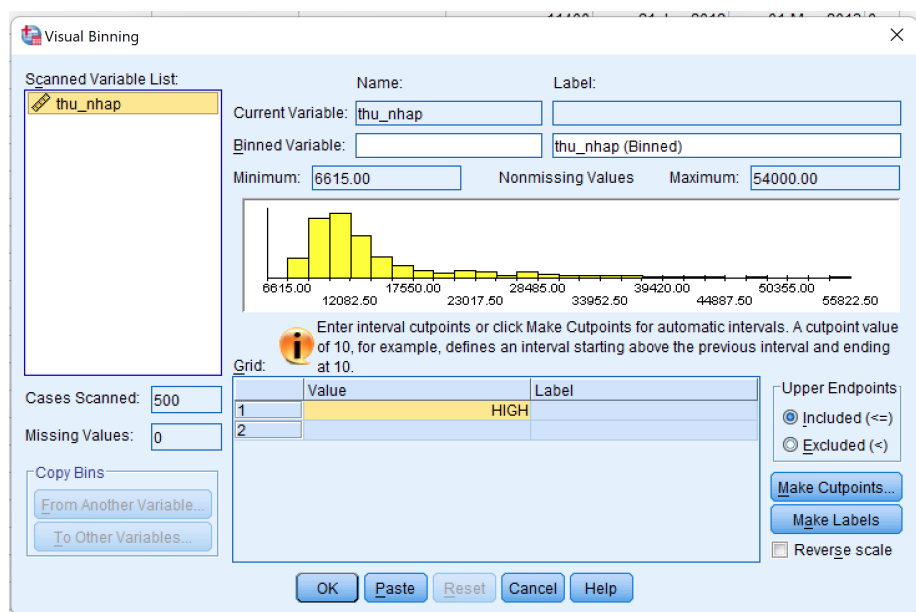
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Miền Bắc	183	36.6	36.9	36.9
	Miền Nam	206	41.2	41.5	78.4
	Miền Trung	107	21.4	21.6	100.0
	Total	496	99.2	100.0	
Missing	System	4	.8		
Total		500	100.0		

- Cột percent tính cho cả missing (thường không có ý nghĩa)
- Cột valid percent thì chỉ tính cho valid

▼ Đối với biến định lượng có nhiều biểu hiện

- Tiền hành tạo biến mới (recode, compute)
- Tạo biến tự động (chỉ dùng được khi phân tổ có khoảng cách = nhau)
 - **Transform** → **Visual Binning**

1. Đặt tên với label



2. Chọn cận trên: Nguyên tắc đối với thống kê (không bao gồm cận trên) → phải chọn **Excluded (<)**

3. Chọn **make cutpoints**, chọn **first cutpoint location** (điểm cắt đầu tiên), **number of cut points** (số điểm cắt), **width** (khoảng cách tổ)

Make Cutpoints

☒ **Equal Width Intervals**

Intervals - fill in at least two fields

First Cutpoint Location:

Number of Cutpoints:

Width:

Last Cutpoint Location:

☐ **Equal Percentiles Based on Scanned Cases**

Intervals - fill in either field

Number of Cutpoints:

Width(%):

☐ **Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases**

☐ +/- 1 Std. Deviation

☐ +/- 2 Std. Deviation

☐ +/- 3 Std. Deviation

Apply will replace the current cutpoint definitions with this specification.
A final interval will include all remaining values: N cutpoints produce N+1 intervals.

Apply Cancel Help

- ▼ Giả sử phân C7 (tiền lương) thành 3 tổ (< 10 , $10 - 15$, $15 - 20$, ≥ 20)

Intervals - fill in at least two fields

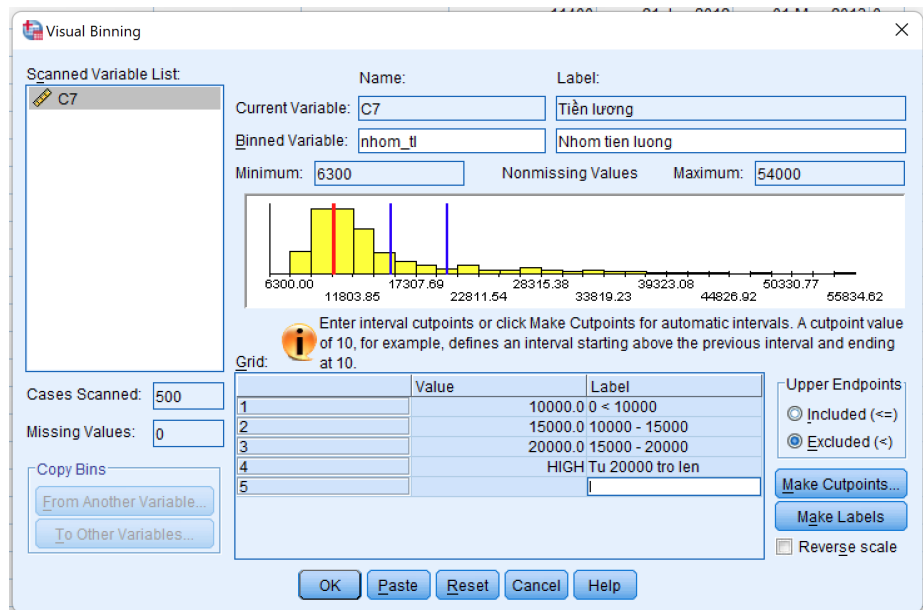
First Cutpoint Location:

Number of Cutpoints:

Width:

Last Cutpoint Location:

4. Sửa lại tổ → OK



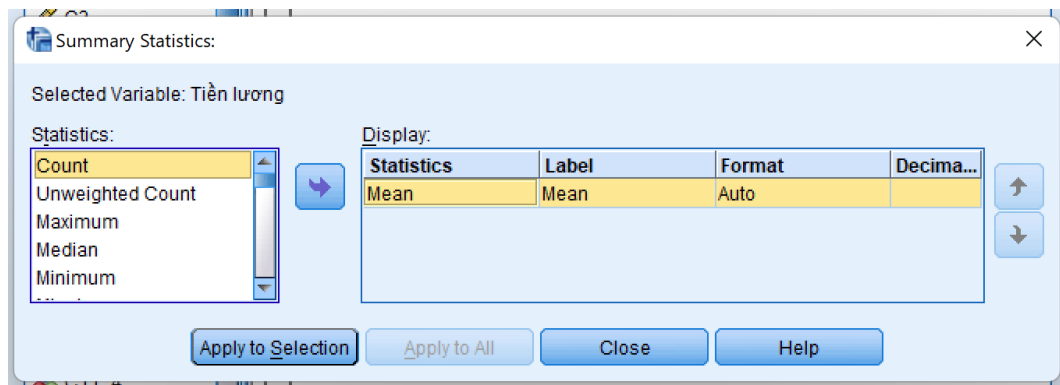
3. Lập bảng thống kê

▼ Lập bảng tần số cho 1 biến thuộc tính

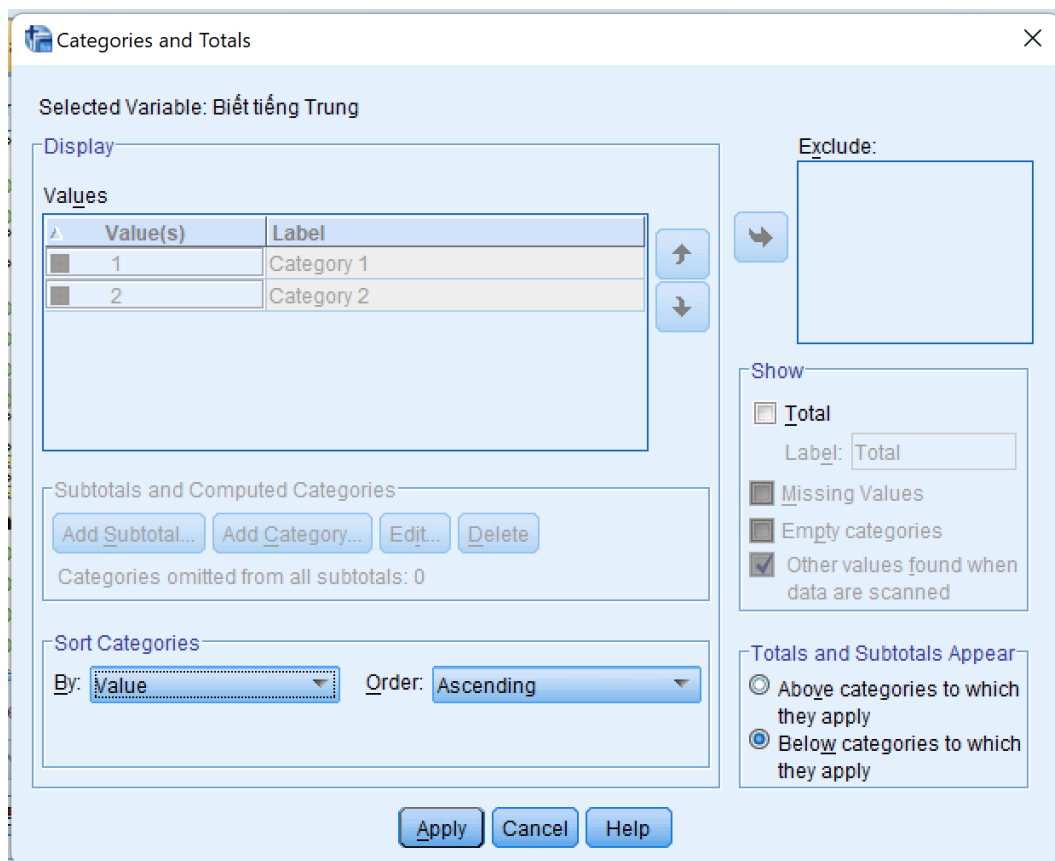
Analyze → Tables → Custom tables ...

Lập bảng theo dòng hoặc theo cột

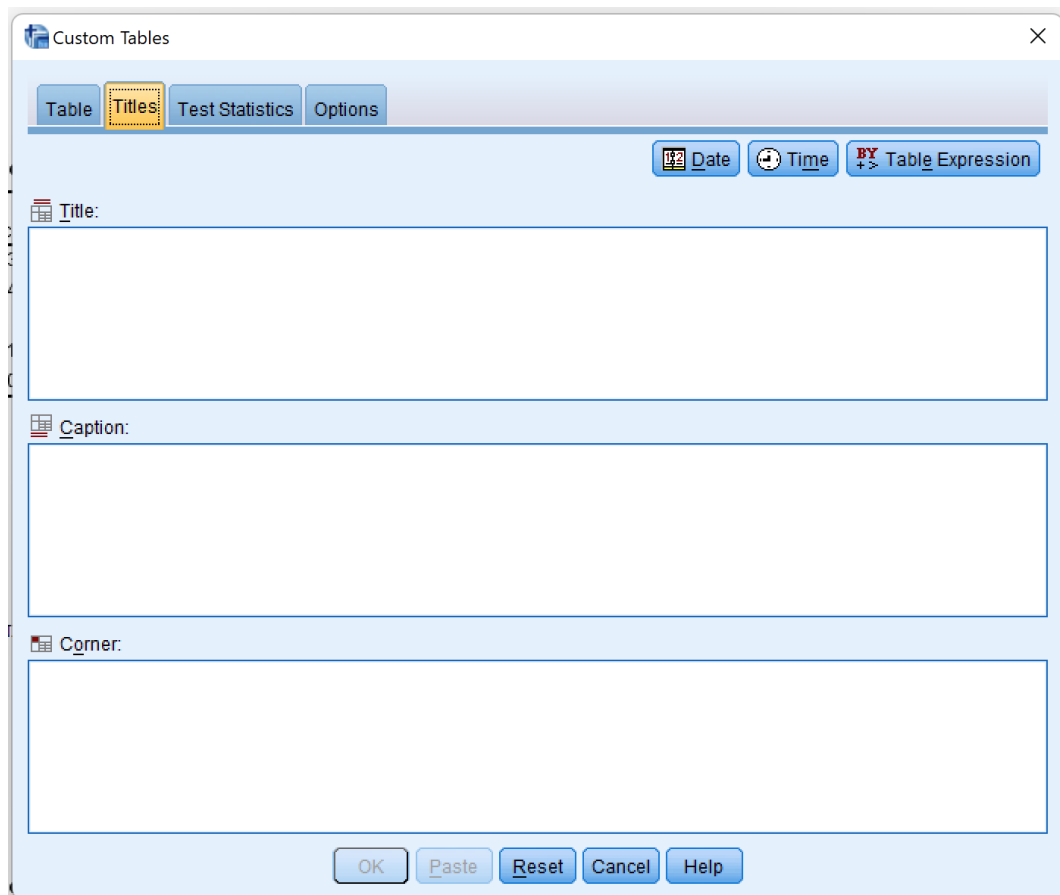
Chọn các thống kê ở Summary Statistics, ấn nhấn tên biến thì chọn Hide



Chọn thống kê, đặt tên cột (dòng) cho thống kê đó, ...



Chọn sum total, thứ tự dòng, bỏ các biểu hiện, ...

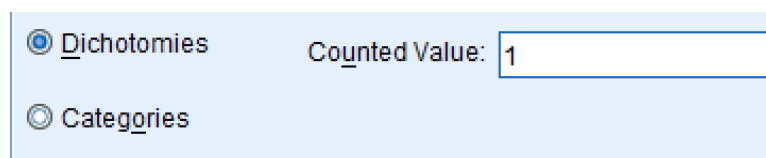
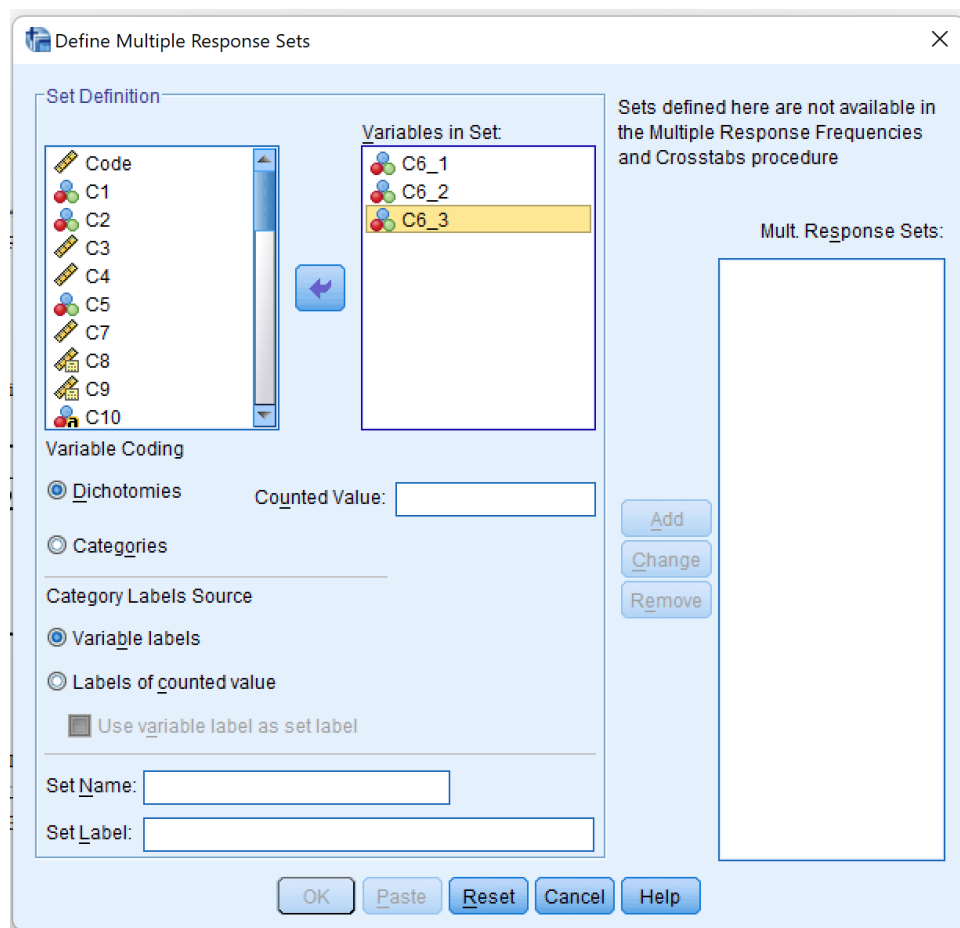


Đổi tên bảng, ghi chú, ...

▼ Lập bảng cho biến tùy chọn (Multiple answer)

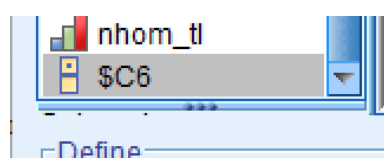
Analyse → Tables → Multiple Response Set

▼ Ví dụ - người biết tiếng anh, trung, pháp



Multiple Response Sets

Name	Label	Coded As	Counted Value	Data Type	Elementary Variables
\$C6	Biet ngoai ngu	Dichotomies	1	Numeric	Biết tiếng Anh Biết tiếng Pháp Biết tiếng Trung

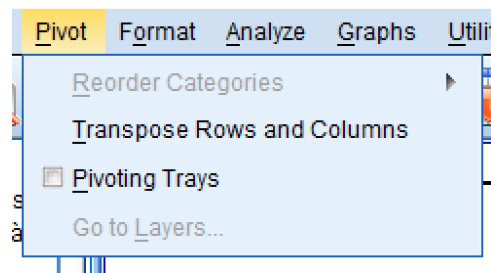


		Count	Column N %
Biet ngoai ngu	Biết tiếng Anh	241	87.0%
	Biết tiếng Pháp	67	24.2%
	Biết tiếng Trung	66	23.8%
	Total	277	100.0%

▼ Xoay bảng

Pivot controls → Pivoting Tray

		Giới tính					
		Nam		Nữ		CHUNG	
		So luong	Ty trong (%)	So luong	Ty trong (%)	So luong	Ty trong (%)
Phan to theo tuoi	Thanh nien	1	.37	27	11.79	28	5.60
	Trung nien	244	90.04	162	70.74	406	81.20
	Cao nien	26	9.59	40	17.47	66	13.20
	CHUNG	271	100.00	229	100.00	500	100.00



▼ Ví dụ

▼ 1. Tính số lượng và tỷ trọng khách hàng theo từng trình độ

Số lượng và tỷ trọng khách hàng theo từng trình độ

		So luong	Ty trong (%)
Bang cap cua khách hàng	Chung	500	100.00
	Chua qua dao tao	257	51.40
	Cu nhan / ky su	129	25.80
	Tren dai hoc	114	22.80

Nguồn: Tổng cục thống kê

▼ 2. Tính số lượng theo giới tính và nhóm tuổi, tính tỷ trọng

		Giới tính					
		Nam		Nữ		CHUNG	
		So lương	Tỷ trọng (%)	So lương	Tỷ trọng (%)	So lương	Tỷ trọng (%)
Phân to theo tuổi	Thanh niên	1	.37	27	11.79	28	5.60
	Trung niên	244	90.04	162	70.74	406	81.20
	Cao niên	26	9.59	40	17.47	66	13.20
	CHUNG	271	100.00	229	100.00	500	100.00

▼ 3. Lập bảng thống kê biểu hiện tiền lương trung bình phân theo giới tính và khu vực

		Giới tính		
		Nam	Nữ	CHUNG
		Tiền lương	Tiền lương	Tiền lương
Khu vực sinh sống	Miền Bắc	11.043,34	10.466,85	10.750,42
	Miền Nam	20.468,26	15.786,00	19.586,10
	Miền Trung	8.745,00	8.142,92	8.209,81
	CHUNG	16.784,32	10.398,51	13.859,62

▼ 4. Lập bảng thống kê biểu hiện số lượng khách hàng và tiền lương trung bình phân theo giới tính và trình độ

		Giới tính											
		Nam				Nữ				CHUNG			
		Tiền lương				Tiền lương				Tiền lương			
		Tiền lương BQ	So lương	Median	Mode	Tiền lương BQ	So lương	Median	Mode	Tiền lương BQ	So lương	Median	Mode
Bảng cấp của khách hàng	Chưa qua đào tạo	11.678,11	90	11738	12300	9.439,84	167	9240	9780	10.223,67	257	9900	12300
	Cu nhân / kỹ sư	13.289,18	93	12480	10380	10.786,11	36	10440	9660	12.590,65	129	12060	11100
	Tên đại học	25.700,32	88	24875	16080	16.019,38	26	14880	9240	23.492,39	114	22250	16080
	CHUNG	16.784,32	271	12960	12300	10.398,51	229	9660	9780	13.859,62	500	11520	12300

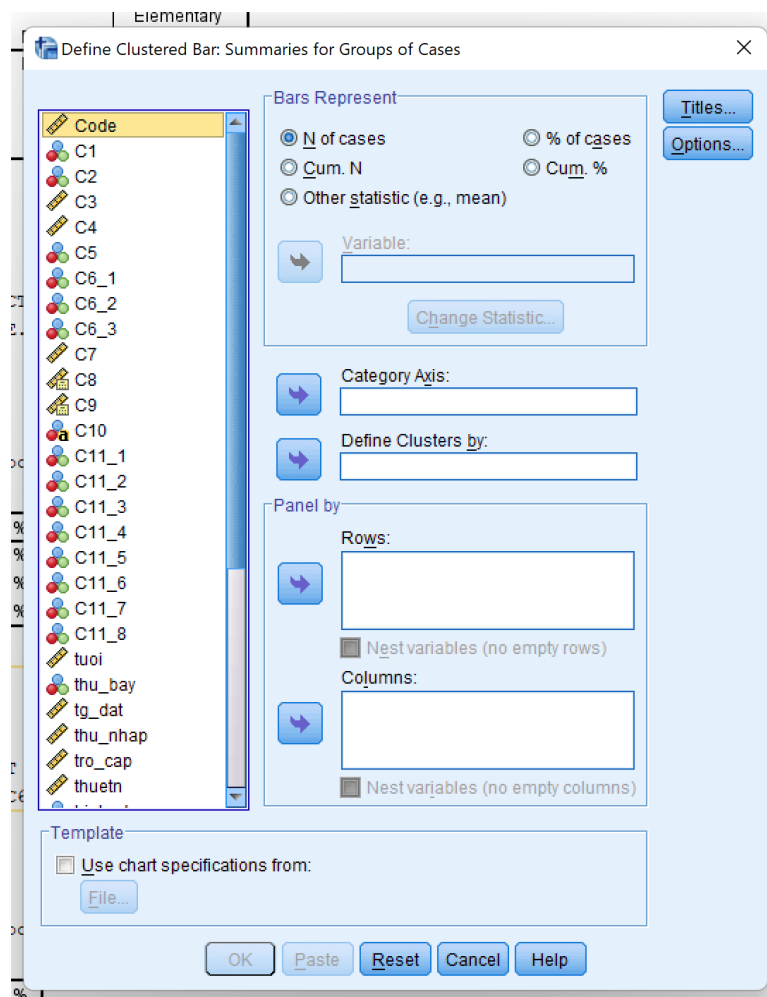
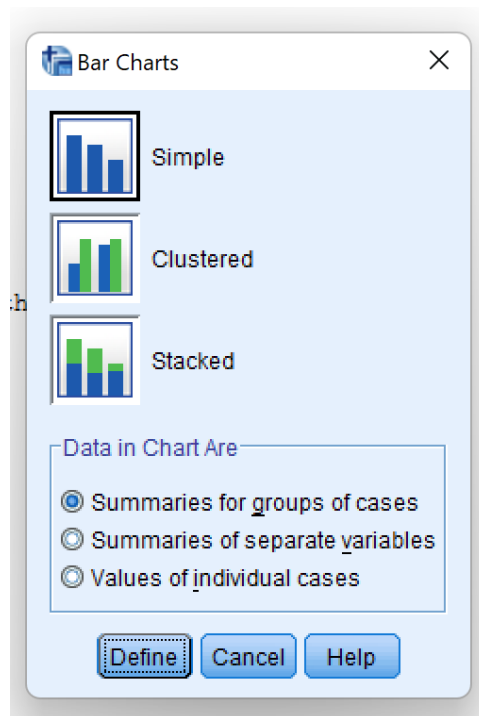
▼ 5. Lập bảng thống kê biểu hiện mức trợ cấp (số lượng khách hàng, tỷ trọng và mức trợ cấp trung bình) theo trình độ của khách hàng

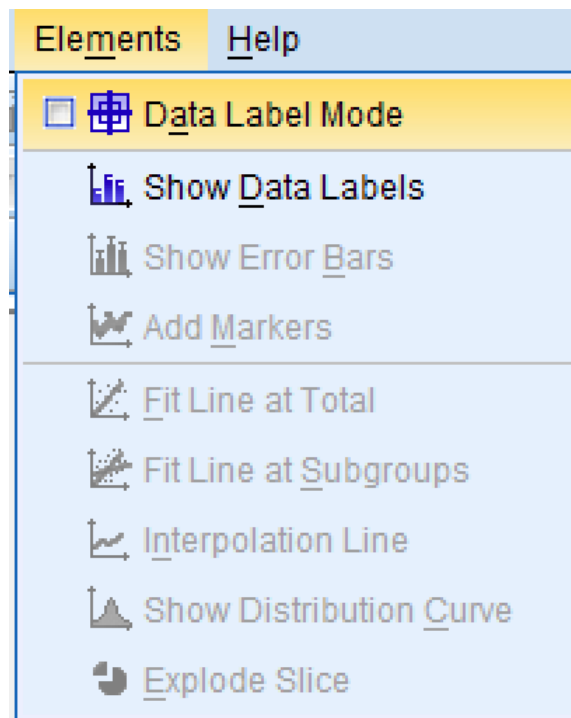
		tro_cap		
		Tro cap BQ	So lương	Tỷ trọng %
Bảng cấp của khách hàng	Chưa qua đào tạo	471.99	167	72.9%
	Cu nhân / kỹ sư	539.31	36	15.7%
	Tên đại học	800.97	26	11.4%
	Chung	519.93	229	100.0%

4. Đồ thị thống kê

▼ Đồ thị

Graphs → Legacy Dialogs → Bar ...

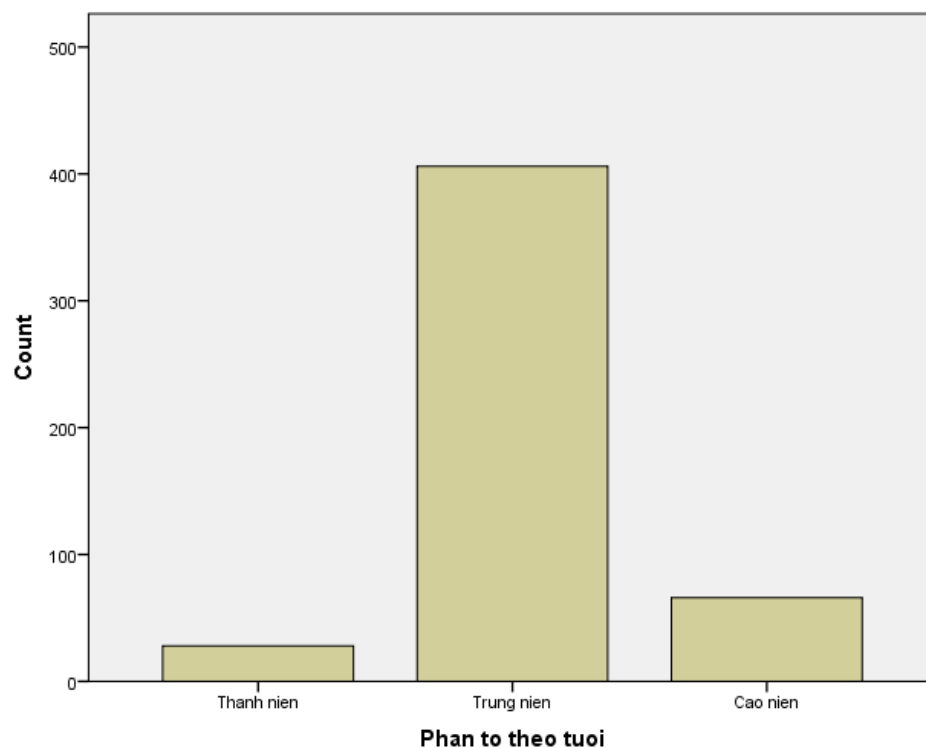




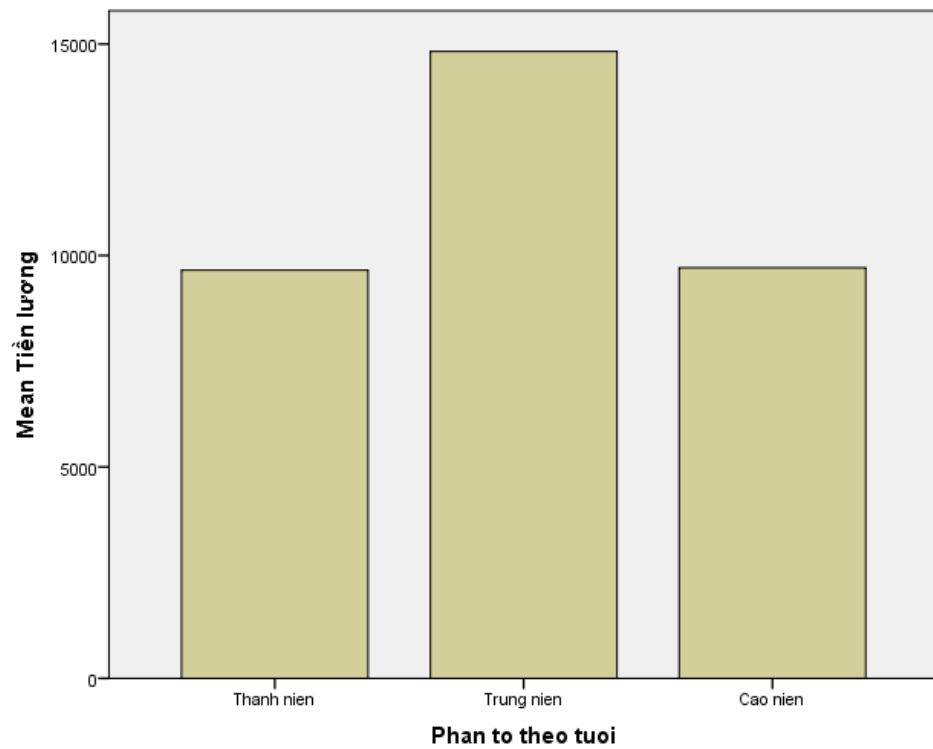
Thêm data label

▼ Ví dụ

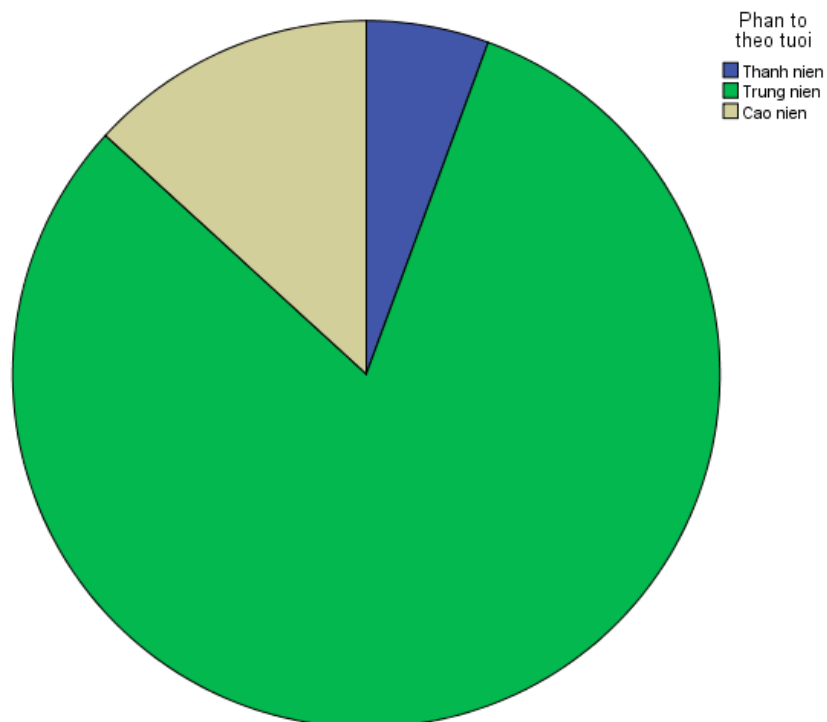
▼ Số lượng khách hàng theo nhóm tuổi



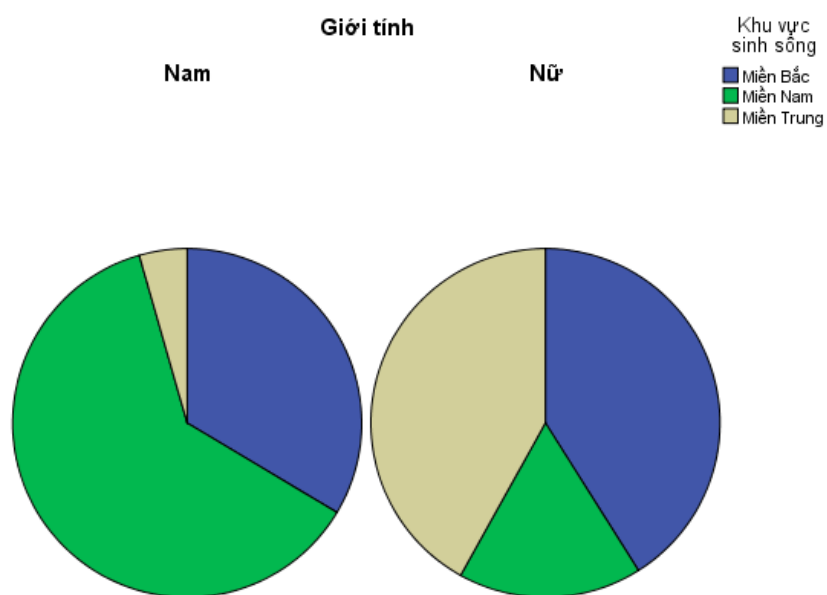
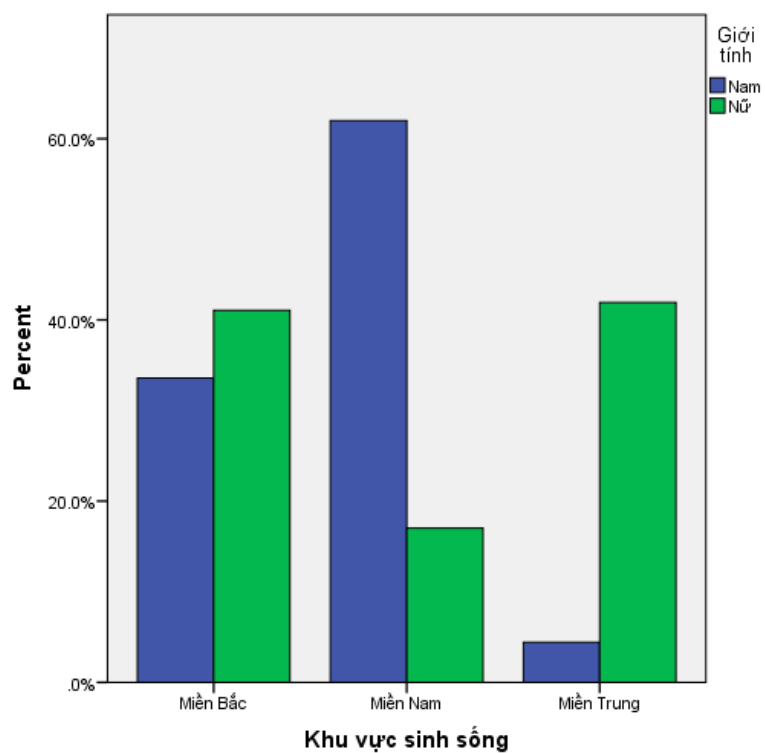
▼ Đồ thị biểu diễn tỷ trọng khách hàng theo nhóm tuổi



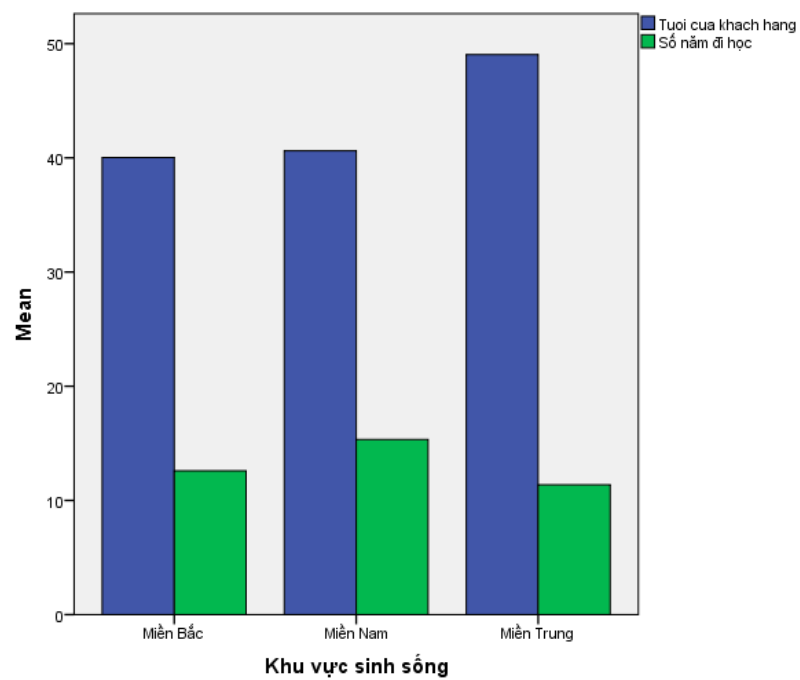
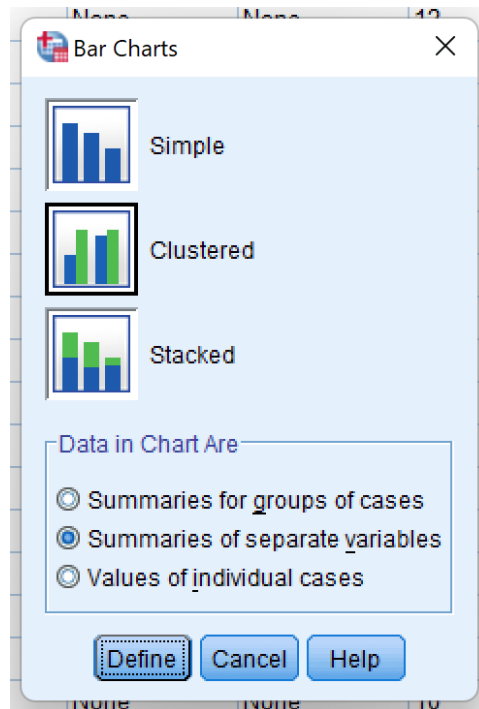
▼ Đồ thị biểu diễn tỷ trọng khách hàng theo nhóm tuổi



▼ Vẽ đồ thị cơ cấu khách hàng theo khu vực của từng giới



▼ Vẽ đồ thị độ tuổi, số năm đi học trung bình theo khu vực



▼ Tháp dân số theo giới tính

